Predicting Covid-19 Infections Using Multi-layer Centrality Measures

Christine Hedde - von Westernhagen[†]

May 8, 2023

Abstract

One of the most pressing problems of the recent past has been to understand the spread of the Covid-19 virus within and across populations. A popular strategy in investigating spreading phenomena have been network models, as they are able to represent complex social interactions. While previous studies have shown that network centrality measures are able to identify influential spreaders, it is not yet known if they can also be used to predict individual infection risks. However, information about individual risks is vital for the design of targeted interventions, and can give citizens more autonomy over their decisions. This study therefore focuses on the predictive abilities of centrality measures for individual infections. I investigate this issue leveraging the framework of multi-layer networks, which allows to explicitly consider human interaction to take place in different contexts simultaneously. Drawing on large-scale administrative data from the Netherlands, I find that existing centrality measures offer good predictions of relative infection risks, but are only weakly informative about the timing of individual infections. I therefore introduce a new Degree-based multi-layer centrality measure that takes into account the transmission rate of Covid-19 in different contexts. The new measure shows similar predictive performance as existing centrality measures, indicating that centrality measures alone are limited in their potential to predict the timing of infections, but could be used to complement other prediction approaches.

Keywords: Covid-19, multi-layer networks, centrality measures, spreading, infections

[†]*Department of Methodology and Statistics, Utrecht University, Padualaan 14 3584 CH Utrecht, The Netherlands.* Email: c.hedde-vonwesternhagen@students.uu.nl, ORCiD: 0000-0002-4417-9255.

1 Introduction

Since the onset of the Covid-19 pandemic in late 2019, a myriad of studies has attempted to adequately model how the virus spreads within and across populations in order to predict outbreaks and assess interventions (see [1] for an overview). Like for all epidemic diseases, models of Covid-19 have to take into account how members of populations are interconnected. To this end, it is important to consider that human interaction simultaneously takes place within and *across* different areas of life. One way of accommodating this structure is the use of multi-layer network models, where each layer presents a context of interaction [2].

A relevant indicator of spreading capacity has shown to be the centrality of a node within the network. This has been studied for networks representing a single type of interaction [3, 4][5, ch. 10.3], as well as for multi-layer networks where interactions can be of multiple types, modelled by the individual layers [6]. While focusing on spreading capacity enables identification of influential spreaders in a network, the risk of infection for individual nodes cannot directly be inferred. However, the individual infection risk is of practical interest for epidemic scenarios. It can facilitate government interventions specifically targeted at high-risk groups, or allow citizens to receive personalized risk assessments. In the context of Covid-19, several studies already employed multi-layer networks in modelling the spread of the virus [7, 8, 9], but none of them have used the framework to predict individual infections. This study aims at filling this gap, answering the research question:

How well can multi-layer centrality measures predict the infection of individuals with epidemic diseases like Covid-19?

In the past years, definitions have been developed for multi-layer versions of, e.g., PageRank [10], Eigenvenctor [11], and Betweenness centrality [12]. De Domenico et al. [13] then showed that node rankings based on the respective single-layer and multi-layer centralities can differ substantially. Differences between the measures can be attributed to the increased structural complexity of multi-layer networks, resulting in spreading dynamics that cannot be observed in single-layer networks [14, 15].

To assess the performance of a selection of these developed measures in predicting infections, I made use of administrative micro-data of the Dutch population. Administrative population networks bear a novel opportunity to researchers studying social processes, in that they do not suffer from common drawbacks of studies based on surveys or digital trace data [16]. Tapping this source of information thus presents another contribution of this study. Specifically, I used the administrative data to construct a network consisting of a family, household, school, and work layer. On the basis of this network, I simulated the infection process of a Covid-19 epidemic.

The results of the prediction tasks indicate that multi-layer centrality measures can provide a good assessment of a node's relative infection risk. However, the multi-layer measures do not perform substantially better than their single-layer counterparts. Furthermore, neither type of measure exhibits high predictive performance concerning the exact time point of infection. Therefore, I introduced an adaption of multi-layer Degree centrality which incorporates layer-specific weights. Testing this measure on the simulated infection data, I found only little improvement over the existing measures. Applying the measure on PCR-test data of actual Covid infections showed no explanatory power, demonstrating that predicting infections in this epidemic scenario cannot be done based on administrative network data alone.

The paper proceeds as follows: Section 2 provides a theoretical background on the role of centrality in spreading processes, and introduces concepts and notation of multi-layer networks. Section 3 describes the data used in this study in more detail. In Section 4, I outline the methodological procedure to answer the research question. In Section 5, results of the analyses are presented and discussed. I close the paper with concluding remarks, as well as possible opportunities for future research in Section 6.

2 Theoretical Background

2.1 The role of network centrality in spreading processes

Centrality measures are an integral part of characterizing nodes within a network, especially in the social sciences where the concept originated in the late 1940's [17]. Each of the of the measures developed since then intends to capture the relevance of a node in a theoretically distinct way.

When investigating spreading processes in a network of individuals, the outcomes of interest are often aggregate results of the process, such as the final size of the spread, or the spreading capacity of each individual. These outcomes are indispensable in understanding the scale of spreading phenomena. Accordingly, the role of centrality measures in applications of disease, information, or behaviour spreading has mostly been to detect individuals who influence changes in those target variables [3, 4, 5, 6].

Considering single-layer network representations, correlation-based analyses from several studies [3, 4, 5, 6] suggest that most types of centrality measures are associated with final outbreak sizes as well as the spreading caused by individual nodes. However, the findings are highly dependent on the type and size of the network, as well as the type of spreading process which was investigated, implying that there is no general conclusion about which centrality measures perform best. A comprehensive study conducted by De Arruda et al. [4] indicates that for epidemic spreading Degree centrality is usually most highly associated with final outbreak size. [3] also find evidence for the influence of PageRank and Katz centrality on this outcome. That

random walk based measures like PageRank perform well is further corroborated by [5, ch. 10.3]. For the case of multi-layer networks, research on centrality and spreading outcomes is more limited. [6] demonstrated that, generally, measures based only on the direct neighborhood of a node are more predictive of its spreading capacity than measures like PageRank or Betweenness that also incorporate links at longer distances.

While previous studies give a notion about which centrality measures are generally useful in explaining spreading processes, the spreading capacity of a node does not directly translate to its risk of becoming infected. Research on this topic is very limited, with a notable exception being [18], who find quasi equal performance of Degree, Betweenness, and Closeness centrality in predicting infection probabilities. However, the study only looked at a small single-layer network, in which the centrality measures were highly correlated.

Clearly, there is no straightforward guidance on how to relate centrality with Covid-19 infections in a multi-layer network. I therefore look at three prominent measures which have shown to be important in some of the findings presented above: Degree, Eigenvector, and PageRank centrality. This leaves only measures based on calculating the shortest path between all pairs of nodes (e.g., Betweeness and Closeness) unconsidered in this paper, since it is infeasible to compute them on a network of the size presented here. Before addressing the mathematical definitions of the studied centrality measures, the concept of multi-layer networks and their notation is introduced.

2.2 Multi-layer networks and tensorial notation

Multi-layer networks have emerged simultaneously from a variety of research areas, each with the aim to model real-world systems composed of relationships which may differ in their type, strength, and direction. Depending on the restrictions imposed on the network structure, one can distinguish a broad range of sub-variants of the most general multi-layer network. Examples of such restrictions are the ordering of the layers, the presence or absence of nodes in specific layers, the presence of edges between layers, or the weighting of certain edges. Kivela et al. [2] have introduced a unifying mathematical framework for multi-layer networks and provide a thorough introduction to their properties and applications.

