Robustness of Complex Supply Chain Networks to Targeted Attacks

by

María Ximena Rueda Guerrero

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Master of Engineering in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Author ...............................................................

Department of Electrical Engineering and Computer Science

May 29, 2018

Certified by ..........................................................

Matthias Winkenbach
Research Scientist
Thesis Supervisor

Accepted by ..........................................................

Katrina LaCurts
Chair, Master of Engineering Thesis Committee
Robustness of Complex Supply Chain Networks to Targeted Attacks

by

María Ximena Rueda Guerrero

Submitted to the Department of Electrical Engineering and Computer Science on May 29, 2018, in partial fulfillment of the requirements for the degree of Master of Engineering in Computer Science and Engineering

Abstract

In this thesis, we study the robustness of complex supply chain systems from a network science perspective. Through the simulation of targeted attacks to nodes and edges using different hierarchical measures from network science to select the most relevant components, we evaluate the extent to which local centrality measures can estimate the relevance of a node in maintaining the connectivity and the efficient communication across the network. We perform the experiments on two real-world supply chain data sets, and on an ensemble of networks generated from network growth models that share simple topological properties with the real-world networks. It is found that all models produce more robust networks than the data sets of choice. In addition, the removal of high average neighbor degree nodes seems to have little impact on the connectivity of the network, and a highly varying impact on the efficiency of the network. Finally, robustness against targeted node and edge removal is found to be more associated to the number of nodes and links in the network than to more complex network measures such as the degree distribution.

Thesis Supervisor: Matthias Winkenbach
Title: Research Scientist
Acknowledgments

I would like to thank my thesis advisor, Dr. Matthias Winkenbach, for introducing me to the field of applied network science and sharing his expertise in the area of supply chain management. This work would not have been possible without his constant guidance, insightful feedback and support throughout the development.
Contents

1 Introduction .................................................. 13

2 Literature Review ........................................... 21
   2.1 Local and Global Network Metrics ......................... 21
   2.2 Network Growth Models .................................. 26
   2.3 Failure, Attacks and Robustness ......................... 30

3 Methodology .................................................. 35
   3.1 Dataset Processing ....................................... 35
   3.2 Models and Metrics for Data Analysis ...................... 38
   3.3 Networks .................................................. 39
      3.3.1 Real-World Supply Networks ......................... 39
      3.3.2 Erdős-Renyi Networks ................................. 40
      3.3.3 Watts-Strogatz Networks .............................. 41
      3.3.4 Barabási-Albert Networks ............................ 42
      3.3.5 Scale-free Networks with Tunable Clustering .......... 42
      3.3.6 Degree-Locality Based Attachment (DLA) .............. 43
   3.4 Data Visualization ....................................... 44

4 Experimentation and Results ............................... 45
   4.1 Network Topologies ...................................... 46
4.2 Simulation Results for Targeted Node Attacks .................................. 50
  4.2.1 Targeted Attacks by the Degree, Betweenness Centrality and
          Load Centrality ........................................................................ 52
  4.2.2 Targeted Attacks by the Average Neighbor Degree ...................... 53
  4.2.3 Load and Betweenness Centrality .............................................. 54
4.3 Simulation Results for Targeted Edge Attacks ................................. 54
4.4 Monotonic Measures of Network Robustness .................................. 55

5 Conclusion .................................................................................... 61
  5.1 Collective Behavior for Supply Chain Robustness ............................ 62
  5.2 Future Work ................................................................................ 63
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Attitude towards supply chain risk management priority</td>
<td>14</td>
</tr>
<tr>
<td>1-2</td>
<td>Frequency of risks in supply chains</td>
<td>15</td>
</tr>
<tr>
<td>1-3</td>
<td>Methodological framework for robustness modeling from network science</td>
<td>17</td>
</tr>
<tr>
<td>2-1</td>
<td>Effect of clustering and edge range in networks</td>
<td>24</td>
</tr>
<tr>
<td>3-1</td>
<td>Real-world Supply Chain</td>
<td>36</td>
</tr>
<tr>
<td>4-1</td>
<td>comp ensemble of networks degree distribution</td>
<td>48</td>
</tr>
<tr>
<td>4-2</td>
<td>tools ensemble of networks degree distribution</td>
<td>49</td>
</tr>
<tr>
<td>4-3</td>
<td>DLA_comp degree distribution</td>
<td>50</td>
</tr>
<tr>
<td>4-4</td>
<td>comp ensemble of networks robustness to targeted node attacks</td>
<td>51</td>
</tr>
<tr>
<td>4-5</td>
<td>tools’s ensemble of networks robustness to targeted node attacks</td>
<td>52</td>
</tr>
<tr>
<td>4-6</td>
<td>comp’s ensemble of networks robustness to targeted edge attacks</td>
<td>56</td>
</tr>
<tr>
<td>4-7</td>
<td>tools’s ensemble of networks robustness to targeted edge attacks</td>
<td>57</td>
</tr>
<tr>
<td>4-8</td>
<td>Growth of $\frac{N-2}{N-1}$</td>
<td>59</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Recent works in SCN robustness ........................................ 32

3.1 Real-world Supply Chain Stage Classification ...................... 36

3.2 Comparison robustness related network-level metrics of two real-life
supply networks of different product lines. .............................. 39

3.3 Robustness related network-level metrics of computer equipment man-
ufacturer supply network. .................................................... 40

3.4 Robustness related network-level metrics of a power-driven handtools
manufacturer supply network. ................................................ 41

4.1 Global measures for the networks in the ensemble generated for the
comp supply chain. ............................................................... 46

4.2 Global measures for the networks in the ensemble generated for the
tools supply chain............................................................... 46
Chapter 1

Introduction

Supply chains are complex systems that regulate the flow of goods and services, as well as the related information flows and financial transactions. They comprise several processes involving a variety of products between diverse entities such as suppliers, manufacturers, distributors, and retailers. Modern supply chains involve entities that operate individually, but are still largely interdependent. It is rather uncommon for one entity to be able to exercise control over all the components it is dependent upon [27]. As a result, each entity needs to evaluate the risks that others are subject to in order to prepare for potential interruptions in the flow of goods and services.

According to a survey conducted in 2009 with over 1,400 supply chain professionals from 70 countries, there is a general preference to allocating risk management resources towards risk prevention rather than risk mitigation (see Figure 1-1). In addition, respondents were asked to quantify the frequency at which certain internal supply chain risks occurred, and three out of the four most risks with the highest average frequency involve an external entity such as a supplier or transportation carrier (see Figure 1-2) [3]. Companies that participate in complex supply chains are also subject to risks that are unpredictable and outside the control of any decision makers [23]. Retail supply chains, for example, have previously been affected by
terrorist attacks, port lockouts, natural disasters (floods, hurricanes, earthquakes or break-outs of pandemic diseases), and economic crises [21]. These unexpected disruptions have a significant impact on the short-term financial performance of the involved firms, and carry potential long-term consequences such as loss of sales and stock price fall [15].

Recent publications called for attention in modeling supply chains as complex adaptive systems in order to quantify their robustness. There is no clear consensus on a simple way to measure robustness or even on one definition of robustness across literature. Perera et al. (2017) review a number of contributions that study supply chain robustness and summarize the definition to the ability of a system to maintain adequate performance over internal or external perturbations. As supply chains become more globalized, the likelihood of a firm being able to alter the general organization of the system decreases, and decision makers in each firm should therefore attempt to understand the system in order to position themselves in it [23].
There has been a special focus in modeling supply chain systems from a network science perspective, since this field offers a macro-level analysis that focuses on the topological properties of the relationships between the entities of a system. In this context, an entity is modeled as a node, and the relationships between a pair of entities as a link, and the entire system is referred to as a Supply Chain Network (SCN). Depending on the perspective of the problem, the intrinsic relationships can be considered unidirectional or bidirectional, in which cases we can represent the supply chain as a directed or undirected network, respectively.

Network theory provides measures on the level of individual components (nodes or links) that quantify the connectivity of these components and their impact on the overall structure, as well as global properties of the network. These global properties are often an aggregate of the local properties of all elements, such as an average or minimum, and provide a more general sense of the topology of the network [25] (see Section 2.1). Both types of metrics come useful in evaluating the performance of systems as complex as supply chains [26]. In addition, previous research has shown a high impact of the topology of a SCN on its robustness [18, 20, 29, 31].
Aside from helping describe the structure of existing supply chains, network science is also useful for modeling the evolution of SCNs through well-known growth models that describe different mechanisms and rules according to which nodes enter the system and links are created between nodes [27] (Section 2.2). It is known that many values of network measures are greatly influenced by basic network properties such as the number of nodes, the density (the ratio between the number of links and the number of nodes), and especially the degree distribution, that is, the frequency of occurrence for each possible number of connections a node could have.

