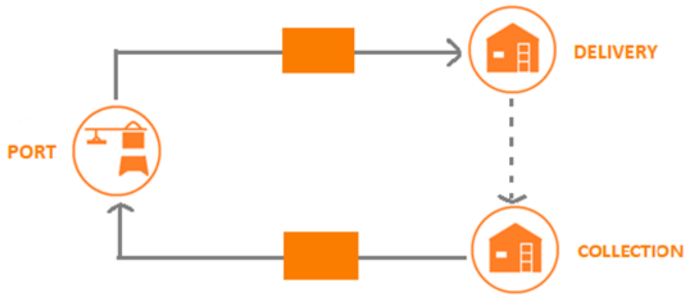
**Erasmus University Rotterdam**

**MSc in Maritime Economics and Logistics**

**Christos Fotis 590391**

26/04/2022




# Network optimisation based in order matching

## Acknowledgements

I would like to thank my supervisor Rob Zuidwijk for providing me with a very interesting topic and valuable feedback regarding the content of my thesis. On a personal note, I want to thank my friends for being there for me, especially Lydia Kakampakou and Electra Kaltsoni who assisted with the language and the academic outlook of the thesis. I also wish to thank my girlfriend for supporting me emotionally and motivating me during the whole stressful period of the thesis. Finally, I owe a great debt of gratitude to my mother who supported me financially and emotionally during the whole Master.

## Abstract

The current study aimed to investigate whether and to what extent the quality of a network platform can be quantified. The analysis resulted to the construction of a model quantifying the quality of a network platform and additionally proved that in the event of two hauler companies willing to join a network, if one is willing to collaborate with all the members while the other only works together with just one other member of the network, the best option would be the company willing to collaborate with everyone. Of course, this would be the case if both companies are of the same size, otherwise the model should be used to identify the best option for the network. A company with a scale of 30 orders that is unwilling to collaborate with the other members of the network is still a better addition compared to a company that is willing to collaborate with everyone but only has a size of 10 orders. Finally, it was also proven that when a company is willing to work only with one other member , then the optimal choice is to coordinate with the least connected member of a network. The model used in this thesis, quantify the quality of a network based on the calculation of the expected maximum matchings, the higher the number of matchings the bigger the cost reduction meaning the higher the amount of money generated/saved by the network effect , which represent a better quality network .

## Table of contents

## List of Tables



## List of Figures

# 1. Introduction

The economical advances of the past few decades caused a pressing need for international collaborations. The benefits of working collaboratively in the shipping sector are numerous, especially when it comes to ordering transportation, and can lead to improved management of the sector as well as cost reduction associated with order shipping and transportation. Such benefits are even more valuable when the economy relies heavily on the shipping industry, as even small collaborations can save the industry a significant amount of money. Collaborative efforts which come in many forms could collectively help reduce the associated transportation cost. However, for the purpose of this thesis, the focus will be placed on the general category of the so-called *vehicle routing* problems and, in particular, on the problem of *order matching*.

Let us consider the situation where a truck hauler receives import and export orders. When an import order arrives at a seaport, the truck hauler must pick up a container from the port and transport it to the consignee's distribution centre (see Figure 1, top plot). Conversely, in the case of an export order, the truck hauler must pick up a container from the consignor's distribution centre and deliver it to the seaport (see Figure 1, bottom plot). The matching of an import and an export order (see Figure 2) requires compatibility of the import and export location, often referred to as *triangulation* (i.e., the two locations must either coincide or be nearby), as well as the alignment of the time window of delivery for the consignee and the time window of pickup for the consignor. Successful matching would result in cost savings due to the combined execution of the import and export order. Such savings reflect the difference between executing the orders separately, as opposed to executing the orders combinatorically.

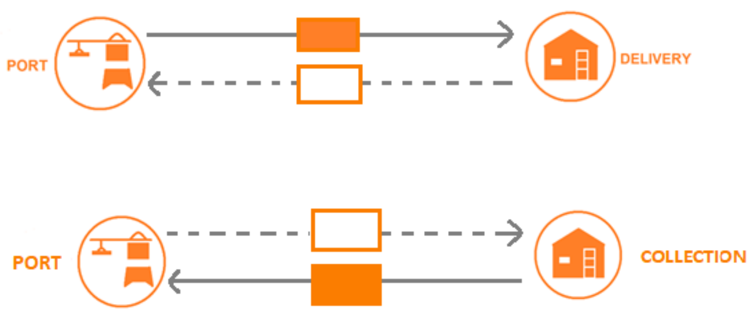



*Figure 1: Traditional order transportation*

Order matching is crucial for reducing transportation costs in a business network. During the past few years, matching strategies have gained an ever-increasing popularity. Given the fact that a business network desires to minimize cost in any case, strategies that can contribute to this effort play a significant role in ensuring the prosperity of the network as well as in encouraging its further future development. The undermentioned research question aims to contribute to the research of existing cost-reducing strategies.

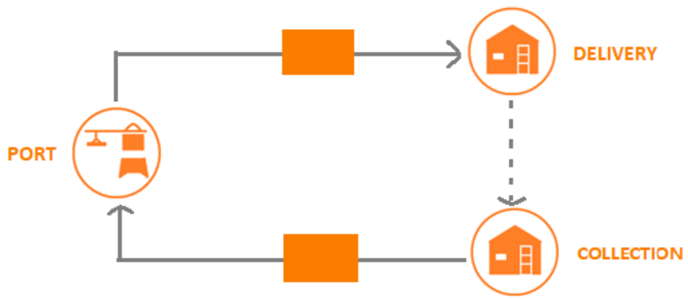



*Figure 2: Order transportation with matching*

> This thesis is assessing the value of a network based on order matching.
>
> The research question was formulated as to conclude which addition would be optimal for the expansion of a network. Between two hauler companies both willing to join a network, which one would be the most beneficial addition for the network: the company willing to collaborate with everyone already existing in the network or the company that demands to be affiliated only with one specific member of the network?
>
> This scientific inquiry naturally leads to two hypotheses that should be further explained and answered in the following segments, in order to collect data and satisfy the research question. The answers given are sufficiently well-documented and the results of the experiment contribute to this research topic in general.
>
> Finally, for the purpose of assessing the value of a network, the creation of a quantitative model was considered of major importance.

Thus, in order to answer the research, question an optimization problem should be addressed. More precisely, there is a need to estimate which of the aforementioned hypotheses will give rise to a higher expected maximum number of favourable matches. Extensive research on the problem of order matching has resulted in myriads of algorithms designed to tackle the matching problem. The expected maximum matchings were calculated with the use of random bipartite graphs (Janson, Rucinski., & Luczak, 2011) (to be defined in the following sections). The expected maximum matching for two different hauler companies in relation with two different network settings was particularly studied, and subsequently the business networks as a whole were simulated using Monte Carlo methods in MATLAB (Leonid Peshkin, 2007).The results of both scenarios were rigorously compared in order to decide on the most beneficial company setting.

The subsequent sections of this thesis are structured as follows. Section 2 provides a thorough literature review of the use of order matching in business networks. Section 3 describes in detail the conceptualisation and the theoretical setting of this research. Section 4 focuses on the particular methodology and the associated technicalities involved in the algorithmic part of it, and Section 5 presents my key findings and results. The thesis concludes with a brief discussion on the topic under study and provides also further recommendations for future research.

# 2.Literature Review

## 2.1 Business networks

In the modern highly competitive economic environment, companies are networked and clustered in various types of cooperative formations so as to maintain or enhance their competitiveness. These actions stem from the awareness on the part of entrepreneurs that each company individually is unable to face the growing competition from both domestic and foreign companies. The literature on networking structures has led to the establishment of definitions that cover a wide range of different business structures, depending on the characteristics of the participants, the goals of networking, the operational structure, etc. The variety of conditions prevailing in each networking initiative obscures the boundaries between the various theoretical structures, often creating confusion. The following are definitions of the dominant networking categories, Clusters and Business Networks, and a basic subcategory of networks (Molina-Moraleset al., 2015).

### 2.1.1 Clusters

Clusters are defined as the geographical concentration of interconnected companies and bodies (of institutional nature), which are associated with common technologies and capabilities. They are usually located in a geographically limited area facilitating communication between companies, the movement of goods and raw materials and allowing interpersonal contacts (Ballandet al., 2016).

### 2.1.2 Business Networks

Business Networks are groups of companies, which work together to achieve specific goals, and the results of this activity will have a recognizable and measurable impact on their members. They have a limited number of members who have agreed to work together in some way to achieve specific business goals, which will likely lead to enhanced competitive advantage and/or mutual economic benefit (Belso-Martínezet al., 2017).

According to the "Conclusions of the Literature Review and Presentation of Good International Practices" of the Institute of Small Businesses of the Hellenic Confederation of Professionals, Craftsmen & Merchants (IME GSEVEE) it is stated: Today, a common and accepted definition of clusters is formulated by the EU according to which the "clusters are groups of independent companies and related bodies that:

- cooperate and compete,
- are geographically concentrated in one or more regions, although the cluster is likely to have international dimensions,
- specialize in a specific field - business sector and are interconnected by common technologies and capabilities,
- their interest may be based on scientific research e.g. scientific research and development that contributes to the creation of new products, production processes and is based on mixed forms of cooperation between companies and knowledge production institutions (such as universities and research centers) or is of a conventional nature where joint promotional cooperation actions are developed as well as promotion, supply, development and utilization of productive infrastructure, etc.,
- they can have either a formal form (i.e. they can also have an institutional nature) or an informal form".

Clusters and business networks are concepts that are often seen as synonymous. This is because networks are a key structural element of cooperative formations. The existence of networks within the clusters facilitates the cooperation in various issues such as education, financing, technological development, product design, marketing, exports and distribution (Piperopoulos, 2016).

The key differences identified between clusters and business networks are as follows:

- Clusters consist of a much wider range of businesses compared to networks.
- Clusters have a specific geographical designation as opposed to networks.
- Clusters have a broader strategic goal than networks. Many different networks at different levels can operate in parallel within a cluster.

However, the main difference between business networks and clusters is that the former is not purely aimed at strengthening businesses, but is mobilized to achieve broader goals related to regional and national economic development for the benefit of society as a whole. The predominant benefits of business networking are achieving economies of scale, access to new markets, promoting innovation, specialization, increasing brand name and competitiveness (and the region), and flexibility in adapting to market changes" (Buchnea, 2016).

In more detail, the differences between the two concepts are shown in the table below:

Table 2.1: Differences between Clusters and Networks

| **Clusters** | **Business Networks** |
|---|---|
| Usually, member companies are neighbors | There are no geographical restrictions |
| The more members the better | The number of members is determined from the beginning and usually does not change |
| The participation of the members is equal | Members' participation is not always equal |
| Relationships are flexible | The relations of the members are defined |
| They are a separate entity that is evolving (new company) | They are the activity of the companies that make them up |
| Member companies usually have competing products / services | Network member companies usually have complementary products / services |

| | |
|---|---|
| They encourage the provision of specialized services in a specific geographical area | Networks allow the development of specialized services at a lower cost |
| They may contain business networks | They may not contain clusters |
| The objectives may be varied and yet different from the companies that make them up (e.g. serving the common good) | The goals of the networks are similar to the goals of the companies that make them up |

Source: IME GSEVEE (2013) p. 4

In the present study, the term networking is used to refer to the structures of business alliances, on one hand because it expresses the networks of Small and Medium Enterprises (SME) (small structures with business orientation) in a better way, and on the other hand, because networking refers to the general process of cooperation and concluding alliances that take place in the creation of networks and clusters.

### 2.1.3 Main subcategory of business networks

According to the "Study for the sectoral specialization of business networking actions" of the Special Secretariat for Planning and Implementations of the 3rd CSF, the main subcategory of business networks is based on their degree of verticality and includes the following three types (Nyström and Mustonen, 2017):

1. Vertical networks: members develop some degree of specialization in a specific sector of the production chain. These networks are based on input-output relationships, where each member develops a specific specialization in its field by serving the efficient production and distribution of a specific product category to the final markets.

2. Horizontal networks: consist of companies that produce the same or similar products and compete with each other. They are usually recommended for the joint promotion, research and development of new products, the joint supply of raw materials, etc.

3. Complementary or diagonal networks: consist of companies that do not compete with each other, nor are they connected to each other in a production chain. They are usually recommended for the promotion of common interests of different business sectors, for the formation of integrated packages of products and services, the creation of information centers, etc.

### 2.1.4 Process of implementing networking initiatives

In general, the process of implementing a networking initiative is divided into the following three stages:

1. Networking: at this stage there is the launch of the initiative by organizations or entrepreneurs, the recognition of candidate member companies, the awareness and commitment of members, the creation of the network, the definition of the coordination structure, objectives, action planning, financing and mode of operation.

2. Network operation: at this stage the network operates normally, joint training actions, implementation of new products, advertising, etc. are promoted.

3. Strategy review: the course and effectiveness of the network are evaluated, as well as the new perspectives and opportunities and if necessary, its course is redesigned by promoting new actions, expanding or reducing the number of members, or even terminating the program.

The predominant problems that arise from business networking have to do with how the structures work and their efficiency. The operation depends on the correct choice of a number of factors, such as the structure of the coordination group, the number of members, the stringency of the terms of participation, the financing model, the required contribution of the members, the cooperation style, the relations with external bodies etc. Respectively, the effectiveness of the networking depends on the degree of commitment of the members, the level of trust, the effectiveness of the cooperation, the correct planning of the actions, the capacity of the coordinating body/facilitator, the existence of a critical mass of companies, etc. (Silvaet al., 2017).

The commitment of the members for a systematic collective effort is the cornerstone of the matter and is strengthened by the concentration of the effort in dealing with specific problems faced by the companies and the creation of a common orientation and goals, the emphasis on the correct formulation of the message for the multiplier results brought by the collaboration, the reference to specific successful examples, the creation of a climate of esprit de corps (joining forces) since “Others do it and it succeeds, why not us”, and finally the creation of training programs on group management and group work, so that they can gradually begin to develop codes of communication and cooperation with each other.

Undoubtedly, the establishment of a clear organizational structure and mode of administration in advance can significantly help the networks to adopt clear rules and methods of cooperation from the beginning, so that problems do not arise due to lack of communication. The provision of a central governing body and regional - independent audit bodies, which will be elected by a majority of the members, largely ensures common acceptance of transparency at all levels of operation as well as the greatest possible flexibility (Turba, Breimo and Lo, 2019).

Best practices indicate that prior to the adoption of a network financing program, actions have been taken to identify and record “emerging networks” that meet the requirements for the development of partnerships (Cluster Mapping).

