Predicting Covid-19 infections using multi-layer centrality measures

Master Thesis

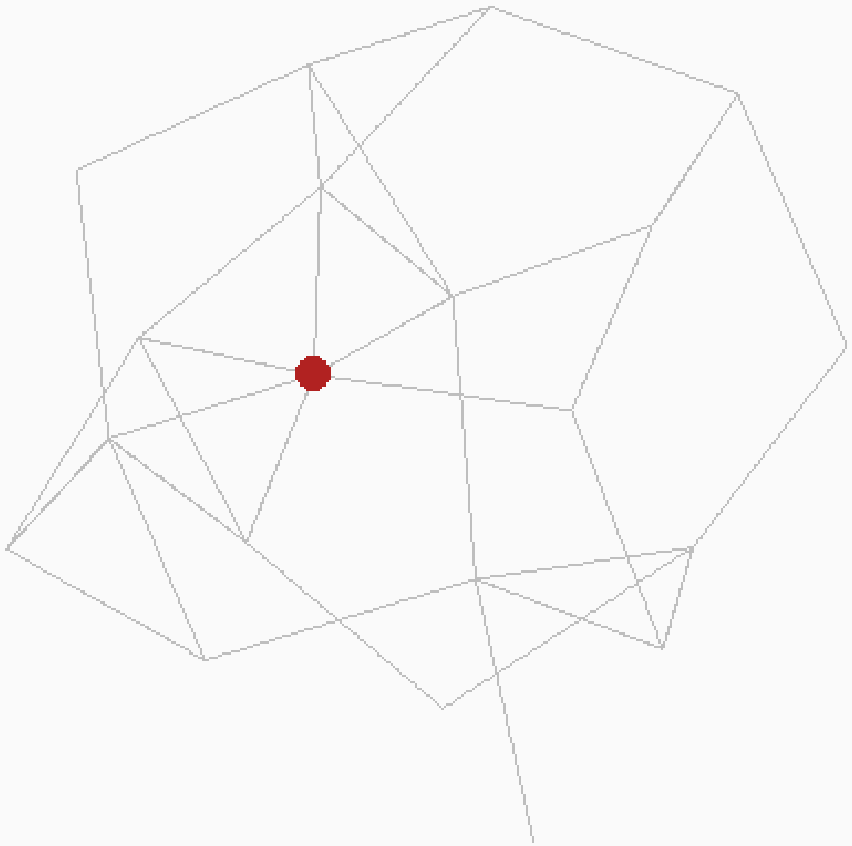
Christine Hedde - von Westernhagen

22-05-2023

Program MSBBSS, Department of Methodology and Statistics, Utrecht University

Supervisors: Javier Garcia-Bernardo, Ayoub Bagheri

Predicting Covid-19 infections using multi-layer centrality measures

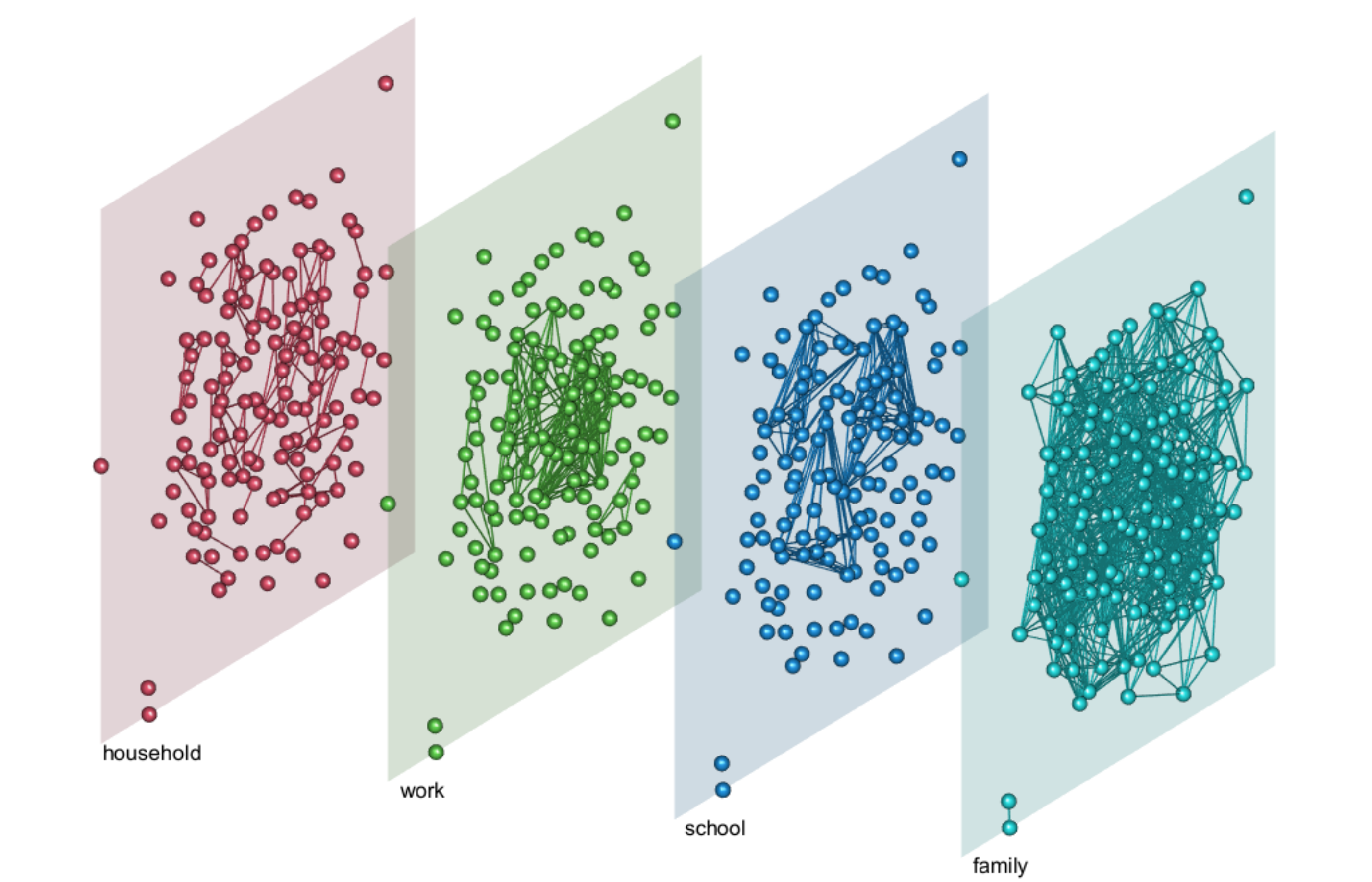