For this reason, when exploring specific measures of networks, it is common to generate an ensemble of networks that preserve the number of nodes, density or degree distribution, but have diverse internal organizations, in order to obtain insights on the measures of interest [25]. As an example, networks where the occurrence of a number $k$ of connections can be approximated to $c \cdot k^{-\gamma}$ (also known as the scale-free distribution) for all possible values $k$ and for two fixed constants $c, \gamma$ are known to have the small-world property, meaning that the expected minimum number of nodes and relationships separating any two nodes is very small compared to the size of the network [4].

An ensemble of networks can be created through network growth models, or probabilistic algorithms that determine how nodes enter a system and create links with other existing nodes. Such algorithms typically require simple parameters such as the final number of nodes and edges, and each of them is able to produce a large number of different networks that satisfy these properties. See Section 2.2 for a thorough review on the most relevant network growth models.

Evaluating the robustness of networks from a topological perspective has been a common research interest in network theory for years, yet there is no consensus on one way to measure this property [23]. Perera et al. (2017) summarize the methodology for studying supply chain robustness and topology from a network science perspective (see Figure 1-3), grouping them into two types of approaches:
Figure 1-3: General methodological framework for studying robustness from the perspective of topology in networks. [23]
analytical and simulation-based. The former consists in using global measures that numerically describe the organization of links along the nodes, while the latter involves generating failure scenarios and analyzing the impact of each of them on global measures that describe the capacity for communication across the nodes [23].

The analytical approach provides static methods to assess the robustness of the system in its current state by summarizing the complexity of the network into a numerical value. The simulation approach, in contrast, provides non-deterministic answers based on a larger set of potential failure scenarios. The behavior of modern supply networks is rarely controlled by one entity, but is rather the result of many decisions made locally by a few firms [23]. Given this lack of power to largely modify the system as a whole, decision makers in SCNs need to focus on understanding how the other entities in the supply chain can affect their local performance. The simulation approach appears more suitable in this case, since it provides an opportunity for decision makers to understand the risks of the system they are a part of but that they can not entirely control. For this reason, we choose to explore on the simulation approach in this study.

In existing SCN literature, with the exception of Chen and Lin (2012) [6], failure scenarios are typically generated by simulating node removals. There exist two types of simulations: *random failures*, where the nodes to be removed are randomly selected, and *targeted attacks*, where iteratively the most relevant node is removed. SCN studies generally identify the most relevant node as the one with the largest *degree*, or number of connections [23]. Chen and Lin (2012) identify the most relevant edge as the one that connects the two nodes with the largest *degree* [6]. In the context of supply chains, a node removal represents a situation that prevents the underlying entity to perform their role in the chain appropriately. Example situations include a riot in a factory, a natural disaster in the location of the facility, a company going out of business, among others. In this study we extend the notion of failure to include edge removals. Real-life scenarios that could be represented as an edge
removal include the alteration of a contract between a supplier and a manufacturer or a failure in the transportation activity between two nodes (see Figure 1-2).

In addition to considering both node and edge disruptions, this study also explores different techniques for determining how relevant a component is in order to simulate targeted attacks. Identifying the sources of vulnerability in the performance of a network is important to determine which components to protect [12]. A straightforward approach to this problem would be to simulate the failure of every possible group of components, however, this becomes infeasible for SCNs involving a large number of entities and relationships between them [28]. Therefore, we point out the importance of developing and evaluating different methods for identifying the components in the network responsible for a lack of robustness. To this end, we propose the use of local measures that have been broadly studied in network theory as a heuristic to determining the most relevant components in maintaining the performance of an SCN. We evaluate each proposed measure by simulating targeted attacks where the most relevant node or edge to be removed is determined by the ordering imposed by each of these metrics.

In accordance with the general methodological framework for studying supply chain network robustness from a topological perspective (Figure 1-3), we take two real-world supply chains [30], generate distinct network topologies that resemble certain properties of the original data set, define global robustness measures for the network, and record the change in these measures after the failure scenarios to evaluate their impact on the network.

In addition to the analysis of the impact of component removal across different orderings, we propose and implement a proof of concept for an interactive visual interface that displays a supply chain network, highlighting critical components determined by different metrics. This is intended as a tool for supply chain managers to help them understand the topology of the network of their interest and explore design alternatives to increase the robustness of their systems.
Chapter 2

Literature Review

This literature review focuses on three areas that are fundamental to this research: the use of network measures to describe the topology of supply chain systems (see Section 2.1), the use of growth models to generate an ensemble of networks that share certain topological properties (see Section 2.2), and a review the different methods for generating alternative scenarios where certain components have failed and measuring the impact of these changes on the network robustness (see Section 2.3).

2.1 Local and Global Network Metrics

Studying the functionality of an SCN requires understanding the overall structure as well as the individual components of the network. Measures for nodes or links quantify the relevance of each component in different aspects of the network structure. These component-specific measures can be used to create a hierarchy among the components, but also to generalize the functionality of the network, usually by studying their average, minimum, maximum, and distribution, or the correlation between pairs of measures. This section provides an overview of the metrics that are commonly used in existing literature to assess vulnerability, robustness or resilience,
either as component-specific or as network-level metrics. Costa et al. (2007) and Rubinov and Sporns (2009) provide a thorough overview of most metrics used to characterize complex networks [8, 25].

The most fundamental metric for a node \(i\) is the number of links \(k_i\) it belongs to, referred to as the *degree*. The probability \(p_k\) that a node has degree \(k\) is computed by the ratio between the number of nodes with degree \(k\) and the total number of nodes, and the set of values \(p_k\) for all possible degrees \(k\) make up what is called the probabilistic *degree distribution* [4]. The shape of this distribution is commonly used to characterize networks, and several previous studies suggest that SCNs usually follow a *power-law distribution*, meaning \(p_k \sim k^{-\gamma}\) for some constant \(\gamma\) [16]. This distribution implies that there exist few agents with a large number of connections and a large number of agents with few connections. As a result, the connectivity across the network decreases significantly with the simulation of few targeted attacks. On the contrary, SCNs tend to maintain their connectivity after several random node removals, since there is a high probability that low-degree nodes are removed, which rarely play a vital role in maintaining the network’s connectivity.

Many times, we use links in networks to represent transactions of any kind (for example money, goods, or information) between the agents represented as nodes. A useful notion in this context is the *path* between a pair of nodes, which is a sequence of links such that the target node of each link in the sequence is the source node of the following link. The *length* of a path is the number of links in the sequence, and the *distance* or *geodesic distance* \(d_{ij}\) between two nodes \(i\) and \(j\) is the shortest length of any path that connects them. A *shortest path* between two nodes \(i\) and \(j\) is a path from \(i\) to \(j\) whose length is \(d_{ij}\) [4]. By definition, if there exists no path between \(i\) and \(j\), then \(d_{ij} = \infty\). The *largest connected component* (LCC) is the subset of nodes in the graph such that there exists a path between any two pairs of nodes in the subset. The *diameter* of a network is the largest distance between any two nodes. The *average path length* is the average of all the distances between all
pairs of nodes [4]. Long distances between nodes in an SCN are indicative of many stages in production and a greater difficulty in intervening in the supply chain [23].

The diameter and the average path length give a sense of how closely connected nodes are to one another, however they diverge if the graph is not fully connected, that is, if there is at least one pair of nodes that are not reachable from each other [5]. An alternative for capturing this property is the average inverse geodesic distance, sometimes referred to as the (global) efficiency of the network. It is given by

\[
E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}
\]

where \(1/d_{ij} = 0\) when there is no path from \(i\) to \(j\) [17]. This metric quantifies how efficient the SCN is in transmitting information, assuming that the efficiency to which information is exchanged between two nodes is inversely proportional to the distance separating them.