A concise way of notation for multi-layer networks is to represent them as tensors[1]. Following the notation of [13], the *adjacency tensor* of rank 4 is defined as $M_{j\beta}^{i\alpha}$, which captures the link of node i in any layer α to node j in any layer β. This formulation allows to explicitly consider multiple types of connections between nodes, which would otherwise get lost in the aggregation of the layers, but are vital to the dynamical properties of the network [19, 20].

In this study, a *multiplex network*[2] is employed to model interpersonal relationships. This variant of a

[1]Helpful visualizations of tensor representations can be found at `https://github.com/manlius/muxViz/blob/master/gui-old/theory/README.md`.

[2]See [2], Table 1, for detailed definitions of various multi-layer network variants.

multi-layer network contains the same set of nodes in each layer. Furthermore, it does not allow for ties between different nodes in different layers, but only between a node in one layer and its copy in all other layers, so called categorical couplings. Concerning the adjacency tensor M, this property can be expressed as $\alpha = \beta$, except when $i = j$. All edges in the network are undirected, representing symmetric relationships. A schematic plot of the network in the context of this study is given in Figure 1.

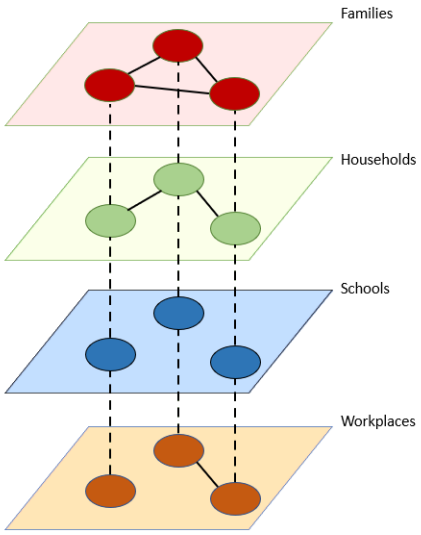

Figure 1: Schematic plot of four-layer network of families, households, schools, and workplaces. The network contains nodes of the same three individuals that are present in each layer. Ties within layers are displayed as bold lines, ties between layers as dashed lines. Dashed inter-layer edges are categorical couplings, i.e., every node is tied to all of its representations in the other layers.

2.3 Centrality in multi-layer networks

Considering the long tradition of centrality measures in network analysis, attempts have been made to establish the concept of centrality in the relatively new area of multi-layer networks. Using a tensor representation as introduced above, [13, Supplementary Note 3] show how some commonly used single-layer measures can be generalized to obtain analogous multi-layer measures.

In the following, I provide the equations for the multi-layer centrality measures investigated in this study. Throughout, I follow the notation of [13] with slight deviations. For more details on the mathematical formulation of multi-layer networks and their properties see also [21].

Degree Centrality. The most commonly inspected property of a node is its degree, i.e., the number of incoming or outgoing ties of a node. Using Einstein notation for conciseness[3], the multi-layer degree centrality (or 'multidegree') of a node i across all layers can be defined as

[3]The notation in Equation 1 is equivalent to $\kappa_i = \sum_{\alpha=1}^{L} K_{i\alpha} u^{\alpha}$.

$$\kappa_i = K_{i\alpha} u^{\alpha}, \tag{1}$$

where $K_{i\alpha}$ is the degree of node i in layer α and u^{α} is a vector of ones used to contract the layer index, i.e., to sum up the degree over all layers. In a multiplex network as used here, this measure is analogous to the strength of a collapsed single-layer network because of the exclusively categorical couplings between layers.

Eigenvector Centrality. Another popular measure of node centrality quantifies a node's importance based on the importance of the nodes it is connected to. In the domain of multi-layer networks this is achieved by finding the eigentensor $\Theta_{i\alpha}$ that satisfies

$$M^{i\alpha}_{j\beta} \Theta_{i\alpha} = \lambda_1 \Theta_{j\beta}, \tag{2}$$

where λ_1 is the leading eigenvalue of $M^{i\alpha}_{j\beta}$. Solving this eigenvalue problem and contracting the layers as in the case of degree centrality gives the overall centrality for each node:

$$\theta_i = \Theta_{i\alpha} u^{\alpha}. \tag{3}$$

PageRank Centrality. PageRank is likely one of the centralities most often applied in practice, since it is underlying to Google's search algorithm [22]. It is based on a random walk with the transition tensor

$$R^{i\alpha}_{j\beta} = r T^{i\alpha}_{j\beta} + \frac{(1-r)}{NL} u^{i\alpha}_{j\beta}, \tag{4}$$

where r is the rate of jumping to a neighboring node and $1-r$ is the rate of the walker being teleported to any other node in the network. The transition tensor $T^{i\alpha}_{j\beta} = M^{k\gamma}_{j\beta} \tilde{D}^{i\alpha}_{k\gamma}$, with $\tilde{D}^{i\alpha}_{j\beta}$ being the inverse of the non-zero entries of the strength tensor of the network, meaning, $\tilde{D}$ normalizes the adjacency tensor M. The number of nodes and layers are given by N and L, respectively. $u^{i\alpha}_{j\beta}$ is a rank-4 tensor of ones. Finally, the PageRank centrality of node i is given by

$$\omega_i = \Omega_{i\alpha} u^{\alpha}, \tag{5}$$

with $\Omega_{i\alpha}$ being the eigentensor of the transition tensor $R^{i\alpha}_{j\beta}$.

3 Data

3.1 Construction of the network dataset

Administrative data provided by *Statistics Netherlands* (CBS) was used to construct a multi-layer network based on the Dutch population. The data provides direct information on different kinds of relationships between individuals registered as residents in the Netherlands. This allowed to assemble an undirected multiplex network that consists of four layers, namely, families, households, schools, and workplaces. Each node within a layer represents a person, and per definition of the multiplex structure, each layer contains the same set of nodes.

Unlike other approaches to obtain social network datasets, like surveys or digital trace data, this source does not suffer from the common problems of non-response bias, selection bias, or social desirability effects. However, it should be pointed out that edges in this network do not measure social interactions directly, but only represent structural relations between individuals. Nonetheless, an edge indicates 'a highly increased probability that two individuals interact socially' [16, p. 146].

While country-level population data bears a highly valuable resource, it also drives up computational demands. I therefore conducted analyses on a regional subset, including all primary schools from the Dutch capital city Amsterdam[4]. Initiating the network construction at the school level was motivated by this context's role as bridge between different households in potential disease transmission [5] [24].

All students within the selected schools became nodes in the final network, where members of the same school year are completely interconnected. For the family relations of the network, I added all parents and siblings of the school students as nodes, and, again, families are completely interconnected. The household layer added nodes living in the same household as the students, regardless of their family relationship. I then retrieved information on the colleagues of all nodes included so far, and added them as new nodes to the network. Ties were established between people working at the same workplace address.

[4] By the beginning of 2020, the municipality (Dutch: gemeente) of Amsterdam has housed 872 thousand people, around 5 percent of the Dutch population [23].

[5] While the same argument could be made for work relationships, workplaces in this dataset often encompass up to a hundred employees, so the contact probability of work colleagues is on average lower than between members of a school year.

