Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Robustness on the $R\&D$ alliance network: A multi-layer approach

Semester Project

Jiduan Wu

D-INFK, jiduwu@student.ethz.ch

Supervisors:
Dr. Giacomo Vaccario
Dr. Giona Casiraghi
Prof. Dr. Dr. Frank Schweitzer
Prof. Dr. Siyu Tang

April 1, 2021 - August 31, 2021

Abstract

We delve into an empirical dataset of firms and their cooperators and build a multi-layer firm network from them. Each sector of firms is represented with a layer of the network. We study the robustness of this empirical network against removing nodes with different strategies. In different attacking strategies, we remove nodes of the attacked layer in the order of increasing or decreasing coreness values in the k-core decomposition algorithm. We extend our cascading model from the two-layer scenario to the ten-layer empirical network. In the two-layer scenario, we study the relationship between sparsity and the removal of nodes. We study the nestedness of the network utilizing its correlation with the modularity property. We study the influence of the multi-layer structure by differentiating inter-layer connections and intra-layer connections in the cascading process and make comparisons with the unipartite setting. Our discovery on the fragile periphery, i.e., the nodes that are not in the largest connected component (LCC) of the network reveal the uncommon structure of our network, and it also gives a reminder to other researchers on the usage of the LCC approximation and the defense of their economical system.

Keywords: multi-layer network, empirical study, firm network, cascading model, periphery, sparsity

Contents

Chapter 1

Introduction

1.1 Motivation

Understanding the robustness of a network against removing nodes involves analyzing its topology and plays a significant role in improving its resilience. Most researches focus on robustness analysis in unipartite graphs such as [1]. Exceptions for the two-layer network in [2] reveal how differentiating nodes in different layers and the multi-layer structure influences the robustness of the network. Here the multi-layer structure refers to inter-layer connections and intra-layer connections. In this project, we extend this question to the multi-layer scenario: "What about the robustness of multi-layer networks with three or more layers?". We study the influence of the multi-layer structure by differentiating inter-layer connections and intra-layer connections in the cascading process and make comparisons with the unipartite setting.

Our empirical dataset contains various information of firms and alliances they participate. Firms create alliances together in order to share capabilities in the costly process of knowledge production. Each firm has its capabilities that depend not only on the sector where it operates but also some other features ranging from its human capital to its geographical location. We conduct an empirical study on the empirical dataset by building a multi-layer firm network. The intra-sector and inter-sector cooperation are captured by the intra-layer and inter-layer connections in our multi-layer empirical network. Under these settings, we try to gain more insights into questions such as "Will more layers add to the robustness of our network or boost the process of spreading cascades?", "What are the special properties of our network and why?" and so on.

Our main contributions are: 1) We analyze the statistics of the empirical network and summarize its characterizing properties. 2) We reproduce the results for the synthetic two-layer network in [2]. 3) We quantify the robustness of the multi-layer network under different attacking strategies. 4) We explain the relationship between the sparsity of the network (filling rate of the adjacency matrix) and the node removal ("drops" in the simulation curves). 5) We extend the cascading model to the multi-layer scenario and observe interesting differences between the simulation results of the whole network and LCC, and we give our explanation: the "fragile periphery".

This report is divided into four parts, first, we discuss the dataset and how we build the firm network in the first section, then we present how we adapt the model for the two-layer scenario to the cascading model for the multi-layer network in the second section. In the third section, we present the results and interesting observations along with our explanations. The third section is divided into two parts: the results for the synthetic two-layer network and the empirical multi-layer network. For the multi-layer network case, we adopt two settings when simulating the cascades: unipartite setting and multi-layer setting. In the last section, we summarize our findings, point out their limitations, and clarify their meaning for future researches.

1.2 Empirical datasets

SDC Database: SDC databases offer over 200 data elements such as the names, SIC codes, and nationality of participants, the term of the deal, deal synopsis, and more for each alliance agreement. SIC codes consist of 11 main divisions (including unclassifiable ones), see figure A.1 from [3] in the appendix for an illustration.

SDC tracks a very wide range of agreement types, including joint ventures, strategic alliances, research and development agreements, sales and marketing agreements, manufacturing agreements, supply agreements, and licensing and distribution pacts. SDC database covers the widest range of sectors. Most of the time the coding is highly accurate as [4] claims. Our empirical dataset of firm alliances is just one slice of the SDC database. In our empirical dataset, we have 14535 valid firms, and each firm belongs to a sector. There are 10 sectors such as "services", "manufacturing", "public administration", and so on. There's another widely used codes for firm classification - NAICS (North American Industry Classification System), see figure A.2 from [3] in the appendix for an illustration.

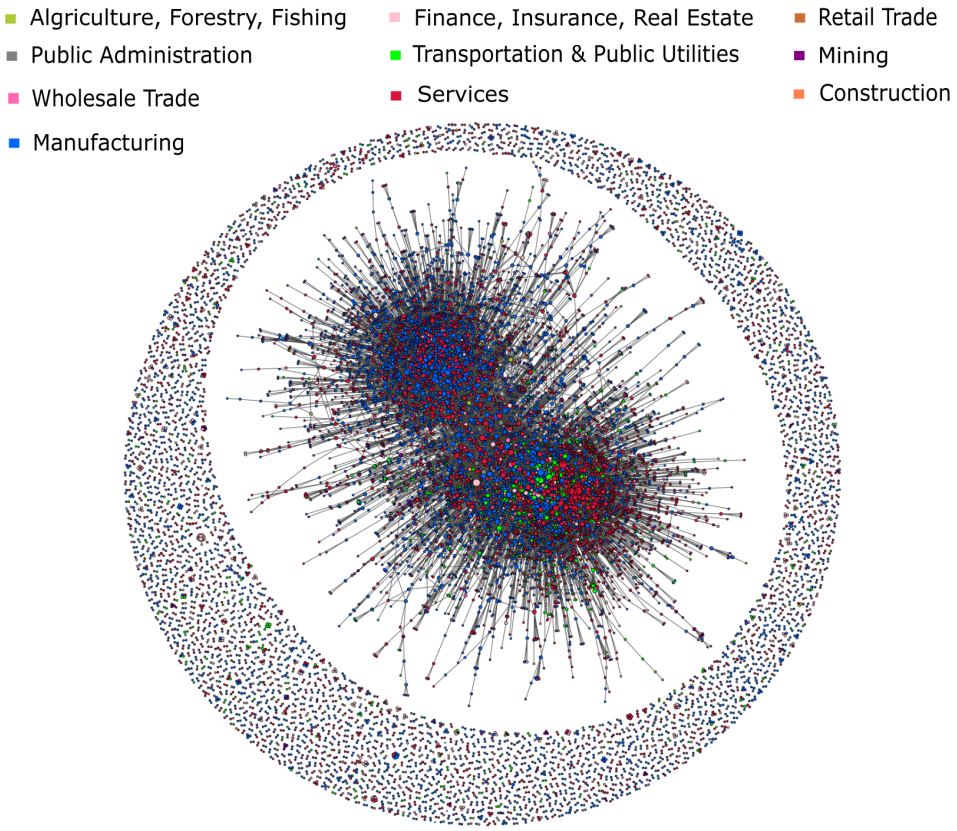

Figure 1.1: Firm network

Each firm can participate in several alliances as illustrated, and alliance size can be any positive number: in figure 1.2, MICROSOFT company participate in both alliance 1 and alliance 2, and alliance 1 has a size of two while alliance 2 has a size of three. The complete alliance size distribution is visualized in figure 1.3, we can observe that the distribution is right-skewed and the majority of firms have size two. And we assume a link exists between any pair of firms in each alliance.

