# Social-Network (*Facebook*) Connectedness and COVID-19 Outbreak in German Municipalities

Working Paper

******Do not cite nor circulate without previous author's authorization******

**Dr. Cruz-Aceves, Victor**

Research Associate at the Sociology Department at The University Of Kiel

Twitter: @vic_dca

**January 5, 2021**

Using aggregated data from one global social network (*Facebook*), I provide preliminary evidence that, even though non-geographical, social-network based connectedness between German municipalities and a given *COVID-19 hotspot* (Heinsberg) is significantly associated with COVID-19 prevalence, COVID-19 was not more likely to spread across municipalities with stronger social-network connections. Locations with a higher number of social ties to Heinsberg generally had more confirmed COVID-19 cases by March 30, 2020 only after a certain threshold is reached; otherwise, this relationship is actually negative. These associations are robust to the inclusion of controls for physical distance to the hotspot municipality, per capita income, population density and to the addition of data from late November.

## Introduction

One way in which we can begin to understand the intricacies of the spread of pandemic diseases such as COVID-19 is by finding out which persons are more likely to experience physical interactions with each other (Pastore y Piontti et al. 2019; Kuchler et al. 2020b). Namely, given that social connections affect patterns of physical interaction, the degree of social connectedness between locations is crucial as to determine the risk of experiencing future outbreaks. However, the geographic composition of social connections has proven to be hard to quantify, especially in comparable manners (Bailey et al. 2018). I confront this challenge in this article by using recently released, aggregated data from one of the biggest global social networks (*Facebook*) to quantify social connections between German municipalities. The objective of this article is to assess the outbreak of the 2020 COVID-19 pandemic at the municipal level in Germany and to test the generalizability of previous results that build upon the USA and Italy (Bailey et al. 2018; Aref et al. 2020; M. Bailey et al. 2019; Charoenwong et al. 2020).

In particular, I use the *index for social connectedness* between NUTS-3 locations (municipalities) in Germany. That covariate measures the probability that users of this social network in municipality-dyads are *Facebook friends* with each other (Bailey et al. 2018). I hypothesize that locations connected through a number of *Facebook friendship* links are more likely to have more physical interactions between their inhabitants, which will eventually lead to more opportunities for contagion of communicable diseases. Recent works have shown the potential of this novel data for the study of phenomena, for instance, travel patterns across European (NUTS-2) regions and within New York (M. Bailey et al. 2019; Aref et al. 2020), and trade between all countries and Europe (Bailey et al. 2020a). Another group of authors found out that counties more connected  to New York had higher likelihood of being destinations for people that fled the city during the COVID-19 pandemic (J. Coven und A. Gupta 2020). It was also found out that US municipalities with higher (Facebook-based) social connectedness to China and Italy complied to a higher extent with mobility restriction ordinances and suggested that social connections are channels of information about the pandemic, in the end affecting compliance with, and impact of, mobility restrictions (Charoenwong et al. 2020).

I proceed as follows. I first review relevant literature, followed by the description of the data that I will use throughout the article, afterwards I provide a description of the method of analysis, last, I present and discuss the results. In the end, I provide evidence that municipalities in Germany with stronger social connections to *COVID-19 hotspot municipality* Heinsberg actually had less confirmed COVID-19 cases per inhabitant as of March 30, 2020. This result holds after controlling for physical distance, demographic characteristics, and even after expanding the period of observation to November 2020. In line with previous works, the results here presented highlight that social connectedness has indeed non-negligible explanatory value – in addition to geographical distance and other epidemiological factors – to understand the spread of COVID-19 across Germany in general and related to one early hotspot in particular. However, the results also indicate that the positive association between COVID-19 prevalence and Facebook-based social connectedness identified in US municipalities and Italian provinces might not be generalizable to all countries.

## Literature Review

More generally, my results contribute to a scholarship that has applied *network theory* to construct spatial epidemiological models (Keeling und Eames 2005; M. J. Keeling und P. Rohani 2011; Danon et al. 2011). This literature goes beyond the basic assumption that people (within a population) are equally likely to interact with each other; instead, these scholars provide a more accurate picture of the dynamics of real-world connections (Chaoqi Yang et al. 2020; Mossong et al. 2008; Newman 2002). In particular, this article attempts to contribute to a growing literature that focuses of how social media and Internet-based communication can be useful for tracking communicable diseases. This literature has been divided in at least three general research agendas (Kuchler et al. 2020c). Some researchers interested in explaining public health outcomes have used content from other platforms, for instance *Instagram* (CORREIA et al. 2016), *Wikipedia* (Generous et al. 2014) and *Twitter* (Garzon-Alfonso und Rodriguez-Martinez 2018). Another group of scholars have used surveys and *crowd-sourced* data in order to track possible disease symptoms and identify potential outbreaks (Smolinski et al. 2015; Paolotti et al. 2014). Finally, another group of investigators made use of geo-located data in order to monitor motion patterns of individuals and to forecast the spread of diseases (Bengtsson et al. 2015; Wesolowski et al. 2015; Peixoto et al. 2020). More comprehensive reviews on it has been done elsewhere (P Giuliano und I Rasul 2020; Aiello et al. 2020).

There are a couple reasons that suggest that *social connections* to *early hotspots* may provide relevant information for tracking the COVID-19 spread in Germany and other countries, similar to what has been hypothesized for the US (M. Bailey et al. 2019). In other words, a number of articles reported that wealthy inhabitants from the New York area fled to other parts of the U.S. (T Tully und S Stowe 2020), which could have served as a vector that could have potentially spread the disease. In fact , both geneticists and epidemiologists reported that trips starting in New York seeded much of the first wave of COVID-19 outbreaks in the USA (B. Carey und J. Glanz 2020). More important Coven and Gupta (2020) found that connectedness to New York predicted travel patterns from the city early in the pandemic. Therefore, *social connections* to *early hotspots* may thus provide relevant information for tracking the COVID-19 spread.

**Analysis**

In this section, I explore how the domestic spread of confirmed COVID-19 cases is related to social connectedness to an early COVID-19 *hotspot* in Germany, in other words, I analyze the relationship between COVID-19 prevalence and social ties to Heinsberg municipality. Heinsberg is a location on the far west side of Germany which allegedly experienced the first major outbreak of this disease (Robert Koch Institut; LandNRW.de; Focus Online; Frankfurter Allgemeine; Hamburger Abendblatt).

Figure 1 shows heat maps with the distribution of COVID-19 cases per 10,000 residents across German municipalities as of as of  March 30  2020 (Badr et al. 2020), with darker colors corresponding to higher COVID-19 prevalence (for a map with data as of November 22, 2020, refer to Figure 4). It stands out that a second location located to the Eastern opposite side of the country, in the state of Bavaria, right at the border with Czech Republic, recorded a high number of confirmed COVID-19 by late March. Interestingly, two municipalities adjacent to that second hotspot in East Germany experienced higher prevalence rates than any of the locations (inside Germany) that border Heinsberg, as illustrated by the lighter colors of municipalities next to Heinsberg. One more interesting fact about that graph is that Heinsberg is also located at the border with another country. Nonetheless, that neighboring country is closer to Germany regarding a myriad of development indicators compared to Czech Republic.

Two more municipalities report high numbers of confirmed cases. These two locations are also located at the border with Austria and in the state of Bavaria. These are adjacent municipalities. However, these two experienced a lower rate of COVID-19 confirmed cases per 10,000 residents. One last thing that stands out from Figure 1 is that also the West-southern state of Baden-Württemberg also hosted to municipalities with high rates of COVID-19 confirmed cases by March 30, 2020. It is noteworthy that in that state, those two most affected municipalities do not border (the two neighboring, namely, France on the west and Switzerland in the South) other countries.

Figure 2 shows a heat map of the social connectedness Heinsberg to all other German municipalities; darker colors correspond to stronger social ties.

*Figure 1. COVID-19 Cases in German municipalities with data as of March 30, 2020*

