2021

A Network Pandemic: Exploring the effects of Social Connectedness on the spread of COVID-19 in the United States

Mrinalini Bhushan

Claremont McKenna College

A Network Pandemic:

Exploring the effects of Social Connectedness on the spread of COVID-19 in the United States

submitted to
Professor Jessamyn Schaller, Ph.D.

by
Mrinalini Bhushan

For Senior Thesis

Spring 2021

May 3rd, 2021

Acknowledgment

I would like to start by thanking Professor Jessamyn Schaller for being my advisor during an atypical semester. Professor Schaller pushed me to think out of the box and challenged me to question and defend the decisions I made in my research process. I would not have produced a work I am proud of without her support.

I would also like to thank all my Professors in Economics, Psychology, and Literature at CMC who over the years have taught me lessons not just in writing and analyzing, but also in life. A special thank you to Professor Laura Grant for challenging me over the years to perfect my work and think creatively and to Professor Janet Smith for guiding me through this thesis writing process.

To my family members, who have always pushed me to achieve my dreams and expand my knowledge, thank you for giving me the opportunity to attend such a prestigious institution and for having my back through these past four years and beyond. I would not be where I am without you.

To Maxine Baghdadi and Matiss Ozols, my friends and supporters, thank you for providing a space for me to bounce my ideas, for helping me think in the right direction, and for spending all those late nights and early mornings writing thesis with me.

Thank you to my housemates, the Jamestown Jokers and Walnut Wizards, for making senior year as special as it has been and for supporting me through all the ups and downs. Lastly, to all my other peers at CMC, I will miss the classroom discussions and debates, that have taught me to think critically, for years to come.

Abstract

Social interactions influence the way we think and act. Recent literature on COVID-19 and social connectedness explores how social interactions influence people's perceptions of the risk from COVID-19 and their behaviors. This paper seeks to investigate how social connectedness, political ideologies, and physical interaction are associated with local COVID-19 case and death rates at the US county level. Social connectedness, as defined by (Bailey et al, 2018) measures connectedness between US counties based on Facebook friendship links. I examine whether a county's average social connectedness to other counties, as determined by the Facebook index, has an impact on its own COVID-19 cases and deaths per 100,000 people. I also examine whether a county's social connectedness to other counties that are in the top quartile for Republican voter percent or top quartile for COVID-19 deaths per 100,000 people has a positive impact on its own COVID-19 case and death rates. My results suggest that a county's overall social connectedness to other counties has no significant impact on its cases and deaths per 100,000. I also find that for every 10,000 unit increase in a county's social connectedness to top quartile Republican counties, cases per 100,000 decrease by 0.00276 and deaths per 100,000 increase by 0.0000380. Additionally, for every 10,000 unit increase in a county's social connectedness to counties in the top quartile for COVID-19 deaths per 100,000, a user county's cases per 100,000 increase by 0.00637 and deaths per 100,000 increase by 0.000478.

Contents

1 Introduction

Social media and network connectedness are integral to advancement in today's virtually-run world. The information we consume, because of the people we are connected to, on our social platforms informs the way we think and influences our ideas (Anagnostopoulos et al, 2008). More specific is the niche wherein social media impacts the health behavior and choices of individuals (Centola et al, 2013). In a world ravaged by COVID-19, it has become especially important to investigate health-related behaviors and statistics. To understand the connection between our interactions on social media and our health behavior, I build on existing research that explores how social connectedness and interaction influence the way people make decisions about their health (Bailey et al, 2018). While there is a lot of existing research on how health behavior and social relationships are interconnected (Centola et al, 2013), little is known yet about how the spread of COVID-19 is affected by social connectedness.

I use the Facebook Social Connectedness Index (SCI) dataset that Bailey et al, 2018 constructed, to examine how social connectedness on Facebook, among users across US counties, has impacted COVID-19 cases and deaths at the county level. More specifically, I examine how COVID-19 cases and deaths per 100,000 people vary for users' counties based on how connected they are to friends in counties that fall in the top quartile for Republican percentage and for COVID-19 deaths rates. I run multiple selection-on-observables regressions, controlling for own-county demographics and political leanings. I create interactions between Facebook's SCI and top quartile Republican and death rate counties. This allows me to create summary indices of connectedness with top quartile Republican counties and high death rate counties and estimate their effects on counties' own rates of illness and death from COVID-19.

By including a variety of controls in my regressions, I identify whether different levels of social connectedness amongst users' home counties that are otherwise similar in terms of demographics, political status, and employment statistics, are associated with variation in COVID-19 cases and deaths per 100,000 for the period between March 2020 and April 2021. I hypothesize that overall average social connectedness for a user's county to all other counties will have a positive impact on COVID-19 cases and deaths in their county, and as a result, increase cases and deaths. This is because individuals in highly socially connected areas are also more physically socially connected, especially in urban areas (Bailey et al, 2020). As a result, I hypothesize that this increased physical interaction, because of high social interaction, will lead to more COVID-19 cases and deaths given the opportunity for higher spread of the disease.

I further hypothesize that being friends with people in top quartile Republican percentage counties will be associated with higher COVID-19 cases and deaths in a user's home county. These hypotheses are based on studies by van Holm et al, 2020 and Grossman et al, 2020 that point to how political leanings influence behavior and concern related to COVID-19. Van Holm et al, 2020 find that liberals and moderates are less likely than conservatives to go on trips and are more likely to change behavior based on government suggestions as a result of how high their perceived risk of COVID-19 is. This is influenced by the information they receive from the media and their political ideologies. Grossman et al, 2020 also find that government leaders' stay-at-home mandates are more effective in reducing individual mobility in Democrat counties than in Republican counties.

I hypothesize that being socially connected with people in counties that fall in the top quartile for COVID-19 deaths per 100,000, will be associated with an increase in a user county's COVID-19 cases and deaths per 100,000. In this case, it is difficult to predict the expected effects because there are two opposing possibilities. These hypotheses are based on the premise that by being close to people in counties with high death rates, people in home counties also have a higher chance of contracting COVID-19 and suffering from it. An alternative hypothesis is that there would be a negative, or decreasing, effect on a user county's cases and deaths. This would be true if people in counties that are connected with friend counties that have high death rates are more careful with their health behavior. If people become more aware of the risks posed by COVID-19 through their friendships, they may choose to act carefully because of the fear assocaited with them contracting it.

I find that average social connectedness has no significant impact on COVID-19 cases and deaths per 100,000. I further find that being friends with people in top quartile Republican percentage counties is associated with significant negative effects on cases and significant positive effects on deaths. Lastly, I find that being friends with people in top quartile COVID-19 death counties is associated with significant and positive effects on both cases and deaths per 100,000 people for a home county.

These findings indicate that there is no significant impact of overall social connectedness on physical connectedness. This could be the result of the different lockdown rules implemented across counties. Furthermore, they point to the fact that social connectedness can act as a proxy for physical connectedness when considering strong interaction to counties with high death rates. Also, social connectedness can also act as a proxy for information flow; being friends with people in highly Republican counties increases death rates in a home county. However, the results from this regression show conflicting suggestions and hence, it is important to consider the discrepancy in the reporting of cases and deaths because of limited testing, hesitancy for people to test, and how that varies across counties.

