Ca' Foscari
University
of Venice

Master's Degree

Data Analytics for Business and Society

Final Thesis

Spatial Proximity and Network Effects on Firm Growth in the Context of Network Contracts in Italy

Academic Year: 2023/2024

Supervisor:

Massimiliano Nuccio

Graduand:

Luca Xie

Matriculation Number: 873424

Contents

Abstract

This thesis explores the influence of spatial proximity and network effects on firm growth for firms participating in Italian Network Contracts from 2016 to 2023.

The datasets include data such as company and network identifiers, geographical details, network characteristics, and a range of firm level economic indicators. Key aspects of this analysis are the computation of a Localized Density metric, assessing the spatial concentration of firms within networks, and Closeness Centrality metrics, which evaluate a firm's centrality both in terms of geographic location and its position within the network structure.

An essential phase of the research involved rigorous data cleaning processes to ensure the accuracy and integrity of the dataset. Special attention was given to removing duplicates caused by discrepancies in naming conventions and data format inconsistencies and refining the data for precision in analysis. This cleaning process was pivotal in preparing the dataset for robust econometric modeling, which facilitates a deeper understanding of the roles and effectiveness of network contracts in enhancing firm-level economic performance.

Through this study, insights into the operational dynamics of network contracts and their strategic utility for business growth are explored, highlighting the relationships between spatial proximity characteristics and economic outcomes.

List of Tables

List of Figures

Chapter 1

Introduction and Data Overview

1.1 Network Contracts

Italy has a long tradition of inter-firm cooperation through industrial districts that thrived in the 1970s and 1980s [2]. However, globalization and rapid economic changes showed the flaws of these structures [3]. As a response, the Italian government introduced Network Contracts with the Decree law n.33/2009 and subsequent modifications to strengthen Italian firms' competitiveness and innovation through enhanced collaboration and synergy among firms.

Network contracts, known in Italy as "contratti di rete", represent a business framework designed to enable firms to pursue common projects and objectives without the need for merging their assets or identities with the main goal of increasing competitiveness on the market. More specifically, a network contract can be stipulated between firms with no limitations regarding dimensions, activity sector, geography, numbers of participants and legal nature (corporations, sole proprietors, partnerships, cooperatives, consortia etc.).

The networks can be of two types: "Rete Contratto" or "Rete Soggetto". The first allows companies to enhance their competitiveness and innovation capacity through collaborations without giving up their legal autonomy but bound contractually on a agreement to cooperate based on shared objectives. The agreement can include sharing resources, knowledge, and market strategies with the main focus as achieving mutual benefits. The latter can be considered as an evolution of the "Rete Contratto" where participants establish a new legal subject with its own legal personality beyond the mere contractual agreement. This subject can act independently under a unified management and is capa-

ble of owning assets in the name of the network, entering contracts binding the network as a whole rather than individual members, incur liabilities by taking debt or other financial obligations, and legal representation in legal proceedings, protecting the individual legal identities of its members.

In the regulatory framework, the parties involved in the stipulation of a network contract have to first prepare a network program, which is a general action plan aimed at increasing innovation capacity and competitiveness, and then concretely execute the activities outlined in the program, which can be of 3 types:

1. **Collaboration** between parties in areas relevant to the operation of their businesses
2. **Exchange** of information or services of any kind (industrial, commercial, technical and technological)
3. **Joint operation** among the parties of one or more activities that are part of the object of their businesses.

The primary advantage of entering into a network contract is the between firm collaboration, enabling smaller firms to operate at a scale similar to larger organization by pooling their resources and capabilities. Participants in a network can promptly face both domestic and foreign markets performing activities that may be too costly or risky for them to undertake alone like expanding their offerings, sharing costs, accessing to funding and non-repayable grants, enjoying tax benefits and competing to win public contracts.

In summary, the objective of networks contracts is to foster inter-firm collaboration, driving innovation and enhancing competitiveness while offering a legally secure and flexible framework that allows businesses of all sizes and sectors to pool resources achieving greater market success.

1.2 Background Literature

In recent years, studies on Italian Network Contract landscape has emerged providing a better understanding of the benefits and challenges associated with this framework.

Rubino and Vitolla (2018) examine the structural characteristics of network contracts and their impact on the performance of small firms in Italy, explicating the positive

impact of network size on financial performance although with low significance. Network diversity and geographical openness present challenges particularly due to coordination and integration between firms, highlighting the importance of strategic and managed network formations.

Network contracts also facilitate the internationalization of Italian firms by providing a structured framework for cooperation (Rubino, Vitolla, Garzoni, 2018). In this paper the importance of network characteristics are confirmed as influential factors on internationalization, key outcome is the importance of effective coordination given by the difficulties of managing large and geographically dispersed networks.

Leoncini, Vecchiato and Zamparini (2019) investigates on whether and how the Italian Network Contract Law has facilitated the fomation of networks among firms and improved their performance, concluding that it is effective in promoting the formation of cooperative networks leading to improved performance, particularly in innovation and competitiveness. However, also here the role of effective management, strategic partner selection are emphasized.

Cisi, Devicienti, Manello, Vannoni (2018) examine the benefits of network contracts on Italian SMEs highlighting the influence of size, geographical dispersion, sector diversity on firm performance. They found that participation in networks contracts leads to an increase in firm's gross margin ratios and export propensity but with no significant effect on profits.

The benefits of forming networks/clusters are confirmed also in contexts outside of the Italian Network Contract framework. Abdesslem and Chiappini (2019) provide a case study on the French optic photonic industry examining the impact of competitiveness clusters policies on firms performance. Using a difference-in-differences estimation they found that cluster policies significantly improve the financial and innovation performance of firms.

The empirical evidence on Italian network contracts converges on the benefit of formalizing networks on firm performance but shows varied impacts for network diversity and geography, emphasizing the importance of better coordination, hinting that spatial proximity of firms within a network can play an important role for better coordination and management.

Oerlemans and Meeus (2005) analyzed how organizational and spatial proximity

within networks affects firm performance in the Netherland's context. The paper's questions focused on understanding the extent to which proximity influences coordination capacity, managerial skill enhancement, and innovation through knowledge spillovers. Their findings showed that spatial proximity positively impacts these dimensions, thereby improving overall firm performance. Specifically, the study demonstrated that firms located closer to each other benefit from more frequent and informal interactions, leading to better coordination, enhanced managerial skills due to easier exchange of tacit knowledge, and increased innovation through facilitated knowledge spillovers.

While prior studies have demonstrated the benefits of network contracts and the challenges associated with geographical dispersion, there is a gap in understanding specifically how the spatial proximity of Italian firms participating in network contracts influence their economic performance.

Building on this framework, this thesis aims to explore the findings of the above empirical studies, addressing the research question: How does spatial proximity within network contracts affect firm growth in Italy?

By constructing a comprehensive panel dataset and analyzing the firm's performance using spatial and organizational proximity as proxies for network concentration and firm positioning within a network, as well as reflecting levels of interaction and knowledge exchange, while controlling for network characteristics and firm characteristics, the study tries to assess the influence of these elements on economic outcomes. The analysis will incorporate variables such as:

- Spatial proximity: Localized Density is used as a proxy reflecting the concentration of firms within a network, indicating the geographical closeness of member firms and its impact on performance through facilitated interactions and knowledge spillovers.
- Organizational proximity: Closeness Centrality as a proxy reflecting the positioning of a firm within a network, which measures how centrally a firm is positioned within the network structure, influencing its access to information and resources.
- Network characteristics: to mirror the structure of a network, variables such as network age, network participants, network hub (reference company), network legal status can be used as proxies reflecting the maturity, size, central connectivity, and formal legal engagement of the networks.

- Firm characteristics: firm-level proxies can be firm size, age, sector helping to control for differences between firms that might influence performance.

Integrating spatial proximity and organizational characteristics into the analysis of network effects on firm performance can provide deeper insights into how geographical distance impacts coordination and growth in the Italian context. Moreover, by interacting spatial and network characteristics it may be possible to gain more understanding of their combined effects on firm performance.

To achieve this, a comprehensive panel dataset is constructed including network-firm characteristics, firm financial and economic performance, geographic characteristics.
Three main datasets are utilized:

- Network Contract Dataset (DF): this dataset provides a detailed view of firms and their participation in network contracts, serving as the primary source for analyzing individual firm behavior within networks.
  - A secondary network dataset (DN) is derived from DF, this dataset solely focus on the characteristics and dynamics of each unique network.
- AIDA Dataset (AIDA): this dataset supplements the analysis with financial, registration, and commercial information about firms, enabling a comprehensive assessment of economic performance.
- Spatial Measures: derived from the cleaning and processing of the DF dataset, this collection includes geographical variables such as latitude and longitude, Localized Density, Average Distance and Centrality Measures. These metrics facilitate a deeper exploration of geographical and structural aspects of networks and their impact on firm performance.

Successive sections will explore each dataset and their variables, providing what each variable represent, how it is calculated and some descriptive statistics.

1.3 Network Contract Dataset

This section introduces the variables included in the Network Contract Datasets (DF and DN). Understanding these variables is pivotal for subsequent analyses, as they provide

insights into the legal, geographical, operational, and structural aspects of the networks and their member firms. The descriptions highlight the significance of each variable in exploring the dynamics of network contracts and their economic implications for firms in Italy from 2016 to 2023.

DF variables and their missing counts and percentages are presented in the following table:

Table 1.1: DF Table Variables and Their Missing Values

Variable	Description	NA	NA%
firm_taxcode	Unique tax code for each firm.	0	0%
network_name	Name or identifier of the network to which a firm belongs.	0	0%
year	Year of data entry, indicating when the information was recorded.	0	0%
municipality_firm	Municipality where the firm is located.	0	0%
province_firm	Province where the firm is located.	0	0%
region_firm	Region in Italy where the firm is located.	0	0%
region_group	Group of regions characterizing macro-areas in Italy.	0	0%
reference_company	Indicates if the firm is the main reference within the network.	0	0%
ateco_2007	ATECO 2007 code indicating the primary economic activity of the firm.	506	0.2%
Sector	Sector categorization of the firms main activity.	520	0.21%
liquidation_firm	Indicates whether the firm is in liquidation or not.	0	0%
identification	Unique identifier used for internal data management.	0	0%
number_of_networks	Number networks a firm is participating in.	0	0%
years_in_network	How many years a firm is in the network.	0	0%

The variable firm_taxcode uniquely identifies each firm , it also serves as identifier how analysis purposes, together with network_name and year can uniquely identify the composition of each network by year. In this dataset each firm can be present multiple times in the same year if they are participating in multiple networks in the same year.

The municipality variables is used for geographical purpose and it is also the key variable used to calculate the spatial measures for each firm. The reference company indicator indicates which firm in the network is the reference within network and that uploaded the bureaucratic documents into the "registro d'imprese" platform. Sector indicates which sector category each firm belongs to, using the ATECO code it is possible to retrieve the Standard Industry Classification code from the ISTAT platform. A futher categorization is employed unifying all the Sectors with less than 2% of the data. Identification identifies wether a firm was retrieved from the "Elenco" dataset or "Soggetto Giuridico" dataset, from this a differentiation between Legal Networks and non is performed. number_of_networks is calculated by counting the unique networks associated to each firm for each year. years_in_network is then calculated using act_date - year if the first appearance of the firm is before the year interval of our dataset (2016-2023), or first year of appearance - year if it entered in a network in the period between 2016 and 2023.

DN variables are the following:

Table 1.2: DN Table Variables and Missing Values

Variable	Description	NA	NA%
network_name	Name of the network.	0	0%
act_date	Official establishment date of the network contract.	0	0%
year	Year of the data entry.	0	0%
network_members	The count of members within a network.	0	0%
network_age	Representing the age of the network.	0	0%
liquidation_network	Indicates whether the network is on liquidation.	2809	6.66%
Legal_network	Identifies whether a network has autonomous legal subjectivity.	0	0%

To uniquely identify each network the variable network_name is used, it is the derived using the actual name of the network the value is not missing in the original dataset, in the case it is missing it is dervied using a combination of network taxcode and act number.

Network age is calculated as the difference between year and the act date, representing the age of the network. While network_members counts the members participating in a

network for each year.

Liquidation_network indicates if the network is on liquidation, it assumes value 1 when the name of the network originally possessed the regex "INLIQUID[A-Z]*", the reason of the 2809 missing values is for the missing values in the network name initially. Legal_network variable is defined as network that is funded with autonomous legal subjectivity, data retrieved by identifying any network that has at least one firm originating from "Soggetto Giuridico" dataset.

Overview of Network Contract Data Trends

Following the introduction of the datasets, this section explores the descriptive statistics, analyzing the data used in the research spanning from 2016 to 2023, highlighting the changes in behavior and evolution of firms and networks.

The tables presented in this section summarise the key characteristics of Network Contract environment from years 2016 to 2023, providing a quantitative overview of its key characteristics.

Table 1.3: Summary Statistics for the Network Contract Dataset

Statistic	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Cumulative Total Observations	13,754	31,659	56,543	89,887	127,233	168,636	214,166	266,437
Total Observations	13,754	17,905	24,884	33,344	37,346	41,403	45,530	52,271
Total Observations % change		30.18%	38.96%	34%	12%	10.66%	9.97%	14.61%
Unique Firms	13,002	16,875	23,319	31,363	34,841	38,276	42,076	47,243
Unique Firms % change		29.79%	38.19%	34.5%	11.09%	9.86%	9.93%	12.28%
Unique Networks	2,500	3,195	4,153	4,920	5,601	6,295	7,047	8,459
Unique Networks % change		27.8%	29.98%	18.47%	13.64%	12.39%	11.95%	20.04%

The Network Contract Dataset presents a total of 266,437 observations recorded from 2016 to 2023, it displays a clear upward trend over the years. Starting with 13,754 observations in 2016, the dataset expands annually reaching 17,905 in 2017, 24,884 in 2018, and 33,344 by 2019. The growth continues with 37,346 observations in 2020, further rising to 41,403 in 2021 and 45,530 in 2022. By 2023, the number of observations peaks at 52,271, marking a substantial increase from the initial count. The upward trend is an obvious effect of the growth of unique firms and unique networks engaged in this network contracts ecosystem. The count of unique firms participating in networks shows a steady increase from 13,002 firms to 47,243 firms participating in 2023. Unique networks being

created during this period also increases from 2,500 networks in 2016 to 8,459 networks in 2023.

The percentage changes show that there is a robust initial surge until 2019, seeing a moderation in the pace of growth from 2020 onwards, coinciding with the beginning of the COVID-19 pandemic and its economic impacts. The slight rise in 2023 hints at a gradual recovery of the activities after the slowdown caused by the pandemic.

Firms and Networks in Liquidation

Liquidations reflects a consistent increase (minimal) in both number and percentage of firms and networks in liquidation from 2016 to 2023, with a rise after 2019. This trend could be linked to the delayed economic impacts of the COVID-19 pandemic. Initial government support may have temporarily cushioned firms, but as the assistance waned, the true financial repercussions began to manifest, leading to increased liquidations. The spike in liquidation rates in later years, particularly 2021 and 2023, suggests that the cumulative effects of the pandemic, possibly compounded by Italy's slow bureaucratic processes, led to a more pronounced economic toll on firms and networks.

Table 1.4: Firms and Networks in Liquidation from 2016 to 2023

Statistic	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Firms in Liquidation	207	334	460	620	815	1,083	1,374	1,820
% of Firms in Liquidation	1.59%	1.98%	1.97%	1.98%	2.34%	2.83%	3.27%	3.85%
Networks in Liquidation	1	3	6	7	6	8	13	26
% of Networks in Liquidation	0.04%	0.094%	0.144%	0.142%	0.107%	0.127%	0.184%	0.307%

Network Participation

The Average Networks per Company indicates the mean number of networks that companies are part of each year. Starting from 2016 with 1.06, it shows a relatively stable trend with slight increases over time, reaching 1.11 in 2023.

The standard deviation represents the variability in the number of networks per company. From an initial low of 0.26 in 2016, meaning lower levels of disparity in terms of network participation, we see a peak in 2018 at 0.48 and the highest level reached in 2023 with 0.56. This indicates that during this period, some companies became more involved in multiple networks than others.

The median number of networks per company remains consistently at 1 throughout the years. This supports the idea that despite the increase in the number of networks for some companies, the most common scenario is that companies are involved in a single network. The sharp rise in standard deviation from 2017 to 2018 might indicate that in 2018 there was a formation or expansion of multiple networks, which pushed the average and variability higher. This is confirmed by the Max Networks per Company value, which spikes in 2018 from 12 to 48 and stabilizes at 54 from 2019 onwards.

A notable rise in multi-network engagement is evident after 2020, exceeding 6% in multi-network presence in 2020 and reaching levels as high as 8.34% in 2023. This behavior highlights a shift in how firms are leveraging networks for business operations, it might represent be a response to the challenging business climate during the pandemic, with firms seeking to strengthen their business through broader network connections.

Table 1.5: Network Participation per Firm from 2016 to 2023

Statistic	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Avg. Networks per Company	1.06	1.06	1.07	1.06	1.07	1.08	1.08	1.11
SD Networks per Company	0.26	0.29	0.48	0.47	0.47	0.5	0.49	0.56
Median Networks per Company	1	1	1	1	1	1	1	1
Min Networks per Company	1	1	1	1	1	1	1	1
Max Networks per Company	5	12	48	54	54	54	54	54
Multi-Network Firms (%)	5.26%	5.36%	5.57%	5.19%	5.71%	6.42%	6.6%	8.34%

Network Size

In terms of networks size there is a positive trend over the years. On the average, networks have 5.5 members in 2016 reaching a peak level of 6.78 members in 2019. However, a modest downturn is observed during and after the COVID-19 period with the average falling at 6.18 firms per network in 2023.

The median network size, consistently holding at 4 members across the years, demonstrates that the typical or central tendency of network composition has remained constant, largely unaffected by the broader economic shifts.

The standard deviation, on the other hand, increase from 5.48 to a peak of 10.24 signals growing variability, indicating that alongside many networks maintaining a size close to the average, there are networks that are either significantly smaller or larger.

This interpretation is further reinforced by the maximum network size, which expands

significantly from 87 members in 2016 to 271 in 2023. The growth in the maximum size confirms that, alongside the overall trend of network size consolidation as indicated by the average and median, there is also a concurrent trend of expansion within certain networks.

Table 1.6: Network Members Statistics from 2016 to 2023

Statistic	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Avg. Network Members	5.5	5.6	5.99	6.78	6.67	6.58	6.46	6.18
SD Network Members	5.48	6.27	7.82	10.24	10.12	10.04	10.15	9.85
Median Network Members	4	4	4	4	4	4	4	3
Min Network Members	1	1	1	1	1	1	1	1
Max Network Members	87	124	137	149	146	194	232	271

Network Age

In the data there is an evident progression in network ages, with the average age of networks increasing from 2.57 in 2016 to 5.06 years in 2022, denoting a general trend of network maturation. However, in 2023, a slight decrease in the average age to 5.01 years is observed. This marginal reduction could suggest a contained influx of newer networks or a shift in the lifecycle dynamics of the networks post-pandemic. The median age increase from 3 to 5 years over these years suggests a steady presence of more mature networks. An increase in the standard deviation suggests a diversifying age profile within the network landscape, with a increasing blend of emerging and well-established networks. The introduction of networks with an age of 0 in 2023 indicates the inclusion of networks founded in the same year as the data collection, while the progressive increase in the maximum network age, peaking at 13 years, highlights the sustained presence and potential influence of long-standing networks. The shift in minimum network age from 1 to 0 in 2023 may indicate a methodological update in data collection, transitioning from capturing networks with a minimum one-year establishment to including those formed within the same reporting year.

Sector

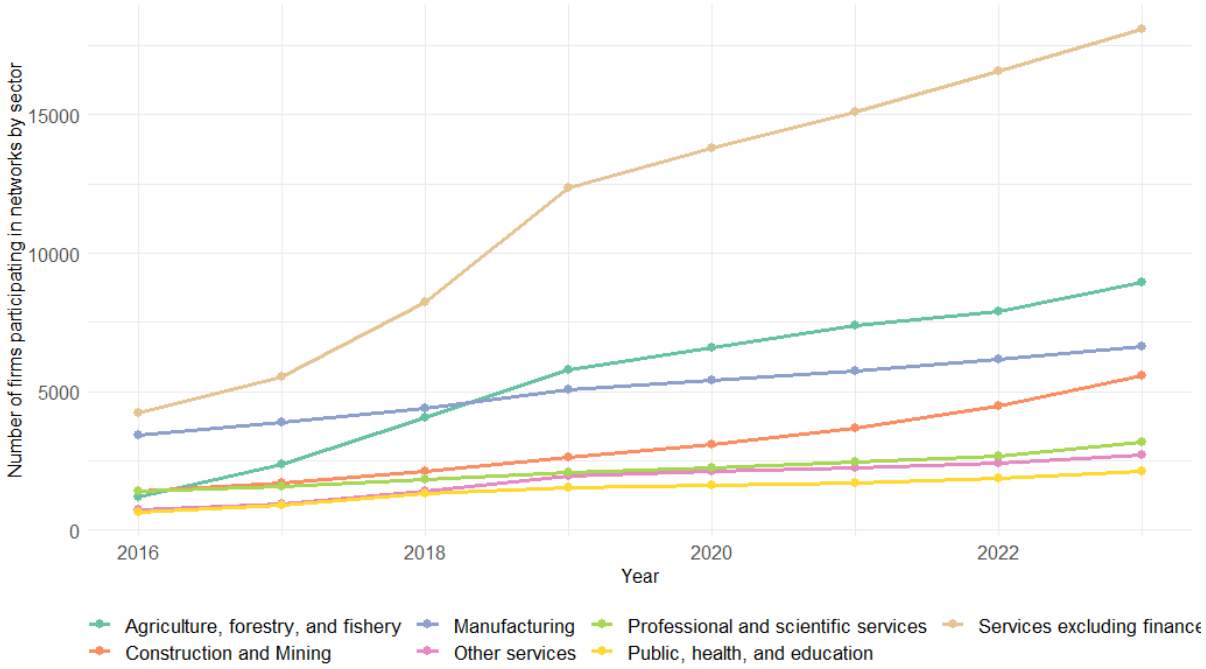
As shown in Figure 1.1 and Table 1.8, the distribution of firms participating in network contracts has seen a increasing trend over the years for all the sector. Especially for firms

Table 1.7: Network Age Statistics from 2016 to 2023

Statistic	2016	2017	2018	2019	2020	2021	2022	2023
Avg. Network Age	2.57	2.94	3.19	3.63	4.14	4.64	5.06	5.01
SD Network Age	1.25	1.53	1.82	2.03	2.25	2.51	2.8	3.33
Median Network Age	3	3	3	3	4	4	5	5
Min Network Age	1	1	1	1	1	1	1	0
Max Network Age	6	7	8	9	10	11	12	13

in the sectors of "Services excluding finance" and "Agriculture, forestry, and fishery" we can note a significant increase in the period between 2016 to 2019, with frequency raising from 4243 to 18120 and 1178 to 8955 respectively, reaching together more than half of the total firms participation in network (57,31%).

Figure 1.1: Yearly Distribution of firms per Sector



Firms in "Services excluding finance" is consistently the biggest sector in terms of participation in networks occupying from 32,63% of the total firms in 2016 with the peak at 39,42% in 2021 and 2022.

Firms in "Manufacturing" instead see a shrinking of the data portion from 26,26% of total data in 2016 to 14,05% in 2023.

Overall, all sectors face a increasing trend in terms of absolute number of firms participating in networks but some sectors shrink in terms of data proportion.

Table 1.8: Yearly Distribution of Firms Across Sectors Table

Sector	2016	2017	2018	2019	2020	2021	2022	2023
Agriculture, forestry, and fishery	1178	2384	4063	5792	6599	7392	7901	8955
	(9.06%)	(14.13%)	(17.42%)	(18.47%)	(18.94%)	(19.31%)	(18.78%)	(18.96%)
Construction and Mining	1414	1697	2109	2624	3094	3693	4483	5565
	(10.88%)	(10.06%)	(9.04%)	(8.37%)	(8.88%)	(9.65%)	(10.65%)	(11.78%)
Manufacturing	3414	3888	4391	5049	5407	5725	6171	6639
	(26.26%)	(23.04%)	(18.83%)	(16.10%)	(15.52%)	(14.96%)	(14.67%)	(14.05%)
Other services	736	917	1388	1932	2098	2223	2406	2692
	(5.66%)	(5.43%)	(5.95%)	(6.16%)	(6.02%)	(5.81%)	(5.72%)	(5.70%)
Professional and scientific services	1378	1584	1823	2071	2238	2449	2668	3151
	(10.60%)	(9.39%)	(7.82%)	(6.60%)	(6.42%)	(6.40%)	(6.34%)	(6.67%)
Public, health, and education	639	883	1321	1513	1606	1705	1860	2121
	(4.91%)	(5.23%)	(5.66%)	(4.82%)	(4.61%)	(4.45%)	(4.42%)	(4.49%)
Services excluding finance	4243	5522	8224	12382	13799	15089	16587	18120
	(32.63%)	(32.72%)	(35.27%)	(39.48%)	(39.61%)	(39.42%)	(39.42%)	(38.35%)

Sectors such as "Professional and scientific services" or "Manufacturing," which traditionally have higher levels of patents and innovation activities, might not exhibit as steep an increase in network participation. This could be due to these sectors already benefiting from strong internal capabilities and competitive advantages driven by their technological advancements and intellectual property. In contrast, sectors like "Services excluding finance" and "Agriculture, forestry, and fishery," which might have fewer opportunities for internal innovation or less access to cutting-edge technologies, may seek out network contracts as a strategic approach to access shared resources, new technologies, and market opportunities. Network contracts in these sectors could serve as crucial mechanisms to boost competitive parity and foster collaborative innovations that individual firms might not be able to achieve alone.

Geographical variables

The Table 1.9 shows the top five regions in Italy by the number of firms participating in network contracts for each year from 2016 to 2023.

It is possible to observe that from 2018 onwards Lazio leads the table by a a wide margin showing significant involvement of regional firms in networks. Lombardia and Veneto consistently appear in the top three throughout most of the years covered, indi-

cating strong and stable network contract environments in these regions. But the strong growth in numbers of network participation in Lazio is not replicated by any other region as they shower a slower and more linear growth.

It's also worth noting that starting from 2019, the Campania region surpasses both Emilia-Romagna and Toscana in terms of firms involved in network contracts, with 2,472 firms, despite not being in the top five before that year.

Table 1.9: Table by year of Top 5 Regions by firms in networks

Pos.	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
1	Lombardia	Lombardia	Lazio	Lazio	Lazio	Lazio	Lazio	Lazio
	(2444)	(2835)	(4008)	(8303)	(8900)	(9566)	(9960)	(10393)
2	Emilia	Toscana	Lombardia	Lombardia	Lombardia	Lombardia	Lombardia	Lombardia
	(1315)	(1675)	(3051)	(3308)	(3575)	(3915)	(4390)	(5069)
3	Toscana	Lazio	Veneto	Veneto	Veneto	Veneto	Veneto	Veneto
	(1279)	(1618)	(2026)	(2404)	(2751)	(3031)	(3327)	(3958)
4	Veneto	Emilia	Toscana	Campania	Campania	Campania	Campania	Campania
	(1104)	(1589)	(1823)	(2380)	(2636)	(2865)	(3155)	(3558)
5	Lazio	Veneto	Emilia	Toscana	Toscana	Toscana	Toscana	Toscana
	(1094)	(1467)	(1788)	(2075)	(2372)	(2642)	(2966)	(3301)

Figure1.2 compares the distribution of firms involved in network contracts across Italian regions in 2016 and 2023 providing a visualization of the geographical shifts and trends showcased in the 1.9, but also including all the rest of the regions. It is possible to observe immediately the concentration in Lazio as showed in the above table. In general all regions showed a increase in network participation, but other than the Lazio, Lombardia and Veneto most of the regions share similar frequency in terms of firm participation.

