Institut für Informatik
Praktische Informatik und Bioinformatik
Prof. Dr. Ralf Zimmer
Dr. Evi Berchtold, Armin Hadziahmetovic, Dr. Markus Joppich
Samuel Klein, Felix Offensperger Chuqiao Pan

SoSe 2025 based on Corona/Online Summer Term 2020
Computational (Food) Systems Biology

Assignments 2025-1
Modeling the COVID-19 Pandemics

Working period: Three weeks (29.4.-20.5.2025)
Hand-in anytime or in any exercise class
Please hand-in only reproducible results, answers, figures, tables, simulations, ...
Report due May 20, 2025

This project analyses the current COVID-19[1] pandemics, a world-wide crisis caused by a new corona virus, the SARS-CoV-2[2]. It investigates some basic dynamical systems models and the employed methods. It also introduces some technical issues and frameworks for visualizing results and making models available for parametrization and simulations.

As a first teaser into the situation and their mathematical modelling please watch the short video of Dr. Mai Thy Ngyuen Kim (https://youtu.be/3z0gnXgK8Do).

(Remember: in the SoSe 2020 the pandemics had recently started was in full swing, universities were closed, and the further development was completely unforeseeable - both biologically medically, and politically):
The goal of this first class of the summer term 2020 is to understand as much as possible about an ongoing and pressing disease from a bioinformatics/system biology perspective. The class should also enable to better understand the information, myths, and fakes about the disease. It also should set the stage for a already quite broad range of techniques (from text mining, knowledge extraction, to network reconstruction, high-throughput analysis, and systems simulation) in current bioinformatics/systems biology research, which will be of use in other research projects. Thereby, the COVID-19 analysis introduces already quite some concepts which we will discuss in more detail during the course. We also setup a platform for implementing and communicating current and forthcoming programming tasks.

[1] COVID-19 = COronaVIrus Disease 2019

[2] SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2, "Schweres akutes Atemwegssyndrom Coronavirus"; vormals 2019-nCoV, 2019-novel Corona virus, neuartiges Coronavirus 2019 sowie Wuhan-Coronavirus

Task 1 (Tools: github and the python/scipy/numpy/panda/matplotlib environment)

Many tools have been and will be used in systems biology modeling. A broad variety of modelling tools has been developed and are publicly available implemented on a wide spectrum of programming platforms and environments. Moreover it is important to have some tools available to prepare and postprocess data for the use in the various modelling tools.

Therefore, we recommend for this course to familiarize with a python based framework for basic (and tool/library-based) systems biology programming , gitlab/github for documenting and communicating results and programs, and the scipy, numpy, Panda, and matplotlib extensions for more involved scientific computations.

As a warm up, implement a discrete approximation of small differential equation system.

(a) The pendulum: the movement of a pendulum can be described via the system of first order ordinary differential equations:

$$\begin{pmatrix} \dot{\varphi} \\ \dot{\omega} \end{pmatrix} = \begin{pmatrix} \omega \\ -\frac{g}{l} \sin \varphi \end{pmatrix}$$

Compare and visualize your results with a standard solver (readily available in your platform). Discuss any differences if necessary. Make a notebook, e.g. with Jupyter, to document and communicate your work. Document and publish your tool via git*.

(b) A so called "feed-forward" network motif is described as follows:

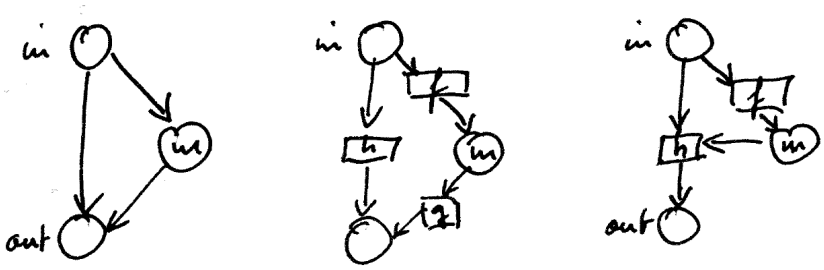

Abbildung 1: Representation and simulation of a simple network motif

You can model the network sketched on the left side using several "update" functions (middle) and

(right). Discuss the differences of independent influence of *in* and *m* on *out* as compared to *in* and *m* combined onto *out*. Document your analysis e.g. using Jupyter.

