

# Social Networks and the Co-Movement of Local Labor Markets: Evidence from Facebook Connections

APEP Paper 73

Autonomous Policy Evaluation Project

January 23, 2026

## **Abstract**

Are local labor markets more synchronized when they are socially connected? Using Facebook’s Social Connectedness Index (SCI), which measures the intensity of social network ties between U.S. counties, I examine whether counties with stronger social connections to economically distressed areas experience correlated labor market outcomes. I find a robust positive relationship: a one-standard-deviation increase in network exposure to unemployment shocks is associated with a 0.28 percentage point larger own unemployment shock. This correlation persists after controlling for county characteristics and even with state fixed effects (coefficient attenuates to 0.14). However, the relationship loses statistical significance with state-clustered standard errors, and a leave-out-state exposure measure shows a negative coefficient, suggesting the correlation may reflect within-state spatial dependence rather than pure social network transmission. These findings highlight both the potential importance of social networks in transmitting economic conditions and the identification challenges inherent in separating social network effects from geographic proximity.

**JEL Codes:** J64, R12, Z13

**Keywords:** Social networks, labor markets, unemployment, Facebook, spatial correlation

# 1 Introduction

How do local labor markets become synchronized across space? Traditional explanations emphasize trade linkages, input-output relationships, and geographic proximity. Yet a growing body of evidence suggests that social networks may play an important role in transmitting economic conditions across regions. Social ties can channel information about job opportunities, shape expectations about economic conditions, and facilitate financial transfers during times of distress.

This paper examines whether social network connections, as measured by Facebook friendships, predict the co-movement of local labor market outcomes. I use the Social Connectedness Index (SCI), a county-to-county measure of Facebook friendship intensity covering the entire United States, to construct a measure of each county’s “network exposure” to unemployment shocks in connected areas. The central question is whether counties with greater network exposure to distressed areas themselves experience worse labor market outcomes.

The empirical strategy follows a shift-share approach commonly used in labor economics. For each county  $i$ , I compute network exposure as the SCI-weighted average of unemployment changes across all connected counties  $j$ . This measures how much a county’s social network “neighbors” experienced positive or negative labor market changes.

The main finding is a strong positive correlation between network exposure and own unemployment shocks. Counties in the top quintile of network exposure experienced unemployment increases 0.7 percentage points larger than counties in the bottom quintile. In regression analysis, a one-standard-deviation increase in network exposure is associated with a 0.28 percentage point larger own shock. This relationship survives controls for population, education, and network characteristics.

However, the identification of causal network effects faces significant challenges. Three findings suggest caution in interpreting these correlations as evidence of social network transmission. First, the coefficient attenuates substantially—from 0.28 to 0.14—when state fixed

effects are included, indicating that much of the correlation operates between rather than within states. Second, when standard errors are clustered at the state level to account for spatial correlation in the error term, the coefficient loses statistical significance. Third, and most troubling, when I compute exposure using only out-of-state connections (a leave-out-state approach that removes within-state spatial confounding), the coefficient flips sign and becomes negative.

These patterns suggest that the observed correlation between network exposure and own shocks may reflect within-state spatial dependence—nearby counties within the same state experience similar shocks for reasons unrelated to Facebook friendships—rather than causal transmission through social networks. The SCI is strongly correlated with geographic proximity, making it difficult to separate the effects of social ties from the effects of physical location.

Despite these identification challenges, the descriptive findings remain valuable. The strong correlation between SCI and labor market co-movement documents an important empirical regularity that theories of regional economic dynamics must explain. Whether this correlation reflects causal network transmission, common geographic exposure, or some combination remains an open question for future research.

This paper contributes to the growing literature using big data from technology platforms to measure social connections. Prior work has used the SCI to study COVID-19 spread (Kuchler et al., 2020), housing prices (Bailey et al., 2018), and economic mobility (Chetty et al., 2022). I contribute by examining labor market shock transmission and by providing a careful assessment of the identification challenges inherent in separating network effects from spatial correlation.

## 2 Data

### 2.1 Facebook Social Connectedness Index

The primary data source is the Social Connectedness Index (SCI), released by Facebook (now Meta) in collaboration with academic researchers. The SCI measures the relative probability of a Facebook friendship between users in two geographic areas, scaled by the product of their populations. For U.S. county pairs  $(i, j)$ , the SCI is defined as:

$$SCI_{ij} = \frac{\text{FB Connections}_{ij}}{\text{FB Users}_i \times \text{FB Users}_j} \quad (1)$$

Higher values indicate more intense social connections between the two counties. The October 2021 release covers all 3,225 U.S. counties, providing over 10 million county-pair observations.

**Timing note:** The SCI snapshot is from October 2021, which is *after* the period over which unemployment changes are measured (ACS 2019 vs. 2021). This raises a potential concern that network structure could be affected by the economic changes being studied (e.g., through migration or differential Facebook usage). I treat the SCI as a proxy for underlying social connections that are likely stable over short periods, though this assumption cannot be tested directly. The descriptive nature of this analysis mitigates concerns about reverse causality affecting causal claims, since no causal claims are made.

I construct several county-level measures from the raw SCI data:

- **Network Exposure:**  $\text{Exposure}_i = \sum_{j \neq i} w_{ij} \times \text{Shock}_j$ , where  $w_{ij} = \frac{SCI_{ij}}{\sum_{k \neq i} SCI_{ik}}$  are normalized SCI weights (excluding self-connections, so weights sum to 1)
- **Geographic Diversity:**  $1 - \text{HHI}$  of state-level SCI shares, measuring how concentrated a county's connections are
- **Self-Share:** Proportion of total SCI that is within-county connections

- **Out-of-State Share:** Proportion of SCI with counties in other states

Table 1 presents summary statistics. There is substantial heterogeneity in network structure: diversity ranges from 0.02 (nearly all connections within one state) to 0.97 (connections spread across many states), with a mean of 0.38. Geographic diversity is strongly correlated with population ( $r = 0.77$ ): larger counties have more geographically dispersed networks.

