Dissertation Type: simulation modelling and data science

University of BRISTOL

DEPARTMENT OF COMPUTER SCIENCE

Cascading Failures in interdependent network

Junoh Park

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Bachelor of Science in the Faculty of Engineering.

Friday 21st May, 2021

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of BSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Junoh Park, Friday 21st May, 2021

Acknowledgements

I would like to thank my supervisor, Professor. Seth Bullock for his support and guidance throughout this project. The feedback and comments make me progress. Also, I would like to thank my parents. This project could not have been completed without the support and encourage that they have given.

Abstract

Nowadays, the world cannot be explained with independent single networks. There is more and more interaction between various networks as time goes by. Beyond the interaction, there is a dependency between the networks. We call the network depends on another network if the network cannot operate normally without it. The dependency can be bidirectional and the networks depend on each other form a huge networks which is the interdependent network. It can be easily understood as a network of networks. The degree of interdependence between the networks is getting higher. So, if the attack or failure occurs in the interdependent network, it could pose a serious risk. In 2003, Italy, the failure of one power line causes failures of trains, underground, and airplanes nationally. We call this sequential failure as cascading failure.

Buldyrev et al (2010) introduced cascading failure in the interdependent network based on the Erdős–Rényi (ER) Graph. Following the model, there are many other models based on the random interdependent network. However, there are not many cases that focus on the spatial-based interdependent network, while the networks in real world are based on the space. Therefore, our thesis considers both random-based and spatial-based interdependent network. Then, the comparison of the models is mainly focused on this thesis. The spatial based network uses Random Geometric Graph (RGG) as a component. First, the replication of the Buldyrev's model is built to reinspect the paper and get the verification of our model. Then, the model is upgraded with spatial feature. After applying the feature, there are cases: (RGG / ER) graphs are (spatially / randomly) interdependent. The latter is the supporting system. Hence there is total 4 unique models.

Our model reaches the analysis that the RGG requires more edges to build giant cluster than ER graph because there is a distance limitation. However, in the supporting system, spatial based system shows strength. The distance limitation make the failure not to spread out.

Contents

Chapter 1

Introduction

Networks in the real world cannot be isolated from other networks nowadays [1, 2, 3, 4, 5]. Not only one network depends on another network, but also many networks depend on each other and have connections with them. Therefore, they are composing various interdependent network together.

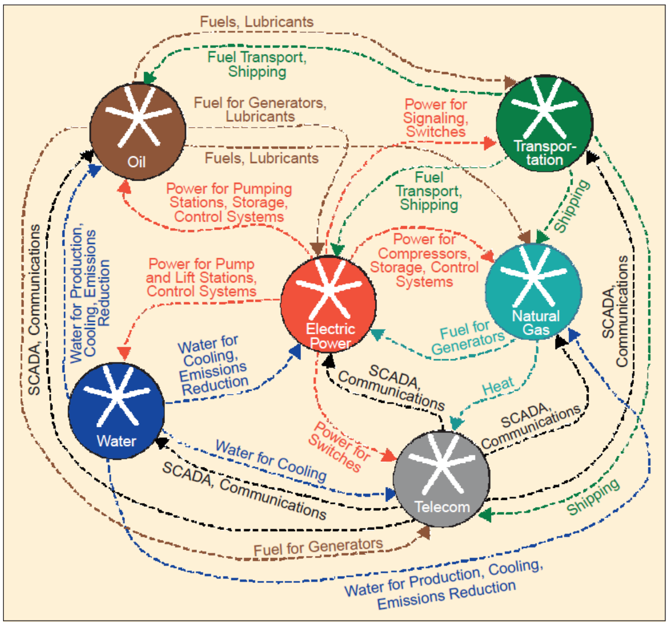


Figure 1.1: The interdependent network system in infrastructure [4]

It could be shown well in Figure 1.1. The figure represents the interdependent network system of infrastructure in our world. In the figure, the water network supplies water to the electricity network, the oil network and the telecommunication network for cooling and production by the transport network [4]. Simultaneously, the electricity network supplies electricity to water network for power, and the telecommunication network provides communication service to both water and electricity networks.

As above, the degree of interdependence of today's networks is very high. Therefore, the failure occurred in one network can be a serious threaten to others, even though the scale of the initial failure is not much big. We can see it in the 2003 Italy Blackout [6] and 2012 India Blackout [7]. In 2003 Italy Blackout, the failure of the power line due to storm causes the cancellation of trains, undergrounds, and airplanes [6]. In 2012 India Blackout, the traffic signals were not working, trains and airplanes are stopped, and even hospitals and water plants are suspended few hours because of the circuit breakers on one electricity support line [7, 8]. These are the cascading failures from the initial failure. The initial failure causes more failures sequentially, and the failures causes much more serious failures across inter-network systems.

Therefore, understanding the interdependent networks and the cascading failure on it is the key point to interpret the structure of our society system and infrastructure [1, 9]. Of these networks, many of them are spatially embedded. The nodes in the networks are located in space, and the connections also are likely to follow the distance between nodes. The closeness is important factor in connecting edges. Transportation network and power grid are the obvious networks that rely on the distance. Also, the telecommunication network is based on the spatial feature, since the devices have a restriction on radio range [10]. Even the social network is related with the spatial feature too. Many people tends to minimise the effort to maintain the social ties, and it lead them to connect with their spatial neighbours mainly [11]. Moreover, many online communities show that the members tend to make connection with the neighbours [12].

By contrast, lots of models of network feature non-spatial graphs [1, 13, 14, 15, 16]. It raises the question of whether there will be a difference from cascading failure in random network if spatial elements are included in the network. Hence, the interdependent network with both non-spatial and spatial features are going to be modelled, and the simulations of cascading failures are going to be explored in this thesis. For non-spatial graph, replicating the previous prominent paper (Buldyrev et al 2010) is going to be presented. Since the Erdős–Rényi Graph is used in the paper, all of non-spatial graphs in this thesis is also going to use Erdős–Rényi Graph. And for spatial graph, the Random Geometric Graph is going to be used, since it is one of the simple and easy graph to express the spatial model.

In chapter 2, the background information of the thesis such as graph theory, technical background and the previous paper is going to be introduced first. Then, in chapter 3, the first focus of this thesis follows. The replicating of the paper (Buldyrev et al 2010) is the content of this chapter. In the replicating, the re-inspection of building an interdependent network model and the cascading failure introduced on it. In chapter 4, the second focus of the thesis is going to be proposed. It is adding a spatial feature into the model and simulating the cascading failure for comparison of non-spatial and spatial random networks. The result and analysis of each part are included in each chapter. Therefore, chapter 3 and chapter 4 have their own methods and analysis sections. Lastly in chapter 5, the conclusion will follow.

Chapter 2

Background Information

2.1 Graphs and Networks

2.1.1 Erdős–Rényi Graph

The Erdős–Rényi (ER) Graph is one of the random graph model. It is named after the Hungarian mathematicians, Paul Erdős and Alfréd Rényi [17]. The graph model consists of nodes and edges, and it is based on the probabilistic method. The essential factors for constructing the graph are the number of nodes (N) and the probability (p) used on edge creation. Therefore, the expression of the Erdős–Rényi Graph is $G(N, p)$ [18].

The specific explanation of the p is that every edge between every node pair has probability p to be added into the graph independently from every other edges. Therefore, if $p = 0$, there is no edge on the graph and if $p = 1$, the graph is fully connected. With this feature the ER graph could be a stochastic model. Then, the probability that the graph, $G(N, p)$, has M edges is:

$$P(M) = p^M (1-p)^{\binom{N}{2}-M} \tag{2.1}$$

Then, from the equation (2.1), the expected number of edges can be calculated sequentially. By counting the number of graphs that contain M edges, the expected number of edges is calculated as:

$$Exp(M) = \binom{\binom{N}{2}}{M} p^M (1-p)^{\binom{N}{2}-M} \tag{2.2}$$

The probability distribution of M, $P(M)$, is a binomial distribution. Therefore, with the equation of calculating the mean value in the binomial distribution, the expected number of edges is [19]:

$$Exp(M) = \sum_{m=0}^{\binom{N}{2}} mP(m) = \binom{N}{2} p \tag{2.3}$$

From the equation (2.3), the mean degree of the graph could be deduced. The mean degree, $Exp(k)$ is the mean number of edges that one node in the graph has. Therefore, it is a value that $Exp(M)$ is divided by $\frac{n}{2}$. The reason why the value is divided by $\frac{n}{2}$ is that the graph is undirected. Hence, it is:

$$Exp(k) = \sum_{m=0}^{\binom{N}{k}} \frac{2m}{N} P(m) = \frac{2}{N}\binom{N}{2} p = (N-1)p \tag{2.4}$$

Next, the degree distribution of the ER graph could be derived from the following equations. Simply, if a node is connected with k other nodes, then the node is not connected with $N-k-1$ nodes. Therefore, the probability, p_k, that any node is connected with other k nodes is calculated as follows:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-k-1} \tag{2.5}$$

The result indicates that the degree distribution of ER graph, $G(N, p)$, is a binomial distribution. With large N, the equation could be expressed with an approximate equations [20]. The first part of the equation 2.5 could be approximated as:

$$\binom{N-1}{k} = \frac{(N-1)!}{(N-k-1)!k!} \simeq \frac{(N-1)^k}{k!} \tag{2.6}$$

And the second part of the equation 2.5 could construct the new equation:

$$ln[(1-p)^{N-k-1}] = (N-k-1)ln(1-\frac{Exp(k)}{N-1}) \simeq -(N-k-1)\frac{Exp(k)}{N-1} \simeq Exp(k) \tag{2.7}$$

with the equation 2.4 and the Taylor expansion in logarithm [20]. Therefore, $(1-p)^{N-k-1} \simeq e^{-Exp(k)}$. By substituting equation 2.6 and equation 2.7 into the equation 2.5, the p_k becomes:

$$p_k \simeq \frac{(N-1)^k}{k!} p^k e^{-Exp(k)} = \frac{(N-1)^k}{k!}(\frac{Exp(k)}{N-1})e^{-Exp(k)} \simeq e^{-Exp(k)}\frac{Exp(k)^k}{k!} \tag{2.8}$$

Hence, with the large N, the degree distribution of the ER graph is a Poisson distribution [17]. And also, since the probability p is $\frac{Exp(k)}{N-1}$, the clustering coefficient of the graph is also $\frac{Exp(k)}{N-1}$. As N grows infinitely, $p \to 0$. Thus, the ER graph has very low clustering coefficient [21].

2.1.2 Random Geometric Graph

The Random Geometric Graph (RGG) model is a random network model which havs a spatial feature. It is the base of the spatial graphs that are used in this thesis. The Random Geometric Graph consists of nodes and undirected edges. The graph can be denoted as $G = (V, E)$ where V represents the set of nodes and E represents the set of edges [22]. The factors required to build the Random Geometric Graph are N, the number of nodes, and r, the threshold. The threshold r is used on connecting edges.

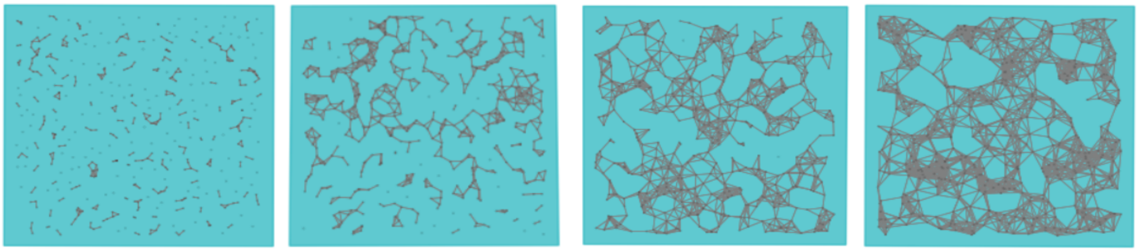
Since the RGG model is the sptial graph, the nodes have their own location in the graph. So, distributing the nodes is the first task on building the RGG model. The node distribution is the independent uniform distribution. Therefore, the nodes are scattered randomly in the graph space [23].

```
G = (V, E)

for i ∈ V do
    for j ∈ V \ {i} do
        if distance(i, j) < r then
            E ← edge(i, j)

```

Algorithm 2.1: Connecting Edges in RGG

The Algorithm 2.1 shows the pseudo code of connecting edges. For all node pairs except the node and itself, if the distance between two nodes is smaller than the threshold r, the edge is going to be connected between two nodes.

Figure 2.1 shows the random geometric graph with same number of nodes, 500, and different threshold values. As the threshold grows, the graph contains more edges than before. Also, with the lower



(a) threshold: 0.03 (b) threshold: 0.05 (c) threshold: 0.07 (d) threshold: 0.09

Figure 2.1: RGG graphs (N=500) in 2-dimension with different threshold

thresholds, there are more than one components in the graph. However, with higher thresholds, the components are getting connected each other and they make up large components. So, the number of components is getting smaller. Hence, when the threshold becomes 1, the graph is fully connected.

Since edge connection in RGG is based on the distance, the edges are concentrated in the community of node neighbours. So RGG has relatively higher clustering coefficient rather than ER graph [23]. And the feature that the nodes are distributed in the space in RGG makes that ther could not be an exact theoretical expression of largest component in the RGG model. However, with practicies, it is revealed that RGG requires more edges to build giant component as ER graph. [23, 24]

2.1.3 Interdependent Network and Cascading Failure

Interdependent network is the collection of the one-layer networks. Each network consists each layer, and the layers are connected each other. There are different kinds of interdependent networks. In this paper, the 'Multilayer Interdependent Network' is going to be used. The Multilayer network is the interdependent network with the property that there is not a same node in each layer [25, 26]. Therefore, every node in the network is unique. Figure 2.2 is the example of the multilayer interdependent network. The unique nodes in layer A, B, and C build unique graphs in each layer, and the graphs build one big interdependent network with the connections of each other.

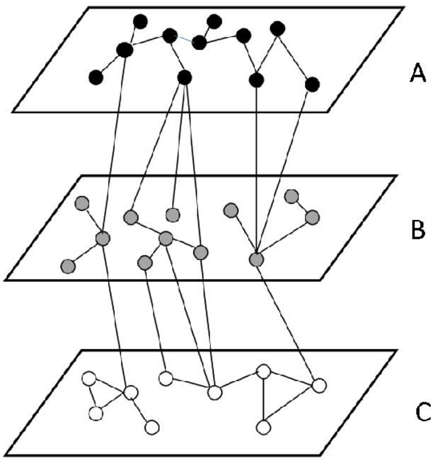


Figure 2.2: Multilayer Interdependent Network Model [27]

The edges between each layer are considered as the edges between nodes and their supporting nodes [28]. The supporting node system is suggested in the interdependent network model. In the system, the node depends on its supporting node in different layer that decides the survive of the node. And since the edges are undirected, they support each other [29]. This demonstrates the *interdependence* of the entire network. The example of electricity network and water network in the chapter 1 is one case of this system. The interdependent networks with random graph without spatial feature have constructed and simulated in various papers [1, 13, 14, 15].

The cascading failure is based on the percolation theory. The percolation theory in the network science is observing the outcome after removing nodes or edges in the network [30]. Especially, the percolation theory in interdependent network is introduced by Buldyrev et al (2010) [1], which is the paper that is going to be replicated in this thesis. In the paper, when the initial attack kills nodes in the network, all edges that are connected to the dead node are removed. Then, the failures of nodes follow sequentially as a chain reaction. This is the cascading failure. Detailed conditions for determining the death of a node is going to be introduced in the next part. It is visualised well in the appendix A.

2.2 Pymnet Library

To implement the simulation with the interdependent network, the 'pymnet' library is used [25]. This is the specialise library for multilayer interdependent network. With this library, it could be possible to realise the interdependent network model and visualise it. Figure 2.3 is the example interdependent network model which is built with 'pymnet' library. The interdependent network consists of 2 layers, 10 nodes in each layer and the edges are undirected. Each node in the layer has 1 unique supporting node in another layer. Since the network has undirected edge, nodes support each other.