1.3 Firm network

In our firm network, we use a node to represent a firm and a sector of firms is represented by a layer in the network. We only consider the 10 sectors in table 1.1. The sizes of layers are listed in table 1.1. An undirected and unweighted edge exists between two firm nodes if they share common alliance(s) since the cooperation is mutual. We ignore the durations of the alliances and treat our empirical network as a *static network*. The *supra-adjacency matrix* is a distinct tool for the representation of a multi-layer network. We define the *periphery* as the nodes that are not in the largest connected component (LCC) of the network.

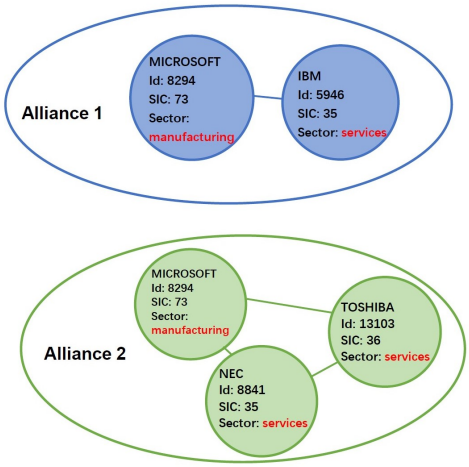

Figure 1.2: Dataset snapshot

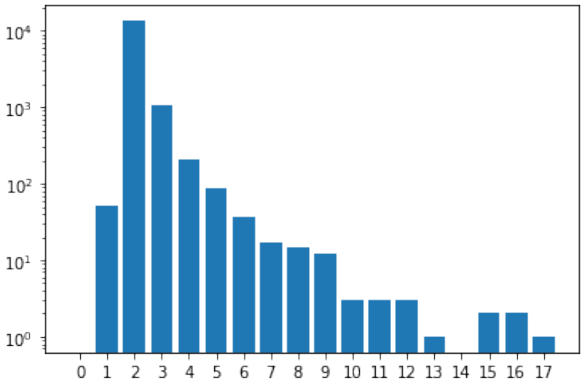

Figure 1.3: Alliance size distribution

To interpret our figures clearly, we name the sectors with their real names instead of indices and sort them by their sizes from the largest to the smallest (see table 1.1). In some figures in this report, we use the percentage to represent the layers instead of their real names, see figures 3.24, 3.25.

Table 1.1: Sector labels and sector sizes

Name	Size	Size percentage	Size in LCC
Manufacturing	7344	0.5053	4576
Services	4508	0.3101	2679
Transportation & Public Utilities	813	0.0559	461
Finance, Insurance, Real Estate	795	0.0547	359
Wholesale Trade	371	0.0255	180
Mining	256	0.0176	95
Public Administration	195	0.0134	110
Retail Trade	103	0.0071	41
Agriculture, Forestry, Fishing	76	0.0052	35
Construction	74	0.0051	29

We visualize the whole network in figure 1.1, we can see a core consists of two major parts and a strip of periphery floating around. The periphery consists of 41.07% of nodes.

Preliminary statistics: There are 14881 alliances in the empirical dataset. The maximum, median, and minimum of an alliance size are 17, 2, and 1. The distribution of the alliance sizes is

right-skewed, see figure 1.3. We don't consider *self-loops* in our network [1] and only a single edge can exist between two firms[2] while there might be several links between two firms. We have 14535 valid firms, 21543 edges in the network. The inter-layer edges distribution is visualized in figure 1.4. From the heat map, we can see that both the Manufacturing layer and Services layer have lots of interactions with each other and within themselves.

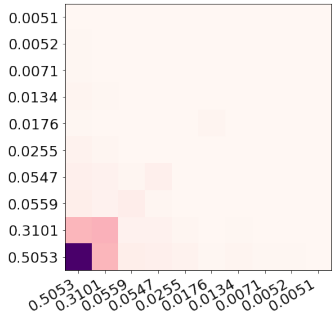

Figure 1.4: Heat Map of Inter-layer Links

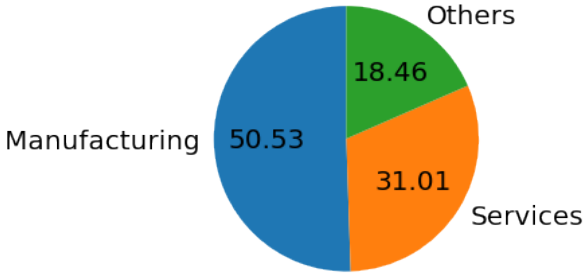

Figure 1.5: Node-layer distribution

Table 1.2: Statistics of firm connections

	Minimum	Median	Maximum	Mean
number of alliances per firm	1	1	284	2.2017
number of links per firm	0	1	421	2.9700

Table 1.3: Statistics of firm node degree

	Minimum	Median	Maximum	Mean
node degree	0	1	421	2.9643
number of alliances per firm	1	1	284	2.2017

The degree distribution of firm nodes, see table 1.3[3]. We can also observe that most of the firms participate in only one alliance, and a small part of active firms contribute to most of the alliance number.

Considering the heterogeneity of sector sizes, we look into the Manufacturing layer and Services layer, which have considerable sizes. We try to learn the structure within them and the reason why they have intense interactions with each other and also within themselves.

For the Services sector (SIC code: 79-89) and the Manufacturing sector (SIC code: 20-39), we apply finer sic codes to see the firm distribution, see figure 1.6, 1.7. We know that two subsectors (73: Business Services and 87: Engineering, Accounting, Research, Management, and Related Services) almost contribute to all the alliances involving the service sector. The manufacturing sector is more evenly distributed compared to the service sector, where a leading subsector (28: Chemicals and Allied Products) exists. See table A.1 from [5] in the appendix for the complete reference. We also plot the alliance distributions in figures 1.8, 1.9 where y values are the number of alliances

[1] Alliances of one firm can happen when the alliance is made up by subsidiaries of one big company

[2] We clarify in advance that the multi-edge network produces the same results as the corresponding single-edge network under our settings.

[3] It is worth mentioning here that the reason why the minimum number of links per firm is 0 is that we don't consider 1-alliances or self-loops in our networks.

that contains participants in the corresponding subsector. We can see the alliance distributions have perfect correspondence to the firm distributions both in the Service sector and Manufacturing sector.

