

Networked Systems as Witnesses

Association Between Content Demand, Human Mobility and an Infection Spread

Sana Asif
Northwestern University

Fabián E. Bustamante
Northwestern University

Byungjin Jun
Northwestern University

John P. Rula
Akamai

ABSTRACT

While non-pharmaceutical interventions (NPIs) such as stay-at-home, shelter-in-place, and school closures are considered the most effective ways to limit the spread of infectious diseases, their use is generally controversial given the political, ethical, and socioeconomic issues they raise. Part of the challenge is the non-obvious link between the level of compliance with such measures and their effectiveness.

In this paper, we argue that users' demand on networked services can serve as a proxy for the social distancing behavior of communities, offering a new approach to evaluate these measures' effectiveness. We leverage the vantage point of one of the largest worldwide CDNs together with public-available datasets of mobile users' behavior, to examine the relationship between changes in user demand on the CDN and different interventions including stay-at-home/shelter-in-place, mask mandates, and school closures. As networked systems become integral parts of our everyday lives, they can act as witnesses of our individual and collective actions. Our study illustrates the potential value of this new role.

CCS CONCEPTS

• Networks → Public Internet; Network measurement;

KEYWORDS

COVID-19, pandemic, CDN, user demand, user mobility, Internet measurement

ACM Reference Format:

Sana Asif, Byungjin Jun, Fabián E. Bustamante, and John P. Rula. 2021. Networked Systems as Witnesses: Association Between Content Demand, Human Mobility and an Infection Spread. In *ACM Internet Measurement Conference (IMC '21), November 2–4, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3487552.3487831>

1 INTRODUCTION

The COVID-19 pandemic caused by infection of SARS-CoV-2 has radically altered daily life. Starting in the early weeks of 2020 the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IMC '21, November 2–4, 2021, Virtual Event, USA

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9129-0/21/11...\$15.00

<https://doi.org/10.1145/3487552.3487831>

virus rapidly spread around the world, triggering a number of large-scale interventions to control this, starting with the full lockdown of Wuhan, China, where the virus was first reported [15]. Countries around the world quickly followed with measures that range from school closures to large-scale lock-downs [28].

In the US, the first confirmed case was reported on January 20, 2020, in Snohomish County, Washington State. By March 23rd, every state had reported cases of the virus [21]. By early April of 2020, several state, city, and county governments had started to adopt similar social-distancing interventions, including shelter-at-home and temporarily closing of non-essential businesses and schools, although with variable levels of enforcement [53].

Social distancing measures aim, through a variety of means, to decrease or interrupt virus transmission by minimizing contact between individuals or population groups [31]. Non-pharmaceutical interventions such as these are considered the most effective way to limit the spread of infectious diseases like COVID-19 [45, 60, 63]. Indeed, even when or if vaccines become globally available - and only 32.7% of the world population has been vaccinated as of September 2021 [40] - recent studies suggest this alone will not be sufficient to end the pandemic [16].

Despite their benefits, the use of NPIs is generally controversial given the political, ethical, and socioeconomic issues they raise [6]. Part of the challenge is the non-obvious link between the level of compliance with such measures and their effectiveness [49, 68], particularly when the severity of the illness associated with the infection renders controlled-human infection studies controversial [22].

We argue that users' demand for networked services can serve as a proxy for the social distancing behavior of communities, offering a new approach to evaluate these measures' effectiveness.

The Internet is the fastest growing and more transformative technology of our times. In just over 50 years, Internet penetration has grown to surpass 60% of the globe, connecting nearly 5 billion people [34]. The networked systems we have built over it, from search engines to social networks and content distribution networks (CDN), have become an integral part of our lives and witnesses to our individual and community actions. Our work is motivated by a simple observation – as people comply with social distancing measures, one would expect to see a significant change in networked systems' demand as people attend to their education, work, and entertainment remotely.

In this paper, we leverage the vantage point of one of the largest worldwide CDNs to study the relationship between user demand on networked systems, social distancing, school closures, and the

rate of COVID-19 cases, and to extend a prior study on the positive impact of mask mandates.

We focus our study in several counties¹ in the US, where the highly variable mitigation response and its enforcement, and the varying intensity of the outbreak, challenges the analysis of these responses' effectiveness.

We first use Google's Community Mobility Reports (CMR) [32] together with logs of resource demand on the CDN (*CDN demand*), to characterize the relationship between user mobility across different counties in the US and user demand on the CDN. Google CMR provides the percentage change in activity at each of six location categories (such as retail and recreation, workplaces, and residential) compared to that of a baseline days before the outbreak. We use this to define a metric of mobility as a proxy for social distancing, on the basis of the assumption that when individuals make fewer trips, they physically interact less. This usage aligns with prior work on epidemiology and public health (e.g., [2, 4, 11, 43, 59]). We demonstrate that changes in CDN demand are positively correlated with this metric. For 20 of the top US counties based on population density and Internet penetration, the average distance correlation is 0.54 (max = 0.74).

We then employ CDN demand to examine the relationship between social distancing and the growth rate of COVID-19 infections in different counties in the US. Following Badr et al. [4], we focus on the 25 counties with the highest number of reported cases by April 2020. Our analysis shows that CDN demand is strongly correlated with lower COVID-19 case growth rates for the most affected counties in the US, with distance correlations over 0.65 for 20 out of the 25 counties (avg. 0.71).

Some social distancing NPI, such as shelter-at-home and closing of non-essential businesses, results in changes to intra-region or intra-county mobility which are well captured by the Google CMR and its categories. University campus closures, another commonly adopted NPI, should have a very different relationship to demand. As campuses close and people relocate *outside* the region, one would expect to see a correlated drop in CDN demand. To understand the applicability of CDN demand as a proxy of social distancing for interventions that result in inter-region mobility, we use university campus closures in the 20 largest *college towns* in the US [29]. Our analysis shows a strong correlation (0.71) between decreasing CDN demand, as students move away from school, and the incidence of COVID-19 cases in the respective counties, providing further evidence of the effectiveness of this mitigation strategy.

Besides social distancing, one of the most popular if contentious NPIs is mask-wearing [1, 47]. A key factor fueling the debate has been the relatively limited direct evidence on how much widespread community use would affect the spread of COVID-19 [30]. In the HHS and CDC weekly report of November 27, 2020, Van Dyke et al. [67] provided some of the needed evidence with a study comparing trends in county-level incidence of COVID-19 in counties with and without a mask mandate in the state of Kansas. Their analysis shows that the increasing trends of infection reverse in mandated counties, after the mandate became effective on July 3, 2020.

¹A county in the US is an administrative or political subdivision of a state that consists of a geographic region with specific boundaries and usually some level of governmental authority [10].

While the study provides evidence that mask-wearing decreases the spread of COVID-19, the authors point to the absence of social distancing information, a confounding factor, as one of the main limitations of their work. We use CDN demand as a proxy of social distancing to address this limitation, extending their study to explore the combined effects of social distancing and mask mandates. Our analysis shows the clear benefits of the combined interventions. The combination of mandated use of masks and increased social distancing resulted in a clear decrease in COVID-19 incidence (incidence case trend with a slope of -0.71) for these counties, compared with counties with only mask mandates (slope of 0.05) and those counties with neither mask mandates nor social distancing (slope of 0.19).

Our findings are subject to several limitations (§8), including the following. First, while we control for known confounding factors using natural experiment designs [8, 20, 23, 44], there may be additional confounding factors for which we have not accounted. Second, the data we rely on does not include sociodemographic information and it might not representatively capture all groups, such as low-income groups for whom Internet access may be hard to afford. Despite this, the consistency of our findings with those of prior work (e.g., [4, 67]) increases confidence in our approach. Third, our analysis is limited to counties in the US and relies on the perspective of a single CDN. The US presents an interesting and challenging case study. The varied response to the pandemic across the country, combined with different progression and intensity of outbreaks, complicates the already challenging task of evaluating the effectiveness of different NPIs [67]. In total, our study focuses on 163 counties across 21 states. The CDN we rely on is one of the largest in the world, operating more than 240,000 servers in over 1,700 networks, serving nearly 3 trillion HTTP requests daily. In addition, our analysis is descriptive looking at the association between mobility, CDN demand, and COVID-19 transmission across different datasets from different sources. Deriving statistical models that could be used for prediction is left as future work [4].

1.1 Contributions

This work makes the following contributions:

- We demonstrate the strong relationship between CDN demand and user mobility as a proxy of social distancing (§4).
- Relying on CDN demand as a new proxy of social distancing, we show the strong correlation between demand and the growth rate of COVID-19 cases in a selected set of counties in the US (§5).
- We investigate the effectiveness of school closures as an NPI by exploring the relation between inter-region mobility, CDN demand, and the incidence of COVID-19 infections before and after university campus closures (§6).
- Finally, we extend prior work studying the effectiveness of mask mandates using CDN demand to show the combined values of social distancing and mask-mandates (§7).

1.2 Ethical Considerations

This work does not raise any ethical issues.

2 BACKGROUND AND MOTIVATION

Non-pharmaceutical interventions have been proven effective in reducing the spread of infectious diseases including COVID-19 [45, 60, 63]. NPIs such as physical distancing, school closures, stay-at-home or shelter-at-home orders, travel restrictions, and the use of masks in public spaces, have been put in place around the world to reduce the transmission of the virus. As new and potentially more infectious strains of the virus continue to appear, many countries have reinstated these large-scale interventions or were considering doing so by May 2021, well over a year since the beginning of this pandemic [26, 39].

Even if vaccines were to become widely available, there is a growing consensus that these social distancing interventions as well as the mandated use of masks should remain in place [5]. While multiple effective COVID-19 vaccines have been under production, vaccination varies widely around the world with only 2% of the population in low-income countries having received at least one vaccine dose compared with about two-thirds (65%) in high-income countries as of September 2021 [49, 51, 57]. As the virus continues to circulate in these countries, new variants will emerge, threatening the world and slowing global economic recovery.

Social distancing measures, while generally considered one of the most effective public health measures to limit the spread of infectious diseases, remain controversial partially due to the non-obvious link between the level of compliance with these measures and their effectiveness [6]. Webster et al. [68] argues that adherence to these measures is mainly influenced by the knowledge people had about the disease and the measure, social norms, perceived benefits of the intervention, and perceived risk of the disease, among other factors. They are effective if adherence to them results in large-scale behavioral changes that reduce the close contacts and mobility patterns that facilitate the spread of the disease, but quantifying these effects, particularly on the spatial, temporal, and population scales to inform public health responses, is challenging [33, 43].

Prior work [4, 15, 42, 54] has proposed the use of mobility as captured by anonymized mobile users data to generate social distancing metrics and relies on this, together with epidemiological data, to evaluate how social distancing affects the rate of new infections. The use of mobile phone data for evaluating COVID-19 responses has raised issues with the potential sources and implications of selection bias in mobile phone data [33]. The aggregated view of networked system demand offers a promising alternative.

Our work is motivated by a simple observation – as people comply with social distancing measures, one would expect to see a significant change in networked systems' demand as parents and children work, study, and entertained themselves remotely. We explore the potential use of CDN demand as a proxy of social distancing, offering an alternative approach to evaluating the effectiveness of certain NPIs, and rely on natural experiment designs to control for known confounding factors [8, 18, 20, 44].

3 DATASETS

In this section, we describe the three main datasets we rely on for this study: COVID-19 infection data from the Johns Hopkins University CSSE real-time dashboard [21], mobility data from Google Community Mobility Reports [32], and aggregated data on requests issued or demand for resources hosted by a CDN. Additional datasets are described as part of later sections.

