Disaggregation of a SIRD model for policy testing in targeted groups of the population

C.Cortes Balcells, R. Krueger, M. Bierlaire

PhD Student,
Ecole Polytechnique Federale de Lausanne (EPFL)

ETH Workshop

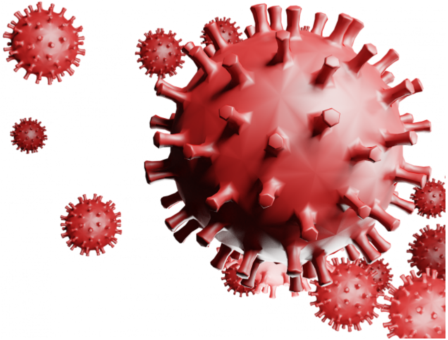
Agenda

- Introduction and Motivation
- State of the Art
- SIRD Disaggregated Model
- Results
- Conclusion

Motivation

Challenges

- Accounting for individual behaviour through an epidemic outbreak by using **large scale models**.
- Difficulty of finding **disaggregated data** to **validate** the model.
- Capturing **spread of the disease** through **daily activities**.
- Allows to **assess the impact** that a certain policy has on **different segments of the population**.
- Epidemiological datasets are becoming available.



Research gaps

Limitations

- Lack of data leads to add aggregated parameters inside the agent-based models, [TYK+20].
- Agent-based models in order to define more targeted and less disruptive interventions. Results are achieved using real-time data-driven analysis, [AMCB+20].
- Clear methodology to know which variables are meaningful inside an epidemiological model, for example income or residence place, [CPK+21].
- Make the probabilities time dependant, since an early adoption can potentially allow to contain the epidemics, [MBCV20].

Outline of this talk

1. Added value of using disaggregate models for modelling SARS-CoV-2 spreading. [1]
2. Description of the preliminary considerations and presentation of a model that accounts for virological and socio-economic variables.[2]
3. Potential of these models to study SARS-CoV-2 policy decision making.[3]

Literature:

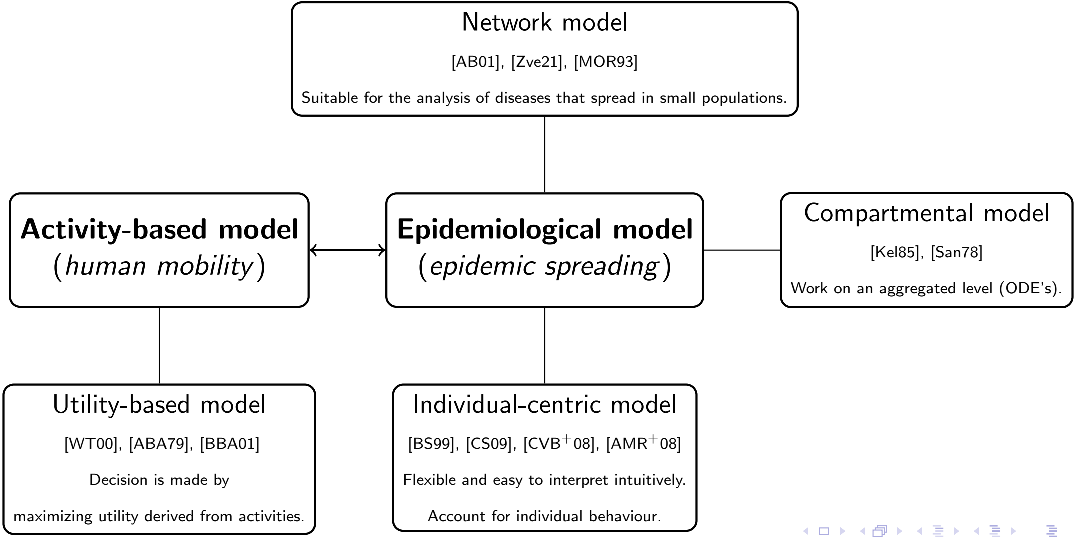
[1] A. Aleta, D. Martin-Corral, M. Bakker, A. Piontti, M. Ajelli, M. Litvi-nova, M. Chinazzi, N. Dean, M. Halloran, I. Longini, A. Pentland,A. Vespignani, Y. Moreno, and E. Moro. Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas, 12 2020.