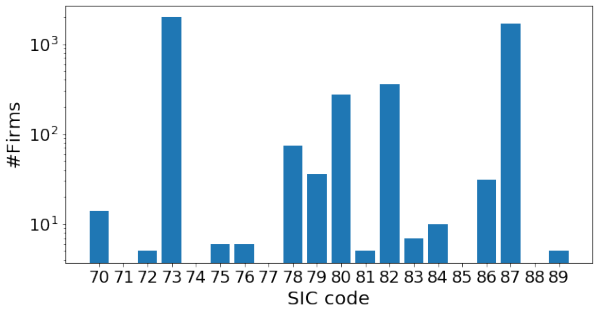

Figure 1.6: Firm distribution of the Service sector

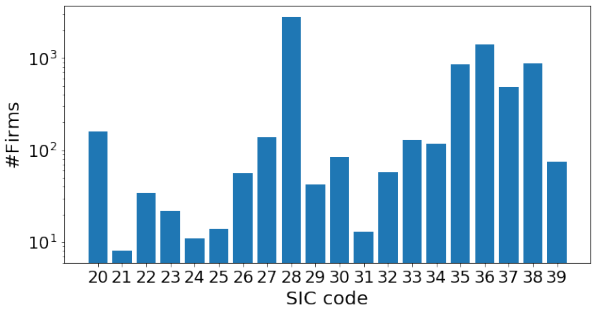

Figure 1.7: Firm distribution of the Manufacturing sector

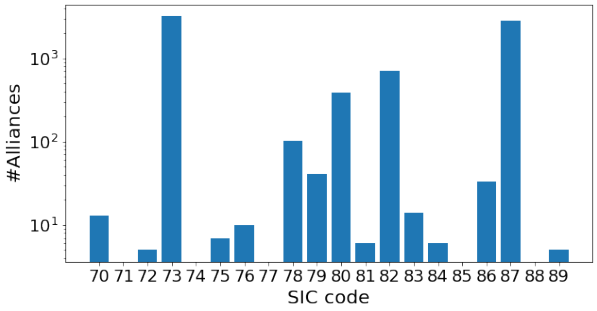

Figure 1.8: Alliance distribution of the Service sector

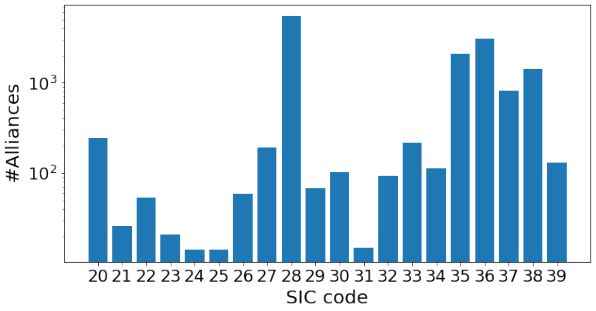

Figure 1.9: Alliance distribution of the Manufacturing sector

Characterising properties: We summarize the important properties of our network as follows

- **Large-scale**: We are studying a large network with over fourteen-thousand nodes while normal networks only have less than a few thousand of nodes.
- **Sparsity**: Firstly, we define the filling rate of a matrix M as follows:

$$\text{Filling rate}(M) = \frac{\#\text{non-zero entries}}{\#\text{rows} \times \#\text{columns}} \tag{1.1}$$

  The filling rate of our empirical network is less than two ten thousandths.

- **Dominant layers**: As figure 1.1 suggests, there are two main parts in the core of the network. They are the Services layer and Manufacturing layer. In the whole network, the fraction of nodes in these two layers is around 81.54%. In the LCC, the fraction of nodes in these two layers is 84.71% and the fraction of edges involving nodes in these two layers is 94.15%. Readers can better observe their dominance in the pie plot 1.5. These numbers show the possibility of approximating the LCC or even the whole network with these two layers.

Chapter 2

Method

2.1 Cascading Model

With the cascading model inspired by [2], the cascades are triggered by removing nodes of the attacked layer with three strategies: remove nodes randomly (random), with increasing coreness values (min), and with decreasing coreness values (max). The coreness values are assigned by applying the *k-core decomposition* algorithm in [6] to the attacked layer. Coreness values describe the centrality of nodes in the network: the higher the coreness value is, the more centrally the node locates, the better the node is integrated into the network.

We also visualize and summarize basic statistics of the coreness distribution of the services layer and manufacturing layer in figures 2.2, 2.3 and table 2.1. The top 10 firms with the highest coreness values in the services layer and manufacturing layer are listed in table 2.2 (some firms with the same coreness values are omitted here due to the limited space.)

Table 2.1: Coreness value distribution

	min	median	max	var	mean
Services layer	0	1	15	0.8418	1.5746
Manufacturing layer	0	1	14	2.2837	1.2990

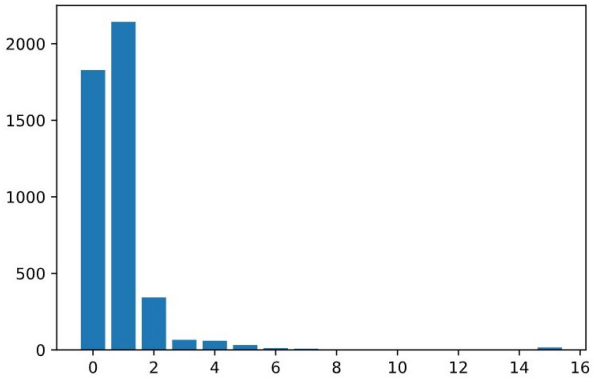

Figure (2.2) Services layer

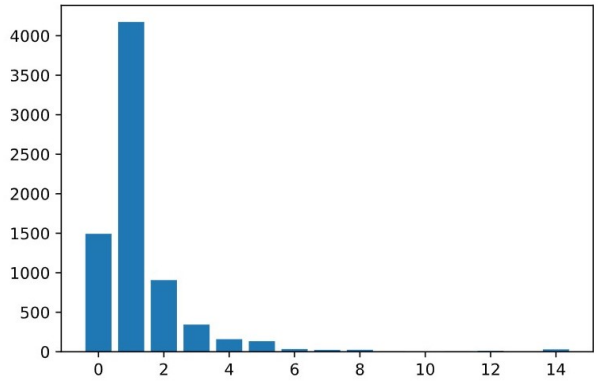

Figure (2.3) Manufacturing layer

Figure 2.3: Coreness value distribution in Services layer and Manufacturing layer

In the cascading model in [2], synthetic two-layer networks are considered. It is assumed that a node is removed when disconnected from the attacked layer. Researchers apply different coupling

Table 2.2: Firms with the highest coreness values

	Services layer	**Manufacturing layer**
1	TELERATE	MATSUSHITA ELECTRIC IND
2	IBM	MITSUBISHI ELECTRIC
3	MICROSOFT	MOTOROLA
4	NOVELL	MTI TECH
5	AISCORP	NAT SEMICONDUCTOR
6	ARBORTEXT	NEC
7	AVALANCHE DEV	NORTH TELECOM
8	COMPUTER TASK	OKI ELECTRIC IND
9	DATABASE PUBLISHING SYS	PHILIPS ELECTRONICS
10	EBT	QLOGIC
11	INFORMATION DESIGN	QUANTUM
12	INFORMATION DIMENSIONS	SANYO ELECTRIC
13	INTERGRAPH	SEAGATE SOFTWARE
14	INTERLEAF	SONY
15	OBJECT DESIGN	SUN MICROSYSTEMS
16	OFFICESMITH CTMG	TEXAS INSTR
17	OPEN TEXT	TOSHIBA
18	ORACLE SYS	UNISYS
19	SOFTQUAD	WESTERN DIGITAL
20	XSOFT	FULCRUM

strategies, i.e., inter-layer connections of two layers to determine the degree of nestedness [1] of the network. Figure 2.11 shows an example of *perfect nestedness*. Nodes that interact with a large number of nodes of the other layer are called *generalists* (G), while nodes that interact only with a small number of nodes of the other layer are called *specialists*. Thus the perfect nestedness is also called *"generalists to generalists" coupling*.

In the multi-layer scenario, we assume a node is removed when disconnected with all the other layers.