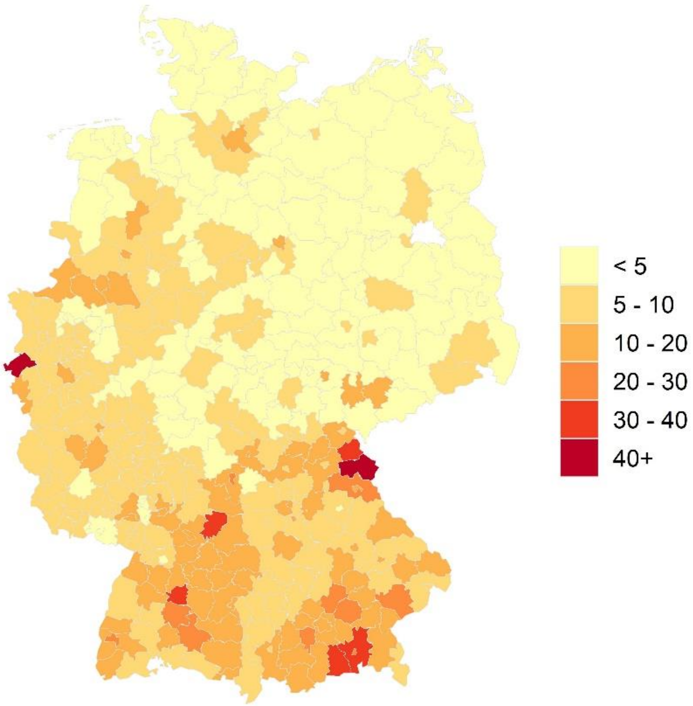
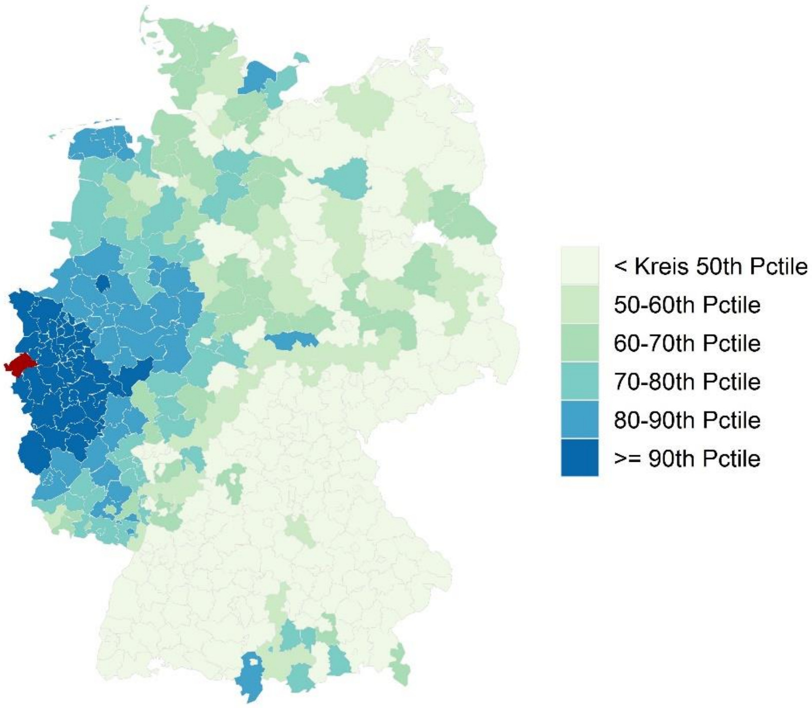



*Figure 2. Social connectedness of Heinsberg*




**Method and data**

In this section, I explore the relationship between confirmed COVID-19 cases as of March 30, 2020 and non-geographical *social ties* to Heinsberg in a regression framework. Specifically, I make use of the so-called *social connectedness index*. That measure is based on a snapshot of all active Facebook users from one month in particular from year 2019, namely July, namely those that reside in Germany and that interacted through Facebook over the 30 days prior to the date of the snapshot[1]. Having said that, the *relative probability of social connections* between German municipalities is measured according to the following Equation 1 (Bailey et al. 2018):

*Equation 1: (Relative Probability of) Social Connectedness*

$$SocialConnectedness_{ij} = \frac{FB\_Connections_{ij}}{FB\_Users_i * FB\_Users_j}$$

The upper part of the division is the total number of connections between persons residing in municipality *i* and individuals living in municipality *j*. The two elements of the denominator represent the number of eligible Facebook users in each German municipality. The division by the product of the number of Facebook users in the two German municipalities allows controlling for the fact that we will observe more *friendship* links between municipalities with higher numbers of Facebook users. In other words, this index proxies for the probability that two random users of this platform across the two locations are *friends* with each other, namely if $SocialConnectedness_{ij}$ is, for instance, twice as large, a Facebook user in location *i* is about twice as likely to be connected with a given user of this platform in location *j*.

In comparison to previous work that used data from other social media platforms (Ginsberg et al. 2009; Garzon-Alfonso und Rodriguez-Martinez 2018; Gittelman et al. 2015; Generous et al. 2014; CORREIA et al. 2016), the network-based and stable variable of interest in this article is less likely to suffer from changes in internet behavior or seasonality. Moreover, the social connectedness index does not require people to have experienced symptoms, which potentially allows identifying municipalities at-risk even before disease transmission. last, given that the main explanatory variable from this article is based only on aggregated connections (vis-a-vis individual movement), it is easily accessible to the public and consistently available for a large number of regions around the world. While some of the above-mentioned studies build their analysis with information on local networks, I am unaware of any work that uses a measure with the (high) level of coverage, granularity and thus comparability that the index used in this article offers

in order to explore the above-mentioned relationship between COVID-19 prevalence and social connectedness to Heinsberg, I estimate for each municipality *i* the following equation:

*Equation 2: Empirical Specification*

$$COVID19\ cases\ per\ 10\mathrm{k}_{\ i} = \beta_0 + \beta_1 \log\left(SocialConnectedness_{ij}\right) + X_i + \delta_{ij} + \varepsilon_i$$

[1] refer to "Appendix: Social Connectedness versus physical distance across Europe" for a graphical representation of this variable across Europe, and for a short analytical illustration about how it is associated with geographical distance

The vector $X_i$ includes demographic measures, particularly population density and GDP per inhabitant. The parameter $\delta_{ij}$ consists of dummy variables for the quantile of the geographic distance from each German municipality to the identified hotspot location. For summary statistics, refer to Table 5

**Results**

Column number one of the following Table 1 shows the Association between purely non-geographical social distance between German municipalities and the COVID-19 hotspot, and the number of COVID-19 confirmed cases per 10,000 residents in Germany. I exclude those municipalities within 50 miles of Heinsberg while those areas have strong social links to Heinsberg, they are also close enough geographically such that their populations might interact physically with Heinsberg residents even in the absence of social links (e.g., in supermarkets, churches). One concern with interpreting these initial relationships is that they might be picking up other factors that affect the spread of COVID-19 and that are associated with social connectedness. Namely, even after dropping municipalities within 50 miles of Heinsberg, the correlations might be primarily showing geographic distance to Heinsberg (which is related to the number of friendship links to Heinsberg). The next model adds geographic distance to the picture (measured with 20 dummy variables[2] ; I omit the coefficients corresponding to dummy variables for space purposes. Refer to Table 3 in the appendix for full results). We observe that both the sign (negative) and the level of statistical significance of our variable of interest does not vary between these two models. Only the magnitude of the estimated coefficient differs. In column number three, I include demographic factors to the equation. Even after controlling for geographic distance, per capita income and population density, the coefficient for social connectedness is still negative, highly statistically significant and slightly of a smaller dimension compared to the estimated coefficient of model number two.

[2] In Table 3, I implement alternative parametrizations of geographic distance, specifically, I estimate models with 40, 60, 80 and 100 dummies. These numbers of dummy variables are based on results from the US and Italy, to which I referred to in this text, accounted for geography with different numbers for each case. Namely, 20 dummies for the quantile of the province distance to the province of interest for Italy and 100 dummies for the percentile of the county distance to the County of interest for the US. Those results, there is no sufficient evidence to believe that the way in which geographic distance is parametrized affects dramatically the size/sign/statistical significance of either the main variable nor of the controls.

Table 1. Main models

| | | | |
|---|---|---|---|
| log_sci | -2.34 | -6.01 | -5.33 |
| | (0.73)*** | (1.61)*** | (1.64)*** |
| gdp_per_hab | | | 0.00 |
| | | | (0.00)*** |
| pop_per_km | | | -0.00 |
| | | | (0.00)* |
| Constant | 27.62 | 65.51 | 57.69 |
| | (6.17)*** | (15.87)*** | (16.46)*** |
| Chi2 | . | . | . |
| Df_M | 1 | 20 | 22 |
| P | . | . | . |
| Ll | -1,260.10 | -1,241.58 | -1,235.14 |
| *N* | 391 | 391 | 391 |

