Final Project: Catastrophic cascade of failures in interdependent networks

Eleonora Pura (17-732-678), Pietro Bonazzi (17-200-635), Songyi Han (18-796-847), Kim Hyeongkyun (21-732-797)

Network Science

Faculty of Business, Economics and Informatics

17.12.2021

ABSTRACT

Although past research has focused mainly on independent networks, modern systems such as water supply, transportation, and fuel and power stations are increasingly coupled together. These systems are properly represented by interdependent networks, where nodes from different networks are linked. This connection means that the failure of one or more nodes in a network can cause failures and fragmentation of another system. Following the approach presented in Buldyrev et al. 2010, this paper simulates a cascade of failures and analyse the robustness of some synthetic and real networks. First, cascade failure on random Erdős–Rényi and Scale-free networks is tested. Then, the same procedure is used on the Paris Multilayer Transport Network (Asgari et al. 2016). Results show some phase change around theoretical p_c for ER Networks, but confirmed the results of the paper for Scale-free and the Real world network.

1 INTRODUCTION

Research about complex networks usually focuses on the study of individual non-interacting networks. However, nowadays networks are increasingly coupled and dependent on each other. These types of networks are called interdependent networks. Examples can be found in the fields of water supply, transportation, and fuel and power stations. Given their high interconnectivity, these types of networks are very sensitive to random failures of individual nodes, to the extent that the removal of just a small part of them can cause an iterative cascade of failures in different networks (Buldyrev et al. 2010).

If, for example, a power failure occurs as a result of a catastrophe, the failure of a single microgrid may affect the failure of other dependent nodes in different networks, ultimately resulting in a cascade of failures (Amini et al. 2020).

The aim of this paper is to replicate a study previously presented by Buldyrev et al. 2010, in which the robustness of interdependent networks is investigated. In order to reach that goal, we are initially analyzing random Erdős–Rényi and Scale-free networks, and subsequently the Paris Multilayer Transport Network real dataset, consisting of subway, trains and roads components.

2 THEORY

In order to better understand this work, it is necessary to introduce the model previously presented in Buldyrev et al. 2010. The model considers two interdependent networks A and B with general topologies. Interdependent means that the functioning of some nodes in A depends on the ability of one or more nodes of B to provide an essential resource to the nodes in A and vice versa. The main intuition from the percolation theory is that after nodes have failed, the network must have a spanning cluster of functional nodes in order to remain functional.

Instead, nodes that are not included in the spanning cluster are considered to be nonfunctional along with all the nodes connected with them from other networks.

Let's consider two networks A in B with N randomly and independently connected nodes and degree distributions $P_A(k) and P_B(k)$. Each node in A is connected and dependent on one node in B and vice versa. If 1-p nodes are removed from A we have to remove all the edges connected to them and all their dependent nodes and edges in B. With the progressive removal of edges only the nodes p remain functional, and the networks break into clusters/connected components. A mutually-connected cluster is a set of nodes in A which belong to a cluster in A and also have their dependent nodes in B belonging to a single cluster in B (or vice versa). The network core is defined as the giant component: mutually-connected elements inside it are of particular interest for the model because they are still functional. Instead, all the nodes disconnected to the giant component are considered to be non-functional and are therefore removed.

The fundamental questions that this model wants to answer are the following:

1. What is the critical $p = p_c$ below which no mutual giant component exists (all the mutual clusters are only an infinitesimal fraction of the network)?
2. What is the probability $p_\infty(p)$ for a node to belong to the mutual giant component?

The cascade of failure is a recursive process in which the model can be solved. Let's define a1-clusters in A after only a number p of nodes are left. The first stage in the cascade of failures consists in treating each a1-cluster as a subset of network B: in this way all the B-edges connected with the a1-cluster are removed. In the second stage of the cascade of failures it is possible that each a1-cluster defined earlier has split into many b2-clusters or that the a-1-clusters coincide with the b2-clusters (mutually connected). Similarly to step one, this time all edges of A that are connected to the different b2-clusters will be removed. This continues with the definition of a3-clusters in the third stage and b4-clusters in the fourth and so on, until no additional splitting and removal of edges occur. A visualisation of the model is given in figure 1.