2.1.1 Replications

In this section, we compare the simulation results produced by our model and the codes provided by the author of [2] to validate our model. The results comparison are shown in figures 2.11, 2.15, 2.19.

The incidence matrices of the two-layer network specify different coupling strategies, the rows represent the layer 1 and columns represent layer 2. The yellow part is the 1s in the incidence matrix. In [2], the core-periphery topology is used within each layer. CGCG means the core nodes from layer 1 that are generalists are coupled to core nodes from layer 2 that are also generalists, see figure 2.5. CGCS means that nodes from layer 1 that are generalists are coupled to core nodes from layer 2 that are specialists, see figure 2.6.

Because of the symmetry, all the cascades are triggered from layer 1. Results produced by our model and original codes are shown below, see figure 2.11. We also reproduce the results in the case where the filling rate is 0.015, see figures 2.15, 2.19. The similarities between the results validate our model convincingly.

[1] Nestedness is a property that describes the tendency for nodes to interact with subsets of the interaction partners of better-connected nodes, see [7], page 1.

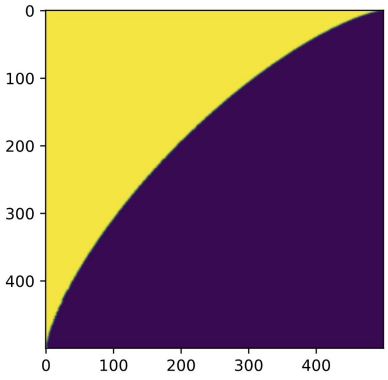

Figure (2.5) CGCG

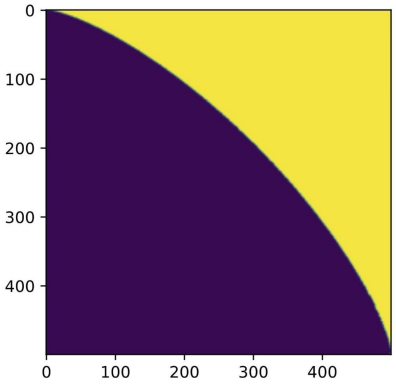

Figure (2.6) CGCS

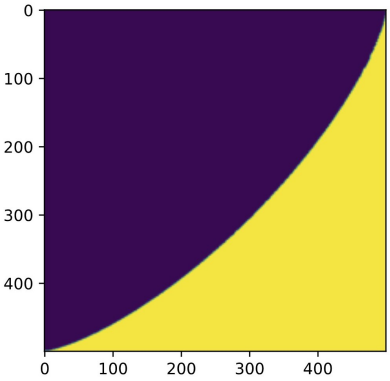

Figure (2.7) CSCS

Figure 2.7: Visualization of a 500 × 500 perfectly nested matrix

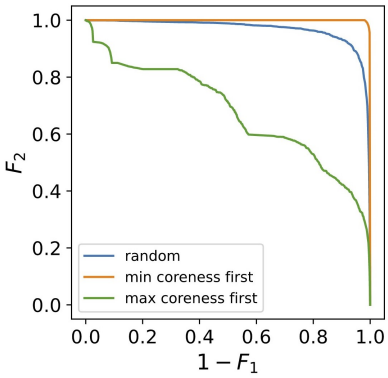

Figure (2.9) Reproduction of CGCG

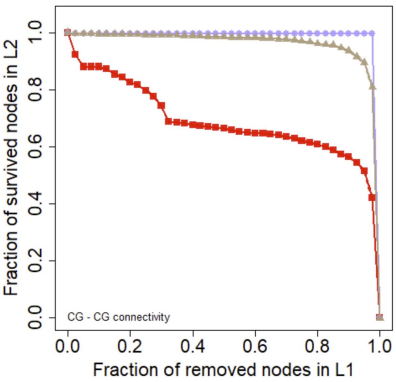

Figure (2.10) Original CGCG

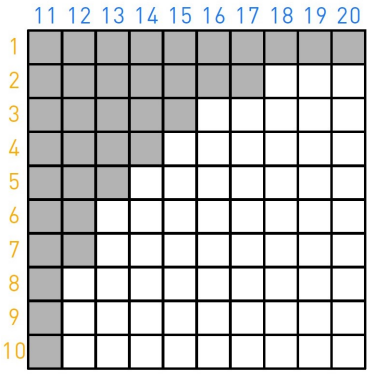

Figure (2.11) Perfect nestedness

Figure 2.11: Model validation (filling rate= 0.35)

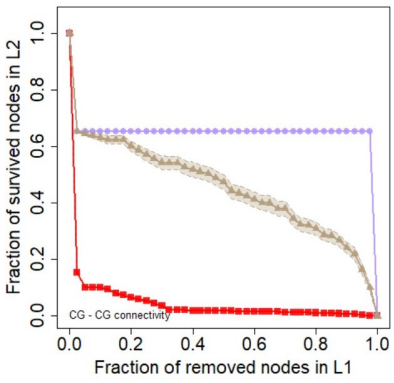

Figure (2.13) Original CGCG

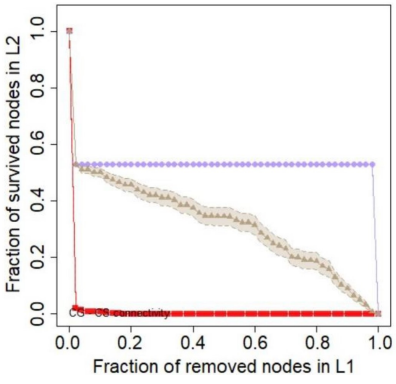

Figure (2.14) Original CGCS

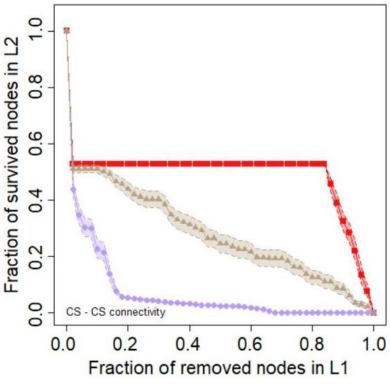

Figure (2.15) Original CSCS

Figure 2.15: Results produced by original codes (filling rate= 0.015)

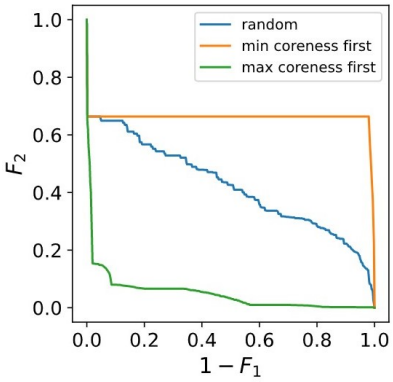

Figure (2.17) Reproduction of CGCG

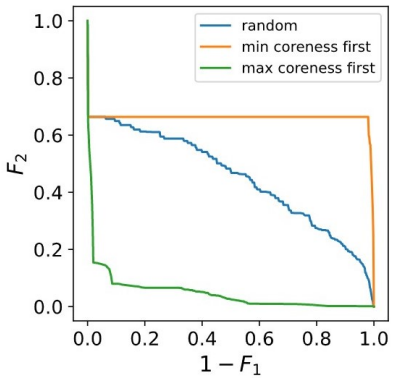

Figure (2.18) Reproduction of CGCS

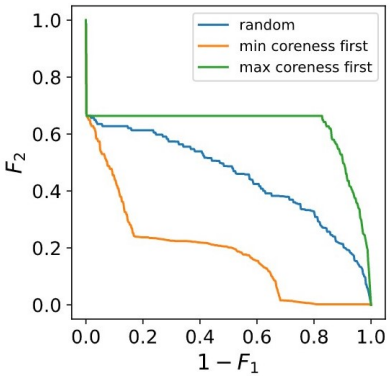

Figure (2.19) Reproduction of CSCS

Figure 2.19: Results produced by our model (filling rate= 0.015)

2.1.2 Nestedness analysis for large-scale and sparse network

Inspired by the basic statistics, we use two dominant layers, the Manufacturing layer and the Services layer, to approximate the LCC or the whole network. And even for the two-layer approximation of Services layer and Manufacturing layer (over ten thousand nodes), popular algorithms for nestedness analysis such as NODF, MT, and BR are computationally inefficient as suggested in [7], [8].