* $p$<0.1; ** $p$<0.05; *** $p$<0.01

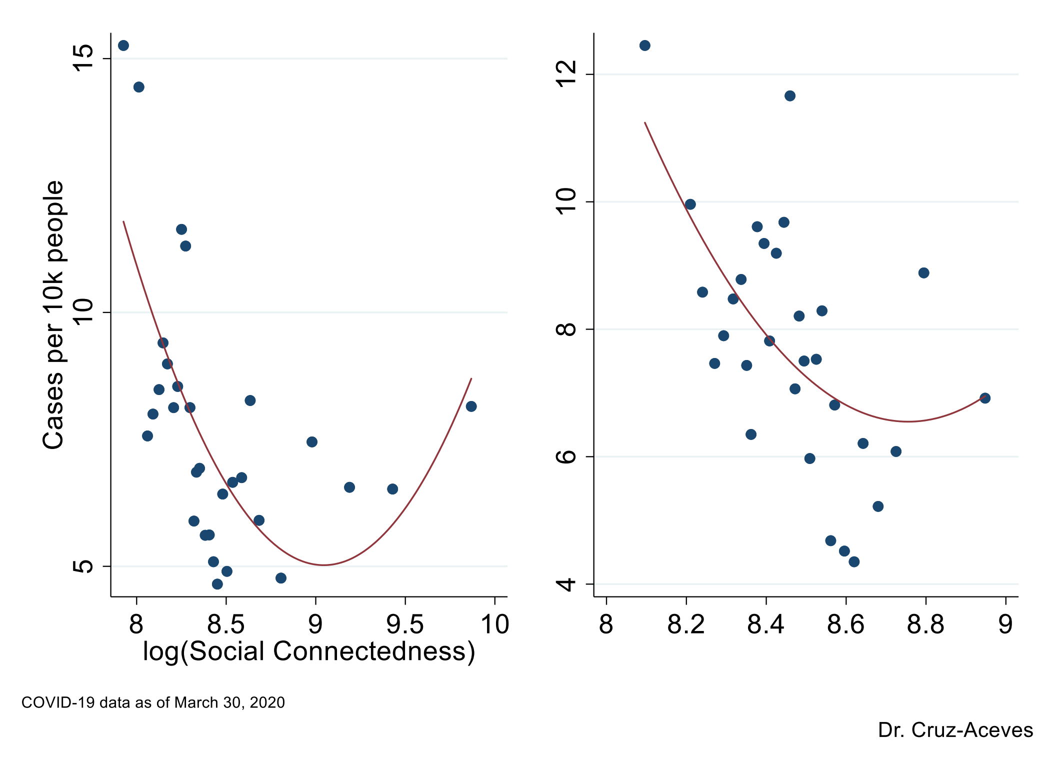
Figure 3 visualizes results from columns number two and three from Table 2 using binned binscatter plots[3] with municipalities more than 50 kilometers from Heinsberg as the unit of observation. To generate that plot I group the main explanatory variable into equal-sized bins and graph the average against the corresponding *average case density*. The Red line follows *quadratic fit* regressions. The right panel is constructed similarly, nonetheless, I first regress the social connectedness index and COVID-19 confirmed cases per 10,000 inhabitants on the above-mentioned control variables, next, I plot the *residualized values* on each axis. In the left panel we observe a U-shaped relationship between COVID-19 prevalence and social ties to Heinsberg (refer to **Appendix. Results with data from late November 22, 2020** for results using more recent data). In the right Panel of Figure 3, I plot of the relationship between social connectedness to Heinsberg and COVID-19 cases that controls for a number of these possible confounding covariates (in addition to excluding geographically adjacent municipalities)[4]. Conditional on these other factors, the right Panel of Figure 3 shows a negative relationship between COVID-19 cases as of March 20, 2020 and social connectedness to Heinsberg (refer to **Appendix. Results with data from late November 22, 2020** for results using more recent data). With these controls, a 1% increase of a municipality's social

[3] This type of graphs groups the explanatory variable into equal-sized bins, then, it computes the average of the x-axis and y-axis covariates within each bin, and finally, a scatterplot of these data points is generated. The result is a non-parametric visualization of the conditional expectation function

[4] I control for the geographic distance between each municipality and Heinsberg non-parametrically by including 20 dummies for percentiles of that distance.

connectedness to Heinsberg is associated with a decrease of about -0.0533 COVID-19 cases per 10,000 residents. [5]

Figure 3. Binscatter (Heinsberg as unit of observation, data from March 30, 2020)




## Discussion

Whereas the results, regardless of whether we use data from the beginning of the pandemic in Germany in late March or from late November 2020, highlight the relevance of accounting for not only geographical but also – and perhaps more important – none geographical *connectedness* amongst individuals (illustrated by *Facebook friendships* in this paper), the analysis referred to a distinct *shape* of the Association between non-graphical social connectedness and the prevalence (rate per 10,000 people) of COVID-19 than the one previously registered in US municipalities. Even controlling for geography and key demographic factors

[5] We can interpret these results for different values (University of Virginia Library). In order to calculate the figure, namely a 10% increase, I multiplied the coefficient by log(1.10). In other words, for every 10% increase in the independent variable, our dependent variable increases by about -5.33 * log(1.10) = -0.22. Alternatively, if I divide the estimated coefficient by 100, I obtain that for every 1% increase in the independent variable, our dependent variable decreases by about -0.0533. An alternative interpretation is that, controlling for certain demographic factors, a one standard deviation increase in municipalities' social connectedness to the hotspot is associated with a decrease of about -5.33 Covid 19 confirmed cases per 10,000 inhabitants (Chetty et al. 2013).

(per capita income and population density, for comparability purposes), the relationship between social connectedness and confirmed COVID-19 cases in Germany follows a U-shape, in other words, high numbers of COVID-19 confirmed cases seem to have been related to low non-geographical social connectedness. However, after a given point of social connectedness is reached, the above-mentioned association inverts and starts to grow, which can be interpreted as preliminary evidence that a high Facebook based social connectivity is related to high COVID-19 prevalence values after a given threshold has been reached. The latter seems more in line with the manifestation of this phenomenon in Italian provinces up until March 30, 2020 (Kuchler et al. 2020a). In Italy, when authors do not control for the covariates above-mentioned, the relationship between Facebook based social connectedness and the rate of confirmed COVID-19 cases per 10,000 inhabitants shows a more subtle U-shaped.

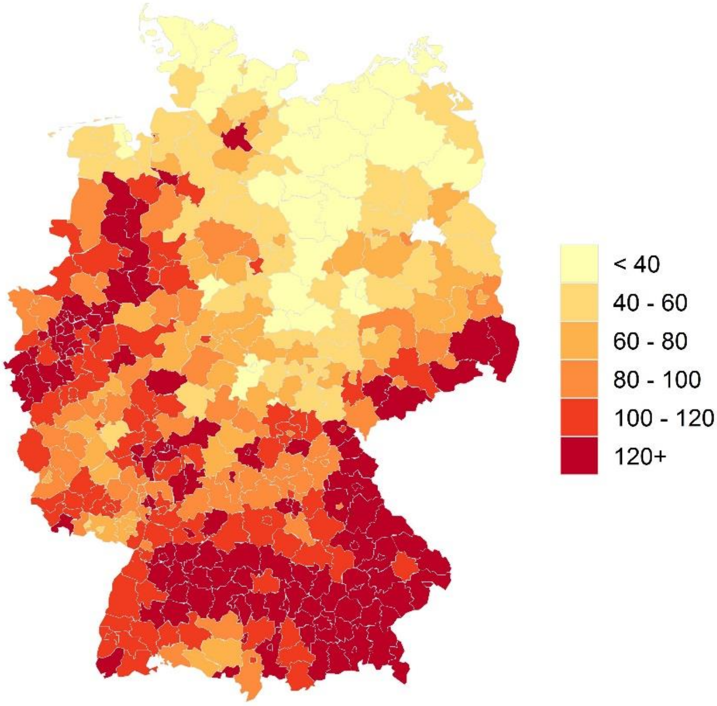
## Conclusion and Avenues For For Future Research

The main question that emerges from the results here presented is why social connectedness between German municipalities and a COVID-19 hotspot location in that country is negatively associated with COVID-19 prevalence, contrary to what has been found in the USA and Italy (Kuchler et al. 2020b), and also opposed to the relationship between Facebook-based connectedness and international trade (Bailey et al. 2020a). That question can begin to be explored by looking at the determinants of social connectedness in Germany, similar to what has been done already for the USA at the municipality level (Bailey et al. 2018) and Europe at the regional level (Aref et al. 2020).

Epidemiological scholarship can build on the above-mentioned works and the evidence provided in this. At a broader level, I results can motivate and hopefully contribute to the literature on the determinants/effects of social networks (Bailey et al. 2020b; Mossay und Picard 2011; Brueckner und Largey 2008; Bailey et al. 2020a, 2020a).

**Appendix *COVID-19 Cases in German municipalities with data as of November 22, 2020***

*Figure 4.COVID-19 Cases in German municipalities with data as of November 22, 2020*




**Appendix. Robustness checks**

When we only estimate a univariate model of distance in kilometers between the hotspot municipality and the rest of locations, one could easily think based on the (positively signed coefficient) result that larger distances between hotspot and the rest of municipalities is actually associated with a bigger prevalence of COVID-19, as can be seen in column number one of Table 2. This is probably due to a wrong measurement of distance. In fact, if we look at Figure 1, there are municipalities far away from the hotspot, especially in the two states located south from Germany; actually, we observe in Figure 1 that handful of municipalities right at the border with Czech Republic (East) and Austria (without) registered a high COVID-19 prevalence, comparable to that of the hotspot. For that reason, in the next column from Table 2 I estimate a model that measures distance between municipalities and the hotspot with 20 dummies for the quantile of the municipality distance to the hotspot location. In that model, we actually observe that the way in which we account for distance between hotspot and other locations is not trivial.

The third model shows the result of regressing our non-geographical, Facebook based social connectedness index on the rate of confirmed cases of COVID-19 per 10,000 inhabitants. Both the negative sign and the magnitude of the estimated coefficient of our *social connectedness index* corresponding to this univariate model and a bivariate one where geographical distance (not represented by dummies) is controlled for, are (almost) identical; interestingly, though, the size of estimated coefficient of the geographic distance covariate and the sign of it do change when social connectedness is accounted for in the same model through a geographical and a non-geographical variable (column number four). If we add demographic controls to the latter (column number six), none of the distance variables are statistically significant.