[2] S. Chang, E. Pierson, P. Koh, J. Gerardin, B. Redbird, D. Grusky, andJ. Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. Nature, 589:1-6, 01 2021.

[3] M. Mancastroppa, R. Burioni, V. Colizza, and A. Vezzani. Active and inactive quarantine in epidemic spreading on adaptive activity-driven networks. Physical Review E, 102, 08 2020.

State of the Art

State of research



Compartmental models

The SIR epidemic model can be written in the following way(c.f [KMS17]):

- The transitions at each time step Δt are:

$$\frac{\partial S}{\partial t}(t) = -\lambda I(t)\frac{S(t)}{N}$$
$$\frac{\partial I}{\partial t}(t) = \lambda I(t)\frac{S(t)}{N} - \gamma I(t)$$
$$\frac{\partial R}{\partial t}(t) = \gamma I(t)$$

Notation	Parameters
λ	Contagion rate between S and I.
$1/\gamma$	Length of the infectious period for population I.

- S: Susceptible
- I: Infected
- R: Recovered

Compartmental model

[Kel85], [San78]

Work on an aggregated level (ODE's).

SIRD Disaggregated Model

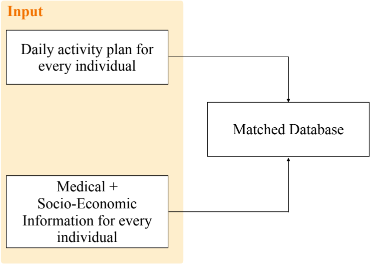
SIRD Disaggregated Model

Input

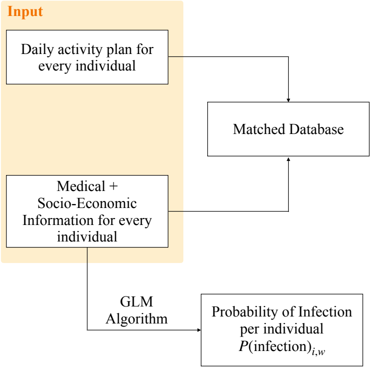
Daily activity plan for every individual

