Nursing Home Staff Networks and COVID-19*

M. Keith Chen

UCLA Anderson School of Management

Judith A. Chevalier

Yale School of Management and NBER

Elisa F. Long

UCLA Anderson School of Management

Revised: October 4, 2020

Abstract

Nursing homes and other long-term care facilities account for a disproportionate share of COVID-19 cases and fatalities worldwide. Outbreaks in U.S. nursing homes have persisted despite nationwide visitor restrictions beginning in mid-March. An early report issued by the Centers for Disease Control and Prevention identified staff members working in multiple nursing homes as a likely source of spread from the Life Care Center in Kirkland, Washington to other skilled nursing facilities. The full extent of staff connections between nursing homes—and the role these connections serve in spreading a highly contagious respiratory infection—is currently unknown given the lack of centralized data on cross-facility employment. We perform the first large-scale analysis of nursing home connections via shared staff and contractors using device-level geolocation data from 50 million smartphones, and find that 5.1 percent of smartphone users who visit a nursing home for at least one hour also visit another facility during our 11-week study period—even after visitor restrictions were imposed. We construct network measures of connectedness and estimate that nursing homes, on average, share connections with 7 other facilities. Controlling for demographic and other factors, a home's staff-network connections and its centrality within the greater network strongly predict COVID-19 cases. Traditional federal regulatory metrics of nursing home quality are unimportant in predicting outbreaks, consistent with recent research. Multivariate regressions comparing demographically and geographically similar nursing homes suggest that 49 percent of COVID cases among nursing home residents are attributable to staff movement between facilities.

---

*keith.chen@anderson.ucla.edu, judith.chevalier@yale.edu, elisa.long@anderson.ucla.edu. The authors thank Veraset for access to anonymized smartphone data, and acknowledge research support from the Tobin Center at Yale University. We are grateful for excellent research assistance from Jun Chen, Anna Schickele, and Sabrina Yihua Su, and for the forbearance of Seneca Longchen. This research was found to be not human subjects by the Yale IRB.

1 Introduction

Linked to more than forty percent of all U.S. fatalities as of August 31, 2020, nursing homes and other long-term care facilities have been disproportionately afflicted by the ongoing coronavirus pandemic (Conlen et al., 2020; Kaiser Family Foundation, 2020; Girvan, Gregg and Roy, Avik, 2020).[1] With an elderly resident population, many with underlying chronic medical conditions, congregate living quarters, and routine contact with staff members and outside visitors, nursing homes are particularly vulnerable to outbreaks of respiratory pathogens strausbaugh2003infectious, lansbury2017influenza. The U.S. Centers for Medicare and Medicaid Services (CMS), the primary federal regulator of nursing homes, estimates that more than 30 percent of all nursing home residents in New Jersey, Connecticut, and Massachusetts had contracted SARS-CoV-2 as of June 28, 2020 and that more than 9 percent of the entire nursing home population died in these states (Centers for Medicare and Medicaid Services, 2020a).

Evidence from the early outbreak at the Life Care Center in Kirkland, Washington demonstrated that nursing homes and other congregate facilities face extremely elevated risks of virus spread (D'Adamo et al., 2020; McMichael, 2020). CMS guidance issued on March 13, 2020 significantly restricted visitor access to long-term care facilities—effectively locking down nursing homes to residents, staff, and contractors (Centers for Medicare and Medicaid Services, 2020b). Nevertheless, many COVID outbreaks subsequently occurred in nursing homes, suggesting the unwitting introduction of the virus into homes by staff and contractors as one potential channel. The practice of employing nursing home staff across multiple facilities may play a key role in the spread of SARS-CoV-2, as a U.S. Centers for Disease Control and Prevention (CDC) report issued on March 18, 2020 identified staff working in multiple nursing homes as a likely source of spread from the Life Care Center to other skilled nursing facilities in Washington State (McMichael, 2020). Of the first four nursing homes with COVID outbreaks following this initial outbreak, two facilities received patient transfers from Life Care and two facilities employed staff working in both places (McMichael et al., 2020).

Despite this early recognition of cross-traffic between congregate settings as a potentially important transmission mode, the extent of connections between nursing homes remains unknown due to lack of systematic data. Furthermore, although the CDC identified staff members working in multiple long-term care facilities as a key high-risk group, CMS has not provided any specific guidance on this practice nor on reducing contacts between homes more generally (Centers for Medicare and Medicaid Services, 2020b,c,d).

Using novel device-level geolocation data for 501,503 smartphones observed in at least one of the 15,307 nursing homes in the continental U.S., we find that 5.1 percent of individuals who spend at least one hour in a nursing home also spend at least one hour in one or more other nursing homes in the 11-week period following the March 13th nationwide restriction on nursing home visitors. We construct several measures from network theory to characterize nursing-home connectedness,

---

[1] As of September 24, 2020, the Centers for Medicare and Medicaid Services documents 55,845 COVID-19 nursing home deaths, fewer than other cited sources. As discussed below, nursing home cases and deaths occurring before May 1, 2020 were not required to be reported to CMS.

and examine whether such connectivity predicts confirmed and suspected COVID-19 cases. To our knowledge, this is the first effort to measure and map the network structure of non-social visitors to nursing homes. These data are anonymized, but, given the prohibition of social visitors, this cross-traffic between homes is likely traceable to staff and contractors. We find that the number and strength of connections between nursing homes—and a home's centrality within the greater network—strongly predict COVID cases, even after controlling for location, demographic factors, number of beds, for-profit status, and CMS quality ratings. Consistent with recent research (Abrams et al., 2020; Konetzka, 2020; White et al., 2020), we observe that traditional federal regulatory metrics of nursing home quality are unimportant in predicting outbreak size.

2 COVID-19 in Congregate Facilities

The high case count and death toll in long-term care facilities demonstrates the urgent need to understand how transmission mechanisms within these facilities are distinct from broader community spread, to guide targeted policy initiatives and testing strategies (Pillemer et al., 2020; Ferguson et al., 2020). Given the incomplete case reporting by CMS, extant studies of nursing home cases typically rely upon researcher-compiled state data. Three studies (Abrams et al., 2020; Konetzka, 2020; White et al., 2020) examine the relationship between cases, home location, home demographics, and CMS quality ratings for facilities in a number of states. No study finds CMS ratings to be significant explanators of cases, although demographics and urban location are predictive of cases. Two studies of individual states (He et al., 2020; Li et al., 2020) find that higher CMS-rated nursing homes report fewer cases. One analysis finds no evidence that for-profit status significantly predicts nursing home cases (Konetzka, 2020), yet a study of Connecticut facilities does find for-profit status to be a predictor of cases (Rowan et al., 2020). While all of these papers provide careful statistical analysis of COVID in nursing home settings, no study directly measures connections amongst homes.

The importance of connections between congregate settings in SARS-CoV-2 spread has largely been identified through case studies rather than large-scale analysis. The CDC's evaluation of the Kirkland, Washington outbreak pointed specifically to staff employed at multiple nursing homes as a factor in spreading the initial outbreak to additional homes (McMichael, 2020). A study of four nursing homes in London (Ladhani et al., 2020) finds that 11 percent of staff worked in multiple homes and these workers were three times as likely to be infected than workers in a single home. Further, (Ladhani et al., 2020) also show that whole-Genome Sequencing of positive samples from residents and staff indicated cross-infection between residents and staff as well as multiple introductions of the virus into individual care homes. In a different congregate setting, movement of staff and residents across three affiliated homeless shelters likely contributed to outbreaks in each location (Tobolowsky et al., 2020). Employees at food processing plants are at increased risk of contracting SARS-CoV-2 given their proximate working conditions and frequent use of shared transportation between crowded, communal housing and the workplace (Dyal et al., 2020).