Table1.10 focus on the top 5 municipalities for number of firms participating in networks, it reveals that Rome, Milan, and Naples consistently appear among the top five for network participation, serving as regional capitals for Lazio, Lombardia, and Campania, respectively. This aligns with the broader regional data. However, the remaining prominent municipalities —mainly Turin (Piemonte region), Genoa (Liguria), and Bari (Puglia)— are from regions that do not consistently rank in the regional top five. This suggests that network participation in these areas might be heavily concentrated in these cities, whereas in other regions, such participation might be better distributed across multiple municipalities.

Figure 1.2: Regional distribution of firms in networks

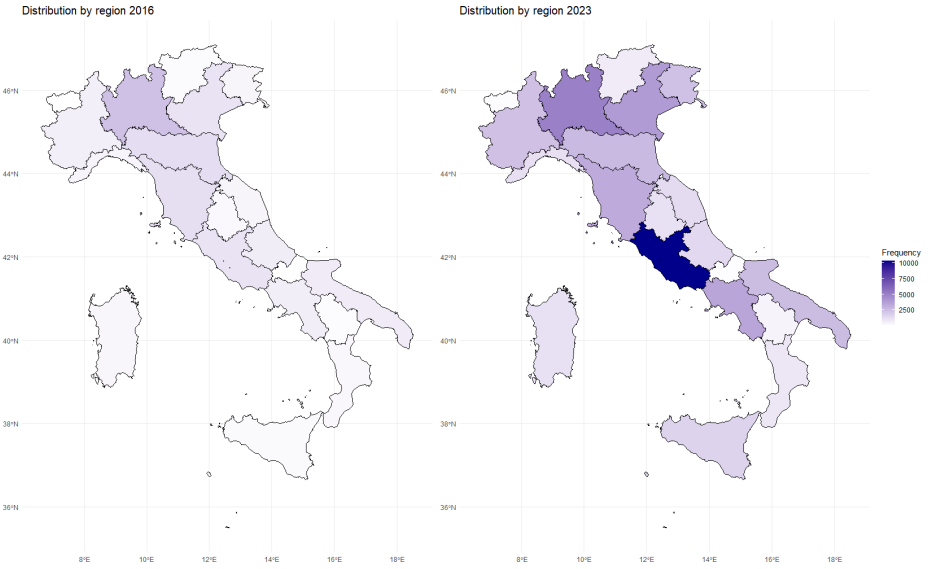


Table 1.10: Table by year of Top Municipalities by firms in networks

Pos.	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
1	Roma	Roma	Roma	Roma	Roma	Roma	Roma	Roma
	(666)	(863)	(1243)	(2808)	(3128)	(3571)	(3929)	(4888)
2	Milano	Milano	Milano	Milano	Milano	Milano	Milano	Milano
	(549)	(640)	(725)	(821)	(915)	(1040)	(1199)	(1481)
3	Torino	Torino	Torino	Torino	Torino	Torino	Torino	Torino
	(166)	(189)	(228)	(287)	(340)	(426)	(461)	(553)
4	Napoli	Napoli	Napoli	Napoli	Napoli	Napoli	Napoli	Napoli
	(156)	(207)	(284)	(342)	(396)	(439)	(514)	(630)
5	Genova	Genova	Genova	Genova	Genova	Genova	Genova	Genova
	(132)	(186)	(209)	(240)	(281)	(293)	(349)	(418)

Registered Firms

One important distinction working with this data set is the distinction between registered firms and non-registered firms, as the non-registered firms are not included in the AIDA database and will be excluded from further empiric analysis.

Table 1.11 highlights the proportion of unique registered firms for each year in DF dataset. Column "Total" is shows the total number of unique firms for each year, "Registered" represents the number of registered firms and "Proportion" is the percentage over the total firms, and Non_Registered is the rest of the firms. The proportion of registered firms participating in network contracts is more than half of the total number of firms, especially in the 2016, 2017 when the proportion was as high as 72.8% and 68.6%. The percentage shrinked in 2019 to 28.9% and recovered to 2018 levels in 2023 at 64.2%.

Table 1.11: Proportion of Registered and Non-Registered Firms by Year

Year	Total	Registered	Proportion (%)	Non-Registered
2016	13,754	10,007	72.8%	3,747
2017	17,905	12,284	68.6%	5,621
2018	24,884	15,542	62.5%	9,342
2019	33,844	19,374	58.1%	13,970
2020	37,346	22,005	58.9%	15,341
2021	41,403	24,928	60.2%	16,475
2022	45,530	27,954	61.4%	17,576
2023	52,271	33,559	64.2%	18,712

1.4 AIDA dataset

The AIDA dataset, sourced from the Analisi Informatizzata delle Aziende (Computerized Analysis of Companies) and managed by Bureau van Dijk, provides detailed information on Italian companies. This support dataset includes financial and commercial data about firms involved in the Network Contract Datasets. AIDA offers historical series of balance sheets for up to 10 years, covering the period from 2013 to 2022, and provides detailed demographic and financial information specifically on registered Italian firms, excluding unregistered ones.

Table 1.12 shows the various metrics included in the AIDA dataset, complete with a brief description and details on missing values, presented both as absolute numbers and percentages.

The percentages of NA values are higher respect to the Network Contracts dataset, this is due to the fact that the AIDA database provided balance sheets with non disclosed values in certain years for a subset of firms.

The high percentage of missing values of certain variables like 'Patent_Rights' and 'ROI' (46.38% and 44.62% respectively) suggest it is prudent to exclude these variables from the econometric modeling, as their inclusion might introduce bias. Instead, there are other firm performance variables with more complete data such as Liquid_Assets, Intangible_Assets, ROS, ROE, ROA.

The number of employees presents 11.43% of missing values, but the information loss is somehow mitigated by creating the variable AverageFirmSize bringing the missing value to 2.38% of the observations. AverageFirmSize is calculated with the average firm size of firms using n_employee considering the whole time interval from 2013 to 2022 assigning

Table 1.12: AIDA Variables and Missing Values

Variable	Description	NA	NA%
firm_taxcode	Unique tax identifier for each firm.	0	0%
year	Year of data entry.	0	0%
n_employee	Number of employees in the firm.	29652	11.43%
EBITDA	EBITDA.	2130	0.82%
Sales_Revenue	Total sales revenue.	428	0.16%
Net_Income	Net income after all expenses.	460	0.18%
Liquid_Assets	Total liquid assets.	1727	0.67%
Intangible_Assets	Total value of non-physical assets.	1717	0.66%
Patent_Rights	Value of patent rights held by the firm.	120299	46.38%
ROI	A measure of investment profitability.	115748	44.62%
ROS	A measure of operational performance.	26442	10.19%
ROE	A measure of financial performance.	22567	8.7%
ROA	A measure of asset profitability.	127	0.05%
innovative_startup	Innovative Startup = 1, otherwise 0.	0	0%
innovative_sme	Innovative SME = 1, otherwise 0.	0	0%
Year of foundation	Year of establishment of the firm.	40	0.02%
FirmAge	Age of the firm.	142	0.05%
FirmGrowth	Firms' growth rate from .	45470	17.53%
LabProd	Labor productivity.	32010	12.34%
AverageFirmSize	Average firm size of the firms.	6183	2.38%

"Micro" when the average size is smaller or equal to 10, "SMEs" when it is between 250 and 10, and "Large" when the average firm size is higher than 250. This way it is possible to assess the size of a firm assuming that firms do not change size abruptly from one year to another.

EBITDA, Sales_Revenue, Net_Income, Liquid_Assets, Patent_Rights and Intangible_Assets are all represented in thousand of Euros.

The founding year, which has only 0.02% missing values, is employed to calculate a firm's age, a variable critical for empirical analysis. This year represents the most recent registration, often updated during restructuring or re-registration of a company. Consequently, this can result in negative FirmAge values. To address this, firms with negative ages are assigned a value of NA. The assignment results in 142 missing values

for Firm Age, which is 0.05% of observations in the whole dataset.

Other variables that can be used as dummies in the empirical analysis (innovative_startup, innovative_sme) show 0% missing values ensuring reliable identification and categorization of innovative firms.

`FirmGrowth`: Defined as the difference in the logarithm of Sales Revenue between two consecutive periods, which measures the growth rate of sales revenue from one period to the next.

$$\Delta \log(\text{SalesRevenue})_t = \log(\text{SalesRevenue}_t) - \log(\text{SalesRevenue}_{t-1})$$

FirmGrowth has higher percentage of missing values (17.53%) than Sales_Revenue (0.16%) due to the fact that it is calculated on the difference between logarithm of Sales Revenue at time t minus logarithm of Sales Revenue at time t-1, so we lose the first year for all firms.

`LabProd`: Labor productivity, measured as revenue per number of employees. It presents higher missing values respect to n_employee because some firms present 0 employees.

$$\text{LabProd} = \frac{\text{Sales Revenue}}{\text{n\_employee}}$$

`AverageFirmSize`: average firm size of the firms, retrieved by calculating the average firm size of firms in the time interval and assigning "Micro" when the average size is smaller or equal to 10, "SMEs" when it is between 250 and 10, and "Large" when the average firm size is higher than 250.

By merging the dataset to the Network Contract Dataset, high number of missing values are composed by: Non-Registered firms from 2016 to 2023, all firms in 2023 (AIDA dataset does not have data about 2023) and variable specific NAs.

Table 1.13: Network AIDA Missing Values

Variable	NA	NA%
n_employee	159525	64.48%
EBITDA	149151	60.39%
Sales_Revenue	148246	60.02%
Net_Income	148266	60.03%
Liquid_Assets	148674	60.27%
Intangible_Assets	148674	60.27%
Patent_RightsR	191322	77.46%
ROI	191966	77.72%
ROS	157968	63.96%
ROE	157628	63.82%
ROA	148060	59.94%
innovative_startup	147967	59.91%
innovative_sme	147967	59.91%
year of foundation	147983	59.91%
FirmAge	148024	59.93%
FirmGrowth	153251	62.05%
LabProd	159525	64.48%
AverageFirmSize	149969	60.73%

Overview of AIDA dataset

Originally AIDA dataset is composed by 35.110 unique firms and 350.983 observations, of which 91.588 have missing values in all economic measures due to non-disclosure in the database or firms that closed that got brought up any way by the system, so the dataset shrinks to a total of 259.395 observations and 33.664 unique firms.

In this dataset the number of observations coincides with the number of unique firms, unlike the Network Contract dataset where firms could be part of multiple networks thus creating cross-combinations of the two.

The Table 1.14 shows that the number of firms participating in networks being registered in AIDA is growing yearly until reaching the pandemic years, possibly reflecting impact of the liquidations as saw in Table 1.4.

By combining the AIDA dataset with the DF dataset it is possible investigate the summary statistics of firms participating in networks from 2016 to 2022 (and not the

Table 1.14: Total Observations and Percentage Change by Year

Year	Total Observations	% change	Cum. Obs.
2013	21,326	NA	21,326
2014	23,274	9.13%	44,600
2015	24,545	5.46%	69,145
2016	25,561	4.14%	94,706
2017	26,597	4.05%	121,303
2018	27,380	2.94%	148,683
2019	28,022	2.34%	176,705
2020	28,130	0.39%	204,835
2021	28,036	-0.33%	232,871
2022	26,524	-5.39%	259,395

average metrics for firms entering networks atleast once ever). Here, the measures of year 2023 will be gone as there is no data of the year from AIDA.

Some of the key metrics statistics are described as followed:

Firm Age

Table 1.15 shows the descriptive statistics for the age of firms (in years) from 2016 to 2022, calculated annually.

- avg_FirmAge: This is the average age of firms for each year. The average age has gradually increased from 19.58 years in 2016 to 21.02 years in 2022. This increase could suggest either a growth in the longevity of firms over time or a decline in the number of new firms entering the market, making the average age of existing firms higher.
- sd_FirmAge: This represents the standard deviation of firm ages each year, which measures the variability or dispersion of firm ages from the average. The values fluctuate slightly but remain around 15 years, indicating a consistent spread in the ages of firms throughout the period.
- median_FirmAge: This is the median age of firms, showing the middle value of firm age when all are listed in order. It has stayed relatively constant at 16 or 17 years, suggesting that despite some variability in average ages, the central tendency of firm ages hasn't shifted dramatically.
- min_FirmAge and max_FirmAge: These values show the minimum and maximum ages of firms each year. The minimum age is consistently 0, indicating the presence

of newly established firms each year. The maximum age increases from 143 years in 2016 to 194 years in 2022, showing that some very old firms continue to operate, increasing the overall age range.

Table 1.15: Firm Age Statistics Over the Years

Year	Avg. Firm Age	Std. Dev.	Median Age	Min. Age	Max. Age
2016	19.58	14.65	16	0	143
2017	20.02	14.95	16	0	144
2018	20.14	15.31	17	0	190
2019	19.96	15.41	16	1	191
2020	20.27	15.42	17	0	192
2021	20.51	15.48	17	0	193
2022	21.02	15.68	17	0	194

Sales Revenue

The table presents revenue metrics for each year from 2016 to 2022, expressed in thousands of euros. The metrics provided for each year include: The average revenue, showing a fluctuating trend over the years with a notable increase in 2022 reaching 18.7 million euro. Considering also the median revenue, which is the middle value of revenue data, has followed a similar trend, showing the increase in 2022 is not only an outlier but a more generalized effect. The standard deviation of revenue varies significantly across the years, suggesting changes in revenue distribution among the sampled firms. The minimum revenue recorded, which is consistently 0 across all years, suggesting that there are firms or instances within each year that reported no revenue.

Table 1.16: Firm Age Statistics Over the Years (thousand euro)

Year	Avg. Revenue	SD Revenue	Median Revenue	Min Revenue	Max Revenue
2016	12,368.68	296,582.7	1,275.50	0	26,18,274
2017	12,858.71	295,675.1	1,252.34	0	28,575,407
2018	12,064.71	256,143.2	1,187.15	0	27,198,084
2019	11,253.49	210,796.4	1,084.83	0	24,370,111
2020	10,210.55	168,375.1	937.41	0	19,957,465
2021	12,161.66	164,348.2	1,185.97	0	21,923,105
2022	18,699.99	313,643.1	1,496.47	0	24,034,555

Table 1.17: Firm Growth by Year

Year	Avg. Growth	SD Growth	Median Growth	Min Growth	Max Growth
2016	-0.01	0.66	0.03	-12.37	9.47
2017	0.01	0.60	0.04	-11.12	11.03
2018	0.02	0.61	0.04	-9.66	6.19
2019	0.00	0.61	0.02	-12.96	11.81
2020	-0.21	0.73	-0.10	-9.62	8.39
2021	0.20	0.68	0.18	-10.65	11.77
2022	0.16	0.62	0.13	-11.23	9.03

Firm Growth

The table1.17 provides an overview of various statistical measures of firm growth for each year from 2016 to 2022 as percentage values.

- Avg. Growth: This represents the average growth rate each year. The data show fluctuations, with a noticeable dip in 2020 (-0.21 %) indicating a likely downturn, possibly due to external factors like economic recessions or pandemics. However, there is a recovery in the following years, particularly in 2021 (0.20%), suggesting a rebound.
- SD Growth: The standard deviation of growth shows the variability of growth rates around the average. A higher standard deviation in 2020 (0.73%) correlates with the significant negative average growth, highlighting a year with high volatility in growth rates among the observed entities.
- Median Growth: The median growth values are relatively stable over the years, slightly fluctuating around 0.03% to 0.04%, except in 2020 where it dips to -0.10%, matching the negative trend seen in the average growth.
- Min Growth: The minimum growth figures show the worst growth rates each year. The lowest point is -12.96% in 2019, while other years also show significant negative minimum growth.
- Max Growth: Conversely, the maximum growth rates provide insight into the best-performing firms. There's a wide range of maximum growth, peaking at 11.81% in 2019, which suggests that despite some challenging conditions, there were high-performing outliers every year.

Firm Size

Table 1.18: Employee Statistics by Year

Year	Avg. Employees	SD Employees	Median Employees	Min Employees	Max Employees
2016	53.71	459.03	12	1	33494
2017	55.61	455.07	12	1	32988
2018	56.40	452.18	12	1	32737
2019	55.32	416.83	11	1	31984
2020	56.28	437.97	11	1	37036
2021	57.24	429.98	12	1	36433
2022	71.29	1001.95	13	1	106653

Looking at the statistics of number of employees (table 1.18), the average number of employees has gradually increased from 53.71 in 2016 to 71.29 in 2022. The standard deviation of the number of employees shows some fluctuations, indicating variability in firm sizes. It peaked at 1001.95 in 2022, suggesting a significant increase in the disparity of firm sizes that year. The median number of employees has remained relatively stable, ranging between 11 and 13, indicating that most firms are small to medium-sized. The minimum number of employees per firm has consistently been 1, showing the presence of very small firms throughout the period. The maximum number of employees per firm saw a notable rise from 33494 in 2016 to 106653 in 2022, indicating substantial growth in the largest firms. In 2022, there was a significant variation in employee statistics, likely caused by outliers. Specifically, the entry of the Italian firm Poste Italiane into the dataset with its 106,653 employees dramatically impacted the statistics, skewing the average and standard deviation upwards.

The table1.19 shows the distribution of registered firm sizes from 2016 to 2022. The count for all firm sizes has roughly doubled, while their proportions have remained stable:

- Large Firms: The count increased from 228 in 2016 to 509 in 2022. The proportion has remained around 2.6-2.8%.
- SMEs: The count rose from 4089 in 2016 to 8911 in 2022. Their proportion fluctuated slightly, staying around 47-51%.
- Micro Firms: The count went from 3779 in 2016 to 9311 in 2022. Their proportion varied between 46.6% and 50.1%.

Table 1.19: Firm Size Distribution by Year

Year	Firm Size	Count	Proportion
2016	Large	228	0.0286
	SMEs	4089	0.5060
	Micro	3779	0.4668
2017	Large	263	0.0268
	SMEs	4912	0.5091
	Micro	4630	0.4728
2018	Large	323	0.0268
	SMEs	5881	0.4884
	Micro	5837	0.4848
2019	Large	388	0.0256
	SMEs	6862	0.4688
	Micro	7387	0.5046
2020	Large	439	0.0270
	SMEs	7606	0.4730
	Micro	8034	0.4960
2021	Large	474	0.0269
	SMEs	8308	0.4718
	Micro	8228	0.5013
2022	Large	509	0.0272
	SMEs	8911	0.4754
	Micro	9311	0.4970

In summary, although the percentage increase is small, micro firms have grown from 46.68% of the dataset in 2016 to 49.7% in 2022. Large firms have decreased from 2.86% in 2016 to 2.72% in 2022, with the lowest at 2.56% in 2019. SMEs have shrunk from 50.60% in 2016 to 47.54% in 2022.

1.5 Spatial Measures

The Spatial Measures dataset supplements the main dataset by providing distance and geographic variables that are key for the calculation of closeness and localized density metrics. The dataset provides information on distances between firms and their geographical coordinates, allowing us to assess how proximity affects the density of connections and network centrality. Importantly, these measures are calculated at the network level, focusing on firm-network combinations rather than individual firms.

This approach helps us analyze the collective behavior of firms within networks and understand the impact of spatial factors on business performance. The table below details the network and centrality variables in the dataset, along with any missing values. These

variables will be used as independent variables in our research.

The following variables are calculated:

Table 1.20: Network and Centrality Variables with Missing Values

Variable	Description	NA	NA%
firm_taxcode	Unique tax code identifier for each firm.	0	0%
network_name	The identifier of the network to which a firm belongs.	0	0%
year	The year the data was recorded.	0	0%
lat	Latitude coordinate of the firms location.	0	0%
lon	Longitude coordinate of the firms location.	0	0%
avg_distance	Average distance between the firms within the same network.	223	0.08%
LD	Measures the density of a firm's connections within its network based on geographic proximity.	0	0%
centrality_star	Measures the centrality of a firm in the network assuming a star network configuration.	73505	27.59%
centrality_complete	Measures centrality assuming a complete network configuration.	223	0.08%

Average Distance

The average distance between firms within each network is calculated to assess the geographical spread. The foundation of this calculation, as well as the calculations for Localized Density (LD) and Closeness Centrality in complete networks, is the distance matrix. The distances are calculated using the latitude and longitude metrics retrieved by matching the municipality of the firm and the ISTAT provided coordinates. All distances are converted in kilometers and the average is calculated by the sum of distances between each pair of firms divided by the number of pairs.

Localized Density

Localized Density (LD) is a measure used to evaluate the concentration of firms within a network based on their geographical proximity. It helps to understand the spatial

distribution of firms and how closely they are located to each other. The formula for calculating the Localized Density for each firm is given by:

$$LD_i = \sum_{j=1}^{n} \frac{1}{1 + D_{ij}}$$

Where:

- LD_i = Localized Density for firm i
- D_{ij} = distance (in km) from node i to node j
- n = total number of nodes/firms in the network

A high LD value indicates that network i is surrounded by many other firms at close distances within the network. Implying strong local connectivity or clustering. A low value of indicates firm i is surrounded by fewer firms or at greater distances. Low localized density, implies sparse network geographical presence.

Centrality Measures

Centrality measures are fundamental metrics in network analysis, used to identify the most important nodes within a network. For the purpose of this research Closeness Centrality will be used, assuming in two types of network settings: star network and complete network. Closeness Centrality measures how close a node is to all other nodes in the network. A node with high closeness centrality can quickly interact with all other nodes, making it influential and central within the network structure.

- **Closensess Centrality - Star Network**

  The star network is a specific type of network configuration where a central node, known as the star node, is connected to all other nodes, referred to as non-star nodes. This star node is identified using the reference firm variable from the Network Contract dataset. In this structure, the star node acts as a hub, directly connected to every other node, while the non-star nodes are only connected to each other through the star node.

  - `Star nodes closeness centrality:` For the star node, the closeness centrality is calculated based on its direct connections to all other nodes. The

formula for the star node's closeness centrality is:

$$C(v_s) = \frac{1}{\sum_{i \neq s} 1}$$

– `Non-Star nodes closeness centrality:` For non-star nodes, the closeness centrality reflects their indirect connections through the star node. The formula for non-star nodes' closeness centrality is:

$$C(v_i) = \frac{1}{2n - m - 2}$$

Where:

– i indicates if the node v is a non-star node.

– s indicates if the node v is a star node.

– n is the number of nodes in the network.

– m is the number of stars in the network.

- **Closensess Centrality - Complete Network** The second setting used in this research is the complete network configuration, where every node is directly connected to every other node. Unlike the star network, in a complete network, the edges between nodes are weighted by the actual distance between them. Closeness centrality in a complete network measures how easily a node can reach all other nodes, considering the physical distances involved.

  The formula is:

$$C(v) = \frac{1}{\sum_{u \in V \setminus \{v\}} w(v, u)}$$

  Where:

  – V is the set of nodes.

  – v is the node we are considering.

  – u is a node in the set of nodes.

  – $w(v, u)$ is the distance (in km) between node v and node u.

  If two firms are located in the same municipality, a distance of 1 km is assigned instead of 0 km to ensure calculations can be performed.

The closeness star configuration presents a higher number of missing values (73505) due to the fact that in the original dataset not all networks had a reference firm probably due to imputation errors. While this problem does not concert the complete network configuration as it only needs the distance matrix and the network setup.

Average Network Distance

The table1.21 displays the average network distance metrics for each year from 2016 to 2023.

It shows a decreasing mean average network distance from 2016 (73.73km) to 2019 (62.14km), and an increasing trend from the 2020 period to 2023 with a peak at 83.21km. This may be explained by the adoption of digital communication tools during the covid video that facilitated the collaborations with more distant firms. Same trends are reflected by the median distance.

The central measure's trend are the same for the standard deviations, suggesting that before covid firms were getting closer to each other physically and after covid the variability increased, firm within the network are becoming more spread out.

The minimum distance is 0 for each year, a network has 0 average distance when the firms participating in the network are from the same municipality.

The maximum average distance of networks confirms the increase in the geographical spread of the network or inclusion of new, more distant members.

Table 1.21: Average Network Distance Metrics by Year

Year	Mean Avg Network Dist.	SD Distance	Median Distance	Min Distance	Max Distance
2016	73.73	122.89	27.01	0	1032.59
2017	73.28	123.79	25.93	0	1047.48
2018	67.86	121.20	22.85	0	1086.57
2019	62.14	117.18	19.24	0	1081.04
2020	66.71	123.00	20.22	0	1081.79
2021	70.98	127.15	21.46	0	1081.79
2022	76.20	131.83	23.50	0	1081.04
2023	83.21	138.48	26.06	0	1277.57

Localized Density

The average localized density has shown significant changes over the years. Starting at 3.45 in 2016, it peaked at 10.60 in 2019 before decreasing to 7.18 in 2023. This fluctuation indicates periods of increased and decreased clustering of firms within the network. The peak in 2019 suggests that during this year, firms were more densely clustered, enhancing local interactions and connectivity. The median localized density also increased from 1.01 in 2016 to 2.00 in 2019, then decreased to 1.31 in 2023. This measure represents the central point of localized density and follows the trend of the average LD, the changes in the mean are not caused by entrance of outliers but is a general trend.

The minimum localized density remained consistently at 0 throughout the years. This consistency implies that there were always some firms with no neighboring firms in close proximity, indicating persistent isolation within the network. Most common case is the network with only 2 firms from the same municipality.

The maximum localized density increased from 86.00 in 2016 to 131.75 in 2023. This rise indicates that the most clustered firms became even more densely packed over time, suggesting the development of very dense clusters in the network, in line with the increasing network members statistics showed in table 1.6.

Table 1.22: Localized Density Metrics by Year

Year	Avg. LD	SD LD	Median LD	Min LD	Max LD
2016	3.45	9.24	1.01	0	86.00
2017	3.65	8.86	1.03	0	85.00
2018	6.08	13.21	1.15	0	86.00
2019	10.60	18.85	2.00	0	95.11
2020	9.83	18.22	1.72	0	92.11
2021	8.95	17.05	1.58	0	92.31
2022	8.21	15.99	1.46	0	87.33
2023	7.18	15.17	1.31	0	131.75

Closeness Centrality - Star

The average closeness centrality shows a increasing trend but stayed quite consistent at around 0.21-0.22, indicating a stable overall network structure where the star node consistently maintains its central position. The standard deviation increased from 0.2464 in 2016 to 0.3030 in 2023, reflecting growing variability among nodes' centrality. This

suggests that while some nodes have maintained or increased their centrality, others have become less central over time. This growing disparity can be attributed to the expanding network size, where the central position of the star node remains fixed, but the distances to non-star nodes may vary more widely. The median closeness centrality decreased from 0.11 to 0.09, implying a slight decrease in the centrality of the typical node, can be explained by the increasing network size, but only a few firms are a star node. The minimum closeness centrality is very low with decreasing trend, from 0.0093 to as low as 0.0019. This indicates presence of networks with elevated number of participating firms but only 1 or few star nodes. The maximum closeness centrality was consistently 1, structural centrality value for star nodes.