Medical + Socio-Economic Information for every individual

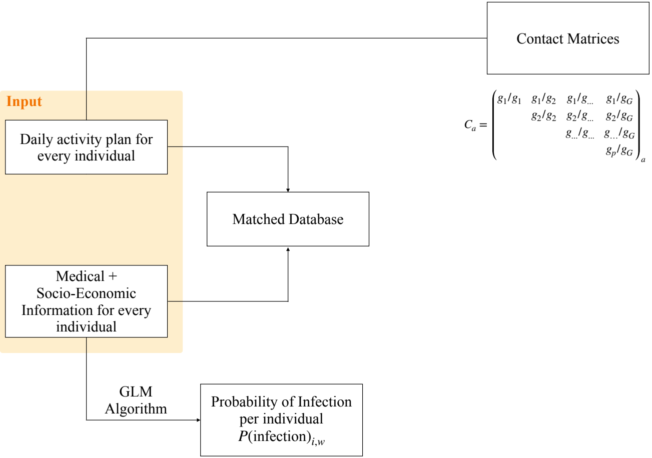
SIRD Disaggregated Model



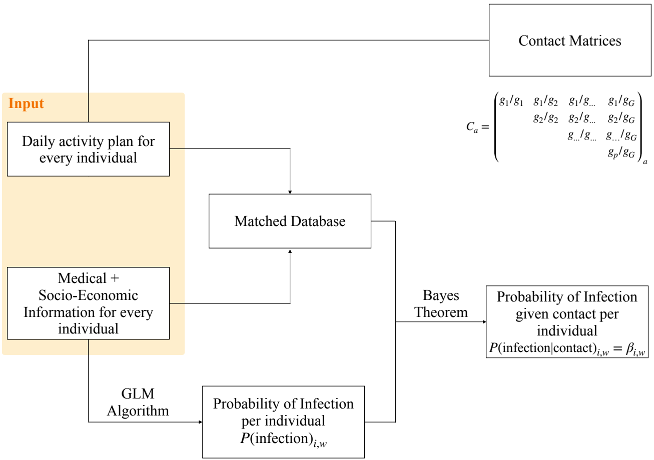
SIRD Disaggregated Model



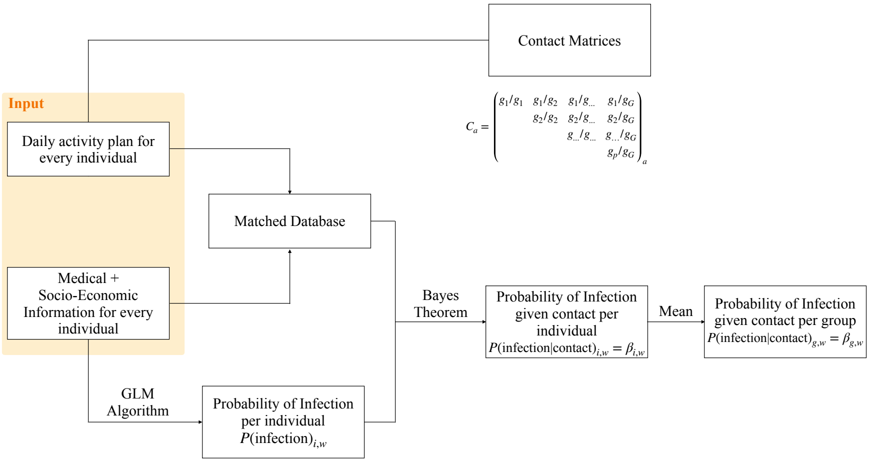
SIRD Disaggregated Model



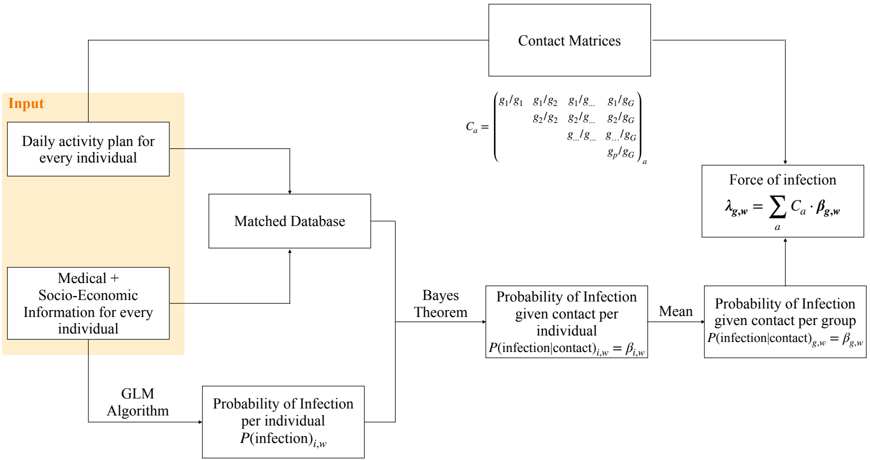
SIRD Disaggregated Model



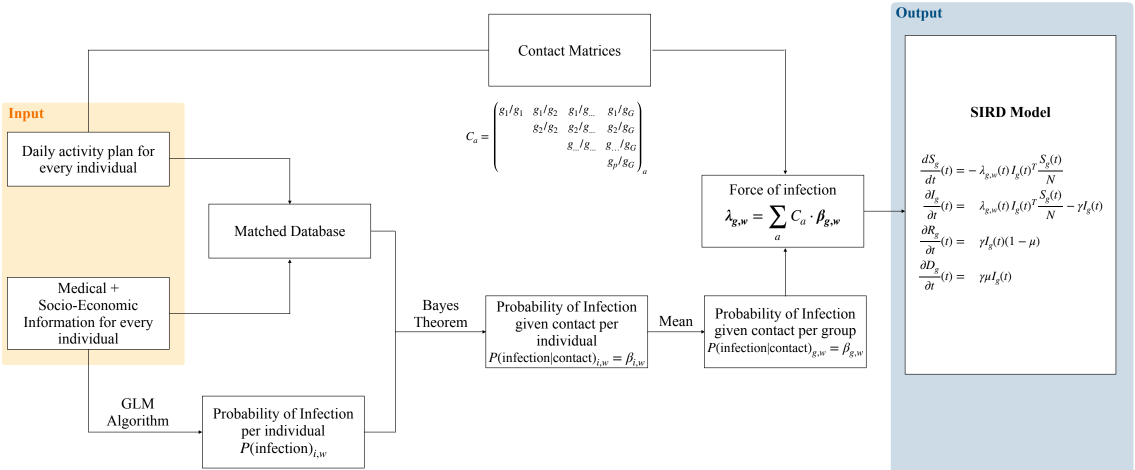
SIRD Disaggregated Model



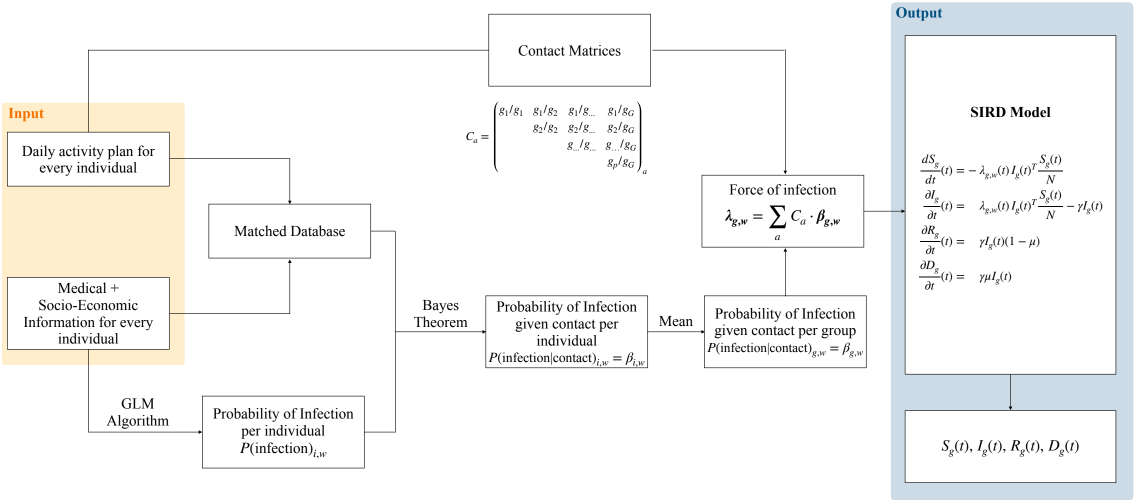
SIRD Disaggregated Model



SIRD Disaggregated Model

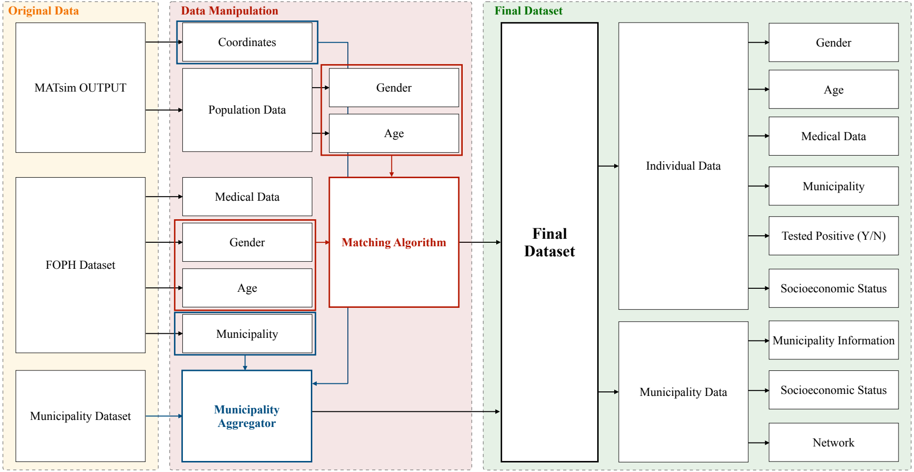


SIRD Disaggregated Model

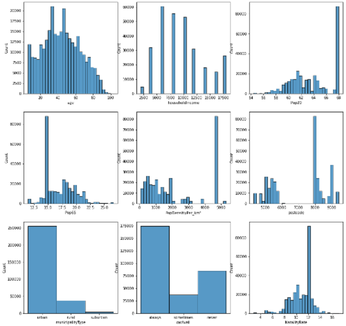


