Long ties, disruptive life events and economic prosperity

Eaman Jahani^{1,2,3,†}, Samuel P. Fraiberger⁴, Michael Bailey⁵, and Dean Eckles^{1,2,†}

¹Sloan School of Management, Massachusetts Institute of Technology; ²Institute for Data, Systems, and Society, Massachusetts Institute of Technology; ³Department of Statistics, University of California, Berkeley; ⁴World Bank; ⁵Meta Platforms

Social networks shape and reflect economic life. Prior studies have identified long ties, which connect people who lack mutual contacts, as a correlate of individuals' success within firms and places' economic prosperity. However, we lack population-scale evidence of the individual-level link between long ties and economic prosperity, and why some people have more long ties remains obscure. Here, using a social network constructed from interactions on Facebook, we establish a robust association between long ties and economic outcomes and study disruptive life events hypothesized to cause formation of long ties. Consistent with prior aggregated results, we show that administrative units with a higher fraction of long ties tend to have higher income and economic mobility. Individuals with more long ties live in higher income places and have higher values of proxies for economic prosperity (e.g., using more Internet-connected devices and making more donations). Furthermore, having stronger (i.e. higher intensity of interaction) long ties is associated with better outcomes, consistent with an advantage from the structural diversity constituted by long ties, rather than them being weak ties *per se*. Second, we uncover the role of disruptive life events in the formation of long ties later in life. Individuals who have migrated between US states, have transferred to a different high school, or have attended college outside their home state have a higher fraction of long ties among their contacts many years after the event. They also exhibit better outcomes along the three economic well-being proxies. Overall, these results suggest that long ties are robustly associated with economic prosperity and highlight roles for important life experiences in developing and maintaining long ties.

Keywords: social networks | weak ties | structural diversity

E.J. was a contractor of Meta Platforms, which operates Facebook, while conducting this research. D.E. has received funding from Meta for other research. Meta has sponsored a conference that D.E. organizes. D.E. was until recently a consultant to Twitter.

[†] To whom correspondence should be addressed. E-mail: eaman@berkeley.edu, eckles@mit.edu

Social networks can shape access to information and opportunities (1–4) and, thus, economic prosperity for individuals and places. Likewise, economic activity can affect the formation and maintenance of social ties. Such reciprocal causal relationships can produce robust and even self-reinforcing associations between network structure and economic prosperity, for individuals and geographic regions. Indeed, recent empirical work relying on digital interaction data at a population-wide scale has documented some notable associations between the characteristics of social connections in a region and economic outcomes in that region. In particular, places with a higher structural and geographic diversity of connections have better economic outcomes (5–8). For example, the relative frequency of Facebook friendship links between every county-pair in the United States is a strong predictor of socioeconomic characteristics such as income, education, teenage birth rate, or life expectancy at the county level (8).

These aggregate associations between network structure and economic propensity are consistent with a broader theoretical and empirical literature on network structure. Long ties, which connect individuals who do not share any mutual contact, are thought to facilitate the circulation of novel and valuable information between multiple communities (9–14). Structurally diverse networks, which are comprised of numerous long ties (i.e. local bridges), can provide individuals with an information advantage which leads to better managerial performance (11, 15–18), a higher propensity to find jobs (19, 20), more innovative teams (21–23), and faster adoption of new technologies (13, 24–26). Despite the extensive literature arguing for benefits of structural diversity (i.e. having a network with more long ties) in specialized professional contexts, the relationship between such network structures and general measures of economic prosperity is under-explored. Population-scale evidence has largely relied on ecological regressions with network structures and economic outcomes aggregated to telecommunication (6) or administrative regions (8), and association with network structure has often been characterized with univariate correlations and/or by combining multiple measures of network structure into a composite index (6). Furthermore, the question of who has more long ties and what makes them to develop these ties is largely unexplored despite their importance.

We have two main contributions: establishing a robust association between long ties and important measures of economic prosperity and linking major, disruptive life events to formation of long ties. First, we present population-scale evidence on the relationship between structural diversity and socioeconomic outcomes. Individual Facebook users in the United States with more long ties have better outcomes along four proxies of socio-economic well-being, including living in richer zip codes, having more internet-connected devices, more expensive phones, and donating more often to charitable fundraising campaigns. Furthermore, we examine the additional role of tie strength, which captures the intensity of interactions between individuals. Individuals with stronger long ties tend to have better economic outcomes, even after accounting for the prevalence of these ties and their total intensity of interactions. Both of these results are also reflected in geographically aggregated results in the US and Mexico. Structural diversity appears to importantly characterize the networks of higher income people and places.

Given the robust association between long ties and economic outcomes, we then probe some potential origins in specific life events. Few studies have focused on the determinants of long ties. Existing works have shown that brokers, who connect two otherwise disconnected communities by filling a structural hole, are less risk-averse, more entrepreneurial, and more prone to change (10, 27, 28). They have a higher ability to integrate with various social groups and form structurally diverse networks because they are less likely to define themselves and others in terms of socially distinct categories (29, 30). While these studies suggest that personality traits are associated with the presence of long ties, personality traits are affected by life experience (31) and network measures related to long ties have large non-heritable components (32, 33). We therefore study life events that expose individuals to multiple communities and push them to establish connections with individuals who might be very different from their previous contacts: (i) migrating to another state, (ii) transferring between high schools, and (iii) attending an out-of-state college. In all three cases, experiencing such disruptions is

Fig. 1. (A) Fraction of long ties, measured across counties in the United States, after controlling for average degree. The adjusted (residualized) fraction of long ties across administrative units in each country is binned into 10 deciles, with lower values shown in purple and higher values shown in green. Administrative units with less than 200 Facebook users sufficiently active over the measurement period or with less than 500 individuals are shown in white. **(B)** The fraction of long ties and **(C)** the tie-strength weighted fraction of long ties predict median income in US zip codes. The binned regression plots are generated according to models (6) and (7) in *Materials and Methods*, which adjust for number of nodes in the zip code, its racial composition, and average degree; the regression with weighted fraction of long ties also adjusts for the fraction of long ties and the sum of the weights. Solid lines are local smoothers of second degree; bars are 95% confidence intervals, which are cluster-robust at the county level.

associated with having more structurally diverse networks years later, even excluding the ties more directly implicated (e.g., ties to alumni of the involved schools). Though we flexibly adjust for covariates, we cannot attribute these associations entirely to causal effects of these events; however, in the case of transferring between high schools, a quasi-experimental analysis focusing on school closures yields similar results. Together these results provide a view of long ties as consistently linked to disruptive life events, even long after those events.

Long Ties and Economic Outcomes

How does structural diversity relate to economic outcomes? To answer this question, we first count all the public comments posted on Facebook between December 2020 and June 2021 between users living in the United States and Mexico to form a network of interactions between users; this is the network used throughout the paper (see *Materials and Methods* for details). As a first measure of structural diversity, we compute the fraction of each individual's ties that are long ties (i.e. lack any mutual network neighbors).

Pooling the networks of all Facebook users living in the same region provides an overview of variation in this measure of network structure (Fig. 1A and S3). The least structurally diverse regions with available data, such as Appalachia in the US or Zacatecas in Mexico, are economically disadvantaged. Zip codes in the US with more long ties have higher household incomes (Fig. 1B), even after flexibly controlling for their overall population and racial composition, the number of Facebook users residing in the zip code, and their average degree (see *Materials and Methods* for details). This positive relationship holds for additional measures of economic prosperity in the US, including social mobility and unemployment rates, and for a wealth index in Mexico (Fig. S4).

Our primary analyses examine associations at the level of individuals', rather than places', networks. As typical measures of economic outcomes (e.g., income, wealth) are not available at an individual level linked to Facebook users, we rely on proxy measures of economic prosperity (34). In the six most populous US states, we divide the fraction and weighted fraction of long ties each into 12 equal-sized categories and match on degree, level of daily activity, gender, and age (see *Materials and Methods* for details). Individuals with a higher fraction of long ties live in zip codes with higher median family

Fig. 2. Fraction of long ties (**top row**) and tie-strength weighted fraction of long ties (**bottom row**) in ego networks and individual level variables that are positively correlated with proxies for economic outcomes for users who reside in the top 6 most populous states in the US. Users are matched on degree, average degree of their contacts, daily login activity, average daily login activity of their contacts, age, and gender; see *Materials and Methods*. The matching for the bottom row additionally incorporates the fraction of long ties and average edge weight. Values in the y-axis correspond to the difference with the baseline represented with a value of zero. Bars are 95% confidence intervals for these differences assuming independent observations; robustness to within-state dependence is illustrated by the multiple state-specific results.

income, use more distinct devices to access Facebook, and donate more often to fundraising campaigns on Facebook (Fig. 2), consistent with previous findings that richer people donate more often (35). They similarly use more expensive mobile phones and have visited more countries in short trips (Fig. S6).

