How to Stop Contagion: Applying Network Science to Evaluate the Effectiveness of Covid-19 Vaccine Distribution Plans

Olga V. Chyzh* Mark David Nieman†

June 19, 2021

Abstract

Network theory and data are crucial for understanding any political outcomes resulting from actor interdependence. The interdependent nature of network data, however, often impedes empirical causal inference. Variation in Covid-19 vaccine distribution plans in the US sets up unique conditions for testing theories of network contagion using a quasi-experimental design. Despite clear agreement on the goal, state-formulated vaccine distribution plans diverged beyond initial priority groups: some continued prioritizing vaccinations based on mortality risks, while others determined eligibility based on risks of occupational exposure. We leverage this divergence as an identification strategy, given that vaccine prioritization plans were formulated prior to conclusive evidence that vaccines reduce transmission. According to network theory, reducing contagion requires disabling the transmission potential of the most connected actors—a strategy consistent with occupation-based vaccine prioritization. Our analysis of daily Covid-19 data in a matched sample of Oregon and California counties show strong support for the theory.

*Assistant Professor, University of Toronto, olga.chyzh@utoronto.ca
†Assistant Professor, University of Toronto, mark.nieman@utoronto.ca

Political scientists apply network theory to model diffusion and its pathways in all areas of study, from policy adoption (Desmarais, Harden, and Boehmke 2015) and elite collaboration (Kirkland and Williams 2014) to the spread of conflict (Dorff, Gallop, and Minhas 2020; Nieman et al. 2020). Despite their wide application, the bulk of network insights are purely theoretical. Empirical causal inference with network data is complicated by the interdependence among network topologies (e.g., density, actor centrality). For example, while we can compare the rate of transmission of policy, disease, or information in networks with different densities, we cannot causally attribute a change in transmission to a change in density, as changing network density necessarily changes other topologies, e.g., actor centrality.[1] Hence, network data are especially poorly suited for the application of the statistical principle of "holding all else constant."

In this letter, we argue that President Trump's haphazard decision to delegate Covid-19 vaccine distribution to US states sets up unique conditions for testing network theory within a quasi-experimental design. Given rather broad and non-binding guidelines from the Center for Disease Control (CDC), states defined vaccine priority groups using their own discretion. As a result, prioritization lists diverged as soon as vaccination campaigns moved beyond groups with the highest mortality risk.[2] Two broad strategies emerged. The first was to continue prioritizing based on mortality risk (i.e., using age), whereas the second strategy was to assign priority based on the risks of occupational exposure (e.g., grocery employees).[3]

Crucially for implementing a quasi-experimental design, most of the early decisions were made under conditions of very limited information, as the first vaccines were given emergency approval before various aspects of their effects were fully evaluated. In particular, early distribution plans were made in the absence of reliable information on whether vaccines

[1] Density, calculated as the total number of connections in the network, is a network-level metric, whereas centrality is a measure of connectedness for an individual actor (e.g., the sum of direct connections), and is actor-specific. Since changing density requires removing/adding connections to the network, it necessarily results in changes of centrality for at least some actors.

[2] The most at-risk groups were defined as medical workers, the elderly, individuals with pre-existing conditions or residing in congregated settings.

[3] In contrast to partisan divisions regarding other pandemic-related policies (Adeel et al. 2020; Neelon et al. 2021), Covid-19 vaccine distribution plans did not exhibit clear partisan divides.

reduced transmission of the virus, prevented disease, or merely ameliorated the symptoms; a sufficient body of evidence showing that vaccines were effective at reducing transmission did not become available until mid-March 2021 (Christie, Mbaeyi, and Walensky 2021; CDC 2021).[4]

Given the now available evidence that vaccines are effective at reducing transmission, the divergence in state-formulated vaccine priority plans sets up conditions for testing a key network theory insight—that the bulk of transmission through a network is disproportionately channeled through only a handful of highly connected or *central* actors (Granovetter 1973; Padgett and Ansell 1993; Box-Steffensmeier et al. 2018). In a service-oriented economy, such as the US, grocery store employees are the largest occupational group engaged in the highest number of face-to-face interactions.[5] The handful of US states that opted for prioritizing vaccine eligibility based on occupational risks essentially reduced or disabled the virus transmission potential of the most central actors in the networks of human interactions. Therefore, we expect to see lower rates of contagion in states that opted to prioritize vaccine eligibility based on occupational risks.

We test this prediction using a matched sample of counties from two contiguous, Democratic-governed states—Oregon and California. While both states implemented similar policy responses to the Covid-19 pandemic in terms of school closures, stay-at-home orders, and mandatory mask mandates (Adeel et al. 2020), California's Phase 1b of vaccine distribution included grocery employees, along with other CDC-defined *essential frontline workers*[6], while Oregon adopted age-based prioritizion. Consistent with the predictions

[4]Though there was a working hypothesis among medical researchers that vaccinations may reduce transmission—owing to fewer of the more contagious symptomatic cases and reduced viral loads among vaccinated individuals (Mallapaty 2021; Levine-Tiefenbrun et al. 2021)—there was little direct evidence. Moreover, there was no consensus opinion: some medical researchers offered a countering view, arguing that vaccinated individuals could still carry and spread the virus, even if they themselves were largely protected (Bleier, Ramanathan, and Lane 2021).

[5]In contrast to other service sectors, grocery stores had to remain open throughout the pandemic, even when most other retail businesses and even schools reduced hours or switched to remote operation. Though some grocers were able to limit exposure through offering pick-up and delivery services, most found these options unfeasible due to added costs, a lack of resources and experience with online platforms, or customer preference for in-store shopping.

[6]The CDC defines frontline essential workers as distinct from essential healthcare workers. Specifically,

of network theory, our results show that counties prioritizing grocery store employees experienced fewer new Covid-19 cases and this effect grew stronger over time.

Beyond the Covid-19 pandemic, this research contributes to the general understanding of contagion and its pathways. The resulting insights are equally applicable to other areas of contact contagion research—e.g., the spread of information, growth of extremist movements, or diffusion of policy innovation. Unlike many types of contact contagion that are not amenable for easy tracing and data collection, the Covid-19 vaccine distribution and contagion data give us a unique opportunity to empirically evaluate network theory.

Network Theory and Vaccine Prioritization

From the perspective of network theory, the nature of contagion—its speed, reach, and main pathways of transmission—depends on the local structures within the network (Chyzh and Kaiser 2019). Rather than simply a function of network density, the rate of contagion depends on the presence of a few highly connected actors with cross-cutting connections to otherwise disconnected parts of the network (Granovetter 1973; Feld 1991; Padgett and Ansell 1993). What ultimately determines an individual actor's transmission potential is its level of connectedness or *centrality*.[7] It follows that, as long as there is some level of heterogeneity in actors' centrality (i.e., some actors have more direct or indirect connections than others), disabling the transmission potential of the network's most central actors is the fastest and most effective way to reduce or stop contagion.

By implication, individuals who are the most likely to limit their social interactions as to protect themselves against the virus (e.g., due to a pre-existing condition), effectively self-select themselves into less central network positions. Since the essential-service providers may

frontline essential workers are "the subset of essential workers likely at highest risk for work-related exposure [...] because their work-related duties must be performed on-site and involve being in close proximity (<6 feet) to the public or to coworkers" (Dooling et al. 2021, 1657). Examples include grocery and manufacturing workers.