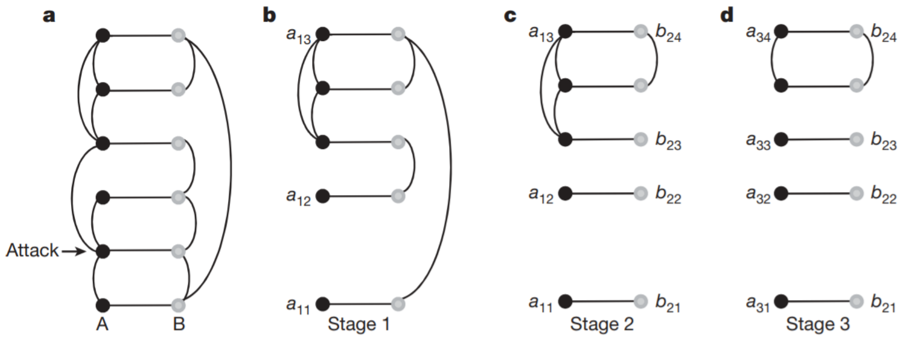

Fig. 1. Iterative process of a cascade of failures from Buldyrev et al. 2010.

At the beginning each node in network A is linked to one node in network B with an horizontal line, showing the dependence. In **stage 1**, when a node in network A is attacked, both the A node and the dependent node in B are removed together with all their direct links. This has the consequence that network A breaks into three a1-clusters, denoted with a11, a12 and a13. In **stage 2** all the B links that connect sets of B-nodes linked to separate a1-clusters are eliminated. As a consequence, network B breaks in four clusters b21, b22, b23 and b24. Finally, in **stage 3** all links in A that connect sets of A-nodes linked to separate b2-clusters are removed, causing the network A to break in four additional a3-clusters a31, a32, a33 and a34. No further link elimination is needed, each b2/a3-clusters connected pair is a mutually connected cluster, where b24 and a34 are the largest ones and therefore constitute the giant mutually connected component.

3 DATASET DESCRIPTION

For the scope of our project, we generated two types of random networks, namely Erdős–Rényi and Scale Free Networks. In 2, an example for both type of random networks with $N = 50$ and $\langle k \rangle = 4$ are showed. Successively, we wanted to apply and analyse the cascade failure into the real world Network (Paris transportation).

3.1 Random Network Model

3.1.1 *Erdős–Rényi Network Dataset*

Erdős–Rényi (ER) model is a random network generating model which has a Poisson degree distribution. We used the networkx library to generate this ER model graph. At first, three sets of two ER networks (Network A and Network B) were generated. Each set has different node sizes, N=500/1000/2000, but all ER networks in every set have a fixed average degree $\langle k \rangle = 4$. To unify Network A and Network B as an interdependent network, we took the simple assumption that each node in network A is dependent on one randomly selected node in network B and vice-versa.

3.1.2 *Scale Free Network Dataset*

The second random networks are Scale-free Netowrks (SF). Compared to ER model, SF model has a power-law degree distribution $P(k) \sim k^{-\lambda}$ and we wanted to see the results of cascading failure based on the different $\lambda = 2.7/2.8/2.9$. To demonstrate this, the node size was fixed to 1000 and the average degree was also fixed to $\langle k \rangle = 4$. Because there was no built-in function in networkx library to generate SF network under this specific condition, we created the SF generation function which was able to get λ and $\langle k \rangle$ as input parameters.