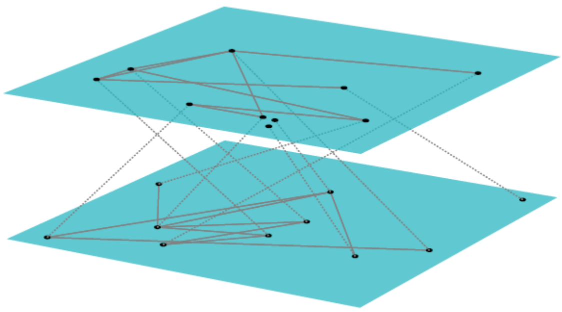


Figure 2.3: 2 layers, 10 nodes in each layer, undirected edges, 1 unique supporting nodes in another layer

There are many methods and functions in the library which are used in this thesis to build the network and simulate the failure. First, the frame of the interdependent network is built with 'MultilayerNetwork()' funtion. Then, with 'add_layer()' and 'add_node()' methods, the frame can be filled with the components of the network. Next, the edge between node i and j can be connected with the method, $network\_name[i, j, layer\ of\ i, layer\ of\ j] = 1$. By changing the value from 1 to 0, the edge can be removed. Lastly, the visualisation of network is completed with 'draw()' function. The detailed options could be set by adjusting the parameters. For example, setting directed graph, changing aspects of the graph, and coloring the graph are possible with setting the parameters.

However, there are limitations on 'pymnet' library too. It is only specialised in constructing the interdependent network and visualising. Therefore, 'networkx' package must be suggested as a supplement to use other methods [31]. For example, the ER graph is implemented through 'networkx' module first and moved into the interdependent network frame built with the 'pymnet' library. Since 'pymnet' does not provide any applications of various networks, additional steps like 'networkx' package are needed for constructing various networks. Also, the 'networkx' package is used to analyse the graphs in each layer before and after cascading failure respectively. The graphs are rebuilded in the package and analysed with various methods such as 'average_clustering()' methods for calculating the clustering coefficient.

2.3 Statistical Tests

2.3.1 Kolmogorov–Smirnov Test

Kolmogorov–Smirnov test (K-S test or KS test) is a non-parametric test that could be used both continuous and discrete datasets. There are two types of tests which are: 1) one-sample K-S test and 2) two-sample K-S test. In one-sample K-S test, the test checks whether the sample is drawn from the reference distribution or not. The reference distributions can be every continuous distributions such as normal distribution, exponential distribution and so on. The two-sample K-S test checks whether the two sample is drawn from the same distribution or not [32]. However, two-sample K-S test has a serious weakness. It is that the test is designed to react sensitively on every possible types of differences between samples, so it is not very powerful. Hence, one-sample K-S test is going to be used in this thesis.

The one-sample K-S test is based on the cumulative distribution function. And it uses difference between sample and the reference model in terms of the cumulative distribution function. And the 'Null-hypothesis' is the main method to uncover the relationship between the sample and the model. In this test, null hypothesis is H_0*: The sample is derived from the reference model.*

Then, the sample and reference model is converted into the cumulative distribution function type. The $F_n(x)$ represents the reference based model and $F_n(x)$ represents the sample based model. Then, calculating the maximum difference between two models follow [33]:

$$D_n = \max x|F_n(x) - F_n(x)| \tag{2.9}$$

The equation 2.9 shows how to calculate the difference that is used in this test. From the difference, the p-value for 'null-hypothesis' is derived. If the p-value is larger than 0.05, it is enough that the sample is regarded as coming from the reference model. If p-value is smaller than 0.05, it means that there is no association between the sample and the model.

2.3.2 Mann-Whitney U Test

Mann-Whitney U test is a non-parametric test which is based on the null hypothesis. The condition that make the test available is that the samples are ordinary or the samples are not normally distributed [34]. Therefore, the Mann-Whitney U test is usually used with Komogorov-Smirnov test together. Mann-Whitney U test checks whether the two samples are derived from same distribution or not. Therefore, the samples must be independent each other. In this thesis, the distribution of samples are verified through K-S test and if they are not normally distributed, Mann-Whitney U test is going to follow for further verification.

The null hypothesis in this test is that H_0*: The distributions of samples are equal.* To verify this null hypothesis, the U statistic is used in the Mann-Whitney U test. However, before introducing the U statistics, the rankings of observations from each sample must be calculated. In the process of ranking the observations, if there are tied values, their ranks are set as a midpoint of the unadjusted ranking [35]. For example, when the observations are (3, 4, 7, 7, 9), the ranks of each observations become (1, 2, 3.5, 3.5, 5). After this process, the U statistic follows with the equation 2.10:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}, \qquad U_2 = R_2 - \frac{n_2(n_2+1)}{2} \tag{2.10}$$

where R_1 and R_2 are sum of each rank, and n_1 and n_2 are the size of each sample [36]. Then, he standard normalisation is applied on these U_1 and U_2 values to get p-values. If the p-value is larger than 0.05, the null hypothesis is adopted and the samples are regarded as came from same distribution. However, if p-value is smaller than 0.05, it shows that the samples are not came from same distribution.

2.4 Previous Related Work

In this section, the paper 'Catastrophic Cascade of Failures in Interdependent Networks' (Buldyrev et al, 2010) is going to be introduced. This is the prominent paper that is going to be replicated in this thesis. The paper introduced cascading failures in interdependent network model with ER graph. The background in section 2.1.3 is also based on this paper. Since modelling the interdependent network is presented enough in previous section, the cascading failure and result from theoretical base are handled in this part.

As introduced, the cascading failure is the sequential outcome of the initial attack. After some nodes are dead and edges are removed in initial attack, there will be new nodes that satisfy the death conditions. There are 2 death conditions. First one is when the node is not in the giant mutually connected component. A component in the network theory is the induced subgraph with any two nodes connected by path. And the giant mutually connected component is the mutually connected component covering a significant proportion of the entire network [1, 37]. In this thesis, the node satisfies this condition is going to be called 'isolated'. The second condition is when the supporting node of its node dies. The model is built as every node depends on its own supporting node. Therefore, if support is lost due to the death of its supporting node, the node will also die. In this thesis, the node meets this condition is going to be called 'unsupported'. Each step has a newly dead node that meets the above conditions. And if there is no newly dead node, that is the moment when cascading failure ends.

The result that paper reached with these models is: there is a unique threshold that decide the survival of the giant component. It could be proved in theoretical way too and the paper introduced it. Let p is the proportion of nodes which are not the targets of initial attack. Then, by using the characteristic that the degree distribution of the ER graph follows the Poisson distribution (proved in chapter 2.1.1), it could be revealed that there is unique value, $p = p_c$, which is the threshold. The value that is proved in the paper is $p_c = 2.4554/\langle k \rangle$ where $\langle k \rangle$ is the mean degree of the graph. Therefore, the point, where the multiplication of mean degree and proportion of survived node from initial attack is 2.4554, is the threshold that decide the survival of giant component. There is a comparison of theoretical value and graphs from model simulations in the paper. It will be analysed together with our model in the analysis part of next chapter.

Chapter 3

ER Interdependent Network Model

Our focus in this chapter is building the Erdős–Rényi Random interdependent network, simulating the cascading failure on the network, and comparing the result with the Bulyrev's model [1]. It is undertaken to build the frame of the interdependent network and the cascading failure system, and check the validation of our model by comparing it with the data from the Buldyrev's model. Therefore, to compare them exactly, the condition of our model must be maintained similar with the Bulyrev's model. It is achieved by setting the same environment for the interdependent network and following the same cascading failure logic with the Buldyrev's model.

3.1 Methods

Before introducing the details of the methods, the brief explanation of the entire logic is going to be proposed in this paragraph. The logic of making interdependent network model is composed with three parts. First part is designing the environment for interdependent network. In this part, diverse default parameters for the following networks are set, and the layers in the interdependent network are constructed and named. Next part is building the intra-layer details which is constructing independent graph in each layer. And lastly, completing the inter-layer details, which constructs the interdependent connections between the nodes in different layers, is carried on. The second part can be seen as the local side of the interdependent network and the third part can be seen as the global side. After this steps, there will be a multilayer interdependent network such as figure 2.2.

Then, the cascading failure proceeds. The first part is the initial attack. The selected nodes as first targets are going to be killed in this part. The edges related with the dead nodes are removed as the nodes are dying. Then, the cascading failure starts. As introduced in chapter 2.4, there are two conditions which decide whether the nodes will be died or not. They are 'isolated' and 'unsupported'. If the nodes satisfy one of the conditions, then the node will be killed and all edge that connected with the node is removed. In each step, there is new targets after the nodes are died and edges are cut. The next step will be carried on with the new targets, and the cascading failure will be stopped if there is not any new target.

3.1.1 Building Interdependent Network

A. Design the Environment

In the designing the environment part, the layers and the default parameters which are needed for manufacturing the network should be set. The number of nodes in each graph (N), the number of layers (L), and the probability (t) are the set parameters for the environment of the model. In the Buldyrev's model, there are two layers. Thus, even in our model, the value of L is fixed at 2. Also, since each graph in each layer in Buldyrev's model has same features with others, all parameters such as N and t applies to all layers. With the variable L, each layer in the interdependent network is constructed by 'add_layer' method in the 'pymnet' library. Their names are based on the English characters. Therefore, the first layer's name is 'a', the second one is 'b' and so on. The variables N and t are used in building the Erdős–Rényi graphs in each layer.

B. Build Intra-Layer Graph

In this part, the details of designing the intra-layer graphs are going to be presented. The steps for building each graph are: making nodes in each layer first and connecting the edges between nodes next. The nodes are named as numbers. And since the multilayer network system is used, there are unique nodes in each layer. So, there are total $N \times L$ nodes, and their names are from 0 to $(N \times L - 1)$. Since the Erdős–Rényi graph is not based on the space, there is not a step for distributing nodes. Then, the nodes are added in the frame of the interdependent network that is built in the previous part.

Next, edge connection step follows. First, with the variable 'N' and 't', the Erdős–Rényi graph must be built by using the 'erdos_renyi_graph' function in the 'networkx' module in python. Since the 'pymnet' library does not have a function of building various types of graphs, the edges that make up the graph from the 'networkx' module are added to the interdependent network frame manually as presented. In the Buldyrev's paper, the models are constructed with graphs having the exact number of edges. Therefore, it is necessary in our model to ensure that the graphs in all layers have the same number of edges. As explained in chapter 2.1.1, the mean degree of the ER graph is based on the binomial distribution. So, with the parameter N and t, it is impossible to set the exact number of edges. Hence, additional works is needed to adjust the numbers.

G = networkx.erdos_renyi_graph(N, t)

Interdependent Network Frame ← G.edge

expected_number = $\frac{N(N+1)}{2}$p

if *the number of edges* $<$ *expected_number* **then**
    Add new random edges to the Interdependent Network Frame()

if *the number of edges* $>$ *expected_number* **then**
    Delete random edges in the Interdependent Network Frame()

Algorithm 3.1: Connecting Intra-Layer Edge

The algorithm 3.1 shows the pseudo code of building the edge connections. First, the edges in the ER graph created through 'networkx' module are added to the interdependent network frame. Then, compare the number of edges with the expected number. The expected number is based on the mean number of edges of the ER graph (chapter 2.1.1). Then, if the number is smaller than the expected number, new random edges are added until the number is equal to the expected number. Contrariwise, if the number is lager than the expected number, random edges in the interdependent network frame are deleted until the number is equal to the expected number. Since the ER graph is the random graph, adding and deleting the edges randomly do not affect the property of the graph.

C. Build Inter-Layer Part

In this part, connecting the supporting edges between the graphs in each layer are going to be introduced. The first step is finding unique supporting node of each node and the second step is connecting the pairs by edges. Let $a_1, a_2, a_3, ...a_N$ are the nodes in layer 'a', and $b_1, b_2, b_3, ...b_N$ are the nodes in layer 'b'. In the Buldyrev's model, one node is supported by one unique node, and one node supports one unique node. Since the model takes undirected edge system, the support is made in both directions on an edge. Therefore, if there is an edge between a_i and b_i, it means that they are each other's supporting node. In this thesis, it is expressed as (a_i, b_i) , and named as 'supporting pair'. The supporting edges also should be connected based on the randomness. Since the intra-graphs are ER graph which is non-spatial random graph, there is not any logic or order on $a_1, a_2, a_3, ...a_N$ and $b_1, b_2, b_3, ...b_N$. Therefore, pairing the nodes as (a_1, b_1), (a_2, b_2), (a_3, b_3), ... does not affect on the randomness of the graph. The name of nodes are set as a number at the beginning. Therefore, supporting pairs are mated as

$(0, N), (1, N+1), (2, N+2), ..., (N-1, 2N-1)$.

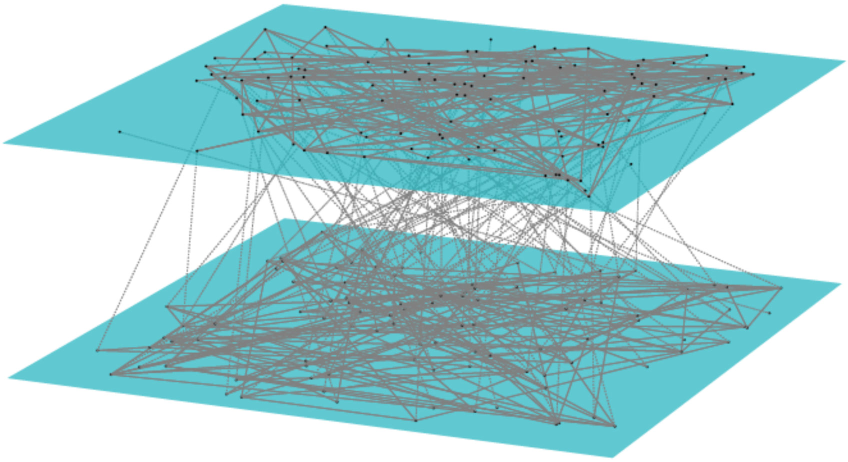


Figure 3.1: ER Interdependent Network Model

After the supporting pair is decided, the edge connection starts. Through the pyment library, edges can be connected between nodes present in different layers. So, based on the supporting pairs, supporting edges are established. This is the last part of the building the ER interdependent network model. The figure 3.1 shows the image of the network model.

3.1.2 Cascading Failure

In this part, the simulation of cascading failure on the interdependent network is explained. For the failure, the new parameter, attack size (K), should be set. It determines the size of the 'initial attack', which happens ahead of the cascading failure and causes a cascading failure. So, the number of nodes are killed before the cascading failure starts. Therefore, K must be smaller than N. Also, the new dictionary, *state*, is declared to store the status of nodes. Before the attack and failure start, the statuses of all nodes are initialised as 'Alive'.

```
initialise_status(state)

step = 0

targets = random_select(K)
initial_attack(targets)
killed = targets

step += 1

while len(killed) > 0 do
    killed = cascade_attack()
    if len(killed) > 0 then
        step += 1

```

Algorithm 3.2: Initial attack and Cascading failure of the Interdependent Network

Algorithm 3.2 shows the steps of initial attack and cascading failure. At the beginning, the states are initialised as explained in previous paragraph. Then, the initial attack starts. The first step is selecting the target for initial attack. With the attack size (K), the nodes which are going to die are selected. Since the network system is based on the random graph, selecting the targets is also based on the randomness. And as the model in the Buldyrev's paper, the targets are selected in one of the layer. After selecting the targets, the initial attack begins. In the attack, the statuses of targeted nodes are changed into 'Initial Dead'. Also, every edge that are connected with at least one targeted node is removed.

After the initial attack, the targeted nodes become killed nodes and the step increases. Then, the cascading failure begins. The specific algorithm of the cascading failure follows in next algorithm:

```
killed = []

for node ∈ G.node do
    if state[node] is 'Alive' then
        if node ∉ G.largest_component then
            state[node] = 'Dead from isolation'
            killed ← node
        else if node is not supported then
            state[node] = 'Dead from unsupported'
            killed ← node

for target ∈ killed do
    remove_all_edges(target)

return killed
```

Algorithm 3.3: Specific Algorithm of the Cascading Failure

The algorithm 3.3 contains the details of the cascading failure. First of all, a step of verifying that the node belongs to the giant component is taken for all alive nodes. If the node is isolated, it is killed and the status of node changes to 'Dead from isolation'. After the node is killed, it is added to the *'killed'* list. Everything is same in the next condition, unsupported. It goes through a step to verify that the node is supported. It is decided with the existence of the edge between the node and its supporting node. If there is an edge, if means the supporting is continuing. If there is not an edge, it means the supporting is broken. If the node is not supported, the node dies and the status is changes to 'Dead from unsupported'. Also, the node is added to the *'killed'* list. Lastly, every edge related with the nodes in the *'killed'* list are removed. Then, the *'killed'* list is returned to determine whether the cascading failure is finished or not.