[7] Measures of centrality include *degree*, *eigenvector*, *closeness*, and *betweenness centrality* (Bonacich 1972; Patty and Penn 2017). The Online Appendix contains formal definitions. The results of our analysis do not reveal a meaningful difference in outcome, based on the choice of measure.

be the toughest social links to eliminate, giving vaccine priority to essential frontline workers is also an effective way to reduce the risk of infection for the most vulnerable individuals. Conversely, age-based vaccine prioritization is less likely to reduce or stop the spread of the virus, as the key transmitters—individuals in entry-level positions, such as front-facing grocery employees—tend to fall within a younger age range.

To explore how these well-known theoretical insights work within this specific application, we also compare the effects of vaccine prioritization plans (centrality- vs. vulnerability-based) using a simulation experiment. To keep this letter within the space constraints, we describe the full details of the simulation in the Online Appendix. Our simulation, based on a real-world interaction network, confirms that, in terms of reducing contagion, centrality-based vaccination strategies substantially outperform the strategy of allocating vaccines based solely on vulnerability. Specifically, in simulated networks, allocating vaccines based on vulnerability prevented the spread of the virus in only about 46% of the actors, any of the centrality-based strategies increased this number to about 65%. The centrality-based strategies fare rather well in terms of protecting the individuals defined as vulnerable, albeit none guarantee protection for 100% of these individuals.

The simulation highlights the effectiveness of vaccinating essential frontline workers for reducing the spread of Covid-19, compared to other vaccine prioritization strategies, under the conditions of vaccine scarcity or distributional constraints. States that prioritized vaccine access for essential frontline workers were able to reduce transmission by eliminating a key source of contagion—those individuals that came into contact with both the greatest number of people and societal groups. This leads to the following hypothesis:

Research Hypothesis: Administrative units that prioritize individuals with higher centrality will experience fewer infection cases.

Research Design

To test the research hypothesis, we leverage variation in vaccine priority lists between California and Oregon. While early on the two states followed a similar vaccine distribution strategy as the rest of the US (i.e., prioritizing the elderly, medical workers, and individuals with pre-existing conditions), California was among the first states to extend vaccine eligibility to grocery store employees (on March 1), whereas in Oregon, grocery store employees did not become eligible for vaccination until almost a month later (on March 29).

The unit of analysis is the county-day. Our dependent variable is a logged 7-day rolling average of the number of new Covid-19 cases, obtained from the John Hopkins University Center for Systems Science and Engineering (Gassen 2021).[8] The independent variable is an interaction between *California* and *Day of Treatment*, in which *California* is a binary variable that equals 1 for California and 0 for Oregon, and *Day of Treatment* is a count variable; the count starts at 1 on March 14, 2 weeks since grocery employees became eligible for vaccination in California—the date when those who had received the first dose would have achieved partial immunity (between 50–80%) (Polack et al. 2020; Bernal et al. 2021). The estimation equation is:

$$\begin{aligned} log(New\ Cases/1000) = \beta_0 + \delta_0 Day\ of\ Treatment + \beta_1 California \\ + \delta_1 California * Day\ of\ Treatment + other\ factors. \end{aligned}$$

This, of course, is a textbook example of the difference-in-difference design (Wooldridge 2015, 407–12). The *treatment* here is measured as the day since the first grocery workers had developed partial immunity in California (*Day of Treatment*). The estimation parameter δ_0 is the average difference between the periods *before* and *after* the start of the treatment for the control group; β_1 is the average difference between the two groups *prior* to the treatment; $\beta_1 + \delta_1$ is the average difference between the two groups *after* the start of the treatment, and

[8] The John Hopkins data contain cumulative data by day and US county. We calculated daily cases by first differencing cumulative cases. In a small number of cases, first differencing resulted in negative numbers of cases, due to data corrections. Any such negative values were recoded to 0.

δ_1 is the difference-in-difference coefficient that gives the average difference attributable to the treatment (i.e., the average effect of expanding vaccine eligibility to grocery employees in California).

We control for a number of county-level demographic variables that may influence Covid-19 contagion, including the daily cumulative number of administered vaccine doses per 1000 residents, the daily cumulative number of reported Covid-19 cases per 1000 residents, GDP (2019 USD), population, unemployment rate, percent of population that hold at least a Bachelor's degree, urbanization, percentage of black and other racial minorities, percentage of Hispanic/Latino population, percentage of foreign population, and Biden's percentage margin in the two-party vote in the 2020 election. Data on vaccine distribution were obtained from the official state websites.[9] Data on county-level economic outcomes were obtained from the Bureau of Economic Analysis. The demographic variables are from the most recent US Census American Community Survey (2015–19 averages). Unemployment data are 2019 numbers obtained from the most recent decennial census. To address temporal autocorrelation in the data, we include a lagged dependent variable.[10]

We estimated our model using OLS regression on a full sample, as well as a sample of matched observations (Iacus, King, and Porro 2012). The full sample consists of daily observations for all counties (58 in California, 36 in Oregon) between December 17, 2020 (the first day of vaccine administration in both states) and April 12, 2021 (two weeks after Oregon authorized giving vaccines to food processing employees, including grocery employees) for a total of 10,271 non-missing observations.

A Matched Design

Despite geographical proximity, California and Oregon counties differ in many ways that are not easy to control for within the OLS framework (e.g., Los Angeles county in California

[9] California's vaccine distribution data come from https://covid19.ca.gov, while Oregon's data are from https://covidvaccine.oregon.gov.

[10] Diagnostics support using a two-period lag.

is not comparable to any county in Oregon). To tighten our causal claims, we implement a matched-sample design. We matched California counties (treatment group) with those in Oregon (control group) on all independent variables (other than *California* and its interaction with *Treatment Day*) by day using coarsened exact matching (CEM) (Iacus, King, and Porro 2012).

CEM consists of identifying exact matches (observations with the exact same values on all covariates) after coarsening (dichotomizing or multichotomizing variables into discrete categories) any continuous and ordinal variables. Per Iacus, King, and Porro (2012), we selected the number of cutpoints to multichotomize our variables using our empirical knowledge of the data, whenever possible. To maximize sample size, we allowed for multiple matches for each observation. The resulting matched sample consists of 1,454 observations (753 from California and 701 from Oregon).

Results

Figure 1 shows the temporal trends in the raw data (on the left) and in the matched sample (on the right).[11] Both subfigures show that the daily number of new Covid-19 cases followed a similar declining trend in the two states between the start of vaccinations in December 2020 and the middle of March, 2021. As California expanded vaccine eligibility to grocery store employees on March 1, 2021, whereas Oregon did not, the two trends diverge. The point of divergence falls somewhere between March 1 and March 15—the date when the grocery employees who were able to get the vaccine on March 1 would have achieved between 50–80% immunity. After this time, the trend of new Covid-19 cases in California continues to decline, whereas Oregon starts observing an increase.

The relationship is even starker in the matched sample, where the two states' trends are almost exactly the same between February and the middle of March—the start date of the treatment—after which Oregon's trend shoots upwards, while California continues

[11]The daily values bounce around in the matched data, as different counties enter and exit the matched sample based on the day-to-day variation in vaccine doses and cumulative Covid-19 cases.