2 Literature Review

There are three different avenues of literature that this paper examines: the impact of social network information on behavior, the relationship between social connectedness and COVID-19, and the association between political ideology and leanings on COVID-19 behavior.

Social network information, such as mobile data and social platform data, has many different uses such as tracking location, mobility, responsive behaviors, and understanding the spread of information. There is a large amount of research regarding social network information and the tracking of how diseases spread within a given population. A study by Wesolowski et al, 2012 illustrates how human mobility contributes to the spread of malaria across regions in Kenya. Here, the daily locations of 15 million individuals were monitored over the course of one year to create a trackable pathway of the disease. The results of this study are important to note as they reveal how the human travel network is denser than the parasite travel network in Kenya. While some studies look at indirect information, such as inbuilt location data from phones, others look at more direct sources, such as search engine entries (Ginsberg et al, 2009, Zimmer et al, 2018). Ginsberg et al's (2009) study focuses on how Google search queries can predict Influenza-like illness in multiple health regions around the US.

Another study by Klovdahl et al, 1994 shows how social network data can provide insight into human behavior, and not just movement, in light of a pandemic or the spread of a disease. Here, different pathogens such as HPV and HIV were examined with regards to certain population demographics they are more associated with. Researchers created a network design, connecting over 600 individuals, both directly and indirectly, and found that susceptible individuals were within a distance of 7 steps from a person infected with HPV or HIV. In this study, social network information is used to understand the transmission of pathogens according to participants' behavior. These studies provide crucial information and can often be used for early detection much faster than information provided by other governmental authorities. The estimates of the study by Ginsberg et al, 2009 were published 1-2 weeks ahead of CDC influenza-like illness surveillance reports. Hence, social data can be an extremely beneficial indicator to curb the spread of such pathogens and viruses before they result in a pandemic.

The Social Connectedness Index (SCI) that I use in this paper was created by Bailey et al, 2018 using Facebook data. It measures the connectedness of two local areas in different parts of the United States (US) or the world, through friendships on Facebook (Bailey et al, 2018). Previous research has used the SCI to illustrate that areas with more social ties to two COVID-19 hotspots in Westchester County, NY and Lodi province, Italy had more confirmed cases in their home county by the end of March 2020 (Kuchler et al, 2020). This literature sets the foundation for me to explore specifically US county relations and understand how political ideology and aggregate death rates throughout the pandemic are associated with own county case and death rates. This ties into other literature that suggests how health interventions can be made more effective by targeting individuals who have higher social reach (Kim et al, 2015). Hence, social data can be used both positively and negatively due to the large spillover effects that occur as a result of social connectedness. Previous research has focused mostly on tracking and forecasting the spread of diseases, whereas this paper seeks to analyze trends in relation to social categories, such as behavior, political leanings, and COVID-19 death rates.

COVID-19, caused by the coronavirus SAR-CoV-2, resulted in a pandemic that became severe in the US in approximately March 2020. The virus spread quickly after cases were first reported in Wuhan, China in December 2019, and resulted in serious threats to health systems all over the world (Vaughan et al, 2021). Countries responded differently to the spread of this virus, with some governments enforcing much stricter lockdowns and mobility restrictions than others (Haldar et al, 2020). Due to the devastating economic and social impacts (Nicola et al, 2020), COVID-19 has been widely researched to understand the most effective ways health officials and authorities should react in future scenarios (Chakraborty et al, 2020). Within this field, there is a subsection of research that looks particularly at how mobile and social network connectedness data can be used to understand the patterns of movement of populations, the impact of social interactions on ideologies, and the risks associated with such behavior (Piexoto et al, 2020, So et al, 2020). There are two possibilities that emerge here. This first is that social contact acts as a proxy for physical contact, and the second is that it acts as a proxy for information flow. Therefore, I choose to examine social connectedness to counties with high death rates and high Republican percentage. The first sheds light on how social connectedness acts as a proxy for physical contact and the second sheds light on how acts as a proxy for information flow.

A study by Piexoto et al, 2020 analyzes mobile geological data of people within Brazil to show which cities were at higher risk of infection based on a simulation of metro populations. Another study uses the same framework of network analysis to graph the degree of connectedness among people and its relation to confirmed COVID-19 cases (So et al, 2020). Such research is important for policy building regarding resource allocation and planning, especially since these analyses allow researchers to predict the risk of a potential outbreak much earlier and, thus, potentially save thousands of lives. It is important to note that social connectedness does not only predict the spread of disease, but also impacts the spread of diseases such as COVID-19. A study by Fritz et al, 2021 examines the spread of COVID-19 in Germany and illustrates how when there is less social activity and connectedness between people in German federal administrative districts, as measured through Facebook data, there are lower weekly cross-infections between those regions.

An important subset of COVID-19 specific literature investigates how political leanings and partisanship impact people's responses to COVID-19 restrictions, and in turn, the case and death rates in their hometowns. A study by Grossman et al (2020), which examines how partisanship and political leanings play a role in people following government restriction orders in the US, sets the foundation for the selection of my independent variables. The study found that state government leader's restrictions were better followed in Democrat leaning counties compared to Republican leaning counties. In another study, van Holm et al (2020) found that political ideological commitments influence beliefs on COVID-19 in the US. They also found that liberals were less likely to make trips during the pandemic than conservatives, were more likely to adapt their behavior based on government guidelines, and that their beliefs about other people's behavior had some influence in predicting their own behavioral changes.

Another study by Bursztyn et al (2020) explores how social connectedness can act as a proxy for information flow. Researchers studied how early misinformation on mass media influences health outcomes. They directly compared Fox News, CNN, and MSNBC on COVID-19 content and compared their users' preventative techniques. Researchers found that areas with high exposure to TV shows that downplayed COVID-19 threats experienced higher COVID-19 cases and deaths. In this study, that channel was Fox News, a primarily Republican supporting and watched channel (DellaVigna et al, 2006). These three studies compel me to explore how social connectedness to counties with Republican status impacts own counties' COVID-19 cases and deaths given the evidence that political ideologies impact health behaviors.

All the literature described above lays the framework for my study which seeks to explore how social connectedness informs behaviors and outcomes related to COVID-19. The first set of literature, which focuses on how network connectedness impacts human behavior has informed my choice to explore the Facebook Social Connectedness dataset in my regressions. The second set of literature, which points to how health behavior is associated with social connectedness, has influenced my decision to explore a very relevant and pressing issue in the world today, the COVID-19 pandemic. The third set of literature on political ideologies and how they bias health behavior and how more socially connected regions are also more physically connected, motivated me to create interactions between social connectedness, high death rate counties, and highly Republican counties.

Influenced by the three previous strands of literature, my study examines three things: The first is how overall social connectedness affects COVID-19 cases and deaths for a US county and I hypothesize a positive effect; the second is how a county's overall social connectedness to counties in the top quartile for Republican percentage impacts its own cases and deaths per 100,000; I hypothesize a positive effect; the third is how a county's overall social connectedness to counties in the top quartile for deaths per 100,000 impacts its own cases and deaths per 100,000. In this instance I hypothesize a positive effect but acknowledge that an alternative hypothesis could cause a negative effect if people get scared and choose to act cautiously when they have friendships in high death-rate counties.

