NURSING HOME STAFF NETWORKS AND COVID-19

M. Keith Chen
Judith A. Chevalier
Elisa F. Long

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 crucial role these connections serve in spreading a highly contagious respiratory infection---is currently unknown given the lack of centralized data on cross-facility nursing home employment. In this paper, we perform the first large-scale analysis of nursing home connections via shared staff using device-level geolocation data from 30 million smartphones, and find that 7 percent of smartphones appearing in a nursing home also appeared in at least one other facility---even after visitor restrictions were imposed. We construct network measures of nursing home connectedness and estimate that nursing homes have, on average, connections with 15 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. Results suggest that eliminating staff linkages between nursing homes could reduce COVID-19 infections in nursing homes by 44 percent.

M. Keith Chen
UCLA/Anderson School
110 Westwood Plaza
Cornell Hall, Suite C513
Los Angeles, CA 90095
keith.chen@anderson.ucla.edu

Elisa F. Long
UCLA/Anderson School
110 Westwood Plaza
Gold Hall, Suite B508
Los Angeles, CA 90095
elisa.long@anderson.ucla.edu

Judith A. Chevalier
Yale School of Management
165 Whitney Ave
New Haven, CT 06511
and NBER
judith.chevalier@yale.edu

1   Introduction

Linked to more than 316,000 COVID-19 cases and 57,000 deaths—nearly half of all U.S. fatalities—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). 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 (Strausbaugh et al., 2003; Lansbury et al., 2017). 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 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 of 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 only residents, staff, and contractors (Centers for Medicare and Medicaid Services, 2020b). Nevertheless, infections have subsequently broken out in nursing homes, suggesting the unwitting introduction of the virus into homes by staff as one potential channel. In particular, the practice of employing nursing home workers across multiple care facilities may play an important role in the spread of COVID-19, 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).

Despite this early recognition of cross-traffic between congregate settings as an important potential 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).

Using novel device-level geolocation data for 509,603 smartphones observed in at least one of the 15,307 nursing homes in the continental U.S., we find that 7 percent of individuals entering a nursing home also entered at least one other nursing home in the six-week period following the March 13th nationwide restriction on nursing home visitors. We construct several measures from network theory to characterize nursing-home connectedness, 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, inspectors, 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, 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 which homes suffered large outbreaks.

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. 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). Movement of staff and residents across three affiliated homeless shelters likely contributed to outbreaks in each location (Tobolowsky et al., 2020). Employees at meat and other 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 partial 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 hindered by the fact that CMS did not require nursing homes to submit data on COVID-19 cases and fatalities until May 2020.

For our main data analysis, we use the disclosures of individual state Departments of Public Health to determine cumulative nursing home COVID cases. From the 23 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 Supplementary Information, 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.

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 period March 13 to April 23, 2020. CMS had announced guidelines preventing nursing home social visitors starting March 13, 2020. Thus, over the period for which smartphone activity was measured, the smartphones detected in nursing homes would likely belong to staff, contractors, and residents. Summary statistics of nursing homes for both the 23 states for which we have assembled state data and the set of facilities regulated by CMS in the continental U.S. for which we have complete data are found in Table 1.

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

Variable	State reporting facilities	CMS reporting facilities
Number of nursing homes	6,644	12,775
Demographics		
High proportion (>25%) of Black residents, %	16.1	12.7
High proportion (>50%) on Medicaid, %	33.0	28.3
Urban location, %	80.5	72.5
Regulatory measures		
Number of beds	114 (58.5)	110 (60.6)
CMS quality rating (1-5)	3.18 (1.42)	3.15 (1.42)
Has infection violations, %	75.3	75.7
Network metrics		
Node degree	15.6 (17.5)	14.3 (16.5)
Node strength	20.9 (26.3)	19.7 (30.8)
Weighted average neighbor degree	24.4 (20.7)	24.5 (33.6)
Eigenvector centrality in state	0.14 (0.22)	0.13 (0.22)

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.

Smartphone Location Data

We estimate staff 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 smartphone user has consented to the use of their anonymized data. Previous studies of these data have found them to be highly representative of the U.S. on numerous demographic dimensions (Chen and Rohla, 2018). Of the roughly 30 million smartphones we include in our U.S. sample, we identify 509,603 smartphones that appear to enter at least one U.S. nursing home in the six-week period of our study.