As introduced in the Algorithm 3.2, the cascading failure is repeated until no newly dying node exists. This is because if there no new nodes have died during the step, then no edges are going to be removed and failure can no longer affect the network model. With these algorithms, the interdependent models are attacked. The attacked model is going to be analysed in next section.

3.2 Analysis

This section presents comparison between the Buldyrev's model and our model. The further result and analysis of chapter 2.4 is going to be introduced with our model. Figure 3.2 shows the analysis from the Buldyrev's paper [1]. This figure represents the probability of existence of the giant component after cascading failure. With different number of the nodes, N, it is shown as function of $p\langle k\rangle$. As introduced in chapter 2.4, p is the ratio of nodes that survive the initial attack and $\langle k\rangle$ is the mean degree of the network. Also, there is a unique value, p_c, that can determine the aliveness of the giant component. The value of p_c is about $2.4554/\langle k\rangle$, and it is shown as a upper arrow in the figure.

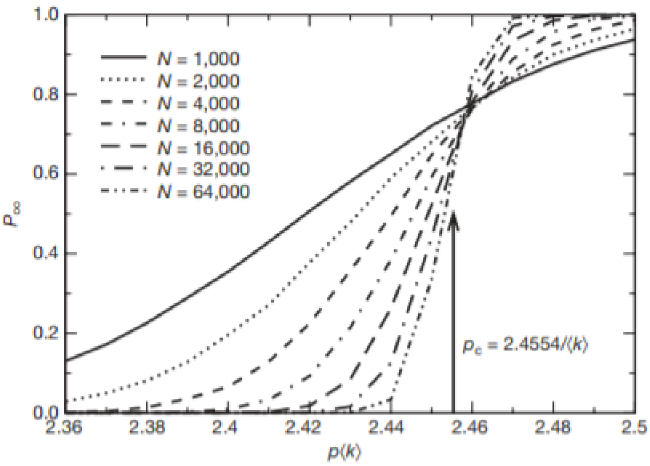
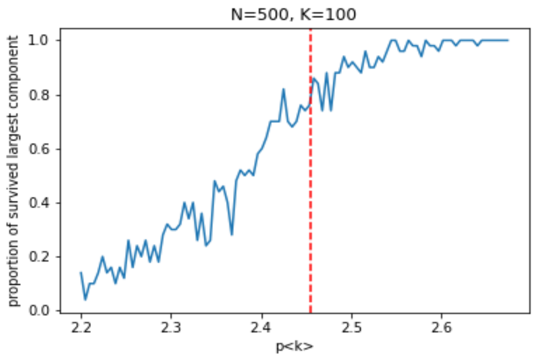
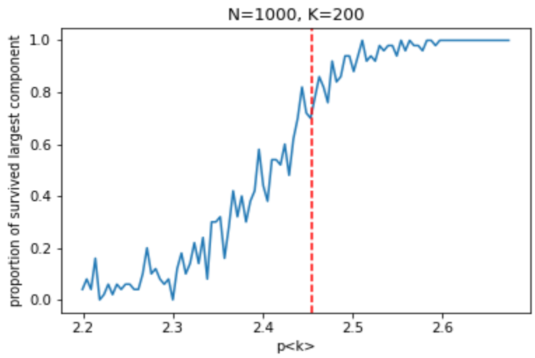


Figure 3.2: Survival ratio of giant components in the Buldyrev's model [1]

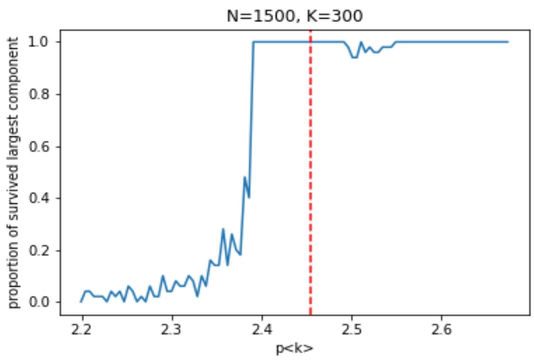
In the figure 3.3, there are same functions that are introduced in the Buldyrev's model, fig 3.2. Each graph is the result from different parameters. Graph 3.3a is from N=500 and K=100. Graph 3.3b is from N=1000 and K=200, and graph 3.3c is the result of N=1500 and K=300. The values of N and K are changed and the value, p is fixed as 0.2. Since the cascading failure part takes huge running time, it is impossible to run simulations with the same number of nodes as the Buldyrev's model. Especially, verifying the isolation condition requires massive running time. Since there is not methods for components of graph in 'pymnet' library, the graph must be copied to the 'networkx' module and analysed. These procedures were always to be carried out at the beginning of the step. Therefore, huge running time is required in verifying the isolation condition.



(a) Nodes=500, Attack Size = 100



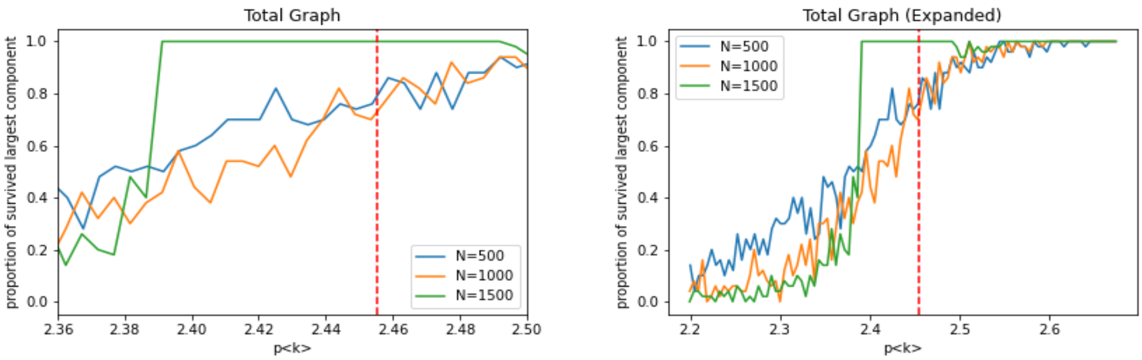
(b) Nodes=1000, Attack Size = 200



(c) Nodes=1500, Attack Size = 300

Figure 3.3: Survival ratio of giant components in our model

In comparison between the Buldyrev's model and our model, there are two points that should be focused. The first point is the place where the ratio shows a sharp increase. In theory, $p\langle k\rangle = 2.4554$ is the value that determine the survive of the giant component. So in the Buldyrev's model, the y-value shows dramatic increases near $p\langle k\rangle = 2.4554$. In our model, there is a sudden growth of the probability near $p\langle k\rangle = 2.4554$ in every model. Hence, our model could be decided as well implemented in this point. However in figure 3.3c, the graph is nonsmooth near $p\langle k\rangle = 2.4$. In terms of the general tendency of the graph, it looks like some outliers affect the graph.



(a) The range of x-axis: [2.35, 2.5]

(b) The range of x-axis: [2.2, 2.65]

Figure 3.4: Survival ratio of giant components in our model (Total)

The second point is that the gradient of the sharp increases varies by the number of nodes. In figure 3.2, the graph with small number of nodes shows relatively gradual growth while the graph with large number of nodes shows relatively sudden growth. Figure 3.4 shows two figures with 3 graphs on one plane. Although figure 3.4a includes the graphs with the same x-axis range as the Bulyrev's model, it is hard to determine the sharp increases. This is because the number of nodes are smaller than the Buldyrev's model, and so the increases are relatively gradual.

Therefore, the range of x-axis is expanded on the figure 3.4b to observe the change better. On the left side of figure 3.4b, the graph with $N = 500$ shows the highest y-value. The graph with $N = 1000$ has the second highest proportion while the graph with $N = 1500$ has the lowest one. However, after the increase, the order is reversed on the right side of the figure. It can be observed in $2.4554 < p\langle k\rangle < 2.55$. However, since the difference of the graphs is smaller than the left side, it is not as clear as left side. It is same in the Buldyrev's model. With these points, the final result could be derived from this chapter. The result is that our model is valid to explain the cascading failure in the interdependent network.

Chapter 4

Spatial and Non-Spatial based Interdependent Network Model

In this chapter, the focus is comparing the ER graph based interdependent network and Random Geometric Graph based interdependent network. These two types of graphs are applied into the intra-layer part of the interdependent network. Similarly, there are also two types of connecting system for inter-layer part. First one is the random system, which is same as the Buldyrev's model. And the second one is the spatial system, which is RGG based system. Therefore, there are 4 unique models that are going to be compared each other. The details of the systems that are applied on each part are going to be described in following sections.

Additionally, the model with spatial features in every part is going to be implemented to be considered as the real-world based spatial model. In the model, one of the importance is that the layers indicates the same region even though they are separated independently. It could be easier to understand with the example. Let the interdependent network would like to be modeled the electricity and water networks in same city. Then, the electricity and water networks are separated into different layers while they are located in same region. Therefore, it could be supposed that all layers are overlapped in real world. Hence, when calculating the distances between the nodes in different layers, it could be assumed that nodes are on the same space. Also, the attacks with spatial features may affect layers together. All of these assumption and methods are going to be explored in following sections.

4.1 Methods

As introduced in chapter 1, the RGG interdependent network model is the upgraded model of the Buldyrev's model (ER interdependent network model). The upgraded feature in the interdependent network is the spatial feature, and it is updated for both intra-layer and inter-layer parts. Therefore, while the basic structures are similar, some details for geometrical and spatial system are additional in the methods section.

There are 4 unique models as introduced. In the intra-layer part, 2 of them contain spatial-based graph (RGG), while others contain non-spatial-based graph (ER). Also, in the inter-layer part, 2 of the models have spatial-based system, while others have random-based system. In this thesis, the models are named as: (*intra-layer graph*)-(*inter-layer system*). Therefore, the names of 4 models are: RGG-Spatial, RGG-Random, ER-Spatial and ER-Random. So, for example, the RGG-Random graph uses the RGG graph as intra-layer part and the Random system for inter-layer system, i.e,. the RGG graphs are Randomly interdependent. The naming system is arranged in the table 4.1.

Inter-layer \ Intra-layer	Random Geometric Graph	Erdos-Renyi Graph
Spatial System	RGG-Spatial	ER-Spatial
Random System	RGG-Random	ER-Random

Table 4.1: Names of 4 model types

4.1.1 Building Interdependent Network Model

A. Design the Environment

In the designing the environment part, most of the part is similar with chapter 3.1.1. However, there are differences on the part of set parameters. Although there is the number of nodes in each graph (N) and the number of layers (L) still, the probability is removed and the new parameters are added. Also, for fair comparison with the Buldyrev's model, the value of L is also set as 2. Primarily, the parameters used in building the interdependent network are included. First parameter is the intra-threshold ($intra\_T$), which is going to be used for RGG graph in intra-layer part. Next one is the inter-threshold ($inter\_T$), which is used for spatial system in inter-layer part. Both are the thresholds of the distance. The final updated parameter is the number of supporting nodes per one node (S). From $S = 1$, the new analysis can be carried by increasing the value.

Next, many lists and dictionaries for storing the information of the graphs are declared in this part. Although there is only one interdependent network model in the chapter 3, there are total 4 models in this chapter. And for fair comparison, the conditions and environments of the graphs, such as the number of edges and supporting nodes, must be same. So, the information of the early designed graph is stored, and used for constructing the next graphs. The order of building the models is: 'RGG-Spatial' → 'RGG-Random' → 'ER-Spatial' → 'ER-Random'. The stored information and used information in each step of modelling could be shown in both table 4.2 and table 4.3. Since the table is too long, it is divided into two tables. The simple example of handling the information from the tables is that: Once the node is distributed in the first step, building RGG-Spatial model, the node distribution (coordinates of nodes in the layer) is stored in the dictionary. Then, it will be used in step 2, 3 and 4. The more specific explanations about usage of the information are going to follow in next parts.

	Step 1: RGG-Spatial	Step 2: RGG-Random
Stored Information	1-1. Node distribution 1-2. Number of intra-edges 1-3. Edges of RGG graph 1-4. Supporting pairs in Spatial system	2-1. Supporting paris in Random system
Used Information		1-1. Node distribution 1-3. Edges of RGG graph

Table 4.2: Stored / Used information in the steps of building RGG-Spatial and RGG-Random

	Step 3: ER-Spatial	Step 4: ER-Random
Stored Information	3-1. Edges of ER graph	
Used Information	1-1. Node distribution 1-2. Number of intra-edges 1-4. Supporting pairs in Spatial system	1-1. Node distribution 2-1. Supporting pairs in Random System 3-1. Edges of ER graph

Table 4.3: Stored / Used information in the steps of building ER-Spatial and ER-Random

B. Build Intra-Layer Graph

In this part, the graphs in each layer are going to be constructed. As introduced in the previous parts, there are two types of graphs that are going to be implemented in the intra-layer part. For both of the graphs, nodes distributing method is introduced. Although the ER graph which is non-spatial is used in 'ER-Spatial' and 'ER-Random' models, the supporting system in 'ER-Spatial' and the initial attack in every model are based on the spatial feature in this simulation. Therefore, the step of node distributing in the layer space is included in this chapter. To maintain the same environment in every model, the node distribution is basically same in all models. Therefore, once the nodes are distributed in the first model, 'RGG-Spatial' model, the coordinates of the nodes are saved and used for the three alternative model. As presented in chapter 2.1.2, the node distribution in the Random Geometric Graph is based on the independent uniform distribution. And the layer that pymnet library provides in the interdependent network is the unit square. Thus, the locations of nodes are chosen randomly, uniformly, and independently in the unit square.

Once the nodes are distributed, the step of connecting the edges follows. First for RGG graph, the edge connection is based on the intra-threshold, ($intra\_T$). From the definition of the RGG graph in chapter 2.1.2, the edge between any node i and node j is linked if the distance between the nodes is smaller than ($intra\_T$). It is applied to this model exactly. So, for every pair of nodes in the same layer, the distance is calculated and the edge connection is determined. Then, the number of intra-edges and the data of edges (start and end point of all edges) are stored for future usage. The data of edges is used in 'RGG-Random' model. Therefore, model gets exactly same intra-layer edges as the 'RGG-Spatial' model to make the inter-layer system the only difference between 'RGG-Spatial' and 'RGG-Random'.

Next, for the ER graph, the edge connection is based on the number of intra-edges from the 'RGG-Spatial' model in previous paragraph. To make the two systems to be fairly comparable, the number of edges in two types of graphs are set to be the same. From the chapter 2.1.1, the expected number of edges of the ER graph, $G(N,t)$ is $\binom{N}{2}t$ where N is the number of nodes and t is the probability of edge creation. Let the number of edges from the 'RGG-Spatial' model as E. Then, from $\binom{N}{2}t = E$, the probability t could be obtained as: $t = \frac{2E}{N(N-1)}$. With the probability, the ER graph, $G(N, \frac{2E}{N(N-1)})$, is built through 'networkx' module. Then, delivering the edges to the interdependent network model and adjusting the number of edges are same logic as the chapter 3.1.1. Also, with the same reason, the data of edges (start and end point of all edges) is stored for following 'ER-Random' model.

C. Build Inter-layer Part

In this part, selecting the supporting pairs and connecting the supporting edges are going to be introduced. Since this part also contains the spatial feature, selecting the supporting pairs is different with previous chapter. For the 'Spatial' system, selecting is based on the RGG graph and inter-threshold ($inter\_T$) is used as a threshold.