3.1 JHU CSSE Dashboard

Since January 22, 2020, the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University has been hosting an interactive web-based dashboard and associated GitHub repository with the location and number of confirmed COVID-19 cases, deaths, and recoveries for all affected countries. The dashboard reports cases at the province level in China, county level in the US, state level in Australia and Canada, and at the country level otherwise.

Initially, the data was provided by DXY, an online platform run by members of the Chinese medical community, and as the diseases spread around the world, the data sources have correspondingly expanded to include 10 aggregated data sources and nearly 200 at the country, region level and, for the US, at the state or county/city level, including county and state health departments and multiple national government health departments.

We gathered the data for all counties in the US from the CSSE COVID-19 repository [41] for our analysis. Throughout the paper, we use different months of 2020 for different parts of the study, as needed to address the question at hand. For the analysis on demand and infection cases (§5) we use April and May 2020, for the study on mask mandates that extends Van Dyke et al. [67] (§7) we use June and July 2020, and for the analysis of campus closures (§6) we use data for November 2020.

3.2 Google Community Mobility Reports

Also in response to the COVID-19 outbreak, Google started to release aggregated data collated from users accessing their applications from mobile and handheld devices who allow recording of 'location history'. These *Community Mobility Reports* (CMR) show changes in activity and mobility at different location types, compared to before the spread of COVID-19 [32].

To generate the reports, individual user presence and time spent at specific location categories are collated to indicate activity. The data is categorized into six discrete classes, which can be summarized as *retail and recreation*, *parks*, *groceries and pharmacies*, *workplaces*, *transport transit hubs* and *residential* areas. CMR provides the percentage change in activity at each location category compared to that on baseline days before the outbreak (January 3, 2020, to February 6, 2020). Baseline day figures are calculated for each day of the week for each country and are calculated as the median value. Daily activity changes are compared to the corresponding baseline figure day (i.e., data on a Monday is compared with a baseline Monday) and thus represent the relative change in percentages compared to baseline days, not the absolute number of visitors. Missing values were returned if the activity was too low upon a specific day and thus failed to achieve the anonymity

threshold set by Google [25]. Increases in *residential* indicate decreased mobility, while the other categories are more indicative of general mobility as they are related to activity around workplaces, retail outlets, and the use of public transportation.

We use Google CMR to define a metric of mobility patterns which we use as a proxy for social distancing. This is done on the basis of the assumption that when individuals make fewer trips, they physically interact less. This usage aligns with prior work on epidemiology and public health (e.g., [2, 4, 11, 43]).

3.3 CDN Dataset

Our third dataset lets us observe changes in networked systems' demand during the period of analysis. For this, we leverage the vantage point of one of the largest worldwide content delivery networks, which receives trillions of requests per day, and use their resource demand aggregated by the client's AS number and location. The dataset combines the view from 17,878 autonomous systems across 3,026 counties in all 50 states in the US. We utilize the request logs of the CDN from January 1, 2020, to December 31, 2020, as hourly request counts (e.g. hits) of all combined CDN traffic. These logs accumulate all requests received across the CDN's entire platform, covering all types of protocols and devices. To generate this, all daily request statistics are aggregated by /24 subnets for IPv4 and /48 subnets for IPv6. These requests are normalized across the platform into unit-less Demand Units (DU). Demand Units are normalized out of 100,000, with each DU representing 0.001% of global request demand (i.e. 1,000DU = 1%).

4 USER MOBILITY AND CDN DEMAND

We first explore the association between CDN demand and user mobility leveraging data on demand users impose on the CDN and the Google's Community Mobility Reports [32] of mobile users' behavior (§3).

Google's CMR captures the mobility of users of mobile and handheld devices. The data is presented as a percentage difference value from a baseline day before the pandemic outbreak (the median value of a 5 week period from January 3 - February 6, 2020) for the different categories of locations. Hence, the data is meant to show how the visitors mapping to categorized places change over time as compared to pre-pandemic. For instance, the end of March 2020 sees a drop of almost 50% in the number of people visiting workplaces, transit stations, and retail. Whereas, parks, and grocery stores see a drop of more than 10% [56].

In our analysis, we determine the average mobility across parks, transit, grocery, recreation, and workplace per day to quantify the number of people moving outside their house (residence). This in turn will allow us to gauge the degree of social distancing. Higher values of mobility would mean lesser social distancing and vice versa.

The formula for percentage difference of mobility is quoted below:

$$M_j^t = \frac{parks_j^t + transit_j^t + grocery_j^t + recreation_j^t + workplace_j^t}{5}$$

Where M_j^t is the percentage difference of mobility in county j on day t .

Our goal is to determine whether demand on a content delivery network (CDN) - CDN demand - can be used as a proxy for the social distancing behavior of communities.

We hypothesize that a decrease in user mobility from people staying at homes such as in the case of lockdowns and work-from-home, will result in an increase in demand. This may be because those users are more likely to rely on the Internet more, for the purposes of communication, or entertainment because of restrictions on other outdoor activities. Whereas, when user mobility increases and people get out of the house more, we expect to see a decrease in demand.

For this analysis, we normalize CDN demand by calculating the percentage difference of demand with respect to the same baseline period as Google's CMR reports, using in this case the median value of demand for a 5 week period between January 3 and February 6, 2020.

We employ *distance correlation* [61] to measure how well network demand witnesses human mobility and the spread of the pandemic. Distance correlation measures the dependency between two vectors, including both linear and non-linear association, and is obtained by dividing their distance covariance by the product of their distance standard deviations. Distance correlation has a number of advantages over classical Pearson correlation coefficient as it can detect nonlinear associations that are undetectable by Pearson correlation, it is applicable to random variables of any dimension, and it is zero if and only if the variables are independent. Given the non-linearity of the change in mobility and network demand, distance correlation is a better choice than classical Pearson correlation for this analysis.

We focused our analysis in 20 US counties during the months of April and May 2020. We concentrated on the top 20 counties with the highest population density and Internet penetration according to the US Census data [65]. To select this set, we started with the top 100 counties with highest density and the top 100 with the highest Internet penetration and selected those with highest population density if they are among the highest Internet penetration counties. We found, perhaps not surprisingly, a high overlap between the two sets.

Table 1 lists the correlation between percentage difference of mobility and percentage difference of CDN demand across these counties, in descending order. The results show that changes in CDN demand are positively correlated (moderate and high) with increasing social distancing with a median correlation of 0.56 and a maximum correlation of 0.74.

We generate graphs for 30-day periods for the months of April and May 2020 to visualize the trends of mobility and demand. Figure 1 highlights four counties, Fulton County in Georgia, Montgomery County in Pennsylvania, Fairfax County in Virginia, and Suffolk County in New York (in bold in Table 1). We inverted the mobility axis to better visualize the alignment of the trends.

Even in the case of Montgomery County, with 0.39 correlation, the alignment of both curves is apparent with demand and mobility following closely opposite trends. The timings of the peaks and

troughs in both the trends are similar. The complete set of graphs is included in Appendix A.1.

County	State	Correlations
Fulton	Georgia	0.74
Norfolk	Massachusetts	0.71
Bergen	New Jersey	0.7
Montgomery	Maryland	0.66
Fairfax	Virginia	0.61
Arlington	Virginia	0.59
Franklin	Ohio	0.58
Gwinnett	Georgia	0.58
Cobb	Georgia	0.57
Middlesex	Massachusetts	0.56
Delaware	Pennsylvania	0.54
Allegheny	Pennsylvania	0.53
Alameda	California	0.49
Macomb	Michigan	0.47
Suffolk	New York	0.43
Multnomah	Oregon	0.4
Hudson	New Jersey	0.4
Orange	California	0.39
Montgomery	Pennsylvania	0.39
Nassau	New York	0.38

Table 1: Correlations between Percentage difference of mobility and Percentage difference of CDN demand, arranged in descending order. The average correlation (StdDev) is 0.54 (0.1453). Figure 1 includes a graph for each of the highlighted counties; the full set of graphs can be found in the Appendix.

Limitations. Some of the limitations of this part of our study result from the datasets we rely on and possible confounders. Google CMR only captures mobility information from users who have access to smartphones and turn on their location settings, so the data is a sample of Google Maps users which may or not be representative of the wider population. The CDN demand dataset only captures users' requests for resources hosted by the particular CDN we rely on which, again, may not be a representative sample of the user population.

5 DEMAND AND INFECTION CASES

In the previous section, we established that CDN demand can be used as a proxy for social distancing. In this section we explore the relationship between CDN demand and COVID-19 reported cases. More specifically, we analyze whether there is a relationship between CDN demand and cases, to what degree, and the optimum lag of days after which the impact of people's behavior on the rate of confirmed COVID-19 cases is most apparent. We focus our analysis on the US, where the absence of centralized policy and highly variable outbreak mitigation responses challenges efforts at assessing the effectiveness of NPIs.

We model our analysis in this section after Badr et al. [4] where the authors explore the association of mobility patterns and COVID-19 infections from Jan 1 to April 20, 2020. They used epidemiological data to compute the COVID-19 growth rate ratio per county/day to evaluate how social distancing, measured by the relative change in mobility (calculated based on aggregated and anonymized cell phone data), affected the rate of new infections in counties in the US.

We examined the relationship of CDN demand to the rate of transmission using the same set of 25 US counties with the highest number of cases by April 16, 2020, according to the JHU CSSE dataset [21]. We note that five of these 25 counties are also among the 20 counties included in the previous analysis on the relationship between mobility and CDN demand (§4); the counties in both sets are Nassau, Middlesex, Suffolk, Bergen, and Hudson.²

As in Badr et al. [4], we use daily new cases to derive growth rate ratio (GR) [4] defined as the logarithmic rate of change (number of newly reported cases) over the previous 3 days relative to the logarithmic rate of change over the previous week. GR for any county j on day t is as follows:

$$GR_j^t = \frac{\log(\sum_{t-2}^t \frac{C_j^t}{3})}{\log(\sum_{t-6}^t \frac{C_j^t}{7})}$$

where C_j^t is the number of new cases in county j on a day t

GR is a non-negative value and is defined only when the average number of reported cases per day is greater than one over any period (3-day or 7-day moving averages). It takes a value below 1 if the growth rate during the last 3 days was lower than that of the last week; a value greater than 1 represents a growth rate increase in the last 3 days relative to the last week [66].

Determining Lag. As part of the analysis we determine the *lag* which is the period separating the beginning of social distancing and onset of case growth reduction [4]. We do this by finding out when demand affects cases the most using cross-correlation.

We conduct the cross-correlation analysis separately for each county because the lag value is dependent on a number of factors. These include the incubation period (the period between exposure to an infection and the appearance of the first symptoms) which can be between 2 to 14 days [37], the day the subject decides to get tested, and the number of days it takes for the test results to be generated. A rapid test gives results within the hour where a common PCR-test available at the time can take 72hrs and, depending on the backlogs, up to 7 days. Although during April and May 2020, the availability of testing across the US became high, states still varied substantially in their testing times and eligibility criteria for tests [36], thereby increasing the randomness associated with the return date of test results.

We further cater to the randomness associated with the lags by taking small windows of 15 days in the span of two months. This prevents us from using a single value of lag for the duration of two whole months. It should be noted here that we are not claiming that we were able to completely eradicate the issue of the randomness

²The overall distance correlation between mobility and CDN demand of these 25 counties is slightly lower than that of the 20 counties with highest population density and Internet penetration used in the previous section, ranging between 0.14 and 0.67.