Data on individual Covid infections was also provided by CBS, and could be linked to the network data. The coverage of the Covid data ranges from June 2020 until September 2021, which led to the decision to also base the network dataset on the administrative records from 2020. Ethical approval for the use of all datasets was obtained by the Ethical Review Board of the Faculty of Social and Behavioural Sciences of Utrecht University on November 8, 2022 (case numbers 22-1886, 22-1887, 22-1888).

3.2 Network characteristics

After assembling the dataset from the official records, I arrived at a network consisting of about 1 million nodes, with 175 thousand ties in the family layer, 163 thousand ties in the household layer, 1.4 million ties in the school layer, and 4.9 million ties in the workplace layer. Table 1 summarizes further network characteristics.

Table 1: Descriptive network measures by layer and for the aggregate network.

				Degree Distribution				
Layer	Nodes	Ties	Clustering	Percentile 5	Mean	Median	Percentile 95	Std. Dev.
Family	172,448	175,213	0.47	1	2.03	2	4	1.27
Household	164,036	163,834	0.49	1	2.00	2	4	1.26
School	58,087	1,386,862	1.00	16	47.75	45	91	24.08
Work	942,386	4,874,205	0.19	1	10.34	3	31	25.68
Aggregate	1,067,879	6,442,695	0.33	1	12.07	3	62	26.80

Note: Instead of minima and maxima, the lower and upper five percentiles are given for the degree distribution as to not disclose information at the individual level. The median value applies in each layer to > 10 individuals and does thus not pose a disclosure risk.

Comparing the network sizes across layers, one can clearly observe the dominance of work relationships, which is also present in the complete population data, and results from people being employed at large companies. In assembling the original population dataset, the number co-worker ties in such companies was limited to a maximum of 100 which were sampled based on living in the same (or a close-by) location [25, pp. 7–8]. This procedure resulted partly in directed relationships, i.e., i being the co-worker of j but not the other way around. I reconstructed these reciprocities to obtain an undirected network, which lead to some nodes having a degree > 100.

Accordingly, the degree distribution of the work layer shows large variation and is highly right-skewed. The distributions in the household, family, and school layers are much less skewed than in the work layer, but the school layer exhibits a similar amount of variation.

The perfect clustering of the school layer is caused by the design that all students within a year are completely connected. For the other layers, there is a medium degree of clustering present, meaning, in about half of the cases where two people are linked to a third person, these two people are also connected.

Overall, the individual layers show quite distinct structural characteristics. This insight highlights how aggregation can discard a lot of structural variation that may be vital to spreading behaviour.

4 Methods

4.1 Epidemic modeling

In order to obtain the outcome of interest for this study—individual infections—I first simulated a Covid-19 epidemic on the network data. As commonly employed in epidemiology, I used a SIR model (see, e.g., [15, Section 3]), which captures the infection states in a population over time across three groups: **S**usceptible, **I**nfected, or **R**ecovered, where recovered individuals are removed from the population. The model can be expressed as a set of differential equations:

$$\frac{dS}{dt} = -\beta IS, \tag{6}$$

$$\frac{dI}{dt} = \beta IS - \gamma I, \tag{7}$$

$$\frac{dR}{dt} = \gamma I, \tag{8}$$

where β indicates the infection rate, i.e., the transition rate $S \to I$, and γ the recovery rate, i.e., the transition rate $I \to R$. The state variables S, I, R define the total number of individuals in the respective state, together summing up to the population size N.

While, traditionally, the model assumes a homogeneous mixing of the population, when applied on a network, it is evaluated at each individual node based on the current state of its neighbors. This represents a more realistic scenario than the possibility of getting in contact with any individual of the population at any time. Instead of a fixed infection rate β, the multi-layer framework furthermore allows that each layer gets assigned its own infection rate τ_l, with $l \in \{family, household, school, work\}$. These infection rates were then applied in generating new infections at weekly time steps t.

The values for τ_l were based on the so called secondary attack rate. This rate expresses the share of infections originating from an initial infection among the number of possibly infected individuals in a group over fixed a period of time, usually one to two weeks. Based on the literature on secondary attack rates of Covid-19 in household contexts, $\tau_{household}$ was set to 0.20 [26, 27, 28]. This means that within one week after one household member got infected, 20 percent of the household members would be infected. The same studies suggest that infections originating from children are less common, and τ_{school} was then set to 0.10.

Since family relations in this network do not necessarily imply living in the same household, $\tau_{family} = 0.15$ was set somewhat lower than $\tau_{household}$. Finally, infections in workplaces were assumed to be on average the least likely, expressed by $\tau_{work} = 0.05$.

To arrive at a vector of individual infection probabilities for the nodes in a layer l in week t based on the previously infected direct neighbors, I used the Reed-Frost (or chain-binomial) model [29, 30]:

$$\boldsymbol{\tau'}_{lt} = 1 - (1 - \tau_l)^{\mathbf{\Gamma_{t-1}}}, \tag{9}$$

where $\mathbf{\Gamma_{t-1}}$ is a vector with the number of infected neighbors of each node. Note that if a node became infected in one layer, it also changed its infection status in all other layers. This also implies that infection probability was higher for individuals with links in multiple layers.

The recovery time γ did not differ by layer and was sampled from a Weibull distribution with shape $\eta = 1$ and scale $\lambda = 5$ on a scale of days [31]. Since the model was evaluated at weekly intervals, the recovery status $R_{it} \in \{0, 1\}$ of a node was changed to $R_{it} = 1$ if $\sum_{t=1}^{T} I_{it-1} \times 7 > \gamma$, i.e., if the number of days infected was greater than the recovery time. In the following analyses, the model was run $k = 500$ times with differing random seed nodes from which the spread started. I set the number of seed nodes to 10 per simulation, since it is realistic to assume that the virus entered the Dutch population via more than a single infected person. This also decreased the dependence of the epidemic process on the position of the seed nodes in the network.

It should be noted that the model was constructed for a setting in the early phases of the Covid-19 pandemic, without the possibility of vaccination, or considering differing infectiousness of virus subvariants. It was also outside the scope of this study to account for the effects of interventions like mobility restrictions. Descriptive results of the simulations are given in Appendix A.

4.2 Prediction of infections using centrality measures

Before relating the centrality measures to the timing of Covid-19 infections, I inspected the Pearson correlations between the different measures. This enabled a first insight into qualitative differences between the measures and allowed to identify potential problems of multicollinearity in the subsequent analyses.

I then obtained Spearman's rank correlations ρ of the time point of infection of the nodes and the value of their centrality for the respective multi-layer and single-layer measures. This analysis was analogous to the procedure of most studies investigating the relationship of centrality measures and outbreak size or spreading capacity. In this study, however, it presented merely the first step in relating centrality to spreading outcomes. The correlation coefficient for one simulation k of the epidemic was defined as

$$\rho_k = \frac{cov(R(TTI_k), R(Cent))}{\sigma_{R(TTI_k)}\sigma_{R(Cent)}}, \tag{10}$$

with $cov(R(TTI_k), R(Cent))$ being the covariance of the *ranks* of the time to infection and the respective centrality, and σ being their standard deviations. To arrive at an overall estimate of ρ, I retrieved correlation coefficients for each of the k simulations, which were then averaged after applying a Fisher z-transformation [32]. This transformation was desirable since sampling distributions of correlations do not follow normality. The transformation rescaled the estimates such that they approximated a normal distribution, which then allowed to construct appropriate confidence intervals.