The algorithm 4.1 shows the selecting method of supporting nodes for every node. As the Buldyrev's model introduced that one node is supported and supporting one unique node, our model also takes the system. So, the supporting pairs are also going to be used in this chapter too. First part is finding all possible pairs. The nodes in layer 'a' are a_is and the nodes in layer 'b' are b_js. Then, finding 2D lists of all possible [a_i, [all possible $b_j s$]] where the distance between a_i and all $b_j s$ are smaller than $inter\_T$ follows. In this step, the layers are supposed overlapped, so nodes a_i and b_j are supposed in the same layer. 'all_poss_pairs' is then sorted by the length of the list, [all possible $b_j s$] as a key. Sorting the 2D lists is an important step in order to give priority to nodes that lack supporting node options.

Secondly, the supporting node for each node is selected through for loop. In the loop, if the number of possible $b_j s$ for current a_i is bigger than 0, the method of selecting b_j starts. Selecting b_j is also based on the number of options. The number of a_is that selected each b_j is recorded in the variable, 'target_num'. Then, the b_j which has a small number in 'target_num' is going to be given a chance to be selected. Giving a chance to node with minimal options first is absolutely important because it could decide minimal $inter\_T$ value. However, in the first strategy (strategy 1) of modelling, the method of introducing the variable 'target_num' and giving opportunities primarily to b_j with fewer options are missed. There was sorting system for a_i only. In the second trial (strategy 2), the algorithms are updated

```
supporting_pairs = []

cur_nodes = the nodes in layer 'a', where a_i ∈ layer 'a'
target_nodes = the nodes in layer 'b', where b_j ∈ layer 'b'

all_poss_pairs = lists [a_i, [all possible b_js]] where all of distance(a_i, b_j) ≤ inter_T
s_all_poss_pairs = sorted(all_poss_pairs, key = the number of possible b_js)
target_num = list of the number of a_i selected for each b_j.

for _ ∈ cur_node do
    if the number of all possible b_js of s_all_poss_pairs[0][0] > 0 then
        supp_pair = (a_i, choose one of the b_js which has the least a_i chosen)

        Unselected b_js reduce their numbers by 1 in the target_number()

        remove(s_all_poss_pairs[0])
        remove(chosen b_j in all other lists in s_all_poss_pairs)

        s_all_poss_pairs = sorted(s_all_poss_pairs, key = the number of possible pairs)
    else
        supporting_pairs ← (a_i, −1)

```

Algorithm 4.1: Algorithm of selecting the supporting nodes with Spatial system

as it is now and selecting the supporting nodes becomes more efficient than before. Therefore, there are two analysis parts in this chapter. First analysis is with strategy 1 and second one is with strategy 2.

Then, selected b_j form a supporting pair, (a_i, b_j). Then, the list with the current a_i, which is s_all_poss_pairs[0] is removed. Also, the chosen b_j is removed in all other lists. Literally, the possibility of selecting b_j as a supporting node is eliminated for all other a_is. Else, the supporting node of a_i becomes -1 that represents there is not a supporting node. In the analysis parts, since the size of the *inter_T* could affect the possibility of forming the supporting pairs, selecting appropriate *inter_T* is going to be discussed.

Selecting the supporting node in 'Random-based' system are much easier. Since it is based on the randomness, pairing the a_i in layer 'a' and b_j in layer 'b' randomly is the only step. And then, in both 'Spatial-based' and 'Random-based' systems, the inter-edges are connected by using the supporting pairs. Additionally, in both systems, the data is stored for future models. The data which is going to be stored is the supporting pairs. The data from 'RGG-Spatial' model is used in 'ER-Spatial' model and the data from 'RGG-Random' model is used in 'ER-Random' model as introduced in the table 4.2 and table 4.3. With the data of supporting pairs, edges are connected in 'ER-Spatial' model and 'ER-Random' model. Appendix B shows the 4 models built with both strategy 1 algorithms and strategy 2 algorithms, introduced in above parts.

4.1.2 Cascading Failure

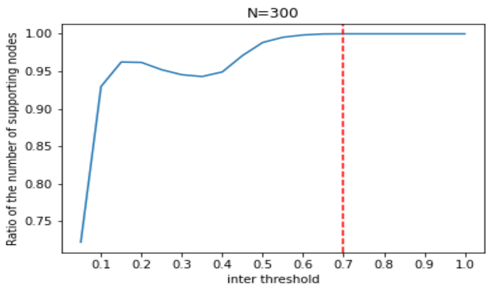
The initial attack and the cascading failure are going to be simulated in this part. The parameters for cascading attack should be also updated for this step. Since the spatial system is included in the network, it is also added into the attack. So, the new type of attack is built, which is the '*spatial_attack*'. It is the attack that affects a specific region, such as a earthquake and a bomb in the real world. Therefore, the attack point $(a = (i, j))$ where $0 \leq i, j \leq 1$, which is the centre of the spatial attack, is the new parameter for the new attack type. The size of attack (K) remains same, so the initial attack kills nearest K nodes from the attack point a.

In the chapter 3.1.2, the targets are selected only in one layer. However, the attack in the real world like earthquake does not affect only one layer because the layers indicates the same region in real world. Hence, K nodes in every layer are going to set as targets. Once all targets are selected, the attack and failure are going to be simulated. The simulation and methods are exactly same with the chapter 3.1.2.

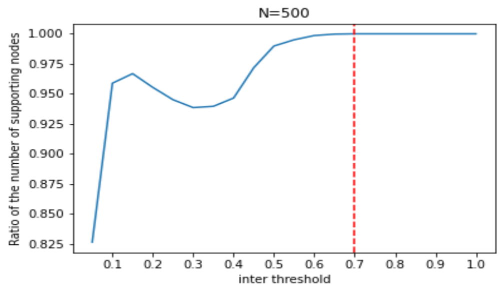
4.2 Analysis: Strategy 1

4.2.1 Finding Inter-Threshold

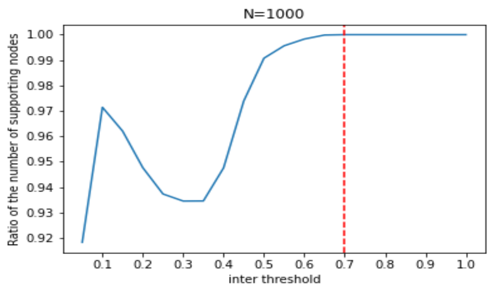
In the analysis section, the simulations with the models and the results are going to be introduced. The first step for starting simulations is finding appropriate *inter_T* that affects the supporting node system. As introduced in the chapter 4.1.1, the *inter_T* affects the number of supporting nodes' options for each node. So, if *inter_T* is too small, there could be nodes without the supporting nodes. And if *inter_T* is too big, the spatial property of the Random Geometric Graph could be diluted.



(a) Nodes=300



(b) Nodes=500



(c) Nodes=1000

Figure 4.1: The ratio of the number of supporting nodes

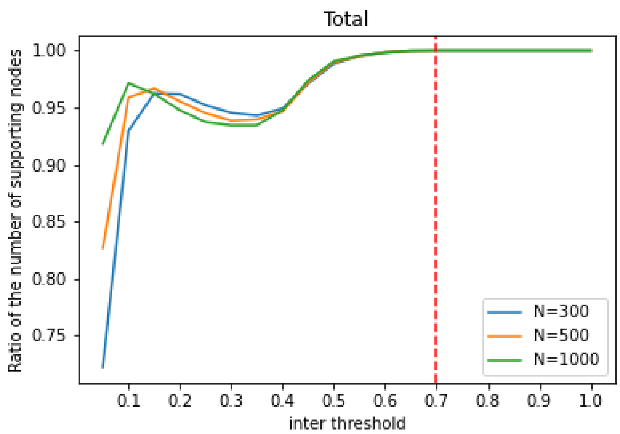


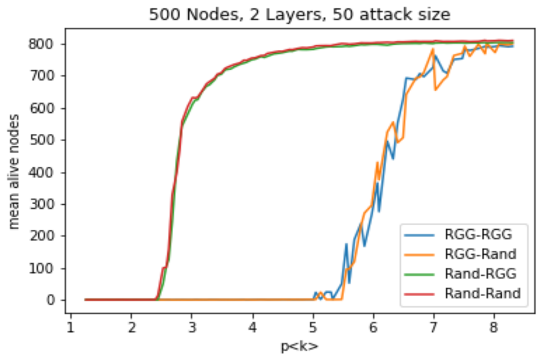
Figure 4.2: The ratio of the number of supporting nodes (total)

Figure 4.1 shows the ratio of the number of supporting nodes for each *inter_T* value. In the graph, the ratio exactly means the value, $the\_number\_of\_supporting\_pairs/N$. Each fig 4.1a, fig 4.1b, and fig 4.1c is the graph based on the number of nodes, 300, 500 and 1000 for each. To analyse them easily, the figure 4.2 is the sum of all 3 graphs. In the total figure, it can be observed that the number of nodes does not affect the tendency of the ratio of the number of supporting pairs. Although the starting points are different, they immediately develop similar tendencies. In the figure, when *inter_T* is smaller than 0.7, the value is unstable and it is also smaller than 1.0 in all of 3 graphs. Especially, near the threshold 0.2, the graph shows temporal decreases. The reason of decreasing could be interference between the nodes. When the value of *inter_T* is small, nodes can only consider the nodes that are very close to them as options. However, as *inter_T* getting larger, the options for each node is getting duplicated. So some nodes could still others' options that will affect the ratio of the number of supporting pairs. If the value of *inter_T* is huge enough, then the graph is stabilised since there are more options that could offset the interference. From 0.7, which is shown as a red dotted line, the ratio is stabilised and it always shows 1.0 in all of 3 graphs. Hence, for next parts of the analysis, *inter_T* is fixed as 0.7.

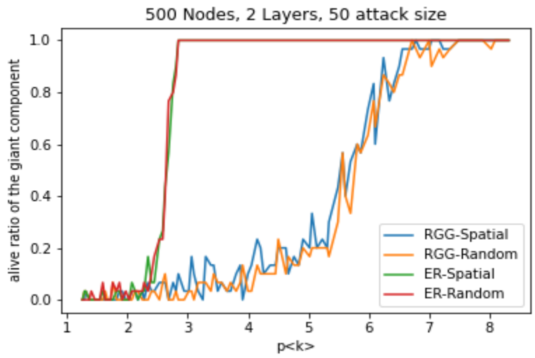
Since the algorithm of selecting the supporting pairs are not efficient as demonstrated in chapter 4.1.1, the value of *inter_T* is too big to show the spatial feature well. However, it is revealed after extracting and analysing the datasets from that algorithm and *inter_T* value. Therefore, in this section, the analysis with the inefficient algorithm follows first.

4.2.2 Relationship with the Mean Degree

From the chapter 3, which is the Buldyrev's model, the attacked networks show relationship in terms of the mean degree. Precisely, they have relations on the value $p\langle k\rangle$, where p is the proportion of the alive nodes after the initial attack and $\langle k\rangle$ is the mean degree of the initial graph. Therefore, the same test is progressed in this part.



(a) N=500, K=50, Mean alive nodes

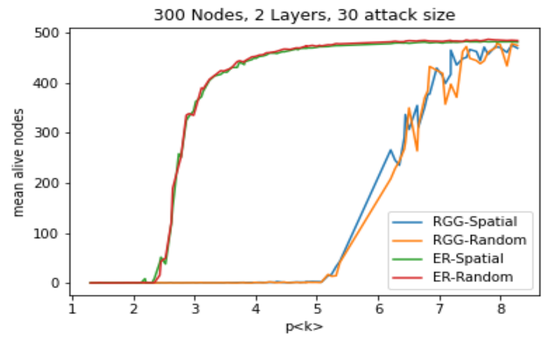


(b) N=500, K=50, Alive ratio

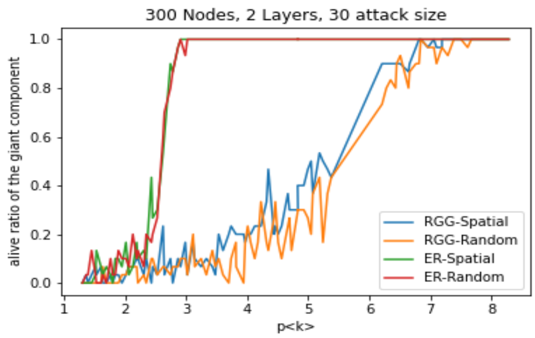
Figure 4.3: Mean alive nodes and alive ratio on each $p\langle k\rangle$

In the figure 4.3, the left graph shows the mean number of alive nodes and the right graph shows the alive ratio of the giant component. This model is simulated with $N = 500$ and $K = 50$. The 'ER-Random' model is exactly same model with the Buldyrev's one except one thing, the initial attack. This is the reason why the point of sharp growth in our 'ER-Random' model is slightly different with the Buldyrev's one. Generally, the figure presents that 'RGG-Spatial' and 'RGG-Random' models have similar graph and 'ER-Spatial' and 'ER-Random' models have similar graph. It represents that the intra-layer system affects more on the general consequence of cascading failure than inter-layer system. Specifically, the RGG graph in intra-layer system is more vulnerable than Random graph. To normalise it, more simulations are carried out with different number of nodes and size of initial attacks.

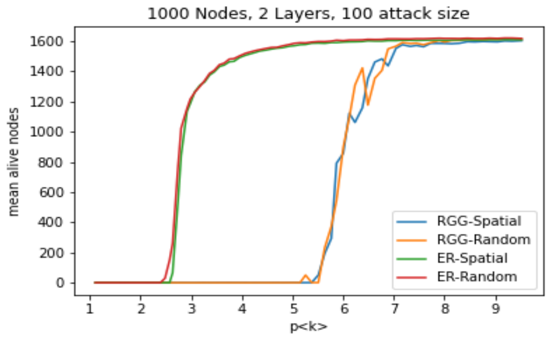
First, the simulations with different number of nodes are conducted. The percentage of dead nodes by initial attack is remained as same with above. Figure 4.4 shows the graphs with the simulations. In the fig 4.4a and fig 4.4b, there are graphs with $N = 300$. The graphs with $N = 1000$ are in fig 4.4c and fig 4.4d. There are total NxL nodes in each model and this is the reason why the value of alive node is larger than N. Next, the size of attack is changed while the number of nodes is same. Figure 4.5 contains graphs from simulations with different attack size. The results of simulation with increased attack size, $K = 100$ are in fig 4.5a and fig 4.5b. And in fig 4.5c and fig 4.5d, there are graphs with the attack size $K = 200$.