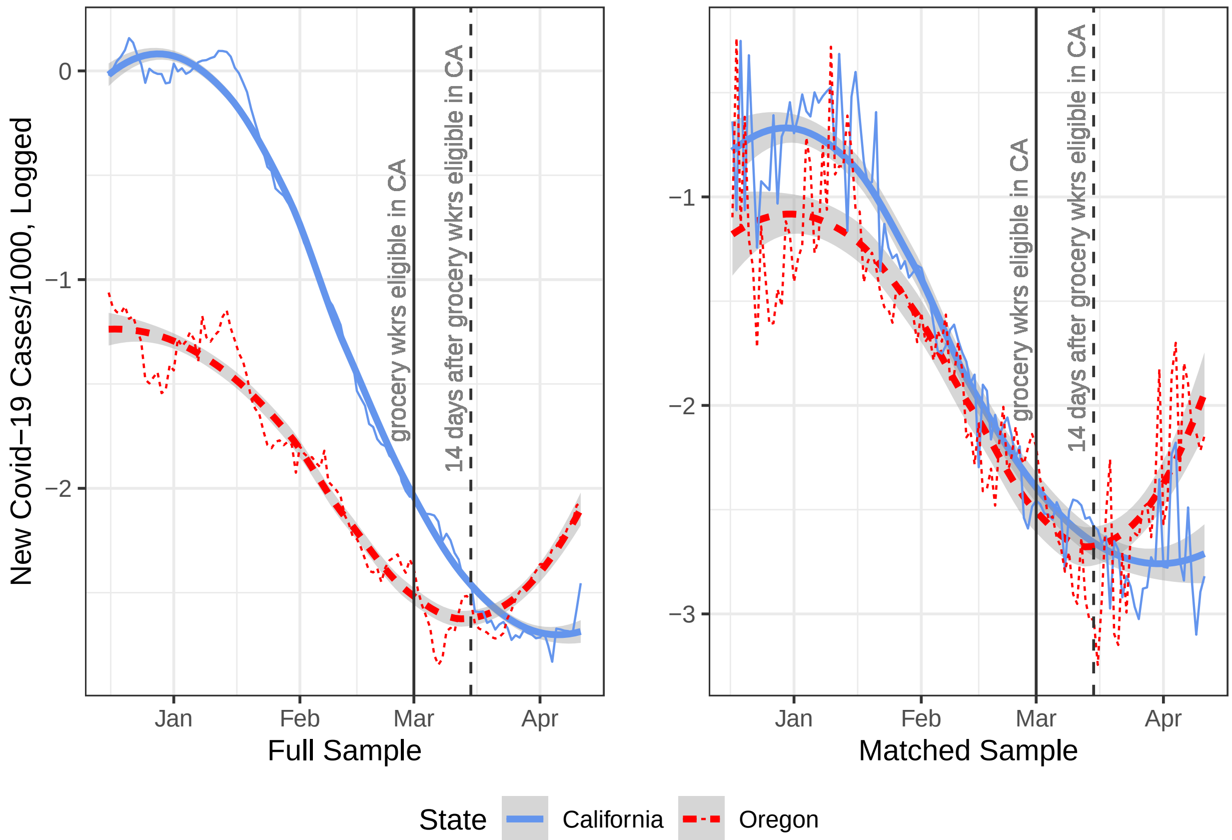

Figure 1: Temporal Trends in Covid-19 Cases, December, 2020–April, 2021. Thin lines show the trends in the raw data. Thick lines show loess estimates. Gray bands show the 95% CIs.

experiencing an, albeit slower, decrease.

Table 1 shows the results of the statistical analysis for the full sample (Model 1) and the matched sample (Model 2). Since the dependent variable, *New Covid-19 Cases* is measured on a logged scale, the model coefficients have a percentage interpretation (Wooldridge 2015). Thus, the coefficient of 0.021 on *California* indicates that, *prior* to opening up Covid-19 vaccination to grocery store employees, the state of California has had, on average, about 2.1 (0.021*100) percent more new daily Covid-19 cases than the state of Oregon. This effect is statistically significant in the full sample, but not in the matched sample.

The coefficient of 0.004 (0.005 in the matched sample) on the *Day of Treatment* gives the average difference in new Covid-19 cases in Oregon counties before and after California opened up vaccination to grocery employees. This coefficient is statistically significant, indi-

Table 1: The Effect of Vaccine Eligibility to Grocery Employees on New Daily Covid-19 Cases/1000 residents (logged)

	Full Sample	Matched Sample
Day of Treatment	$0.004^{***}(0.001)$	$0.005^{***}(0.001)$
California	$0.021^{***}(0.005)$	$0.022\ (0.015)$
California∗Day of Treatment	$-0.004^{***}(0.001)$	$-0.004^{***}(0.001)$
Cum. Doses/1000 res., logged	$-0.034^{***}(0.002)$	$-0.031^{***}(0.006)$
Cum. Covid-19 Cases/1000 res., logged	$0.002\ (0.006)$	$0.013\ (0.022)$
County GDP, logged	$-0.003\ (0.006)$	$0.042^{**}(0.017)$
County Population, logged	$-0.001\ (0.005)$	$-0.039^{***}(0.014)$
Unemployment Rate	$-0.001^{**}(0.001)$	$-0.001\ (0.002)$
Percent BA Degree	$0.001\ (0.001)$	$-0.002\ (0.002)$
Urbanization	$0.016^{*}(0.008)$	$0.027\ (0.033)$
Percent Black	$0.001\ (0.001)$	$-0.007\ (0.011)$
Percent Latino	$0.001^{**}(0.001)$	$0.001\ (0.002)$
Percent Other Race	$0.001^{***}(0.001)$	$0.002\ (0.002)$
Percent Foreign	$-0.002^{***}(0.001)$	$-0.002\ (0.003)$
Biden's Margin	$0.001{**}(0.001)$	$0.001\ (0.001)$
Lagged DV	$0.912^{***}(0.004)$	$0.913^{***}(0.012)$
Constant	$0.073^{*}(0.042)$	$-0.197\ (0.164)$
Num.Obs.	10271	1454
R^2	0.947	0.908
R^2 Adj.	0.947	0.907

cating that Oregon experienced a 0.4 percent (0.5 percent in the matched sample) increase, on average, in new daily Covid-19 cases in the period since March 14, 2021 compared to the period between December 17, 2020 and March 14, 2021.

The coefficient of -0.004 on the interaction term is the difference-in-difference coefficient. It indicates a -0.4 percent difference in new Covid-19 cases between the treatment and control groups in the periods before and after California opened vaccine eligibility to grocery employees. This coefficient, of course, gives us only the average daily effect, as *Day of Treatment* is measured on an integer scale.[12] The marginal effect of *California* by *Day of Treatment*, is shown in Figure 2. As one can see, the effect of vaccinating grocery employees grows (in absolute value) as time advances, reaching about -0.8 percent 27 days after the start

[12]Results are robust to coding *Day of Treatment* on a nominal scale or including polynomials. Diagnostics favor the model presented in Table 1.

