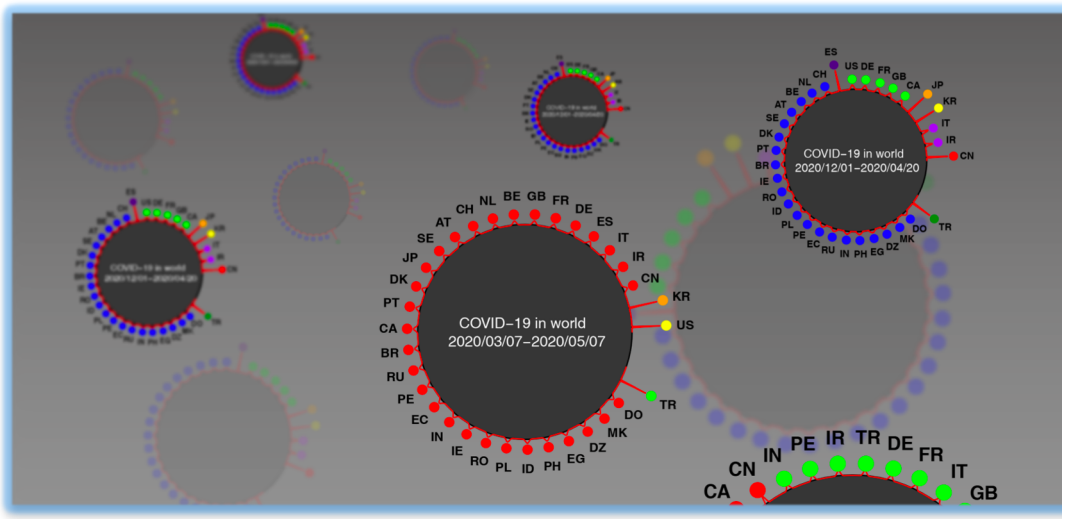
Working paper on graph clustering

Help documentation on using the features offered by the graph clustering system

Graph Clustering System (GCS)

Supplemental document

Version 1.0
December 24, 2020



Xiaoping Shi, Department of Mathematics and Statistics,
Thompson Rivers University, Kamloops, BC, Canada V2C0C8.
Meiqian Chen, Center for Network Big Data and Decision-Making,
Business School, Sichuan University, Chengdu, China 610065.
Yucheng Dong, Center for Network Big Data and Decision-Making,
Business School, Sichuan University, Chengdu, China 610065.

Supplemental documents on the GCS

Version 1.0

Xiaoping Shi, Department of Mathematics and Statistics,
Thompson Rivers University, Kamloops, BC, Canada V2C0C8.
Meiqian Chen, Center for Network Big Data and Decision-Making,
Business School, Sichuan University, Chengdu, China 610065.
Yucheng Dong, Center for Network Big Data and Decision-Making,
Business School, Sichuan University, Chengdu, China 610065.

The GCS system and working paper are updated frequently, please check the latest version of GCS at http://graph-clustering-system.com/.
Detailed documentation of the corresponding data and code can be found on the GitHub repository: https://github.com/Meiqian-Chen/GraphCpClust.

Abstract

Graph Clustering System (GCS) is a free platform that provides users with graph clustering analysis, change point analysis and data fitting based on the clustering-segmented autoregressive sigmoid (CSAS) model. The platform not only reproduces the results of data analysis based on multiple datasets, but also serves as a tool to provide users with the corresponding data analysis.

Please cite: *SHI, X., CHEN, M.Q. & DONG, Y.C. Exploring the space-time pattern of log-transformed infectious count of COVID-19: a clustering-segmented autoregressive sigmoid model. arXiv. https://arxiv.org/abs/2102.13287*

Analysis of three datasets

This section shows graph clustering analysis, change point analysis and fitting analysis based on three COVID-19 datasets: 1. Our world in Data at https://ourworldindata.org/covid-data-switch-jhu , 2. WHO at https://covid19.who.int/, 3. Wuhan-2019-nCoV at https://github.com/canghailan/Wuhan-2019-nCoV). In our paper, only the last dataset, Wuhan-2019-nCoV, is analyzed in our paper.