By using a network weighted by number of interactions, we examine the interplay between structural diversity and tie strength. Weak ties, which connect people who infrequently interact with one another, are often described as providing access to more novel — and thus incrementally valuable — information than strong ones (9). However, while weak ties tend to be long ties in general, there is substantial variation, and even some very long ties spanning distant parts of a network can be strong (36). This motivates an explicit distinction between tie strength and tie length (10, 18, 37). So as a second measure of structural diversity, in addition to the fraction of long ties, we compute a tie-strength weighted fraction of long ties; see *Materials and Methods* for details. When comparing individuals with similar fraction of long ties, those with stronger long ties (i.e. higher values of weighted fraction of long ties) live in higher income zip codes, use more unique devices, and make more donations (Fig. 2B). Similar results hold for additional proxies (Fig. S6B), and at a geographically aggregated level for median income in the US (Fig. 1C) and other economic outcomes in the US and Mexico (Fig. S4).

To jointly examine the contribution of the fraction and weighted fraction of long ties, we fit random forests (38) that use these fractions and other covariates to predict each of the economic outcomes. The average marginal change in the predicted outcomes from increasing the fraction and weighted

Fig. 3. Estimated average marginal change in measures of individual economic prosperity given a 1 percentage point (0.01) increase in the fraction or weighted fraction of long ties. These are computed from random forest models fit to each state and DC and including degree, average degree of contacts, daily login activity of the user, average daily login activity of their contacts, average weight of user's edges, age, and gender. Bars are 95% confidence intervals that are cluster-robust at the state level.

fraction of long ties are positive and are of similar magnitude to each other (Fig. 3) where averages are computed over US states, covariates, and the distribution of the long tie measures. These results are quite consistent in direction when considering US states individually (Fig. S7) suggesting that having long and strong ties provides a robust signal for predicting economic prosperity.

Taken together, we find evidence that people and places with more long ties and stronger long ties are more economically prosperous. Having stronger long ties, rather than weak ties *per se*, predicts economic prosperity.

Life Events and Long Ties

Having established the existence of a robust cross-sectional relationship between the prevalence of long ties and economic prosperity, we turn our focus to processes responsible for their formation. We study three life events, each involving mobility between institutions and/or places, that we hypothesized would enable individuals to build connections to disparate communities: migrating to a new state, attending an out-of-state college, and transferring to a different high school. In all three cases, we compare the structural diversity of the network — many years later — surrounding individuals who experienced such events with those who did not.

Interstate migration. To study interstate migration and structural diversity, we examine Facebook users born in the US and who have lived in the same state since 2012. We compare the fraction of long ties for individuals who were born in a different state from the one they currently live to that for individuals living in the state where they were born, matching on degree, age, gender, and home county income (see *Materials and Methods* for details). Interstate migrants have on average 13% more long ties relative to non-migrants (33.7% vs. 29.7% in absolute percentage terms of all their contacts with difference-in-means $p < 10^{-10}$ using cluster-robust standard errors at the county level). While this difference is in part driven by migrants having accumulated long ties in their previous state, we also find that migrants have 4.0% more long ties with residents of their current state relative to non-migrants (28.7% vs. 27.6% in absolute percentage terms with difference-in-means $p = 0.011$ using county-level

Fig. 4. Conditional on degree, people who (A) are interstate migrants or (B) attended multiple high schools have more long ties than matched controls. Qualitatively similar relationships hold when restricting the analysis to less-directly implicated ties — those within the current state (A inset) and those outside of high school (B inset). All estimates are post-stratified by gender, age, and hometown county income bins.

cluster-robust standard errors). To account for differences in ego network sizes, we compare migrants' propensity to form long ties with that of non-migrants who have a similar degree. Migrants consistently have more long ties than non-migrants with a similar degree (Fig. 4A). This difference in structural diversity between migrants and non-migrants increases with the number of contacts and persists when we only consider long ties to contacts living in their current state. Thus, interstate migrants form long ties at a higher rate than similar non-migrants independently of their number of contacts, and even in the state where they currently live.

Out-of-state college. To analyse attending an out-of-state college or university, we study Facebook users born in the US and who attended college outside of their home state, and compare them with users of the same cohort who attended college within their home state. Attending an out-of-state college is associated with having a higher fraction of long ties years after attending college, and the effect persists even when we focus on connections to friends made outside of college and when we match on whether they are currently interstate migrants (Fig. S12). See SI Section D.4 for more detailed analysis of this life event.

Transferring high schools. Adolescence involves learning how to expand one's social network (39), and therefore disruptive experiences during this period are likely to shape an individual's network structure later in life (40). We focus on Facebook users who were born between 1961 and 1991 in the US, went to high school, and are currently living in the US. We then compare the ego networks of users who report attending multiple high schools within the same state with those who attended a single high school (thus ruling out the effect of interstate relocation), while matching on degree, age, gender, and income of their home county (see *Materials and Methods* for details). People who transferred to a different high school on average have 10% more long ties relative to those who attended a single high school (35.2% vs. 31.9% in absolute percentage terms of all their contacts with difference-in-means $p < 10^{-10}$ using cluster-robust standard errors at the county level). Some of this difference could be due to an increase in the immediate opportunity of meeting new friends. We therefore conduct an analysis restricted to long ties formed with contacts outside of high school, finding a fraction of long ties for users who transferred to a different high school that remains 9% larger relative to those who

Fig. 5. People who attended a high school during its closure have more long ties than those who attended the same high school prior to closure. The estimates are post-stratified by school and age. Bars are bootstrap 95% confidence intervals that are cluster-robust at the school level.

did not transfer (34.8% vs. 31.8% in absolute terms with $p < 10^{-10}$ using county-level cluster-robust standard errors). Users who attended multiple high schools consistently form more long ties across all values of the ego network degree (Fig. 4B), even when we only consider contacts outside of high school.

Even conditional on covariates, any effects of transferring high schools (like the other two life events) on network structure may be confounded with other unobserved variables. Here we consider some plausibly exogenous variation in transferring high schools by identifying high schools that closed prior to 2016. Among attendees of these high schools around the time of the closure, later cohorts report that they have attended multiple high schools 20% more relative to earlier cohorts (17.3% vs 14.3% in absolute percentage terms); both fractions are higher, likely reflecting various factors affecting transfers prior to outright closure. Within this restricted sample, and matching on school and age, we find people in the closure-affected cohorts have 6.1% more long ties relative to before-closure cohorts (46.4% vs. 43.7% in absolute percentage terms, Fig. 5). In a linear model with fixed effects for school and age, we estimate that being in a closure-affected cohort increases the fraction of long ties by 1.2 percentage points (95% CI [0.4, 1.9], Table S1). Similar results hold when considering only the ties made outside high school. Together these findings suggest that transferring to a different high school shapes the proclivity to connect with different communities.

Finally, in *SI* Section D.6, we provide some evidence that individuals with each of these life events above are also associated with better economic outcomes measured with the same proxy variables as in prior analyses (Figs. S14, S15, and S16).

Discussion

There is an extensive literature on the role played by social network structures in constituting social capital (2, 34, 41–43). A long tie is potentially a source of information not as directly available through any other contact, therefore individuals with structurally diverse networks ought to have an informational advantage (10, 11, 15–18, 27). Our study documents a strong cross-sectional relationship between structural diversity and measures of economic prosperity at both geographically aggregated and individual levels. The literature has sometimes equated weak ties and long ties, for example, by treating number of mutual contacts as a measure of tie strength (44–46), such that long ties are all

counted as “weak”. Our measure of tie strength allows us to clarify that it is long ties and stronger long ties (rather than weak ties per se) that are particularly associated with prosperity, consistent with the mechanisms posited by both Granovetter (9) and subsequent work (10, 18).

Prompted by this link between long ties and economic outcomes, we consider the origins of long ties and factors that enable people to create and maintain these ties. Seminal work by Simmel (47) suggests that individuals with long ties are mediators who can resolve potential conflicts between disconnected communities due to their unique involvement with both communities. Simmel discussed the ability of these mediators as a natural consequence of their structural position in the network, however it is perceivable that certain individuals have a higher tendency to occupy these network positions (Fig S8). The existing literature on the origins of long ties has characterized the personality of individuals with long ties to be entrepreneurial, risk-taking, and thriving on change (28, 29, 48).

The studies mentioned above treat the determinants of long ties as a matter of personality or even genetic heritability (32), rather than a skill or disposition to be acquired, even though many of the personality correlates of long ties can develop through experiences. We examine the link between experiencing major life disruptions — interstate migration, transferring high schools, and attending an out-of-state college, and the formation of these ties. The experience of these major disruptions are similar in the sense that they expose individuals to diverse communities, require them to form new ties with others who are different than their old network and potentially acquire the social skills to do so. We find that individuals who experienced such disruptions have more structurally diverse networks years later, in ways that is not mechanically due to geographic mobility or exposure to a new community during high school or college. These findings suggest that the tendency to form long ties is not only a matter of innate personality but also life experiences. Shaping one’s personality and acquiring the necessary social skills is one potential mechanism behind the effect of these disruptive life experiences.