The movement of incarcerated individuals and the cross-usage of staff across prisons have been identified as risk factors for COVID-19 outbreaks; incoming inmate transfers were the probable source of the San Quentin Prison outbreak (Kinner et al., 2020; Williams and Griesbach, 2020). While we focus on SARS-CoV-2, the importance of linkages between congregate settings has been identified in case studies of prior disease outbreaks. Each of the three flu outbreaks at San Quentin during the 1918 influenza pandemic were linked to the introduction of a single transferred prisoner from a facility where flu was prevalent (Stanley, 1919).

In principle, if a congregate setting were completely closed to the outside, infection could not enter. A key challenge in isolating nursing homes derives from their reliance on staff who live in the community. A study by the State of New York (New York State Department of Health, 2020) concluded, largely based on the timing of infections, that through no fault of their own, nursing home workers were likely the main source of SARS-CoV-2 transmission in nursing homes. They find that roughly one-quarter of nursing home workers in New York State tested positive for the virus. Below, we describe briefly nursing home staffing practices and how they may exacerbate disease spread.

3 Nursing Home Staffing Practices and Regulation

Even in non-pandemic times, nursing home staffing presents challenges. Resident census and health conditions fluctuate from day to day, altering staffing needs on a daily basis with unpredictable absences, complicating the staffing problem (Slaugh et al., 2018). Understaffing leads to poor service and regulatory violations while overstaffing increases costs. To help manage this trade-off, care facilities often rely on staffing agencies to employ nurses and nurse aides and provide them on an on-call basis (Slaugh et al., 2018; Lu and Lu, 2017). While data are limited, a 2009 study suggests that 60 percent of nursing homes use a staffing agency for some of their staffing (Castle, 2009). Given this widespread reliance on staffing agencies and the recent growth in nursing home chain affiliates (Cadigan et al., 2015), many nurses and nursing assistants commonly work in multiple facilities. Nursing homes also receive services from hospice workers, dialysis technicians, clinicians, medical transporters, and other non-nursing staff that visit multiple homes. In addition to this planned cross-usage, nursing home workers may combine employment across multiple nursing homes as well as other jobs. Survey data from 2012 indicate that 19 percent of nursing assistants and 13 percent of registered nurses hold a second job of some type (Van Houtven et al., 2020). According to the Bureau of Labor Statistics, the median nursing assistant earned $28,980 in May 2019, which makes a willingness to work multiple jobs unsurprising. However, extant regulatory data at the nursing home level do not track the degree to which healthcare workers work in more than one nursing home or other healthcare setting.

4 Data and Methodology

Examination of the nursing home COVID-19 crisis is further hindered by the fact that CMS did not require nursing homes to submit data on COVID-19 cases and fatalities until May 2020. Thus, for our main data analysis, we use the disclosures of individual state Departments of Public Health

Tab. 1: Summary statistics of U.S. nursing homes.

Variable	State reporting facilities	CMS reporting facilities
Number of nursing homes	6,337	13,165
Cases per home	15.1	11.5
Demographics		
High proportion (>25%) of Black residents, %	16.7	12.7
High proportion (>50%) on Medicaid, %	32.9	28.1
Urban location, %	81.2	72.5
For profit home, %	74.4	71.3
Regulatory measures		
Number of beds	115 (59.1)	109 (60.3)
CMS quality rating (1-5)	3.18 (1.42)	3.15 (1.42)
Has infection violations, %	75.3	75.7
Network metrics		
Node degree	7.08 (8.38)	6.42 (7.89)
Node strength	8.82 (12.4)	8.11 (14.4)
Weighted average neighbor degree	10.21 (8.33)	9.42 (8.22)
Eigenvector centrality in state	0.095 (0.19)	0.087 (0.19)

CMS facilities include all continental U.S. nursing homes that report demographic and regulatory data. Binary variables are % of nursing homes; continuous variables are mean values with standard deviations in parentheses.

to determine cumulative nursing home COVID cases. From the 22 states for which home-level resident case data are available, we collected data on cumulative resident cases as of May 31, 2020 (or closest reporting period). In the Supplement, we repeat our analyses using the cumulative case data reported by CMS for homes nationwide, with the caveat that CMS instructions for reporting cumulative cases allowed nursing homes to not report cases occurring before May 2020. For example, the nation's first congregate COVID outbreak, the Life Care Center of Kirkland, Washington, is recorded in CMS data as having a cumulative zero COVID-19 cases, while the CDC report (McMichael, 2020) states that, as of March 18, 2020, 81 residents of the facility had contracted the virus and 23 persons had died.

Using the CMS address of record for each facility, we merge the nursing home-level COVID-19 case data with nursing home staff-network connections measured using anonymized device-level smartphone data for the continental U.S. over the 11-week period March 13 to May 31, 2020.

4.1 Smartphone Location Data

We estimate staff and contractor networks across nursing homes using anonymized smartphone-location data provided by Veraset, a company that aggregates location data across several apps on both the Apple and Android platforms after the user consents to the use of their anonymized data. Previous studies with these data have found them to be highly representative of the U.S. on numerous demographic dimensions (Chen and Rohla, 2018). A smartphone typically reports ("pings") a user's location every ten minutes throughout the day. We filter these data to estimate user/nursing-home visits by, first, excluding visits with fewer than three user pings inside that home that day, then further excluding visits whose first and last pings are separated by less than one hour. This helps reduce staff false-positives due to GPS error or users who briefly enter a home (like a delivery person). Under this definition, of the more than 50 million smartphones in our U.S. sample, we identify 501,503 smartphones that visit at least one U.S. nursing home between March 13 and May 28, 2020, and a visitor to a home visits that home an average of 16 days over our 11-week study period.

We match all U.S. nursing homes with a shapefile delineating each facility's rooftop boundary. To do so, we match a nursing home's CMS-provided street address to a latitude-longitude location using the Google Maps API, and then match that location to a satellite image machine-learned geofence of the convex-hull of the building's rooftop (provided by Microsoft / Open Streetmaps). Using these rooftop geofences, we find all times that a sampled smartphone spends more than one hour in a U.S. nursing home during our study period, when visitor restrictions were in effect. By identifying smartphones that entered more than one nursing home, we measure the nursing home staff-contact network.

4.2 Network Metrics

The contact structure among nursing homes within a state is represented by an undirected network consisting of n nodes (the facilities) and $n(n-1)/2$ possible edges (pairs of facilities). We construct a symmetric $n \times n$ adjacency matrix $\mathbf{A}$, where $a_{ij} = 1$ if at least one smartphone is observed in both facilities i and j, and 0 otherwise. Edge weights w_{ij} correspond to the number of smartphones observed in both facilities. State-level summaries of each network measure are in the Supplement Table S1.

A facility's *degree* k_i equals the total number of other nursing homes connected to facility i (i.e., the number of node i's *neighbors*).

$$k_i = \sum_{j=1}^{n} a_{ij} \tag{1}$$

Strength s_i is the weighted sum of *contacts* with other facilities (i.e., the total number of smartphones that appear in facility i and some other nursing home).

$$s_i = \sum_{j=1}^{n} w_{ij} a_{ij} \tag{2}$$

Weighted average neighbor degree $\bar{k}_i^w$ is the average degree of node i's neighbors (i.e., the neighbors' connections to other facilities), weighted by the number of connections w_{ij} shared with node i, as previously defined (Barrat et al., 2004).

$$\bar{k}_i^w = \frac{1}{s_i}\sum_{j=1}^{n} w_{ij}a_{ij}k_j \tag{3}$$

Eigenvector centrality v_i measures the extent to which node i is connected to other highly connected nodes in the network.

$$v_i = \frac{1}{\lambda}\sum_{j=1}^{n} a_{ij}v_j \tag{4}$$

This measure is computed using the principal eigenvector of the adjacency matrix, rewritten in matrix notation as $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. We normalize v_i to range between 0 and 1 within each state.