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 through the use of the Google Maps API, and then match that location to a satellite-image machine-learned shapefile of the convex-hull of the building's rooftop (provided by Microsoft / Open Streetmaps). Using these rooftop shapefiles, we find all smartphones in our data that have ever recorded more than two location traces in the same nursing home facility during our study period, when visitors were explicitly barred. By identifying smartphones that entered more than one nursing home, we estimate a nursing home staff-contact network.

Network Metrics

The contact structure among nursing home facilities 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 estimate 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.

A facility's *degree* k_i equals the total number of other nursing home facilities with a connection to facility i (i.e., the number of *neighbors* of node i).

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

Strength s_i is the weighted sum of connections to 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 centrality measure is computed using the principal eigenvector of the adjacency matrix, rewritten in matrix notation as $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. We then normalize these values to range between 0 and 1 within each state's network.

Empirical Specification

Our main specification examines the predictors of the number of cases in a nursing home as a function of various explanatory variables. Specifically, we include the home's *demographic* characteristics, including the number of beds and the squared number of beds to allow for nonlinearities. Following previous literature, 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 also include *CMS quality* measures of nursing homes as done previously (Abrams et al., 2020). CMS rates nursing homes on a five-point scale. Our specification includes indicator variables for one-star, two-stars, three-stars, and four-stars, with the omitted category being five-stars (the highest possible rating). We include a binary indicator variable that equals one if the home had *infection control violations* in its most recent inspection. 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).

We examine whether nursing home connectivity predicts COVID-19 cases using 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-

19 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 facility's network connectivity, as described in the previous section. The vector $\mathbf{F}_i$ is a set of fixed effects, for the geographical unit of the home (state or county), and 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. In the Supplement, we include coefficients for all variables in Table S1. Table S2 replaces the dependent variable with a binary indicator variable if the nursing home has had any COVID-19 cases. Table S3 repeats our analysis using county fixed effects. This robustness check is important given the finding in (White et al., 2020) that county-level SARS-CoV-2 prevalence predicts case counts in nursing facilities. Finally, Table S4 replaces the data for 23 states with the larger CMS dataset for the continental U.S.

5 Network Measures

Nursing homes display a wide range of connectedness to other facilities. Average *degree*—the number of other facilities a nursing home shares at least one smartphone connection with—across all U.S. nursing homes is $\langle k \rangle = 14.3$, but this ranges from a state-average degree below 2 in Montana, South Dakota, Vermont, and Wyoming, to an average exceeding 20 in Florida, Illinois, Maryland, New Jersey, and Texas. Among nursing homes with confirmed or suspected cases, as reported to CMS, average degree is 17.3 compared to 11.8 among homes with no documented cases ($t = 19.2$, $p < 0.0001$), with a similar effect across the entire degree distribution (see Fig. 1). Average *strength*—the total number of smartphones appearing in a nursing home and one of its neighbors—is also greater in COVID-positive homes (23.7 vs. 16.4, $t = 13.4$, $p < 0.0001$).

To illustrate how degree, strength, and other 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. Here, nodes represent 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 contact with the hub) in dark grey and its neighbors' neighbors in light grey. Node size corresponds to the number of confirmed and suspected COVID-19 cases (from CMS data to allow for continental U.S. coverage) and edge color corresponds to the number of smartphones observed in that 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 Connecticut sub-network (Fig. 2A),

for instance, the focal nursing home had a large outbreak of 102 cases, yet is directly connected to only six other homes (*degree* = 6), with one smartphone observed in each pair (*strength* = 6), well below our sample mean or Connecticut's average degree of 13. However, the hub's neighbors are highly connected themselves, resulting in a *weighted average neighbor degree* of 26, well above the state's average.

The hub node in Georgia (Fig. 2B) has average degree, but with strength double the state's average, an indicator that many staff members are likely working at this facility and its 16 neighbors. Crucially, these individuals are potential conduits of SARS-CoV-2 transmission given the home's 189 cases by the end of May, one of the largest outbreaks in the state. Likewise, the hub node in the Illinois sub-network (Fig. 2C) is not the most connected home in the state with a *degree* of 42, yet this facility can play an important role in viral spread by acting as a bridge between geographically distant nursing homes. In Massachusetts (Fig. 2D), the selected hub node is quite central to this network, with an eigenvector centrality of 0.75, substantially higher than the state average of 0.16, because this particular facility is directly connected to many other central nodes. The North Carolina network (Fig. 2E) illustrates a hub node connected to each of four other nursing homes, which share several contacts among themselves. Here, the focal facility faced a large outbreak of 181 cases, potentially increasing the risk of transmission to its neighboring homes. Lastly, with more