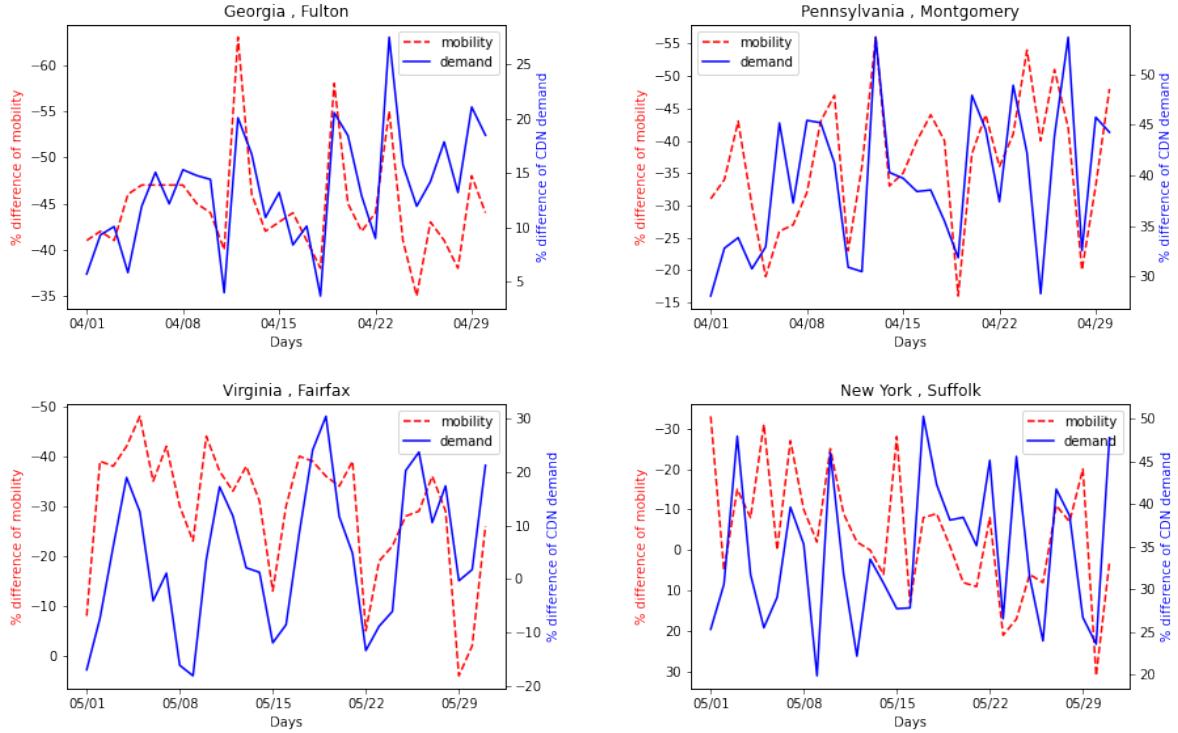


Figure 1: Four examples of the similarity between the trends of percentage difference of mobility and percentage difference of demand for the month of April 2020 - Fulton County (Georgia) and Montgomery (Pennsylvania), and May 2020 - Fairfax County (Virginia), and Suffolk County (New York). We use the best y-axis range that illustrates this similarity for each county.

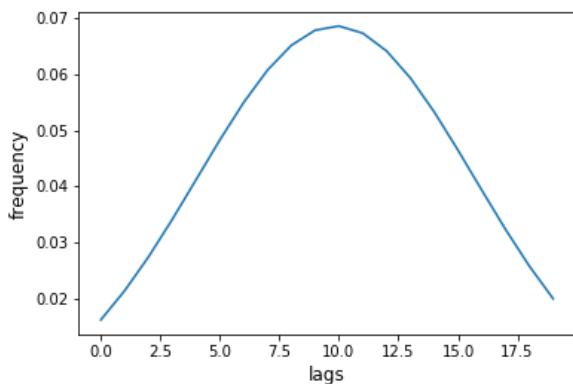


Figure 2: Distribution of lags. Mean=10.2 (Std Dev =5.6). Badr et al. [4] estimates a 11-day lag.

associated, with each subject's case on a particular day having likely different days when they were infected. Reducing the window size merely reduces the distribution of these lags, enabling us to observe a better correlation at the mean lag. We use a 15-day window of demand and growth rate ratio (GR) of cases, and cross correlate it to find the lag. Cross correlation allows us to shift the demand

trend back by days within the range of 0 and 20 and see which lag gives the best negative Pearson correlation. We use Pearson correlation for this purpose because it gives us both positive and negative values, and we want a lag that gives a negative correlation depicting opposing trends of GR and demand.

We determine four offsets for each of the 25 counties and plot the distribution of these offsets in Figure 2. The distribution of lags with an average of 10 days, is consistent with the incubation period of the SARS-CoV-2 virus, plus the additional time it would take to get the test results [58]. Furthermore, our mean offset rounds up to the same lag as used by other studies when they measure the correlation between mobility and cases [4]. A related study by Badr et al. [4] uses a lag of 11 when determining the correlation between mobility and GR of infection cases.

Demand and Infection. Using the aforementioned metrics and the determined lags, we evaluate if and how well demand affects the rate of new infections in these 25 counties in the US, calculating the distance correlations of GR and CDN demand. Table 2 lists these correlations across the counties, in descending order. The results show that changes in CDN demand are positively correlated (moderate and high) with increasing social distancing with correlation ranging from 0.58 to 0.83, with an average correlation of 0.71.

We generate lineplots that show the trends of GR and demand with the demand trend shifted during the observation period. Figures 3 highlights four counties, Wayne County in Michigan, Passaic

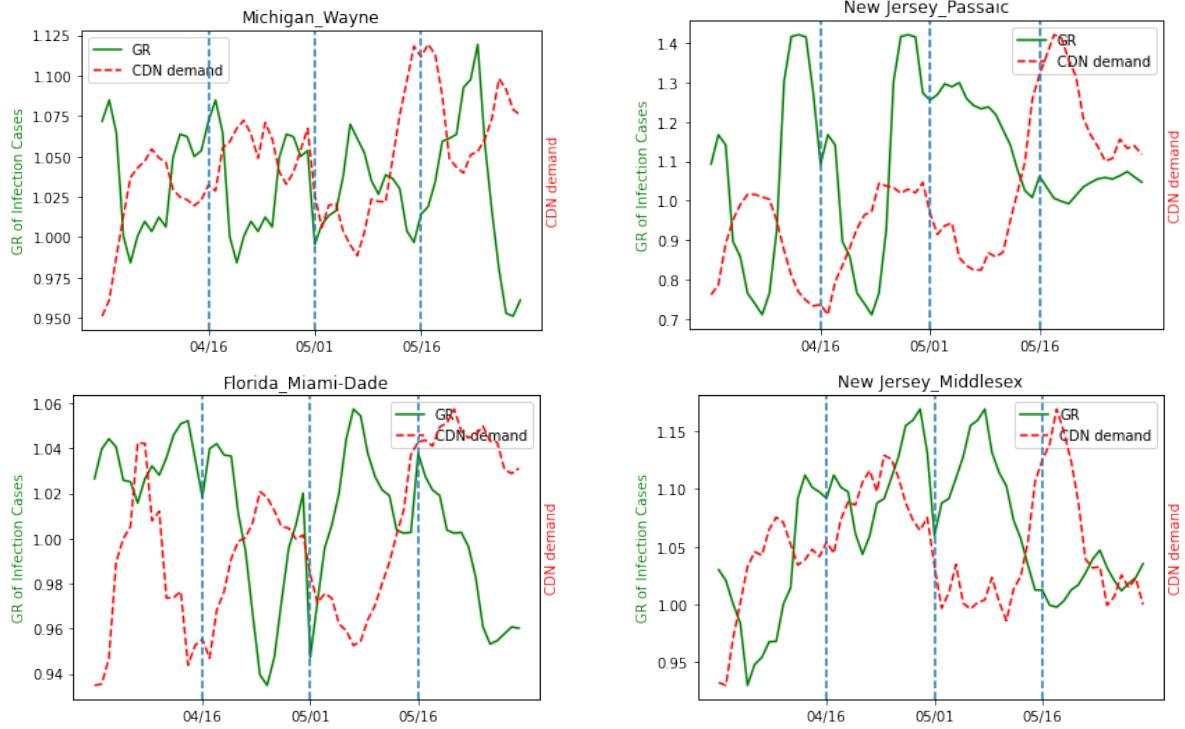


Figure 3: Four examples of the opposing trends of GR of Infection Cases and CDN demand for April and May 2020 for Wayne County (Michigan), Passaic County (New Jersey), Miami-Dade County (Florida) and Middlesex (New Jersey).

County in New Jersey, Miami-Dade County in Florida, and Middlesex County in New Jersey. We highlighted these counties in Table 2. Each plot has four 15 day windows with each window separated by the black dotted line. This is to show that the offset was determined for each window separately. Appendix A.2 includes the full set of graphs.

Limitations. This part of our study has limitations, some inherited from the original study [4]. First, as we focus on the relationship between CDN demand and case growth rate, our study does not account for the role of other potentially mitigating factors (e.g., mask wearing and handwashing) that could have contributed to the observed decline in cases. Second, changes in CDN demand data do not differentiate the reasons for such changes (e.g., connectivity problems, exercising or low-risk trips). Although these topics are outside the scope of our work, they are critical to an understanding of the risk of COVID-19 infections. Third, as Badr et al. [4] we use GR as a representative metric of the degree of transmission among local populations; future work should explore replacing this variable with other transmission indexes used in epidemiology. Fourth, although the analysis focused on 25 counties, which might be a biased sample of locations, we believe these locations accurately capture the relationship between CDN demand and case growth rates. The consistency of the correlations found at the state level (counties in the same state) increases confidence in our results.

6 UNIVERSITY CAMPUS CLOSURES

Some social distancing NPI, such as shelter-at-home and closing of non-essential businesses, results in changes to intra-region or intra-county mobility which are well captured by the Google CMR data and its categories (e.g., *retail*). We have shown that increased social distancing, as captured by this mobility data, is correlated with CDN demand (§4), and that CDN demand as a proxy of social distancing is related to growth rates of infection (§5).

Despite some prior work on the effectiveness of school closures in reducing the spread of viral respiratory illnesses (e.g., [3, 13]), the value of this NPI is still disputed. In this section, we use university campus closures in large *college towns* in the US³, to understand the applicability of CDN demand in evaluating the effectiveness of interventions that result in inter-region mobility.

The closure of university campuses results in a different type of mobility which, we expect will have a very different relationship with demand. As campuses close and people potentially relocate *outside* the region, one would expect to see a correlated drop in demand (rather than an increase, as when people stay home). For this analysis, we focus on the 20 biggest college towns [29] (and their counties), with metro areas between 100,000 and 1 million people, and under 100,000 people (we do not include towns with over 1 million people, as large metro includes multiple schools). The percentage of the population made up of students at these

³College towns are communities, often a separate town or city, where students make up a significant share of the population [35].