4.3 Empirical Specification

Our main specification examines predictors of nursing home resident COVID cases as a function of several explanatory variables. We include the home's *demographic* characteristics, including linear and quadratic terms for the number of beds. Following previous literature (Abrams et al., 2020), we include indicator variables for whether a nursing home has a large proportion ($> 50\%$) of residents on Medicaid and a large proportion ($> 25\%$) of Black residents. We include *CMS quality* measures, as done previously (Abrams et al., 2020), which rate nursing homes on a five-point scale; we include this as a categorical variable, with the omitted category being five-stars (the highest possible rating). We include an indicator variable if the home had *infection control violations* in its most recent inspection. Finally, we define an indicator variable for whether a home is in an *urban* location based on the CDC's urban-rural classification (Centers for Disease Control and Prevention, 2020).

To examine whether nursing home connectivity predicts COVID-19 cases, we use the following regression model:

$$\begin{aligned} sinh^{-1}(Cases_i) = {} & \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 NodeDegree_i + \boldsymbol{\beta}_2 NodeStrength_i \\ & + \boldsymbol{\beta}_3 WeightNeighDeg_i + \boldsymbol{\beta}_4 EigenCentrality_i \\ & + \boldsymbol{\gamma}_0 \mathbf{X}_i + \boldsymbol{\gamma}_1 \mathbf{F}_i + \varepsilon_i \end{aligned}$$

where $sinh^{-1}(x) = \ln(x + \sqrt{1 + x^2})$ is the inverse hyperbolic-sine of a nursing home's COVID cases. All reported semi-elasticities are adjusted for the $sinh^{-1}$ functional form. We include as independent variables the four network measures that characterize a home's connectivity, as described in the previous section. The vector $\mathbf{X}_i$ includes demographic, geographic, and regulatory controls for nursing home i. To control for reporting and other differences across states, we include state fixed effects denoted by $\mathbf{F}_i$. Coefficients for all variables are in the Supplement Table S2, with results for individual states reported in Table S3. Table S4 replaces the dependent variable with a binary indicator variable if the nursing home has had any COVID cases. Table S5 repeats our analysis using county fixed effects, an important robustness check given that county-level SARS-

CoV-2 prevalence predicts case counts in nursing facilities (White et al., 2020). Finally, Table S6 replaces the data for 22 states with the larger CMS dataset for the continental U.S.

For the time series analysis, we use daily nursing home resident case counts provided by the Florida, Connecticut, and Colorado Departments of Health (Florida Department of Health, 2020; Connecticut Department of Public Health ; Colorado Department of Public Health and Environment , 2020). Because the states did not all report cumulative cases, we construct a binary variable *first outbreak* to indicate the first week in which each nursing home appears in the database. The case data for all three states begins by the week ending April 19; thus, we construct network measures from 5 weeks of smartphone data, beginning with the visitor lockdown after March 13 until April 19, 2020.

We use a linear probability model with each observation a nursing home-week. For each home i in week t, we regress the "first outbreak" indicator on county-week interaction fixed effects and on $NewOutbreakDegree_{i,t-2}$, the number of nursing homes directly connected to home i that experienced their first recorded case in week $t-2$. For example, if home A is only connected to home B, and home B experiences its first outbreak during the week of May 3, 2020, the variable $NewOutbreakDegree_{A,t-2}$ equals zero in all weeks except the week of May 17, 2020. $NewOutbreakDegree_{i,t-2}$ is similar to our previous measure $NodeDegree_i$, but is calculated only for homes with reported outbreaks and is lagged by 2 weeks. With the inclusion of county-week fixed effects, the specification is effectively estimating whether, a home i in county j in week t is differentially likely to have an outbreak in week t relative to other homes in county j and week t if i is connected to homes that had first outbreaks in week $t-2$. We obtain the following specification:

$$FirstOutbreak_{i,t} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 NewOutbreakDegree_{i,t-2} + \boldsymbol{\gamma}_1 \mathbf{F}_{i,t} + \varepsilon_{i,t}$$

We repeat this specification using connections to homes in the prior week, $NewOutbreakDegree_{i,t-1}$, and contemporaneous connections, $NewOutbreakDegree_{i,t}$ as alternative right hand side varables. We exclude from our analysis any homes that have connections to homes out-of-state as we cannot observe detailed case timing for homes in states neighboring these three states.

5 Network Measures

Nursing homes display a wide range of connectedness with other homes. Average *degree*—the number of facilities a nursing home shares at least one smartphone connection with—across the U.S. is $\langle k \rangle = 6.6$, but ranges from an average degree below 1 in South Dakota, Vermont, and Wyoming, to an average exceeding 10 in Florida, Maryland, and New Jersey (Supplement Table S1). Among nursing homes with confirmed or suspected cases reported to CMS, average degree is 7.8 compared to 5.6 among homes with no documented cases ($t = 16.1$, $p < 0.0001$), with a significant difference across the entire degree distribution (see Fig. 1). Average *strength*—the total number of smartphones appearing in a home and its connections—is also greater in homes with COVID (9.7 vs. 7.5, $t = 7.9$, $p < 0.0001$). For example, the correlation between a home's *degree* in the first 5 weeks of our data collection and the last 5 weeks is 0.58; the correlation for *strength* is even higher at 0.82.

To illustrate how network measures differ across nursing homes, we present network diagrams for a subset of homes in six states as depicted in Fig. 2 and summarized in Table 2. Nodes denote individual nursing homes and edges represent connections between nodes (i.e., at least one smartphone observed in both homes). More connected nodes are generally towards the center of each diagram and nodes with fewer connections are on the periphery. In each sub-network, a focal nursing home or “hub” is shown in blue, with its direct neighbors (homes with at least one shared contact) in dark grey and its neighbors’ neighbors in light grey. Node size denotes CMS-reported confirmed and suspected COVID cases among residents as of May 31, 2020. Edge color corresponds to the number of unique smartphones observed in each pair of homes.

A major challenge facing nursing homes is that every connection is a potential link to other connections—and to SARS-CoV-2 transmission. In the Alabama sub-network (Fig. 2A), for instance, the focal nursing home reported eight COVID cases among residents and 30 confirmed or suspected cases among staff, and this facility is directly connected to another Alabama nursing home with 68 resident and 48 staff cases (the larger grey node). Both facilities are highly connected to other homes, including one nursing home that shared 43 smartphones with the focal home—after visitor restrictions were imposed in March. Although California nursing homes have average degree of 6.0 and average strength of 7.3, both slightly below U.S. averages, one Los Angeles facility (Fig. 2B) has *degree* of 9 and *strength* of 83, implying that homes connected to this hub share, on average, nine staff members, each of whom may be a potential conduit of SARS-CoV-2 transmission given the home’s 63 reported cases by May 31st.

With an *eigenvector centrality* of 1.0, the selected hub node is the most “connected” nursing home in Florida (Fig. 2C). Not only is this facility directly linked to 52 other homes—substantially higher than the state’s average of 11.4—many of these direct connections are themselves highly connected,

Fig. 1: Degree distribution of nursing homes with and without COVID cases (reported to CMS as of May 31, 2020).