Why did we analyze the Wuhan-2019-nCoV dataset? There are two main reasons here.

(1) We conducted a study of COVID-19 back in early March 2020. At that time, there were very few datasets on COVID-19, and especially few containing timely outbreak data from Chinese provinces. This Wuhan-2019-nCoV dataset was updated very timely at the beginning of COVID-19 transmission.

(2) The data quality of Wuhan-2019-nCoV dataset is trustworthy and has been included in the "Open Source Wuhan" data resource. The data about COVID-19 in each country are from WHO's epidemic reports, while the data about COVID-19 in each province of China are from the daily epidemic reports of provincial health and family planning commissions.

Missing data handling:

For these three datasets, we took the following measures for missing data. For some countries with missing early data, we replaced them with zeros; however, in the case of other missing data, the previous data of the series was used instead.

Data intervals:

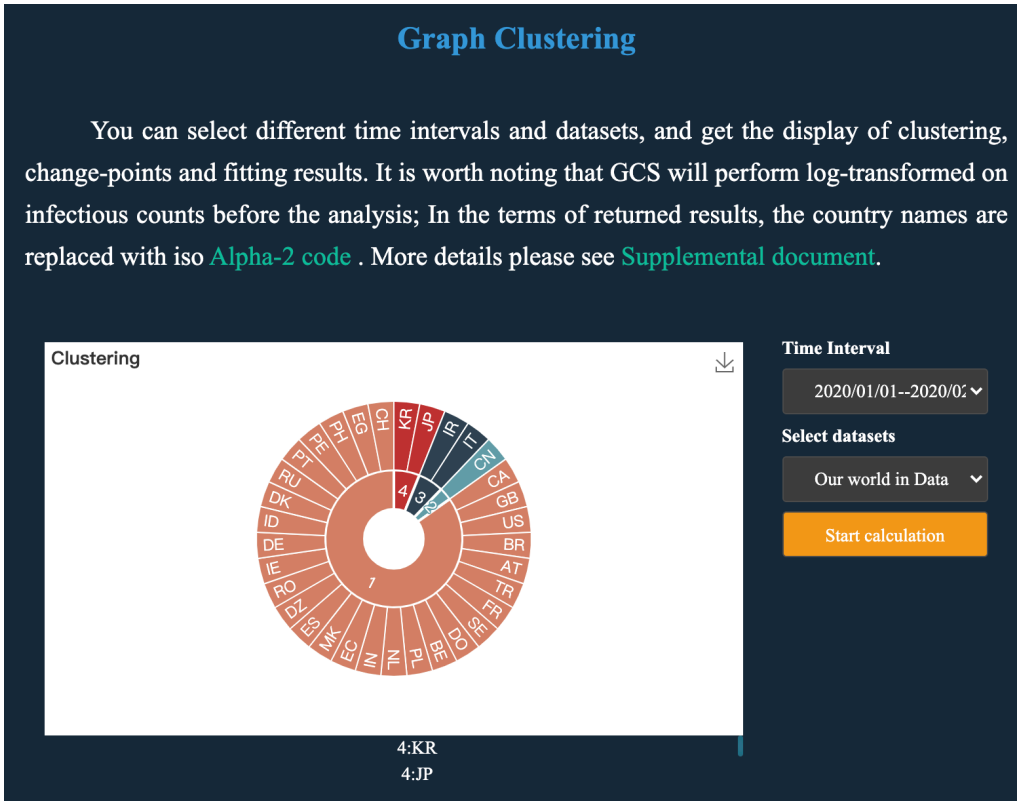
The data intervals for the three datasets considered are as follows.