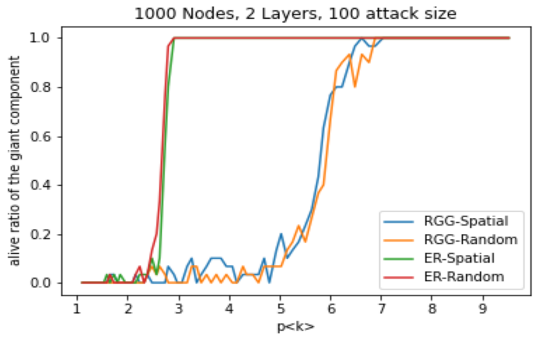
(a) N=300, K=30, Mean alive nodes



(b) N=300, K=30, Alive ratio

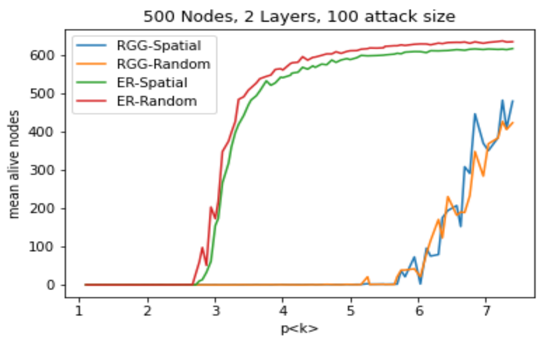


(c) N=1000, K=100, Mean alive nodes

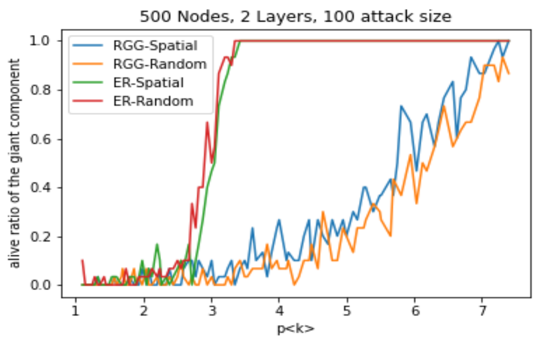


(d) N=1000, K=100, Alive ratio

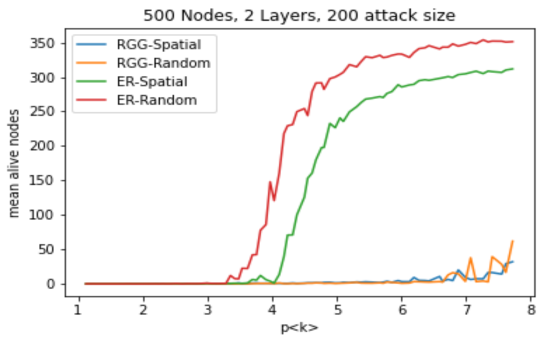
Figure 4.4: Same proportion of initial dead nodes, Different number of nodes



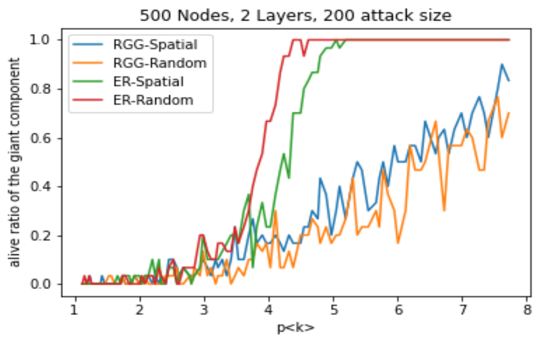
(a) N=500, K=100, Mean alive nodes



(b) N=500, K=100, Alive ratio



(c) N=500, K=200, Mean alive nodes



(d) N=500, K=200, Alive ratio

Figure 4.5: Same number of nodes, different proportion of initial dead nodes

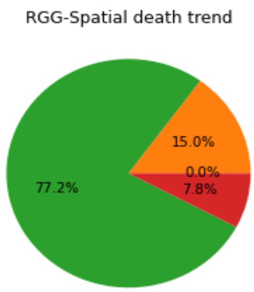
In these figures, the correlation between p, $\langle k \rangle$ and the alive ratio of the giant component from the models could be found. The value p is highly related with the survival of the giant component. In figure 4.4, graphs representing that same models show similar degree of growing on the similar point regardless of the number of nodes. However, if the p value is changed while the number of nodes is same, the shape of the graphs are also changed. In figure 4.5, $\langle k \rangle$ value is maintained as same condition with fig 4.3 since the number of nodes is fixed. As the size of the initial attack grows, the point that determines whether the largest component is alive is pushed to the right. This is different with the Buldyrev's model in chapter 3. Thus, the correlation between p, $\langle k \rangle$, N and the network type could be summarised as:

1. The interdependent network that contains RGG graph in intra-layer requires higher mean degree, $\langle k \rangle$, than ER graph to make giant component survive from the cascading failure.
2. The number of nodes, N, in interdependent network system does not affect the property of the cascading failure on the network much when the percentage of the initial attack is kept same.
3. As the ratio of the nodes survive initial attack, p, increases, the number of required edges to make the largest component survive after the cascading failure is also increased.

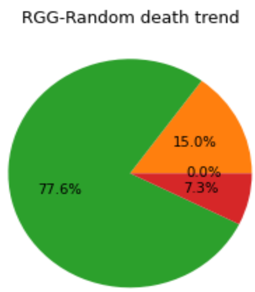
Additionally, two more points are observed with these result. The first point is that as initial attack size grows, the difference between the graphs with same intra-layer system is getting larger. In fig 4.3 and fig 4.4, the graphs with same intra-layer system do not show big difference. However, in fig 4.5, it is detected that the difference is getting bigger. The second point is that the models that have different system on intra- and inter-layer tends to be weaker than the network consists of same system. Figure 4.5c and figure 4.5d shows the gap between 4 models mostly in those results. In the figures, the green and yellow graphs are skewed more to the right and more gentle than the red and blue graph respectively. Hence, the largest components in RGG-Rand model and Rand-RGG model are less likely to survive in same condition after cascading failure compared to RGG-RGG model and Rand-Rand model respectively. Deeper analysis for these points follow in next sub-section.

4.2.3 Validating Four Models are Different

In this part, the analysis of whether the four models are really different or not. First of all, an analysis of the reason why there is a big difference between RGG and ER intra-layer systems is introduced. In previous subsection, the models contain the RGG graph as a intra-layer system require more edges to survive. The reason could be explained with following figure:



(a) N=500, K=50, intra_T=0.45



(b) N=500, K=50, intra_T=0.45

Figure 4.6: greeen: Dead from isolation, red: Dead from unsupported, orange: Initial attack, blue: Alive

Figure 4.6 and figure 4.7 show the pie graphs for each model. The interdependent networks are built with $N = 500$, $K = 50$ and $intra\_T = 0.45$ which make $2.5 < p\langle k \rangle < 3$. Therefore, with these parameters, RGG-Spatial and RGG-Random model does not have any alive components and ER-Spatial and ER-Random models only have alive giant components. In the pie graph, green color means the ratio of dead nodes from isolation, red color means the ratio of dead nodes from unsupported, the orange color means the ratio of initial dead nodes and the blue color means the ratio of alive nodes. From the graph, it could be discovered that the biggest reason that nodes die in RGG-Spatial and RGG-Random model

(a) N=500, K=50, intra_T=0.45 (b) N=500, K=50, intra_T=0.45

Figure 4.7: greeen: Dead from isolation, red: Dead from unsupported, orange: Initial attack, blue: Alive

is 'isolated' which is not in the giant component in the layer. The reason of this phenomena could be a different graph shape between the Spaitial-based model and Random-based model.

Figure 4.8 represents the size of the initial giant component for each model. In two cases, there is a very large difference in the size. The size of the largest component in RGG-Spatial and RGG-Random model is smaller than 20% of the total graph. And it leads to the result that more than 80% of nodes die in initial attack or the first step of cascading failure. It could be discussed that these difference is derived from the properties of RGG and ER graphs. Since the nodes in the RGG graph make local connections with neighbour nodes while the nodes in the Random graph make global connections, the size of the largest component increases more easily in the Random graph. Therefore, larger mean degree is needed for RGG-Spatial and RGG-Random model to make their largest component survive from the cascading attack.

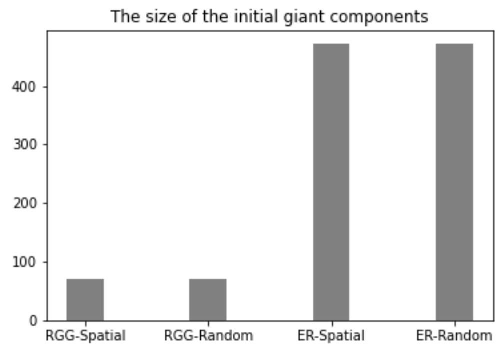


Figure 4.8: The size of the initial giant component

Next, the analysis of the growing differences between RGG-Spatial and RGG-Random and between ER-Spatial and ER-Random when the size of initial attack increases is carried out. To examine it, the dataset with $N = 500$ and $K = 200$ is used since large p value shows bigger differences. In previous part, there is only intuitive interpretations based on graphs. Therefore, at the beginning of the part, analysing whether there is a significant difference between the graphs based on statistical interpretations is presented. To analyse two graphs for comparison, two statistcal tests which are Kolmogorov–Smirnov test and Mann–Whitney U test are used. The Kolmogorov–Smirnov test is used to test whether the distribution is normal distribution or not. And Mann–Whitney U test is used to compare two sets which are not normally distributed. The alive ratio of the giant component is selected as the samples used in both statistical tests.

Table 4.4 shows the result from the statistical tests on 'ER-Spatial' and 'ER-Random' samples. For Kolmogorov–Smirnov test it could be said that the distribution is normal if the p-value is greater than

	ER-Spatial	ER-Random
Kolmogorov–Smirnov test	1.213e-23	1.460e-82
Mann–Whitney U test	1.373e-19	

Table 4.4: p-values. $N = 500$, $K = 200$ and $intra\_T = 0.0685$

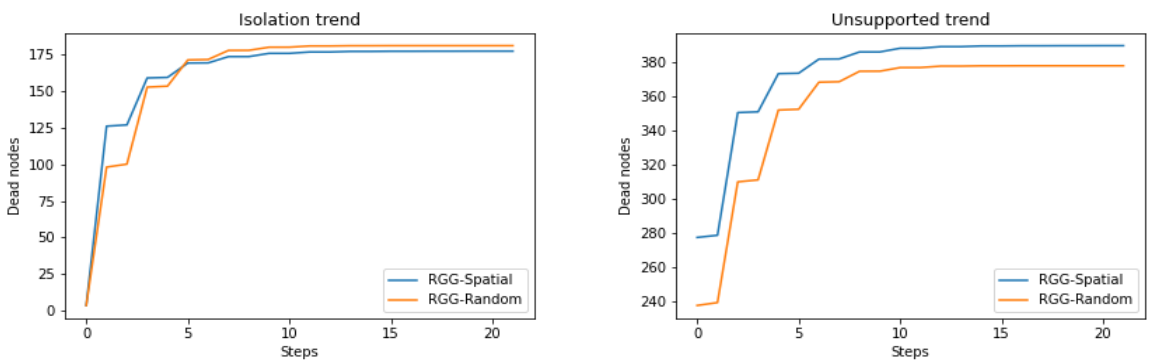
0.05. However, since the p-values on both graphs are smaller than 0.05, they are not normally distributed. Therefore, Mann–Whitney U test could be used to compare the two graphs. If the result of the test is greater than 0.05, it could be said that the two sets are derived from the same distribution. However, the result of the Mann-Whitney U test is much smaller than 0.05. Thus, it is revealed that ER-Spatial and ER-Random graphs are giving the alive ratio based on the different distribution.

	RGG-Spatial	RGG-Random
Kolmogorov–Smirnov test	2.761e-74	2.272e-46
Mann–Whitney U test	0.0009	

Table 4.5: p-values. $N = 500$, $K = 200$ and $intra\_T = 0.095$

Table 4.5 shows the result of the tests on RGG-Spatial and RGG-Random models. The figure 4.4 and figure 4.5 shows that the models with RGG as an intra-layer system and ER graph as an intra-layer system are significantly different. Therefore, the $intra\_T$ value is adjusted for statistical test to the point. So in both table 4.4 and table 4.5, alive ratio is about 0.5, which is the near range of the decision point. In this table, the p-values from the Kolmogorov-Smirnov test with both model much are smaller than 0.05. Therefore, they are not normally distributed. Also, in the Mann-Whitney U test, the p-value is smaller than 0.005. So, it can be revealed that two graphs of the alive ratio are not derived from the same distribution. Hence, the above results give an outcome in this section. The observations from the graphs in figures 4.6, 4.7, 4.8 could lead to the feature that models with RGG graph in intra-layer and ER graph in intra-layer are not from same distribution. And the tables 4.5 and 4.4 shows that different inter-layer systems make the models differently. Hence, the final outcome of this section is that all 4 models are derived from the different distributions, so they are definitely different.

4.2.4 Specific Differences between Models



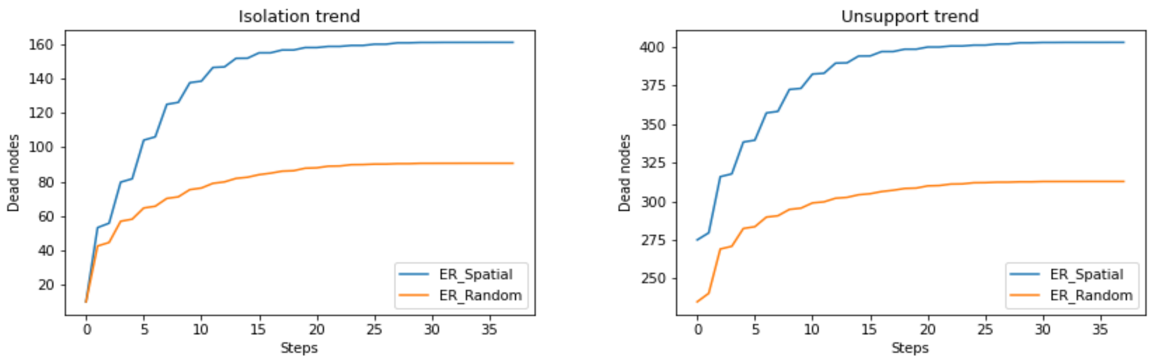
(a) Cumulative dead node by Isolation

(b) Cumulative dead node by Unsupported

Figure 4.9: N=500, K=200, inter_T=0.095, death trends in each model

Now, the specific differences between the models are analysed. Finding the difference between RGG-Spatial model and RGG-Random model is presented first. Figure 4.9 shows the trends of the death. In fig 4.9a, it can be observed that the number of dead node from isolation is similar in both models. The difference is mostly based on the death from unsupported. More nodes in the RGG-Spatial model tends to die from unsupported in cascading failure. So, when the network model does not collapse completely, the mean surviving nodes in RGG-Random model is larger than RGG-Spatial model. However, from the fig

4.5d, the largest component in RGG-Spatial model survives better than RGG-Random model. Also the dataset that is used in fig 4.9 shows that 82%of largest component in RGG-Spatial model survives while only 68%of largest component in RGG-Random model survives. Therefore, it could be demonstrated that the RGG-Rand model has more extreme results. This phenomena could be derived from the feature of the Random graph. Although the effect of the one node's death in RGG-Spatial model does not spread beyond the *inter_T* in every simulation, the effect in RGG-Random model does not have lower or upper bound. Therefore, the result also could not have any bounds, and it makes the extreme results.

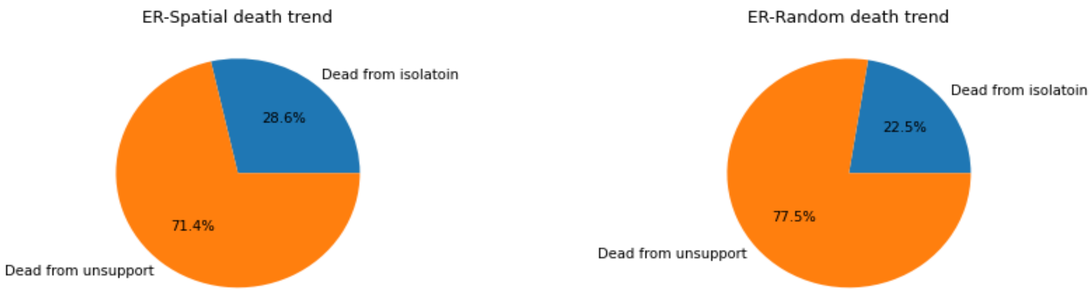


(a) Cumulative dead node by Isolation

(b) Cumulative dead node by Unsupported

Figure 4.10: N=500, K=200, inter_T=0.0685, death trends in each model

For ER-Spatial and ER-Random model, the difference between the models could be observed in fig 4.5c and fig 4.5d. Figure 4.10 shows the number of dead nodes for each reason. In this case, ER-Spatial always have more number of dead nodes in both of the two reason. However, from the detection in figure 4.11, it could be found that there is not a big difference of the ratio of death.



(a) ER-Spatial: real death trend

(b) ER-Random: real death trend

Figure 4.11: N=500, K=200, inter_T=0.0685, the ratio of the death

Therefore, without the ratio of the reasons of dead node, the thing should be focused is the number of dead nodes. For ER-Spatial nodes, the rapid raises of the dead node from isolation and unsupported are both happened in earlier step. It means that isolation and unsupported are happening simultaneously. Figure 4.12 shows the reason why it could be happened. The difference between ER-Spatial and ER-Random is more serious than between RGG-Spatial and RGG-Random. This is because the degree to which two type of models are clustered is different. The clustering coefficient of initial network for each 4 models could be detected in figure 4.12. The clustering coefficients of ER-Spatial and ER-Random models are fairly small values. It means that the degree of clustering is very low, and so the clusters could be break down easily. In this situation, in ER-Spatial model, the failures after the initial attack are happened relatively locally in the network then ER-Random. Then, the failures are concentrated, and so break the giant component into a number of clusters well rather than ER-Random graph. Therefore, more serious cascading failures could follow it.