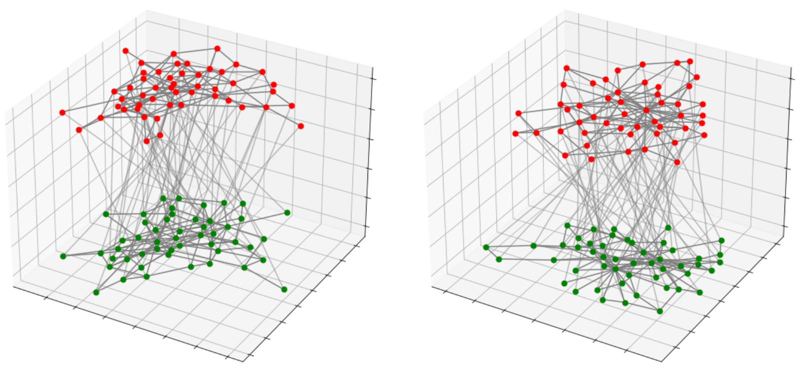

Fig. 2. Random Network Examples visualization. L: ER(N=50, $\langle k \rangle$ = 4), R: SF(N=50, λ =3, $\langle k \rangle$ = 4)

3.2 Real-World Network Dataset

3.2.1 *Paris Transportation System*

In Buldyrev et al. 2010 a close-source power network was tested in relation with a close-source internet network. Instead, in this study, we focus on the open-source Paris Transport Network presented by Asgari et al. 2016. Thus, all results shown in the paper can be easily reproduced [1].

The dataset is a multimodal transportation network of Ile-de-France, composed by separate graphs each indicating a different transportation mode. These networks are interconnected into a multi-layer graph G. In this paper, the rail transport network for train and metro, extracted from OpenStreetMap (OSM) [2], were aggregated to build an interdependent graph. The resulting graph G comprises nodes that are either a rail station or a metro station. An important characteristic of the transportation network is that the transitions between different transport modes during a trip are modeled so that CrossLayer edges are added between layers, which enable to design and model trajectories with different transportation modes on one hand, but increases complexity of the network on the other (Asgari et al. 2016).

[1] https://github.com/pbonazzi/21-network-cascade-failure/

[2] http://www.OpenStreetMap.org/

In these networks, the nodes represent the stops and the edges the way in which the different stops are connected. If a stop can directly be reached by another one, these two are connected and It was stated as a crosslayer in the edgelist dataset. In image 3 it can be seen how the metro and train networks have been mapped: it is clear that the two networks are overlapping and therefore have a relationship with each other.

To sum up, the metro sublayer of the network is composed of 303 nodes and 356 edges and the train sublayer 241 nodes and 244 edges. And the crosslayer has 64 edges between train and metro. The vertices have a latitude and longitude property called Lat, Lon formatted to the standard coordinate system WSG 84 and a layer properties which can either be metro or train. The edges properties are name, direction and layer. The name property takes the metro or train line name, direction is either “TwoWay” or “OneWay” and layer is a value among “train”, “metro” and “crosslayer” (Asgari et al. 2016).