Some local metrics quantify the relevance of a component in the network for the communication between other nodes. The betweenness centrality of a node \(i\) measures the intervention of \(i\) in the interactions of other nodes. Put differently, it can measure the extent to which a node can spread transacted values originated in other nodes to the rest of the network. The betweenness centrality is calculated as

\[
C_b(i) = \frac{\sum_{s \neq t \neq i} \sigma_{s,t}(i)}{\sigma_{s,t}},
\]

where \(\sigma_{s,t}\) is the total number of shortest paths from node \(s\) to node \(t\) and \(\sigma_{s,t}(i)\) are the number of such paths that pass through \(i\). The same notion of betweenness centrality can be applied to edges. The load centrality is a closely related measure that describes the same behavior, however is computed as the fraction of all shortest paths that pass through \(i\) [10]. In the context of SCNs, specifically, a high betweenness centrality measures the extent through which a firm intervenes in the
Figure 2-1: The effect of clustering and range on the survivability and communication of a simple supply chain network. The existence of links between the neighbors of a Supplier allow Supplier to find an alternative path to Retail_1 that is only one edge longer. More generally, this means that for all nodes that Supplier is connected to through Manuf_1 are not strictly dependent on the existence of the edge (Supplier, Manuf_1). Also, the short range of the edge (Supplier, Manuf_1) means that the communication between Supplier and Manuf_1 can not only be reestablished, but also is still efficient after the re-routing.

interactions between other firms in the network [23]. This means that the failure of a firm with a high betweenness centrality might compromise the communication between several pairs of firms.

In many real-life examples of networks it is common to encounter the formation of communities, or subsets of nodes that are highly connected to other members of the community but poorly connected to nodes outside their own community. Some graph measures are related to the formation of communities and the role of nodes or edges in the existence of such communities. The clustering coefficient of a node $i$ is the ratio of its neighbors that are connected to each other. Given by the equation

$$C_i = \frac{\sum_{j \in N(i)} \sum_{l \in N(i) - \{j\}} (A)_{jl}}{k_i(k_i - 1)},$$

where $A_{ij} = 1$ if there exists a link between $j$ and $l$, and $A_{ij} = 0$ otherwise. A high level of clustering is good for flexibility in operations in an SCN by guaranteeing the existence of short alternate paths in case of edge disruptions (see Figure 2-1).

The clustering coefficient can also be computed for an edge $(i, j)$ with the equa-
\[
\hat{\mathcal{C}}_{i,j} = \frac{\Delta_{i,j} + 1}{\min(k_i - 1, k_j - 1)},
\]

where \(\Delta_{i,j}\) is the number of triangles \(i, j\) to which the edge \((i, j)\) belongs [8]. Edges with a low clustering coefficient have been found to serve as connectors for pairs of nodes in different communities [24], indicating contracts in the SCN that, if broken, might easily compromise the connectivity of the entire network.

Some measures quantify the impact of a component in the connectivity of the network by directly simulating the failure of the component and studying the impact it has on the network. One example is the range, which is defined for an edge \((i, j)\) as the distance between nodes \(i\) and \(j\) in the absence of edge \((i, j)\). The notion of range is important to measure the complexity in re-routing in the case of a disruption in the flow of goods between to adjacent nodes, since it can be thought of as a measure of the efficiency in communication between two nodes in the case of failure.

In some cases, it is useful to classify nodes by type and analyze the assortative mixing, or the tendency of nodes to link with others of the same type. The degree assortativity measures the tendency of nodes to link with others of similar degree. It is also possible to define local assortativity measures, meaning the tendency of a specific node to have networks of its same type. The average neighbor degree is a measure for local assortativity, and is computed by

\[
k_{nn,i} = \frac{\sum_{j \in N(i)} k_j}{k_i}.
\]

Typically, nodes who score lowly in this measure are the most critical for maintaining the network together [25].
2.2 Network Growth Models

In network science, a random network is one where the probability that each pair of nodes is connected is a fixed value \( p \). The \textbf{Erdős-Renyi (ER)} model generates directed or undirected random networks given two positive integer parameters \((N, L)\) as described in Algorithm 1 [4].

\begin{algorithm}
\begin{algorithmic}
  \State \textbf{input} : \( N, L \in \mathbb{Z}^+ \), \texttt{directed} \in \{ \texttt{True}, \texttt{False} \}
  \State \textbf{output}: A random network with \( N \) nodes and \( L \) edges
  \State Initialize a graph with nodes 1, 2, \ldots, \( N \) and no edges;
  \If {\texttt{directed}}
      \State \( S = \{ (i, j) \mid 1 \leq i, j \leq N \text{ and } i \neq j \} \);
  \Else
      \State \( S = \{ (i, j) \mid 1 \leq i < j \leq N \} \);
  \EndIf
  \State Choose \( T \subseteq S \) uniformly at random such that \( |T| = L \);
  \State Add an edge between each pair \( (s, t) \in T \);
\end{algorithmic}
\end{algorithm}

\textbf{Algorithm 1:} The Erdős-Renyi (ER) growth model algorithm.

Networks generated by this model represent the most basic way to randomize attachment between nodes so that the probability of an edge existing between any pair of nodes is uniform. In SCNs, the ways in which nodes join the network and create links is highly non-randomized. A node joining a network represents a firm positioning itself along the supply chain, and the creation of links represents partner selection. Partner selection is a strategic decision that is far from random, and is rather a multi-objective problem that takes into account many different factors such as costs, location, ease of transportation, size of the partners, to name a few [23, 13].

For this reason, we consider other models that generate networks that display certain real-world properties of SCNs that have been found desirable for the survivability of the system. The first of these properties is node \textit{clustering}, or the general tendency of the neighbors of a node to form connections with each other. Another real-life pattern of SCNs is called the \textit{small world effect}, that is observed when the expected distance between any pair of networks is low with respect to the number
of nodes in the network. A good metric for “small-worldness” is that the maximum or the average distance of a network with \( N \) edges will be \( \sim \log(N) \) for really large values of \( N \). The **Watts-Strogatz (WS)** growth model generates random networks that probabilistically reflect for clustering and small-worldness by arranging nodes in a circle, connecting them to their closest neighbors, and randomizing edge rewiring (see Algorithm 2).

```plaintext
input : \( N, 2L \in Z^+, p \in [0, 1] \)
output: A random network with \( N \) nodes, \( 2LN \) edges, a high clustering coefficient and generally small distances between nodes

1. Initialize a graph with nodes 1, 2, \ldots, \( N \);
2. for \( i \leftarrow 1 \) to \( N \) do
   3. for \( j \leftarrow 1 \) to \( L \) do
      4. Connect \( i \) with its \( 2L \) closest neighbors
         5. if \( i - j < 1 \) then
            6. Add edge \((i, i - j + N)\);
         7. else
            8. Add edge \((i, i - j)\);
         end
      9. if \( i + j > N \) then
         10. Add edge \((i, i + j - N)\);
      11. else
         12. Add edge \((i, i + j)\);
      end
   end
3. for \( i \leftarrow 1 \) to \( N \) do
   4. for \( t \in N(i) \) do Rewire
      5. Generate \( r \in [0, 1] \) uniformly at random;
      6. if \( r < p \) then
         7. Choose node \( t' \) uniformly at random;
         8. Remove edge \((i, t)\);
         9. Add edge \((i, t')\);
      end
   end
```

The degree distribution of a random node resembles a Poisson distribution, which
means that the probability $p_k$ of a node having degree $k$ can be approximated to

$$\frac{\lambda^k e^{-\lambda}}{k!}$$

for some constant value $\lambda$. However, many real-world networks such as the World-Wide Web [1], the Internet [9], metabolic networks [14], display a scale-free distribution. Mathematically, this means that the probability $p_k$ of a node having degree $k$ follows $p_k \sim k^{-\gamma}$ for a constant $\gamma$.

The Barabási-Albert (BA) model generates scale-free networks following the preferential attachment principle, also known as “rich gets richer”. The growth model consists in iteratively adding a new node to the graph in each step, and creating $M$ links between the new node and existing ones with a probability proportional to the existing nodes’ degree (see Algorithm 3).

```plaintext
input : $N, M \in \mathbb{Z}^+$
output: A scale-free network with $N$ nodes and $(N - M)M$ edges
1 Initialize a graph with nodes 1, 2, $\ldots$, $M$ and no edges;
2 $i \leftarrow M + 1$;
3 while $i < N$ do
4     Choose node $t \leftarrow \{1, 2, \ldots, i - 1\}$ with probability $\sim k_t$;
5     if edge $(i, t)$ does not exist then
6         Add edge $(i, t)$;
7         $i \leftarrow i + 1$;
8     end
9 end
Algorithm 3: The Barabási-Albert (BA) growth model algorithm.
```

Scale-free networks exhibit the existence of hubs, that is, nodes found in small frequency that concentrate a large number of links. They also exhibit small-worldness for any values of $\gamma$. In particular, values of $\gamma < 3$ create what is referred to as the ultra-small property, meaning that the diameter of the network is $\sim \log \log N$ for networks with a large number of nodes $N$ [7].

