

ReadMe_Hsuan Lee

h.lee@uu.nl

Research Archive Structure

- Simulation_study_1

- Pre_processing

- [\[1\]Synthetic_Data](#) → the code used to simulate synthetic social network data.
- [\[2\]MW_Parameters](#) → the code for fitting the MW-REM to each synthetic dataset, and extracting the model parameters from each window.

- Data

- [Synthetic_Data](#) → the stored synthetic social network data.
- [Window_Parameters](#) → the stored extracted parameters along windows of each effect after fitting the MW-REM on each synthetic dataset. These parameters represent the estimated values of the effects over time, and were used as input for the changepoint detection algorithms.

- Scripts

- [\[3\]Changepoint_Detection](#) → the code for conducting the changepoint detection task of three algorithms on the extracted parameters.
- [\[4\]Comparison_of_CPDs](#) → the code for comparing the three algorithms' performance.

- Output

- [CP_Detection](#) → the stored detected changepoint locations from the [\[3\]Changepoint_Detection](#).
- [CPD_Comparison](#) → the stored comparison results from the [\[4\]Comparison_of_CPDs](#).

- Real_Life_study_2

- Pre_processing

- [\[5\]RealLife_MW](#) → the code for fitting the MW-REM to the Apollo 13 data, and extracting the model parameters from each window.

- Data

- [RealLife_data](#) → the Apollo 13 voice-loop data.
- [Window_Parameters](#) → the stored extracted parameters along windows of each effect after fitting the MW-REM.

- Scripts

- [\[6\]Changepoint_Detection](#) → the code for conducting the changepoint detection task of three algorithms on the extracted parameters.

- [Output](#) → the stored detected changepoint locations from the [\[6\]Changepoint_Detection](#).

Introduction to the design of the study

All of the analyses in this study were conducted using R version 4.2.2.

1. Simulation_study_1:

The objective of this study was to introduce and compare three state-of-the-art changepoint detection algorithms in the context of the Moving Window Relational Event Model (MW-REM), a social network

model that captures time-stamped social interactions. We started by simulating synthetic datasets with no changepoints, one changepoint, and two changepoints, following the process outlined in the "[1]Synthetic_Data" file in the "Pre_processing" folder. The resulting data was stored in the "Synthetic_Data" folder within the "Data" directory. Next, we applied the MW-REM to each synthetic dataset and extracted the parameters of the variables, also known as effects, involved in the model. These parameters were used as input to the changepoint detection algorithms. The process of extracting the parameters is described in the "[2]MW_Parameters" file in the "Pre_processing" folder, and the results were stored in the "Window_Parameters" folder within the "Data" directory.

Having obtained the synthetic data and window parameters, we fed the parameters of each effect of each dataset to the three changepoint detection algorithms we utilized: Binary Segmentation, Bayesian Online Changepoint Detection, and Pruned Exact Linear Time. The process of detection for each algorithm can be found in "[3]Changepoint_Detection" in the "Scripts" folder, and the results of detected changepoint locations were stored in the "CP_Detection" folder under "Output."

To compare the performance of the algorithms, we employed the confusion matrix, mean square error, and mean signed difference. The comparison was based on the performance of each algorithm on datasets with no changepoint, one changepoint, and two changepoints, respectively. The comparison processes can be found in "[4]Comparison_of_CPDs" in the "Scripts" folder, and the comparison results were stored in the "CPD_Comparison" folder under "Output."

***Note:**

(1) In the "Pre_processing" folder, two users have recently reported **issues with downloading** the packages ("REHdynamics", "remstats", and "remify") that are required to run the simulation and fit the MW-REM. Fortunately, we have already stored the simulation and extracted parameters in the "Data" directory. Therefore, it is not necessary to run the "pre_processing" folder.

(2) In the "[1]Synthetic_Data" file located in the "Pre_processing" folder, even if you successfully download the necessary packages, the simulation process takes **50 to 55 hours** to complete. Similarly, in the "[2]MW_Parameters" file located in the same folder, fitting the MW-REM for each dataset takes an additional 10 hours.

Therefore, we recommend against running the files in the "Pre_processing" folder due to the reported issues with downloading packages, and the long simulation and fitting times. These factors may make it impractical or infeasible for most users to run these files.

2. Real_Life_study_2:

To validate our simulation study with real-life data, we applied the MW-REM and changepoint detection algorithms to the Apollo 13 voice loop data. We hypothesized the location of the changepoint and tested whether all the algorithms could detect it. The data can be found in the "RealLife_data" folder of the "Data" directory. Similar to the simulation study, we fitted the MW-REM to the data and extracted the parameters. The entire process can be found in the "[5]RealLife_MW" file of the "Pre_processing" folder, and the extracted parameters were stored in the "Window_Parameters" folder of the "Data" directory.