Column number five accounts for geographical distance with 20 dummies for the quintiles of the municipality distance to the hotspot location and for non-geographical social distance as well. We observe that few geographical dummies are not statistically significant; nonetheless, they are all negatively signed, in line with the coefficient for social distance, which is also statistically significant. This statistical characteristics are robust to the inclusion of demographic factors (both per capita income and population density are statistically significant, too), as we can see in the last column of Table 2. The relationship between COVID-19 prevalence and social distance (accounting for geographic distance – with dummy variables, as above described – and demographics) is shown in Figure 3.

Table 2. Full (Main) Models

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| log_dist | 1.46 | | | -0.36 | | -0.06 | |
| | (0.58)** | | | (1.07) | | (1.10) | |
| 1bn.dist_group | | -0.31 | | | -2.98 | | -2.80 |
| | | (2.27) | | | (2.35) | | (2.33) |
| 2.dist_group | | 2.11 | | | -1.94 | | -2.70 |
| | | (2.27) | | | (2.48) | | (2.58) |
| 3.dist_group | | 0.34 | | | -6.81 | | -7.22 |
| | | (2.27) | | | (2.94)** | | (3.05)** |
| 4.dist_group | | -0.46 | | | -8.02 | | -8.25 |
| | | (2.27) | | | (3.02)*** | | (3.09)*** |
| 5.dist_group | | -0.57 | | | -8.42 | | -8.67 |
| | | (2.27) | | | (3.07)*** | | (3.16)*** |
| 6.dist_group | | -0.37 | | | -8.14 | | -8.60 |
| | | (2.27) | | | (3.05)*** | | (3.18)*** |
| 7.dist_group | | -0.95 | | | -9.50 | | -9.70 |
| | | (2.27) | | | (3.20)*** | | (3.32)*** |
| 8.dist_group | | 0.27 | | | -8.22 | | -8.68 |
| | | (2.27) | | | (3.19)** | | (3.31)*** |
| 9.dist_group | | 1.93 | | | -7.23 | | -7.47 |
| | | (2.27) | | | (3.32)** | | (3.44)** |
| 10.dist_grou | | 2.40 | | | -7.00 | | -7.60 |

| | | | | | | |
|---|---|---|---|---|---|---|
| p | | | | | | |
| | (2.27) | | | (3.37)** | | (3.52)** |
| 11.dist_group | 2.66 | | | -7.00 | | -7.15 |
| | (2.27) | | | (3.42)** | | (3.57)** |
| 12.dist_group | 1.30 | | | -8.31 | | -8.30 |
| | (2.27) | | | (3.41)** | | (3.52)** |
| 13.dist_group | 1.31 | | | -8.05 | | -8.20 |
| | (2.27) | | | (3.36)** | | (3.49)** |
| 14.dist_group | -0.14 | | | -9.58 | | -9.49 |
| | (2.27) | | | (3.37)*** | | (3.53)*** |
| 15.dist_group | 4.05 | | | -5.59 | | -6.16 |
| | (2.27)* | | | (3.41) | | (3.54)* |
| 16.dist_group | -1.06 | | | -10.02 | | -9.93 |
| | (2.27) | | | (3.28)*** | | (3.42)*** |
| 17.dist_group | 5.00 | | | -4.16 | | -4.45 |
| | (2.27)** | | | (3.32) | | (3.43) |
| 18.dist_group | 3.84 | | | -5.29 | | -5.74 |
| | (2.27)* | | | (3.31) | | (3.47)* |
| 19.dist_group | 2.00 | | | -7.49 | | -7.57 |
| | (2.27) | | | (3.38)** | | (3.54)** |
| log_sci | | -2.34 | -2.73 | -6.01 | -1.72 | -5.33 |
| | | (0.73)*** | (1.35)** | (1.61)*** | (1.36) | (1.64)*** |
| gdp_per_hab | | | | | 0.00 | 0.00 |
| | | | | | (0.00)*** | (0.00)*** |
| pop_per_km | | | | | -0.00 | -0.00 |
| | | | | | (0.00) | (0.00)* |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Constant | -0.51 | 6.58 | 27.62 | 32.93 | 65.51 | 19.86 | 57.69 |
| | (3.31) | (1.83)*** | (6.17)*** | (16.84)* | (15.87)*** | (17.25) | (16.46)*** |
| Chi2 | . | . | . | . | . | . | . |
| Df_M | 1 | 19 | 1 | 2 | 20 | 4 | 22 |
| P | . | . | . | . | . | . | . |
| Ll | -1,262.09 | -1,248.82 | -1,260.10 | -1,260.04 | -1,241.58 | -1,252.24 | -1,235.14 |
| *N* | 391 | 391 | 391 | 391 | 391 | 391 | 391 |

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

*Table 3. Models with different parameterization of geographic distance*

| | | | | | |
|---|---|---|---|---|---|
| log_sci | -5.33 | -6.12 | -4.68 | -4.92 | -4.73 |
| | (1.64)*** | (1.76)*** | (1.76)*** | (1.86)*** | (1.81)*** |
| 1bn.dist_group | -2.80 | -10.19 | | | |
| | (2.33) | (6.39) | | | |
| 2.dist_group | -2.70 | -11.73 | | | |
| | (2.58) | (6.48)* | | | |
| 3.dist_group | -7.22 | -13.17 | -3.00 | -9.21 | |
| | (3.05)** | (6.69)** | (3.76) | (6.55) | |
| 4.dist_group | -8.25 | -11.85 | -4.45 | -10.74 | -3.25 |
| | (3.09)*** | (6.77)* | (3.77) | (6.58) | (4.45) |
| 5.dist_group | -8.67 | -13.47 | -4.45 | -10.67 | -3.51 |
| | (3.16)*** | (6.84)** | (3.93) | (6.68) | (4.40) |
| 6.dist_group | -8.60 | -17.60 | -1.72 | -11.78 | -4.53 |
| | (3.18)*** | (7.18)** | (4.03) | (6.78)* | (4.51) |
| 7.dist_group | -9.70 | -17.54 | -4.06 | -11.93 | -4.52 |
| | (3.32)*** | (7.16)** | (4.01) | (6.95)* | (4.50) |
| 8.dist_group | -8.68 | -18.23 | -5.08 | -9.26 | -5.95 |
| | (3.31)*** | (7.18)** | (4.12) | (6.92) | (4.61) |
| 9.dist_group | -7.47 | -19.00 | -8.13 | -11.71 | -4.59 |
| | (3.44)** | (7.20)*** | (4.51)* | (7.00)* | (4.79) |
| 10.dist_group | -7.60 | -18.28 | -8.58 | -11.54 | -4.54 |
| | (3.52)** | (7.26)** | (4.49)* | (6.91)* | (4.78) |
| 11.dist_group | -7.15 | -19.88 | -7.00 | -12.44 | -1.40 |
| | (3.57)** | (7.23)*** | (4.37) | (7.18)* | (4.71) |
| 12.dist_group | -8.30 | -18.94 | -7.67 | -15.87 | -4.64 |
| | (3.52)** | (7.26)*** | (4.51)* | (7.42)** | (4.72) |
| 13.dist_group | -8.20 | -19.15 | -9.55 | -15.33 | -5.92 |
| | (3.49)** | (7.28)*** | (4.44)** | (7.41)** | (4.81) |
| 14.dist_group | -9.49 | -20.55 | -9.24 | -14.48 | -4.93 |
| | (3.53)*** | (7.34)*** | (4.50)** | (7.37)* | (4.96) |
| 15.dist_group | -6.16 | -19.91 | -7.62 | -16.41 | -8.76 |
| | (3.54)* | (7.40)*** | (4.64) | (7.40)** | (5.18)* |

| | | | | | |
|---|---|---|---|---|---|
| 16.dist_group | -9.93 | -19.71 | -10.07 | -15.33 | -8.48 |
| | (3.42)*** | (7.38)*** | (4.47)** | (7.39)** | (5.18) |
| 17.dist_group | -4.45 | -18.69 | -9.87 | -16.90 | -9.25 |
| | (3.43) | (7.36)** | (4.53)** | (7.43)** | (5.22)* |
| 18.dist_group | -5.74 | -16.61 | -8.34 | -17.64 | -6.19 |
| | (3.47)* | (7.46)** | (4.51)* | (7.40)** | (5.04) |
| 19.dist_group | -7.57 | -19.56 | -9.25 | -16.06 | -9.69 |
| | (3.54)** | (7.48)*** | (4.66)** | (7.47)** | (5.22)* |
| gdp_per_hab | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.00)*** | (0.00)*** | (0.00)*** | (0.00)*** | (0.00)** |
| pop_per_km | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.00)* | (0.00)** | (0.00)* | (0.00)** | (0.00) |
| 20.dist_group | | -18.86 | -9.88 | -15.82 | -8.62 |
| | | (7.51)** | (4.54)** | (7.52)** | (5.15)* |
| 21.dist_group | | -17.67 | -11.08 | -16.44 | -9.80 |
| | | (7.56)** | (4.68)** | (7.47)** | (5.14)* |
| 22.dist_group | | -18.95 | -10.48 | -18.53 | -9.51 |
| | | (7.48)** | (4.65)** | (7.45)** | (5.24)* |
| 23.dist_group | | -16.77 | -9.23 | -16.78 | -9.75 |
| | | (7.66)** | (4.74)* | (7.47)** | (5.15)* |
| 24.dist_group | | -17.55 | -10.36 | -15.41 | -9.57 |
| | | (7.46)** | (4.65)** | (7.37)** | (5.24)* |
| 25.dist_group | | -20.35 | -7.69 | -18.17 | -10.14 |
| | | (7.58)*** | (4.76) | (7.62)** | (5.28)* |
| 26.dist_group | | -18.88 | -9.68 | -16.32 | -6.94 |
| | | (7.50)** | (4.65)** | (7.54)** | (5.30) |
| 27.dist_group | | -18.82 | -1.73 | -17.63 | -10.78 |
| | | (7.51)** | (4.92) | (7.50)** | (5.16)** |
| 28.dist_group | | -20.89 | -13.02 | -18.69 | -9.98 |
| | | (7.51)*** | (4.68)*** | (7.56)** | (5.22)* |
| 29.dist_group | | -19.48 | -8.13 | -17.80 | -11.36 |
| | | (7.58)** | (4.83)* | (7.61)** | (5.27)** |
| 30.dist_group | | -14.88 | -8.52 | -18.75 | -7.49 |
| | | (7.55)** | (4.83)* | (7.65)** | (5.13) |
| 31.dist_group | | -18.75 | -10.76 | -16.38 | -11.42 |