County	State	Average Correlation
Essex	New Jersey	0.83
Nassau	New York	0.83
Middlesex	Massachusetts	0.79
Suffolk	New York	0.78
Suffolk	Massachusetts	0.77
Cook	Illinois	0.75
Union	New Jersey	0.75
Bergen	New Jersey	0.75
New York	New York	0.72
Bronx	New York	0.72
Richmond	New York	0.7
Rockland	New York	0.7
Passaic	New Jersey	0.7
Wayne	Michigan	0.7
Hudson	New Jersey	0.7
Queens	New York	0.69
Fairfield	Connecticut	0.69
Los Angeles	California	0.67
Orange	New York	0.67
Miami-Dade	Florida	0.66
Philadelphia	Pennsylvania	0.64
Essex	Massachusetts	0.63
Kings	New York	0.62
Middlesex	New Jersey	0.59
Westchester	New York	0.58

Table 2: Correlations between lagged demand and growth rate ratio (GR), arranged in descending order. The average correlation (StdDev) is 0.71 (0.179). Figure 3 includes a graph for each of the highlighted counties; the full set of graphs can be found in the Appendix.

places ranges between 21.4% (Alachua, FL) and 71.8% (Clay, SD).⁴ We excluded Vincennes University (Vincennes, IN) from the list as we were not able to gather network data for its campus.

Following prior work on the analysis of the potential benefits of school closure we use daily COVID-19 incidence per 100,000 residents in each county [3]. We estimate COVID-19 incidence using publicly available data from the Johns Hopkins University CSSE COVID-19 repository [41]. The denominator for the incidence measure was the county population from the 2018 American Community Survey [65].

We focus on days before and after the second closing of university campuses, around the Thanksgiving holiday of November 26th, 2020. The first set of university campus closures happen very early in the pandemic with the University of Washington, the first university closing its campus in the US, announcing its move to online classes on March 6th, 2020 [62]. Given that it was only by the end of March that every state had reported COVID-19 infection cases and that, at that point, the maximum number of confirmed

⁴Although the article was published in September 2016, we have manually validated the reported figures. While population and enrollment numbers have changed, the fraction of the county population corresponding to the student body remains similar.

cases among the set of studied counties was only 8 [21], we do not expect to see a significant impact of these measures [14].

For each campus, we separate demand originated from networks belonging to the school from that of other networks, and related to the number of confirmed COVID-19 cases in the county. Figure 4 uses the campuses of University of Illinois, Urbana-Champaign (UIUC), Cornell University, the University of Michigan and Ohio University, as examples. Note that different schools report either the *End of Fall term* or *End of In-Person Classes* which we use to label these figures.

UIUC is a public land-grant research university in Illinois, located in the twin cities of Champaign and Urbana, with an enrollment of over 44,000 students, representing $\approx 22\%$ of the population for the county. Cornell University is a private institution in New York state with an enrollment of 24,000, representing $\approx 32\%$ of the Tompkins County population. The University of Michigan is another large public institution with 76,000 students, representing $\approx 21\%$ of the population of Washtenaw County. The Ohio University, located in Athens, Ohio, is a relatively small public university with a student enrollment of 25,000, but comprising a large fraction of the total county population, $\approx 38\%$.

As the figure shows, in all cases lagged school CDN demand is closely related to the drop in confirmed COVID-19 cases, suggesting the effectiveness of the measure at controlling the spread of the infection. In the cases of UIUC, cases drop rapidly from a peak right at the time of in-person class ending. At Tompkins County, the number of confirmed COVID-19 cases increases up to the end of the Fall term and drops, together with school network demand on the CDN after closing of campus. Appendix A.3 shows the complete set of plots for all 19 college-town campuses.

Table 3 shows the distance correlation between both school (network) and non-school lagged demands and confirmed cases of COVID-19 in the county (ordered by correlation with school demand). Our results show a strong correlation between campus closures, social distancing quantified by CDN demand, and reduction of COVID-19 incidence. Despite the rapidly increasing trend of new COVID-19 cases in the US, as a whole, during the November–December 2020 period [50], COVID-19 incidence in most college towns we analyzed show similar trends as that of UIUC.

There are three college campuses showing low correlation coefficients (<0.5) between school network demand and confirmed COVID-19 cases: University of Mississippi, Blinn College, and Mississippi State University. We notice that the counties where these universities are located show a sharp increase in confirmed cases before and during the closing of their respective campuses.

Limitations. This part of our study has some specific limitations. By comparing counties to themselves before and after the end of term, we removed many potential confounding factors that could skew the observed trends (e.g., timing of community mitigation measures, testing capabilities, infrastructure and strictness of testing guidelines). Still, there may be additional factors we have not taken into account. As discussed in the related analysis by Auger et al. [3], it is impossible to totally isolate the effects of this or any single non-pharmaceutical intervention as school closures were enacted in close proximity to other measures (e.g., closing of nonessential businesses and prohibiting large gatherings) and the dynamics of

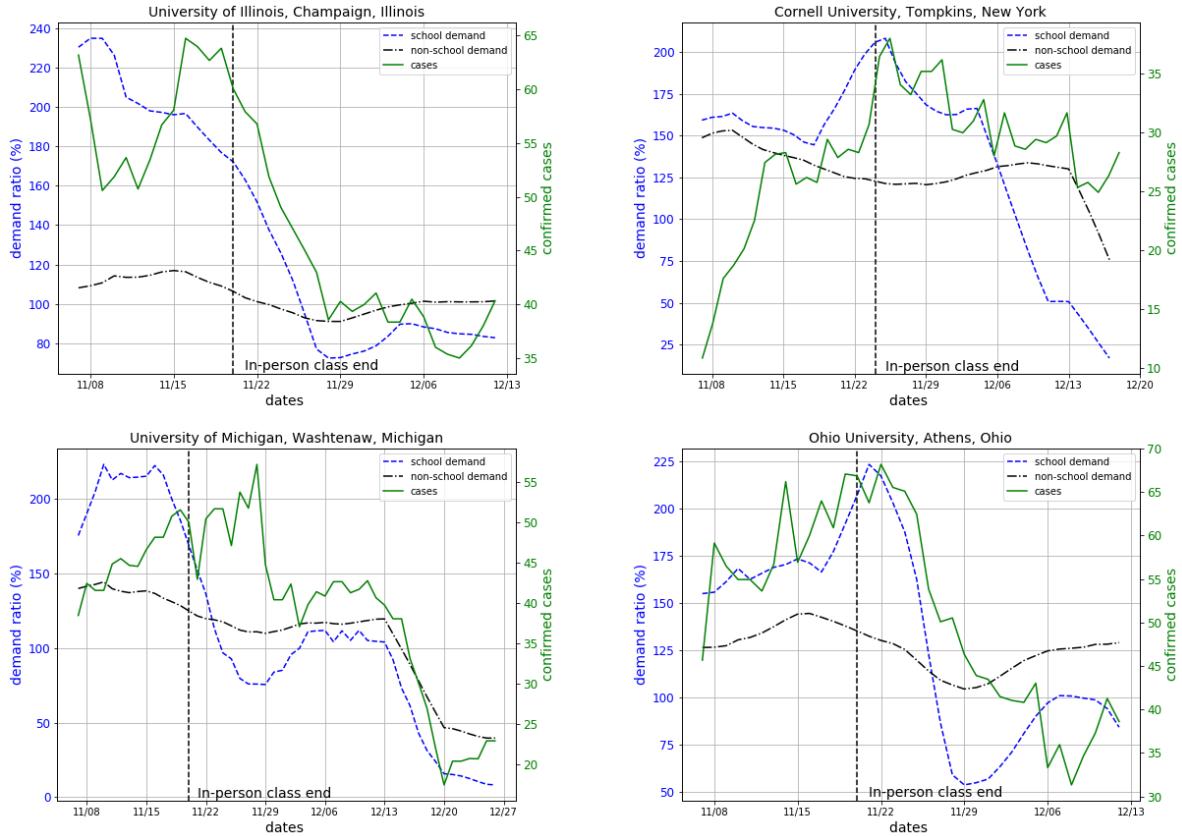


Figure 4: Four examples of school/non-school networks demands and confirmed cases at the University of Illinois, Urbana-Champaign (Champaign County), Cornell University, Ithaca (Tomkins County), University of Michigan (Washtenaw County) and Ohio University (Athens County). Beginning dates on x-ticks correspond to Mondays.

the pandemic vary across counties and even within counties in the US. We use COVID-19 incidence rate which is a function of the population size of the county, but do not adjust to account for people who may have left once the term ends or the campus closes. We are unable to determine the fraction of the population who has actually left the county, and the number of people with/without the virus infection among those who left the area is unclear.

7 MASK MANDATES AND DEMAND

In previous sections, we explored the relationship between CDN demand, social distancing and the rate of COVID-19 infection. Social distance measures, such as stay-at-home, school and non-essential businesses closures, are some of the available NPIs. Numerous studies have suggested face masks or coverings as another NPI effective at curbing the transmission of the virus (e.g., [1, 48, 67]), by preventing the spreading of respiratory droplets from when a person talks, sneezes or coughs, and the inhalation of these droplets by the wearer. Despite some of these studies showing a potential reduction in the infection rates across the US [48, 67] and Germany [52], several US states have lifted their mask mandates, notwithstanding the low vaccination rates among their population.

In the following paragraph we leverage CDN demand as a proxy of social distancing to extend an often-cited study by Van Dyke et al. on the positive impact of mask-wearing [67]. In a Morbidity and Mortality Weekly Report by the US Department of Health and Human Services (HHS) and the CDC [67], the authors investigate trends in COVID-19 incidence in Kansas before and after a governor's executive order requiring masks or other face coverings in public spaces came into effect. Among the main limitations of their study, the authors point to the absence of social distancing information, a clear confounding factor. We build on their work using CDN demand to address this limitation, and explore the combined effects of social distancing and mask mandates.

The governor of Kansas issued an executive order, effective as of July 3, 2020, requiring the wearing of masks in public spaces. A Kansas state law enacted just a month before (June 9, 2020) allows counties to issue less strict public health policies as compared to the ones issued by the state. Therefore, several counties chose to opt out or partially opt out of this mandate. Van Dyke et al. [67] uses the variations in mask mandate adoptions in the state of Kansas as the basis of a natural experiment [18] to evaluate the effectiveness of this NPI.