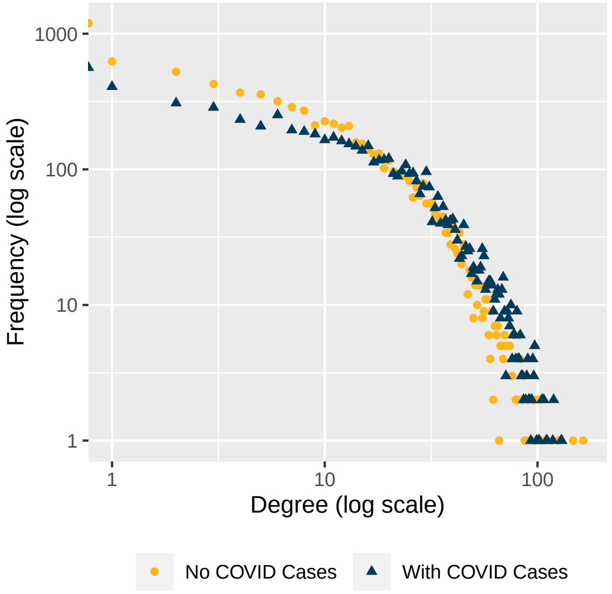
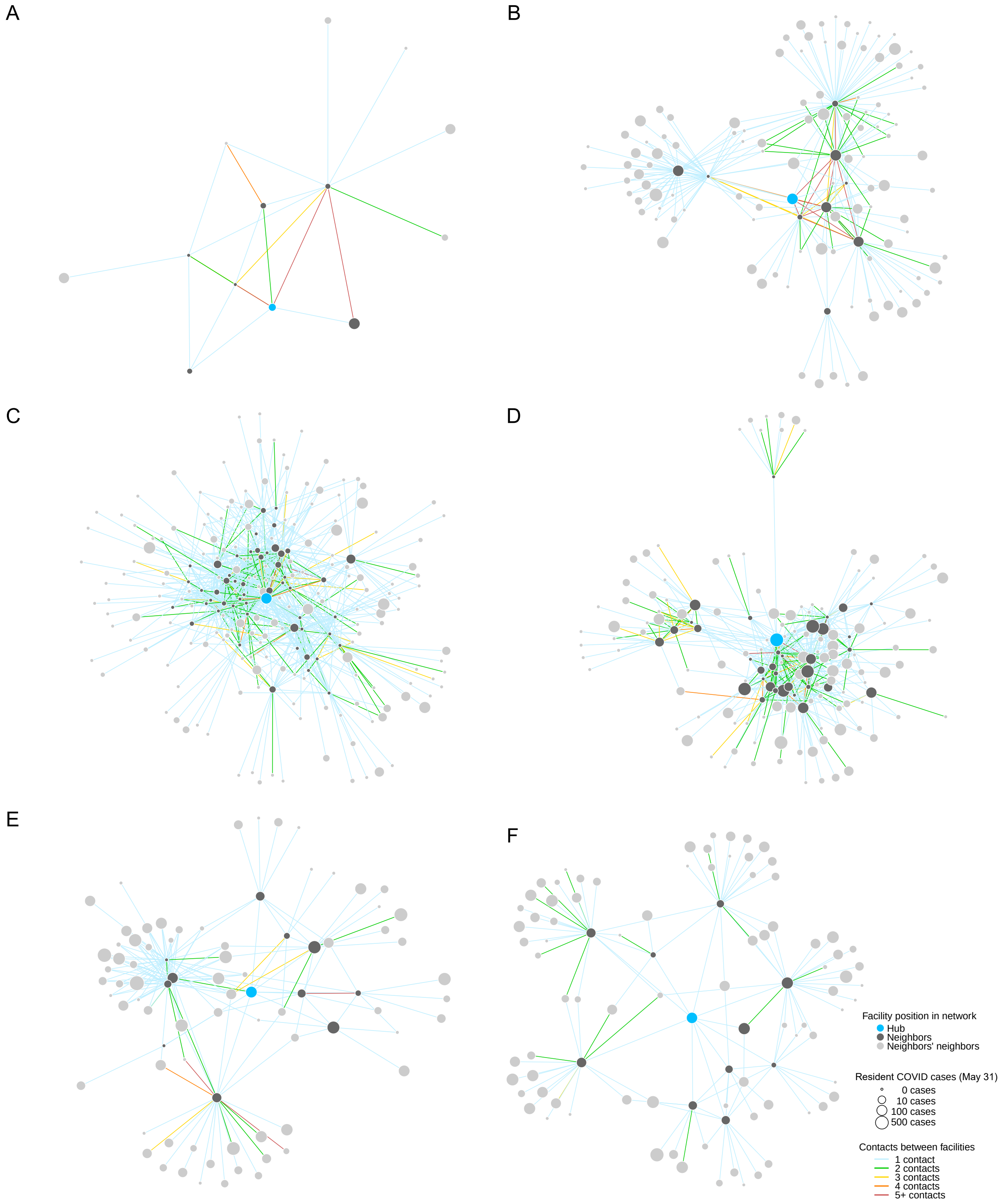


Fig. 2: Network structure of selected nursing home facilities in six U.S. states. Details provided in Table 2.



Tab. 2: Network centrality measures for six selected nursing homes.

Hub Facility	State	COVID Cases	Degree	Strength	Wtd. Avg. Neigh. Deg.	Eigenvector Centrality
A	AL	8	6	56	8.8	<0.01
B	CA	63	9	83	24.1	0.09
C	FL	54	52	81	23.9	1.00
D	GA	220	34	57	24.4	0.56
E	NY	62	5	5	42.4	0.12
F	PA	78	10	10	13.5	0.08

COVID cases are confirmed and suspected cases among residents reported to CMS as of May 31, 2020.

demonstrating the importance of capturing the entire network in these outcome measures. A small number of facilities have disproportionate influence in the overall network in Florida, with only 4% of nursing homes having eigenvector centrality > 0.5. The hub node in Georgia (Fig. 2D) faced a large COVID outbreak with 220 confirmed or suspected cases among residents by May 31st, one of the largest outbreaks in the state. With a *degree* of 34 and relatively high *eigenvector centrality* of 0.56, this node is directly linked to other highly connected facilities, including several with large COVID outbreaks.

Home to more than 600 skilled nursing facilities, New York State has an average *degree* of 7.8. While this illustrative hub facility (Fig. 2E) has only five direct connections, these neighbors are highly connected themselves, resulting in a *weighted average neighbor degree* of 42, well above the state's average. Lastly, a Pennsylvania nursing home (Fig. 2F) has both a *degree* and *strength* of 10, meaning that only one smartphone appears in both the focal home and each connected facility. This particular nursing home illustrates how direct connections act as bridges to other clusters of homes, potentially importing or exporting SARS-CoV-2 infection across different sub-networks.

6 Cross-Sectional Results

Table 3 presents multivariate regressions of cumulative nursing home COVID-19 cases as of May 31 on a set of explanatory variables. Importantly, these regression specifications include state fixed effects to allow for differences in baseline risks and reporting practices across states; we include even finer county fixed effects in the Supplement. We use the inverse hyperbolic-sine of cases as the dependent variable, given its non-negative skewed distribution. Column (1) shows our base specification with our simplest network explanatory variable, *node degree* k_i—the number of "neighbors" or other nursing homes connected to the focal home by at least one smartphone. Results indicate that, if a home adds 1 neighbor (average degree is 7.1), the expected number of COVID-19 cases increases by $3.43\% \times 4.43 = 15.2\%$. Column (2) replaces the degree measure with *node strength* s_i—the total number of "contacts" or smartphones that appear in the facility of interest and in some other nursing homes. This too predicts nursing home cases significantly: if a home adds 1 contact (average strength is 8.8), expected cases of COVID-19 increase by $1.63\% \times 2.14 = 3.48\%$.

Tab. 3: Covariates of COVID-19 cases within nursing homes.

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Node degree	0.0343***			0.0242***	
	(0.00255)			(0.00508)	
Node strength		0.0163***		-0.00610^{+}	
		(0.00166)		(0.00297)	
Weighted average neighbor degree			0.0409***	0.0299***	
			(0.00267)	(0.00344)	
Eigenvector centrality in state					1.044***
					(0.109)
Fixed effects	State	State	State	State	State
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	6,337	6,337	6,337	6,337	6,337
F-stat	123.4	114.9	128.7	112.9	114.5
R^2	0.408	0.400	0.412	0.415	0.399
Within R^2	0.189	0.178	0.195	0.199	0.177

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $^{*}p < 0.01$, $^{**}p < 0.001$, $^{***}p < 0.0001$. Dependent variable is inverse hyperbolic sine of COVID cases using state data. Demographics include number of beds, high proportion of Black residents, and high proportion on Medicaid. CMS quality is a 1-5 categorical rating.

Column (3) replaces the degree and strength measures with *weighted average neighbor degree* $\bar{k}_i^w$—the average degree of a nursing home's neighbors, weighted by strength to the focal node (Barrat et al., 2004). Here, an increase of 1 (mean is 10.2) leads to an expected $4.10\% \times 2.05 = 8.41\%$ increase in cases.