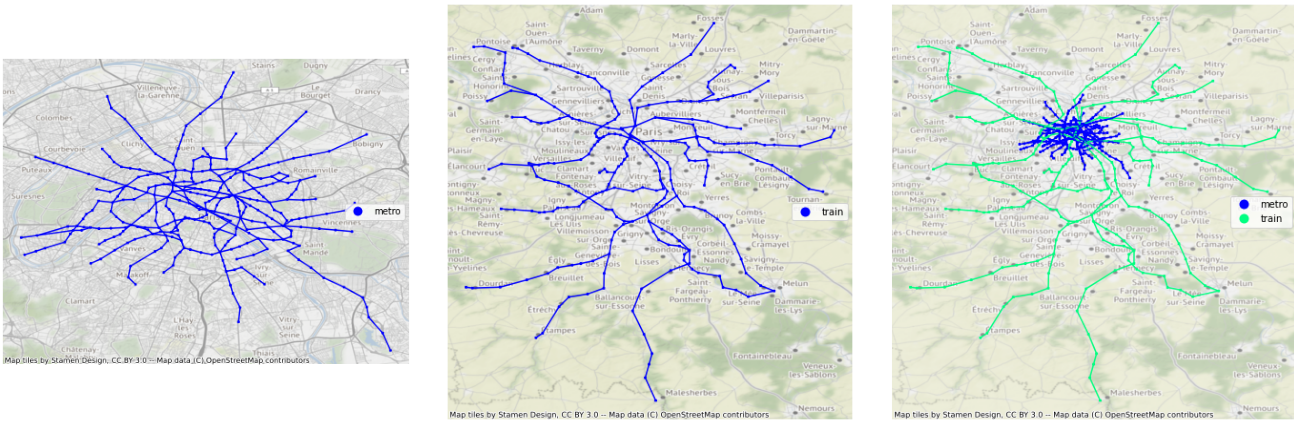

Fig. 3. Paris metro and train networks and their interconnection

4 METHODS

To generate the network data, execute the cascade failure scenario and analyze the mutually connected giant component, this project was led on Python programming environment. In the following section, the implementations are covered in detail.

4.1 Apparatus

In order to generate the random network and the real-world network, the Python library’s NetworkX[3] was found suitable for this project. It makes easy to create the Erdős–Rényi model based on the given nodes and average degree. Furthermore, using the csv file of edge and vertices, Paris Train-Metro interdependent network was formed as networkx.graph datatype. However, the scale-free network based on declared powerlaw distribution is not supported from NetworkX built-in functions. So, for this case, the Python powerlaw package was used to generate a degree sequence following powerlaw method. And, we used it to set up the scale-free network.

All the specific requirements for the Python libraries can be found in out GitHub repository under requirements.txt.

4.2 Experiments

This project cover three different cases of interdependent network, Erdos-Renyi model base Random Network, Scale-Free model base Random Network and Paris Transportation Real world Network. All these cases was tested by following the experiment process above.

Since one of the goals of this paper is to reproduce the results of Buldyrev et al. 2010, the exact procedure presented in their paper will be followed. Namely, our implementation of cascade edges failure will follow the procedure in Buldyrev et al. 2010 where : ”B-links that link sets of B-nodes connected to separate a1- clusters are eliminated” .

[3] https://networkx.org/

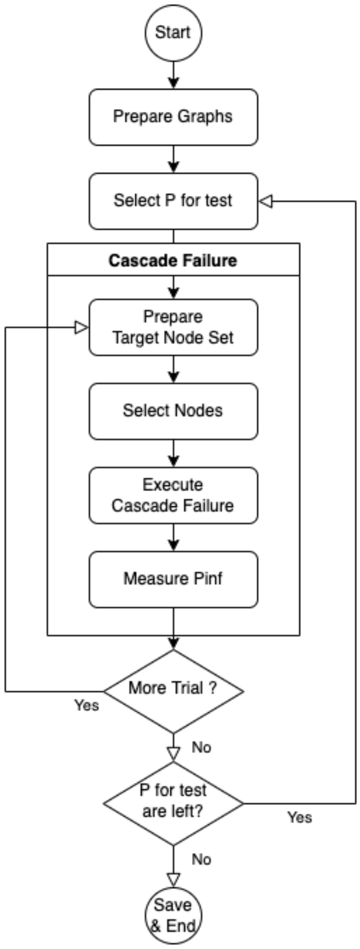

Fig. 4. Process of testing of cascading failure scenario

After that, the Cascade of Failure model explained in the theory section is applied to the networks. As a result, the same visualisation of the ones presented in the reference paper are presented (see section 5).

Additionally, we used the remote computing resource from UZH IT team which has 120 CPUs and 500GB RAM. It improved the computing power for this project but still took a lot of time because the connection to the server was not secured safely. So, most of project time was intensively used to simulate the cascade failure scenario for each cases. Finally, to get all data for this experiments, It took a week, 6 hours a day.

As a final note, we also tried a different approach on cascade failure to see if what was said in the paper was actually implemented differently . This second implementation , maintained all B-links that linked sets of B-nodes connected to at least a common a1- cluster . Even this more logic implementation of cascade of failure did not help us reproduce the results in Buldyrev et al. 2010.

5 RESULTS AND DISCUSSION

5.1 Erdős–Rényi Networks

We first generated an interdependent ER network with node size 500, 1000 and 2000. Then, we measured p_{inf} (the probability of existence of the giant mutually connected component) as a function of p.