These results are consistent with findings showing that the benefits of long ties come from *oscillation* between making deep investments and brokering in multiple communities (49). Brokers with network oscillation have developed the skill for successfully moving in and out of groups, much similar to the experience of migration or relocation during high school:

“Experience with change is preparation for change... [Oscillators] can be expected to be flexible in moving between identities ... and so develop the adaptive self-monitoring associated with network brokers. The image that comes to mind is people who grew up in multiple countries, or in families that frequently moved between cities.” (49, p. 387)

Our findings have important limitations. While the analyses of network structure and economic outcomes flexibly adjust for or match on relevant covariates, some of the observed association can be due to unobserved factors. Furthermore, the observed associations likely reflect causal processes in both directions; that is, social networks affect economic outcomes and also economic prosperity affects network formation and maintenance. For example, wealthier people may be more able to form and maintain long ties (by, e.g., traveling, dining out). There may also be important feedback between processes in these two directions if wealth enables forming long ties, which are then advantageous economically, etc. Overall, our results fit with the growing literature on benefits of long ties in specific settings and outcomes, and some of these studies rely on randomized interventions that affect network structure (46). We regard both causal directions as of interest for further study.

While the three disruptive life events we studied were selected because we hypothesized they would increase structural diversity, there can also be confounding factors between groups here. However, having three different types of events — involving different processes by which people are selected into the disruption — pointing to the same conclusions makes for a more robust and notable pattern. In addition, as we examine the network interactions of an individual many years after experiencing a disruptive event, we rule out endogeneities occurring on a short time scale. Finally, we argue that transferring high schools, in particular due to high school closure, is the most plausibly exogenous event.

Transferring high school is often either a decision made by parents on behalf of students, or results from external factors outside a student’s control (50–52). It may even result from their current school closing, which we also found is associated with forming more long ties. Taken together, we believe that our findings on the role of disruptive life events are unlikely to be fully attributed to confounding.

Understanding the origins and consequences of long ties can have important implications for public policy. Intergenerational economic mobility has been declining in the United States (53), but long ties (Figs. S4 and S5) and ties spanning class boundaries are associated with higher economic mobility (34). Some of those inter-class ties are likely attributable to the design of institutions, as, e.g., smaller high schools have more such ties (43). More generally, we posit that dispositions to form long ties — developed in part through the kind of disruptive life events studied here — are upstream of many ties that cut across groups, including inter-class ties. Policies aimed at fostering structural diversity could have a positive impact on social mobility and economic prosperity more generally. For these reasons, we hope that this study will encourage further research into the mechanisms of formation of long ties, particularly as it pertains to skills that can be taught, and their causal links with economic prosperity.

Materials and Methods

First, we describe the data and methods for examining the link between long ties and economic outcomes at the individual level, with additional notes for the geographically-aggregated level. We then explain the data collection and analysis on the determinants of long ties.

Individual-Level Long Ties and Economic Outcomes.

Ego Networks. For each user currently residing in the US, we construct their ego network from public commenting communication data spanning a period of 6 months from December 2020 to June 2021. An ego network includes all the edges between the user (the ego) and their contacts (alters) and the edges among contacts themselves. In the commenting network, there is a directed edge between two users if one has mentioned or replied to a post made by the other. In constructing the ego networks, we only retain reciprocal edges (with at least one comment exchanged in each direction) to exclude one-off interactions between individuals who would be otherwise unrelated. Furthermore, non-reciprocal edges have been shown to be less important in diffusion and peer influence processes in networks (54, 55). To ensure a minimal level of activity by each user, we discard those with degree less than 10 (have communicated with less than 10 people over the 6 month period). Commenting activities are more appropriate for the construction of the network than the actual friendships on Facebook for two reasons: commenting contacts are more likely to be strong and relevant to each user, as opposed to many friends on Facebook who are only acquaintances without any meaningful interaction over a long period. But more importantly, Facebook users extensively use commenting as a form of communication as there are about 195 million daily active users on the Facebook website or its mobile application in US and Canada*.

Measuring Long Ties. Structural diversity is a purely topological measure and captures the unique number of communities a node is involved in (10). It can be operationalized in different ways, but at its core it relies on measuring the extent to which contacts are clustered together. The fraction of long ties, those without any mutual contacts, is a simple such measure. Long ties are sometimes also called local bridges and have edge embeddedness 0 in the ego network. The long ties closely resemble the widely used concept of structural holes or social bridges and their overall prevalence indicates the extent of novel information received by users residing inside the zip code.

*These numbers are from a recent earnings report: <https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Second-Quarter-2021-Results/>.

Given an ego network, we compute the fraction of long ties among all ties from the ego to their contacts:

$$l_i = \frac{\sum_{j \in N_i} I(\text{mc}_{ij} = 0)}{|N_i|}, \quad [1]$$

N_i is the set of i 's neighbors, $I(\cdot)$ is the indicator function, and mc_{ij} is the number of mutual contacts between i and j .

We also consider what fraction of total interaction by an individual occurs along long ties. The normalized count of exchanged comments is our measure of (normalized) tie strength that is comparable across users with different levels of activity:

$$w_{ij} = \frac{t_{ij}}{t_i}, \quad [2]$$

where t_{ij} is the number of comments i has exchanged with j in the 6 month measurement period and $t_i = \sum_{j \in N_i} t_{ij}$ is the total number of comments user i has exchanged with all their contacts. Similar definitions of tie strength are widely used [e.g., (56, 57)], and are known to correlate with other survey-based measures of tie strength (58, 59).

Given this definition of tie strength, a simple and intuitive measure for the strength of long ties is their weighted sum. The analysis on the strength of long ties uses the weighted fraction of long ties in which each long tie is weighted by its normalized tie strength defined in Equation 2:

$$l_i^w = \sum_{j \in N_i} w_{ij} I(\text{mc}_{ij} = 0). \quad [3]$$

Both l_i and l_i^w can range from 0 to 1.

Economic Measures. Direct measures of economic well-being are not available at the individual level linked to Facebook users. However, various proxy measures that are related to economic outcomes are available at the individual level. Our analysis relies on the following three proxy outcome variables in the main text:

1. The median and mean household income of user's residential zip code: The income value is obtained from 2018 ACS and the user zip code of residence is based on a prediction that incorporates various pieces of information including internet connection. Note that this is not strictly a individual-level outcome.
2. Unique devices: A user connection to Facebook logs their device model. We count the number of unique devices a user has used more than once to connect to Facebook from 2019 to 2021.
3. Number of donations to fundraisers: Facebook has a personal fundraising platform where each user can request for financial assistance either for personal reasons or to a charity organization. Since 2016 when fundraisers launched to 2021, about \$5 billion dollars have been donated through Facebook fundraising platform, involving about 85 million people[†]. We count the number of times a user has made a donation to a fundraising campaign over a 2-year period from June 2019 to June 2021. We expect that users with a higher disposable income tend to make more donations.

In the *SI*, we provide further evidence of this relationship by examining the following two proxy outcomes:

1. The user's phone price: A user connection to Facebook logs their mobile phone model. We match their phone model against a database of phone prices in dollars.
2. The user's international touristic trips: The number of countries, other than Canada, Mexico and Puerto Rico, visited by a user in trips shorter than two weeks since 2012 captures the user's tourism activity. A visit to a country is inferred based on the connection IP of the user while on a trip.

[†] www.facebook.com/business/news/coming-together-to-raise-5-billion-on-facebook-and-instagram-

Matching and Estimation. We compare individuals with different values of the focal structural diversity measures but similar values of other covariates. We discretize the fraction (or weighted) long ties into 12 equal-sized levels. Generalized full matching (60) is an appropriate matching framework here as it allows for more than two treatment arms (12 in our case), assigns all units to exactly one matching group, and is computationally tractable. Fig. 2 shows the difference between each level of fraction of long ties or weighted long ties and the fourth bin which is used as the base level. The matching is done once per each state, rather than the whole country, which uses only the data from users whose predicted current residence is in that state. This allows us to discover potential heterogeneities that might exist between the states.

The Euclidean distance metric in the matching procedure is studentized so that different variables with different ranges contribute equally to the distance metric. It is constructed based on the following variables when the independent variable is the fraction of long ties (top row of Fig. 2):

1. Log user degree and log average degree of their contacts. These two variables ensure comparisons are made between ego networks of similar sizes.
2. Daily login activity of the user, measured in terms of the number of days over a month with a login, and the average daily login activity of their contacts. These two variables ensure the ego and alters have similar levels of activity which might affect the fraction of long ties in the ego network.
3. Age and gender of the user.

When the independent variable is weighted long ties (bottom row of Fig. 2), the distance metric incorporates the following additional variable:

1. Log average weight of user’s edges ($\log(\frac{t_i}{|N_i|})$). This ensures that matched users have similar levels of commenting activity since, as this affects the feasible ranges of weighted long ties l_i^w
2. Fraction of long ties l_i .

Predictive Model. Section B.1 in SI reports on a different estimation approach based on the average marginal effect of fraction of or weighted long ties in a random forest model. Fig. 3 reports summaries of these models fit to data from each state. The results of this alternative procedure are consistent with the matching results presented above.

Geographically-Aggregated Long Ties and Economic Outcomes.

Data on Network Interactions. These analyses use the same communication network as in the individual-level analyses. For each Facebook user, we also observe the predicted city of residence which relies on the user’s information and activity, for example their self-reported city on their profile and their internet connection information. These residential predictions allow us to match users to zip codes, counties, and states and construct the corresponding zip code network, consisting of users residing in the zip code.