Column (4) may be of particular interest to policymakers as it examines the predictive power of local network features, potentially knowable by individual nursing home administrators. Intuitively, this regression compares demographically and geographically situated nursing homes of similar quality, which are thus likely exposed to similar risks of community-spread. Regression 4 suggests that 49 percent of nursing home resident cases are attributable to shared-staff transmitting the virus across multiple nursing homes.

Column (5) uses our final network measure, *eigenvector centrality* v_i—the extent to which a nursing home is connected to other highly connected nursing homes, normalized to range between 0 and 1 within each state. This measure implies that, as we move from an unconnected nursing home in the state ($v_i = 0$) to the most connected ($v_i = 1$), expected cases increase by 195%.

Consistent with other studies (Abrams et al., 2020; Konetzka, 2020; White et al., 2020), we find that CMS ratings of nursing home quality are not predictive of infections, yet facilities in urban locations, those with more beds, a higher share of Black residents, or a higher share of residents on Medicaid are all associated with more COVID-19 cases (details in Supplement Table S2). We find that for-profit homes are associated with more COVID-19 cases, consistent with (Rowan et al., 2020).

7 Time Series Evidence

One potential limitation of our analysis is that we do not explicitly show that SARS-CoV-2 travels from home to home. Given data limitations—particularly the late initiation of CMS and state reporting and the failure of many states to archive early reporting—we do not have consistent time series data for individual nursing homes to examine cases over a long time period. However, we were able to hand-collect weekly home-by-home data on the presence of cases starting in mid-April for three states, Florida, Colorado, and Connecticut. The three states have had very different time patterns of COVID cases. (Supplement Fig. S1).

To investigate whether an initial COVID outbreak in a nursing home is systematically preceded by outbreaks in homes connected to it, we track each nursing home in these states weekly from the week ending April 19, 2020 until the home's first reported COVID case or August 23, 2020, whichever occurs first. Table 4 presents results of a linear probability model with nursing home-week observations. The dependent variable "first outbreak" coded as 0 before a nursing home's first case and 1 in the week of its first case. The independent variable in the first column is the number of homes connected to the nursing home that had a first case two weeks before the examined week. Importantly, county times week fixed effects are included. Column 1 demonstrates that a home with more connections to homes with new outbreaks two weeks prior are more likely to have first outbreak in a given week, relative to other homes in the county that same week. Connections in the previous week are somewhat less predictive, but still statistically significantly different from zero. Connections to homes experiencing their first outbreak contemporaneously are not predictive at all (the coefficient is negative but statistically insignificant). One might expect that a spurious correlation—the possibility that connected nursing homes are alike in unobserved ways—would most likely manifest in the data with a finding of contemporaneous outbreaks, but we find no evidence of this.[2] The coefficient magnitude implies that a shared contact with a nursing home experiencing its first COVID case in week $t-2$ is associated with a 2.5 percentage-point higher probability of a home reporting a first case in week t. This is large relative to the baseline probability of a first case in any given week, as the mean of the first week indicator in the regression specification is 0.07.

---

[2] The exact time lag anticipated by a shared staff mechanism is unclear. Given the weekly reporting cadence, the typical 2-week infectivity period, and the fact that a shared staff member who carries the disease from home A to home B may not be the index case at home A, we expect some lag from measured and reported cases at home A to measured and reported cases at home B.

Tab. 4: Time series evidence

	Dependent variable: First outbreak indicator		
	(1)	(2)	(3)
New outbreak degree$_{t-2}$	0.0245*		
	(0.00810)		
New outbreak degree$_{t-1}$		0.0154^{+}	
		(0.00680)	
New outbreak degree$_{t}$			-0.0128
			(0.00972)
Constant	0.0657***	0.0670***	0.0730***
	(0.00327)	(0.00336)	(0.00349)
Fixed effects	County×Week	County×Week	County×Week
Observations	7,429	7,429	7,429
F-stat	9.142	5.156	1.7415
R^2	0.213	0.212	0.211
Within R^2	0.00309	0.00138	0.000412

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $^{*}p < 0.01$, $^{**}p < 0.001$, $^{***}p < 0.0001$. Dependent variable is a binary variable that equals 1 for the nursing home-week in which a home first reports having a COVID case using state data for Colorado, Connecticut, and Florida.

8 Discussion and Conclusions

Using a large-scale analysis of smartphone location data, we document substantial connections among nursing homes after nationwide visitor restrictions were enacted in March 2020. Consistent with the CDC's conclusion that shared workers were a source of infection for the nursing home outbreak in Kirkland, Washington (McMichael, 2020), our network measures suggest that staff linkages between nursing homes are a significant predictor of SARS-CoV-2 infections. Our general findings are robust to alternative specifications or the use of the case count data available from CMS. Clearly, there are limitations to drawing unambiguously causal inferences from an observational study. However, this is not an environment in which randomized controlled trials are feasible or ethical.

These results provide evidence for a policy recommendation of compensating nursing home workers to work at only one home and limit cross-traffic across homes. While some nursing homes and other long-term care facilities have undertaken actions to create a "staff bubble", this is still not a component of extant regulation (Sudo, 2020; Rodricks, 2020). Absent such regulation, allocation of PPE, testing, and other preventive measures should be targeted thoughtfully, recognizing the current potential for transmission across homes. New CMS testing guidelines as of August 2020 state that a nursing home not experiencing a current outbreak and located in a county with case positivity rates of less than five percent need only test staff members once per month (Centers for Medicare and Medicaid Services, 2020d). If two homes are known to share workers, however, testing could be increased at one home if an outbreak occurs at the other facility. Further, given the greater chance that a highly connected home experiences a new outbreak—and the risk this creates for its connections—more frequent testing of highly connected homes could be warranted, even when county positivity rates are low. While the nursing home population is particularly fragile, this research has implications for cross-linkages in other congregate settings such as assisted living homes, prisons, food-processing plants, and large workplace facilities.

9 References

H. R. Abrams, L. Loomer, A. Gandhi, and D. C. Grabowski. Characteristics of US Nursing Homes with COVID-19 Cases. *Journal of the American Geriatrics Society*, 2020.

A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004.

R. O. Cadigan, D. G. Stevenson, D. J. Caudry, and D. C. Grabowski. Private investment purchase and nursing home financial health. *Health services research*, 50(1):180–196, 2015.

N. G. Castle. Use of agency staff in nursing homes. *Research in Gerontological Nursing*, 2(3): 192–201, 2009.

Centers for Disease Control and Prevention. National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties, 2020. URL https://www.cdc.gov/nchs/data_access/urban_rural.htm.

Centers for Medicare and Medicaid Services. COVID-19 Nursing Home Data as of week ending 6/28/2020, 2020a. URL https://data.cms.gov/stories/s/COVID-19-Nursing-Home-Data/bkwz-xpvg/.

Centers for Medicare and Medicaid Services. Guidance for Infection Control and Prevention of Coronavirus Disease 2019 (COVID-19) in Nursing Homes, 2020b. URL https://www.cms.gov/files/document/3-13-2020-nursing-home-guidance-covid-19.pdf.

Centers for Medicare and Medicaid Services. Guidance for Infection Control and Prevention of Coronavirus Disease 2019 (COVID-19) in Nursing Homes. April 2, 2020, 2020c. URL https://www.cms.gov/files/document/4220-covid-19-long-term-care-facility-guidance.pdf.

Centers for Medicare and Medicaid Services. Interim Final Rule (IFC), CMS-3401-IFC, Additional Policy and Regulatory Revisions in Response to the COVID-19 Public Health Emergency related to Long-Term Care (LTC) Facility Testing Requirements and Revised COVID19 Focused Survey Tool , 2020d. URL https://www.cms.gov/files/document/qso-20-38-nh.pdf.

