Social Ties and Residential Choice: Micro Evidence and Equilibrium Implications*

Martin Koenen†
Drew Johnston‡

November 7, 2024
Link to latest version

Abstract

Why don't more people move to places where they can earn higher incomes? We use individual-level data from Facebook to find that social ties play a crucial role in explaining this puzzle: social ties are concentrated locally and shape migration decisions. On average, individuals live within 100 miles of nearly 80% of their friends, with less-educated individuals having even more concentrated social networks. To establish a causal link between the location of one's friends and migration, we exploit plausibly exogenous variation in the timing of friends' moves around individuals' college graduation. Having one more friend in a given commuting zone at the time of graduation increases one's likelihood of living there by 0.3 percentage points, which is comparable in magnitude to the effect of a \$470 increase in annual wages. We incorporate these findings into a spatial equilibrium model and show that the magnitude of social network effects can explain why people stay in poorer places and why less-educated people are much less responsive to economic shocks. Overall, this study shows that social networks play a first-order role — as important or more important than canonical economic factors such as wages and rents — in determining residential choice at the individual and aggregate level.

*We thank our advisors Raj Chetty, Ed Glaeser, Nathan Hendren, and Jesse Shapiro for their continued support and mentorship. We also want to thank Sahil Chinoy, Benjamin Goldman, Jamie Gracie, Robert Fluegge, Matt Jacob, Larry Katz, Gabriel Kreindler, and Stefanie Stantcheva as well as participants at the Harvard Economics labor and public finance seminar for their helpful feedback, comments and suggestions. The authors thank the Chae Family Foundation for supporting this project through a grant.

†Job Market Paper. Department of Economics, Harvard University, martin_koenen@fas.harvard.edu

‡Meta Platforms, Inc: djohnston@meta.com

1 Introduction

Why don't more people move to places where they could earn higher incomes? Incomes vary substantially across regions, even within countries: average wages in Boston are 50% higher than those in Cleveland.[1] Despite such regional differences, over 70% of people who grow up in a declining area stay in their place of birth as adults. Those who do leave often move to places with similarly poor economic conditions (Bartik, 2009; Sprung-Keyser et al., 2022). The disconnect between economic opportunity and migration is especially pronounced for less-educated individuals. To explain these observations, standard spatial equilibrium models typically resort to extremely high moving costs, search frictions, or strong preferences to live in one's place of birth.

In this paper, we show that the weak link between migration and incomes, especially among the less-educated, is driven in part by the effects of individuals' friends and family, or social networks. In the first part of the paper, we show that social networks are highly spatially concentrated and have causal effects on migration decisions. In the second part, we use a general equilibrium model calibrated using the parameters from part I to demonstrate that network effects can explain the low elasticity of migration with respect to wages, as well as differences in migration patterns between more- and less-educated individuals.

We use individual-level data on Facebook users in the United States where we observe an individual's location every month from 2012 to 2023, as well as detailed information on demographics, socioeconomic status, and social connections. Crucially, unlike much of the existing literature, this data allows us to measure individuals' networks directly. As a result, our data provides us with a more accurate and complete picture of one's social ties.

We begin by highlighting that social networks are spatially concentrated and correlate with migration decisions. On average, 78% of an individual's friends live within 100 miles of their home commuting zone (CZ). But this average masks substantial heterogeneity: the networks of individuals without a college degree are over 30% more spatially concentrated than those of people with a college degree. Consistent with evidence presented in existing research, we find that this type of heterogeneity in the dispersion of individuals' social networks is strongly correlated with their migration decisions. People with two-thirds or more of their friends

[1] A long literature examines regional differences in economic performance, showing that part of these differences reflect causal effects of places and that they cannot entirely be explained by differences in prices (e.g. Topel, 1986; Blackaby and Manning, 1990; Barro and Sala-i Martin, 1991; Blanchard and Katz, 1992; Chetty et al., 2018; Deutscher and Mazumder, 2020; Bilal, 2023; Alesina et al., 2021; Asher et al., 2024; Card et al., 2023; Sprung-Keyser and Porter, 2023; Chetty and Hendren, 2018; Handbury and Weinstein, 2014; Handbury, 2021; Diamond and Moretti, 2021). The average wage numbers discussed above can be found in QCEW (2024).

in their home CZ as of 2012 remain there more than 80% of the time. On the other hand, individuals with one-third or less of their friends in their home CZ remain there in less than 50% of cases. However, these observations do not necessarily imply a causal relationship, as the size of an individual's social network in a given location may be correlated with unobserved factors that influence migration decisions.

To isolate the effect of social networks on location decisions, we exploit differences in the timing of friends' moves around the time when individuals graduate from college. Consider two college graduates — *Alice* and *Bob* — who share similar backgrounds and have similar initial distributions of friends across CZs. Around the time of their graduation, each has one additional friend moving to Austin. But *Alice*'s friend moves to Austin slightly before *Alice* and *Bob* graduate while *Bob*'s friend moves just after *Alice*'s and *Bob*'s graduation. As a result, when *Alice* and *Bob* decide where to live at the point of graduation, *Alice* has one more friend in Austin than *Bob*. Our empirical approach leverages this difference in timing to see if *Alice* is more likely to move to Austin because her friend already lives there. More formally, we compare the predictive power of friends' moves to a given CZ before (like *Alice*'s friend) and after (like *Bob*'s friend) graduation for a college graduate's decision where to live.

We show analytically that friends' moves before graduation — in the "pre-period" — provide an estimate combining the causal effect of the social network and a bias. This bias arises from a potential correlation between having a friend move to a given CZ and unobservable factors that shape location decisions. Friends' moves after an individual's graduation — in the "post-period — only capture a bias because they cannot affect location choices made at graduation. Our key identifying assumption, which we term "symmetric bias", is that the biases are identical for the friends' moves in the pre- and post period. Importantly, this assumption is about the correlation between friends' moves and unobservable factors, rather than unobservable factors directly: we do *not* assume that unobservable factors do not exist or that they do not vary over time; instead we merely assume that the bias arising from these factors is symmetric for pre- and post-period friends' moves.

Friends' moves to a given CZ in the pre-period predict graduates' migration decisions more strongly than the same moves in the post-period. Those (like *Alice*) whose friends move to a given CZ before their graduation are more likely to live in that CZ right after graduation compared with those (like *Bob*) whose friends move to that same CZ slightly later. Under our symmetric bias assumption, the difference in the pre- and post-period coefficient estimates identify the causal effect of the network. Applying this approach to our estimates, we find that having an additional friend in a given CZ increases one's probability of choosing to live in that CZ by around 0.3 percentage points on average. This magnitude is comparable to an estimate of the effect of a $470 increase in annual wages.[2]

Three pieces of evidence support our identifying assumption and address potential concerns that unobservable factors might vary in a way that would violate the symmetric bias assumption. First, we split the sample by cohort: if a major shock in one year—such as a recession—drove the drop in the predictive power of friends' moves, we would expect to see the drop only for one cohort.[3] However, we find a similar drop around graduation for every cohort.

Second, we examine whether friends' moves to "sister-CZs" — regions with similar industry compositions as the focal CZ — affect an individual's decision to live in the focal CZ. By construction, sister CZs (for instance, Cleveland and Detroit) should be similarly affected by local economic shocks as focal CZs.[4] If the observed effects were driven by similar responses to the same shock among college graduates and their friends, then we would expect friends' moves to a sister CZ to be predictive of whether graduates live in the focal CZ. They are not.

Finally, we ask whether individuals' location choices are predicted by moves among people who they eventually befriend, but who they had not yet befriended by graduation. These future friends cannot have a causal effect on one's behavior, but are likely affected by location-specific shocks in a similar way as pre-graduation friends. If unobserved shocks were driving the prior results, we would expect to see similar effects for these friends. Reassuringly, we find no evidence that future friends impact individuals' location decisions at graduation.

So far, our causal analysis has focused on college graduates, but our primary interest lies in understanding migration decisions for the broader population. We next explore how the effects of networks on location choices generalize, again drawing on differences in the timing of friends' moves. Returning to the stylized example from above, in this design, we measure *Alice*'s and *Bob*'s location at two points in time. At the first point in time, the two live in the same place. In between the first and second point in time, *Alice* has one friend moving to Austin; *Bob* has a friend moving to Austin only after the second point in time. In order for the difference between individuals with friends moving earlier (like *Alice*) vs. later (like *Bob*) to identify the causal effect of the network, we need to make a stronger

[2]For this comparison, we use a shift-share approach to instrument for recent local wage changes on the college graduates' decisions of where to live.

[3]In this case, for exactly one cohort, the recession would begin before graduation, while for others, it would begin after, so that the drop in the predictive power should vary in timing.

[4]This intuition underlies the widespread use of shift-share instruments.

assumption, namely that the extent to which unobservable factors vary over time is limited.[5] Under this assumption, we find that the effect of having an additional friend in a CZ on an individual's likelihood of living there is comparable in magnitude to the effect identified for college graduates.

We next examine the mechanisms linking networks to residential choice. To test if information drives the observed network effects, we study whether friends moving *away from* a given CZ affect location choices, arguing that even after their departure friends can still provide information. We find that a friend's departure has a large, negative effect, suggesting that information is not the primary driver of network effects and that networks must provide other benefits that are linked to the contemporaneous presence of one's friends. Consistent with this empirical evidence, we find in a survey that companionship may play an important role: more than 85% of respondents indicated that they primarily enjoyed the company of their social ties rather than needing their help.

In the second part of the paper, we incorporate our reduced form results into a general equilibrium framework to study whether social forces can shed light on some of the observed aggregate patterns of migration. To do this, we generalize a Rosen-Roback style model of spatial equilibrium (Rosen, 1979; Roback, 1982) featuring local production, housing and workers who choose where to live.[6] The key innovation is incorporating a preference for living near one's social network calibrated from our quasi-experimental estimates. Since the size of one's social network in a given CZ is itself an endogenous outcome that directly depends on the utility maximization problem of others, we solve for the model's equilibrium using an iterative, stepwise approach.[7] We compare the predictions of our "network model" to those obtained from an otherwise identical model which does not include network effects.

The two models differ dramatically in their reliance on unexplained residuals to rationalize why people live in the places they do. As is common in Rosen-Roback style models, we rationalize the variation in CZ-level populations that cannot be explained by differences in wages and rents by the presence of unobserved local amenities. The variance of these amenities, or residuals, needed to fit the data falls by more than three quarters when we incorporate networks as a third observable factor that explains residential choice. For example, in the basic model, a relatively large positive residual is required to explain the high population of

[5] We provide evidence supporting this stronger assumption using the placebo tests (sister CZs and future friends) similar to those we employed in the college graduation sample.

[6] For a recent review of quantitative spatial equilibrium models and their empirical estimation, see Redding and Rossi-Hansberg (2017).

[7] We are able to estimate this model in the micro-data we work with, which allows us to exploit individual-level heterogeneity in social networks within a location to compute more realistic counterfactuals.

Detroit; in our framework, these patterns can be explained by the fact that people in Detroit have large social networks there.

Moreover, the inclusion of social networks renders the remaining residuals less opaque, causing them to align more closely with "real amenities". When we correlate the CZ-specific residuals with various measures of place characteristics known to be highly valued — air quality (Chay and Greenstone, 2005), weather (Rappaport, 2007) and retail environment (Glaeser et al., 2001) — we find modest correlations of at most 0.2 between these characteristics and the residuals from the basic model. On the other hand, the correlations between the residuals from the network model and the place characteristic range between 0.4 and 0.6, consistent with the notion that those forces likely do influence an individual's decision of where to live.

Our network model fits the data better for three reasons. First, networks are spatially concentrated. Second, our quasi-experimental evidence highlights that network effects are large, so that every additional friend in a given CZ has a reasonable impact on one's decision where to live. These two forces explain why people are reluctant to move without requiring high moving costs or other frictions. Third, we also find a negative correlation between local wage growth and the share of one's friends living locally. This observation can explain the relatively low out-migration rates for regions characterized by economic hardship without having to assign these areas particularly high amenity terms. Our findings highlight that social networks are critical in explaining why migration rates do not follow economic opportunity more closely.

Finally, we consider whether social networks can explain differences in migration patterns between more- and less-educated individuals. Using a shift-share instrument, we estimate that the observed wage elasticity following the Great Recession was substantially higher for those with a college degree than for those without. We also find that those without a college degree have on average 75% of their friends in their current CZ, compared to 55% among the college-educated. By simulating productivity shocks in our network model, we show that this difference in the concentration of social ties can explain most of the observed gap in migration elasticities between more- and less-educated individuals.

This paper contributes to several strands of literature. First, our work relates to the literature on social networks and migration (see Munshi (2020) for a recent review). Previous studies of social factors in migration have typically relied on proxies for networks, such as common origin locations or workplaces (e.g. Munshi, 2003; McKenzie and Rapoport, 2007; Costa et al., 2018; Mahajan and Yang, 2020; Stuart and Taylor, 2021; Porcher, 2020; Munshi, 2020; Egger et al., 2022; Hansch et al., 2024; Green, 2024). While these community-based approaches have broadened our understanding of the factors that shape migration decisions, they are limited by their reliance on incomplete and coarse measures of social ties. Our individual-level data on social networks provides us with a more accurate and complete picture of networks and enables us capture the vast differences between individuals from the same origin observed in prior work (e.g. McPherson et al., 2001; Chetty et al., 2022a).[8]

Second, our paper relates to the literature studying migration frictions to explain the long-standing observation that migration does not respond very strongly to local economic conditions (e.g. Sjaastad, 1962; Carrington et al., 1996; Chen and Rosenthal, 2008; Autor et al., 2013; Yagan, 2019; Wilson, 2020; Sprung-Keyser et al., 2022). A few noteworthy frictions that have received attention in recent empirical work include moving costs (e.g. Kennan and Walker, 2011; Diamond et al., 2019), search frictions (e.g. Bayer et al., 2016; Bergman et al., 2020), as well as preferences to live in one's place of birth (e.g. Diamond, 2016; Piyapromdee, 2020). We highlight that the size of one's local social network helps explain the limited responsiveness of migration to economic forces, and can also be regarded a "friction".

Portraying network effects as a friction indicates that our findings carry implications for the macroeconomic literature on the effects of local economic shocks, as migration is a key lever through which the economy adjusts to shocks (e.g. Blanchard and Katz, 1992; Hornbeck and Moretti, 2024). This friction can lead to a more drawn-out adjustment process than would be the case if social networks were more dispersed and people were consequently more mobile. The nature of the network-friction also matters for designing policies to respond to local shocks. While traditionally economists have been skeptical of place-based policies (e.g. Glaeser and Gottlieb, 2008; Kline and Moretti, 2014), the welfare costs of leaving a place where one has a large local social network may provide a stronger rationale for investing in declining places (Glaeser and Redlick, 2009). At the same time, the fact that networks are not inherently tied to places may motivate subsidizing coordinated moves of individuals and their friends to "better" areas to allow people to maintain their social ties while harnessing the benefits of more productive regions.

Lastly, our work connects to the literature studying variation in the migration behavior by socioeconomic and demographic characteristics. Prior research has documented large differences in migration behavior by education, race, or immigration status (e.g. Topel, 1986;

[8]A handful of other studies of migration patterns have employed direct measures of social networks, most notably Blumenstock et al. (2023) and Sahai and Bailey (2023). We expand on this prior work in several ways, in particular by employing a novel identification strategy that allows us to weaken several identifying assumptions, and by showing the relevance of these factors in a developed country, where informational and credit-based frictions are less binding.

Bound and Holzer, 2000; Borjas, 2001; Gregg et al., 2004; Wozniak, 2010; Malamud and Wozniak, 2012; Machin et al., 2012; Cadena and Kovak, 2016; Notowidigdo, 2020; Gaubert and Diamond, 2022; Sprung-Keyser et al., 2022). Existing explanations of differences by education often center disparities in preferences, as well as differences in the returns to and costs of migration (e.g. Moretti, 2011; Lkhagvasuren, 2014; Davis and Dingel, 2019; Gaubert and Diamond, 2022; Caldwell and Danieli, 2024; Amior, 2024). Our results highlight that the large discrepancies in the local concentration of social networks between more- and less-educated individuals can account for much of the observed differences in the migration patterns between these groups.[9]

Overall, this study shows that social networks play a first-order role — as important or more important than canonical economic factors such as wages and rents — in determining residential choice at the individual and aggregate level. By taking this force into account, we can explain why many people remain in economically distressed areas and why especially less-educated individuals are reluctant to move away.

The remainder of this paper is structured as follows. Section 2 describes the data in more detail. Section 3 shows descriptive evidence on the role of social networks for residential choice. In Section 4 we describe in detail the empirical framework to study the causal effect of social networks on residential choice, the results of which we present in Section 5. Lastly, Section 6 presents a model of spatial equilibrium which incorporates the role of social networks and discusses how this addition helps to explain observed migration patterns. Section 7 concludes.

2 Data Description

2.1 Facebook Data

Prior studies of the relationship between social networks and migration have been limited by data availability, as few datasets combine longitudinal data on individuals' home location with information about their social network. In this paper, we make extensive use of a de-identified, individual-level data set from Facebook that combines both of these attributes, allowing us to relax several assumptions common in prior research and to examine the impact of social networks on the migration patterns on specific groups of people. As of December 2019, Facebook had 272 million monthly active users and 205 million daily active users

[9]This finding is not necessarily inconsistent with prior research and can provide the theoretical underpinning behind some of the other explanations such as differences in moving costs or a stronger attachment to one's place of birth.

in the U.S. and Canada (Facebook, 2023). For each user, we observe basic demographic information such as their age, the high school and college they attended, and their predicted socioeconomic status.[10,11] Importantly, we also observe their set of on-platform friends and changes in this set of connections over time.[12] Prior research has shown that these friendships are usually between individuals who know each other in person, and capture a more intuitive notion of connection than the unidirectional connections found on other platforms (Jones et al., 2013). For each user in our sample, we observe their predicted CZ of residence for each day between January 2012 and December 2023.[13] Using this dataset, we define two main samples: the Expanded Sample, which we employ in Sections 3 and 5.5 and the College Graduation Sample (which we explore for the majority of our empirical analyses throughout Section 5). We describe each in turn below.

2.1.1 Expanded Sample

We first construct a broad sample of U.S. Facebook users, which we refer to as our Expanded Sample. We require that these users have at least 50 friends on the platform, were active in the last 30 days, and were born between 1985 and 1997. Coverage is particularly high among those cohorts (Chetty et al., 2022a) meaning that our sample is quite representative of individuals in that age group. We present summary statistics about this sample in Panel (a) of Table 1. We present additional descriptive statistics regarding the social networks and migration patterns of these individuals in Section 3.

In Appendix A1 we show that the rates of migration between CZs measured in our sample are very similar to rates reported in Sprung-Keyser et al. (2022), which are derived from linked data from the Department of Housing and Urban Development, the Department of the Census, and the Internal Revenue Service. Our rates of inter-CZ migration have a weighted correlation of 0.958 with those derived from administrative data; the rates of overall migration from a CZ have a weighted correlation of 0.908 with the analogous measures.

2.1.2 College Graduation Sample

We construct our College Graduation Sample as a subset of the Expanded Sample. In addition to the restrictions imposed in the Expanded Sample, we further limit our analysis

[10]This variable is imputed using characteristics such as a user's phone price. For a description of the procedure used to construct this variable, see Appendix B.1 of Chetty et al. (2022c).

[11]Duggan et al. (2016) found that, among U.S. adults, usage rates were relatively constant across income groups, education levels, and race; usage rates were slightly declining in age.

[12]These connections are undirected and an individual can have at most 5000 connections at a given time.

[13]This prediction combines several sources of information, including an individual's self-reported place of residence, their IP address and their on-platform activity. We aggregate these daily predictions by month.

to users who report on their profile that they attended a four-year college in the United States[14] and were born between 1994 and 1997. For reasons discussed in Section 4, in most of the analyses we present in Section 5, we focus on this sample. Note also that we subset each individual's network of friends, considering only friendships that were at least one year before the individual's graduation to other individuals who are predicted to be at least one academic year older than the individual. We infer an individual's expected year of graduation based on their date of birth information.[15] We present summary statistics about these individuals, their migration patterns, and their social networks in Panel (b) of Table 1.

2.2 Labor Market Data

At various points in Sections 3 and 5, we benchmark the effects of social networks on residential choice against the effects on migration probabilities that result from local wage and employment shocks. We also study the interaction between the role played by social networks and these labor market forces. To do this, we use data from the Quarterly Census of Employment and Wages, which publishes quarterly estimates of the level of pay and employment in each county in the United States.[16] We aggregate the county-level unemployment rates to the CZ level, applying weights proportional to the size of each county's labor force.

2.3 Survey Data

In Section 5.6, we analyze a survey that we conducted externally and which we use to gather information about individuals' self-reported motivations behind their location decisions. We fielded this survey online through Qualtrics between April 2nd and April 18th of 2024. Our sample is comprised of 517 American respondents aged 19-45 that we recruited through Prolific. Many of our analyses focus on the 298 individuals who report that they still live within 100 miles of the city they grew up in.

[14] We match these self-reported colleges to administrative data from IPEDS, dropping online colleges and those that do not appear in the Carnegie Classification. Our approach matches that described in more detail in the data appendix of Chetty et al. (2022b).

[15] Our procedure assumes that individuals follow the "standard" timeline prescribed for the date of birth. Violations of this assumption that lead us to mis-measure an individual's time of graduation which would attenuate our main estimates discussed in Section 5.

[16] This data is available for download here.

3 Descriptive Facts

This section motivates our main analyses by documenting several facts about the spatial concentration of social networks as well as the relationships between migration decisions, social networks, and economic opportunity. To do this, we explore the networks and moving decisions of the users in our Expanded Sample.

We begin by showing that social networks are highly spatially concentrated. In Panel A of Figure 1, we highlight that, at the beginning of our sample[17], the average individual lives in close proximity to most of their friends. We find that at the start of our panel in January 2012, nearly two thirds of the average individual's friends live in the same CZ that they do, and around 78% of friends living within 100 miles (similar to the distance between Los Angeles and San Diego). On average, individuals live within 500 miles of over 90% of their friends.[18,19]

Despite these broad patterns, there exists substantial heterogeneity in the dispersion of social networks. For instance, 5% of individuals have fewer than 12% of their friends within 100 miles of them, while another 5% have more than 94% of their friends within that distance. One's educational attainment is strongly correlated with the geographic dispersion of their friends: Panel B of Figure 1 shows that individuals without a college degree have on average 87% of their friends within 100 miles of their location, while individuals with a college degree have only 65% of their friends within that distance.[20] Even more strikingly, individuals attending an elite college (one with an average SAT score over 1300[21]) have, on average, only around 40% of their friends within a 100 mile radius.[22]

Motivated by the large differences in the dispersion of individuals' social networks, we next ask whether this kind of variation can help to explain migration patterns. To that end, we compare an individual's location in January 2012 and December 2019.[23,24]

We find that an individual's likelihood of staying in their home CZ is strongly correlated

[17] As in Section 2, we focus here on individuals and their friends in January 2012.

[18] For reference, the distance between San Diego and San Francisco is around 500 miles.

[19] These patterns are consistent with prior evidence on the spatial concentration of social networks Bailey et al. (2018).

[20] To proxy for whether individuals have a college degree we use self-reported information on whether an individual lists a college on their profile.

[21] We use 2013 average SAT scores of students at enrolled at a given university. For reference, Harvard has an average SAT score of 1505 and Florida International University has an average SAT score of 1080.

[22] Chetty et al. (2022b) presents similar gaps between individuals with higher and lower socioeconomic status.

[23] For the purpose of this exercise, we focus on locations prior to COVID-19 given that the pandemic induced dramatic changes in mobility behavior (Bailey et al., 2024).

[24] A more detailed description of the construction of this sample is available in 2.

with the size of their local social network, a finding that we illustrate in Panel A of Figure 2. While constructing this figure, we include only friendship links that were already formed on the platform by January 2012 and consider friends' locations at that point in time. We find a strong relationship between these two variables—while those with two-thirds of their friends in the home CZ stay there around 80% of the time, less than 50% of those with only one third of their friends living in the CZ remain there in 2019.

We also find that individuals' social networks are predictive not only of their likelihood of staying in their home CZ, but also their likelihood of moving to a given alternative CZ. We present this relationship graphically in Panel B of Figure 2, where a striking, linear relationship emerges between an individual's probability of moving to a given CZ — among one of the top 50 CZs by population[25] — and the share of their social network residing in that CZ[26]. Concretely, an individual has a probability of 0.01% of moving to a CZ where they have no friends, but a probability of 0.5% of moving to a CZ where they have 1% of their friends.

To benchmark the size of the relationship between local network size and migration, we also study how local economic growth shapes migration patterns. Following prior work (e.g. Beaudry et al., 2014; Wozniak, 2010; Dao et al., 2017), we instrument for changes in local wages between 2012 to 2019 using an industry-based shift-share approach.[27] Appendix Figure A3 highlights that migration probabilities respond to wage growth: if wages increase by \$1,000 in one's home CZ, one's probability of staying there increases by 1.3p.p. Correspondingly, if wages increase by \$1,000 in an alternative CZ, one becomes 0.1p.p. more likely to move there.[28] These magnitudes are similar to those found in prior research using tax data to estimate the effect of changing economic conditions on migration (Sprung-Keyser et al., 2022).

Comparing the magnitudes in Figure 2 to the effects of local wage growth suggests that social networks are substantially more predictive of an individual's location decisions than standard economic forces. To illustrate this, consider that an individual whose home CZ is at the 95th percentile in terms of wage growth is around 6 percentage points more likely

[25]Panel A of Appendix Figure A2 presents a map displaying the most populous CZs in the country by user count.

[26]In this graphic, we exclude an individual's home CZs.

[27]In order to construct predicted CZ-level wage growth between 2012-2019, we employ data from the Quarterly Census of Employment and Wages and use 6-digit NAICS industry codes to predict local wage growth by combining 2012 local industry shares with 2012-2019 leave-out industry-specific wage growth rates.

[28]In the regressions for whether an individual stays in the home CZ, we control for 2012 CZ-level wages. For the regressions for alternative destinations we control for predicted local wage growth in one's home CZ.

to stay there than an individual from a home CZ at the 5th percentile of the wage growth distribution. In contrast, someone whose relative network size in the home CZ is at the 95th percentile is 50 percentage points more likely to stay than someone whose local network size is at the 5th percentile, a far larger shift. Appendix Figure A4 highlights that focusing on local growth in employment—as opposed to wages—would lead to a similar conclusion.

We also find evidence of an interaction between social networks and shifting labor market conditions in shaping migration probabilities. In Figure 3, we explore how the relationship between economic shocks and migration varies at an individual-level as a function of a person's connectedness to a given commuting zone. In Panel A, we present separate estimates of the relationship between residential choice and local wage growth in one's home CZ by decile of connectedness to the home CZ, finding a striking pattern: among individuals with 30% or less of their social network in their home CZ, if local wages increase by an additional $1,000, their probability of staying increases by around 1.6p.p. In contrast, for those with 75% or more of their friends in their home CZ, their likelihood of remaining in their home CZ only increases by around 0.8p.p. in response to the same change in wages. In other words, those with fewer friends in their home CZ are far more responsive to changes in local wages than those with more friends there.

In contrast, Panel B of Figure 3 shows that those with larger networks in a given CZ other than their own are more responsive to changes in pay there. For instance, while a $1,000 increase in local wages in a given CZ where one has no friends increases one's odds of moving there by only 0.05p.p., the same change in local wages is associated with a 0.3p.p. shift in migration probability if it occurs in a CZ where one has 2% or more of their social network. Appendix Figure A5 presents similar patterns, focusing on responses to local employment growth.

How can networks make people less responsive to changes in the economy of their home CZ, while at the same time making them more responsive to changes elsewhere? In Appendix Section B, we describe how a standard spatial choice model following a logit structure can provide the theoretical underpinning for these results and match the observed patterns. For the purpose of the present section, we focus on the intuition: a larger share of an individual's friends tend to live in their home CZ than any other, and the probability that one stays in their home CZ is correspondingly higher their probability of moving elsewhere. As a result, those with particularly large networks in the home CZ often stay there "no matter what" and are less responsive to changes in the local economy. On the other hand, individuals are extremely unlikely to move to places where they have few friends, regardless of economic conditions there. Places where individuals have a non-negligible social network, however, are more realistic options, making individuals more sensitive to changes in economic conditions there.

The patterns shown in Figures 2 and 3 make two important points. First, they suggest that individuals' social ties are strongly predictive of their residential choices, with more predictive power than common measures of local economic growth. Second, the figures indicate that the dispersion of an individual's social network is predictive of the extent to which they can harness the economic opportunities offered by different areas: those with few connections outside their home CZs are much less likely to leave, even in the face of negative local shocks. Correspondingly, those with stronger connections to booming CZs are substantially more likely to move to such areas and hence exploit the opportunities there.

In the next section, we investigate the extent to which these cross-sectional relationships are driven by a true causal effect of one's social network on their location decisions.

4 Estimation Strategy

In this section, we use quasi-random variation in the timing of one's friends' moves to estimate the impact of an individual's social network on their decision of where to live.

We begin by describing a hypothetical, "ideal" experiment that could be used to measure such an effect empirically. In this ideal experiment, the experimenter would randomly allocate an individual's existing friends across geographies: for a set of people indexed by i the econometrician would move n of their friends to CZ j. The econometrician could then obtain the causal effect of the location of one's friends on residential choices with the following regression:

$$Y_{ij,t} = \beta n_{ij,t} + \varepsilon_{ij,t} \tag{1}$$

where $Y_{ij,t}$ corresponds to an indicator for whether i lives in j at time t. In this case, β captures the causal effect of i's social network in j on i's likelihood of living there.

In reality, social networks are not randomly assigned across geographies, so estimating equation 1 in cross-sectional data would yield a biased estimate of β. This is because there are likely other, unobserved factors that affect both the size of an individual's network in j and their likelihood of living there. Building on equation 1 above, we can consider a case in which, in the true model of the world, there are other factors $\theta_{ij,t}$, which may be unobservable and correlated with both $Y_{ij,t}$ and $n_{ij,t}$:

$$Y_{ij,t} = \beta n_{ij,t} + \theta_{ij,t} + \varepsilon_{ij,t} \tag{2}$$

where $Cov(\theta_{ij,t}, n_{ij,t}) \neq 0$. In this case, the naïve regression outlined in equation 1 would return a biased estimate of β.

How should we think about unobservable factors included in $\theta_{ij,t}$? Broadly, we can differentiate between permanent factors — such as latent preferences or latent skills — as well as time-varying factors.

We provide examples of these two types of unobservables in turn. To understand the threat to identification presented by permanent factors, imagine a world in which $\beta = 0$—that is, a world where the location of one's social network does not drive migration decisions. Further, imagine an individual i, who works as a software developer in the tech industry. Since that industry is concentrated geographically in places like Austin, i lives in Austin. Due to homophily—that is, the tendency to befriend individuals who share similar characteristics—i has many friends who are also software developers and who for the same reason live in Austin, i.e., $n_{iAustin,t} >> 0$. In this case, i's decision to live in Austin has nothing to do with their large social network there, but is simply driven by their unobserved occupation.

To understand the threat to identification posed by time-varying unobservables, we can modify the above scenario in the following way: while i and their friends did not live in Austin initially—meaning $Y_{iAustin,t-1} = 0$ and $n_{iAustin,t-1} = 0$—, with the recent opening of Tesla's new factory there, i as well as some of i's friends are re-locating there to find better jobs. As a result, $Y_{iAustin,t} = 1$ and $n_{iAustin,t} > 0$. In this case, i and their friends respond in the same way to the same "common shock", rather than i re-acting to their friends' moves. In this way, even if we employed an empirical design, such as an event-study or difference-in difference design, that exploits changes in a user's and their friends' locations over time, we may obtain a biased estimate of β.

Regardless of whether these unobservables are permanent or time-varying, a cross-sectional regression of equation 2 that does not include $\theta_{ij,t}$ would yield a biased coefficient estimate of β, or the causal effect of the network. More formally, we would obtain:

$$b = \beta + \delta \tag{3}$$

where δ is a bias term given by $\frac{Cov(\theta_{ij,t}, n_{ij,t})}{Var(n_{ij,t})}$.

To address concerns over unobservable factors, we focus our analysis on a critical point in time at which young adults choose where to live: the time at which they graduate from college. While we describe below how doing so allows us to address concerns over unobservable factors, in Appendix Figure A6, we present evidence that for college-goers the time of graduation is of crucial importance for residential choices. Panel A presents a time-series of the proportion of college graduates living outside their home CZs—i.e., the CZs in which one went to high school or college—highlighting that the presumed point at which an individual graduates from college is indeed a point in time at which many college graduates decide to leave their home CZ and move to new parts of the country: at the end of May following graduates' senior years, the proportion of individuals living in a CZ other than one's home CZ jumps and continues to increase until October, at which point close to 15% of graduates live in a CZ other than the ones in which they attended college and high school. In Panel B of Appendix Figure A6, we highlight that the year in which college-goers are expected to graduate is the year in which they are most likely to move across CZs among all the years that we observe. Importantly, these location decisions often persist over time, as we demonstrate in Panel C. We demonstrate in the Appendix that the broad patterns of migration among college graduates in our sample align with those of the broader set of users discussed in Section 3.[29]

Having established that the time of college graduation marks a critical point in time for people's decisions where to live, we next discuss how by focusing on this point in time we can isolate the causal effect of social networks under relatively weak assumptions. We begin with a stylized example to build intuition for our empirical design drawing on the graphical illustration in Figure 4.

To motivate our design, imagine two college graduates, *Alice* and *Bob*, who each grew up in Philadelphia and attended college in Boston, graduating at the same time. Between the beginning of their senior year and the summer or early fall after graduation *Alice* and *Bob* are likely to decide where to work and live following graduation.[30] During their first year

[29]Panel A of Appendix Figure A7 shows that the location of an individual's social network measured in the fall of one's senior year is strongly predictive of one's decision to stay in the CZs in which they went to high school or college, paralleling the results of Panel A of Figure 2, which shows similar patterns in the more general sample. Similarly, Panel B of Appendix Figure A7 highlights that one's migration to a given alternative CZ following graduation is strongly predicted by the initial distribution of one's friends, consistent with the results of Panel B of Figure 2, which considers the more general sample. Appendix Figure A8 shows that though local wage and employment growth positively predict an individual's location choice after graduation, these economic forces are less predictive of individual-level migration than the location of the individual's social network. Lastly, Appendix Figure A9 shows that a graduate's responsiveness to local wage and employment growth in a location varies substantially with the size of one's local network there, a finding we replicate in the more general sample in Figure C2. These patterns suggests that migration decisions among graduates are shaped by similar forces as those driving migration in the population at large.