Hence we approach this challenge in another way: we utilize the correlation of nestedness and other properties, e.g., *dissortativity, core-periphery metric, and modularity*. Here we choose the modularity to explain in detail.

It is found in [7] that low-density networks exhibit a positive correlation between nestedness and modularity, while high-density networks exhibit a negative correlation between them. We computed the z-score, -16.67 with 50 null models generated by Bascompte's algorithm. The probability to reject the hypothesis that our empirical bigraph consists of the Manufacturing layer and Services layer has modularity is higher than 99.99%. The conclusion we can draw here is our empirical two-layer approximation has low modularity w.h.p, and this probably (w.h.p) means it has low nestedness w.h.p.

Chapter 3

Results

In this section, we first approximate the whole network with two dominant layers, the Services layer and the Manufacturing layer, and simulate cascades on the two-layer network and its LCC. Then we simulate cascades on the ten-layer network with two cascading settings: unipartite setting and multi-layer setting. In the unipartite setting, a node is removed when it's disconnected from any other nodes, and the whole network is treated as a unipartite graph. In the multi-layer setting, a node is removed when it's disconnected from all the other layers, which utilizes the multi-layer structure[1].

3.1 Two-layer approximation

3.1.1 The importance of filling rates

When we simulate cascades on the empirical network, we observe big "drops" at the beginning, which is not seen in figure 2.11 but figure 2.15. We recall that the filling rate of the incidence matrix of our empirical network is 0.000326, and in figure 2.15 the filling rate is around 0.015, while the filling rate of the incidence matrix of the baseline synthetic network is 0.35.

This phenomenon encouraged us to delve into the relationship between filling rates and "drop" sizes. The relation between the filling rates and the "drops" is visualized in figure 3.3. Figure 3.3 shows that "drops" are insensitive to the attacking strategies and coupling strategies. And when the filling rate is approximately larger than 0.35, the drop converges to 0, which explains why the "drops" were not observed in [2].

Now we can draw the conclusions that "drops" are caused by the *low filling rates* and the assumption that nodes get removed once disconnected with the attacked layer. This finding reminds us to differentiate these "drops" from standard cascading processes.

3.2 Cascades in multi-layer network

3.2.1 Cascades under "unipartite" setting & "multi-layer" setting

The cascades are triggered by removing nodes in the attacked layer until no nodes are left in the attacked layer. Under each attacked layer and attacking strategy, we compute the mean of 50 simulations. We apply two different ways of visualization. From now on, we use N_i to denote the number of remaining nodes in layer i, and $N_i^{(0)}$ to denote the initial size of layer i. N_s and $N_s^{(0)}$

[1] All the results in this project can be reproduced with the codes in [9]

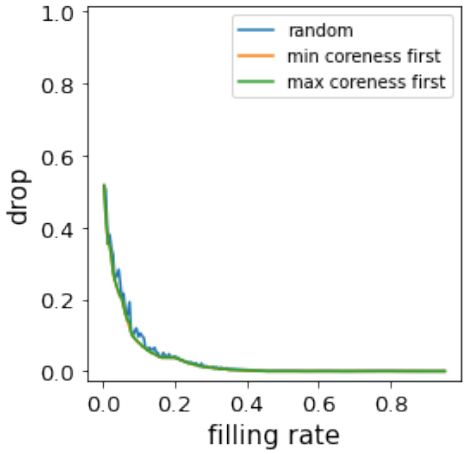

Figure (3.2) Filling rate - "drop" (CGCG)

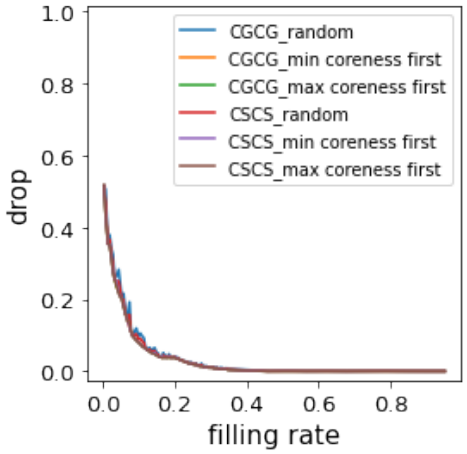

Figure (3.3) Filling rate - "drop" (CGCG and CSCS)

Figure 3.3: Filling rates and "drops"

denote the temporary size and the initial size of the starting layer. N_{LCC} represents the size of the LCC.

1. $1 - F_1$ and F_2 (see figure 3.7 as an example):

$$F_1 = \frac{N_s}{N_s^{(0)}} \tag{3.1}$$

$$F_2 = \frac{N_i}{N_i^{(0)}} \tag{3.2}$$

2. $1 - F_1$ and the number of removed nodes (see figure 3.11 as an example):

$$F_1 = \frac{N_s}{N_s^{(0)}} \tag{3.3}$$

$$N = log(N_i^{(0)} - N_i) \tag{3.4}$$

Cascades under "unipartite" setting: Under the "unipartite" setting, we treat the cascades on the ten-layer network in the same way as the unipartite networks: nodes are removed when they become "isolated", i.e., disconnected with all the other nodes. Here we represent simulation results when the attacked layer is the Manufacturing layer, see figures 3.7, 3.11. Refer to figures A.6, A.10 for the results of the Services layer.

We also do the simulations on the LCC of the whole network, see figure 3.15, A.10. One difference between the simulation results on the whole network and the LCC of the whole network is that there is a section of slope at the beginning in the whole network under min attacks case, while these steeper slopes are not observed in the LCC case.

Cascades under "multi-layer" setting: "Multi-layer" setting assumes a node is removed when it becomes isolated, i.e., disconnected with all the other layers, see figures 3.19, 3.23, A.14, A.18.

Summary and comparison: Figures 3.15, 3.7 3.19, 3.23 reveal three important pieces of information.