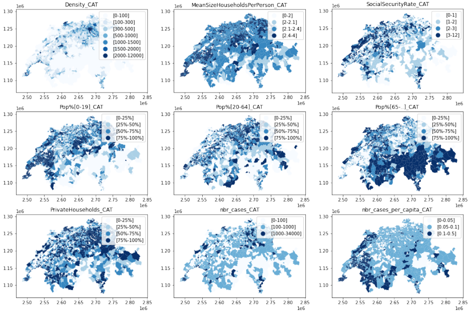
Results

Pre-process [HB21], [RPA$^+$21], [BFS22]



Pre-process





Density_CAT
[0-100]
[100-300]
[300-500]
[500-1000]
[1000-1500]
[1500-2000]
[2000-12000]
MeanSizeHouseholdsPerPerson_CAT
SocialSecurityRate_CAT
Pop%[0-19]_CAT
[0-25%]
[25%-50%]
[50%-75%]
[75%-100%]
Pop%[20-64]_CAT
PrivateHouseholds_CAT
nbr_cases_CAT
[0-100]
[100-1000]
[1000-34000]
nbr_cases_per_capita_CAT

Activity Contact Matrix

- We stratify the model into 4 age groups:
  - C = the individuals younger than 18 years old,
  - A1 = individuals between 19 and 35 years-old,