Task 2 (Information and Data about the COVID-19 pandemics)

Read the Wikipedia Article (german and english version, https://de.wikipedia.org/wiki/COVID-19). What are possible bioinformatics and, more specific, systems biology questions you could identify wrt. the COVID-19? Are there aspects which interests you in particular and which you might be interested to follow up yourself via some computational modelling?

A lot of research is currently being done to understand but also to actually fight the ongoing pandemics. It might be crucial to communicate and publish new findings and hypotheses as early as possible to enable further research and the combination of results, hypotheses, and data. Of course preliminary results imply the risk of being wrong or even if ok being misinterpreted, sometimes with severe political and societal if not medical and health system consequences.

Task 3 (Text mining and fact collection about the COVID-19 pandemics)

In the article, preprints are mentioned, which gain - also in general, i.e. not connected to COVID-19 - ever more attention, importance, and impact. Several preprint servers are in use extending the public databases of published research (such as PUBMED and derivatives). For COVID-19 a particular outlet of the prestigious journal Nature is mentioned (https://www.nature.com/articles/d41586-020-00613-4, https://outbreaksci.prereview.org/), which aims at at least very quick preliminary peer-review of preprints to achieve better quality checks and avoid 'fake facts'.

Check out the preprint servers, e.g. bioRxiv and medRxiv. Search for and collect other sources of preprints. Mine the established preprint servers for COVID-19 and/or SARS-CoV-2 papers. Maybe related papers might also be relevant (lung diseases, related viruses, relevant proteins, relevant (maybe dual use) drugs). Some information is also only be made available via press releases (caution!) and interviews. Can these 'facts' be substantiated via respective evidence, e.g. available data or preprints?

Download the (most) relevant papers and resources (databases, links). Implement a small database to collect and search the downloaded papers and resources. Could relevant facts be extracted (and linked to the source)? https://publons.com/publon/covid-19 and https://pages.semanticscholar.org/coronavirus-research could be interesting resources as well.

Task 4 (Facts about the SARS-CoV-2)

Collect genomic and proteomic information about the SARS-CoV-2 virus (and maybe related coronaviruses), maybe starting again at wikipedia (https://de.wikipedia.org/wiki/SARS-CoV-2).

Collect information about the SARS-CoV-2 life cycle, in particular the interaction with the host(s).

Evolutionary information is also systems relevant.

Task 5 (Dynamics of the COVID-19 infections - I: SIR/SEIR/SECIR models of Flu epidemics)

There are quite a number of systems models around modelling and simulating the development of the COVID-19 pandemics. (Keyword: "Flattening the curve", see also in https://de.wikipedia.org/wiki/COVID-19, but there are *many* around). These models are called and based on variants of SIR/SIER/SECIR (Susceptible, Exposed, Carrier, Infected, Recovered) models. An application and good starting point to understand these models is the position paper of the german society of epidemiology (an der Heiden M, Buchholz U: Modellierung von Beispiel-Szenarien der SARS-CoV-2-Epidemie 2020 in Deutschland. — DOI 10.25646/6571.2) from Mar 20, 2020 with a number of references e.g. to the Imperial college study (Ferguson NM, Laydon D, Nedjati-Gilani G, et al.: on behalf of the Imperial College COVID-19 Response Team. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. DOI: https://doi.org/10.25561/77482, 2020) again with the references to original publications on modelling (flu) pandemics in the early 2000-ds.

Other examples are the Helmholtz model (Khailaie et al, 2020, doi.org/10.1101/2020.04.04.20053637) on which the Helmholtz position paper (Initiative "Systemische Epidemiologische Analyse der COVID-19-Epidemie") is based. Several other models which support their analysis and conclusions are listed there.