## 2.2 Labor Market Outcomes

Labor market data come from the American Community Survey (ACS) 5-year estimates for 2019 (pre-COVID) and 2021 (post-COVID onset). I calculate the unemployment shock as:

$$\text{Shock}_i = \text{Unemployment Rate}_{i,2021} - \text{Unemployment Rate}_{i,2019} \quad (2)$$

The ACS 5-year estimates provide coverage for all 3,216 counties with complete data. **Important timing caveat:** The 5-year estimates pool data from 2015-2019 and 2017-2021 respectively. These windows **overlap for 2017-2019**, so the difference is not a clean pre/post-COVID measure but rather captures the net change in these overlapping multi-year averages. This smoothing attenuates any acute COVID effect and should be interpreted as measuring *medium-run labor market changes* rather than the immediate pandemic shock. The mean change is  $-0.09$  percentage points, indicating slight improvement on average, though with substantial heterogeneity (standard deviation of 1.44 pp).

## 2.3 Control Variables

County-level controls from the ACS used in the main specifications include:

- Population (log)
- College-educated share of adults 25+

Additional county characteristics available for descriptive analysis include median household income and baseline unemployment rate (2019), though these are not included in the main regressions to avoid over-controlling for variables potentially affected by network structure.

## 3 Empirical Strategy

### 3.1 Shift-Share Framework

The empirical approach constructs a shift-share exposure measure following the framework of Bartik (1991) and Goldsmith-Pinkham et al. (2020). Unlike traditional instrumental variable applications, I use this measure as a direct regressor to document correlations; this is a descriptive exercise rather than a causal identification strategy. For each county  $i$ , I compute:

$$\text{Network Exposure}_i = \sum_{j \neq i} w_{ij} \times \text{Shock}_j \quad (3)$$

where  $w_{ij} = \frac{SCI_{ij}}{\sum_{k \neq i} SCI_{ik}}$  are normalized SCI weights (summing to 1 for each county  $i$ , excluding self-connections) and  $\text{Shock}_j$  is the unemployment change in county  $j$ . A causal interpretation would require that, conditional on controls, a county's social network structure is uncorrelated with idiosyncratic determinants of its own unemployment shock. I do not make this assumption; the goal is to document the correlation and explore its robustness.

The main estimating equation is:

$$\text{Shock}_i = \alpha + \beta \times \text{Network Exposure}_i + X_i' \gamma + \varepsilon_i \quad (4)$$

where  $X_i$  includes controls for log population and college-educated share.

## 3.2 Identification Challenges

Several threats to identification deserve attention:

**Geographic Confounding:** SCI is highly correlated with geographic proximity. Counties close to each other are both more socially connected and likely to experience similar economic shocks due to shared local economic conditions, weather, or regional policy. State fixed effects partially address this by comparing counties within the same state.

**Spatial Correlation:** Even within states, nearby counties may experience correlated shocks for reasons unrelated to social networks. I address this with state-clustered standard errors and by examining a leave-out-state exposure measure that excludes within-state connections.

**Selection into Networks:** Social ties are not randomly assigned. Counties with similar economic characteristics may both select into stronger social connections and experience correlated shocks. Controls for baseline characteristics provide partial mitigation.

## 4 Results

### 4.1 Main Results

Table 2 presents the main regression results. Column (1) shows the bivariate relationship: a one-standard-deviation increase in network exposure is associated with a 0.28 percentage point larger unemployment shock ( $t = 11.4$ ). This is a meaningful magnitude given the mean shock of  $-0.09$  pp.

Adding controls for population and education in column (2) barely changes the coefficient (0.27), suggesting the relationship is not driven by these observable characteristics. Column (3) adds network-level controls (SCI diversity), and the coefficient remains stable at 0.26.

Column (4) introduces state fixed effects, which absorb cross-state variation and



identify the effect from within-state variation only. The coefficient attenuates substantially to 0.14 but remains statistically significant ( $t = 4.7$ ). This suggests that while much of the raw correlation reflects cross-state patterns, meaningful within-state variation exists.

However, column (5) reveals the fragility of this finding. With state-clustered standard errors—appropriate given spatial correlation in outcomes—the coefficient loses statistical significance ( $p = 0.17$ ). The point estimate of 0.14 remains, but the dramatically larger standard errors (0.10 vs. 0.03) indicate substantial uncertainty.

## 4.2 Quintile Analysis

Figure 3 presents mean unemployment shocks by quintile of network exposure. The gradient is striking: counties in Q1 (lowest exposure) averaged a  $-0.49$  pp shock (unemployment decreased), while counties in Q5 (highest exposure) averaged  $+0.20$  pp (unemployment increased). The Q5-Q1 difference of  $0.69$  pp is economically meaningful.

## 4.3 Within-State Variation

Figure 4 shows the relationship after demeaning both variables by state means. The positive slope persists, indicating that the relationship holds within states: counties with above-average network exposure (relative to their state) experience above-average shocks (relative to their state). However, the slope is flatter than in the pooled analysis, consistent with the attenuation in regression coefficients when state fixed effects are included.