3 Method

3.1 Data

The Facebook SCI Index measures the strength of Facebook connections between Facebook friends across various geographical locations within the US and across the globe at a county-county, county-country, and country-country level. It is "constructed using aggregated and anonymized information from the universe of friendship links between all Facebook users as of April 2016" (Bailey et al, 2018). Researchers map users to county and country locations to get total friendship links between different geographies. The index only reflects users who were active on Facebook 30 days prior to April 2016 and each link is treated identically (Bailey et al, 2018). The index is constructed between 3,229 US county pairs and between every US county and foreign country. "Relative differences in the index correspond to relative differences in the total number of friendship links" (Bailey et al, 2018). For my study, I use the county-county dataset that is constructed on a scale of 1 - 1,000,000. It has three columns, Facebook User 1's County by FIPS code, Facebook user 2's (friend) county FIPS code, and SCI for each county pair. If a pair of counties has a SCI twice as large as another pair of counties, that means that the Facebook user in the own county, in this pair, is twice as likely to be socially connected to the Facebook user in a friend county as compared to another pair of counties. For example, (See Figure 1) if Greene county, AL has a SCI of 100,357 and Macon county has a SCI of 36,046, then Greene county is 64,311 (100,357 – 36,046) times more socially connected, to all other counties, than Macon county.

I merge this dataset with the New York Times (New York Times) county-level COVID-19 data on reported cases and deaths as of April 6, 2021. This was the date I downloaded the dataset from the NY times website. I also use county level demographics on age (U.S Census Bureau), race (U.S Census Bureau), income (U.S Census Bureau, SAIPE), poverty (U.S Census Bureau, SAIPE), employment rates (U.S Department of Agriculture, Economic Research Service), urban percent (U.S Census Bureau), and political status (McGovern, 2016) to control for factors that are correlated with social connectedness based on the study by (Bailey et al, 2018). This study shows that being white is a determinant of social connectedness and that it is important to adjust for race, education, income, and poverty rates to compare how they relate to social connectedness across counties. I control for percent of urban population (greater than 50,000 people) and urban clusters (greater than 2500 and less than 50,000 people). I create an interaction between User 2 (friend) county's SCI and counties in the top quartile for Republican percentage, and between User 2 (friend) county's SCI and counties in the top quartile for COVID-19 deaths per 100,000. To do this, I identify the counties in the top quartiles for both Republican percentage and COVID-19 deaths and create dummy variables. The number 1 indicates counties in the top quartile and 0 indicates counties in the other three quartiles. I interact the average scaled SCI with the dummy variables to get my two independent variables – friend county SCI interacted with top quartile Republican percentage and friend county SCI interacted with deaths per 100,000. These interactions allow me to investigate two subsets of data from my main dataset that focus on how social connectedness changes with Republican political ideology and with high death rates in friend counties. My final dataset is achieved by collapsing data to get means of all variables by User 1's county FIPS Code. This gives averages of all independent variables and controls that I use in my regression model.

3.2 Descriptive Statistics

Table 1 outlines descriptive statistics for all the variables included in my regression. The means for top quartile Republican percent SCI (85,117) and top quartile Deaths SCI (51,883) are both higher than the mean for average social connectedness Index (46,075). This is likely because rural counties are overrepresented when taking an average across all counties without weighting by population. By taking one observation per county, small county representation is getting upweighted and has a disproportionate impact compared to population density. The means for user county cases per 100,000 (8,146) and deaths per 100,000 (167.9) are important figures to compare my results for economic significance.

Santa Clara County in California (FIPS Code 6085) has the lowest average social connectedness Index of 1470. Arthur County in Nebraska (FIPS 31005) has the highest average social connectedness Index of 657,352. Benton County in Oregon (FIPS 41003) has the lowest cases per 100,000 and Chattahoochee County in Georgia (FIPS 13053) has the highest cases per 100,000. Aleutians West in Alaska (FIPS 02016) has the lowest deaths per 100,000 and Gove County in Kansas (FIPS 20063) has the highest deaths per 100,000. Bayamon in Puerto Rico (FIPS 72021) has the lowest average connectedness to top quartile Republican percent counties and Arthur County in Nebraska (FIPS 31005) has the highest average connectedness to top quartile Republican percentage counties. Saipan in

Table 1: Descriptive Statistics

VARIABLES	N	Mean	sd	Min	Max
Average Social Connectedness Index	3,229	46,075	56,632	1,470	657,352
Top quartile Republican percent SCI	3,229	85,177	210,105	356.9	2.62055
Top quartile Deaths SCI	3,229	51,883	144,132	735.2	2.20436
User Cases Per 100,000	3,229	8,146	3,036	0	27,845
User Deaths Per 100,000	3,229	167.9	94.27	0	834.6
Percent of Population Urban	3,229	42.49	32.03	0	100
Percent of Population Urban Area	3,229	19.99	34.56	0	100
User Democrat percent	3,229	0.65	0.161	0.054	0.962
User High School Dropout Rate	3,229	12.9	6.3	1.1	73.6
User High School Graduation Rate	3,229	34.1	7.2	7.8	57.4
Log Median Household Income	3,229	10.9	0.242	10.1	11.93
Fraction of Population Black	3,229	9.8	14.4	0.124	88.5
Fraction of Population Hispanic	3,229	9.7	13.8	0.648	96.3
Fraction of Population White	3,229	75.8	20.2	2.686	97.8
Poverty percentages	3,229	14.4	5.8	2.7	47.7
Fraction of age > 65	3,229	19.7	4.8	4.9	58.1
Fraction of age 45-64	3,229	26.2	2.6	9.6	38
Fraction of age 20-44	3,229	29.6	4.4	14	57.3
User Unemployment Rate	3,229	4.15	1.795	0.7	19.3

This table reports descriptive statistics of all the variables considered in following regression models. The table notes: Number of Observations (N), Mean, Standard Deviation (sd), Minimum Values, and Maximum Values.

Northern Mariana Islands (FIPS 69110) has the lowest average connectedness to top quartile COVID-19 death counties and Wibaux County in Montana (FIPS 30109) has the highest average connectedness to top quartile COVID-19 death counties.