[30]These months are a common time for graduates to find jobs, with many students beginning to apply

of college, *Alice* and *Bob* had the same number of friends in Austin.[31] In addition, the two each have one additional friend who moves to Austin around the time that they graduate. The key difference between *Alice* and *Bob* is the timing of the friends' moves. While *Alice*'s friend moves to Austin before *Alice* and *Bob* decide where to live — in the "pre-period" — *Bob*'s friend moves only after the two have already decided where to live, far enough in the future that *Bob* cannot anticipate his friend's move — in the "post-period". For instance, *Alice*'s friend may have moved in the fall of their last year in college, whereas *Bob*'s friend moved during the winter after graduation. Our empirical design leverages this difference in timing, asking whether *Alice* is more likely to choose to live in Austin after graduation than *Bob* since, when they each decided where to live, *Alice* had one more friend in Austin than *Bob*.

How does a comparison between *Alice* and *Bob* help to resolve the concerns about unobservable factors? First, since we only compare people who have both the same number of friends in Austin during both their first year in college and at the end of our sample, we avoid comparing people with generally different levels of social networks in a given place, who may have these friends there for permanent unobserved reasons. To elaborate, if *Alice*'s friend's move to Austin is simply a proxy for the fact that *Alice* is a software developer who is already likely to move to Austin, then *Bob*'s friend's move to Austin should have the same predictive power about *Bob*'s unobserved occupation.

Time-varying unobservables present us with a more nuanced concern. In principle, it is possible for Tesla to be hiring at the time when *Alice*'s friend moves as well as when *Alice* and *Bob* decide where to live, but to have stopped hiring by the time that *Bob*'s friend moves. In this case, *Bob*'s friend's move must be due to a different reason, such as a research institute opening at UT Austin which started hiring after *Alice* and *Bob* had already decided where to live. In this case, it would be possible for *Alice* and her friend to respond to the same, common shock, whereas *Bob* and his friend were not able to respond to the same common shock. Thus, rather than being driven by a difference in the social networks of *Alice* and *Bob*, any difference in the location decisions of *Alice* and *Bob* could be due to the fact that *Alice* and her friend were able to respond to the same, common shock while *Bob* and his friend were not.

To address concerns of this sort, we need to make an additional assumption. To isolate the effect of social networks, we assume that *Alice*'s and *Bob*'s friends' moves are equally

well before graduation (LinkedIn, 2023).

[31]Throughout this analysis, we only consider friends that the two made at least one year prior to graduation to mitigate concerns over endogenous friendship formation.

reflective of the types of shocks *Alice* and *Bob* could have responded to when they decided where to live. In other words, if *Alice* and her friend are to a certain degree able to respond to similar conditions (such as Tesla's new factory), then *Bob* and his friend need to be able to respond to similar conditions to the *same degree* as well (such as the new research institute) even if the conditions can have changed between when *Alice*'s friend moved and when *Bob*'s friend moved. Importantly, this assumption is about symmetry and does not require that no shocks exist or that shocks must be the same in the pre- and post-period: Tesla does not need to be hiring in the post-period, and UT Austin does not need to be hiring in the pre-period. Instead, our assumption implies that if Tesla and UT Austin present the only two shocks in this economy then it must be the case that if Tesla hires in the pre-period and when *Alice* and *Bob* decide where to live, then UT Austin must be hiring not only in the post-period but also when the two graduates decide where to live. In other words friends' moves before and after graduation must be equally predictive of the local conditions faced by *Alice* and *Bob* when deciding where to live.

We next formalize the intuition behind this stylized example. We begin by extending equation 1 to include friends' moves, which we measure in terms of changes in the size of one's social network in a CZ between periods. Importantly, throughout this analysis we hold the stock of one's friends constant, focusing on friendships formed at least one year before individual i graduates from college.[32,33] Note that this restriction allows us to sidestep concerns over endogenous friendship formation: it is not possible that people decide to live in j and as a result befriend people living or moving there. In our main analyses, we additionally exclude all friends who are not at least one cohort older than the graduates to ensure that the post-period moves we observe are not driven by subsequent graduates, which would be driven by different factors.[34] Our main estimation equation is therefore given by:

$$Y_{ij,t^*} = \sum_{t=early+1}^{T} \beta_t \Delta n_{ij,t} + X_{ij,t^*} + \varepsilon_{iod,t^*} \tag{4}$$

Here, t^* denotes the October following graduation, the time at which we measure graduates' locations. We choose this point in time in light of the fact that Panel A of Appendix Figure

[32]This ensures that changes in the number of friends in a given CZ are driven by friends' moves rather than by the formation of new friendships.

[33]In Section 5.6 we present evidence from an estimation that is based separately on moves to and away from a given CZ. We find results that are very consistent with the more parsimonious definition employed here.

[34]We do not make this restriction in Section 5.5 where we estimate effects for a more generalized sample and obtain qualitatively and quantitatively similar results.

A6 indicates that by October the year of graduation post-graduation location decisions have largely stabilized.[35] Each time period t corresponds to a quarter, beginning with *early*, which refers to the summer of an individual's first year in college. The difference $\Delta n_{ij,t} = n_{ij,t} - n_{ij,t-1}$ captures changes in the size of one's local social network between adjacent quarters. We present results based on measuring $n_{ij,t}$ in terms of both numbers and proportions of friends, though we focus on the number-based definition given its more intuitive interpretation.[36]

To mimic the same type of narrow comparisons described in the illustrative example, equation 4 also includes a rich set of control variables, $X_{ij,t}$. At baseline, we include childhood-CZ by college-CZ by potential destination j by cohort fixed effects as well as a CZ-specific linear control for the number of friends in j at the beginning of the sample period.

In the above setup, β_t captures the extent to which changes in the size of one's social network in a location at time t predict location decisions in the October following graduation, t^*. In the same way as estimating equation 2 in cross-sectional data would result in biased estimates of β corresponding to $b = \beta + \delta$, estimating equation 4 in cross-sectional data would yield a set of coefficient estimates b_t which would also combine a causal effect (β_t) and a bias component (δ_t), so that $b_t = \beta_t + \delta_t$.

Before showing how we can use our data to sidestep these concerns, it is helpful to introduce some notation. In particular, we distinguish between friends' moves in the pre-period (which we term t^{pre}) before we measure users' location at t^*, and those in the post-period (which we term t^{post}) following t^*. We exclude a narrow period (which we call $\bar{t}$) surrounding t^* from these sets, given that college graduates start deciding where to live after graduation at least a few months before graduation and because graduates likely anticipate friends' moves that happen right after graduation.[37]

Under the assumption that friends' moves in the post-period have no impact on location decisions upon graduation, $\beta_{t^{post}} = 0$, so $b_{t^{post}}$ reduces to $\delta_{t^{post}}$, with later friends' moves identifying a selection term.[38] Under the additional assumption that the biases are constant

[35] In results not shown, we find virtually identical results if we measure location in the adjacent months.

[36] In the structural model presented in Section 6 we use the proportion based measure because of potentially larger differences between the size of people's social networks.

[37] In our preferred specifications, $\bar{t}$ includes the two quarters before t^*, t^* itself, and the one period following it. Our main results presented in Figure 5 highlight that our results are robust to adjusting this window in either direction as the drop in the coefficient estimates happens within a very narrow band around t^*.

[38] In principle, it is possible for there to be a reverse causality issue, if later friends' moves are a response to the graduates' migration decisions. We believe that this is unlikely to be of concern because it affects both $b_{t^{pre}}$ and $b_{t^{post}}$. Since our parameter of interest is the difference between the two, this is unlikely to affect our results. Moreover, if one was concerned that this issue is particularly present for friends' moves in the

over a relatively short period of time, $\delta_{t^{post}} = \delta_{t^{pre}} = \delta$ we can use the $b_{t^{post}}$ coefficient estimate to quantify the constant bias term and subtract it out to from $b_{t^{pre}}$ to recover the causal effect of the network $\beta_{t^{pre}}$:

$$\Delta b_{t^{pre}} = b_{t^{pre}} - b_{t^{post}} = \beta_{t^{pre}} + \delta_{t^{pre}} - \beta_{t^{post}} - \delta_{t^{post}} = \beta_{t^{pre}} + \delta - 0 - \delta = \beta_{t^{pre}} \tag{5}$$

Therefore our main identifying assumption—which we term the "symmetric bias" assumption—is as follows:

Identifying Assumption: conditional on having a friend move to j around $\bar{t}$, the exact timing of such move is orthogonal to potential outcomes.[39]

Importantly, this assumption allows for local conditions to change over time, and for there to be time-varying shocks affecting the local conditions of a given CZ. In other words, our analysis is robust to changes in $\theta_{ij,t}$ between t^{pre} and t^{post}. However, the above assumption rules out that the bias arising from changes in $\theta_{ij,t}$ differs between the pre- and the post-period, requiring that the relationship is symmetrical during the two periods:

$$\delta_{t^{pre}} = \frac{Cov(\theta_{ij,t^{pre}}, \Delta n_{ij,t^{pre}})}{Var(\Delta n_{ij,t^{pre}})} = \frac{Cov(\theta_{ij,t^{post}}, \Delta n_{ij,t^{post}})}{Var(\Delta n_{ij,t^{post}})} = \delta_{t^{post}} \tag{6}$$

As this discussion highlights, our identification strategy can address concerns about most cases of unobservable confounders. Most importantly, the approach is immune to concerns that individuals who have friends living in, or moving to, a given CZ are systematically different from those without such friends in a way that is correlated with their location decisions. Our identifying assumption is hence relatively weak, and we provide evidence for its plausibility in Section 5.2.

While we began this section with the objective of wanting to estimate β, or the causal effect of one's social network on location choice, we have now identified conditions under which we can estimate versions of $\beta_{t^{pre}}$ for various different $t \in t^{pre}$. These $\beta_{t^{pre}}$ describe the effect of friends' moves to a given CZ at various points in time. β can thus loosely be understood

post-period, note that because we are subtracting $b_{t^{post}}$, this will at most create a downward bias. Lastly, the coefficient estimates presented in Section 5.1 highlight that this concern is unlikely to be warranted. A related concern is that friends' moves in the post-period could themselves affect individuals' location decisions by t^*. Equation 5 shows that this would tend to bias our estimates of the effect of networks towards 0.

[39] Given that our measure of friends' moves is based on the difference in the number of friends in a given CZ, in principle it also captures moves away from a given CZ. The analysis in Section 5.6 highlights that this does not give rise to concerns as the effects of friends moving to and away from a given CZ are opposite and symmetric.

as an average over the $\beta_{t^{pre}}$ for which we can obtain an estimate from our empirical design. If the effects of social networks do not differ much with tenure — that is, the amount of time a friend has already spent in a given CZ — each $\beta_{t^{pre}}$ is a good approximation to the average effect β. On the other hand, if tenure matters more, this will be important to take into account when constructing an estimate of the average effect based on $\beta_{t^{pre}}$. We return to this discussion in Section 5.3.

5 Results

We next build on the identification strategy outlined in the preceding section and present reduced form evidence on the causal effect of social networks on residential choice.

In this section, we primarily focus on the three cohorts of college graduates who graduated between 2017 and 2019. For each user, we observe several years of friends' moves happening before and after $\bar{t}$. For computational reasons, we restrict the choice set of potential destination CZs d to the 50 most populous CZs in the country. We highlight these CZs in Panel A of Appendix Figure A2. In Panel B of the same figure, we show that close to 60% of users live in one of those CZs. Finally, at baseline, we cluster standard errors at the individual-level in light of the fact that the treatment, the friends' moves, happens at the level of the individual. We explore the robustness of our results to these restrictions later on.

5.1 Time-Specific Network Effects: β_t

We begin by presenting regression results corresponding to our main estimation equation in Figure 5. The Figure is based on equation 4, so that each dot captures one b_t, corresponding to the extent to which friends moving to j at a time t predict the graduate's probability of living in j after graduation.[40] The gray bar in the middle of the figure highlights $\bar{t}$, or the time period in which college graduates likely decide where to live after graduation or during which they can presumably anticipate future friends' moves happening shortly after graduation. 0 on the horizontal axis corresponds to t^*, i.e., October the year of graduation, which is the time period at which we measure graduates' location. The red coefficient estimates in the figure correspond to the coefficients on friends' moves in the pre-period (before $\bar{t}$), while blue coefficient estimates correspond to friends' moves in the post-period (after $\bar{t}$). Friends' moves during $\bar{t}$ — referred to as the "mid-period"— are shown in gray.

[40] Recall from equation 3 that, since we are unable to observe $\theta_{ij,t}$, we are unable to obtain estimates for β_t, and instead estimate $b_t = \beta_t + \delta_t$, where δ_t represents a bias term.

We begin by discussing the coefficient estimates close to the center of the plot, focusing in particular on those immediately adjacent to $\bar{t}$. The last red coefficient (b_{-3}) captures friends' moves during the fall of the individuals' senior years, while the first blue coefficient (b_{+2}) captures friends moving in the winter following the individuals' graduations. These periods mirror closely the illustrative example discussed in detail in Section 4: we can think of b_{-3} as capturing the extent to which *Alice*'s friend's move predicts *Alice*'s decision of where to live, while b_{+2} captures how predictive *Bob*'s friend's move was for *Bob*'s decision of where to live. The figure highlights that even within this relatively short period of time, the exact timing of a friend's move matters dramatically. While a friend moving to a given CZ in the fall of one's senior year (b_{-3}) is associated with an increase in the odds the individual chooses to live in that same CZ of around 0.32p.p., a friend moving in the winter after graduation (b_{+2}) corresponds to an increase in the same outcome of only 0.06 p.p.[41]

Following the discussion in Section 4, b_{-3} and b_{+2} both include bias terms: δ_{-3} and δ_{+2}, respectively. Under our identifying assumption of symmetric biases, the two biases are the same, so that the difference between the two estimates b_{-3} and b_{+2} provides us with an estimate of the causal effect of having a friend move to a given CZ three quarters before t^*. In the present case, this implies that $\beta_{-3} = b_{-3} - b_{+2} = 0.26$p.p, so that having an additional friend in a CZ at the time an individual decides where to live after graduation increases their odds of choosing that CZ by 0.26p.p.

Zooming out from the two coefficient estimates b_{-3} and b_{+2}, Figure 5 reveals a striking pattern: friends' moves at any point in time before $\bar{t}$ are substantially more predictive of a graduate's location choice than friends' moves happening at any point in time after $\bar{t}$. The coefficient estimates in the pre-period are large, with magnitudes between 0.3p.p. and 0.45p.p., while friends' moves in the post-period are far less predictive of a graduate's location choice, and generally have magnitudes close to 0. The extent to which friends' moves are predictive of location choices declines steeply during $\bar{t}$. Under the symmetric bias assumption, the elevated estimates in the pre-period and the much lower estimates in the post-period suggest that having one more friend move to a given CZ at any point in time before $\bar{t}$ has a robust causal effect on a college graduate's location choice — that is to say, $\beta_t > 0$ for all t in the pre-period. This result is not sensitive to the definition of $\bar{t}$, since the series of coefficient estimates suggests that even if we shifted $\bar{t}$ by a few periods, we would reach a similar conclusion.[42]

[41]Note the fact that the later friends' moves are predictive of contemporaneous location decisions is consistent with the observation in Section 4 that unobservable factors—either permanent or time-varying—may confound simpler empirical approaches even after controlling for a rich set of covariates.

[42]Estimates for the post-period also provide evidence against the potential concerns over reverse causality

5.2 Evidence Supporting Identifying Assumption

The conclusion that the patterns observed in Figure 5 represent network effects and are not driven by time-varying unobservable factors relies on the plausibility of our identifying assumption of symmetric biases in the pre- and post-period. In what follows, we provide evidence supporting this assumption based on three exercises.

Cohort-Specific Estimates

One potential threat to the symmetric bias assumption could be a major shock (such as a recession) that coincides in timing with $\bar{t}$, generating a high correlation between local conditions in the pre-period and those at t^*, but lower correlations between local conditions in the post-period and those at t^*. This concern would be especially warranted if we only had a single cohort of college graduates. However, since our results are pooled over multiple graduating cohorts, this concern is unlikely to hold: while a shock, such as a recession at time $\bar{t}$, implies a break in the correlation around $\bar{t}$ for one cohort, the same example would not imply the same break for other cohorts, as $\bar{t}$ corresponds to different years in those cohorts.[43]

While this line of reasoning makes it unlikely that major shocks drive our results, in order to provide more formal evidence against it, we repeat the estimation of equation 4 separately for each graduating cohort in our sample. We present the corresponding results in Panel A of Figure 6, focusing in particular on the estimates corresponding to b_{-3} and b_{+2} for every cohort. If a major shock around $\bar{t}$ in one particular year generated the pattern in the aggregate series shown in Figure 5, we would expect to see a drop in the coefficient estimates between b_{-3} and b_{+2} for only one cohort. For the two other cohorts, we would not expect to see a decline, and may even see an increase depending on the timing of such shock. On the other hand, if the effects discussed above are truly driven by an effect of social networks on location decisions, we should see a drop for each graduating cohort. Despite the estimates being somewhat more noisily estimated, they display a similar picture for every cohort.[44]

discussed in footnote 38. As explained there, reverse causality might in principle result in an upward bias of the coefficient estimates corresponding to post-period friends' moves, meaning that by taking the difference between pre- and post-period coefficient estimates, we would find a downward-biased estimate of β_t. Not only is this concern unlikely to be warranted given our discussion in footnote 38, but the estimates in Figure 5 also highlight that this is not a serious concern: the coefficients in the post-period are all close to zero, so that even if one was concerned about the coefficients in the post-period suffering from a an upward bias, this would only result in a very small downward bias of the estimates of β_t.

[43] In fact, depending on the exact timing of the recession, it may even imply a break in the opposite direction for other cohorts, such that the correlation would be stronger in the post-period than in the pre-period.

[44] The drop is somewhat larger for the class of 2019, coming from a lower coefficient estimate for b_{+2}. This observation can likely be explained by the fact b_{+2} in this case corresponds to the winter of 2020 when the

This pattern indicates that the effects here and in the aggregated analysis are indeed driven by network effects rather than any major shocks that coincide in timing with the college graduation event.

Sister CZs

In a second exercise, we present evidence from a placebo test in which we examine whether friends' moves to CZs with industry compositions similar to the focal CZ affect an individual's decision to live in the focal CZ. Specifically, for each CZ in our sample, we use a nearest neighbor matching procedure based on industry-specific employment shares to identify "sister CZs", or the CZ most similar to the focal CZ in their industry composition.[45] One example of a pair of focal and sister CZ from this procedure is Detroit and Cleveland. In light of the large literature exploiting industry-based shift-share instruments building on Bartik (1991), we expect these pairs of CZs to be affected similarly by economic shocks. Consistent with that conjecture, we find that measures of recent local economic growth are highly correlated between focal and sister CZs[46], suggesting that sister CZs are indeed likely subject to similar economic shocks as the focal CZs.

Exploiting the similarity between focal and sister CZs, we now repeat the estimation of equation 4 while adding friends' moves to sister CZs as regressors.[47,48] If the effects in Figure 5 were driven by similar responses to the same shock among college graduates and their friends, we would expect to see a similar pattern for friends' moves to the sister CZs. On the other hand, if the previous effects are truly driven by network effects, we would expect to see no effect, or a flat series of coefficient estimates since a friend moving to a sister CZ

COVID-19 pandemic had begun. During this time period moving patterns changed dramatically (Bailey et al., 2024), making them less predictive of the location decisions around graduation in 2019. To some extent this therefore confirms the intuition that in principle major shocks could influence our results for a single year. However, since the coefficient estimates for the 2017 and 2018 graduating cohorts exhibit drops in the coefficient estimates as well we conclude that the overall pattern in Figure 5, which pools over multiple different cohorts is not driven by a major shock.

[45]For the purpose of this exercise, we define industries using 6-digit NAICS codes.

[46]The correlation between local wage and employment growth (2012-2018) is 0.55 and 0.53, respectively and the slopes are 0.77 and 0.63.

[47]The estimation for this exercise is given by:

$$Y_{ij,t^*} = \sum_{t=early+1}^{T} \beta_t^{FOCAL} \Delta n_{ij,t}^{FOCAL} + \sum_{t=early+1}^{T} \beta_t^{SISTER} \Delta n_{ij,t}^{SISTER} + X_{ij,t^*} + \varepsilon_{ij,t^*} \tag{7}$$

[48]We continue to include friends' moves to the focal CZ in the regression in light of Danieli et al. (2023). In results not shown, we find that we obtain very similar and conceptually consistent results when dropping friends' moves to the focal CZ from the specification. In that case, we obtain small negative effects of friends' moves to the sister CZ on the graduates' decision to live in the focal CZ. This is intuitive: network effects suggest that having one more friend in a sister CZ should make you more likely to to live in the sister CZ and, as a result of that, less likely to live in the focal CZ.

should not make one more likely to go to a focal CZ even if the two CZs are very similar. Reassuringly, Panel B of Figure 6 displays a flat series of coefficient estimates with levels close to zero, supporting the interpretation that the main effects shown in Figure 5 are driven by network effects.

Post-Graduation Friends

Lastly, we draw on another placebo test in which we exploit moves among people whom they eventually befriend, but whom they have not yet befriended by graduation. To elaborate, while the results presented so far have focused on friends made at least one year prior to graduation, we now study the effects of moves among friends made at least four years after graduation (who we refer to as"post-graduation" friend) on college graduates' decisions of where to live right after graduation. By definition, such friends cannot exert a direct influence on one's decision where to live right after graduation as the individual and the friend did not know each other at the time of graduation.[49] Put differently, for the post-graduation friends we expect $\beta_{t^{pre}} = 0$. If the effects shown in Figure 5 are truly driven by network effects, we would thus not expect to see any effect for these post-graduation friends. If the coefficients in Figure 5 were instead driven by similar responses to local shocks by an individual and their friend, we would continue to see similar effects for the post-graduation friends, if they also proxy for the type of shocks the graduates might respond to.[50] Consistent with the conjecture that pre- and post-graduation friends are similar, we find that the spatial distribution of these friends is very similar to the pre-graduation friends studied so far.[51]

We present the results corresponding to this placebo test in Panel C of Figure 6.[52,53] For the

[49]In practice, it is possible that an individual and a post-graduation already did know each other at the time of graduation, but that they were not friends on the Facebook platform yet. However, even if that was the case, that would only make it more likely to find an effect of post-graduation friends on an individual's decision where to live right after graduation, meaning that this concern would only lead us to reject the placebo test.

[50]Intuitively speaking, if *Alice* is a software engineer, her friends both pre- and post-graduation are likely to be software engineers as well. Consequently, if the pattern in Figure 5 was driven by *Alice* and her friend responding in the same way to a common shock, then we would expect to see a similar effect for the post-graduation friends.

[51]The correlation of the proportion of pre- and post-graduation friends in a given CZ is 0.73 at the time of the study period among individuals included in the analysis of Panel C of 6.

[52]The estimation for this exercise is given by:

$$Y_{ij,t^*} = \sum_{t=early+1}^{T} \beta_t^{PRE-GRAD} \Delta n_{ij,t}^{PRE-GRAD} + \sum_{t=early+1}^{T} \beta_t^{POST-GRAD} \Delta n_{ij,t}^{POST-GRAD} + X_{ij,t^*} + \varepsilon_{ij,t^*} \quad (8)$$

[53]As in the case of sister CZs, we continue to include friends' moves to the focal CZ in the regression in light of Danieli et al. (2023).

purpose of this exercise we use a proportion-based definition of social networks, rather than a count-based definition, since individuals in our sample tend to have fewer post-graduation (3.9 on average) than pre-graduation friends, making it difficult to otherwise compare the coefficient estimates.[54,55] Interestingly, we find that the coefficient estimates are positive, meaning that the post-graduation friends' moves are highly predictive of an individual's decision where to live right after graduation, well before they befriend these post-graduation friends. This is intuitive: it is far more likely for a given individual to befriend someone who moves to the same CZ as they themselves decide to live in at the end of college. More importantly, the coefficient estimates depict a flat pattern, indicating that there is no effect of these post-graduation friends on the graduates' location choices, once more reinforcing the interpretation that the results in Figure 5 are indeed driven by network effects and not confounded by other unobservable factors.

5.3 Average Effect of Social Networks: β

While the time-specific estimates β_t help us understand how friends' moves at various different points in time impact an individual's decision where to live, we now aim to provide a more parsimonious way to express the effect of social networks on residential choice. The fact that the coefficients in Figure 5 in the pre- and post-period are relatively stable suggests that the β_t are relatively similar for all t in the pre-period. As we laid out in Section 4, if effect sizes do not differ substantially with the timing of the friends' moves, the tenure of one's friends over this range is not an important source of heterogeneity in the effects of social networks on residential choice. As a result, the average effect derived from flows of friends at various points in time is likely a good approximation to the estimand of interest, β.

In Table 2 we present coefficient estimates from a specification that aggregates friends' moves in the pre-, mid- and post-period, respectively, from the regression equation:[56]

[54] We provide a more detailed discussion regarding the proportion-based measure of social networks in Section 5.3. The results discussed there highlight that the baseline results in Figure 5 are robust to using a proportion-based measure of social networks.

[55] Due to the COVID-19 pandemic and the well known differences in migration patterns it sparked (Bailey et al., 2024), we focus on the 2017 graduating cohort for the construction of Panel C of Figure 6 to avoid issues in which the post-period coincides with the beginning of the pandemic. As indicated by Panel A of the same figure, our main results are robust to focusing on the 2017 graduating cohort.

[56] In equation 9, *pre* corresponds to friends' moves at any point in time between 13 and 3 quarters prior to t^*, *mid* represents friends' moves between 2 quarters before and 1 quarter after t^* and post denotes friends' moves happening between 2 and 12 quarters after t^*. Thus, *pre* captures the same moves as the red coefficients in Figure 5, while *mid* and *post* capture the same moves as the gray and blue estimates, respectively. All other terms in equation 9 are defined as in equation 4.

$$Y_{ij,t^*} = \sum_{t \in [pre,mid,post]} \beta_t \Delta n_{ij,t} + X_{ij,t^*} + \varepsilon_{ij,t^*} \tag{9}$$

Consistent with the evidence presented in Figure 5, column 1 of Table 2 shows a highly significant and positive coefficient estimate. The magnitude of 0.0033 suggests that having one more friend in a given CZ increases one's likelihood of choosing to live there after graduating from college by 0.33p.p., in line with the magnitude we deduced from the graphical analysis in Figure 5.

We next explore the sensitivity of this estimate of the average effect of social networks on residential choice by modifying our main empirical specification in various ways. First, in columns 2-4 of Table 2, we further narrow the comparisons we make by including an even richer set of control variables.[57] Despite the finer comparisons, we find estimates that are quite similar to those presented in column 1, both in levels and in terms of statistical significance.[58] Second, we explore robustness to the way in which we parameterize the size of one's local social network: in Appendix Figure A10 and Appendix Table A1 we repeat the analyses from Figure 5 and Table 2, this time measuring friends' moves in proportions of one's total number of friends, rather than in counts.[59,60] Our results are qualitatively identical to the ones shown above for the number-based definition, and the magnitudes are consistent with those presented above.[61] Third, to address concerns that the magnitude of the network effects is influenced by restricting the choice set to the 50 largest CZs in the country, Appendix Table A4 shows that we obtain virtually identical estimates when expanding the choice set further to include smaller CZs as well.[62] Lastly, Appendix Tables A2 and A3 highlight that the significance of our results is unaffected by various alternative

[57]In columns 2 and 3 we interact our set of control variables from column 1 with gender and parent socioeconomic status decile fixed effects, respectively. In column 4, we present results from a specification where we interact the set of fixed effects with indicators for the exact college an individual attended.

[58]The fact that the estimates are somewhat smaller in column 4 when we control for an individual's exact college could be driven by potential spillover effects among college graduates attending the same college.

[59]For the purpose of this exercise, we now parameterize $n_{ij,t} - n_{ij,t-1}$ in equations 4 and 9 as the change in the proportion of i's friends between $t-1$ and t.

[60]We continue to hold the stock of friends constant focusing on friends made at least one year before graduation, so that endogenous friending in response to location decisions cannot affect the results. We also continue to restrict our attention to older friends.

[61]Column 1 of Appendix Table A1 displays a coefficient estimate of around 0.3, indicating that a 10p.p. increase in one's local social network in a given CZ raises the odds of living there by 3p.p. Table 1 shows the mean number of older friends is 112 and the median number is 81. Multiplying the effects estimated in Table A1 by these numbers produces magnitudes similar to those found using the proportion based measure.

[62]In Appendix Table A4, we expand the choice set to the largest 100, and 200 CZs in the country. Panel A of Appendix Figure A2 presents a CZ-level map indicating which CZs are included in each of these analyses. Panel B of the same figure highlights that the proportions of people living in the 50, 100, and 200 largest CZs are 59%, 73% and 87%, respectively.

ways of clustering standard errors.[63]

5.4 Dollar Valuation of a Friend

How does the impact of social networks on residential choice compare to that of other factors influencing individuals' decisions about where to live, such as local wages? To answer this question, we compare our estimates of the effect of one's social network to estimates of the role played by local wages. To do so, we begin by quantifying the effect of local wages for our sample of graduates.

To estimate the effect of local wages on a college graduate's location decision, we employ a shift-share approach to instrument for local wage growth in the lead up to one's graduation. Specifically, we instrument for local wage growth in the years before one's graduation using an industry-based shift-share approach similar to that discussed in Section 3.[64] For the purpose of this exercise, we focus explicitly on the "extensive margin", or the decision to stay in one's college CZ after graduation because Sprung-Keyser et al. (2022) highlight that the effect of local wages on residential choice varies substantially across destinations. Appendix Figure A11 presents the results: if annual wages grow by an additional $\$1,000$, one's likelihood of staying in the CZ of the college after graduation increases by 1 p.p.[65]

To compare the effects of wages and networks, we amend our analysis from Section 5.3 to focus on one's decision to stay in the CZ in which they attended college. Concretely, we estimate equation 9, restricting the set of CZs included in the analysis to the CZ of

[63]At baseline, we cluster standard errors at the individual level. For robustness, we alternatively cluster standard errors at the childhood-CZ-by-destination-CZ (column 2), the college-CZ-by-destination CZ-level (column 3), as well as the intersection of the two, i.e., the childhood-CZ-by-college-CZ-by-destination-CZ level (column 4).

[64]We define industries using six-digit NAICS codes and use a shift-share approach to instrument for CZ-level wage growth in the seven years leading up to one's graduation. For example, for those graduating in 2017, we instrument for CZ-level wage growth between 2010 and 2017. Equipped with a measure of local wage growth, we then use the same data as in Sections 5.1 - 5.3 while subsetting our sample the set of CZs included in our main analyses.

[65]We note that this estimate is somewhat higher than that found in Sprung-Keyser et al. (2022) who find that, if wages grow by around $800 in one's childhood that increases the odds of staying there through age 26 by 0.3p.p. This difference can likely be explained by the difference in the sample (we focus on college graduates rather than all young adults), the time period at which we study location decisions or the empirical strategy. More importantly, the fact that Sprung-Keyser et al. (2022) find a smaller estimate suggests that if we compared our estimate of the network effects to their estimates regarding the role played by local wages, we would find even higher numbers for the "dollar value" of a friend than we obtain with our estimate of the role of wages.

one's college.[66,67]. We obtain an estimate of 0.47, indicating that having one more friend living in one's college CZ increases an individual's odds of staying there after graduation by 0.47p.p.[68]

Comparing the effects of wages and networks, we find that networks play an important role in shaping an individual's decision of where to live: the ratio of the two estimates suggests that having one more friend living in one's college CZ has about the same effect on their propensity to stay after graduation as a $470 increase in average annual wages. This effect underscores that the locations of one's friends may be more important than relatively small differences in local pay across CZs. Since on average 59% of a typical college graduate's friends live in their college CZ at the time of graduation, these magnitudes imply that there is a high cost of leaving one's college CZ. We revisit this conjecture more formally in Section 6.

5.5 Expanded Sample

Our analysis has so far focused on college graduates and their location decisions upon graduation. We next explore how the observed effects generalize to a broader population and to other points in time. Specifically, we present results from an amended version of our estimation strategy, drawing on the Expanded Sample described in Sections 2 and 3. While this amended version of our empirical design requires a stronger identifying assumption, it enables us to estimate the network effects for a broader set of individuals and in a more general context.

As in the estimation strategy for college graduates, we again leverage differences in the timing of friends' moves to study network effects in the expanded sample. To mimic the setup discussed for the sample of college graduates, we compare individuals who initially lived in the same CZ and ask how their decision about where to live three-and-a-half years later is predicted by their friends' moves at various points in time before and after we measure their location again.[69] We return to equation 4 and re-define t^* as the second point in time at

[66] In the current exercise, our data is now uniquely identified by individual. While Y_{ij,t^*} still corresponds to an indicator for whether i lives in j at t^*, j now exclusively captures the CZ of one's college. Likewise, $\Delta n_{ij,t}$ captures friends' moves to the CZ of the college.

[67] In results not shown, we find that the effect of one additional friend on residential choices tends to be substantially larger if we focus on explaining residential choices beyond staying in the CZ of the college. This is driven by the fact that people are found to be relatively inelastic to local wages in those places (Sprung-Keyser et al., 2022).

[68] This effect is noticeably larger than the average effect found in Section 5.3; we present a more systematic analysis of how network effects vary by destination in Appendix Section C.1.1 to highlight that the larger effect observed for the college CZ is consistent with a standard spatial choice model following a logit structure.

[69] We choose a three-and-a-half year gap to mirror the analysis for college graduates, where we studied the

which we measure an individual's location—in this context, this is October of 2017, 2018 or 2019—and *early* corresponds to the first time at which we measure an individual's location. Our measure of social networks now includes all friends made as of early 2012, consistent with our definition for this sample in Section 3.[70] The other terms of equation 4 are unchanged. As before, we are interested in the difference of the coefficient estimates corresponding to friends moves prior to t^* and after t^*. Following the argument in Section 4, the estimates from the latter period primarily capture a bias term while the former combine both a bias term and the causal effect of the network. Under the symmetric bias assumption, we can thus once again identify the causal effect of the network by taking the difference between the coefficients on friends' moves in the pre- and post-period.