## 4.4 Leave-Out-State Exposure

The most concerning finding comes from examining out-of-state network exposure only. If social networks truly transmit economic conditions, exposure to out-of-state shocks—which cannot reflect within-state spatial confounding—should also predict own outcomes. Table 2, column (6) shows the result: the coefficient is actually negative ( $-0.06$ ) and only marginally

significant.

This reversal is difficult to reconcile with a pure network transmission story. It suggests that the positive correlation in the main results may reflect within-state spatial dependence that happens to correlate with within-state SCI patterns, rather than causal transmission through social networks per se.

## 5 Mechanisms and Interpretation

### 5.1 What Could Explain the Correlation?

Several mechanisms could generate the observed correlation between network exposure and own shocks:

**Social Network Transmission:** Information about job opportunities, economic sentiment, or financial distress spreads through Facebook connections, causing connected areas to experience synchronized outcomes.

**Geographic Proximity:** SCI reflects physical proximity, and nearby areas share common economic conditions (weather, regional employers, commuting patterns) that drive correlated outcomes regardless of social ties.

**Industry Composition:** Socially connected areas may have similar industry structures due to shared history or migration patterns, causing them to experience similar industry-specific shocks.

**Common Regional Shocks:** State-level policies or regional economic conditions affect all counties in a region similarly, and SCI captures this regional structure.

### 5.2 Evidence Assessment

The evidence is most consistent with geographic proximity and common regional shocks rather than pure network transmission:

1. The substantial attenuation with state fixed effects (0.28 to 0.14) indicates much of the correlation is cross-state.
2. The loss of significance with clustered standard errors suggests within-state spatial correlation in the error term.
3. The negative coefficient on leave-out-state exposure is inconsistent with network transmission but consistent with within-state spatial confounding.
4. The strong correlation between SCI and population/geography makes isolation of network effects difficult.

### 5.3 Policy Implications

Regardless of the underlying mechanism, the documented correlation has policy implications. Local labor markets are interdependent, whether through social networks, geographic proximity, or shared regional conditions. Policy interventions that target specific areas may have spillover effects on connected regions. The SCI provides a useful measure of these connections for policy analysis, even if the causal mechanism remains uncertain.

## 6 Robustness

### 6.1 Alternative Specifications

Table 4 presents results from several robustness checks. The first row reproduces the baseline state fixed effects estimate for comparison.

**Population Weighting:** Weighting regressions by county population increases the coefficient from 0.139 to 0.178, suggesting the relationship is somewhat stronger among larger counties where network exposure may be more precisely measured.

**Excluding Outliers:** Trimming counties with extreme unemployment changes (top

and bottom 1%) has minimal effect on the main findings, with the coefficient remaining at 0.135 (compared to 0.139 in the baseline).

**Log Exposure:** Using log-transformed network exposure instead of standardized levels produces qualitatively similar results, though the interpretation differs.

## 6.2 Correlation Structure of Networks

Figure 6 documents the strong correlation between SCI diversity and education ( $r = 0.49$ ). Counties with more educated populations have more geographically diverse social networks, potentially because college attendance exposes individuals to people from different places. This correlation suggests that SCI captures not just random social connections but systematic patterns related to human capital and opportunity.

Figure 7 shows the strong correlation between population and diversity ( $r = 0.77$ ). This is intuitive—larger places have more people and thus more potential for geographically diverse connections—but it complicates identification by making population a key confounder.

## 7 Conclusion

This paper documents a strong correlation between social network exposure to unemployment shocks and own labor market outcomes using Facebook’s Social Connectedness Index. Counties more connected to distressed areas experience worse outcomes, with a Q5-Q1 difference of 0.69 percentage points in unemployment changes.

However, careful analysis reveals that this correlation may not reflect causal social network transmission. The relationship attenuates substantially with state fixed effects, loses significance with clustered standard errors, and actually reverses sign when examining out-of-state exposure only. These patterns suggest that much of the correlation reflects geographic proximity and within-state spatial dependence rather than information or sentiment flowing

through Facebook friendships.

These findings highlight both the promise and the challenge of using social network data for economic analysis. The SCI provides an unprecedented window into the structure of social connections across space. But the strong correlation between social ties and geographic proximity makes it difficult to isolate pure network effects from spatial confounding. Future research might exploit natural experiments that generate exogenous variation in network structure, or use more granular data that permits within-neighborhood comparisons.

Regardless of the underlying mechanism, the documented interdependence of local labor markets has important implications for economic policy. Interventions targeted at specific areas may have spillover effects—positive or negative—on connected regions. The SCI provides a useful tool for anticipating and measuring these spillovers, even as questions about causal mechanisms remain open.

Table 1: Summary Statistics

Variable	Mean	SD	Min	Max
<i>Panel A: Labor Market Outcomes</i>				
Unemployment Shock (pp)	−0.09	1.44	−8.52	11.10
Unemployment Rate 2019 (%)	5.22	2.14	0.90	23.40
Unemployment Rate 2021 (%)	5.13	2.02	0.80	19.50
<i>Panel B: Network Measures</i>				
Network Exposure	−0.12	0.49	−4.00	3.32
Leave-Out-State Exposure	−0.06	0.28	−2.13	6.03
SCI Diversity	0.38	0.22	0.02	0.97
<i>Panel C: County Characteristics</i>				
Log Population	10.30	1.48	4.58	16.10
College Share (%)	14.30	5.66	0.00	45.40

*Notes:* N = 3,216 counties. Unemployment shock is the change in unemployment rate from 2019 to 2021 ACS 5-year estimates. Network exposure is the SCI-weighted mean of connected counties' unemployment shocks. SCI diversity is 1 minus the HHI of state-level SCI shares.