Table 1.23: Closeness Centrality - Star Configuration by Year

Year	Avg. Centrality	SD Centrality	Median	Min Centrality	Max Centrality
2016	0.2090	0.2464	0.1111	0.0093	1
2017	0.2162	0.2678	0.1111	0.0062	1
2018	0.2172	0.2837	0.1111	0.0047	1
2019	0.2180	0.2836	0.0909	0.0037	1
2020	0.2194	0.2910	0.0909	0.0035	1
2021	0.2180	0.2920	0.0909	0.0026	1
2022	0.2214	0.2960	0.0909	0.0022	1
2023	0.2236	0.3030	0.0909	0.0019	1

Closeness Centrality - Complete

The closeness centrality statistics for the complete network show varying trends in average, standard deviation, median, minimum, and maximum values from 2016 to 2023.

The average closeness centrality increases from 0.0634 in 2016 to 0.0745 in 2023, indicating that, on average, nodes have become slightly more central over time. The standard deviation also increases from 0.1811 in 2016 to 0.2066 in 2023. This rising variability indicates that there is a growing disparity in how central different nodes are, with some nodes becoming significantly more central while others lag behind.

The median closeness centrality, however, decreases from 0.0059 in 2016 to 0.0043 in 2023. This decrease suggests that while the average centrality has improved, the typical (median) node is becoming less central relative to the overall network. This disparity between the average and median values implies that the increase in centrality is not

uniform across all nodes. Instead, a few nodes are becoming much more central, raising the average, while many nodes remain less central, pulling the median down.

The minimum closeness centrality consistently remains near 0 (smallest closeness complete is 0.00000368), indicating the networks with very high number of firms and distant to the node.

The maximum closeness centrality starts at 1.0000 from 2016 to 2018 and then increases to 1.0921 from 2019 onwards. This change suggests that the most central nodes have become even more central, which might be due to changes in network distances or the addition of highly central nodes, usually small networks with not distant firms have higher closeness centrality.

Table 1.24: Closeness Centrality - Complete Configuration by Year

Year	Avg. Centrality	SD Centrality	Median Centrality	Min Centrality	Max Centrality
2016	0.0634	0.1811	0.0059	0	1.0000
2017	0.0659	0.1880	0.0056	0	1.0000
2018	0.0700	0.1983	0.0060	0	1.0000
2019	0.0635	0.1912	0.0056	0	1.0921
2020	0.0665	0.1957	0.0054	0	1.0921
2021	0.0687	0.1996	0.0051	0	1.0921
2022	0.0708	0.2027	0.0048	0	1.0921
2023	0.0745	0.2086	0.0043	0	1.0921

Chapter 2

Modeling Economic Performance

The second chapter of the thesis is dedicated to modeling the economic permance of firms, exploring various statistical techniques to investigate the impact of network contracts on firm growth in Italy. The goal is to understand how various network attributes, specifically Localized Density and Closeness Centrality, influence business performance.

The analysis begins by examining the key variables from the datasets discussed from the previous chapter. The main goal is to retrieve different categories of control variables: network controls, firm-specific controls, sector controls, and geographic controls. These controls will help isolating the effect of network characteristics from other factors that may influence firm performance.

Successive section of the chapter will go through a series of models with different configurations, starting from Ordinary Least Squares (OLS) models to Least Absolute Deviations (LAD) models. This progression provides an overview of the general effects of network participation on firm performance across the dataset helping to test the relationships between the explanatory variables and dependent variables.

Subsequently, panel data techniques will be used to account for individual differences among firms. This approach leverages the panel structure of the data, capturing entity fixed effects to control for unobserved heterogeneity and and to control for potential correlation between entity-specific effects on predictors.

2.1 Distribution and Correlation

This section focus on the refinement of the dataset to accurately reflect each firm's network environment and on selection of variables accounting for problems like multi-collinearity.

Network Contracts Dataset is used as the base to which firm performance metrics and spatial measures are added creating a dataset with 266,437 observations and 38 columns. An initial part ensures that each firm is associated with unique network values for each year. Particularly critical for firms that participate in multiple networks as their performance metrics must reflect the combined influence of the networks it participates in. To achieve this, the network level variables are transformed into firm level variables, capturing eventual changes in network characteristics:

- **Localized Density:** For firms involved in multiple networks, the Localized Density is calculated as the weighted average of the density values of each network the firm participates in, with weights based on network age to account for the influence of older, more established networks.
- **Closeness Centrality:** For firms in multiple networks, closeness centrality (in both star and complete configuration) is computed as the weighted average of the centrality values across all networks, with weights based on both the age of the networks and the number of connections (firms connected) in each network.
- **Hub:** Hub is assigned to 1 for firms that are reference company of at least 1 network it participates in.
- **Average Network Age:** It is calculated by averaging the ages of each network associated with a firm, weighted by the firm's involvement in these networks.
- **Networked Firms Count:** It is determined by counting the unique firms across all networks a firm is part of, reflecting the breadth of its network connections. This captures the extent of a firm's network and its potential access to diverse resources and information.
- **Legal Network:** Legal Network is assigned to 1 for each firm that is member of at least 1 legal network.

An additional Average Growth variable is created to mitigate the loss of information when there are missing values for firms in certain years. The Average Growth value is

calculated for each firm by the mean of available firm growth from 2016 to 2022 and assigned for all years.

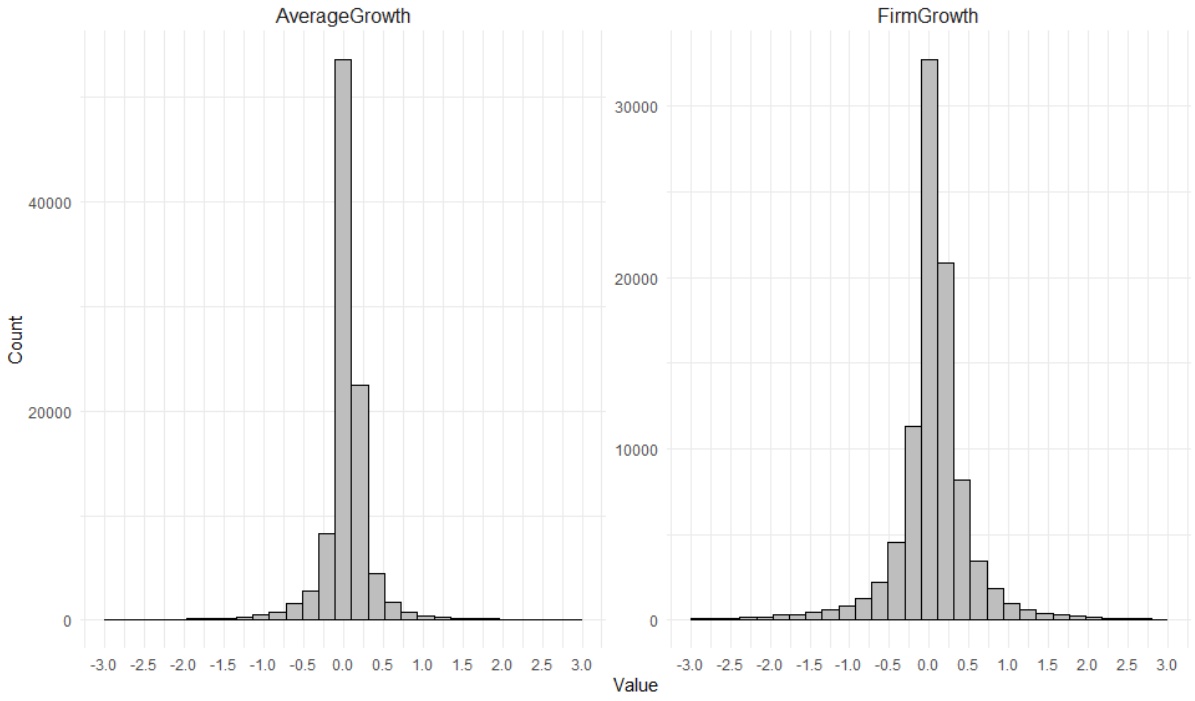
The subsequent step ensures that each firm appears only once per year in the dataset. This step reduces the original dataset from 266,437 observations to 246,995 observations. Additionally, AIDA database did not provide any financial metric for firms for the year 2023, so this year is also excluded from the dataset further reducing the dataset to 199,752. At the end, firms that are categorized under the region group "Estero" (abroad) which only account for 49 observations, are also filtered out.
After the adjustments, the final dataset contains 199,703 observations.

2.1.1 Variables Distribution

Firm Growth

Figure 2.1 provide distribution plots for firm performance variables, the plots are zoomed in to range form -3 to 3 for clearer visual of the main distribution characteristics, as noted in table 1.17 that the most extreme values goes to as low as -12% and as high as 11%.

Figure 2.1: Distribution of Firm Growth (zoomed x axis)



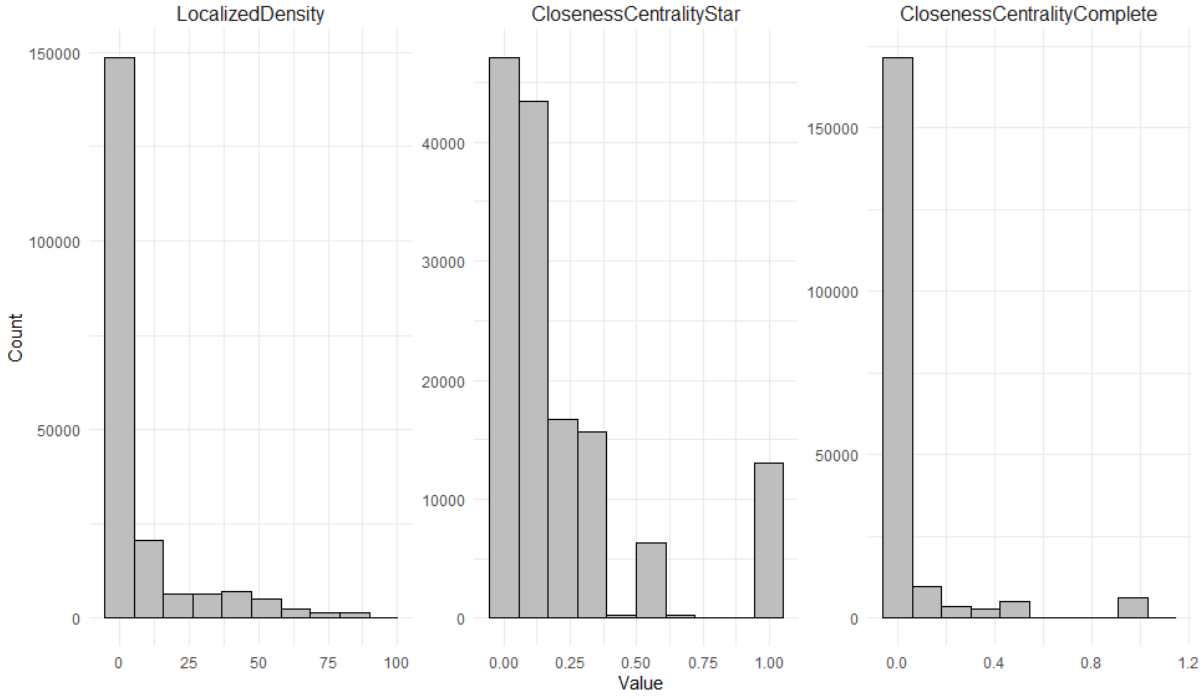
Both "AverageGrowth" and "FirmGrowth" have a pronounced peak at zero and lack of lack of skewness. This indicates that the majority of firms in the dataset have growth

rates close to zero, confirming the table 1.17 , implying minimal growth or decline. The central concentration is very high, suggesting that stable growth is the most common outcome for firms with most growths between -0.5 and 0.5.

Spatial Metrics

Figure 1.8 display the distributions of spatial metrics, in order: Localized Density, Closeness Centrality in Star network configuration and Closeness Centrality in Complete network configuration.

Figure 2.2: Distribution of Spatial Metrics



The majority of firms have very low localized density values, with a large spike around zero. This indicates that most firms are not surrounded by many other firms within close geographical proximity. The high count at zero suggests that many firms are geographically isolated within their networks. The instances with much higher LD values indicate the presence of densely packed clusters with close geographic distance. In Star network configuration there is a clear distinction between non-star nodes with low centrality and star nodes with higher centrality, creating two clusters of firms. The histogram reflects this characteristic and shows two prominent peaks: one at very low values and another at higher values near 1. In complete networks every firm is connected to each other in the same network but the centrality is weighted by the distance between each node. The

pattern in this case in very different from the star network and has some resemblance with the LD patterns. The high number of firms with very low centrality measures means that despite all firms being theoretically connected, geographic distance might reduce the strength of interactions. Consequently, firms tend to have low effective centrality because they are not truly close in a practical sense to all other firms.

Network Variable Distribution

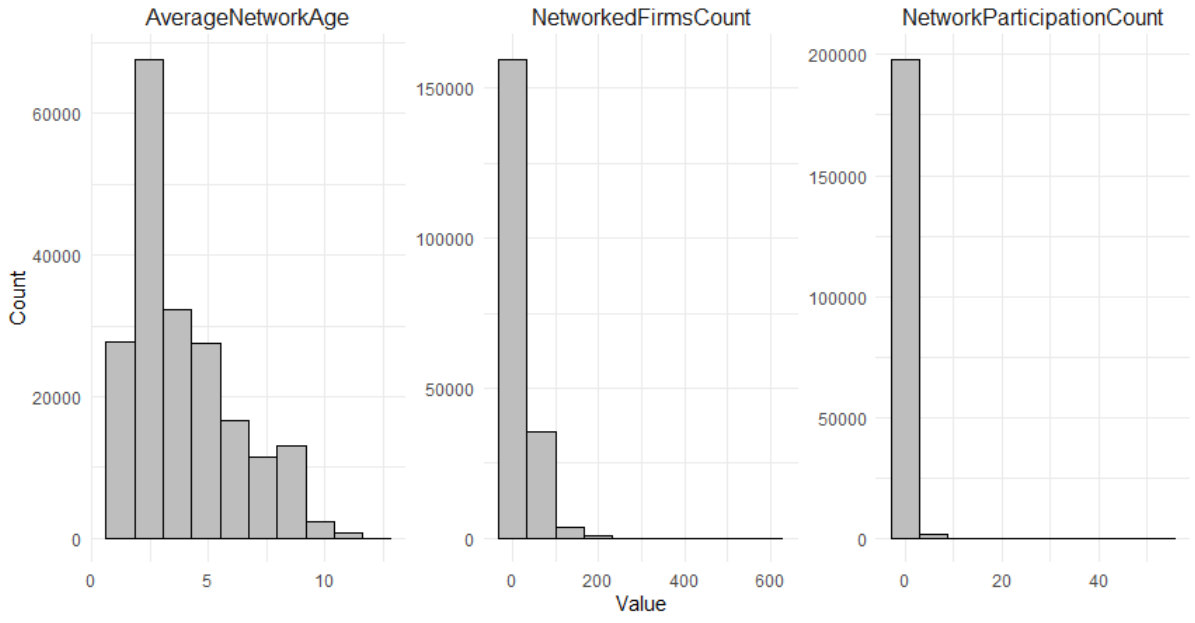
All network variable histogram plots are right skewed, most of the data are clustered in the initial part of the metric with few outliers with higher values.

For Average Network Age the majority of the data clustered towards the lower end of the age range. Most of the networks are young and have below 5 years of age.

In Networked Firms Count, the each firm is connected with a small number of firms. The fewer data points at the higher end indicate that it is less common to have networks with a large number of participating firms. In line with table 1.6 showing the median at around 4 members per network.

The Network Participation Count is extremely skewed, indicating most firms participate in a few networks. This suggests that firms prefer to limit their collaborations, potentially to maintain focus or due to resource constraints (as in table 1.5).

Figure 2.3: Network Variable Distribution plots

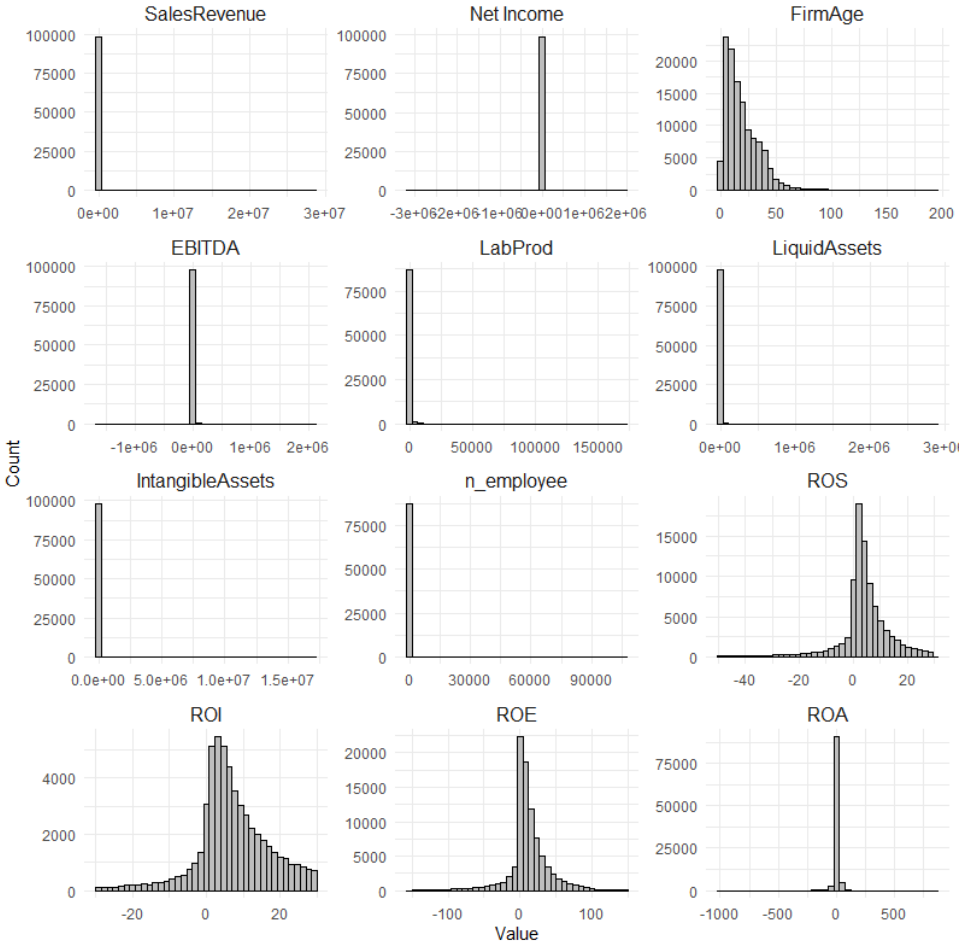


Firm Variable Distribution

Similar to Network variables, Firm variables are also very right skewed. Sales Revenue, Net Income, EBITDA, Labor Productivity, Liquid Assets, Intangible Assets, and Number of Employees, are extremely right-skewed distributions. This means that the majority of firms concentrate on the lower values for these metrics, with only a few outliers achieving significantly higher values.

Firm Age shows a right-skewed distribution, showing that most firms have less than 50 years. The Return on Sales (ROS), Return on Investment (ROI), Return on Equity (ROE), and Return on Assets (ROA) show varying distributions. ROS, ROI and ROE have approximately normal distributions with right skewness, very noticeable for ROI, indicating more firms with higher positive returns and fewer firms with negative returns but with greater magnitude. ROA is extremely peaked around zero, with very low variance.

Figure 2.4: Firm Variable Distribution plots



Categorical Variables

Categorical variables in this dataset are: Hub, AverageFirmSize, InnovativeSME, InnovativeStartup, LegalNetwork, ProvinceCapital, RegionGroup, RegionCapital and Sector. The proportion of these variables are in table 2.1 (the table only accounts for non-NA categories for each variable, excluding the NAs):

Table 2.1: Categorical Variables

Variable	Value	Count	%
Hub	0	172,883	86.6
	1	26,820	13.4
AverageFirmSize	SMEs (employees > 10 and 250)	49,615	49.54
	Micro (employees 10)	47,575	47.50
	Large (employees > 250)	2,960	2.96
InnovativeSME	0	118,844	98.86
	1	1,373	1.14
InnovativeStartup	0	120,044	99.86
	1	173	0.14
LegalNetwork	0	123,933	62.10
	1	75,770	37.90
ProvinceCapital	0	132,320	66.26
	1	67,383	33.74
RegionGroup	Center	68,318	34.21
	North-east	42,540	21.30
	South	39,965	20.01
	Nord-west	38,571	19.31
	Isole	10,309	5.16
RegionalCapital	0	161,388	80.81
	1	38,315	19.19
Sector	Services excluding finance	75,843	37.98
	Agriculture, forestry, and fishery	35,276	17.66
	Manufacturing	34,045	17.05
	Construction and Mining	19,104	9.57
	Professional and scientific services	14,208	7.11
	Other services	11,700	5.86
	Public, health, and education	9,527	4.77

In the final dataset, considering all the years, the total number of observations with Hub characteristic (reference firm) are 26,820, being 13.4% of the total records, while 86.6% are non-hub.

Almost half of the observations with available data are SMEs (49.54%), followed by Micro firms (47.50%), and a small proportion of Large firms (2.96%).
For innovative SME and Startups there is a big unbalance with only 1.14% of the data being innovative SMEs and 0.14% being innovative Startups.
62.10% of the observations are not part of a legal network, while 37.90% participate in at least 1 legal network.
About geography, 33.74% of the observations are located in provincial capitals, 19.19% are in regional capitals. In terms of region groups, most of the data come from the Center with 34.21% followed by North-east and South with 21.3% and 20.01%.
The most populated sector of the dataset are the "Services excluding finance" with 37.98% of the observations, followed by "Agriculture, forestry and fishery" and "Manufacturing" with 17.66% and 17.05%.

2.1.2 Correlation

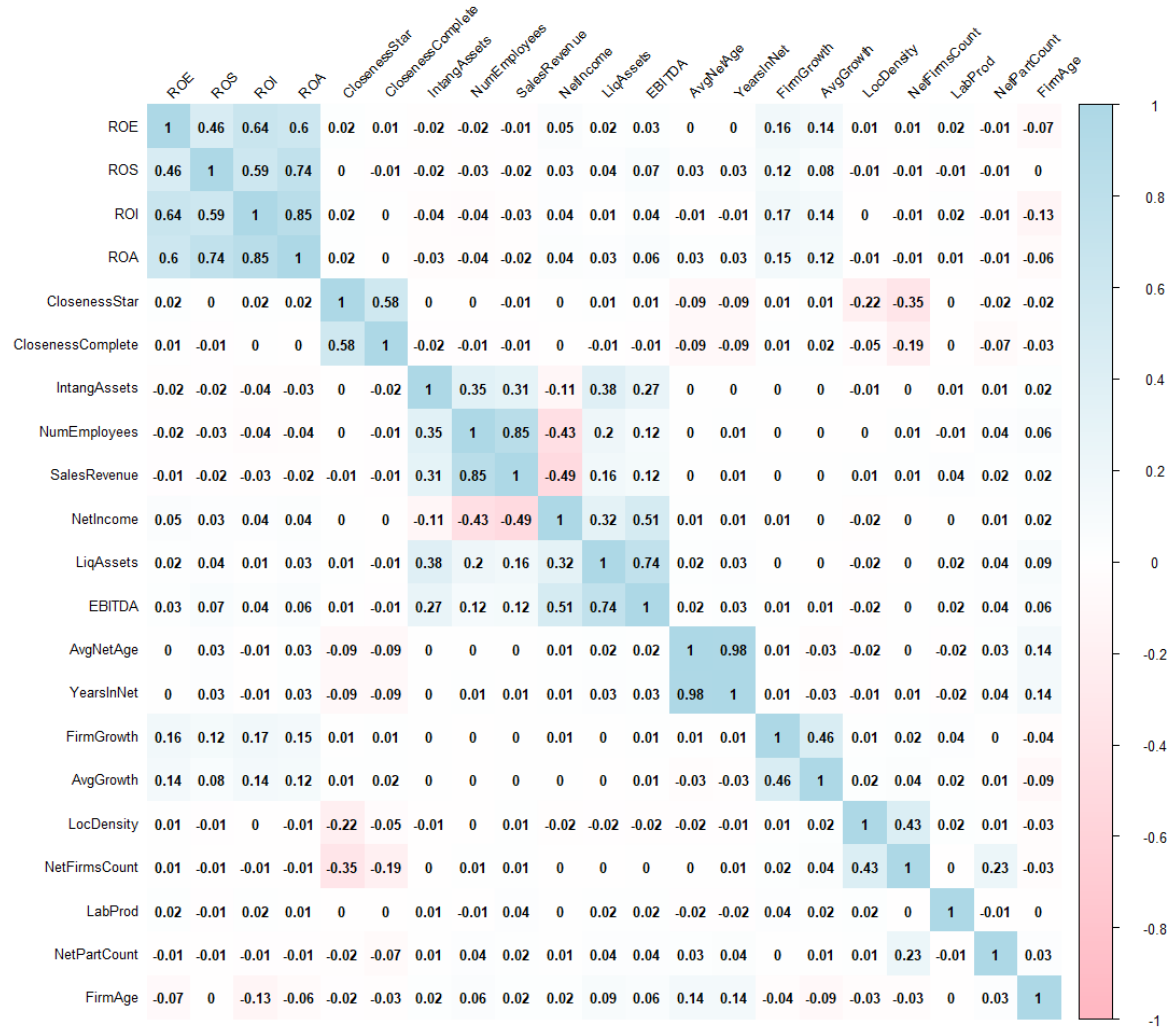
This section presents the correlation plots for the key variables in our dataset, providing a visual representation of the relationships between them and help in identifying multicollinearity issues, guiding the selection of variables for the econometric models.

Numeric Variables

The correlation plot of numeric variables uses the Pearson correlation coefficient to measure the linear relationship between pairs of numeric variables. Each cell in the plot shows the Pearson correlation coefficient, indicating the strength and direction of the correlation, ranging from -1 to 1.

From the figure 2.5, the strongest positive correlation is between NumEmployees and SalesRevenue (0.85), indicating that firms with more employees tend to generate higher sales revenue. This is expected, as larger firms typically have greater production and sales capacities. Although both the number of employees and sales revenue are positively correlated with each other, they are negatively correlated with net income(-0.49 and -0.43, respectively). This could suggest that firms with higher sales revenues have higher expenses or lower margins, leading to reduced net income.
Additionally, the correlation between EBITDA and Liquid Assets (0.74) indicates that firms with higher EBITDA generally mantain higher liquid assets, highlighting the strong

Figure 2.5: Corrplot of Numeric Variables



cash generation ability of profitable firms.

Collectively, these economic performance metrics form a cluster of correlation where IntangAssets, NumEmployees, SalesRevenue, LiqAssets and EBITDA are positively correlated to each other and negatively correlated with Net Income.

Profitability indicators such as ROE, ROI, ROA, ROS also form a cluster exhibiting high positive correlations among each other, ranging from 0.46 to 0.85. This indicates that these profitability metrics are closely related showing how efficiency in one aspect of operations can positively impact overall financial health.

The 0.58 positive correlation between Closeness Centrality in Star and Complete configuration suggests that firms which act as hubs (reference firm in a network) might also be central in terms of geography within the network.

It is somehow expected since both centralities measure node importance within the same

network, albeit with different structures, the star nodes in star configurations have the same number of edges as in the complete network. The inclusion of distance weights for the complete configuration makes it more sensitive to distance variations. Suggesting that central nodes in a star network tend to be central also in complete network in terms of distances.