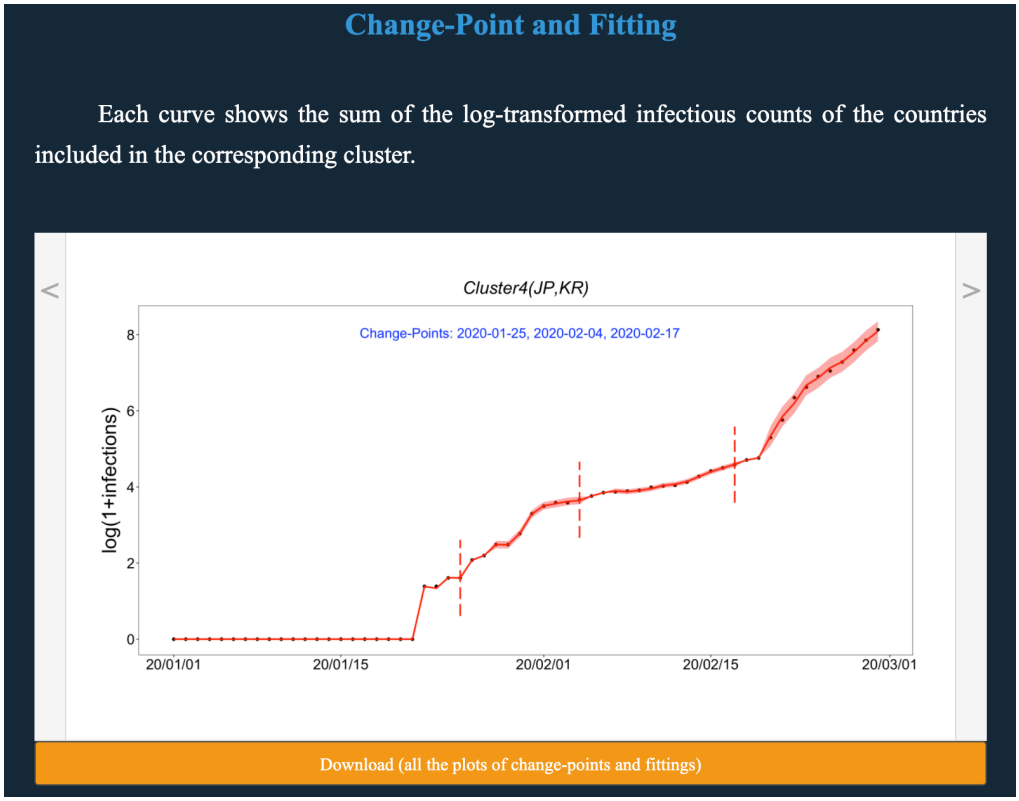
Our world in Data: 2020/01/01 to 2020/12/03.

WHO: 2020/01/01 to 2020/12/03.

Wuhan-2019-nCoV: 2019/12/01 to 2020/12/03.

A showcase for the clustering and change-point analysis of the epidemic data





Step 1: Please select a time interval that you want to show.

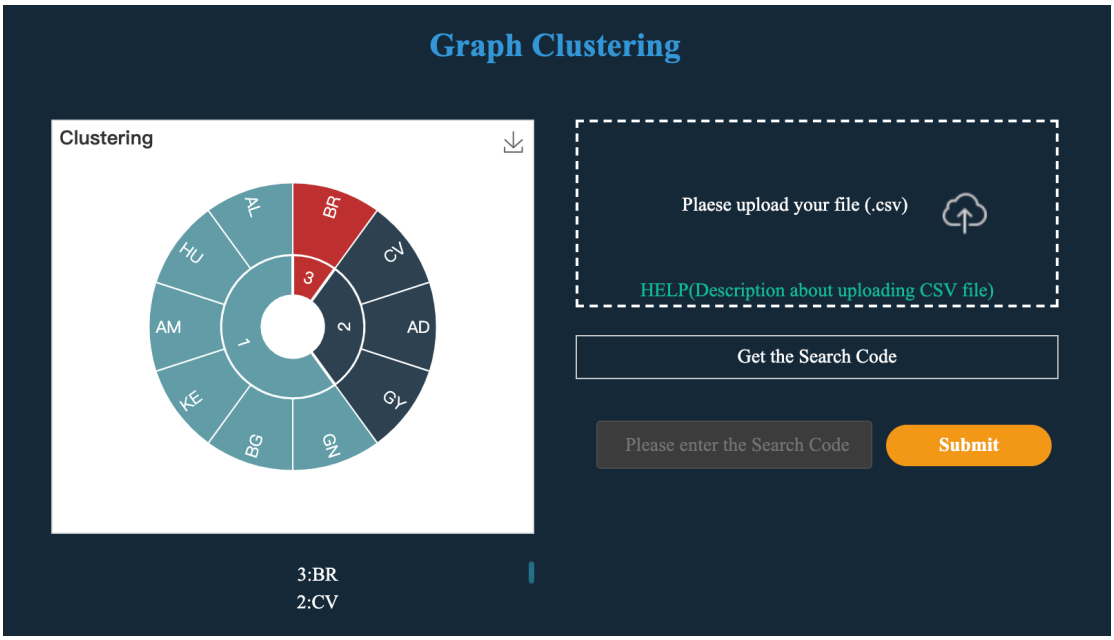
Step 2: Select a dataset from Our World in Data, World Health Organization and Wuhan-2019-nCoV.