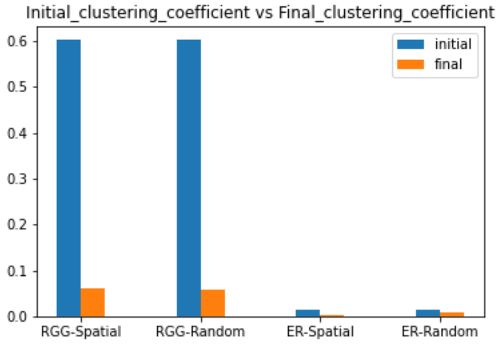


Figure 4.12: Clustering Coefficients for initial and attacked network

With above analysis, the detection from the fig 4.5 that RGG-Random and ER-Spatial models are weaker than RGG-Spatial and ER-Random model respectively is also found in data analysis. And the reasons of the phenomena were also introduced above. With the analysis and reasoning, the effects of spatial and non-spatial on intra- and inter-layer were discussed. In the next part, with the complemented algorithm of finding inter-threshold, more precise results will be analysed.

4.3 Analysis: Strategy 2

Before start the analysis with strategy 2, the summary of the analysis with strategy 1 in previous section is exhibited in this paragraph to compare both strategies well. In previous analysis, it is revealed that:

- In Finding Inter-threshold:
  1. Near the value $inter\_T = 0.2$, it is expected that the interference on selecting the supporting nodes between neighbour nodes could occur. And as the value growth, more number of options offset the interference.

- In Relationship with the Mean Degree:
  1. Higher mean degree, $\langle k \rangle$, is required to the interdependent network with RGG graph in intra-layer than ER graph, to make giant component survive from the cascading failure.
  2. The number of nodes, N, does not affect the property of network and the cascading failure much when the percentage of the initial attack is kept same.
  3. The number of required edges to make the giant component survive after the cascading failure is increased, as the value p increases.

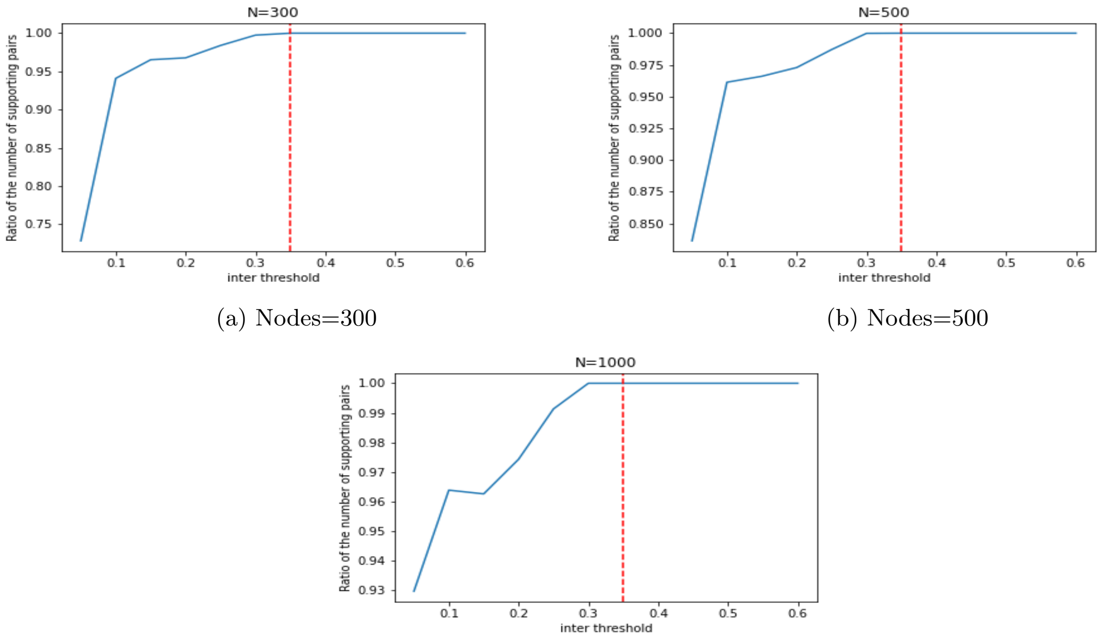
- Additionally in Relationship with the Mean Degree:
  1. The models that have different system on intra- and inter-layer tends to be weaker than the network consists of same system. Between RGG-Spatial and RGG-Random, this is because that there is not any upper or lower bound in Random-based inter-layer system. Between ER-Spatial and ER-Random model, this is because since the clustering coefficient of ER graph is too small, the Spatial graph could break the giant component easily.
  2. From above, it could be known that the difference between RGG and ER graphs could be manifested well in radical situation. Therefore, the difference between the graphs with same intra-layer system is getting bigger as initial attack size grows.

- Spatial and Non-spatial based systems show clear differences on the statistical tests.

Then, the same analysis as chapter 4.2 is going to be introduced. Since the graphs and the methods of analysis is exactly same, duplicated explanations are going to be omitted.

4.3.1 Finding Inter-Threshold

Since the new algorithm shows strong efficiency in finding the inter-threshold value, there might be a big difference with previous part.



(a) Nodes=300

(b) Nodes=500

(c) Nodes=1000

Figure 4.13: The ratio of the number of supporting nodes

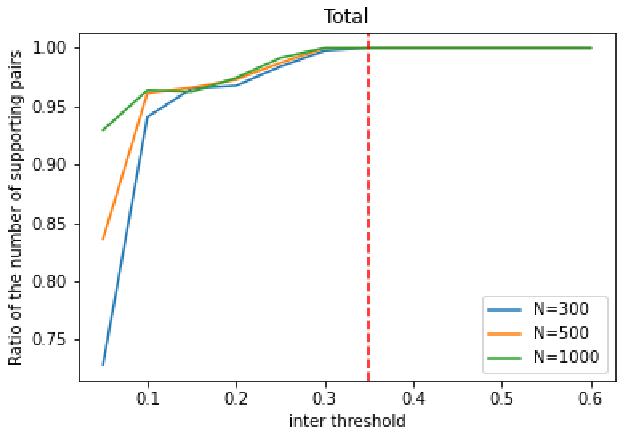


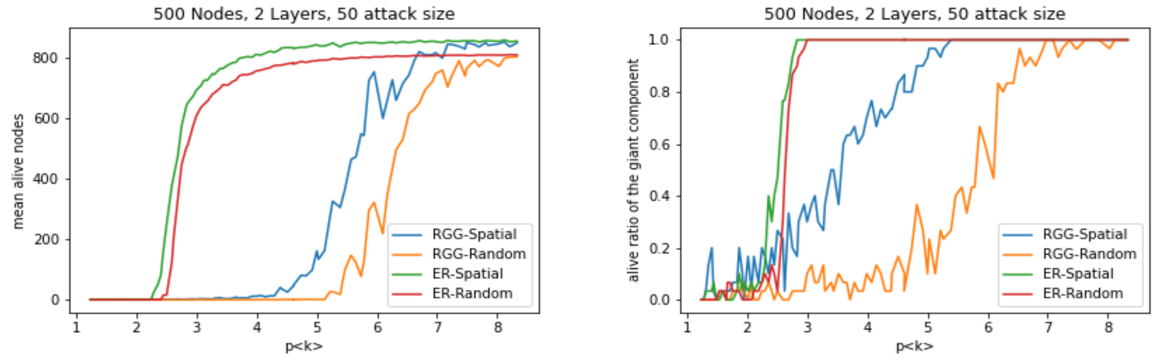
Figure 4.14: The ratio of the number of supporting nodes (total)

Different with the previous part, it could be observed that the minimal value of the inter-threshold is 0.35 rather than 0.7 in figure 4.13. This is the half of the previous one, and it could indicate that not only the efficiency massively grows, but also the spatial feature of inter-layer system is much more strong. Therefore, with the result from this inter-threshold, it could focus on the difference between spatial and non-spatial system well. Moreover, by comparing it with analysis 1, it is possible to fix the wrong analysis in previous part.

These graphs in figure 4.13 also show that the number of nodes, N, does not affect the ratio of the number of supporting pairs on each inter-threshold value except in very small numbers. Also interestingly, there is a common point that the graphs show temporal decreases as previous part. Therefore, the assumption of the interference between nodes could be more persuasive with this result. Next, new analysis with this new $inter\_T = 0.35$ value are going to be demonstrated.

4.3.2 Relationship with the Mean Degree

Finding the relationships with p, $\langle k \rangle$ and the alive ratio of the giant component also proceeds in this section.



(a) N=500, K=50, Mean alive nodes

(b) N=500, K=50, Alive ratio

Figure 4.15: Mean alive nodes and alive ratio on each $p\langle k \rangle$

Figure 4.15 shows both of the similarities and differences from the Analysis: strategy 1. First, the similarities are: 1) Generally, the models with ER graphs in intra-layer shows skewed left graphs and the models with RGG grpahs in intra-layer shows relatively skewed right graphs. and 2) the models with ER graphs shows relatively sudden growth while the models with RGG graphs shows relatively gradual growth. The differences are: 1) ER-Spatial model is more skewed to the left than ER-Random model while it was opposite in previous analysis. and 2) In the graph of the alive ratio, the RGG-Spatial model is so skewed that the point at which it begins to rise is no different from ER-Spatial and ER-Random. These are going to be focused more while the size of node and attack are changed.

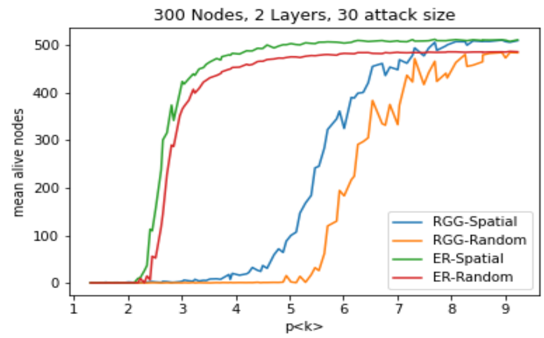
Figure 4.16 presents the models with diffrent number of nodes and same percentage of the size of initial attack. By comparing it with figure 4.15, it could be revealed that the size of node does not affect the relationship between the alive ratio of the giant component and $p\langle k \rangle$. In figure 4.17, there are the graphs with same number of nodes and different initial attack size such as $K = 100, 200$. As the size of attack increase, the RGG-Random and ER-Random models show that the graph moves to the right. However, the RGG-Spatial and ER-RGG models show that there is not any change even as the size of attack grows.

This is the first major difference with the analysis 1. In previous chapter 4.2, every model shows movement to the right side as the size of attack grows. Even ER-Spatial model shows more movement than others, However, with new algorithm, the models with Spatial features in inter-layer system which are ER-Spatial and RGG-Spatial models does not show any movement. Since the spatial features in the new algorithm is much strong, the observe result must be updated from the previous analysis to this analysis. Therefore, the correlations between p, $\langle k \rangle$ and the alive ratio could be summarised as:

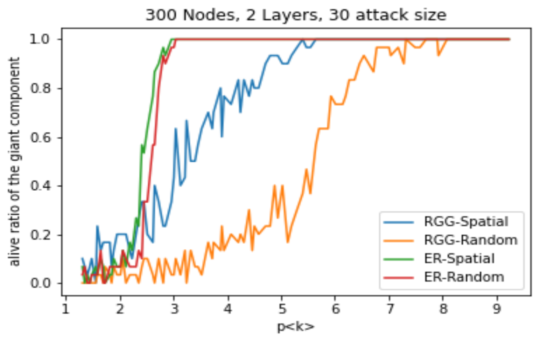
1. The interdependent network that contains RGG graph in intra-layer part requires higher mean degree $\langle k \rangle$ to make giant component survive.

2. The number of nodes, N, does not affect the property of the cascading failure and the relations when the value of p does not changed.

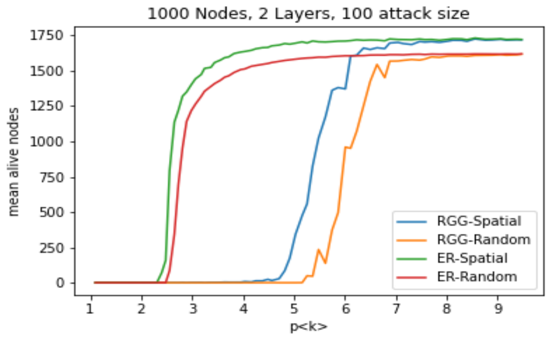
3. As the value of p increases, RGG-Spatial and ER-Spatial models remains same relationships while RGG-Random and ER-Random requires more edges to make the giant component survive.



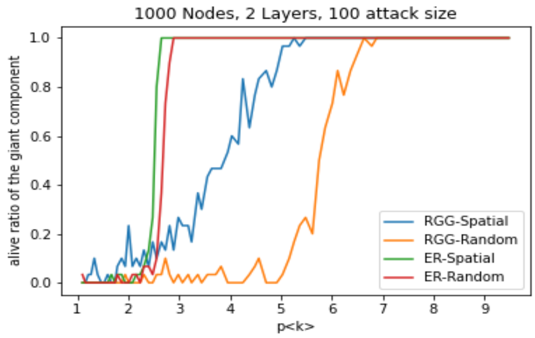
(a) N=300, K=30, Mean alive nodes



(b) N=300, K=30, Alive ratio

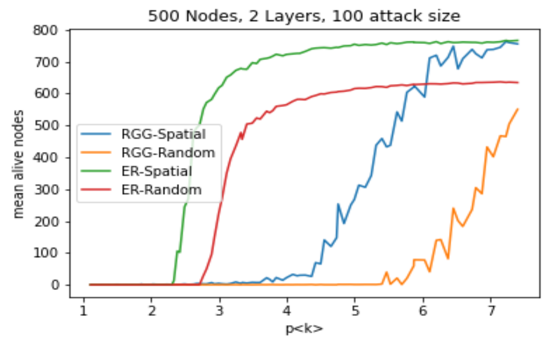


(c) N=1000, K=100, Mean alive nodes

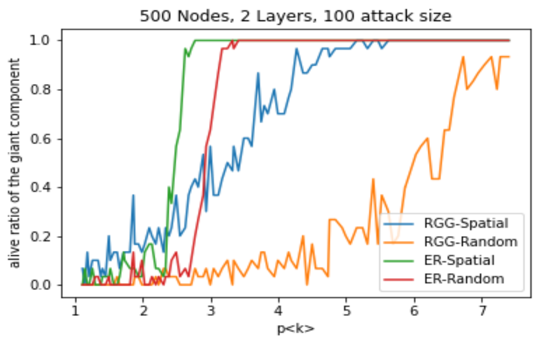


(d) N=1000, K=100, Alive ratio

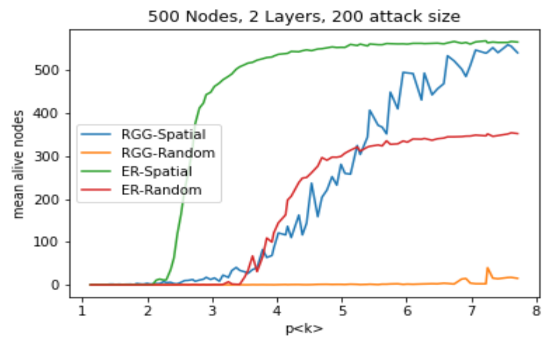
Figure 4.16: Same proportion of initial dead nodes, Different number of nodes



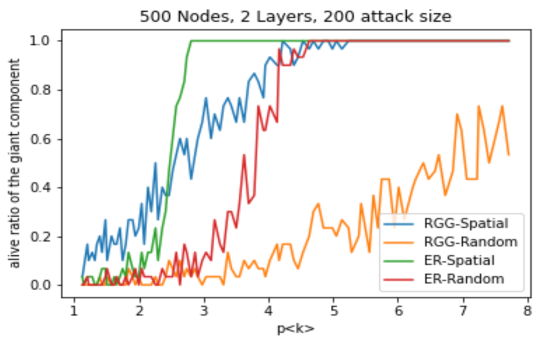
(a) N=500, K=100, Mean alive nodes



(b) N=500, K=100, Alive ratio



(c) N=500, K=200, Mean alive nodes



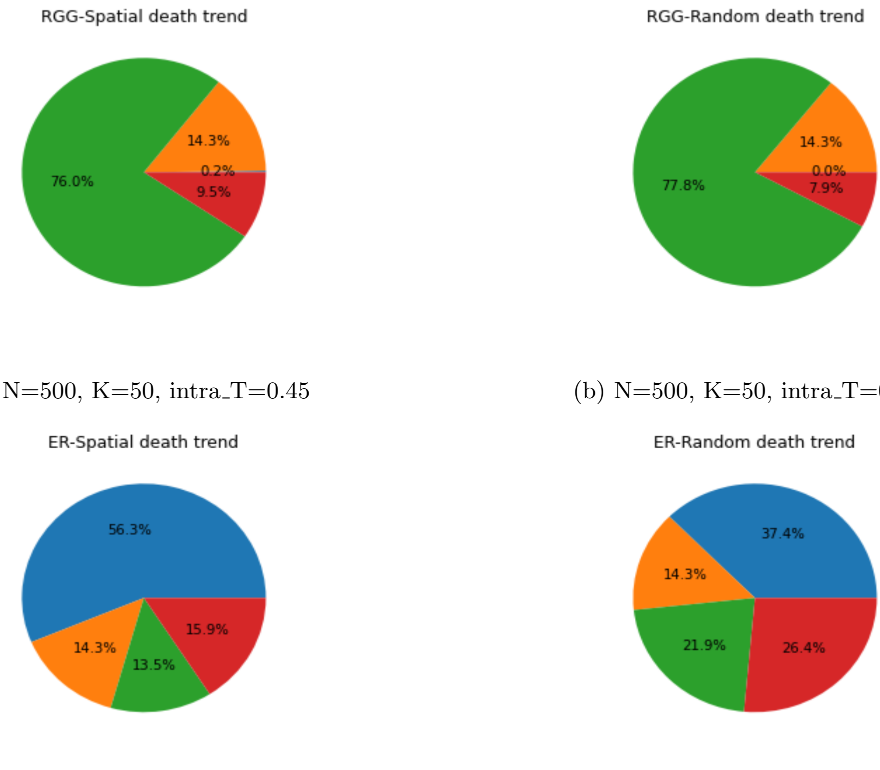
(d) N=500, K=200, Alive ratio

Figure 4.17: Same number of nodes, different proportion of initial dead nodes

The further analysis through the differences between spatial and non-spatial model is going to follow.