  - A2 = individuals between 36 and 55 years-old
  - E = individuals over 56 years-old

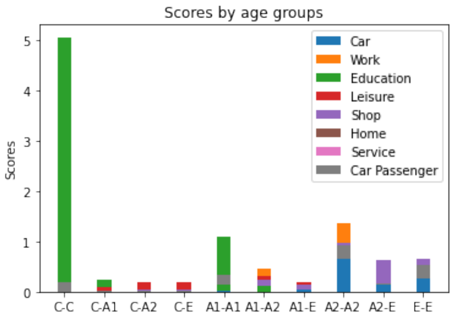
- The segmentation endows our model with high flexibility for policy testing on the different population groups. The contact matrix C_a becomes:

	child (C)	*adult1 (A1)*	*adult2 (A2)*	*elderly (E)*
child	child / child	child / adult1	child / adult2	child / elderly
adult1	-	adult1 / adult1	adult1 / adult2	adult1 / elderly
adult2	-	-	adult2 / adult2	adult2 / elderly
elderly	-	-	-	elderly / elderly

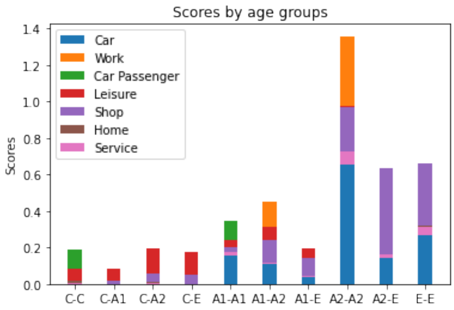
Table: Contact matrix structure for each activity