| | | | | |
|---|---|---|---|---|
| | (7.51)** | (4.91)** | (7.65)** | (5.31)** |
| 32.dist_group | -20.64 | -4.81 | -19.12 | -9.11 |
| | (7.41)*** | (4.91) | (7.50)** | (5.36)* |
| 33.dist_group | -20.44 | -7.89 | -15.74 | -10.60 |
| | (7.49)*** | (4.83) | (7.76)** | (5.35)** |
| 34.dist_group | -15.35 | -7.42 | -14.72 | -10.30 |
| | (7.49)** | (4.88) | (7.60)* | (5.26)* |
| 35.dist_group | -14.72 | -7.40 | -17.99 | -11.37 |
| | (7.41)** | (5.06) | (7.60)** | (5.25)** |
| 36.dist_group | -15.63 | -7.02 | -8.14 | -10.97 |
| | (7.43)** | (4.81) | (7.80) | (5.40)** |
| 37.dist_group | -17.12 | -8.28 | -20.27 | -10.96 |
| | (7.55)** | (4.88)* | (7.63)*** | (5.36)** |
| 38.dist_group | -15.76 | -10.59 | -18.48 | -11.76 |
| | (7.53)** | (4.90)** | (7.77)** | (5.48)** |
| 39.dist_group | -20.78 | -11.39 | -15.68 | -9.04 |
| | (7.57)*** | (4.84)** | (7.70)** | (5.43)* |
| 2bn.dist_group | | -2.37 | -9.75 | |
| | | (3.65) | (6.53) | |
| 40.dist_group | | -6.78 | -18.04 | -11.83 |
| | | (4.91) | (7.67)** | (5.33)** |
| 41.dist_group | | -8.13 | -14.79 | -8.33 |
| | | (4.80)* | (7.88)* | (5.34) |
| 42.dist_group | | -10.15 | -13.23 | -8.07 |
| | | (4.94)** | (7.83)* | (5.53) |
| 43.dist_group | | -8.40 | -17.06 | -9.79 |
| | | (4.80)* | (7.82)** | (5.40)* |
| 44.dist_group | | -11.29 | -14.64 | -10.76 |
| | | (4.94)** | (7.67)* | (5.34)** |
| 45.dist_group | | -10.59 | -18.35 | 2.55 |
| | | (4.91)** | (7.81)** | (5.56) |
| 46.dist_group | | 0.76 | -12.36 | -13.39 |
| | | (4.97) | (7.86) | (5.52)** |
| 47.dist_group | | -10.34 | -15.91 | -13.84 |
| | | (4.78)** | (8.01)** | (5.38)** |

| | | | |
|---|---|---|---|
| 48.dist_group | -10.82 | -14.43 | -7.74 |
| | (4.80)** | (7.68)* | (5.58) |
| 49.dist_group | -10.56 | -15.65 | -10.14 |
| | (4.73)** | (7.75)** | (5.44)* |
| 50.dist_group | -9.95 | -16.72 | -10.52 |
| | (4.82)** | (7.90)** | (5.45)* |
| 51.dist_group | -4.38 | -18.86 | -7.08 |
| | (4.84) | (7.79)** | (5.55) |
| 52.dist_group | -3.51 | -19.43 | -11.99 |
| | (4.67) | (7.69)** | (5.58)** |
| 53.dist_group | -6.70 | -13.34 | -4.24 |
| | (4.86) | (7.83)* | (5.79) |
| 54.dist_group | -4.73 | -15.85 | -8.98 |
| | (4.87) | (7.81)** | (5.62) |
| 55.dist_group | -7.02 | -16.83 | -7.28 |
| | (4.81) | (7.72)** | (5.47) |
| 56.dist_group | -6.60 | -18.27 | -10.30 |
| | (4.83) | (7.77)** | (5.51)* |
| 57.dist_group | -1.17 | -18.59 | -5.90 |
| | (5.01) | (7.77)** | (5.64) |
| 58.dist_group | -11.18 | -14.79 | -8.29 |
| | (4.80)** | (7.78)* | (5.68) |
| 59.dist_group | -10.68 | -19.08 | -9.10 |
| | (4.92)** | (7.90)** | (5.76) |
| 60.dist_group | | -19.40 | -8.32 |
| | | (7.73)** | (5.54) |
| 61.dist_group | | -5.30 | -6.93 |
| | | (7.90) | (5.37) |
| 62.dist_group | | -12.87 | -7.99 |
| | | (7.85) | (5.62) |
| 63.dist_group | | -19.54 | -12.03 |
| | | (7.69)** | (5.63)** |
| 64.dist_group | | -17.86 | -11.31 |
| | | (7.64)** | (5.57)** |
| 65.dist_group | | -18.78 | -12.02 |

| | | |
|---|---|---|
| | (7.69)** | (5.43)** |
| 66.dist_group | -18.42 | -6.47 |
| | (7.78)** | (5.57) |
| 67.dist_group | -17.47 | -10.22 |
| | (7.72)** | (5.67)* |
| 68.dist_group | -11.31 | -7.19 |
| | (7.77) | (5.48) |
| 69.dist_group | -14.49 | -10.31 |
| | (7.73)* | (5.51)* |
| 70.dist_group | -10.40 | -10.77 |
| | (7.46) | (5.59)* |
| 71.dist_group | -14.17 | -11.35 |
| | (7.87)* | (5.53)** |
| 72.dist_group | -11.42 | -6.39 |
| | (7.74) | (5.53) |
| 73.dist_group | -15.04 | -12.03 |
| | (7.63)** | (5.46)** |
| 74.dist_group | -17.23 | -12.47 |
| | (7.82)** | (5.77)** |
| 75.dist_group | -12.07 | -12.36 |
| | (7.80) | (5.46)** |
| 76.dist_group | -7.81 | 4.93 |
| | (7.83) | (5.70) |
| 77.dist_group | -18.72 | -5.86 |
| | (7.76)** | (5.75) |
| 78.dist_group | -16.87 | -8.83 |
| | (7.90)** | (5.51) |
| 79.dist_group | -19.66 | -13.16 |
| | (7.78)** | (5.45)** |
| 3bn.dist_group | | -2.46 |
| | | (4.41) |
| 80.dist_group | | -11.13 |
| | | (5.40)** |
| 81.dist_group | | -13.39 |
| | | (5.52)** |

| | |
|---|---|
| 82.dist_group | -10.36 |
| | (5.40)* |
| 83.dist_group | -10.36 |
| | (5.53)* |
| 84.dist_group | -10.17 |
| | (5.52)* |
| 85.dist_group | -5.19 |
| | (5.54) |
| 86.dist_group | -7.04 |
| | (5.44) |
| 87.dist_group | -3.24 |
| | (5.47) |
| 88.dist_group | -3.59 |
| | (5.36) |
| 89.dist_group | -8.23 |
| | (5.57) |
| 90.dist_group | -6.49 |
| | (5.52) |
| 91.dist_group | -2.69 |
| | (5.44) |
| 92.dist_group | -9.54 |
| | (5.55)* |
| 93.dist_group | -12.08 |
| | (5.60)** |
| 94.dist_group | -3.04 |
| | (5.53) |
| 95.dist_group | 1.93 |
| | (5.57) |
| 96.dist_group | -11.44 |
| | (5.66)** |
| 97.dist_group | -13.11 |
| | (5.47)** |
| 98.dist_group | -7.66 |
| | (5.76) |
| 99.dist_group | -12.95 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | (5.51)** |
| Constant | 57.69 | 75.04 | 52.91 | 62.64 | 54.38 |
| | (16.46)*** | (19.65)*** | (18.04)*** | (20.70)*** | (18.66)*** |
| Chi2 | . | . | . | . | . |
| Df_M | 22 | 42 | 61 | 81 | 100 |
| P | . | . | . | . | . |
| Ll | -1,235.14 | -1,227.79 | -1,204.22 | -1,194.98 | -1,177.50 |
| *N* | 391 | 391 | 391 | 391 | 391 |

* $p$<0.1; ** $p$<0.05; *** $p$<0.01

**Appendix: Social Connectedness versus physical distance across Europe**

The *heat maps* in Figure 5 plot the social network distributions of (the German municipalities of) Kiel in the upper panel and Freiburg im Breisgau in the bottom panel (both marked in red); darker colors refer to higher connectedness/probability of connection. In other words, that graph shows the relative probability of connection (measured by Equation 1) of all European NUTS3[6] locations *j* with two locations *i* (Kiel and Freiburg)[7].