Another popular server provided by the Neher group in Basel (https://covid19-scenarios.org/)

Explain the model and the underlying methods. Reproduce the model and model simulations from the two models for the COVID-19 epidemiology. The Neher group provides a quite good documentation on their website. Moreover, the Neher model and code is available at github (https://github.com/neherlab/covid19_scenarios/).

Task 6 (Dynamics of the COVID-19 infections - II)

A model targeted to the development in France is the Colizza INSERM study (Di Domenico et al, Expected impact of lockdown in Ile-de-France and possible exit strategies, Report #9, medRxiv.org).

Implement the Colizza model.

Compare the Neher and Colizza models via appropriate parametrization and simulation. Discuss the resulting predictions, the different outcomes and the possible consequences.

Task 7 (Sequencing the Corona virus)

What is known about Corona sequencing. There are already about 10.000 sequences of SARS-CoV-2 genomes. What is known about their differences? What is known about the SARS-CoV-1 virus and the MERS virus genomes. What are the differences between these species: genomic organisation, proteins, variations, ...? What are the phenotypic or life cycle differences between these viruses. Can they or the host specificities be attributed to genomic or sequence differences?

Task 8 (Evolution of the Corona virus)

Checkout the nextstrain.org website (also from the Neher lab) and the GSIAID resource (which collects the flu and CoV genome sequences) learn about the genomes of the Corona virus and the regional spread of the SARS-CoV-2 viruses. What can be learned from the genomic phylogenies and their locational patterns. Do the mutations coincide with locational patterns?

Task 9 (Structure of the Corona virus)

The structure of the virus is pretty well-known and several highly resolved structures for important virus proteins are known. How many? Can structures be modelled based on homologous structures?

What are the differences of respective structures from the different viruses and what can be learned from the differences between the viruses.

A neutralizing antibody (from a convalescent SARS patient) in complex with the receptor-binding domain (RBD) of the SARS-CoV-2 spike (S) protein has been resolved to 3.1 A. [Yuan et al., A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV, Science 10.1126/science.abb7269 (2020).]

What is structurally known about the interaction of virus and host proteins [Xu et al, Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modelling of its spike protein for risk of human transmission, March 2020 Vol.63 No.3: 457-460 https://doi.org/10.1007/s11427-020-1637-5]

Task 10 (Statistical and machine learning modeling)

Epidemiologists and statisticians have analysed various parameters and outcomes of the pandemics. Also colleagues from the statistics department of our faculty have constructed such a statistical model (https://www.uni-muenchen.de/forschung/news/2020/kuechenhoff_corona.html, Günther et al, Nowcasting the COVID-19 Pandemic in Bavaria, April 16, 2020). The LMU StaBLab employs generalized additive models and Bayesian "nowcasting" to estimate important system parameters such as the famous reproduction number R_0, here time-dependent numbers $R(t)$. Explain the model and analyse

the model predictions e.g. for Bavaria at different days of forecasting (based on more or less data). Discuss!

Task 11 (First view on Corona virus-host networks)

Networks of virus proteins and virus-host factors have been determined, visualized and analysed. This has been done for various viruses such that differences and similarities could possibly be exploited. E.g. one could try to propose drugs which target one of these proteins in order to disturb these interactions and the associated network.

Bioinformatics professor Jan Baumbach (TU Munich) tweeted: "CoVex went online. It's the first COVID-19 / SARS-CoV-2 drug repurposing prediction online tool using integrated systems and network medicine approaches. We report in our special blog until publication. It's online available here: https://exbio.wzw.tum.de/covex/"

Analyse and visualize the CoV-Host(human) interaction networks.

There are also databases available which connects drug molecules with possible viral target proteins (e.g. CORDITE http://corona.mathematik.uni-Marburg.de from the Heider Lab). Analyse and visualize such data available from web-based resources.