In a next step, I predicted individual infection risk over time based on a person's centrality. Unlike the correlation analysis, this enabled accounting for multivariate relationships between the centrality measures. To this end, Cox proportional hazards models were used to estimate the infection risk over time. The model is commonly expressed as

$$h_i(t|x_i) = h_0(t)e^{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots}, \tag{11}$$

where $h_i(t|x_i)$ is the so called hazard rate of an individual conditional on its values for the predictor variables $x_{i1}, x_{i2}, ...$ This can be interpreted as the instantaneous risk of an individual experiencing an event—here: infection—given that they did not experience it up until this point. The baseline hazard $h_0(t)$ is the hazard function for a person whose predictor variables are all zero. The coefficients $\beta_1, \beta_2, ...$ were estimated using the partial maximum likelihood method [33], which takes into account the number of right-censored observations—i.e., individuals for which no event could be observed in the given time.

The models include the introduced centrality measures as linear combinations up to the third order polynomial as predictors. Across-variable interactions were also considered, resulting in 25 models in total. The exact predictor combinations are displayed together with the results in Table 4. The models were run separately for single- and multi-layer versions of the centraliaty measures to allow for comparisons between those. I evaluated the performance of the models using a variant of the Concordance index (C-index), which is weighted to account for censored observations as proposed by [34]. The C-index expresses the degree to which the estimated individual risks align with the order in which infections occur. The index ranges from 0, meaning the risks are completely inverse to the order of infections, to 1, indicating perfect alignment of infection risk and order of infections.

Secondly, I used linear regression models, including the same predictor combinations as in the Cox models. This modeling approach followed from the intent to obtain an estimate of the actual time until an individual gets infected, as opposed to the infection risk at any point in time, as achieved by the Cox models. Model performance was assessed using adjusted R^2 and root-mean-square error (RMSE) values. Like for the rank correlations ρ, R^2, RMSE values, and C-indices, were calculated for each epidemic simulation and then pooled into one estimate by taking the mean. Appropriate transformations to normality of the R^2 values [35] and C-indices[6] were performed before averaging over simulations, following the same reasoning as mentioned for the rank correlations above.

4.2.1 Exploring τ-weighted multi-layer Degree centrality as alternative predictor

Based on the insights gained from the previous steps, I developed an adaption of multi-layer Degree centrality to increase predictive performance regarding the time until infection. Instead of a simple summation across layers, as expressed by the vector u^α in Equation 1, a layer-dependent weighted sum was calculated by this new measure. Each weight is a scalar that corresponds to the infection rate τ_l in the respective layer. In tensor notation, τ-weighted multi-layer Degree centrality is defined as

$$\kappa_i = K_{i\alpha}\tau^\alpha, \tag{12}$$

where τ^α is a vector containing the layer specific infection rates, replacing the 1-vector u^α in contracting the layer index. This measure thus explicitly accounts for parameters of the epidemic scenario, and leverages the multi-layer structure of the network. By including it as a predictor in linear regression models, I tested the adapted centrality measure on both the previously simulated epidemics as well as using the actual Covid-19 PCR-test data. For completeness, I also ran the Cox regressions again on the simulated data including this new measure. Again, polynomials up to the third order were considered as predictor variables, and performance was evaluated using adjusted R^2, RMSE, and C-index values.

4.3 Use and adaption of software

All data manipulation and analyses were carried out in *R*, version 4.1.3 [36]. Relevant software packages were `igraph` [37] for operations on single-layer networks and `muxViz` [38] for multi-layer networks.

In order to compute the Eigenvector and PageRank multi-layer centralities on this large-scale network, I adapted existing functions of the `muxViz` package. Instead of obtaining the leading eigenvector for these

[6]Since C-indices can be seen as a type of (rank) correlation, the same transformation as for the ρ values was used.

centralities algebraically, I used the Power method, which obtains the largest eigenvalue and corresponding eigenvector in an iterative procedure, and is far more efficient when working with large sparse matrices. The Power method works by multiplying a diagonalizable matrix, here, the supra-adjacency matrix obtained from $M_{j\beta}^{i\alpha}$[7], with an initial guess for the eigenvector of the dominant eigenvalue. The result of the multiplication then replaces the previous approximation of the eigenvector. These steps are repeated until convergence is reached.

A pull-request to incorporate this approach into the original package will be issued on GitHub[8] by May 15, 2023. The code to reproduce the analyses in this study and instructions on how to obtain the data will also be openly available on GitHub[9] by that date.

5 Results

5.1 Correlations between centrality measures

The results of the first correlation analysis revealed some interesting insights into the association between both multi- and single-layer centralities (Table 2). Across both network types, a similar pattern emerged: Neither Degree nor PageRank centrality were correlated notably with Eigenvector centrality. This is likely due to Eigenvector centrality allocating most of the centrality to few highly connected nodes, resulting in little variation within the measure. The correlation of Degree and PageRank was in both cases substantial, with $r_{multi} = 0.78$ and $r_{single} = 0.59$.

Table 2: Pearson correlations between different centrality measures.

		Multi-layer			Single-layer		
		Degree	Eigenvector	PageRank	Degree	Eigenvector	PageRank
Multi-layer	Degree	1.00					
	Eigenvector	0.09	1.00				
	PageRank	0.78	0.01	1.00			
Single-layer	Degree	1.00	0.09	0.78	1.00		
	Eigenvector	0.07	0.96	0.01	0.07	1.00	
	PageRank	0.59	0.01	0.88	0.59	0.01	1.00

Looking at the correlations between the two centrality types (lower left of Table 2), revealed that the respective centrality counterparts are very similar to one another, with $r = 0.96$ for Eigenvector and $r = 0.88$ for PageRank centrality. Since Degree centrality in the multi-layer definition corresponds to the weighted Degree in the single-layer representation, those measures were perfectly correlated.

[7] This is a matrix with dimensions $(N \times L) \times (N \times L)$ and the intra-layer connections on the diagonal, i.e., the result of 'unfolding' the tensor M into two dimensions.

[8] `https://github.com/manlius/muxViz/pulls`

[9] `https://github.com/christine-hvw/thesis_disclosed`

These results indicate that there might not be a great benefit in using the multi-layer versions of centrality, since they seem to measure similar properties as the single-layer versions. However, the later multivariate analyses provided more insight on this question.

5.2 Rank correlations between centralities and time to infection

When correlating the centrality ranks of nodes with their time until infection, all types of multi-layer centralities showed a negative association according to Spearman's ρ (Table 3). That is, a higher centrality on any of the measures is indicative of becoming infected with Covid-19 earlier in the epidemic. Among the measures, Degree and Eigenvector centrality had the largest association with $\rho = -0.33$ and $\rho = -0.24$, respectively. PageRank centrality followed closely after, with $\rho = -0.14$.

For the correlations using the single-layer measures, the same ordering emerged. However, the correlations of Eigenvector and PageRank centrality were slightly larger, with $\rho = -0.32$ and $\rho = -0.24$, respectively. Again, Degree centrality is defined identically for either network type and thus had the same value in both cases. Across all results, confidence intervals were extremely narrow, with the limits being largely the same up to the third decimal. This shows that there was little dependence of the infection process on the seed nodes of the epidemic.

Table 3: Spearman's ρ of centrality measures and time to infection averaged across epidemic simulations.