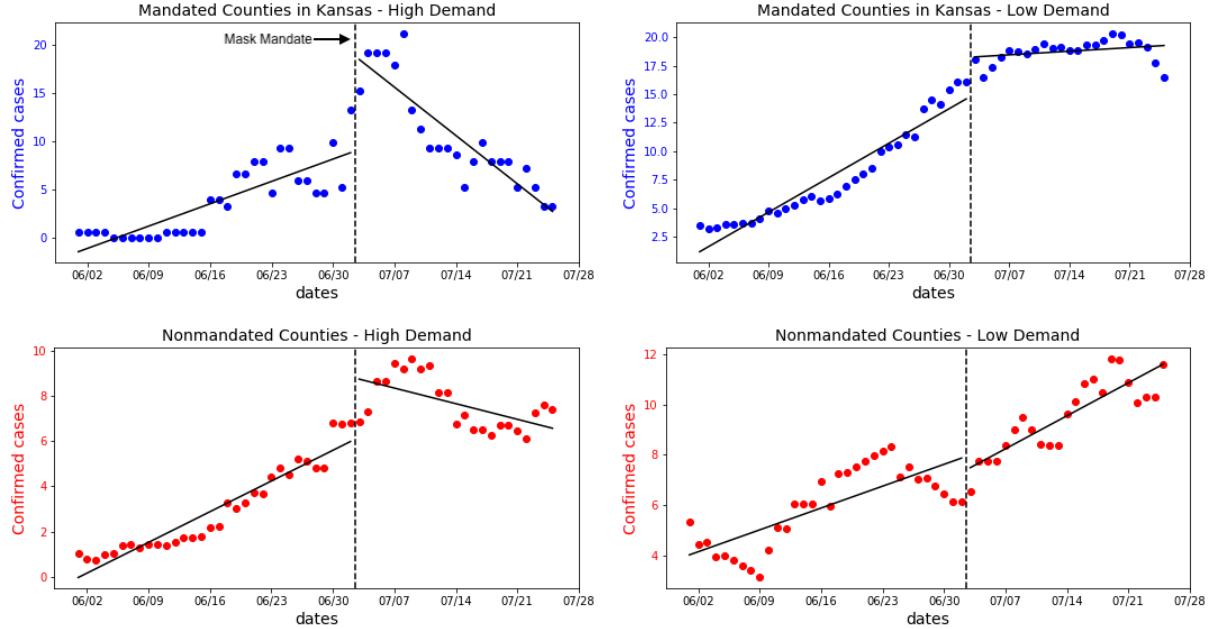


Figure 5: Trends in 7-day average of new COVID-19 cases per 100,000 population among mask-mandated/non-mask-mandated and high-demand/low-demand counties in Kansas. The plot includes cases before and after the governor's executive order requiring masks which took effect on July 3, 2020. Top-left figure shows mask-mandated counties with high-demand on the CDN while bottom-right figure shows non-mask-mandated counties with low demand on the CDN.

School Name	School	Non-school
University of Illinois	0.95	0.49
Indiana University	0.94	0.45
Texas A&M University-Kingsville	0.90	0.49
Ohio University	0.90	0.81
University of Michigan	0.88	0.94
South Plains College	0.88	0.8
Iowa State University	0.86	0.89
University of South Dakota	0.86	0.28
University of Missouri	0.82	0.71
Penn State	0.80	0.35
Virginia Tech	0.79	0.89
Cornell University	0.78	0.58
Washington State University	0.58	0.74
Texas A&M	0.56	0.66
University of Florida	0.55	0.62
University of Kansas	0.54	0.52
University of Mississippi	0.40	0.49
Blinn College	0.37	0.52
Mississippi State University	0.33	0.43

Table 3: Distance correlation coefficients between lagged demand (from school and non-school networks) and COVID-19 incidence (lag is the same for both networks). Schools are ordered by the correlation coefficient, in descending order.

By exploring a single state in the study, the authors expect to remove other potential variables that could skew the trends such as weather [64], racial, and economic inequality [38], the timing and extent of community mitigation measures, diagnostic testing capacity and public health reporting practices [17]. We compared the population density of mask mandated and nonmandated counties considering the importance of density in predicting cumulative COVID-19 cases in the US [69]. While both mask-mandated and nonmandated sets include high and low densely populated counties, most of the mask-mandated ones are among the top-30 most densely populated counties in the state (14 out of 24), with less than 20% of nonmandated counties making it to the list (16 out of 81). Moreover, previous works, e.g. [55], suggest that temperature has an effect on the prevalence of COVID-19 which makes it more suitable to compare two locations with similar average temperatures.

In our analysis, we compare the time periods of June 1 to July 3, 2020 (when the state mask mandate became effective), with that of July 4 to July 31, 2020. The counties are compared to themselves over time thus avoiding the impact of confounding variables on the study, e.g. income or population density. Following Van Dyke et al. [67], we divide the counties between those with and without mask mandates, that is counties that opted out of the state mandate and did not adopt a mask mandate of their own as of August 11, 2020. We further divide the counties in those with high CDN demand and those with low CDN demand. For this, we calculate the percentage difference of CDN demand from a baseline value in January; this yields a value in the range of [-200,200] for each day, with a negative (positive) value

Counties	Before Mandate	After Mandate
Mandated Counties in Kansas - High CDN demand	0.33	-0.71
Mandated Counties in Kansas- Low CDN demand	0.43	0.05
Nonmandated Counties in Kansas - High CDN demand	0.19	-0.1
Nonmandated Counties in Kansas- Low CDN demand	0.12	0.19

Table 4: The slopes of the trends of COVID-19 Incidence in mask-mandated and nonmandated counties in Kansas, with high and low CDN demand, before and after the state mask mandate became effective on July 3, 2020 [67].

meaning a decrease (increase) in CDN demand usage compared to the baseline period. We define high CDN demand as a positive demand value (in $(0 - 200]$) and low CDN demand as a no-positive (negative or zero) demand value (in $[-200, 0]$). We opted for this discretization of CDN demand to better match the study we extend. We obtained the data on county mask mandates from the Kansas Health Institute. As in the previous section and consistent with Van Dyke et al. [67], we use COVID-19 Incidence per county (daily county-level cases per 100,000), from the JHU CSSE COVID-19 repository [41], and apply a rolling 7-day average on it.

Figure 5 presents trends in 7-day average of COVID-19 per 100,000 population. The figure, modeled after Van Dyke et al. [67], gives a visual representation of the trend of cases before and after the mask mandate (July 3, 2020). July 3 is represented by the black dotted line in the figure. We use segmented regression to find changes in the trend of the pandemic before and after the mask mandate. The slopes of the regression lines are mentioned in Table 4. The figure shows a steep drop in the trend of COVID-19 incidence for the mask-mandated counties that have high demand, quantified by the slope of -0.71 obtained through regression. This compared to -0.1 for nonmandated counties with high demand shows the impact that masks have on reducing the spread of the pandemic. Counties with no mask mandates and low demand have the highest Covid Incidence slope of 0.19 for the month of July. The analysis suggests that, while mask mandates can help reduce the rate of infection (top-right figure), the combined interventions (top-left) are clearly effective at significantly reducing the spread. Besides further illustrating the potential value of CDN demand as a proxy of social distancing, the analysis may help explain, at least in part, some of the variability observed in the correlation between CDN demand and infections cases in the previous sections.

Limitations. We inherit some of the limitations of the study we extend. As Van Dyke et al. [67], the experimental design and the limited information on the implementation of the mandates and community mask-wearing behaviors limit the ability to determine the extent to which the mask mandates accounted for the observed declines in COVID-19 incidence. Second, although the design limits potential confounding from constant county-related characteristics, the findings are conditional on the absence of some time-varying factors, such as access to testing, within counties around the effective date of the mandate. Nonetheless, as Van Dyke et al. [67] we find that, while testing rates were observed to increase overall over time, these studies observed decreases in COVID-19 incidence in mandated and high CDN demand counties after July 3.

8 LIMITATIONS

We acknowledge several limitations of this study, to help put our findings in perspective.

- While we control for known confounding factors using natural experiment designs [8, 20, 23, 44], there may be additional confounding factors for which we have not accounted (e.g., party affiliation). Despite this, the consistency of our findings with those of prior work increases confidence in our approach.
- In addition, we would expect that, over time, the introduction of other pharmaceutical and non-pharmaceutical interventions, changes in the level of compliance with different measures, and the growing fraction of people with active immunity (vaccine-induced or through infection with the actual disease) will further complicate this and similar studies.
- Our analysis is limited to counties in the US. Beyond the challenges of analyzing such complex interactions across countries (with different political systems, health infrastructure, cultural norms, etc.), the distributed decision-making process resulted in a highly variable mitigation response which, combined with variable enforcement, challenges this type of analysis while, at the same time, offering the opportunity for natural experiments, as otherwise similar counties opted for diametrically different responses [67].
- The data we rely on does not include sociodemographic information and it might not representatively capture all groups, such as older adults, individuals with chronic diseases, or low-income groups for whom social distancing might be difficult to maintain or Internet access hard to afford.
- In addition, the completeness and accuracy of the JHU database with respect to COVID-19 incidence has not been established. This database aggregates publicly available data whose accuracy may vary from state to state. As with limitations in testing, inconsistencies in reporting are unavoidable limitations of all COVID-19 US population-based studies.
- Finally our analysis is descriptive looking at the association between mobility, CDN demand, and COVID-19 transmission. Deriving statistical models that could be used for prediction is left as future work [4].

9 RELATED WORK

In this section, we examine prior work on COVID-19 with a focus on the interaction between human behavior during the pandemic and the networked systems they use. More specifically, we review prior efforts from the networking community exploring changes in user mobility and networks' and systems' demand during the pandemic.

We also review work from the health sciences that leverages mobile user data to evaluate the effectiveness of different NPIs.

Previous studies have looked at how changes in people's social behavior due to the pandemic have also changed their network usage patterns. The pandemic has lead to unprecedented changes in social interaction and increased pressure on the Internet. Lutu et al. [46] describe how different aspects of network usage, including cellular data volume and average radio load, have changed since the pandemic started. Feldmann et al. [27] present an analysis of the impact of the COVID-19 pandemic from a large and diverse set of vantage points. Boettger et al. [9] explore this effect from the vantage point of a global social network.

Several efforts, particularly in epidemiology and public health, have relied on aggregated and anonymized cell phone data to estimate the degree of social distancing [2, 4, 43] and the effectiveness of related interventions [11, 59]. Asweto et al. [2] use Google CMR data to estimate the level of transmission reduction in 26 African countries as a result of different NPIs aimed at containing the disease, finding that both mobility changes in public spaces and residential are significant predictors of COVID-19 cases. Kephart et al. [43] use aggregated mobile phone location data, confirmed COVID-19 cases, and features of urban and social environments to analyze population mobility and COVID-19 incidence at the sub-city level among 314 cities in Argentina, Brazil, Colombia, Guatemala, and Mexico. Among their findings, the authors report that a 10% lower weekly mobility was associated with 8.6% lower incidence of COVID-19 in the following week. Badr et al. [4] rely on anonymized mobility data to study the correlation between the growth rate of infection and mobility, finding that mobility patterns are strongly correlated with decreased COVID-19 case growth rates for the most affected counties in the USA, with Pearson correlation coefficients above 0.7 for 20 of the 25 counties evaluated. Two closely related studies show that public interest in COVID-19, as captured by relative search popularity in Google Trends, Instagram, and Twitter, can be used in the assessment of social distancing measures [70] and correlates with the number of newly reported COVID-19 cases [24].

Others have suggested the idea of networked systems as witnesses of large-scale global events. Cardona et al. [12] describe the correlation between traffic demand and weather conditions. Bischof et al. [7] explore how the popularity of a peer-to-peer system like BitTorrent is impacted by events like natural disasters (2011 earthquake and tsunami in Tohoku, Japan). Dainotti et al. [19] use BGP routing data, traceroute measurements, and Internet Background Radiation (IBR) traffic to understand how large-scale censorship events took place.

10 CONCLUSION

We argued that the pervasiveness of networked systems in daily life gives them a unique perspective on our individual and collective actions. We leveraged this perspective by studying the relationships between user mobility, content demand and the spread of COVID-19. We used public-available datasets of user mobility to demonstrate that changes in CDN demand are strongly correlated with social distancing behavior. We then built on this, using CDN demand as a proxy of social distancing, to show that increased social distancing is strongly correlated with decreasing growth rates of infection.

We used university campus closures to explore the relationship between inter-region mobility, CDN demand, and the incidence of COVID-19 infections. Last, we extended prior work studying the effectiveness of mask mandates and show the combined benefits of social distancing and mask mandates. Demonstrating the effectiveness of non-pharmaceutical interventions requires a better understanding of compliance with such measures. The use of mobile phone data for evaluating COVID-19 responses raises issues with the potential sources and implications of selection bias in mobile phone data [33]. The aggregated view of networked system demand offers a promising alternative.