### 2.1.5 Benefits of setting up clusters / networks

Regarding the benefits from the operation of clusters/networks for companies, according to the “Conclusions of the Literature Review and Presentation of Good International Practices” of the Institute of Small Businesses of the Hellenic Confederation of Professionals, Craftsmen & Merchants (IME GSVEE) these are the following in detail:

- Increased levels of expertise: Networking helps businesses know in depth their supply chain, which in turn can contribute to inter-corporate training and collaboration.
- Promotion of new products and services. Networking helps the participating companies to design products and services together at such a level of supply that it would be impossible for them to achieve it as individual companies.
- Opportunities for economies of scale. Through stronger specialized production in each cooperating company, through the consolidation of their supplies so that they can achieve greater discounts or through the consolidation of the markets to which they are addressed.
- Increased productivity: Through the provision of specialized resources and access to information and knowledge, businesses can increase their productivity.
- Faster access to innovation: Through specialization and complementarity of resources and skills as well as collaborations with research institutes and universities, companies have faster access to innovation.
- Facilitation in the introduction of new technology: Through partnerships and the dissemination of knowledge, companies are immediately informed about new technologies and are encouraged to adopt them.

- Internationalization: The pooling of resources and capabilities of companies allows the penetration of foreign markets that individually each company could not achieve.
- Attracting demand: Customers and suppliers are attracted by the high concentration of companies in the same industry in clusters.
- Encouraging new entrepreneurship: New businesses are established within clusters that focus on selected markets (islets) or activities. These companies in the first period of their operation receive support from the cluster in supporting parts of their operation.
- Encouraging the relocation of existing businesses: Clusters often create local benefits for their members that are difficult for distant competitors to overcome. Thus, they become poles of attraction for existing companies that move their activities within the cluster's geographical boundaries.
- Strengthening social and other ties: Which in turn can lead to the birth of new ideas and the creation of new business activities.
- Improved flow of information: The members of a network have the opportunity through their cooperation to have more reliable information about the changes taking place in the market such as consumer trends, new emerging markets, consumer/supplier behaviors, etc.
- Possibility to create infrastructure: Networking helps to create common infrastructure for professional, legal, financial and other specialized services.

However, there are benefits from the operation of clusters/cooperative formations that have been also identified at the level of local communities and regions:

- Local economy development: Local communities can leverage existing business potential to attract investment.
- Promoting collective learning and innovation: Combined knowledge contributes to promoting collective learning and innovation
- Reducing unemployment: The workforce of local communities has the opportunity to work in networking/cluster companies.

On the other hand, in certain conditions, clusters can be an inhibitory factor for the further development of their members. For example, in an ever-changing technology environment, networked businesses become more vulnerable if they persist on using outdated practices and technology or if they do not develop the flexibility needed to respond to these changes (del-Corte-Lora, Vallet-Bellmunt and Molina-Morales, 2015).

In addition, when clusters rely on a few buyers or one activity of one large or a few companies, as is usually the case, they may fail if one of the above factors disappears, even if they remain competitive. To avoid these pitfalls, small and medium-sized enterprises should develop flexible skills and adaptability, build collaboration opportunities and participate in knowledge sharing events on an ongoing basis.

However, it should be noted that whether the aforementioned advantages really exist has not been studied with the help of approved research models. Until recently, the majority of studies on clusters had been limited to providing theoretical explanations for their performance. Most of the time, these studies were based on the observation of successful clusters. Thus, the current situation regarding the creation of networks of both successful and failed examples for the further accumulation of valuable and applicable knowledge should be studied (Belso-Martínez, Mas-Tur and Roig-Tierno, 2017).

## 2.2Horizontal container collaboration/ Container transportation

Road transportation is a very competitive market presently. The competition between companies in this market is driven by the increasingly demanding standards of consumers who require prompt, accurate, and timely delivery of their orders at the lowest price possible. However, suppliers are often faced with difficulties when attempting to satisfy such demands individually, a phenomenon that highlights the pressing need for inter-business collaborations.

Collaborations of many sorts could facilitate businesses to optimize their functionalities and could result in the elimination of their supply hurdles. Three main types are typically encountered: vertical, horizontal, and lateral collaborations. According to Simchi-Levi et al. (2000), vertical collaborations refer to a "set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses, and stores, so that merchandise is produced and distributed in the right quantities, in the right locations, and at the right time, in order to minimize system-wide costs while satisfying service level requirements." This definition implies that vertical collaborations aim at installing beneficial partnerships among multiple parties operating at different levels of the supply chain, thus intending to avoid unnecessary logistics costs. Horizontal collaborations, on the other hand, involve partnerships between businesses operating at the same level in a market. Such partnerships involve sharing of private information, facilities, or resources so as to reduce costs or to improve their services.

Ultimately, lateral collaborations involve a mixture of the previous two types. Although vertical and lateral collaborations have been extensively documented in the existing literature (eg for vertical Saenz: 2014 and Danloup: 2013, and for lateralTavasszy: 2003), horizontal collaboration is both documented and practically implemented very scarcely. In fact, most practical examples of horizontal collaboration belong to the maritime shipping and aviation (e.g. Skyteam, Star Alliance) industries, the principal reason being that such partnerships can help overcome operational hurdles related to different cross-border traffic rights. Landside transportations, however, do not share the same characteristics of the aforementioned industries and, therefore, the extension of the respective horizontal collaboration strategies to landside logistics is not as straightforward as one would hope.

Literature written on horizontal collaboration in logistics on the landside is quite limited. Nevertheless, companies have slowly begun to realize the potentially powerful benefits of such collaborations and show increasing interest in formulating and/or joining already existing horizontal collaboration networks. In fact, the advantages are numerous and one could argue highly tempting (Cruijssen et al: 2007a). To name a few, successful collaboration can lead to less routes, les CO2 emissions, less road congestion and, of course, a significant cost reduction for the hauler companies. However, the process of a company toassess whether joining such networks would be in its interest or if and by how much collaboration with existing companies of the alliance would be beneficial is not a trivial matter. Such decisions require deep knowledge of the strategic and organizational capabilities of the potential partners and such information is usually not readily available or easily estimated. A key role in assisting businesses in making such decisions can be played by appropriate logistics service providers (LSPs). For such partnerships to be successful and yield a business performance greater than the one that could be achieved by each firm individually, they must be based on mutual trust, openness, shared risk and shared decision making. However, it is important to note that even with the use of a neutral third party doing the matching, many barriers, primarily related to trust issues, still exist for a full collaboration between everyone in the network (Cruijssen et al: 2007b). This represents an additional issue for the healthy advancement of the network platform, related to the characteristics of the members. More detailed information on the role of trust in the development of horizontal collaborations can be found in Das and Teng (1998). All of the aforementioned considerations will be taken into account in the thesis project under study which aims to provide some insight in the practical implications of such considerations.

## 2.3 Horizontal platform, network effect of platforms

The rapid development of markets around the world and the exponentially growing demands of customers have forced businesses in the supply chain to try to maximize their efficiency, thereby minimizing the cost. However, cost reduction is a painstaking and intricate endeavour which, in complex business environments such as the ones encountered nowadays, can only be achieved individually by a firm to a small extent. On the contrary, collaborative efforts towards cost reduction, such as the formation of horizontal collaboration networks, could lead to much more significant results (Cruijssen et al: 2007b). Unfortunately, neither the creation and maintenance of such networks is straightforward, nor their good function is guaranteed. In fact, such networks are usually so complicated that can only be managed by appropriate LSPs who possess the necessary skills and expertise on the field. Hence, most companies willing to join such networks use third-party LSP services. The rapid growth of the internet and internet-based services have led to most LSPs taking the form of online platforms (Zhao: 2010) which essentially operate as a business-to-business (B2B) electronic environment (e-hub).

Online platforms have been used extensively so as to bridge the gap between suppliers and consumers and, at the same time, to help achieve matching of the needs of two or more sides of users (Loux: 2020). Well known examples are social media platforms (which bring together users, advertisers and content developers e.g. Facebook, Twitter, Instagram), online marketplaces (which unite sellers, buyers and advertisers e.g. Ebay, Amazon), food delivery platforms (which draw together drivers, riders or customers and merchants, e.g. Wolt), online dating platforms (which gather individuals e.g. Tinder), recruitment platforms (which connect firms offering jobs with job seekers e.g. LinkedIn), the list goes on.

Online platforms such as the ones mentioned above can be categorised into the broadest categories of business-to-business (B2B: connects organizations with other organizations), customer-to-customer (C2C: connects various individuals), or a combination of the two i.e. business-to-customer (B2C: connects organizations with individuals) platforms. In the case of landside transportation logistics, which is the focus of this thesis, B2B networks are of the greatest interest. Further details on B2B networks and their characteristics can be found in Kaplan (1999)

One such initiative is the Boxreload (www.boxreload.com), created in 2015 by the PARIS Optimal Transport Planning division of Hutchison Ports with the support of Erasmus

University of Rotterdam. It is a web-based platform functioning by matching an import to an export order and vice versa, resulting in improved truck and driver utilization by using only one truck, one driver, and one container for the transportation of two orders from different companies. The order matching is achieved by a neutral real-time automated planning engine which, after finding the optimal matchings, sends to the respective companies the reload opportunity. If both parties agree, the collaboration will materialise. Furthermore, in order for the automated engine to consider the possibility of matching between two companies, both of them must have already agreed that they are willing to cooperate. Even with the use of a neutral third party so as to do the matching, many barriers still exist for an all-inclusive collaboration between everyone in the network (Cruijssen et al. : 2007a and 2007b).

## 2.4 Matching in mathematics (random bipartite graphs)

As briefly mentioned in the introductory section of this report, answering our research question, i.e., finding the maximum expected number of matches under the different scenarios that were studied ultimately involves solving an optimization problem. For the purposes of this research thesis, so-called *random bipartite graphs* will be employed to arrive to a solution to the maximum matching problem. Let us first describe the matching problem. Consider a business network consisting of a number of import orders and a number of export orders. We are interested in finding the maximum number of favourable matches that can be made in the network given the available import and export orders. That is, we wish to find the maximum number of orders that can be combined and executed together in a single journey of a hauler truck instead of being executed separately. This is called the maximum matching problem and can be solved using a bipartite graph (Janson, Rucinski.,&Luczak, 2011).

Without loss of generality, let the import orders be presented in a column and the export orders be presented in a different column; say imports are presented in the left column and exports in the right column of the bipartite graph (hence the name bipartite, i.e., it is formed by two parts-columns). For given vertex sets $N = \{1, \dots, n\}$ of import orders and $M = \{1, \dots, m\}$of export orders, an edge $(i, j)$between an import vertex $i$ and export vertex $j$ represents that the matching between the corresponding import order and export order is favourable. In this manner, a set of edges between the node$E \subseteq N \times M$ is established. This gives rise to a bipartite graph $B(N, M, E)$ with vertex sets $N$ and $M$ and set of edges $E$. This bipartite graph can have one or more maximum matchings. Because the number of import and export orders is random, the edge between a given import vertex and export vertex is not deterministic but exists with some probability $p$. A graph of which its edges exist with a certain probability is called a random graph (Gilbert 1959). We will denote such a random graph by $B(N, M, p)$. Note that only the sizes of the vertex sets are relevant, so we may identify $B(N, M, p) = B(n, m, p)$ with $|N| = n$ and $|M| = m$. Let us also denote by $k = k(E)$ the number of edges in the graph and by $\mu = \mu(E)$ the size of a maximum matchings. Then, the expected maximum number of matches of a random bipartite graph $B(n, m, p)$ is given by the polynomial

$$R_{n,m}(p) = \sum_{E \in B(n,m,p)} p^k (1-p)^{nm-k} \mu. \ (1)$$

It is too straightforward to generalise the above-mentioned theory for multiple haulers in a business network and, hence, answer the research question under study. However, as $n, m$ increase the sum in (1) becomes overly complicated, consisting of too many terms, thus, making calculations demanding and time-consuming.

A commonly adopted alternative in practice is the use of Monte Carlo integration to approximate the sum above. In statistics, Monte Carlo essentially means simulation. Monte Carlo methods are used primarily to evaluate integrals (for random variables) where analytical solutions are either not possible or extremely laborious. As a procedure, Monte Carlo integration involves simulating a large number of times the quantity of interest and subsequently approximating its expected value with the simulated sample mean. The theoretical foundations for such approximation lie on the law of large numbers, according to which the sample mean tends to the theoretical mean of a population as the sample size increases (tends to infinity). Hence, we can use Monte Carlo integration to estimate the expected maximum number of matchings, by simulating the number of maximum matchings in our network a large number of times and calculating the mean value of the simulated sample. In our case, it was found that adequate convergence occurs after approximately $10^5$ simulations (Kiwi and Loebl, 2002).

A random network is formed by a set V = {$V_1$, $V_2$,…,Vn}, n vertices, to which we add edges at random. There are several ways to randomly select the edges that lead to different models and to different probability distributions in the resulting graphs. Model $G$ ($n$, $M$) Corresponds to an equal probability of all graphs having exactly M edges (0≤$M$≤$N$), where $N$ = $n$ 2. Since the aim is to form a graph with M edges, which are from the existing N, so there are $MM$ ways to select these $M$ edges, which are equal probabilities. So, the sample space ($n$, $M$) contains $NM$elements each with $NM$ ◌̑ ৡ probability M, almost always, it is a function of

$n$, i.e.M = M($n$) and then the model is symbolized as G(n, M$n$ ). This model was originally defined in the work "On Random GraphsI" (1959) byErdősandRényi (Hence the name) (Waclaw et al., 2008).

**The Gilbert model (($n, p$))**

In the $G$ ($n$, $p$) model it is considered that each of the $N = \binom{n}{2}$possible edges that can connect two of the $n$ vertices, is selected with probability $p$, where 0 <$p$<1.

Equivalently we can consider that we have the complete graph $K_n$ and from this we delete its edges with probability q=1−$p$.

A $G_0$ element in the sample space ($n$, $p$) that has $n$ vertices and $m$ edges have a chance of occurring

$$P(\{G_0\}) = P(G = G_0) = p^m.(1 - p)^{N-m}$$

If$p = \frac{1}{2}$then ($n$,$\frac{1}{2}$) Is the sample space that contains all graphs of $n$ vertices $G^n$ with the same probability

$$P(G = G_0) = \left(\frac{1}{2}\right)^N$$

This model was proposed by Edgar Gilbert in his work "Random Graphs" (1959).

**The $G^n$sample space**

Consider a random graph to be formed as follows:

We start with the graph $n$ vertices without edges ($G_0$).

We bring an edge by joining two of the vertices, so the$G_1$ is formed (graph with an edge). We continue, adding an edge at each step, forming the sequence of graphs:

$$G_0, G_1, G_2 \ldots\ldots\ldots G_N$$

Where $G_t$is a graph with exactly $t$ edges.

There areN! such sequences that make up the $G^n$ sample space.

Bollobas (1985) found that the three spaces we described are closely related.