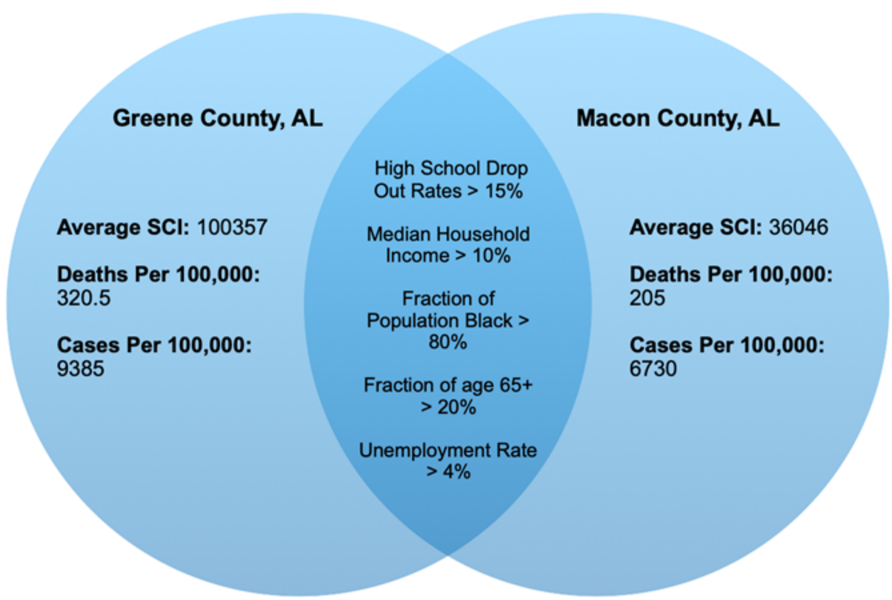
3.3 Motivation for Regression

My paper builds on the foundational premise that two counties in the US, when controlled for similarities in age, education, poverty rates, income levels, urban density, and political leanings, with different SCI's would have diverse case and death rates. Greene County in Alabama (FIPS Code 1063) and Macon County in Alabama (FIPS Code 1087) are two examples of similar counties with varying SCI's, cases and deaths per 100,000. I filtered for these counties based on the following randomly chosen factors:

Table 2: Filters Applied for Motivation for Regression Example

High School Drop Out Rate	>15%
Log of Median Household Income	>10%
Fraction of Population Black	>80%
Fraction of age > 65	>20%
Unemployment Rate	>4%

Figure 1: Findings from Table 2 Filters on County SCI, Cases and Deaths Per 100,000



3.4 Regression Model

I use a selection-on-observables approach to conduct multiple linear regressions. My dataset is cumulative, ranging from January 2020 to April 2021. My sample has 3,229 observations. The units of analyses for my dependent variables are COVID-19 cases and deaths per 100,000 and for my independent variables are the average SCI and average SCI interacted with top quartile Republican percentage and top quartile deaths per 100,000.

To account for county effects, I control for (ln) median household income and percent of the population in poverty; I do this to adjust for income differences related to social connectedness across counties. I also control for race and age composition given existing research that suggests the share of population that is white is a determinant of social connectedness; it has a negative and significant effect on the log of SCI (Bailey et al, 2018). I also control for education levels, including high school graduation and dropout rates to account for educational differences related to social connectedness (Bailey et al, 2018). I include controls for county political leanings, in particular Democrat percentage of a county, to account for how they influence behavior and decisions during COVID-19 (van Holm et al, 2020). Further, I control for population of percent urban ($>$50,000 people) and population of percent urban cluster ($>$2500 people and $<$ 50,000 people) to acknowledge their effect on social connectedness; previous literature has found that residents in urban areas are mostly connected to people located nearby and that urban areas that are well-connected have higher SCI's (Bailey et al, 2020). For ease of inference, I divide SCI by 10,000 to get a more scalable and comparable index and use that index in my regressions.

I run my regressions using interactions between SCI and top quartile deaths per 100,000 and Republican percentage to understand how a subset of my sample interacts with SCI and COVID-19 statistics rather than focusing on every US county pair's interaction.

I have four model specifications (equations) and I run eight regressions, two for each model. The first is a linear regression that examines the overall effect of Average Scaled SCI on COVID-19 cases and deaths per 100,000. In equations (1), (2), (3), and (4), the dependent variable (Covid Outcome) is either COVID-19 cases or deaths per 100,000. α is a constant, β_1 $Average\ Scaled\ SCI_i$ is average social connectedness for a user county, ρX_i is a vector of control variables, ε is a measure of heteroskedasticity.

$$(1) \qquad (Covid\ Outcome\ )_i = \alpha + \beta_1\ Average\ Scaled\ SCI_i + \rho X_i + \varepsilon$$

The second is a linear regression that examines the effect of a county (User 1) being more socially connected with a friend county (User 2) that is in the top quartile for Republican percentage as of April 6^{th}, 2021, and how that affects its (User 1's) COVID-19 cases and deaths per 100,000. β_2 $Top\ quartile\ Republican \times SCI_i$ is an interaction between counties that fall in the top quartile for Republican percentage and average scaled social connectedness.

(2)

$$(Covid\ Outcome\ )_i = \alpha + \beta_1\ Average\ Scaled\ SCI_i + \beta_2\ Top\ quartile\ Republican \times SCI_i + \rho X_i + \varepsilon$$

The third is a linear regression that examines the effect of a county (User 1) being socially connected with a friend county (User 2) that is in the top quartile for COVID-19 deaths per 100,000 as of April 6^{th}, 2021, and how that affects its (User 1's) COVID-19 cases and deaths per 100,000. β_3 $Top\ quartile\ deaths \times SCI_i$ is an interaction between counties that fall in the top quartile for deaths per 100,000 and average scaled social connectedness.

(3)

$$(Covid\ Outcome\ )_{\ i} = \ \alpha \ + \ \beta_1 \ Average\ Scaled\ SCI_i \ + \ \beta_3 \ Top\ quartile\ deaths \times SCI_i \\ + \ \rho X_i \ + \ \varepsilon$$

The fourth is a linear regression that examines the effect of a county (User 1) being socially connected with a friend county (User 2) that is in the top quartile for Republican percentage; this regression also controls for the effect of that county being connected with a friend county that is in the top quartile for COVID-19 deaths per 100,000 and examines how this impacts User 1 county's COVID-19 cases and deaths per 100,000. I include both my independent variable interactions in this regression to account for any correlation that may arise between top quartile Republican percentage counties and top quartile deaths per 100,000 counties.

(4)

$$(Covid\ Outcome\ )_{\ i} = \ \alpha \ + \ \beta_1 \ Average\ Scaled\ SCI_i \ + \ \ \beta_2 \ Top\ Quartile\ Republican \times SCI_i \\ + \ \beta_3 \ Top\ quartile\ deaths \times SCI_i \ + \ \rho X_i \ + \ \varepsilon$$

4 Results

Table 3 reports the results from two correlational linear regressions for the entire sample of user_loc average social connectedness and COVID-19 cases and deaths per 100,000 as of April 6^{th}, 2020. The dependent variable in Model 1 is COVID-19 cases per 100,000 and in Model 2 is COVID-19 deaths per 100,000. The independent variable and controls are defined in the regression analysis section.