ACKNOWLEDGMENTS

The authors thank the anonymous shepherds and reviewers for their detailed and helpful feedback. We are also grateful to Walter Willinger for fruitful discussions in the early stages of this work. This work was supported in part by the U.S. National Science Foundation under Award CNS-2027922. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agencies.

REFERENCES

- [1] Dhaval Adjodah, Karthik Dinakar, Matteo Chinazzi, Samuel P. Fraiberger, Alex Pentland, Samantha Bates, Kyle Staller, Alessandro Vespignani, and Deepak L. Bhatt. 2021. Association between COVID-19 outcomes ad mask mandates, adherence, and attitudes. *PLoS ONE* 16, 6 (June 2021).
- [2] Collins Otieno Asweto, Patrick Ogala Onyango, Mohamed Ali Alzain, and Wei Wang. 2020. Effects of increased residential mobility and reduced public spaces mobility in containing COVID-19 in Africa. *Journal of Global Health Reports* 4 (August 2020).
- [3] Katherine A. Auger, Smair S. Shah, and Troy Richardson. 2020. Association between statewide school closure and Covid-19 incidence and mortality in the US. *Journal of the American Medical Association (JAMA)* 324, 9 (2020), 859–870.
- [4] Hamada Badr, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta Squire, and Lauren Gardner. 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* 20, 11 (2020), 1247–1254.
- [5] Rachel E. Baker, Sang Woo Park, Wenchang Yang, Gabriel A. Vecchia, C. Jessica E. Metcalf, and Bryan T. Grenfell. 2020. The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections. *Proc. of the National Academy of Sciences (PNAS)* 117, 48 (2020), 30547–30553.
- [6] Cecile M. Bensimon and Ross E.G. Upshur. 2007. Evidence and Effectiveness in Decisionmaking for Quarantine. *American Journal of Public Health* 97 (April 2007).
- [7] Zachary Bischof, John Otto, and Fabián E. Bustamante. 2011. Distributed systems and natural disasters: BitTorrent as a global witness. In *Proc. CoNEXT Special Workshop on Internet and Disasters*.
- [8] Zachary S. Bischof, Fabián E. Bustamante, and Rade Stanojevic. 2014. Need, Want, Can Afford - Broadband Markets and the Behavior of Users. In *Proc. of IMC*.
- [9] Timm Boettger, Ghida Ibrahim, and Ben Vallis. 2020. How the Internet reacted to Covid-19: A perspective from Facebook's Edge Network. In *Proc. of IMC*. 34–41.
- [10] The Editors of Encyclopaedia. Britannica. [n. d.]. "County". <https://www.britannica.com/topic/county>. ([n. d.]).
- [11] Caroline O. Buckee, Satchit Balsarijennifer, Chanmercé Crosas, Francesca Dominici, Urs Gassery, Yonatan H. Grad, Bryan Grenfell, Elizabeth Halloran, Moritz U. G. Kraemer, Marc Lipsitch, C. Jessica E. Metcalf, Lauren Ancel Meyers, Alex Perkins, Mauricio Santillana, Samuel V. Scarpino, Cecile Viboud, Amy Wesolowski, and Andrew Schroeder. 2020. Aggregated mobility data could help fight COVID-19. *Science* 368, 6487 (April 2020).
- [12] Juan Camilo Cardona, Rade Stanojevic, and Rubén Cuevas. 2013. On weather and internet traffic demand. In *Proc. of PAM*.
- [13] Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boëlle, Antoine Flahault, and Neil M. Ferguson. 2008. Estimating the impact of school closure on influenza transmission from sentinel data. *Nature* 452, 7188 (2008), 750–754.
- [14] CDC. 2021. COVID-19 Stats: COVID-19 Incidence by Age Group - United States, March 1 - November 14, 2020. *MMWR*. (January 2021).
- [15] Jasper Fuk-Woo Chan, Shuofeng Yuan, Kin-Hang Kok, Kelvin Kai-Wang To, Hin Chu, and Jin Yang et al. 2020. A familial cluster of pneumonia associated with

- the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* 395, 10223 (January 2020).
- [16] Matthew Conlen, Denise Lu, and James Glanz. 2021. Why Vaccines Alone Will Not End the Pandemic. (Jan 2021). <https://www.nytimes.com/interactive/2021/01/24/us/covid-vaccine-rollout.html>
- [17] Team CDC COVID, Bialek Stephanie, Bowen Virginia, Chow Nancy, Curns Aaron, Gierke Ryan, Hall Aron, Hughes Michelle, Pilishvili Tamara, Ritchey Matthew, et al. 2020. Geographic Differences in COVID-19 Cases, Deaths, and Incidence—United States, February 12–April 7, 2020. (2020).
- [18] Peter Craig, Srinivasa Vittal Katikireddi, Alastair Leyland, and Frank Popham. 2017. Natural Experiments: An overview of methods, approaches, and contributions to public health intervention research. *Annual Reviews of Public Health* 38 (March 2017), 39–56.
- [19] Alberto Dainotti, Claudio Squarcia, Emile Aben, Kimberly C Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. 2011. Analysis of country-wide Internet outages caused by censorship. In *Proc. of IMC*.
- [20] J. DiNardo. 2010. Natural Experiments and Quasi-Natural Experiments. *Microeconomics* (December 2010), 139–153.
- [21] Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 20, 5 (February 2020).
- [22] Alexander D. Douglas and Adrian V. S. Hill. 2020. Immunological considerations for SARS-CoV-2 human challenge studies. *Nature Reviews Immunology* 20 (2020), 715–716.
- [23] Thad Dunning. 2012. *Natural Experiments in the Social Sciences – A Design-based Approach*. Cambridge University Press.
- [24] Maria Effenbergera, Andreas Kronbichlerb, Jai Il Shin, Gert Mayer, Herbert Tilga, and Paul Perco. 2020. Association of the COVID-19 pandemic with Internet Search Volumes: A Google TrendsTM Analysis. *International Journal of Infectious Diseases* 95 (2020).
- [25] Ahmet AKtay et al. 2020. *Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.1)*. Technical Report. arXiv.org. <https://arxiv.org/abs/2004.04145>
- [26] Explained Desk. 2021. Explained: A list of countries that have imposed lockdowns to curb COVID-19 spread. The Indian Express. (May 2021).
- [27] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, et al. 2020. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In *Proc. of IMC*.
- [28] Seth Flaxman, Swapnil Mishra, and et al. Axel Gandy. 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020).
- [29] Richard Florida. 2016. The Reality of America's College Towns. (Sep 2016). <https://www.bloomberg.com/news/articles/2016-09-08/america-s-biggest-college-towns>
- [30] Centers for Disease Control and Prevention (CDC). 2021. Scientific Brief: Community Use of Cloth Masks to Control the Spread of SARS-CoV-2. (May 2021). <https://www.cdc.gov/coronavirus/2019-ncov/more/masking-science-sars-cov2.html>
- [31] European Centre for Disease Prevention and Control. [n. d.]. *Considerations relating to social distancing measures in response to the COVID-19 epidemic*. Technical Report. ECDC.
- [32] Google. [n. d.]. COVID-19 Community Mobility Report. <https://www.google.com/covid19/mobility/?hl=en>. ([n. d.]).
- [33] Kyra H. Grantz, Hannah R. Meredith, Derek A. T. Cummings, C. Jessica E. Metcalf, Bryan T. Grenfell, John R. Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, Caroline O. Buckee, and Amy Wesolowski. 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications* 11, 4961 (2020).
- [34] Miniwatts Marketing Group. 2020. World Internet users and 2020 populations stats. Internet usage statistics. (November 2020). <https://www.internetworldstats.com/stats.htm>
- [35] Blake Gumprecht. 2008. *The American College Town*. University of Massachusetts.
- [36] Thomas Hale, Tilbe Atav, Laura Hallas, Beatriz Kira, Toby Phillips, Anna Petherick, and Annalena Pott. 2020. Variation in US states responses to COVID-19. *Blavatnik School of Government*. (2020).
- [37] Harvard Health Publishing. 2021. If you've been exposed to the coronavirus. (August 2021). <https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus>
- [38] Elham Hatef, Hsien-Yen Chang, Christopher Kitchen, Jonathan P Weiner, and Hadi Kharrazi. 2020. Assessing the impact of neighborhood socioeconomic characteristics on COVID-19 prevalence across seven states in the United States. *Frontiers in public health* 8 (2020).
- [39] Jason Horowitz. 2021. Italy imposes lockdown measures as cases spike across Europe. New York Times. (March 2021).
- [40] Global Change Data Lab Our World in Data. [n. d.]. Statistics and Research: Coronavirus (COVID-19) Vaccinations. https://ourworldindata.org/covid-vaccinations?country=OWID_WRL. ([n. d.]).
- [41] Johns Hopkins University CSSE. [n. d.]. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. <https://github.com/CSSEGISandData/COVID-19>. ([n. d.]).
- [42] Mustafa T. Kartal, Özer Depren, and S. Depren. 2021. The relationship between mobility and COVID-19 pandemic: Daily evidence from an emerging country by causality analysis. *Transportation Research Interdisciplinary Perspectives* 10, 100366 (June 2021).
- [43] Josiah L. Kephart, Xavier Delclòs-Alió, Daniel A Rodríguez, Olga L. Sarmiento, Tonatiuh Barrientos-Gutiérrez, Manuel Ramírez-Zea, D. Alex Quistberg, Usama Bilal, and Ana V. Diez Roux. 2021. The effect of population mobility on COVID-19 incidence in 314 Latin American cities: a longitudinal ecological study with mobile phone location data. *The Lancet* (August 2021). Online First.
- [44] S. Shumugha Krishnan and Ramesh K. Sitaraman. 2012. Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. In *Proc. of IMC*.
- [45] Christopher T Leffler, Edsel Ing, Joseph D Lykins, Matthew C Hogan, Craig A McKeown, and Andrzej Grzybowski. 2020. Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. *The American Journal of Tropical Medicine and Hygiene* (2020).
- [46] Andra Lutu, Diego Perino, Marcelo Bagnulo, Enrique Fries-Martinez, and Javad Khangosstar. 2020. A Characterization of the COVID-19 Pandemic Impact on a Mobile Network Operator Traffic. In *Proc. of IMC*.
- [47] Wei Lyu and George L Wehby. 2020. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US. *Health Affairs* (2020).
- [48] Wei Lyu and George L Wehby. 2020. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US: Study examines impact on COVID-19 growth rates associated with state government mandates requiring face mask use in public. *Health affairs* 39, 8 (2020), 1419–1425.
- [49] Edouard Mathieu, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. 2021. A global database of COVID-19 vaccination. *Nature Human Behaviour* (May 2021).
- [50] Christina Maxouris. 2020. December is the deadliest month in the US since the coronavirus pandemic began – and projections for January are ‘nightmarish,’ expert says. (Dec 2020). <https://edition.cnn.com/2020/12/27/health/us-coronavirus-sunday/index.html>
- [51] Mark McClellan, Krishna Udayakumar, Michael Merson, and Gary Edson. 2021. *Reducing Global COVID Vaccine Shortages: New Research and Recommendations for US Leadership*. Policy Brief. Duke Global Health Innovation Center.
- [52] Timo Mitze, Reinhold Kosfeld adn Johannes Rode, and Klaus Wälde. 2020. Face masks considerably reduce COVID-19 cases in Germany. *Proc. of the National Academy of Sciences (PNAS)* 117, 51 (December 2020).
- [53] Candace Norwood. [n. d.]. Most states have issued stay-at-home orders, but enforcement varies widely. *PBS News Hour* ([n. d.]). <https://tinyurl.com/28s9bxhm>
- [54] Pierre Nouvellet and Sangeeta Bhatia et al. 2021. Reduction in mobility and COVID-19 transmission. *Nature Communication* 12, 1090 (February 2021).
- [55] Sangeeta Parihar, Suresh Singh Sankhla, Tarun Gehlot, Krishan Kumar Saini, and Poonam Poonia. [n. d.]. Influence of Temperature on Daily Growth Rate of COVID 19 Cases in Rajasthan, India. ([n. d.]).
- [56] Hannah Ritchie. [n. d.]. Google Mobility Trends: How has the pandemic changed the movement of people around the world? ([n. d.]). <https://ourworldindata.org/covid-mobility-trends>
- [57] Anna Rouw, Jennifer Kates, Adam Wexler, and Josh Michaud. [n. d.]. Tracking Global COVID-19 Vaccine Equity: An Update. <https://www.kff.org/coronavirus-covid-19/issue-brief/tracking-global-covid-19-vaccine-equity-an-update/>. ([n. d.]).
- [58] Tate Ryan-Mosley. [n. d.]. Why some COVID-19 tests in the US take more than a week. *Technology Review* (April [n. d.]).
- [59] Asiyah Sheikh, Zakariya Sheikh, and Aziz Sheikh. 2020. Novel approaches to estimate compliance with lockdown measures in the COVID-19 pandemic. *Journal of Global Health* 10, 1 (2020).
- [60] Matthew A Spinelli, David V Glidden, Efstathios D Gennatas, Michel Bielecki, Chris Beyer, George Rutherford, Henry Chambers, Eric Goosby, and Monica Gandhi. 2021. Importance of non-pharmaceutical interventions in lowering the viral inoculum to reduce susceptibility to infection by SARS-CoV-2 and potentially disease severity. *The Lancet* 21, 9 (September 2021). Personal Opinion.
- [61] Gábor Székely, Maria L. Rizzo, and Nail K. Bakirov. 2007. Measuring and Testing Dependence by Correlation of Distances. *The Annals of Statistics* 35, 6 (December 2007).
- [62] Andy Thomason. 2020. U. of Washington Cancels In-Person Classes, Becoming First Major U.S. Institution to Do So Amid Coronavirus Fears. *The Chronicle of Higher Education* (March 6 2020).
- [63] Eugenia Tognotti. 2013. Lessons from the History of Quarantine, from Plague to Influenza A. *Emerging Infectious Diseases* 19, 2 (February 2013).
- [64] Ramadhan Tosepu, Joko Gunawan, Devi Savitri Effendy, Hariati Lestari, Hartati Bahar, Pitrah Asfian, et al. 2020. Correlation between weather and COVID-19 pandemic in Jakarta, Indonesia. *Science of The Total Environment* 725 (2020), 138436.
- [65] US Census Bureau. [n. d.]. American Community Survey. ([n. d.]). <https://www.census.gov/programs-surveys/acs>