Fig. 1: Degree distribution of nursing homes with and without COVID cases (reported to CMS as of May 31, 2020).
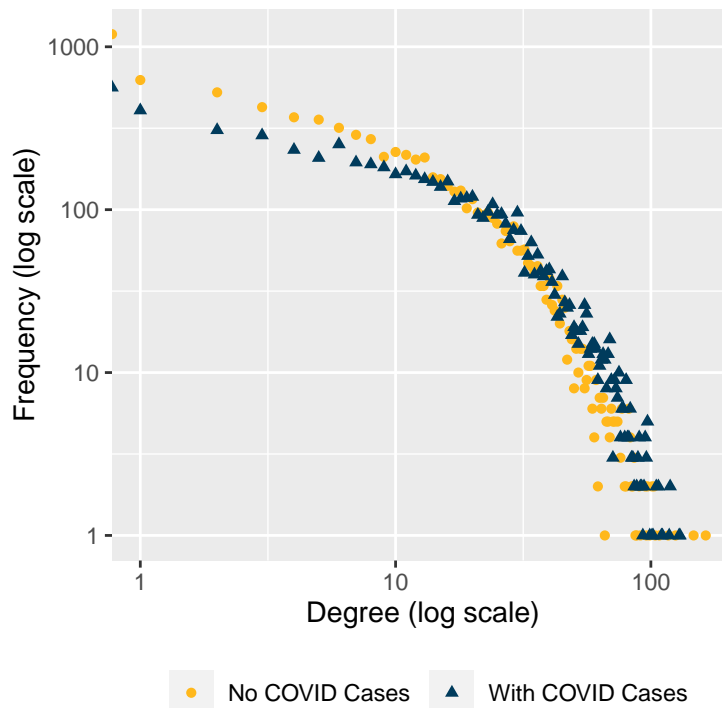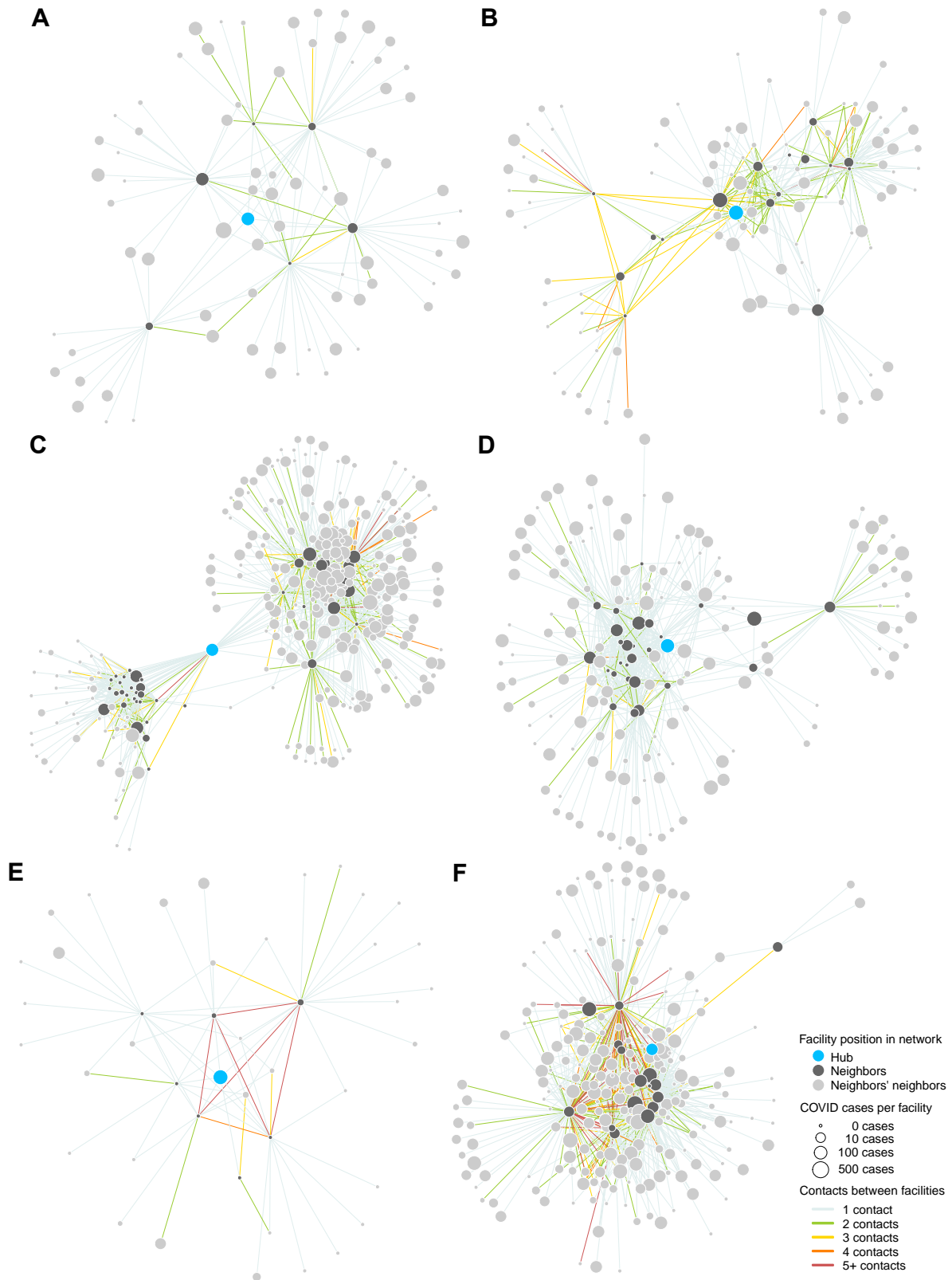
Fig. 2: Network structure of selected nursing home facilities in six U.S. states. Details provided in Table 2.