The negative relationship (-0.22) between Localized Density and Centrality in Star configuration might be due the fact that nodes with high closeness centrality (central nodes) in a star network have many direct connections but their immediate neighborhoods lack interconnections, leading to low localized density. Conversely, nodes with high localized density are likely part of tight clusters where many neighbors are interconnected. These nodes are not as central in terms of closeness centrality because they are part of a local cluster rather than acting as a hub to the overall network.

The positive correlation between Localized Density and Networked Firm Count (0.43) might be interpreted as increase in density when network members increase, this caused by a higher probability of geographical proximity and interaction between the firms.

Years in Network and Average Network Age have nearly perfect correlation as they both are calculated using network act date and year.

Categoric Variables

For categorical variables a Cramér's V correlation plot is created to visually represent the strength of association between pairs of categorical variables. Each cell in the figure shows the Cramér's V value, ranging from 0 (no association) to 1 (perfect association).

The association matrix of categorical variables highlits several key relationships among the variables. The perfect association between Region Group and region_firm (region of each firm) is expected as Region Group is a groups each region in the same area thus each region is associated with a single region group.

LegalNetwork is moderately associated with RegionGroup (0.28), region_firm (0.36) and Sector (0.22), suggesting some geographic area and sectors have more networks with legal subjectivity.

Similarly, hub also shows low/moderate association with LegalNetwork (0.23), RegionGroup (0.1), region_firm (0.13) and Sector (0.1), indicating that hubs are more commonly found in networks with legal subjectivity and in specific geographic areas.

Figure 2.6: Corrplot of Categorical Variables

Sector is also moderately associated with regional capitals and province capitals, this suggests that certain sectors are more likely to be located in capitals, possibly due to the availability of better infrastructure and resources in these areas.

A high correlation between ProvinceCapital and RegionalCapital (0.68) indicating that a significant number of firms located in provincial capitals are also in regional capitals. This is expected as many provincial capitals serve as regional capitals.

InnovativeSME, InnovativeStartup and AverageFirmSize have near 0 association with most of the variables indicating that innovation status and the firm size are not strongly related to each other and to geographical location, sector.

2.2 Variable Selection

This section aims to synthesize the previously conducted analyses (correlation, distribution, and missing values) to finalize the variables that will be used for the modeling of the economic impact of network contracts on firm performance.

The main objective is to investigate the impact of localized density and closeness centrality on firm growth. To ensure a robust analysis and isolate the effects of the proximity measures, controlling for the impact of other characteristics is essential.
Based on the dataset available it is possible to include network characteristics and firm-specific characteristics. To further isolate the effects of the main variables, geography, sector and time effect are included ensuring that regional economic conditions, industry-specific factors and temporal trends do not confound the results. Additionally, all the numeric variables are lagged to time t-1 helping to mitigate endogeneity by ensuring that explanatory and controls variables are determined prior to the current outcome (growth of current year). Also by using t-1 we assume that past values influence current performance aligning the logical direction of causality, strengthening the argument that changes in the independent variable lead to changes in the dependent variable, and not the other way.

Dependent Variables

A dependent variable is the variable/outcome that is affected by changes in the independent variable. In this context, the main objective is to analyze the economic performance of firms, so the variables are:

- `Firm Growth`: primary dependent variable. By analyzing Firm Growth, it is possible to determine whether being in a densely populated network or holding a central position within a network contributes positively to a firm's economic performance.
- `Average Growth`: average growth rate over the specific period, it helps recover missing values for firms in certain years and it is used to check the consistency of the model's behavior with the primary dependent variable.

Independent Variables

An independent variable is the variable that varies to explore its effects on the dependent variable. To study the impact of spatial proximity and network effects on firm growth, the following independent variables are utilized:

- `Localized Density`: measures the geographical density of firms within a network, to see the effect of more or less dense networks on firm growth. A higher localized density indicates a higher concentration of firms in a given area, which can lead to increased collaboration, resource sharing, and competitive advantages. Conversely,

a lower density may suggest isolation or less frequent interactions among firms.

- `Closeness Centrality`: measures the centrality of a firm within the network, Firms that are more central in their networks might benefit from better access to information, resources, and collaboration opportunities, which can enhance their performance.

Network Control Variables

The set of network control variables chosen to account for network characteristics based on the available data are:

- `Hub`: included as it controls for the status of a firm being a reference company for the network.
- `NetworkedFirmsCount`: It is used as a proxy to network size as firms participating in multiple firms don't have one unique network size. The size of the network can influence resource availability, knowledge sharing, and collaborative opportunities, which are crucial for firm growth.
- `AverageNetworkAge`: It is the average age of the networks a firm participates in. The maturity of the network can affect stability, trust among members, and accumulated experience in collaboration, impacting firm growth positively or negatively. It is chosen over "Years in Network" (0.98 correlation) as the latter is estimated not knowing the years in network of firms before 2016.
- `LegalNetwork`: A binary indicator of whether a firm participates in a network that has legal status. Participating in a network with legal personality can influence the formalization of agreements, enforceability of contracts, and access to external funding or benefits, thereby impacting firm growth.

Firm Control Variables

To accurately assess the impact of network characteristics on firm growth, the following controls for firm-specific factors are chosen:

- $FirmAge^2$: The quadratic age of the firm in years. It is used to capture the non-linear relationships between firm age and performance. Younger firms might be more innovative and adaptable as they establish themselves and gain market shares.

As firms age, the growth rate might slow down, indicating a more mature phase with stable performance. In some cases, very old firms might face challenges such as outdated business models or competition, leading to a decline in performance.

- `AverageFirmSize`: The average number of employees in the firm. Firm size can impact economies of scale, resource availability, and market power. This variable is used as a proxy of number of employees as it retrieves missing values over the year inherited from AIDA dataset.
- `InnovativeSME`: A binary indicator of whether the firm is an innovative SME. Innovative SMEs might have different growth trajectories due to their focus on innovation and technology. Controlling for this variable helps in isolating these effects.
- `InnovativeStartup`: Similar to InnvoativeSME but refered to Startups.

For financial performance controls the following are chosen:

- `ROS`: Return on Sales, a measure of a firm's profitability relative to its total sales revenue. The choice of ROS over other profitability metrics in this study is based on several considerations. Firstly, ROS, along with ROE, has lower number of missing values (table 1.12) ensuring a more complete dataset. Secondly, the shape of the distribution seems more normal respect to ROI and ROA. Lastly since the dependent variable in our analysis is the growth in sales revenue, it might be more relevant to focus on how well a firm converts sales into profits.
- `ln_LabProd`: The natural logarithm of labor productivity, Natural logarithm is performed to achieve a more normal distribution. Labor productivity is a key determinant of firm efficiency and competitiveness. Choosing LabProd it is possible to exclude Sales Revenue and number of employees as they are incorporated in this metric.
- `ln_LiquidAssets`: The natural logarithm of liquid assets, representing the firm's cash and cash equivalents. Natural logarithm is performed to achieve a more normal distribution.
- `ln_IntangibleAssets`: The natural logarithm of intangible assets, including intellectual property and goodwill. Intangible assets reflect the value of a firm's

intellectual property and brand equity, which can significantly influence growth potential. Natural logarithm is performed to achieve a more normal distribution.

Time, Sector and Geographic Fixed Effect Controls

It is important to control for time, sector and geographic fixed-effects as it provides more robust results by accounting for unobserved heterogeneity that could introduce bias.

Time fixed-effects control for factors that impact all the firms in the same way over time, including macroeconomic trends, policy changes, technological advancements, economic cycles or shocks like the pandemic. Controlling for time ensures the observed relationships are not confounded by time-related external factors. A dummy variable for each year in the panel data is created to account for this effect.

For sectorial variables there is the variable `Sector` which is also the only variable that describes the firm's industry. It permits to control for industry specific effects that might influence firm performance.

For geographic fixed-effect control, the variable `RegionGroup` is used. As it has fewer categories than `region_firm` it reduces the dimensionality of the model, without creating one dummy for each of the 20 regions of Italy that might lead to overfitting.

Regression Equation

To investigate the impact of the spatial proximity and network effects on the economic performance of firms within the Italian network contracts, regression analysis can be performed as it is able to handle multiple variables, quantify the relationships, model the outcomes, address variability, conduct robustness checks making it suitable method for the analysis of the impact of the chosen variables on firm performance.

The base form regression equation for a model using the above variables can be:

$$\begin{aligned}\text{FirmGrowth}_{it} = \beta_0 + \beta_1 \text{LocalizedDensity}_{i,t-1} + \beta_2 \text{ClosenessCentrality}_{i,t-1} \\ + \delta_t \text{Year}_t + \gamma_j \text{Sector}_j + \theta_k \text{RegionGroup}_k + \epsilon_{it}\end{aligned} \tag{2.1}$$

Where:

- FirmGrowth_{it}: Dependent variable representing the growth of firm i at time t.
- β_0: Intercept term.

- β_1: Coefficient for the effect of localized density on firm growth.
- $\text{LocalizedDensity}_{i,t-1}$: Lagged independent variable representing the localized density for firm i at time $t-1$.
- β_2: Coefficient for the effect of closeness centrality on firm growth.
- $\text{ClosenessCentrality}_{i,t-1}$: Lagged independent variable representing the closeness centrality for firm i at time $t-1$.
- $\delta_t \text{Year}_t$: Time fixed effects, where δ_t are the coefficients for each year dummy variable Year_t.
- $\gamma_j \text{Sector}_j$: Sector fixed effects, where γ_j are the coefficients for each sector dummy variable Sector_j.
- $\theta_k \text{RegionGroup}_k$: Geographic fixed effects, where θ_k are the coefficients for each region group dummy variable RegionGroup_k.
- ϵ_{it}: Error term capturing unobserved factors affecting firm growth for firm i at time t.

By adding the controlled characteristics, the regression equation becomes:

$$\begin{aligned} \text{FirmGrowth}_{it} = \beta_0 &+ \beta_1 \text{LocalizedDensity}_{i,t-1} + \beta_2 \text{ClosenessCentrality}_{i,t-1} \\ &+ \beta_3 \text{NetworkChar}_{it} + \beta_4 \text{FirmChar}_{it} + \beta_5 \text{FirmPerf}_{it} \\ &+ \delta_t \text{Year}_t + \gamma_j \text{Sector}_{ij} + \theta_k \text{Geography}_{ik} + \epsilon_{it} \end{aligned} \tag{2.2}$$

Where:

- FirmGrowth_{it}: Dependent variable representing the growth of firm i at time t.
- β_0: Intercept term.
- β_1: Coefficient for the effect of localized density on firm growth.
- $\text{LocalizedDensity}_{i,t-1}$: Independent variable representing the localized density for firm i at time $t-1$.
- β_2: Coefficient for the effect of closeness centrality on firm growth.
- $\text{ClosenessCentrality}_{i,t-1}$: Independent variable representing the closeness centrality for firm i at time $t-1$.
- β_3: Coefficient for the effect of network characteristic controls on firm growth.

- NetworkChar$_{it}$: Vector of control variables related to the network characteristics: hub, NetworkedFirmsCount, AverageNetworkAge, LegalNetwork.
- β_4: Coefficient for the effect of firm characteristic controls on firm growth.
- FirmChar$_{it}$: Vector of control variables related to firm-specific characteristics: $FirmAge^2$, AverageFirmSize, InnovativeSME, InnovativeStartup.
- β_5: Coefficient for the effect of financial performance controls on firm growth.
- FirmPerf$_{it}$: Vector of control variables related to firm-specific financial performance: ROS, ln_LabProd, ln_LiquidAssets, ln_IntangibleAssets.
- δ_tYear$_t$: Time fixed effects, where δ_t are the coefficients for each year dummy variable Year$_t$.
- γ_jSector$_j$: Sector fixed effects, where γ_j are the coefficients for each sector dummy variable Sector$_j$.
- θ_kRegionGroup$_k$: Geographic fixed effects, where θ_k are the coefficients for each region group dummy variable RegionGroup$_k$.
- ϵ_{it}: Error term capturing unobserved factors affecting firm growth for firm i at time t.

2.3 Regression Models

Regression analysis is fundamental in econometric modeling, it offers insights into the relationships between variables. In this case it helps to investigate the relationship between network spatial metrics and firm performance. In this section Ordinary Least Squares (OLS) and Least Absolute Deviations (LAD) models are analyzed.

The OLS Pooled regression serves as baseline model for its simplicity and efficiency in estimating the average effect of explanatory variables on the dependent variable. It works as a benchmark for a baseline understanding of how network contracts and other characteristics impact firm performance.

While the OLS model is simple to use, its estimates might be affected by outliers and non-constant error variances. Thus, the model is complemented with LAD regression

models, also known as median regression. This method minimizes the sum of absolute residuals making it less sensitive to outliers and more robust in the presence of non-normal error distributions or heteroscedasticity.

Two versions of each regression model are implemented: one using FirmGrowth as the dependent variable and another using AverageGrowth. This approach ensures that the results are consistent across different measures of performance.

Models in this section are pooled regressions, meaning that data from multiple time periods are combined and analyzed together, treating them as one large cross-sectional dataset. This approach allows us to examine the overall effect of the explanatory variables on firm performance across different times without accounting for the specific time-period effects.

To account for heteroscedasticity Robust Standard Errors are employed for OLS models providing consistent estimates of the standard errors.

2.3.1 OLS Regression Models

The Table 2.2 is composed by 4 specifications of pooled regression model using LocalizedDensity and ClosenessCentralityStar (star network configuration) as explanatory varaibles and FirmGrowth as dependent variable. Starting from the baseline model (1) which includes only the explanatory variables. Control variables are progressively added, where the last specification (4) includes all control variables: network-characteristic controls, firm-characteristic controls, firm-performance controls. For each specification robust standard errors are provided in the parentheses and time, sector, geographic fixed-effects are accounted for.

In the baseline model, model specification (1), the impact of Localized Density on firm growth is not significant while Centrality is positive and significant at 1% level. Indicating that firms with higher centrality within their networks experience higher growth. In the second specification network characteristics such as Hub, Networked Firms Count, Average Network Age and Legal Network are added. Localized Density's coefficient remains small and non significant. Centrality continues to show a significant positive effect ($p<0.01$) with a coefficient of 0.026. The newly added variables are not statistically significant except for Average Network Age which has negative and significant impact on firm growth (-0.007 with $p<0.01$), indicating that as the average age of the networks in which

Table 2.2: OLS - FirmGrowth (star configuration)

	Dependent variable:			
	FirmGrowth			
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	0.0003 (0.0004)	0.0002 (0.0004)	0.001* (0.0004)	0.001** (0.0004)
$ClosenessCentralityStar_{t-1}$	0.025*** (0.008)	0.026*** (0.010)	0.027*** (0.010)	0.025*** (0.010)
Hub: 1		−0.006 (0.006)	−0.013 (0.006)	−0.007 (0.006)
$NetworkedFirmsCount_{t-1}$		0.00002 (0.0002)	−0.0001 (0.0002)	0.0001 (0.0002)
$AverageNetworkAge_{t-1}$		−0.007*** (0.001)	−0.006*** (0.001)	−0.005*** (0.001)
LegalNetwork: 1		0.009 (0.007)	0.007 (0.007)	0.001 (0.007)
$FirmAge^2_{t-1}$			−0.00001*** (0.00000)	−0.00000 (0.00000)
AverageFirmSize: Micro			−0.067*** (0.008)	0.018 (0.012)
AverageFirmSize: SME			−0.024*** (0.008)	0.030*** (0.009)
InnovativeSME: 1			0.094*** (0.015)	0.043** (0.017)
InnovativeStartup: 1			0.246*** (0.082)	0.263*** (0.091)
ROS_{t-1}				0.004*** (0.0005)
$\ln(LabProd_{t-1})$				−0.056*** (0.004)
$\ln(LiquidAssets_{t-1})$				0.010*** (0.001)
$\ln(IntangibleAssets_{t-1})$				0.006*** (0.001)
Constant	−0.006 (0.009)	0.015 (0.011)	0.054*** (0.013)	0.202*** (0.022)
Observations	55,270	55,270	52,806	48,443
N. unique firms	15,419	15,419	14,593	13,998
R^2	0.043	0.043	0.047	0.071

Note: OLS estimates and robust standard errors are given in parentheses. All regressions also include year, sector and geographic fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

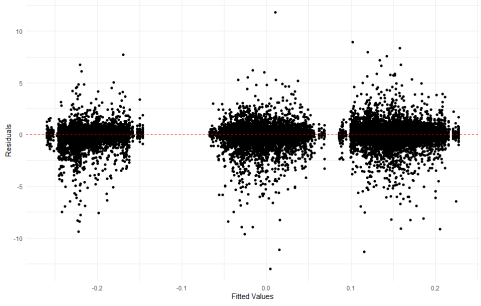
a firm participates increases, firm growth decreases, suggesting that younger networks might be more dynamic or innovative and more capable of stimulating firm growth.

By adding firm characteristics (squared firm age, average firm size, innovative SME and innovative Startup dummies), Localized Density acquires small significance (p<0.1) with marginal coefficient of 0.001. Centrality remains positive and significant at 1% significance. Squared firm age has a significant and negative impact on firm growth, possibly confirming the non-linear relationship where younger firms grow more rapidly and the growth decelerates over time. Being a Micro firm or SME have negative and significant coefficient (-0.064 and -0.024 respectively), indicating that, the growth rate of these two categories of firms on starts on average 0.067 and 0.024 units lower than large firms. While being a innovative SME or Startup brings a strong positive impact on firm growth (0.094 and 0.246 respectively at 1% level).

The final model adds performance controls such as ROS, labor productivity, tangible and intangible assets. Localized density acquires higher significance (p<0.05) with same coefficient. Centrality's significance remains the same with similar coefficient (0.025) con-

firming the importance of being central in a network. Network characteristics do not face substantial changes with Networked Firms Count becoming positive but still not significant. $FirmAge^2_{t-1}$ loses significance and becomes even more marginal. Interestingly, being SME brings positive and significant impact while being a Micro firm also turns to a positive impact although non significant. By controlling factors directly related to the firm's performance, the size of the firm has a clearer positive relationship with firm growth, suggesting that smaller firms have better growth prospects than large firms. Innovative dummies remain both significant and positive. The financial performances at time t-1 all impact the firm growth in time t in a positive and significant way except for $\ln(LabProd_{t-1})$. This implies that firms with strong financial health translates into higher growth. The labor productivity might imply that firms with high productivity are potentially mature enough with optimized processes to an extent that further growth possibilities are limited.

Figure 2.7: Residual plot of model(1) - FirmGrowth(star configuration)



The residual plots show evident clusters and outliers forming in figure 2.7 in the base model specification, indicating potential issues with heteroskedasticity and non-linearity. However, by adding the control variables (model 4), the distribution of residuals becomes more uniform as seen in figure 2.8, indicating better model fit and more reliable estimates.

Figure 2.8: Residual plot of model(4) - FirmGrowth(star configuration)

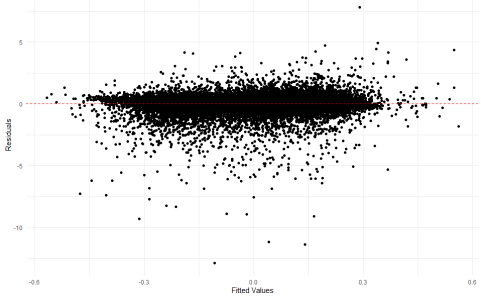


Table 2.3 shows the regression models for the AverageGrowth model in star network

configuration.

Table 2.3: OLS - AverageGrowth (star configuration)

	Dependent variable:			
	AverageGrowth			
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	0.001** (0.0004)	0.0001 (0.0005)	0.001* (0.0005)	0.0003 (0.0003)
$ClosenessCentralityStar_{t-1}$	0.055*** (0.012)	0.063*** (0.015)	0.044*** (0.010)	0.027*** (0.007)
Hub: 1		−0.018** (0.008)	−0.022*** (0.007)	−0.015*** (0.004)
$NetworkedFirmsCount_{t-1}$		0.001** (0.0003)	0.001** (0.0003)	0.0004** (0.0002)
$AverageNetworkAge_{t-1}$		−0.018*** (0.001)	−0.017*** (0.001)	−0.009*** (0.001)
LegalNetwork: 1		−0.006 (0.008)	−0.005 (0.008)	−0.003 (0.005)
$FirmAge^2_{t-1}$			−0.00001*** (0.00000)	−0.00001*** (0.00000)
AverageFirmSize: Micro			−0.086*** (0.009)	−0.051*** (0.009)
AverageFirmSize: SME			−0.026*** (0.008)	−0.019*** (0.007)
InnovativeSME: 1			0.085*** (0.013)	0.054*** (0.011)
InnovativeStartup: 1			0.414*** (0.075)	0.359*** (0.071)
ROS_{t-1}				0.004*** (0.0002)
$\ln(LabProd_{t-1})$				0.005** (0.002)
$\ln(LiquidAssets_{t-1})$				0.001 (0.001)
$\ln(IntangibleAssets_{t-1})$				−0.001* (0.001)
Constant	−0.024*** (0.008)	0.032*** (0.008)	0.086*** (0.011)	0.015 (0.015)
Observations	60,262	60,262	57,409	49,387
N. unique firms	15,483	15,483	14,711	14,133
R^2	0.010	0.020	0.032	0.054

Note: OLS estimates and robust standard errors are given in parentheses. All regressions also include year, sector and geographic fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

Localized Density is moderately significant in specification (1) and (3), significant at 5% and 10% level, with low coefficient. Similar to previous models, Closeness Centrality is always positive and significant, with higher coefficients in the first two models, and decreases in the subsequent specifications, indicating stronger impact of Closeness Centrality on average growth over the years.

In network characteristic controls being a Hub for networks becomes significant across all specifications with -0.015 coefficient when accounting for all control variables, indicating that firms acting as a hub within their networks tend to experience lower average growth rates compared to firms that are not hubs. AverageNetworkAge shows consistent and significant negative impact on average growth, NetworkedFirmsCount becomes significant as main difference to the previous models with Firm Growth as dependent variable.

AverageFirmSize in this case does not become positive but remains negative and significant. Remaining firm characteristic controls remain the same as previous model.

Firm financial performance controls remain mostly similar with decrease in significance.

In table 2.4, the centrality measures in a complete network configuration weighted by

distance are used instead of the star network configuration.

Table 2.4: OLS - FirmGrowth (complete configuration)

	Dependent variable:			
	FirmGrowth			
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	−0.0003 (0.0002)	−0.001*** (0.0003)	−0.0003 (0.0003)	−0.0002 (0.0003)
$ClosenessCentralityComplete_{t-1}$	0.007 (0.010)	0.008 (0.011)	0.013 (0.011)	0.001 (0.010)
Hub: 1		0.0002 (0.006)	−0.007 (0.006)	−0.001 (0.006)
$NetworkedFirmsCount_{t-1}$		0.0003 (0.0002)	0.0003 (0.0002)	0.0003* (0.0002)
$AverageNetworkAge_{t-1}$		−0.006*** (0.001)	−0.006*** (0.001)	−0.004*** (0.001)
LegalNetwork: 1		0.003 (0.005)	0.004 (0.005)	−0.001 (0.005)
$FirmAge^2_{t-1}$			−0.00001*** (0.00000)	−0.00000 (0.00000)
AverageFirmSize: Micro			−0.066*** (0.008)	0.021* (0.011)
AverageFirmSize: SME			−0.022*** (0.007)	0.034*** (0.008)
InnovativeSME: 1			0.098*** (0.014)	0.046*** (0.016)
InnovativeStartup: 1			0.178** (0.090)	0.188** (0.088)
ROS_{t-1}				0.004*** (0.0004)
$\ln(LabProd_{t-1})$				−0.058*** (0.004)
$\ln(LiquidAssets_{t-1})$				0.009*** (0.001)
$\ln(IntangibleAssets_{t-1})$				0.006*** (0.001)
Constant	−0.004 (0.008)	0.012 (0.009)	0.051*** (0.011)	0.213*** (0.019)
Observations	69,909	69,909	66,404	60,859
N. unique firms	20,134	20,134	18,922	18,157
R^2	0.047	0.048	0.052	0.079

Note: OLS estimates and robust standard errors are given in parentheses. All regressions also include year, sector, and geographic fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01.

It is possible note that Localized Density and centrality measures are never significant except for Localized Density in the model (2), interestingly the coefficients of localized density are now negative. All the control variables have consistent significance progression similar to previous models with the star configuration, showing that controlling for financial performance can uncover the growth potential of smaller firms participating in networks.

Using the AverageGrowth as dependent variable uncovers the significance of centrality which is significant in all model specifications. Localized Density becomes consistently significant and negative while it was positive and not significant in the star configuration. Also in the case of complete configuration the control variables of the AverageGrowth models are mostly similar to the control variables of the annual FirmGrowth models with the exception of AverageFirmSize dummies and NetworkedFirmsCount acquiring increasing significance.

The results in both star and complete configurations confirm the the crucial role of centrality in firm growth. However, the Star configuration reflect the centrality of a firm

Table 2.5: OLS - AverageGrowth (complete configuration)

	Dependent variable:			
	AverageGrowth			
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	−0.0003 (0.0002)	−0.001*** (0.0003)	−0.001* (0.0003)	−0.001*** (0.0002)
$ClosenessCentralityComplete_{t-1}$	0.040** (0.018)	0.038* (0.019)	0.027** (0.011)	0.016** (0.007)
Hub: 1		−0.008 (0.007)	−0.016** (0.006)	−0.010** (0.004)
$NetworkedFirmsCount_{t-1}$		0.001** (0.0002)	0.001** (0.0002)	0.0004*** (0.0002)
$AverageNetworkAge_{t-1}$		−0.017*** (0.001)	−0.015*** (0.001)	−0.009*** (0.001)
LegalNetwork: 1		−0.006 (0.007)	−0.003 (0.007)	−0.008** (0.004)
$FirmAge^2_{t-1}$			−0.00001*** (0.00000)	−0.00001*** (0.00000)
AverageFirmSize: Micro			−0.082*** (0.008)	−0.053*** (0.008)
AverageFirmSize: SME			−0.022*** (0.007)	−0.019*** (0.007)
InnovativeSME: 1			0.086*** (0.012)	0.058*** (0.011)
InnovativeStartup: 1			0.345*** (0.068)	0.288*** (0.064)
ROS_{t-1}				0.004*** (0.0002)
$\ln(LabProd_{t-1})$				0.008*** (0.002)
$\ln(LiquidAssets_{t-1})$				0.001 (0.001)
$\ln(IntangibleAssets_{t-1})$				−0.001** (0.001)
Constant	−0.012* (0.006)	0.041*** (0.007)	0.087*** (0.010)	0.010 (0.014)
Observations	76,120	76,120	72,108	62,140
N. unique firms	20,207	20,207	19,063	18,342
R^2	0.009	0.017	0.027	0.053

Note: OLS estimates and robust standard errors are given in parentheses. All regressions also include year, sector, and geographic fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01.

in the network in terms of operational interactions, which is more relevant instead of a geographic centrality where every firm is assumed to communicate with any other firm. In terms of empirical robustness, star configuration has shown consistent and significant results across various model specifications, suggesting it can better capture the network structures respect to the complete configuration. Therefore, Closeness Centrality Star will be used as the measure of network centrality due to its better fit and coherence with the network structure.

2.3.2 LAD Regression models

The inconsistency in the significance of the Localized Density in previous model specifications shows that it can be significant but it may not be captured correctly by the OLS regression due to its sensitivity to outliers and the skewed distribution of the data. Additionally OLS regression assumes that the errors are normally distributed, an assumption that may not hold in this context (as seen in residual plot 2.8), leading to inefficiencies in the estimation process. Therefore, using a LAD regression or Median regression, which minimizes the sum of absolute deviations rather than the sum of squared deviations,

making it more robust to outliers. Unlike OLS, LAD regression does not assume a normal distribution of errors, making it more capable of providing more accurate estimates in presence of non-normal error distributions.