Model spread to inform policy decisions!

Who gets infected and **when**?

Predicting Covid-19 infections using **multi-layer** centrality measures

Our lives are multi-layered.



CBS micro-data allows to construct multi-layer network dataset.

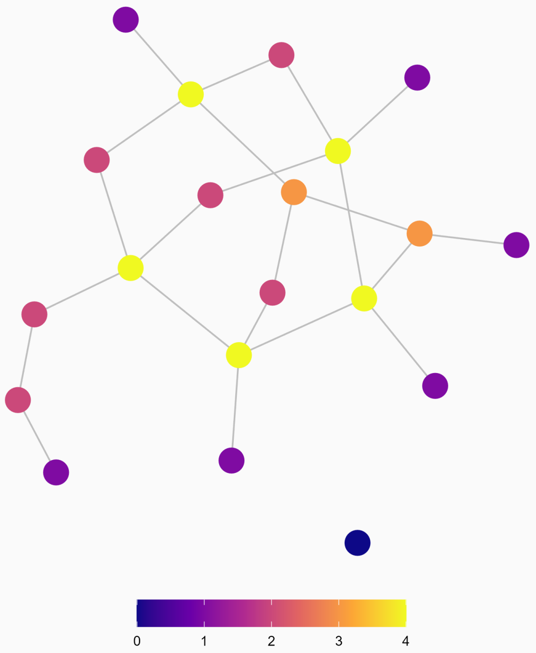
Micro-data can be linked to **PCR test data.**

Analyses conducted on regional **subset** of ~ 1m nodes.