**The distribution of degrees in G ($n, p$)**

If a vertex has a d degree, then it means that d from the n-1 strong neighbors, selected in $\binom{n-1}{d}$ ways, were associated with a $p^d$ probability and the rest were not associated with $(1-p)^{n-1-d}$ probability. The probability of this happening is:

$$P(d)\binom{n-1}{d}p^d(1-p)^{n-1-d}$$

that is, the sq.m. which measures the degree of a vertex follows a binomial distribution with an average value of $d = (n-1)p$ and $\sigma^2 = (n-1)p(1-p)$ dispersion.

If we now let n tend to infinity, then normal tends to Poisson,

$$P(d) = e^{-(n-1)p}\frac{\left((n-1)p\right)^d}{d!}$$

Conversely, if we consider a random network in which the vertex degrees are equal and independent distributions Poisson (λ) with λ=n(p-1) (i.i.d. Poisson (λ) distributions), then it is found that this is equivalent to the Erdős – Rényi model of the n tendency to infinity. For this reason, these models are also called Poisson models.

**Properties of the G (*n, p*) model**

What are the properties (topological characteristics) of the G (*n, p*) model?

- Is it binding?
- Does it have circles or is it a tree-forest?
- Is there a correlation between neighbors's degrees?
- Which vertices play a central role?
- Are there any cliques?
- There are communities;

Many of these properties appear suddenly as the probability p increases from 0 to 1. There is a characteristic value, a threshold, from which a property appears.

**Threshold function for sub-graphs**

If $Q$ the property of a random graph contains a sub-graph with $k$ vertices and $l$ edges, it turns out that the function $p_Q(n) = cn^{-\frac{k}{l}}$

is a threshold function.

# 3.Research methodology

Many business initiatives in order matching have taken place as mentioned in the literature review. Those initiatives usually take the form of network platforms with their main focused being the growth of the network in order to have more efficient matching and higher collaborative profits. Nevertheless, a network is usually burdened with some additional costs for every new member that enters it; this observation emphasizes the importance of including quality members. Based on this knowledge, it is crucial to quantify and research what can be considered a quality member for different networks(Kothari, 2004). In this segment, the methodology that was used to answer the research question is presented.

For this thesis, two instances were created that correspond to two different networks in order to get a wider range of results and to make the research more robust. Both networks have the same amount of total orders, but both networks consist of a different amount of companies as well as a different number of connections (i.e collaborations) between the companies. One new company was added in both instances representing the independent variable while the expected maximum matchings represent the dependent variable. The maximum matchings were calculated with the use of a MATLAB code of Dr. Leonid Peshkin (2007) that constructs a (non-weighted) maximum cardinality matching. With the aim of better anticipating the following experiments necessary constraints can be found in the appendix.

The experiments commenced with the first hypothesis, calculating the expected maximum matchings for a new company that is joining the network and is willing to work together with all the existing members. Those calculations were made for both instances and to further research the quality of a network I took five different sizes (number of orders) of each company. Afterwards I proceeded with the second hypothesis which simultaneously focused on two aspects of the independent variable. The first one being the size (number of orders) and the second one being the connectivity of the added company in relation to the existing network. This was achieved by creating a 5x5 matrix in which the vertical axis included five different scales of the newly added company, whereas the horizontal axis, included five different companies that the new company is willing to collaborate with. It is important to mention that on the contrary of the first hypothesis, in the second one that involve two matrixes (one for each Instance) the independent variable was only connected to one other member of the network every time. This experiment also clarifies the impact of when a company is  only willing to work together with a specific member of the network as opposed to a company that is willing to collaborate with all the other members. After calculating the expected maximum matching for both instances and the two new companies, the answer to which of the two companies would be better for each networking platform(instance) becomes more perspicuous (Goddard and Melville, 2004).

With regards to the first instance, the network consists of five trucking companies {A, B, C, D, E} with A, B, C, willing to collaborate with everyone in the platform, whereas company D only cooperates with B and company E only cooperates with A. All of the above-mentioned companies have the same size of orders -namely-20 orders each- and the network has a total of 100 orders. Ten of the orders are import while the other ten are export.

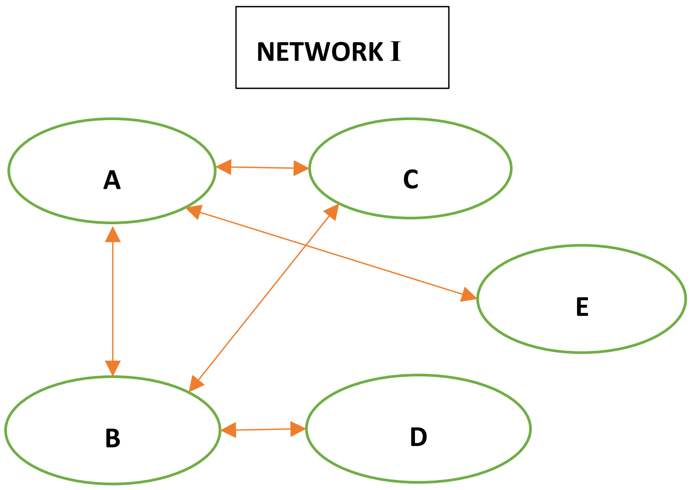



*Figure 1. Network (1)*

The second instance is a network including ten hauler companies {a, b, c, d, e, f, g, h, i, j} in which companies a, b, c, d and e are willing to work together with everyone, while company f is only willing to work with a and b, company g only with a, company h only with c, and d, company I with e and, finally, company j is only willing to cooperate with b and e. All of the companies have the same size of orders (namely 10 each) which makes the network have a total of 100 orders. For each company, five of the orders are import and the remaining five are export.

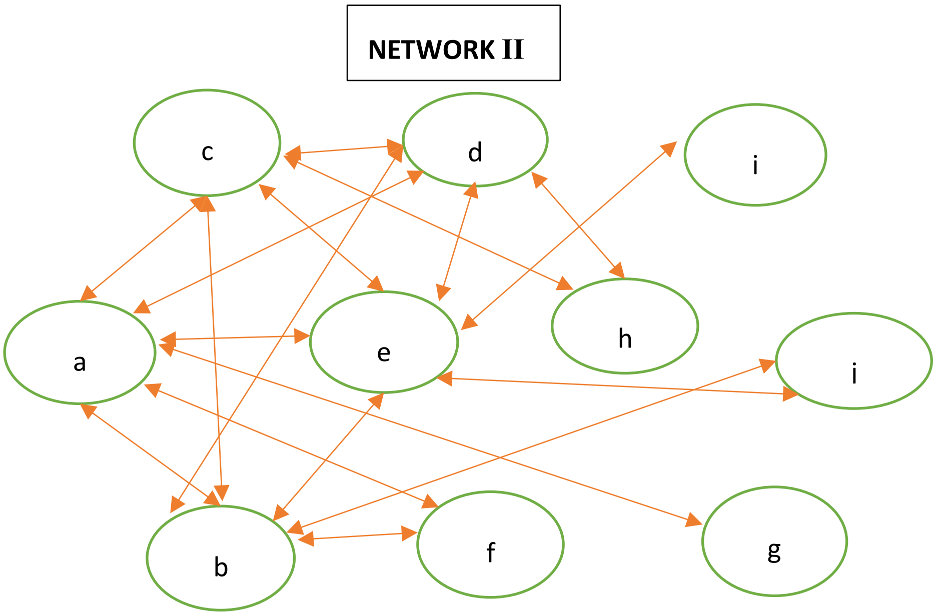



*Figure 2. Network (2)*

To better research the topic, the scale of the two hauler companies that were added to the base network platforms vary in five different sizes: 10%, 20%, 30%, 40%, and 50% of the total orders, referring to 10, 20, 30, 40, and 50 orders respectively. The highest number of orders of the two new companies was set by the author to 50 orders (namely 50% of the total number of orders of the network) in order to also emphasise the connectivity between the members of the networks and not create big disruptions only focused on size.

The Monte Carlo simulation was used for the newly added company that is willing to collaborate with all other members of the network, and the expected maximum matching was calculated quintuple for each of the different sizes as mentioned above. On the other hand, during the experiment with the company that is only willing to cooperate with one member of the network, a matrix 5x5 was created. Both of those techniques were repeated for both instances under study (Kumar, 2018).

Regarding the first instance, the horizontal axis of the drawn matrix consists of all the existing companies of the network named A, B, C, D, and E. Company A is affiliated with three other companies in the network while companies B, C, D, and E are affiliated with three, two, one, and one company/-ies respectively. Additionally, the horizontal axis of the second instance includes the companies a, c, d, f, g with six, five, five, two, and one connection respectively. Even though the second instance could argumentatively end up with a bigger matrix, it should be pointed out that the important aspect of this matrix is the size and the connectivity of the variable. Taking these two aspects under consideration I have chosen the five companies with the biggest variance in their connectivity in order to make my results more diverse and the research more robust, whereas including more companies with the same number of connections would just present similar results.

The results of this experiment created the opportunity to assess the value of two different network platforms based on the connectivity of its members and demonstrated the most optimal option for a network when considering its expansion.

The research was based on the assumptions presented below:

- All companies are active in similar proximity.
- All companies have the same number of orders each day.
- The order matching happens every day at night from the platform and the optimal matches are presented to the companies in the morning before the trucks leave for work.
- The platform always respects the (un)willingness of the companies to work together with each other.
- The total number of existing orders from all the companies is set to 100.
- The probability of a matching is set to 5%.

The scale of the company in the first and second hypothesis was calculated for five different sizes of 10%, 20%, 30%, 40% and 50% of the total orders (Mukul, 2011). The aforementioned assumption presented in the previous segment can be considered one of the most crucial key starting points for this experiment. The probability of matching was set to 5% based on the results of the experiment presented below:

For a random bipartite graph with n=50 and m=50 orders, the Monte Carlo Simulation was executed using K=100.000 for different matching probabilities (p).

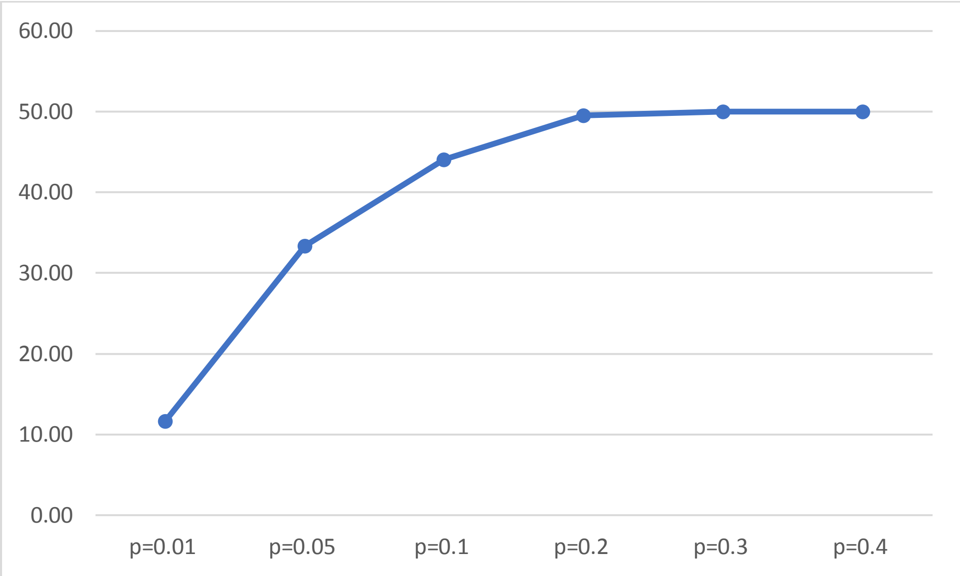



*Figure 3. Expected maximum matchings for different matching probabilities*

Examining the graph above, it can be understood that the expected maximum matching for a matrix of R50,50(p>0.1) is the most optimal matching. Supporting this observation is the consideration that the higher the number of orders is, the higher the available connections will be and the easier for a matching to take place, using those high matching probabilities. In the project under study such high matching probabilities that lead to optimal matching would not be useful and for that reason the matching probability was set to 5% for the following experiments .

## 3.1 Confidence interval calculation

The quantitative model used is the Monte Carlo simulation. The choice of this specific model for our study has already been explained in the Matching in mathematics (random bipartite graphs) segment. As the method used to investigate our research question is the calculation of simulations and not actual data from the industry, the first step was to determine the conversion rate of the actually expected maximum matching from the formula to the simulation (Daniel and Sam, 2011). The formula used is shown below:

$$R_{n,m}(p) = \sum_{E \in B(n,m,p)} p^k (1-p)^{nm-k} \mu.$$

The actual expected maximum matching was manually calculated and afterwards the simulation model was implemented five times for each K(repetition) = 1.000, 10.000, 50.000, 80.000, 100.000 and 150.000 with matching probability p 70% due to the small number of orders , opposite to the p 5% that we used in our R50,50(0.05) experiment as explained in the previous segment. Based on the findings of the small matrixes R2,2(p) -R4,4(p) which can be found in the list of tables, it can be concluded that from K=50.000 we get quite an accurate estimate of the actual expected maximum matchings while for a higher number of repetitions the variance is even smaller; almost perfectly matching the formula value.

Needless to say, it is imperative to prove the robustness of the methodology followed for bigger matrixes and, more specifically, for 50x50 matrixes as those that were used in our experiments. This can be demonstrated by calculating the confidence interval for a sample of n = 70 Monte Carlo simulations with K= 100.000 repetitions each and a matching probability p=0.05. The use of confidence interval for model validation is a quantitative method of validation of an output variable, produced by running a simulation. The sample size was set to n= 70 in order to use the Central Limit Theorem (CLT), as the distribution of the expected maximum matchings is unknown.

According to the CLT, for a sufficiently large sample (the definition of large depends on the size of the population; but, habitually, any sample with n>30 can be considered large) with a finite level of variance, the mean of the population will be approximately equal to the mean of the sampled variables. The experiment that was conducted fulfils both the sample size and the finite variance requirements of the theorem. Furthermore, the CLT states that these samples approximate to a normal distribution, with their variances being roughly equal to the variance of the population as the sample size gets larger, according to the law of large numbers.

