



LibEpidemic: An Open-source Framework for Modeling Infectious Disease with Bigdata

Honghao Shi

School of Computer Science and Engineering, Beihang University, Beijing, China
shihonghao@buaa.edu.cn

Qijian Tian

School of Computer Science and Engineering, Beihang University, Beijing, China
tqj19@buaa.edu.cn

Jingyuan Wang*

School of Computer Science and Engineering, Beihang University, Beijing, China
Pengcheng Laboratory, Shenzhen, China
jywang@buaa.edu.cn

Jiawei Cheng

School of Computer Science and Engineering, Beihang University, Beijing, China
JarvisC@buaa.edu.cn

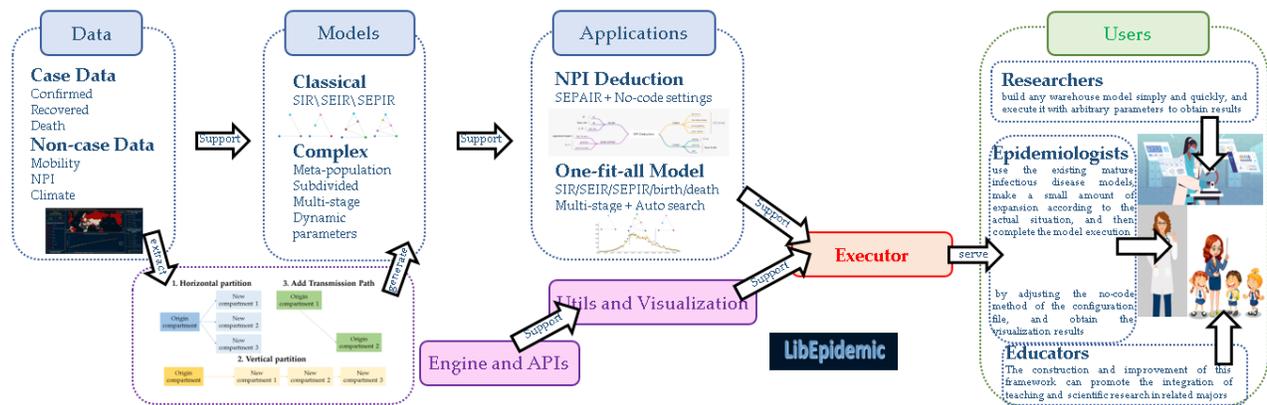


Figure 1: The Overall Structure of LibEpidemic: Elements, Framework and Applications

ABSTRACT

With increased human mobility and the introduction of NPIs, the complex, dynamic spread of COVID-19 has diverged significantly from SEIR's single, static assumption. At the same time, the ability to obtain front-line data also limits the modeling capabilities of SEIR. For researchers who cannot program, they must find suitable collaborators to implement their research. Even for researchers who can program, they need to repeat the principle and application process of the infectious disease model. LibEpidemic provide an open-source framework for modeling infectious disease, especially

* Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CIKM '22, October 17–21, 2022, Atlanta, GA, USA

© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9236-5/22/10...\$15.00
<https://doi.org/10.1145/3511808.3557183>

COVID-19, with bigdata. Researchers can implement subdivided, multi-stage or even metapopulation with the support of LibEpidemic.

CCS CONCEPTS

• Computing methodologies → Simulation tools.

KEYWORDS

epidemic modeling, no-code, open-source framework, bigdata embedding

ACM Reference Format:

Honghao Shi, Qijian Tian, Jingyuan Wang*, and Jiawei Cheng. 2022. LibEpidemic: An Open-source Framework for Modeling Infectious Disease with Bigdata. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22)*, October 17–21, 2022, Atlanta, GA, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3511808.3557183>

1 OVERALL STRUCTURE OF LIBEPIDEMIC AND FEATURES TO BE DEMONSTRATED

The Homepage of LibEpidemic is <https://libepidemic.github.io/#/>, which is also the URL of the screencast video of LibEpidemic. The Github repository of LibEpidemic is <https://github.com/Bigscity-epidemic/Epidemic-Modeling-survey>. The document website for LibEpidemic is <https://libepidemic-docs.readthedocs.io/en/latest/>.