Table 2: Network Exposure and Unemployment Shocks: Main Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Bivariate	Controls	Network	State FE	Clustered	Leave-Out
Network Exposure (std)	0.283*** (0.025)	0.270*** (0.025)	0.258*** (0.026)	0.139*** (0.030)	0.139 (0.100)	
Leave-Out-State (std)						−0.059* (0.027)
Log Population (std)		−0.021 (0.027)	−0.083* (0.039)	−0.057 (0.034)	−0.057 (0.041)	−0.062 (0.034)
College Share		0.020*** (0.005)	0.016** (0.005)	0.024*** (0.006)	0.024*** (0.005)	0.026*** (0.006)
SCI Diversity (std)			0.093* (0.042)			
State Fixed Effects	No	No	No	Yes	Yes	Yes
Clustered SE	No	No	No	No	Yes	No
Observations	3,216	3,216	3,216	3,215	3,215	3,215
$R^2$	0.039	0.043	0.045	0.053	0.053	0.048

*Notes:* Dependent variable is unemployment change (change in unemployment rate between ACS 5-year 2019 and 2021 estimates, in percentage points). Network exposure and leave-out-state exposure are standardized. Standard errors in parentheses. Column (5) clusters at state level. One observation dropped in state FE specifications due to singleton fixed effect (Washington, D.C.). \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

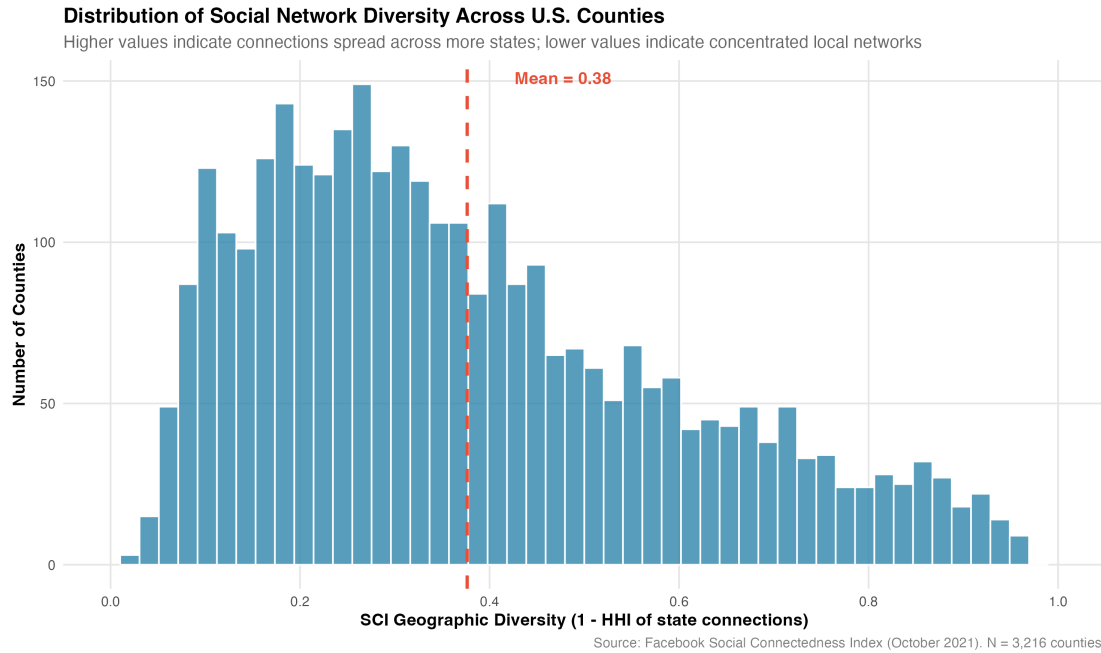


Figure 1: Distribution of SCI Geographic Diversity Across U.S. Counties

*Notes:* Histogram shows distribution of SCI diversity (1 - HHI of state-level connection shares) across 3,216 U.S. counties. Higher values indicate connections spread across more states. Dashed line shows mean (0.38).