The first step in finding the confidence interval during the experiments was to run the Monte Carlo simulation 70 times in order to get a sample of expected maximum matchings. The results can be seen in the table below (Hazra, 2017):

*Table 1. Confidence interval calculation*

| | | | | | | |
|---|---|---|---|---|---|---|
| 43.243 | 43.236 | 43.232 | 43.229 | 43.246 | 43.243 | 43.244 |
| 43.247 | 43.239 | 43.229 | 43.236 | 43.244 | 43.231 | 43.245 |
| 43.234 | 43.236 | 43.237 | 43.242 | 43.239 | 43.247 | 43.230 |
| 43.236 | 43.244 | 43.242 | 43.235 | 43.238 | 43.233 | 43.227 |
| 43.240 | 43.241 | 43.220 | 43.237 | 43.232 | 43.239 | 43.237 |
| 43.244 | 43.243 | 43.231 | 43.241 | 43.231 | 43.247 | 43.245 |
| 43.240 | 43.239 | 43.226 | 43.238 | 43.241 | 43.245 | 43.242 |
| 43.234 | 43.243 | 43.237 | 43.243 | 43.247 | 43.237 | 43.227 |
| 43.242 | 43.239 | 43.232 | 43.245 | 43.231 | 43.237 | 43.238 |
| 43.233 | 43.245 | 43.259 | 43.239 | 43.233 | 43.239 | 43.227 |

Then the mean of all the expected maximum matchings in the n=70 sample was calculated:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 43.238$$

As well as the sample standard deviation:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = 0.006558$$

As mentioned previously in this segment, the distribution of the expected maximum matchings was unknown which meant the standard deviation of the population σ was unknown too. In this case, instead of the normal z distribution the student t distribution was used for the calculation of the confidence interval as seen below:

$$\left[\bar{x} - t_c \frac{s}{\sqrt{n}}, \bar{x} + t_c \frac{s}{\sqrt{n}}\right]$$

With $t_C$being the critical value for the Student t distribution for confidence level c. The degrees of freedom of the sample in this project is d.f= n-1= 69 and, based on the tables, the $t_C$ can be defined according to the confidence level needed. For this experiment c= 0.95 was used, resulting to a $t_C = 1.994$ .

Lastly, it can be confidently stated that in a frequency as high as 95%, for a random bipartite graph with m=50 import and n=50 export orders, and probability of matching p=0.05, the expected maximum matchings calculated by the use of the Monte Carlo simulation fluctuate between [ 43.236, 43.239]

Evidently, the Monte Carlo simulation model used is robust enough to infer conclusions based on it. Of course, several experiments were conducted in the course of this thesis. All of them presented some small differentiations between them and the base experiment, which its confidence interval was calculated in advance. Therefore, since the variance for the base experiment was that little, the calculation of the confidence interval for all experiments was considered of minor importance.

# 4.Results

As explained thoroughly in the methodology segment to examine the topic, a hypothetical case study was conducted. Based on two instances, two different networks were theoretically created, and different MATLAB constraints were dictated. The aim was to investigate whether and which changes in the network would lead to the attainment of the expected maximum matching. For this purpose, different companies with different policies of collaboration were added to the network and the consequence was studied in depth. The symbol A expresses the constraints that were implemented, K signifies the number of stimulations made and p indicates the probability of matching. According to the experiment's results, the augmentation of p leads to higher number of expected maximum matching. Simultaneously, a big network like the one we experimented in R50,50(p) can apparently reach its full matching capacity when the p is over 20%. It is quite understandable that the more options given, the expected maximum matching becomes more possible.

The first two experiments that were conducted can be considered an extension of the robustness section. While in the main matrixes the aforementioned figures of m=50 and n=50 was used, small variations were implemented in both instances. Those include mainly the use of several constrains between the collaboration of some of the companies in the network. For that reason, the simulation was executed for different probabilities of matching p for the base networks.

I instance

| K=100.000 | EMM |
|---|---|
| p=0.01 | 11.64 |
| p=0.05 | 33.39 |
| p=0.1 | 44.07 |
| p=0.2 | 49.53 |
| p=0.3 | 49.96 |
| p=0.4 | 49.99 |

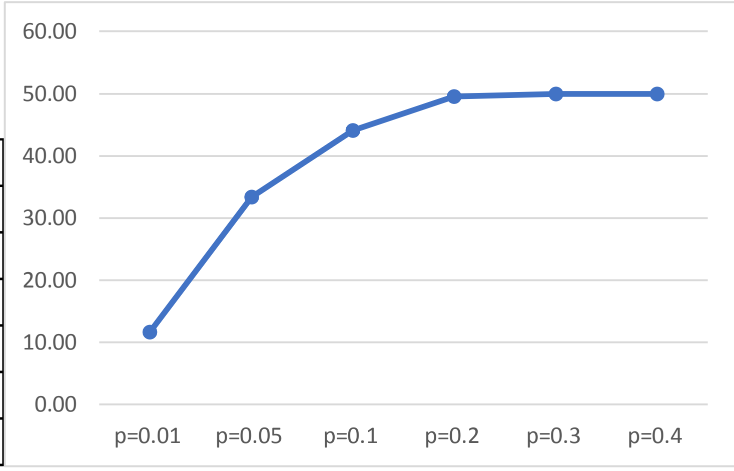



*Figure 4. I Instance*

II instance

| K=100.000 | EMM |
|---|---|
| p=0.01 | 9.23 |
| p=0.05 | 28 |
| p=0.1 | 39.11 |
| p=0.2 | 47.35 |
| p=0.3 | 49.32 |
| p=0.4 | 49.86 |

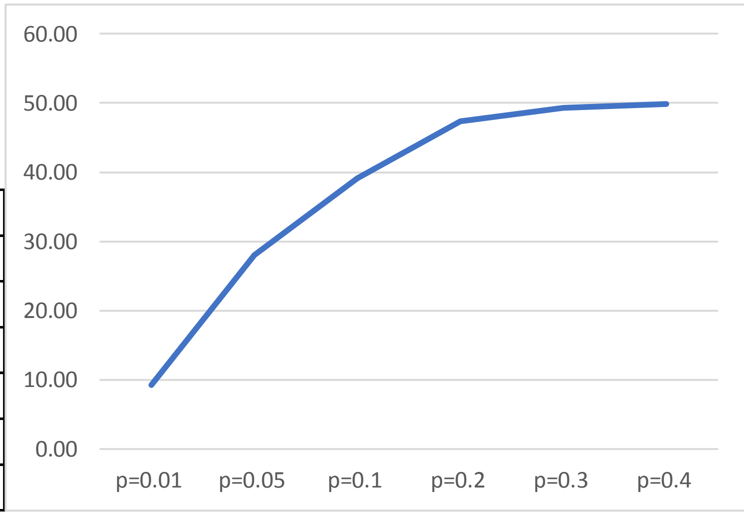



*Figure 5. II Instance*

By examining the two tables and two graphs above, the hypothesis supported in the robustness section can be confirmed. According to the results, taking high matching probability for matrixes of this size does not lead to useful results. For that reason, the matching probability used in the experiments was decided to be 5%.

Thereafter, the calculation of the first hypothesis: the expected maximum matching of the network platforms with the addition of a company willing to collaborate with everyone was conducted as it was explained in the conceptual framework. Certainly, these calculations were done for all five different scales of the newly added company.

| K:100,000&p=0.05 | I Instance |
|---|---|
| N=10 orders | 40.52 |
| N=20 orders | 47.65 |
| N=30 orders | 54.74 |
| N=40 orders | 61.69 |
| N=50 orders | 68.42 |

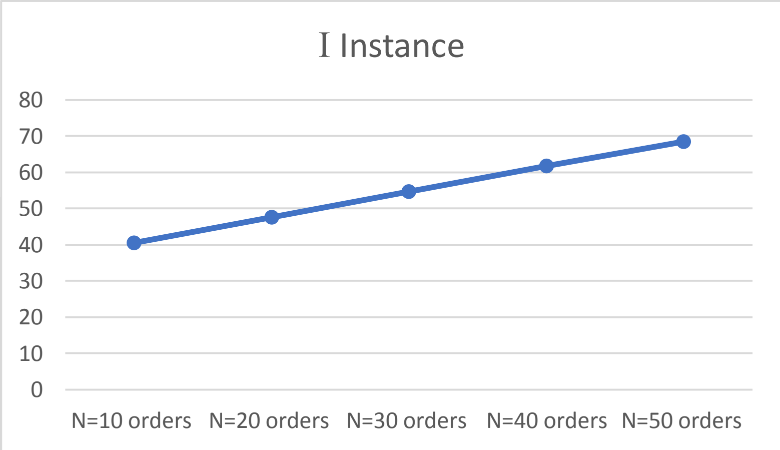



*Figure 6. Hypothesis I*

| K:100,000&p=0.05 | II Instance |
|---|---|
| N=10 orders | 35.62 |
| N=20 orders | 43.23 |
| N=30 orders | 50.78 |
| N=40 orders | 58.25 |
| N=50 orders | 65.57 |

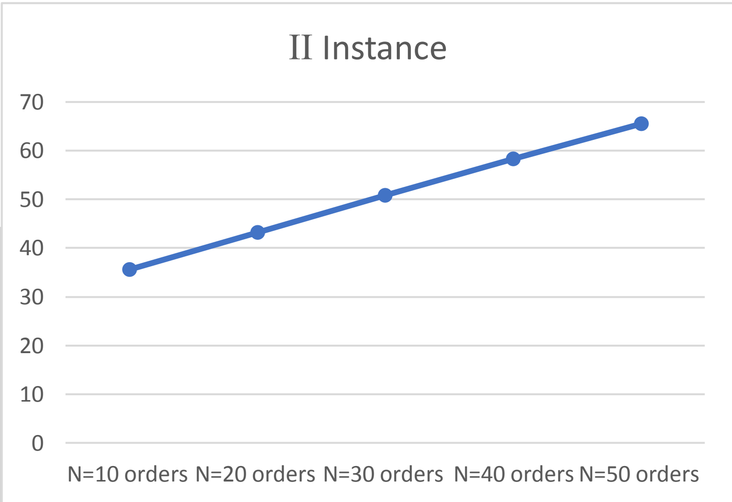



*Figure 7. Hypothesis I*

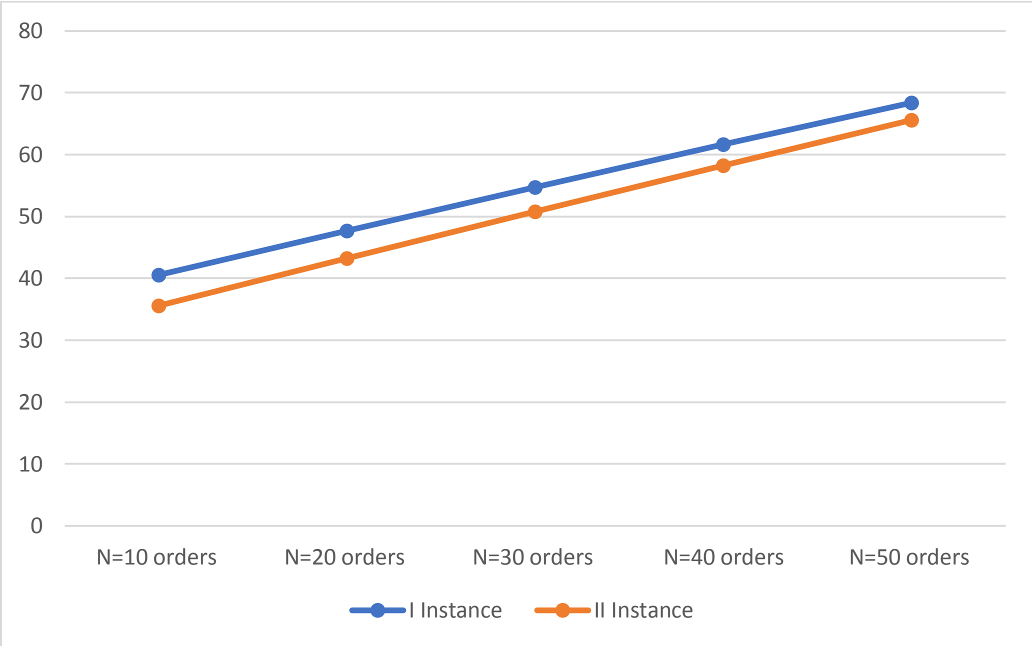



*Figure 8. I & II Instance, increasing number of orders*

The final step of the experiments of this project was to calculate the expected maximum matching of the second hypothesis: the new company only willing to collaborate with only one other company in the network. This calculation was conducted for both instances. At this point, the simulation was executed for the five different scales of the added hauler company but each time, the company was affiliated to a different member of the network. Since all the existing companies of the first instance's network are five, a matrix five by five was created which led to 25 variations. In the second case, the base networks consist of ten companies. In order to make the extraction and comparison of the results easier, only five of those companies were chosen so as to create a five-by-five matrix again. Those companies are named a, c, d, e, and f subsequently (Ruxton and Neuhäuser, 2013).

The simulation executed during the first instance had a matching probability of p = 0.05. Even if it has been already mentioned in the segment of conceptual theory, it is of major importance that the connections of the existing members of each network should be clearly stated so as to better understand the experiments' results. Company A is connected to three other members while companies B, C, D and E are affiliated to three, two, one, and one other member(s) of the network subsequently.

*Table 2. Hypothesis II : 5x5 matrix of first Instance*

| **K=100.000 & p=0.05** | **A** | **B** | C | D | E |
|---|---|---|---|---|---|
| **N=10** | 35,78 | 35,78 | 36,23 | 36,51 | 36,51 |
| **N=20** | 39,14 | 39,15 | 39,89 | 40,38 | 40,37 |
| **N=30** | 43,16 | 43.16 | 44,15 | 44,80 | 44,79 |
| **N=40** | 47,72 | 47,70 | 48,88 | 49,64 | 49,63 |
| **N=50** | 52,66 | 52,67 | 54,01 | 54,85 | 54,85 |

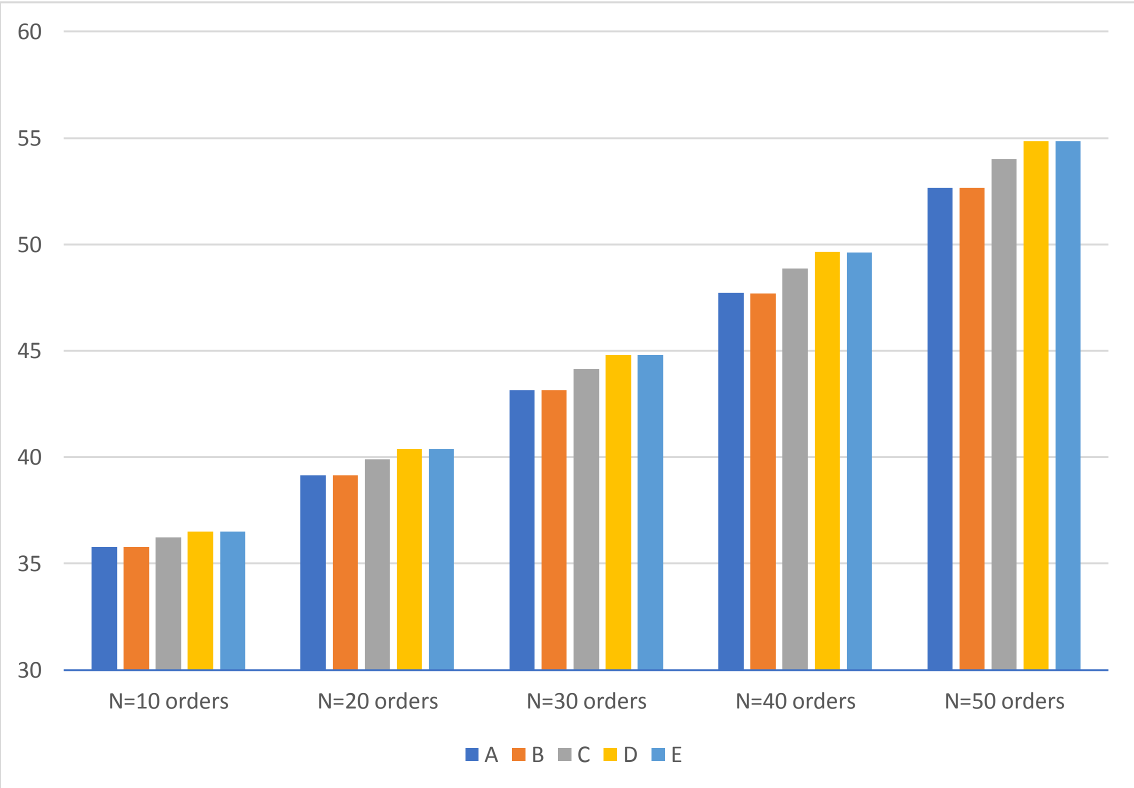



*Figure 9. Hypothesis II*

In the second instance the matching probability used was equal to p=0.05. The companies used in this experiment were chosen based on their connectivity in the network. Company a is connected with six other companies while companies named c, d, e, and f with five, five, two and one other members of the network platform subsequently.