Table 2.6 shows the regression analysis with FirmGrowth as dependent variable. It follows the structure of previous models with 4 specifications and progressive control variables added for each model specification.

Table 2.6: LAD - FirmGrowth

	Dependent variable:			
	FirmGrowth			
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	−0.0003** (0.0001)	−0.0005*** (0.0002)	−0.0002 (0.0002)	−0.0002 (0.0002)
$ClosenessCentrality_{t-1}$	0.008** (0.004)	0.009** (0.004)	0.009** (0.004)	0.010** (0.004)
Hub: 1		0.0002 (0.003)	−0.001 (0.002)	0.0003 (0.002)
$NetworkedFirmsCount_{t-1}$		0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
$AverageNetworkAge_{t-1}$		−0.003*** (0.001)	−0.003*** (0.001)	−0.003*** (0.001)
LegalNetwork: 1		0.006** (0.003)	0.005* (0.003)	0.006** (0.003)
$FirmAge^2_{t-1}$			−0.00001*** (0.00000)	−0.00000*** (0.00000)
AverageFirmSize: Micro			−0.018*** (0.004)	0.014*** (0.004)
AverageFirmSize: SME			0.0002 (0.003)	0.020*** (0.003)
InnovativeSME: 1			0.035*** (0.009)	0.025*** (0.009)
InnovativeStartup: 1			0.236*** (0.050)	0.229** (0.115)
ROS_{t-1}				0.001*** (0.0001)
$\ln(LabProd_{t-1})$				−0.014*** (0.001)
$\ln(LiquidAssets_{t-1})$				0.003*** (0.0005)
$\ln(IntangibleAssets_{t-1})$				0.002*** (0.0004)
Constant	0.042*** (0.003)	0.051*** (0.004)	0.060*** (0.005)	0.081*** (0.007)
Observations	55,270	55,270	52,806	48,443
N. unique firms	15,419	15,419	14,593	13,998

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector, and geographic fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01.

Unlike Pooled OLS 2.2, using LAD regression, Localized Density is negative across all models, significant in Models (1) and Model (2) (p-value < 0.05 and p-value < 0.01, respectively), becoming non-significant in Models (3) and (4). This suggests that when outliers' influence is minimized, the relationship between localized density and firm growth is negative, but the significance is captured by other control variables when added.
Closeness Centrality is consistent with previous results as it is positive across all specifications and gains consistent high level of significance at 5% level.
In terms of network characteristics, being a Hub remains non-significant across all models. Being part of a Legal Network is significant across all models with minimum significance level of 10%, with a positive impact on firm growth. Average Network Age's negative impact on firm growth is confirmed with high level of significance (p<0.01), suggesting the

initial phase of a network is the most dynamic one. Networked Firms Count is positive but non-significant.

Firm-level characteristics are all significant. Quadratic firm age is negative suggesting diminishing returns in firm growth. Micro firms are shown with a significant negative impact in Model (3) (p-value < 0.01) but becomes positive in Model (4) (p-value < 0.01), indicating a change in the effect of firm size with additional controls. While SMEs tend to a positive and significant impact in Model (4) (p-value < 0.01). Innovative SMEs and Startups have consistent positive and significant impact with Startups having a higher impact than SMEs (0.229 against 0.026).

Financial performance controls are all significant. Return on Sales (ROS) is positive and significant in Model (4) with a coefficient of 0.001 (p-value < 0.01). Logarithm of Labor Productivity is negative and significant in Model (4) at -0.014 level (p-value < 0.01). Logarithm of Liquid Assets is positive and significant in Model (4) (p-value < 0.01). Logarithm of Intangible Assets is positive and significant in Model (4) (p-value < 0.01).

Table 2.7: LAD - AverageGrowth

	Dependent variable:			
	AverageGrowth			
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	-0.0003^{***} (0.0001)	-0.0004^{***} (0.0001)	-0.0002^{***} (0.0001)	-0.0004^{***} (0.0001)
$ClosenessCentrality_{t-1}$	0.020^{***} (0.002)	0.022^{***} (0.002)	0.021^{***} (0.002)	0.020^{***} (0.002)
Hub1		-0.002^{*} (0.001)	-0.004^{***} (0.001)	-0.004^{***} (0.001)
$NetworkedFirmsCount_{t-1}$		0.0002^{***} (0.00004)	0.0002^{***} (0.00004)	0.0002^{***} (0.00004)
$AverageNetworkAge_{t-1}$		-0.006^{***} (0.0003)	-0.005^{***} (0.0003)	-0.004^{***} (0.0002)
LegalNetwork: 1		0.002 (0.001)	0.002^{**} (0.001)	0.002 (0.001)
$FirmAge^2_{t-1}$			-0.00001^{***} (0.00000)	-0.00001^{***} (0.00000)
AverageFirmSize: Micro			-0.028^{***} (0.002)	-0.019^{***} (0.003)
AverageFirmSize: SME			-0.007^{***} (0.002)	-0.005^{***} (0.002)
InnovativeSME: 1			0.039^{***} (0.004)	0.032^{***} (0.004)
InnovativeStartup: 1			0.326^{***} (0.007)	0.287^{***} (0.081)
ROS_{t-1}				0.002^{***} (0.0001)
$\ln(LabProd_{t-1})$				0.008^{***} (0.001)
$\ln(LiquidAssets_{t-1})$				0.001^{***} (0.0002)
$\ln(IntangibleAssets_{t-1})$				0.00001 (0.0002)
Constant	0.041^{***} (0.002)	0.056^{***} (0.002)	0.075^{***} (0.003)	0.009^{**} (0.004)
Observations	60,262	60,262	57,409	49,387
N. unique firms	15,483	15,483	14,711	14,133

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector, and geographic fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 2.7 illustrates the regression table using AverageGrowth as the dependent variable.

Also in this case differently from the OLS Pooled models Localized Density is negative

across all models and always significant at 1% level.
Closeness Centrality is consistent as previous models with higher level of significance.
Most notable difference of these specifications is the Networked Firms Count of previous year that becomes highly significant (p-value < 0.01) and positive, suggesting the number of firms a firm is exposed to through networks has a positive impact on the average growth, suggesting benefits from being part of larger and multiple networks.
Firm characteristic controls are consistent with table 2.3, with lower magnitude possibly not influeced by outliers.
Firm financial performance control become positive and significant except for the Intangible assets.

Overall, the LAD regression underscores the importance of accounting for outliers and non-normal error distributions. The more robust results confirm the positive impact for firm of being central in a network, and shows the negative impact of Localized Density on firm performance when the impact of outliers are minimized, highlighting the need for a deeper investigation into the effects of Localized Density due to its inconsistent significance across models.

2.3.3 LAD Interactions

The LAD regression results for FirmGrowth models indicate a consistent negative impact of Localized Density across all models, similar to the AverageGrowth results where the negative impact is also significant across all specifications. This suggests that higher localized density may constrain firm growth due to increased competition and resource constraints when outliers are minimized.
The inconsistent significance levels suggest that its impact on firm growth is complex and might be influenced by other network characteristics. To better understand these relationships an analysis on the interaction between Localized Density and other key network characteristics is performed.
Interaction terms are employed by combining the effect of two or more variables on the dependent variable, highlighting effects that might not be evident when considering each variable independently. Specifically, it is possible to uncover non-linear relationships and joint effects of two variables.
Interactions analyzed in this section are:

- `Density-Centrality interaction`: this term examines how the impact of localized density on firm growth is influenced by the firm's closeness centrality within the network. It captures the combined effect of the spatial concentration of firms in the same area (localized density) and their relative position or influence within the network (closeness centrality) on their economic performance.
- `Brokerage`: this term measures the extent to which a firm acts as a bridge between different parts of the network. The interaction LocalizedDensity*HUB explores how the role of firms as intermediaries (hubs, identified by reference_company) within networks affects their growth, highlighting the importance of brokerage in leveraging network connections for economic performance.
- `Stability`: it refers to the consistency or persistence of firms within a network over time. The interaction term LocalizedDensity*NetworkedFirmsCount examines how the impact of localized density on firm growth changes with the number of firms a company is networked with, reflecting how stable, long-term network participation influences performance in densely connected environments.
- `Network Embeddedness`: degree to which a firm is integrated within its network.. The interaction term LocalizedDensity*AverageNetworkAge analyzes how the combined effect of localized density and the average age of the network influences firm growth, showing how the impact of localized density on firm growth changes with the longevity of the network.
- `WeakStrongTie`: this term differentiates between the strength of connections firms have within the network, where weak ties represent informal network connections, and strong ties indicate formal networks. The interaction term LocalizedDensity*LegalNetwork explores whether the impact of localized density on firm growth varies with the presence of legal networks, indicating how formal, strong ties in dense networks affect performance.

The tables are composed by four specifications, model (1) and model(2) showing models with FirmGrowth as dependent variable, model(3) and model(4) with AverageGrowth as dependent variable. For each pair of models the first shows only the explanatory variables and the interaction term while the second model specification adds all the controls variables.

Density-Centrality interaction

Table 2.8: Density-Centrality interaction

	Dependent variable:			
	FirmGrowth		AverageGrowth	
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	−0.0002 (0.0002)	−0.00002 (0.0002)	−0.0002** (0.0001)	−0.0003*** (0.0001)
$ClosenessCentrality_{t-1}$	0.012*** (0.004)	0.012** (0.005)	0.023*** (0.002)	0.021*** (0.002)
LD_{t-1}*CC_{t-1}	−0.010*** (0.004)	−0.011*** (0.004)	−0.005** (0.002)	−0.004*** (0.001)
Hub: 1		0.002 (0.003)		−0.004*** (0.001)
$NetworkedFirmsCount_{t-1}$		0.0001 (0.0001)		0.0002*** (0.00004)
$AverageNetworkAge_{t-1}$		−0.003*** (0.001)		−0.004*** (0.0002)
LegalNetwork: 1		0.006** (0.003)		0.001 (0.001)
$FirmAge^2_{t-1}$		−0.00000*** (0.00000)		−0.00001*** (0.00000)
AverageFirmSize: Micro		0.014*** (0.005)		−0.019*** (0.003)
AverageFirmSize: SMEs		0.019*** (0.003)		−0.005** (0.002)
InnovativeSME: 1		0.024*** (0.009)		0.032*** (0.004)
InnovativeStartup: 1		0.231** (0.117)		0.286*** (0.081)
ROS_{t-1}		0.001*** (0.0001)		0.002*** (0.0001)
$\ln(LabProd_{t-1})$		−0.014*** (0.001)		0.008*** (0.001)
$\ln(LiquidAssets_{t-1})$		0.003*** (0.001)		0.001*** (0.0003)
$\ln(IntangibleAssets_{t-1})$		0.002*** (0.0004)		0.00001 (0.0002)
Constant	0.043*** (0.003)	0.082*** (0.007)	0.041*** (0.002)	0.010** (0.004)
Observations	55,270	48,443	60,262	49,387
N. unique firms	15,419	13,998	15,483	14,133

Note: LAD estimates and robust standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

The results indicate that localized density negatively impacts average growth, as seen from the significant negative coefficients in models 3 and 4. However, its effect on firm growth is still negative but is not significant.

Closeness centrality is confirmed to positively impact both firm growth and average growth across all models.

The interaction term between localized density and closeness centrality (LD Œ CC) is negative and significant across all models, -0.011 (p<0.01) in model 2. This implies that the positive effect of centrality on growth diminishes for firms with high localized density. The results indicate that while centrality within a network generally promotes firm growth, this effect are reduced when firms are in densely populated network areas. This interaction suggests a trade-off between being central in a densely populated network area and the associated benefits of growth.

Brokerage

The table "Brokerage: LD-Hub Interaction" examines how the interaction between localized density and being a hub firm affects firm growth and average growth.

Table 2.9: Brokerage: LD-Hub Interaction

	Dependent variable:			
	FirmGrowth		AverageGrowth	
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	0.0002 (0.001)	0.0004 (0.0005)	0.0001 (0.0002)	0.0002 (0.0003)
$ClosenessCentrality_{t-1}$	0.008** (0.004)	0.010** (0.004)	0.021*** (0.002)	0.017*** (0.002)
LD_{t-1} ∗Hub:1	−0.0005 (0.0005)	−0.001* (0.0003)	−0.0004** (0.0002)	−0.001** (0.0002)
$NetworkedFirmsCount_{t-1}$		0.0001 (0.0001)		0.0002*** (0.00004)
$AverageNetworkAge_{t-1}$		−0.003*** (0.001)		−0.004*** (0.0002)
LegalNetwork: 1		0.006** (0.003)		0.001 (0.001)
$FirmAge^2_{t-1}$		−0.00000*** (0.00000)		−0.00001*** (0.00000)
AverageFirmSize: Micro		0.014*** (0.004)		−0.019*** (0.003)
AverageFirmSize: SME		0.020*** (0.003)		−0.005*** (0.002)
InnovativeSME: 1		0.025*** (0.009)		0.030*** (0.005)
InnovativeStartup: 1		0.228** (0.113)		0.286*** (0.081)
ROS_{t-1}		0.001*** (0.0001)		0.002*** (0.0001)
$\ln(LabProd_{t-1})$		−0.014*** (0.001)		0.008*** (0.001)
$\ln(LiquidAssets_{t-1})$		0.003*** (0.001)		0.001*** (0.0003)
$\ln(IntangibleAssets_{t-1})$		0.002*** (0.0004)		−0.00003 (0.0002)
Constant	0.042*** (0.003)	0.081*** (0.007)	0.041*** (0.002)	0.010*** (0.004)
Observations	55,270	48,443	60,262	49,387
N. unique firms	15,419	13,998	15,483	14,133

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

The interaction term LD_{t-1} ∗Hub:1 is negative in all models, being significant for firm growth when all controls are added, although at a 10% level, but is significant in both base and controlled specifications for average growth. This suggests that when localized density is high, being a hub might slightly reduce the positive effect on firm growth.

Stability

The interaction term $LD_{t-1} * NetworkedFirmsCount_{t-1}$ shows to not be significant for FirmGrowth while it is positive and significant for AverageGrowth.
Compariing the results with tables 2.6 and 2.7 shows that the interaction term does not significantly alter the coefficients and the significance of the model. Possibly due to the fact that Localized Density is positively correlated with NetworkedFirmsCount (0.43).

Table 2.10: Stability: LD-NetworkedFirmsCount Interaction

	Dependent variable:			
	FirmGrowth		AverageGrowth	
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	−0.001** (0.0004)	−0.0005 (0.0004)	−0.002*** (0.0002)	−0.002*** (0.0002)
$ClosenessCentrality_{t-1}$	0.008** (0.004)	0.008* (0.005)	0.019*** (0.002)	0.016*** (0.002)
$LD_{t-1} * NetworkedFirmsCount_{t-1}$	0.00001 (0.00001)	0.00001 (0.00001)	0.00004*** (0.00001)	0.00003*** (0.00001)
Hub: 1		0.0002 (0.003)		−0.004*** (0.001)
$AverageNetworkAge_{t-1}$		−0.003*** (0.001)		−0.004*** (0.0002)
LegalNetwork: 1		0.006** (0.003)		0.003*** (0.001)
$FirmAge^2_{t-1}$		−0.00000*** (0.00000)		−0.00001*** (0.00000)
AverageFirmSize: Micro		0.013*** (0.005)		−0.019*** (0.003)
AverageFirmSize: SME		0.019*** (0.003)		−0.005** (0.002)
InnovativeSME: 1		0.024*** (0.009)		0.032*** (0.005)
InnovativeStartup: 1		0.228** (0.115)		0.282*** (0.083)
ROS_{t-1}		0.001*** (0.0001)		0.002*** (0.0001)
$\ln(LabProd_{t-1})$		−0.014*** (0.001)		0.008*** (0.001)
$\ln(LiquidAssets_{t-1})$		0.003*** (0.001)		0.001*** (0.0002)
$\ln(IntangibleAssets_{t-1})$		0.002*** (0.0004)		−0.00003 (0.0002)
Constant	0.042*** (0.003)	0.083*** (0.007)	0.042*** (0.002)	0.013*** (0.004)
Observations	55,270	48,443	60,262	49,387
N. unique firms	15,419	13,998	15,483	14,133

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

Embeddedness

Table 2.11: Embeddedness: LD - NetworkAge

	Dependent variable:			
	FirmGrowth		AverageGrowth	
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	0.001** (0.0003)	0.001** (0.0004)	0.001*** (0.0002)	0.001*** (0.0002)
$ClosenessCentrality_{t-1}$	0.009** (0.004)	0.012*** (0.005)	0.020*** (0.002)	0.024*** (0.002)
$LD_{t-1} * AverageNetworkAge_{t-1}$	−0.0002*** (0.00004)	−0.0003*** (0.0001)	−0.0003*** (0.00003)	−0.0003*** (0.00003)
Hub: 1		0.001 (0.003)		−0.005*** (0.001)
$NetworkedFirmsCount_{t-1}$		0.0001 (0.0001)		0.0003*** (0.00005)
LegalNetwork: 1		0.006** (0.003)		0.001 (0.001)
$FirmAge^2_{t-1}$		−0.00000*** (0.00000)		−0.00001*** (0.00000)
AverageFirmSize: Micro		0.013*** (0.005)		−0.020*** (0.003)
AverageFirmSize: SME		0.019*** (0.003)		−0.007*** (0.002)
InnovativeSME: 1		0.025*** (0.009)		0.029*** (0.005)
InnovativeStartup: 1		0.237** (0.120)		0.280*** (0.055)
ROS_{t-1}		0.001*** (0.0001)		0.002*** (0.0001)
$\ln(LabProd_{t-1})$		−0.013*** (0.001)		0.008*** (0.001)
$\ln(LiquidAssets_{t-1})$		0.003*** (0.001)		0.001** (0.0003)
$\ln(IntangibleAssets_{t-1})$		0.002*** (0.0004)		−0.0001 (0.0002)
Constant	0.041*** (0.003)	0.072*** (0.007)	0.040*** (0.002)	−0.003 (0.004)
Observations	55,270	48,443	60,262	49,387
N. unique firms	15,419	13,998	15,483	14,133

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

When including the $LD_{t-1} * AverageNetworkAge_{t-1}$ interaction in the model the Localized Density's impact on firm growth becomes positive with a coefficient of 0.001 for all models.

The coefficients of the interaction term on FirmGrowth is -0.0002 in model (1) and -0.0003 in model (2), both significant at the 1% level. For AverageGrowth the coefficient is -0.0003 for both model (3) and (4) at 1% level.

The positive effect of localized density indicates that firms in higher localized density area tend to grow more, but the negative and significant interaction term indicates that the positive effect of localized density on growth diminishes as the average age of networks to which a firm participates in increases. Suggesting older networks might not be as effective in leveraging the benefits of high localized density, possibly due to a stagnation of the network where benefits provided by density can no longer be translated into significant growth.

Weak Strong Tie

Table 2.12 explores how formal, strong ties in dense networks affect performance.

Table 2.12: WeakStrongTie: LD - LegalNetwork

	Dependent variable:			
	FirmGrowth		AverageGrowth	
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	−0.0002 (0.0005)	−0.0001 (0.0004)	−0.0002 (0.0002)	−0.0004 (0.0003)
$ClosenessCentrality_{t-1}$	0.008** (0.004)	0.009* (0.005)	0.020*** (0.002)	0.019*** (0.002)
LD_{t-1} ∗LegalNetwork:1	−0.0001 (0.0003)	−0.00005 (0.0003)	−0.00004 (0.0001)	0.00002 (0.0002)
Hub: 1		0.0004 (0.003)		−0.004*** (0.001)
$NetworkedFirmsCount_{t-1}$		0.0001 (0.0001)		0.0002*** (0.00004)
$AverageNetworkAge_{t-1}$		−0.003*** (0.001)		−0.004*** (0.0002)
$FirmAge^2_{t-1}$		−0.00000*** (0.00000)		−0.00001*** (0.00000)
AverageFirmSize: Micro		0.013*** (0.004)		−0.019*** (0.003)
AverageFirmSize: SME		0.019*** (0.003)		−0.005** (0.002)
InnovativeSME: 1		0.026*** (0.009)		0.032*** (0.005)
InnovativeStartup: 1		0.228** (0.111)		0.285*** (0.079)
ROS_{t-1}		0.001*** (0.0001)		0.002*** (0.0001)
$\ln(LabProd_{t-1})$		−0.014*** (0.001)		0.008*** (0.001)
$\ln(LiquidAssets_{t-1})$		0.003*** (0.001)		0.001*** (0.0003)
$\ln(IntangibleAssets_{t-1})$		0.002*** (0.0004)		0.00003 (0.0002)
Constant	0.042*** (0.003)	0.081*** (0.007)	0.041*** (0.002)	0.009** (0.004)
Observations	55,270	48,443	60,262	49,387
N. unique firms	15,419	13,998	15,483	14,133

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

The interaction between localized density and legal network membership does not produce a significant impact on firm or average growth (not significant for all models). This suggests that while being part of a legal network and having high centrality within

a network can be beneficial for firm performance, the density of the network does not significantly enhance or detract from these benefits. The strong formal ties provided by legal networks and the advantages of centrality appear to operate independently of the network's localized density.

In summary, the analysis of interaction terms produced some precious findings about the dynamics of how network characteristics interact with localized density to impact economic performance. Key insights are that while centrality promotes firm growth its positive effects are mitigated in densely populated networks. The network embeddedness has a negative impact of firm growth showing how the positive effects of localized density diminishes as the average age of the network increases. The insignificant interaction between localized density and legal network membership suggests that advantages provided by strong formal ties are possibly independent of the localized density of each firm.

2.3.4 Time Split Analysis

To further investigate the determinants of firm performance and enhance the robustness of the analysis, a time split analysis of the various model specifications are employed.
To analyze the time split effects the dataset is separated in 2 time periods: 2016-2019 and 2020-2022, useful to understand to differences in the impact of the explanatory variables on the economic performance.
The pre-pandemic period (2016-2019) represents a relatively stable economic environment, while the pandemic period (2020-2022) introduces unique challenges and uncertainties. This temporal segmentation allows for a comparative analysis that highlights how firms and networks adapt to changing conditions.
To achieve more precise and robust results on Average Growth in this splitting, the measure is recalculated within each period.
InnovativeStartup control variable is removed due to non sufficient amount of cases, not allowing a correct computation and comparison of the models.

Table 2.13 presents four different model specificications, model (1) and model (2) are the models specific to the time period 2016-2019, model (3) and model (4) are related to the time period 2020-2022. For each time period the first column represents the specifications with FirmGrowth as dependent variable and the second column represents the specifications with AverageGrowth as dependent variable.

Table 2.13: Time Split Analysis

	Dependent variable:			
	FirmGrowth	AverageGrowth	FirmGrowth	AverageGrowth
	2016-2019		2020-2022	
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	0.00004 (0.0002)	−0.0004*** (0.0001)	−0.0002 (0.0004)	−0.0003*** (0.0001)
$ClosenessCentrality_{t-1}$	0.016*** (0.006)	0.020*** (0.004)	0.003 (0.007)	0.010*** (0.003)
LD_{t-1}*CC_{t-1}	−0.001 (0.004)	0.003*** (0.001)	−0.013*** (0.004)	−0.004*** (0.001)
Hub: 1	0.0001 (0.003)	−0.007*** (0.002)	0.004 (0.004)	−0.003 (0.002)
$NetworkedFirmsCount_{t-1}$	0.0001 (0.0002)	0.0002* (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
$AverageNetworkAge_{t-1}$	−0.002*** (0.001)	−0.003*** (0.0004)	−0.003*** (0.001)	−0.005*** (0.0004)
LegalNetwork: 1	0.004 (0.004)	0.006*** (0.002)	0.004 (0.005)	−0.001 (0.002)
$FirmAge^2_{t-1}$	−0.00001*** (0.00000)	−0.00001*** (0.00000)	−0.00000** (0.00000)	−0.00001*** (0.00000)
AverageFirmSize: Micro	−0.018*** (0.006)	−0.031*** (0.004)	0.050*** (0.008)	−0.007 (0.005)
AverageFirmSize: SME	0.005 (0.004)	−0.007* (0.004)	0.036*** (0.006)	0.0001 (0.004)
InnovativeSME: 1	0.030** (0.012)	0.037*** (0.006)	0.011 (0.021)	0.033*** (0.008)
ROS_{t-1}	0.002*** (0.0002)	0.003*** (0.0001)	−0.0004* (0.0002)	0.002*** (0.0001)
$\ln(LabProd_{t-1})$	−0.011*** (0.001)	0.004*** (0.001)	−0.017*** (0.002)	0.011*** (0.001)
$\ln(LiquidAssets_{t-1})$	0.001** (0.001)	0.001** (0.0004)	0.005*** (0.001)	−0.00004 (0.0005)
$\ln(IntangibleAssets_{t-1})$	0.001** (0.0005)	−0.0003 (0.0003)	0.003*** (0.001)	0.0004 (0.0003)
Constant	0.086*** (0.009)	0.017*** (0.006)	−0.045*** (0.012)	0.020*** (0.006)
Observations	19,312	19,951	29,131	30,328
N. unique firms	8,921	9,227	12,373	12,682

Note: LAD estimates and standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

The effect of LocalizedDensity on FirmGrowth, shows a non-significant positive association in 2016-2019 and a non-significant negative association in 2020-2022. However, it is negatively and significantly (p<0.01) associated with AverageGrowth of firms in both periods. This indicates that higher localized density consistently constrains average growth, but its impact on firm growth is not statistically significant in either period.

ClosenessCentrality is positively and significantly associated with both FirmGrowth and AverageGrowth during 2016-2019, suggesting that centrality within the network leads to higher growth during stable economic conditions.In the 2020-2022 period, Closeness-Centrality becomes non-significant for FirmGrowth but remains positively significant for AverageGrowth at the 1% level, indicating that centrality continues to benefit average growth even during the pandemic, though its effect on firm growth diminishes.

To study the dynamic effects of LocalizedDensity, the interaction with ClosenessCentrality is included. LD_{t-1}*CC_{t-1} has insignificant negative impact (-0.001) on FirmGrowth and is significant and positive (0.003) for AverageGrowth in 2016-2019, implying a synergistic effect in stable periods. However, this interaction becomes significantly negative in 2020-2022 for both FirmGrowth (-0.013) and AverageGrowth (-0.004) at 1% level, indicating that the benefits of centrality are reduced in high-density areas during the

pandemic.

Micro-sized firms exhibit significant negative effects on both FirmGrowth and AverageGrowth during 2016-2019. However, from 2020 to 2022, Micro-firms show a significant positive association with FirmGrowth, while their impact on AverageGrowth becomes non-significant, suggesting that smaller firms participating in networks performed better than larger firms in less stable periods.

SMEs show a marginal positive effect on FirmGrowth (not significant) and a significant negative effect on AverageGrowth ($p<0.1$) in 2016-2019. During 2020-2022, SMEs show a significant positive association with FirmGrowth but no significant impact on AverageGrowth, indicating that SMEs participating in networks also adapted well to the unstable environment.

The remaining characteristic controls show consistent impacts aligning with previous models.

The analysis reveals that while localized density consistently constrains average growth, its impact on firm growth varies with economic conditions. Closeness centrality is beneficial during stable times but less impactful during crises, and the interaction between density and centrality highlights a complex dynamic where the advantages of centrality diminish in densely populated networks during unstable periods.