Figure 1 shows the elements, framework and applications of LibEpidemic. LibEpidemic provides Engine and APIs for building, executing and visualizing epidemic models. LibEpidemic integrating Case and non-Case data, then use engine and APIs to make models from it. LibEpidemic preset 12 models, while users can build their own models with the support of engine and APIs in LibEpidemic. Researchers, epidemiologists and educators are targeted users of LibEpidemic.

2 INTRODUCTION AND BACKGROUND

After the outbreak of COVID-19, the SEIR model has shown two limitations as an infectious disease model: it cannot model realistic scenarios and dynamics.

Although the SEIR model models the whole process of contact-exposure-onset-removal of the development of infectious diseases, it is too ideal for the assumptions of compartments and individuals. This is reflected in:

- Individuals in the same compartment are identical. For example, infected individuals transmit the disease to susceptible individuals at an average rate, and each individual has the same importance in the transmission chain.
- Each individual is a "guinea pig" without subjective initiative. Individuals will not change their action strategies or formulate non-pharmacological interventions(NPIs, similarly hereinafter) according to the development of the epidemic.
- The compartment is set according to the principle of the epidemic, not the actual observation data. For example, the infected person's compartment is set, but only the confirmed data can be obtained in reality, and the error of approximate substitution is unignorable.

In order to solve these limitations, the researchers improved the infectious disease model represented by SEIR from two perspectives: Modeling the dynamics and Modeling more realistic scenarios.

Modeling the dynamics means that the epidemic model varies with external factors such as spatial factors, temporal factors, and characteristics, while modeling more realistic scenarios requires designing the settings and transformations between of the compartments. Compared with traditional infectious disease models, these methods have a more complex structure of the compartment, and the parameters and models are not single, static. Therefore, a modeling framework like this became a requirement:

- The compartment structure should be determined through several APIs, even in no-code way
- The differential equation system in the compartment model should be determined by the user in the form of natural language, not coding these discipline
- After the structure, equations and parameters of the model are determined, the model should be executed automatically

and one-click until the data results are given and the visual feedback is completed

In response to the above needs, LibEpidemic was born. On the basis of satisfying all the above three requirements, he built 12 finished models and 4 teaching models. At the same time, LibEpidemic also provides mature, out-of-the-box applications for scholars in another field: epidemiologists.

3 ENGINE AND APIS OF LIBEPIDEMIC

3.1 Determine Structure

There are 3 APIs in LibEpidemic for building the compartment model structure: vertical divide, horizontal divide and add path. With these three APIs, It is easy to determine strcuture of model that built by LibEpidemic.

Vertical divide. Vertical divide can subdivide compartments by information relevant to the epidemiological process, such as asymptomatic, presymptomatic, confirmed, isolated, etc. And such subdivide supports the modeling of individual differences within the same compartment. Figure 2 is a general view of vertical divide.

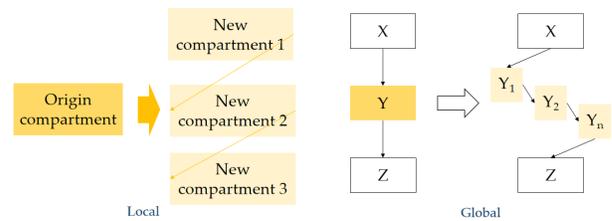


Figure 2: Vertical Compartment Subdivision: it expands the original single compartment into a chain, the chain's head connects the original prefix compartments, and the tail connects the original suffix compartments

From a local perspective, vertical divide subdivides a compartment into three new compartments. From a global perspective, vertical divide adds stages in the epidemiological process.