		Mean	95% CI
Multi-layer	Degree	-0.33	[-0.333,-0.333]
	Eigenvector	-0.24	[-0.243,-0.242]
	PageRank	-0.14	[-0.144 -0.144]
Single-layer	Degree	-0.33	[-0.333,-0.333]
	Eigenvector	-0.32	[-0.322,-0.322]
	PageRank	-0.24	[-0.240,-0.240]

Aligning with the correlations between centrality measures presented above, these results showed no support for multi-layer measures being better predictors of infections than their single-layer versions—but rather the opposite. Furthermore, while a relationship between the respective centralities and the time to infection could be detected in both network structures, the correlations were far lower than what has previously been observed for outbreak size or spreading capacity of a node, where ρ often ranges between $0.7-0.9$ (see references Section 2.1). This is an important finding, as it could indicate that the goals of possible interventions—preventing spread or preventing infections—cannot necessarily be achieved by the same means. It also hints at possible limitations of previous studies in the types and sizes of networks considered. In the following, I present findings for the predictive ability of the centrality measures regarding risks and timing of infections, taking into account also multivariate associations between the centrality measures.

5.3 Cox proportional hazards models

Upon inspecting the results of the Cox models, more differentiated insights into individual infection risks could be gained. Table 4 presents the C-indices for all models. Since the estimates' confidence intervals resulting from the simulations were again very narrow, they were omitted from the main text, and included in Appendix B, Table 8 instead.

Table 4: Concordance indices for Cox proportional hazards models by type and order of included centrality measure.

	Degree			Eigenvector			PageRank			C-index	
Model	x^1	x^2	x^3	x^1	x^2	x^3	x^1	x^2	x^3	Multi-layer	Single-layer
1	x									0.84	0.84
2	x	x								0.83	0.83
3	x	x	x							0.84	0.84
4				x						0.73	0.78
5				x	x					0.73	0.78
6				x	x	x				0.74	0.78
7							x			0.47	0.73
8							x	x		0.47	0.73
9							x	x	x	0.47	0.73
10	x			x						0.84	0.84
11	x						x			0.83	0.84
12				x			x			0.47	0.73
13	x			x			x			0.83	0.84
14	x	x		x	x					0.83	0.84
15	x	x					x	x		0.84	0.84
16				x	x		x	x		0.47	0.73
17	x	x		x	x		x	x		0.84	0.84
18	x	x	x	x	x	x				0.84	0.84
19	x	x	x				x	x	x	0.85	0.85
20				x	x	x	x	x	x	0.47	0.73
21	x	x	x	x	x	x	x	x	x	0.85	0.85
22	x	i		x	i					0.84	0.84
23	x	i					x	i		0.84	0.81
24				x	i		x	i		0.47	0.73
25	x	i	i	x	i	i	x	i	i	0.84	0.81

Note: An x denotes terms composed of a single variable, i denotes two- and three-way interactions between multiple variables.

For the multi-layer measures, the best performance according to a C-index of 0.85 was achieved by Models 19 and 21, including the third-order polynomials of Degree and PageRank centrality, plus those of Eigenvector centrality, respectively. However, almost all models including some term of Degree centrality performed quasi equally, with a C-index of $0.83-0.84$. This means, when choosing two random individuals in the network, the models could predict correctly who gets infected first in the epidemic in more than 80 percent of the cases. Models 4-6 including only transformations of Eigenvector centrality showed similar results with a C-index of $0.73-0.74$ For models including PageRank, or PageRank and Eigenvector centrality, performance was

substantially worse, with a C-index of 0.47 indicating a risk prediction slightly worse than choosing the the ordering of infections at random.

The models including the single-layer versions of the centralities resulted in a very similar performance as compared to the multi-layer measures. However, here, even PageRank centrality had a C-index of ~ 0.7. Also Eigenvector centrality fared slightly better than in the multi-layer case. While the good performance in risk predictions across most of the measures is a promising result, these risks are only *relative* and might be of limited use in practice. In contrast, the following findings from linear regression uncovered whether also the timing of the infections could be predicted accurately.

5.4 Linear regression models

As the results of the linear regression models in Table 5 illustrate, predicting the exact timing of infections was a more difficult endeavour than predicting relative infection risks. Again, confidence intervals of all estimates can be found in Appendix B, Table 9.

Table 5: Adjusted R-squared and root-mean-square error for linear regression models by type and order of included centrality measure.

	Degree			Eigenvector			PageRank			Multi-layer		Single-layer	
Model	x^1	x^2	x^3	x^1	x^2	x^3	x^1	x^2	x^3	Adj. R2	RMSE	Adj. R2	RMSE
1	x									0.06	2.50	0.06	2.50
2	x	x								0.11	2.44	0.11	2.44
3	x	x	x							0.12	2.43	0.12	2.43
4				x						< 0.01	2.59	< 0.01	2.59
5				x	x					< 0.01	2.59	< 0.01	2.58
6				x	x	x				< 0.01	2.59	< 0.01	2.58
7							x			0.01	2.58	0.01	2.58
8							x	x		0.03	2.56	0.02	2.57
9							x	x	x	0.04	2.54	0.03	2.55
10	x			x						0.06	2.50	0.06	2.50
11	x						x			0.11	2.44	0.08	2.49
12				x			x			0.01	2.58	0.01	2.58
13	x			x			x			0.11	2.44	0.08	2.49
14	x	x		x	x					0.11	2.44	0.11	2.44
15	x	x					x	x		0.14	2.40	0.13	2.42
16				x	x		x	x		0.03	2.55	0.02	2.56
17	x	x		x	x		x	x		0.14	2.39	0.13	2.41
18	x	x	x	x	x	x				0.12	2.43	0.12	2.43
19	x	x	x				x	x	x	0.15	2.39	0.14	2.40
20				x	x	x	x	x	x	0.04	2.53	0.04	2.54
21	x	x	x	x	x	x	x	x	x	0.15	2.39	0.14	2.40
22	x	i		x	i					0.07	2.50	0.07	2.50
23	x	i					x	i		0.12	2.43	0.09	2.47
24				x	i		x	i		0.01	2.58	0.01	2.58
25	x	i	i	x	i	i	x	i	i	0.12	2.43	0.09	2.47

Note: An x denotes terms composed of a single variable, i denotes two- and three-way interactions between multiple variables.

According to R^2 and RMSE values, the best performing models using the multi-layer measures were again 19 and 21, including the third order polynomials of at least Degree and PageRank centrality as predictors. Nonetheless, in absolute terms they only explained 15 percent variance in the outcome, despite being among the most complex models. A range of simpler models obtained quasi the same performance as those models, all of which including a squared term of Degree centrality and a higher order combination of other the variables. PageRank centrality alone showed low explanatory power, with a maximum of $R^2 = 0.06$. Even worse was the performance of Eigenvector centrality alone, manifested in all $R^2 < 0.01$. Multiplicative terms between the predictors (Models 22-25) did not arrive at better results than using within-predictor transformations. Looking at the RMSE, the prediction error across models was found to range from 2.40 to 2.59 weeks on average.

When comparing these findings to the results for the single-layer measures, prediction performance was essentially the same, with only slight differences for the case of models including PageRank centrality. In general, it could be seen that Eigenvector and PageRank did not contribute much to the prediction on top of Degree centrality. Based on this insight, I tested a new version of this measure which made use of the multi-layer network representation in this epidemic scenario.