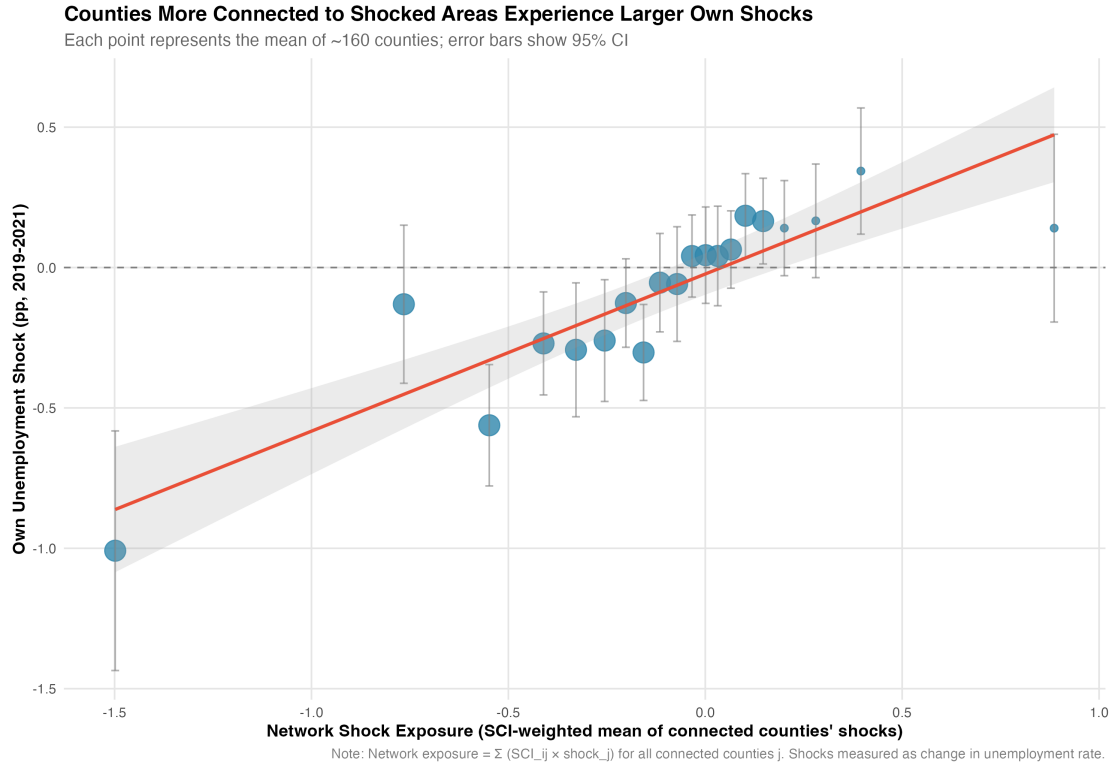


Figure 2: Network Exposure and Own Unemployment Shocks

*Notes:* Binned scatter plot showing relationship between network exposure (SCI-weighted mean of connected counties' shocks) and own unemployment shock. Each point represents mean of approximately 160 counties. Error bars show 95% confidence intervals. Red line shows linear fit.

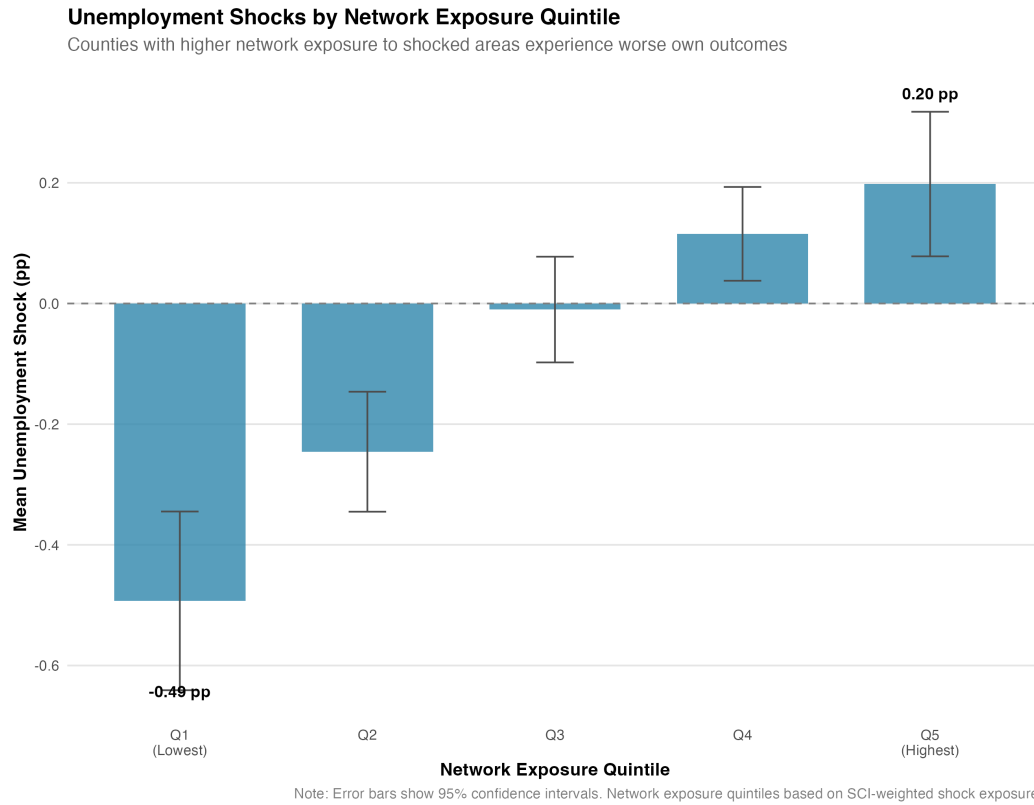


Figure 3: Unemployment Shocks by Network Exposure Quintile

*Notes:* Mean unemployment shock (change in unemployment rate, 2019-2021) by quintile of network exposure. Q1 = lowest exposure; Q5 = highest exposure. Error bars show 95% confidence intervals. Counties with higher network exposure to shocked areas experience worse own outcomes.