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

Hub Facility	State	COVID-19 Cases	Degree	Strength	Wtd. Avg. Neigh. Deg.	Eigenvector Centrality
A	CT	102	6	6	26.2	0.13
B	GA	189	16	30	18.2	0.15
C	IL	72	42	51	45.1	0.12
D	MA	134	27	27	32.5	0.75
E	NC	181	7	7	23.0	0.02
F	NY	57	16	36	61.1	0.17

than 600 nursing homes in New York State (Fig. 2F), average degree is higher than in most states. With a weighted average neighbor degree of 61, this particular hub node is connected to several other central nodes, but unlike the Massachusetts network, many nursing home pairs in New York appear to share multiple contacts.

6 Regression 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 Supplementary Table S3. 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 10 neighbors (average degree is 15), the expected number of COVID-19 cases increases by $13.7 \times 1.95 = 26.7\%$. 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 10 contacts (average strength is 21), expected cases of COVID-19 increases by $6.79 \times 1.78 = 12.1\%$. Column (3) replaces the degree and strength measures with *weighted average neighbor degree* $\bar{k}_i^w$—the weighted average degree of a nursing home's neighbors (Barrat et al., 2004). Here, an increase of 10 (mean is 25.6) would lead to an expected $14.7 \times 2.09 = 30.7$ percent increase in cases. Column (4) uses our final network measure, *eigenvector centrality* v_i—which measures the extent to which the 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 ($v_i = 1$), the total increase in expected cases exceeds 190%.

Column (5) is our final specification, which includes both weighted average neighbor degree and eigenvector centrality, summarizing both the local and global connectivity of a nursing home. Intuitively, regression 5 compares demographically and geographically situated nursing homes of similar quality, which are thus likely exposed to similar risk of community-spread. When infections co-vary with differences in staff-network connectivity and centrality, these increased infections are likely attributable to shared-staff transmitting the virus across multiple nursing homes. As a useful counterfactual, regression 5 suggests that severing all staff-linkage transmission between facilities could reduce nursing home resident COVID-19 cases by 43.6%.

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

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

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Urban indicator	0.649***	0.686***	0.609***	0.650***	0.574***
	(0.054)	(0.054)	(0.053)	(0.053)	(0.054)
Node degree	0.0137***				
	(0.0012)				
Node strength		0.00679***			
		(0.00079)			
Weighted average neighbor degree			0.0147***		0.0116***
			(0.0010)		(0.0011)
Eigenvector centrality in state				1.073***	0.643***
				(0.091)	(0.099)
Fixed effects	State	State	State	State	State
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	6,644	6,644	6,644	6,644	6,644
F-stat	123.8	117.9	132.0	125.0	125.3
R^2	0.412	0.407	0.419	0.413	0.422
Within R^2	0.169	0.163	0.179	0.171	0.184

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.