Construction of Geographically-Aggregated Networks. The nodes in each zip code network consists of all users who reside in that zip code and all their contacts. (Throughout we focus on zip codes, though the maps in Figs. 1A and S3 use aggregation to counties with otherwise identical logic.) As with the individual-level analyses, we only consider reciprocal ties to filter out one-off communication events that don’t signify a meaningful relationship. Thus, the existence of an edge in the zip code network means there has been at least one communication event in each direction. As opposed to the individual-level results which used directed networks, edges in the aggregate zip code networks are undirected with weights that correspond to the total number of communication events in either direction from December 2020 to June 2021. Fig. S1 shows a schematic view of the zip code networks. In our analysis, we only consider zip codes with at least 150 users who reside in that zip code and for which we have census data available. This leaves us with 26,429 zip codes in the US (excluding PO Boxes or businesses) and 15,157 zip codes in Mexico for which we have both network and economic indicator data. Fig. S2 shows

the histogram of some basic characteristics of these networks in the US, along with a few economic indicators.

Measuring Long Ties in Geographically-Aggregated Networks. We use the fraction of long ties among all edges that involve at least one node inside the zip code as our primary measure of zip code structural diversity:

$$l_z = \frac{\sum_{i \in Z} \sum_{j \in N_i} I(\text{mc}_{ij} = 0)}{\sum_{i \in Z} |N_i|} \quad [4]$$

where Z is the set of nodes residing inside the zip code. These long ties are shown with dotted black lines in Fig. S1.

As with the individual-level analysis, we also define a version of this measure weighted by tie strength:

$$l_z^w = \frac{\sum_{i \in Z} \sum_{j \in N_i} w_{ij} I(\text{mc}_{ij} = 0)}{|Z|} \quad [5]$$

Since the measure of tie strength is normalized, l_z^w ranges between 0 and 1 and can be treated as *weighted fraction of long ties*. Figs. S4 and S5 compare both l_z and l_z^w measures against zip code outcomes.

Data on Economic Indicators. In the US, the zip code outcome variables consists of a set of economic indicators obtained from the census 5-year estimates of the 2018 American Community Survey (ACS) at the level of zip code tabulation areas (ZCTA). The economic indicators we consider here are median household income, fraction of households with income below \$25K, and the unemployment rate. We also consider measures of social mobility provided by the Atlas of Opportunity project at the ZCTA level (61):

1. the probability of a child born in the zip code and from the bottom 25% of the income distribution in the US reaching the top 20% of the income distribution and
2. the percentile income rank of a child born in the zip code from the bottom 25% of the income distribution.

Finally, as a robustness check we replicate these analyses in Mexico as a second country. Due to lack of income data at the zip code level, we constructed a wealth index for each zip code using the first principal component of the following variables from the 2020 census: Percentage of illiterate population aged 15 or more, population aged 6-14 not attending to school, population aged 15 or more without complete elementary school, population with no access to health services, log of average occupants per room, and percentage of households with dirt floor, no toilet, no piped water, no sewer system, no electricity, no washing machine, or no fridge. The wealth index for each zip code is the value of the first component re-scaled to between 0 and 1. Effectively, our outcome indicator in Mexico is a summary measure on the intensity of economic, health, and education prosperity.

Covariates. There is a lot of variation in the size of zip code networks in ways that can influence the range of reasonable values for l_z and affect its statistical relationship with the outcomes. Thus, we ensure to control for the following variables in our analysis: population of the zip code, number of network nodes residing in the zip code and mean number of edges from a node inside the zip code (mean degree). All the economic indicators we consider are highly correlated with racial composition of the zip code. Thus to ensure our network measure explain any outcome variation above and beyond racial composition, we also control for the racial composition (percent white) in each zip code. The data from Mexico does not include zip code-level racial composition, so the model for Mexico simply controls for the network covariates mentioned above.

Our second goal is to establish the relationship between the strength of long ties and the economic outcomes given a fixed number of long ties. Since l_z^w is highly correlated with the unweighted measure of structural diversity, l_z , our analysis must control for it. Thus, our second analysis controls for the fraction of long ties, l_w , in addition to those mentioned above.

Model and Estimation. To enable discovering potential nonlinearity in relationships between the zip code outcome and network structure, we use a modern binscatter method, binsreg (62). In addition to visualizing the data with cluster-robust confidence intervals at each bin, binscatter provides a principled way to test for shape restrictions and most importantly adjust for our control variables such that it does not suffer from problems of residual-based approaches (63).

Binscatter regression with covariate adjustment fits the data according to a linear model on the covariates. In order to avoid any assumption on the functional form of the covariates, they are binned and the model includes an indicator for each bin. For example, the model on the relationship between the a zip code outcome and the fraction of long ties is:

$$\begin{aligned} y_z &= \mu(l_z) + \alpha p_z + \beta n_z d_z + \theta r_z + \epsilon_z \\ E[\epsilon_z | l_z, p_z, n_z, d_z, r_z] &= 0 \end{aligned} \quad [6]$$

where y_z is an outcome measure (e.g. zip code mean household income), l_z is the fraction of long ties, p_z is the zip code population bin, n_z is the binned number of nodes inside the zip code, d_z is the binned average degree of nodes inside the zip code, and r_z is the percent white in the zip codes. The model attempts to make comparisons between networks of similar sizes by controlling for the combination of n_z and d_z , such that d_z bins are determined conditional or within the n_z bin. Binscatter regression reveals the shape of function $\mu(\cdot)$, and computes the standard errors of each bin. Overall the model above includes 87 indicators as covariates.

We employ a similar approach for the model on the relationship between zip code outcomes and the strength of long ties, with the additional controls for the fraction of long ties as formulated below.

$$\begin{aligned} y_z &= \mu'(l_z^w) + \delta' l_z + \alpha' p_z + \beta' n_z d_z + \theta' r_z + \epsilon'_z \\ E[\epsilon'_z | l_z^w, l_z, p_z, n_z, d_z, r_z] &= 0 \end{aligned} \quad [7]$$

where l_z^w is the weighted fraction of long ties, l_z is the binned unweighted fraction of long ties. The model above include 94 indicators as covariates.

Life Events and Long Ties.

Data on Ego Networks. The same network described in the previous sections are used to measure the fraction of long ties.

Data on Inter-State Migration. The current state of each user can be inferred based on Facebook’s internal prediction of residential location (based on activity and self-reported location information on their profile) as also used above in construction of zip-code-level networks. Given this data, we construct a monthly panel of predicted state location for each user for 9 years starting from January 2012. To determine whether a user has experienced inter-state migration, we also need information on their hometown state, for which we rely on the optionally listed hometown on the user’s Facebook profile and ensure it can be matched against a known location. Overall, 72% of the active users in the data have provided this self-reported information on their profile. Given the monthly residence and hometown states, we restrict our attention to users with a hometown city in the US, aged between 30 and 60 who have not migrated since 2012 (i.e. their monthly residence state has remained the same since January 2012). We then divide these users to migrants or those with a different hometown state than their current state and non-migrants with the same current and hometown states. Therefore in our analysis, migrants are users who moved to their current state prior to 2012 and have lived in their current state continuously since 2012. Similarly, non-migrants are users who have continued residing in their home state since 2012. The final data consists of 20.6 million users, out of which 16.7 million users (81%) are local or non-migrant and 3.9 million users (19%) are migrant.

Data on High School Attendance. We obtain high school attendance data on each user based on the self-reported high school names and pages on their profiles. Multiple pages referring to the same high school by different users are consolidated to a single high school and then matched against the database of high schools in the US provided by the National Center for Education Statistics (NCES) to ensure the validity of these self-reported schools. About 65% of the users we examined have reported their high school information. Our analysis is based on users aged between 30 and 60 who currently live in the US, have indicated their hometown city in the US, and have attended their high schools in the US and all within a single state. We divide this population of users to two groups of single or multiple high schools and join each user with their ego network constructed from communication data. The final population consists of 24.4 million users, out of which 23.8 million (97%) attended a single high school and 673 thousand (3%) attended multiple high schools all within a single state.

Analysis. Our goal is to compare the average fraction of long ties over the population of two groups: migrants vs. locals, out-of-state vs. in-state college attendees, or multiple vs. single high school attendees. A difference in means within each degree bin, however, does not account for differences between the two populations. In our analysis, we account for observable differences using post-stratification (i.e. many-to-many matching). In our post-stratification, the combination of binned age (3 bins), gender (2 bins), and hometown county income (15 bins) constitute a stratum (a total of 90 strata). Figs. 4A, 4B, and S12 show these post-stratified means within each degree bin (with a total of 30 equal-size degree bins). We perform the post-stratification conditional on degree for two reasons: first, to show how the range of common values for the fraction of long ties varies by degree and second, to show that the difference between two groups (migrants vs. locals or single vs. multiple high schools attendance) is consistent across the whole range of ego degrees.

While migrant and local populations have the similar degree and gender distribution, they vary significantly by their age, mean household income of where they grew up and current state (Fig. S9), hence the need for stratification. Similarly, users who have attended multiple high schools tend to have slightly higher degree and be of younger age than the single high school users (Fig. S10). Our analysis does not include migration status in the strata definition for two reasons: first, as we see in Fig. S10, there is no notable difference between the two high school groups in terms of migration. Second, our analysis requires that neither group experiences an inter-state migration event during high school.