Table 3: Linear Regression of Average Scaled SCI effects on COVID-19 cases and deaths

	Cases Per 100,000	Deaths Per 100,000
VARIABLES	Model 1	Model 2
Average SCI	0.138	-0.385
	(18.28)	(0.518)
User percent Democrat	-0.512***	-0.00588**
	(0.118)	(0.00272)
User High School Drop Out Rate	-0.618	-0.0151
	(0.774)	(0.0181)
User High School Graduation Rate	1.813	0.121***
	(1.368)	(0.0294)
Log Median Household Income	-0.318**	-0.0302***
	(0.128)	(0.00295)
Fraction of Population Black	-0.276**	-0.00241
	(0.118)	(0.00257)
Fraction of Population Hispanic	0.315***	0.00394*
	(0.0942)	(0.00215)
Fraction of Population White	0.105	-0.0179***
	(0.125)	(0.00283)
Poverty percentages	-2.089**	-0.0227
	(0.853)	(0.0191)
Fraction of age > 65	-0.907***	-0.0132***
	(0.143)	(0.00253)
Fraction of age 45-64	-0.563***	0.0113***
	(0.132)	(0.00292)
Fraction of age 20-44	-0.785***	-0.0127***
	(0.183)	(0.00338)
User Unemployment Rate	-21.58***	-0.0748
	(5.225)	(0.113)
Percent of Population Urban	6.902	0.263**
	(4.778)	(0.109)
Percent of Population Urban Cluster	-6.629*	0.222***
	(3.776)	(0.0793)
Constant	11,604***	161.3***
	(468.3)	(10.14)
Observations	3,229	3,229
R-squared	0.076	0.11

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

I find in my first regression, Model 1, SCI has an insignificant effect on cases per 100,000 and in Model 2 the variable also has an insignificant effect on deaths per 100,000. These findings go against my initial hypothesis that average social connectedness has a positive, significant impact on cases and deaths per 100,000. This is likely because of the effects of lockdown and travel that aren't accounted for in these models. As a result, I don't find anything significant from these results. However, user unemployment rate has a negative and significant impact on cases per 100,000. This is likely because higher unemployment leads to more people staying home based on how lockdown policies were implemented in different counties. These results are simply correlational and set the foundation for my other regressions where I hope to find associations between social connectedness, cases, and deaths per 100,000.

Tab1e 4 reports the results from two linear regressions examining the effects of friend counties being top quartile Republican % on user or home county COVID-19 cases and deaths per 100,000 as of April 6^{th}, 2020. The dependent variable in Model 3 is COVID-19 cases per 100,000 and in Model 4 is COVID-19 deaths per 100,000. The independent variable and controls are defined in the regression analysis section. ***, **, * indicate that the parameters estimate is significantly different from zero at 1 percent, 5 percent, and 10 percent respectively.

In my second regression, in Model 3, top quartile Republican percent SCI has a negative and statistically significant impact on cases per 100,000. This model suggests that a 10,000 unit increase in a user county's average social connectedness to friend counties that are higher in the top quartile Republican scale is associated with a decrease in that user county's COVID-19 cases per 100,000 by -0.00276 ($p<0.01$).

Table 4: Linear Regression of top quartile Republican County Average Scaled SCI effects on COVID-19 cases and Deaths

VARIABLES	Cases Per 100,000 Model 3	Deaths Per 100,000 Model 4
Top Quartile Republican Percent SCI	-0.00276***	3.80e-05*
	(0.000936)	(0.0000222)
Average SCI	87.68**	-1.591**
	(35.41)	(0.76)
User percent Democrat	-0.359***	-0.00800***
	(0.125)	(0.00291)
User High School Drop Out Rate	-0.681	-0.0142
	(0.773)	(0.018)
User High School Graduation Rate	0.705	0.136***
	(1.426)	(0.0312)
Log Median Household Income	-0.328**	-0.0301***
	(0.128)	(0.00295)
Fraction of Population Black	-0.287**	-0.00226
	(0.118)	(0.00257)
Fraction of Population Hispanic	0.299***	0.00416*
	(0.0946)	(0.00216)
Fraction of Population White	0.125	-0.0182***
	(0.126)	(0.00281)
Poverty percentages	-2.409***	-0.0183
	(0.856)	(0.0191)
Fraction of age > 65	-0.967***	-0.0123***
	(0.144)	(0.00257)
Fraction of age 45-64	-0.548***	0.0111***
	(0.132)	(0.00292)
Fraction of age 20-44	-0.846***	-0.0119***
	(0.184)	(0.00341)
User Unemployment Rate	-22.28***	-0.0651
	(5.289)	(0.112)
Percent of Population Urban	9.083*	0.233**
	(4.796)	(0.11)
Percent of Population Urban Cluster	-4.94	0.199**
	(3.785)	(0.0786)
Constant	11,585***	161.6***
	(472.9)	(10.05)
Observations	3,229	3,229
R-squared	0.079	0.111

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

These findings go against my initial hypothesis that being friends with top quartile Republican percentage Counties would have a positive impact on cases and deaths per 100,000 for a user county. This discrepancy could be because my initial hypothesis was incorrect or because my model doesn't fully account for unobservable causal linkages or confounding effects through my list of control variables. My hypothesis is based on findings from existing studies which suggest political status impacts behavior and Republican or conservative leaning individuals respond less positively to COVID-19 restrictions (Grossman et al, 2020, van Holm et al, 2020, DellaVigna et al, 2006). Model 3's findings indicate the opposite and suggests that increasing social connectedness between a user county and a friend county that falls in the top quartile Republican percentage leads to a statistically significant decrease in user county's COVID-19 cases per 100,000.

This could be the case because people in user counties actively chose to not follow in their friend counties' footsteps and chose to be careful about their whereabouts and actions during the pandemic. On the other hand, Model 4 suggests an increase in COVID-19 deaths because of increased social connectedness between a user county and a highly Republican. A 10,000 unit increase in Average Social Connectedness leads to an increase in deaths per 100,000 by 0.0000380. The disparity in the fact that cases decrease, and deaths increase can be explained by the fact that cases reported can be misleading and are influenced by people's willingness to test and the availability of tests. While case level data can be fudged and unreliable, once a death occurs it is more likely to get reported. These figures may also not account for COVID-19 diagnoses that occur after a death occurs.

Table 5 reports the results from two linear regressions examining the effects of friend counties having top quartile COVID-19 deaths per 100,000 on user or home county COVID-19 cases and deaths per 100,000 as of April 6^{th}, 2020. The dependent variable in Model 5 is COVID-19 cases per 100,000 and in Model 6 is COVID-19 deaths per 100,000. The independent variable and controls are defined in the regression analysis section. ***,

**, * indicate that the parameters estimate is significantly different from zero at 1 percent, 5 percent, and 10 percent respectively.

In my third regression, in Models 5 and 6, top quartile deaths per 100,000 SCI has a positive and statistically significant effect on cases and deaths per 100,000 respectively. In Model 5, a 10,000 unit increase in a user county's average social connectedness to counties in the top quartile for deaths per 100,000, leads to an increase in the user county's COVID-19 cases per 100,000 by 0.00637 ($p<0.01$). In Model 6, a 1 unit increase in a user county's average social connectedness to counties in the top quartile for deaths per 100,000, leads to an increase in the user county's COVID-19 deaths per 100,000 by 0.000478. This could likely be explained by mobility, travel, and physical contact between counties that are highly socially connected.