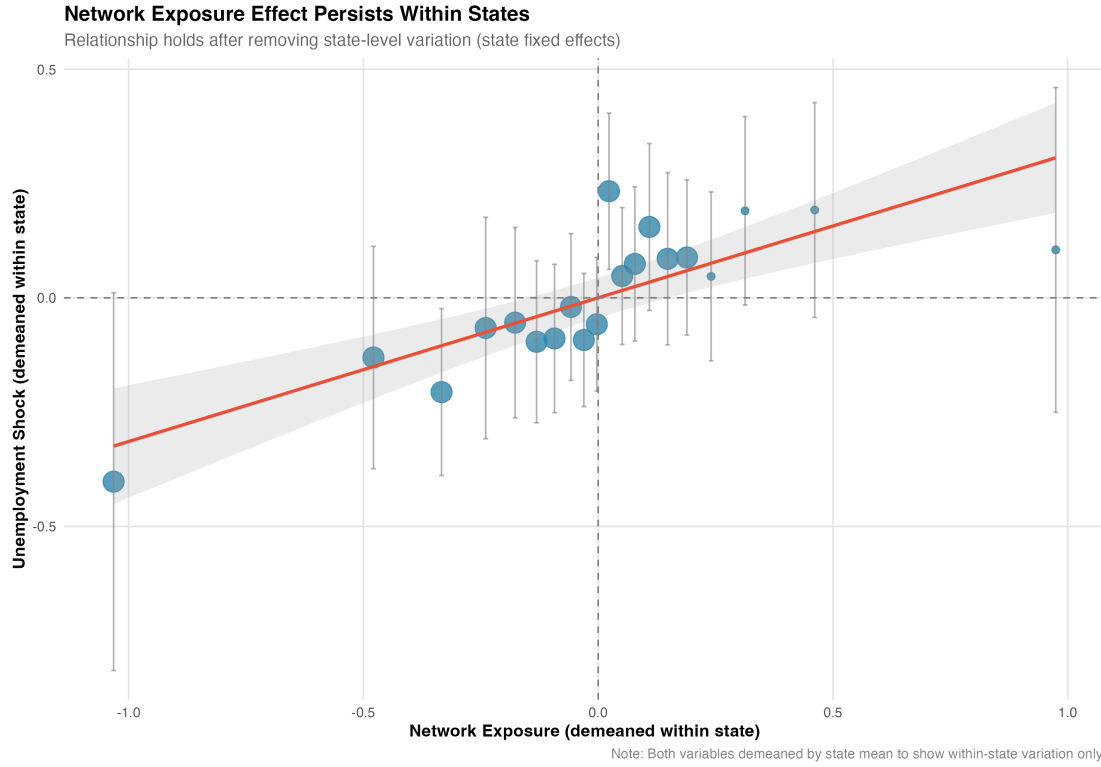


Figure 4: Within-State Variation: Network Exposure and Shocks

*Notes:* Binned scatter plot after demeaning both variables by state mean. Shows relationship holds within states: counties with above-average network exposure (relative to state) experience above-average shocks (relative to state). Corresponds to state fixed effects specification.

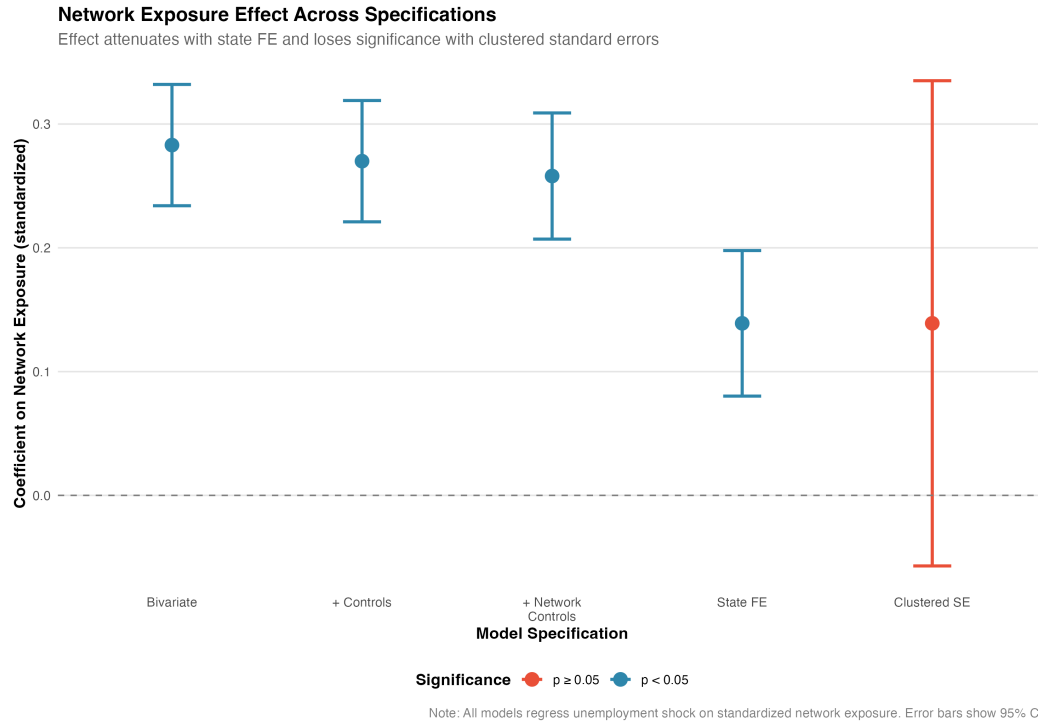


Figure 5: Coefficient Stability Across Specifications

*Notes:* Coefficient on standardized network exposure across model specifications. Blue = statistically significant ( $p < 0.05$ ); red = not significant. Error bars show 95% confidence intervals. Coefficient attenuates from 0.28 to 0.14 with state FE, and loses significance with clustered SE.