4.3.3 Validating Four Models are Different

First of all, the difference between the RGG graph and ER graph in intra-layer system is going to be focused. In various previous chapters, it is shown that the model with RGG graph requires more edges to make giant component survive. In the following paragraph and figure, there will be the step of checking to see if previous analyses are correct.



(a) N=500, K=50, intra_T=0.45 (b) N=500, K=50, intra_T=0.45

(c) N=500, K=50, intra_T=0.45 (d) N=500, K=50, intra_T=0.45

Figure 4.18: greeen: Dead from isolation, red: Dead from unsupported, orange: Initial attack, blue: Alive

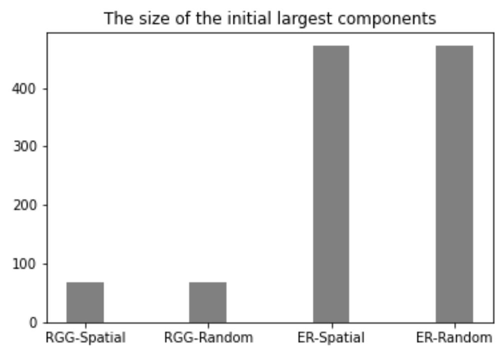


Figure 4.19: The size of the initial giant component

Figure 4.18 shows the death trend in each model. The pie graphs of RGG-Spatial and RGG-Random looks similar with the pie graphs in chapter 4.2. It indicates that most of the nodes are dead from isolation. Also, with figure 4.19 it could be shown that models with ER graphs tend to construct the giant component much faster. Therefore, in the models with RGG graphs, more number of edges are needed to make giant component basically.

	RGG-Spatial	RGG-Random
Kolmogorov-Smirnov test	1.213e-23	1.213e-23
Mann-Whitney U Test	9.6895e-11	

Table 4.6: p-values. N = 500, K = 100 and *intra_T* = 0.05

Table 4.6 shows the result of Kolmogorov-Smirnov test and Mann-Whitney U Test with the datasets from RGG-Spatial and RGG-Random models. The attack size, K, is chosen as 100 since there is appropriate gap between two models. And the value of *intra_T* is determined since it is the moderate value learned from previous analysis. The sampled datasets are the alive ratio of the giant component. The result says that both RGG-Spatial and RGG-Random models are not normally distributed first. Since the p-values derived from Kolmogorovv-Smirnov test is lower than 0.05. Also, the p-value from the Mann-Whitney U test is also lower than 0.05, which means that two models are not based on the same distribution.

	ER-Spatial	ER-Random
Kolmogorov-Smirnov test	2.5353e-170	0.000
Mann-Whitney U Test	1.1331e-34	

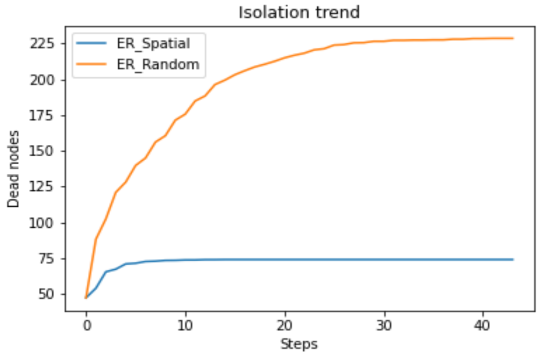
Table 4.7: p-values. N = 500, K = 100 and *intra_T* = 0.073

Table 4.7 presents the result of Kolmogorov-Smirnov test and Mann-Whitney U Test with the datasets from ER-Spatial and ER-Random models. The parameters, K and *inter_T* are chosen in similar way of table 4.6. The result of Kolomogorov-Smirnov test shows that both models are not normally distributed since the p-value is extremly smaller than 0.05. Even, the p-value of ER-Random model is 0. Additionally, the p-value of Mann-Whitney U test, which is also smaller than 0.005, represents that the models are not came from same distribution. Hence, with intuitive results in the graphs and the results from the statistical test, it is proven that all 4 models are unique.

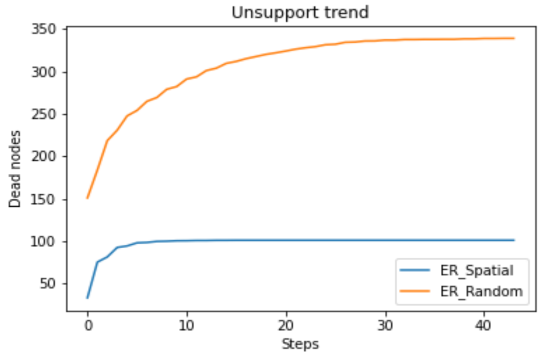
4.3.4 Specific Differences between Models

Since the most differences with the chapter 'Analysis: Strategy 1' are in this part, the validation and examination are introduced in this part. The analysis and result of this section is based on more powerful algorithm and models with better spatial feature, It is going to be assumed that there could be errors in previous chapter, 'Analysis: Strategy 1'. In previous result and current result, the difference of RGG-Spatial and RGG-Random model is not a serious problem since the difference is the size of the gap between two models. Therefore, it is possible to upgrade the assumption in chapter 4.2.4. The assumption is that since there is not a upper or lower bounds in inter-layer system of RGG-Random model, the failure could affect all over the networks or just a small part of the network. And this is the reason why RGG-Random models shows lower alive ratio of the giant component while the number of nodes that survive is higher. However, as the value of *inter_T* decreases, there is a serious change.

Although the size of alive ratio of the giant component still shows same tendency, the order of the number of survive node is reversed. Figure 4.20 shows that there are more dead nodes in RGG-Random model. Therefore, with decreased *inter_T*, RGG-Spatial model has the large alive ratio of the giant component with higher number of alive nodes. It can be explained based on the previous assumption. As *inter_T*, the area that failure in RGG-Spatial can affect is also decreased. So, the number of nodes affected has decreased. And it also increased the chances of the giant component surviving, and increased the mean number of surviving nodes. In the perspective of RGG-Random model, the probability that failure based on the random edge affect a wider range than RGG-Spatial model is increase as *inter_T* decreases. Therefore, the gap of alive ratio gradually widen and the number of alive nodes is also reversed.



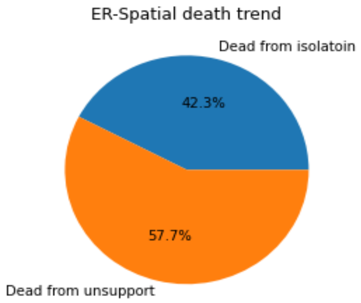
(a) Cumulative dead node by Isolation



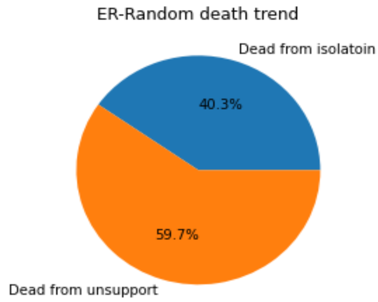
(b) Cumulative dead node by Unsupported

Figure 4.20: N=500, K=100, inter_T=0.05, death trends in each model

The model ER-Spatial and ER-Random give opposite result in this part. In previous chapter 4.2.4, the ER-Random model shows larger alive ratio of the giant component and higher number of mean alive nodes. However, it is totally reversed in this section. Figure 4.17 shows that the ER-Spatial model shows higher alive ratio and mean alive nodes than ER-Random model. The graphs that indicates the characteristics of both models are also similar.



(a) ER-Spatial: real death trend



(b) ER-Random: real death trend

Figure 4.21: N=500, K=100, inter_T=0.05, the ratio of the death

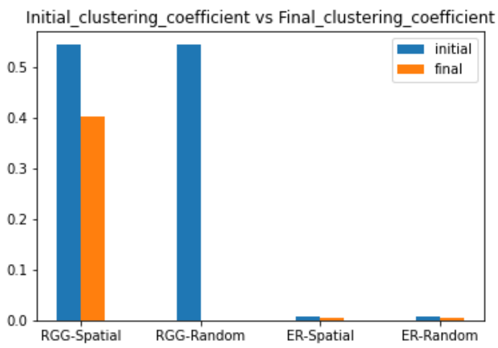


Figure 4.22: Clustering Coefficients for initial and attacked network

Figure 4.21 and figure 4.22 shows the proportion of the reasons of death and clustering coefficient of the graph. Similar with previous chapter 4.2.4, the ratio of the reasons are similar and clustering coefficient is also extremely small. In previous chapter, the assumption was that: since ER-Spatial graph generates the cascading failure in relatively small region, it could easily break some points of giant component and lead them to die. If the assumption is right, there must be more dead node in ER-Spatial graph with small size of $inter\_T$ value. However, the result is not. Hence, the assumption in previous analysis 4.2.4 should be terminated.

Then, without the previous result, it is possible to bring new assumption that applies on the current result. It is also based on the reachable area through $inter\_T$. Since the reachable area of attack in ER-Spatial model is smaller than ER-Random model, there will be more alive nodes than ER-Random model. However, this assumption has many weakness because it could not explain the result in previous analysis and why the alive ratio of the giant component is not affected by p.

Chapter 5

Conclusion

5.1 Achievement and Critique

The aim of this thesis is replicating the Buldyrev's interdependent network and cascading failure model first, and upgrading it with spatial features. Then, comparing the spatial and non-spatial models with simulations of cascading failure to observe the characteristics and differences. Therefore, the first main focus is building a replicated model and cascading failure system. On constructing the network, the Erdős–Rényi Graph is used for intra-layer graph. The achievement of this focus is checked through the result of the simulations. Due to lack of time and devices, running the simulations with relatively small number of nodes is the only option. However, it shows that our model follows the Buldyrev's model and also the theoretical base introduced in the background information part. Near the value $p\langle k\rangle = 2.4554$, the models shows significant raises that follows both theory and the Buldyrev's model.

The next focus is implanting the spatial feature into the network and failure model. To add a spatial feature, the new type of intra-layer graph is introduced, which is the Random Geometric Graph. Also, the spatial system is introduced in inter-layer part for choosing the selecting pairs. The achievement of this focus could be observed by visualising the network model. Initially, there was an obstacle on building the algorithm of spatial feature in inter-layer part. For this reason, the first results are not well-developed and the spatial feature on the results are already diluted more than expected. Hence, the upgraded algorithm is developed by adding a priority system. With the algorithm, efficiency increases a lot and also the spatial feature in the model becomes stronger than before.

After constructing both non-spatial and spatial network models, analysing the dataset from the simulations becomes a new focus. There are total 4 models which are: RGG-Spatial model, RGG-Random model, ER-Spatial model and ER-Random model. The models are named as (*intra-layer system*)-(*inter-layer system*) model. So for example, RGG-Random model is the interdependent network that RGG graphs are randomly interdependent. Since the Buldyrev's model shows that the models have relationship based on $p\langle k\rangle$, the same analysis proceeds in this chapter too. Then, there are the results which are:

1. The interdependent model with RGG graph in intra-layer requires higher mean degree, $\langle k\rangle$, to make giant component survive the cascading failure.
2. The number of nodes, N, does not affect the cascading failure model and the relations between the graphs and $p\langle k\rangle$ if the value of p is maintained.
3. As the value of p increases, models with Random based inter-layer system requires more edges to make the largest component survive. However, models with Spatial based int-layer system such as RGG-Spatial and ER-Spatial does not show and movement while the value of p changes.

Next focus is comparing each models to check whether the models are comes from same distribution or not. Although there are differences on the graph, the differences could be an observational error. Therefore, analysing the graph and adopting statistical tests are steps for this focus. Analysing the graph is based on various graphs comes from datasets. And by applying the logical assumption, it could be possible to get meaningful result. The statistical tests used in this thesis are 'Kolmogorov–Smirnov Test' and 'Mann-Whitney U Test'. By verifying that the distributions are not normally distributed through

'Kolmogorov–Smirnov Test', 'Mann-Whitney U Test' is applied to get the result which is whether the chosen two datasets are derived from same distribution or not. With those verification, it is revealed that the 4 models are unique and not came from the same distribution.

Lastly, from the assumptions in previous focus, developing the assumptions and finding the supporting evidence of it is the partially achieved part. This part is selected as a focus to check the reason that makes the differences of spatial and non-spatial based graph. Some of the assumption is strong enough while others not. For example, there is an assumption from the phenomena, RGG-Spatial graph tends to have higher alive ratio of the giant component and the number of mean alive nodes than RGG-Random graph. The assumption is that it is because each node in spatial based inter-layer system has limited area that could affect in this step, while random based system does not have any upper or lower bound. The reason why this assumption is strong is that there is supporting evidences. It is supported by the outcome of comparing the result with initial algorithm with weak spatial feature and upgraded algorithm with strong spatial feature. As the area that one node could affect in a step becomes smaller, the alive ratio and the number of alive nodes gets higher. With these supports, the assumption could be regarded as strong. By contrast, there are also weak assumptions. This is the reason why this part is partially achieved.

To summarise the thesis, it is revealed that RGG graphs in intra-layer requires more edges to build giant cluster than ER graph. It is because there is a limitation of distance. However, it could be analysed that once the giant component is built, it is hard to break it down into small components since the connection between near nodes are dense. In the inter-layer system, spatial-based system shows strengths since the distance limitation makes the failure not to spread out. Therefore, although the local part could be wiped out, the possibility that whole graph could survive is much higher than random-based system. In the real world, most of the networks are based on the spatial network. Therefore, it could be shown that selecting appropriate $inter\_T$ value is important. If it is too small, it is much more vulnerable than the Random-based graphs. If it is too large, it takes too much capital. In the middle point, the intra-layer graph could be strong enough since edges between neighbours are dense and the failure could not be spread out much. Hence, finding the balanced position between both side is essential in real world.

5.2 Further Work

There are some points that would like to be completed in further work.