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 Supplementary Table S1).

7   Discussion and Conclusions

Using a large-scale analysis of smartphone location data, we document substantial connections amongst 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 individuals moving between nursing homes is 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 any observational study. However, this is not an environment in which randomized controlled trials are feasible or ethical.

These results provide evidence of the magnitude of the benefits that would derive from compensating nursing home workers to work at only one home and limiting 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 preventative measures should be targeted thoughtfully, recognizing the current potential for cross-transmission across homes. While the nursing home population is particularly fragile, this research has implications for cross-linkages amongst other congregate settings such as assisted living homes, prisons, or large workplace facilities such as food-processing plants.

8   Bibliography

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.

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

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?

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.

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.

L. E. Lansbury, C. S. Brown, and J. S. Nguyen-Van-Tam. Influenza in long-term care facilities. *Influenza and Other Respiratory Viruses*, 11(5):356–366, 2017.

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.

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.

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.

L. J. Strausbaugh, S. R. Sukumar, C. L. Joseph, and K. P. High. Infectious disease outbreaks in nursing homes: an unappreciated hazard for frail elderly persons. *Clinical Infectious Diseases*, 36(7):870–876, 2003.

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.

9   Supplementary Materials

In this Supplement, we provide tables with additional detail from our Table 3 and three robustness checks of our main empirical specification. First, we show all coefficient estimates for our main analysis (Table S1). Second, we repeat our main analysis, replacing the inverse hyperbolic sine of the number of cases in the nursing home with a binary variable that takes the value of 1 if the nursing home has cases (Table S2). 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 3 percentage points. In the data overall, 42 percent of homes have cases. The network measures are all statistically significant in this alternative specification. Third, we repeat our main analysis replacing the state fixed effects with county fixed effects (Table S3). 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 using data from CMS rather than data from the individual states (Table S4). 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 45 percent of nursing homes 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.

Finally, we report mean values and standard deviations of our four network measures, for all 48 contiguous U.S. and the District of Columbia (Table S5).

Tab. S1: Detailed covariates of COVID-19 cases within nursing homes.

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Beds	0.00920***	0.00974***	0.00983***	0.00948***	0.00933***
	(0.000721)	(0.000721)	(0.000709)	(0.000717)	(0.000711)
Beds^2	-0.00000496*	-0.00000566**	-0.00000527*	-0.00000517*	-0.00000475*
	(0.00000167)	(0.00000167)	(0.00000165)	(0.00000166)	(0.00000165)
High proportion on Medicaid	0.105$^+$	0.112*	0.0833	0.112*	0.0888$^+$
	(0.0433)	(0.0434)	(0.0431)	(0.0432)	(0.0430)
High proportion of Black residents	0.533***	0.551***	0.519***	0.515***	0.485***
	(0.0575)	(0.0577)	(0.0571)	(0.0576)	(0.0572)
CMS rating 1	0.0292	0.0359	0.00298	0.0253	0.0105
	(0.0660)	(0.0663)	(0.0657)	(0.0660)	(0.0655)
CMS rating 2	0.0587	0.0681	0.0402	0.0573	0.0436
	(0.0602)	(0.0604)	(0.0599)	(0.0601)	(0.0597)
CMS rating 3	0.110	0.119$^+$	0.0935	0.114	0.0982
	(0.0604)	(0.0606)	(0.0600)	(0.0603)	(0.0598)
CMS rating 4	0.0132	0.0210	0.00337	0.0184	0.0100
	(0.0560)	(0.0562)	(0.0557)	(0.0559)	(0.0555)
Has infection violations	-0.0709	-0.0676	-0.0511	-0.0681	-0.0573
	(0.0485)	(0.0487)	(0.0482)	(0.0484)	(0.0481)
Urban indicator	0.649***	0.686***	0.609***	0.650***	0.574***
	(0.054)	(0.054)	(0.053)	(0.053)	(0.054)
Node degree	0.0137***				
	(0.0012)				
Node strength		0.00679***			
		(0.00079)			
Weighted average neighbor degree			0.0147***		0.0116***
			(0.0010)		(0.0011)
Eigenvector centrality in state				1.073***	0.643***
				(0.091)	(0.099)
Fixed effects	State	State	State	State	State
Observations	6,644	6,644	6,644	6,644	6,644
F-stat	123.8	117.9	132.0	125.0	125.3
R^2	0.412	0.407	0.419	0.413	0.422
Within R^2	0.169	0.163	0.179	0.171	0.184