High School Closures and Long Ties.

Data. The National Council for Education Statistics provides data on high schools that closed in each year from 1995 to 2016. The analysis on high school closures involves Facebook users who had reported attending any of these closed high schools and have provided the exact time period of their attendance. The users whose reported end date at the school is within one year of its closure are considered to be in the “During Closure” condition and those whose reported end date at the school is between 2 and 5 years prior to its closure constitute the “Before Closure” condition. The final data consists of 33,813 users who had attended 285 closed high schools, out of which 11,309 (22,504) users attended the high schools during (before) their closure.

Analysis. The main result reported above is based on table S1 which uses a two-way fixed effect model with both age in whole years and schools as fixed effects and cluster robust standard errors at the school level. The average fractions of long ties in Fig. 5 are estimated using post-stratification with the combination of school and age in 4-year bins as the strata.

ACKNOWLEDGMENTS. We thank A. Pentland, M. O. Jackson, D. G. Rand, R. Reagans, C. Roth, and seminar participants at University of Michigan, University of Pennsylvania, Stanford University, and the Massachusetts Institute of Technology for their helpful comments. We specially appreciate Linxi Wang for her help in our analysis. E.J. was supported by a grant to D.E. from the World Bank. S.F. was supported by the SDG Partnership Fund.

1. M Granovetter, Economic action and social structure: The problem of embeddedness. *Am. J. Sociol.* **91**, 481–510 (1985).
2. MO Jackson, An overview of social networks and economic applications. *Handb. social economics* **1**, 511–585 (2011).
3. C García-Jimeno, A Iglesias, P Yildirim, Information networks and collective action: Evidence from the women's temperance crusade. *Am. Econ. Rev.* **112**, 41–80 (2022).
4. A Banerjee, AG Chandrasekhar, E Duflo, MO Jackson, The diffusion of microfinance. *Science* **341** (2013).
5. Y Zhang, L Wang, JJ Zhu, X Wang, AS Pentland, The strength of structural diversity in online social networks. *Research* **2021** (2021).
6. N Eagle, M Macy, R Claxton, Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
7. S Luo, F Morone, C Sarraute, M Travizano, HA Makse, Inferring Personal Economic Status from Social Network Location. *Nat. Commun.* **8** (2017).
8. M Bailey, R Cao, T Kuchler, J Stroebel, A Wong, Social connectedness: Measurement, determinants, and effects. *J. Econ. Perspectives* **32**, 259–80 (2018).
9. M Granovetter, The strength of weak ties. *J. Sociol.* **78**, 1360–1380 (1973).
10. RS Burt, *Structural holes: The social structure of competition*. (Harvard University Press, Cambridge, MA), (1992).
11. S Rodan, Structural holes and managerial performance: Identifying the underlying mechanisms. *Soc. Networks* **32**, 168–179 (2010).
12. SP Borgatti, Structural holes: Unpacking Burt's redundancy measures. *Connections* **20**, 35–38 (1997).
13. J Ugander, L Backstrom, C Marlow, J Kleinberg, Structural diversity in social contagion. *Proc. Natl. Acad. Sci.* **109**, 5962–5966 (2012).
14. J Su, K Kamath, A Sharma, J Ugander, S Goel, An experimental study of structural diversity in social networks. *Proc. Int. AAAI Conf. on Web Soc. Media* **14**, 661–670 (2020).
15. R Reagans, EW Zuckerman, Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organ. Sci.* **12**, 502–517 (2001).
16. MT Hansen, The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Adm. Sci. Q.* **44**, 82–111 (1999).
17. R Reagans, B McEvily, Network structure and knowledge transfer: The effects of cohesion and range. *Adm. Sci. Q.* **48**, 240–267 (2003).
18. S Aral, M Van Alstyne, The diversity-bandwidth trade-off. *Am. J. Sociol.* **117**, 90–171 (2011).
19. V Yakubovich, Weak ties, information, and influence: How workers find jobs in a local Russian labor market. *Am. sociological review* **70**, 408–421 (2005).
20. EJ Castilla, GJ Lan, BA Rissing, Social networks and employment: Mechanisms (part 1). *Sociol. Compass* **7**, 999–1012 (2013).
21. A Mehra, AL Dixon, DJ Brass, B Robertson, The social network ties of group leaders: Implications for group performance and leader reputation. *Organ. Sci.* **17**, 64–79 (2006).
22. D Obstfeld, Social networks, the tertius iungens orientation, and involvement in innovation. *Adm. Sci. Q.* **50**, 100–130 (2005).
23. M de Vaan, D Stark, B Vedres, Game changer: The topology of creativity. *Am. J. Sociol.* **120**, 1144–1194 (2015).
24. I Kloumann, L Adamic, J Kleinberg, S Wu, The lifecycles of apps in a social ecosystem in *Proceedings of the 24th International Conference on World Wide Web*. (IW3C2), pp. 581–591 (2015).
25. S Aral, D Walker, Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Manag. Sci.* **60**, 1352–1370 (2014).
26. M Bailey, DM Johnston, T Kuchler, J Stroebel, A Wong, Peer effects in product adoption, (National Bureau of Economic Research), Technical report (2019).
27. RS Burt, et al., Structural Holes and Good Ideas. *Am. J. Sociol.* **110**, 349–99 (2004).
28. RS Burt, JE Jannotta, JT Mahoney, Personality correlates of structural holes. *Soc. Networks* **20**, 63–87 (1998).
29. Y Kalish, G Robins, Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. *Soc. Networks* **28**, 56–84 (2006).
30. Y Kalish, Bridging in social networks: Who are the people in structural holes and why are they there? *Asian J. Soc. Psychol.* **11**, 53–66 (2008).
31. J Specht, B Egloff, SC Schmukle, Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *J Pers Soc Psychol.* **101**, 862–82 (2011).
32. JH Fowler, CT Dawes, NA Christakis, Model of genetic variation in human social networks. *Proc. Natl. Acad. Sci.* **106**, 1720–1724 (2009).
33. M Neugart, S Yildirim, Heritability in friendship networks. *J. Econ. Behav. & Organ.* **194**, 41–55 (2022).
34. R Chetty, et al., Social capital I: Measurement and associations with economic mobility. *Nature* **608**, 108–121 (2022).
35. P Wiepking, R Bekkers, Who gives? A literature review of predictors of charitable giving. Part Two: Gender, family composition and income. *Volunt. Sect. Rev.* **3**, 217–245 (2012).

36. PS Park, JE Blumenstock, MW Macy, The strength of long-range ties in population-scale social networks. *Science* **362**, 1410–1413 (2018).
37. D Lyu, Y Yuan, L Wang, X Wang, A Pentland, Investigating and modeling the dynamics of long ties. *Commun. Phys.* **5**, 1–9 (2022).
38. L Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
39. J Saramäki, et al., Persistence of social signatures in human communication. *Proc. Natl. Acad. Sci.* **111**, 942–947 (2014).
40. C Wrzus, M Hänel, J Wagner, FJ Neyer, Social network changes and life events across the life span: a meta-analysis. *Psychol. bulletin* **139**, 53 (2013).
41. MO Jackson, BW Rogers, Y Zenou, The economic consequences of social-network structure. *J. Econ. Lit.* **55**, 49–95 (2017).
42. M Granovetter, The Impact of Social Structure on Economic Outcomes Social Networks and Economic Outcomes: Core Principles. *The J. Econ. Perspectives* **19**, 33–50 (2005).
43. R Chetty, et al., Social capital II: Determinants of economic connectedness. *Nature* **608**, 122–134 (2022).
44. LK Gee, J Jones, M Burke, Social networks and labor markets: How strong ties relate to job finding on Facebook's social network. *J. Labor Econ.* **35**, 485–518 (2017).
45. LK Gee, JJ Jones, CJ Fariss, M Burke, JH Fowler, The paradox of weak ties in 55 countries. *J. Econ. Behav. & Organ.* **133**, 362–372 (2017).
46. K Rajkumar, G Saint-Jacques, I Bojinov, E Brynjolfsson, S Aral, A causal test of the strength of weak ties. *Science* **377**, 1304–1310 (2022).
47. G Simmel, *The Sociology of Georg Simmel*. (Simon and Schuster) Vol. 92892, (1950).
48. P Totterdell, D Holman, A Hukin, Social networkers: Measuring and examining individual differences in propensity to connect with others. *Soc. Networks* **30**, 283–296 (2008).
49. RS Burt, J Merluzzi, Network oscillation. *Acad. Manag. Discov.* **2**, 368–391 (2016).
50. RO Welsh, School hopscotch: A comprehensive review of K–12 student mobility in the United States. *Rev. Educ. Res.* **87**, 475–511 (2017).
51. RW Rumberger, GJ Palardy, Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *Am. educational research journal* **42**, 3–42 (2005).
52. Q Brummet, The effect of school closings on student achievement. *J. Public Econ.* **119**, 108–124 (2014).
53. R Chetty, N Hendren, P Kline, E Saez, Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Q. J. Econ.* **129**, 1553–1623 (2014).
54. A Almaatouq, L Radaelli, A Pentland, E Shmueli, Are you your friends' friend? poor perception of friendship ties limits the ability to promote behavioral change. *PLOS ONE* **11**, 1–13 (2016).
55. YX Zhu, et al., Influence of reciprocal links in social networks. *PLOS ONE* **9**, 1–8 (2014).
56. E Bakshy, D Eckles, R Yan, I Rosenn, Social influence in social advertising: Evidence from field experiments in *Proceedings of the 13th ACM Conference on Electronic Commerce*. pp. 146–161 (2012).
57. RM Bond, et al., A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).
58. JJ Jones, et al., Inferring tie strength from online directed behavior. *PLOS One* **8**, e52168 (2013).
59. M Burke, RE Kraut, The relationship between Facebook use and well-being depends on communication type and tie strength. *J. computer-mediated communication* **21**, 265–281 (2016).
60. F Sävje, MJ Higgins, JS Sekhon, Generalized full matching. *Polit. Analysis* **29**, 423–447 (2021).
61. R Chetty, JN Friedman, N Hendren, MR Jones, SR Porter, The opportunity atlas: Mapping the childhood roots of social mobility, (National Bureau of Economic Research), Working Paper 25147 (2018).
62. MD Cattaneo, RK Crump, MH Farrell, Y Feng, Binscatter regressions (2021).
63. MD Cattaneo, RK Crump, MH Farrell, Y Feng, On binscatter (2021).
64. DW Apley, J Zhu, Visualizing the effects of predictor variables in black box supervised learning models. *J. Royal Stat. Soc. Ser. B (Statistical Methodol.)* **82**, 1059–1086 (2020).