- [66] US Department of Health and Human Services. 2012. Principles of Epidemiology in Public Health Practice. (May 2012). <https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section1.html>
- [67] Miriam E Van Dyke, Tia M Rogers, Eric Pevzner, Catherine L Satterwhite, Hina B Shah, Wyatt J Beckman, Farah Ahmed, D Charles Hunt, and John Rule. 2020. Trends in county-level COVID-19 incidence in counties with and without a mask mandate - Kansas, June 1–August 23, 2020. *Morbidity and Mortality Weekly Report* 69, 47 (2020), 1777.
- [68] R. Webster, S. Brooks, L. Smith, L. Woodland, S. Wessely, and J. Rubin. 2020. How to improve adherence with quarantine: Rapid review of the evidence. *Public Health* 182 (May 2020).
- [69] David W. S. Wong and Yun Li. 2020. Spreading of COVID-19: Density matters. *PLoS ONE* 15, 12 (December 2020).
- [70] Joseph Younis, Harvey Freitag, Jeremy S Ruthberg, Jonathan P Romanes, Craig Nielsen, and Neil Mehta. 2020. Social Media as an Early Proxy for Social Distancing Indicated by the COVID-19 Reproduction Number: Observational Study. *JMIR Public Health and Surveillance* 6, 4 (October 2020).

A ADDITIONAL RESULTS

This appendix presents the complete list of graphs for the results discussed in previous sections, including those illustrating the relationship between CDN demand and mobility for 20 of the top US counties based on population density and Internet penetration (§A.1), between demand and growth rate of infection for the 25 counties with the highest number of reported COVID-19 cases by April 2020 (§A.2), and between demand and incidence of COVID-19 cases before and after campus closing in college towns around the US (§A.3).

A.1 Demand and Mobility

Figures 6 and 7 show the correlation between human mobility (Google CMR) and CDN demand during April and May 2020. See Sec. 4 for a discussion of the results.

A.2 Demand and infection GR

Figure 8 shows the correlation between CDN demand and Infection GR in 25 US counties. The reader is referred to Sec. 5 for the discussion of key findings.

A.3 Demand and infection ratio in college towns

Figure 9 shows the correlation between CDN demand and COVID-19 incidence. Table 5 lists the different campuses and student enrollments, their college town (county) and their population, as well as the population ratio [29]. The reader is referred to Sec. 6 for the discussion of key findings.

School Name	Region	Enrollment	Population	Ratio
University of Illinois	Champaign, IL	51,660	237,199	21.8%
Texas A&M University-Kingsville	Kleberg, TX	11,619	32,593	35.7%
Ohio University	Athens, OH	24,358	64,702	37.6%
Iowa State University	Story, IA	32,998	94,035	35.1%
University of Michigan	Washtenaw, MI	76,448	356,823	21.4%
University of South Dakota	Clay, SD	9,998	13,921	71.8%
Texas A&M	Brazos, TX	60,137	242,884	24.8%
Penn State	Centre, PA	47,823	158,728	30.1%
Indiana University	Monroe, IN	44,564	164,233	27.1%
Cornell University	Tompkins, NY	33,451	104,606	32.0%
South Plains College	Hockley, TX	8,534	23,577	36.2%
University of Missouri	Boone, MO	41,057	172,703	23.8%
Washington State University	Whitman, WA	25,823	46,808	55.2%
University of Kansas	Douglas, KS	29,512	116,559	25.3%
Blinn College	Washington, TX	17,707	34,437	51.4%
Virginia Tech	Montgomery, VA	45,150	181,555	24.9%
University of Mississippi	Lafayette, MS	21,482	52,921	40.6%
University of Florida	Alachua, FL	58,453	273,365	21.4%
Mississippi State University	Oktibbeha, MS	18,159	49,403	36.8%

Table 5: College town (county) and population ratio [29] of 19 of the largest college towns in the US.

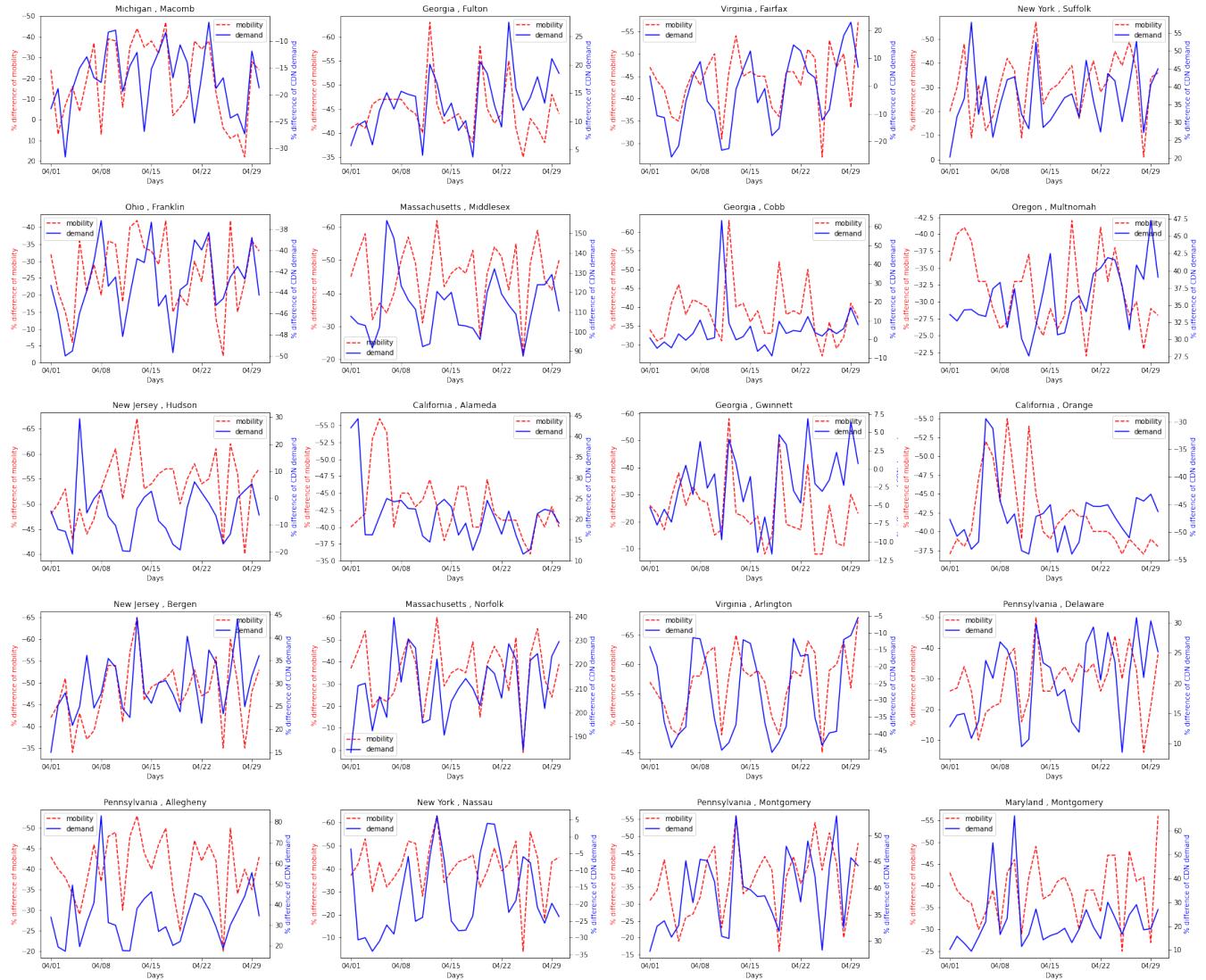


Figure 6: Correlation between CDN demand and mobility derived from Google's CMR for the top 20 counties with the highest Internet penetration and population density (April 2020).