Horizontal divide. Horizontal Divide can subdivide compartments by information not relevant to the epidemiological process, such as age, gender, country, etc. And such subdivide supports the modeling of individual differences within the same compartment. Figure 3 is a general view of horizontal divide.

From a local perspective, vertical divide subdivides a compartment into three new compartments. From a global perspective, vertical divide subdivides different compartments in one stage of the epidemiological process.

Add path. Add Path can add a transmission, assign a formula and parameters to it. It is usually used in combination with the two methods above. Figure 4 is a general view of horizontal divide:

3.2 Multi-stage model and Dynamic Parameter

Multi-stage model refers to the use of different models to model the epidemic according to certain rules. The dynamic properties of multi-stage models are usually reflected at the time point when NPIS changes significantly. In libepidemic, different models can be

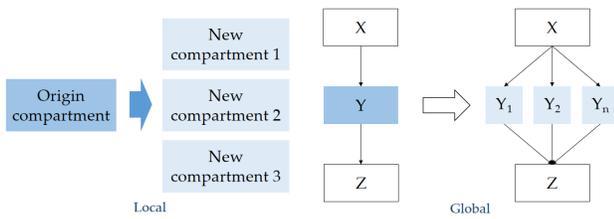


Figure 3: Horizontal Compartment Subdivision: it expands the original compartment into several compartments with the same status, all of which are connected to the original prefix and suffix compartments

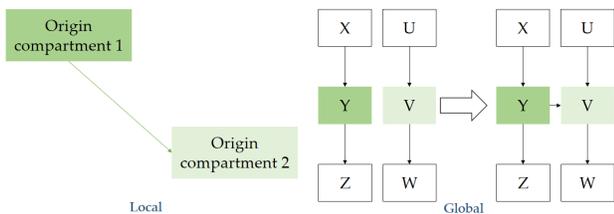


Figure 4: Add Path: it assists the first two APIs

built according to the changes of NPIS, and these models can form a multi-stage model by stages that meet the modeling requirements with obvious stage differences.

Dynamic parameters refers to the use of external data to transform the parameters of infectious disease model from a fixed value to an array or even a matrix. Dynamic parameters is the basis of modeling. It focuses on reviewing and analyzing the past epidemic situation, and carries out model experiments with the model. In LibEpidemic, dynamic parameters is realized by embedding. Embedding can replace the fixed parameters in the model with an array. The values in the array can be set in advance or changed in the process of model calculation to realize the effect of dynamic parameters.

3.3 Workflow of LibEpidemic

In LibEpidemic, the APIs determine the structure of the model first. In this process, the compartments contained in the model and the transfer paths between compartments should be determined. And the corresponding transfer equations for the transfer paths need to be set to determine the states of compartments. LibEpidemic’s interpreter will translate the equations into the transfer paths between compartments. And the structure of the model can be established. Next, LibEpidemic need epidemic data data to fit the model parameters and set the initial values. And the whole model will be executed by the executor. LibEpidemic will continue to deduce and iterate on the basis of the initial values, and finally complete the prediction.

4 MODELING STEPS IN LIBEPIDEMIC

LibEpidemic is a general and simple modeling framework. By using LibEpidemic, users can easily build personalized models and it has various usages.

4.1 How to construct SEIR with LibEpidemic

To use LibEpidemic to build a model, first initialize a directory and create a python file. If users only need to experience the most basic functions, such as building a basic SEIR model to familiarize themselves with the usage of LibEpidemic, users can also simply create a single file. However, doing so will clutter the file structure and interfere with the subsequent use. Therefore, we do not recommend this. Users need to make sure to have successfully installed the environment dependencies in requirements. To make sure the environment is well, users can type a print information. Then a response without errors can be seen., which means that the environment is configured correctly and users can start building infectious disease models using APIs.