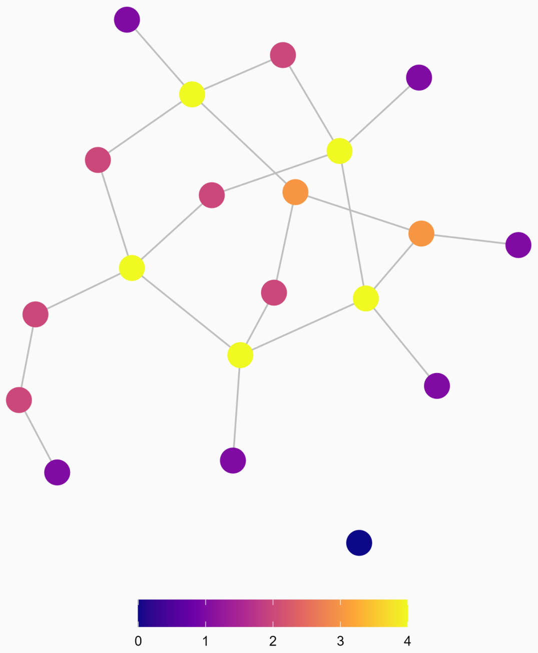
Predicting Covid-19 infections using multi-layer

centrality measures

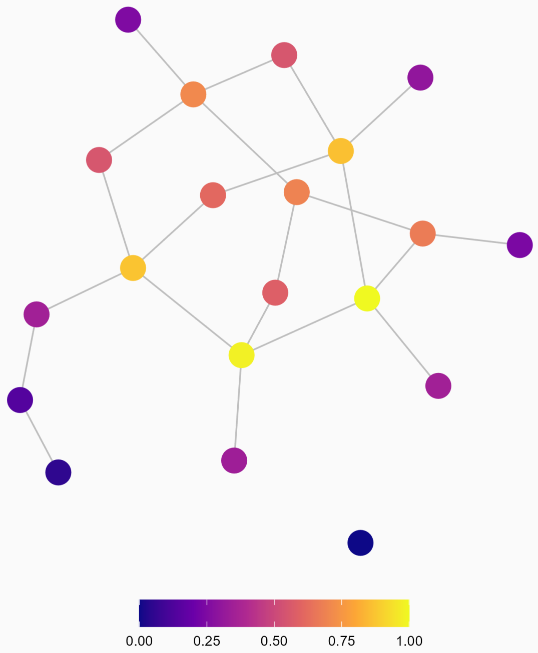
Degree

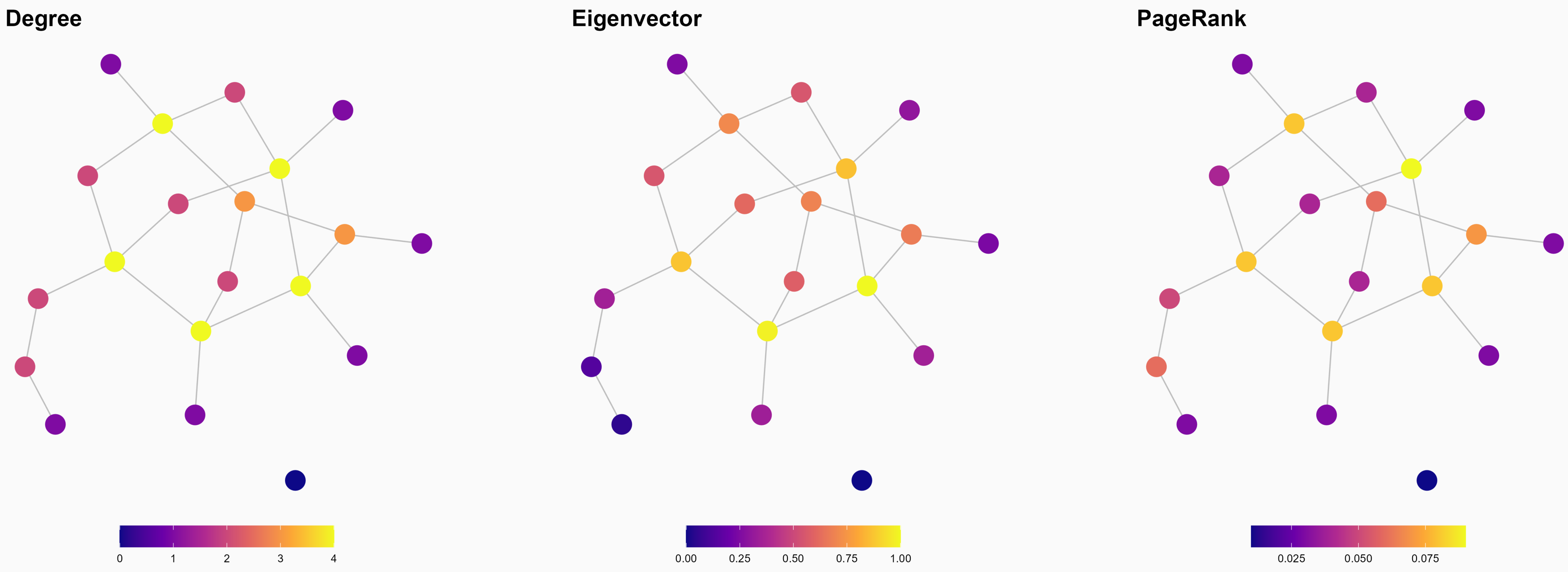


Degree



Eigenvector

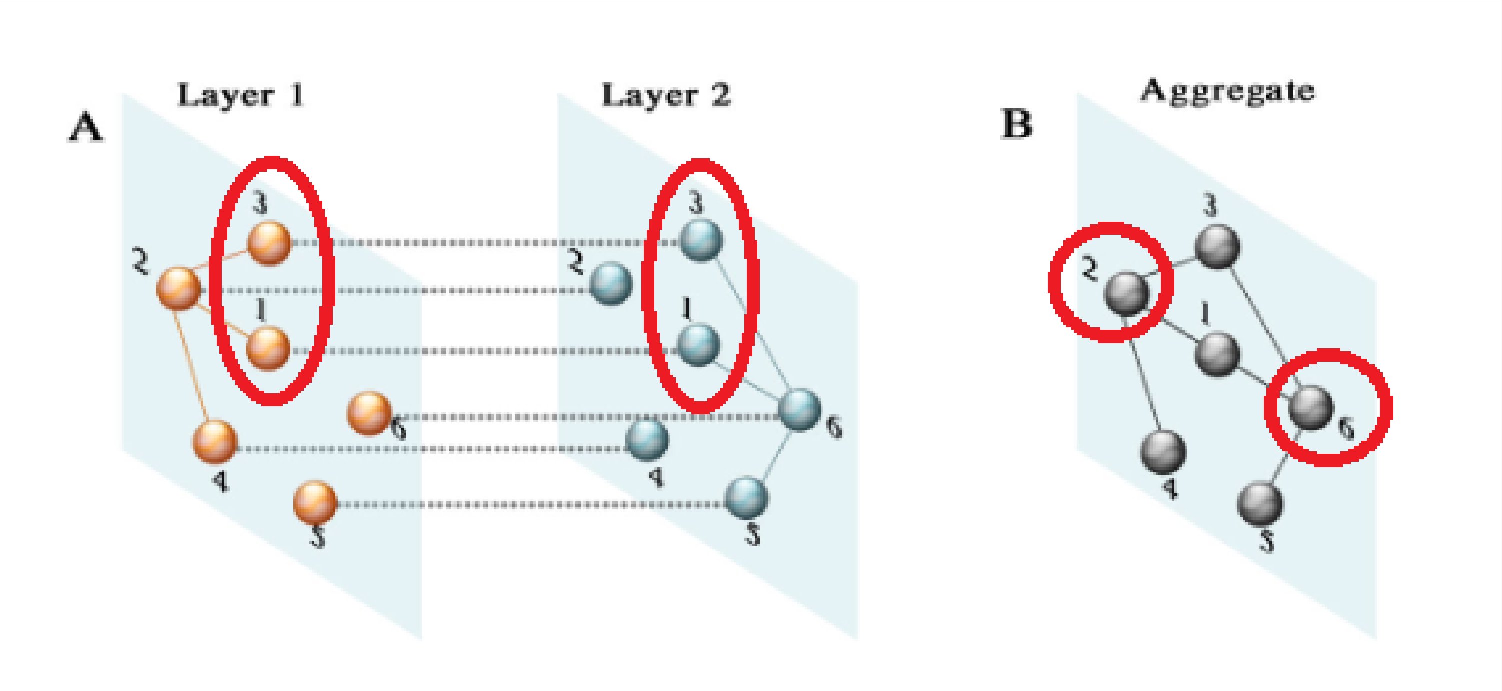



Degree
0
1
2
3
4
Eigenvector
0.00
0.25
0.50
0.75
1.00
PageRank
0.025
0.050
0.075

Single-layer centrality

≠

Multi-layer centrality



De Domenico et al. 2015, Supplementary Figure 1

De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., & Arenas, A. (2015). Ranking in interconnected multilayer networks reveals versatile nodes. Nature Communications, 6(1), 6868. https://doi.org/10.1038/ncomms7868

Bringing it all together:

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

Analytic Strategy

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

---

1. Simulate an epidemic on the network k = 500

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

1. Simulate an epidemic on the network k = 500
2. Get time until infection for the nodes in each simulation

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

1. Simulate an epidemic on the network k = 500
2. Get time until infection for the nodes in each simulation
3. Run prediction models for each of the 500 data sets and compute estimates