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

Tab. S2: Covariates of the existence of nursing home COVID-19 cases

	Dependent variable: Nursing home has > 0 cases				
	(1)	(2)	(3)	(4)	(5)
Urban indicator	0.180***	0.190***	0.164***	0.183***	0.159***
	(0.0142)	(0.0142)	(0.0141)	(0.0142)	(0.0141)
Node degree	0.00307***				
	(0.000320)				
Node strength		0.00135***			
		(0.000208)			
Weighted average neighbor degree			0.00399***		0.00358***
			(0.000270)		(0.000297)
Eigenvector centrality in state				0.218***	0.0860*
				(0.0240)	(0.0262)
Fixed effects	State	State	State	State	State
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	6,644	6,644	6,644	6,644	6,644
F-stat	78.46	73.36	91.23	77.51	84.65
R^2	0.345	0.340	0.357	0.344	0.358
Within R^2	0.114	0.107	0.130	0.113	0.132

Standard errors in parentheses. Significance levels: $^{+}p < 0.05$, $^{*}p < 0.01$, $^{**}p < 0.001$, $^{***}p < 0.0001$. Dependent variable in our linear probability regressions using individual state data is a binary indicator that equals 1 if COVID cases are reported in that nursing home.

Tab. S3: Covariates of nursing home COVID-19 cases with county fixed effects

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Node degree	0.00385* (0.00131)				
Node strength		0.00187$^+$ (0.000843)			
Weighted average neighbor degree			0.00286$^+$ (0.00136)		0.00167 (0.00142)
Eigenvector centrality in state				0.379** (0.110)	0.342* (0.114)
Fixed effects	County	County	County	County	County
Home demographics	Yes	Yes	Yes	Yes	Yes
CMS quality rating	Yes	Yes	Yes	Yes	Yes
Observations	6,255	6,255	6,255	6,255	6,255
F-stat	60.90	60.50	60.44	61.28	55.84
R^2	0.575	0.574	0.574	0.575	0.575
Within R^2	0.0987	0.0981	0.0980	0.0992	0.0993

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.

Tab. S4: Covariates of COVID-19 cases within nursing homes using CMS data

	Dependent variable: $sinh^{-1}(Cases)$				
	(1)	(2)	(3)	(4)	(5)
Urban indicator	0.310***	0.349***	0.362***	0.319***	0.309***
	(0.0336)	(0.0334)	(0.0333)	(0.0335)	(0.0337)
Node degree	0.0103***				
	(0.000919)				
Node strength		0.00331***			
		(0.000463)			
Weighted average neighbor degree			0.00234***		0.00137*
			(0.000433)		(0.000443)
Eigenvector centrality in state				0.664***	0.617***
				(0.0634)	(0.0652)
Fixed effects	State	State	State	State	State
Observations	12,775	12,775	12,775	12,775	12,775
F-stat	152.8	145.3	143.1	151.3	139.6
R^2	0.248	0.244	0.242	0.247	0.248
Within R^2	0.116	0.111	0.109	0.115	0.116

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.