To construct SEIR with LibEpidemic, the first step is create structure of the model. APIs for determining structure can be used. Then, users need to assign a formula to each edge in the struct to implement the dynamics in the infectious disease model. The interpreter formally transforms the previously established SEIR model structure into a complete SEIR model, including kinetic mechanisms and parameters. First, compare the differential equation form of the SEIR model. LibEpidemic’s engine and compilation system support arbitrary expressions consisting of multiplication and addition. This can cover the vast majority of scenarios in infectious disease models.

To get the final results, the model need to be executed by LibEpidemic’s engine. If things go well, users can get the following results from standard output.

4.2 Users’ need and LibEpidemic’s function

LibEpidemic is a general modeling framework, which can be applied to scientific research and realize complex epidemic modeling. Because of the user-friendly design, it is convenient for epidemiologists who do not write programs to use, which has strong practical value. In addition, libepidemic can also be used to build a simple and classic model, which can be applied in the field of education.

Function for Research. Researchers can use the framework to build any warehouse model simply and quickly, and execute it with arbitrary parameters to obtain results. They need to learn the whole engine-model-application three-level framework, including engine, engine-model, model structures and model parameters.

Function for Epidemiologist. Epidemiologists need to use the existing mature infectious disease models, make a small amount of expansion according to the actual situation, and then complete the model execution by adjusting the no-code method of the configuration file, and obtain the visualization results.

Based on this model, epidemiologists can introduce a subdivided model to model any factors that may affect the spread of the epidemic, such as age, income, country, population mobility, etc. , or review or deduce the effects of arbitrary policies(More than 20 categories in total, covering the vast majority of specific policies for all countries) by writing code-free configuration files.

Function for Education. The construction and improvement of this framework can promote the integration of teaching and scientific research in related majors. The framework development team has cooperated with the School of Computer Science and Engineering, Beihang University to carry out the experimental

course “scientific research classroom”, using the framework to carry out the teaching work.

5 APPLICATIONS IN LIBEPIDEMIC

5.1 NPI Modeling and Deduction in LibEpidemic

Non-pharmaceutical interventions (NPIs) are the factors by which humans, mainly governments or rulers, proactively propose measures to intervene in epidemic.

LibEpidemic use SEPIAR model and the deduction library to help you modeling NPI. SEPAIR has 6 compartments: S(susceptible), E(exposed), P(presymptomatic), A(asymptomatic), I(infectious) and R(removed). Among them, P, A and I have the ability to transmit diseases. The deduction library consists of several items, each of which contains a policy, a 0/1 flag, several data and several parameters. Among them, the policy is described by a natural language string, and the 0/1 mark indicates whether the policy participates in the deduction.

Occasional and sudden outbreaks have the characteristics of “short, fast, small, and hidden”. For this type of epidemic, epidemiologists prefer to know intuitively and concretely what the results of the policies they have released will be[5]. LibEpidemic divides all policies into social control, nucleic acid test and isolation categories according to the principle of policy action.

5.1.1 Social Control. LibEpidemic subdivides social management and control policies based on the classification of the Oxford government response index. Considering the actual application scenarios, 6 of them are reserved for modeling:

- School Closing
- Workplace Closing
- Public Events Cancelling
- Restrictions on Gatherings
- Public Transport Closing
- Staying at Home

LibEpidemic uses a simple and trained GBRT model to obtain 6-dimensional policy, with covariates such as population and GDP, for changes in $\Delta\beta$, where β means the transmission parameter in SEPAIR model.

Taking Beijing as an example, when Beijing launched the secondary response-related policy group, $\beta=-0.85$, which means that the secondary response policy group in Beijing will control the comprehensive population activity intensity to 15% of the previous level.