One limitation of the BA model is that it produces networks with a low clustering
coefficient. As mentioned before, a high level of clustering is important for SCNs for easy rerouting. Holme and Kim’s (HK) propose a model for scale-free networks with a tunable parameter \( p_t \) that adds an extra level of clustering. The algorithm is based off of the BA model, with an extra triad formation step after the preferential attachment (see Algorithm 4) [11].

\[
\text{input: } N, M \in \mathbb{Z}^+, p_t \in [0, 1] \\
\text{output: } \text{A scale-free network with } N \text{ nodes and } (N - M)M \text{ edges and high clustering}
\]

1. Initialize a graph with nodes 1, 2, \ldots, \( M \) and no edges;
2. \( i \leftarrow M + 1 \);
3. while \( i < N \) do
   4. Choose node \( t \leftarrow \{1, 2, \ldots, i - 1\} \) with probability \( \sim k_t \);
   5. if edge \((i, t)\) does not exist then
      6. Add edge \((i, t)\);
      7. \( i \leftarrow i + 1 \);
   end
9. \( M' \leftarrow 1 \);
10. while \( M' < M \) do
   11. Generate \( r \in [0, 1] \) uniformly at random;
   12. if \( r < p_t \) then
      13. \( N' = \{(s, s') \in N(i) \mid \text{edge (s, s') does not exist}\} \);
      14. if \( |N'| > 0 \) then
         15. Choose \((s, s') \in N(i)\) uniformly at random;
         16. Add edge \((s, s')\);
         17. \( M' \leftarrow M' + 1 \)
      end
   19. else
      20. Execute preferential attachment (lines 4-8);
   end
22. end
23. end

Algorithm 4: Holme and Kim’s (HK) scale-free with tunable clustering growth model algorithm.

The last three models mentioned above achieve expectedly short distances between nodes as a consequence of the topology of the networks generated. As mentioned before, a company’s capacity of intervening in the performance of the other
entities that are part of their chain decreases as the distance from them increases [23]. As a result, it would make sense for a firm to actively try to minimize distances from other firms as they form their connections along the supply chain. The Degree and Locality Based Attachment (DLA) models a preferential attachment based on both degree and distance (see Algorithm 5) [31].

\begin{algorithm}
\begin{algorithmic}
\State \textbf{input} : $N, M \in \mathbb{Z}^+, p_t \in [0,1]$
\State \textbf{output}: A scale-free network with $N$ nodes and $(N - M)M$ edges and high clustering
\State Initialize a graph with nodes $1, 2, \ldots, M$ and no edges;
\State $i \leftarrow M + 1$;
\While {$i < N$}
\State Choose node $t \leftarrow \{1, 2, \ldots, i - 1\}$ with probability $\sim k^u_t$;
\While {edge $(i, t)$ exists}
\State Reassign $t \leftarrow \{1, 2, \ldots, i - 1\}$ with probability $\sim k^u_t$;
\EndWhile
\State Add edge $(i, t)$;
\State $M' \leftarrow 1$;
\EndWhile
\While {$M' < M - 1$}
\State Choose node $j \leftarrow \{1, 2, \ldots, i - 1\}$ with probability $\sim d^r_{ij}$;
\If {edge $(i, j)$ does not exist}
\State Add edge $(i, j)$;
\State $i \leftarrow i + 1$
\EndIf
\EndWhile
\EndWhile
\end{algorithmic}
\caption{Algorithm 5: The Degree-Locality Based Attachment model algorithm.}
\end{algorithm}

2.3 Failure, Attacks and Robustness

This section reviews the existent literature on using failure simulation to measure robustness in real-life supply networks, or in networks generated by different growth models (see Table 2.1). In particular, we focus on contributions that consider targeted attacks on the components of the network. Perera et al. (2017) summarize the concept of SCN robustness as the “extent to which a given SCN can withstand
the loss of its components, without losing its basic function" (p. 11) [23]. This definition of robustness will be adopted in this paper. It is worth noting that some works reviewed in this section refer to this notion as *invulnerability* or *resilience*, but preserve the methodological framework of interest described in Figure 1-3.

Albert et al. (2000) explore the degree of tolerance to failure of complex networks, using both real-world data sets and generated networks from growth models [2]. The two real-world networks used (the World-Wide Web and the Internet) are said to follow scale-free degree distribution. Through failure simulation, the authors effectively demonstrate that the response to targeted attacks of the three scale-free networks is identical, despite different values of clustering coefficient, power-law exponent and average degree.

Motter et al. (2002) does not directly aim to capture the robustness of networks to failure, but indirectly addresses this by studying the influence of short vs. long-range edges in the small-worldness property. One thing to observe is that the small-worldness is defined as short average/maximum path length. Even though the efficiency of the network is also a generalization for the distances between nodes across the network, there is no proved correlation between the efficiency and the average/maximum path length in the networks studied.

Holme et al. (2002) provide a very thorough evaluation of targeted attacks by the degree and betweenness centrality for both edges and nodes, accompanied by a correlation analysis between the two metrics. Something unique about this work is that it distinguishes between initial and recalculated centrality values for attack simulations. Simulations that use initial values calculate the centrality measure for each node (edge) in the original graph, sort the nodes (edges) by the largest value, and remove them in that order. On the contrary, simulations with recomputed centrality measures compute the centrality measures for each component before every single removal. Some studies [2, 19] do not specify how these values are calculated, and studies who do specify do not justify the choice [31, 6].
Table 2.1: Summary of recent works that use failure simulation to measure robustness in real-life supply networks, or in networks generated by different growth models

Chen et al. (2012) is the only work that studies a real supply chain data set. While it provides a thorough analysis of the data set described, the fact that it only studies one network limits the extent to which the conclusions drawn in the study can be extrapolated to other networks. This work can be thought of as a demonstration of how a supply chain decision maker may go about analyzing the topology of the underlying SCN, rather than a study with any significant contribution to applied network science.

As can be seen, in context of SCNs there is only one study that has explored
targeted edge removals [6] despite the recurrent calls for action in this field [18, 23, 32]. No known SCN studies explore node removal by a criteria different than degree. The only known network science study that explores node removal by a criteria different than degree uses only betweenness centrality. Other node centrality measures such as the load centrality and the average neighbor degree have not been explored as a hierarchy for targeted node removal, and the edge clustering coefficient also has not been used as a criteria for targeted edge removal. Some authors argue that targeted attacks are only considered by degree because they correspond to adversary attacks, and any other centrality measure would require knowledge of the entire topology of the network, which is rarely available to the attacker [2, 31, 32]. However, Home et al. raise the importance of vulnerability studies for the sake of protection, in order to identify the most important components and protect them accordingly [12].
Chapter 3

Methodology

The focus of this thesis is to provide a framework for the assessment of supply chain robustness from a network science perspective through the analysis of failure scenarios. In a scientific review Perera et. al [23] summarize the general methodological options into two types of approaches: analytical and simulation-based (see Figure 1-3). This work focuses and expands on the simulation approach, and proposes the use of visual tools to aid the understanding of the topological features of supply chain networks that are relevant for the simulations proposed. The data set under study and the computational aspect of the analysis is documented in Section 3.1. A description of the models and metrics chosen is presented in Section 3.2. Section 3.3 provides a through description of the two real-world supply chains that were chosen as the subject of this study, as well as the parameters used to generate their corresponding ensembles of networks using network growth models. Finally, Section 3.4 contains a description of the visualization tools proposed.

3.1 Dataset Processing

The data chosen for this study is a collection of 38 real world supply chains modeled by either company analysts or consultants. The industries represented include
food, organic chemicals, computer parts, electrical appliances, heavy machinery and freight transportation. Each of the 38 data sets consists of *stages* which represent a supply chain activity (see Table 3.1) and *arcs* between stages which denote the precedence of an activity to the other one (see Figure 3-1).