Supplementary Information for “Long ties, disruptive life events and economic prosperity”

Contents

A Geographically-Aggregated Associations	17
B Individual-level Associations	18
B.1 Average Marginal Effects with Random Forests	19
C Demographics of Long Ties	21
D Life Events and Long Ties	22
D.1 Inter-State Migration	22
D.2 Switching High Schools	22
D.3 High School Closure	22
D.4 Out-of-state College Attendance	25
D.4.1 Data and Methods	25
D.4.2 Results	26
D.5 Self-Reporting as Potential Confounder	27
D.6 Associated Economic Outcomes	27

A. Geographically-Aggregated Associations

We use each zip code in the US or Mexico as a unit of analysis and construct a network for each zip code corresponding to the communication patterns of its residents. Fig. S1 illustrates how each zip code network is constructed. It includes all the users residing in a zip code and all their contacts as the nodes and all the ties among them. Fig. S2 shows the histogram of some basic characteristics of the zip code networks in the US, along with a few economic indicators obtained from census. The main approach is to examine the relationship between the overall network structure of each zip code and an array of economic indicators on that zip code provided by the census bureau. In the main text, we present our results on the link between structural diversity, measured as the fraction of long ties, and three important economic outcomes. In particular, the independent variable is the fraction of long ties: the fraction of dotted black edges over all black edges, where dotted pattern indicates a long tie and black color indicates a tie involving at least one user residing in the zip code. The second histogram in second row of Fig. S2 indicates that the majority of edges that are adjacent to a blue node in the zip code are to red nodes outside the zip code. Even a larger majority of long edges that are adjacent to nodes in the zip code are to nodes outside the zip code (histogram not shown). Our secondary analysis examined the strength of long ties which is the sum of edge weights on dotted black edges normalized by the total edge weight on black edges. As explained in the main text, the edge weight incorporates the number of exchanges in either direction.

Figs. S4 and S5 present the relationship of the fraction of long ties and the weighted long ties with zip code outcome variables other the median income presented in the main text. The binned regression models used to generate these plots are the same as one described in *Materials and Methods*. The main

Fig. S1. A schematic representation of a network for a single region (zip code). Blue nodes correspond to individuals residing in the zip code, the purple shaded area indicates the zip code boundary, and the red nodes are contacts who reside in other zip codes. Black edges involve at least one node inside the zip code, and gray edges involve two nodes outside the zip code. Dotted lines correspond to long ties involving at least one node inside the zip code, to a contact either inside or outside the zip code. Our primary analysis considers such long ties represented with a dotted pattern.

conclusion from Fig. S4 is that the residents in zip codes with better outcomes tend to have more long ties. Fig. S4 shows how the zip code outcomes change as the long ties in the zip code network become stronger, conditional on the fraction of long ties. These plots suggest that given a fixed number of long ties in a zip code network, the zip code outcomes tend to be better if the long ties are stronger. This finding points to the opportunity and information value of long ties since the stronger a tie is, the more likely valuable resources and information will be transmitted across the tie.

B. Individual-level Associations

In the main text, we showed the relationship between (weighted) fraction of long ties and log median family income of residence zip code, number of unique devices and number of donations. Those results were obtained using (generalized full) matching of ego networks per state. Fig. S6 provides further evidence on the this relationship by using three other proxy variables for economic well-being at the individual level: log mean family income of residence zip code (as opposed to the median in the main text), number of unique countries other than Canada, Mexico and Puerto Rico visited since 2013 in trips shorter than 3 weeks and price of main mobile device used to login to Facebook. These results generally confirm the findings in Fig. 2 that higher fraction of long ties or stronger long ties tend to be associated with better economic proxies. The relationship between strength of long ties and mobile phone price is the notable exception as it exhibits largely a null or inverted U-shape behavior with large standard errors. This could be due to either due to measurement errors on phone price or observations on weighted long ties with extremely large values which might behave differently than the rest of the population. The observations with fraction (weighted) of long ties larger than 0.4 account for small fraction of the population as indicated in Fig. S8 and did not appear in the aggregate results due to averaging at the zip code level.

Examining this relationship per state has multiple advantages over a single country-level relationship. First, it provides insights on the existence of heterogeneity of the relationship, if any between the states. Second, it provides a good understanding on the variability of the estimated effect if there is cluster-level dependence between the units. Third, it is computationally tractable as opposed to a single country-level analysis with tens of millions of data points. The remainder of this section explores the robustness of our findings in the main text by employing a different methodology for estimating the marginal effect of (weighted) fraction of long ties on either outcome.

Fig. S2. Histogram of some basic metrics from US zip code networks, along with their household income and unemployment rate from census, and social mobility measures obtained from Atlas of Opportunity project. Mean degree/total activity refers to average degree of/number of comments sent by nodes inside the zip code. Fraction out-edges is the fraction of total edges that originate from a node inside the zip code to one outside the zip code. % rank income from bottom 25% indicates the percentile income rank of a child born in the bottom 25% income group. Similarly, the probability top 20% from bottom 25% refers to the probability that a child born in the bottom 25% income group can make it to the top 20% income group.

B.1. Average Marginal Effects with Random Forests. As opposed to the matching approach which generates multiple effect sizes, one per a discretized level of long ties, here we estimate a single quantity that captures the average marginal effect of fraction of long ties or weighted long ties on the outcome. In particular when the independent variable is the fraction of long ties, the estimand is the following quantity:

$$\mu_l = E\left[\frac{\partial f(l_i, l_i^w, C_i)}{\partial l_i}\right]. \quad (\text{S8})$$

where l_i is user i fraction of long ties as defined in Equation 1, l_i^w is user i weighted long ties as defined in Equation 3, C_i is the set of other covariates, $f(\cdot)$ is a function that maps the covariates to the outcome value and the expectation is taken over the joint distribution of all covariates. The covariates in C_i are the same as those used to obtain the matching results for weighted long ties in the main text: user degree, average degree of user’s contacts, daily login activity of the user, average daily login activity of their contacts, average weight of user’s edges, age and gender of the user. The quantity in Equation 8 can be interpreted as the average marginal effect of fraction of long ties on the outcome. A similar quantity captures the average marginal effect of weighted long ties on the outcome using the

Fig. S3. Proportion of Facebook ties that are long, measured across across municipalities in Mexico, and after controlling for the average number of Facebook contacts. The adjusted (residualized) fraction of long ties across administrative units in each country is binned into 10 deciles, with lower values shown in purple and higher values shown in green. Administrative units with less than 200 Facebook users sufficiently active over the measurement period or with less than 500 individuals are shown in white.

same function $f(\cdot)$:

$$\mu_l^w = E\left[\frac{\partial f(l_i, l_i^w, C_i)}{\partial l_i^w}\right]. \quad (\text{S9})$$

The estimation of the marginal effects above can use an approach similar to that used for Accumulated Local Effects (64). First, we train the prediction function $f(\cdot)$ using a random forest and obtain $\hat{f}(\cdot)$. Second in order to estimate the local marginal effect, we divide the variable of interest (either l_i or l_i^w) into many small intervals and compute the difference in $\hat{f}(\cdot)$ prediction along each interval end points. This procedure estimates the local derivative or marginal effect. The estimated average marginal effect is the average of such local marginal effects over the empirical distribution of other observed covariates:

$$\hat{\mu}_l = \frac{1}{n} \sum_{k=1}^K \sum_{\{i: l_{k-1} \leq l_i < l_k\}} \left[\frac{\hat{f}(l_k, l_i^w, C_i) - \hat{f}(l_{k-1}, l_i^w, C_i)}{l_k - l_{k-1}} \right] \quad (\text{S10})$$

where k denotes the k^{th} grid interval on l_i or fraction of long ties from l_{k-1} to l_k , K is the total number of intervals and n is the total number of points. The inner sum is the total marginal effect over all the points that fall in the k^{th} interval. A similar estimator also exists for the average marginal effect of weighted long ties:

$$\hat{\mu}_l^w = \frac{1}{n} \sum_{k=1}^K \sum_{\{i: l_{k-1}^w \leq l_i^w < l_k^w\}} \left[\frac{\hat{f}(l_i, l_k^w, C_i) - \hat{f}(l_i, l_{k-1}^w, C_i)}{l_k^w - l_{k-1}^w} \right] \quad (\text{S11})$$