5.4.1 τ-weighted multi-layer Degree centrality

Applying the new measure as defined in Section 4.2.1 in the same regression tasks as above yielded the results presented in Table 6. Confidence intervals are supplied in Appendix B, Table 10.

Table 6: Adjusted R-squared, root-mean-square error, and C-index for models including tau-weighted multi-layer Degree centrality.

Model	x^1	x^2	x^3	Adj. R2	RMSE	C-index
1	x			0.10	2.45	0.85
2	x	x		0.11	2.45	0.85
3	x	x	x	0.11	2.43	0.85

I found that the linear version of the measure (Model 1) lead to better performance in predicting the timing of infections compared to the existing version of Degree centrality. However, the difference was small, with $R^2 = 0.10, RMSE = 2.45$ instead of $R^2 = 0.06, RMSE = 2.50$. Furthermore, no additional improvement was found when including the polynomials of the measure as predictors, rendering it quasi equivalent to the existing measure. Accordingly, also the C-index from the Cox regressions was of similar size (0.85 instead of 0.84).

Nonetheless, I tested τ-weighted multi-layer Degree centrality also on the actual Covid-19 test data. This resulted in $R^2 < 0.005$ and $RMSE = 10.88$ across all models, which were equally defined as in Table 6. This

result does not lend support to the new centrality measure being suited to predict Covid-19 infections as they occurred in the Dutch population, which is likely also the case for the other centrality measures investigated in this paper. This implies that administrative network data may not be not sufficient to derive the actual infection process.

6 Conclusion

In this study, I set out to investigate the predictive ability of centrality measures regarding the timing of Covid-19 infections. Drawing on large-scale administrative data, I found that, among the investigated measures, Degree centrality was best suited to predict the timing of infections. However, in absolute terms, its performance showed to be quite limited, with the predicted time of infection deviating on average around 2.5 weeks from the actual time point in the simulated epidemic scenario. This result could not be substantially improved by an adaption of Degree centrality which applied different weights to the network layers corresponding to their specific disease transmission rates. Nonetheless, relative infection risks could be identified well, especially using Degree and Eigenvector centrality.

Across all analyses, findings did not provide evidence for the multi-layer versions of centrality measures being better predictors of infections than their single-layer definitions. This raises the question in how far the more complex multi-layer representation of administrative social networks can contribute to understanding epidemic infection scenarios better.

An additional analysis using data of actual Covid-19 infections derived from positive PCR-tests demonstrated that the epidemic model as specified here to simulate infections did not offer a good representation of the real infection process. While this could lie partly in the specification of the model, it likely resulted from the administrative data not reflecting social contacts accurately enough. Future research should investigate this issue more closely before conducting similar analyses based on this data.

Given that a more representative contact network can be established, centrality measures still present an opportunity to augment models of infection prediction. Combining them with other individual characteristics related to infection risk, e.g., age or occupation, might result in a more powerful tool to inform targeted government interventions or provide guidelines for individuals.

Furthermore, only three prominent centrality measures out of many possible ones were studied. Especially, shortest-path-based measures like Betweenness or Closeness centrality should be investigated as well. However, to do this at the scale of population data may not be feasible, or would at least require substantial computational resources.

Another avenue for improving predictions lies in the use of more flexible methods like Neural Networks.

These could also discover other important predictor transformations and combinations of the centrality measures, even though it may come at the cost of reduced interpretability and more computational effort.

Apart from improving on the prediction task, more fundamental work is needed to understand why certain measures perform better or worse than others. For the simulated epidemics, the higher performance of Degree centrality can partly be attributed to the explicit inclusion of an individual's direct neighbourhood in generating new infections. Still, it would be valuable to derive exactly how the centrality measures relate to epidemic outcomes.

Appendix

A: Descriptive statistics of epidemic simulations

In total, 500 simulations of a SIR epidemic were performed on the network (details in Section 4.1). In 14 out of 500 simulations (2.8%), the epidemic threshold was not reached, i.e., only a marginal number of the nodes (here max. 24) was infected before the spreading stopped. These simulations were not considered in the main analyses outlined in Section 4.2, nor in the statistics in Table 7 below.

Table 7: Descriptive statistics of simulated epidemic SIR process on the network.

	Min.	Mean	Median	Max	Std. Deviation.
Duration (weeks)	29	35.32	35	50	3.28
No. infected (%)	29.34	29.54	29.54	29.79	0.07
Avg. time to infection (weeks)	11.70	14.24	13.98	20.55	1.32
Infections per node (%)	0.00	29.54	15.64	100.00	31.15

Note: The 14 removed simulations are not accounted for in the statistics.

The minimum of 0 infections per node means that some nodes were not infected in any of the simulations. However, this was the case only for 349 nodes, i.e., 0.03% of all nodes.

B: Additional statistics of prediction models

Table 8: Mean and confidence intervals of Concordance indices of Cox proportional hazard models across simulations.

	Multi-layer		Single-layer	
Model	Mean	95% CI	Mean	95% CI
1	0.84	[0.834, 0.834]	0.84	[0.838, 0.838]
2	0.83	[0.832, 0.832]	0.83	[0.832, 0.832]
3	0.84	[0.838, 0.839]	0.84	[0.838, 0.838]
4	0.73	[0.726, 0.726]	0.78	[0.778, 0.778]
5	0.73	[0.732, 0.732]	0.78	[0.778, 0.778]
6	0.74	[0.737, 0.738]	0.78	[0.778, 0.778]
7	0.47	[0.471, 0.471]	0.73	[0.727, 0.727]
8	0.47	[0.470, 0.470]	0.73	[0.726, 0.726]
9	0.47	[0.471, 0.471]	0.73	[0.725, 0.726]
10	0.84	[0.838, 0.838]	0.84	[0.841, 0.841]
11	0.83	[0.835, 0.835]	0.84	[0.844, 0.844]
12	0.47	[0.471, 0.471]	0.73	[0.728, 0.728]
13	0.83	[0.834, 0.834]	0.84	[0.844, 0.844]
14	0.83	[0.832, 0.832]	0.84	[0.836, 0.836]
15	0.84	[0.843, 0.843]	0.84	[0.840, 0.840]
16	0.47	[0.470, 0.471]	0.73	[0.727, 0.727]
17	0.84	[0.843, 0.843]	0.84	[0.840, 0.840]
18	0.84	[0.839, 0.839]	0.84	[0.842, 0.842]
19	0.85	[0.846, 0.846]	0.85	[0.847, 0.847]
20	0.47	[0.471, 0.471]	0.73	[0.726, 0.726]
21	0.85	[0.846, 0.846]	0.85	[0.847, 0.847]
22	0.84	[0.838, 0.838]	0.84	[0.841, 0.841]
23	0.84	[0.840, 0.840]	0.81	[0.808, 0.808]
24	0.47	[0.471, 0.471]	0.73	[0.728, 0.728]
25	0.84	[0.839, 0.840]	0.81	[0.808, 0.808]

Note: Full model specifications are displayed in main text Table 4.

Table 9: Mean and confidence intervals of adjusted R^2 and RMSE values of linear regression models across simulations.