<table>
<thead>
<tr>
<th>Classification</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist</td>
<td>a stage that distributes an item</td>
</tr>
<tr>
<td>Manuf</td>
<td>a stage that manufactures or assembles an item</td>
</tr>
<tr>
<td>Part</td>
<td>a stage that procures an item</td>
</tr>
<tr>
<td>Retail</td>
<td>a stage that acts as a demand origination point</td>
</tr>
<tr>
<td>Trans</td>
<td>a stage that transports an item between stages</td>
</tr>
</tbody>
</table>

Table 3.1: Classifications used to label stages in the chains [30]

Figure 3-1: A real-life supply chain from a semiconductors and related devices company

The included supply chains have a wide range of complexity, and are composed of anywhere from 8 to 2025 stages, and 10 to 16255 arcs [30]. This data has been the subject of study of two SCN-related publications. Mari et al. (2015) use one of the provided supply chain data sets to show the applicability of network growth models to design resilient supply chains. Their results show that, even though the best fitting model is the scale-free growth model, the networks generated by the BA growth model do not represent an efficient and resilient supply chain. The paper calls for the development of new network growth models that are both robust to targeted attacks and random failures [18]. Perera et al. select 26 out of the 38 data
sets to compare their topology with those produced by different network growth models, finding that the data set is generally lowly centralized, either slightly or strongly dissassortative, either moderately or highly modular, and robust against random failures. The study also concludes that most networks are scale-free with frequent occurrences of values of $\gamma > 2$ [23].

For the purpose of demonstrating the use of our proposed methodology, we experiment on only two of the 38 datasets, which we consider the most complex. In the description of the dataset, Willems et al. (2008) point out that the complexity of the contained supply chains does not necessarily increase with the size, since large chains sometimes exhibit redundant patterns because they tend to replicate the bill of materials in an enterprise planning system. It also suggested that “chains of 150 and 350 stages demonstrate the greatest complexity" (p. 21) [30]. While there is only one dataset with approximately 350 stages, there are about five with close to 150 stages, so it is unclear which one(s) the authors are referring to. Aside from reportedly observed complexity, another requirement for selecting a supply chain is the presence of specific patterns in the topology of the network. For instance, some of the proposed metrics appeal to the existence of more than one path between pairs of nodes, and only one of the five chains with approximately 150 stages have this property. The datasets eventually selected for our analysis correspond to a computer equipment manufacturer (156 stages) and a producer of power-driven hand tools (334 stages). They are identified as data sets number 19 and 24 in the data source, respectively.

The data is publicly available as a table containing the list of the directed arcs between stages (see Figure 3-1). The 38 chains were parsed using Python’s xrl1d library for reading .xls files and was later built into instances of the Graph class of Python’s NetworkX package for network analysis.
3.2 Models and Metrics for Data Analysis

The supply chains we analyze are represented as undirected homogeneous networks with unweighted links. Although the relationship between stages in supply chains is by nature directed, there are very few known network growth models to generate directed graphs that preserve the information about the directionality. Therefore it would not be possible to produce a varied ensemble of networks to compare the impact of failures on the network robustness. Similarly, most network growth models and network metrics used in the SCN literature correspond to homogeneous networks -where all nodes are regarded the same way-, and those few that propose methods for heterogeneous networks -where nodes have types that the model distinguishes-use attachment rules that are specific to the classification of nodes in the real-world data set being experimented \cite{29, 31, 18}. By ignoring node types in our data set, we provide an analysis that is purely based on the internal organization of the relationships between entities in the supply chain, and the results can be extrapolated to applications of network science methods in other fields.

For each of the two chosen supply chains, we generated an ensemble of networks consisting of the original network, and five networks generated using each of the five growth models described in Section 2.2, for a total of twelve networks to be analyzed. The Erdos-Renyi (ER), Watts-Strogatz (WS) and Barabasi-Albert (BA) are considered the three benchmark models for in SCN literature \cite{23}. We also included Holme and Kim’s scale free model with tunable clustering (HK) given the significant role of clustering in robustness \cite{20, 29}, and the Degree and Locality-based Attachment (DLA) model given the importance of short distances in order for a firm to exercise influence in the internal performance of other entities along the supply chain \cite{23}.

For each of these twelve networks we measure their robustness by studying the change after failures in the efficiency and the size of the LCC. We simulate tar-
geted node attacks picking the nodes iteratively by the following centralities: degree, betweenness, load and average neighbor degree; and edge attacks chosen by betweenness centrality, clustering coefficient and range. We believe that each sub-network that remains after each targeted attack represents a possible realization of an SCN. We believe that using the initial (as opposed to recalculated) centrality measures would limit our analysis to only reflect properties of the initial network. Recalculating the measures after each simulation is a way of recognizing the leftover components of the original network as a coherent structure with valuable information.

3.3 Networks

This section presents the main topological properties of the two real-world of the networks that are analyzed in this study, as well as the choice of parameters to generate the ensemble of networks with different growth models presented in Section 2.2 (see Table 3.2).

<table>
<thead>
<tr>
<th>Growth Model</th>
<th>Computer Peripheral Equipment</th>
<th>Power-Driven Handtools</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>$N = 156, M = 263$</td>
<td>$N = 334, M = 1245$</td>
</tr>
<tr>
<td>WS</td>
<td>$N = 156, K = 3, p = 0.006410$</td>
<td>$N = 334, K = 7, p = 0.00299$</td>
</tr>
<tr>
<td>BA</td>
<td>$N = 156, M = 2$</td>
<td>$N = 334, M = 4$</td>
</tr>
<tr>
<td>HK</td>
<td>$N = 156, M = 2, p = 0.3$</td>
<td>$N = 334, M = 4, p = 0.3$</td>
</tr>
<tr>
<td>DLA</td>
<td>$N = 156, M = 2$</td>
<td>$N = 334, M = 4$</td>
</tr>
</tbody>
</table>

Table 3.2: Comparison robustness related network-level metrics of two real-life supply networks of different product lines.

3.3.1 Real-World Supply Networks

As mentioned before, the real-world supply chains we analyze correspond to two different companies and two different product types: computer peripheral equipment and power-driven handtools. The computer equipment supply chain has 156 stages
### Table 3.3: Robustness related network-level metrics of computer equipment manufacturer supply network.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># Nodes</td>
<td>156</td>
</tr>
<tr>
<td># Links</td>
<td>263</td>
</tr>
<tr>
<td>Avg. # Links/Node</td>
<td>3.37</td>
</tr>
<tr>
<td>Max. Chain Length</td>
<td>8</td>
</tr>
<tr>
<td>Chains Description</td>
<td>Part (\rightarrow) {Manuf, Trans} (\rightarrow) Dist</td>
</tr>
<tr>
<td>Degree Assortativity</td>
<td>-0.2542</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

classified as \(Dist, Manuf, Part\) or \(Trans\), and its chains all consist of an initial \(Part\) stage, a chain of alternated \(Manuf\) and \(Trans\) stage, and a \(Dist\) stage at the end. The handtools chain has 334 stages classified as \(Dist, Manuf\) or \(Part\), and its chains consist of a \(Part\) stage followed by zero or more \(Manuf\) stages and two final sequential \(Dist\) stages. For a comparison of some relevant topological properties between the two chains, we refer the reader to Tables 3.3 and 3.4. Both supply chains represent fully connected networks, meaning that the undirected representation there exists a path between any two nodes. This also implies that the initial size of the LCC is equal to the number of nodes in the network.

### 3.3.2 Erdős-Rényi Networks

The ER networks are generated with two parameters, \(N\) and \(L\), corresponding to the number of nodes and links, respectively. These parameters are simply set to the
<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># Nodes</td>
<td>334</td>
</tr>
<tr>
<td># Links</td>
<td>1245</td>
</tr>
<tr>
<td>Avg. Links/Node</td>
<td>7.45</td>
</tr>
<tr>
<td>Max. Chain Length</td>
<td>5</td>
</tr>
<tr>
<td>Chains Description</td>
<td>Part → {Manuf} → Dist → Dist</td>
</tr>
<tr>
<td>Degree Assortativity</td>
<td>-0.6038</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.0079</td>
</tr>
</tbody>
</table>

Table 3.4: Robustness related network-level metrics of a power-driven handtools manufacturer supply network.

original number of stages and arcs of the supply chains. For this generation, the `gnm_random_graph()` method of the NetworkX Python \(^{1}\) library is used.