Activity Contact Matrix

All activities



Education removed



Generalized Linear Model Regression

$$P(\text{infection})_{i,1} \sim \beta_\Lambda \log(\Lambda) + \beta_\chi \log(\chi) + \beta_\Upsilon \log(\Upsilon) + \beta_\kappa \log(\kappa). \quad (1)$$

$$P(\text{infection})_{i,2} \sim \beta_\Lambda \log(\Lambda) + \beta_\chi \log(\chi) + \beta_\Upsilon \log(\Upsilon) + \beta_\kappa \log(\kappa) + \beta_\Phi \Phi. \quad (2)$$

Where:

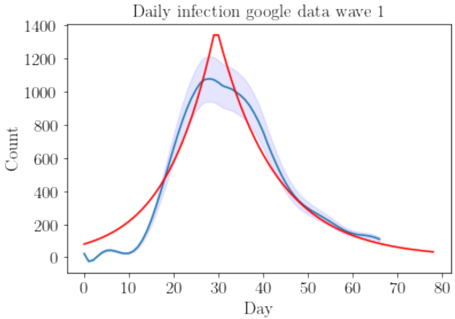
- The age of the individual: Λ.
- The percentage of the population above 65 years old for a specific municipality: χ.
- The percentage of the population between 20 and 65 years old for a specific municipality: Υ.
- The population density per km: κ.
- The income: Φ.

SIRD Disaggregated Model Output

Assumptions:

We reduce the activities the 24^{th} of March 2020 to:

- 100% Education.
- 90% Leisure and Services.
- 80% Work.
- 70% Shop.
- 60% Car and car passenger.



Conclusions and future work

- The most significant contributions are:
  - We capture how the socio-economic characteristics of an individual define the force of infection
  - We obtain a self-explanatory model, defined by the estimates of the variables that characterize the spreading event
  - We obtain high goodness of fit of our model with Google data.
- The future work includes:
  - Improve and validate the current model
  - Scale it up to more groups.
  - Include it in an optimization framework to use it for policy analysis.

Thank you

References I

[AB01] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 06 2001.

[ABA79] Thomas Adler and Moshe Ben-Akiva. A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological*, 13:243–257, 09 1979.

[AG92] Kay Axhausen and Tommy Gärling. Activity-based approaches to travel analysis: Conceptual frameworks, models and research problems. *Transport Reviews - TRANSP REV*, 12:323–341, 10 1992.

[AMCB+20] Alberto Aleta, David Martín-Corral, Michiel Bakker, Ana Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie Dean, M. Halloran, Ira Longini, Alex Pentland, Alessandro Vespignani, Yamir Moreno, and Esteban Moro. Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas, 12 2020.

[AMR+08] Marta Atti, Stefano Merler, Caterina Rizzo, Marco Ajelli, Marco Massari, Piero Manfredi, Cesare Furlanello, Gianpaolo Tomba, and Mimmo Iannelli. Mitigation measures for pandemic influenza in italy: An individual based model considering different scenarios. *PloS one*, 3:e1790, 02 2008.

References II

[BBA01] J.L. Bowman and Moshe Ben-Akiva. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35:1–28, 01 2001.

[BFS22] BFS. Bfs municipality open data, 2022.

[BS99] Michael Boots and Akira Sasaki. ‘small worlds’ and the evolution of virulence: Infection occurs locally and at a distance. *Proceedings. Biological sciences / The Royal Society*, 266:1933–8, 11 1999.

[CPK+21] Serina Chang, Emma Pierson, Pang Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589:1–6, 01 2021.

[CS09] Connie Carpenter and Lisa Sattenspiel. The design and use of an agent-based model to simulate the 1918 influenza epidemic at norway house, manitoba. *American journal of human biology : the official journal of the Human Biology Council*, 21:290–300, 05 2009.