What is the main difference between the design for college graduates and the one for the Expanded Sample? In the case of the college graduates, there is a clearly defined period during which individuals decide where to live. Moreover, college graduates cannot move before they finish college. These feature created a well-defined pre- and post-period. In contrast, in the case of the Expanded Sample, people are in principle able to move at any point in time between t^{early} and t^*. As a result, the relative timing of friends' moves and an individual's potential move is less clear: prior to t^*, friends and individuals could move at the same point in time, for instance because of a common shock that both individuals respond to. In contrast, this is not the case after t^*, where by definition, friends' moves happen after the point in time at which we measure the individuals' locations.

To illustrate how the present setup complicates the identifying assumption in this context, we return to the stylized example of *Alice* and *Bob*. In the more general setup, we are unable to distinguish between the following two situations, (a) and (b). In situation (a) *Alice*'s friend moves before t^* and *Alice* follows her right around t^*. In situation (b) both *Alice* and her friend move at a similar point in time before t^*, perhaps because they both start working at Tesla simultaneously. Note that in the case of college graduates, (b) was not possible because *Alice* still was attending college when her friend moved. This example highlights that for the Expanded Sample, the relative timing of *Alice* and her friend's move is unclear. On the other hand, the relative timing for *Bob* and his friend is clear: since we measure *Alice*'s and *Bob*'s location at t^*, *Bob*'s friend's move — which happens after t^* — must happen after *Bob* decided where to be at t^*. Put differently, while *Alice* and her friend were in principle able to decide where to live based on the same local conditions, the same is not necessarily true for *Bob*. Consequently, in order for the symmetric-bias assumption

effects of friends' moves up to three-and-a-half years prior to our measurement of graduates' locations.

[70] In this section, we no longer restrict the networks we observe to friends older than the individual.

to hold here, we require that local conditions are relatively constant over time. Thus, while shocks may exist, they must be relatively slow-moving, so that the later friends' moves (*Bob*'s friend) can capture the same bias as the earlier friends' moves (*Alice*'s friend).

With this caveat in mind, Appendix Figure A12 presents the estimation results.[71] Reassuringly, the patterns shown for both the number-based definition of social networks (Panel A) and for the proportion-based definition (Panel B) are remarkably similar to the patterns observed for the college graduation sample: friends' moves prior to t^* yield positive, highly significant and stable coefficient estimates. Coefficient estimates then steeply decline for friends' moves right around t^* before stabilizing again soon after t^* at much lower levels than before. Under our identifying assumption, the observed pattern indicate substantial effects of social networks on residential choice for this more general sample.

To arrive at a more parsimonious estimate of the effects of social networks on location decisions for this more general sample, in Table 3 we present coefficient estimates similar to those shown in Table 2.[72,73] We find point estimates around 0.003, which are robust to changing the set of fixed effects in a similar fashion as for the sample of college graduates. These estimates hence suggest effects that are quite similar to those obtained for the sample of college graduates.

To provide evidence in support of the stronger identifying assumption underlying our approach, in Appendix Figure A13 we show results from similar placebo exercises as those discussed in Section 5.2. Panel A highlights the extent to which friends' moves to sister CZs exert an effect on an individual's likelihood to live in a given focal CZ.[74] Reassuringly, the series of coefficient estimates displays a flat pattern, suggesting no effect of friends' moves to sister CZs. Panel (b) shows effects of "late friends", or people who an individual eventually befriends but has not yet done so by t^*.[75] Again, we observe a flat series of coefficient estimates, meaning that such friends have no effect on an individual's decision where to live. Taken together, we interpret these results as suggesting that the network effects identified for college graduates generalize to a broader set of individuals and periods of life.

[71] For computational reasons, our estimation uses a random sample of two percent of individuals.

[72] Appendix Table A1 presents corresponding results using the proportion-based definition of social networks.

[73] To learn more about potential differences in the effect sizes across groups, in Appendix Section C.1.2 we present estimates by subgroup, drawing on a multinomial logit approach explained in Section 6.2 that allows us to more directly compare the magnitudes. We find that the effects are relatively similar in size for those with and without a college degree.

[74] We use the same approach to identify sister CZs as discussed for the sample of college graduates. The regressions are based on equation 7, amended for the Expanded Sample.

[75] We define late friends as individuals an individual befriends in 2022 or after.

5.6 Mechanisms

Having demonstrated that social networks shape residential choices in important ways, we next ask about the drivers behind these observed effects. In particular, we contrast the role played by information with the relevance of other benefits the network may provide such as an amenity value of friendships.

5.6.1 Information Effects

We begin by testing for the extent to which the networks effects are driven by information shared between individuals and their friends. Given the extensive prior literature highlighting the important role played by social networks in job referrals and in finding housing (e.g Rees, 1966; Ioannides and Datcher Loury, 2004; Bayer et al., 2008; Desmond, 2012), the observed effects could be driven by friends providing various types of information about a place. In turn, this information might cause people to choose to live in places where they have a larger network (e.g. Nelson, 1959; Porcher et al., 2024). To test for this mechanism, we differentiate explicitly between friends' moves *to* and *away from* a given CZ, reasoning that even if friends move away from a given CZ, they are still able to provide many kinds of information such as about amenities and opportunities there. Concretely, we amend equation 4 in the following way:

$$Y_{ij,t^*} = \sum_{t=early+1}^{T} \beta_t^{TO} n_{ij,t}^{TO} + \sum_{t=early+1}^{T} \beta_t^{AWAY} n_{ij,t}^{AWAY} + X_{ij,t^*} + \varepsilon_{ij,t^*} \tag{10}$$

Here, $n_{ij,t}^{TO}$ captures the number of friends moving to j between $t-1$ and t and $n_{ij,t}^{AWAY}$ represents the number of friends moving away from j between $t-1$ and t. All other terms are unchanged. In this setup, we generally expect the β_t^{AWAY} coefficients to be weakly negative since having one friend less in a given CZ likely lowers one's odds of living there. However, if information plays an important role for explaining the observed effects, we would expect the estimates of β_t^{AWAY} to be smaller in absolute magnitude than β_t^{TO}, if friends who have moved away from a place are still able to provide information about it. On the other hand, if the social network effects are primarily driven by the continuing presence of friends locally, we would expect the estimates of β_t^{AWAY} to roughly equal those of β_t^{TO} in absolute magnitude.[76]

[76] One caveat to the above reasoning is that the act of a friend moving away might itself new information about a given place; for instance, a friend might tell someone they are moving away because it is difficult to find jobs there. While we cannot directly address this concern, since it typically tends to take some amount of time for people to relocate, a friend will likely already provide this information to those in their network

Panel A of Figure 7 shows the corresponding results in our sample of college graduates.[77] As in Figure 5, the coefficient estimates to the left in shades of red represent friends' in the pre-period and coefficient estimates to the right in shades of blue depict friends' moves in the post-period. Friends' moves *to* a given CZ are shown in lighter shades and use circles, while friends' moves *away from* a given CZ are shown in darker shades and use diamonds.

The series of coefficient estimates corresponding to friends moving to and away from a given CZ depict very different patterns. The series for the friends moving to a given CZ looks virtually identical to the one shown before in Figure 5. In stark contrast, the series corresponding to friends' moves away exhibits relatively stable, large, and negative coefficient estimates in the pre-period, a steep increase around $\bar{t}$, and stable coefficient estimates that are close to zero in the post-period. Following the same approach as before, in which we take the difference between the estimates corresponding to pre- and post-period moves, we find that having a friend move away from a given CZ in the pre-period *lowers* an individual's probability of choosing to live there by 0.33p.p. Thus, the effect of friends moving away is essentially identical to the effect of friends moving to a given CZ in absolute magnitude. Under the assumption that friends moving away can still provide information about a given CZ even after their departure, this provides evidence against the hypothesis that the effects are predominantly informational. Instead, Figure 7 supports the interpretation that the effects are driven by benefits linked directly to the concurrent presence of one's friends in a given CZ.[78]

before they actually move. Thus, by examining the exact shape of the series of coefficient estimates we can assess this concern to some extent.

[77]We only conduct this exercise for the sample of college graduates since for the Expanded Sample used in Section 5.5, our empirical approach does not allow us to observe whether friends' moves happen before or after an individual's move. In the present case, this artifact could make the shape of the β_t^{AWAY} estimates difficult to interpret, but the shape is important for reasons described in footnote 76.

[78]One caveat to the above interpretation of Panel A of Figure 7 is that the present analysis pools over all top 50 CZs included in the choice set. The analyses in Section C.1.1 highlight that the average effect skews towards places in which college graduates already have more friends to begin with. In those places, one additional friend may only provide very limited additional information, so that one friend moving away does not account for a substantial difference in the amount of information provided by the social network. To assess the validity of this concern, in Appendix Figure A15, we present estimates of the effects of friends moving to and away from a given CZ separately by whether the potential CZ is one's home CZ, as well as by their initial connectedness to the CZ at the end of their first year in college. If networks provide important information about a place but each additional friend only provides a limited amount of new information, we would expect to see the effect of friends moving to and away to be roughly the same in absolute magnitude in places where college graduates have many friends. Conversely, in cases where college graduates only have few initial connections, we would expect smaller absolute effects for friends moving away from a given CZ than for friends moving to a CZ. Panel B suggests that there is no evidence in support of that conjecture; regardless of the restrictions we impose on the destinations included in the analysis, we find that the effect of having a friend move to a given CZ is similar in absolute magnitude to the effect of having a friend move away from a CZ.

5.6.2 Social Amenity vs. Other Benefits

While the above exercise helps us to narrow down the potential number of mechanisms behind the observed effects, there could still be various reasons for why the effects are linked to the concurrent presence of friends. For one, having a larger local social network could provide a direct social amenity if people enjoy spending time with those in their social circle and therefore choose to live where they have a larger social network. Alternatively, having a larger social network could also provide other important benefits such as child care, or help with tasks in the household.

Since the Facebook data does not allow us to directly test for the mechanisms, we conduct an off-platform survey to further explore the reasons behind the network effects.[79] Specifically, we present results from 298 respondents who indicate that they live in the city in which they grew up and who are between 19 and 45 years old[80] about the role of their social network for their decision where to live.[81]

Consistent with our findings from the Facebook data, survey respondents indicate that social networks play a crucial role in their decision to stay in their home city. Panel A of A14 highlights that over 80% of respondents agree that friends and family are one of the main reasons that they live in their current city. Moreover, when asked about whether and for what level of pay raise respondents would move to a city where they didn't know anyone in Panel B, around 50% of respondents say they would only do so for a pay raise of at least 50% their current income or more, if at all.[82] These magnitudes are consistent with the evidence shown in Section 5.4, suggesting a large valuation in dollar terms for having at least one friend in a given CZ.

Using the survey data to explore the underlying mechanisms, our findings strongly support the importance of the social amenity value of having friends and family nearby. When asked about the primary reason respondents see the people they deem most important in their

[79] The survey was administered on Prolific in April 2024.

[80] This age range parallels the set of individuals included in the analyses we have conducted in the Facebook data.

[81] We focus on individuals still living in their home city since we are particularly interested in understanding the workings of social network effects for these individuals given that people often continue to live in their home CZ despite the fact that they could access greater economic opportunities in other parts of the country. In order to be included in the survey sample it is not necessary that respondents have always lived in their home CZ. In results not shown, we find that individuals who report that they have moved away from their home town also indicate that social networks are important to them in their decision where to live: around 50% of such individuals say that friends and family have been rather or very important reasons for them to move.

[82] Note that we only presented the hypothetical of moving to a city without prior connections to the sample of stayers, so in Panel B, we subset the analysis to stayers.

decision where to live, leisure overwhelmingly stands out: Panel A of Figure 8 shows that over 70% report that they see those influential people always or most of the time for leisure. Consistently, Panel B shows that the practical help related to jobs or housing offered by these influential people is relatively limited.[83] Finally, nearly 85% of respondents agree that they value the company of these individuals far more than any practical assistance they provide (Panel C). Together, these results highlight the significant role of the social amenity value that social networks offer in explaining the observed effects.

6 Model

Section 5 highlights that social networks play a key role in shaping an individual's decision of where to live. In the following section, we explore whether this quasi-experimental — or partial equilibrium — evidence can help us understand aggregate — or general equilibrium — patterns of residential choice. Specifically, we draw on the reduced form approach to study whether the effects of social networks can shed light on why many people live in places with low levels of economic opportunity and why less-educated individuals are less responsive to local economic shocks than more-educated individuals. To do so, we generalize a Rosen-Roback style model of spatial equilibrium to incorporate social networks.

6.1 Model Setup

Our model comprises three parts: local production, a local housing market, and workers who decide where to live. Our model incorporates standard approaches to modeling production and housing markets from the discrete choice literature, but expands on the standard approach by incorporating a preference for colocation with ones friends into individuals' utility function.[84] The way in which we model and quantify this preference follows directly from

[83]Interestingly, the same results do indicate that help in relation to child care is more important: around 50% of those with children indicating that they have received a decent amount or a lot of help with this in recent years. This result may help to explain the somewhat larger network effects observed for older individuals in Section 5.5.

[84]Our approach is similar to that of Zabek (2024), who presents an model of how preferences to stay in one's home state influence migration decisions. In the context of that paper, these preferences are assumed to approximate the importance of one's local ties. Our approach uses data on social ties directly. In addition, we use the well-identified, quasi-experimental approach from Sections 4 and 5 to quantify the role of social networks rather than assuming a preference parameter in a more adhoc way. The two approaches to modeling local ties also have distinct implications. Since individuals are mobile, in our network-based explanation, one's attachment to a given CZ can vary based on the location of one's friends, which can change over time. This can also lead to different policy implications: for instance, if a policy helps move some of *Alice*'s friends to leave their declining home CZ, that might induce *Alice* to leave the place as well. An explanation based on fixed preferences across locations would not capture this dynamic.

our quasi-experimental approach in Sections 4 and 5.

6.1.1 Production

Each CZ j has a representative, perfectly competitive firm producing a local good variety Y_j. While all CZs share the same Cobb-Douglas production function for Y_j, local productivities (θ_j) can vary across space:

$$Y_j = \theta_j K_j^{\alpha^Y} N_j^{1-\alpha^Y} \tag{11}$$

Capital K_j is supplied at rate ρ, and labor N_j is paid in the form of wages w_j.

At the national level, a perfectly competitive firm produces consumption goods Y using the local good varieties as inputs in a CES production function:

$$Y = \left[\sum_j (Y_j)^{\frac{\eta^Y - 1}{\eta^Y}} \right]^{\frac{\eta^Y}{\eta^Y - 1}}$$

where η^Y corresponds to the Armington elasticity of substitution (Armington, 1969). The presence of national production relates the local goods Y_j to each other, as local goods are bought at price p_j. The Armington elasticity governs the strength of the relationship between the individual good varieties and is hence an important factor determining the extent to which productivity expansions in one CZ spill over to other places.[85]

6.1.2 Housing

We model housing supply using the following function:

$$H_j^S = \nu_j r_j^{\eta^H} \tag{12}$$

where r_j is the local rental rate and η^H captures the national housing supply elasticity and ν_j is a location-specific supply shifter. ν_j captures factors such as regional differences in local construction productivities. As a result, the local housing supply can differ even when holding other factors constant.

[85] Note that another reason for why productivity expansions can spillover is migration. However, the above argument highlights that even in the absence of migration there can be spillovers to other regions.

6.1.3 Worker's Problem

Lastly, we describe the worker's problem. Workers freely choose where to live, with a utility of individual i living in CZ j given by:

$$U_{ij} = \alpha_w \ln(w_j) - \alpha_h \ln(r_j) + A_j + \tilde{\beta} n_{ij} + \tilde{\delta} \bar{n}_{ij} + \epsilon_{ij} \tag{13}$$

where $\ln(w_j)$ captures the natural log of the average wage rate in j, $\ln(r_j)$ is the natural log of rent prices in j, and A_j corresponds to the level of local amenities, which we describe in more detail in Sections 6.2 and 6.3. Everyone living in j has equal access to these amenities and values them the same amount. ϵ_{ij} captures idiosyncratic preferences, which follow a type I extreme value distribution, giving the model a standard logit structure. In Appendix Sections B and C.1.1, we provide evidence in support of the logit structure being a reasonable assumption in this context.[86]

The main innovation of our approach relative to prior work is the addition of $\tilde{\beta} n_{ij}$ and $\tilde{\delta} \bar{n}_{ij}$. These factors capture, respectively, the role played by social networks and individual's place-specific preferences that are correlated with the size of their social network in a given place. We discuss each of these terms in turn.

We begin with n_{ij}, which captures the proportion of i's friends who live in CZ j. More formally, we can write n_{ij} as:

$$n_{ij} = \frac{\sum_{u \neq i}^{I} g_{iu} \times \psi_{uj}}{\sum_{u \neq i}^{I} g_{iu}} \tag{14}$$

where g_{iu} is an indicator for whether i and u are friends, and ψ_{uj} is the probability that u lives in j. Given the logit structure of our setup, ψ_{uj} is given by:

$$\psi_{uj} = \frac{\exp(V_{uj})}{\sum_{j'} \exp(V_{uj'})}$$

with V_{uj} being the indirect utility function corresponding to U_{ij} (McFadden, 1974).[87]

[86] We provide two pieces of evidence. First, in Appendix Section C.1.1 we demonstrate that the effects identified in Section 5 scale with the extent to which individuals have prior connections to a given CZ in a way that is consistent with the implications of a logit model. Second, in Appendix Section B we show that the extent to which the responsiveness to local wage and employment growth discussed in Section 3 varies with an individual's degree of connections to a given CZ is consistent with the implications of a logit model.

[87] The expected number of people in j, or the labor supply, is then given by:

Equation 14 suggests that the size of one's social network in CZ j is an endogenous object for two reasons. For one, individuals can decide whom they are friends with, meaning that g_{iu} is endogenous. Second, ψ_{iu} is itself an equilibrium outcome and follows from the maximization problem of one's friends. For these reasons, there exists a direct interdependence between individuals and their friends: i's utility maximization problem directly depends on their friends' actions and vice versa. To simplify the complexity added by these features, we treat g_{iu} as exogenous, meaning that we take as given who is connected to whom, which eliminates the first source of endogeneity in n_{ij}. Nonetheless, n_{ij} continues to be endogenous given that ψ_{iu} is endogenous.

Next, we turn to $\bar{n}_{ij}$, which follows from the discussion of unobservables in Section 4. $\bar{n}_{ij}$ captures an individual's unobserved place-specific preferences that are correlated with the size of their social network. Returning to the stylized example from Section 4, $\bar{n}_{ij}$ could for instance represent the fact that *Alice* is a software engineer working in the tech industry in Austin due to the industry's concentration there and that, because of homophily, many of *Alice*'s friends are software engineers working in Austin too. $\tilde{\delta}\bar{n}_{ij}$ hence corresponds to the "biases" discussed in Section 4. We can then leverage the fact that, as part of our estimation strategy, we quantified the relevance of these place-specific preferences.[88] Unlike n_{ij}, $\bar{n}_{ij}$ is not an endogenous object and is not treated that way: we take $\bar{n}_{ij}$ as given and hold it constant in the counterfactual exercises of Section 6.3.2.

Even if $\bar{n}_{ij}$ is taken to be exogenous, the fact that n_{ij} is endogenous implies that the distribution of networks is an equilibrium outcome: each worker's utility maximization problem must be consistent with their friends' utility maximization problem. This generates a very large number of additional equilibrium conditions. While we provide a more detailed discussion of the model's equilibrium conditions in Appendix Section D, due to the presence of the endogenous network in the utility function, we do not have closed form solutions for the equilibrium conditions. Therefore, we use an iterative approach to solve for the model's parameters, which we describe in more detail below.

$$N_j = \sum_{u \in I} \psi_{uj}$$

[88] While we consider the fact that we can identify the role played by individual-and-place specific preferences that are correlated with the size of one's network as a feature, in results not shown we explore robustness to omitting $\tilde{\delta}\bar{n}_{ij,t}$ from the worker's problem described in equation 13. This omission decreases our ability to match the network moments of the data discussed in Section 6.2, though we continue to obtain results that are qualitatively consistent with those presented in Section 6.3.

6.2 Calibration

To estimate the model, we use the micro-data of individuals in our Expanded Sample. Using micro-data to estimate the model marks a departure from much of the existing literature in this space, which typically draws on aggregated data. In our case, using the micro-data on the billions of social connections in our sample is essential as doing so allows us to more accurately capture the fact that there are vast differences between people's networks, which drive important differences in behavior and generate social spillovers. In our calibration, we continue to focus on the largest 50 CZs, matching our approach in Section 5.[89]

We parameterize the model such that it closely aligns with the data in terms of (a) local wages and rents, (b) local populations, (c) the observed distribution of social networks, and (d) the average wage elasticity observed for this sample of individuals. Exploiting the fact that the quasi-experimental approach discussed in Sections 4 and 5 enables us to use a well-identified micro-estimate for the role of social networks, in Section 6.2.1, we explain how we leverage the same quasi-experimental design to quantify $\tilde{\beta}$ and $\tilde{\delta}$. In Section 6.2.2, we describe how we find values for the vectors θ_j, ν_j, and A_j and the parameters α_w, and α_h. We rely on prior research for the parameterization of η^Y, η^H, ρ, and α^Y. We present an overview of the model parameters in Appendix Table A8.

6.2.1 Network Parameters

In Section 5, we identified an estimate of the role of social networks on an individual's location decision, which we referred to as β. By virtue of our empirical design, we also quantified the role played by other factors that are correlated with the size of one's local network, which we call δ. While closely related, the parameters $\tilde{\beta}$ and $\tilde{\delta}$ in the worker's utility function in equation 13 are somewhat distinct[90]: Rather than capturing the effect on one's decision where to live, $\tilde{\beta}$ and $\tilde{\delta}$ capture the effect of one's network — or its correlates — on the level of utility associated with living in a given CZ. Fortunately, given the assumptions imposed on ϵ_{ij}, the model follows a standard logit structure (McFadden, 1974). As a result, we can obtain an estimate of $\tilde{\beta}$ and $\tilde{\delta}$ from the same quasi-experimental design as described in Sections 4 and 5 if we estimate the two parameters using a multinomial logit regression model

[89] Given our focus on the 50 largest CZs, we restrict the Expanded Sample to individuals living in one of those 50 CZs both in 2012 and in 2019. This is necessary for the estimation of the wage elasticity described in Section 6.2.2. We did not impose this restriction in Section 3.

[90] We use tildes to clearly denote the difference in the definitions.

rather than OLS, as we did in Section 5.[91] We are therefore able to use the same approach to identification as before to obtain a well-identified estimate of the relevant parameters for this context.[92]

We next provide more details on the multinomial logit estimator of our quasi-experiment. For the purpose of this exercise, we return to the sample of college graduates which allowed us to identify the causal effect under weak assumptions; at the same time, we emphasize that we obtained similar results for the Expanded Sample in Section 5.5, suggesting that the effects found for the college graduates likely generalize. While the approach looks generally quite similar, we note a few differences as a result of the multinomial logit approach. First, we can only estimate effects for college graduates who choose to live in one of the top 50 CZs by population after graduation.[93,94] Second, the logit framework does not allow us to include the high-dimensional control vectors we included in Figure 5 and Table 2. We therefore approximate the regressions of equation 9 using the following logit equation:

$$\tilde{Y}_{ij,t^*} = \mu_j + \tilde{\beta}_{early} n_{ij,early} + \sum_{t \in [pre,mid,post]} \tilde{\beta}_t \Delta n_{ij,t} + dist_{kj} + dist_{kj}^2 + I(k=j) + \varepsilon_{ij,t^*} \quad (15)$$

Here, $\tilde{Y}_{ij,t^*}$ corresponds to the log odds ratio of living in j rather than a reference CZ $\bar{j}$. μ_d denotes an intercept specific to each potential destination j, $n_{ij,early}$ corresponds to the size of one's initial network in j, $dist_{kj}$ and $dist_{kj}^2$ correspond to the distance, and distance squared, respectively, between the home CZ k and the potential destination CZ j. $I(k=j)$ is an indicator for whether an individual's home CZ matches the potential destination CZ.[95]

[91] The logit structure implies that the utility function yields the following:

$$\log(\frac{\psi_{ij}}{\psi_{i\bar{j}}}) = \alpha_w[\ln(w_j) - \ln(w_{\bar{j}})] - \alpha_h[\ln(r_j) - \ln(r_{\bar{j}})] + \tilde{\beta}[\ln(n_{ij}) - \ln(n_{i\bar{j}})] + \tilde{\delta}[\ln(\bar{n}_{ij}) - \ln(\bar{n}_{i\bar{j}})]$$

where $\bar{j}$ denotes a reference CZ.

[92] As an alternatively, one could also find parameter values for $\tilde{\beta}$ and $\tilde{\delta}$ using a simulated methods of moments (SMM) approach. We refrain from doing so here because of two reasons. First, the scale and the infrastructure of the Facebook data highly constrain the number of iterations we are able to do to find parameter values that would allow us to match the data. Second, by using the approach outlined here, we are able to use a well-identified, micro-estimate of $\tilde{\beta}$ and $\tilde{\delta}$.

[93] For a more detailed discussion of the theoretical background for this restriction see Train (2003).

[94] In Panel B of Appendix Figure A2, we show that close to 60% percent of users live in one of the largest 50 CZs.

[95] Recall that the term home CZ combines a college graduate's college CZ as well as the CZ in which they grew up. In practice, we therefore include separate terms for both the distance—raw and squared—between an individual's college CZ and the potential destination CZ as well as the CZ they grew up in and the potential destination CZ. Similarly, we include two indicator terms for $I(k=j)$, one for whether the

Taken together, these four terms thus aim to present a parsimonious version of X_{ij,t^*} from equations 4 and 9. Appendix Tables A6 and A7 present evidence showing that our OLS results are largely robust to the changes implemented here.[96]

How can we use a multinomial logit estimate of equation 15 to recover the parameters of interest, $\tilde{\beta}$ and $\tilde{\delta}$? Following the discussion in Section 4, estimating equation 15 in cross-sectional data yields coefficient estimates $\tilde{b}_{tpre}$ and $\tilde{b}_{tpost}$, where $\tilde{b}_{tpre} = \tilde{\beta}_{tpre} + \tilde{\delta}_{tpre}$ and $\tilde{b}_{tpost} = \tilde{\delta}_{tpost}$. Under the symmetric-bias assumption outlined in 4, we have that $\tilde{\delta}_{tpre} = \tilde{\delta}_{tpost}$, and hence $\tilde{b}_{tpre} - \tilde{\delta}_{tpost} = \tilde{\beta}_{tpre}$. In words, we can quantify $\tilde{\delta}$ using the coefficient estimates from friends' moves in the post-period, after college graduates decided where to live. We can then obtain $\tilde{\beta}$ by taking the difference between the coefficient estimates corresponding to friends' moves in the pre- and the estimates corresponding to those in the post-period to recover $\tilde{\beta}$.

Table 4 presents the regression results corresponding to the multinomial logit estimation. The coefficient estimate for $\tilde{\beta}$ in column 1 is 3.2, indicating that if one's stock of friends in a given CZ increases by 10p.p, this increases the log odds ratio of living in that CZ rather than the reference CZ[97] by 0.32. In addition, we find $\tilde{\delta} = 1.8$, suggesting that other unobserved factors correlated with the proportion of one's friends in a given CZ further increase one's likelihood to live there.[98] The relative size of these coefficients however indicates that the causal effect is around twice the magnitude of the correlated unobserved factors.[99] In what follows, we use these two coefficients in the parameterization of the model.

6.2.2 Non-Network Parameters

Having found parameter values for $\tilde{\beta}$ and $\tilde{\delta}$, we proceed with a description of how we identify parameter values for the non-network terms: the vectors θ_j, ν_j, and A_j and the parameters

individual attended college in j, and one for whether the individual grew up in j.

[96]Appendix Tables A6 and A7 present OLS estimates from a specification implementing the same changes to the sample and control variables as described above. While the estimates change to some extent, they are still largely in the same range as our baseline estimates. For the proportion-based definition of social networks used here the coefficient estimate goes up from 0.31 to 0.36. We interpret this as evidence that although these slight deviations are important to bear in mind, the alternative sample and set of controls are unlikely to have far reaching implications for the size of the effects.

[97]We choose Dallas, TX as reference CZ.

[98]Note that this findings differs qualitatively from our earlier OLS results which, as indicated by Figure 5, found estimates of δ that were very close to zero. We attribute this difference to the fact that the multinomial logit approach does not allow for the same kind of high-dimensional control vectors as discussed in more detail above.

[99]Column 2 presents qualitatively consistent results for the number-based definition of local social networks, again showing that having a larger network in a given CZ greatly increases one's probability of choosing to live there, and that the causal effect is substantially larger than the effect of unobserved correlated factors.

α_w, and α_h. Since we do not have closed form solutions for our equilibrium conditions, we use an iterative, stepwise approach to solve for these parameter values. While we discuss the details of this approach in more detail in Appendix Section E, we highlight a few key features here.

First, to simplify our approach and to reduce the dimensionality, we impose $a_h = \frac{1}{2}$, so that households spend a third of their income on housing.

Second, we find a vector of A_j so that the model-implied labor supply from the workers' problem matches the observed population of each CZ. Thus, rather than necessarily reflecting "real amenities", these terms can be understood as residuals. We discuss the relevance of these residuals in greater detail in Section 6.3.1.

Third, when finding the vector A_j, we treat n_{ij} as an exogenous object, taking the observed distribution of one's social network as given. Once we have found the A_j's, we endogenize n_{ij} by drawing on equation 14, and let workers re-optimize until we have found an internally consistent equilibrium in which moves are minimal. Importantly, we let workers optimize asynchronously, so that while they can react to their friends' location choices, they cannot anticipate their friends' decisions, nor can they directly coordinate.[100]

Fourth, for the individuals in our sample, we empirically observe a wage elasticity of 0.93.[101] To arrive at this number, we use a shift-share approach similar to the one described in Sections 3 and 5.4 which allows us to study how populations (among those in our sample) changed between 2012-2019 in response to local wage shocks.[102] Appendix Figure A16 presents the results of this empirical exercise. To quantify the model-implied elasticity for a given set of parameters, we simulate local productivity shocks in our model and study the resulting equilibrium responses to find an estimate of the model implied elasticity. We repeat the above steps until we find a set of parameters that aligns our model closely with the data.

In Appendix Figure A17, we highlight that our model matches the data well in terms of the

[100] We believe this procedure is reasonable for multiple reasons. First, networks tend to be large, with the average individual in our sample having over 200 friends, making it implausible that individuals will anticipate all their friends' location decisions. Second, this approach is similar in spirit to the widespread use of Poisson processes constraining the extent to which individuals are able to optimize and preventing simultaneous or coordinated decisions (Arnott, 1989). Third, the coefficient estimates shown in Figure 5 for friends' moves after $\bar{t}$ are very close to zero, consistent with a lack of anticipation or coordination.

[101] Note that this estimate falls within the range of a large number of empirical studies examining the wage elasticity such as Beaudry et al. (2014); Wozniak (2010); Hornbeck and Moretti (2024).

[102] Specifically, we instrument for local wage growth in each of CZs under study using a shift-share instrument and regress changes in the local populations among those in our sample on the predicted wage growth rates. Note that the time period 2012-2019 is consistent with the sample period considered in Section 3.

target parameters. Beginning with Panel A, we find a tight relationship between the model implied wages (net of local rents) and the observed wages (again net of rent). Similarly, Panel B shows that the labor supply implied by the model matches the true population distribution. Panel C highlights that the distribution of social networks (n_{ij}) aligns closely with the observed distribution of a given individual's friends. In all these three Panels, the slope is very close to one suggesting that not only are these relationships very tight, but that the levels also line up closely. Lastly, in Panel D of the same Figure we present estimates of the wage elasticity from the exercise based on simulated shocks. We find an elasticity of 0.9, close to the observed elasticity for this population.

6.3 Model Results

6.3.1 Why People Live Where They Live

Having calibrated the model to match the data, we next assess to what extent incorporating the role played by social networks can help us understand why people live where they do. To do so, we benchmark the performance of our "network model" to a parsimonious spatial equilibrium model, to which we refer as the "basic model", which shares the same local production and housing supply functions, but does not include social ties in the worker's utility function.[103]

We intentionally compare the network model to a basic model rather than to one incorporating further elements that feature prominently in contemporaneous research — such as preferences to stay (e.g. Diamond, 2016; Piyapromdee, 2020) or moving costs (e.g. Kennan and Walker, 2011) — to highlight how the addition of a single factor, social networks, for which we can draw on well-identified reduced form evidence based on Sections 4 and 5 can help to understand patterns of migration without the need to incorporate other forces. We calibrate the basic model so that it also matches the data in terms of (a) local wages and rents, (b) local populations, and (c) the observed wage elasticity for individuals in our sample. Due to the absence of network effects in the basic model, we do not match those moments. The corresponding parameter values are shown in Appendix Table A9[104]

[103]While the basic model shares equations 11-12 with the network-model, instead of equation 13, a worker's utility function is given by:

$$\tilde{U}_{ij,t} = \tilde{\alpha}_w \ln(w_{j,t}) - \tilde{\alpha}_h \ln(r_j) + \tilde{A}_{j,t} + \tilde{\epsilon}_{ij,t} \tag{16}$$

where the tilde's indicate that the parameters can differ relative to those shown in equation 13 due to the absence of the network terms.

[104]We use an iterative stepwise approach to match the basic model to the data, similar to that used for the network model. However, due to the absence of the network effects, our approach here is abbreviated relative to one used for the network model. Appendix Section E provides additional details.

To compare the performance of the network and the basic models in their ability to explain why people live where they do, we focus our attention on the relevance of residuals in these models. As described in Section 6.2, as part of the model calibration, we find vectors of local amenities (A_j for the network model, and $\tilde{A}_j$ for the basic model) that allow us to match the populations implied by the worker's utility maximization problem to the observed values. Rather than necessarily reflecting "real amenities", these terms can therefore be thought of residuals, as their primary purpose is to help us rationalize local population sizes.