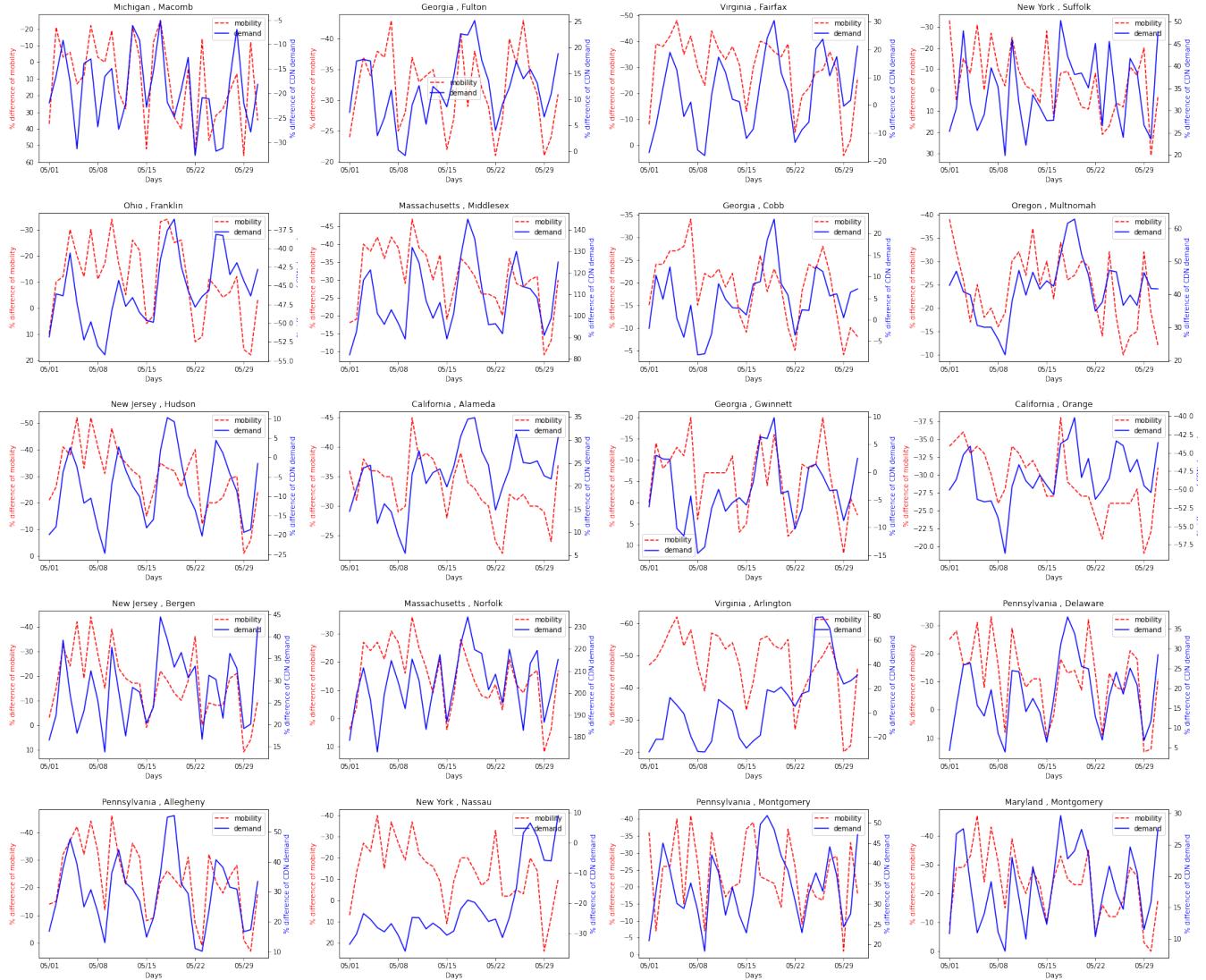


Figure 7: Correlation between CDN demand and mobility derived from Google's CMR for the top 20 counties with the highest Internet penetration and population density (May 2020).

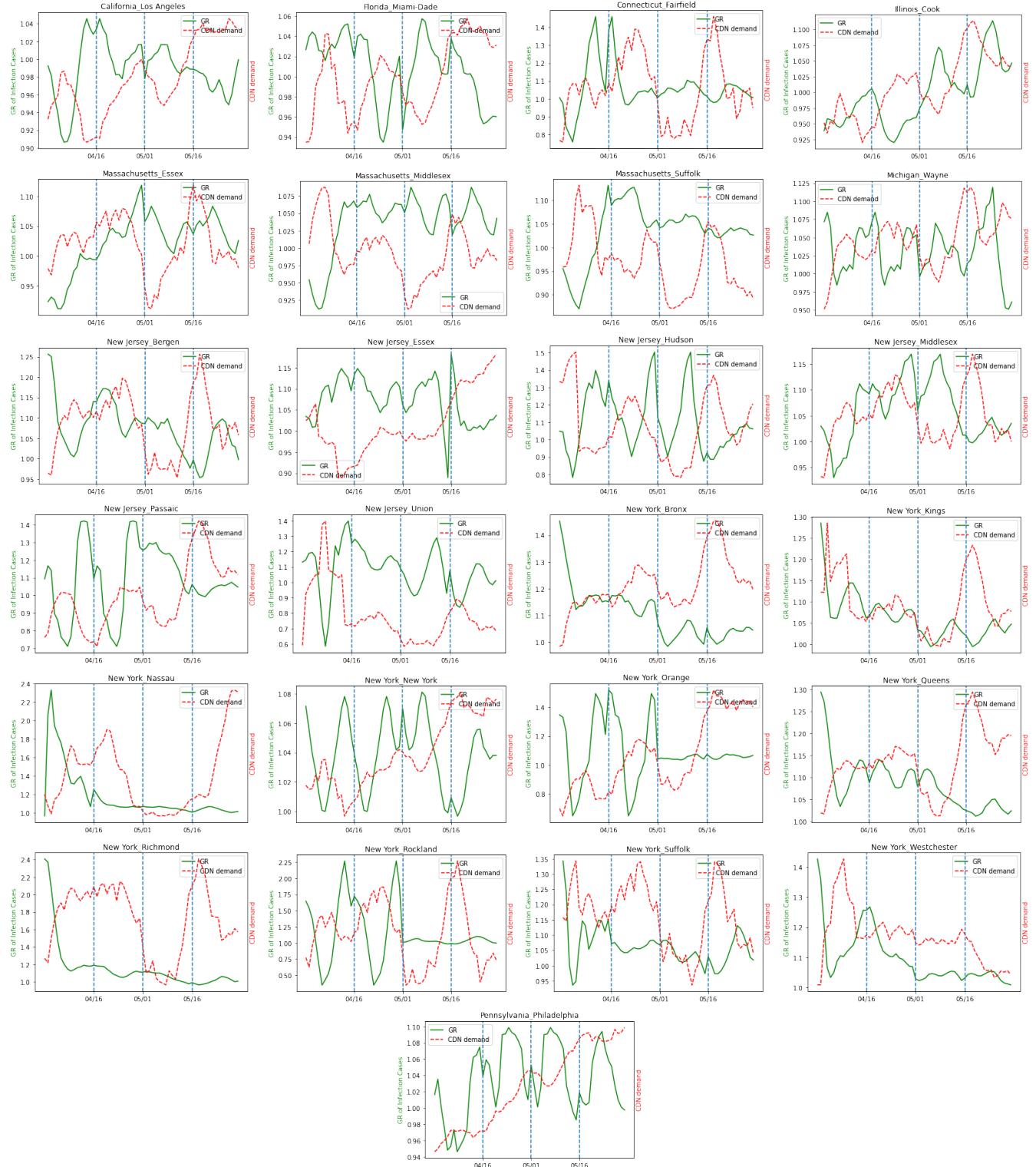


Figure 8: Correlation between CDN demand and COVID-19 infection Growth rate ratio in the 25 US counties with the highest number of cases by April 16, 2020.

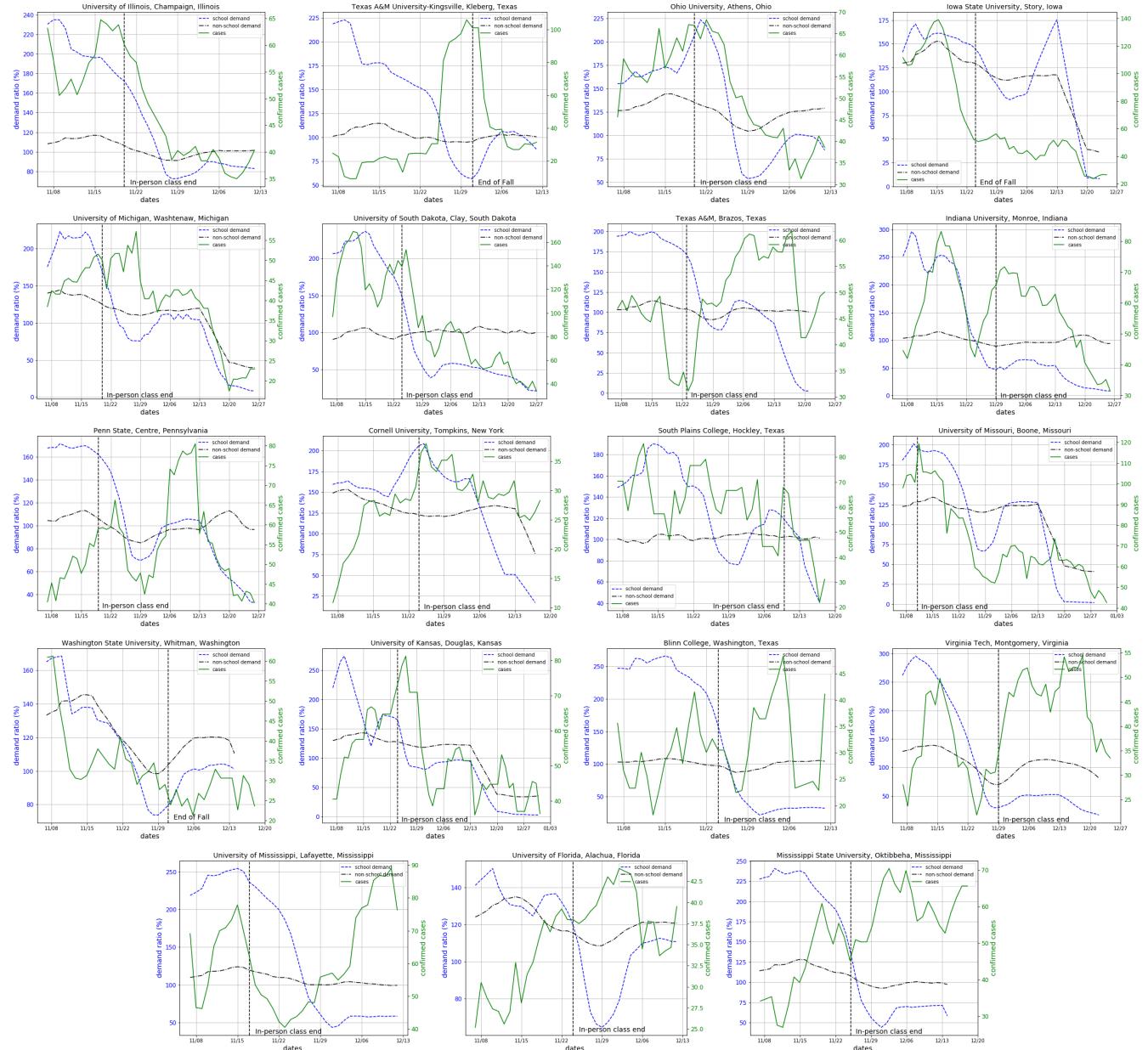


Figure 9: Correlation between CDN demand (school or non-school networks) and the number of confirmed COVID-19 cases per 100,000 in 19 of the largest college towns in the US, before and after the second campus closure in November 2020.