[CVB+08] Simon Cauchemez, Alain-Jacques Valleron, Pierre-Yves Boelle, Antoine Flahault, and Neil Ferguson. Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452:750–4, 05 2008.

References III

[HB21] Sebastian Hörl and Milos Balac. Synthetic population and travel demand for paris and Île-de-france based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, 130:103291, 2021.

[Kel85] A.W. Kelman. Compartmental models and their application. *International Journal of Bio-medical Computing - INT J BIO MED COMPUT*, 16:294–295, 05 1985.

[KMS17] Istvan Kiss, Joel Miller, and Péter Simon. *Mathematics of Epidemics on Networks*, volume 46. 01 2017.

[MBC+20] Sebastian Müller, Michael Balmer, Billy Charlton, Ricardo Ewert, Andreas Neumann, Christian Rakow, Tilmann Schlenther, and Kai Nagel. Using mobile phone data for epidemiological simulations of lockdowns: government interventions, behavioral changes, and resulting changes of reinfections, 07 2020.

[MBCV20] Marco Mancastroppa, Raffaella Burioni, Vittoria Colizza, and Alessandro Vezzani. Active and inactive quarantine in epidemic spreading on adaptive activity-driven networks. *Physical Review E*, 102, 08 2020.

[MOR93] MARTINA MORRIS. Epidemiology and social networks:: Modeling structured diffusion. *Sociological Methods and Research*, 22:99–126, 08 1993.

References IV

[RPA+21] Julien Riou, Radoslaw Panczak, Christian Althaus, Christoph Junker, Damir Perisa, Katrin Schneider, Nicola Criscuolo, Nicola Low, and Matthias Egger. Socioeconomic position and the cascade from sars-cov-2 testing to covid-19 mortality: Population-based analysis of swiss surveillance data. *SSRN Electronic Journal*, 01 2021.

[San78] Irwin Sandberg. On the mathematical foundations of compartmental analysis in biology, medicine, and ecology. *Circuits and Systems, IEEE Transactions on*, CAS-25:273 – 279, 06 1978.

[Smi09] Timo Smieszek. A mechanistic model of infection: Why duration and intensity of contacts should be included in models of disease spread. *Theoretical biology and medical modelling*, 6:25, 11 2009.

[TYK+20] J. T. Tuomisto, Juha Yrjölä, M. Kolehmainen, Juhani Bonsdorff, J. Pekkanen, and Tero Tikkanen. An agent-based epidemic model reina for covid-19 to identify destructive policies. *medRxiv*, 2020.

[WT00] Donggen Wang and Harry Timmermans. Conjoint-based model of activity engagement, timing, scheduling, and stop pattern formation. *Transportation Research Record*, 1718:10–17, 01 2000.

[Zve21] Vadim Zverovich. *Modern Applications of Graph Theory*. 04 2021.

Generalized Linear Model Regression

Summary of statistics list of covariates

Table: Summary statistics of the list of covariates

	1^{st} WAVE			2^{nd} WAVE		
Stratified by infection	0	1	SMD	0	1	SMD
n	269642	414		281576	14499	
$\Lambda(mean(SD))$	44.42 (20.99)	50.91 (20.24)	0.315	44.43(21.05)	43.33(19.68)	0.054
$\Upsilon(mean(SD))$	3.89 (3.27)	65.17 (3.07)	0.0400	63.63(3.30)	63.91(3.25)	0.086
$\kappa(mean(SD))$	2399.74 (1760.49)	3123.01 (1733.89)	0.414	2222.49(1771.15)	2401.70(1751.82)	0.102
$\chi(mean(SD))$	16.89 (2.50)	16.26 (2.29)	0.0264	17.04(2.56)	16.84(2.48)	0.077
$\Phi(mean(SD))$				0.02(0.13)	0.02(0.12)	0.002