Both the findings in this regression align with my initial hypotheses that COVID-19 cases and deaths per 100,000 would increase as social connectedness between user county and a friend county in the top quartile for deaths per 100,000 increased. In Model's 5 and 6, the statistically significant findings both suggest that increased social connectedness has a positive effect on cases and deaths per 100,000. By being friends with people in counties with the highest quartile of death rates, user county case rates could go up through social interaction. Simultaneously, deaths could also go up. These findings do not align with the findings from Model 3, where we see user county cases go down when they are friends with top quartile Republican counties. This difference could likely be explained by the fact that the sample of counties considered in is different; this is because the top quartile Republican percentage counties are not all the same as top quartile death counties. Thus, we see disparity in two regressions.

Table 5: Linear Regression of top quartile COVID-19 deaths' counties Average Scaled SCI effects on COVID-19 cases and deaths

	Cases Per 100,000	Deaths Per 100,000
VARIABLES	Model 5	Model 6
Top quartile Deaths SCI	0.00637***	0.000478***
	(0.000601)	(0.0000303)
Average SCI	-55.72***	-4.578***
	(16.11)	(0.532)
User percent Democrat	-0.479***	-0.0034
	(0.116)	(0.0021)
User High School Drop Out Rate	-0.994	-0.0434***
	(0.761)	(0.0139)
User High School Graduation Rate	1.286	0.0810***
	(1.328)	(0.0238)
Log Median Household Income	-0.149	-0.0175***
	(0.126)	(0.00244)
Fraction of Population Black	-0.312***	-0.00518**
	(0.116)	(0.00212)
Fraction of Population Hispanic	0.297***	0.0026
	(0.0921)	(0.00169)
Fraction of Population White	0.201*	-0.0107***
	(0.122)	(0.00227)
Poverty Percentages	-2.258***	-0.0354**
	(0.838)	(0.0151)
Fraction of age > 65	-0.863***	-0.00985***
	(0.141)	(0.00204)
Fraction of age 44-64	-0.607***	0.00803***
	(0.13)	(0.00237)
Fraction of age 20-44	-0.755***	-0.0105***
	(0.181)	(0.00261)
User Unemployment Rate	-19.30***	0.0966
	(4.947)	(0.0875)
Percent of Population Urban	6.91	0.264***
	(4.622)	(0.0803)
Percent of Population Urban Cluster	-6.037	0.267***
	(3.713)	(0.0642)
Constant	11,132***	125.9***
	(449.1)	(9.332)
Observations	3,229	3,229
R-squared	0.111	0.453

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6 reports the results from two linear regressions examining the effects of friend counties having top quartile Republican percentage or COVID-19 deaths per 100,000 on user or home county COVID-19 cases and deaths per 100,000 as of April 6^{th}, 2020. The dependent variable in Model 7 is COVID-19 cases per 100,000 and in Model 8 is COVID-19 deaths per 100,000. The independent variable and controls are defined in the regression analysis section. ***, **, * indicate that the parameters estimate is significantly different from zero at 1 percent, 5 percent, and 10 percent respectively.

In my 4^{th} regression, in Model 7, top quartile Republican SCI has a negative and statistically significant effect on a user county's COVID-19 cases per 100,000. Contrastingly, top quartile deaths SCI has a positive and significant effect on a user county's COVID-19 cases per 100,000. In Model 8, both top quartile Republican percentage SCI and top quartile deaths SCI have positive and statistically significant effects on a user county's COVID-19 deaths per 100,000.

In Model 7, a 10,000 unit increase in a county's average social connectedness to counties in the top quartile for Republican percentage, leads to a decrease in the user county's COVID-19 cases per 100,000 by 0.00262 ($p<0.01$). These findings go against my initial hypothesis that there would be a positive effect on cases per 100,000 of the top quartile Republican percentage independent variable. This could be because my hypothesis was incorrect or because the variables I have controlled for do not fully capture causal linkages or acknowledge confounding factors. Also in Model 7, a 10,000 unit increase in a county's average social connectedness to counties in the top quartile for deaths per 100,000 leads to an increase in the user county's COVID-19 cases per 100,000 by 0.00634 ($p<0.01$). Model 8, a 10,000 unit increase in a county's average social connectedness to counties in the top quartile for Republican percentage leads to an increase in the user county's COVID-19 deaths per 100,000 by 0.0000487. Additionally, a 10,000 unit increase in a user county's average social connectedness to counties in the top quartile for deaths per 100,000 leads to an increase in the user county's COVID-19 deaths per 100,000 by 0.000479 ($p<0.01$).

Table 6: Linear Regression of top quartile Republican % and Deaths SCI on COVID-19 cases and deaths

VARIABLES	Cases Per 100,000 Model 7	Deaths Per 100,000 Model 8
Top quartile Republican percent SCI	-0.00262***	4.87e-05**
	(0.000894)	(0.0000191)
Top quartile Deaths SCI	0.00634***	0.000479***
	(0.000586)	(0.0000308)
Average SCI	27.58	-6.126***
	(35.78)	(0.619)
User Democrat percent	-0.334***	-0.00610**
	(0.12)	(0.00246)
User High School Drop Out Rate	-1.053	-0.0423***
	(0.759)	(0.0139)
User High School Graduation Rate	0.236	0.101***
	(1.382)	(0.0252)
Log Median Household Income	-0.159	-0.0173***
	(0.127)	(0.00247)
Fraction of Population Black	-0.323***	-0.00499**
	(0.117)	(0.00211)
Fraction of Population Hispanic	0.283***	0.00288*
	(0.0923)	(0.0017)
Fraction of Population White	0.219*	-0.0111***
	(0.123)	(0.00225)
Poverty percentages	-2.561***	-0.0298*
	(0.84)	(0.0152)
Fraction of age > 65	-0.921***	-0.00878***
	(0.143)	(0.00205)
Fraction of age 44-64	-0.592***	0.00775***
	(0.13)	(0.00236)
Fraction of age 20-44	-0.813***	-0.00941***
	(0.182)	(0.00265)
User Unemployment Rate	-19.97***	0.109
	(4.994)	(0.0876)
Percent of Population Urban	8.979*	0.226***
	(4.66)	(0.0784)
Percent of Population Urban Cluster	-4.437	0.237***
	(3.725)	(0.0647)
Constant	11,116***	126.2***
	(451.7)	(9.365)
Observations	3,229	3,229
R-squared	0.113	0.454

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

These findings align with my initial hypothesis that cases and deaths per 100,000 increase as a user county's social connectedness to top quartile deaths per 100,000 counties increases. They also align with my hypothesis that deaths per 100,000 would increase as a user county's social connectedness to top quartile Republican percentage counties increases. One possible explanation for this could be that social connectedness acts as a proxy for physical connectedness and for information flow. Thus, people with friends in counties with high death rates, through their physical interactions, see an increase in cases and deaths in their own counties. Also, people with friends in counties that are Republican, through the information they receive on media (DellaVigna et al, 2006) underplay COVID-19 threats and see an increase in deaths in their counties. The decrease in cases can be explained by potential discrepancies in testing rates, testing availability, and people's inclination to test varying across counties. This is just one possible hypothesis and does not imply economically significant causality. There are several other factors that could influence these results such as the number of nursing homes in the highly Republican counties that had high death rates throughout the pandemic or the number of people with pre-existing conditions.