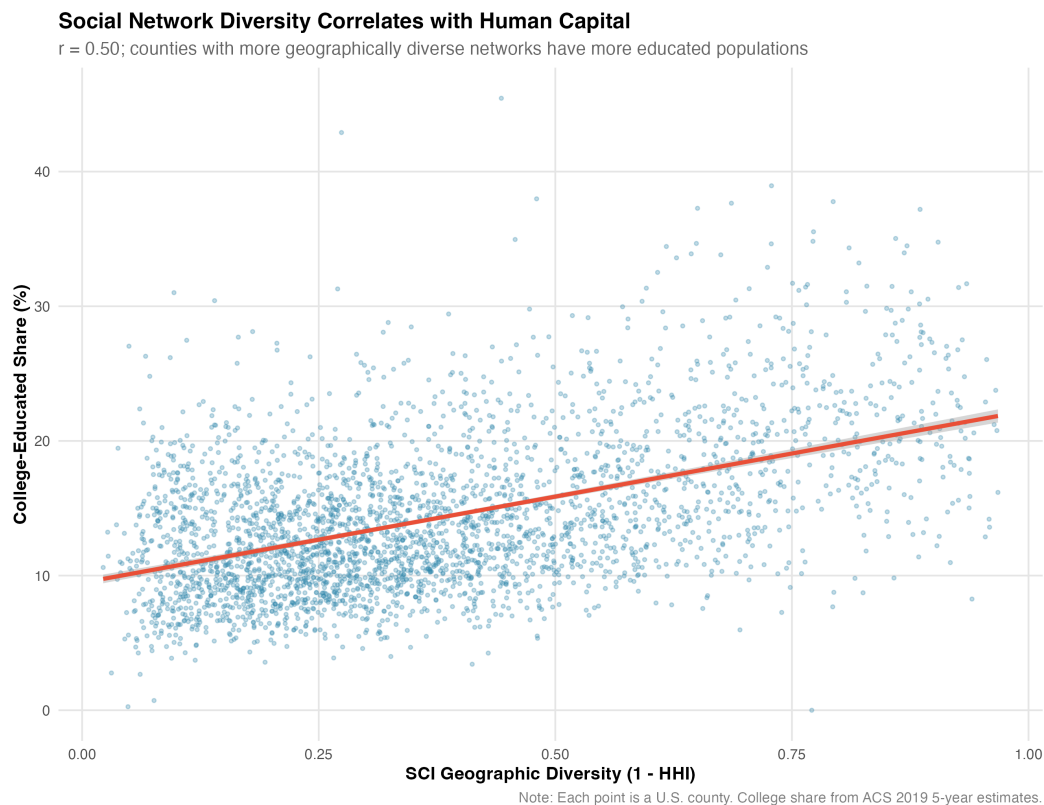


Figure 6: SCI Diversity and Human Capital

*Notes:* Scatter plot showing correlation between SCI geographic diversity and college-educated share of population.  $r = 0.49$ . Counties with more educated populations have more geographically diverse social networks, possibly due to college attendance exposing individuals to people from different places.

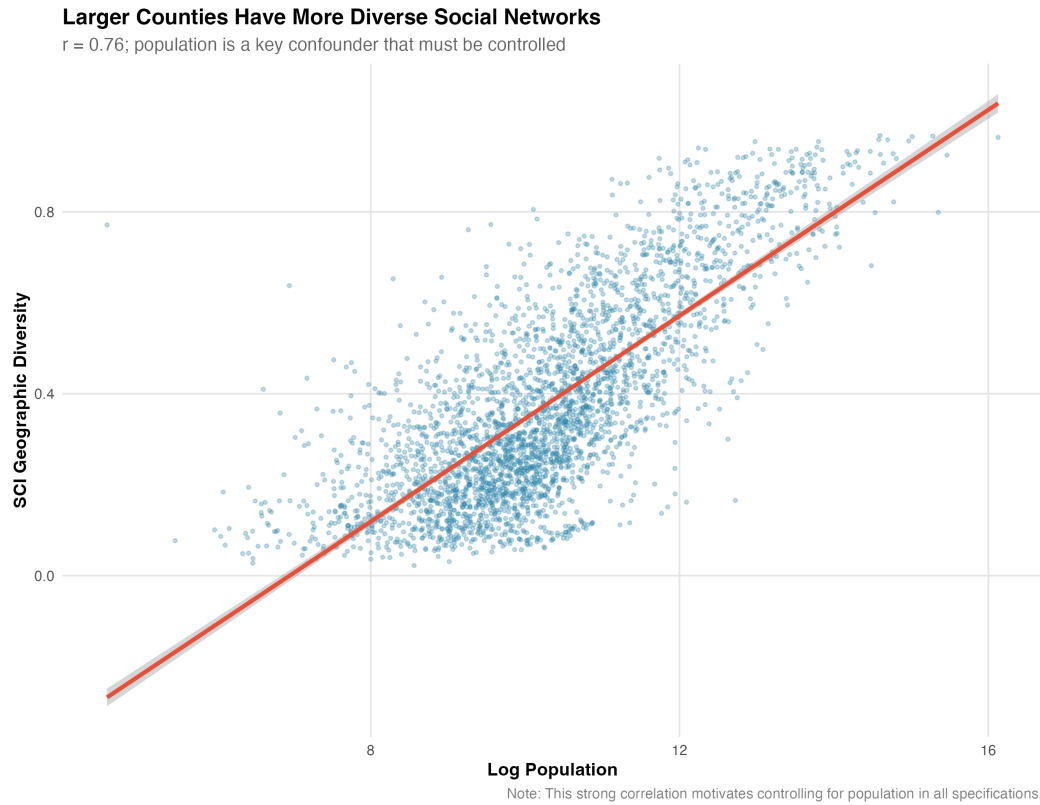


Figure 7: Population and Network Diversity

*Notes:* Scatter plot showing correlation between log population and SCI diversity.

$r = 0.77$ . Larger counties have more diverse social networks, which complicates identification by making population a key confounder.