2.4 Panel Regression

The dataset used in this research is a panel dataset or longitudinal dataset, it is a multi-dimensional dataset containing time series observations of multiple entities (in this case, firms) across different dimensions (firm growth, firm age, region group, average size etc.). The characteristics of a panel dataset allows the study of changes over time while considering differences among firms and other characteristics using panel regression.

A panel regression leverages the characteristics of the panel structure allowing the study of cross-sectional variations (across entities) and time-series variations (across time).

There are different types of panel regression models:

- **Fixed Effects Model:** it controls for time-invariant characteristics of the entities using entity-specific intercept.
- **Random Effects Model:** it assumes random entity-specific effects that are uncorrelated with all predictors included in the model.

In the context of this research, fixed effect models are used to control for potential correlation between entity-specific effects on predictors, it is reasonable to assume such correlations because of the firm-specific characteristics, firms are likely to have unobserved characteristics that could influence their performance and be correlated also with predictors. For example, management quality, firm culture, financial strategy and other qualities that are different for each firm influence firm growth and can be correlated with ROS, liquidity or intangible assets.

The regression equation for a fixed effect model would be:

$$\begin{aligned}\text{FirmGrowth}_{it} = \alpha_i + \beta_1 \text{LocalizedDensity}_{i,t-1} + \beta_2 \text{ClosenessCentrality}_{i,t-1} \\ + \delta_t \text{Year}_t + \gamma_j \text{Sector}_j + \theta_k \text{RegionGroup}_k + \epsilon_{it}\end{aligned} \tag{2.3}$$

Where:

- FirmGrowth_{it}: Dependent variable representing the growth of firm i at time t.
- α_i: The firm-fixed effect, capturing time-invariatn characteristics of firm i.
- β_0: Intercept term.

- β_1: Coefficient for the effect of localized density on firm growth.
- $\text{LocalizedDensity}_{i,t-1}$: Lagged independent variable representing the localized density for firm i at time $t-1$.
- β_2: Coefficient for the effect of closeness centrality on firm growth.
- $\text{ClosenessCentrality}_{i,t-1}$: Lagged independent variable representing the closeness centrality for firm i at time $t-1$.
- $\delta_t \text{Year}_t$: Time fixed effects, where δ_t are the coefficients for each year dummy variable Year_t.
- $\gamma_j \text{Sector}_j$: Sector fixed effects, where γ_j are the coefficients for each sector dummy variable Sector_j.
- $\theta_k \text{RegionGroup}_k$: Geographic fixed effects, where θ_k are the coefficients for each region group dummy variable RegionGroup_k.
- ϵ_{it}: Error term capturing unobserved factors affecting firm growth for firm i at time t.

By controlling for network and firm specific controls would be:

$$\begin{aligned} \text{FirmGrowth}_{it} = \alpha_i &+ \beta_1 \text{LocalizedDensity}_{i,t-1} + \beta_2 \text{ClosenessCentrality}_{i,t-1} \\ &+ \beta_3 \text{NetworkChar}_{it} + \beta_4 \text{FirmChar}_{it} + \beta_5 \text{FirmPerf}_{it} \\ &+ \delta_t \text{Year}_t + \gamma_j \text{Sector}_{ij} + \theta_k \text{Geography}_{ik} + \epsilon_{it} \end{aligned} \tag{2.4}$$

Where:

- FirmGrowth_{it}: Dependent variable representing the growth of firm i at time t.
- α_i: The firm fixed effect, capturing time-invariant characteristics of firm i.
- β_1: Coefficient for the effect of localized density on firm growth.
- $\text{LocalizedDensity}_{i,t-1}$: Independent variable representing the localized density for firm i at time $t-1$.
- β_2: Coefficient for the effect of closeness centrality on firm growth.
- $\text{ClosenessCentrality}_{i,t-1}$: Independent variable representing the closeness centrality for firm i at time $t-1$.
- β_3: Coefficient for the effect of network characteristic controls on firm growth.

- NetworkChar$_{it}$: Vector of control variables related to network characteristics.
- β_4: Coefficient for the effect of firm characteristic controls on firm growth.
- FirmChar$_{it}$: Vector of control variables related to firm-specific characteristics.
- β_5: Coefficient for the effect of financial performance controls on firm growth.
- FirmChar$_{it}$: Vector of control variables related to firm-specific financial performance.
- δ_tYear$_t$: Time fixed effects, where δ_t are the coefficients for each year dummy variable Year$_t$.
- γ_jSector$_j$: Sector fixed effects, where γ_j are the coefficients for each sector dummy variable Sector$_j$.
- θ_kRegionGroup$_k$: Geographic fixed effects, where θ_k are the coefficients for each region group dummy variable RegionGroup$_k$.
- ϵ_{it}: Error term capturing unobserved factors affecting firm growth for firm i at time t.

2.4.1 Panel Regression Models

In this section panel regression with fixed effect models are analyzed. Using the fixed effect model we account for individual fixed effects, accounting for unobserved heterogeneity across firms that may influence the dependent variable. The models include year, sector, and region fixed effects to control for unobserved heterogeneity across time, industries, and regions.

Table 2.14 presents the results from fixed effect regression models with depedent variable FirmGrowth and AverageGrowth.

The models are employed progressively starting from a basic fixed effect model which only accounts for the effect of the independent variables (localized density and centrality) on firm growth and average growth (model 1 and 4 respectively).

In model (2) and (5) control for network characteristic controls are added. At last, in Model (3) and (6) firm level controls are added.

Campared to previous OLS and LAD models firm specific controls such as AverageFirmSize, InnovativeStartup and InnovativeSMEs are excluded as they are accounted for by

the firm-fixed effect.

Table 2.14: Fixed Effect Models

	Dependent variable:					
	FirmGrowth			AverageGrowth		
	(1)	(2)	(3)	(4)	(5)	(6)
$LocalizedDensity_{t-1}$	−0.005 (0.004)	−0.005 (0.005)	−0.003 (0.003)	0.0003 (0.004)	0.00005 (0.005)	−0.0004 (0.003)
$ClosenessCentrality_{t-1}$	0.051 (0.051)	0.044 (0.054)	0.026 (0.044)	0.007 (0.051)	0.008 (0.054)	0.012 (0.044)
Hub: 1		−0.005 (0.038)	−0.013 (0.030)		0.004 (0.038)	0.011 (0.030)
$NetworkedFirmsCount_{t-1}$		−0.0003 (0.001)	−0.0002 (0.001)		0.0001 (0.001)	0.0002 (0.001)
$AverageNetworkAge_{t-1}$		0.005 (0.006)	0.001 (0.005)		−0.003 (0.006)	0.0004 (0.005)
LegalNetwork: 1		0.013 (0.018)	0.010 (0.014)		0.003 (0.018)	0.003 (0.014)
$FirmAge^2_{t-1}$			0.0003*** (0.0001)			0.00001 (0.0001)
ROS_{t-1}			0.001** (0.001)			0.0001 (0.001)
$\ln(LabProd_{t-1})$			−0.485*** (0.018)			−0.008 (0.018)
$\ln(LiquidAssets_{t-1})$			0.001 (0.003)			−0.0002 (0.003)
$\ln(IntangibleAssets_{t-1})$			0.013*** (0.003)			0.0003 (0.003)
Observations	49,536	49,536	43,387	53,934	53,934	43,946
N. unique firms	11,414	11,384	10,445	11,188	11,162	10,317
R^2	0.050	0.050	0.196	0.008	0.009	0.010

Note: Robust standard errors are given in parentheses. Standard errors are clustered at firm level. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

LocalizedDensity shows negative but insignificant effect on FirmGrowth, on Average-Growth the effect is also non-significant and becomes negative when all control variables are added.

ClosenessCentrality exhibit positive coefficients across all models but are not statistically significant.

Differently from previous models, none of the network characteristics have a significant impact on growth rates. Firm controls show higher level of significance on FirmGrowth, while are insignificant for the smoothed AverageGrowth. This is possibly due to the nature off AverageGrowth being an aggregate measure, which tends to have lower variability making it less sensitive to firm operational controls.

The reason of the non significance of all network characteristics contrary to all previous models could be found in the variability of these variables showed in table 2.15, each row represents the network variables of interest with variance values represented as minimum, first quantile, median, mean, third quantile and maximum variance.

The median variance for almost all variables is near 0, except for AverageNetworkAge, which is expected to vary as networks grow older each year. The first and third quantile variances for both Localized Density and Closeness Centrality are very low (0.3 and 0, respectively), indicating that the median firm, once part of one or more networks, rarely changes its position within the network.

Additionally, for the remaining network variables, once a firm joins a set of networks,

Table 2.15: Variances of Network Characteristic Variables

Variable	Min	1st Qu.	Median	Mean	3rd Qu.	Max
LocalizedDensity_var	0.0000	0.0000	0.0002	8.5682	0.3333	2097.3779
ClosenessCentrality_var	0.000	0.000	0.000	0.003	0.000	0.395
Hub_var	0.000000	0.000000	0.000000	0.001177	0.000000	0.333333
NetworkedFirmsCount_var	0.000	0.000	0.000	48.925	4.267	15492.000
AverageNetworkAge_var	0.000	1.000	1.667	2.159	3.500	19.000
LegalNetwork_var	0.00000	0.00000	0.00000	0.02799	0.00000	0.50000

the composition tends to remain stable with only a few outliers showing significant variance, impacting the mean (NetworkedFirmsCount and Localized Density). This stability suggests that firms typically do not experience much change in network composition over time.

Given this context, it is reasonable to understand why the FE models show non-significant results for these network characteristics. The lack of significant variability in these variables over time means that participation in networks with certain characteristics depend on firm-specific unobservable characteristics like managerial skills and decision making capacity, which are accounted for in the FE model when controlling for time-invariant characteristics. In contrast, the LAD regression effectively captures the influence of these variables.

Chapter 3

Methodology

This chapter outlines the methodologies employed to investigate the impact of network contracts and spatial proximity on economic performance on participating firms, it includes the data preparation and cleaning process, the construction of the panel dataset, the computation of spatial measures and the statistical techniques employed.

In the initial stage of the research, the Network Contract Dataset was subjected to cleaning and preparation processes to ensure consistency. This was accomplished using the R programming language, specifically using libraries such as `readxl` for reading Excel files, `data.table` for high-performance data processing, and `stringi` for string operations, alongside base functions of R.

3.1 Data Collection and Preparation

The first step of the research is the cleaning and preparation of the data (Network Contract Dataset) aiming to achieve a consistent panel dataset. This is performed using the R programming language, utilizing libraries such as `tidyverse` for data manipulation, `data.table` for high-performance data processing, and `stringi` for string operations, alongside base functions of R.

The process addresses several critical issues, among which are:

- Naming convention and format inconsistencies: the naming of variables, columns, or categories were non-consistent across different years and between the `Elenco` and `Soggetto Giuridico` datasets.
- Missing values: the presence of missing values within variables across different years

due to human error during data entry or database inconsistencies.

- Data duplications: duplicated records occur within the dataset due to overlapping firms between the `Elenco` and `Soggetto Giuridico` datasets or inadvertent duplication of network entries caused by naming errors or entry duplication when registering network information.

3.1.1 Data Integration and Standardization

The datasets are provided by Confindustria, and are structured in Excel files for each year from 2016 to 2023. Each Excel file is composed by two sheets, `Elenco` and `Soggetto Giuridico`. The latter representing a table of firms and networks which formed a legal contract.

The data tables are loaded separately and column names are standardized to avoid inconsistencies for merging purposes. A column `'identification'` is created and 0 is assigned for each record present in the `Elenco` dataframe, 1 for records in `Soggetto Giuridico`.

Following the column name standardization, 3 functions are created and used for the cleaning process:

- `clean_text`: text cleaning function, which cleans any input element from punctuation, blanks, quotation marks.
- `clean_descriptions`: a column cleaning function which cleans and standardizes text within specified column of a data frame adding the cleaned column to the original data frame.
- `merge_and_process`: a merge and process function which combines data frames `Elenco` and `Soggetto Giuridico` for specified year and applies above functions to the columns. After the process it creates a year column specifying which year merged dataset is originated from.

Combining the merged datasets for each year creates the effective panel data structure that will be used, containing the time series for each firm and network.

Further processing is performed on the combined panel by removing the rows with missing firm taxcode values as these cannot be recovered anywhere, patterns such as

'RETE', 'RETI', 'NETWORK', 'NET' specifying networks are removed from network names. Old names are kept for consistency checks in successive steps.

The process above handles complex data merging, cleaning and processing tasks sequentially, preparing the data for the next step on duplication detection.

3.1.2 Duplicates detection

The main issue of the dataset is the problem of duplicated records within the dataset caused by naming errors or entry duplication when registering network information. Example of naming errors can be:

- Simple typing errors: where one letter is missing, or some letters have reversed orders and other kind of typos. (i.e. 'ARIANNAILFILODELLARICOSTRUZIONE', 'ARIANNILFILODELLARICOSTRUZIONE')
- Insertion errors: where one word or multiple extra words are included with the result that a network can be present twice in the same year with the two different names. (i.e. 'RETEIMPRESEBALNEARIVIAREGGIO', 'RETEIMPRESEBALNEARIVIAREGGIORIVA')

Initially the idea was to use network identifiers such as `act_number`, `network_taxcode`, `repertoire_number` and `network_repertoire_number`. But the codes are not unique identifiers due to input errors: a single identification is attributed to multiple different network names. For example, table 3.1 shows the top 5 act numbers and unique network names to which it is associated.

Table 3.1: Top 5 Act Numbers associated to different networks

act_number	n_networks
3	11
1	7
149	6
166	6
1197	5

In total there are 1103 act number out of 6987 associated with multiple networks. Similarly, 17 out of 1608 network taxcodes, 344 out of 7960 repertoire number and 30 out of 1493 network repertoire number associated with more than 1 network.

Considering non reliability of the identifiers, and the presence of insertion errors in the name, to achieve robustness and non duplicated network-firm records, similarity matching is performed on networks names.

Similarity Matching

To find matching name pairs, Jaro-Winkler similarity scores is calculated on pair of network name strings.

Jaro-Winkler is a metric used for measuring the similarity between two strings. It combines the Jaro Similarity measure where the similarity is measured based on the number and order of the matching characters to the Winkler adjustment which provides better ratings to strings with matching initial characters, useful for typographical errors.

The formula for Jaro similarity is:

$$J = \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) \tag{3.1}$$

Where:

- $|s_1|$ and $|s_2|$ are the lengths of the two strings.
- m is the number of matching characters.
- t is the number of transpositions.

The Winkler adjustment is an enhancement to the Jaro similarity that gives more favorable ratings to strings that match from the beginning. This is useful in cases where typographical errors are more likely to occur at the end of the string.

The adjustment formula is:

$$W = J + (l \cdot p \cdot (1 - J)) \tag{3.2}$$

Where:

- J is the Jaro similarity score.
- l is the length of the common prefix at the start of the string, up to a maximum of 4 characters.
- p is a scaling factor, set to 0.1.

The Jaro-Winkler score of the example cited before('ARIANNAILFILODELLARICOSTRUZIONE' and 'ARIANNILFILODELLARICOSTRUZIONE') is 0.96, correctly identifying the matching names. The total number of unique network names before cleaning is 8711. To perform the similarity matching of these unique network names more than 30.000.000 pairs need to be created, which requires to maximize performance and efficiency through parallel processing to perform the computations leveraging multiple cores to distribute the workload. Jaro-Winkler score is computed for each pairs using the libraries `RecordLinkage`, `doParallel`, `foreach`.

```
combinations <- CJ(String1 = unique_net, String2 = unique_net)[String1 <= String2]
setDT(combinations)

no_cores <- detectCores() - 1
cl <- makeCluster(no_cores)

string1 <- combinations$String1
string2 <- combinations$String2
clusterEvalQ(cl, library(RecordLinkage))
clusterExport(cl, varlist = c("string1", "string2"))

JaroWinklerSimilarity <- parLapply(cl, seq_along(string1), function(i) {
  jarowinkler(string1[i], string2[i])})
combinations[, JaroWinklerSimilarity := unlist(JaroWinklerSimilarity)]
combinations <- combinations %>% arrange(desc(JaroWinklerSimilarity))

stopCluster(cl)
registerDoSEQ()

high_similarity <- combinations[JaroWinklerSimilarity >= 0.90]
```

The above R code generates combinations of network names and ensures each pair is unique and not repeated, then it setup the parallel computing environment by finding the number of CPU cores available and initializes a cluster for parallel computing. Package `RecordLinkage` and the string pairs are loaded into each cluster, after this Jaro-Winkler similarity is calculated and stored in column JaroWinklerSimilarity in the data table with the string pairs. In the end clusters are stopped and only similarity scores above 90% are saved, creating a table with 5400 pairs of similar network names.

To identify the paired networks, an ID code identifying the pair is created, network names associated to more than 1 pair need to be attributed the same ID, the process is

coded as the following:

```
similarities <- df_similarities %>%
  arrange(desc(JaroWinklerSimilarity)) %>%
  mutate(new_index = row_number()) %>%
  select(new_index, String1, String2)%>%
  pivot_longer(cols = c(String1, String2), names_to = "StringType", values_to = "StringValue")
     %>%
  select(new_index, StringValue) %>%
  rename(network_name = StringValue, ID = new_index) %>%
  group_by(network_name) %>%
  slice(1) %>%
  ungroup()%>%
  group_by(ID) %>%
  mutate(count = n()) %>%
  ungroup()

singles <- similarities %>% filter(count %% 2 == 1)
multiples <- similarities %>% anti_join(singles)
```

An ID column is created using the index number for each network pair, the table is then converted to a long format so each network name is associated with the number grouping similar network names with the same ID. The long format conversion is used also the keep only one for each unique network without repetition, identifying 2687 unique networks.

There are cases of IDs with only a single network associated, this happens in cases when one network is paired with multiple networks, so the paired table is saved as dataframe `multiples` (1690 records) and a table with IDs associated with single networks is saved as `singles` (997 records).

```
singles <- singles %>%
  rowwise() %>%
  mutate(new_ID = {
    scores <- sapply(multiples$network_name, function(multiple) {
      jarowinkler(network_name, multiple)
    })
    if (length(scores) > 0 && max(scores) > 0.95) {
      multiples$ID[which.max(scores)]
    } else {
      ID
    }
  }) %>%
  select(-ID) %>%
```

```
  rename(ID=new_ID)

similarities <- bind_rows(singles, multiples) %>%
  select(ID, network_name)
```

The above step improves the accuracy of matching. For network name entries in 'singles', it tries to find a corresponding entry in multiples that is very similar (similarity score above 0.95). This helps in identifying and consolidating near-duplicate entries.
In the end the two datasets are recombined into one achieving a more accurate grouping of similar network names. The IDs are joined into the main dataset using network name as key for the operation. A total number of 84,900 records over 277,739 have an ID number.

A high similarity score alone is not insufficient to determine whether the networks are the same, additional control checks are performed prior to network name uniforming.
A first check consists in controlling if networks with the same ID have the same firms associated with them across multiple years. The function `check_same_firms` performs this operation and returns two dataframes: `df_trues` containing networks with identical firm member list across all years and `df_false` containing networks with non-matching firm lists.

```
check_same_firms <- function(df) {
  # Create a unique list of firm tax codes for each ID, network name, and year
  df_unique_imp <- df %>%
    filter(!is.na(ID)) %>%
    group_by(ID, network_name, year) %>%
    summarise(unique_cf_impresa = list(sort(unique(firm_taxcode))),
              .groups = 'drop')

  # Filter IDs with more than one distinct network name
  df_filtered <- df_unique_imp %>%
    group_by(ID) %>%
    filter(n_distinct(network_name) > 1) %>%
    ungroup()

  # Helper function to check if all lists of firm tax codes are identical
  all_lists_equal <- function(lists) {
    all(map2_lgl(lists[-1], lists[-length(lists)], ~identical(.x, .y)))
  }

  # Check if all lists of firm tax codes are identical for each ID and year
```

```
  df_filtered <- df_filtered %>%
    group_by(ID, year) %>%
    mutate(all_lists_equal = ifelse(n_distinct(network_name) > 1 & all_lists_equal(unique_cf_
    impresa), TRUE, FALSE)) %>%
    ungroup()

  # Filter IDs where all lists are equal
  df_trues <- df_filtered %>%
    group_by(ID) %>%
    filter(all(all_lists_equal == TRUE)) %>%
    ungroup()

  # Get unique network names where all lists are equal
  unique_network_names <- unique(df_trues$network_name)

  # Filter out the rows where network names are in the unique list
  df_false <- df_filtered %>%
    filter(!network_name %in% unique_network_names)

  return(list(df_trues = df_trues, df_false = df_false))
}
```

Networks in `df_trues` are already identified as same network as they have high similarity score and all years with the same network participants, a total of 38 observations (8 network names) meet these requirements. For the remaining networks (saved in `df_false, with 7696 observations, 1733 unique network names`), the function `high_common` evaluates and filters network names based on the proportion of common firm participants, if all the participants in one network are the same as the bigger network with same ID. It identifies the networks with high commonality or full containment as the same network, while others with lower commonality are considered different. The rationale is that if most of the firms still match, it is reasonable to believe it is a duplication, if very low percentage of participants match then it is a different network.

```
high_common <- function(df_trues, df_false) {
  # Create a dataframe to preserve network_name information
  denominazione_info <- df_false %>%
    group_by(ID, year) %>%
    summarise(network_name = list(unique(network_name)), .groups = 'keep') %>%
    ungroup()

  # Combine the information to have similarity ID, year, common_elements, and name of the
    networks
```

```
  intersections <- df_false %>%
    group_by(ID, year) %>%
    summarise(common_elements = list(Reduce(intersect, unique_cf_impresa)), .groups = 'keep')
    %>%
    ungroup() %>%
    left_join(denominazione_info, by = c("ID", "year"))

  # Join with df_false and calculate the proportion of common elements in each network
  df_proportions <- df_false %>%
    left_join(intersections, by = c("ID", "year")) %>%
    rename(network_name = network_name.x) %>%
    filter(map_int(network_name.y, length) > 1) %>%
    select(-all_lists_equal, -network_name.y) %>%
    filter(map_lgl(common_elements, ~ length(.x) > 0)) %>%
    group_by(network_name, year) %>%
    mutate(common_percent = length(unlist(common_elements))/length(unlist(unique_cf_impresa)))
     %>%
    ungroup() %>%
    # Calculate the mean common percentage and relevant flags
    group_by(ID, year) %>%
    mutate(mean_perc = sum(common_percent)/n(),
           above_threshold = all(common_percent >= 0.6),
           contained = any(common_percent == 1)) %>%
    ungroup()

  # Filter networks with high commonality
  df_common <- df_proportions %>%
    group_by(ID) %>%
    filter(above_threshold == TRUE | contained == TRUE) %>%
    ungroup()

  # Determine networks considered the same
  df_same <- bind_rows(df_common, df_trues) %>%
    select(-all_lists_equal) %>%
    distinct(ID)

  # Determine networks considered not the same
  not_same <- df_false %>%
    filter(!ID %in% df_same$ID) %>%
    distinct(ID)

  return(list(df_same = df_same, not_same = not_same))
}
```

The function returns a list containing two dataframes: `df_same`, containing the ID numbers from `df_true` and the ID numbers of networks that satisfy the commonality thresholds (61 distinct ID groups and 123 networks distinct network names), and dataframe

`not_same`, containing all the firms excluded by the functions.
A unique network name is assigned to all networks with same ID within `df_same`. Networks classified as `not_same` retain their original network name.

Jaro-Winkler similarity score is very at matching strings with similar lengths, but there are situations where one name is substring of another one and are clearly refering to the same network.
For example: 'PSVANTINCENDIOCONILMARCHIOPSVANTINCENDIO' and 'PSVANTINCENDIO', the two network names have a similarity score of 0.87 and gets excluded by the initial filter of similarity score above 90%.

To find network names with this characteristic the function `find_pairs` is created, looking for pairs of network names where one is substring of the second one.

```
find_pairs <- function(list_denom) {
  # Create all combinations of indices
  combos <- expand.grid(i = seq_along(list_denom), j = seq_along(list_denom), stringsAsFactors
     = FALSE)

  # Filter out same-element pairs and where i's string is not shorter than j's
  filtered_combos <- subset(combos, i != j & nchar(list_denom[i]) < nchar(list_denom[j]))

  # Check if i's string is contained in j's
  contained <- str_detect(list_denom[filtered_combos$j], fixed(list_denom[filtered_combos$i]))

  # Subset the combos where containment is true
  valid_combos <- filtered_combos[contained, ]

  # Create the final data frame of pairs
  pairs <- data.frame(
    container = list_denom[valid_combos$j],
    contained = list_denom[valid_combos$i],
    stringsAsFactors = FALSE
  )
  return(pairs)
}
```

The function returns a data frame of pairs where each shorter string (contained) is found within a longer string (container). Before applying the function a preprocessing is performed by grouping for firm tax code, finding effectively cases when firms are registered with both network (contained string and container string). Groups with only 1 network name are removed as they are irrelevant and to save computational resources.

Similar to previous process an ID is assigned to pairs of network names, after joining the IDs to the main dataset, 7409 records are affected, a total number of 105 distinct IDs and 206 network names associated to each other. The same process as before is performed with functions `check_same_firms` and `high_common`.
Four pairs (4 IDs) of networks are found with same members across all years:

Table 3.2: Network Names by ID

ID	**network_name**
16	NEST
16	NESTDIIMPRESE
67	DIIMPRESETEAMWORKTHEPARTNERSHIP
67	TEAMWORKTHEPARTNERSHIP
90	BARICENTROCONFIDINSIEME
90	CONFIDINSIEME
96	CINEMA
96	DIIMPRESECINEMA

After the `high_common` function, a total number of 40 IDs are identified as same affecting 80 network names.

Over the 277,739 records, 10,246 have missing network name (after removing generic network names like CONTRATTODIRETE, RETE, IMPRESEINRETE.

Over these 10,246 two main patterns are discovered in the data as all data has a pair of act_number and repertoire_number or a pair of network_taxcode and network_repertoire_number. Network names are grouped in 3 groups (non-NA act_number and repertoire_number as group 0, non-NA network_taxcode and network_repertoire_number as group 1, all other as group 0) to check the distribution of NA values: 10,237 over the total missing networks names are part of group 1, and group t has 9 observations without network name, this means by using a combination of identifying codes it is possible to retrieve all the missing networks.