*Table 3. Hypothesis II : 5x5 matrix of second Instance*

| **K:100,000 & p=0,05** | a | c | d | f | G |
|---|---|---|---|---|---|
| N=10 orders | 29,78 | 29,93 | 29,94 | 30,29 | 30,42 |
| N=20 orders | 32,73 | 32,99 | 32,99 | 33,55 | 33,73 |
| N=30 orders | 36,46 | 36,80 | 36,78 | 37,53 | 37,77 |
| N=40 orders | 40,79 | 41,18 | 41,18 | 42,06 | 42,35 |
| N=50 orders | 45,59 | 46,01 | 46,00 | 46,97 | 47,31 |

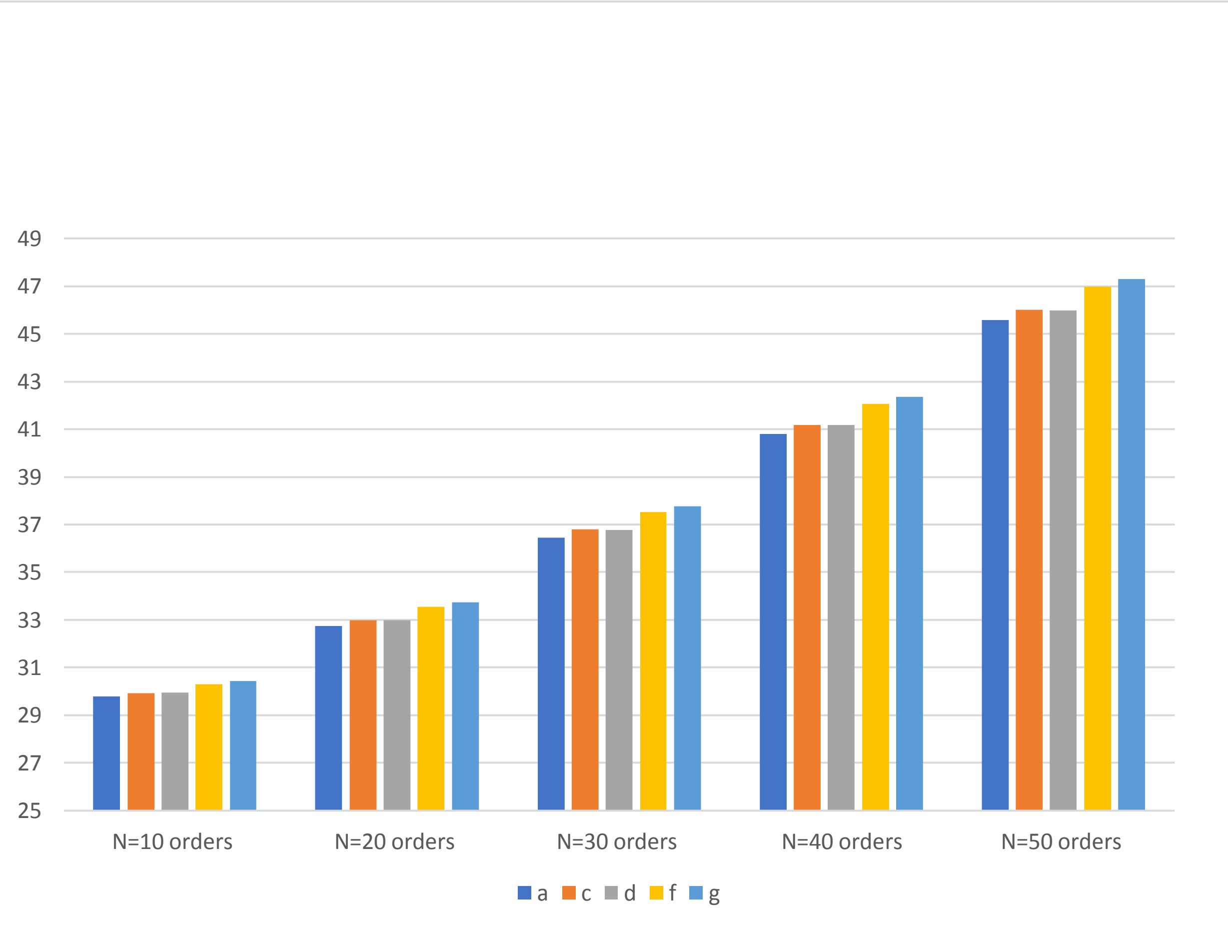



*Figure 10. Network by the number of orders, p = .05*

*Table 4. Comparison of hypothesis I & II both for the most and least connected company of instance II .*

| **K:100,000 & p=0,05** | a | willing to work with all | **K:100,000 & p=0,05** | g | willing to work with all |
|---|---|---|---|---|---|
| N=10 orders | 29,78 | 35,62 | N=10 orders | 30,42 | 35,62 |
| N=20 orders | 32,73 | 43,23 | N=20 orders | 33,73 | 43,23 |
| N=30 orders | 36,46 | 50,78 | N=30 orders | 37,77 | 50,78 |
| N=40 orders | 40,79 | 58,25 | N=40 orders | 42,35 | 58,25 |
| N=50 orders | 45,59 | 65,57 | N=50 orders | 47,31 | 65,57 |

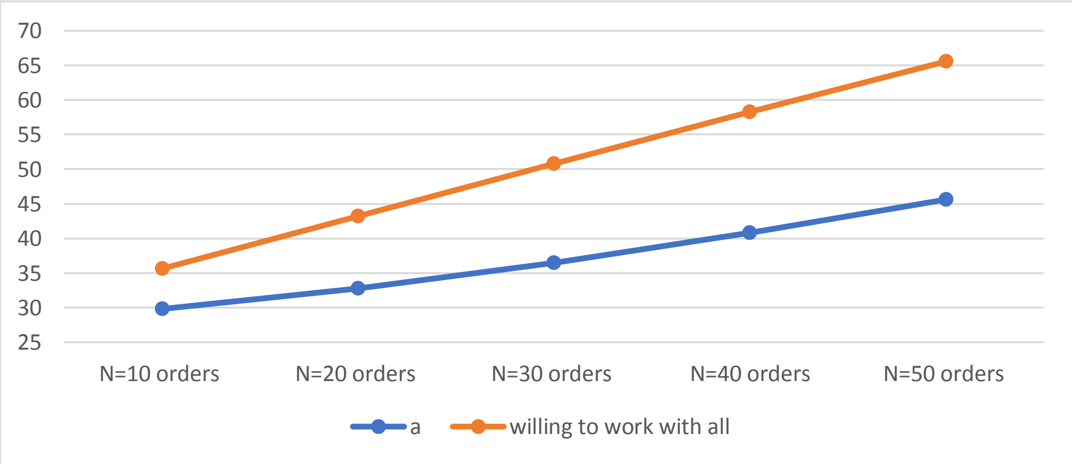



*Figure 11. Comparison of hypothesis I & II for the most connected company of instance II .*

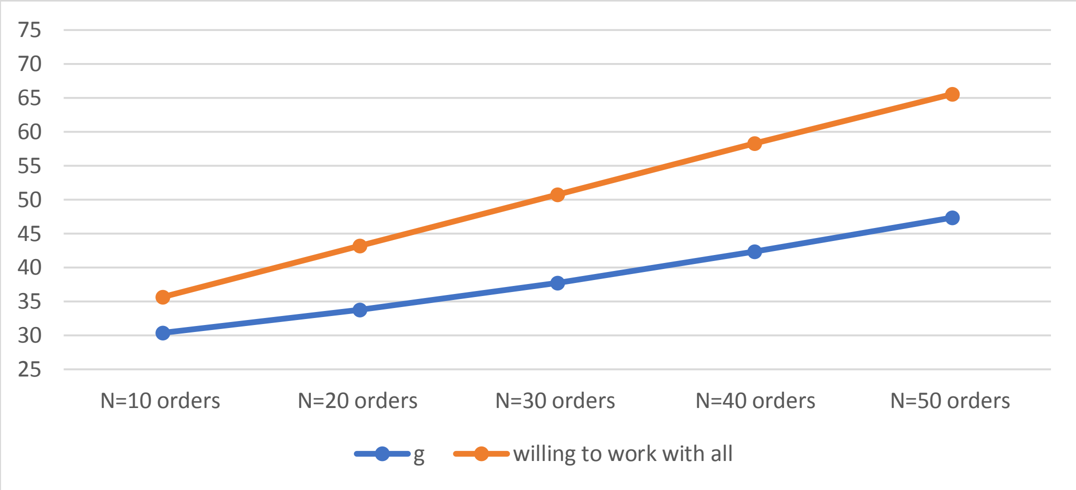



*Figure 12. Comparison of hypothesis I & II for the least connected company of instance II .*

## 5.Discussion and Conclusion

The focal metric for the success or not of a network platform ,is mainly based on the extra profit generation or cost reduction for the network's members and the network itself. The accomplishment of this extra profit generation depends heavily on the number and the cooperation between the members . This project is an attempt to create a quantitative method that could possibly assess the value of a network platform based on the different choices for collaboration of its members and extract some useful information that could help optimize network expansion. This was achieved by calculating with the use of random bipartite graphs the expected maximum matching of a network in different scenarios according to the business scenarios that were created by the author.

Many conclusions could be extracted from the experiments that were conducted while the observations made definitely create further room for future research and discussion. This thesis contributes by providing a model that could quantify the quality of a network platform and assess which company is a better fit depending on its number of orders and willingness to cooperate. It also proved that a company willing to work with everyone is a better option than one that only wants to cooperate with just another member. Of course, that is the case only if all other things are equal. For example, the results of this research present a quantifying degree of how big a company unwilling to collaborate with the other members should be in order to still be considered an optimal addition to a network. According to the experiments (table 4), a company with a scale of N=30 orders unwilling to collaborate with the other members of the network is still a better addition than a company willing to collaborate with everyone but only having a size of N=10 orders.

Finally, another observation that can be made is that the bigger the difference between the connectivity of the member the newly added company is connected with, the bigger the impact it will have. The better option for the network would be the company willing to collaborate with the least connected member of that network. This observation can be verified by the examination of tables (2 and 3) in both instances. The experiments made obvious the fact that when a company willing to work with the least connected members of a network is chosen to join it, higher expected maximum matchings are achieved which can be translated to bigger cost reduction.

These conclusions hopefully give a clear direction for better optimization of the network platforms and the model created may also allow the extraction of much more in-depth information; this thesis consists only of a small contribution to the research of this matter considering the time frame available. An interesting topic for further research based on the model used in this project could be the study of order matching connected with the cost savings and incorporating different independent variables like distance between members.

# List of Tables

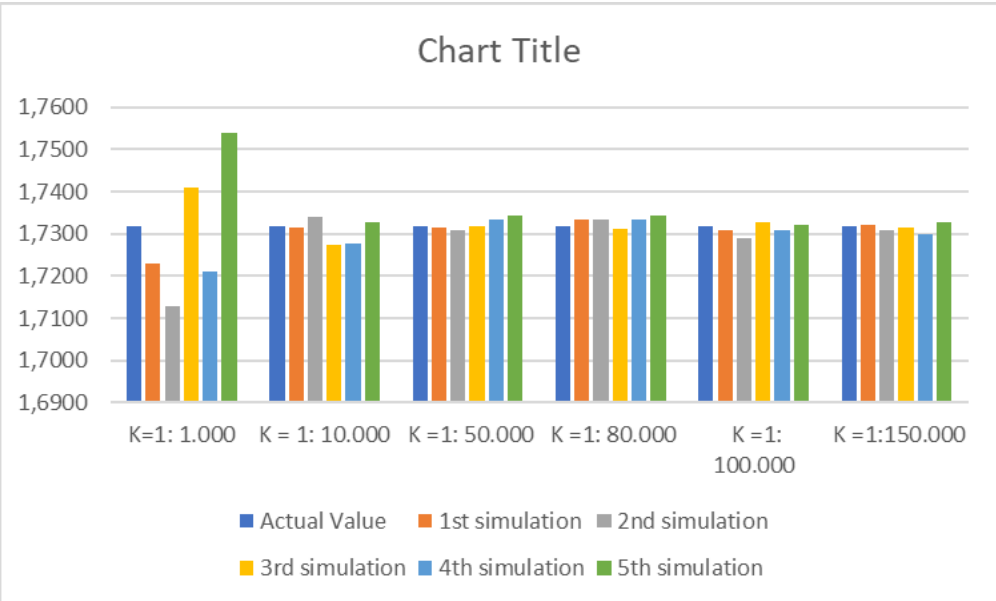
Conversion rate

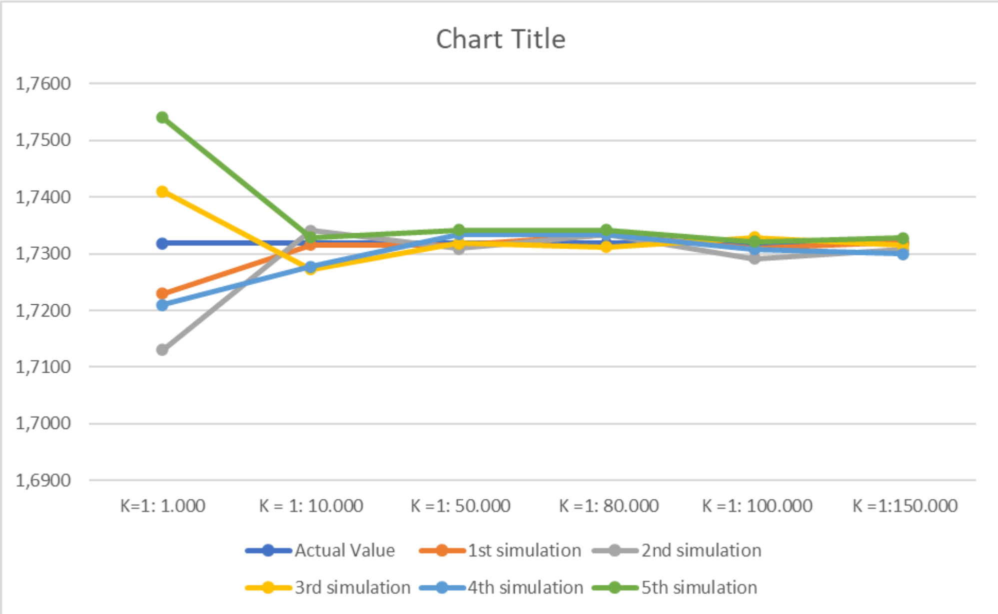
Rn,1(p) = 1 − (1 − p)^ n

P=0.7

R2,2(p) = 1,7318

In the following section, a total of six examples of comparison between the conversion rate of a suggested equation and the rate generated by a number of stimulations is presented. In each example, both the equation and a chart of the simulations are shown.