Based on the result from Buldyrev et al. 2010 for coupled ER models with a finite number of nodes N, the giant mutually connected component exists at p near a theoretical p_c. As N goes infinite, it is expected that p_{inf} converges to a step function at $p = \frac{2.4554}{<k>}$, which changes value to zero.

The drastic drop near the p_c as in the theory was not shown in our result. Nevertheless, similar to the paper, we observe at theoretical p_c that the steepness of slope is flipping in the exact way as in the paper. In the case of the network with N = 2000, our highest node size, if we consider the region where steepest slope appears as p_c, the steepest slope was observed in the range of 2.4 to 2.6, which includes the theoretical p_c.

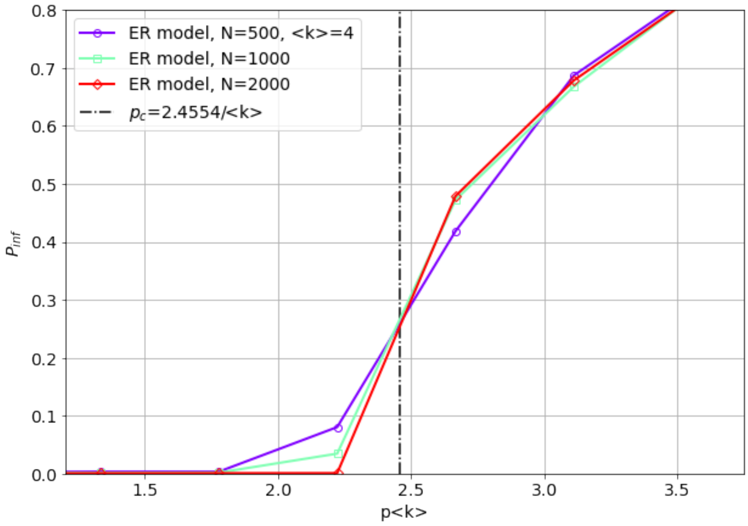

Fig. 5. Results for ER-Networks for N=500, 1'000, 2'000 and $\langle k \rangle = 4$

5.2 Scale Free Networks

When it comes to the broadness of distribution, a broader distribution means there are more nodes with low degree to compensate for high degree nodes, especially for the network with a fixed average degree. Since the low degree nodes are more likely to be disconnected, having a broad distribution can reduce the chance of destroying the giant component which is composed of high degree in single networks. However, this property can not be applied to interdependent networks.

As it can be noticed from 6, we found out that coupled SF networks with a broad distribution are more vulnerable to random attack compared to networks with a narrow distribution. This confirms the results previously presented by Buldyrev et al. 2010. The reason behind it is that in the case of an interdependent network, high degree nodes from one network might be connected to low degree nodes from the other network. Therefore, the giant components that play an important role in the robustness of single networks become vulnerable when random attack occurs in interdependent networks.

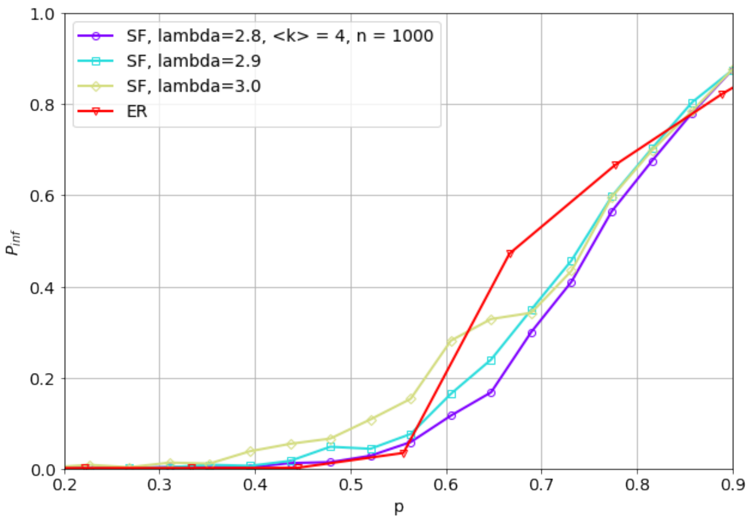

Fig. 6. Results for ER network and SF-Networks with $\lambda = 2.8, 2.9, 3.0$. Both networks are fixed by N = 1000 and $\langle k \rangle = 4$