We find that we can greatly reduce the relevance of these residuals by taking into account the role played by social networks. In Panel A of Figure 9, we compare the distributions of the A_j terms from the network model and the $\tilde{A}_j$ terms from the basic-model. For interpretability, we convert these terms to their dollar equivalents, so that they can be understood as a willingness to pay (WTP) for local amenities (or residuals).[105] The two distributions exhibit a strikingly different image: the distribution of the of WTP for $\tilde{A}_j$ ranges from \$-50,000 to \$+50,000, suggesting that a frictionless spatial equilibrium model requires very highly valued amenities besides wages and rents to reconcile the distribution of populations with this model in which people are free to choose where to live. In contrast, the distribution of the WTP for A_j's is much tighter, ranging from around \$-20,000 to \$+15,000, so that the variance of the A_j terms is less than a quarter that of the $\tilde{A}_j$ terms. This highlights that a large fraction of the unobserved residuals — or high WTP for other local factors — are social networks. Once we account for networks directly, the relevance of other factors besides wages and rents is greatly diminished.

The inclusion of social networks in the spatial equilibrium model not only helps to reduce the relevance of other factors, but also helps to render the remaining residuals less opaque, aligning them more closely with "real amenities". In Panel B of Figure 9, we correlate the values for A_j and $\tilde{A}_j$ with measures of place characteristics commonly thought of as amenities: air quality, weather and retail availability (e.g. Chay and Greenstone, 2005; Rappaport, 2007; Glaeser et al., 2001). For the basic model we find at most weak correlations ranging between just under 0 and 0.2. This seems counterintuitive since these factors are commonly considered important amenities that influence an individual's decision where to live. In contrast, the purple bars indicate that for the network model the correlations between the amenity measures and the A_j terms are quite large, consistently above 0.4 and in some cases even above 0.6, consistent with the notion that those forces likely influence an individual's decision where to live.

[105] For the scaling, we divide the raw values for A_j and $\tilde{A}_j$ by the α_w and $\tilde{\alpha_w}$ values from the utility function, and then multiply them with the average local wages. We further de-mean the resulting vectors.

What drives the reduction in the residuals and the fact that they align more closely with standard measures of local amenities? We argue that these results are driven by three factors. First, as shown in Section 5, network effects are large: every additional friend in a given CZ has a meaningful impact on one's decision where to live. Second, as discussed in Section 3, networks are highly spatially concentrated, with the typical individual in our sample having around 70% of their friends in their home CZ.[106] This creates a strong pull factor to stay in one's home CZ. Third, we find evidence that this pull factor is particularly strong among individuals living in places with a less favorable economy as people in those places have especially concentrated networks. Specifically, the extent to which one's friends live in the same CZ as oneself is negatively correlated with the distribution of wages and rents: for wages, the correlation between the share of one's friends living in the same CZ is -0.38; for rents, that same correlation is -0.71. As an illustrative example, individuals living in Detroit have on average 81% of their friends in the same CZ whereas those living in Austin only have 47% of their friends there.[107]

Taken together, these findings highlight that social networks are critical in explaining why people live where they do, and are particularly important for understanding why people stay in places with low economic prospects. Due to the spatial concentration of social networks, migrating to a new place and leaving behind one's ties comes at a high cost, one that is even higher for those living in places with a weaker economy, since individuals in those places tend to have stronger local ties. Conversely, the income gains from moving to a CZ without social ties need to be relatively large to make such a move worthwhile.

6.3.2 Why More-Educated People Respond More Strongly to Shocks

A large literature has highlighted that migration responses to changing economic conditions vary substantially across socioeconomic groups (e.g. Bound and Holzer, 2000; Gregg et al., 2004; Wozniak, 2010; Sprung-Keyser et al., 2022). In light of this literature, we estimate empirically how individuals with a college degree ("more-educated") and those without a college degree ("less-educated") respond differently to local wage shocks, and contextualize these results using our network model.

[106]This number differs slightly from our earlier statistics presented in Section 3 because for the purpose of the present exercise we focus on friends as of 2019 among individuals living in one of the 50 largest CZs.

[107]This also helps to explain why the correlations in Panel B of Figure 9 are substantially larger and in some cases "flip" for the network model relative to the basic model. The fact that the correlations are low for the basic model is not merely a result of measurement error or random noise; rather, it reflects in part the fact that distribution of social networks is negatively correlated with local wages and rents. Accounting for the role played by social networks and the variation in the local concentration of networks therefore enables us to "correct" for this fact, so that residuals A_j align more closely with common measures of amenities.

We begin by investigating empirically how more- and less-educated individuals responded to local wage growth in the aftermath of the Great Recession. More specifically, we follow the approach used in the calibration of the model (see Appendix E), and employ a shift-share instrument to estimate group-specific wage elasticities among individuals in our sample in response to local wage growth between 2012-2019.[108] As shown in the green bars in Figure 10, we find a substantially higher wage elasticity for the more-educated (1.15) than we do for the less-educated (0.6).

The differences in the elastiticies between the two groups may in part be driven by differences in the dispersion of their social networks. Consistent with that conjecture, recall from Section 3 that the less-educated have a substantially more locally concentrated social network than the more-educated: for those without a college degree the average share of friends living in the same CZ is 75%, compared with just over 55% for those with a college degree.[109]

To test for the plausibility of our explanation, we simulate the effects of shocks in our network model. Concretely, we replicate the exercise conducted to match the model to the data in terms of the average wage elasticity, this time studying specifically how the elasticity differs between more- and less-educated individuals. Importantly, as in all parts of Section 6, we constrain the more- and less-educated to share the same parameters (α_w, α_h, β, δ, and A_j) in the utility function. In Appendix Sections C.1.2 we demonstrate that this assumption is at least to first order reasonable as there do not appear to be large differences in how college- and non-college-educated individuals respond to their friends' locations when deciding where to live.[110] Moreover, more- and less-educated individuals face identical environments, so that any differences in the elasticities between those groups must come from differences in their networks.[111]

[108]Specifically, we use the same shift-share instrument as in Section 6.2.2, but study population responses separately for those with and without a college degree.

[109] Our hypothesis that differences in the concentration of social networks between more- and less-educated individuals drive part of the observed gap in the wage elasticities for these groups is to some extent consistent with prior explanations, yet has somewhat different implications. Existing work studying gaps in the wage elasticity of more- and less-educated individuals has argued that these gaps are driven by different preferences among the two groups and that the less-educated have a stronger desire to stay close to one's home (e.g. Diamond, 2016). In contrast, our network-based explanation is *not* about ties to a given place, but rather about ties to individuals. Since most of one's social network lives in one's home CZ, these explanations are clearly connected. Yet, because individuals are mobile, in the network-based explanation, one's attachment to a given CZ can vary depending on the actions of one's friends. Footnote 84 discusses how further implications of these differences.

[110]Appendix Section C.2 further highlights that there does not seem to be a meaningful interaction effect between the role of social networks and local economic growth consistent with our model setup which abstracts from such interactions.

[111]We emphasize that more- and less-educated individuals share the same parameters (α_w, α_h, β, δ, and A_j) in the utility function and that other factors are also the same across the two groups. This may spark questions over why we would still expect differences in the elasticity. While we do not have a closed form

The results from the simulation exercise suggest that much of the elasticity gap between college and non-college educated individuals can indeed be attributed to differences in the dispersion of their networks. As shown in the purple bars in Figure 10, the simulated shocks generate a sizeable gap in the elasticities of those with and without a college degree. While the empirical estimates lack the precision to let us quantify exactly how much of the observed gap can explained by differences in the networks, the point estimates suggest that networks account for a large share of the gap, despite the fact that, by construction, these two groups have identical preferences and only differ with respect to the dispersion of their social networks. Figure 10 thus highlights that networks play a critical role in explaining why more-educated individuals are more elastic than the less-educated: since the former have a smaller relative network in their home CZ, they are less inclined to stay and thus respond more strongly to economic shocks. In turn, the differences in the distribution of social networks might therefore also drive differences in the extent to which more- and less-educated bear the burden from local downturns, an issue we aim to investigate in future research.

7 Conclusion

Despite vast differences in economic opportunities across commuting zones, few Americans move and those who do rarely go to places offering much greater opportunities. These patterns are especially true for less-educated individuals. This paper highlights that social networks play a crucial role in explaining these observations.

Using individual-level data from Facebook on social networks and locations of millions of users in the United States, we highlight that social networks are highly spatially concentrated with the average individual living within 100 miles of nearly 80% of their friends. We then present descriptive and quasi-experimental evidence showing that having an additional friend in a given region significantly increases an individual's likelihood of choosing to live there, whether it involves relocating or remaining in their current location.

By incorporating our findings on the effects of social networks and their spatial concentration into a model of spatial equilibrium, we demonstrate that networks can explain why many

---

solution for the wage elasticity, in a model without network spillovers we would obtain a standard logit elasticity of:

$$\epsilon_{ij}^{w} = \alpha_w \times (1 - \psi_{ij})$$

Even in the absence of spillovers, ψ_{ij} is a function of one's social network with $\frac{\partial \psi_{ij}}{\partial n_{ij}} > 0$, so that $\frac{\partial \epsilon_{ij}^{w}}{\partial n_{ij}} < 0$ since $\psi_i j \leq 1$. This highlights how differences in social networks can drive differences in elasticities.

individuals in economically distressed areas choose not to relocate to regions with significantly better economic opportunities. Each additional friend in a given region has a notable impact on one's decision to reside there. However, since most friends are local, individuals encounter a high social cost when considering a move to areas with better economic prospects but lacking social connections. This dynamic is particularly pronounced for those living in economically challenged regions and among less-educated individuals, as these groups tend to have more concentrated social networks.

Overall, our paper highlights the importance of considering the role of social networks in shaping residential choices when thinking about access to economic opportunities. By driving our migration decisions, social networks shape the economic opportunities we have access to. While networks can keep individuals in distressed areas, they also offer pathways to regions with greater opportunities, as those with larger networks in high-growth areas are more likely to relocate. Social networks can thus both facilitate and inhibit access to local economic opportunities. In future research, we aim to study how policy makers can leverage these results when thinking about the incidence of adverse economic shocks and how to promote more equitable access to economic opportunities.

References

Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou, "Intergenerational Mobility in Africa," *Econometrica*, 2021, *89* (1), 1–35.

Amior, Michael, "Education and Geographical Mobility: The Role of the Job Surplus," *American Economic Journal: Economic Policy*, November 2024, *16* (4), 341–81.

Armington, Paul S., "A Theory of Demand for Products Distinguished by Place of Production (Une théorie de la demande de produits différenciés d'après leur origine) (Una teoría de la demanda de productos distinguiéndolos según el lugar de producción)," *Staff Papers (International Monetary Fund)*, 1969, *16* (1), 159–178.

Arnott, Richard, "Housing vacancies, thin markets, and idiosyncratic tastes," *The Journal of Real Estate Finance and Economics*, 1989, *2* (1), 5–30.

Asher, Sam, Paul Novosad, and Charlie Rafkin, "Intergenerational Mobility in India: New Measures and Estimates across Time and Social Groups," *American Economic Journal: Applied Economics*, April 2024, *16* (2), 66–98.

Autor, David H. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, August 2013, *103* (5), 1553–97.

_ , _ , and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, October 2013, *103* (6), 2121–68.

Bailey, Michael, Drew Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel, "Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing during the COVID-19 Pandemic," *Journal of Political Economy Microeconomics*, 2024, *2* (3), 463–494.

_ , Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, August 2018, *32* (3), 259–80.

Barro, Robert and Xavier Sala i Martin, "Convergence across States and Regions," *Brookings Papers on Economic Activity*, 1991, *22* (1), 107–182.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?*, W.E. Upjohn Institute, 1991.

— , "What Proportion of Children Stay in the Same Location as Adults, and How Does This Vary Across Location and Groups?," Upjohn Working Papers 09-145, W.E. Upjohn Institute for Employment Research February 2009.

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins, "A Dynamic Model of Demand For Houses and Neighborhoods," *Econometrica*, 2016, *84* (3), 893–942.

— , Stephen L. Ross, and Giorgio Topa, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 2008, *116* (6), 1150–1196.

Beaudry, Paul, David A. Green, and Benjamin M. Sand, "Spatial equilibrium with unemployment and wage bargaining: Theory and estimation," *Journal of urban economics*, 2014, *79*, 2–19.

Bergman, Peter, Eric W Chan, and Adam Kapor, "Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment," Working Paper 27209, National Bureau of Economic Research May 2020.

Bilal, Adrien, "The Geography of Unemployment*," *The Quarterly Journal of Economics*, 03 2023, *138* (3), 1507–1576.

Blackaby, D H and D N Manning, "The North-South Divide: Questions of Existence and Stability?," *Economic Journal*, 1990, *100* (401), 510–27.

Blanchard, Olivier Jean and Lawrence F. Katz, "Regional Evolutions," *Brookings Papers on Economic Activity*, 1992, *23* (1), 1–76.

Blumenstock, Joshua E, Guanghua Chi, and Xu Tan, "Migration and the Value of Social Networks," *The Review of Economic Studies*, 12 2023, p. rdad113.

Borjas, George J., "Does Immigration Grease the Wheels of the Labor Market?," *Brookings papers on economic activity*, 2001, *2001* (1), 69–119.

Bound, John and Harry J. Holzer, "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s," *Journal of labor economics*, 2000, *18* (1), 20–54.

Cadena, Brian C. and Brian K. Kovak, "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession," *American Economic Journal: Applied Economics*, January 2016, *8* (1), 257–90.

Caldwell, Sydnee and Oren Danieli, "Outside Options in the Labour Market," *The Review of Economic Studies*, 01 2024, p. rdae006.

Card, David, Jesse Rothstein, and Moises Yi, "Location, Location, Location," Working Paper 31587, National Bureau of Economic Research August 2023.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath, "Migration with Endogenous Moving Costs," *The American Economic Review*, 1996, *86* (4), 909–930.

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Chen, Yong and Stuart S. Rosenthal, "Local amenities and life-cycle migration: Do people move for jobs or fun?," *Journal of Urban Economics*, 2008, *64* (3), 519–537.

Chetty, Raj and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

_ , John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Income Segregation and Intergenerational Mobility Across Colleges in the United States*," *The Quarterly Journal of Economics*, 02 2020, *135* (3), 1567–1633.

_ , _ , Nathaniel Hendren, Maggie R Jones, and Sonya R Porter, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," Working Paper 25147, National Bureau of Economic Research October 2018.

_ , Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt, "Social capital I: measurement and associations with economic mobility," *Nature (London)*, 2022, *608* (7921), 108–121.

_ , and _ , "Social capital II: determinants of economic connectedness," *Nature*, 2022, *608* (7921), 122–134.

_ , Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin,

Matthew Jacob, Drew Johnston, Martin Koenen et al., "Social capital I: measurement and associations with economic mobility," *Nature*, 2022, pp. 1–14.

Costa, Dora L., Matthew E. Kahn, Christopher Roudiez, and Sven Wilson, "Persistent social networks: Civil war veterans who fought together co-locate in later life," *Regional Science and Urban Economics*, 2018, *70*, 289–299.

Danieli, Oren, Daniel Nevo, Itai Walk, Bar Weinstein, and Dan Zeltzer, "Negative Control Falsification Tests for Instrumental Variable Designs," Papers 2312.15624, arXiv.org December 2023.

Dao, Mai, Davide Furceri, and Prakash Loungani, "Regional Labor Market Adjustment in the United States: Trend and Cycle," *The Review of Economics and Statistics*, 05 2017, *99* (2), 243–257.

Davis, Donald R. and Jonathan I. Dingel, "A Spatial Knowledge Economy," *American Economic Review*, January 2019, *109* (1), 153–70.

Desmond, Matthew, "Eviction and the Reproduction of Urban Poverty," *American Journal of Sociology*, 2012, *118* (1), 88–133.

Deutscher, Nathan and Bhashkar Mazumder, "Intergenerational mobility across Australia and the stability of regional estimates," *Labour Economics*, 2020, *66*, 101861.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, March 2016, *106* (3), 479–524.

_ **and Enrico Moretti**, "Where is Standard of Living the Highest? Local Prices and the Geography of Consumption," Working Paper 29533, National Bureau of Economic Research December 2021.

_ **, Tim McQuade, and Franklin Qian**, "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, September 2019, *109* (9), 3365–94.

Duggan, Maeve, Shannon Greenwood, and Andrew Perrin, "Social Media Update 2016," Report, Pew Research Center 2016.

Egger, Dennis, Daniel Auer, and Johannes Kunz, "Effects of migrant networks on labor market integration, local firms and employees," 2022.

Facebook, "Facebook Form 10-K, 2023 Annual Report," 2023.

Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Katheryn N. Russ, "In Search of the Armington Elasticity," *The Review of Economics and Statistics*, March 2018, *100* (1), 135–150.

Gaubert, Cecile and Rebecca Diamond, "Spatial Sorting and Inequality," *Annual Review of Economics*, 2022, *14* (1), 795–819.

Glaeser, Edward L. and Charles Redlick, "Social Capital and Urban Growth," *International Regional Science Review*, 2009, *32* (3), 264–299.

_ and Joshua D. Gottlieb, "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, *2008* (1), 155–253.

_ , Jed Kolko, and Albert Saiz, "Consumer city," *Journal of Economic Geography*, 01 2001, *1* (1), 27–50.

Green, Allison, "Networks and Geographic Mobility: Evidence from World War II Navy Ships," 2024.

Gregg, P, S Machin, and A Manning, "Mobility and joblessness," in "Seeking a Premier Economy: The Economic Effects of British Economic Reforms 1980-2000" Univ Chicago Press CHICAGO 2004, pp. 371–410.

Handbury, Jessie, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," *Econometrica*, 2021, *89* (6), 2679–2715.

_ and David E. Weinstein, "Goods Prices and Availability in Cities," *The Review of Economic Studies*, 09 2014, *82* (1), 258–296.

Hansch, Michelle, Jan Nimczik, and Alexandra Spitz-Oener, "Workplace Connections and Labor Migration: The Role of Information in Shaping Expectations," Rationality and Competition Discussion Paper Series 490, CRC TRR 190 Rationality and Competition January 2024.

Hornbeck, Richard and Enrico Moretti, "Estimating Who Benefits from Productivity Growth: Local and Distant Effects of City Productivity Growth on Wages, Rents, and Inequality," *The Review of Economics and Statistics*, 05 2024, *106* (3), 587–607.

Ioannides, Yannis M. and Linda Datcher Loury, "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, December 2004, *42* (4), 1056–1093.

Jones, Jason J, Jaime E Settle, Robert M Bond, Christopher J Fariss, Cameron Marlow, and James H Fowler, "Inferring tie strength from online directed behavior," *PloS one*, 2013, *8* (1), e52168.

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kline, Patrick and Enrico Moretti, "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," *Annual Review of Economics*, 2014, *6* (Volume 6, 2014), 629–662.

LinkedIn, "Graduating Next Year? The Time is Now to Begin Your Job Search," https://www.linkedin.com/pulse/graduating-next-year-time-now-begin-your/ 2023. Last Accessed: 2024-07-15.

Lkhagvasuren, Damba, "Education, mobility and the college wage premium," *European Economic Review*, 2014, *67* (C), 159–173.

Machin, Stephen, Kjell G. Salvanes, and Panu Pelkonen, "Education and Mobility," *Journal of the European Economic Association*, 2012, *10* (2), 417–450.

Mahajan, Parag and Dean Yang, "Taken by storm: Hurricanes, migrant networks, and US immigration," *American Economic Journal: Applied Economics*, 2020, *12* (2), 250–277.

Malamud, Ofer and Abigail Wozniak, "The Impact of College on Migration: Evidence from the Vietnam Generation," *Journal of Human Resources*, 2012, *47* (4).

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," *Conditional Logit Analysis of Qualitative Choice Behavior, Frontiers in Econometrics*, 1974.

McKenzie, David and Hillel Rapoport, "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of development Economics*, 2007, *84* (1), 1–24.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 2001, *27* (Volume 27, 2001), 415–444.

Moretti, Enrico, "Local Labor Markets," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, 1 ed., Vol. 4B, Elsevier, 2011, chapter 14, pp. 1237–1313.

Munshi, Kaivan, "Networks in the modern economy: Mexican migrants in the US labor market," *The Quarterly Journal of Economics*, 2003, *118* (2), 549–599.

__ , “Social Networks and Migration,” *Annual Review of Economics*, 2020, *12* (Volume 12, 2020), 503–524.

Nelson, Phillip, “Migration, Real Income and Information,” *Journal of Regional Science*, 1959, *1* (2), 43–74.

Notowidigdo, Matthew J., “The Incidence of Local Labor Demand Shocks,” *Journal of Labor Economics*, 2020, *38* (3), 687–725.

Piyapromdee, Suphanit, “The Impact of Immigration on Wages, Internal Migration, and Welfare,” *The Review of Economic Studies*, 08 2020, *88* (1), 406–453.

Porcher, Charly, “Migration with costly information,” *Unpublished Manuscript*, 2020, *1* (3).

__ **, Eduardo Morales, and Thomas Fujiwara**, “Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach,” Working Paper 32413, National Bureau of Economic Research May 2024.

QCEW, “Quarterly Census of Employment and Wages,” U.S. Bureau of Labor Statistics 2024.

Rappaport, Jordan, “Moving to nice weather,” *Regional Science and Urban Economics*, 2007, *37* (3), 375–398.

Redding, Stephen J. and Esteban Rossi-Hansberg, “Quantitative Spatial Economics,” *Annual Review of Economics*, 2017, *9* (Volume 9, 2017), 21–58.

Rees, Albert, “Information networks in labor markets,” *The American Economic Review*, 1966, *56* (1/2), 559–566.

Roback, Jennifer, “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Sherwin, “Wage-based indexes of urban quality of life,” *Current issues in urban economics*, 1979, pp. 74–104.

Sahai, Harshil and Mike Bailey, “Social Networks and Internal Migration: Evidence from Facebook in India,” *Working Paper*, 12 2023.

Saiz, Albert, “The Geographic Determinants of Housing Supply*,” *The Quarterly Journal of Economics*, 08 2010, *125* (3), 1253–1296.

Sjaastad, Larry A., "The Costs and Returns of Human Migration," *Journal of Political Economy*, 1962, *70* (5), 80–93.

Sprung-Keyser, Ben and Sonya Porter, "The Economic Geography of Lifecycle Human Capital Accumulation: The Competing Effects of Labor Markets and Childhood Environments," Working Papers 23-54, Center for Economic Studies, U.S. Census Bureau November 2023.

_, Nathaniel Hendren, and Sonya Porter, "The Radius of Economic Opportunity: Evidence from Migration and Local Labor Markets," Working Papers 22-27, Center for Economic Studies, U.S. Census Bureau July 2022.

Stuart, Bryan A. and Evan J. Taylor, "Migration Networks and Location Decisions: Evidence from US Mass Migration," *American Economic Journal: Applied Economics*, July 2021, *13* (3), 134–75.

Topel, Robert H., "Local Labor Markets," *The Journal of political economy*, 1986, *94* (3), S111–S143.

Train, Kenneth, *Discrete choice methods with simulation*, Cambridge, U.K. ; New York: Cambridge University Press, 2003.

Wilson, Riley, "Moving to Economic Opportunity: The Migration Response to the Fracking Boom," *Journal of Human Resources*, 2020.

Wozniak, Abigail, "Are college graduates more responsive to distant labor market opportunities?," *The Journal of human resources*, 2010, *45* (4), 944–970.

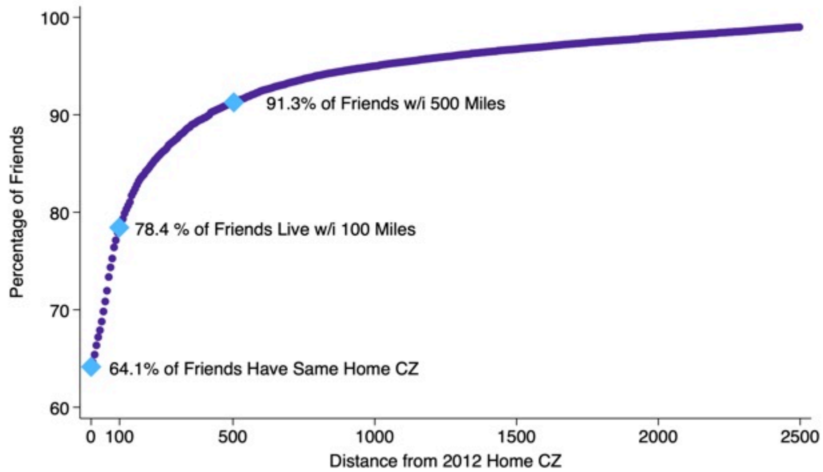
Yagan, Danny, "Employment Hysteresis from the Great Recession," *The Journal of political economy*, 2019, *127* (5), 2505–2558.

Zabek, Mike, "Local Ties in Spatial Equilibrium," *American Economic Journal: Macroeconomics*, April 2024, *16* (2), 287–317.

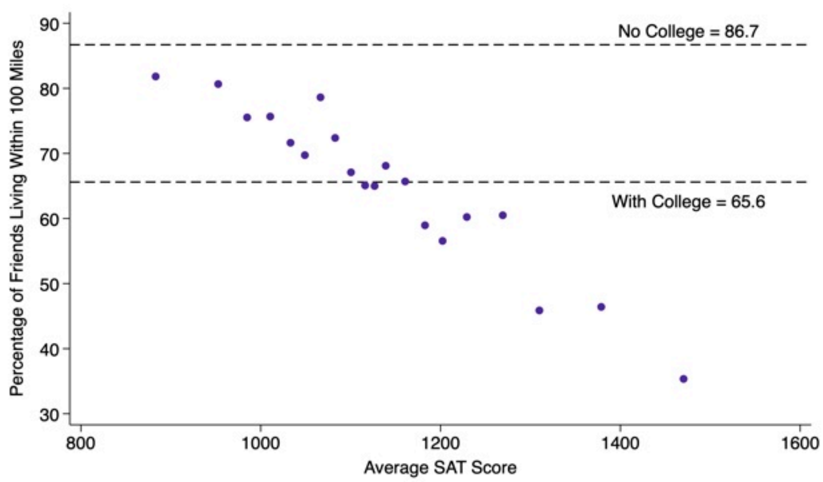
Figures

Figure 1: Spatial Concentration of Social Networks

(a) Proportion of Friends by Distance



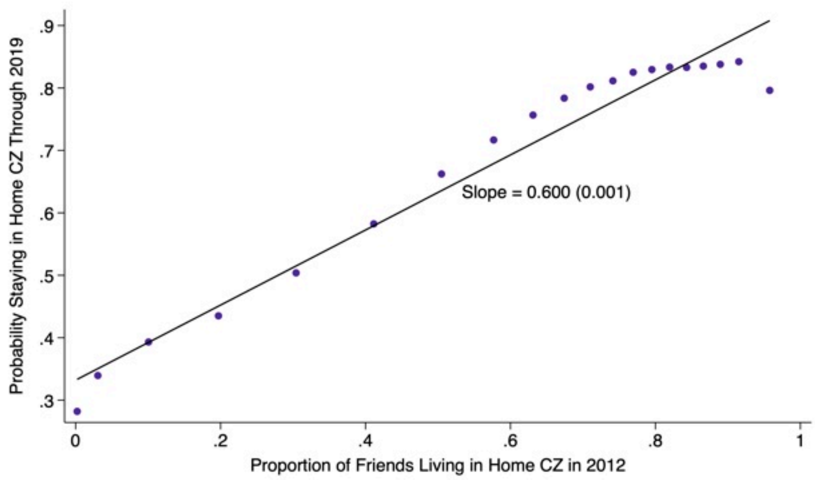
(b) Dispersion of Social Networks By Educational Attainment



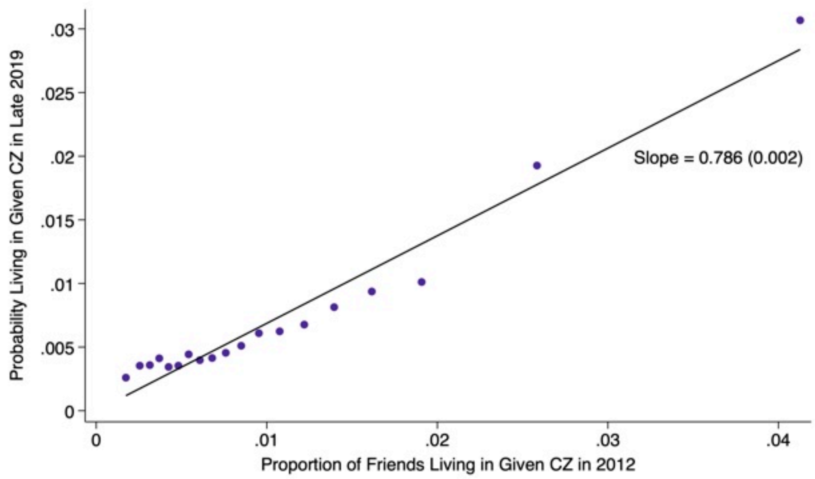
Notes: Figure presents statistics regarding the spatial concentration of individuals in the Expanded Sample. Panel (a) shows the cumulative percentage of friends by distance from the individuals' home CZ, or the CZ they lived in as of January 2012. Analysis subsets to friends made as of that point in time. Panel (b) shows heterogeneity in the spatial concentration of friends by educational attainment. The SAT score measure on the horizontal axis corresponds to 2013 average SAT scores of an individual's self-reported college. Panel (b) includes friendships through 2019.

Figure 2: Residential Choice and Location of Social Networks

(a) Probability of Staying in Home CZ



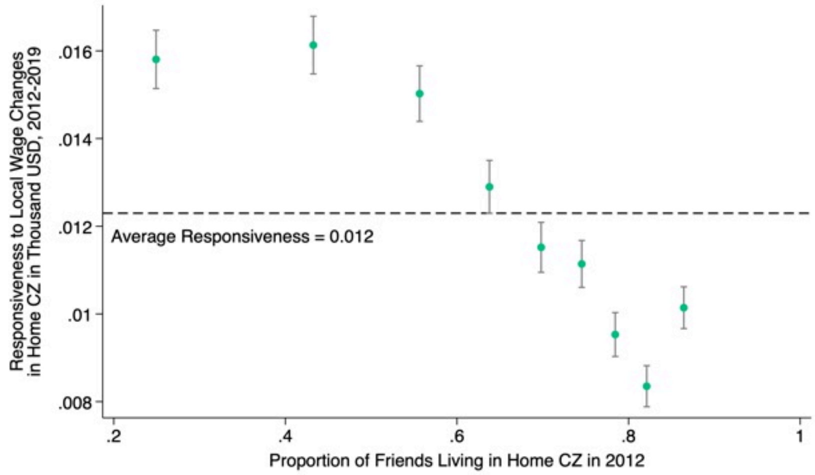
(b) Probability of Living in Given Alternative CZ



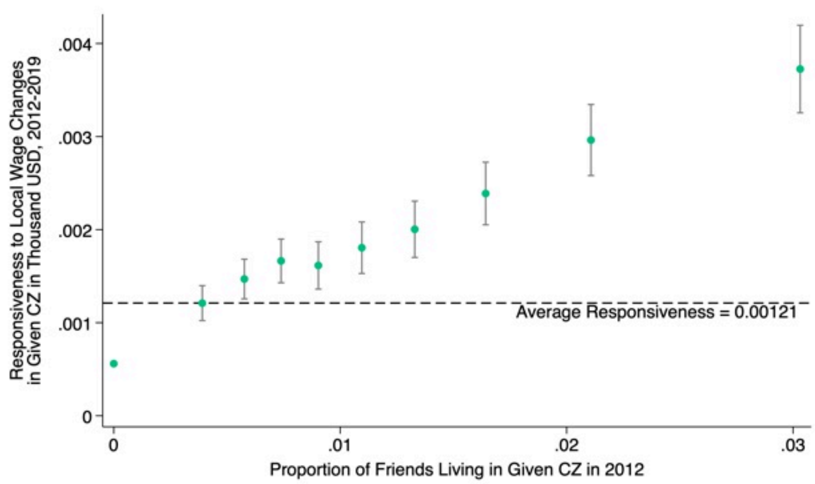
Notes: Figure presents descriptive statistics on the relationship between social networks and migration from the Expanded Sample discussed in 2. Panel (a) shows a binned scatter plot where the vertical axis corresponds to the probability a given user resides in the same CZ in December 2019 as in January 2012. The horizontal axis displays the proportion of friends a user has in the 2012 CZ as of the beginning of 2012, i.e. using the friends' location in January 2012, and restricting to friendships already formed at that point in time. Panel (b) shows a binned scatter plot where the vertical axis corresponds to the probability a given user resides in a given CZ, excluding a user's 2012 CZ and focusing exclusively on CZs that are among the 50 largest CZs shown in Figure A2. The horizontal axis displays the proportion a user has in a given CZ as of early 2012. The sample of users included in both Panels is defined in more detail in Section 2. Panel (b) is estimated based on a two-percent random sample for computational reasons. The black lines in both Panels represent the best-fit from a linear regression on the binned series.