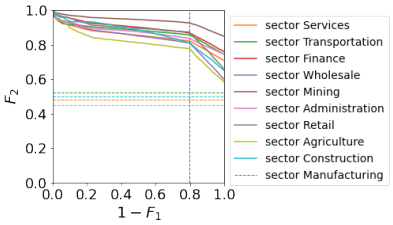

Figure (3.5) Manufacturing layer under max attacks

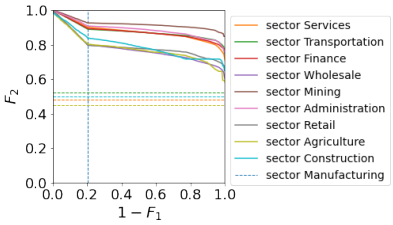

Figure (3.6) Manufacturing layer under min attacks

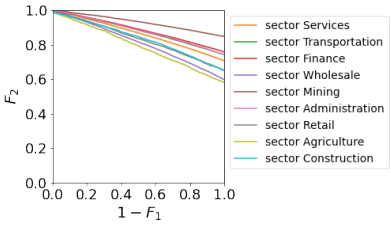

Figure (3.7) Manufacturing layer under min attacks

Figure 3.7: Attacked layer: Manufacturing layer (whole network, unipartite setting)

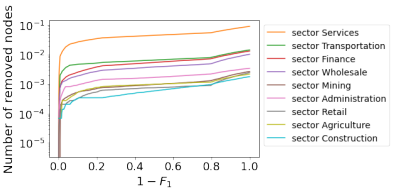

Figure (3.9) $(1 - F_1)$ - node num max attacks

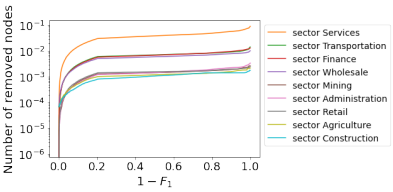

Figure (3.10) $(1 - F_1)$ - node num min attacks

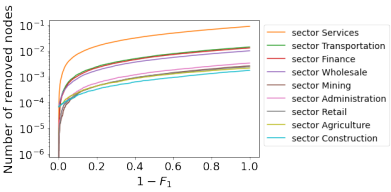

Figure (3.11) $(1 - F_1)$ - node num random attacks

Figure 3.11: Attacked layer: Manufacturing layer (whole network, unipartite setting)

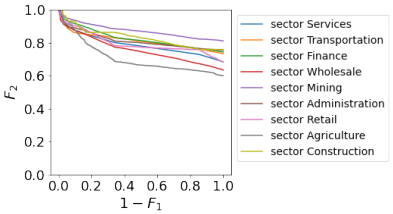

Figure (3.13) Manufacturing layer under max attacks

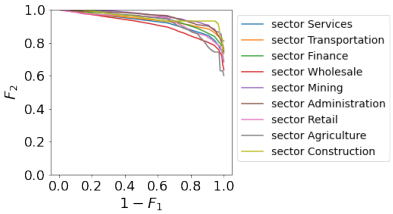

Figure (3.14) Manufacturing layer under min attacks

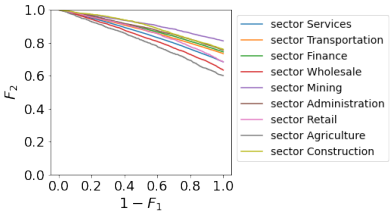

Figure (3.15) Manufacturing layer under random attacks

Figure 3.15: Attacked layer: Manufacturing layer (LCC, unipartite setting)

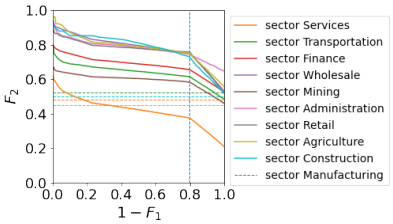

Figure (3.17) Manufacturing layer under max attacks

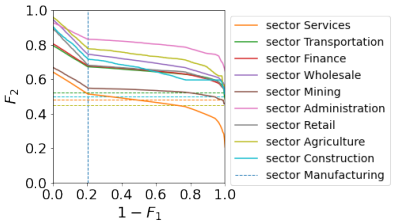

Figure (3.18) Manufacturing layer under min attacks

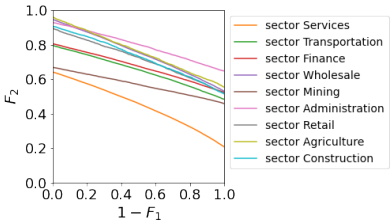

Figure (3.19) Manufacturing layer under random attacks

Figure 3.19: Attacked layer: Manufacturing layer (whole network, multi-layer setting)

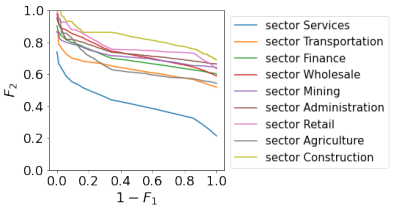

Figure (3.21) Manufacturing layer under max attacks

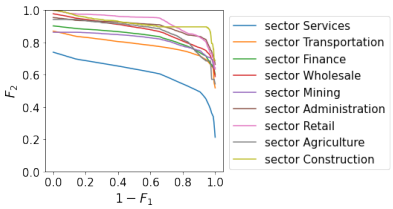

Figure (3.22) Manufacturing layer under min attacks

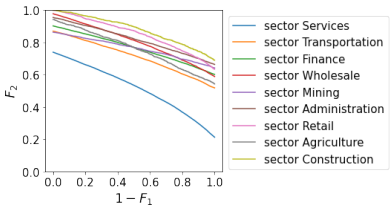

Figure (3.23) Manufacturing layer under random attacks

Figure 3.23: Attacked layer: Manufacturing layer (LCC, multi-layer setting)

1. The "drops" at the beginning are observed under the multi-layer setting. It is due to the fact that many nodes don't have inter-layer connections in our sparse network with the filling rate of the adjacency matrix be 0.000188.

2. The network shows robustness against attacks, even attacking the largest layer, which contains half of the nodes doesn't cause disastrous cascades to other layers. Our network behaves more robustly under the unipartite setting judging from the numbers of final remaining nodes in layers.

3. Under both settings, linear faster decays are observed at the beginning of cascading curves of the whole network under min attacks while LCC case not. In other words, the whole network is more fragile when we first attack nodes with low coreness, this difference leads us to the question that "why is the periphery important when we analyze the robustness of our network?"

3.2.2 Why is the periphery important?

In this section we only consider the multi-layer setting. As its name suggests, a considerable part of nodes with low coreness in each layer are in the periphery. In the Manufacturing layer, 55.46% of nodes with coreness 0 are in the periphery, and in the Services layer, 42.07% of nodes with coreness 0 are in the periphery. For coreness value distribution in the Manufacturing and Services layer, see table 3.1. In our model, the nodes with the same coreness values have the same priorities to be removed. Hence it is reasonable to make assumptions that the turning points of the slopes correspond to variations of coresness values.

Table 3.1: Coreness value fractions

Sector \ Coreness	0	1	2	3	4
Manufacturing	0.2033	0.5684	0.1235	0.0468	0.0217
Services	0.4055	0.4754	0.0146	0.0133	0.0071

Moreover, after we looked back at the original dataset and delved into the periphery, we find that the periphery is well-mixed. There are 3161 alliances in the periphery of the network in total. The max size of the alliance is 11 and the min size is 1. The median of the alliance sizes is 2, and 92.3% of alliances have the size of 2. Furthermore, in these 2-alliances, 54.8% of them are consist of two firms from the same sectors, and 45.2% are consist of two firms from two different sectors. This well-mixed periphery explains the linear faster decay we find in the whole network under min attacks.