Tab. S5: State-level network summary statistics

	COVID-19 Cases		Degree		Strength		Weighted average neighbor degree		Eigenvector Centrality	
State	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AL	9.1	18.9	13.3	11.1	21.3	20.5	20.8	13.1	0.19	0.25
AR	4.3	14.7	8.3	8.5	11.5	12.7	12.2	9.7	0.15	0.25
AZ	4.7	9.8	14.6	14.8	22.6	27.6	23.0	18.5	0.28	0.33
CA	10.2	26.1	14.2	17.1	16.3	20.6	22.3	18.1	0.08	0.15
CO	7.9	15.8	8.4	9.6	10.4	13.0	13.8	11.3	0.16	0.25
CT	35.2	49.3	13.2	10.1	15.5	12.7	18.6	9.1	0.24	0.23
DC	37.3	40.1	7.6	3.0	11.6	5.9	9.9	2.0	0.60	0.25
DE	23.6	38.4	6.6	5.1	10.8	10.7	10.3	7.1	0.34	0.40
FL	5.3	14.1	26.1	20.5	37.5	32.3	37.5	19.9	0.11	0.24
GA	15.4	37.4	14.2	13.5	21.4	26.1	22.4	20.1	0.17	0.23
IA	4.0	10.7	5.6	6.1	7.3	9.1	9.9	7.8	0.08	0.16
ID	1.4	4.9	2.4	2.1	3.1	3.0	3.6	3.0	0.17	0.26
IL	15.6	32.0	21.8	27.0	27.7	37.1	34.6	30.2	0.13	0.23
IN	7.4	20.3	16.7	17.3	23.7	28.6	26.1	19.8	0.13	0.23
KS	1.7	6.9	12.4	15.1	16.8	24.3	19.7	18.2	0.16	0.31
KY	5.8	18.4	8.9	8.6	13.7	17.3	15.5	12.5	0.17	0.23
LA	12.5	22.2	11.4	9.9	18.6	20.5	18.1	13.1	0.11	0.23
MA	32.4	46.7	12.9	11.4	15.2	13.9	18.8	10.2	0.15	0.20
MD	20.2	31.8	28.7	26.2	37.2	38.1	43.6	25.6	0.28	0.31
ME	3.8	13.1	2.8	3.8	3.6	5.0	3.7	4.2	0.13	0.25
MI	10.9	21.5	12.9	14.4	16.8	20.3	19.8	16.6	0.12	0.21
MN	5.8	14.9	6.3	7.8	8.2	11.8	12.1	14.6	0.10	0.19
MO	4.3	15.6	15.9	16.4	23.0	28.6	25.6	20.0	0.10	0.23
MS	7.0	17.7	7.6	7.7	10.9	12.8	13.0	10.6	0.13	0.26
MT	1.1	4.7	2.2	3.0	3.6	5.4	3.1	4.3	0.15	0.32
NC	5.2	14.8	11.7	11.0	14.9	17.0	17.4	13.4	0.10	0.21
ND	3.1	8.1	4.1	4.9	7.4	12.7	9.8	13.4	0.20	0.31
NE	3.5	11.8	4.7	5.9	5.5	7.4	6.8	6.3	0.14	0.23
NH	12.6	28.2	3.1	2.8	3.5	3.3	4.0	3.1	0.15	0.28
NJ	36.2	50.0	26.1	19.2	35.9	30.3	38.1	18.6	0.24	0.22
NM	3.4	12.4	3.6	3.8	5.6	7.9	4.9	4.8	0.20	0.33
NV	21.9	38.6	6.9	6.4	10.9	11.9	10.0	7.3	0.31	0.37
NY	25.6	46.1	19.4	20.5	28.4	39.3	41.3	36.8	0.12	0.19
OH	5.2	15.4	18.1	15.8	23.9	24.9	29.2	17.5	0.10	0.17
OK	2.3	8.6	11.4	10.6	15.0	15.1	17.0	10.7	0.12	0.23
OR	3.1	7.5	4.0	4.2	4.9	5.3	5.8	4.7	0.13	0.26
PA	16.9	32.3	14.7	14.5	20.9	29.9	25.3	24.1	0.08	0.17
RI	22.8	38.7	12.3	10.9	15.3	15.5	21.5	12.1	0.34	0.33
SC	8.9	23.5	9.2	7.2	14.0	12.8	13.2	8.0	0.16	0.24
SD	1.9	6.1	1.2	1.7	1.8	3.0	2.1	2.7	0.05	0.20
TN	2.6	8.1	13.7	10.0	21.5	20.5	22.1	12.2	0.11	0.17
TX	3.7	16.0	20.2	19.5	29.8	68.9	53.4	93.8	0.05	0.13
UT	1.3	3.5	5.1	5.5	7.2	8.8	8.9	6.9	0.21	0.32
VA	9.0	23.5	12.8	10.9	17.4	16.9	18.8	10.7	0.17	0.22
VT	8.4	18.3	0.5	1.0	0.6	1.1	0.6	1.2	0.10	0.29
WA	6.1	18.7	6.7	6.8	8.5	9.5	9.5	7.8	0.17	0.26
WI	3.9	10.4	6.2	9.4	8.3	13.9	10.6	12.9	0.11	0.25
WV	3.5	11.4	6.1	6.5	10.6	15.4	13.8	11.8	0.18	0.26
WY	1.1	3.0	0.5	0.7	0.6	0.9	0.7	0.9	0.09	0.29

COVID cases include confirmed and suspected cases reported to CMS as of May 31, 2020.

Degree refers to the number of nursing homes that a particular home is connected to through a smartphone observed in both facilities.

Strength refers to 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.