Figure 3: Heterogeneity in Responsiveness to Changes in Local Wages

(a) Δ Pay in Home CZ

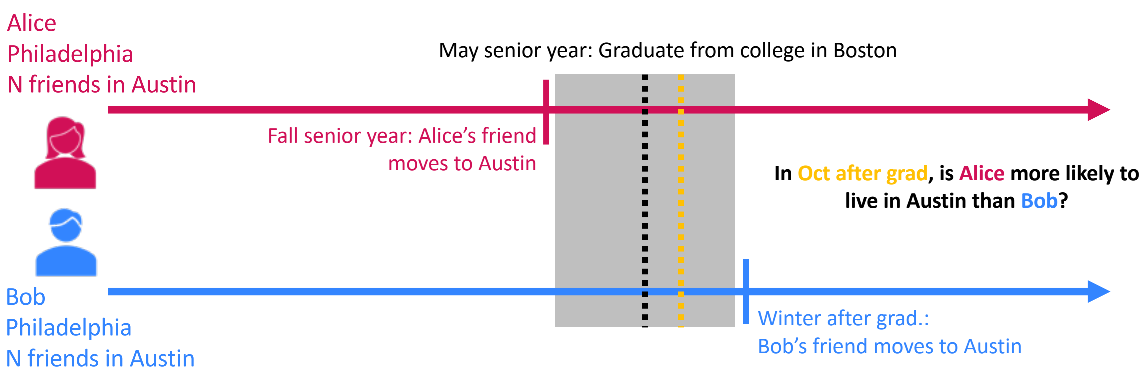


(b) Δ Pay in Alternative CZ



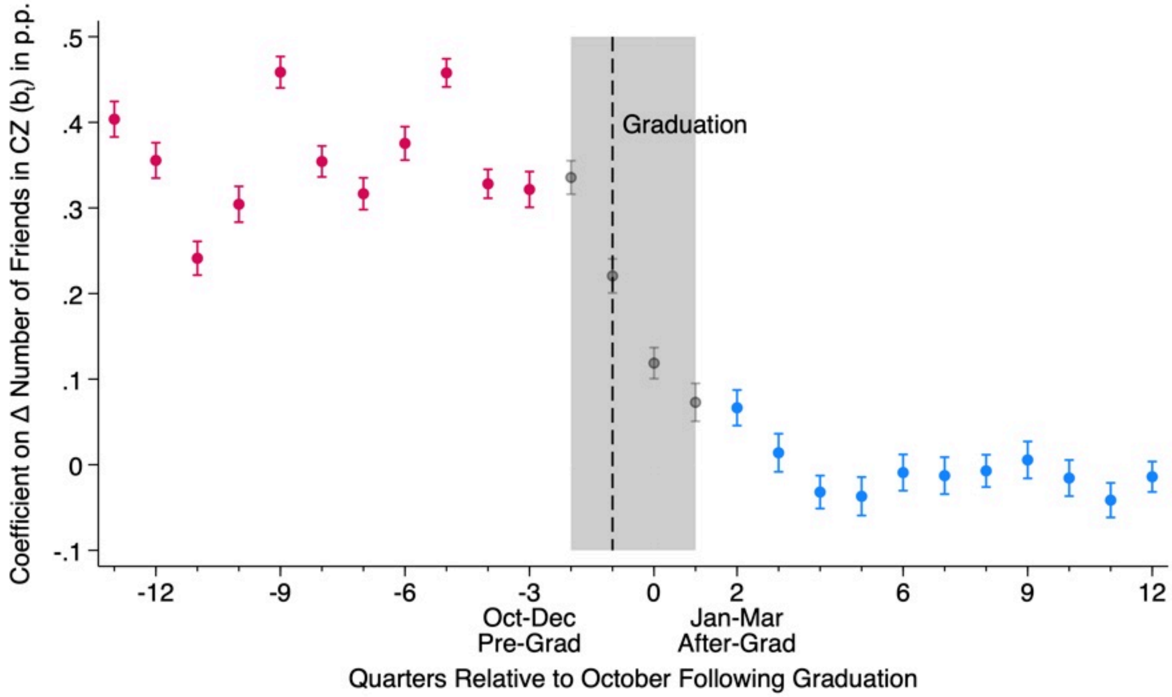
Notes: Figure shows heterogeneity in an individual's responsiveness to recent local wage growth by level of social connectedness. In both Panels, our measure of recent local wage growth corresponds to the predicted wage growth (in thousand dollars) between 2012 and 2019. We predict wage growth using a shift-share instrument based on 6-digit NAICS codes and drawing on data from the Quarterly Census of Employment and Wages described in Section 2. Panel (a) presents coefficient estimates from a regression of whether a given individual remains in their 2012 home CZ through December 2019 on local wage growth in the 2012 CZ, separately for individuals with different levels of proportions of their friends in the 2012 CZ. A binned scatter plot of the relationship between staying in the 2012 CZ and wage growth for the full sample is shown in Panel (a) of Appendix Figure A3. Panel (b) presents coefficient estimates from a regression of whether, as of December 2019, a given individual lives in a given CZ (excluding their 2012 CZ) on recent local wage growth in that CZ, separately for individuals with different levels of proportions of their friends in said CZ as of early 2012. A binned scatter plot of the relationship between moving to a given and wage growth for the full sample is shown in Panel (b) of Appendix Figure A3. In both Panels we omit the coefficient estimate corresponding to individuals in the bottom decile of the x-axis measure for expositional reasons.

Figure 4: Illustrative Example of Identification Strategy



Notes: Figure presents a stylized example to illustrate the identification strategy described in Section 4. In this example, there are two individuals, *Alice* and *Bob*, who both grew up in Philadelphia and graduate from a college in Boston at the same point in time. The black dashed line indicates the point in time at which *Alice* and *Bob* graduate from college, and the yellow dashed line corresponds to the point in time at which we measure their post-graduation location. The gray bar around the two dashed lines stands for the time period during which the two graduates are assumed to decide where to live after graduation or during which they can anticipate friends' moves shortly after graduation. In this example, *Alice* and *Bob* decide whether they want to live in Austin after graduation. At the beginning of the study period, the two individuals had the same number of friends (N) in Austin. While both *Alice* and *Bob* have one friend moving to Austin around the point in time at which they decide where to live after graduation, the timing of their friends' moves differs. *Alice*'s friend moves just before *Alice* and *Bob* decide whether to move to Austin, while *Bob*'s friend moves slightly after.

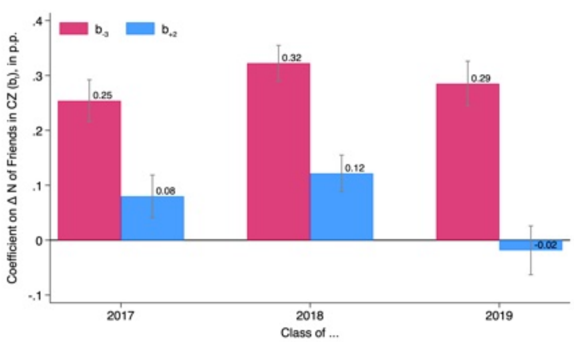
Figure 5: Effect of Friends' Moves on Residential Choice



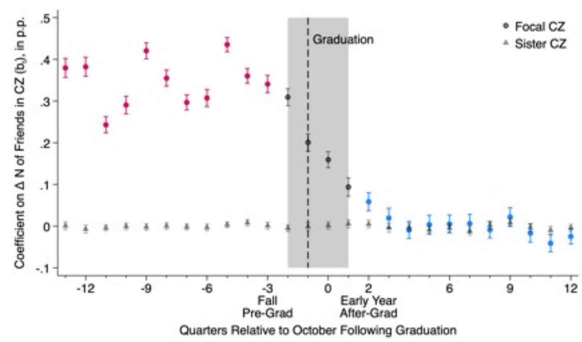
Notes: Figure presents coefficients of equation 4. Red coefficient estimates on the left-hand-side of the graph correspond to friends' moves happening between three and 13 quarters before we measure the college graduates' location. Blue coefficient estimates on the right correspond to friends' moves happening between two and 12 quarters after we measure the college graduates' location. Gray coefficient estimates in the middle of the plot correspond to friends' moves happening around the time the college goers graduate, i.e., between two quarters before and one quarter after we measure the college graduates' location. Vertical lines going through coefficient estimates show 95% confidence intervals. The gray dashed line indicates the time at which college graduates are assumed to graduate, i.e., in between May and July of senior year. The gray bar that spans between two quarters before and one quarter after we measure the college graduates' location denotes the window of time during which college graduates are assumed to decide where to live, or during which they can likely anticipate friend moves' happening shortly after graduation.

Figure 6: Evidence Supporting Identifying Assumption

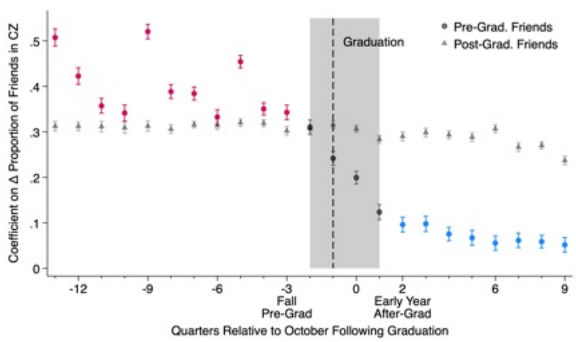
(a) Effects by Cohort



(b) Sister CZs

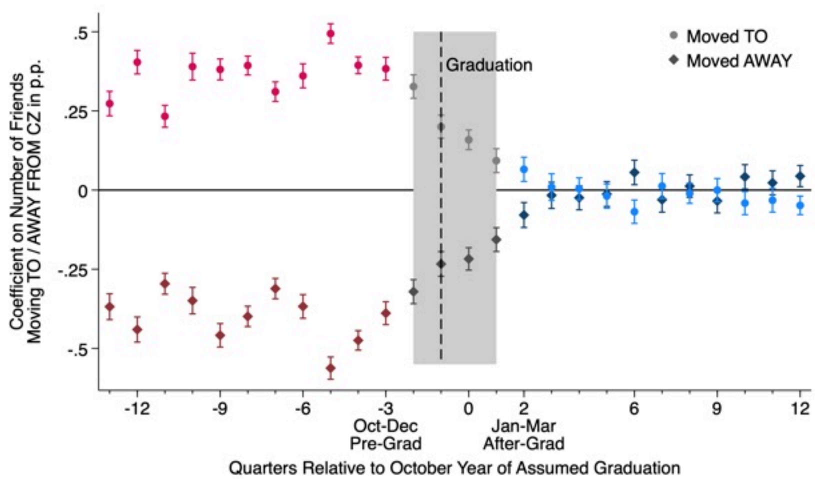


(c) Post-Graduation Friends



Notes: Figure presents evidence supporting the identifying assumption. Panel (a) presents coefficient estimates corresponding to equation 4, estimated separately for individuals graduating in 2017, 2018 and 2019. Red bars show estimates for friends' moves three quarters before we measure graduates' locations, or b_{-3}. Blue bars show estimates for friends' moves two quarters after we measure graduates' locations, or b_{+2}. Panel (b) presents estimates of equation 7 regarding the influence of friends' moves to focal and sister CZs on a graduates' likelihood of living in the focal CZ. Sister CZs are CZs with very similar industry compositions as the focal CZs. To identify sister CZs, we employ a nearest neighbor matching procedure based on industry-specific employment shares using 6-digit NAICS codes. The gray triangles correspond to coefficient estimates for sister CZs, while the colored circles represent coefficient estimates for the focal CZ. Panel (c) presents estimates of equation 8 regarding the influence of pre- and post-graduation friends' moves to a given CZ on a graduates' likelihood of living in that CZ. Pre-graduation friends follow the definition of friends in Figure 5. Post-graduation friends are people who graduates befriend 4 years or more after graduation, so that they had not yet befriended these people by graduation. The gray triangles correspond to coefficient estimates for post-graduation friends, while the colored circles represent coefficient estimates for the pre-graduation friends. Panel (c) uses the proportion-based definition of social networks to account for large differences in the average number of pre- and post-graduation friends. Panel (c) restricts to individuals graduating in 2017 and friends' moves happening at most nine quarters after t^* due to concerns of the COVID-19 pandemic otherwise potentially affecting the results. The layout of Panels (b) and (c) follows that of Figure 5. Vertical lines in all panels indicate 95% confidence intervals.

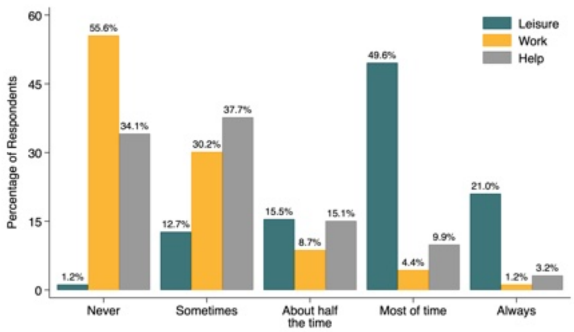
Figure 7: Network Effects for Friends Moving To and Away From CZ



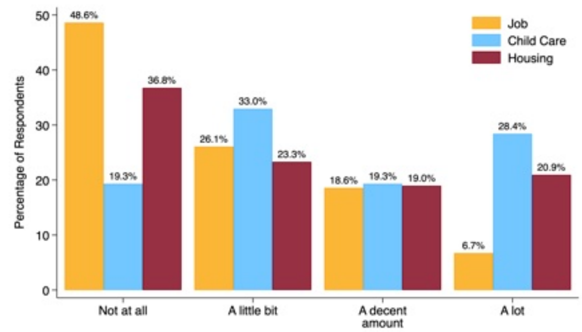
Notes: Figure presents estimates of effects of friends moving to and away from a given CZ based on equation 10. The layout follows that of Figure 5. Friends' moves to a given CZ are shown in lighter shades and use circles, while friends' moves away from a given CZ are shown in darker shades and use diamonds.

Figure 8: Survey Evidence on Mechanisms of Network Effects

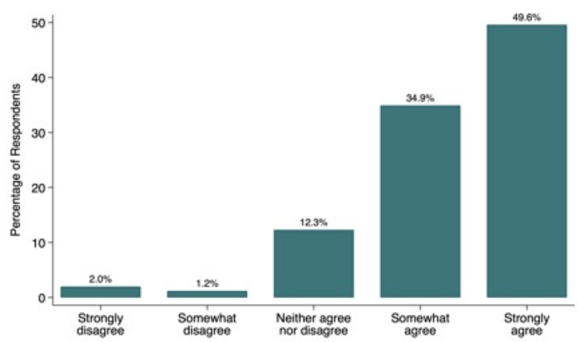
(a) Primary purpose for seeing friends



(b) Help received from friends in recent years



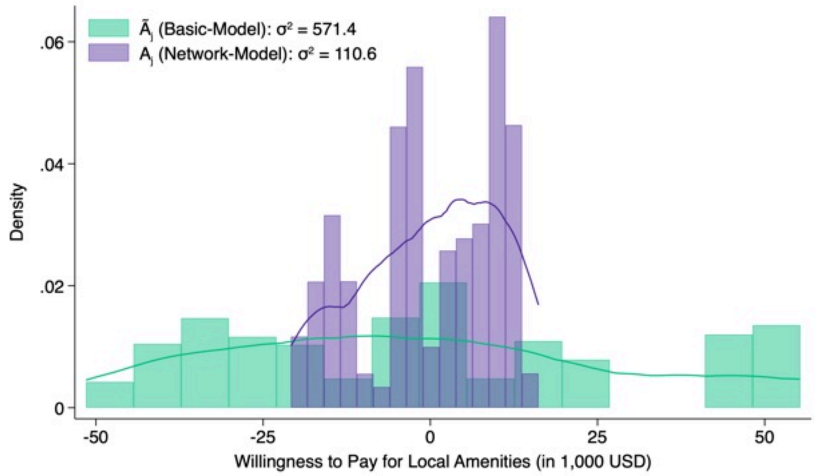
(c) Enjoy friends' company more than needing their help



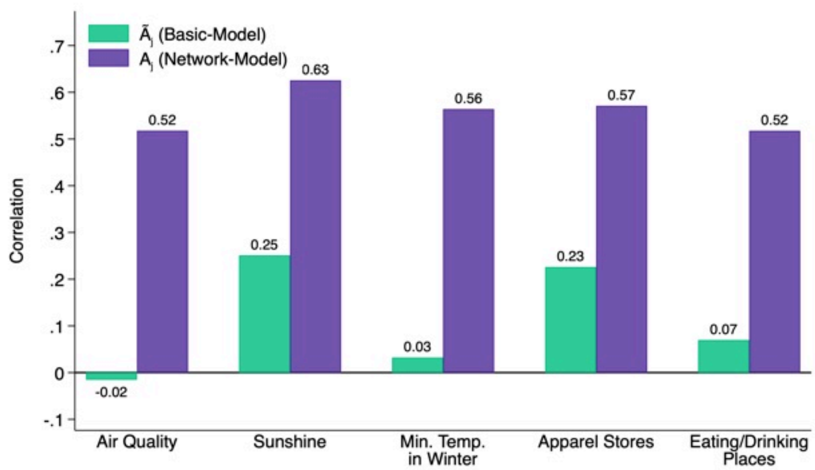
Notes: Figure presents evidence from survey regarding the role played by social networks for residential choice. For more details regarding this sample, see Section 2. For the construction of this figure, we limit attention to individuals who currently live in the same city as they grew up in while including individuals who may have lived elsewhere during some point of their life. Panel (a) shows the frequency with which respondents see the people that are most influential for their decision of where to live for reasons related to leisure (in turquoise), work (in yellow) and other forms of help (in gray). Panel (b) shows distribution of the extent to which respondents recently received help from their local friends and family with respect to one's job (in yellow), child care (in light blue) and housing (in dark red). For the values corresponding to child care, we subset to individuals with children. Panel (c) shows the distribution of respondents agreeing with the statement that they enjoy the company of their local social networks much more than that they need their help.

Figure 9: Local Amenities: A_j, $\tilde{A}_j$

(a) Distribution of Willingness to Pay For A_j and $\tilde{A}_j$

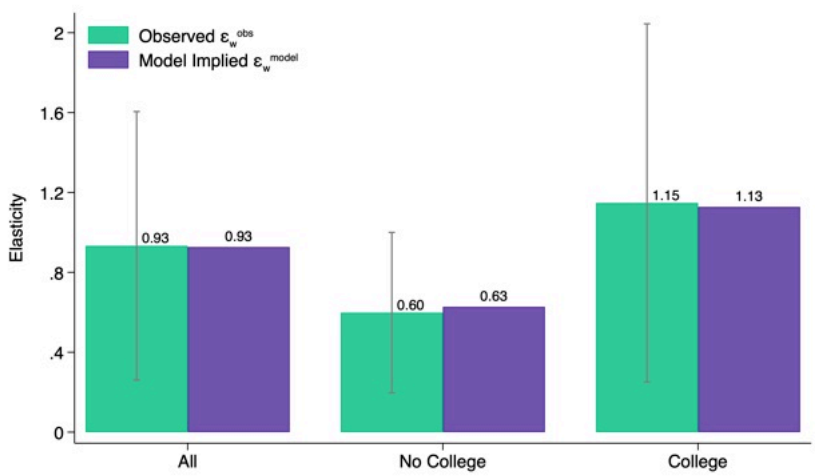


(b) Correlation: A_j, $\tilde{A}_j$ vs. Place Characteristics



Notes: Figure presents model results regarding the role of local amenities in the network model (A_j) and in the basic model ($\tilde{A}_j$). For details regarding the estimation of A_j and $\tilde{A}_j$, see Section 6.2 and Appendix Section E. Panel (a) shows the distribution of the willingness to pay for A_j and $\tilde{A}_j$ in 1,000 USD. To scale A_j and $\tilde{A}_j$ in terms of willingness to pay from the utility framework described in equation 13, we multiply the ratio of A_j (and $\tilde{A}_j$) and α_w with average local wages. The lines in Panel (a) present the kernel density plots while the bars present histograms. Panel (b) shows correlations between A_j and $\tilde{A}_j$ and various measures of place characteristics. Air Quality corresponds to the negative of PM2.5 (24-hr, $\mu g/m3$), sunshine corresponds to average daily sunlight in kjm (1979-2011), minimum temperature in winter is the average daily minimum temperate in February (1979-2011), and apparel stores as well as eating/drinking places are the numbers of establishments per capita with NAICS codes: 448, 7224, and 7225.

Figure 10: Differences in Wage Elasticities Between More- and Less-Educated Individuals



Notes: Figure shows empirical estimates as well as model predictions of the migration elasticity with respect to wages for more- and less-skilled individuals. More-skilled are individuals who report a college on their profile, while less-skilled individuals are those who do not report a college. For the empirical estimates, we use a shift-share approach predicting local wage growth between 2012-2019 based on 6-digit NAICS industry composition. Analysis focuses on individuals in Expanded Sample included in the model analysis. Vertical lines denote 95% confidence intervals. The model predictions are based on elasticity estimates obtained from simulating a range of local productivity shocks. We simulate shocks of various different sizes (-15% to 15% of productivity) shocking a single CZ a time. We then find the implied equilibrium wages, rents, and populations construct an estimate of the elasticity.

Tables

Table 1: Summary Statistics: Expanded Sample and College Graduates

Panel (a): Expanded Sample

	Mean	SD	P10	P25	P50	P75	P90	P99
Inter-CZ Moving Rate 2012-2019	0.32	0.46	0	0	0	1	1	1
Number of Friends	206.41	153.16	75	102	161	260	393	766
Number of CZs w/ 1+ Friend	23.42	15.84	8	13	20	30	43	77
Lists College	0.62	0.48	0	0	1	1	1	1
Is Female	0.61	0.49	0	0	1	1	1	1

Panel (b): College Graduation Sample

	Mean	SD	P10	P25	P50	P75	P90	P99
Moved After Graduation	0.31	0.46	0	0	0	1	1	1
Number of Older Friends	112.52	105.24	21	40	81	151	243	504
Share of Older Friends in College CZ (for those in top 50 CZs)	0.48	0.29	0.09	0.2	0.52	0.74	0.83	0.94
Number of Top 50 CZs w/ 1+ Friend	7.64	5.55	2	4	6	10	15	26
Lists College	1.00	0.00	1	1	1	1	1	1
Is Female	0.58	0.49	0	0	1	1	1	1

Note: Table shows summary statistics about the demographics, networks, and moving behavior of individuals in our two samples. Panel (a) captures individuals in our Expanded Sample, which is described in Section 2.1.1. Panel (b) captures individuals in our College Graduation Sample, which is described in Section 2.1.2.

Table 2: Effect of Friends' Moves on Likelihood Living in CZ

	In CZ After Grad.			
	(1)	(2)	(3)	(4)
Number Friends in Dest. CZ	0.0033***	0.0032***	0.0029***	0.0023***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Number of Observations	64,596,900	62,606,550	28,041,250	61,254,150
Number of College Graduates	1,291,938	1,252,131	560,825	1,225,083
Dependent variable mean	0.012	0.012	0.013	0.012
R^2	0.677	0.699	0.821	0.721
Cohort X Coll-CZ X HS-CZ X Dest-CZ FEs	✓			
... X Gender FEs		✓		
... X Income FEs			✓	
... X College FEs				✓

Notes: Table presents estimates of equation 9 with various set of fixed effects. The main coefficient of interest corresponding to β, or the effect of having one more friend in a given CZ on one's likelihood to live there after graduation is shown in the first row. Each observation is a individual-by-CZ combination and the analysis is conducted for the largest 50 CZs in the country only. In the first row, we include fixed effects for one's college cohort fully interacted with one's college CZ, one's childhood CZ as well as the potential destination CZ. In the second and third column, we further interact said fixed effects with gender fixed effects and the parental income (in deciles), respectively. In the last column, we control for the exact college an individual attended to. All columns also include controls for the number of friends in a given CZ as of the first year in college and the relationship between that control and the outcome is allowed to vary by potential destination CZ. Standard errors are clustered at the individual-level and are shown in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table 3: Effect of Friends' Moves on Likelihood Living in CZ - Expanded Sample

	In CZ At Time of Measuring Location			
	(1)	(2)	(3)	(4)
Number Friends in Dest. CZ	0.0033***	0.0034***	0.0030***	0.0029***
	(0.0001)	(0.0001)	(0.0003)	(0.0002)
Number of Observations	28,517,150	27,469,200	13,707,900	18,839,600
Number of Individuals	198,299	198,299	198,299	198,299
Dependent variable mean	0.012	0.012	0.015	0.012
R^2	0.718	0.746	0.887	0.852
Year X Cohort X Initial-CZ X Dest-CZ FEs	✓			
... X Gender FEs		✓		
... X Income FEs			✓	
... X College FEs				✓

Notes: Table presents coefficient estimates of equation 9 adjusted for the Expanded Sample as discussed in Section 5.5. Each column presents estimates for a different set of fixed effects. The main coefficient of interest corresponding to β, or the effect of having one more friend in a given CZ on one's likelihood to live there at the time of measuring an individual's location (t^*) is shown in the first row. Each observation is a individual-by-CZ combination and the analysis is conducted for the largest 50 CZs in the country only. In the first row, we include fixed effects for one's birth cohort fully interacted with the time at which we measure their location, the CZ they lived in initially (at t^{early}), as well as the potential destination CZ. In the second and third column, we further interact said fixed effects with gender fixed effects and the parental income (in deciles), respectively. In the last column, we control for the exact college an individual attended to. Individuals without a college are grouped in one composite group for this exercise. All columns also include controls for the number of friends in a given CZ as of the first point in time we measure location (t^{early}) and the relationship between that control and the outcome is allowed to vary by potential destination CZ. Standard errors are clustered at the individual-by-year-level and are shown in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table 4: Effect of Friends' Moves on Likelihood Living in CZ - Logit Estimates

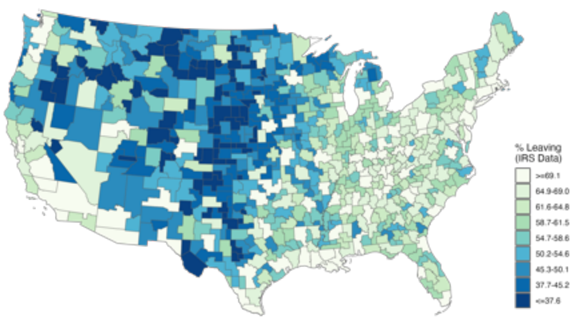
	In CZ After Grad.	
	(1)	(2)
Prop Friends: $\tilde{b} = \tilde{\beta} + \tilde{\delta}$	5.0932***	
	(0.0293)	
Prop Friends: $\tilde{\beta}$	3.2171***	
	(0.0400)	
Prop Friends: $\tilde{\delta}$	1.8761***	
	(0.0355)	
N Friends: $\tilde{b} = \tilde{\beta} + \tilde{\delta}$		0.0439***
		(0.0003)
N Friends: $\tilde{\beta}$		0.0389***
		(0.0006)
N Friends: $\tilde{\delta}$		0.0050***
		(0.0005)
Number of Observations	44,356,900	44,356,900
Number of College Graduates	887,138	887,138
Pseudo-R^2	0.734	0.733

Notes: Table presents estimates of equation 15. Each observation is a individual-by-CZ combination and the analysis is conducted for the largest 50 CZs in the country only. We restrict the sample to individuals who live in one of those CZs at the point in time at which we measure post-graduation location. Column 1 presents coefficient estimates for the proportion-based definition of social networks. Column 2 shows estimates for the number-based definition of social networks. In both cases, we control for the size of one's initial network in given CZ, the distance and distances squared between an individual's college CZ and the potential destination CZ as well as distance between childhood CZ and the potential destination CZ. The regressions also include indicators for whether the individual attended college in the given CZ, and one for whether the individual grew up in said CZ. Standard errors are shown in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

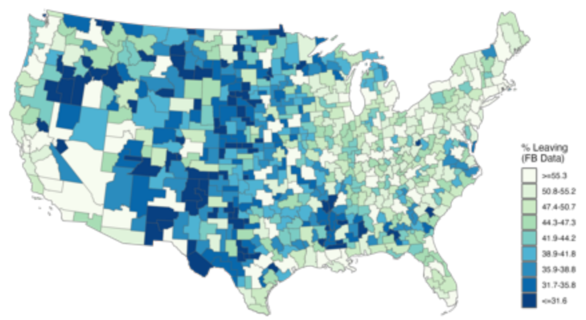
A Appendix Figures

Figure A1: Migration Validation Tests

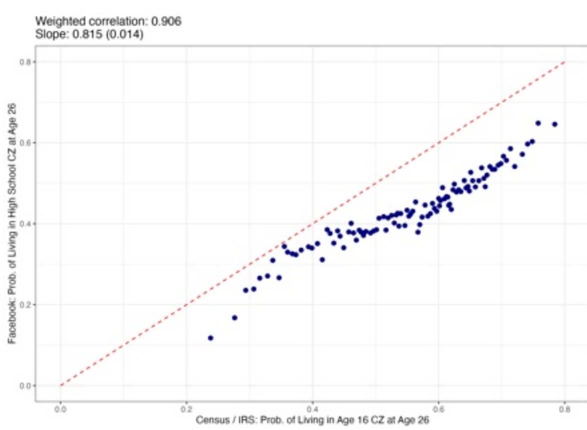
(a) Probability of Leaving Childhood CZ (IRS)



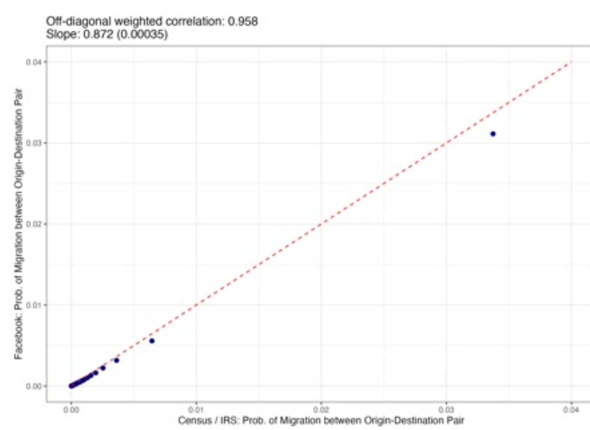
(b) Probability of Leaving Childhood CZ (FB)



(c) Probability of Living in Childhood CZ



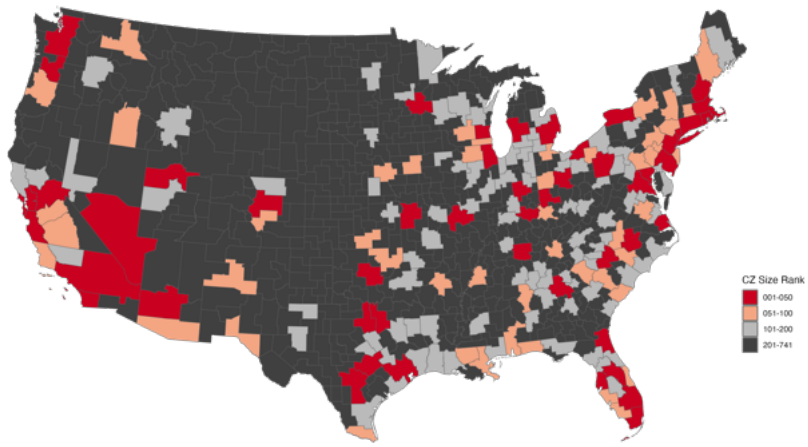
(d) Probability of Moving to a Specific CZ



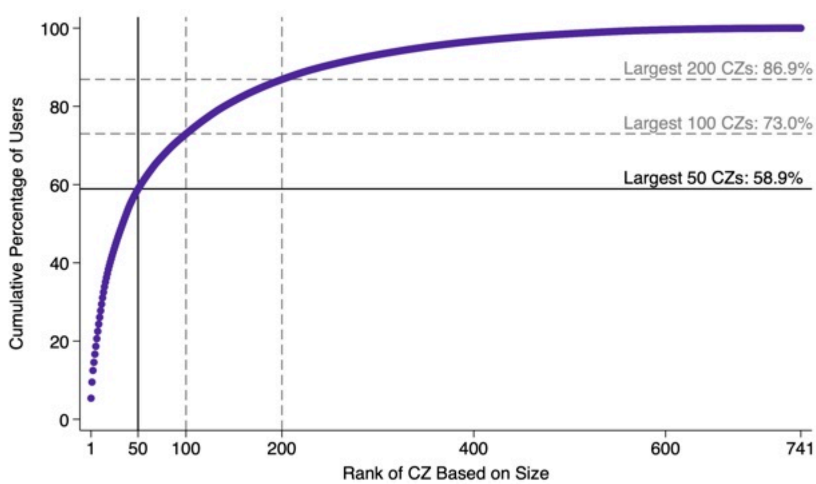
Notes: Figure presents validation tests of the migration patterns observed in the Facebook data. Panel A plots the share of individuals who lived in a given commuting zone at age 16 who no longer live in that commuting zone at age 26. The data is drawn from Sprung-Keyser et al. (2022), who observe the location of individuals born between 1984 and 1992 using linked data from the Census, American Community Survey, and Internal Revenue Service. In Panel B, we produce a similar plot, which captures the share Facebook users in our sample who were born in 1991 and who live in a commuting zone different than the one they attended high school in October 2017. We group users according to the commuting zone of their high school. In Panel C, we present a binned scatter plot of these two quantities, showing that the two series are highly correlated, though we find somewhat higher rates of out-migration. In Panel D, we present a binned scatter plot of the share of individuals who move between each pair of origins in the two data sets. We exclude individuals who live in the same commuting zone in which they attended high school. The correlations and slopes we report in Panels C and D are weighted using the number of individuals in each origin in the data from Sprung-Keyser et al. (2022).

Figure A2: CZ Size

(a) Map of CZ Size in Ranks



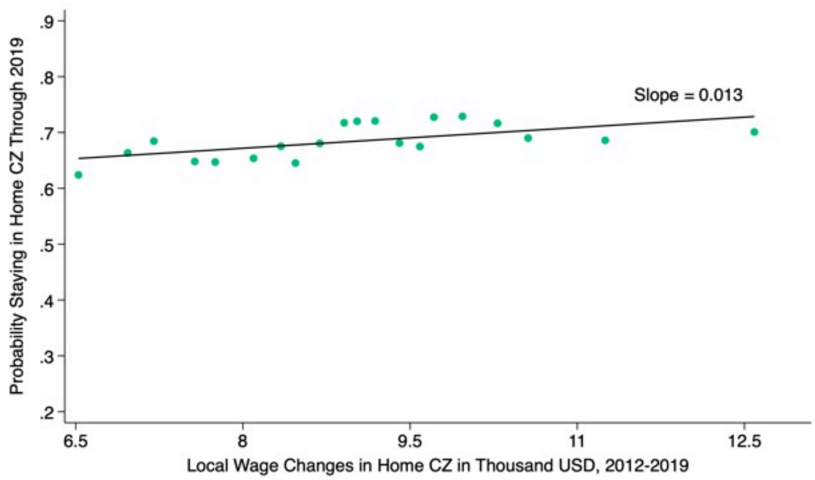
(b) Cumulative Percentage of Users



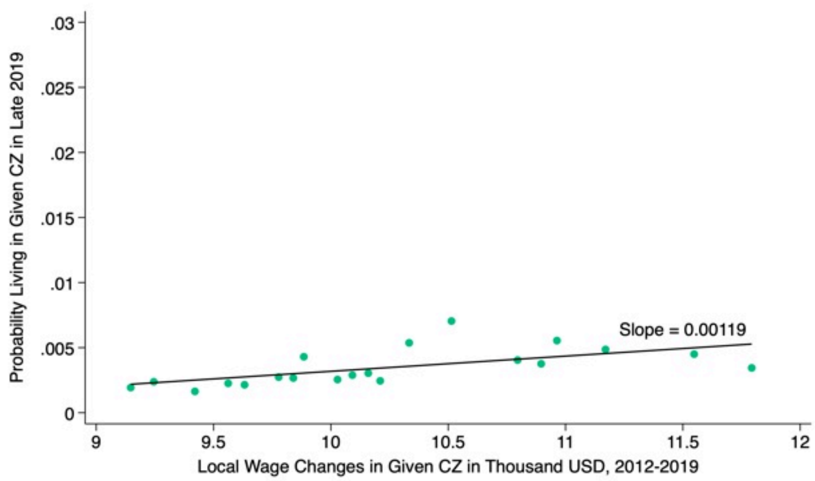
Notes: Panel A presents a map of commuting zones by size following the 1990 definition of CZs (Autor and Dorn, 2013). CZs are ranked based on 2012 Facebook user counts. Red CZs belong to the 50 largest CZs in the country. Salmon-colored CZs rank between the 51th to the 100th largest CZs while light gray CZs are those ranked between the 101th and 200th largest CZs. Dark gray colors indicate all other CZs. Panel B presents a cumulative distribution function of users by CZ size.