M. K. Chen and R. Rohla. The Effect of Partisanship and Political Advertising on Close Family Ties. *Science*, 360(6392):1020–1024, 2018.

Colorado Department of Public Health and Environment . Outbreak Data Archive, 2020. URL https://covid19.colorado.gov/covid19-outbreak-data.

M. Conlen et al. More Than 40% of U.S. Coronavirus Deaths Are Linked to Nursing Homes. *The New York Times*, 2020. URL https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html?

Connecticut Department of Public Health . Nursing homes with residents positive for covid-19.

H. D'Adamo, T. Yoshikawa, and J. G. Ouslander. Coronavirus Disease 2019 in Geriatrics and Long-term Care: The ABCDs of COVID-19: COVID-19 in Geriatrics and Long-Term Care. *J Am Geriatr Soc*, 2020.

J. W. Dyal et al. COVID-19 Among Workers in Meat and Poultry Processing Facilities-19 States, April 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69, 2020.

N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, 2020.

Florida Department of Health. COVID-19 Report Archive, 2020. URL `http://ww11.doh.state.fl.us/comm/_partners/covid19_report_archive/`.

Girvan, Gregg and Roy, Avik. Foundation on Research on Equal Opportunity: Long Term Care Facilities Reported COVID-19 Deaths by State, 2020. URL `https://docs.google.com/spreadsheets/d/17JmyFzOd3ZEYCGpP0mK51S_P10yPgCuYF8PYALKuTs8/edit#gid=0`.

M. He, Y. Li, and F. Fang. Is there a Link between Nursing Home Reported Quality and COVID-19 Cases? Evidence from California Skilled Nursing Facilities. *Journal of the American Medical Directors Association*, 2020.

Kaiser Family Foundation. State data and policy actions to address coronavirus, 2020. URL `https://www.kff.org/health-costs/issue-brief/state-data-and-policy-actions-to-address-coronavirus/`.

S. A. Kinner, J. T. Young, K. Snow, L. Southalan, D. Lopez-Acuña, C. Ferreira-Borges, and É. O'Moore. Prisons and custodial settings are part of a comprehensive response to COVID-19. *The Lancet Public Health*, 5(4):e188–e189, 2020.

R. T. Konetzka. Senate Testimony before the Senate Special Committee on Aging, 2020. URL `https://www.aging.senate.gov/imo/media/doc/SCA_Konetzka_05_21_20.pdf`.

S. N. Ladhani, J. Y. Chow, R. Janarthanan, and et. al. Increased risk of SARS-CoV-2 infection in staff working across different care homes; enhanced CoVID-19 outbreak investigations in London care Homes. *Journal of Infection*, 2020.

Y. Li, H. Temkin-Greener, S. Gao, and X. Cai. Covid-19 infections and deaths among connecticut nursing home residents: facility correlates. *Journal of the American Geriatrics Society*, 2020.

S. F. Lu and L. X. Lu. Do mandatory overtime laws improve quality? Staffing decisions and operational flexibility of nursing homes. *Management Science*, 63(11):3566–3585, 2017.

T. M. McMichael. COVID-19 in a long-term care facility — King County, Washington, February 27–March 9, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69, 2020.

T. M. McMichael, D. W. Currie, S. Clark, S. Pogosjans, M. Kay, N. G. Schwartz, J. Lewis, A. Baer, V. Kawakami, M. D. Lukoff, et al. Epidemiology of Covid-19 in a long-term care facility in King County, Washington. *New England Journal of Medicine*, 382(21):2005–2011, 2020.

New York State Department of Health. Factors Associated with Nursing Home Infections and Fatalities in New York State During the COVID-19 Global Health Crisis, 2020. URL https://health.ny.gov/press/releases/2020/docs/nh_factors_report.pdf.

K. Pillemer, L. Subramanian, and N. Hupert. The Importance of Long-term Care Populations in Models of COVID-19. *JAMA*, 324(1):25–26, 2020.

D. Rodricks. How A West Baltimore Nursing Home has Zero COVID Infections. *Baltimore Sun*, 2020. URL https://www.baltimoresun.com/opinion/columnists/dan-rodricks/bs-md-rodricks-0619-20200618-re2obahhbzbddojazrkf65gdoq-story.html.

P. Rowan, R. Lester, M. Levere, J. Libersky, D. Lipson, J. Robison, et al. A study of the covid-19 outbreak and response in connecticut long-term care facilities. Technical report, Mathematica Policy Research, 2020.

V. W. Slaugh, A. A. Scheller-Wolf, and S. R. Tayur. Consistent staffing for long-term care through on-call pools. *Production and Operations Management*, 27(12):2144–2161, 2018.

L. L. Stanley. Influenza at San Quentin Prison, California. *Public Health Reports (1896-1970)*, 34 (19):996–1008, 1919. ISSN 00946214. URL http://www.jstor.org/stable/4575142.

C. Sudo. How Three Senior Living Communities Bubbled Communities by Having Staff Live on Site. *Senior Housing News*, 2020. URL https://seniorhousingnews.com/2020/05/11/how-3-senior-living-providers-bubbled-communities-by-having-staff-live-on-site/.

F. A. Tobolowsky, E. Gonzales, J. L. Self, C. Y. Rao, R. Keating, G. E. Marx, T. M. McMichael, M. D. Lukoff, J. S. Duchin, K. Huster, et al. Covid-19 outbreak among three affiliated homeless service sites-king county, washington, 2020. *Morbidity and Mortality Weekly Report*, 69(17):523, 2020.

C. H. Van Houtven, N. DePasquale, and N. B. Coe. Essential long-term care workers commonly hold second jobs and double- or triple-duty caregiving roles. *Journal of the American Geriatrics Society*, 2020. doi: 10.1111/jgs.16509.

E. M. White, C. M. Kosar, R. A. Feifer, C. Blackman, S. Gravenstein, J. Ouslander, and V. Mor. Variation in SARS-CoV-2 Prevalence in US Skilled Nursing Facilities. *Journal of the American Geriatrics Society*, 2020.

T. Williams and R. Griesbach. San Quentin Prison Was Free of the Virus. One Decision Fueled an Outbreak. *The New York Times*, 2020. URL https://www.nytimes.com/2020/06/30/us/san-quentin-prison-coronavirus.html.

10 Supplementary Materials

In this Supplement, we provide tables with network summary statistics by state, as well as full results from our main regression specification and three robustness checks. In Table S1, we report mean values and standard deviations of our four network measures, for all 48 contiguous U.S. and the District of Columbia.

Next, we show all coefficient estimates for our main analysis in Table S2. Table S3 show results for our main network measures, *Degree* and *Strength*, by individual state, for the 22 states with detailed COVID-19 case counts prior to the mandated reporting by CMS in late-May.

In Table S4, we repeat our main analysis replacing the inverse hyperbolic sine of the number of cases in the nursing home with a binary variable that equals 1 if the nursing home has reported COVID cases among its residents. Here, for example, the results in column (1) suggest that, when 10 additional nursing home connections are added, the probability that a home has cases increases by 6.6 percentage points. In the overall data, 42 percent of homes have cases. The network measures are all statistically significant in this alternative specification. In Table S5 we repeat our main analysis replacing the state fixed effects with county fixed effects . This allows a smaller number of units within which variation can be measured. The urban variable, which is measured at the county-level, is omitted. Results are qualitatively similar to our state fixed effects data though, as expected, significance levels diminish somewhat.

As a final robustness check, we repeat our main analysis in Table S6 using data from CMS rather than data from the individual states. This allows us to examine the 48 continental United States plus the District of Columbia, but these data are subject to the reporting limitation that homes were not required to add cases prior to May into their cumulative case totals. The CMS data reports cases in 48 percent of nursing homes by week-ending May 31, 2020 but reports overall fewer cases than the individual state data. This is expected since the CMS data did not require homes to report cases in the cumulative total that had resolved before May 2020. In this robustness specification, point estimates for the network variables are slightly smaller than in our base specifications but qualitatively extremely similar.