	Multi-layer				Single-layer			
Model	Adj. R2	95% CI	RMSE	95% CI	Adj. R2	95% CI	RMSE	95% CI
1	0.06	[0.064, 0.065]	2.50	[2.50, 2.51]	0.06	[0.064, 0.064]	2.50	[2.504, 2.505]
2	0.11	[0.110, 0.112]	2.44	[2.44, 2.45]	0.11	[0.111, 0.111]	2.44	[2.441, 2.442]
3	0.12	[0.118, 0.120]	2.43	[2.42, 2.44]	0.12	[0.119, 0.119]	2.43	[2.430, 2.430]
4	< 0.01	[0.002, 0.002]	2.59	[2.58, 2.59]	< 0.01	[0.002, 0.002]	2.59	[2.587, 2.587]
5	< 0.01	[0.003, 0.003]	2.59	[2.58, 2.59]	< 0.01	[0.003, 0.003]	2.58	[2.585, 2.585]
6	< 0.01	[0.003, 0.003]	2.59	[2.58, 2.59]	< 0.01	[0.003, 0.003]	2.58	[2.585, 2.585]
7	0.01	[0.007, 0.007]	2.58	[2.58, 2.59]	0.01	[0.006, 0.006]	2.58	[2.582, 2.582]
8	0.03	[0.026, 0.026]	2.56	[2.55, 2.56]	0.02	[0.018, 0.018]	2.57	[2.566, 2.566]
9	0.04	[0.041, 0.042]	2.54	[2.53, 2.54]	0.03	[0.033, 0.033]	2.55	[2.547, 2.547]
10	0.06	[0.065, 0.065]	2.50	[2.50, 2.51]	0.06	[0.065, 0.065]	2.50	[2.504, 2.504]
11	0.11	[0.111, 0.112]	2.44	[2.44, 2.45]	0.08	[0.075, 0.075]	2.49	[2.490, 2.490]
12	0.01	[0.009, 0.009]	2.58	[2.57, 2.58]	0.01	[0.007, 0.007]	2.58	[2.580, 2.580]
13	0.11	[0.111, 0.113]	2.44	[2.44, 2.45]	0.08	[0.076, 0.076]	2.49	[2.489, 2.489]
14	0.11	[0.112, 0.114]	2.44	[2.43, 2.44]	0.11	[0.113, 0.113]	2.44	[2.438, 2.438]
15	0.14	[0.143, 0.145]	2.40	[2.39, 2.40]	0.13	[0.128, 0.129]	2.42	[2.417, 2.417]
16	0.03	[0.028, 0.028]	2.55	[2.55, 2.56]	0.02	[0.021, 0.021]	2.56	[2.562, 2.562]
17	0.14	[0.143, 0.146]	2.39	[2.39, 2.40]	0.13	[0.130, 0.130]	2.41	[2.415, 2.415]
18	0.12	[0.121, 0.123]	2.43	[2.42, 2.43]	0.12	[0.122, 0.122]	2.43	[2.427, 2.427]
19	0.15	[0.147, 0.149]	2.39	[2.39, 2.40]	0.14	[0.142, 0.142]	2.40	[2.399, 2.399]
20	0.04	[0.043, 0.044]	2.53	[2.53, 2.54]	0.04	[0.035, 0.035]	2.54	[2.543, 2.543]
21	0.15	[0.147, 0.150]	2.39	[2.38, 2.39]	0.14	[0.143, 0.143]	2.40	[2.396, 2.397]
22	0.07	[0.065, 0.065]	2.50	[2.50, 2.51]	0.07	[0.066, 0.066]	2.50	[2.503, 2.503]
23	0.12	[0.115, 0.117]	2.43	[2.43, 2.44]	0.09	[0.088, 0.088]	2.47	[2.473, 2.473]
24	0.01	[0.009, 0.009]	2.58	[2.57, 2.58]	0.01	[0.007, 0.007]	2.58	[2.580, 2.580]
25	0.12	[0.115, 0.117]	2.43	[2.43, 2.44]	0.09	[0.089, 0.089]	2.47	[2.471, 2.471]

Note: Full model specifications are displayed in main text Table 5.

Table 10: Mean and confidence intervals of adjusted R^2, RMSE, and Concordance index values across simulations for models including τ-weighted multi-layer Degree centrality.

Model	Adj. R2	95% CI	RMSE	95% CI	C-index	95% CI
1	0.10	[0.100, 0.102]	2.45	[2.45, 2.46]	0.85	[0.853, 0.853]
2	0.11	[0.105, 0.106]	2.45	[2.44, 2.45]	0.85	[0.849, 0.849]
3	0.11	[0.112, 0.114]	2.43	[2.43, 2.44]	0.85	[0.851, 0.852]

Note: Full model specifications are displayed in main text Table 6.

References

[1] Kong, Lingcai et al. (2022). "Compartmental Structures Used in Modeling COVID-19: A Scoping Review". In: *Infectious Diseases of Poverty* 11.1, p. 72. DOI: 10.1186/s40249-022-01001-y. URL: https://idpjournal.biomedcentral.com/articles/10.1186/s40249-022-01001-y.

[2] Kivela, M. et al. (2014). "Multilayer Networks". In: *Journal of Complex Networks* 2.3, pp. 203–271. DOI: 10.1093/comnet/cnu016. URL: https://academic.oup.com/comnet/article-lookup/doi/10.1093/comnet/cnu016.

[3] Bucur, Doina and Petter Holme (2020). "Beyond Ranking Nodes: Predicting Epidemic Outbreak Sizes by Network Centralities". In: *PLoS Computational Biology* 16.7, e1008052. DOI: 10.1371/journal.pcbi.1008052.

[4] De Arruda, Guilherme Ferraz et al. (2014). "Role of Centrality for the Identification of Influential Spreaders in Complex Networks". In: *Physical Review E* 90.3, p. 032812. DOI: doi.org/10.1103/PhysRevE.90.032812.

[5] Rodrigues, Francisco Aparecido (2019). "Network Centrality: An Introduction". In: *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*. Ed. by Elbert E. N. Macau. Vol. 22. Cham: Springer International Publishing, pp. 177–196. DOI: 10.1007/978-3-319-78512-7_10. URL: http://link.springer.com/10.1007/978-3-319-78512-7_10.

[6] Basaras, Pavlos et al. (2019). "Identifying Influential Spreaders in Complex Multilayer Networks: A Centrality Perspective". In: *IEEE Transactions on Network Science and Engineering* 6.1, pp. 31–45. DOI: 10.1109/TNSE.2017.2775152. URL: https://ieeexplore.ieee.org/document/8114211/.

[7] Bongiorno, Christian and Lorenzo Zino (2022). "A Multi-Layer Network Model to Assess School Opening Policies during a Vaccination Campaign: A Case Study on COVID-19 in France". In: *Applied Network Science* 7.1, p. 12. DOI: 10.1007/s41109-022-00449-z. URL: https://appliednetsci.springeropen.com/articles/10.1007/s41109-022-00449-z.

[8] Chung, N. N. and L. Y. Chew (2021). "Modelling Singapore COVID-19 Pandemic with a SEIR Multiplex Network Model". In: *Scientific Reports* 11.1, p. 10122. DOI: 10.1038/s41598-021-89515-7. URL: http://www.nature.com/articles/s41598-021-89515-7.

[9] Nande, Anjalika et al. (2021). "Dynamics of COVID-19 under Social Distancing Measures Are Driven by Transmission Network Structure". In: *PLOS Computational Biology* 17.2. Ed. by Alex Perkins, e1008684. DOI: 10.1371/journal.pcbi.1008684. URL: https://dx.plos.org/10.1371/journal.pcbi.1008684.