We then ran the changepoint detection task using each of the algorithms. In the "[6]Changepoint_Detection" file located in the "Scripts" folder, you can see the entire process we used to detect the changepoints. The results of the detected locations of the changepoints for each algorithm were stored in the "Output" folder.

***Note:**

Similar to the simulation study, we do not recommend users to run the files in the "Pre_processing" folder due to package issues. However, as we have stored the necessary files in the "Data" directory, the script can be run independently.

Data Information

1. Simulation_study_1:

- Synthetic_Data (*FETC Case Number: 22-1870*)

The "*Synthetic_Data*" folder stores the time-stamped social network data that we simulated. It contains synthetic data with no changepoint, one changepoint, and two changepoints, respectively.

- Window_Parameters

The "*Window_Parameters*" folder contains the extracted parameters of each effect (i.e., variable) after fitting the MW-REM on the synthetic data. We selected four effects for inclusion in the model: sender effect, difference effect, inertia effect, and outdegree sender effect. You can observe the parameter fluctuations of each effect under different changepoint conditions in the synthetic data. These files were used as input for the changepoint detection algorithms.

2. Real_Life_study_2:

- RealLife_data (*FETC Case Number: 22-1871*)

The "RealLife_data" folder contains the Apollo 13 voice loop data that we used for validation. The complete Apollo 13 voice loop data is publicly available and can be found at: <https://github.com/issa-tseng/apollo13rt>.

- Window_Parameters

The "*Window_Parameters*" folder contains the extracted parameters of each effect (i.e., variable) after fitting the MW-REM on the Apollo 13 data. We selected seven effects for inclusion in the model: inertia effect, outdegree sender effect, indegree receiver effect, total degree sender effect, psABBA effect, psABBY effect, and psABXA effect. You can observe the parameter fluctuations of each effect under different changepoint conditions in the Apollo 13 data. These files were used as input for the changepoint detection algorithms.

Scripts Guide

1. Simulation_study_1:

- [3]Changepoint_Detection

This file contains the code for detecting changepoints using each of the changepoint detection algorithms on the synthetic data. Running these codes produced the detected changepoint locations for each dataset, and we stored the results in the "*CP_Detection*" folder within the "*Output*" directory.

Line 18 – 413: This section includes the code for each algorithm used to detect changepoints in synthetic data with no changepoint.

Line 414 – 762: This section includes the code for each algorithm used to detect changepoints in synthetic data with one changepoint in the inertia effect.

Line 763 – 1111: This section includes the code for each algorithm used to detect changepoints in synthetic data with one changepoint in all effects.

Line 1112 – 1460: This section includes the code for each algorithm used to detect changepoints in

synthetic data with two changepoints in the inertia effect.

Line 1461 – 1812: This section includes the code for each algorithm used to detect changepoints in synthetic data with two changepoints in all effects.

Line 1813 – 1934: This section presents a visualization of one of the synthetic datasets we generated, which helps readers better understand how the changepoint detection algorithms work on fluctuations in the effects parameters.

- [4]*Comparison_of_CPDs*

This file shows the code for comparing the performance of the three changepoint detection algorithms. We first created functions for the confusion matrix, mean squared error, and mean signed difference. Then, we computed the confusion matrix, mean squared error, and mean signed difference for each dataset. Finally, we gathered the results from each dataset to compute the overall performance of each algorithm on the confusion matrix, mean squared error, and mean signed difference in the no changepoint, one changepoint, and two changepoint conditions. we stored the results in the "*CPD_Comparison*" folder within the "*Output*" directory

Line 7 – 252: This section contains the code for creating functions to calculate the confusion matrix, mean squared error, and mean signed difference.

Line 53 – 2188: This section includes the code for computing the confusion matrix, mean squared error, and mean signed difference for each synthetic dataset. However, the primary goal of this section is to collect the results of each changepoint detection algorithm under different changepoint conditions for later analysis.

Line 2190 – 2432: This section includes the code for aggregating the results from each dataset to compute the overall performance of each algorithm on the confusion matrix, mean squared error, and mean signed difference for the no changepoint, one changepoint, and two changepoint conditions.

Note: We only reported and concerned the part from line 2190 to 2432.

2. *Real_Life_study_2:*

- [6]*Changepoint_Detection*

This file contains the code for detecting changepoints using each of the changepoint detection algorithms on the Apollo 13 data. We hypothesized that the changepoint location was between window 9 to 11. Running these codes produced the detected changepoint locations for each dataset, and we stored the results in the "*CP_Detection*" folder within the "*Output*" directory.

Line 18 – 290: This section contains the code for detecting changepoints using each of the changepoint detection algorithms on each effect of the Apollo 13 data.

Line 293 – 672: This section contains the code for visualizing the detected changepoint locations on each effect for each algorithm, making it easier to observe whether the algorithm captured the hypothesized changepoints or not.