### 3.3.3 Watts-Strogatz Networks

The WS model takes three parameters \(N, K\) and \(p\) corresponding to the number of nodes, the initial degree of each node at the simulation setup, and the probability that each edge is rewired during randomization (see Algorithm 2 lines 16-25). We want to preserve the number of nodes and links from the original network, so we set \(K\) to be the average degree. In addition, according to Holme et al. (2002) a probability \(p \sim O(1/N)\) creates a network that displays a high clustering coefficient as well as the small-world effect [12]. Since \(N\) is small, we choose explicitly \(p = 1/N\). For this generation the `connected_watts_strogatz_graph()` method of the

\(^{1}\)https://networkx.github.io/documentation/networkx-1.10/reference/generators.html#module-networkx.generators.random_graphs
NetworkX package is used, which guarantees that the network will be connected after rewiring. This method was chosen above the standard `watts_strogatz_graph()` since the size of the LCC is one of the parameters used to analyze robustness, and it was desirable to keep it consistent with the original size of the LCC in the real-life network.

### 3.3.4 Barabási-Albert Networks

The simulation of this growth model takes two parameters $N$ and $M$ corresponding to the number of nodes and the number of links added to each new node in the simulation. For this growth model, the `barabasi_albert_graph()` method of the NetworkX library is used. The parameter $M$ was set to the closest integer so that $MN$ equals the number of links in the original network.

### 3.3.5 Scale-free Networks with Tunable Clustering

This algorithm takes three parameters $N, M$ and $p$ where $N$ is the number of nodes and $M$ is the number of links added to each new node in the simulation. The parameter $M$ is set the same way it is set for the BA simulation in order to guarantee that the density of the generated network is at least the density of the original. The control parameter $m_t$ that corresponds to the average number of triad formation trials per added node and is computed as $m_t = m(1-p)$ is set to 0.3 which is expected to raise the clustering coefficient by $\sim 10\%$ [12]. A small value is chosen because the clustering coefficient of these supply chain networks is rather small (see Table 3.3 and 3.4). The NetworkX library includes a built-in method `powerlaw_cluster_graph()` which is used for this simulation.
3.3.6 Degree-Locality Based Attachment (DLA)

This growth model takes four parameters $N, M, u,$ and $r$, corresponding to the number of nodes in the network, the number of links added to each node as it joins the network, and two parameters for tuning the preferential attachment (see Algorithm 5) which are set to 1 for consistency in the way preferential attachment is performed in BA and HK. The parameters $M, N$ were set the same way as for the BA algorithm (see Section 3.3.4). This algorithm is implemented as shown in 6.

```python
def DLA_graph(n, m):
    # Initialize an empty graph with m initial nodes
    G = nx.empty_graph(m)
    # Add the first m links from the m+1 node to all the rest
    G.add_edges_from(zip([m]*m, range(m)))
    source = m
    while source < n:
        # The first edge connects with degree-based attachment
        first_node = pick_from_probabilities(nx.degree_centrality(G))
        G.add_edge(source, first_node)
        targets = set([first_node])
        # Add each of the m-1 links
        for new_link in range(m-1):
            # Recalculate inverse distances between source and targets
            # to choose preferential
            distances = nx.single_source_shortest_path_length(G, source)
            locality_preferences = {node: 1/dist
                                        for node,dist in distances.items()
                                        if node not in targets and dist != 0}
            normalization_value = sum(locality_preferences.values())
            attachment_probabilities = {node: val/norm
                                            for node,val in locality_preferences.items()}
            target = pick_from_probabilities(attachment_probabilities)
            G.add_edge(source, target)
            targets.add(target)
            source += 1
    return G
```

**Algorithm 6:** Implementation for Degree-Locality Based Attachment growth model.
3.4 Data Visualization

The purpose of data visualization is to help supply chain managers understand the sources of vulnerability to attacks in different aspects of the overall performance of the network that can be determined from the network topology. The proposed software would also be useful for managers to study possible reconfiguration strategies and their impact on the robustness of the network.²

The software proposed uses three principles for visual analytics systems particularly useful for the study of supply networks [22], contextualized in the current study as follows:

1. **Integration of multiple views in an interface.** The interface has two side-to-side network visualizations: one corresponding to the original network, and a second one displaying the modified network with applied changes.

2. **Interactive investigation.** The network visualizations have dragging and zooming capabilities. Hovering over a node identifies it and provides its type of node. Clicking over a node or an edge displays all the relevant properties. In addition, the user has the choice to select a vulnerability measure for network components, and the nodes and edges change color and thickness based on this measure.

3. **Data-driven analytics capabilities.** The proposed tool allows investigation of alternative scenarios. In the context of this project, the visualization supports adding and removing nodes and edges and reflecting the changes in the properties of each component.

²A proposed implementation of this system can be found at https://github.mit.edu/mxruedag/meng_thesis
Chapter 4

Experimentation and Results

This section describes the main findings from the analysis of the simulation of targeted attacks on nodes and edges by using different measures to identify the next most relevant component in the network. Section 4.1 provides some insights on the topology of the networks generated with the predetermined growth models and discusses how they resemble the structure of the original supply chains (SCs). Sections 4.2 and 4.3 describes the general results of the simulations of targeted attacks using different centrality measures on the nodes and edges of the network, respectively. Section 4.4 discusses the mathematical nature of the efficiency measure for networks with respect to the desired properties of a global measure of robustness.

Along this section, we will use the notation \( \text{original}_{\text{dataset}}_{\text{growth_model}} \) to refer to the network generated by \( \text{growth_model} \) with the parameters from the real-life supply chain (SC) denoted by \( \text{original}_{\text{dataset}} \). The growth models are \( \text{ER}, \text{WS}, \text{BA}, \text{HK}, \text{DLA} \) for the original data set, the Erdős-Renyi model, the Watts-Strogatz, the Barabási-Albert model, the preferential attachment with tunable clustering model, and the degree-locality based attachment model. The original dataset is denoted by orig. The supply chains are denoted \( \text{comp} \) and \( \text{tools} \) for the Computer Equipment and Power-Driven Handtools SCs, respectively. For example, \( \text{orig}_{\text{comp}} \) refers to the network that represents the Computer Equipment supply
chain, and ER_tools refers to the network generated by the ER model parametrized on the topology of the Power-Driven Handtools SC.

4.1 Network Topologies

This section provides an analysis of the diverse topologies obtained in the ensemble of networks generated from the real-life supply chain networks. See Tables 4.1 and 4.2 for some relevant network metrics on the networks of the ensemble generated by the Computer Equipment and the Power-Driven Handtools SCs, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>orig</th>
<th>ER</th>
<th>WS</th>
<th>BA</th>
<th>HK</th>
<th>DLA</th>
</tr>
</thead>
<tbody>
<tr>
<td># Nodes</td>
<td>156</td>
<td>156</td>
<td>156</td>
<td>156</td>
<td>156</td>
<td>156</td>
</tr>
<tr>
<td># Links</td>
<td>263</td>
<td>263</td>
<td>156</td>
<td>308</td>
<td>308</td>
<td>308</td>
</tr>
<tr>
<td>Density</td>
<td>1.69</td>
<td>1.69</td>
<td>1.0</td>
<td>1.97</td>
<td>1.97</td>
<td>1.97</td>
</tr>
<tr>
<td>Min Degree</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Max Degree</td>
<td>28</td>
<td>10</td>
<td>3</td>
<td>25</td>
<td>19</td>
<td>119</td>
</tr>
<tr>
<td>Average Degree</td>
<td>3.37</td>
<td>3.37</td>
<td>2.0</td>
<td>3.95</td>
<td>3.95</td>
<td>3.95</td>
</tr>
<tr>
<td>Global Clustering Coefficient</td>
<td>0.003</td>
<td>0.010</td>
<td>0.000</td>
<td>0.088</td>
<td>0.277</td>
<td>0.571</td>
</tr>
</tbody>
</table>

Table 4.1: Global measures for the networks in the ensemble generated for the comp supply chain.