[10] Halu, Arda et al. (2013). "Multiplex PageRank". In: *PLoS ONE* 8.10. Ed. by Yamir Moreno, e78293. DOI: 10.1371/journal.pone.0078293. URL: https://dx.plos.org/10.1371/journal.pone.0078293.

[11] Solá, Luis et al. (2013). "Eigenvector Centrality of Nodes in Multiplex Networks". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23.3, p. 033131. DOI: 10.1063/1.4818544. URL: http://aip.scitation.org/doi/10.1063/1.4818544.

[12] Solé-Ribalta, Albert et al. (2014). "Centrality Rankings in Multiplex Networks". In: *Proceedings of the 2014 ACM Conference on Web Science - WebSci '14*. Bloomington, Indiana, USA: ACM Press, pp. 149–155. DOI: 10.1145/2615569.2615687. URL: http://dl.acm.org/citation.cfm?doid=2615569.2615687.

[13] De Domenico, Manlio et al. (2015). "Ranking in Interconnected Multilayer Networks Reveals Versatile Nodes". In: *Nature Communications* 6.1, p. 6868. DOI: 10.1038/ncomms7868. URL: http://www.nature.com/articles/ncomms7868.

[14] De Domenico, Manlio et al. (2016). "The Physics of Spreading Processes in Multilayer Networks". In: *Nature Physics* 12.10, pp. 901–906. DOI: 10.1038/nphys3865. URL: http://www.nature.com/articles/nphys3865.

[15] Salehi, Mostafa et al. (2015). "Spreading Processes in Multilayer Networks". In: *IEEE Transactions on Network Science and Engineering* 2.2, pp. 65–83. DOI: 10.1109/TNSE.2015.2425961. URL: http://ieeexplore.ieee.org/document/7093190/.

[16] van der Laan, Jan et al. (2023). "A Whole Population Network and Its Application for the Social Sciences". In: *European Sociological Review* 39.1, pp. 145–160. DOI: 10.1093/esr/jcac026. URL: https://academic.oup.com/esr/article/39/1/145/6605763.

[17] Freeman, Linton C (1978). "Centrality in Social Networks Conceptual Clarification". In: *Social networks* 1.3, pp. 215–239.

[18] Christley, R. M. et al. (2005). "Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals". In: *American Journal of Epidemiology* 162.10, pp. 1024–1031. DOI: 10.1093/aje/kwi308. URL: http://academic.oup.com/aje/article/162/10/1024/65086/Infection-in-Social-Networks-Using-Network.

[19] De Domenico, Manlio et al. (2014). "Navigability of Interconnected Networks under Random Failures". In: *Proceedings of the National Academy of Sciences* 111, pp. 8351–8356.

[20] Gómez, S. et al. (2013). "Diffusion Dynamics on Multiplex Networks". In: *Physical Review Letters* 110.2, p. 028701. DOI: 10.1103/PhysRevLett.110.028701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.028701.

[21] De Domenico, Manlio et al. (2013). "Mathematical Formulation of Multilayer Networks". In: *Physical Review X* 3.4, p. 041022. DOI: 10.1103/PhysRevX.3.041022. URL: https://link.aps.org/doi/10.1103/PhysRevX.3.041022.

[22] Page, Lawrence (2001). "Method for Node Ranking in a Linked Database". US6285999B (Stanford, CA). URL: https://worldwide.espacenet.com/publicationDetails/biblio?locale=de_EP&CC=US&NR=6285999B1&FT=D&KC=B1.

[23] CBS (2023). *Bevolkingsontwikkeling; Regio per Maand*. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37230ned/table?ts=1678644956300.

[24] Van Iersel, Senna C.J.L. et al. (2023). "Empirical Evidence of Transmission over a School-Household Network for SARS-CoV-2; Exploration of Transmission Pairs Stratified by Primary and Secondary School". In: *Epidemics* 43, p. 100675. DOI: 10.1016/j.epidem.2023.100675. URL: https://linkinghub.elsevier.com/retrieve/pii/S1755436523000117.

[25] van der Laan, D.J. (2022). "A Person Network of the Netherlands". In: *CBS Discussion Paper*. Ed. by Statistics Netherlands. URL: https://www.cbs.nl/-/media/_pdf/2022/20/person_network_netherlands.pdf.

[26] Lei, Hao et al. (2020). "Household Transmission of COVID-19-a Systematic Review and Meta-Analysis". In: *Journal of Infection* 81.6, pp. 979–997. DOI: 10.1016/j.jinf.2020.08.033. URL: https://linkinghub.elsevier.com/retrieve/pii/S0163445320305715.

[27] Madewell, Zachary J. et al. (2022). "Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis". In: *JAMA Network Open* 5.4, e229317. DOI: 10.1001/jamanetworkopen.2022.9317. URL: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2791601.

[28] Telle, Kjetil et al. (2021). "Secondary Attack Rates of COVID-19 in Norwegian Families: A Nation-Wide Register-Based Study". In: *European Journal of Epidemiology* 36.7, pp. 741–748. DOI: 10.1007/s10654-021-00760-6. URL: https://link.springer.com/10.1007/s10654-021-00760-6.

[29] Abbey, Helen (1952). "An Examination of the Reed-Frost Theory of Epidemics". In: *Human Biology* 24.3, pp. 201–233.

[30] Draief, Moez (2006). "Epidemic Processes on Complex Networks". In: *Physica A: Statistical Mechanics and its Applications* 363.1, pp. 120–131. DOI: 10.1016/j.physa.2006.01.054. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378437106000975.

[31] Dekker, Mark M et al. (2023). "Reducing Societal Impacts of SARS-CoV-2 Interventions through Subnational Implementation". In: *eLife* 12, e80819. DOI: 10.7554/eLife.80819. URL: https://elifesciences.org/articles/80819.

[32] Schafer, Joseph L. (1999). *Analysis of Incomplete Multivariate Data*. 1, 1. CRC Press reprint. Chapman & Hall/CRC.

[33] Cox, D. R. (1975). "Partial Likelihood". In: *Biometrika* 62.2, pp. 269–276. DOI: 10.1093/biomet/62.2.269. URL: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/62.2.269.

[34] Uno, Hajime et al. (2011). "On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data". In: *Statistics in Medicine* 30.10, pp. 1105–1117. DOI: 10.1002/sim.4154. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.4154.

[35] Harel, Ofer (2009). "The Estimation of R^2 and Adjusted R^2 in Incomplete Data Sets Using Multiple Imputation". In: *Journal of Applied Statistics* 36.10, pp. 1109–1118. DOI: 10.1080/02664760802553000. URL: https://www.tandfonline.com/doi/full/10.1080/02664760802553000.

[36] R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

[37] Csárdi, Gábor and T Nepusz (2006). "The Igraph Software Package Complex Network Research". In: *InterJournal* Complex Systems.1695. URL: https://igraph.org.

[38] De Domenico, M., M. A. Porter, and A. Arenas (2015). "MuxViz: A Tool for Multilayer Analysis and Visualization of Networks". In: *Journal of Complex Networks* 3.2, pp. 159–176. DOI: 10.1093/comnet/cnu038. URL: https://academic.oup.com/comnet/article-lookup/doi/10.1093/comnet/cnu038.

Disclaimer

Results are based on calculations by the author using non-public micro-data from Statistics Netherlands. Under certain conditions, these micro-data are accessible for statistical and scientific research. For further information contact microdata@cbs.nl. Please see also [16, pp. 149–150] for more explanation on the data itself as well as access and usage regulations.