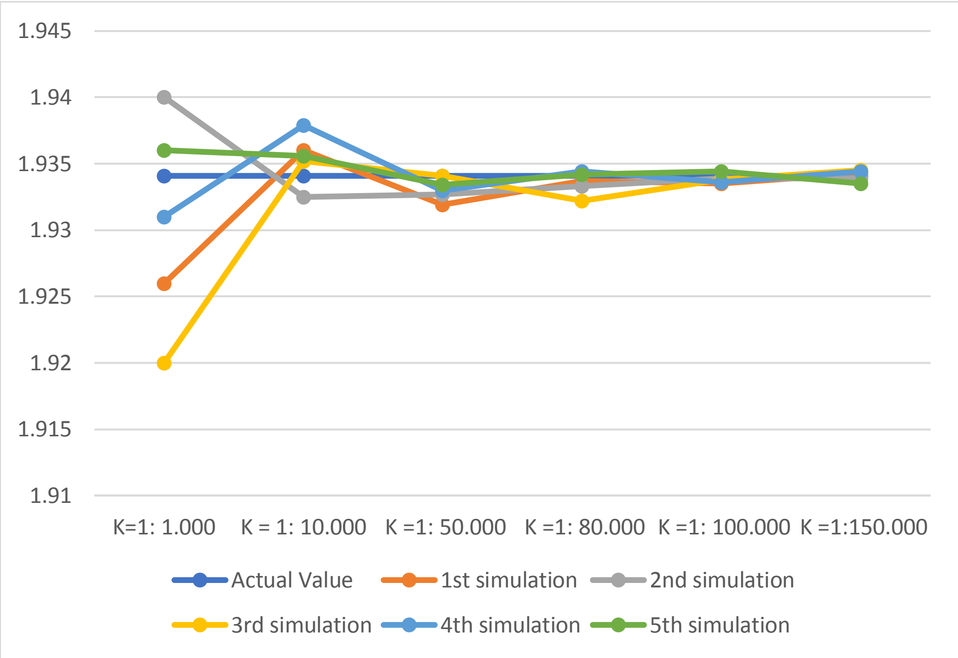

The following equation expresses the expected maximum matching in a network of a total of 5 companies.

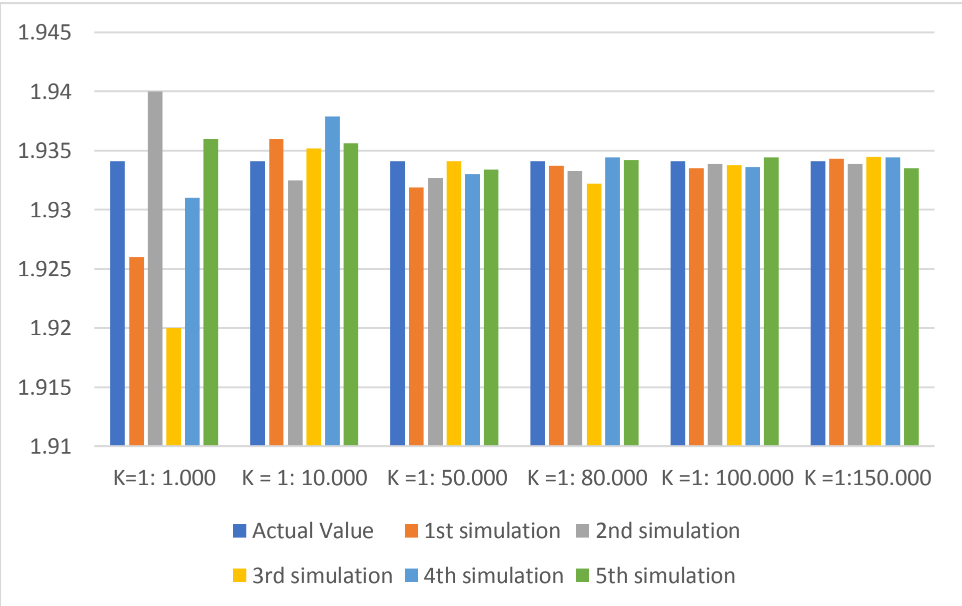
$$R_{3,2}(p)=21p^2(p-1)^4-38p^3(p-1)^3+30p^4(p-1)^2+2p^6-6p(p-1)^5-12p^5(p-1)$$

In order to conclude which number of simulations' conversion rate coincides with the result of the formula, six attempts of simulations were conducted. The total number of the simulations was at first 1.000, 10.000 and subsequently, 50.000, 80.000, 100.000, and 150.000 times. As it can be easily understood by the undermentioned figures and graph, the conversion rate starts to coincide with the one suggested by the formula after the 50.000 simulations. The differences in the computation are infinitesimal. In the following examples of comparison between the formula's result and the simulations', the aforementioned conclusion can be verified.

R3,2(p) = 1,93409

| K 1: 1.000 | K 1: 10.000 | K 1: 50.000 | K 1: 80.000 | K 1: 100.000 | K 1:150.000 |
|---|---|---|---|---|---|
| 1.9260 | 1.9360 | 1.9319 | 1.9337 | 1.9335 | 1.9343 |
| 1.9400 | 1.9325 | 1.9327 | 1.9333 | 1.9339 | 1.9339 |
| 1.9200 | 1.9352 | 1.9341 | 1.9322 | 1.9338 | 1.9345 |
| 1.9310 | 1.9379 | 1.9330 | 1.9344 | 1.9336 | 1.9344 |
| 1.9360 | 1.9356 | 1.9334 | 1.9342 | 1.9344 | 1.9335 |



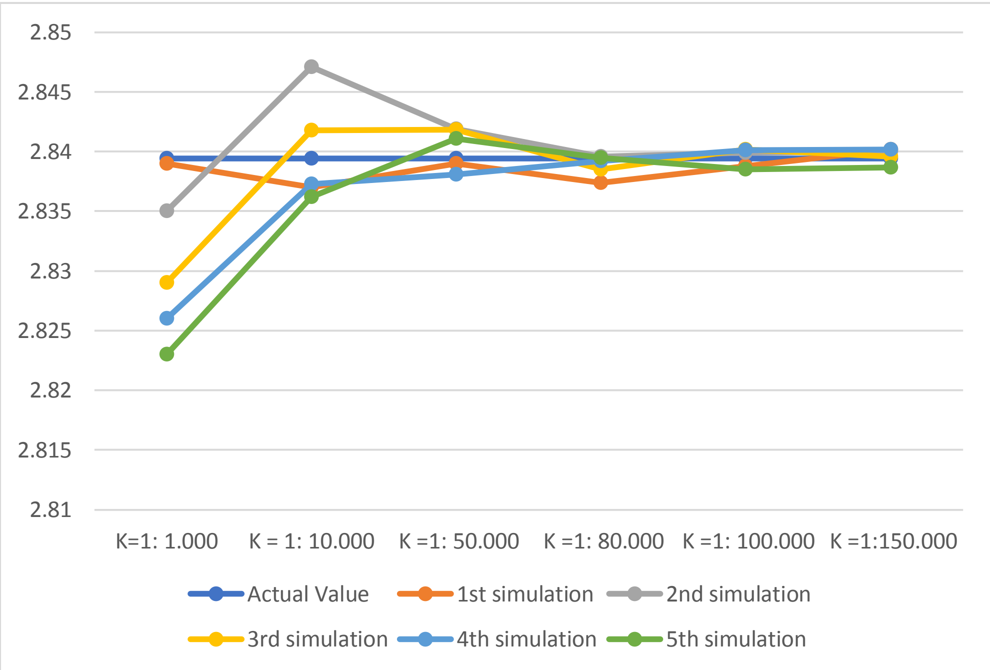

In this case, the formula is implemented on a six-member network and is, thus, modified. See above:

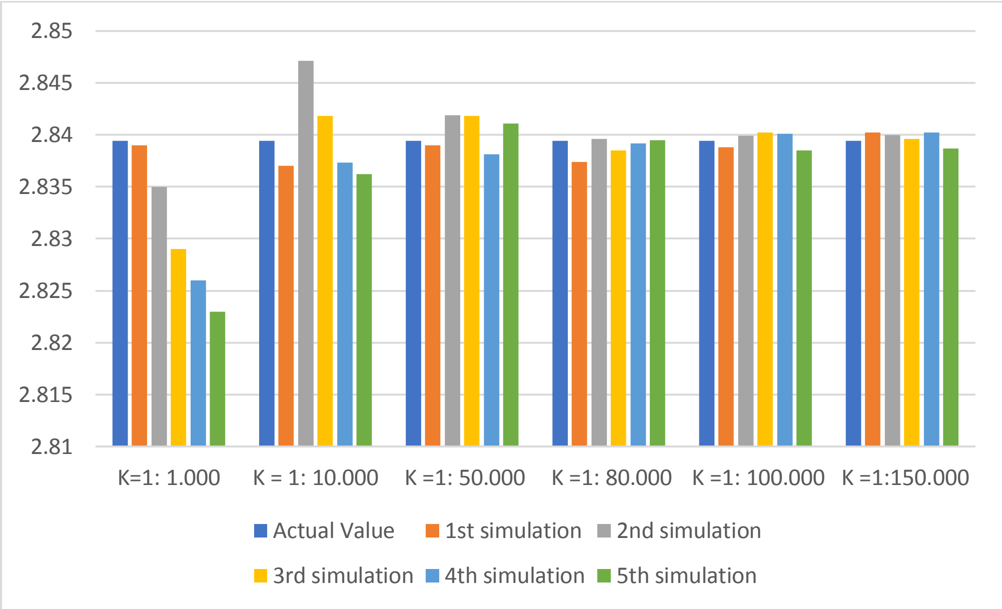
$$R_{3,3}(p)=168p^3(p-1)^6-54p^2(p-1)^7-288p^4(p-1)^5+333p^5(p-1)^4-246p^6(p-1)^3+108p^7(p-1)^2+3p^9+9p(p-1)^8-27p^8(p-1)$$

R3,3(p) = 2,8394, time needed: 8546.185 s

In this case, the same method of calculation was followed, by conducting six simulations with a different total number of repetitions (1.000, 10.000, 50.000, 80.000, 100.000, and 150.000respectively). Confirming the conclusion drawn in the first example of comparison, the conversion rate starts to coincide with the one suggested by the formula after the 50.000 simulations, with insignificant differences.

| K 1: 1.000 | K 1: 10.000 | K 1: 50.000 | K 1: 80.000 | K 1: 100.000 | K 1:150.000 |
|---|---|---|---|---|---|
| 2.8390 | 2.8370 | 2.8390 | 2.8374 | 2.8388 | 2.8402 |
| 2.8350 | 2.8471 | 2.8419 | 2.8396 | 2.8399 | 2.8400 |
| 2.8290 | 2.8418 | 2.84184 | 2.8385 | 2.84019 | 2.8396 |
| 2.8260 | 2.8373 | 2.8381 | 2.8392 | 2.8401 | 2.8402 |
| 2.8230 | 2.8362 | 2.8411 | 2.8395 | 2.8385 | 2.8387 |



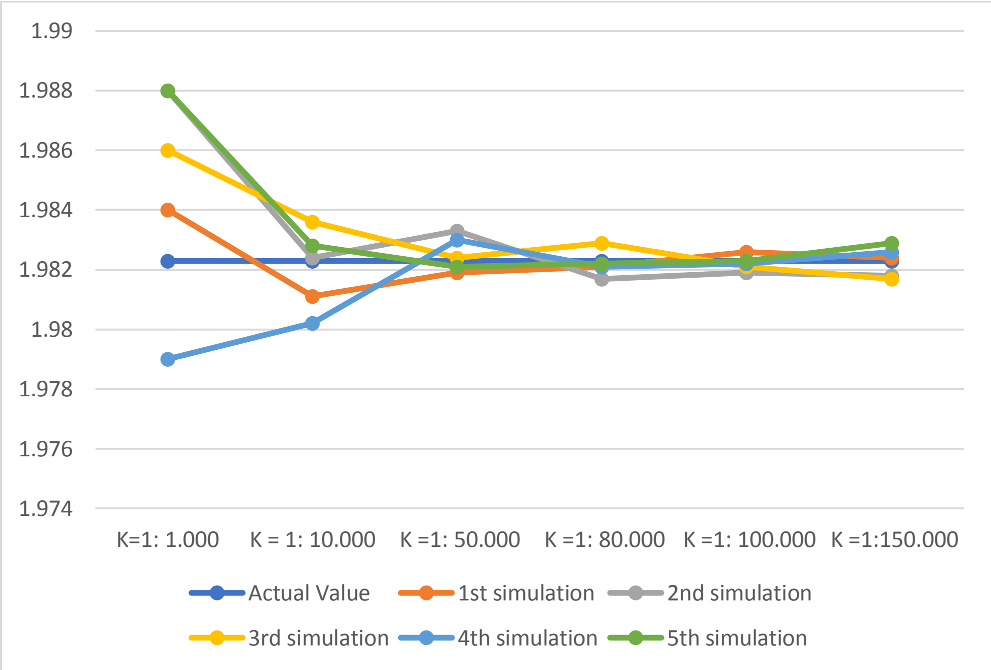


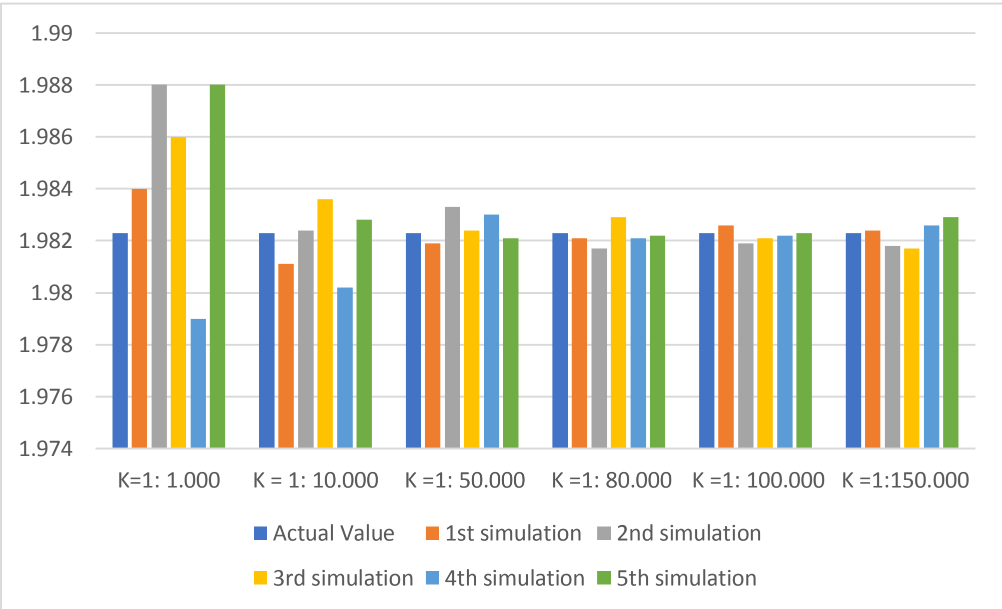

$$R_{4,2}(p) = 40\,p^2\,(p-1)^6 - 104\,p^3\,(p-1)^5 + 138\,p^4\,(p-1)^4 - 112\,p^5\,(p-1)^3 + 56\,p^6\,(p-1)^2 + 2\,p^8$$
$$-8\,p\,(p-1)^7 - 16\,p^7\,(p-1)$$