Tab. S1: State-level network summary statistics, mean and standard deviation

	Cases per Nursing Home		Degree		Strength		Wt. Avg. Nbor Degree		Eigenvector Centrality	
State	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AL	10.0	19.9	5.8	4.8	8.3	8.5	7.9	4.1	0.15	0.21
AR	4.5	16.1	3.4	3.2	4.5	5.0	4.6	3.3	0.09	0.18
AZ	6.1	14.2	4.9	5.0	6.4	7.6	7.1	5.2	0.14	0.21
CA	11.0	26.3	6.0	8.2	7.3	13.0	9.5	8.5	0.04	0.13
CO	8.0	16.9	3.8	4.4	4.5	5.2	5.5	4.7	0.11	0.18
CT	36.5	54.6	6.6	6.7	7.3	7.6	9.6	6.0	0.13	0.21
DC	26.8	39.7	6.3	2.4	9.6	4.7	7.4	0.9	0.65	0.23
DE	23.8	37.8	3.0	2.5	3.5	3.4	4.4	2.6	0.27	0.26
FL	5.6	15.4	11.4	8.9	14.4	12.1	15.4	6.7	0.09	0.15
GA	17.9	42.5	8.6	8.6	12.4	17.7	11.7	8.0	0.10	0.18
IA	4.1	11.2	3.7	3.6	4.6	5.2	5.4	3.8	0.10	0.19
ID	4.6	30.7	1.5	1.7	2.0	2.3	2.0	1.9	0.15	0.26
IL	15.8	32.5	9.3	11.6	10.6	13.6	13.9	11.6	0.07	0.18
IN	6.8	20.5	7.4	8.2	9.2	11.5	11.4	8.0	0.08	0.15
KS	1.8	7.1	5.3	5.9	6.6	7.9	7.8	5.9	0.08	0.19
KY	5.2	17.2	5.1	5.3	7.2	11.6	7.4	5.4	0.13	0.20
LA	13.8	24.1	5.7	5.6	7.3	7.4	7.8	5.1	0.10	0.23
MA	33.4	48.2	6.5	6.4	7.2	7.4	9.8	5.9	0.10	0.18
MD	20.3	36.2	12.9	13.6	15.3	17.5	19.3	11.4	0.19	0.28
ME	3.7	12.1	1.3	1.7	1.6	2.3	1.6	1.9	0.09	0.24
MI	11.9	24.5	4.5	4.7	5.4	6.0	6.2	4.7	0.05	0.17
MN	6.6	16.5	3.6	4.2	4.2	5.1	5.6	5.2	0.09	0.18
MO	4.7	16.8	6.0	7.0	7.7	10.5	9.0	7.0	0.07	0.16
MS	7.9	19.6	4.1	4.2	5.5	6.5	6.2	4.3	0.10	0.24
MT	1.1	4.5	1.5	2.0	1.9	2.7	1.9	2.2	0.13	0.24
NC	7.0	20.4	6.8	6.7	8.6	11.0	9.7	5.9	0.08	0.18
ND	3.3	8.2	1.9	3.1	3.6	7.1	6.7	8.1	0.11	0.15
NE	4.1	13.2	2.4	2.5	2.7	3.0	3.4	2.7	0.10	0.17
NH	17.4	46.7	2.8	3.2	3.1	3.9	4.1	3.6	0.15	0.31
NJ	33.9	49.8	10.4	8.2	12.6	11.0	15.0	6.0	0.23	0.20
NM	3.2	11.7	3.8	4.1	5.4	7.8	5.0	4.3	0.17	0.23
NV	17.7	34.9	3.3	3.1	4.0	4.4	4.2	3.1	0.29	0.35
NY	24.3	46.3	7.8	9.4	10.6	14.7	12.6	9.6	0.07	0.20
OH	5.6	15.9	8.8	8.0	10.9	11.3	13.3	7.6	0.05	0.10
OK	2.2	8.2	5.6	6.5	6.9	8.6	7.8	6.4	0.09	0.19
OR	2.8	7.2	3.0	3.4	3.4	3.9	4.2	3.6	0.10	0.21
PA	19.1	39.2	7.2	7.9	9.8	15.6	10.6	7.8	0.07	0.16
RI	22.5	38.8	4.6	3.8	5.0	4.4	6.7	3.5	0.23	0.23
SC	9.7	23.7	5.1	4.4	7.1	6.8	6.7	4.0	0.10	0.20
SD	2.1	6.2	1.0	1.2	1.2	1.8	1.4	1.6	0.05	0.18
TN	2.9	8.4	5.7	4.9	7.9	8.1	8.6	4.7	0.11	0.17
TX	3.8	18.8	9.2	10.4	13.4	50.2	14.3	10.6	0.04	0.12
UT	4.4	25.8	2.8	3.3	3.8	5.0	4.2	3.6	0.13	0.22
VA	9.3	23.4	6.8	7.4	8.3	9.3	9.8	6.8	0.11	0.22
VT	9.2	19.9	0.5	1.1	0.6	1.3	0.6	1.2	0.12	0.31
WA	6.3	18.6	4.3	5.8	4.7	6.4	6.0	6.0	0.11	0.27
WI	4.0	10.9	3.2	4.4	4.1	6.8	4.8	5.1	0.06	0.19
WV	3.3	10.8	3.6	3.7	5.3	8.0	5.8	4.5	0.14	0.24
WY	1.1	3.1	0.5	0.9	0.8	1.8	0.5	1.0	0.11	0.29

COVID-19 cases include confirmed and suspected cases among residents reported to CMS as of May 31, 2020. *Degree* is the number of nursing homes that a particular home is connected to through a smartphone observed in both facilities. *Strength* is the total number of smartphones observed in a nursing home and other connected homes. *Weighted average neighbor degree* is the average number of connections a nursing home's neighbor has, weighted by the pair strength. *Eigenvector centrality* measures the extent to which a nursing home's neighbors are highly connected, and is calculated within each state and ranges from 0 to 1.

Tab. S2: Detailed covariates of COVID case count specifications

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Beds	0.00957***	0.0103***	0.00981***	0.00936***	0.0103***
	(0.000744)	(0.000744)	(0.000737)	(0.000740)	(0.000744)
$Beds^2$	-0.00000554**	-0.00000635**	-0.00000533**	-0.00000501**	-0.00000621**
	(0.00000171)	(0.00000172)	(0.00000170)	(0.00000170)	(0.00000172)
High proportion on Medicaid	0.0891	0.0925^{+}	0.0783	0.0797	0.0936
	(0.0456)	(0.0459)	(0.0428)	(0.0453)	(0.0459)
High proportion of Black residents	0.526***	0.547***	0.491***	0.489***	0.563***
	(0.0588)	(0.0591)	(0.0587)	(0.0586)	(0.0591)
CMS rating 1	-0.0656	-0.0615	-0.0685	-0.0689	-0.0650
	(0.0701)	(0.0706)	(0.0698)	(0.0697)	(0.0706)
CMS rating 2	-0.0032	0.00262	-0.00924	-0.0101	-0.00423
	(0.0633)	(0.0637)	(0.0630)	(0.0629)	(0.0637)
CMS rating 3	0.0593	0.0681	0.0592	0.0558	0.0609
	(0.0634)	(0.0638)	(0.0631)	(0.0630)	(0.0638)
CMS rating 4	-0.0204	-0.0141	-0.0193	-0.0216	-0.0215
	(0.0585)	(0.0589)	(0.0583)	(0.0582)	(0.0589)
Infect Violations	-0.0750	-0.0691	-0.0677	-0.0725	-0.0783
	(0.0505)	(0.0508)	(0.0503)	(0.0502)	(0.0508)
Urban indicator	0.734***	0.782***	0.662***	0.656***	0.770***
	(0.0563)	(0.0564)	(0.0568)	(0.0566)	(0.0566)
For profit indicator	0.224***	0.249***	0.207***	0.200***	0.234***
	(0.0495)	(0.0497)	(0.0493)	(0.0492)	(0.0498)
Node degree	0.0343***			0.0242***	
	(0.00255)			(0.00508)	
Node strength		0.0163***		-0.00610^{+}	
		(0.00166)		(0.00297)	
Weighted average neighbor degree			0.0409***	0.0299***	
			(0.00267)	(0.00344)	
Eigenvector centrality in state					1.044***
					(0.109)
Fixed effects	State	State	State	State	State
Observations	6,337	6,337	6,337	6,337	6,337
F-stat	123.4	114.9	128.7	112.9	114.5
R^2	0.408	0.400	0.412	0.415	0.399
Within R^2	0.189	0.178	0.195	0.199	0.177

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $*p < 0.01$, $**p < 0.001$, $***p < 0.0001$. Dependent variable is the inverse hyperbolic sine of COVID cases in a nursing home.