Before assigning the network names using network identifiers checking whether a record may be part of networks but the network name was assigned as a generic one thus becoming NA in the processing is needed. The remaining missing network names can be addressed by combination of act numbers and other network identification codes.

```
filter0 <- panel %>%
  filter(gruppo == 0) %>%
  group_by(act_number, repertoire_number) %>%
```

```
  mutate(object = if(any(!is.na(object)) & n_distinct(na.omit(object)==1))
    first(object[!is.na(object)]) else NA) %>%
  ungroup() %>%
  mutate(object = coalesce(object, paste0(act_number, repertoire_number))) %>%
  group_by(firm_taxcode, act_number, repertoire_number) %>%
  mutate(unique = n_distinct(na.omit(network_name))) %>%
  mutate(new_name = ifelse(any(is.na(network_name)) & unique == 1,
    first(na.omit(network_name)), NA_character_)) %>%
  mutate(new_name = ifelse(any(is.na(network_name)) & unique > 1,
    names(sort(table(na.omit(network_name)), decreasing = TRUE))[1], new_name)) %>%
  mutate(new_name = ifelse(unique > 1,
    names(sort(table(na.omit(network_name)), decreasing = TRUE))[1], new_name)) %>%
  ungroup() %>%
  group_by(network_name) %>%
  mutate(new_name = ifelse(!is.na(network_name) & any(is.na(new_name)) &
    n_distinct(new_name) == 1, first(na.omit(new_name)), new_name)) %>%
  mutate(network_name = ifelse(!is.na(new_name), new_name, network_name))

sum(is.na(filter0$network_name))

filter1 <- panel %>%
  filter(gruppo == 1) %>%
  group_by(network_taxcode, network_repertoire_number) %>%
  mutate(object = if(any(!is.na(object)) & n_distinct(na.omit(object)==1))
    first(object[!is.na(object)]) else NA) %>%
  ungroup() %>%
  mutate(object = coalesce(object, paste0(act_number, repertoire_number))) %>%
  group_by(firm_taxcode, network_taxcode, network_repertoire_number) %>%
  mutate(unique = n_distinct(na.omit(network_name))) %>%
  mutate(new_name = ifelse(any(is.na(network_name)) & unique == 1,
    first(na.omit(network_name)), NA_character_)) %>%
  mutate(new_name = ifelse(any(is.na(network_name)) & unique > 1,
    names(sort(table(na.omit(network_name)), decreasing = TRUE))[1], new_name)) %>%
  mutate(new_name = ifelse(unique > 1,
    names(sort(table(na.omit(network_name)), decreasing = TRUE))[1], new_name)) %>%
  ungroup() %>%
  group_by(network_name) %>%
  mutate(new_name = ifelse(!is.na(network_name) & any(is.na(new_name)) &
    n_distinct(new_name) == 1, first(na.omit(new_name)), new_name)) %>%
  mutate(network_name = ifelse(!is.na(new_name), new_name, network_name))

sum(is.na(filter1$network_name))

filter2 <- panel %>%
  filter(gruppo == 2) %>%
  mutate(comb = coalesce(network_taxcode, repertoire_number)) %>%
  group_by(comb) %>%
  mutate(object = if(any(!is.na(object)) & n_distinct(na.omit(object)==1))
    first(object[!is.na(object)]) else NA) %>%
```

```
  ungroup() %>%
  mutate(object = coalesce(object, paste0(act_number, repertoire_number))) %>%
  group_by(firm_taxcode, comb) %>%
  mutate(unique = n_distinct(na.omit(network_name))) %>%
  mutate(new_name = ifelse(any(is.na(network_name)) & unique == 1,
    first(na.omit(network_name)), NA_character_)) %>%
  mutate(new_name = ifelse(any(is.na(network_name)) & unique > 1,
    names(sort(table(na.omit(network_name)), decreasing = TRUE))[1], new_name)) %>%
  mutate(new_name = ifelse(unique > 1,
    names(sort(table(na.omit(network_name)), decreasing = TRUE))[1], new_name)) %>%
  ungroup() %>%
  group_by(network_name) %>%
  mutate(new_name = ifelse(!is.na(network_name) & any(is.na(new_name)) &
    n_distinct(new_name) == 1, first(na.omit(new_name)), new_name)) %>%
  mutate(network_name = ifelse(!is.na(new_name), new_name, network_name)) %>%
  select(-comb)

sum(is.na(filter2$network_name))

union <- bind_rows(filter0, filter1, filter2) %>%
  group_by(old_network_name) %>%
  mutate(
    new_name = if(n_distinct(na.omit(new_name)) == 1 && any(is.na(new_name))) {
      unique(na.omit(new_name))[1]  # Assign the unique non-NA new_name if there's exactly one
    } else {
      new_name  # Otherwise, keep the original new_name values
    },
    network_name = ifelse(!is.na(new_name), new_name, network_name)
  ) %>%
  ungroup()

panel <- union %>%
  select(-new_name, -gruppo, -unique)
```

The above code tries to standardize and clean network_name column. Grouping by the pair of act_number and repertoire_number or network_taxcode and network_repertoire_number. Grouping by firm and network identification code it is possible to find whether the firm was participating in certain network and the data got lost during processing, the script computes the number of distinct network_name values and conditionally creates a new_name column: if a single unique network_name exists amidst NAs, it assigns this name to all. If multiple unique names exist, the most common one is assigned. In the final part of each script the data is grouped again by network_name, any NAs in new_name are replaced with the most common name if consistency is confirmed within

the group. If new_name is valid (there is only 1 unique non-NA new name), it replaces network_name.

The process is performed for each group created before, ensuring the standardization of network names. Moreover, 143 observations were recovered in group 0.

After standardization and the attempt to recover NA values, remaining missing values are filled by concatenating act_number and repertoire_number or network_taxcode and network_repertoire_number, if those values are available.

```
panel <- panel %>%
  mutate(network_name = ifelse(is.na(network_name),
                               ifelse(!is.na(act_number) & !is.na(repertoire_number),
                                      paste(act_number, repertoire_number, sep = "-"),
                                      ifelse(!is.na(network_taxcode) & !is.na(network_
    repertoire_number),
                                             paste(network_taxcode, network_repertoire_
    number, sep = "-"),
                                             network_name)),
                               network_name))
```

The process effectively results in 0 missing values.

Before eliminating duplicated values, act date is uniformed, first any network with missing act date are filled with non-missing act date of that specific network, if network remains missing the average time difference between act date and first year in network is estimated which is 1.43, and attributed (only 1 single network has no act date in any year).

```
panel <- panel %>%
  group_by(firm_taxcode, network_name, year) %>%
  filter(!duplicated(firm_taxcode) | row_number() == 1) %>%
  ungroup()
```

Duplications occur when records share the same firm tax code, network name and year. By grouping by these columns and keeping only 1 row consistently eliminates duplicated firm-network records. Eliminating 11,302 observations achieving the final cleaned dataframe with 266,437 observations.

3.2 Spatial measures creation

This section focus on the creation of spatial measures that are used as independent variables. Measures include the Localized Density and Closeness Centrality measures, which differentiate in structural centrality in the star network configuration and geographical centrality in the complete network configuration.

Libraries used for the computation of spatial measures are `tidyverse` for data manipulation, `geosphere` for calculating geographical distances and `igraph` for network analysis.

The dataset used is the cleaned version from the previous section, which has had duplicates removed. Although the dataset includes the municipalities of each firm, it lacks of latitude and longitude data, key metrics for the correct calculation of spatial measures. Therefore a first steps involves cleaning and processing the geographical data.

```
panel <- panel %>%
  arrange(firm_taxcode, year) %>%
  group_by(firm_taxcode) %>%
  fill(municipality_firm, .direction = "down") %>%  # Fill NA downwards first
  fill(municipality_firm, .direction = "up") %>%
  ungroup() # Then fill NA upwards

na <- panel %>% filter(is.na(municipality_firm)) %>% distinct(firm_taxcode, firm_name,
    municipality_firm)
#write.xlsx(na, "Comuni/na_comuni.xlsx")
na <- read_excel("Comuni/na_comuni.xlsx")

panel <- panel %>%
  left_join(na %>% select(1,3), by = "firm_taxcode") %>%
  rename(municipality_firm = municipality_firm.x) %>%
  mutate(municipality_firm = coalesce(municipality_firm, municipality_firm.y)) %>%
  select(-municipality_firm.y)
```

In this part of the script fills the missing municipalities mixed in years for each firm, attributing the previous year's municipalities, if no present it attributes the next year's municipality. Then if some firm still have missing municipalities it's saved in a separate excel file (14 firms) these firms' municipality is manually added after looking for the municipality on Italian's firm registry.

```
# Loading coordinates
```

```
comuni_coord <- read_excel("Comuni/coord_comuni.xlsx", skip = 1) %>% select(1, 4,8,9) %>%
  rename(comune = denominazione_ita, sigla = sigla_provincia) %>%
  mutate(comune_cleaned = iconv(comune, from = "UTF-8", to = "ASCII//TRANSLIT"),
         comune_cleaned = trimws(tolower(comune_cleaned)),
         comune_cleaned = gsub("-", " ", comune_cleaned),
         comune_cleaned = gsub("[[:punct:]]", "", comune_cleaned)) %>%
  rename(old_name = comune_cleaned,
         province_firm = sigla)

# Add region code
cod_regioni <- read.csv("Comuni/comuni.csv") %>%
  select(1,4,5,6) %>%
  mutate(sigla = ifelse(is.na(sigla), "NA", sigla)) %>%
  mutate(comune = iconv(comune, from = "UTF-8", to = "ASCII//TRANSLIT"),
         comune = trimws(tolower(comune)),
         comune = gsub("-", " ", comune),
         comune = gsub("[[:punct:]]", "", comune),
         backup = comune) %>%
  rename(old_name = comune,
         province_firm = sigla) #%>%
# left_join(soppress, by="old_name") %>%
# mutate(old_name = ifelse(!is.na(new_name), new_name, old_name))

comuni_coord <- comuni_coord %>%
  left_join(cod_regioni %>% select(1:4), by = c("old_name", "province_firm")) %>%
  distinct(province_firm, comune, lat, lon, old_name, cod_reg, den_reg)

soppress <- read_excel("Comuni/Elenco_comuni_soppressi.xls") %>% select(5,9) %>%
  rename(old_name = "Denominazione Comune",
         new_name = "Denominazione Comune associato alla variazione")
soppress <- soppress %>%
  mutate(old_name = iconv(old_name, from = "UTF-8", to = "ASCII//TRANSLIT"),
         new_name = iconv(new_name, from = "UTF-8", to = "ASCII//TRANSLIT")) %>%
  mutate(old_name = trimws(tolower(old_name)),
         new_name = trimws(tolower(new_name)),
         old_name = gsub("-", " ", old_name),
         new_name = gsub("-", " ", new_name),
         old_name = gsub("[[:punct:]]", "", old_name),
         new_name = gsub("[[:punct:]]", "", new_name))
soppress = soppress %>% filter(!old_name %in% comuni_coord$old_name)
```

A second part of the cleaning process loads the coordinates associated to each Italian municipality, cleaning and uniforming characters and converting all characters to lower case, then it addresses the suppressed firms and their new name treating the cases where municipalities cannot be found in the dataset.

```
comuni_panel <- panel %>% select(-firm_taxcode) %>%
  distinct(municipality_firm, region_firm) %>%
  filter(!is.na(municipality_firm)) %>%
  mutate(cleaned = iconv(municipality_firm, from = "UTF-8", to = "ASCII//TRANSLIT"),
         cleaned = trimws(tolower(cleaned)),
         cleaned = gsub("-", " ", cleaned),
         cleaned = gsub("[[:punct:]]", "", cleaned)) %>%
  rename(old_name = cleaned,
         cod_reg = region_firm) %>%
  left_join(soppress, by="old_name") %>%
  mutate(old_name = ifelse(!is.na(new_name), new_name, old_name)) %>%
  distinct(municipality_firm, cod_reg, old_name) %>% arrange(municipality_firm)

no_match <- comuni_panel[!comuni_panel$old_name %in% comuni_coord$old_name, , drop = FALSE]
    %>%
  filter(!is.na(municipality_firm))
newmatch <- read_excel("Comuni/nomatch.xlsx")
no_match <- no_match %>%
  left_join(newmatch, by="old_name") %>%
  mutate(old_name = iconv(old_name, to = "ASCII//TRANSLIT"),
         old_name = trimws(tolower(old_name)),
         old_name = gsub("-", " ", old_name),
         old_name = gsub("[[:punct:]]", "", old_name))

comuni_panel <- comuni_panel %>%
  left_join(no_match[c(2,3,4)], by = c("old_name", "cod_reg")) %>%
  mutate(old_name = ifelse(!is.na(new_name) & new_name != "ESTERO", new_name, old_name)) %>%
  select(-new_name) %>%
  distinct(municipality_firm, cod_reg, old_name) %>%
  mutate(cod_reg = as.integer(cod_reg)) %>%
  left_join(comuni_coord %>% select(old_name, cod_reg, den_reg, province_firm, lat, lon), by =
     c("old_name", "cod_reg")) %>%
  distinct(municipality_firm, cod_reg, den_reg, old_name, lat, lon) %>%
  left_join(comuni_coord %>% select(old_name, cod_reg, province_firm, lat, lon), by = c("old_
    name")) %>%
  mutate(lat.x = ifelse(is.na(cod_reg.y), lat.y, lat.x),
         lon.x = ifelse(is.na(cod_reg.y), lon.y, lon.x)) %>%
  rename(lat = lat.x,
         lon = lon.x,
         cod_reg = cod_reg.x) %>%
  select(municipality_firm, cod_reg, den_reg, province_firm, old_name, lat, lon) %>%
  distinct(municipality_firm, province_firm, cod_reg, den_reg, old_name, lat, lon)
```

Municipalities in the panel dataset is then retrieved and cleaned using the same methodologies as the coordinate dataset. Finally the municipalities are joined using as key the municipality name and region code, accounting for municipalities with same

name but in different regions.

Latitude and longitude of firms in foreign cities are retrieved using the `OpenCage` library which performs API calls to retrieve the data when municipality name is fed into the function:

```
no_match <- no_match %>% mutate(new_name = ifelse(is.na(new_name), "ESTERO", new_name))
estero <- no_match %>% filter(new_name == "ESTERO")
geocode_location <- function(old_name) {
  geocoded <- oc_forward_df(old_name)
  if (!is.null(geocoded) && nrow(geocoded) > 0) {
    list(latitude = geocoded$oc_lat, longitude = geocoded$oc_lng)
  } else {
    list(latitude = NA, longitude = NA)  # Return NA values if geocoding fails
  }
}
geocoded_results <- map(estero$old_name, geocode_location)
geocoded_data <- do.call(rbind, geocoded_results)%>%
  as.data.frame() %>%
  mutate(latitude = as.numeric(latitude),
         longitude = as.numeric(longitude))
estero <- bind_cols(estero, as.data.frame(geocoded_data))
```

The last part joins the dataset with foreign city coordinates to the panel dataset achieving the final dataset with each firm having latitude and longitude values.

```
comuni_panel <- comuni_panel %>%
  left_join(estero, by = "municipality_firm")%>%
  rename(old_name = old_name.x,
         cod_reg = cod_reg.x) %>%
  mutate(
    lat = coalesce(latitude, lat),
    lon = coalesce(longitude, lon),
    province_firm = case_when(
      new_name == "ESTERO" ~ "estero",
      TRUE ~ province_firm
    ),
    cod_reg = case_when(
      new_name == "ESTERO" ~ "estero",
      TRUE ~ as.character(cod_reg)
    )
  ) %>%
  select(-latitude, -longitude, -cod_reg.y,-old_name.y, -new_name) %>%
  distinct(municipality_firm, lat, lon, .keep_all = TRUE) %>%
  filter(!is.na(lat) & !is.na(lon)) %>%
  group_by(municipality_firm) %>%
```

```
  filter(!duplicated(municipality_firm) | row_number() == 1)

panel <- panel %>%
  left_join(comuni_panel, by = "municipality_firm")%>%
  select(-c("province_firm.x", "region_firm","municipality_firm")) %>%
  rename(municipality_firm = old_name,
         province_firm = province_firm.y,
         region_firm = den_reg)
```

3.2.1 Localized Density computation

Localized Density metric is computed with the function `calculate_LD_for_year`, which use as input the dataframe and reference year. The function processes columns representing year, longitude, latitude, network and firm identifiers.

```
calculate_LD_for_year <- function(df, target_year) {
  # Define the distance matrix calculation function
  calculate_distance_matrix <- function(df) {
    coords <- df[, c("lon", "lat")]
    dist_matrix <- as.matrix(distm(coords, fun = distVincentySphere) / 1000) # Convert to km
    diag(dist_matrix) <- NA # Set the diagonal to NA
    return(dist_matrix)}

    # Define the LD calculation function
  calculate_LD <- function(dist_matrix) {
    LD_values <- apply(dist_matrix, 1, function(distance) {
      sum(1/(1 + distance), na.rm = TRUE)
    })
    return(LD_values)}

  # Group the dataframe, calculate distances, and compute LD
  df %>%
    filter(year %in% target_year) %>%
    group_by(network_name, year) %>%
    do({
      dist_matrix <- calculate_distance_matrix(.)
      LD_values <- calculate_LD(dist_matrix)
      data.frame(., LD = LD_values)
    }) %>%
    ungroup() %>%
    group_by(network_name, year) %>%
    mutate(network_members = n()) %>%
    ungroup()
```

First step of the function is the creation of a distance matrix, defined with a function `calculate_distance_matrix` which calculates the pairwise distances between points represented by geographic coordinates latitude and longitude. Distances are calculated using the Vicenty formula and converted to kilometers.

Another internal function `calculate_LD` calculates the Localized Density values. For each row in the distance matrix created by the previous function, it computes the sum of 1/(1+distance) for all distances. '1+distance' ensures a minimal distance value is present when two firms are resided in the same municipality, avoiding to have 0 in the denominator position.

The main function filters the data frame for target year, and performs the previous functions on each network, creating a dataframe with original data and the calculated LD values. In the process number of network participants are also calculated.

The following code calls the function specifying the dataframe and target years and also computes the total number of networks for each year and the network age by the difference between target year and act date:

```
panel <- calculate_LD_for_year(panel, c(2016:2023))%>%
  group_by(firm_taxcode, year) %>%
  mutate(number_of_networks = n()) %>%
  ungroup() %>%
  mutate(network_age = year - act_date)
```

3.2.2 Centrality measures calculation

To calculate the centrality measures two different functions are created, `calculate_centrality` and `calculate_star_centrality`. For both function library `igraph` is used.

Steps for the computation of closeness centrality in the complete network configuration are the following:

1. Graph construction: it creates a fully connected graph ('graph.full') with a number of nodes equal to the number of rows in the input data frame (the function is called for each network by year, so the input for the function are the firms for each network by year).

2. Geodesic Distance Calculation: it computes the pairwise geodesic distance between

nodes based on their geographic coordinates (longitude, latitude). The distances are converted from meters to kilometers.

3. Distance adjustments: to handle cases where the distance is 0 (when two nodes/-firms are form the same municipality) the function sets these values to 1, and the the diagonal of the matrix is set to 0 as the diagonal represents the distance of a node to itself.

4. Weight assignment: the distance matrix constructed by previous points are attributed to the edges in the graph, representing the distances between nodes.

5. Closeness Centality calculation: centrality measure is calculated using the `closeness` function from `igraph` with the constructed graphs and specifying the weights.

The function returns the dataframe with each firm taxcode and its corresponding closeness centrality.

```
calculate_complete_centrality <- function(data) {

  g <- graph.full(nrow(data))

  num_edges <- ecount(g)
  max_edges <- nrow(data) * (nrow(data) - 1) / 2

  # Calculate the geodesic distance between all pairs of nodes (firms)
  distances <- distm(data[, c("lon", "lat")], fun = distGeo)/1000
  # Replace 0 distances (self loops) with a small positive number, here set to 1
  distances[distances == 0] <- 1
  diag(distances) <- 0

  E(g)$weight <- distances[lower.tri(distances)]
  # Calculate closeness centrality for all nodes
  closeness_centrality <- closeness(g, weights = E(g)$weight, normalized = FALSE)
  }
  result <- data.frame(firm_taxcode = data$firm_taxcode,
                       closeness_centrality = closeness_centrality,
)
  return(result)
}
```

Constructing the star configuration graph is a more complicated process, as it requires specifying each star node and creating edges that connect all other nodes exclusively to

these star nodes, the steps are the following:

1. Graph construction: an empty graph is first created with number of nodes equal to the number of rows in the input ('graph.empty').

2. Star nodes identification: each firm marked as reference company (reference_company == 1) in the dataset are treated as central nodes in a star configuraiton.

3. Handling networks with no reference company: if no reference companies are found an empty dataframe is returned with firm identification and NA for closeness centrality.

4. Create Edges: edges are created between each star node and all other nodes. 'setdiff' is used to differentiate non-star nodes and constructs a matrix of edges where each star node is connected to each non-star node. The edges are then added to the graph using 'add_edges' function.

5. Connect Star Nodes: in case more than one star node is present, they are connected to each other.

6. Closeness Centrality computation: centrality is calculated using 'closeness' function.

The output is a dataframe containing each firm identification, reference_company dummy and the closeness centrality value.

```
calculate_star_centrality <- function(data) {
  # Create an empty graph
  g <- graph.empty(n = nrow(data), directed = FALSE)

  # Identify the reference company nodes
  star_nodes <- which(data$reference_company == 1)

  # Check if there is at least one reference company identified
  if(length(star_nodes) == 0) {
    return(data.frame(firm_taxcode = data$firm_taxcode,
                      reference_company = data$reference_company,
                      closeness_centrality = NA))
  }

  # Create edge pairs between each reference company and all other nodes
  other_nodes <- setdiff(1:nrow(data), star_nodes)
```

```
  edges <- c(rbind(rep(star_nodes, each=length(other_nodes)), other_nodes))

  # Add edges to the graph
  g <- add_edges(g, edges)

  # If there is more than one star node, connect them as well
  if(length(star_nodes) > 1) {
    star_edges <- t(combn(star_nodes, 2))  # Create unique combinations of star nodes
    g <- add_edges(g, t(star_edges))  # Add star edges to the graph
  }

  # Calculate closeness centrality for each node
  closeness_centrality <- closeness(g, mode = "all", normalized = FALSE)

  # Combine results into a data frame
  result <- data.frame(firm_taxcode = data$firm_taxcode,
                       reference_company = data$reference_company,
                       closeness_centrality = closeness_centrality)

  return(result)
}
```

Both above functions are called using similar pipeline:

```
panel <- panel %>%
  group_by(network_name, year) %>%
  do(calculate_centrality(.))

panel <- panel %>%
  group_by(network_name, year) %>%
  do(calculate_star_centrality(.))
```

Grouping by `network_name` and `year` allows for the computation of centrality metrics for each network annually.

3.3 Additional Features

Additional features are obtained from AIDA dataset, which are computed after joining the AIDA dataset and Network Contract panel dataset, and manipulating already present columns.

The process includes the calculation of `FirmGrowth`, `AverageGrowth`, `FirmAge`, `RegionGroup`,

`Sector` etc.

For the `FirmGrowth` and `LabProd` calculation a simple dyplr pipeline is created, first by arranging the dataset by firm_taxcode and year, then grouping by each firm_taxcode both metrics are calculated, assigning NA to all infinite and NaN values:

```
AIDA <- AIDA %>% arrange(firm_taxcode, year) %>%
  group_by(firm_taxcode) %>%
  mutate(FirmGrowth = log(Sales_Revenue_million_EUR) - log(lag(Sales_Revenue_million_EUR)),
         LabProd = Sales_Revenue_million_EUR/n_employee) %>%
  mutate(across(c(FirmGrowth, LabProd), ~replace(., . == Inf | . == -Inf | is.nan(.), NA)))
```

Variable `AverageGrowth` is calculated as the mean of non-missing FirmGrowth from year of entrance in network to last year in network.

```
 final <- final %>%
   group_by(firm_taxcode, network_name) %>%
   mutate(Avg_growth = mean(log_diff_Sales_Revenue, na.rm = TRUE)) %>%
   mutate(Avg_growth = ifelse(is.nan(Avg_growth) | is.infinite(Avg_growth), NA, Avg_growth))
```

`AverageFirmSize` is created by calculating the average number of employees (`mean_empl`) for each firm across all years considered, assuming the firm size do not have sudden changes during the years, a firm is assigned as `Micro` when the `mean_empl` is less or equal than 10, when it is between 10 and 250 (included) it is assigned as `SMEs` and `Large` when it has more than 250 employees.

```
AIDA <- AIDA %>%
  group_by(firm_taxcode) %>%
  mutate(mean_empl = mean(n_employee, na.rm = TRUE)) %>% ungroup() %>%
  mutate(AverageFirmSize = ifelse(mean_empl <= 10, "Micro",
                ifelse(mean_empl > 10 & mean_empl <= 250, "SMEs", "Large"))) %>%
  select(-mean_empl)
```

The variable `FirmAge` is calculated as the difference between the current year and the founding year of the firm. However, due to instances where firms go through restructuring or re-registration leading to negative `FirmAge` values. Since the AIDA portal lacks of a historical record of firm founding dates, it is not possible to retrieve the original founding year of these cases, consequently in such situations FirmAge is assigned to NA.

```
AIDA <- AIDA %>%
  mutate(FirmAge = year - year_founding)%>%
  mutate(FirmAge = ifelse(FirmAge < 0, NA, FirmAge))
```

`RegionGroup` column is created by assigning "Piemonte", "Liguria", "Lombardia", "Valle d'Aosta" to the North-West region group, "Veneto", "Trentino-Alto Adige", "Friuli-Venezia Giulia", "Emilia-Romagna" to the North-East region group, "Toscana", "Umbria", "Marche", "Lazio" to the Center, "Abruzzo", "Molise", "Campania", "Puglia", "Basilicata", "Calabria" to the South, "Sicilia", "Sardegna" to Islands and any other case as "Abroad".

```
DF <- DF %>%
  mutate(region_group = case_when(
    region_firm %in% c("Piemonte", "Liguria", "Lombardia", "Valle d'Aosta") ~ "North-West",
    region_firm %in% c("Veneto", "Trentino-Alto Adige", "Friuli-Venezia Giulia", "Emilia-
    Romagna") ~ "North-East",
    region_firm %in% c("Toscana", "Umbria", "Marche", "Lazio") ~ "Center",
    region_firm %in% c("Abruzzo", "Molise", "Campania", "Puglia", "Basilicata", "Calabria") ~
    "South",
    region_firm %in% c("Sicilia", "Sardegna") ~ "Islands", TRUE ~ "Abroad"))
```

The `Sector` variable is retrieved by combining ATECO codes from ISTAT and the equivalent ISIC category (International Standard Industrial Classification). The ISTAT ATECO-Sector classification is structured as in the table 3.3, for each category a letter represents the ISIC classification (A, B, C etc.), followed the ATECO code and its specifications.

Table 3.3: ISTAT Classification of Economic Sectors

Code	Description
A	**AGRICOLTURA, SILVICOLTURA E PESCA**
01	COLTIVAZIONI AGRICOLE E PRODUZIONE DI PRODOTTI ANIMALI, CACCIA E SERVIZI CONNESSI
011	COLTIVAZIONE DI COLTURE AGRICOLE NON PERMANENTI
0111	Coltivazione di cereali (escluso il riso), legumi da granella e semi oleosi
...	
B	**ESTRAZIONE DI MINERALI DA CAVE E MINIERE**
05	ESTRAZIONE DI CARBONE (ESCLUSA TORBA)
051	ESTRAZIONE DI ANTRACITE
...	

The table is processed by grouping detailed classification entries under their respective

ISIC categoization, ensuring each entry has a consistent sector identifier.

```
settore_ateco <- settore_ateco %>%
  mutate(SectorCategory = ifelse(nchar(ateco_2007) == 1, ateco_2007, NA)) %>%
  fill(SectorCategory, .direction = "down") %>%
  # Create a new column to hold the settore corresponding to the letter
  mutate(SettoreLetter = ifelse(!is.na(SectorCategory) & nchar(ateco_2007) == 1, settore, NA))
     %>%
  fill(SettoreLetter, .direction = "down") %>%
  # Remove rows where ateco_2007 is a single letter as we only want the subsequent rows
  filter(nchar(ateco_2007) > 1) %>%
  rename(macro_settore = SettoreLetter)
```

Result of this data transformation is represented in figure 3.1.