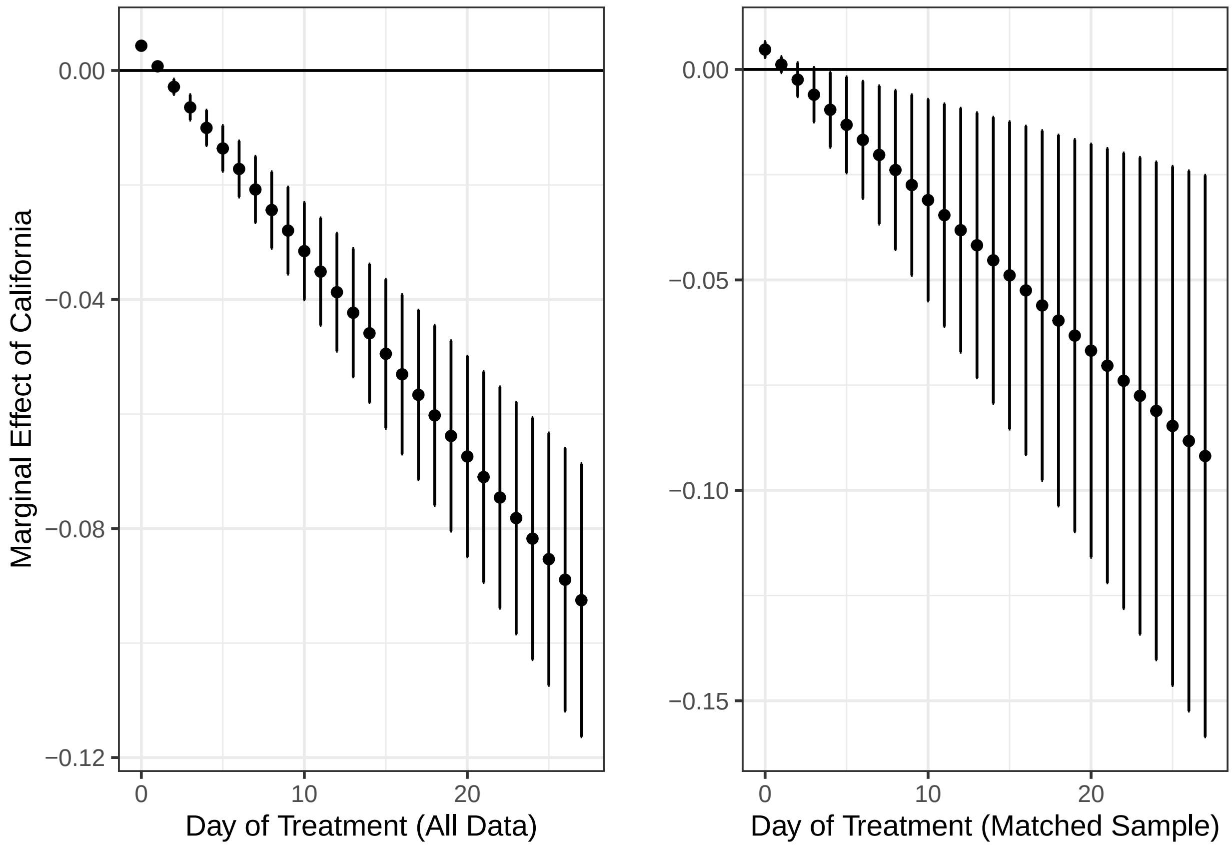

Figure 2: Marginal Effect of Vaccine Eligibility to Grocery Workers. Error bars represent 95% CIs.

of the treatment period. These results provide strong support for the research hypothesis.

The coefficient on *Cumulative Doses, logged* is negative and statistically significant. Interpreted as an elasticity, it indicates that a 10 percent increase in the cumulative number of administered vaccine doses is associated, on average, with a 0.3 percent decrease in new Covid-19 cases. The weakness of this effect is further indicative of our theoretical argument—*who* is vaccinated has a greater impact than *how many* people are vaccinated. Vaccinating individuals that do not engage in a large number of daily interactions has a meager effect on containing the spread of the virus. The remaining control variables act as expected and are discussed in the Online Appendix.

Conclusion

We perform one of the few tests of network theory within a causal inference framework. Our research design allows us to isolate the causal effect of a handful of highly connected actors on the transmission rate within the network by varying the central actors' transmission ability between the treatment group (networks with the vaccinated central actors) and the control group (networks with the unvaccinated central actors), before and after the start of the treatment (vaccine eligibility for grocery employees), while holding all else constant (via matching). This research has wide-ranging implications for the study of any type of social interdependence, such as legislative collaboration, international trade and conflict, terrorist networks, or social movements.

Within the application to vaccine prioritization plans, our results shine light on the central ethical dilemma associated with any vaccine priority plan, as long as vaccines remain scarce. On one hand, society has a moral obligation to protect its most vulnerable members. On the other hand, from the perspective of reducing the virus contagion, allocating the scarce vaccines to protect the vulnerable is equivalent to a scenario in which no vaccine is available at all. The ethical trade-off is further complicated in the view of unexplored long-term effects of contracting a novel virus by otherwise healthy individuals. By fleshing out these differences, this letter contributes to the informed conversation about the benefits, costs, and trade-offs of public health policies, as they relate to political trust and participation (Mattila 2020), inequality (Lynch 2020), and long-term institutional development (Gingerich and Vogler 2021).

References

Adeel, Abdul Basit, Michael Catalano, Olivia Catalano, Grant Gibson, Ezgi Muftuoglu, Tara Riggs, Mehmet Halit Sezgin, et al. 2020. "COVID-19 Policy Response and the Rise of the Sub-National Governments." *Canadian Public Policy* 46 (4): 565–84.

Bernal, Jamie Lopez, Nick Andrews, Charlotte Gower, Julia Stowe, Chris Robertson, Elise Tessier, Ruth Simmons, et al. 2021. "Early Effectiveness of COVID-19 Vaccination with BNT162b2 mRNA Vaccine and ChAdOx1 Adenovirus Vector Vaccine on Symptomatic Disease, Hospitalisations and Mortality in Older Adults in England." *MedRxiv*.

Bleier, Benjamin S., Murugappan Ramanathan, and Andrew P. Lane. 2021. "COVID-19 Vaccines May Not Prevent Nasal SARS-CoV-2 Infection and Asymptomatic Transmission." *Otolaryngology–Head and Neck Surgery* 164 (2): 305–7.

Bonacich, Phillip. 1972. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2 (1): 113–20.

Box-Steffensmeier, Janet M., Benjamin W. Campbell, Dino P. Christenson, and Zachary Navabi. 2018. "Role Analysis Using the Ego-ERGM: A Look at Environmental Interest Group Coalitions." *Social Networks* 52: 213–27.

CDC. 2021. "Science Brief: Background Rationale and Evidence for Public Health Recommendations for Fully Vaccinated People." 2021. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fmore%2Ffully-vaccinated-people.html.

Christie, Athalia, Sarah A. Mbaeyi, and Rochelle P. Walensky. 2021. "CDC Interim Recommendations for Fully Vaccinated People." *Journal of the American Medical Association* 325 (15): 1501–2.

Chyzh, Olga V., and Mark S. Kaiser. 2019. "Network Analysis Using a Local Structure Graph Model." *Political Analysis* 27 (4): 397–414.

Desmarais, Bruce A., Jeffrey J. Harden, and Frederick J. Boehmke. 2015. "Persistent Policy

Pathways: Inferring Diffusion Networks in the American States." *American Political Science Review* 109 (2): 392–406.

Dooling, Kathleen, Mona Marin, Megan Wallace, Nancy McClung, Mary Chamberland, Grace M. Lee, H. Keipp Talbot, Josè R. Romero, Beth P. Bell, and Sara E. Oliver. 2021. "The Advisory Committee on Immunization Practices' Updated Interim Recommendation for Allocation of COVID-19 Vaccine—United States, December 2020." *MMWR. Morbidity and Mortality Weekly Report* 69 (5152): 1657–60.

Dorff, Cassy, Max Gallop, and Shahryar Minhas. 2020. "Networks of Violence: Predicting Conflict in Nigeria." *Journal of Politics* 82 (2): 476–93.

Feld, Scott L. 1991. "Why Your Friends Have More Friends Than You Do." *American Journal of Sociology* 96 (6): 1464–77.