In both examples, the strongest social connections are to geographically adjacent units. however , it also stands out that Freiburg shows stronger connections with the majority of nuts3 locations from neighboring Switzerland and Austria than with (at least three locations) in Easter Germany and just as many connections compared to neighboring states within Germany. more interesting , some nuts3 locations in the Balkans share more connections with this location than, for instance, neighboring Poland and Czech Republic, or Southeast Austria and Hungary. in the upper panel we observe , interestingly, that one location in the Balkans is as connected to that northern German city as to adjacent nuts3 locations in Germany and has even more social connections compared to most of nuts3 locations in Germany.

I will now assess the relationship between geographical distance and the variable of interest between European nuts three locations using Equation 3:

*Equation 3: Empirical Specification*

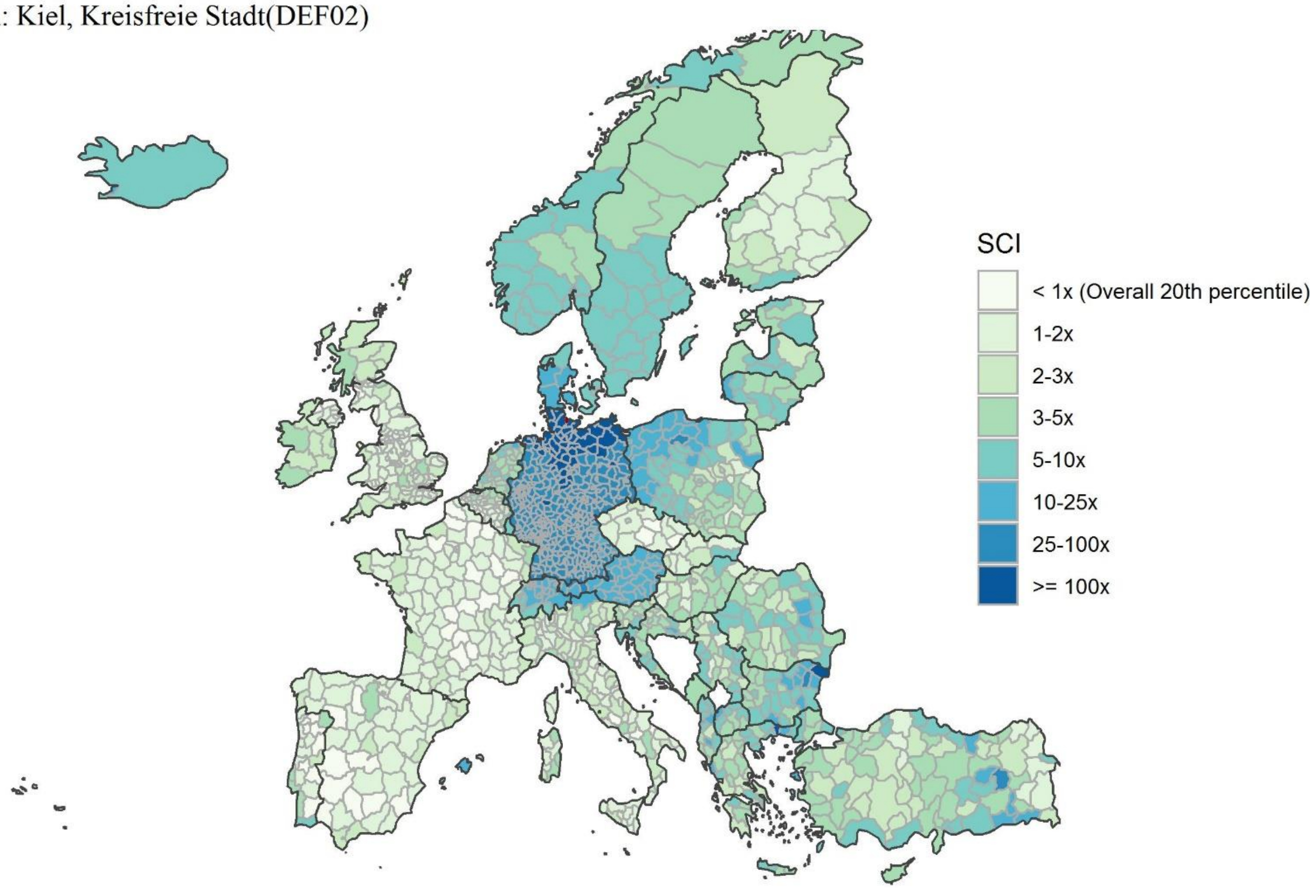
$$\log(SocialConnectedness_{ij}) = \beta_0 + \beta_1 \log\left(d_{ij}\right) + X_{ij} + \delta_i + \delta_j + \varepsilon_{ij}$$

[6] NUTS3 locations outside of Germany might not be municipalities

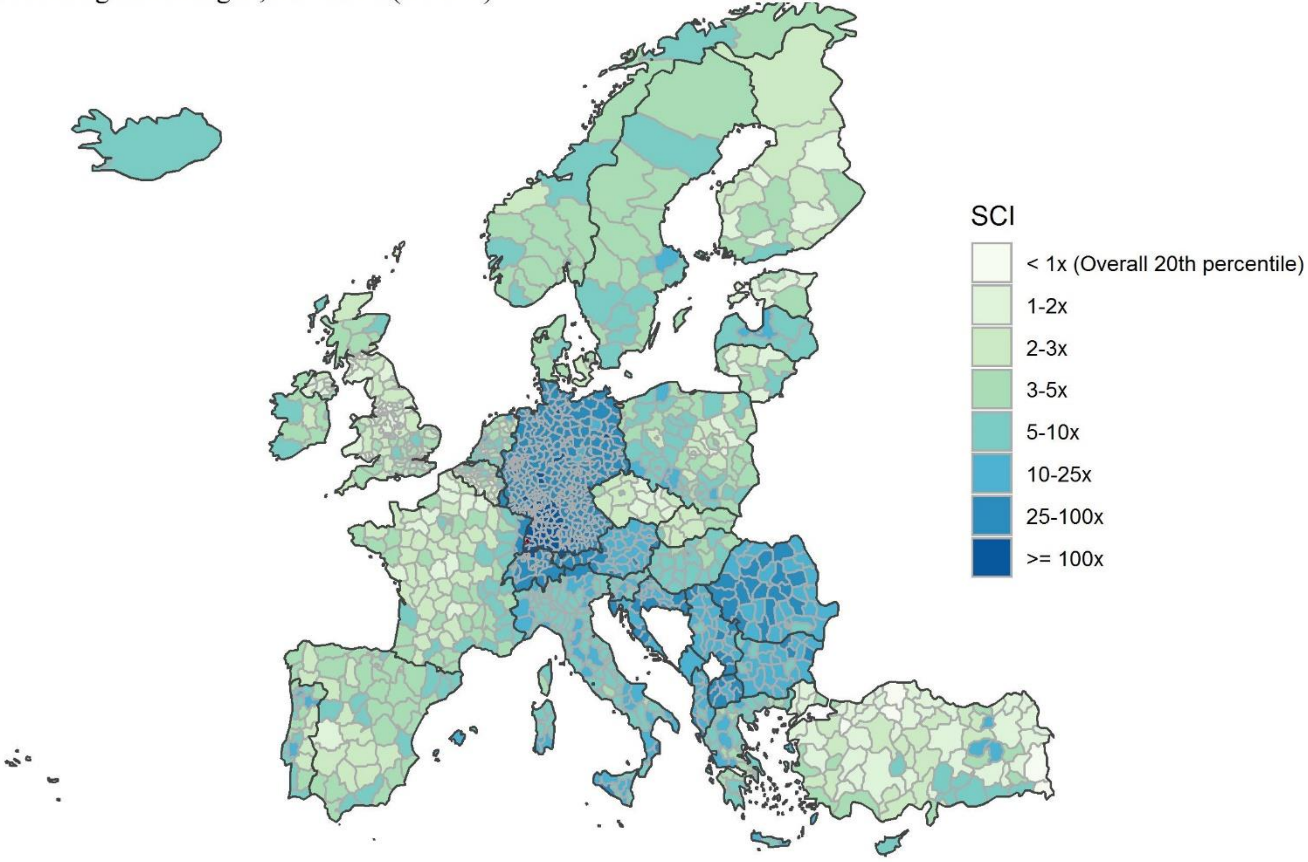
[7] The measures are scaled from the 20th percentile of all *i*, *j* pairs (in Europe).