<table>
<thead>
<tr>
<th>Model</th>
<th>orig</th>
<th>ER</th>
<th>WS</th>
<th>BA</th>
<th>HK</th>
<th>DLA</th>
</tr>
</thead>
<tbody>
<tr>
<td># Nodes</td>
<td>334</td>
<td>334</td>
<td>334</td>
<td>334</td>
<td>334</td>
<td>334</td>
</tr>
<tr>
<td># Links</td>
<td>1245</td>
<td>1245</td>
<td>1002</td>
<td>1320</td>
<td>1316</td>
<td>1320</td>
</tr>
<tr>
<td>Density</td>
<td>3.73</td>
<td>3.73</td>
<td>3.0</td>
<td>3.95</td>
<td>3.94</td>
<td>3.95</td>
</tr>
<tr>
<td>Min Degree</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Max Degree</td>
<td>68</td>
<td>19</td>
<td>7</td>
<td>79</td>
<td>61</td>
<td>305</td>
</tr>
<tr>
<td>Average Degree</td>
<td>7.46</td>
<td>7.46</td>
<td>6.0</td>
<td>7.90</td>
<td>7.88</td>
<td>7.90</td>
</tr>
<tr>
<td>Global Clustering Coefficient</td>
<td>0.008</td>
<td>0.028</td>
<td>0.591</td>
<td>0.079</td>
<td>0.187</td>
<td>0.311</td>
</tr>
</tbody>
</table>

Table 4.2: Global measures for the networks in the ensemble generated for the tools supply chain.

As explained in Sections 3.3.2 through 3.3.6, networks in one ensemble roughly preserve the number of nodes and links from the original data set. However, depending on the model used to generate them, their internal organization will largely
vary. This is evident from the global measures presented in Tables 4.1 and 4.2. The maximum degrees in the random networks of each ensemble is reduced significantly from the original data set. The BA and HK generate values of maximum degree that are fairly close to the original ones ($<20\%$ change). The DLA model concentrates a large share of the links in one hub node and therefore the maximum degree is very large in both cases. The results of the WS modeling highly misrepresent the global network measures of the original data sets. The number of links of the WS\textsubscript{comp} and WS\textsubscript{tools} networks are not even correspondent to the number of links the WS model was parametrized with. This might be a result of using NetworkX’s connected\textunderscore watts\textunderscore strogatz\textunderscore graph() as opposed to watts\textunderscore strogatz\textunderscore graph(). The first method might not be able to converge to a connected network after several attempts of rewiring for small $L$ parameters, in which case the structure will be misrepresented.

We also analyze the degree distributions in each ensemble of networks since it provides a more thorough overview of the sources of discrepancy between the degree-related global properties of the networks. Both original SCNs display a large concentration of low-degree nodes, with almost no intermediate-degree nodes and a small number of high-degree nodes (see Tables 3.3 and 3.4). This phenomenon is only captured by the BA, HK and DLA models (see Figures 4-1(d,e,f) and 4-2(d,e,f)), which are the only models of choice that produce networks with a scale-free distribution. Networks produced by both the BA and the HK models display a gradual decrease in the frequency as the degree grows, and produce a few hub nodes that concentrate most of the links (see Figures 4-1(e,f) and 4-2(e,f)). The DLA model, however, generates for both data sets only one hub concentrating a great share of the links: 119 out of 308 in DLA\textsubscript{comp}, and 305 out of 1320 in DLA\textsubscript{tools}. In DLA\textsubscript{comp} this means that the average degree of the non-hub nodes is 3.19. In fact, there are 80 out of 156 nodes with degree 2, which means that more than half the nodes did not acquire any links other than the 2 they were initialized with, so
Figure 4-1: Degree distribution of ensemble of networks corresponding to the Computer Equipment supply chain network.

both preferential attachment steps (degree and locality-based) quickly converge to highly favor a small set of nodes which end up with higher degree values.

In the case of the handtools supply network, the DLA growth model outputs a network with a more diverse distribution. A closer look into the histogram from
Figure 4-2: Degree distribution of ensemble of networks corresponding to the Computer Equipment supply chain network.

Figure 4-2(f), ignoring the hub node, reveals a smoother decrease in the frequency of the degree power-law distribution in the degree (Figure 4-3) than the decrease in the DLA-generated network for the computer equipment supply chain. The explanation for this lies in the difference in density. More specifically, since more edges are being
Figure 4-3: Degree distribution of the network generated with the DLA growth model based off of the power-driven handtools supply chain network, excluding the node with the highest degree.

added per node, then we can expect more realizations of links to be attached to those nodes who are assigned a low probability for the preferential attachment step.

4.2 Simulation Results for Targeted Node Attacks

In this section we analyze the results of the simulation of targeted attacks on nodes. Figure 4-4 and (4-5) illustrate the change of values of robustness when facing node removals in the ensemble of networks generated from the Computer Equipment and Power-Driven Handtools supply chain, respectively. The top right plot corresponds to the original network, and the other five plots correspond to networks generated by a growth model. The $x$-axis displays the fraction of nodes removed, and the $y$ axis displays the robustness metric after the removal of a certain fraction of nodes. Values where the robustness measure is greater than the original graph are excluded, since the purpose of this analysis is to spot the vulnerabilities in the system, that is, the components that would significantly worsen the network performance.
Figure 4-4: Robustness to node removal of the ensemble of networks generated from the Computer Equipment supply chain. The values for the efficiency and relative size of the LCC are displayed in blue and green, respectively. The values for these measures after targeted attacks by betweenness centrality, degree and load centrality are displayed with circles, triangles and squares, respectively. The fraction of nodes is multiplied by 50 in the $x$-axis.
Figure 4-5: Robustness to node removal of the ensemble of networks generated from the Power-Driven Handtools supply chain. The values for the efficiency and relative size of the LCC are displayed in blue and green, respectively. The values for these measures after targeted attacks by betweenness centrality, degree and load centrality are displayed with circles, triangles and squares, respectively. The fraction of nodes is multiplied by 50 in the x-axis.

4.2.1 Targeted Attacks by the Degree, Betweenness Centrality and Load Centrality

As mentioned in Section 4.1, the DLA model produces networks where there is one hub node that concentrates a large number of links, and all the other nodes have
a significantly smaller degree. We explore the correlation between the degree and other measures of centrality in this network configuration. For the betweenness centrality and load centrality, the fact that a node $s$ concentrates a large share of the links means that there is a large probability that, for random edge $(x, y)$ either of the nodes $x = s$ or $y = s$. This means that the probability that an edge in a shortest path has an end in $s$ is also considerably large. In the case of the network with one large hub, the correlation between the degree and the betweenness and load centrality is strong. This explains why all the local centrality metrics used to simulate a targeted attack yield the same robustness measures after the first attack, since the targeted attack was performed on the same node for different centrality measures.

After the first node removal, centrality measures do diverge and produce different robustness results after the removal of the next most relevant nodes. With the exception of the WS model, the robustness of all networks is largely affected by the removal of the first 10% nodes for the orig_comp, BA_comp, HK_comp, DLA_comp. In this sense, the three scale-free models were able to accurately reflect the robustness of the network. The ER_comp network generally presents a more robust alternative to targeted attacks by node than the original network.

### 4.2.2 Targeted Attacks by the Average Neighbor Degree

The behavior of the robustness metrics in targeted attacks by the betweenness centrality and load centrality different is very similar: they decrease monotonically as more components are removed from the network -with the exception of the efficiency in the WS generated networks-. They also decrease significantly in the first few removals.

Targeted removals of nodes by the highest average neighbor degree, on the contrary, display a constant change in the size of the LCC. This means that the size of the LCC is being reduced by roughly 1 when one node is removed, that is, the
removal of a node is not disconnecting the LCC into two clusters but rather just removing, one by one, nodes from the LCC that do not disconnect the structure. This discards the idea of a necessarily strong correlation between a low average neighbor degree and the relevance of a node in maintaining the connectivity of the network is not satisfied in all networks. Further research in the correlation between these two properties is needed in order to establish the conditions under which it is satisfied.

It is also worth observing that there is no clear pattern in the efficiency of the network as we perform targeted attacks on nodes by average neighbor degree: at times it increases monotonically, at other times it decreases monotonically, and it generally oscillates. We conclude that the average neighbor degree provides no information on whether a node is a source of vulnerability in a SCN.

4.2.3 Load and Betweenness Centrality

Although calculated differently, load centrality and betweenness centrality attempt to capture the same behavior of nodes as participants of shortest paths between pairs of networks across the network. The close relationship between these two measures is also reflected in the nearly identical impact of targeted attacks by load and betweenness centrality in both efficiency and size of the LCC. The only case when they displayed a different behavior was in the WS_tools network, where the values were exactly the same until \( \sim 60\% \) of the nodes have been removed. After that, the removal of the highest betweenness centrality nodes generates a slightly smaller increase in the efficiency of the network.