Gassen, Joachim. 2021. *Tidycovid19: Download, Tidy and Visualize Covid-19 Related Data.*

Gingerich, Daniel W., and Jan P. Vogler. 2021. "Pandemics and Political Development: The Electoral Legacy of the Black Death in Germany." *World Politics.* doi: 10.1017/S0043887121000034.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): l360–1980.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20 (1): 1–24.

Kirkland, Justin H., and R. Lucas Williams. 2014. "Partisanship and Reciprocity in Cross-Chamber Legislative Interactions." *Journal of Politics* 76 (3): 754–69.

Levine-Tiefenbrun, Matan, Idan Yelin, Rachel Katz, Esma Herzel, Ziv Golan, Licita Schreiber, Tamar Wolf, et al. 2021. "Decreased SARS-CoV-2 Viral Load Following Vaccination." *medRxiv.* https://doi.org/10.1101/2021.02.06.21251283.

Lynch, Julia. 2020. *Regimes of Inequality.* Cambridge University Press.

Mallapaty, Smriti. 2021. "Can COVID Vaccines Stop Transmission? Scientists Race to Find Answers." *Nature.* doi: 10.1038/d41586-021-00450-z.

Mattila, Mikko. 2020. "Does Poor Health Mobilize People into Action? Health, Political Trust, and Participation." *European Political Science Review* 12 (1): 49–65.

Neelon, Brian, Fedelis Mutiso, Noel T. Mueller, John L. Pearce, and Sara E. Benjamin-Neelon. 2021. "Associations Between Governor Political Affiliation and COVID-19 Cases, Deaths, and Testing in the US." *American Journal of Preventive Medicine.* https://doi.org/10.1016/j.amepre.2021.01.034.

Nieman, Mark, Carla Martinez Machain, Olga Chyzh, and Sam Bell. 2020. "An International Game of Risk: Troop Placement and Major Power Competition." *Journal of Politics.* doi/10.1086/711716.

Padgett, John F., and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology* 98 (6): 1259–319.

Patty, John W., and Elizabeth Maggie Penn. 2017. "Network Theory and Political Science." In *The Oxford Handbook of Political Networks*, 147. Oxford University Press.

Polack, Fernando P, Stephen J Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L Perez, et al. 2020. "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine." *New England Journal of Medicine* 383 (27): 2603–15.

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach, 6th Edition.* Cengage Learning.

Online Appendix for 'How to Stop Contagion: Applying Network Science to Evaluate the Effectiveness of Covid-19 Vaccine Distribution Plans'

June 19, 2021

Appendix A: Definitions of Centrality Metrics

Formally, define the nodes in a network as a set of n actors $i \in \{1, 2, \dots, n\}$, and one or more interactions between each pair of nodes i and j as a link, $l_{ij} = 1$. *Degree* is simply the total number of links for each node, or $\sum_{j:j \neq i} l_{ij}$. *Eigenvector centrality* is a more sophisticated analog of degree centrality that assigns higher values to nodes that are connected to other central nodes in the network.[1] If we define d_{ij} as the shortest path between i and each other node j in the network,[2] then each node's *closeness centrality* is the inverse of the sum of the shortest paths that separate each node from all other nodes, or $\frac{1}{\sum_{j:j \neq i} d_{ij}}$ (Bavelas 1948). Finally, for all pair of nodes j and k, we can define the number of shortest paths between them as $g_{j,k}$. Then, for each node i, we can define *betweenness centrality*, $g_{j,k}(i)$, as the number of the shortest paths $g_{j,k}$ that go through i (Freeman 1977; Padgett and Ansell 1993).

[1] A node's eigenvector centrality, c_i^e is calculated as $c_i^e = \lambda^{-1} \sum_{j:j \neq i} l_{ij} c_j^e$, where λ is the largest eigenvalue of the adjacency matrix that represents a given network.

[2] If i and j share a link, the shortest path d_{ij} is equal to 1; if i and j are not directly connected, but are each connected to a third node k, then the shortest path $d_{ij} = 2$, and so on.

Appendix B: The Simulation

Consider the canonical interaction data for 73 boys in a small high school in Illinois in Spring 1958 (Coleman 1964). In this network, two individuals are connected by a link if at least one of them named the other in response to the question "What fellows here in school do you go around with most often?"[3] Links in this network are measures of frequent elective interactions among individuals, which makes these data a perfect fit for deriving predictions related to contagion and the strategies of its containment. Figure 1 shows the distributions of individuals' degree, closeness, betweenness, and eigenvalue centrality measures in these data.

In Figure 2, the node with the second highest value on each centrality measure is shown in dark red.[4] Without loss of generality, assume that the virus spreads on contact with certainty. If the dark red nodes are the initial carrier (Patient 0), then its direct neighbors (shown in red) are next catch the virus, after which the virus spreads to the nodes that are reachable through a shortest path of length 2 (shown in orange) as a part of the second round. The figure shows that the choice of Patient 0 has substantial implications for contagion speed and pathways. Counter-intuitively, in this example, infecting the node based on degree centrality does not lead to the largest number of infected individuals at the end of two rounds. The scenarios with the highest rate of infected are the ones in which Patient 0 is chosen based on maximizing betweenness centrality (38 infected) and closeness centrality (36 infected).

Since the local structures in the network affect contagion's speed and reach, we can leverage these structures to reduce contagion. Suppose there are 10 available vaccines. Without loss of generality, assume getting a vaccine makes an individual both immune to the virus and unable to transmit it.

[3]We transformed the original directed network data into a non-directed symmetric network.

[4]The same node happens to get the highest score on all four measures on centrality discussed here. To introduce variation into the demonstration, we therefore chose to show the node with the second highest value on each centrality measure.