Step 3: Click on the "Start Calculation" button to get the display of the clustering results,

simultaneously, a sigmoid curve of each cluster is also presented that share the form of multiple stages and multiple change-points.

Offline requests

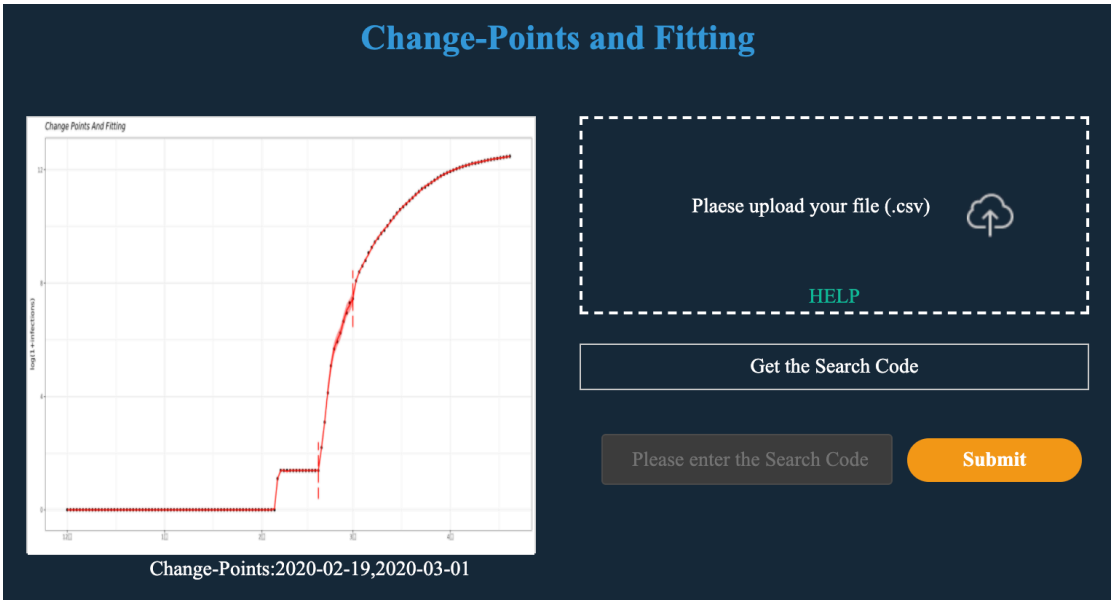
Steps for offline requests



Step1: Please upload your file, which must be in csv format. For details about uploading files, please see "Help (Description about uploading CSV file)".

Step2: After uploading the file, the "Search Code" for this request will appear in the "Get Search Code" window.

Step3: Then enter the search code in the "Please enter the search code" window and submit it to get the display of the clustering results.



Step1: Please upload your file, which must be in csv format. For details about uploading files, please see "Help (Description about uploading CSV file)".

Step2: After uploading the file, the "Search Code" for this request will appear in the "Get Search Code" window.

Step3: Then enter the search code in the "Please enter the search code" window and submit it to get the display of the change-points and fitting results.

Note: It may take longer to calculate the change points and fit the data, so you may have to wait a few minutes to get the corresponding Search Code.