Figure A3: Residential Choice and Changes in Local Wages

(a) Δ Wages in Home CZ



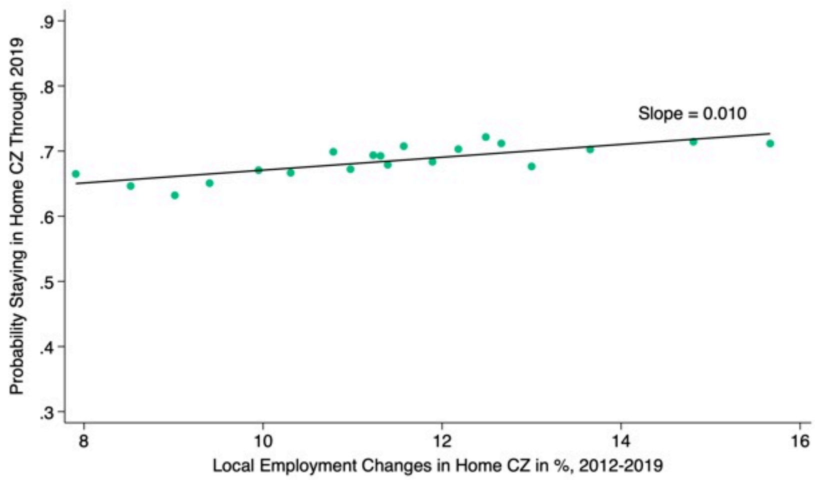
(b) Δ Wages in Alternative CZ



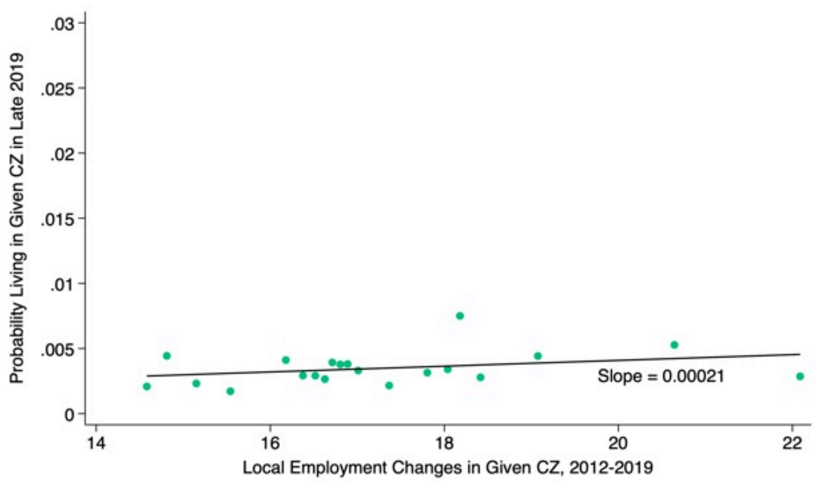
Notes: Figure presents evidence on the relationship between local wage growth and residential choice. We instrument for CZ-level wage growth between 2012-2019 using an industry-based shift-share approach discussed in more detail in Section 3. Panel (a) presents a binned scatter plot of the relationship between local wage growth in the 2012 home CZ—on the horizontal axis—and the likelihood to stay in the home CZ through December 2019 on the vertical axis. The regression underlying the plot controls for local wages (in USD) in 2012 in the home CZ. Panel (b) presents a binned scatter plot of the relationship between local wage growth in a given alternative CZ—on the horizontal axis—and the likelihood to move to said alternative CZ by December 2019 on the vertical axis. 2012 home CZs are excluded from the analysis and we restrict the set of potential alternative CZs to the 50 largest CZs. The regression underlying the plot controls for local wage growth in the home CZ. We present best fit lines in both panels together with standard errors in parentheses.

Figure A4: Residential Choice and Changes in Local Employment

(a) Δ Employment in Home CZ



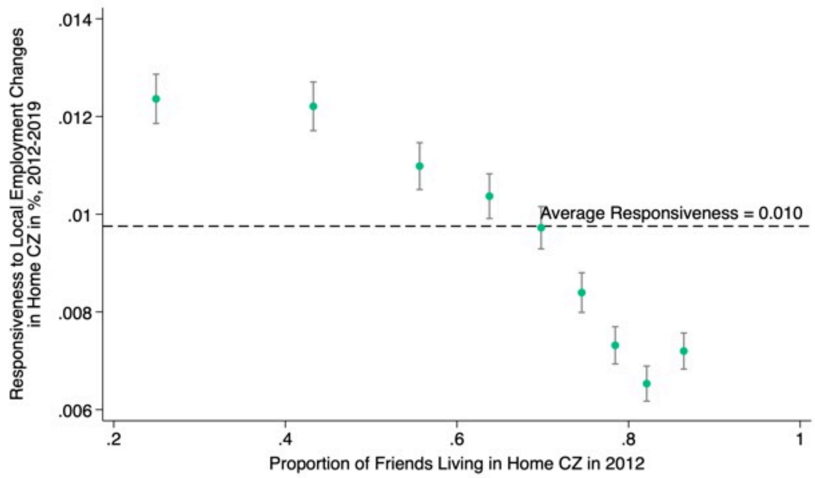
(b) Δ Employment in Alternative CZ



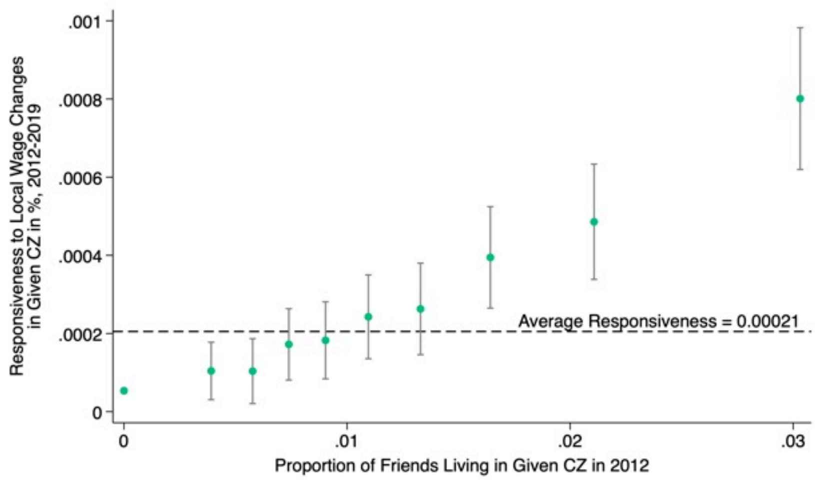
Notes: Figure presents evidence on the relationship between local employment growth and residential choice. We instrument for CZ-level employment growth between 2012-2019 in % using an industry-based shift-share approach discussed in more detail in Section 3. Panel (a) presents a binned scatter plot of the relationship between local employment growth in the home CZ—on the horizontal axis—and the likelihood to stay in the 2012 home CZ through December 2019 on the vertical axis. The regression underlying the plot controls for local employment in 2012 in the home CZ. Panel (b) presents a binned scatter plot of the relationship between local employment growth in a given alternative CZ—on the horizontal axis—and the likelihood to move to said alternative CZ by December 2019 on the vertical axis. 2012 home CZs are excluded from the analysis and we restrict the set of potential alternative CZs to the 50 largest CZs. The regression underlying the plot controls for local employment growth in the home CZ. We present best fit lines in both panels together with standard errors in parentheses.

Figure A5: Heterogeneity in Responsiveness to Changes in Local Employment

(a) Δ Employment in Home CZ



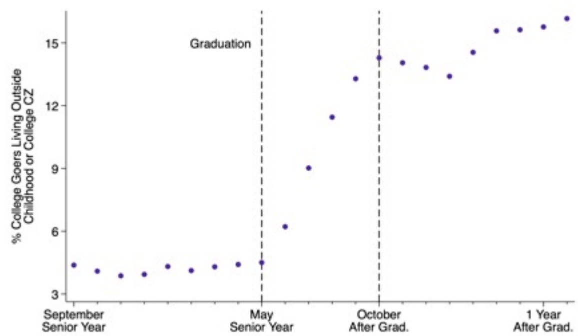
(b) Δ Employment in Alternative CZ



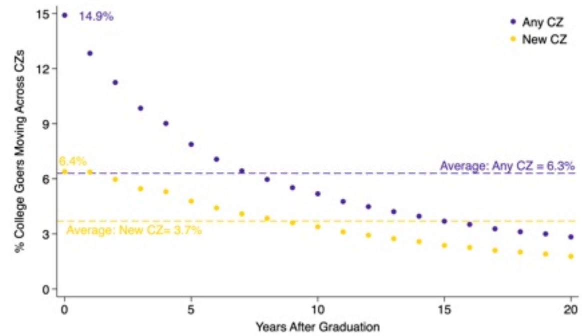
Notes: Figure shows heterogeneity in an individual's responsiveness to recent local employment growth by level of social connectedness. Figure is analogous to Figure 3 for wage growth. For more details regarding the construction of our measures of local growth, see Section 3. Panel (a) presents coefficient estimates from a regression of whether a given remains in their 2012 home CZ through December 2019 on recent local employment in the 2012 CZ, separately for individuals with different levels of proportions of their friends in the 2012 CZ as of early 2012. A binned scatter plot of the relationship between staying in the 2012 CZ and employment growth for the full sample is shown in Panel (a) of Appendix Figure A4. Panel (b) presents coefficient estimates from a regression of whether, as of December 2019, a given individual lives in a given CZ (excluding their 2012 CZ) on recent local employment growth in that CZ, separately for individuals with different levels of proportions of their friends in said CZ as of early 2012. A binned scatter plot of the relationship between moving to a given and employment growth for the full sample is shown in Panel (b) of Appendix Figure A4. The dashed lines in both Panels correspond to the average responsiveness, also shown in Appendix Figure A4. In both Panels we omit the coefficient estimate corresponding to individuals in the bottom decile of the x-axis measure for expositional reasons.

Figure A6: Residential Choices Among College Graduates

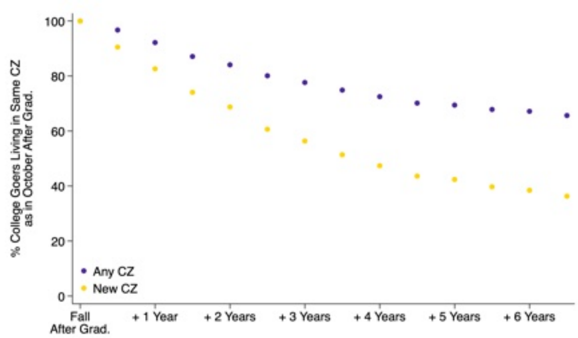
(a) Prob. Living in CZ Other Than Home CZ



(b) Prob. Moving Across CZs



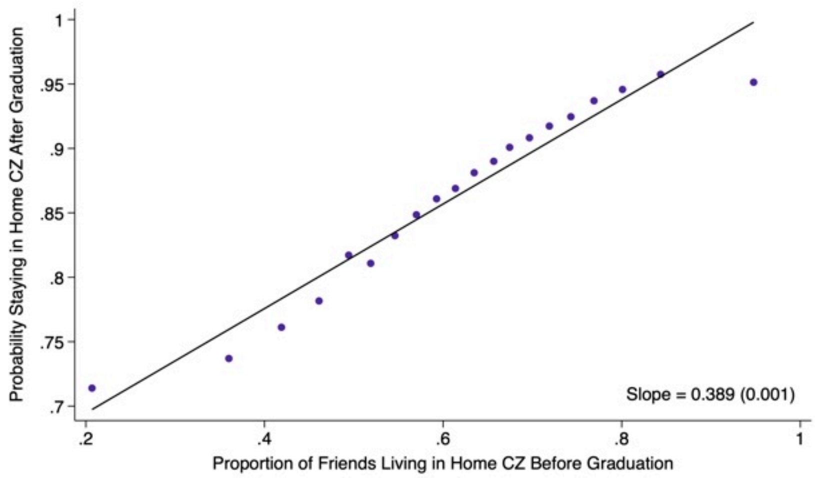
(c) Prob. Living in Post-Graduation CZ



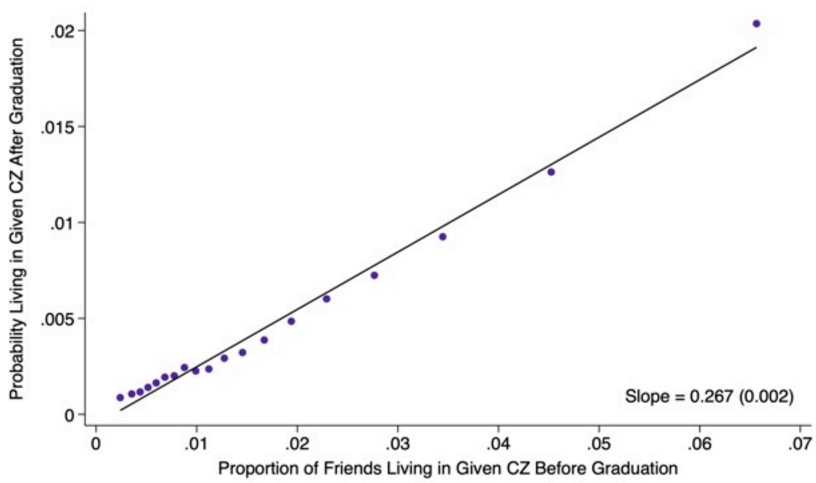
Notes: Figure presents descriptive statistics about the residential choices made by college graduates in our sample. For more details on the construction of college graduates, see Section 2. Panel (a) presents a time-series plot showing the likelihood of living in the home CZ among the sample of college graduates over time. The series is at a monthly frequency and expands from September of an individual's senior year through June of the year after graduation. The first dashed line in May of senior year denotes the point in time at which college goers are likely to graduate from college. The second dashed line highlights the following October, i.e., five months after graduation which corresponds to the point in time at which we measure an individual's location after graduation in subsequent analyses. Panel (b) shows series likelihood of indiviuals moving to a new CZ by number of years after graduation. To construct this figure, we extend our sample of college graduates to include all indiviuals listing a college and who are assumed to graduate between 1997 and 2018. For those indiviuals, we then show the likelihood that they moved between October 2017 and October 2018. The "0" on the horizontal axis thus corresponds to indiviuals graduating in 2018. The purple series includes any type of move while the yellow series displays moving rates for CZs that are neither one's college CZ nor the CZ one grew up in. Panel (c) presents a time series of the likelihood to still live in the CZ one was observed in in October after graduation. The purple series includes all locations while the yellow series subsets to individuals who, after graduation, moved to a CZ that was neither their college CZ nor the CZ one grew up in.

Figure A7: Residential Choice and Location of Social Networks Among College Graduates

(a) Probability of Staying in Home CZ



(b) Probability of Living in Specific Alternative CZ



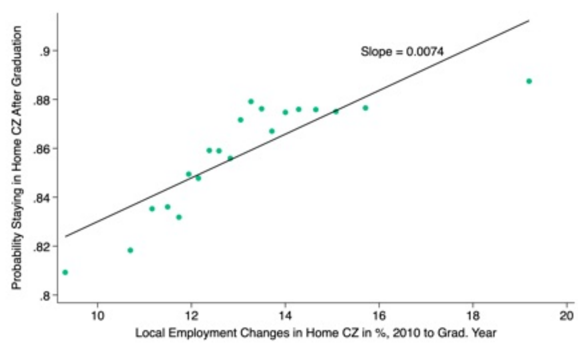
Notes: Figure presents descriptive statistics on the relationship between residential choice and the location of one's social network for the sample of college graduates. Figure is analogous to Figure 2. An individual's location is measured in October the year of graduation. Panel (a) focuses on the extent to which indiviuals stay in their home CZ—i.e., the CZ of the college CZ or the CZ they grew up in—while Panel (b) presents statistics for alternative CZs excluding the home CZ. The locations of an individual's friends are measured in the summer before senior year. We only include friends who are at least one cohort older than the individuals. See Section 4 for additional details.

Figure A8: Residential Choice and Changes in Local Economic Conditions Among College Graduates

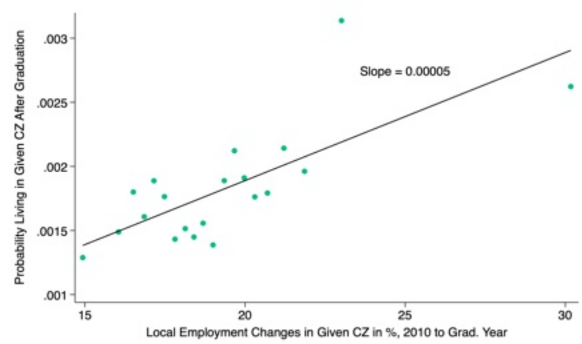
(a) Δ Pay in Home CZ

(b) Δ Pay in Alternative CZ

(c) Δ Employment in Home CZ



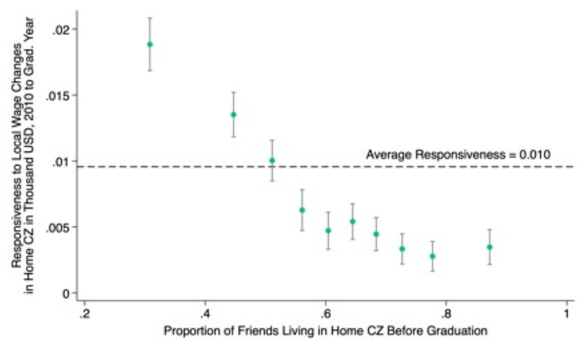
(d) Δ Employment in Alternative CZ



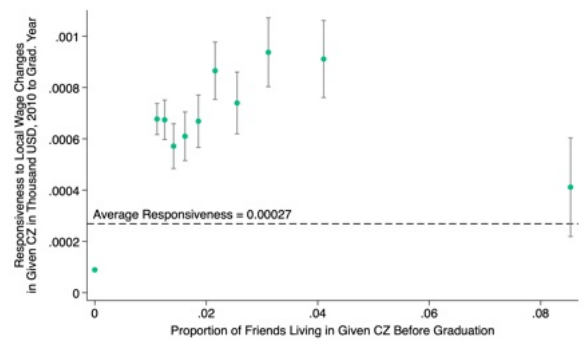
Notes: Figure presents statistics analogous to those shown in Appendix Figures A3 and A4 for the sample of college graduates. Panel (a) and (b) focus on local wage growth in the home CZ and in alternative CZs, respectively. Panels (c) and (d) present results for local employment growth. Values corresponding to the home CZ are averages over an individual's college CZ and the CZ they grew up in. All measures of growth correspond to growth between 2010 and the year in which a indiviual graduates from college.

Figure A9: Heterogeneity in Responsiveness to Economic Conditions by Size of Social Network Among College Graduates

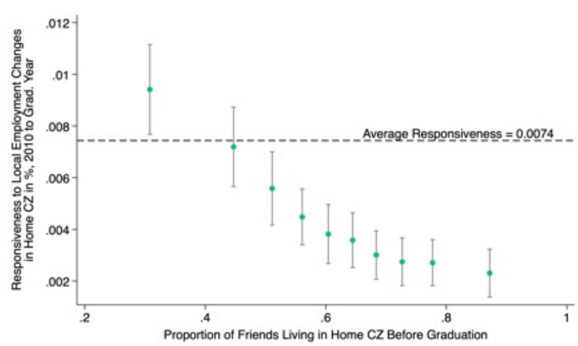
(a) Δ Pay in Home CZ



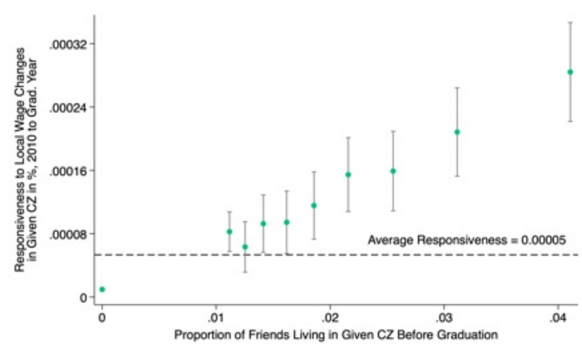
(b) Δ Pay in Alternative CZ



(c) Δ Employment in Home CZ

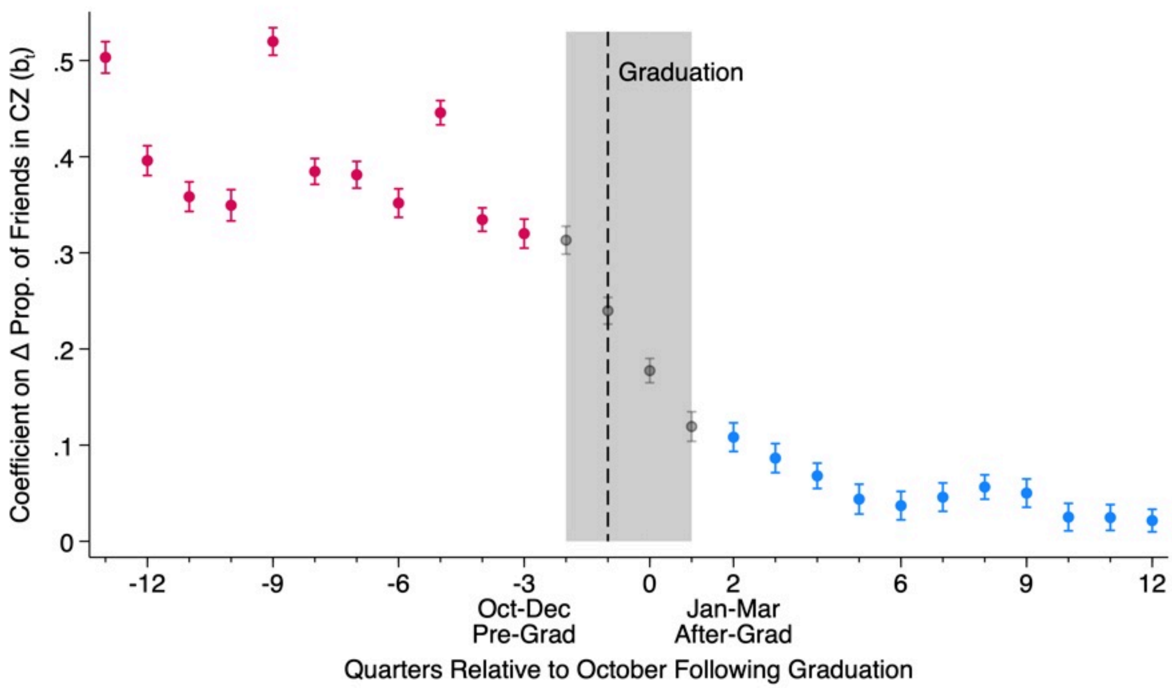


(d) Δ Employment in Alternative CZ



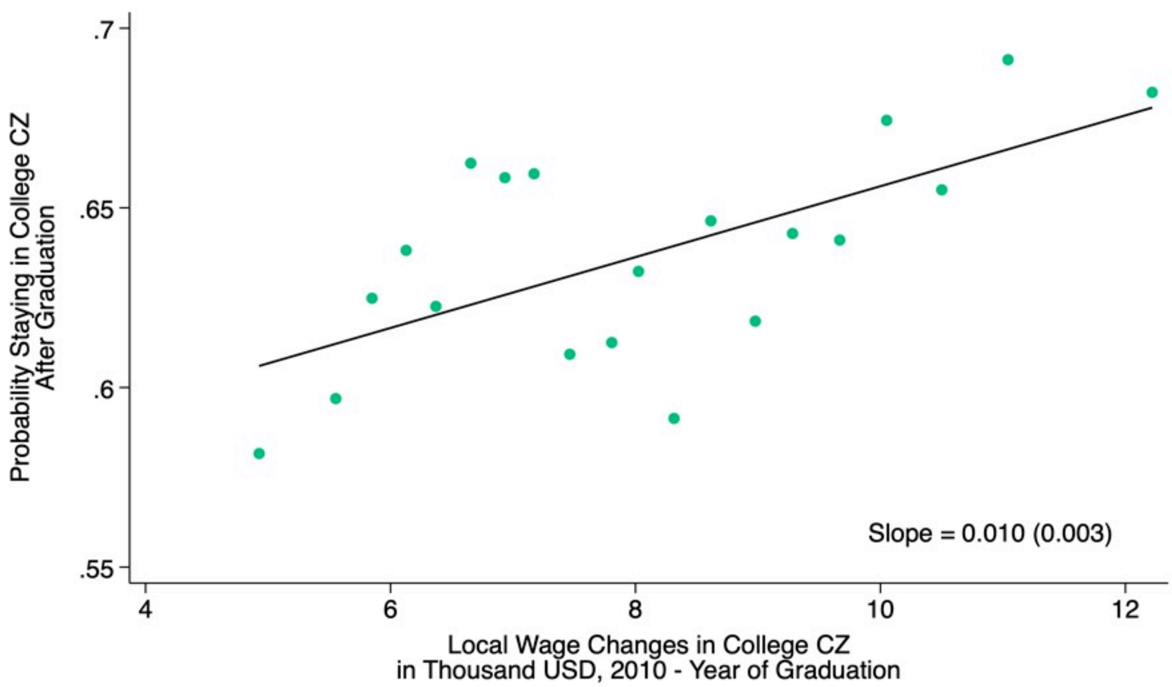
Notes: Figure presents statistics analogous to those shown in Figure 3 and Appendix Figure A5 for the sample of college graduates.

Figure A10: Effect of Friends' Moves on Residential Choice



Notes: Figure presents coefficient estimates of equation 4 while using the share-based definition of social networks. Figure is otherwise analogous to Figure 5.

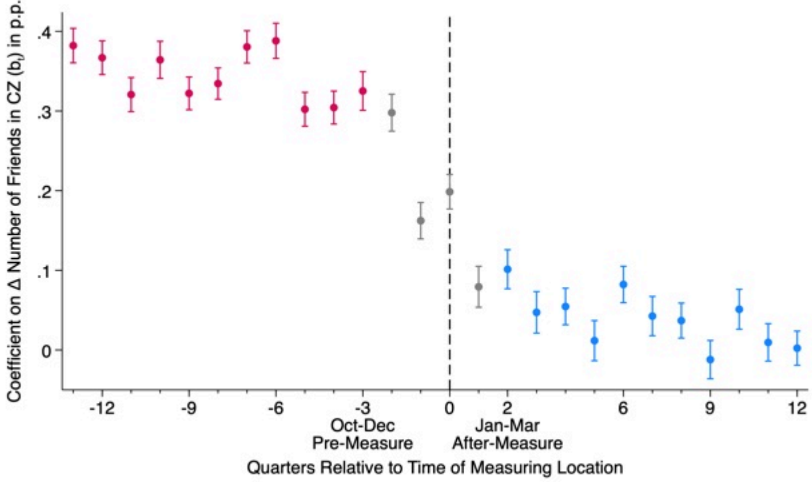
Figure A11: Effect of Local Wage Growth in College CZ on Out-Migration



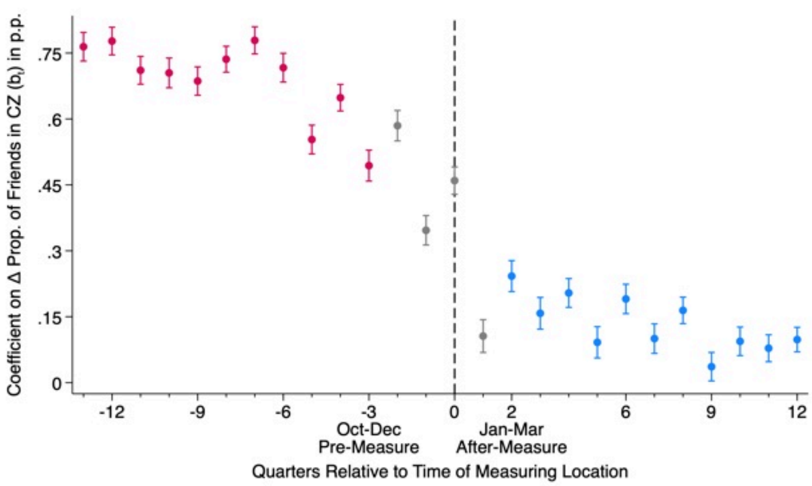
Notes: Figure presents binned scatter plot of the relationship between local wage growth in the CZ of the college and the extent to which graduates leave that CZ after graduation. To instrument for local wage growth in the college CZ we follow the approach described in Section 3. In the present case, we focus specifically on wage growth between 2010 and the year of graduation for the college CZ. We measure the location of college graduates in October following graduation. The analysis is most similar to the one shown in Panel (a) of Appendix Figure A8.

Figure A12: Effect of Friends' Moves on Residential Choice - Expanded Sample

(a) Number of Friends



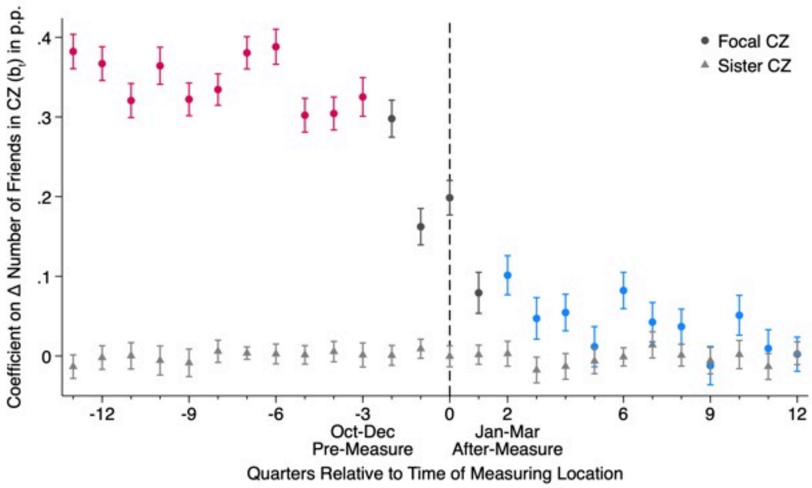
(b) Proportion of Friends



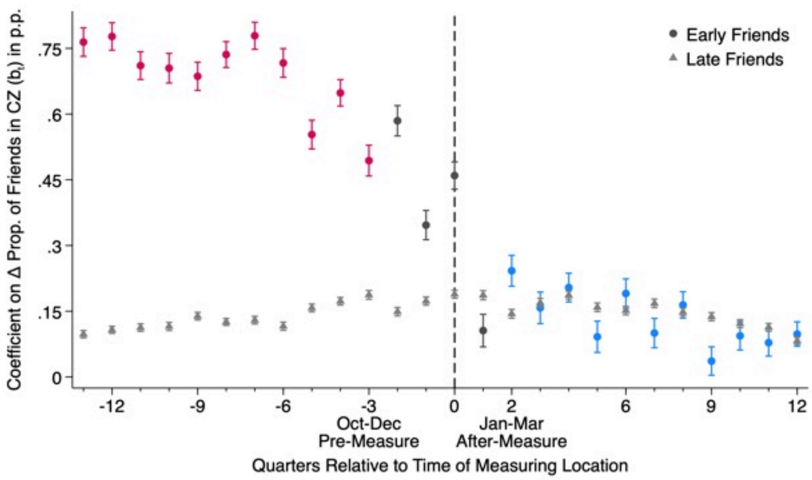
Notes: Figure presents coefficient estimates of equation 4 modified for the Expanded sample. Relative to the definitions in Section 5.1 and Figure 5, we re-define t^* as the second point in time at which we measure an individual's location—in this context, this is October of 2017, 2018 or 2019—and *early* corresponds to the first time at which we measure an individual's location. Our measure of social networks includes all friends made as of early 2014. The other terms are unchanged and the Figure is otherwise analogous to Figure 5 and Appendix Figure A10. Panel (a) uses the number-based definition of social networks while Panel (b) employs the proportion-based definition.