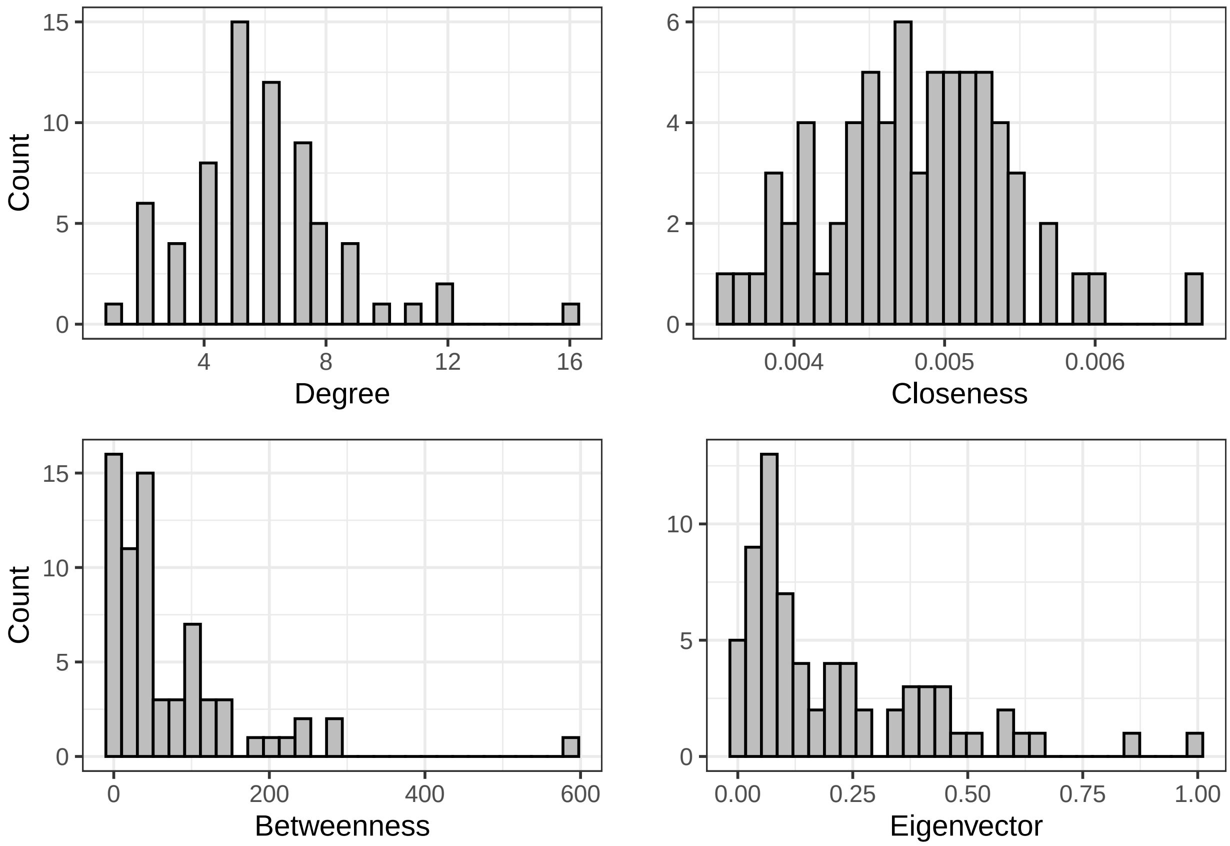

Figure 1: Centrality in the Interaction Network

Figure 3 shows the spread of the virus under five vaccination scenarios:

1. *Low Degree Scenario*—distribute vaccine to the nodes with the lowest degree centrality;
2. *High Degree Scenario*—distribute vaccine to the nodes with the highest degree centrality;
3. *High Closeness Scenario*—distribute vaccine to the nodes with the highest closeness centrality;
4. *High Betweenness Scenario*—distribute vaccine to the nodes with the highest betweenness centrality;
5. *High Eigenvector Scenario*—distribute vaccine to the nodes with the highest eigenvector centrality.[5]

The *Low Degree Scenario* scenario mimics the strategy of prioritizing vaccinations by

[5] In case of a tie, we randomly choose nodes with the same centrality until we reach the required number.

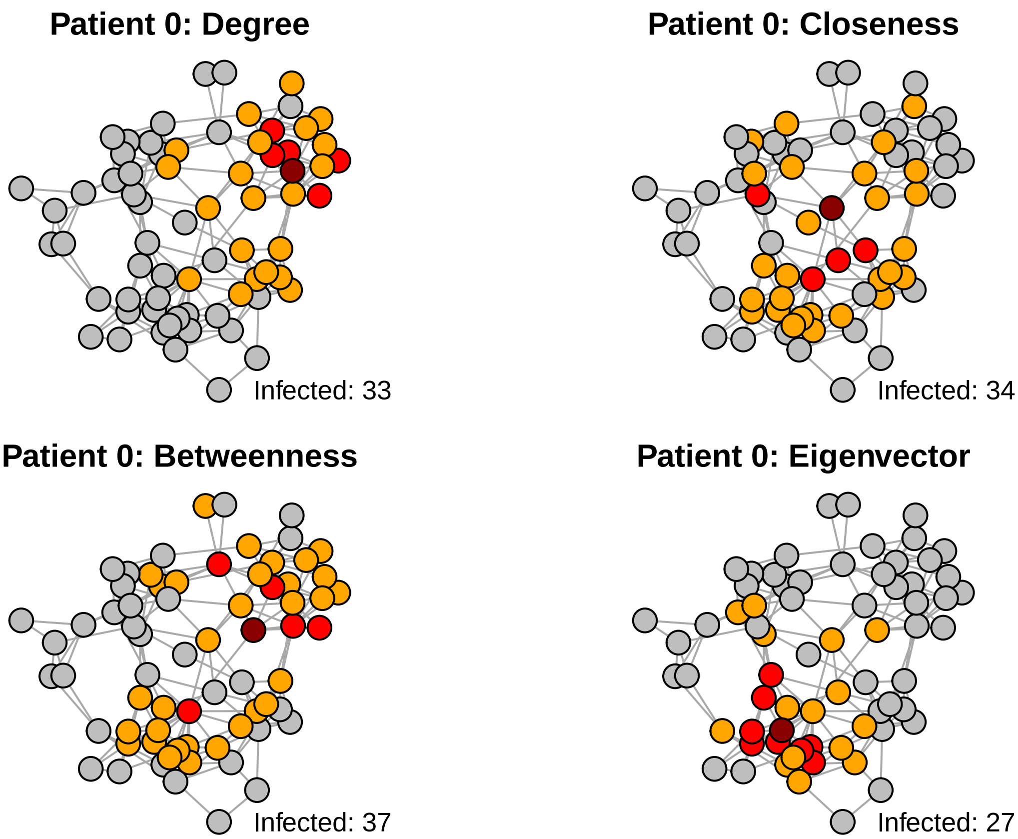

Figure 2: Contagion in the Interaction Network

mortality risk, adopted by many US states. In practice, this strategy focused on introducing vaccines by age groups, starting with the oldest. By defining *vulnerable individuals* as those with the lowest degree centrality, we assume that these are high-risk individuals who chose to limit their risk of exposure by reducing their number of social interactions (i.e., they are willing and able to do so).[6]

The *High Degree Scenario* scenario may be thought of as one, in which vaccine priority goes to individuals with the highest number of face-to-face interactions. These individuals include essential frontline workers, such as medical staff, post office employees, and individuals employed at pharmacies or grocery stores.[7]

[6]The network science focus limits our insights to groups of individuals that are identifiable as a function of their network connections. Making separate inferences for vulnerable individuals that are either unable or unwilling to self-isolate by reducing their number of direct contacts are treated is beyong the scope of this article.

[7]In addition to the individuals who are unable to reduce their number of face-to-face interactions due to the essential nature of their job, this group will also include individuals who are simply unwilling to