*Figure 5. Heat maps of the social network distributions of Kiel and Freiburg im Breisgau*

A: Kiel, Kreisfreie Stadt(DEF02)




B: Freiburg im Breisgau, Stadtkreis(DE131)



The unit of observation is a pair of German municipalities (a.k.a. NUTS3 locations). The response variable is the log of Social Connectedness between locations $i$ and $j$ (see Equation 1). The – population weighted – geographic distance is denoted by $\log(d_{ij})$, $\delta_i$ and $\delta_j$ denote fixed effects for locations $i$ and $j$ [8] , which allows to control for average differences of Facebook usage patterns across NUTS3 units, population levels and any other characteristics that vary at the county level. The vector $X_{ij}$ will include measures of dis-/similarity along (at least) the following demographic and socioeconomic factors: education, income, unemployment, language, and industry similarity. For now, I will only look at the relationship with distance.

In Table 4 we observe that a 10% increase in the distance between two locations is associated with a 13.3% decline in the connectedness between those locations[9]. This relationship is comparable to that observed for U.S. county pairs – 14.8% – (Bailey et al. 2018)[10] .

*Table 4. Regression of Social- on Geographic (distance) Connectedness*

| | (1) |
|---|---|
| log_distance | -1.327*** |
| | (0.022) |
| _cons | 15.055*** |
| | (0.151) |
| NUTS3 FEs | Y |
| Number of Observations | 2.311.920 |
| $R^2$ | 0.560 |

Standard errors are double clustered by each region i and region j in a region pair.
Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

[8] The log-linear, the double standard error clustering and population-weighted distance specifications follow previous evidence/scholarship (Bailey et al. 2018, Bailey et al. 2020b)

[9] Similar to *gravity equations* from the trade literature, that type of analysis estimates the equilibrium relationship between geographic distance and social connectedness, in other words, not necessarily a causal effect of one on the other ( Bailey et al. 2018).

[10] however, in order to determine whether distance is a less or more important determinant of social connectedness in Europe than it is in the United States, one needs to compare the percentage of the variation in social connectedness that distance by itself explains, in other words, the percentage not explained by the fixed effects

## Appendix. Results with data from late November 22, 2020

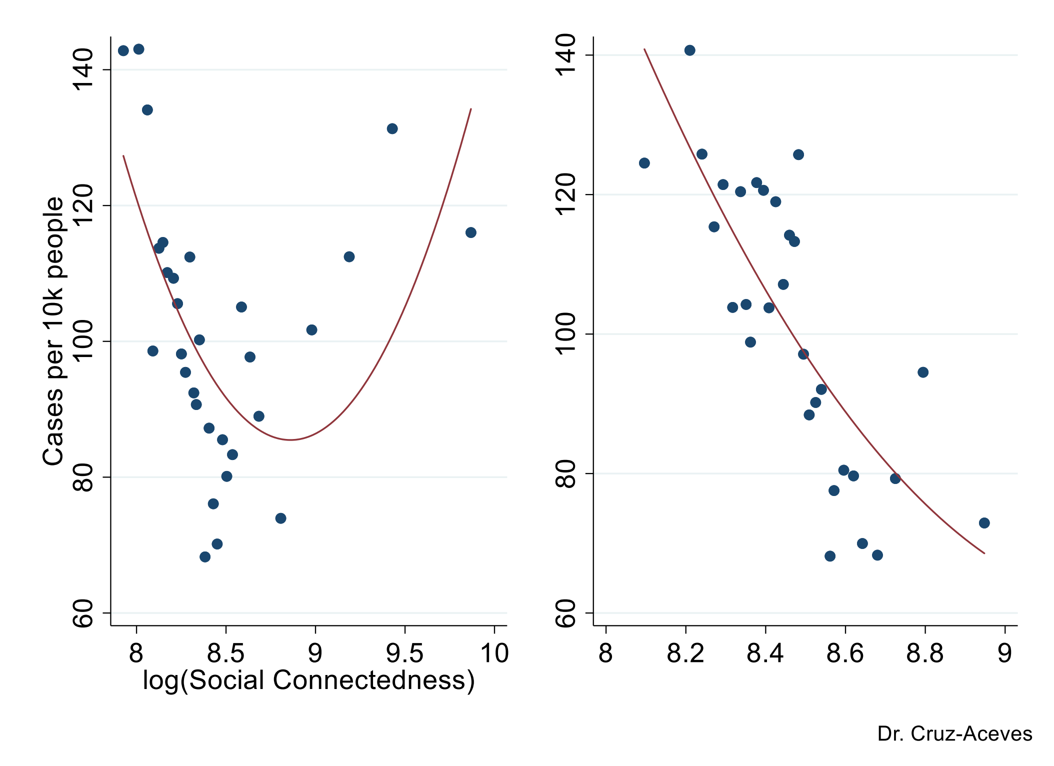
*Figure 6* visualizes results of analysis with data from late November 2020 using binned binscatter plots[11]. In the left panel we observe a U-shaped relationship between COVID-19 prevalence and social ties to Heinsberg (when we use data from March 30, 2020, the direction of the estimated effect seems unchanged, as can be seen in Figure 3).

One concern with interpreting these initial relationships is that they might be picking up other factors that affect the spread of COVID-19 and that are associated with social connectedness. Namely, even after dropping municipalities within 50 miles of Heinsberg, the correlations might be primarily showing geographic distance to Heinsberg (which is related to the number of friendship links to Heinsberg). In the same vein, including social connectedness might not improve predictive power for models that already control for some of these other factors. Having said that, In the right Panel of Figure 6. Binscatter (Heinsberg as unit of observation, data from November 22, 2020), I present a binscatter plot of the relationship between social connectedness to Heinsberg and COVID-19 cases that controls for a number of these possible confounding covariates (in addition to excluding geographically adjacent municipalities)[12]. Specifically I control for income and population density. Conditional on these other factors, the right Panel of *Figure 6* shows a negative relationship between COVID-19 cases as of November 22, 2020 and social connectedness to Heinsberg (again, when we use data from March 30, 2020, the direction of the estimated effect seems unchanged, as can be seen in Figure 3).

[11] this type of graph groups the variable from the X axis into bins of equal size, next, the mean of the x-axis and y-axis variables within each bin is calculated, and finally a scatterplot of these data points is generated, resulting in a non-parametric visualization of the conditional expectation function

[12] I control for the geographic distance between each municipality and Heinsberg non-parametrically by including 20 dummies for percentiles of that distance.

Figure 6. Binscatter (Heinsberg as unit of observation, data from November 22, 2020)




## Appendix. Summary statistics

Table 5. Summary statistics

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| total_cases | 391 | 156.6982 | 231.8249 | 4 | 2756 |
| cases_pe~10k | 391 | 7.776851 | 6.161148 | 0.9172388 | 64.48508 |
| log_sci | 391 | 8.470694 | 0.4239693 | 7.715569 | 10.35587 |
| scaled_sci | 391 | 5356.437 | 3507.68 | 2243 | 31441 |
| pop_2018 | 391 | 193829.4 | 170605.4 | 34270 | 1830584 |
| pop_per_km | 391 | 522.913 | 687.8328 | 36.2 | 4721.9 |
| gdp_per_hab | 391 | 36683.38 | 15965.12 | 16200 | 168000 |
| dist | 391 | 333.1581 | 138.3681 | 57.95338 | 614.6484 |
| log_dist | 391 | 5.691568 | 0.5350272 | 4.059639 | 6.421051 |
| dist_group | 391 | 9.71867 | 5.654373 | 0 | 19 |
| dist_group | | | | | |
| 1 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 2 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 3 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 5 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 6 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 7 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 8 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 9 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 10 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 11 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 12 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 13 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 14 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 15 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 16 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 17 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 18 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |
| 19 | 391 | 0.0511509 | 0.2205877 | 0 | 1 |

## References


Literaturverzeichnis

Aiello, Allison E.; Renson, Audrey; Zivich, Paul N. (2020): Social Media– and Internet-Based Disease Surveillance for Public Health. In: *Annual Review of Public Health* 41 (1), S. 101–118. DOI: 10.1146/annurev-publhealth-040119-094402.

Aref, Samin; Bontcheva, Kalina; Braghieri, Marco; Dignum, Frank; Giannotti, Fosca; Grisolia, Francesco et al. (Hg.) (2020): The Determinants of Social Connectedness in Europe. Social Informatics. Cham: Springer International Publishing.

B. Carey; J. Glanz (2020): Travel from new york city seeded wave of u.s. outbreaks. In: *New York Times*.

Badr, H. S.; B. F. Zaitchik; G. H. Kerr; J. M. Colston; P. Hinson; Y. Chen et al. (2020): Unified COVID-19 Dataset. Online verfügbar unter https://github.com/CSSEGISandData/COVID-19_Unified-Dataset/blob/master/COVID-19.RData, zuletzt geprüft am 24.11.2020.

Bailey, Michael; Cao, Rachel; Kuchler, Theresa; Stroebel, Johannes; Wong, Arlene (2018): Social Connectedness.Measurement, Determinants, and Effects. In: *Journal of Economic Perspectives* 32 (3), S. 259–280. DOI: 10.1257/jep.32.3.259.

Bailey, Michael; Gupta, Abhinav; Hillenbrand, Sebastian; Kuchler, Theresa; Richmond, Robert; Stroebel, Johannes (2020a): International trade and social connectedness. In: *Journal of International Economics*, S. 103418. DOI: 10.1016/j.jinteco.2020.103418.