Figure A13: Evidence Supporting Identifying Assumption - Expanded Sample

(a) Sister CZs



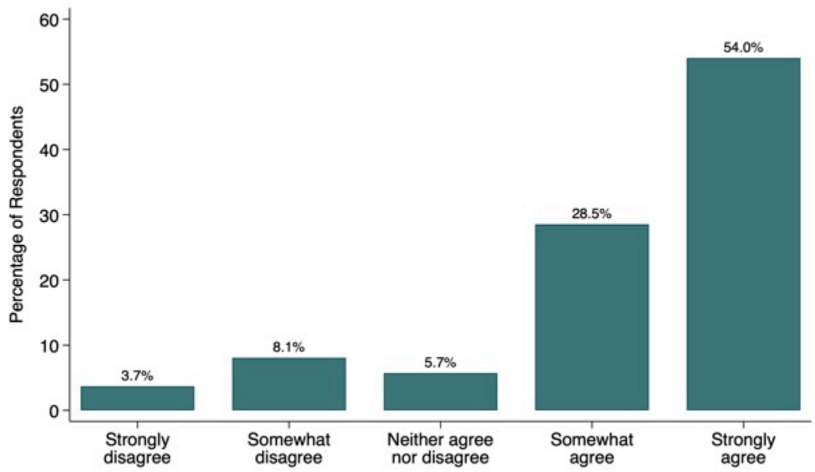
(b) Late Friends



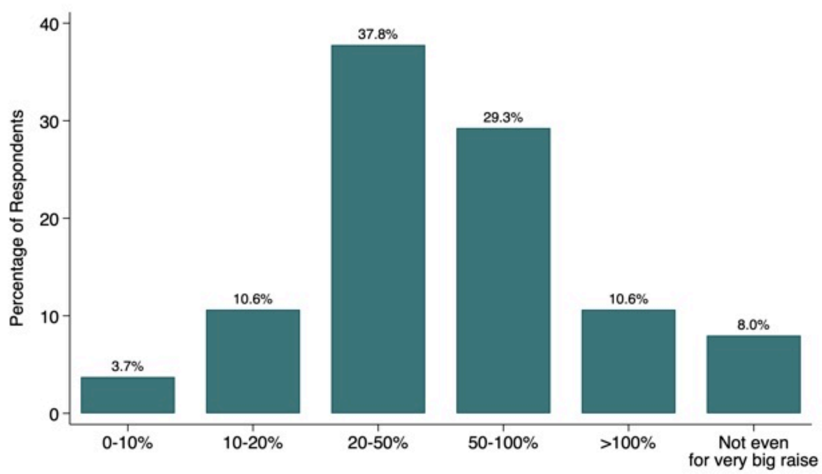
Notes: Figure presents evidence supporting the identifying assumption for the Expanded Sample and follows the types of analyses in Panels (b) and (c) of Figure 6. Panel (a) presents estimates of equation 7 adjusted for the Expanded Sample regarding the influence of friends' moves to focal and sister CZs on an individual's likelihood of living in the focal CZ. See Figure 6 and Section 5.2 for details regarding the construction of sister CZs. The gray triangles correspond to coefficient estimates for sister CZs, while the colored circles represent coefficient estimates for the focal CZ. Panel (b) presents estimates similar to those shown in Panel (c) of Figure 6 focusing on the effects of people who the individual eventually befriends but had not yet done so at the time of measuring location (t^*), or "late" friends. Late friends are those that an individual befriends in 2022 or after. The gray triangles correspond to coefficient estimates for friends an individual had made prior to t^* ("early" friends), while the colored circles represent coefficient estimates for the late friends. Panel (b) uses the proportion-based definition of social networks to account for large differences in the average number of earlier and later friends. The layout follows that of Figure A12. Vertical lines in all panels indicate 95% confidence intervals.

Figure A14: Survey Evidence on Importance of Friends for Residential Choice

(a) Friends + Family one of Main Reasons to Live in City

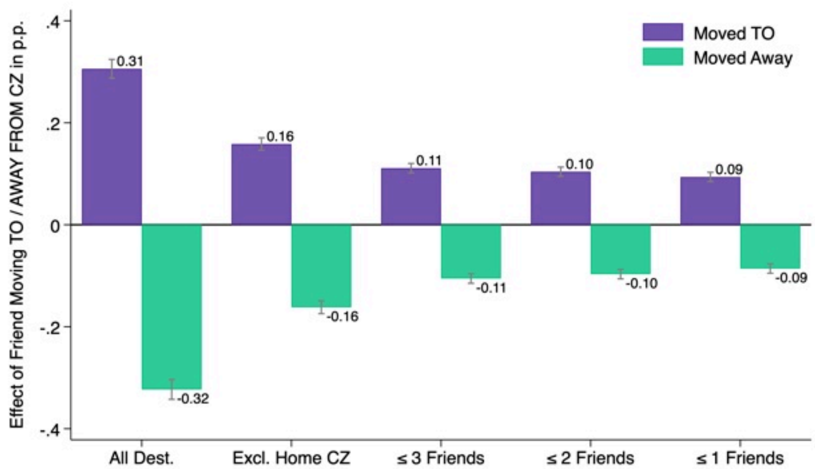


(b) Pay Raise to Move to City without Friends



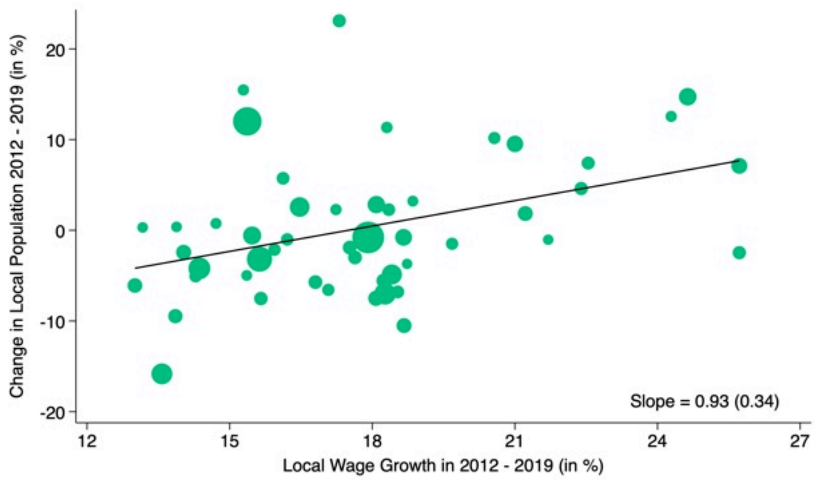
Notes: Figure presents survey evidence on the importance of social networks for residential choice. For more details regarding this sample, see Section 2. For the construction of this figure, we limit attention to individuals who currently live in the same city as they grew up in while including individuals who may have lived elsewhere during some point of their life. Panel (a) shows the distribution of respondents agreeing with the statement that friends and family are one of the main reasons for them to live in the current city. Panel (b) shows the distribution of pay raises respondents say they would have to receive in order to consider moving to a city without any prior connections. The sample for Panels (b) is restricted to respondents who have never lived in a city other than the one they grew up in due to the design of the survey.

Figure A15: Network Effects for Friends Moving To and Away From CZ By Initial Connectedness



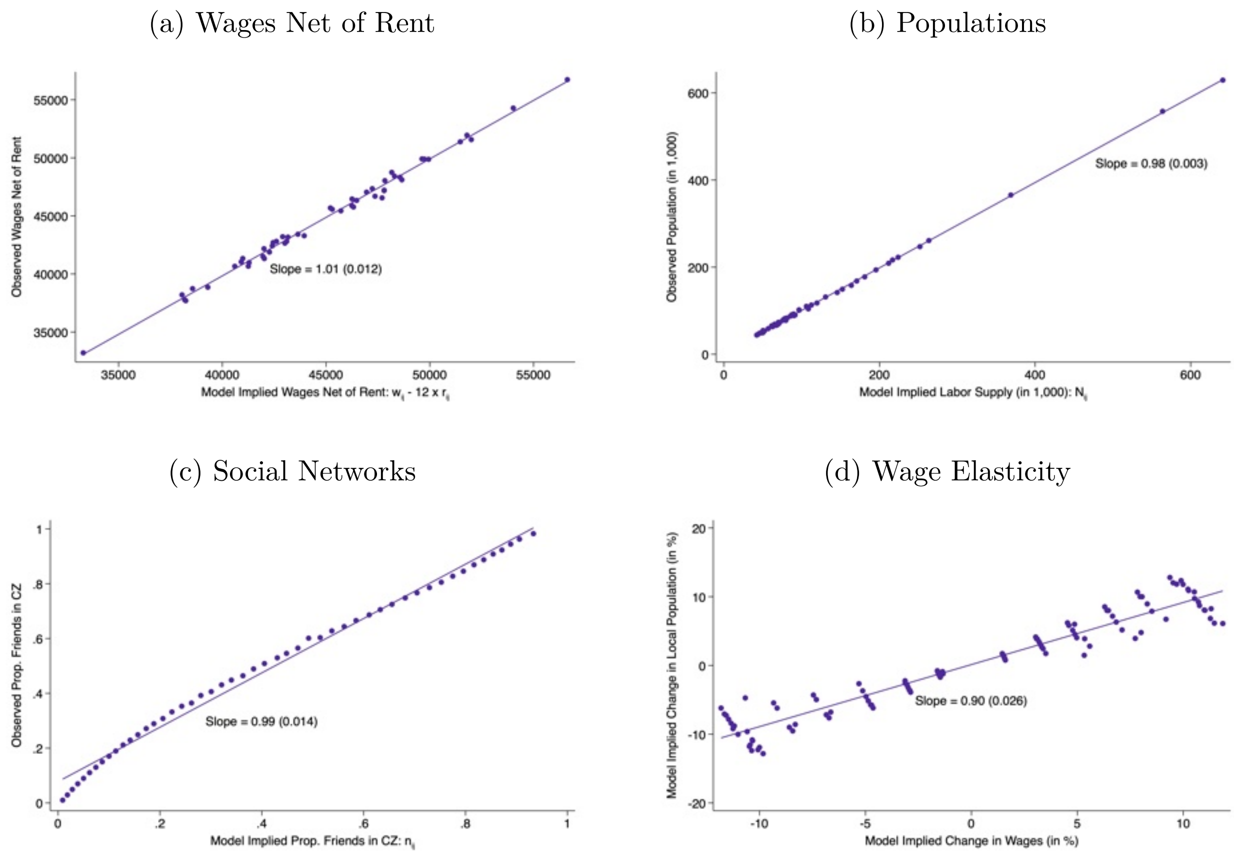
Notes: Figure presents estimates of effects of friends moving to and away from a given CZ based on equation 10 separately for CZs with different baseline levels of connectedness. The results are obtained from estimating parsimonious versions of equation 10, in the same way that equation 9 presented a parsimonious version of equation 4. Purple bars correspond to the effect of friends moving to a given CZ while green bars correspond to the effect of friends moving away. Vertical lines show 95% confidence intervals. The first set of bars presents estimates for all 50 largest CZs. The second set of bars excludes an individual's home CZ from the choice set. The third, fourth and fifth sets of bars focus on CZs where an individual has at most three, two or one initial connection, as of the end of their first year in college.

Figure A16: Observed Wage Elasticity for Model Sample



Notes: Figure shows estimate of the migration elasticity with respect to wages among the sample used in Section 6. For the construction of this Figure, we follow the approach discussed in Section 3 to instrument for local wage growth between 2012-2019 using a shift-share approach based on local industry compositions. The horizontal axis shows local wage growth in percent. The vertical axis shows the change in the local population between 2012 and 2019 for the 50 largest CZs. Each dot is one CZ where the size of the dot is proportional to the CZ population. The best fit line and the slope are weighted by population. For more details, see Section 6.2.2.

Figure A17: Model Calibration



Notes: Figure presents statistics on the extent to which the model matches the data. For more details regarding the calibration of the model, see Section 6.2 as well as Appendix Section E. In the first three Panels, the horizontal axis corresponds to the model implied moments and the vertical axis shows the observed moments. Panel (a) displays the relationship between observed wages net of rent and the respective outcome in the model. Panel (b) shows the relationship for local populations. Each dot in Panels (a) and (b) is a CZ. Panel (c) shows a binned scatter plot of the congruence between the model implied distribution of social ties and those observed in the data. Panel (d) presents an estimate of the model implied elasticity. To construct Panel (d) we simulate numerous productivity shocks of varying sizes and in different CZs. We then find the new equilibrium wages and populations. Each dot in this Figure corresponds to one simulated shock and the resulting equilibrium. The horizontal axis shows the model implied change in wages in % and the vertical axis shows the model implied change in populations. The best fit line corresponds to the model-implied elasticity.

Appendix Tables

Table A1: Robustness: Social Networks and Residential Choice, Proportions of Friends

	In CZ After Grad.			
	(1)	(2)	(3)	(4)
Prop. Friends in Dest. CZ	0.3099***	0.3028***	0.2720***	0.2267***
	(0.0046)	(0.0048)	(0.0088)	(0.0048)
Number of Observations	64,596,900	62,606,550	28,041,250	61,254,150
Number of College Graduates	1,291,938	1,252,131	560,825	1,225,083
Dependent variable mean	0.012	0.012	0.013	0.012
R^2	0.678	0.701	0.824	0.722
Cohort X Coll-CZ X HS-CZ X Dest-CZ FEs	✓			
... X Gender FEs		✓		
... X Income FEs			✓	
... X College FEs				✓

Notes: Table presents coefficient estimates of equation 9 while using the proportion-based definition of social networks. Table is otherwise analogous to Table 2.

Table A2: Robustness: Social Networks and Residential Choice, Clustering of SE's

	In CZ After Grad.			
	(1)	(2)	(3)	(4)
Number Friends in Dest. CZ	0.0033***	0.0033***	0.0033***	0.0033***
	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Number of Observations	64,596,900	64,596,900	64,596,900	64,596,900
Number of College Graduates	1,291,938	1,291,938	1,291,938	1,291,938
Dependent variable mean	0.012	0.012	0.012	0.012
R^2	0.677	0.677	0.677	0.677
Cohort X Coll-CZ X HS-CZ X Dest-CZ FEs	✓	✓	✓	✓
SE's Clustered at User Level	✓			
SE's Clustered at HS-CZ X CZ Level		✓		
SE's Clustered at Coll-CZ X CZ Level			✓	
SE's Clustered at HS-CZ X Coll-CZ X CZ Level				✓

Notes: Table presents coefficient estimates of equation 9 while presenting results for various different approaches to cluster standard errors. Estimation approach is otherwise unchanged relative to column 1 of Table 2. For reference, in column 1, we cluster standard errors at the individual-level, as in column 1 of Table 2. In column 2, we cluster standard errors at the childhood-CZ-by-destination-CZ level. In column 3, standard errors are clustered at the college-CZ-by-destination-CZ level. Column 4 presents the combination of the two, i.e., standard errors are clustered at the childhood-CZ-by-college-CZ-by-destination-CZ level.

Table A3: Robustness: Social Networks and Residential Choice, Clustering of SE's, Proportions of Friends

	In CZ After Grad.			
	(1)	(2)	(3)	(4)
Prop. Friends in Dest. CZ	0.3099***	0.3099***	0.3099***	0.3099***
	(0.0046)	(0.0099)	(0.0135)	(0.0095)
Number of Observations	64,596,900	64,596,900	64,596,900	64,596,900
Number of College Graduates	1,291,938	1,291,938	1,291,938	1,291,938
Dependent variable mean	0.012	0.012	0.012	0.012
R^2	0.678	0.678	0.678	0.678
Cohort X Coll-CZ X HS-CZ X Dest-CZ FEs	✓	✓	✓	✓
SE's Clustered at User Level	✓			
SE's Clustered at HS-CZ X CZ Level		✓		
SE's Clustered at Coll-CZ X CZ Level			✓	
SE's Clustered at HS-CZ X Coll-CZ X CZ Level				✓

Notes: Table is analogous to Table A2 while using the proportion-based definition of social networks.

Table A4: Robustness: Social Networks and Residential Choice, Expanded Choice Set

	In CZ After Grad.					
	(1)	(2)	(3)	(4)	(5)	(6)
Number Friends in Dest. CZ	0.0030***	0.0029***	0.0031***			
	(0.0002)	(0.0001)	(0.0001)			
Prop. Friends in Dest. CZ				0.2947***	0.2967***	0.3080***
				(0.0115)	(0.0104)	(0.0096)
Number of Observations	11,710,950	23,421,900	46,843,800	11,692,250	23,384,500	46,769,000
Number of College Graduates	234,219	234,219	234,219	233,845	233,845	233,845
Number of CZs in Choice Set	50	100	200	50	100	200
Dependent variable mean	0.012	0.008	0.005	0.012	0.008	0.005
R^2	0.739	0.740	0.736	0.741	0.742	0.738

Notes: Table presents coefficient estimates of equation 9 while varying the size of the choice set, i.e., the number of CZs included in the analogous. The analysis is otherwise identical to the analyses presented in column 1 of Table 2 and column 1 of Appendix Table A1. The first three columns of the present table show results for the number-based measure of social networks, while the latter three columns show results for the proportion-based measure. For reference, column 1 and 4 replicate the analysis of column 1 of Table 2 and column 1 of Appendix Table A1, respectively. Columns 2 and 5 present results for the 100 largest CZs and Columns 3 and 6 expand the choice set to the 200 largest CZs. For reference, Appendix Figure A2 shows a map of the largest CZs as well as a cumulative distribution of the proportion of users they account for.

Table A5: Robustness: Social Networks and Residential Choice in Expanded Sample, Proportions of Friends

	In CZ At Time of Measuring Location			
	(1)	(2)	(3)	(4)
Prop. Friends in Dest. CZ	0.6322***	0.6423***	0.6552***	0.5134***
	(0.0153)	(0.0169)	(0.0425)	(0.0260)
Number of Observations	28,517,150	27,469,200	13,707,900	18,839,600
Number of Individuals	198,299	198,299	198,299	198,299
Dependent variable mean	0.012	0.012	0.015	0.012
R^2	0.728	0.755	0.893	0.860
Year X Cohort X Initial-CZ X Dest-CZ FEs	✓			
... X Gender FEs		✓		
... X Income FEs			✓	
... X College FEs				✓

Notes: Table presents coefficient estimates of equation 9 adjusted for the Expanded Sample using the proportion-based definition of social networks. Table is otherwise analogous to Table 3.

Table A6: Robustness: Social Networks and Residential Choice, Logit Controls and Sample

	In CZ After Grad.			
	(1)	(2)	(3)	(4)
Number Friends in Dest. CZ	0.0033***	0.0032***	0.0046***	0.0048***
	(0.0001)	(0.0001)	(0.0000)	(0.0001)
Number of Observations	64,596,900	37,600,500	75,977,150	44,356,900
Number of College Graduates	1,291,938	752,010	1,519,543	887,138
Dependent variable mean	0.012	0.020	0.012	0.020
R^2	0.677	0.768	0.588	0.646
Cohort X Coll-CZ X HS-CZ X Dest-CZ FEs	✓	✓		
Logit Control Variables			✓	✓
Logit Sample		✓		✓

Notes: Table presents coefficient estimates similar to those presented in Table 2 while varying the set of controls as well as the set of users included in the analysis. The objective of these alternative controls and samples is to replicate the approach of the multinomial logit estimates presented in Table 4 while using an OLS estimator. For reference, column 1 replicates the analysis of column 1 of Table 2. Column 2 subsets the analysis to users who choose to live in one of the 50 largest CZs after graduation. In column 3, we replace the rich set of fixed effects by a more parsimonious version, i.e., we control for the distance and distances squared between an individual's college CZ and the potential destination CZ as well as distance between childhood CZ and the potential destination CZ. The regressions also include indicators for whether the individual attended college in the given CZ, and one for whether the individual grew up in said CZ. Column 4 combines these two changes and focuses on the set of users included in column 2 while also using the more parsimonious vector of control variables as in column 3.

Table A7: Robustness: Social Networks and Residential Choice, Logit Controls and Sample, Proportions of Friends

	In CZ After Grad.			
	(1)	(2)	(3)	(4)
Prop. Friends in Dest. CZ	0.3099***	0.3028***	0.3174***	0.3671***
	(0.0046)	(0.0051)	(0.0028)	(0.0035)
Number of Observations	64,596,900	37,600,500	75,977,150	44,356,900
Number of College Graduates	1,291,938	752,010	1,519,543	887,138
Dependent variable mean	0.012	0.020	0.012	0.020
R^2	0.678	0.770	0.588	0.646
Cohort X Coll-CZ X HS-CZ X Dest-CZ FEs	✓	✓		
Logit Control Variables			✓	✓
Logit Sample		✓		✓

Notes: Table is analogous to Table A6 while using the proportion-based definition of social networks.

Table A8: Model Parameters - Network Model

Parameter	Description	Value	Reasoning
α_w	Utility of Wages	1.8	6.2.2
α_h	Utility of Rents	0.9	$:= \frac{1}{2}\alpha_w$
η^H	Housing supply elasticity	2	Saiz (2010)
α^Y	Capital share	0.33	Standard
η^Y	Armington elasticity	4	Feenstra et al. (2018)
ρ	Real interest rate	0.025	Standard
β	Effect of social networks	3.2	Section 6.2.1
δ	Effect of preferences correlated w/ networks	1.8	Section 6.2.1

Notes: Table presents overview of model parameters used for the network model in Section 6.3. For details regarding the calibration of the model, see 6.2 and Appendix Section E.

Table A9: Model Parameters - Basic Model

Parameter	Description	Value	Reasoning
α_w	Utility of Wages	0.75	6.2.2
α_h	Utility of Rents	0.375	$:= \frac{1}{2}\alpha_w$
η^H	Housing supply elasticity	2	Saiz (2010)
α^Y	Capital share	0.33	Standard
η^Y	Armington elasticity	4	Feenstra et al. (2018)
ρ	Real interest rate	0.025	Standard
6.2.1			

Notes: Table presents overview of model parameters used for the basic model in Section 6.3. For details regarding the calibration of the model, see 6.2 and Appendix Section E.

B Logit Framework

The patterns shown in Section 3 indicate that an individuals' responsiveness to local economic growth depends on their connectedness to that CZ. Specifically, for home CZs, the patterns suggest that the more connected a given individual is, the less responsive they are to wage and employment growth there. In contrast, for other alternative CZs, a more connected individual tends to be more responsive to wage and employment growth. In this Section we discuss how those patterns are consistent with a standard discrete choice model following a logit structure.

Consider the utility function of our model framework in Section 6

$$U_{ij} = \alpha_w \ln(w_j) - \alpha_h \ln(r_j) + A_j + \tilde{\beta} n_{ij} + \tilde{\delta} \bar{n}_{ij} + \epsilon_{ij}$$

Under the model assumptions made in 6 (that is, that it follows a logit structure), this utility function implies heterogeneous responses to local economic forces such as w_j. To see this, note that the probability that i individual lives in j is given by:

$$\psi_{ij} = \frac{\exp(V_{ij})}{\sum_{j'} \exp(V_{ij'})} = \frac{\exp(\alpha_w \ln(w_j) - \alpha_h \ln(r_j) + A_j + \tilde{\beta} n_{ij} + \tilde{\delta} \bar{n}_{ij}}{\sum_j' \exp(\alpha_w \ln(w_{j'}) - \alpha_h \ln(r_{j'}) + A_{j'} + \tilde{\beta} n_{ij'} + \tilde{\delta} \bar{n}_{ij'})} \tag{17}$$

That is, the probability that i lives in CZ j is given by the exponentiated indirect utility associated with living in j divided by the sum of the exponentiated indirect utility for all potential destinations. Taking the partial derivative with respect to w_j yields:

$$r_{ij} := \frac{\partial \psi_{ij}}{\partial w_j} = \alpha_w(\psi_{ij}(1 - \psi_{ij})) \tag{18}$$

Since $\tilde{\beta}$ enters into ψ_{ij}, networks affect an individual's responsiveness to w_j. This helps to explain why in some cases network size is correlated with an increase in responsiveness to local economic conditions (as in Panel A of Figure 3), while in other cases it is correlated with a decrease in responsiveness (as in Panel B of Figure 3). Concretely, responses to changes in economic factors w_j depend on the value of $\psi_{ij}(1 - \psi_{ij})$, which is non-monotonic in the baseline probability of living in j, ψ_{ij}. For values of ψ_{ij} above 0.5 (which are generally found when considering an individual's likelihood of staying in their initial region), the responsiveness will decrease with additional friends, while for areas with ψ_{ij} below 0.5 (usually found when considering regions other than one's initial region), responsiveness increases with the size of one's social network in j.

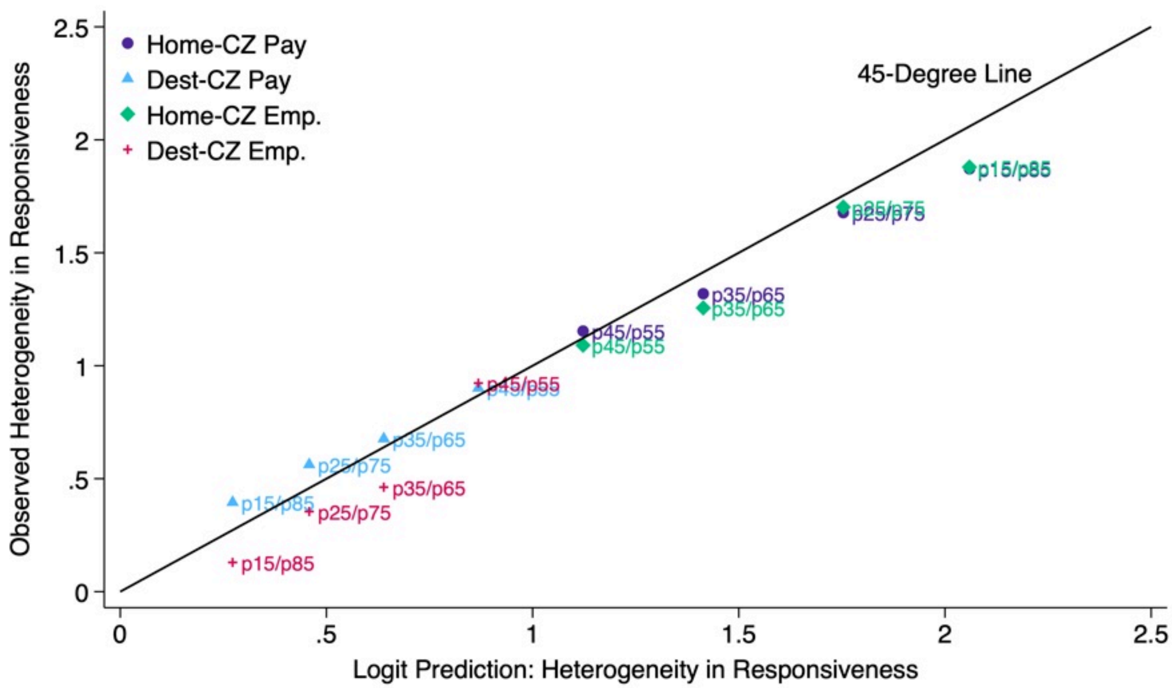
In Figure B1 we highlight how the heterogeneous responsiveness to changing economic conditions we observe in our data matches the predictions given by the logit functional form. In this figure, we use equation 18 to construct ratios of how responsive an individual at a given percentile of the distribution of local social networks is predicted to be relative to an individual at a different percentile, drawing on the cross-sectional relationship between the probability of living in a given CZ and the size of one's local social network shown in Figure 2.[B1] We then compare those predicted levels of heterogeneity to the corresponding ratios stemming from the estimates in Figure 3. The result of this comparison is shown in Figure B1 with different colors and symbols indicating the different measures of changes in local economic conditions—employment and pay growth—as well as whether the results pertain to changes in the home CZ or in other CZs. Reassuringly, all points are close to the 45-degree line suggesting that the logit framework yields predictions about the heterogeneity in the responsiveness similar to those we observe in the data. This helps to reconcile the observed patterns where in some cases people are more responsive to local economic conditions when they have larger networks there while in other cases they become less responsive with greater connectedness.

[B1] For instance, for the p75 to p25 ratio, equation 18 implies that

$$\frac{r^{p75}}{r^{p25}} = \frac{\psi^{p75}(1-\psi^{p75})}{\psi^{p25}(1-\psi^{p25})}$$

where we omit the subscripts ij. Using the magnitude of the cross-sectional relationship between these probabilities of residential choice and the size of one's local social network from Figure 2 therefore enables us to construct these ratios.

Figure B1: Heterogeneity in Responsiveness in Logit Model and Observed



Notes: Figure compares heterogeneity in responsiveness to economic forces observed in 3 to those predicted by multinomial model discussed in Section B. All dots show ratios of responsiveness for those at various percentiles of the connectedness distribution. Ratios observed empirically are shown on the vertical axis, and ratios predicted by the logit framework are displayed on the horizontal axis. The observed ratios are constructed based on the estimates presented in Figure 3 and Appendix Figure A5. The predicted ratios are constructed using equation 18 for those at various percentiles of the connectedness distribution and their probabilities of living in a given CZ. For instance, for the p75 to p25 ratio, equation 18 implies that $\frac{r^{p75}}{r^{p25}} = \frac{\psi^{p75}(1-\psi^{p75})}{\psi^{p25}(1-\psi^{p25})}$. Values for local wage growth are shown in purple circles for the home CZ and blue triangles in alternative CZs, respectively. Values for local employment growth are shown in green diamonds for the home CZ and red pluses in alternative CZs, respectively.

C Heterogeneity and Interactions of Network Effects

In this Section we present additional evidence examining the heterogeneity of the network effects as well as how those interact with other forces driving location decisions. We begin with a discussion of heterogeneity.

C.1 Heterogeneity

C.1.1 Heterogeneity By Initial Connectedness

The estimates regarding the network effects presented in Section 5 correspond to treatment effects *averaged* over all 50 CZs included in the choice set. The weights in this average effect are determined by the frequency of friends' moves to a given location among the sample of individuals we observe. These weights are sensible for deriving the average treatment effects among the individuals we observe, but may conceal heterogeneities in the effects by commuting zone, which could be important in a context in which the distribution of friends' moves differed from the setting we observe. Given the discussion of the logit framework in Section B and the motivating patterns in Section 3, it seems plausible that the strength of the network effect varies with an individual's ex ante probability of locating in a given CZ. Moreover, places that college graduates are more likely to live in are potentially also places where they have more friends to begin with and where they may experience more friends moving to. To that end, Appendix Figure C1 displays the relationship between the absolute number of friends' moves in the pre-period for a given CZ — on the vertical axis — and the initial level of connectedness, or the proportion of one's friends in said CZ on the horizontal axis. Consistent with the above conjecture, college graduates are substantially more likely to see changes in the number of friends in places in which they have many ex ante connections than in places where they have relatively few friends initially. Consequently, the average effect of social networks on residential choice discussed so far likely skews towards the effect for places in which college graduates already have relatively many connections.

Against this backdrop, in Figure C2 we investigate how the effect of social networks on residential choice varies depending on one's initial stock of friends in a given CZ. In Panel A of Figure C2, we repeat the baseline analysis conducted in column 1 of Table 2, estimating the effects separately by quantile of initial connectedness, shown on the horizontal axis. The dashed line indicates the baseline average effect. The figure shows a striking pattern with substantially larger network effects for places to which individuals have stronger ties to begin with: while having one additional friend in a place where an individual has no or almost no friends has an effect of around 0.1p.p., for places that college graduates have around 20-50%

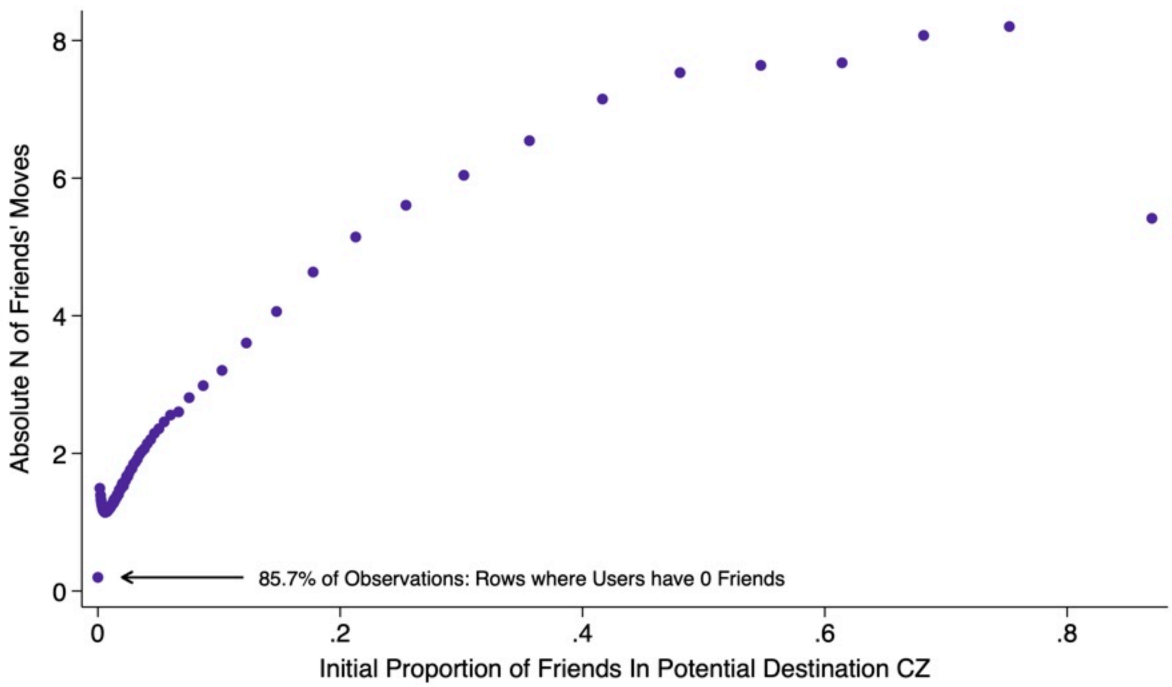
of their local ties—often their home CZs—the effect is around six times as large.

While this suggests that network effects are much larger for places with many friends, it is only one side of the equation. As Panel B of the same Figure shows, college graduates are also dramatically less likely to end up living in CZs with few or no prior connections to begin with: for places with less than 1% of prior connections, the likelihood of a college graduate choosing to live there is well below 0.5%. On the other hand, for CZs where college graduates have around 5% of their friends, that same probability jumps to around 3% and for places with 40% of friends, the probability is over 40%. In Panel C, we therefore re-scale the magnitudes in Panel A by these baseline probabilities to express the effect of social networks on location choice in percent rather than percentage points, we find that having one additional friend in a CZ in which one has less than 1% of one's prior connections leads to a more than 20% increase in the probability of living in that CZ upon graduation. In contrast, the same change in the number of friends for a place in which an individual has about 5% of their friends only increases one's odds of living in that CZ by approximately 5%. Put differently, for places to which a college graduate has few initial connections and for which the probability of living there is very close to zero, one additional friend makes a large proportional difference even if the probability remains quite low. On the other hand, for places with many initial connections and a high baseline rate of living there, one more friend only makes a small proportional difference even if shifting the probability by larger absolute amount.

The analysis in Figure C2 hence suggests that the average effect discussed in Figure 5 and Table 2 is driven by effects for places that college graduates are rather likely to live in to begin with and in which they have a fair number of connections several years before graduating. While this skewed average may very well be the coefficient of interest given that we may be more interested in the extent to which friends' location choices impact an individual's decision where to live for places that both the friends and the individuals have at least a "decent chance" of living in, it is important to bear in mind this dimension of heterogeneity when thinking about the magnitudes.