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

1. Simulate an epidemic on the network k = 500
2. Get time until infection for the nodes in each simulation
3. Run prediction models for each of the 500 data sets and compute estimates
   a. Rank correlations ρ of centrality measures and outcome

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

1. Simulate an epidemic on the network k = 500
2. Get time until infection for the nodes in each simulation
3. Run prediction models for each of the 500 data sets and compute estimates
   a. Rank correlations ρ of centrality measures and outcome
   b. Uno's Concordance Index C of Cox models including linear combinations of centralities as predictors

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

1. Simulate an epidemic on the network k = 500
2. Get time until infection for the nodes in each simulation
3. Run prediction models for each of the 500 data sets and compute estimates

   a. Rank correlations ρ of centrality measures and outcome

   b. Uno's Concordance Index C of Cox models including linear combinations of centralities as predictors

   c. R^2 of linear models including linear combinations of centralities as predictors

Outcome: **Time until infection**

Predictors: Degree, Eigenvector, PageRank

1. Simulate an epidemic on the network k = 500
2. Get time until infection for the nodes in each simulation
3. Run prediction models for each of the 500 data sets and compute estimates
   a. Rank correlations ρ of centrality measures and outcome
   b. Uno's Concordance Index C of Cox models including linear combinations of centralities as predictors
   c. R^2 of linear models including linear combinations of centralities as predictors
4. Average the estimates across simulations

Results

a. **Medium negative correlations** of all centrality measures with outcome

a. **Medium negative correlations** of all centrality measures with outcome

b. **Concordance index ~ 0.85** for predictor combinations including Degree centrality and/or PageRank

a. **Medium negative correlations** of all centrality measures with outcome

b. **Concordance index ~ 0.85** for predictor combinations including Degree centrality and/or PageRank

c. ~ **15% of variance** explained in outcome by (at least) Degree [2]

a. **Medium negative correlations** of all centrality measures with outcome

b. **Concordance index ~ 0.85** for predictor combinations including Degree centrality and/or PageRank

c. ~ **15% of variance** explained in outcome by (at least) Degree [2]

Also: Only small differences between single-layer and multi-layer measures.

Is that it?

Multiple layers, multiple transmission rates.

Multiple layers, multiple transmission rates.

➔ **Weighted Degree centrality**

Multiple layers, multiple transmission rates.

➔ **Weighted Degree centrality**

...minimal performance increase

...what about the real infection data?

...what about the real infection data?

$$R^2 \leq 0.01$$

Lessons learned

> Centrality measures can predict **relative infection risks** well, but are limited in predicting the **timing of infections**.

> Centrality measures can predict **relative infection risks** well, but are limited in predicting the **timing of infections**.
>
> Multi-layer measures don't seem to offer great benefits.

Centrality measures can predict **relative infection risks** well, but are limited in predicting the **timing of infections**.

Multi-layer measures don't seem to offer great benefits.

Administrative data is not sufficient to model social contacts.

Centrality measures can predict relative infection risks well, but are limited in predicting the timing of infections.

Multi-layer measures don't seem to offer great benefits.

Administrative data is not sufficient to model social contacts.

Centrality measures could still **complement** prediction models

> Centrality measures can predict relative infection risks well, but are limited in predicting the timing of infections.
>
> Multi-layer measures don't seem to offer great benefits.
>
> Administrative data is not sufficient to model social contacts.

Centrality measures could still **complement** prediction models

… **if** used with representative contact networks.

Backup slides

Behind the scenes: **Simulating an epidemic** on a large-scale network using a SIR model.

$$\frac{dS}{dt} = -\beta I S$$

$$\frac{dI}{dt} = \beta I S - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Each node can be in one of the **three states**, and **transitions stochastically** based on *network contacts*, *transmission rate*, and *recovery time*.

$$S \rightarrow I$$

$$I \rightarrow R$$

Where the transmission probabilities for the nodes in layer l at time t is given by:

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

Do you have...

👩👴👶👦...many close relatives? (Degree Centrality)

👩👴👶👦 👭 👧👴👵👦...many close relatives and also a partner with a big family? (Eigenvector Centrality)

✈️🌍 👧👴👵👦... but that partner's family lives in a different country? (Betweenness Centrality)

... then you could be a **super-spreader**.

High centrality ⇝ High spreading capacity

But: High spreading capacity ≠ High infection risk