To further validate our explanation, we plot the dotted vertical lines representing the fractions of nodes with coreness 0, and the horizontal lines representing the fractions of disconnected nodes in figures 3.19, A.14. The rigorous definitions are shown in equations 3.5, 3.6. The vertical lines match perfectly with the turning points of the curves. As for the horizontal lines, a unique color is assigned to each layer, and most of the curves are close to the corresponding horizontal lines, which shows a limited spreading of the cascades, in other words, the robustness of the empirical network.

$$x_i = \frac{|\{n \in layer_i | coreness_{subgraph_i}(n) == 0\}|}{|\{n \in layer_i\}|} \tag{3.5}$$

$$y_i^j = 1 - \frac{|\{n \in layer_j | n \in neighbor(layer_i)\}|}{|\{n \in layer_j\}|} \tag{3.6}$$

3.2.3 Robustness analysis

In this section, we plot two sets of figures to summarize the robustness analysis of the network.

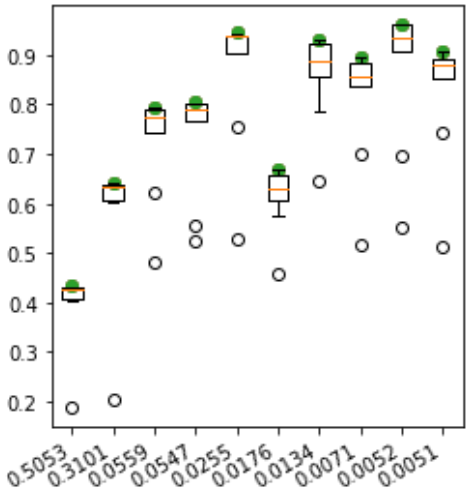

Figure 3.24: Final fractions of remaining nodes

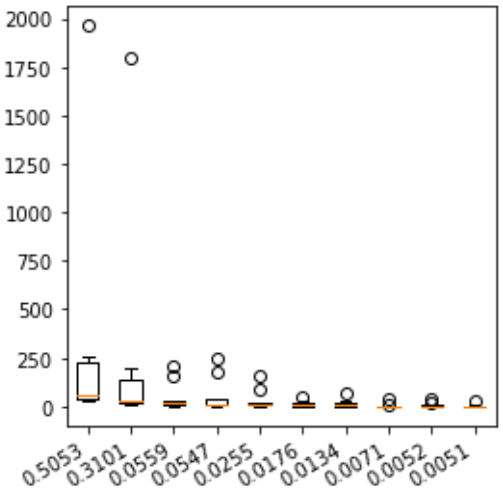

Figure 3.25: Damages caused by different layers

Final remaining fractions of remaining nodes: We plot the final survival fractions of each layer when attacks start from the other nine layers in a box plot, see figure 3.24. In the x-axis, we use layer proportions to represent each layer, e.g. 0.5053 represents the Manufacturing layer. The y-axis represents the remaining fractions of x when the attacked layer is one of the other nine layers. The green scatters are the real initial fractions excluding "drops" at the beginning.

We can observe that removing any layer won't damage the majority of the remaining parts of other layers, which further displays the robustness of our network. The outliers are all caused by the manufacturing layer and services layer, two dominant layers.

Damages caused by the attacked layer: We also show the effects of removing each layer through remaining nodes fractions of other layers, see figure 3.25. In x-axis, we still use layer proportions to represent each layer, and y-axis is the number of removed nodes in the other nine layers excluding the "drops".

More outliers and decreasing interquartile ranges are observed because small layers are more unevenly connected with other layers, which means they can cause considerable damages to some layers while other layers stay uninfluenced.

Chapter 4

Conclusions

This project mainly contributes to two parts. The first part contains some insights into our empirical dataset. Firstly, we observe a considerable part of the network remains under attacks when we consider all the ten layers instead of using the two-layer approximation. Moreover, the ten-layer network with the multi-layer setting has a similar number of removed nodes to the results with the unipartite setting during the cascading process (initial "drops" excluded). This shows the multi-layer structure adds to the robustness of our network instead of spreading more cascades and causing an overall network failure. Inter-layer connections protect some nodes against attacks, and at the same time, inter-layer connections are not too dense thus they constrain the propagation of cascades and prevent an overall cascade. Differentiating inter-layer connections and intra-layer connections provides insights into the influence of multi-industry cooperation on the robustness of our network. Secondly, we generalize the cascading model of the two-layer network to the multi-layer scenario and observe interesting differences between the simulation results of the whole network and LCC. We show the "well-mixed" structure of the periphery makes our network more fragile under the min coreness first attacks. This reminds us that if we protect the periphery from the beginning, then some instant huge losses can be avoided.

From the perspective of methodologies, we find out that "drops" at the beginning can be caused by the sparsity of the network. Our results show these "drops" cease to appear when the filling rate of the incidence matrix is larger than 0.35. The empirical bigraphs can be much more complicated, but this finding still calls attention to the sparsity of networks. These drops should be differentiated from the standard cascading processes. Also, we point out that sometimes focusing on the LCC of the network is not enough. Though LCC contains significant "messages" of the network, the periphery of networks can expose unexpected fragilities and weaknesses. Hence researchers need to make sure the important information is maintained when we choose to use the LCC to approximate the whole network. Making comparisons between the whole network scenario and LCC scenario might provide a brand new perspective to studied problems like our project.

Last but not least, we point out the limitations of our findings and results. First of all, we ignore the durations of our alliances, integrate the data altogether, and treat the network as a static network. Through this pre-processing, we abandon all the dynamical properties of our empirical network. Secondly, we analyze the nestedness and modularity of our network but we don't fully utilize these properties to help us better understand our simulation results. Most of our analysis still relies on the statistical characteristic of our dataset. Thirdly, triggering the cascades by removing nodes of the attacked layer considering coreness values is the most natural and reasonable choice, here we still list some other measures describing centralities such as degrees or the ranking value produced by the NODF algorithm in [2], [10] that can be adopted. The last point, we try to use our results to show interesting possibilities that are not often observed instead of giving theorems in a definite way.

Bibliography

[1] G. Casiraghi and F. Schweitzer, "Improving the robustness of online social networks: A simulation approach of network interventions," *Frontiers in Robotics and AI*, vol. 7, p. 57, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2020.00057

[2] G. Casiraghi, A. Garas, and F. Schweitzer, "Probing the robustness of nested multi-layer networks," 2019.

[3] "Wikipedia, standard industrial classification." [Online]. Available: https://en.wikipedia.org/wiki/Standard_Industrial_Classification

[4] M. Schilling, "Understanding the alliance data," *Strategic Management Journal*, vol. 30, pp. 233 – 260, 03 2009.

[5] "Sic codes lookup - standard industrial classification." [Online]. Available: https://siccode.com/sic-code-lookup-directory

[6] M. AU Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, pp. 888–893, 11 2010. [Online]. Available: https://doi.org/10.1038/nphys1746

[7] M. S. Mariani, Z.-M. Ren, J. Bascompte, and C. J. Tessone, "Nestedness in complex networks: Observation, emergence, and implications," *Physics Reports*, vol. 813, pp. 1–90, 2019, nestedness in complex networks: Observation, emergence, and implications. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037015731930119X

[8] M. Cantor, M. M. Pires, F. M. D. Marquitti, R. L. G. Raimundo, E. Sebastián-González, P. P. Coltri, S. I. Perez, D. R. Barneche, D. Y. C. Brandt, K. Nunes, F. G. Daura-Jorge, S. R. Floeter, and P. R. Guimarães, Jr., "Nestedness across biological scales," *PLOS ONE*, vol. 12, no. 2, pp. 1–22, 02 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0171691

[9] Google Drive Link.