Interestingly, the patterns shown in C2 are consistent with a the predictions from a discrete choice model with a logit structure. Appendix Section B demonstrates that the extent to which wage effects scale with an individual's connectedness to a given CZ are consistent with the predictions from a logit model like the one discussed in Section 6. The patterns in Figure C2 are also in line with a logit model: in Panel D, we show how the ratios of network effects for individuals at various percentiles of the distribution of initial connections

Figure C1: Friends' Moves vs. Baseline Connectedness



Notes: Figure shows statistics on the relationship between the likelihood of having a friend moving to or away from a given CZ and one's initial level of connectedness to said CZ. To construct this Figure, we calculate the sum of all friends' moves by individual by CZ and then take these absolute values of these sums. We then group rows of individuals-by-CZ into 99 equal-sized bins based on the initial proportion of friends one has in a given CZ, i.e., based on one's stock of friends as of the summer after the first year in college. We then plot the average absolute number of friends moves for each of those bins as well as for all individual-by-CZ rows with no initial connections.

compare with corresponding predicted ratios obtained from a multinomial logit.[C1] The figure thus parallels the analysis conducted in Figure B1. While the ratios observed empirically generally lie above the 45-degree line and thus exceed the heterogeneity predicted by a logit model, the deviations are relatively small and, more importantly, the two series are very highly correlated with a slope that is close to one. Thus, the extent to which the logit model predicts larger or smaller network effects is broadly consistent with the patterns observed in the data. Therefore, given that the logit model allows us to compare more directly the magnitudes we obtain for different groups of people, to study additional dimensions of heterogeneity we use a multinomial logit framework below.

[C1] Concretely, we use the coefficient estimates of Panel A of the same figure to construct ratios of effect sizes for those with different baseline connections. These ratios are shown on the vertical axis of Panel B of Figure C2. In a second step, we then use equation 17 to construct corresponding ratios based on the logit framework discussed in Section B using the observed baseline probabilites of living in a given CZ for those with various levels of connections. These ratios are shown on the horizontal axis of Panel B of Figure C2.

C.1.2 Heterogeneity By individual Background

College Graduates

The above discussion suggests that the multinomial logit approach is more suitable to compare effect sizes as it mitigates concerns that any potential differences are driven by underlying differences in the baseline probabilities of living in different CZs. With this in mind, from here forward we will use the logit approach to further explore heterogeneity along various dimensions.

We begin by studying heterogeneity by characteristics of the college graduates. We are particularly interested in the degree to which the effects vary with individual's socio-economic background. Though our primary research design only allows us to consider heterogeneities within the set of college graduates, substantial differences in socioeconomic status exist even within this group (Chetty et al., 2020). To that end, in the first set of bars in Panel A of Figure C3 we present results for those attending a college in the bottom and top quintile of the distribution of college-level SAT score.[C2] The figure suggests that the effects of social networks on residential choice are significantly larger for those attending a college with lower average scores: those at institutions in the bottom quintile of the score distribution exhibit effects of just under 0.05 while the effects of social networks for those in the top quintile is 0.03, or 40% smaller. In the second set of bars, we replicate this analysis studying heterogeneity by parental socioeconomic status. That is, we use our sample of college graduates whom we can link to their parents to estimate separate regressions for the top and bottom quintile of parental socioeconomic status. While the differences are not quite as striking as in Panel A, the pattern again suggests a modest degree of heterogeneity, with larger effects for those whose parents have lower SES. In sum, these observations suggest that the effects of social networks on residential choice are larger for those at the bottom of the distribution for both achievement and income.

In the same vein, in Panel B of Figure C3 we explore heterogeneity in the effects of social networks by the economic circumstances in one's home CZ. We ask to what extent the effects differ for people coming from places where the economy is doing comparatively well, relative to those in which it is doing poorly. Building on the evidence presented in Section 3 and Figure 3, we study heterogeneity by changes in local employment and average pay for which we instrument using a shift-share approach based on a CZ's industry composition.[C3] The first set of bars compares results for those coming from places with low (bottom quintile)

[C2] We use college-level 2013 average SAT scores obtained from Chetty et al. (2020).

[C3] For more details on the construction of the shift-share instrument and the predicted changes in employment and average pay, see Section 3.

predicted employment growth to those coming from CZs with high (top quintile) employment growth, while the second set of bars draws a similar comparison based on predicted average pay. Again, we find that the network effects are somewhat larger for those at the lower end of the distribution: for those coming from a CZ with bottom quintile employment growth, the effect is more than 50% larger than for those from the top quintile. While the difference is slightly less stark in the case of pay changes, it is again the case that the network effects are larger for those in the bottom quintile than for those in the top quintile.

Expanded Sample

We also explore heterogeneity of the network effects in the Expanded Sample. In Table C1 we present coefficient estimates similar to those shown in Table 4 or Figure C3 from a multinomial logit approach. In the first column of Table C1 we find a point estimate of close to 0.05 suggesting that having one more friend moving to a given CZ prior to t^* increases the log odds ratio of living in that CZ relative to the reference CZ by 0.05. This estimate is relatively similar in magnitude to that presented for the sample of college graduates in Table 4.

To learn more about differences in the effect sizes across groups, in the subsequent columns we replicate our analysis on different subsets of our sample. To begin, building on the patterns observed in Figure C3, we ask whether the effects differ for those with and without a college education. Consistent with our previous results, if anything, the effects are larger for those without a college background even if the differences are small. In columns 4-6 we cut the sample by a user's birth cohort. We find that older individuals exhibit substantially larger effects than younger individuals, with effects close to 70% larger for those born between 1985-89 relative to those estimated in our baseline regressions in Table 4. Interestingly, for those born between 1995-97—the same set of cohorts included in our primary analyses focusing on college graduates—we find coefficient estimates that are remarkably similar to those found in Table 4. This is true despite of the somewhat different empirical design and the inclusion of non-college graduates in the analysis.

C.2 Interaction Effects of Social Networks and Economic Factors

To better understand the economic implications of the network effects we observe, in this section, we study how an individual's responsiveness to local economic conditions varies with the size of one's local social network. Recall from Figure 3 that descriptively the size of one's local social networks predicts greatly the degree to which an individual responds to local

Table C1: Effect of Friends' Moves on Likelihood Living in CZ - Expanded Sample, Logit Estimates

	In CZ At Time of Measuring Location					
	(1)	(2)	(3)	(4)	(5)	(6)
Number Friends in Dest. CZ	0.0497***	0.0481***	0.0502***	0.0642***	0.0419***	0.0382***
	(0.0005)	(0.0006)	(0.0010)	(0.0018)	(0.0008)	(0.0006)
Number of Observations	24,621,100	15,217,550	9,403,550	9,144,950	10,244,550	5,231,600
Number of Users	166,459	102,664	63,795	61,702	69,255	35,502
$Pseudo - R^2$	0.693	0.647	0.766	0.748	0.680	0.642
Sample	Full Sample	College	No College	Born 1985-89	Born 1990-94	Born 1995-97

Notes: Table shows estimates of the effect of social networks on residential choice for the Expanded Sample. The estimates are obtained from regressions described in equation 15 and are analogous to those shown for $\tilde{\beta}$ in Table 4. Each column presents estimates for a different set of individuals. Column 1 shows estimates for the full, generalized sample discussed in Section 2. Columns 2 and 3 present separate for those listing a college on their profile as well as those who do not list a college. Columns 4-6 make different cohort restrictions with column 4 restricting to individuals born between 1985 and 1989, column 5 subsetting to individuals born between 1990 and 1994, and column 6 focusing on individuals between 1995 and 1997. Standard errors are shown in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

economic forces in their home CZ and elsewhere. The framework described in Section B suggested one mechanism for this effect, suggesting that the effect of one's network on their migration decisions scales in proportion to their baseline probability of moving to a given place, following the intuition of a logit model. In this section, we examine a potential additional channel, where individuals additionally become more responsive to changing economic conditions in a place if they have a larger pre-existing network there. In Figure C4, we test for the presence of such an interaction by estimating equations of the form:

$$\begin{aligned} Y_{ij,t^*} = \mu_j + \beta_{early} n_{ij,early} + \sum_{t \in [pre,mid,post]} \beta_t (n_{ij,t} - n_{ij,t-1}) + \\ \sum_{t \in [pre,mid,post]} \gamma_t [(n_{ij,t} - n_{ij,t-1}) \times w_{j,t^*}] + dist_{kj} + dist_{kj}^2 + I(k=j) + \varepsilon_{ij,t^*} \end{aligned} \tag{19}$$

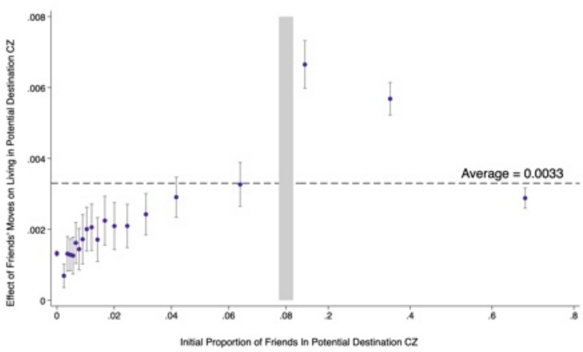
In words, we interact changes in the size of one's local network with local economic characteristics w_{j,t^*}. In the same way in which we construct the difference between β_{pre} and β_{post} to estimate the effect of social networks on residential choice generally, we also construct the difference between γ_{pre} and γ_{post}—denoted γ—in order to capture the interaction between the role played by social networks and other economic characteristics. We parameterize w_{j,t^*} based on our two measures of local economic conditions, predicted changes in local employment and average pay. In each pair of bars of Figure C4 we show estimates of β (or

$\beta_{pre} - \beta_{post}$) on the left and γ (or $\gamma_{pre} - \gamma_{post}$) on the right.

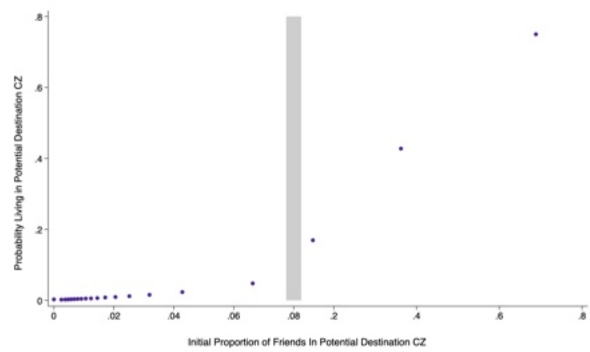
In the first four bars, we interact the number of friends with an indicator for whether the predicted employment growth in the potential destination lies above the median, or the 75th percentile, respectively. Our estimates suggest that, γ is rather small in comparison to β. While the point estimates of γ are negative and significant suggesting that networks effects are smaller for CZs with high employment growth, the magnitude of this result is small and the pattern is non-monotonic with an estimate close to zero for places with employment growth above the 75th percentile. In the next four bars we interact the effects of social networks with indicators for whether the potential destination exhibits above median or top quartile changes in pay, respectively. The conclusions are very similar: the effect on the interaction tends to be negative but small, implying that the effect of social networks does not vary substantially with the local economic conditions of a potential destination.

Figure C2: Heterogeneity by Baseline Connectedness

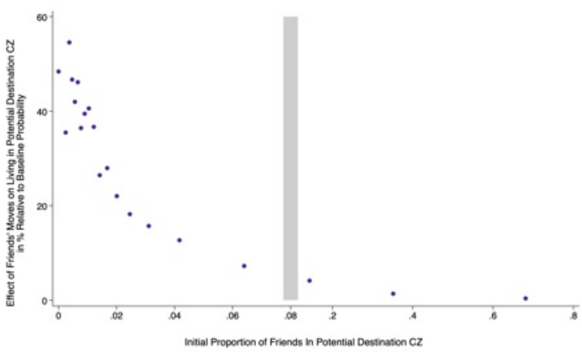
(a) Heterogeneity in Network Effects



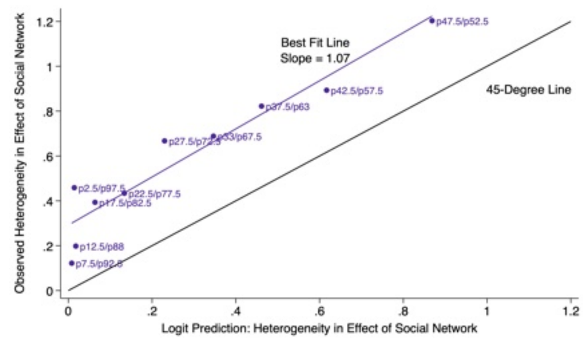
(b) Probabilities of Living in CZ



(c) Social Network Effects in % of Baseline Probability



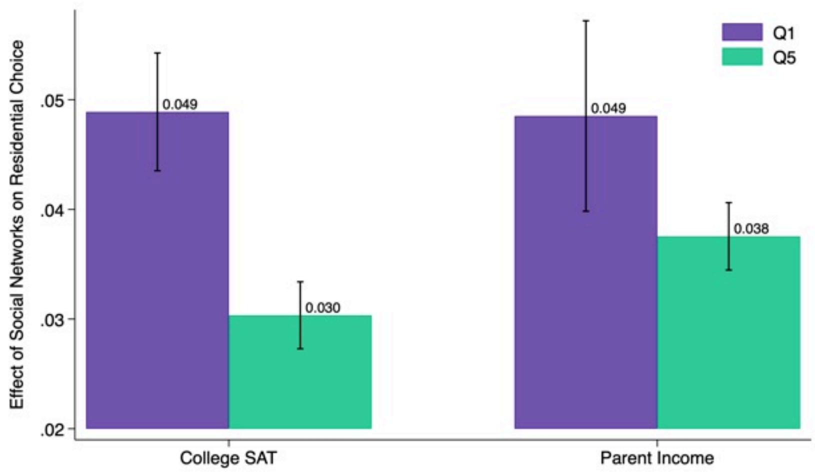
(d) Predicted vs. Observed Heterogeneity



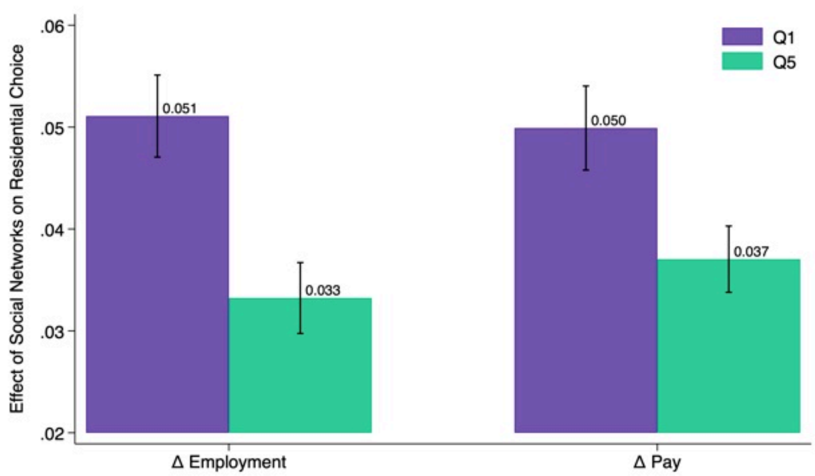
Notes: Figure presents heterogeneity in the networks effects by initial connectedness to a given CZ. Panel (a) shows coefficient estimates from regressions of the form described in equation 9 conducted separately for different levels of connectedness. We group individual-by-CZ observations into ventiles based on the proportion of friends a given individual has in a CZ prior to senior year. For each of these ventiles, we then present estimates of $\tilde{\beta}$ on the vertical axis and the average proportion of friends on the horizontal axis. Vertical lines going through coefficient estimates show 95% confidence intervals. The dashed horizontal line indicates the average effect of social networks on location decisions presented in Table 2. Panel (b) presents descriptive statistics on the relationship between the likelihood of locating in a given CZ and one's initial level of connectedness. Panel (c) presents the effects of social networks in % relative to one's baseline probability of locating in a given CZ. To construct this figure, we divide the coefficient estimates observed in Panel (a) of by the values observed in Panel (b). The gray bar in the middle of Panels (a)-(c) indicates a break in the scaling of the horizontal axis which we make for expositional reasons. Panel (d) contrasts the heterogeneity observed in Panel (a) with predictions about the heterogeneity stemming from the framework described in more detail in Appendix Section B. Each dot corresponds to a ratio of effects for one ventile relative to another ventile as is labeled in the graph using a percentile scale. Observed ratios constructed based on the estimates in Panel (a) are shown on the vertical axis while the predicted ratios are shown on the horizontal axis. Values above the 45-degree line shown in black indicate that the observed ratios exceed the predicted ratios. The purple line highlights the best line of the observed ratios vs. the predicted ratios.

Figure C3: Heterogeneity in Effects of Social Networks by individual Background

(a) Socio-Economic Background

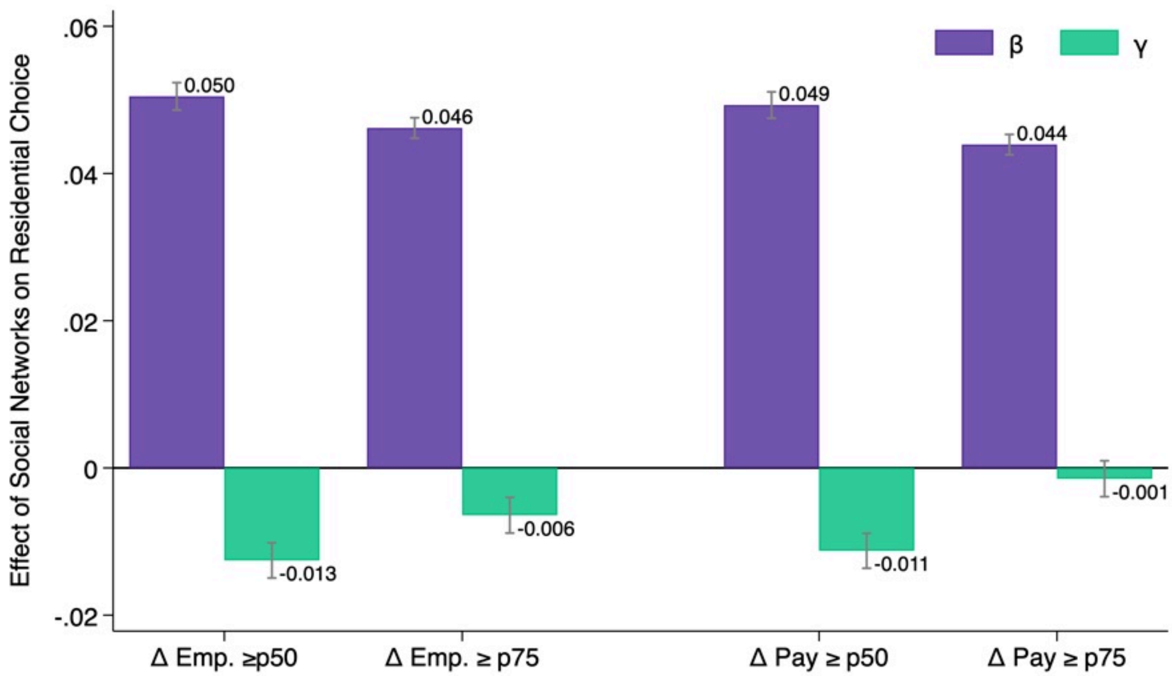


(b) Local Economic Condition in Home CZ



Notes: Figure shows heterogeneity in effects of social networks by individual background with Panel (a) focusing on a individual's socio-economic background and Panel (b) showing heterogeneity by the economic conditions in a individual's home CZ. For each of the analysis shown, we subset the sample of individuals in the described way and estimate equation 15. The bars then show estimates of β, or the effect of having one additional friend in a given CZ on one's probability of living there after college. Purple bars correspond to individuals in the bottom quintile of the various distributions while green bars show values for individuals in the top quintile. Vertical lines show 95% confidence intervals. The left two bars in Panel (a) show separate estimates for individuals attending a college in the bottom and top quintile of the distribution of college-level average SAT scores in 2013. The data on college-level SAT scores comes from Chetty et al. (2020). The next two bars in Panel (a) show separate estimates for individuals in the bottom and top quintile of the parental income distribution. More details on the construction of these income measures can be found in Section 2. Panel (b) shows heterogeneity for those in the bottom and top quintile of the distributions of predicted local employment and wage growth. For more details regarding the construction of these predicted growth measures, see Section 3.

Figure C4: Interaction Between Network Effects and Local Economic Factors



Notes: Figure presents estimates of equation 19. Purple bars correspond to estimates of β, or $\beta_{pre} - \beta_{pre}$ in equation 19 while green bars stand for estimates of γ, or $\gamma pre - \gamma pre$. The first two sets of bars present results from specifications in which we interact friends' moves with local employment growth. The second set of bars present analogous results while interacting friends' moves with local wage growth. Bars 1 and 2 as well as 5 and 6 interact friends' moves with indicators for whether local growth in the potential destination is at or above the median, while the remaining bars are interactions with indicators for whether local growth in the potential destination is at or above the 75th percentile. Vertical lines show 95% confidence intervals.

D Model Derivations and Equilibrium

In this section, we provide additional details and derivations of the equilibrium conditions of the model presented in Section 6.

In equilibrium, the housing and labor market must clear. Given the fact that social networks are an input to an individual's utility maximization problem, and that social networks are itself an endogenous object, we cannot fully characterize the equilibrium in closed form. Below, we detail the equilibrium conditions which we can derive analytically and also discuss which conditions we cannot obtain a closed form solution for.

Housing

Due to the Cobb-Douglas nature of the utility function, households spend a constant fraction of their income on housing. All households together therefore spend $\frac{\alpha_h}{\alpha_h+\alpha_w} w_j N_j$ on housing. We divide by the rental rate to express this in terms of "units" of housing. The resulting housing demand is given:

$$H_j^D = \left(\frac{\alpha_h}{\alpha_h + \alpha_w} w_j N_j\right)/r_j$$

Setting housing demand and supply equal, we can thus write the rental rate as follows:

$$r_j = \left[\frac{\alpha_h}{\alpha_h + \alpha_w} \frac{w_j N_j}{\nu_j}\right]^{\frac{1}{1+\eta^H}}$$

We can also re-arrange this to back out the local rental rate shifter ν_j:

$$\nu_j = \frac{\alpha_h}{\alpha_h + \alpha_w} \frac{w_j N_j}{r_j^{1+\eta^H}} \tag{20}$$

In Appendix Section E, we discuss how we use equation 20 to find local housing supply shifters ν_j, so that the distribution of rents in the model matches the observed distribution.

Production and Labor Demand

From the cost minimization problem of local producers in each CZ, it follows that local wages rate are given by:

$$w_j = (1-\alpha^Y)(p_j\theta_j)^{\frac{1}{1-\alpha^Y}}\left[\frac{\alpha^Y}{\rho}\right]^{\frac{\alpha^Y}{1-\alpha^Y}} \tag{21}$$

In addition, the optimality conditions at the national level imply that the ratio of the marginal products of good j and any j' equals their price ratio. That is

$$\frac{\frac{\partial Y}{\partial Y_j}}{\frac{\partial Y}{\partial Y_j'}} = \frac{p_j}{p_j'}$$

The marginal product $\frac{\partial Y}{\partial Y_j}$ is given by:

$$\frac{\partial Y}{\partial Y_j} = \frac{\eta^Y}{\eta^Y - 1}\Big[\sum_{j\in J}(Y_j)^{\frac{\eta^Y-1}{\eta^Y}}\Big]^{\frac{\eta^Y}{\eta^Y-1}-1}\frac{\eta^Y-1}{\eta^Y}[(Y_j)^{\frac{\eta^Y-1}{\eta^Y}-1}]$$

Note that $\Big[\sum_{j\in J}(Y_j)^{\frac{\eta^Y-1}{\eta^Y}}\Big]^{\frac{\eta^Y}{\eta^Y-1}-1} = Y^{\frac{1}{\eta^Y}}$, so that we can write:

$$\frac{\partial Y}{\partial Y_j} = \left[\frac{Y}{Y_j}\right]^{\frac{1}{\eta^Y}}$$

Therefore, making use of the above optimality condition and normalizing the price in reference CZ $\bar{j}$ to 1, we find that:

$$p_j = \frac{Y_{\bar{j}}}{Y_j}$$

From the optimality conditions of local producers it follows that $\frac{K_j}{N_j} = \left[\frac{\rho}{\alpha^Y\theta_j p_j}\right]^{\frac{1}{\alpha^Y-1}}$. Thus $Y_j = \theta_j N_j\left(\frac{p_j\theta_j\alpha^Y}{\rho}\right)^{\frac{\alpha^Y}{1-\alpha^Y}}$. As a result, that after cancelling ρ and α^Y we obtain:

$$p_j = \Big[\frac{\theta_{\bar{j}}N_{\bar{j}}\theta_{\bar{j}}^{\frac{\alpha^Y}{1-\alpha^Y}}}{\theta_j N_j(p_j\theta_j)^{\frac{\alpha^Y}{1-\alpha^Y}}}\Big]^{\frac{1}{\eta^Y}} \tag{22}$$

which we can re-arrange for p_j to find:

$$p_j = \left(\frac{N_{\bar{j}}}{N_j}\right)^{\frac{1-\alpha^Y}{(1-\alpha^Y)\eta^Y+\alpha^Y}}\left(\frac{\theta_{\bar{j}}}{\theta_j}\right)^{\frac{1}{(1-\alpha^Y)\eta^Y+\alpha^Y}}$$

We can next use this expression to recover the local productivity parameters θ_j as function of parameters and wages w_j. To do that, we first re-arrange equation 21 for w_j:

$$\theta_j p_j = \left(\frac{w_j}{1-\alpha^Y}\right)^{1-\alpha^Y} \left(\frac{\rho}{\alpha^Y}\right)^{\alpha^Y}$$

Combining the above with equation 22 and re-arraning yields:

$$\theta_j = \left(\frac{w_j}{1-\alpha^Y}\right)^{\frac{(1-\alpha^Y)\eta^Y+\alpha^Y}{\eta^Y-1}} \left(\frac{\rho}{\alpha^Y}\right)^{\frac{\alpha^Y(1-\alpha^Y)\eta^Y+\alpha^Y}{(1-\alpha^Y)(\eta^Y-1)}} \left(\frac{N_j}{N_{\bar{j}}}\right)^{\frac{(1-\alpha^Y)\eta^Y+\alpha^Y}{[(1-\alpha^Y)\eta^Y+\alpha^Y](\eta^Y-1)}} \left(\frac{1}{\theta_{\bar{j}}}\right)^{\frac{1}{(1-\alpha^Y)(\eta^Y-1)}} \tag{23}$$

In Appendix Section E, we discuss how we use equation 23 to find local productivities θ_j, so that the distribution of wages in the model matches the observed distribution.

Utility Maximization and Labor Supply

Given the standard logit structure of the model, the utility maximization problem implies that the probability that i lives in j is given by:

$$\psi_{ij} = \frac{\exp(V_{ij})}{\sum_{j'} \exp(V_{ij'})} = \frac{\exp(\alpha_w \ln(w_j) - \alpha_h \ln(r_j) + A_j + \tilde{\beta} n_{ij} + \tilde{\delta} \bar{n}_{ij}}{\sum_j' \exp(\alpha_w \ln(w_{j'}) - \alpha_h \ln(r_{j'}) + A_{j'} + \tilde{\beta} n_{ij'} + \tilde{\delta} \bar{n}_{ij'})}$$

Labor supply is therefore given by:

$$N_j = \sum_{i \in I} \psi_{ij}$$

Importantly, as discussed in Section 6, the size of i's social network in j (n_{ij}) is itself an endogenous object as it depends on the utility maximization problem of one's friends:

$$n_{ij} = \frac{\sum_{u \neq i}^{I} g_{iu} \times \psi_{uj}}{\sum_{u \neq i}^{I} g_{iu}}$$

where g_{iu} is an indicator for whether i and u are friends, and ψ_{uj} is the probability that u lives in j.

$$\psi_{uj} = \frac{\exp(V_{uj})}{\sum_{j'} \exp(V_{uj'})} = \frac{\exp(\alpha_w \ln(w_j) - \alpha_h \ln(r_j) + A_j + \tilde{\beta} n_{uj} + \tilde{\delta} \bar{n}_{uj}}{\sum_j' \exp(\alpha_w \ln(w_{j'}) - \alpha_h \ln(r_{j'}) + A_{j'} + \tilde{\beta} n_{uj'} + \tilde{\delta} \bar{n}_{uj'})}$$

This makes an important point: an individuals' utility maximization problem depends on their friends' actions (through n_{ij}) and vice versa (through n_{uj}). We therefore can not analytically derive a closed form solution for ψ_{ij}, and in turn labor supply (N_j). Consequently, we solve for the equilibrium using an iterative, stepwise approach which we discuss in more detail in Appendix Section E.

E Model Calibration

We parameterize the model such that it matches the data in terms of (a) local wages and rents, (b) local populations, (c) the average wage elasticity, and (d) the distribution of local social networks. As discussed in Section 6.2, we draw on prior research for the parameterization of η^Y, η^H, ρ, and α^Y and use the quasi-experimental setup of Sections 4 and 5 to find values for the network parameters $\tilde{\beta}$ and $\tilde{\delta}$. Since we cannot analytically derive the economy's equilibrium — see Section D for details — we employ an iterative, step-wise approach to find the vectors θ_j, ν_j, and A_j as well as the parameters α_w, and α_h. In this approach we take η^Y, η^H, ρ, α^Y, $\tilde{\beta}$ and $\tilde{\delta}$ as given. We explain the approach in more detail below.

1. **Find θ_j, ν_j to match wages and rents**

   We use equations 23 and 20 to solve for local productivities (θ_j) and local housing supply shifters (ν_j), so that for given levels of populations the model matches the distribution of 2019 wages and rents. Importantly, we use skill-adjusted wages[E1] and rents for two-bedroom accommodations to mitigate concerns that regional differences in the composition of skills or housing type might drive differences in average wages and rents.

2. **Guess α_w, α_h and find A_j to match populations**

   Next, we guess a value for α_w and impose that $\alpha_h = \frac{1}{2}\alpha_w$, so that households spend a third of their income on housing given the Cobb-Douglas nature of the utility function. We then iteratively solve for a vector of local amenities A_j's so that the number of people living in a given CZ implied by the worker's utility maximization problem matches the observed number of people living there. Rather than necessarily reflecting "real amenities", the A_j terms can thus be thought of residuals, their primary purpose being that they help to rationalize local population sizes. We discuss the distribution of the A_j's in greater detail in Section 6.3.

3. **Given 2., find implied n_{ij} and let workers re-optimize to confirm equilibrium**

   In step 2, we treat n_{ij} as an exogenous object using the size of one's social network as observed in the micro-data; we now endogenize it. Specifically, given the values from step 2. for α_w, α_h and A_j, we now find the model implied values for n_{ij} (based on ψ_{iu} implied from step 2) and let all workers re-optimize until we arrive at an equilibrium where the number of moves is small. This point constitutes the model's equilibrium

[E1] To construct a measure of skill-adjusted wages, we use the CZ-specific wages by skill level, and reweight those using the national distribution of skill levels.

for a given guess of α_w and α_h.

Importantly, in this step as well as in the counterfactual exercises of Section 6, we let workers optimize asynchronously. Thus, while workers can react to their friends' location choices, they cannot anticipate their friend's decisions, nor can they directly coordinate. We believe this procedure is reasonable for multiple reasons. First, networks tend to be large with the average individual in our sample having over 200 friends making it implausible that individuals will anticipate all their friends' location decisions. Second, this approach is similar in spirit to the widespread use of Poisson processes constraining the extent to which individuals are able to optimize and preventing simultaneous or coordinated decisions (Arnott, 1989). Third, the coefficient estimates shown in Figure 5 for friends' moves after $\bar{t}$ are very close to zero and are hence consistent with the lack of anticipation or coordination.

4. **Simulate shocks, find new equilibrium and implied wage elasticity**

   Steps 1.-3. are concerned with matching the data in terms of wages, rents, populations and the distributions of network. We now turn towards matching the average wage elasticity. To do so, we begin by empirically estimating the wage elasticity for the individuals in our sample. Concretely, we use a shift-share approach similar to the one described in Sections 3 and 5.4 which allows us to study how populations (among those in our sample) have changed between 2012-2019 in response to local wage shocks during that time period.[E2] Appendix Figure A16 presents the results of this empirical exercise and indicates an elasticity of 0.93, which is comparable to estimates documented by prior research.

   In order to find the wage elasticity implied by the model, we simulate numerous local productivity shocks (based on θ_j) and find the new equilibrium levels of wages, rents and populations for each shock and each CZ. We do this starting from the equilibrium described in the step 3. We then construct the implied elasticity for each shock and average the elasticity over all shocks. Lastly, we compare this average elasticity to the empirically observed elasticity.

5. **Iterate over steps 2-4. until implied elasticity matches observed elasticity**

   We repeat steps 2.-4. until we arrive at an elasticity that closely matches the empirically observed elasticity. At the end of this procedure, as shown in Appendix Figure

[E2]Specifically, we instrument for local wage growth in each of CZs under study using a shift-share instrument and regress changes in the local populations among those in our sample on the predicted wage growth rates. Note that the time period 2012-2019 is consistent with the sample period considered in Section 3.

A17 we have found values for α_w, α_h, and the vectors of θ_j, ν_j and A_j, so that the model aligns closely with the data in terms of (a) local wages and rents, (b) local populations, (c) an average elasticity we can observe for this sample, and (d) the distribution of local ties. Appendix Table A8 provides an overview of all parameters.

In order to also calibrate the basic-model, we use an analogous approach in which we repeat the above steps with the exception of steps 1, and 3. By construction, the local productivities (θ_j) and housing supply shifters (ν_j) are identical across the two models which renders step 1 unnecessary. In addition, since the basic-model does not include the role played by local networks, step 3 is no longer needed. As the result of this procedure, we thus obtain values for $\tilde{\alpha}_w$, $\tilde{\alpha}_h$, and the vector of $\tilde{A}_j$ so that the basic-model can match the data in terms of (a) local wages and rents, (b) local populations, and (c) an average elasticity we can observe for this sample. Appendix Table A9 provides an overview of the corresponding parameters for the basic-model.