(Newman 2005) helped for the plotting of the Scale-Free Networks

5.3 Real world Networks

Finally, we analysed the real world dataset, namely the Paris metro network (n=304, $\langle k \rangle$=2.35), Paris train network (n=241, $\langle k \rangle$=2.02) and Paris metro-train interdependent network (n=544, $\langle k \rangle$=2.44).

In order to see the difference between single network and interdependent network, we compared the Paris single metro network to the interdependent one. As we expected, we found that a network becomes more vulnerable when it is interdependent to other networks. In figure 7 is clearly visible that there is a region where p_{inf} of the interdependent network is lower than the one of the single network at the same p. This confirms the findings of the reference paper, which argued that independent networks are more robust than interdependent ones.

There are two possible ways to generate a cascade of failures: we can either start attacking the train network or the metro network. For this reason, it was interesting to visualise both cases to see the differences (see 8). We found that attacking the metro network first triggers more drastic cascade failure and this is due to the properties of the given real world dataset. Our dataset described that one train node can be interdependent with several nodes of the metro network. This means that if we start attacking from the metro network there are more chances of the two networks being disconnected and that failure happens earlier, even if a smaller number of nodes is attacked. Moreover, the average degree of the metro network is higher than the train network, which can also make failure faster.

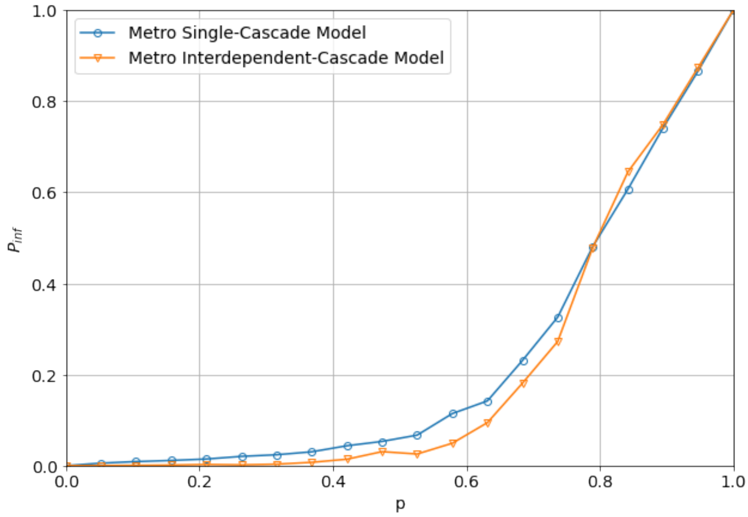

Fig. 7. Result for Paris metro single network and interdependent network

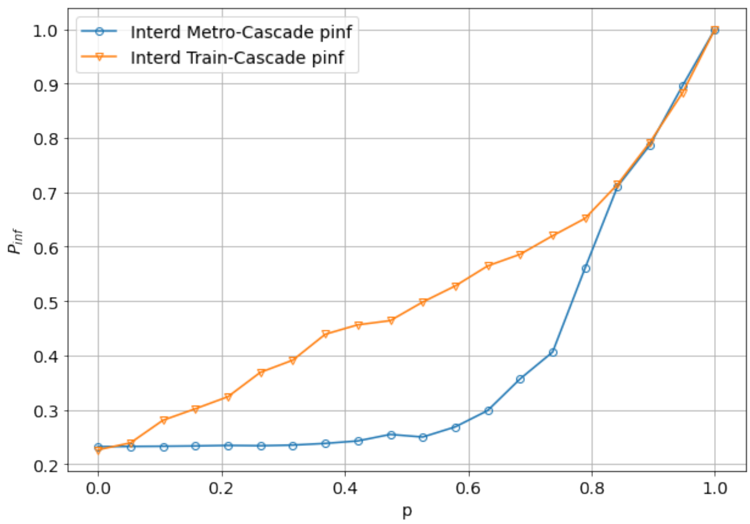

Fig. 8. Results for Paris metro-train interdependent network with different random attack