Figure 3.1: Sector Classification post transformation

ateco_2007	settore	SectorCategory	macro_settore
01	COLTIVAZIONI AGRICOLE E PRODUZIONE DI PRODOTTI ANI...	A	AGRICOLTURA, SILVICOLTURA E PESCA
011	COLTIVAZIONE DI COLTURE AGRICOLE NON PERMANENTI	A	AGRICOLTURA, SILVICOLTURA E PESCA
0111	Coltivazione di cereali (escluso il riso), legumi da granella e s...	A	AGRICOLTURA, SILVICOLTURA E PESCA
01111	Coltivazione di cereali (escluso il riso)	A	AGRICOLTURA, SILVICOLTURA E PESCA
011110	Coltivazione di cereali (escluso il riso)	A	AGRICOLTURA, SILVICOLTURA E PESCA
01112	Coltivazione di semi oleosi	A	AGRICOLTURA, SILVICOLTURA E PESCA
011120	Coltivazione di semi oleosi	A	AGRICOLTURA, SILVICOLTURA E PESCA
01113	Coltivazione di legumi da granella	A	AGRICOLTURA, SILVICOLTURA E PESCA
011130	Coltivazione di legumi da granella	A	AGRICOLTURA, SILVICOLTURA E PESCA
01114	Coltivazioni miste di cereali, legumi da granella e semi oleosi	A	AGRICOLTURA, SILVICOLTURA E PESCA
011140	Coltivazioni miste di cereali, legumi da granella e semi oleosi	A	AGRICOLTURA, SILVICOLTURA E PESCA

Each ISIC classification code is then assigned to firms based on its ATECO code, additional aggregation is performed assigning all sectors below 2% to the category `Other services`.

```
DF <- DF %>%
  mutate(Group = case_when(
    SectorCategory %in% c("A") ~ "Agriculture, forestry, and fishery",
    SectorCategory %in% c("B", "F") ~ "Construction and Mining",
    SectorCategory %in% c("C") ~ "Manufacturing",
    SectorCategory %in% c("G", "H", "I", "J", "L", "N") ~ "Services excluding finance",
    SectorCategory %in% c("M") ~ "Professional and scientific services",
    SectorCategory %in% c("O", "P", "Q") ~ "Public, health, and education",
    SectorCategory %in% c("D", "E","K","R","S") ~ "Other services",
    TRUE ~ NA # Default case to keep the original Sector name if no match is found
  ))
```

The models to be employed will analyze firm performance at the firm level. Since the dataset retains a firm-network configuration, firms participating in multiple networks appear multiple times. Consequently, network measures will be aggregated to the firm level.

```
df_connections <- DF %>%
  select(firm_taxcode, network_name, year) #%>%  distinct()
setDT(df_connections)
# Create a new column 'firms_in_network' that lists all firms in the same network for each
    year
df_connections[, firms_in_network := list(list(firm_taxcode)), by = .(network_name, year)]
df_connections <- df_connections[, .(unique_firms_list = list(unique(unlist(firms_in_network))
    )), by = .(firm_taxcode, year)]
df_connections[, connections_count := lapply(unique_firms_list, length), by = .(firm_taxcode,
    year)]
panel <- panel %>% left_join(df_connections %>% select(1,2,4), by = c("firm_taxcode", "year"))
panel <- panel %>% group_by(firm_taxcode, year) %>% mutate(avg_network_age = mean(network_age)
    )
rm(df_connections)
panel <- panel %>% mutate(connections_count = connections_count - 1) %>% rename(
    NetworkedFirmsCount = connections_count)

panel <- panel %>%
  group_by(firm_taxcode, year) %>%
  mutate(across(c(LD, LD_perc), ~weighted.mean(.x, network_age, na.rm = TRUE)),
         across(c(closeness_centrality_complete, closeness_centrality_star),
                ~weighted.mean(.x, network_age * network_members, na.rm = TRUE))) %>%
  ungroup()
```

The code creates a list of all firms in the same network for each year and generates a unique list of firms for each `firm_taxcode` and year.
It counts the number of unique firms each firm is connected to within its network for each year creating firm level variable `NetworkedFirmsCount` it is later adjusted by subtracting one excluding the firm itself, this metric will substitute the network size metric. It also calculates the average age of networks a firm is part of, substituting the network age metric.
Spatial measures are weighted by network age and network size measures.

After the above process, the dataset is transformed to a firm-level dataset, ensuring each firm has only one record per year avoiding duplications within the models. Additionally, firms located abroad are filtered out as they represent only 0.038% of the dataset.

As a result, the number of observations is reduced from 266,437 to 199,703.

3.4 Model Setup

This section explicates the process of econometric models fitting using R, starting from the data loading to variable lagging, Pooled OLS, LAD models creation, extended LAD model analysis and Panel regression models setup.

The libraries utilized are `data.table` and `tidyverse` for data wrangling and manipulation, `plm` for panel data econometric modeling, `sandwich` for robust standard error estimation, `lmtest` for diagnostic testing, `quantreg` for quantile regression analysis and `stargazer` for creating well-formatted tables of regression results.

First step of model creation is the loading of the data, here we load the cleaned data with the variables chosen for modeling and convert the categorical variable to factor so that dummies of these variables are automatically generated when performing regressions.

```
final <- fread("dati_modelli_finali.csv", colClasses = list(character = c("firm_taxcode")), na
    .strings = c("NA",""))

final <- final %>%
  mutate(
    year = as.factor(year),
    Sector = factor(Sector),
    RegionGroup = as.factor(RegionGroup),
    LegalNetwork = as.factor(LegalNetwork),
    hub = as.factor(hub),
    AverageFirmSize = as.factor(AverageFirmSize),
    InnovativeSME = as.factor(InnovativeSME),
    InnovativeStartup = as.factor(InnovativeStartup),
  )%>%
  mutate(Sector = relevel(Sector, ref = "Manufacturing"),
         RegionGroup = relevel(RegionGroup, ref = "North-East"))
```

The dataset is loaded ensuring the taxcode numbers are read as characters and missing values are treated correctly. Categorical variables such as year, Sector, RegionGroup, LegalNetwork, hub, AverageFirmSize, InnovativeSME and InnovativeStartup are converted to factor to ensure they are treated correctly during analysis. Manufacturing sector is used as reference level for Sector. North-East region area is used as reference level for the region groups.

To create the regression tables multiple models are fitted by adding progressively the control variables:

```
OLS1 <- plm(FirmGrowth ~ LD + CCS + Sector+ factor(year)+ RegionGroup, data = final, model = "
    pooling")
OLS2 <- plm(FirmGrowth ~ LD + CCS + hub + NetworkedFirmsCount+ AverageNetworkAge +
    LegalNetwork+ Sector+ factor(year)+ RegionGroup, data = final, model = "pooling")
OLS3 <- plm(FirmGrowth ~ LD + CCS + hub + NetworkedFirmsCount + AverageNetworkAge +
    LegalNetwork + I(FirmAge^2) + AverageFirmSize + InnovativeSME+ InnovativeStartup+ Sector+
    factor(year)+ RegionGroup, data = final, model = "pooling")
OLS4 <- plm(FirmGrowth ~ LD + CCS + hub + NetworkedFirmsCount+ AverageNetworkAge+ LegalNetwork
    + I(FirmAge^2)+ AverageFirmSize+ InnovativeSME+ InnovativeStartup+ ROS+ ln_LabProd+ ln_
    LiquidAssets+ ln_IntangibleAssets+ Sector+ factor(year)+ RegionGroup, data = final, model
    = "pooling")

# Calculate robust standard errors
robust_se1 <- vcovHC(OLS1, type = "HC1")
robust_se2 <- vcovHC(OLS2, type = "HC1")
robust_se3 <- vcovHC(OLS3, type = "HC1")
robust_se4 <- vcovHC(OLS4, type = "HC1")

# Apply coeftest to each model
coeftest_OLS1 <- coeftest(OLS1, vcov = robust_se1)
coeftest_OLS2 <- coeftest(OLS2, vcov = robust_se2)
coeftest_OLS3 <- coeftest(OLS3, vcov = robust_se3)
coeftest_OLS4 <- coeftest(OLS4, vcov = robust_se4)

# Extract coefficients and robust standard errors for stargazer
se_OLS1 <- coeftest_OLS1[, 2]
se_OLS2 <- coeftest_OLS2[, 2]
se_OLS3 <- coeftest_OLS3[, 2]
se_OLS4 <- coeftest_OLS4[, 2]
```

After fitting the models, robust standard errors are calculated using `vcovHC()`, which takes in the first position the model and the type input stands for the type of estimation, `'HC1'` is the most commonly used apporach for linear models, applies a degrees of freedom-based correction, $(n1)/(nk)$ where n is the number of observations and k is the number of predictor variables in the model.
All Pooled OLS models are created following the same process by changing the dependent variable.

Tables used in this thesis are created using stargazer function, specifying the `'latex'` and by regulating the spacing using stargazer internal functions:

```
stargazer(OLS1, OLS2, OLS3, OLS4, type = "latex",
          coef = list(coef_OLS1, coef_OLS2, coef_OLS3, coef_OLS4),
          se = list(se_OLS1, se_OLS2, se_OLS3, se_OLS4),
          header = FALSE, column.sep.width = "-70pt", font.size = "normalsize",
          align = TRUE, no.space = TRUE, single.row = TRUE, omit.stat = c("f", "adj.rsq", "ser
    "))
```

LAD regression models are created using the `quantreg` package instead of `plm`. Quantile regression for the median targets the 50th percentinle, effectively minimizing the same absolute differences as LAD. Thus quantile regression function is used.
Similar to the OLS models these models are expanded step-by-step including additional control variables.
The function `quantreg` requires the model formula as input, the data table used, and a tau () parameter specifying the quantile of the dependent variable of interest for modeling, which is 0.5 in this case

```
LAD1 <- rq(FirmGrowth ~ LD +CCS + Sector+ factor(year)+RegionGroup, data = final, tau = 0.5)
LAD2 <- rq(FirmGrowth ~ LD + CCS +hub + NetworkedFirmsCount+AverageNetworkAge+ LegalNetwork+
    Sector+ factor(year)+RegionGroup, data = final, tau = 0.5)
LAD3 <- rq(FirmGrowth ~ LD + CCS + hub +NetworkedFirmsCount+ AverageNetworkAge+ LegalNetwork+
    I(FirmAge^2)+ AverageFirmSize+InnovativeSME+InnovativeStartup+Sector+ factor(year)+
    RegionGroup, data = final, tau = 0.5)
LAD4 <- rq(FirmGrowth ~LD + CCS + hub + NetworkedFirmsCount+ AverageNetworkAge+ LegalNetwork+
    I(FirmAge^2)+ AverageFirmSize+ InnovativeSME+ InnovativeStartup+ ROS+ ln_LabProd+ ln_
    LiquidAssets+ ln_IntangibleAssets+ Sector+ factor(year)+ RegionGroup, data = final, tau =
    0.5)

stargazer(LAD1,LAD2,LAD3,LAD4, type="latex",
          dep.var.labels = "FirmGrowth",header = FALSE, column.sep.width = "-70pt", font.size
    = "normalsize",
          align = TRUE, no.space = TRUE, single.row = TRUE, omit.stat = c("f", "adj.rsq", "ser
    "))
```

Similarly, LAD models with interaction terms are created by adding the interaction term in the regression formula of the Quantile Regression model specification, the following example refers to model (1) in Table 2.8:

```
int1 <- rq(FirmGrowth ~ LD +  CCS + I(LD*CCS) + Sector+ factor(year)+ RegionGroup, data =
    final, tau=0.5, na.action = na.omit)
```

All interaction regression are created in the same way by substituting or including additional variables.

The Time Split analysis table 2.13 is created through splitting the dataset in two parts: 2016-2019 and 2020-2022.

```
final_split1 <- final %>%
  filter(year %in% c(2016:2019)) %>%
  group_by(firm_taxcode) %>%
  mutate(AverageGrowth = mean(FirmGrowth, na.rm = TRUE)) %>%
  mutate(AverageGrowth = if_else(is.nan(AverageGrowth), NA_real_, AverageGrowth))

final_split2 <- final %>% filter(year %in% c(2020:2022)) %>%  group_by(firm_taxcode) %>%
  mutate(AverageGrowth = mean(FirmGrowth, na.rm = TRUE)) %>%
  mutate(AverageGrowth = if_else(is.nan(AverageGrowth), NA_real_, AverageGrowth))
```

The code filters and saves the two time periods separately and recalculates the Average-Growth metric on each period.
Following the splitting the regressions are performed just like previous models but referencing the newly created dataframe.

The panel regression requires to handle the unbalanced to perform the analysis on firms with similar exposure to network contracts. Therefore, the first step is to filter and save only the firms with atleast 3 years in networks:

```
filtered_data <- final %>%
  group_by(firm_taxcode) %>%
  arrange(firm_taxcode, year) %>%
  mutate(year = as.numeric(as.character(year))) %>%
  mutate(entrance = min(year)) %>%
  mutate(last = max(year)) %>%
  ungroup() %>%
  mutate(years_in_network = last-entrance) %>%
  filter(years_in_network >= 3)
```

The code calculates the entrance year in networks for each firm (year of appearance in the dataset). Subsequently the last year in the network is saved, if there are atleast 3 years of difference between the last year and the entrance year the firms is saved, every other firm is filtered out.

To leverage the panel structure of the data and apply within effects using the `'plm'`

package there is the necessity of creating a panel data frame. Using `pdata.frame()` function the filtered dataset effectively gets converted in panel structure.

```
pdata <- pdata.frame(filtered_data, index = c("firm_taxcode", "year"))
```

In thee case of panel regression instead of four specification for each model, only three are created, as time-invariant firm-specific characteristics are absorbed and cancelled out by the entity fixed effect.

```
FE1 <- plm(FirmGrowth ~ LD + CCS + Sector+ factor(year)+ RegionGroup, data = pdata, model = "
    within")
FE2 <- plm(FirmGrowth ~ LD + CCS + hub + NetworkedFirmsCount+  AverageNetworkAge+ LegalNetwork
    + Sector+ factor(year)+ RegionGroup, data = pdata, model = "within")
FE3 <- plm(FirmGrowth ~ LD + CCS+ hub + NetworkedFirmsCount+ AverageNetworkAge+ LegalNetwork+
    I(FirmAge^2)+ ROS+ ln_LabProd+ ln_LiquidAssets+ ln_IntangibleAssets+ Sector+ factor(year)
    + RegionGroup, data = pdata, model = "within")

# Compute robust standard errors
robust_se_FE1 <- vcovHC(FE1, type = "HC1", cluster = "group")
robust_se_FE2 <- vcovHC(FE2, type = "HC1", cluster = "group")
robust_se_FE3 <- vcovHC(FE3, type = "HC1", cluster = "group")

# Get coefficients and clustered standard errors
coef_se_FE1 <- coeftest(FE1, vcov = robust_se_FE1)
coef_se_FE2 <- coeftest(FE2, vcov = robust_se_FE2)
coef_se_FE3 <- coeftest(FE3, vcov = robust_se_FE3)
```

Just like Pooled OLS, robust standard error are calculated using `vcovHC` function, with the exception that standard errors are clustered at firm level.
The tables are then created using `stargazer()` function.

3.5 Summary of Methodology

This chapter presented a detailed methodology for the analysis of the impact of spatial proximity measures and network contract effect on economic performance of participating firms.
The approach includes a initial data cleaning process adressing naming convention, missing values and duplication issues, ensuring consistency across the dataset.

Spatial measures, including Localized Density and Closeness Centrality are then computed using the cleaned dataset, key measures used in this study of the spatial proximity's impact on firm perforamnce.

The construction of econometric models required some data wrangling by handling unbalanced panel data, creation of interaction terms and application of robust standard error techniques.

The results of the final models will be discussed in the next chapter. The methodology outlined in this chapter provides a clear framework for future replication and further application of this research.

Chapter 4

Conclusion

This thesis has undertaken a thorough exploration of the factors influencing firm growth, employing a robust analytical framework across various econometric models. The analysis is centered around the usage of spatial proximity measure to proxy the benefits of geographic vicinity, such as more frequent interactions, better coordination and exchange of knowledge, while controlling for network, and firm characteristics as well as time, sector, region group fixed effects.

The analysis was conducted on network and firm economic data from 2016 to 2022. By lagging the variables, the time interval was reduced from 2017 to 2022. Lagging the variables helped address potential endogeneity issues and ensured that the independent variables were not contemporaneously correlated with the error term, thus providing more reliable estimates of the causal impact of network and spatial characteristics on firm performance.

Initially, an Ordinary Least Squares (OLS) is performed as baseline. However, the results from the Pooled OLS were not robust as they were sensitive to outliers and the method did not account for the panel structure of the data. To address these limitations, a Least Absolute Deviations (LAD) regression is computed to provide more robust results to outliers. Successively, the complex dynamics of spatial proximity were explored using LAD regression with interaction terms and by splitting time periods. Finally, a Fixed Effect model was employed to leverage the panel structure and control for unobserved heterogeneity.

Table 4.1 presents the results of the Pooled OLS, LAD, LAD with interaction effects, and fixed effects (FE) models in columns (1), (2), (3), (4). The table provides the models

including all control variables conducted in previous analysis.

Table 4.1: FirmGrowth - Regression Models

	Dependent variable:			
	FirmGrowth			
	Pooled OLS	LAD	LAD interaction	FE
	(1)	(2)	(3)	(4)
$LocalizedDensity_{t-1}$	0.001** (0.0004)	−0.0002 (0.0002)	−0.00002 (0.0002)	−0.003 (0.003)
$ClosenessCentrality_{t-1}$	0.025*** (0.010)	0.010** (0.005)	0.012** (0.005)	0.026 (0.044)
$LD_{t-1} * CC_{t-1}$			−0.011*** (0.004)	
Hub: 1	−0.007 (0.006)	0.0003 (0.002)	0.002 (0.003)	−0.013 (0.030)
$NetworkedFirmsCount_{t-1}$	0.0001 (0.0002)	0.0001 (0.0001)	0.0001 (0.0001)	−0.0002 (0.001)
$AverageNetworkAge_{t-1}$	−0.005*** (0.001)	−0.003*** (0.001)	−0.003*** (0.001)	0.001 (0.005)
LegalNetwork: 1	0.001 (0.007)	0.006** (0.003)	0.006** (0.003)	0.010 (0.014)
$FirmAge^2_{t-1}$	−0.00000 (0.00000)	−0.00000*** (0.00000)	−0.00000*** (0.00000)	0.0003*** (0.0001)
AverageFirmSize: Micro	0.017 (0.012)	0.014*** (0.004)	0.014*** (0.005)	
AverageFirmSize: SME	0.030*** (0.009)	0.020*** (0.003)	0.019*** (0.003)	
InnovativeSME: 1	0.043** (0.017)	0.025*** (0.009)	0.024*** (0.009)	
InnovativeStartup: 1	0.263*** (0.091)	0.229** (0.115)	0.231** (0.117)	
ROS_{t-1}	0.004*** (0.0005)	0.001*** (0.0001)	0.001*** (0.0001)	0.001** (0.001)
$\ln(LabProd_{t-1})$	−0.056*** (0.004)	−0.014*** (0.001)	−0.014*** (0.001)	−0.485*** (0.018)
$\ln(LiquidAssets_{t-1})$	0.010*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.001 (0.003)
$\ln(IntangibleAssets_{t-1})$	0.006*** (0.001)	0.002*** (0.0004)	0.002*** (0.0004)	0.013*** (0.003)
Constant	0.202*** (0.022)	0.081*** (0.007)	0.082*** (0.007)	
Observations	48,443	48,443	48,443	43,387
N. unique firms	13,998	13,998	13,998	10,445
R^2	0.071			0.196

Note: For OLS and FE models robust standard errors are given in parentheses. All regressions also include year, sector and region area fixed-effects. Asterisks denote significance levels: *p<0.1; **p<0.05; ***p<0.01

Key findings of this study are:

- **Localized Density and Closeness Centrality:** $LocalizedDensity_{t-1}$ is positive and significant (p<0.05) in the Pooled OLS model but becomes negative and insignificant when controlling for outlier in the LAD regression model and when accounting for firm specific effects. The inconsistency suggests there might be more complex interactions within the network framework. By considering the interaction effect with centrality ($LD_{t-1} * CC_{t-1}$) the coefficient of Localized Density reach almost 0 and shows a significant (p<0.01) and negative effect (-0.011) of the interaction term on firm growth.

  $ClosenessCentrality_{t-1}$ shows consistent moderate significant effect (p<0.05) on firm growth when minimizing the influence of outliers. The coefficient increases when considering the interaction effect with localized density.

- **Network-Level Control Variables:** Being a hub does not significantly impact firm growth across models. $NetworkedFirmsCount_{t-1}$ consistently demonstrates a

positive but marginal and non-significant influence, indicating that simply being in a network with many firms does not guarantee growth benefits. $AverageNetworkAge_{t-1}$ shows a consistent negative and highly significant effect on firm growth, suggesting that older networks might be less effective in fostering firm growth, possibly due to rigidity or diminished innovative capacity over time. Being part of a legally formalized network shows a positive and significant impact on firm growth in the LAD model, highlighting the importance of formalized network structures for central firms.

- **Firm-Level Control Variables:** The negative impact of squared firm age is significant in the LAD regression, suggesting that younger firms benefit more from network participation. Positive and significant coefficients for AverageFirmSize: Micro and AverageFirmSize: SME in the LAD model highlight substantial growth benefits for smaller firms within networks. Innovative SMEs and Startups show significantly higher growth rates, emphasizing the importance of innovation capacity and the potential enhancement of innovation through network participation. Significant positive impacts of ROS and intangible assets, while labor productivity has a negative impact across all models, possibly indicating over-employment or inefficiencies in labor utilization.

In the Fixed Effects model, both Localized Density and Closeness Centrality, as well as all network-related variables, become non-significant. This absorption of significance is likely due to the model controlling for unobserved firm-specific characteristics, such as interaction frequency, managerial capacity, and coordination capacity. It suggests that the effects of these variables are differenced out by the fixed effects, indicating that network composition and firm positioning do not vary significantly within firms over time as shown in table 2.15.

The time splitting analysis 2.13, which examined the impact of variables over different periods, reveals that the effects of network characteristic variables are not static over time. During certain periods, the influence of localized density and centrality varies, indicating that temporal factors play a role in determining the effectiveness of network participation on firm growth. This temporal variability highlights the need for dynamic strategies that adapt to changing economic conditions and network dynamics.

In conclusion, Localized Density exhibits negative but non-significant effect on firm

growth in more robust models (LAD and FE). Conversely, firms with higher centrality within their networks tend to experience better growth outcomes, as centrality facilitates access to information and resources, enhancing the firm's strategic positioning. The significant negative interaction effect between localized density and closeness centrality implies that the benefits of geographic proximity depend on both physical positioning and organizational position within the network. Firms that are both geographically proximate and central in their networks might experience diminishing returns due to overcrowding and coordination difficulties.

Overall, this thesis contributes to the understanding of the role of spatial proximity and network effects in influencing firm growth in the context of Italian Network Contracts. The findings highlight the complexity of network dynamics and the importance of considering both spatial and organizational dimensions in designing policies and strategies to optimize network benefits for firm growth.

Future research could explore the effects of network contracts and spatial proximity by integrating network diversity, different sectors may have unique characteristics and dynamics that influence how network participation affects firm growth. Extending the time frame of the data would provide a better understanding of the long-term effects of network and spatial proximity, capturing trends and evolving network interactions over time. Addressing these aspects can offer a deeper understanding of the complex dynamics of spatial proximity and network effects on firm growth.

References

[1] Abdesslem, A. B., and Chiappini, R. Cluster policy and firm performance: a case study of the french optic/photonic industry. *Regional Studies 53*, 5 (2019), 692–705. https://econpapers.repec.org/RePEc:taf:regstd:v:53:y:2019:i:5:p:692-705.

[2] Becattini, G. The marshallian industrial district as a socio-economic notion. *Industrial Districts and Inter-Firm Cooperation in Italy 1*, 1 (1990), 37–51.

[3] Becattini, G. From industrial districts to local development: An itinerary of research. *Handbook of Industrial Districts 1*, 1 (2009), 3–17.

[4] Duranton, G., and Overman, H. G. Testing for localization using micro-geographic data. *The Review of Economic Studies 72*, 4 (2005), 1077–1106.

[5] Hothorn, T., and Zeileis, A. *lmtest Package: coeftest Function Documentation*, 2023. https://www.rdocumentation.org/packages/lmtest/versions/0.9-40/topics/coeftest (Accessed: 12 June 2024).

[6] Infocamere. Contratti di rete, n.d. https://contrattidirete.registroimprese.it/reti/ (Accessed: 11 May 2024).

[7] Informatica, G. Database comuni italiani, n.d. https://www.gardainformatica.it/database-comuni-italiani (Accessed: 28 Mar 2024).

[8] ISTAT. Dcumentazione tecnica e classificazioni, classificazione ateco 2007 aggiornamento 2022, 2022. https://www.istat.it/it/archivio/266993 (Accessed: 23 Mar 2024).

[9] ISTAT. Codici statistici delle unitÀ amministrative territoriali: Comuni, cittÀ metropolitane, province e regioni, 2024. https://www.istat.it/it/archivio/6789 (Accessed: 23 Mar 2024).

[10] Koenker, R. quantreg: Quantile regression r package, 2023. https://cran.r-project.org/web/packages/quantreg/quantreg.pdf (Accessed: 12 June 2024).

[11] Lee, S. Y., and Arai, K. I. Fixed effects models, 2023. https://theeffectbook.net/ch-FixedEffects.html (Accessed: 26 May 2024).

[12] Leoncini, R., Vecchiato, G., and Zamparini, L. Triggering cooperation among firms: an empirical assessment of the italian network contract law. *Economia Politica 37*, 2 (2020), 357–380. https://doi.org/10.1007/s40888-019-00141-z.

[13] Millo, G., and Croissant, Y. Panel data econometrics in r: The plm package, n.d. https://cran.r-project.org/web/packages/plm/vignettes/A_plmPackage.html (Accessed: 26 May 2024).

[14] Nguyen, N. Approximate string matching in r using jaro-winkler similarity, 2023. https://blog.devgenius.io/approximate-string-matching-in-r-using-jaro-winkler-similarity-a93436ecf38f (Accessed: 12 June 2024).

[15] Oerlemans, L. A., and Meeus, M. T. Do organizational and spatial proximity impact on firm performance? *Regional Studies 39*, 1 (2005), 89–104. https://econpapers.repec.org/RePEc:taf:regstd:v:39:y:2005:i:1:p:89-104.

[16] Rubino, M., and Vitolla, F. Implications of network structure on small firms performance: Evidence from italy. *International Journal of Business and Management 13*, 4 (2018), 46–53. https://doi.org/10.5539/ijbm.v13n4p46.

[17] Rubino, M., and Vitolla, F. The impact of formal networking on the performance of smes. *International Journal of Business and Management 14*, 4 (2019), 112–125. https://doi.org/10.5539/ijbm.v14n4p112.

[18] SimpleMaps. Italy cities database, n.d. https://simplemaps.com/data/it-cities (Accessed: 30 May 2024).

[19] Stulp, F., and Sigaud, O. Many regression algorithms, one unified model: A review. *Neural Networks 69* (2015), 60–79. https://www.sciencedirect.com/science/article/pii/S089360801500117X.

[20] Torres-Reyna, O. Panel data analysis using r, 2010. https://swh.princeton.edu/~otorres/Panel101R.pdf (Accessed: 12 June 2024).

[21] Zeileis, A., and Hothorn, T. *Sandwich Package: vcovHC Function Documentation*, 2023. https://www.rdocumentation.org/packages/sandwich/versions/3.1-0/topics/vcovHC (Accessed: 12 June 2024).