5 Limitations

There are several limitations to my dataset. The original Facebook dataset on SCI between county pairs is only representative of counties where there is a strong presence of Facebook Users. As a result, my findings may not fully represent the influence of political leanings on social connectedness and on COVID-19 case and death rates for all US counties. Also, counties with low social connectedness indexes may still have frequent social interaction on other platforms and in person that aren't taken into account in this Facebook dataset.

A significant problem in my regressions is that the dependent variable of cases per 100,0000 may not be accurate given that there are many unreported cases. Due to the significant differences across counties and states in the number of tests conducted, the number of test kits that were distributed and available, the costs of the kits, and the differences in people's proclivity to test, it is important to consider underreporting when analyzing my results. Death rate results are more accurate and provide more meaningful regressions, where I find little or no significance in key variables. This calls into question how truly representative my significant findings on the case rate regressions are.

In the regressions where top quartile deaths and Republican percentage interact with SCI to form independent variables, the model does not control for the effect of the two interactions on each other. This is because the samples of counties that fall under top quartiles for COVID-19 deaths per 100,000 and for Republican percentage are different. As a result, these two interactions cannot be directly compared with one another. Thus, my final regression that includes both interactions to test effects of cases and deaths per 100,000 may not fully account for any correlation between the two variables. Additionally, when assessing for economic significance, my findings are weak. This is because the average cases per 100,000 people for all US counties is 8,146 and the average deaths per 100,000 is 167.9 (See Table 2) and the significant effects for my regressions are smaller than one case or death per 100,000 people.

Additionally, even though some of my results are significant, I do not test for where the significant differences are coming from. For example, in my regression on how COVID-19 deaths per 100,000 are affected by a user county's social connectedness to top quartile Republican counties, my findings that deaths increase can only partially be explained by my hypothesis that social connectedness acts as a proxy for information flow. There could be several omitted variables or unconsidered factors that my model does not account for such as the presence of nursing homes, government mandates implemented, and how news was different at local versus national levels.

Lastly, it is important to note that my study only considers top quartile Republican and top quartile COVID-19 deaths per 100,000 at the county level. As a result, my findings only represent two possible COVID-19 statistics of many others that could be examined with this data. While this is a limitation, this also sets scope for future research on other potential factors and indicators that may be worth investigating in terms of their impact on COVID-19 behaviors linked with social connectedness.

6 Conclusion

In this study, I seek to examine how social connectedness on Facebook influences COVID-19 trends such as cases and deaths per 100,000 at the US county level. My initial hypothesis was that average social connectedness would have a positive impact on COVID-19 cases and death rates in a user county. I further hypothesized that a user county that was strongly socially connected to top quartile Republican percentage counties would see a positive effect on its own cases and deaths per 100,000 people. I also hypothesized that a user county that is strongly connected to top quartile deaths per 100,000 people counties would see a positive effect on its own cases and deaths per 100,000 people. I find results inconsistent with my hypothesis; average social connectedness has no significant impact on COVID-19 cases and deaths per 100,000 people. Average social connectedness for a user county to counties with top quartile deaths per 100,000 also has a positive effect on COVID-19 cases and deaths per 100,000 people in a user county. However, my findings on the effect of average social connectedness for a user county to counties in the top quartile for Republican percentage only partially align with my initial hypothesis. These findings suggest social connectedness to top quartile Republican percentage counties has a negative impact on cases but a positive effect on deaths per 100,000 people.

My motivation for this study comes from copious research that points to how network connectedness and social interaction can influence the way we think and act. More specifically, previous literature on COVID-19 behavior being influenced by social connectedness and political status lays the framework for my investigation. While some of my results are significant, they yield little to no true significance given how small the increases and decreases of COVID-19 case and death rates are when associated with the social connectedness index. Additionally, while my R-squared increases as I include new independent variables into my second, third, and fourth regression models, this does not indicate where the variation in the dependent variable is best explained and does not account for the fact that the two interacted independent variables represent different samples.

While my paper builds on a growing work of literature on social connectedness and political behavior's impact on COVID-19, there is a lot more to be explored. My sample only touched on a small sub-sample of the original social connectedness dataset and there are many other relationships that can be explored to understand how political leanings impact COVID-19 behavior and other health related behavior. My models can be improved by adding more interaction variables related to other controls such as top quartile diabetes rates or top quartile obesity (The Economist, 2021) interacted with SCI to form other independent variables. These interactions would increase the likelihood of a point of comparison between the results from my two independent variables because they more holistically cover the US county sample more holistically.

It could also be distinctive to test how these regressions run differently for top quartile Democrat counties now that President Biden has taken over the White House to see if there are any discrepancies in expected hypotheses for those counties. It could also be useful for future researchers to delve more into the way political information sources and credibility varies across counties. My hypothesis focuses on Republican counties due to findings from my literature review, Grossman et al, 2020 and Bursztyn et al, 2020, that pointed specifically to how Republican counties are less likely to respond to COVID-19 mitigation measures due to the information they receive through mass media and news channels (DellaVigna et al, 2006). My study focuses only on a collapsed dataset and primarily controls for user county demographics. Future studies could adopt a model that directly compares every county pair, once cleaned for duplicates, to work with a larger and more representative sample size. Understanding how unique pairs work together and pairing that with time series data could shed light on unique findings as the pandemic continues to unfold.

7 Citations

Anagnostopoulos Aris, Ravi Kumar and Mohammad Mahdian. “Influence and Correlation in Social Networks.” *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2008, pp. 7–15, doi:10.1145/1401890.1401897.

Bailey Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. “Social Connectedness: Measurement, Determinants, and Effects.” *Journal of Economic Perspectives* 32, no. 3 (August 2018): 259–80 and the Facebook Data for Good Program, Social Connectedness Index (SCI). https://dataforgood.fb.com/, Accessed 10th April 2021." https://doi.org/10.1257/jep.32.3.259.

Bailey Michael, Patrick Farrell, Theresa Kuchler, Johannes Stroebel, “Social connectedness in urban areas.” *Journal of Urban Economics*, Volume 118 (2020), 2020,103264, ISSN 0094-1190, https://doi.org/10.1016/j.jue.2020.103264.

Bengtsson Linus, Jean Gaudart, Xin Lu, Sandra Moore, Erik Wetter, Kankoe Sallah, Stanislas Rebaudet and Renaud Piarroux. “Using Mobile Phone Data to Predict the Spatial Spread of Cholera.” *Scientific Reports*, vol. 5, Mar. 2015, p. 8923, doi:10.1038/srep08923.

Bureau, US Census. “County Population by Characteristics: 2010-2019.” *The United States Census Bureau*, US Census Bureau, 20 Apr. 2021, www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html.

Bureau, US Census. “SAIPE State and County Estimates for 2019” *The United States Census Bureau*, US Census Bureau, 8 Dec. 2020, https://www.census.gov/data/datasets/2019/demo/saipe/2019-state-and-county.html

Bureau, US Census. “Urban and Rural” *The United States Census Bureau*, US Census Bureau, 26 April. 2021, https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html

Bursztyn, Leonardo, Aakaash Rao, Christopher P. Roth and David H. Yanagizawa-Drott. *Misinformation during a pandemic*. No. w27417. National Bureau of Economic Research, 2020.