6 CONCLUSION

The aim of this project was to reproduce the study by Buldyrev et al. 2010 for understanding the robustness of interdependent random and real-world networks. In order to do so, we analysed two types of random networks, namely Erdős–Rényi and Scale-free Networks and a Real world network derived from a Paris metro and train dataset. After applying the Cascade of failures model on each of them, we plotted the probability that the giant component still exists for an infinitesimal number of failures. Although some time and resource constraints did not allow us to fully replicate Buldyrev et al. 2010 study in terms of parameters such as the number of nodes, the results are still very similar.

From ER models with different node size, we observed some phase change around theoretical p_c. For Scale-free networks, the results of the reference paper were confirmed, showing that coupled scale-free networks are more vulnerable to random attack compared to networks with a narrow distribution.

Paris transportation Network metro and train interdependent networks showed that interdependent networks are more prone to cascade of failures in contrast to single independent ones. This demonstrates that the application of network theories based on single networks are no longer sufficient to analyse modern complex networks, but that instead the use of interdependent networks would be more appropriate.

7 CONTRIBUTIONS

Bonazzi implemented and documented all the stages in the cascade of failure ($src/attack.py$). He has written several scripts ($scripts/$) and run them on laptop and on the server cluster to create the results. He organized the structure of the repository, wrote the instructions to reproduce results ($README.md$ and $01reproduce.md$) and defined the library requirements. He has written three notebooks ($notebooks/$) used for the experiments (i.e. archive/ example,test1 and maps). After that, he created three figures of the Paris networks on the topological map. Jointly with other authors he wrote $report.ipynb$. He wrote some paragraphs in the report and modified the slides on cascade of failure . Out of curiosity, he invented an additional method to eliminate edges for Stage 2 and 3 of the cascade failure process, which was not used on the final results . Finally, he participated in the discussion about the final results and documentation.

Han was mainly in charge of measuring ($src/measures.py$) and analysising ER/SR/Real world networks. She wrote some scripts ($scripts/$) and run them on laptop and on the cluster to create the results. She has contributed to four notebooks (archive/ ... test1, test2, test3 and $report.ipynb$). She modified the slides on Results Discussions and wrote the Discussion session in report.

Hyeongkyun was mainly in charge of the generation ($src/create.py$) and analysis of dataset for ER, SF and Real world networks. He has written three notebooks (archive/ ... gentestnet, test1, test2) and jointly worked on $report.ipynb$. After that, he contributed to run the test scenario of SF and Real world dataset ($scripts/$) on laptop and cluster. He modified the slides on methods and wrote some sections of the report. He edited the video recording. Finally, he participated in the discussion about the final result and documentation.

Pura took care of writing the report, especially the chapters on theory and methods, the abstract, introduction, and conclusion. Furthermore, she created the presentation.

REFERENCES

Amini, M. Hadi, Ahmed Imteaj, and Panos M. Pardalos (2020) "Interdependent Networks: A Data Science Perspective". In: *Patterns* 1.1. doi: `10.1016/j.patter.2020.100003`, pp. 1–9.

Asgari, Fereshteh, Alexis Sultan, Haoyi Xiong, Vincent Gauthier, and Mounîm A. El-Yacoubi (2016) "CT-Mapper: Mapping sparse multimodal cellular trajectories using a multilayer transportation network". In: *Computer Communications* 95. doi: `10.1016/j.comcom.2016.04.014`, pp. 69–81.

Buldyrev, Sergey V Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin (2010) "Catastrophic cascade of failures in interdependent networks". In: *Nature* 464.7291. doi: `10.1038/nature08932`, pp. 1025–1028.

Newman, MEJ (2005) "Power laws, Pareto distributions and Zipf's law". In: *Contemporary Physics* 46.5. doi: `10.1080/00107510500052444`, pp. 323–351.