Generalized Linear Model Regression

Summary of the estimates using Causal Inference

Table: Coefficients using Matching score algorithm

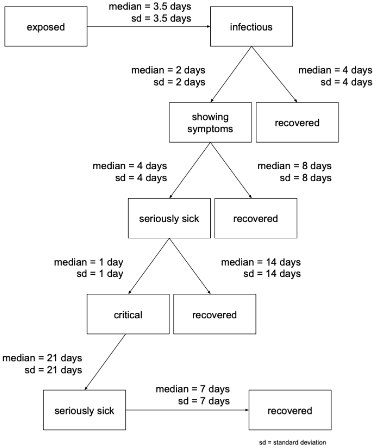
Variable	1^{st} WAVE Est.	SE	z-val.	p-val.	2^{nd} WAVE Est.	SE	z-val.	p-val.
$\log(\Lambda)$	0.641	0.101	6.321	$2.590\mathrm{e}-10$	0.0301	0.0143	2.100	0.0358
$\log(\chi)$	1.395	0.773	1.806	0.0701	-0.386	0.107	-3.608	0.000309
$\log(\Upsilon)$	7.388	2.893	2.554	0.0106	-2.0381	0.391	-5.211	$1.880e-07$
$\log(\kappa)$	0.281	0.109	2.594	0.00950	0.187	0.0141	13.246	$< 2\mathrm{e}-16$
Φ					-0.0556	0.0261	-2.131	0.0331

SIR model used in Episim, [MBC+20]

- The probability for person n to become infected by this process in a time step t in [MBC+20]:

$$P_{n,t} = 1 - \exp\left[-\Theta \sum_{m} q_{m,t} \cdot ci_{nm,t} \cdot in_{n,t} \cdot \tau_{nm,t}\right]$$

- **Main issues of this probability:**
  - Parameters unknown for COVID-19, so set to $value = 1$,
  - All multiplying so it might be independent modifying one or another,
- We want to create dependence on not only individual and time but also on the location.



Activity-based model [AG92]

- They allow more complex policies to be evaluated.
- The phenomena are understood as the result of the interaction of multiple agents, each guided by individual norms or intelligence.
- This interaction results in a complex system, consisting of many sub-systems and agents (applicable to many disciplines: ecology, economics, computer simulation...).
- It uses microscopic simulation.
- Examples: MATsim, TRANSIMS...

Epidemiological models

Issues:

- SEIR models work on an aggregate level: neglects the imperfect mixture.
- Not transferable to different epidemics.
- Crutial parameters might not be available.
- Exponential is a strong assumption.

Network models

The graph G is defined by n vertices, and m edges:

$\{G_1, G_2, \ldots, G_n\}$, where $n = \binom{M}{m}$ with $M = N(N-1)/2$.

The probability of picking each graph is the same:

$1/n$.

Network models

Issues:

- Complex to find the correct adjacency matrix.
- Difficult to use them in densely populated areas.
- Quality of the contacts between two individuals. The adjacency matrices are binary.
- Static character of network models.

Individual-centric models

If an individual is susceptible and it has contact with an infected agent, it becomes infected with a probability: p.

This probability can be defined as desired. For example in [Smi09], the probability for person n to become infected by this process in a time step t is defined as:

$$P_{n,t} = 1 - \exp\left[-\Theta \sum_{m} q_{m,t} \cdot ci_{nm,t} \cdot in_{n,t} \cdot \tau_{nm,t}\right]$$

Main issues of this probability:

- Parameters unknown for COVID-19, so set to $value = 1$,
- All multiplying so it might be independent modifying one or another,
- We want to create dependence on not only individual and time but also on the location.

Individual-centric models

- $\rightarrow$ insight on transmission and intervention that will complete what can be obtained with usual compartmental models (SIR).
- Added value of these models . . .
  - The interactions between agents are nonlinear, discontinuous or complex..
  - When the space is crucial and we do not have fixed positions.
  - Population is heterogenous with different socioeconomic characteristics.
  - Agents have complex behaviour.
  - Topology of interactions is complex.