Fig. S7 shows the estimated marginal effects of fraction of long ties and weighted long ties on five individual-level outcomes variables: median family income of residence zip code, number of unique devices used more than once to login to Facebook, number of donations on the Facebook fundraising platform, number of countries visited in potentially troustism trips, and phone price. There is an estimate per each state, obtained solely based on users who live in that state, and an overall estimate

Fig. S4. Fraction of long ties and geographically aggregated economic outcomes in US (A-D) and Mexican (E) zip codes. The binned regression plots are generated according to model (6) in *Materials and Methods*, which adjusts for the (binned) size of the zip code, its racial composition, and average degree in the network. Solid lines and bars correspond to local smoothers of second degree and 95% confidence intervals in each bin respectively. Standard errors are cluster robust at the county level.

which is the average of all state estimates. The marginal effects correspond to 1 percentage point increase (or 0.01) in fraction of or weighted fraction of long ties (l_i and l_i^w). The findings are consistent with the results in Fig. 2 of the main text as the marginal effect of both fraction of long ties and weighted long ties are consistently positive for all five outcome variables suggesting that people with structurally more diverse networks tend to have better outcomes along these variables. Fig. S7 also shows how the marginal effect estimates will change if the model accounts for more covariates. Model 1, represented by blue points, adjusts for the set of covariates C_i mentioned above. Model 2, represented by red points, also adjusts for highest education level (e.g. high school, college, graduate, ...) and median income of home county where user is from. Model 2 provides a basic assessment on the sensitivity of model 1 to potential unobservables. The estimated coefficient remain positive and highly significant in model 2. It should be noted that the standard errors used to construct 95% confidence interval around the overall average marginal effect is cluster robust at the state level since it is the standard error of 51 independent state (including Washington DC) estimates.

C. Demographics of Long Ties

Fig. S8 illustrates the distribution of long ties fraction over individual user ego-networks across various demographic attributes, including gender, age, highest level of education, population density of residence zip code, census region and division. For each attribute, the first plot compares the mean fraction of long ties over different groups and ego-network degree brackets. In contrast to the post-stratified means in Figs. 4A and 4B, per degree bracket means in Fig. S8 are not stratified. The second plot for

Fig. S5. Interaction-count weighted fraction of long ties and geographically aggregated economic outcomes in US (A-D) and Mexican (E) zip codes. The binned regression plots are generated according to model (7) in *Materials and Methods*, which adjusts for the (binned) size of the zip code, its racial composition, average degree in the network, and unweighted fraction of long ties. Solid lines and bars correspond to local smoothers of second degree and 95% confidence intervals respectively constructed using cluster robust standard errors at the county level.

each demographic attribute compares the conditional distribution of long ties fraction per group, such that the columns sum up to 1. The results suggest that younger, higher educated individuals living in more dense urban areas tend to have more long ties than older, lower educated individuals in less dense rural areas. Furthermore, individuals living in the western states tend to have more structurally diverse networks than the rest of the country.

D. Life Events and Long Ties

D.1. Inter-State Migration. In the main text, we show how inter-state migrants in the US have more long ties than locals. Our analysis involves comparing the post-stratified mean of fraction of long ties between the two groups. Fig. S9 illustrates the distribution of the control variables used in post-stratification (gender, age, home county income) in addition to degree and current state in the US by each group.

D.2. Switching High Schools. Fig. S10 compares the distribution of post-stratification variables in addition to degree, inter-state migration status and current state between the population of users with single or multiple high schools attended.

D.3. High School Closure. The main text describes the data collection around closure of high schools. Our analysis compares users who attended a high school during its closure to those who left the high school immediately before its closure. Fig. 5 compares the stratified mean fraction of long ties across the two high school closure conditions where the strata is defined as the combination of high school

Fig. S6. Fraction of long ties (**top row**) and the tie-strength weighted fraction of long ties (**bottom row**) in ego networks and individual level variables that are positively correlated with socio-economic status for users who reside in the top 6 most populated states in the US. The plots are generated based on the procedure in *Materials and Methods* which matches individuals based on their degree, average degree of their contacts, their daily login activity, average daily login activity of their contacts, their age and gender for the top row. The matching for the bottom row also incorporates the fraction of long ties and average edge weight in addition to the parameters mentioned before. Bars correspond to 95% confidence intervals. Values in the y-axis correspond to the difference with the baseline represented with a value of zero.

and age bracket. This design effectively compares similar cohorts from the same school who only differ on the period they attended the school by a maximum of 5 years. The age stratification attempts to account for differences in fraction of long ties that might be due to age. However, this natural experiment itself makes it difficult to stratify on a single year since cohorts in the two conditions will be different in age by design which implies some age strata will not include both closure conditions under a fine grained age stratification. Hence, Fig. 5 uses a coarse age stratification with brackets of 4 years. This method leads to many strata which contain both conditions and retains a large fraction of the data, nevertheless it might falsely attribute any age-related variation in fraction of long ties within an age strata to high school closure since those affected by the closure tend to be younger.

In order to fully account for age variation, we use the following two-way linear fixed effect model:

$$Y_i = \beta I_{i,\text{closure}} + \gamma_{S_i} + \lambda_{A_i} + \epsilon_i \quad (\text{S12})$$

where Y_i is the i^{th} subject fraction of long ties in their ego-network, γ_{S_i} is the school fixed effect of the i^{th} subject's school S_i , λ_{A_i} is the age fixed effect for age in integer years, and $I_{i,\text{closure}}$ is an indicator variable that takes a value of 1 if the i^{th} subject attended the school during its closure (i.e. their self-reported end date is within one year of the school closure). In contrast to Fig. 5 which uses stratification, this linear model does not require both conditions to be present within each school-age combination, thus it allows for fine-grained adjustments for age. The third column in Table S1 shows the estimates from this model where standard errors are cluster-robust at the school level. The other

Fig. S7. The average marginal effect of fraction of long ties (**Top row**) and weighted long ties (**Bottom row**) on median family income of residence zip code (**First column**), number of unique devices used to login to Facebook (**Second column**), number of fundraising donations (**Third column**), number of countries visited in short trips (**Fourth column**), and phone price (**Fifth column**). Model 1 accounts for the baseline covariates discussed in the main text. Model 2 also accounts for education level and home county income. Each dot corresponds to the estimated marginal effect of 1% (0.01) increase in fraction or weighted long ties for residents of a single state. The triangle corresponds to the overall marginal effect as the average of all state-level estimates. There is no estimated standard error for state-level estimates, but the standard error of the mean can be used for the overall estimate. The bars corresponds to 95% confidence interval for the overall estimate that is cluster robust at the state level.

columns exclude either one or both fixed effects. The results from this table are consistent with the findings in Fig. 5 in the main text: Facebook users who attended a high school during its closure tend to have more long ties in their networks many years after the event compared to similar cohorts who

Fig. S8. The distribution of long ties across several demographic variables, including gender, age, education, population density of residence zipcode, residence census region and census division. Panels A, C, E, G, I and K compare the average fraction of long ties versus the ego network degree within each demographic category. Panels B, D, F, H, J, L show the conditional density of fraction of long ties within each level of the category.

attended the same school prior its closure.

D.4. Out-of-state College Attendance. In the main text, we discuss how the experience of inter-state migration and switching high schools within a single state is associated with higher fraction of long ties years after the event. In this section, we discuss the experience of another disruption, out-of-state college attendance, and its link to network structure years later. In particular, we examine US-born Facebook users who attended a college outside their hometown state and compare them with the population of users who attended a college inside their home state. We show that out-of-state college attendance is associated with higher fraction of long ties years later after the college, and this relationship persists if one looks at friends made out of college or account for the inter-state migration effects of out-of-state college attendance.

D.4.1. Data and Methods. We obtain college attendance data on each user based on their self-reported higher education institution. Overall, more than 47% of the active users in our data have attended a college and have optionally reported the name of the higher education institute attended in their Facebook profile. The self-reported information is generally trustworthy, since we have resolved them

Table S1. The effect of high school closure on fraction of long ties in the ego-network. The first column regresses fraction of long edges only on an indicator variable with value of 1 when the user has attended the high school during its closure as defined in *Materials and Methods*. The second column also includes school fixed effects. The third column is a two-way fixed effects model with both high school and age fixed effects where age is measured in years. All standard errors are cluster-robust at the school level.