Bailey, Michael; Gupta, Abhinav; Hillenbrand, Sebastian; Kuchler, Theresa; Richmond, Robert; Stroebel, Johannes (2020b): International Trade and Social Connectedness. Cambridge, MA: National Bureau of Economic Research.

Bengtsson, Linus; Gaudart, Jean; Lu, Xin; Moore, Sandra; Wetter, Erik; Sallah, Kankoe et al. (2015): Using Mobile Phone Data to Predict the Spatial Spread of Cholera. In: *Scientific Reports* 5 (1). DOI: 10.1038/srep08923.

Brueckner, Jan K.; Largey, Ann G. (2008): Social interaction and urban sprawl. In: *Journal of Urban Economics* 64 (1), S. 18–34. DOI: 10.1016/j.jue.2007.08.002.

Chaoqi Yang; Ruijie Wang; Fangwei Gao; Dachun Sun; J. Tang; T. Abdelzaher (2020): Quantifying Projected Impact of Social Distancing Policies on COVID-19 Outcomes in the US. In: *ArXiv*. Online verfügbar unter https://www.semanticscholar.org/paper/ccff2671cd0d02cf0b56ff7cb5edd0703b748429.

Charoenwong, Ben; Kwan, Alan; Pursiainen, Vesa (2020): Social connections with COVID-19-affected areas increase compliance with mobility restrictions. In: *Science advances* 6 (47). DOI: 10.1126/sciadv.abc3054.

Chetty, Raj; Friedman, John; Rockoff, Jonah (2013): Measuring the Impacts of Teachers II. Teacher Value-Added and Student Outcomes in Adulthood. National Bureau of Economic Research.

CORREIA, RION BRATTIG; LI, LANG; ROCHA, LUIS M. (2016): MONITORING POTENTIAL DRUG INTERACTIONS AND REACTIONS VIA NETWORK ANALYSIS OF INSTAGRAM USER TIMELINES. In: Russ B. Altman (Hg.): Pacific Symposium on Biocomputing 2016. Kohala Coast, Hawaii, USA, 4-8 January 2016. Biocomputing 2016, 2015/11/18. Pacific Symposium on Biocomputing; PSB. New Jersey, London, Singapore: WORLD SCIENTIFIC.

Danon, Leon; Ford, Ashley P.; House, Thomas; Jewell, Chris P.; Keeling, Matt J.; Roberts, Gareth O. et al. (2011): Networks and the Epidemiology of Infectious Disease. In: *Interdisciplinary Perspectives on Infectious Diseases* 2011, S. 1–28. DOI: 10.1155/2011/284909.

Focus Online: China um Hilfe gebeten: Heinsberg steht für das größte Problem in der Corona-Krise. Online verfügbar unter https://www.focus.de/politik/deutschland/es-macht-mich-wuetend-china-um-hilfe-gebeten-heinsberg-steht-fuer-das-groesste-problem-in-der-corona-krise_id_11813014.html, zuletzt geprüft am 22.12.2020.

Frankfurter Allgemeine: Am Beispiel von Heinsberg die Pandemie verstehen. Online verfügbar unter https://www.faz.net/aktuell/politik/inland/coronavirus-am-beispiel-von-heinsberg-die-pandemie-verstehen-16699780.html, zuletzt geprüft am 22.12.2020.

Garzon-Alfonso, Cristian C.; Rodriguez-Martinez, Manuel (2018): Twitter Health Surveillance (THS) System. In: 2018 IEEE International Conference on Big Data (Big Data). 2018 IEEE International Conference on Big Data (Big Data), 2018/12: IEEE.

Generous, Nicholas; Fairchild, Geoffrey; Deshpande, Alina; Del Valle, Sara Y.; Priedhorsky, Reid (2014): Global Disease Monitoring and Forecasting with Wikipedia. In: *PLoS Computational Biology* 10 (11), e1003892. DOI: 10.1371/journal.pcbi.1003892.

Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette; Smolinski, Mark S.; Brilliant, Larry (2009): Detecting influenza epidemics using search engine query data. In: *Nature* 457 (7232), S. 1012–1014. DOI: 10.1038/nature07634.

Gittelman, Steven; Lange, Victor; Gotway Crawford, Carol A.; Okoro, Catherine A.; Lieb, Eugene; Dhingra, Satvinder S.; Trimarchi, Elaine (2015): A New Source of Data for Public Health Surveillance. Facebook Likes. In: *Journal of Medical Internet Research* 17 (4), e98. DOI: 10.2196/jmir.3970.

Hamburger Abendblatt: CORONAVIRUS, zuletzt geprüft am 22.12.2020.

J. Coven; A. Gupta (2020): Disparities in mobility responses to covid-19. In: *NYU Stern Working Paper*.

Keeling, Matt J.; Eames, Ken T.D (2005): Networks and epidemic models. In: *Journal of The Royal Society Interface* 2 (4), S. 295–307. DOI: 10.1098/rsif.2005.0051.

Kuchler, Theresa; Russel, Dominic; Stroebel, Johannes (2020a): The Geographic Spread of COVID-19 Correlates with Structure of Social Networks as Measured by Facebook. Münich: CESifo Working Paper Series (CESifo Working Paper Series, CESifo Working Paper Series), zuletzt geprüft am 28.12.2020.

Kuchler, Theresa; Russel, Dominic; Stroebel, Johannes (2020b): The Geographic Spread of COVID-19 Correlates with Structure of Social Networks as Measured by Facebook. Cambridge, Mass.: National Bureau of Economic Research (NBER working paper series, no. w26990).

Kuchler, Theresa; Russel, Dominic; Stroebel, Johannes (2020c): The Geographic Spread of COVID-19 Correlates with Structure of Social Networks as Measured by Facebook. Cambridge, Mass.: National Bureau of Economic Research (NBER working paper series, no. w26990).

LandNRW.de: Wissenschaftsteam erforscht Infektionsgeschehen des Corona-Virus in Heinsberg. Online verfügbar unter https://www.land.nrw/de/pressemitteilung/kreis-heinsberg-wird-zur-erstregion-wissenschaftsteam-um-prof-hendrik-streeck, zuletzt geprüft am 22.12.2020.

M. Bailey; P. Farrell; T. Kuchler; J. Stroebel (2019): Social connectedness in urban areas. Working Paper 26029. In: *National Bureau of Economic Research*.

M. J. Keeling; P. Rohani (2011): Spatial models. In Modeling infectious diseases in humans and animals, pages 232–290.

Mossay, P.; Picard, P. M. (2011): On spatial equilibria in a social interaction model. In: *Journal of Economic Theory* 146 (6), S. 2455–2477. DOI: 10.1016/j.jet.2011.06.009.

Mossong, Joël; Hens, Niel; Jit, Mark; Beutels, Philippe; Auranen, Kari; Mikolajczyk, Rafael et al. (2008): Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. In: *PLoS Medicine* 5 (3), e74. DOI: 10.1371/journal.pmed.0050074.

Newman, M. E. J. (2002): Spread of epidemic disease on networks. In: *Physical Review E* 66 (1). DOI: 10.1103/physreve.66.016128.

P Giuliano; I Rasul (2020): Compliance with social distancing during the covid-19 crisis.

Paolotti, D.; Carnahan, A.; Colizza, V.; Eames, K.; Edmunds, J.; Gomes, G. et al. (2014): Web-based participatory surveillance of infectious diseases. The Influenzanet participatory surveillance experience. In: *Clinical Microbiology and Infection* 20 (1), S. 17–21. DOI: 10.1111/1469-0691.12477.

Pastore y Piontti, Ana; Perra, Nicola; Rossi, Luca; Samay, Nicole; Vespignani, Alessandro (2019): Charting the Next Pandemic: Springer International Publishing.

Peixoto, Pedro S.; Marcondes, Diego; Peixoto, Cláudia; Oliva, Sérgio M. (2020): Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. In: *PLOS ONE* 15 (7), e0235732. DOI: 10.1371/journal.pone.0235732.

Robert Koch Institut: SARS-CoV-2: Fallzahlen in Deutschland, China und weltweit. Online verfügbar unter https://web.archive.org/web/20200303083152/https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html, zuletzt geprüft am 22.12.2020.

Smolinski, Mark S.; Crawley, Adam W.; Baltrusaitis, Kristin; Chunara, Rumi; Olsen, Jennifer M.; Wójcik, Oktawia et al. (2015): Flu Near You. Crowdsourced Symptom Reporting Spanning 2 Influenza Seasons. In: *American Journal of Public Health* 105 (10), S. 2124–2130. DOI: 10.2105/ajph.2015.302696.

T Tully; S Stowe (2020): The wealthy flee coronavirus. vacation towns respond. Stay away. In: *New York Times*.

University of Virginia Library. Online verfügbar unter https://data.library.virginia.edu/interpreting-log-transformations-in-a-linear-model/, zuletzt geprüft am 21.12.2020.

Wesolowski, Amy; Qureshi, Taimur; Boni, Maciej F.; Sundsøy, Pål Roe; Johansson, Michael A.; Rasheed, Syed Basit et al. (2015): Impact of human mobility on the emergence of dengue epidemics in Pakistan. In: *Proceedings of the National Academy of Sciences* 112 (38), S. 11887–11892. DOI: 10.1073/pnas.1504964112.