1. The comparison between RGG-Spatial model and ER-Random model in terms of the property of networks such as the degree centre, the survived far node could be the further work.
2. The reasons that make differences between spatial based and non-spatial based interdependent network could be the further work. By changing the parameters in the model and running various simulations, the relationship of each feature that could be the key of the reason could be found.
3. In the real world, there are more than one supporting nodes in most of the system. Therefore, there is a limit to express the real world in our model. Therefore, changing the number of supporting nodes could be the further work.
4. With 2 layers interdependent network, the supporting system is monotonous. Therefore, bring more layers into the interdependent network could be a further work. With 3 or more layers, the supporting system could be varied such as layer i supports layer j while not support layer k.

Bibliography

[1] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

[2] Michael M Danziger, Amir Bashan, Yehiel Berezin, Louis M Shekhtman, and Shlomo Havlin. An introduction to interdependent networks. In *International Conference on Nonlinear Dynamics of Electronic Systems*, pages 189–202. Springer, 2014.

[3] Frédéric Petit. Critical infrastructure interdependency analysis: Operationalising resilience strategies.

[4] Steven M Rinaldi, James P Peerenboom, and Terrence K Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE control systems magazine*, 21(6):11–25, 2001.

[5] Jianan Zhang et al. *Modeling, analysis, and control of interdependent networks*. PhD thesis, Massachusetts Institute of Technology, 2018.

[6] UCTE Investigation Committee et al. Final report of the investigation committee on the 28 september 2003 blackout in italy. *Union for the Co-ordination of Transmission of Electricity*, 2004.

[7] Guidong Zhang, Zhong Li, Bo Zhang, and Wolfgang A Halang. Understanding the cascading failures in indian power grids with complex networks theory. *Physica A: Statistical Mechanics and its Applications*, 392(15):3273–3280, 2013.

[8] Yong Tang, Guangquan Bu, and Jun Yi. Analysis and lessons of the blackout in indian power grid on july 30 and 31, 2012. In *Proceedings of the CSEE*, volume 32, pages 167–174, 2012.

[9] Gaihua Fu, Richard Dawson, Mehdi Khoury, and Seth Bullock. Interdependent networks: vulnerability analysis and strategies to limit cascading failure. *The European Physical Journal B*, 87(7):1–10, 2014.

[10] Marc Barthelemy. Discussion: social and spatial networks. *Les nouvelles de l'archéologie*, (135):51–61, 2014.

[11] George Kingsley Zipf. *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books, 2016.

[12] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33):11623–11628, 2005.

[13] Dong-Hoon Shin, Dajun Qian, and Junshan Zhang. Cascading effects in interdependent networks. *IEEE Network*, 28(4):82–87, 2014.

[14] Matias Alberto Di Muro, Lucas Daniel Valdez, HH Aragão Rêgo, SV Buldyrev, HE Stanley, and Lidia Adriana Braunstein. Cascading failures in interdependent networks with multiple supply-demand links and functionality thresholds. *Scientific reports*, 7(1):1–10, 2017.

[15] Liang Tang, Ke Jing, Jie He, and H Eugene Stanley. Complex interdependent supply chain networks: Cascading failure and robustness. *Physica A: Statistical Mechanics and its Applications*, 443:58–69, 2016.

[16] Dror Y Kenett, Jianxi Gao, Xuqing Huang, Shuai Shao, Irena Vodenska, Sergey V Buldyrev, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Network of interdependent networks: overview of theory and applications. *Networks of Networks: The Last Frontier of Complexity*, pages 3–36, 2014.

[17] Paul Erdös and Alfréd Rényi. On the evolution of random graphs. In *The structure and dynamics of networks*, pages 38–82. Princeton University Press, 2011.

[18] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[19] Rob Kaas and Jan M Buhrman. Mean, median and mode in binomial distributions. *Statistica Neerlandica*, 34(1):13–18, 1980.

[20] Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Random graphs with arbitrary degree distributions and their applications. *Physical review E*, 64(2):026118, 2001.

[21] Natarajan Meghanathan. A random network model with high clustering coefficient and variation in node degree. In *2015 8th International Conference on Control and Automation (CA)*, pages 54–57. IEEE, 2015.

[22] Zhenning Kong and Edmund M Yeh. Correlated and cascading node failures in random geometric networks: A percolation view. In *2012 Fourth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 520–525. IEEE, 2012.

[23] Jesper Dall and Michael Christensen. Random geometric graphs. *Physical review E*, 66(1):016121, 2002.

[24] Ge Chen, Changlong Yao, and Tiande Guo. The asymptotic size of the largest component in random geometric graphs with some applications. *Advances in Applied Probability*, 46(2):307–324, 2014.

[25] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[26] Zaynab Hammoud and Frank Kramer. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Analytics*, 5(1):1–18, 2020.

[27] Lingna Wang, Guanghu Zhu, Huiyan Kang, and Xinchu Fu. Epidemic spreading on three-layer interdependent networks. *Journal of Biological Systems*, 24(04):469–494, 2016.

[28] Jean-Francois Castet and Joseph H Saleh. Interdependent multi-layer networks: Modeling and survivability analysis with applications to space-based networks. *PloS one*, 8(4):e60402, 2013.

[29] Jianan Zhang and Eytan Modiano. Connectivity in interdependent networks. *IEEE/ACM Transactions on Networking*, 26(5):2090–2103, 2018.

[30] Ming Li, Run-Ran Liu, Linyuan Lü, Mao-Bin Hu, Shuqi Xu, and Yi-Cheng Zhang. Percolation on complex networks: Theory and application. *Physics Reports*, 2021.

[31] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[32] George Marsaglia, Wai Wan Tsang, Jingbo Wang, et al. Evaluating kolmogorov's distribution. *Journal of statistical software*, 8(18):1–4, 2003.

[33] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

[34] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

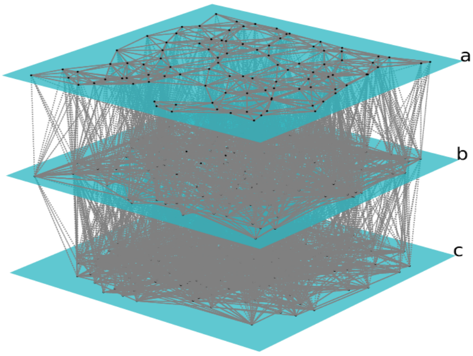
[35] Patrick E McKnight and Julius Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[36] Jerrold H Zar. *Biostatistical analysis*. Pearson Education India, 1999.

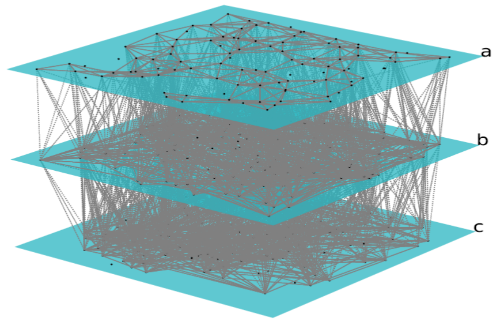
[37] Jia Shao, Sergey V Buldyrev, Reuven Cohen, Maksim Kitsak, Shlomo Havlin, and H Eugene Stanley. Fractal boundaries of complex networks. *EPL (Europhysics Letters)*, 84(4):48004, 2008.

Appendix A

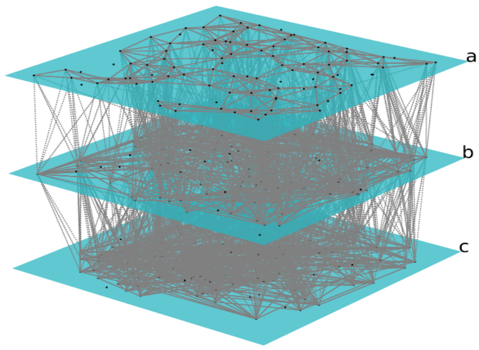
Cascading Failure in Interdependent Network



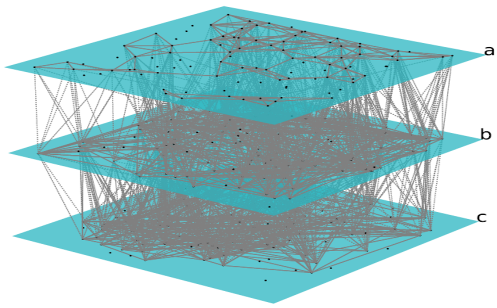
(a) Step: 01



(b) Step: 02

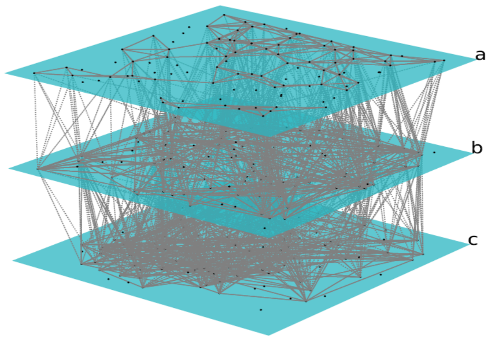


(c) Step: 03

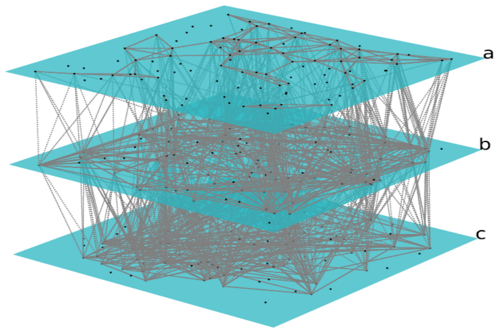


(d) Step: 04

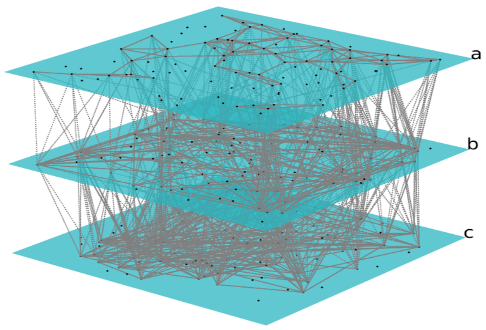
Figure A.1: Example of Cascading Failure in Interdependent Network



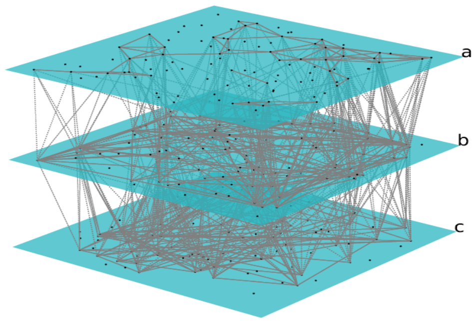
(a) Step: 05



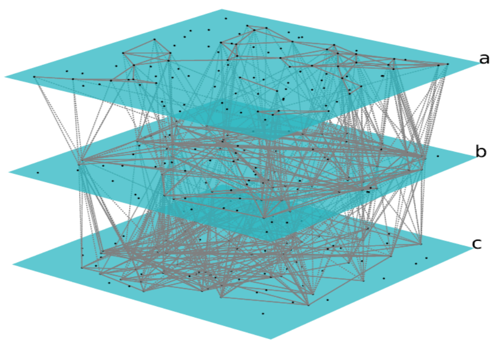
(b) Step: 06



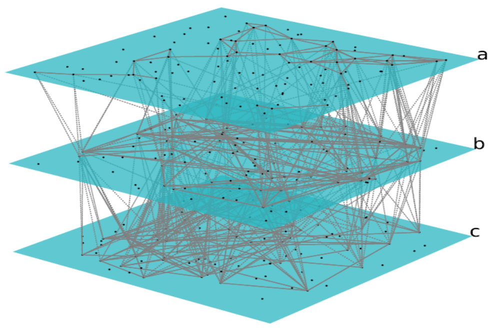
(c) Step: 07



(d) Step: 08



(e) Step: 09



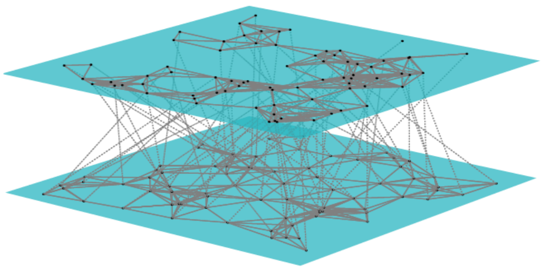
(f) Step: 10

Figure A.2: Example of Cascading Failure in Interdependent Network

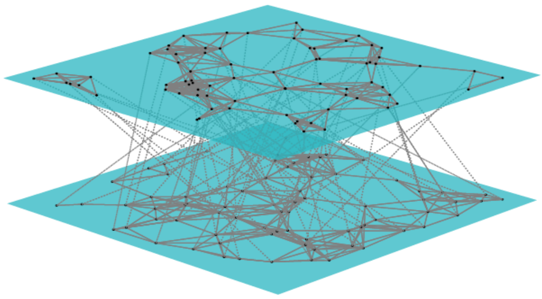
Appendix B

Visualising Four Interdependent Models

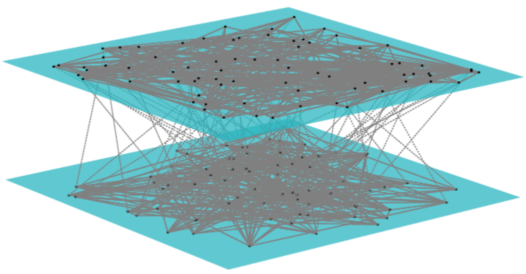
B.1 Strategy 1



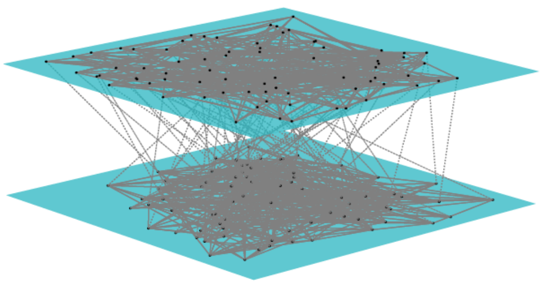
(a) RGG-Spatial network



(b) RGG-Random network



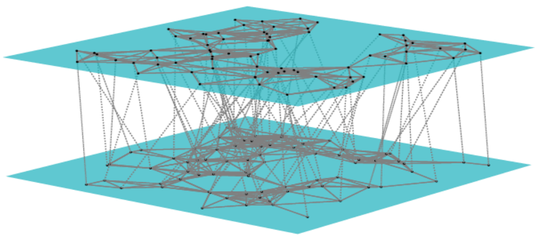
(c) ER-Spatial network



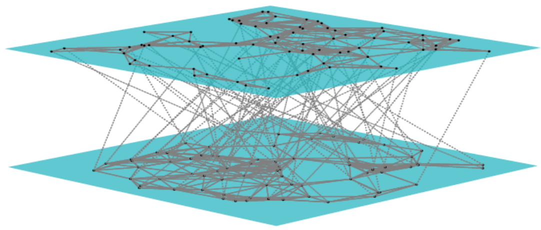
(d) ER-Random network

Figure B.1: $N = 75$, $intra\_T = 0.2$, $inter\_T = 0.7$

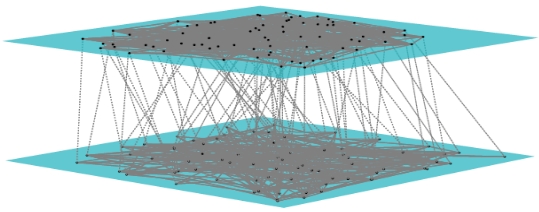
B.2 Strategy 2



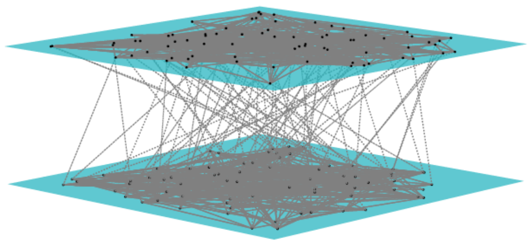
(a) RGG-Spatial network



(b) RGG-Random network



(c) ER-Spatial network



(d) ER-Random network

Figure B.2: $N = 75$, $intra\_T = 0.2$, $inter\_T = 0.35$