	Fraction of Long Edges		
	(1)	(2)	(3)
During School Closure	0.017*** (0.005)	0.022*** (0.006)	0.012*** (0.004)
School Fixed Effects	No	Yes	Yes
Age Fixed Effects	No	No	Yes
Observations	33,813	33,813	33,790
R ²	0.001	0.064	0.067
Adjusted R ²	0.001	0.056	0.057
Residual Std. Error	0.299 (df = 33811)	0.290 (df = 33527)	0.290 (df = 33421)
F Statistic	23.308*** (df = 1; 33811)	8.034*** (df = 285; 33527)	6.560*** (df = 368; 33421)
<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01

against known higher education institutions. Our analysis will examine the subset of users aged between 30 and 60 who have attended at least one college, and divide them into two groups of in-state and out-of-state college attendance. Comparing the state where the user attended a college with their hometown state, which is also self-reported as also used and explained in the case of inter-state migration, we can determine whether at least one attended college is outside the user home state. The users with at least one such college constitute our out-of-state population and the rest are considered in-state. In a manner similar to the cases of inter-state migration and switching high schools, we match each individual in our population with their 6 month communication ego network and discard individuals with low levels of activity (degree less than 10 over the 6 month period). The final data consists of 23.5 million users, out of which 15.3 million (65%) attended only colleges within their home state and 8.2 million (35%) attended at least one out-of-state college.

In order to compare users by the location of college attendance, we employ the same approach as explained in the *Materials and Methods*. We compute the post-stratified mean of the fraction of long ties both conditional and unconditional on the degree bin and for each college attendance group. All combinations of gender, age and hometown county income constitute the 90 strata of the analysis. Fig. S11 shows the distribution of these control variables for both groups. An important observation is that the degree and age distributions of in-state and out-state college groups do not have noticeable differences. In contrast, males tend to attend out-of-state college at a higher rate, and users from richer counties are also more likely to attend a college out of their home state. Any difference in fraction of long ties between in-state and out-of-state user groups could partly be attributed to the effect of migration too since users who attended out-of-state college are more likely to permanently move out of their home state, as clearly verified in Fig. S11. However, it is not clear whether one should control for migration status by including it as a stratum, since the two mechanisms of out-of-state college attendance and inter-state migration only partly overlap and even if they were not distinct, it is difficult to separate them. Nevertheless in our secondary analysis, we include the binary inter-state migration status as another stratum in the post-stratification, which leads to 180 total strata. We will present results comparing the two groups controlling for either stratum definition.

D.4.2. Results. Overall, we find that the out-of-state group has about 9% more long ties relative to in-state users among all ties (34.1% vs. 31.3% in absolute terms with difference-in-means $p < 10^{-10}$

using cluster-robust standard errors at the county level) and 10.0% more long ties relative to in-state users among ties outside college (34.3% vs. 31.3% in absolute terms with $p < 10^{-10}$ using county-level cluster-robust standard errors). Fig. S12 further conditions on user degree and compares the fraction of long ties between users by location of college attendance, using two different stratum definition. The left plot compares the overall fraction of long ties. Similar to the case of inter-state migration, the higher fraction of long ties might be due to exposure to college friends from a different state, rather than an acquired skill. However, the inset plot shows that users with out-of-state college attendance have more long ties even among contacts who did not attend any of their colleges. This suggests that their current higher likelihood to have long ties is not directly due to connections with diverse college friends in the past. Out-of-state users are much more likely to be inter-state migrants as shown in Fig. S11, hence the right plot attempts to control for the migration status, even though this might underestimate the effect of out-of-state college attendance. Accounting for migration status reduces the absolute difference between the two groups (marginal over degree distribution) from 2.8% to 1.4%, however users with the experience of out-of-state college have consistently more (statistically significant) long ties in all degree bins. This suggests the higher structural diversity in the out-of-state group is not solely due to inter-state migration.

D.5. Self-Reporting as Potential Confounder. The differences on the fraction of long ties for those with and without the events we study might be partially attributable to other differences between the two populations. This is particularly true since assignment to either group (migrant vs. local or multiple high schools vs. single high school) is based on self-reported information on user’s profile. For example, it might be that users who tend to reveal more information about themselves (e.g., both of their high schools rather than only the last one) also happen to hold more long ties. We cannot completely rule out this concern; however, we can compare the two populations by how much information they share in their Facebook profiles and examine whether the differences in the fraction of long ties is driven by the differences in propensity to self-report personal information. In this section, we compare migrants versus non-migrants, out-of-state versus in-state college attendees and multiple high schools versus single high school students by how many of the following fields are completed on their Facebook profile: relationship status, current city, employer information and mobile phone number. For each user, we compute the count of such fields (out of 4) reported in their profile and compare the fraction of long ties between the groups with different number of disclosed fields. In terms of overall propensity to disclose information, we observe that migrants, multiple high school attendees and out-of-state college students on average disclose 4.0%, 5.8% and 1.8% more fields relative to non-migrants, single high school attendees and in-state college students respectively.

But are the differences in fraction of long ties due to differences in propensity of self-reporting personal information? We try to address this question by comparing the post-stratified mean of long ties between the groups conditioned on the the number of these fields reported in the Facebook profile (minimum of 0 and maximum of 4). Fig. S13 compares the fraction of long ties between the two groups in each type of life event (inter-state migration, switching high schools, and attending out-of-state college) across different number of disclosed fields within each group. These results do not differ much compared to the ones obtained when post-stratification is conducted on the whole group as reported in the text and shown in Figs. 4A, 4B and S12 rather than subgroups with a fixed number of disclosed fields. Furthermore, we observe that migrants, multiple high school attendees and out-of-state college students consistently have more long ties than non-migrants, single high school attendees and in-state college students no matter which sub-groups based on the number of disclosed fields we compare. Thus, we believe the differences in the fraction of long ties are not confounded due to the propensity of self-reporting personal information, at least given the analysis described above.

D.6. Associated Economic Outcomes. The main results on determinants of long ties show that three major disruptive events are associated with higher fraction of long ties later on in life. But they don't address whether those events lead to better outcomes too, specially since we showed the link between long ties and outcomes at an aggregate scale in the first part of the article. Establishing the link between these events and outcomes is challenging due to lack of important outcome data at the individual level. For illustrative purposes and to provide preliminary evidence of this link, we focus on two outcomes: the mean household income of current county and the mobile phone price. The first outcome allows to answer whether people with the experience of those events live in richer counties. The current county is determined using Facebook data based on location signals as explained previously and mean household income of US counties is obtained from 2018 ACS census. Mobile phone model, matched against estimated prices, is available from the user connection logs.

Fig. S14 shows the mean difference in the outcomes of migrant and local populations (positive values indicate migrants live in richer counties or have more expensive mobile phones) broken down by the racial composition of the hometown county where the user grew up. Hometown county is determined based on the self-reported hometown city in the user's profile. Both plots indicate that inter-state migration is increasingly associated with better outcomes in predominantly white counties, consistent with previous findings (53, 61). The upper limit of racial composition brackets approximately correspond to 20th, 40th, 60th, 75th, 90th and 100th percentiles of white population in US counties as indicated in census data. Figs. S15 and S16 show similar results comparing the population of multiple high schools versus single high school attendees and out-state versus in-state college students. In particular considering the case of switching high schools, users who attended multiple high schools have more expensive phones than those who attended a single high school regardless of the racial composition of their home town counties. However, only those users who grew up in predominantly white counties (with white population beyond the 75th percentile of all US counties) currently live in richer counties.

Fig. S9. The relative histogram of degree, hometown county income, gender, age and current state by migration status. The frequencies sum to 1 within each migration group. The degree distribution is shown with a log-log scale. The solid line in the age and income distributions correspond to a LOESS smoother.

Fig. S10. The relative histogram of degree, hometown county income, migration status, gender, age and current state by number of high schools attended. The frequencies sum to 1 within each high school attendance group. The solid line in the age and income distributions correspond to a LOESS smoother.

Fig. S11. The relative histogram of degree, hometown county income, migration status, gender, age and current state by location of college attendance. The frequencies sum to 1 within each college attendance group. The solid line in the age and income distributions correspond to a LOESS smoother.

Fig. S12. The stratified fraction of long ties over all ties and within the ties who did not attend the same high college as the user (inset) by degree and location of college attendance (**left**). The combination of gender, age and hometown county income bins constitute the stratification strata for the left and middle plots. The **right** plot also controls for the binary inter-state migration status by including it in the strata definition.

Fig. S13. The stratified fraction of long ties over all ties broken down by the number of fields disclosed in the user profile. The combination of degree, gender, age and hometown county income bins constitute the stratification strata for all plots. Bars correspond to 95% confidence intervals.

Fig. S14. The difference in current zip code mean family income (**left**), user mobile phone price (**middle**) and number of donations to fundraisers (**right**) of migrants over locals (migrant - local). The differences are stratified in both plots by strata defined as the combination of gender, age and hometown county income bins. Bars correspond to 95% confidence intervals.

Fig. S15. The difference in current zip code mean family income (**left**), user mobile phone price (**middle**) and number of donations to fundraisers (**right**) of multiple over single high school attendance groups (multiple - single). The differences are stratified in both plots by strata defined as the combination of gender, age and hometown county income bins. Bars correspond to 95% confidence intervals.

Fig. S16. The difference in current county mean household income (**left**), user mobile phone price (**middle**) and number of donations to fundraisers (**right**) of out-state over in-state college attendance groups (out-state - in-state). The differences are stratified in both plots by strata defined as the combination of gender, age and hometown county income bins. Bars correspond to 95% confidence intervals.