R4,2(p) = 1,9823

The result of the equation presented above expresses a formula implemented on a network of six members. Following the same method discussed in the previous examples, six attempts of simulations were conducted, and the number of the simulations were, as previously said, 1.000, 10.000 and 50.000, 80.000, 100.000, and 150.000 times. By examining the undermentioned figures and graph, it can be verified again that the conversion rate starts to coincide with the one suggested by the formula after the 50.000th time of simulations.

| K 1: 1.000 | K 1: 10.000 | K 1: 50.000 | K 1: 80.000 | K 1: 100.000 | K 1:150.000 |
|---|---|---|---|---|---|
| 1.9840 | 1.9811 | 1.9819 | 1.9821 | 1.9826 | 1.9824 |
| 1.9880 | 1.9824 | 1.9833 | 1.9817 | 1.9819 | 1.9818 |
| 1.9860 | 1.9836 | 1.9824 | 1.9829 | 1.9821 | 1.9817 |
| 1.9790 | 1.9802 | 1.9830 | 1.9821 | 1.9822 | 1.9826 |
| 1.9880 | 1.9828 | 1.9821 | 1.9822 | 1.9823 | 1.9829 |



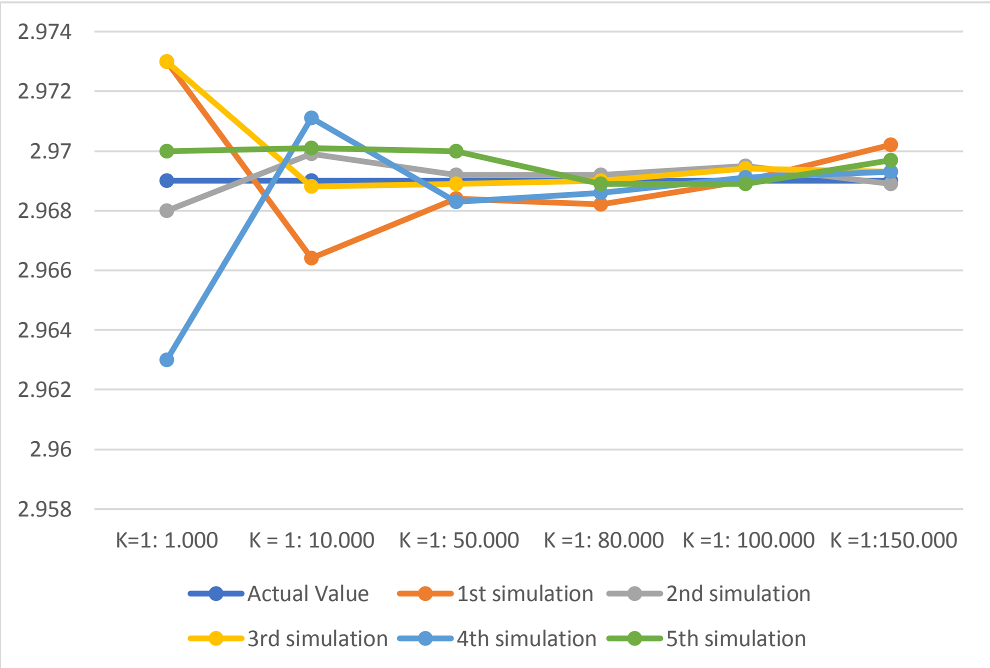



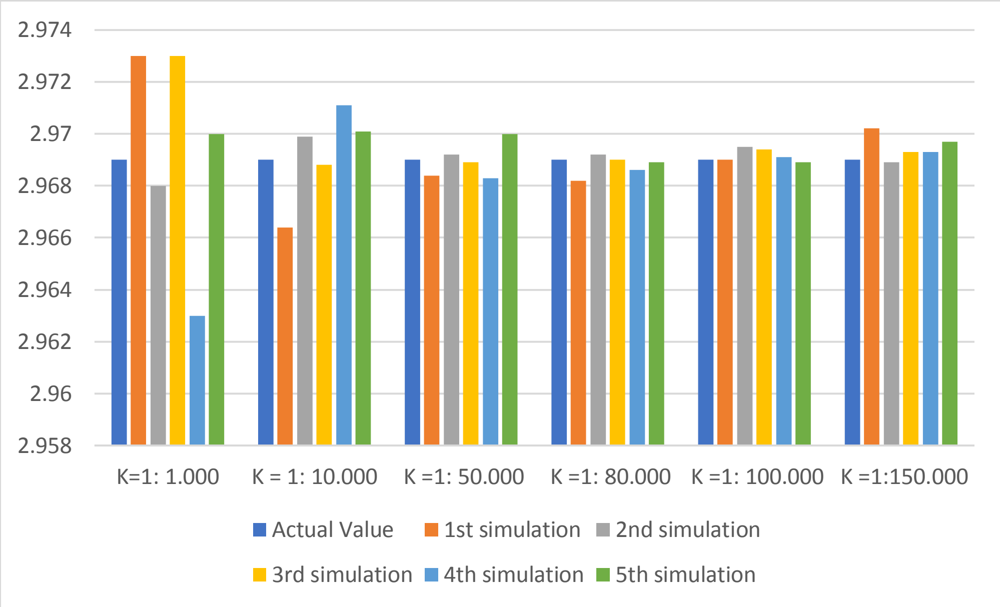

Finally, graphs of simulations that can be compared to the formula's conversion rate for examples of a network with combinations of R4,3(p), R4,4(p), and R5,4(p) are given below. The method implemented is the one explained extensively in the first three examples of comparison.

The conclusion that the conversion rate suggested both by the formula and by stimulations made over 50.000 times is approximately the same is, thus, ascertained.

R4,3(p) = 2,9690 , time needed 8934.306 s

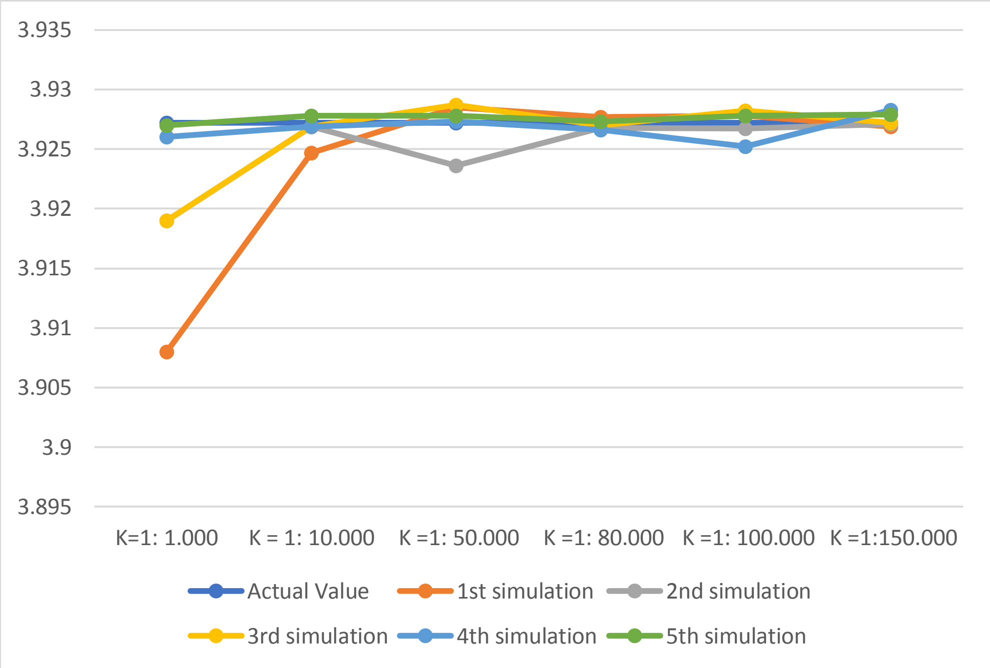
| K 1: 1.000 | K 1: 10.000 | K 1: 50.000 | K 1: 80.000 | K 1: 100.000 | K 1:150.000 |
|---|---|---|---|---|---|
| 2,9730 | 2,9664 | 2,9684 | 2,9682 | 2,9690 | 2,9702 |
| 2,9680 | 2,9699 | 2,9692 | 2,9692 | 2,9695 | 2,9689 |
| 2,9730 | 2,9688 | 2,9689 | 2,9690 | 2,9694 | 2,9693 |
| 2,9630 | 2,9711 | 2,9683 | 2,9686 | 2,9691 | 2,9693 |
| 2,9700 | 2,9701 | 2,9700 | 2,9689 | 2,9689 | 2,9697 |

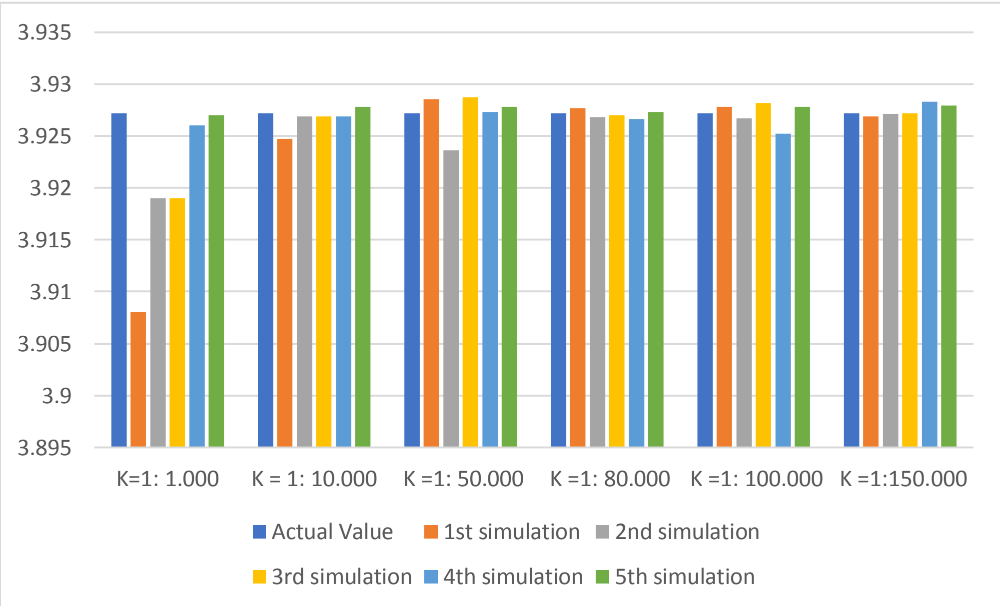





R4,4(p) = 3,9272 , time needed 7229 s

| K 1: 1.000 | K 1: 10.000 | K 1: 50.000 | K 1: 80.000 | K 1: 100.000 | K 1:150.000 |
|---|---|---|---|---|---|
| 3,9080 | 3,9247 | 3,92852 | 3,9277 | 3,9278 | 3,9269 |
| 3,9190 | 3,9269 | 3,9236 | 3,9268 | 3,9267 | 3,9271 |
| 3,9190 | 3,9269 | 3,9287 | 3,9270 | 3,9282 | 3,9272 |
| 3,9260 | 3,9269 | 3,9273 | 3,9266 | 3,9252 | 3,9283 |
| 3,9270 | 3,9278 | 3,9278 | 3,9273 | 3,9278 | 3,9279 |