[10] M. Almeida-Neto, P. Guimarães, P. R. Guimarães Jr, R. D. Loyola, and W. Ulrich, "A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement," *Oikos*, vol. 117, no. 8, pp. 1227–1239, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0030-1299.2008.16644.x

Appendix A

Extra tables and figures

Range of SIC Codes	Division
0100-0999	Agriculture, Forestry and Fishing
1000-1499	Mining
1500-1799	Construction
1800-1999	not used
2000-3999	Manufacturing
4000-4999	Transportation, Communications, Electric, Gas and Sanitary service
5000-5199	Wholesale Trade
5200-5999	Retail Trade
6000-6799	Finance, Insurance and Real Estate
7000-8999	Services
9100-9729	Public Administration
9900-9999	Nonclassifiable

Figure A.1: SIC Codes

Sector #	Description
11	Agriculture, Forestry, Fishing and Hunting
21	Mining, Quarrying, and Oil and Gas Extraction
22	Utilities
23	Construction
31-33	Manufacturing
41/42	Wholesale Trade (41 in Canada,[3] 42 in the United States[2])
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information[notes 1]
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
91/92	Public Administration (91 in the United States, 92 in Canada[4])

Figure A.2: NAICS Codes

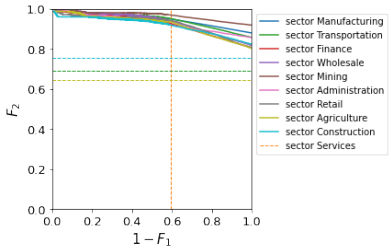

Figure (A.4) Services layer under max attacks

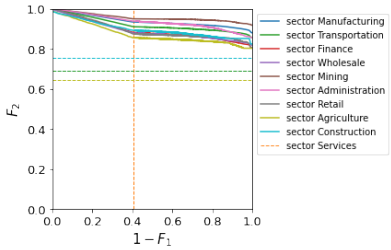

Figure (A.5) Services layer under min attacks

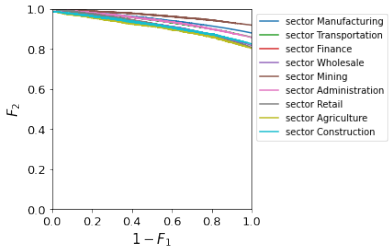

Figure (A.6) Services layer under random attacks

Figure A.6: Attacked layer: Services layer (whole network, unipartite setting)

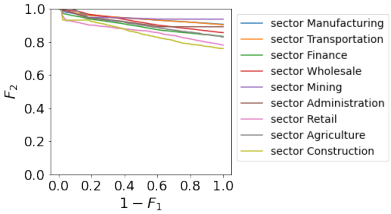

Figure (A.8) Services layer under max attacks

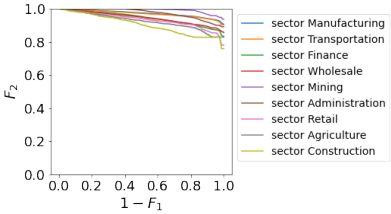

Figure (A.9) Services layer under min attacks

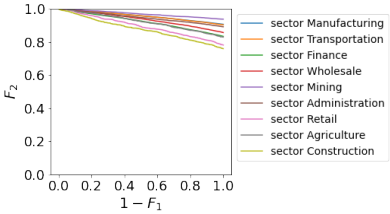

Figure (A.10) Services layer under random attacks

Figure A.10: Attacked layer: Services layer (LCC, unipartite setting)

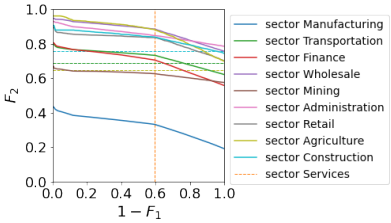

Figure (A.12) Services layer under max attacks

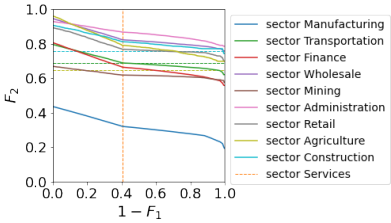

Figure (A.13) Services layer under min attacks

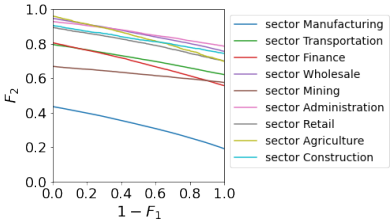

Figure (A.14) Services layer under random attacks

Figure A.14: Attacked layer: Services layer (whole network, multi-layer setting)

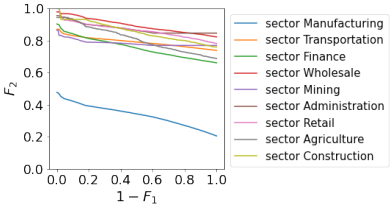

Figure (A.16) Services layer under max attacks

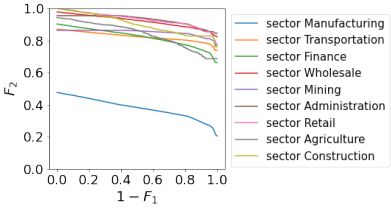

Figure (A.17) Services layer under min attacks

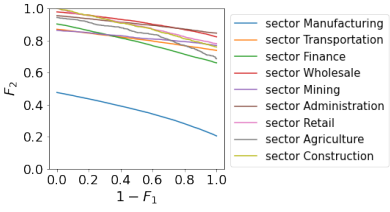

Figure (A.18) Services layer under random attacks

Figure A.18: Attacked layer: Services layer (LCC, multi-layer setting)

Table A.1: Finer SIC Code List

Manufacturing			
Code	Name	Code	Name
20	Food and Kindred Products	21	Tobacco Products
22	Textile Mill Products	23	Apparel and other Finished Products Made from Fabrics and Similar Materials
24	Lumber and Wood Products, except Furniture	25	Furniture and Fixtures
26	Paper and Allied Products	27	Printing, Publishing, and Allied Industries
28	Chemicals and Allied Products	29	Petroleum Refining and Related Industries
30	Rubber and Miscellaneous Plastics Products	32	Stone, Clay, Glass, and Concrete Products
33	Primary Metal Industries	34	Fabricated Metal Products, except Machinery and Transportation Equipment
35	Industrial and Commercial Machinery and Computer Equipment	36	Electronic and other Electrical Equipment and Components, except Computer Equipment
37	Transportation Equipment	38	Measuring, Analyzing, and Controlling Instruments; Photographic, Medical and Optical Goods; Watches and Clocks
39	Miscellaneous Manufacturing Industries		
Services			
70	Hotels, Rooming Houses, Camps, and other Lodging Places	72	Personal Services
73	Business Services	75	Automotive Repair, Services, and Parking
76	Miscellaneous Repair Services	78	Motion Pictures
79	Amusement and Recreation Services	80	Health Services
81	Legal Services	82	Educational Services
83	Social Services	84	Museums, Art Galleries, and Botanical and Zoological Gardens
86	Membership Organizations	87	Engineering, Accounting, Research, Management, and Related Services
88	Private Households	89	Miscellaneous Services

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

ROBUSTNESS AND INTERVENTION ON THE R&D ALLIANCE NETWORK: A MULTI-LAYER APPROACH

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Wu	Jiduan

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zurich, August 26th, 2021	Jiduan Wu

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.