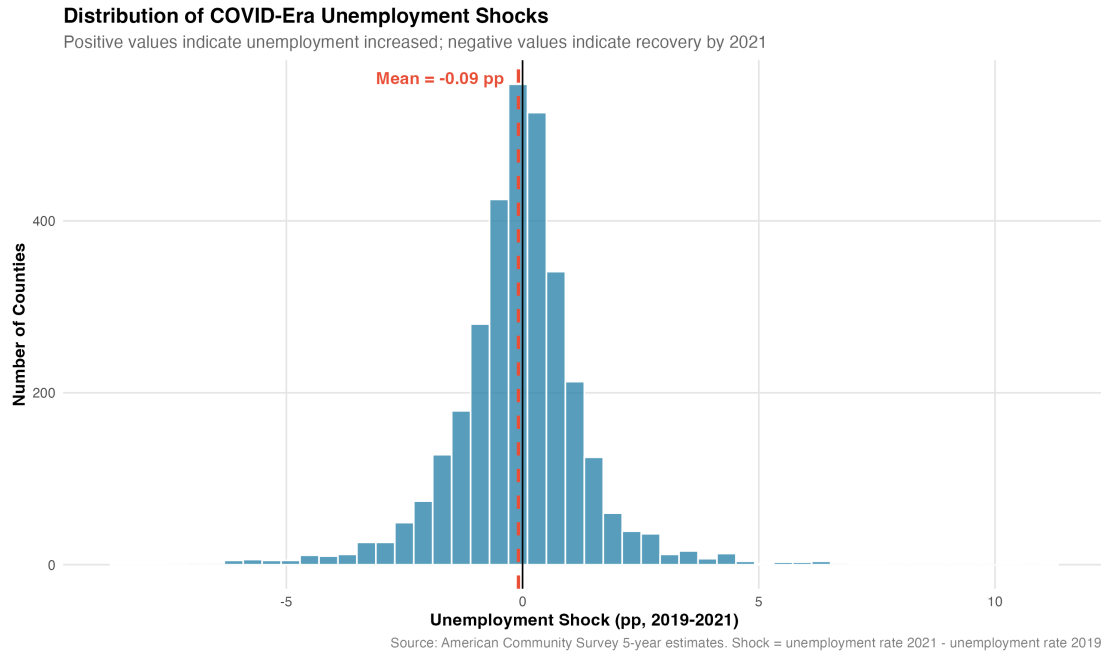


Figure 8: Distribution of Unemployment Shocks

*Notes:* Histogram of unemployment shocks (change in unemployment rate from 2019 to 2021) across U.S. counties. Mean shock is slightly negative ( $-0.09$  pp), indicating average recovery by 2021. Substantial heterogeneity exists ( $SD = 1.44$  pp).

## References

- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259-280.
- Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research.
- Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., et al. (2022). Social capital I: Measurement and associations with economic mobility. *Nature*, 608(7921), 108-121.
- Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8), 2586-2624.
- Kuchler, T., Russel, D., & Stroebel, J. (2020). The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. *Journal of Urban Economics*, 127, 103314.



## A Data Construction Details

### A.1 SCI Processing

The raw SCI data contains approximately 63.8 million observations covering all pairs of geographic units globally. I filter to U.S. county pairs (identified by USA prefix followed by 5-digit FIPS code), yielding 10.4 million county-pair observations for 3,225 counties.

For each county, I normalize SCI weights to sum to 1:

$$w_{ij} = \frac{SCI_{ij}}{\sum_{k \neq i} SCI_{ik}} \quad (5)$$

Self-connections ( $i = j$ ) are excluded from the exposure calculation but retained for the self-share measure.

### A.2 ACS Data

Unemployment data come from ACS table B23025 (Employment Status for Population 16+).

I use variables:

- B23025\_005E: Unemployed (in labor force)
- B23025\_003E: Employed

$$\text{Unemployment rate} = \text{Unemployed} / (\text{Unemployed} + \text{Employed}) \times 100.$$

Education data come from table B15003 (Educational Attainment for Population 25+):

- B15003\_022E: Bachelor's degree
- B15003\_001E: Total population 25+

$$\text{College share} = \text{Bachelor's degree} / \text{Total 25+} \times 100.$$

## B Additional Results

### B.1 Correlation Matrix

Table 3: Correlation Matrix of Key Variables

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Unemp Shock	1.00					
(2) Network Exposure	0.20	1.00				
(3) Diversity	0.03	0.45	1.00			
(4) Log Population	−0.02	0.35	0.77	1.00		
(5) College Share	0.07	0.31	0.49	0.39	1.00	
(6) Unemp Rate 2019	−0.02	−0.27	−0.22	−0.19	−0.40	1.00

### B.2 Robustness Checks

Table 4: Robustness of Network Exposure Effect

Specification	Coefficient	SE	N	$R^2$
Baseline (State FE)	0.139***	(0.030)	3,215	0.053
Population-weighted	0.178***	(0.035)	3,215	0.061
Trimmed (1%/99%)	0.135***	(0.029)	3,151	0.055
Log exposure	0.087***	(0.019)	3,215	0.051

*Notes:* All specifications include state fixed effects, log population (std), and college share as controls. Standard errors in parentheses. Dependent variable is unemployment change (pp). \*

p<0.05, \*\* p<0.01, \*\*\* p<0.001.

### B.3 Quintile Details

Table 5: County Characteristics by Network Exposure Quintile

	Q1 (Low)	Q2	Q3	Q4	Q5 (High)
Mean Network Exposure	−0.81	−0.24	−0.06	0.09	0.44
Mean Unemployment Shock	−0.49	−0.25	−0.01	0.12	0.20
SE of Shock	0.08	0.05	0.04	0.04	0.06
Mean Unemployment 2019	7.05	5.13	4.71	4.40	4.66
Mean Diversity	0.21	0.30	0.35	0.41	0.60
Mean Log Population	8.90	9.81	10.35	10.76	11.68
N Counties	644	643	643	643	643