5.1.2 Nucleic Acid Test and Isolation. Nucleic acid testing and isolation policies include two categories: standing policies and emergency response policies

The impact of the standing policy is reflected in the scale and speed of potential hidden transmission before the first confirmed case is diagnosed. In LibEpidemic, the following standing policies are modeled:

- Periodic testing of high-risk groups
- Close isolation time
- How long it takes to track the close connection
- If asymptomatic infected persons quarantined

Emergency response policies are used to deduce the results of different possible policies after the outbreak. Such policies are dynamic and user configurable, and LibEpidemic provides a no-code adjustment method. LibEpidemic models the following emergency response policies:

- Global nucleic acid detection
- Block Closure
- Isolation of Sub-Close and Space-Time Companions

5.2 Existing Models Implemented by LibEpidemic

5.2.1 “One-fits-all” Model. LibEpidemic has 5 complete basic models built in: SEIR, SIR with natural birth/death, SEIR with natural birth/death, and SEPIR. They come from textbooks or cups for early COVID-19 predictions. Users can view and modify the code of these basic models, fill in any parameter settings, complete the model simulation, and view the results

Considering that the inputs to the underlying SEIR model are similar, LibEpidemic provide Mixed-strategy Model to Epidemiologists. The mixed-strategy model is used with the multi-stage method, and the optimal model in the model library is automatically selected through the given loss function in each stage.

Mixed-strategy models can be used to build a “one fits all model”. Using this model, combined with multi-stage global predictions (tests), the total error is less than or equal to any single model.

5.2.2 Complex Models from Academic Papers. In order to better help users get started with LibEpidemic, and to verify the simplicity and efficiency of LibEpidemic for infectious disease modeling, LibEpidemic reproduced the compartment model used in 5 papers[1–4, 6], and built 4 SEIR-class teaching models. Figure 5 shows the compared results from the origin paper(left) and LibEpidemic(right).

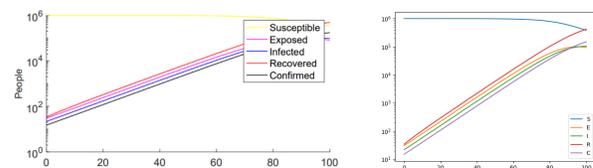


Figure 5: Compared Results: the trend of the curve and the magnitude of the absolute value

6 ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2021ZD0111201), the National Natural Science Foundation of China (Grant No. 82161148011, 72171013, 72222022), and the Fundamental Research Funds for the Central Universities (Grant No. YWF-22-L-838).

REFERENCES

- [1] Rachel E Baker, Wenchang Yang, Gabriel A Vecchi, C Jessica E Metcalf, and Bryan T Grenfell. 2021. Assessing the influence of climate on wintertime SARS-CoV-2 outbreaks. *Nature communications* 12, 1 (2021), 1–7.
- [2] Tamma Carleton, Jules Cornetet, Peter Huybers, Kyle C Meng, and Jonathan Proctor. 2021. Global evidence for ultraviolet radiation decreasing COVID-19 growth rates. *Proceedings of the National Academy of Sciences* 118, 1 (2021).
- [3] Mario Castro, Saúl Ares, José A Cuesta, and Susanna Manrubia. 2020. The turning point and end of an expanding epidemic cannot be precisely forecast. *Proceedings of the National Academy of Sciences* 117, 42 (2020), 26190–26196.
- [4] Rahul Subramanian, Qixin He, and Mercedes Pascual. 2021. Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proceedings of the National Academy of Sciences* 118, 9 (2021).
- [5] Xiao-Li Wang, Xin Lin, Peng Yang, Zun-You Wu, Gang Li, Jennifer M McGoogan, Zeng-Tao Jiao, Xin-Jun He, Si-Qi Li, Hong-Hao Shi, et al. 2021. Coronavirus disease 2019 outbreak in Beijing’s Xinfadi Market, China: a modeling study to inform future resurgence response. *Infectious diseases of poverty* 10, 1 (2021), 1–10.
- [6] Joshua S Weitz, Sang Woo Park, Ceyhan Eksin, and Jonathan Dushoff. 2020. Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proceedings of the National Academy of Sciences* 117, 51 (2020), 32764–32771.