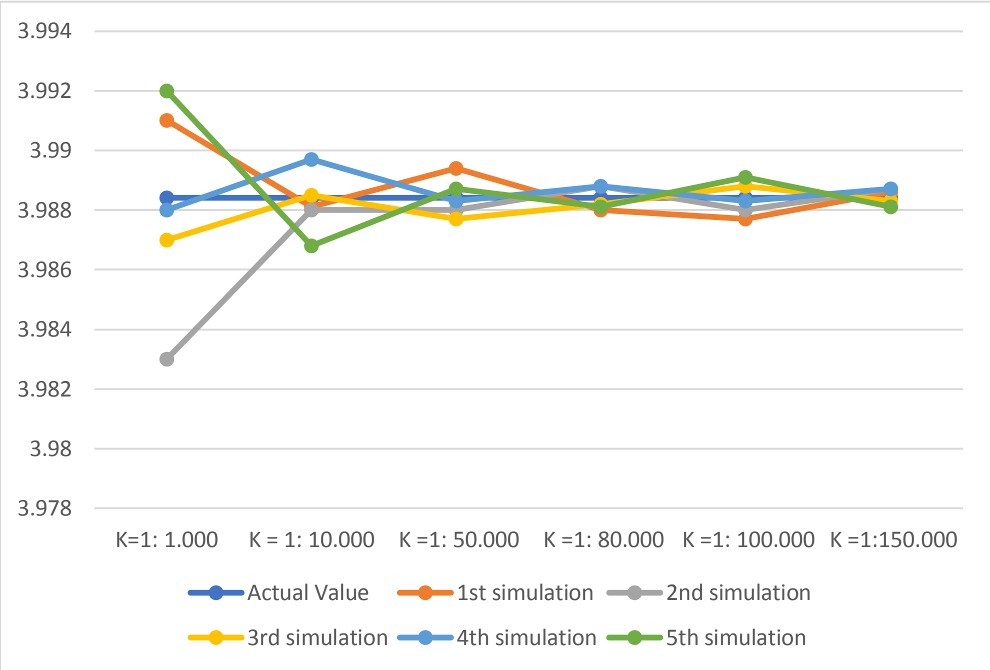


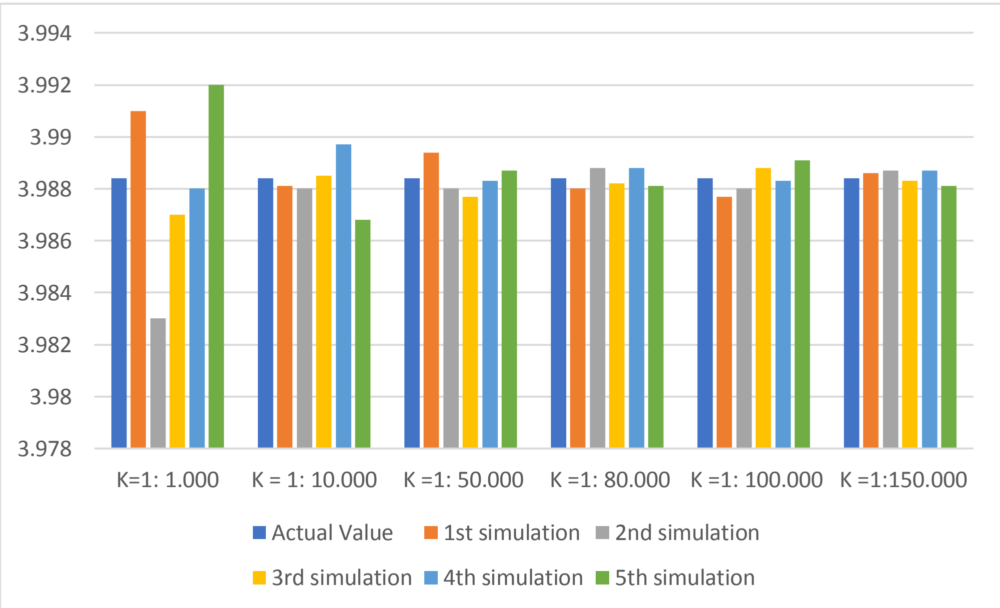




R5,4(p) =3,9884 , time needed 4552.719 S

| K 1: 1.000 | K 1: 10.000 | K 1: 50.000 | K 1: 80.000 | K 1: 100.000 | K 1:150.000 |
|---|---|---|---|---|---|
| 3,9910 | 3,9881 | 3,9894 | 3,9880 | 3,9877 | 3,9886 |
| 3,9830 | 3,9880 | 3,9880 | 3,9888 | 3,9880 | 3,9887 |
| 3,9870 | 3,9885 | 3,9877 | 3,9882 | 3,9888 | 3,9883 |

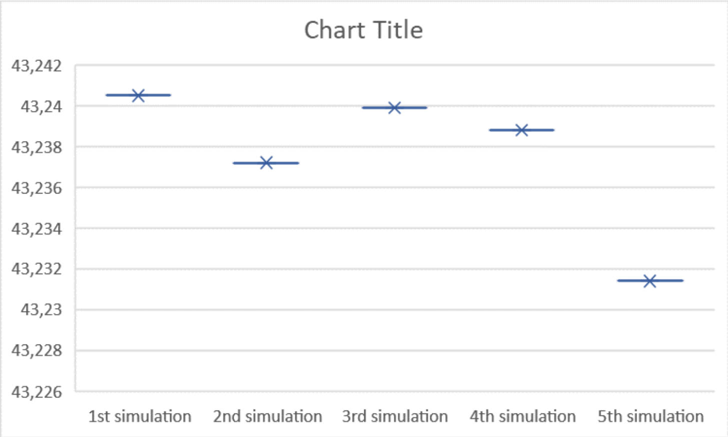
| | | | | | |
|---|---|---|---|---|---|
| 3,9880 | 3,9897 | 3,9883 | 3,9888 | 3,9883 | 3,9887 |
| 3,9920 | 3,9868 | 3,9887 | 3,9881 | 3,9891 | 3,9881 |



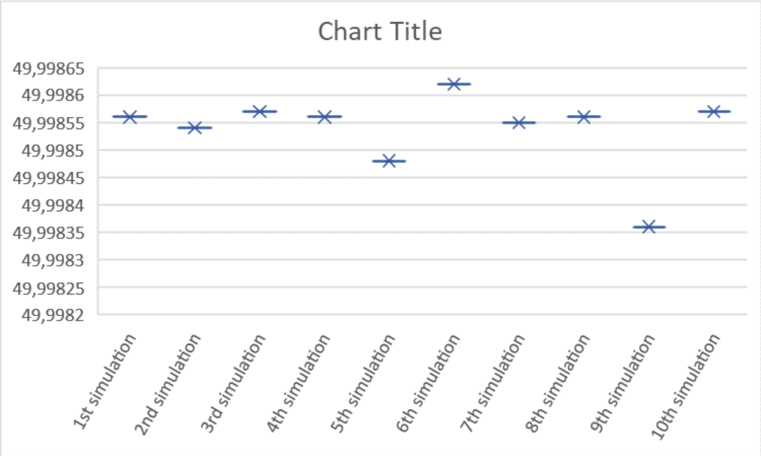
M=50, N=50, P=0.05, time needed: 7981,341s

| K =500.000 |
|---|
| 43,2405 |
| 43,2372 |
| 43,2399 |
| 43,2388 |
| 43,2314 |




M=50, N=50, P=0.2, time needed: 14622,951s

| K =500.000 |
|---|
| 49,99856 |
| 49,99854 |
| 49,99857 |
| 49,99856 |
| 49,99848 |
| 49,99862 |
| 49,99855 |
| 49,99856 |
| 49,99836 |
| 49,99857 |



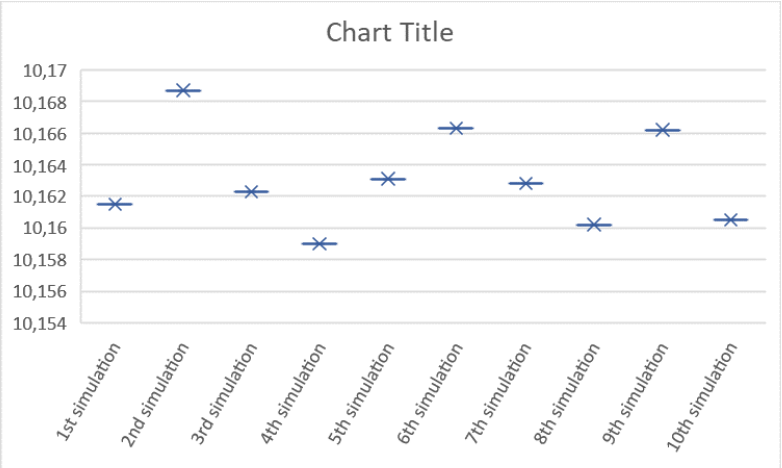

M=50, N=50, P=0.4, time needed: 14622,951s

| K =500.000 |
|---|
| 50 |
| 50 |
| 50 |
| 50 |
| 50 |
| 50 |
| 50 |

| |
|---|
| 50 |
| 50 |
| 50 |

M=50, N=50, P=.005, time needed: 14622,951s

| K =500.000 |
|---|
| 10,1615 |
| 10,1687 |
| 10,1623 |
| 10,1590 |
| 10,1631 |
| 10,1663 |
| 10,1628 |
| 10,1602 |
| 10,1662 |
| 10,1605 |




.

# Appendix

Appendix /MATLAB code and constrains used

First Instance

MATLAB Constraints used.

A(31:40,1:10)=0; A(31:40,21:30)=0;A(31:40,41:50)=0;

A(41:50,11:20)=0; A(41:50,21:30)=0; A(41:50,31:40)=0;

A(1:10,31:40)=0; A(21:30,31:40)=0; A(41:50,31:4vf0)=0;

A(11:40,41:50)=0;

Second instance

MATLAB constraints

A(26:30,11:25)=0; A(26:30,31:50)=0;

A(31:35,6:30)=0; A(31:35,36:50)=0;

A(36:40,1:10)=0; A(36:40,21:35)=0; A(36:40,41:50)=0;

A(41:45,1:20)=0; A(41:45,26:40)=0; A(41:45,46:50)=0;

A(46:50,1:5)=0; A(46:50,11:20)=0; A(46:50,26:45)=0;

A(11:25,26:30)=0; A(31:50,26:30)=0;

A(6:30,31:35)=0; A(36:50,31:35)=0;

A(1:10,36:40)=0; A(21:35,36:40)=0; A(41:50,36:40)=0;

A(1:20,41:45)=0; A(26:40,41:45)=0; A(46:50,41:45)=0;

A(1:5,46:50)=0; A(11:20,46:50)=0; A(26:45,46:50)=0;

## References


Baker, M.J., 2000. Selecting a research methodology. *The marketing review*, *1*(3), pp.373-397.

Balland, P.A., Belso-Martínez, J.A. and Morrison, A., 2016. The dynamics of technical and business knowledge networks in industrial clusters: Embeddedness, status, or proximity?. *Economic Geography*, *92*(1), pp.35-60.

Belso-Martínez, J.A., Mas-Tur, A. and Roig-Tierno, N., 2017. Synergistic effects and the co-existence of networks in clusters. *Entrepreneurship & Regional Development*, *29*(1-2), pp.137-154.



Belso-Martínez, J.A., Mas-Tur, A. and Roig-Tierno, N., 2017. Synergistic effects and the co-existence of networks in clusters. *Entrepreneurship & Regional Development*, *29*(1-2), pp.137-154.



Buchnea, E., 2016. Networks and clusters in business history. In *The Routledge companion to business history* (pp. 273-287). Routledge.

Cruijssen, F., Cools, M., and Dullaert, W., 2007. “Horizontal Cooperation in Logistics: Opportunities and Impediments”, *Transportation Research Part E: Logistics and Transportation Review*, 43(2): 129-142.

Cruijssen, F., Dullaert, W., and Fleuren,H., 2007. “Horizontal Cooperation in Transport and Logistics: A Literature Review”, *Transportation Journal*, 46(3): 22-39.

Daniel, P.S. and Sam, A.G., 2011. *Research methodology*. Gyan Publishing House.

Danloup,N., Allaoui, H., and Goncalves,G., 2013 "Literature Review on OR Tools and Methods for Collaboration in Supply Chain," *Proceedings of 2013 International Conference on Industrial Engineering and Systems Management (IESM)*, 1-7.

Das, T.K., and Teng, B.Sh., 1998. “Between Trust and Control: Developing Confidence in Partner Cooperation in Alliances”, *The Academy of Management Review*, 23(3): 491-512.

del-Corte-Lora, V., Vallet-Bellmunt, T. and Molina-Morales, F.X., 2015. Be creative but not so much. Decreasing benefits of creativity in clustered firms. *Entrepreneurship & Regional Development*, *27*(1-2), pp.1-27.

Dunnigan, K., 2008. Confidence interval calculation for binomial proportions. In *MWSUG Conference, Indianapolis, IN*.

Gilbert, E.N., 1959. “Random Graphs”, *Ann. Math. Statis*t, 30(4): 1141-1144.

Goddard, W. and Melville, S., 2004. *Research methodology: An introduction*. Juta and Company Ltd.

Hazra, A., 2017. Using the confidence interval confidently. *Journal of thoracic disease*, *9*(10), p.4125.

Janson, S., Rucinski, A., &Luczak, T. (2011). *Random graphs*. John Wiley & Sons.

Kaplan, S., and Sawhney, M., 1999, “B2B E-Commerce Hubs:Towards a Taxonomy of Business Models”, *Harvard Business Review*, 79(3): 97-100.

Kiwi, M. and Loebl, M., 2002. Largest planar matching in random bipartite graphs. *Random Structures & Algorithms*, *21*(2), pp.162-181.

Kothari, C.R., 2004. *Research methodology: Methods and techniques*. New Age International.

Kumar, R., 2018. *Research methodology: A step-by-step guide for beginners*. Sage.

Leonid Peshkin, 2007. *"Structure induction by lossless graph compression"*.

Loux,P., Aubry,M., Tran, S., and Baudoin, E., 2020. “Multi-sided Platforms in B2B Contexts: The Role of Affiliation Costs and Interdependencies in Adoption Decisions”, *Industrial Marketing Management*, 84: 212-223.

Molina-Morales, F.X., Belso-Martínez, J.A., Más-Verdú, F. and Martínez-Cháfer, L., 2015. Formation and dissolution of inter-firm linkages in lengthy and stable networks in clusters. *Journal of Business Research*, *68*(7), pp.1557-1562.

Mukul, G., 2011. *Research methodology*. PHI Learning Pvt. Ltd..

Noor, K.B.M., 2008. Case study: A strategic research methodology. *American journal of applied sciences*, *5*(11), pp.1602-1604.

Nyström, A.G. and Mustonen, M., 2017. The dynamic approach to business models. *AMS Review*, *7*(3), pp.123-137.

Piperopoulos, P.G., 2016. *Entrepreneurship, innovation and business clusters*. Routledge.

Ruxton, G.D. and Neuhäuser, M., 2013. Review of alternative approaches to calculation of a confidence interval for the odds ratio of a 2× 2 contingency table. *Methods in Ecology and Evolution*, *4*(1), pp.9-13.

Saenz, M., Ubaghs, E., and Cuevas, A., 2012, "Vertical Collaboration and Horizontal Collaboration in Supply Chain", *Enabling Horizontal Collaboration Through Continuous Relational Learning*, Springer, Cham, 7-10.

Shinkuma, R., Kasai, H., Yamaguchi, K. and Mayora, O., 2012, January. Relational metric: A new metric for network service and in-network resource control. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)* (pp. 352-353). IEEE.

Silva, G., Reis, L., Terceiro, A., Meirelles, P. and Kon, F., 2017, August. Implementing federated social networking: Report from the trenches. In *Proceedings of the 13th International Symposium on Open Collaboration* (pp. 1-10).

Simchi-Levi, D., Kaminsky, Ph., and Simchi-Levi, E., 2008. *Designing and Managing the Supply Chain : Concepts, Strategies, and Case Studies*, Boston: McGraw-Hill Education.

Tavasszy, L.A., Ruijgrok, C.J., and Thissen,M., 2003, "Emerging Global Logistics Networks: Implications for Transport Systems and Policies", *Growth and Change*, 34(4):456-472.

Turba, H., Breimo, J.P. and Lo, C., 2019. Professional and organizational power intertwined: Barriers to networking?. *Children and Youth Services Review*, *107*, p.104527.

Waclaw,B., Bogacz, L., Burda, Z., and Janke, W., 2008, "Monte Carlo Methods for Generation of Random Graphs", *Path Integrals — New Trends and Perspectives*, 342-345.

Wit, H.A.B. de., 2015. "Horizontal Collaboration: Road Transport of Containers", *Supply Chain Management*.

Zhao, X., Zhao, H., and Hou, J., 2010, "B2B e-hubs and Information Integration in Supply Chain Operations", *Management Research Review*, 33(10): 961-979.