Tab. S3: Covariates of COVID-19 cases within nursing homes, by individual state

	Dependent variable: $sinh^{-1}(Cases)$									
	CA		CT		CO		DC		FL	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Node degree	0.0262***		0.0409$^+$		0.0188		0.0872		0.0123	
	(0.00620)		(0.0193)		(0.0320)		(0.227)		(0.00545)	
Node strength		0.0141**		0.0362$^+$		0.0207		0.0190		0.00573
		(0.00381)		(0.0172)		(0.0263)		(0.122)		(0.00402)
Observations	1034	1034	199	199	178	178	17	17	575	575
F-stat	12.66	12.27	6.905	6.898	3.489	3.518	1.898	1.847	3.83	3.561
R^2	0.130	0.126	0.308	0.308	0.202	0.204	0.760	0.755	0.076	0.0707

	Dependent variable: $sinh^{-1}(Cases)$									
	GA		IL		KY		LA		MA	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Node degree	0.0474***		0.0269***		0.031		0.0146		0.0480***	
	(0.0120)		(0.00592)		(0.0168)		(0.0202)		(0.0113)	
Node strength		0.0125$^+$		0.0219***		0.00340		0.00950		0.0403***
		(0.00576)		(0.00509)		(0.00782)		(0.0155)		(0.00986)
Observations	338	338	637	637	256	256	202	202	363	363
F-stat	6.395	5.315	49.02	48.69	2.647	2.349	1.682	1.668	8.234	8.114
R^2	0.191	0.164	0.485	0.484	0.116	0.104	0.097	0.0958	0.220	0.218

	Dependent variable: $sinh^{-1}(Cases)$									
	MD		MI		MS		NC		ND	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Node degree	0.0195$^+$		0.0325		-0.0539*		0.0131		-0.026	
	(0.00833)		(0.0208)		(0.0204)		(0.0110)		(0.0438)	
Node strength		0.0140$^+$		0.0227		-0.0296$^+$		0.000479		-0.00984
		(0.00635)		(0.0162)		(0.0136)		(0.00680)		(0.0196)
Observations	150	150	367	367	141	141	401	401	64	64
F-stat	6.16	6.082	12.2	12.15	1.454	1.253	1.994	1.871	4.473	4.455
R^2	0.350	0.348	0.293	0.292	0.120	0.105	0.058	0.0547	0.486	0.485

Tab. S3: Covariates of COVID-19 cases within nursing homes, by individual state (continued)

	Dependent variable: $sinh^{-1}(Cases)$									
	NH		NJ		PA		RI		SC	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Node degree	0.200*		0.00629		0.0778***		0.264**		0.0175	
	(0.0605)		(0.0102)		(0.00971)		(0.0740)		(0.0227)	
Node strength		0.180**		0.000707		0.0188***		0.208*		0.00129
		(0.0506)		(0.00752)		(0.00472)		(0.0646)		(0.0143)
Observations	66	66	332	332	645	645	74	74	155	155
F-stat	3.152	3.359	10.8	10.76	25.86	20.41	4.108	3.775	0.68	0.629
R^2	0.391	0.406	0.271	0.270	0.329	0.279	0.395	0.375	0.054	0.0505

	Dependent variable: $sinh^{-1}(Cases)$			
	TN		WV	
	(1)	(2)	(1)	(2)
Node degree	0.00275		0.0105	
	(0.0432)		(0.0281)	
Node strength		0.00349		0.0119
		(0.0310)		(0.0133)
Observations	40	40	103	103
F-stat	0.792	0.793	1.203	1.273
R^2	0.260	0.261	0.127	0.133

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $*p < 0.01$, $**p < 0.001$, $***p < 0.0001$.
Dependent variable is inverse hyperbolic sine of COVID cases in the nursing home.
Demographics include number of beds, high proportion of Black residents, and high proportion on Medicaid.
CMS quality is a 1-5 categorical rating.

Tab. S4: Covariates of the existence of nursing home COVID-19 cases

	Dependent variable: Nursing home has > 0 cases				
	(1)	(2)	(3)	(4)	(5)
Node degree	0.00703*** (0.000679)			0.00410* (0.0013)	
Node strength		0.00319*** (0.000442)		-0.00167$^+$ (0.000790)	
Weighted average neighbor degree			0.00981*** (0.000710)	0.00847*** (0.000915)	
Eigenvector centrality in state					0.180*** (0.0289)
Fixed effects	State	State	State	State	State
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	6,337	6,337	6,337	6,337	6,337
F-stat	76.02	70.87	83.88	72.64	69.60
R^2	0.326	0.3146	0.335	0.336	0.319
Within R^2	0.125	0.117	0.136	0.137	0.115

Standard errors in parentheses. Significance levels: $^+p < 0.05$, $*p < 0.01$, $**p < 0.001$, $***p < 0.0001$.
Dependent variable is a binary indicator that equals 1 if COVID cases are reported in a nursing home.
Demographics include number of beds, high proportion of Black residents, and high proportion on Medicaid.
CMS quality is a 1-5 categorical rating.

Tab. S5: Covariates of nursing home COVID-19 cases with county fixed effects

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Node degree	0.0193*** (0.00264)			0.0217*** (0.005088)	
Node strength		0.00937*** (0.00171)		-0.00304 (0.003066)	
Weighted average neighbor degree			0.0148*** (0.00313)	0.00318 (0.00373)	
Eigenvector centrality in state					0.684*** (0.119)
Fixed effects	County	County	County	County	County
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	5,967	5,967	5,967	5,967	5,967
F-stat	61.78	59.40	58.63	52.41	59.73
R^2	0.569	0.567	0.566	0.569	0.567
Within R^2	0.113	0.109	0.108	0.113	0.110

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $*p < 0.01$, $**p < 0.001$, $***p < 0.0001$. Dependent variable is inverse hyperbolic sine of COVID cases in the nursing home using individual state data. Demographics include number of beds, high proportion of Black residents, and high proportion on Medicaid. CMS quality is a 1-5 categorical rating.

Tab. S6: Covariates of COVID-19 cases within nursing homes using CMS data

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Node degree	0.0207*** (0.00187)			0.0164*** (0.00300)	
Node strength		0.00639*** (0.000964)		-0.00219 (0.00135)	
Weighted average neighbor degree			0.0204*** (0.00190)	0.0120*** (0.00241)	
Eigenvector centrality in state					0.623*** (0.0759)
Fixed effects	State	State	State	State	State
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	13,165	13,165	13,165	13,165	13,165
F-stat	157.3	149.8	156.6	137.0	152.0
R^2	0.258	0.253	0.257	0.259	0.254
Within R^2	0.125	0.120	0.125	0.127	0.121

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $*p < 0.01$, $**p < 0.001$, $***p < 0.0001$.
Dependent variable is inverse hyperbolic sine of COVID cases in the nursing home using CMS data.
Demographics include number of beds, high proportion of Black residents, and high proportion on Medicaid.
CMS quality is a 1-5 categorical rating.

Fig. S1: Fraction of state nursing homes with a reported COVID-19 outbreak over time.