Centola Damon. “Social Media and the Science of Health Behavior.” *Circulation*, vol. 127, no. 21, May 2013, pp. 2135–44, doi:10.1161/CIRCULATIONAHA.112.101816.

Chakraborty Indranil and Prasenjit Maity. “COVID-19 Outbreak: Migration, Effects on Society, Global Environment and Prevention.” *Science of The Total Environment*, vol. 728, Aug. 2020, p. 138882, doi:10.1016/j.scitotenv.2020.138882.

DellaVigna, Stefano, and Ethan Kaplan. *The Fox News Effect: Media Bias and Voting*. w12169, National Bureau of Economic Research, 2006, p. w12169, doi:10.3386/w12169.

Fritz, Cornelius, and Göran Kauermann. “On the Interplay of Regional Mobility, Social Connectedness, and the Spread of COVID-19 in Germany.” *ArXiv:2008.03013 [Physics, Stat]*, February 17, 2021. https://doi.org/arXiv:2008.03013v2.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski and Larry Brilliant. “Detecting influenza epidemics using search engine query data.” *Nature* vol. 457,7232 (2009): 1012-4. doi:10.1038/nature07634.

Gough, Aisling, Ruth F Hundter, Oluwaseun Jurek, Anna Jurek, Gary McKeown, Jon Hong, Eimear Barrett, Marbeth Ferguson, Gerry McElwee, Miriam McCarthy and Frank Kee. “Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages.” *JMIR Public Health and Surveillance*, vol. 3, no. 1, Mar. 2017, doi:10.2196/publichealth.6313.

Grossman, Guy, Soojong Kim, Jonah M. Rexer, and Harsha Thirumurthy. “Political Partisanship Influences Behavioral Responses to Governors’ Recommendations for COVID-19 Prevention in the United States.” *Proceedings of the National Academy of Sciences*, vol. 117, no. 39, Sept. 2020, pp. 24144–53, doi:10.1073/pnas.2007835117.

Haldar, Anasuya, and Narayan Sethi. "The effect of country-level factors and government intervention on the incidence of COVID-19." *Asian Economics Letters* 1.2 (2020): 17804. https://doi.org/10.46557/001c.17804

Kim A. David, Alison R Hwong, Derek Stafford, D Alex Hughes, A James O’Malley, James H Fowler, Nicholas A Christakis, Social network targeting to maximize population behaviour change: a cluster randomised controlled trial, The Lancet, Volume 386, Issue 9989, 2015, Pages 145-153, ISSN 0140-6736, https://doi.org/10.1016/S0140-67361560095-2.(https://www.sciencedirect.com/science/article/pii/S0140673615600952)

Klovdahl, A. S., J.J Potterat, D.E Woodhouse, J.B Muth, S.Q Muth and W.W Darrow "Social Networks and Infectious Disease: The Colorado Springs Study." *Social Science & Medicine (1982)*, vol. 38, no. 1, Jan. 1994, pp. 79–88, doi:10.1016/0277-95369490302-6.

Kuchler, Theresa, et al. *The Geographic Spread of COVID-19 Correlates with the Structure of Social Networks as Measured by Facebook*. w26990, National Bureau of Economic Research, 2020, p. w26990, doi:10.3386/w26990.

McGovern, Tony. *Tonmcg/US_County_Level_Election_Results_08-20*. 2016. 2021, https://github.com/tonmcg/US_County_Level_Election_Results_08-20.

Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikalojczyk, Marco Massari, Stefania Salmaso, Gianpoaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinksa, and John W. Edmunds. "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases." *PLOS Medicine*, vol. 5, no. 3, Mar. 2008, p. e74, doi:10.1371/journal.pmed.0050074.

Nicola Maria, Ziad Alsafi, Catrin Sohrabi, Amhed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha and Riaz Agha. "The Socio-Economic Implications of the Coronavirus Pandemic (COVID-19): A Review." *International Journal of Surgery*, vol. 78, June 2020, pp. 185–93, doi:10.1016/j.ijsu.2020.04.018.

New York Times./"Covid-19-Data". *GitHub*, https://github.com/nytimes/covid-19-data.Accessed 6 April 2021

Peixoto PS, Marcondes D, Peixoto C, Oliva SM (2020) Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. PLOS ONE 15(7): e0235732. https://doi.org/10.1371/journal.pone.0235732

So Mike K.P., Agnes Tiwari, Amanda M.Y. Chu, Jenny T.Y. Tsang, Jacky N.L. Chan, Visualizing COVID-19 pandemic risk through network connectedness, International Journal of Infectious Diseases, Volume 96, 2020, Pages 558-561, ISSN 1201-9712, https://doi.org/10.1016/j.ijid.2020.05.011.

Bureau of Labor Statistics"Unemployment." *USDA: Economic Research Service: U.S Department of Agriculture*, https://data.ers.usda.gov/reports.aspx?ID=17828. Accessed 6 April 2021

"See How Age and Illnesses Change the Risk of Dying from Covid-19." *The Economist*, 11 Mar. 2021, https://www.economist.com/graphic-detail/covid-pandemic-mortality-risk-estimator.

van Holm, Eric Joseph, Jake Monaghan, Dan C. Shahar, J. P Messina and Chris Surprenant. *The Impact of Political Ideology on Concern and Behavior During COVID-19*. SSRN Scholarly Paper, ID 3573224, Social Science Research Network, 11 Apr. 2020, https://papers.ssrn.com/abstract=3573224.

Vaughan, Adam. "Origins of Covid-19." *New Scientist (1971)*, vol. 249, no. 3318, Jan. 2021, p. 7, doi:10.1016/S0262-40792100079-8.

Wesolowski, Amy, Taimur Qureshi, Maciej F. Boni, Pål Roe Sundsøy, Michael A. Johansson, Syed Basit Rasheed, Kenth Engø-Monsen and Caroline O. Buckee. "Impact of Human Mobility on the Emergence of Dengue Epidemics in Pakistan." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 38, Sept. 2015, pp. 11887–92, doi:10.1073/pnas.1504964112.

Wesolowski, Amy, Nathan Eagle, Andrew J. Tatem, David L. Smith, Abdisalan M. Noor, Robert W. Snow and Caroline O Buckee. "Quantifying the Impact of Human Mobility on Malaria." *Science (New York, N.Y.)*, vol. 338, no. 6104, Oct. 2012, pp. 267–70, doi:10.1126/science.1223467.

Zimmer, Christoph, Sequoia I. Leuba, Reza Yaesoubi and Ted Cohen. "Use of daily Internet search query data improves real-time projections of influenza epidemics." *Journal of the Royal Society, Interface* vol. 15,147 20180220. 10 Oct. 2018, doi:10.1098/rsif.2018.0220 https://www.researchgate.net/publication/24091984_The_Fox_News_Effect_Media _Bias_and_Voting