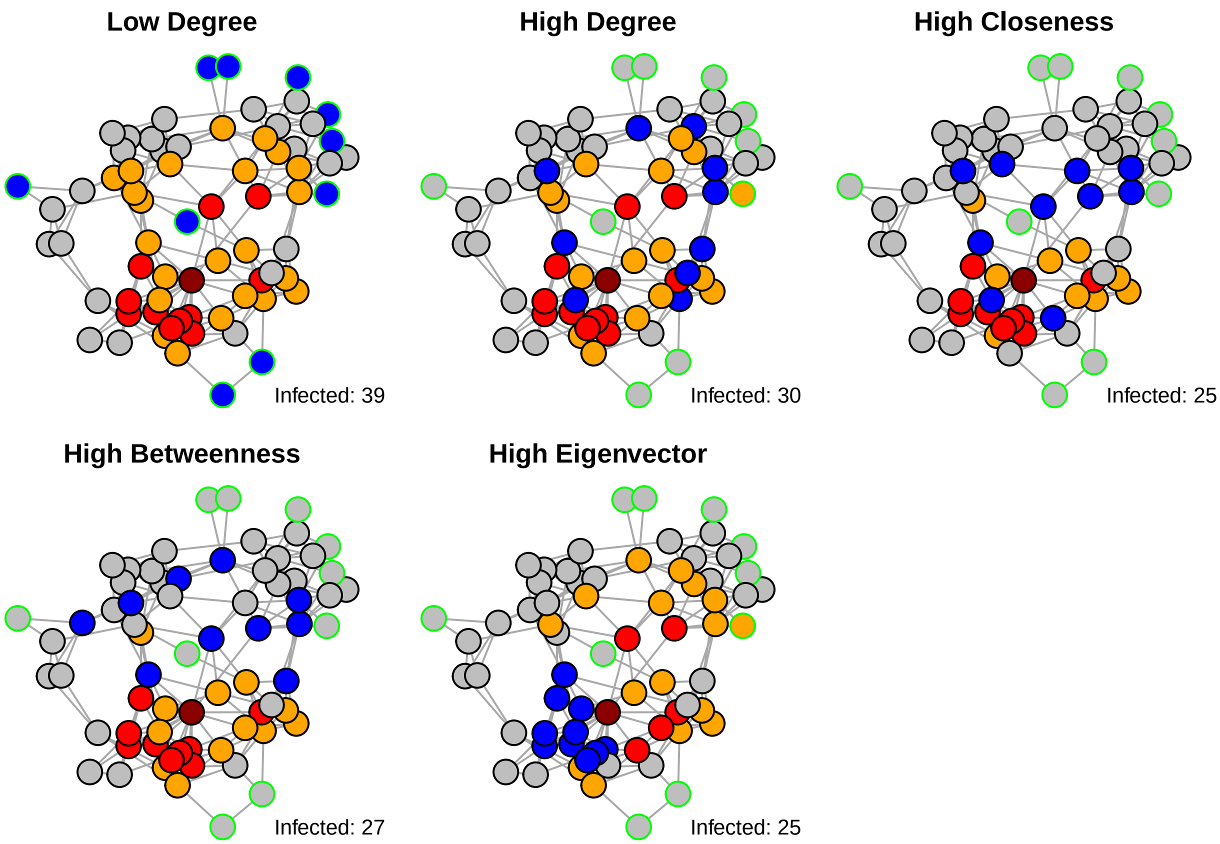

Figure 3: Vaccination Scenarios

The differences between the categories of individuals that would be prioritized under the three remaining scenarios are less obvious. Under the *High Closeness Scenario*, the vaccine priority would go to individuals that are reachable by every other individual in the network though the smallest number of contacts. These individuals may include a subset of essential employees, such as doctors and nurses. Under the *High Betweenness Scenario*, vaccination would start with the individuals that are most important for connecting the network, such as nurses that work at multiple hospitals or care-takers that work at multiple retirement communities. The *High Eigenvector Scenario* would prioritize vaccinating clusters of individuals who all engage in large numbers of face-to-face interactions. Such individuals may include

limit their social interactions. These may include individuals with low information (e.g., pandemic deniers), or groups with high risk acceptance (e.g. young adults who perceive the cost of social isolation as greater than that of possible exposure). The treatment of the former is beyond the scope of this research. The policy implications for treating latter group are similar to the ones for essential frontline workers, albeit the guidelines for identifying these individuals are less clear in practice.

essential frontline workers that service group living communities, such as retirement homes or college dorms.

In order to make comparisons among these scenarios, we designate Patient 0 as the node with the highest scores on all centrality measures. Since, by definition, nodes with no connections cannot be infected by contact, we remove these from the network prior to the simulation. The demonstration shows that all but the first vaccination scenario leads to a drastic reduction of contagion in the network. Compared to the fist vaccination scenario, vaccination scenarios 2–5 help reduce contagion after the second round by anywhere between 9 and 14 cases of infection. In a network of 69 individuals (after removing the isolates), this is a difference between infecting 57 vs. only 36–29 percent of the nodes in the network.

If we think of the nodes with the lowest degree centrality (marked with green border color in the figure) as vulnerable individuals, this demonstration also shows that all of the vaccination scenarios fare extremely well at protecting these individuals. Only 1 such node gets exposed to the virus in two of the scenarios.

As stark as these results may look, they are based on analyzing a single real-world interaction network, and therefore provide no insights regarding the uncertainty around the estimates of infection rates. To estimate the uncertainty around the estimates, we perform the following Monte Carlo experiment. We start by estimating an exponential random graph model (ERGM) (Wasserman and Faust 1994), in which the interaction network is the dependent variable, and the network parameters of interest are the baseline link probability (*edges*), the tendency towards open triangles (*2-stars*), and closed triangles (*gwesp*).[8] The estimates of this model are shown in Table 1.

We then use the estimates from this model to simulate the interaction network 10,000 times, and repeat the analysis done on the original interactions network on these simulated

[8]ERGMs are a state-of-the-art estimation approach for modeling the probability of observing a network with a given set of endogenous statistics, such as the total number of edges, open or closed triangles, or other network features. For a detailed formal description, see Hunter et al. (2008); Robins et al. (2007); Robins and Pattison (2001); Snijders et al. (2006).

Table 1: ERGM Fit of the Interaction Network

Edges	-3.73	0.32
k-Star(2)	-0.14	0.02
Gwesp (decay=0.5)	1.98	0.17

Table 2: Round 2 Summary for 10,000 Simulations of the Interaction Network among 73 Individuals

	Vaccinated	Infected	Not Infected	Vulnerable Infected
No Vaccine	0	41.87 (5.59)	30.13 (5.59)	3.23 (1.52)
Low Degree	10	38.56 (5.31)	23.44 (5.31)	0
High Degree	10	26.61 (4.99)	35.38 (4.99)	2.42 (1.38)
High Closeness	10	24.63 (4.89)	37.37 (4.89)	2.43 (1.37)
High Betweenness	10	26.80 (4.96)	35.20 (4.96)	2.22 (1.29)
High Eigenvector	10	30.13 (6.00)	31.87 (6.00)	2.53 (1.42)

Note:
Cell values are means over 10,000 simulations. Numbers in parentheses are standard deviations. *Not Infected* does not include *Vaccinated*.

networks.[9] Table 2 and Figure 4 provide summaries of these simulations.

Table 2 shows the mean (standard deviation) of the number of individuals infected/not infected after the second round of contagion (individuals connected to patient 0 directly or via one intermediary) in 10,000 simulations of the interaction network. Just as before, we denote the 10 individuals with the lowest number of direct (lowest degree centrality) connections as the "vulnerable" individuals. These are the individuals vaccinated in the *Low Degree* scenario. The last column of 2 shows the mean (standard deviation) for the number of these individuals that are infected under each of the vaccination scenarios. We also perform a simulation for the *No Vaccine* scenario as a reference for comparison.

Under the *Low Degree* scenario of giving vaccine priority to the vulnerable individuals, the number of individuals infected after the second round is only slightly lower than that under the *No Vaccine* scenario (roughly 39 vs. 42). Under each of the alternative scenarios, the number of infected is substantially lower both in absolute terms and as a percent decrease.

[9]For a similar simulation approach, see Boehmke et al. (2017).