4.3 Simulation Results for Targeted Edge Attacks

The impact of removals by edge clustering coefficient is only observed in DLA_comp, WS_tools and DLA_tools. The reason for this is that simulations were not run after no triangles are left in the system. From that point onward, the edge clustering
coefficient is identically the inverse degree of the least connected node, and since there is no clustering remaining, this formula does not provide the same perspective on the edge's relevance in the robustness of the network. We provide an analysis of the simulations based on this centrality measure in the few networks where it was observed. The results in the WS network display a close to uniform impact, where the efficiency of the network decreases at the same rate as the percentage of edges removed. In the DLA_tools model, the network is very robust against attacks until about half of the edges have been removed, and then the efficiency of the network quickly decays until about only 20% of the edges remain, in which case it hits the end of the simulation.

With the exception of the networks generated by the WS growth model, the centrality robustness of targeted attacks seems uniform across the different networks in one ensemble. All networks display a significant change in the size of the LCC after around 20% of the edges have been removed, and stabilize at low values when the remaining number of edges with respect the original number is about 60%. In general terms, the robustness measured by the efficiency decreases more quickly than measured by the size of the LCC.

4.4 Monotonic Measures of Network Robustness

Measures of robustness recorded as targeted attacks are performed generally display a monotonically decreasing behavior, or increase monotonically until they peak and decrease monotonically afterwards [2, 12, 31]. This can be mathematically proved for several robustness measures such as the average or maximum distance between nodes in the network, the size of the LCC, and the average size or number of remaining connected components.

The removal of a node (edge) can only make paths longer, because it does not affect the distance between nodes whose shortest path did not cross this node (edge),
Figure 4-6: Robustness to edge removal of the ensemble of networks generated from the Computer Equipment supply chain. The values for the efficiency and relative size of the LCC are displayed in blue and green, respectively. The values for these measures after targeted attacks by betweenness centrality, range and clustering coefficient are displayed with triangles, squares and diamonds, respectively. The fraction of nodes is multiplied by 50 in the $x$-axis.

yet it does increase the distance between nodes all of whose shortest paths crossed it. As a result, individual distances between pairs of nodes either increase or stay the
Figure 4-7: Robustness to edge removal of the ensemble of networks generated from the Power-Driven Handtools supply chain. The values for the efficiency and relative size of the LCC are displayed in blue and green, respectively. The values for these measures after targeted attacks by betweenness centrality, range and clustering coefficient are displayed with triangles, squares and diamonds, respectively. The fraction of nodes is multiplied by 50 in the x-axis.

same, and so do the average and maximum distance across the network. In terms of the connected components of the network, the removal of a node (edge) can either
disconnect a cluster and create two or more strictly smaller clusters, or leave the same number of clusters and decrease (preserve) the size of the cluster it used to belong to. Therefore, all measures mentioned above regarding the LCC necessarily display a monotonic behavior with the loss of components in the network.

This is not always the case with the average inverse distance, in fact, it is not the case with some of the simulations performed (Figures 4-4 and 4-5), where the value of the efficiency increases in the networks produced by the Watts-Strogatz growth model when facing targeted node removals by betweenness centrality and load centrality. The efficiency is a ratio between a sum of inverse distances and (roughly) the square of the number of nodes in the network $N^2$ (see Equation 2.1). Every time a node is removed, the denominator is reduced by a factor of $\frac{N^2}{N-1}$, which is fairly large ($\sim 0.95$ for values of $N > 20$ (Figure 4-8). The efficiency in $WS\_comp$ starts growing when nearly 40% of the nodes have been removed, and then starts growing much faster after about 80% of the nodes have been removed. The efficiency in $WS\_tools$ starts growing when nearly 60% of the nodes have been removed, and then starts growing much faster after roughly less than 90% of the nodes have been removed. The turning points indicate a moment when the numerator (the sum of average inverse geodesic distances) is increasing, which means that the distances among the graph are significantly being reduced.

With the removal of a node $s$, the inverse distance between two nodes $i, j$ decreases if $s$ was part of all the shortest paths between $i$ and $j$, stays the same if there were more than one shortest paths between nodes $i$ and $j$ or if node $s$ was not part of the unique shortest path, or become exactly 0 if $s$ was $i$ or $j$ or all paths from $i$ to $j$ used to pass through $s$. Prioritizing node removal by the highest load or betweenness centrality directly targets those nodes who are part of most shortest paths, and therefore one would expect that the numerator significantly decreases. The explanation for this behavior might lie in the a potential redundancy of shortest paths in the networks produced by the Watts-Strogatz model.
Figure 4-8: Growth of $\frac{N-2}{N-1}$ with respect to $N$. This is the ratio that the denominator of the global efficiency is multiplied by when a node removal is performed (Equation 2.1).

This counter-intuitive behavior of the efficiency raises the question of how effective it is in displaying the robustness of the network. Although we consider it superior to the maximum or average distance because it captures the pairs of nodes that are not reachable from each other, it might not be the most suitable measure to record changes in robustness as nodes are removed from the network.
Chapter 5

Conclusion

Previous works that study the robustness of the SCNs from a topological perspective focus on quantifying the robustness using a limited set of global measures, or simulate the loss of components as random failures or targeted attacks by highest degree, focusing primarily on node removal. In this thesis, we assess the robustness of networks to targeted attacks by simulating node removals prioritized by degree, load centrality, betweenness centrality and average neighbor degree, and by simulating edge removal prioritized by range, clustering coefficient and betweenness centrality. We perform this analysis on two real-world supply chain networks, and ten networks generated by five different network growth models, and parametrized by basic topological properties of the two real-world data sets. Our results show that nodes with high values of degree, betweenness centrality and load centrality are crucial to maintaining the connectivity and efficiency of the network. The relevance of nodes with high average neighbor degree in the efficiency of the network is still unclear, we found no clear correlation between the relevance of a node in maintaining a low efficiency for the graph. Furthermore, the removal of nodes with a high average neighbor degree does not impact the connectivity of the network measured by the relative size of the largest connected component with respect to the initial number of nodes.
We also evaluate the accuracy of various network growth models in resembling complex topological properties (such as global clustering coefficient and degree distribution) of a real-world SCN when parametrized with simple topological properties (number of nodes and links) of the same real-world network. We concluded that SCNs should not be modeled as random networks, but rather as scale-free networks. In particular, we found the best model to describe the emergence of SCNs to be those with preferential attachment by degree.

5.1 Collective Behavior for Supply Chain Robustness

In Section 4.1 we discussed how few network growth models were able to reproduce a degree distribution that resembled the ones real-life supply chain networks’, and how none of them captured them very accurately. From the perspective of using network theory to accurately describe the emergence and organization of supply chain systems, there is still room for future work. However, one thing to observe is that most growth models produced networks with higher robustness to node and edge removals in almost all forms of targeted attacks. Whatever the conditions that generated the two real-world supply chains have been, the growth patterns described by most network growth models considered in this study are able to produce more robust configurations.

Even though there is usually little room for changing the overall structure of a SCN from the perspective of a decision maker (such as a supply chain manager), collective awareness of different models capable of generating more robust structures can contribute towards better supply chain design. As an example, we have seen that the BA and HK model outperform the DLA model in robustness, which suggests that a general tendency for new firms towards maximizing relationships with the
most highly connected existing firms might generate a more robust SCN structure than a general tendency for new firms to reduce the length of the supply chain.

5.2 Future Work

As mentioned earlier, a SCN can be modeled as a directed or undirected network depending on whether the relationships between the entities are considered unidirectional or bidirectional. This work models the real-world supply chains under consideration as undirected networks, and analyzes the robustness of networks generated from growth models as undirected networks as well. We suggest future works to be directed towards modeling complex supply chain systems as directed networks, and to also analyze the contrast between measuring robustness with respect to directed versus undirected paths.

Finally, we acknowledge that relationships between entities in a supply chain are more complex than we have represented them in our methodology. For instance, the data we use as a case study provides information on the demand, the capacity, the cost and the time delay of each process that we represented as a network. We suggest that the robustness of SCNs should also be assessed by considering local and global measures that incorporate such numerical properties that are intrinsic to the underlying systems.
Bibliography