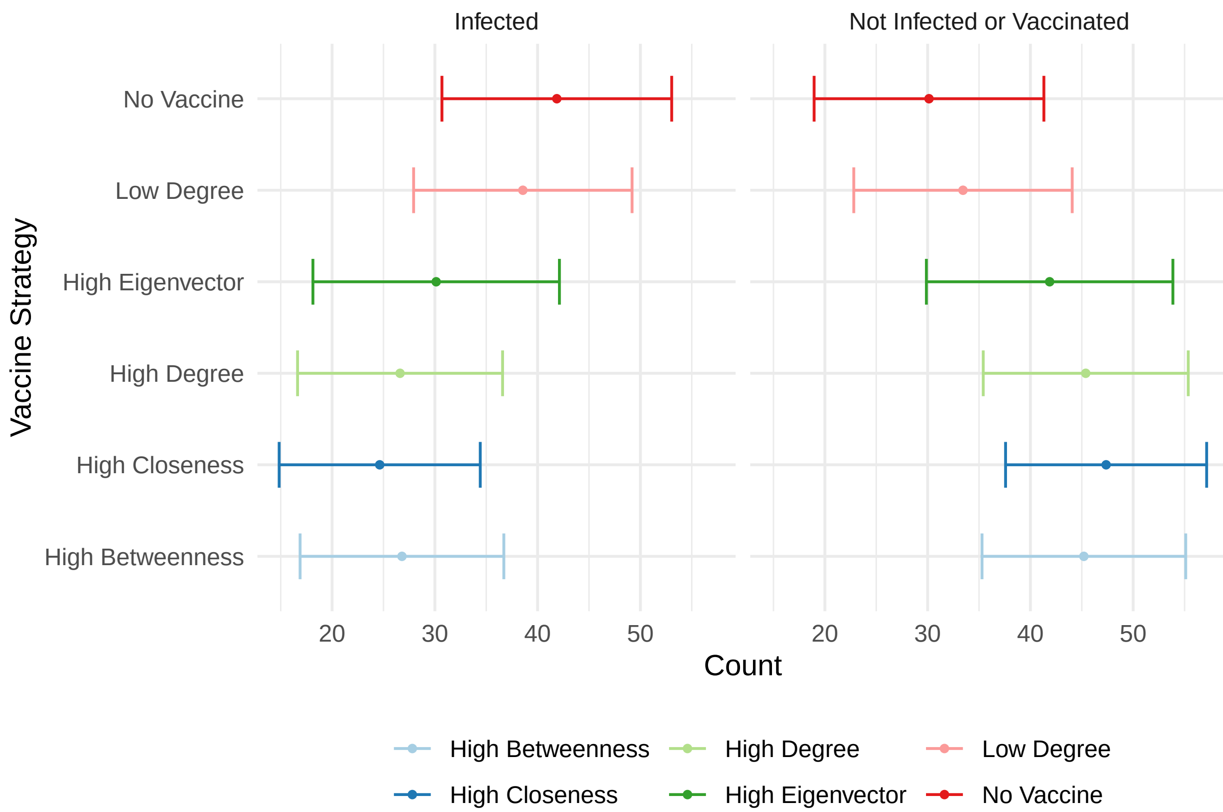

Figure 4: Average Infection Rates over 10,000 Simulations

In absolute terms, any of the alternative vaccination scenarios reduce the number of infected individuals by between about 9 (*High Eigenvector*) and 15 individuals (*High Closeness*). Accounting for the overall size of the simulated networks, these are the differences between protecting only 46 percent of individuals in the network (33 out of a total of 72 individuals are vaccinated or not infected) in the *Low Degree* scenario and 65 percent in the *High Closeness* scenario.

Among the four high centrality-based vaccination scenarios, the spread of infection is minimized under the *High Closeness* scenario, although the differences are not very large. The *High Eigenvector* scenario is the least efficient at reducing the spread of the infection.

Importantly, under all but the *Low Degree* scenario, a small number of vulnerable individuals (around 2.5) are infected. A rather significant reduction in the spread of infection, therefore, comes at a price of failing to protect a small number of vulnerable individuals, though this number is still smaller than under the *No Vaccine* scenario.

Figure 4 provides a visual summary of the same results. The first subfigure shows the infection rates for each scenario, whereas the second subfigure shows the combined numbers of individuals that are not infected or vaccinated. We can again see that the *Low Degree* strategy leads to the outcome that is very similar to that of *No Vaccine*, whereas all of the alternative scenarios lead to substantial decreases in virus spread.

Appendix C: Discussion of Control Variables

County GDP, logged is negative and statistically significant in the matched sample, indicating that more affluent counties experience, on average, few new cases. This adverse effect of the pandemic on lower income communities has been well documented (Mollalo, Vahedi, and Rivera 2020; Scannell, Oronce, and Tsugawa 2020). The same goes for the positive and statistically significant effects of *Percent Latino*, *Percent Other Race*—the pandemic's disproportionate effects on ethnic and racial minorities have also been well-established elsewhere (Rossen et al. 2020; Tai et al. 2021).

Unemployment Rate is negative and statistically significant in the full sample. Tempting as it is to interpret this as support for the first economic stimulus bill that provided generous unemployment benefits to employees that were laid off or let go due to the pandemic, we cannot make this interpretation, as the unemployment data are based on the pre-pandemic 2019 numbers.

County Population is negative and statistically significant in the matched sample, but not significant in the full sample. The matched sample, by construction, has lower median population than the full data (i.e., there are no counties in Oregon to match the size of Los Angeles or San Francisco counties). Therefore, this result may mean that moderate sized counties have fewer cases than smaller, rural counties (perhaps due to better infrastructure, such as hospitals).

Percent Foreign is negative and statistically significant, which may be a function of foreign nationals choosing to live in larger urban centers with better infrastructure. *Biden's Margin* is negative and statistically significant, indicating that counties that voted for Biden by larger margins have, on average, fewer Covid-19 cases. This likely reflects reluctance to adhere to public health policies (e.g., wearing a mask) and vaccine-skepticism among the supporters of former President Trump. Lastly, the coefficient on the *Lagged DV* is positive and statistically significant, which indicates a high temporal autocorrelation in new Covid-19 cases.

References

Bavelas, Alex. 1948. "A Mathematical Model for Group Structures." *Applied Anthropology* 7 (3): 16–30.

Boehmke, Frederick J., Abigail Matthews Rury, Bruce A. Desmarais, and Jeffrey J. Harden. 2017. "The Seeds of Policy Change: Leveraging Diffusion to Disseminate Policy Innovations." *Journal of Health Politics, Policy and Law* 42 (2): 285–307.

Coleman, James Samuel. 1964. *Introduction to Mathematical Sociology.* New York: Free Press.

Freeman, Linton C. 1977. "A Set of Measures of Centrality Based on Betweenness." *Sociometry*, 35–41.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "ERGM: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24 (3): 1–29.

Mollalo, Abolfazl, Behzad Vahedi, and Kiara M. Rivera. 2020. "GIS-Based Spatial Modeling of COVID-19 Incidence Rate in the Continental United States." *Science of the Total Environment* 728: 138884.

Padgett, John F., and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology* 98 (6): 1259–319.

Robins, Garry, and Philippa Pattison. 2001. "Random Graph Models for Temporal Processes in Social Networks." *Journal of Mathematical Sociology* 25 (1): 5–41.

Robins, Garry, Philippa Pattison, Yuval Kalish, and Dean Lusher. 2007. "An introduction to exponential random graph (p^*) models for social networks." *Social Networks* 29: 173–91.

Rossen, Lauren M., Amy M. Branum, Farida B. Ahmad, Paul Sutton, and Robert N. Anderson. 2020. "Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity—United States, January 26–October 3, 2020." *Morbidity and Mortality Weekly Report* 69 (42): 1522.

Scannell, Christopher A., Carlos Irwin A. Oronce, and Yusuke Tsugawa. 2020. "Association Between County-Level Racial and Ethnic Characteristics and COVID-19 Cases and Deaths in the USA." *Journal of General Internal Medicine* 35 (10): 3126–28.

Snijders, Tom AB, Philippa E Pattison, Garry L Robins, and Mark S Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36 (1): 99–153.

Tai, Don Bambino Geno, Aditya Shah, Chyke A. Doubeni, Irene G. Sia, and Mark L. Wieland. 2021. "The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States." *Clinical Infectious Diseases* 72 (4): 703–6.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.