doi.org/10.26398/asaproc.00142

Geography of business alliances and spatial network complexity: The backbone of formal collaboration agreements

Andrea Enrico Vurro[a], Alessio Bumbea[a], Annamaria Giuffrida[a], Andrea Mazzitelli[b], Giuseppe Espa[c]

[a] Department of Engineering and Sciences, Universitas Mercatorum, Rome, Italy.
[b] Department of Economics, Statistics, and Business, Universitas Mercatorum, Rome, Italy.
[c] Department of Economics and Management, University of Trento, Italy.

1. Introduction

To enhance the competitiveness of small businesses and foster enduring network relationships, the Italian government introduced in 2009 the *network contract*, a legal tool that governs and ensures strategic collaboration among firms (Gulati et al., 2000; Dickson et al., 2021).

A formal collaboration agreement (FCA) is a strategic choice for SMEs to form alliances with other companies. Early empirical studies have indicated that firms' involvement in network contracts is positively associated with the enhanced economic performance of firms (Burlina, 2020; Cisi et al., 2020). Notably, from analyses of the initial network contracts, smaller firms reaped the benefits of this instrument in terms of sales growth (Costa et al., 2017). Collaborative networks can help firms overcome strategic challenges and gain competitive advantages (Gulati et al., 2000).

These alliances allow firms to quickly access new assets, knowledge, capabilities, and technologies while maintaining flexibility and autonomy (Cohen and Levinthal, 1990; Latham and Le Bas, 2006). Increased access to external information and expertise (Malecki and Veldhoen, 1993) can strengthen a network member's market position or help enter new market segments (Laurell et al., 2017). However, measuring the long-term impacts of network involvement is difficult due to the scarcity of data. (Rosenfeld,1996). Since the legislation is relatively new, strategic business alliances generally require time to yield substantial benefits (Huggins, 2001), such as the absorption of new knowledge and improved learning capabilities. FCAs are crucial for SMEs as they enable firms to exchange information and experiences for mutual benefit by co-producing, co-marketing, co-purchasing, or collaborating on product or market development. (Dickson et al., 2021).

2. Data and variables

Our analysis, conducted with meticulous attention to detail, was based on a comprehensive dataset from the business register provided by InfoCamere, the Italian Chambers of Commerce IT company. Observing a direct relationship between networking and firm performance is difficult because performance is influenced by other factors that over time can produce value for the alliance between firms. These variables include the characteristics of the partners participating in a network agreement, such as the size of the network, geographic distance between firms (Cisi et al., 2020), i.e., geographic location, and sectoral diversity of its members (Schoonjans et al., 2013).

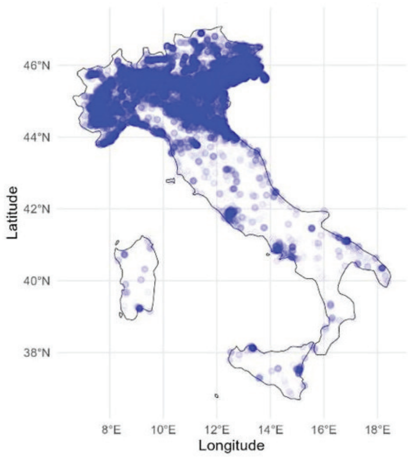
We specifically focused on network contracts without legal subjectivity, a type that is more prevalent than those with legal subjectivity. From this, we derived an initial sample of 5,607 network contracts without legal subjectivity that were active as of June 2024. Subsequently, we delved into the details of 3,035 network contracts, all of which involved firms with headquarters and local units in Northern Italy.

For each contract, the following information is available: the name of the contract, the date of joining the contract, the aims of the agreement, the tax codes, and the names of the participating enterprises.

The analysis is not just a surface-level examination. We have gone to great lengths to delve into additional demographic and economic information for the participating enterprises. This includes their regional (NUTS 2) and provincial (NUTS 3) location, prevailing economic activity (ATECO code), legal nature, date of birth of the enterprise, number of employees, and years of participation in the network contract. This comprehensive approach ensures a thorough understanding of the participating enterprises, providing a solid foundation for our research.

Thus, the sample includes about 11,300 of the 26,800 unterminated enterprises that have participated in at least one active network contract as of June 2024. For this sample, we considered not only the registered offices of the firms but also the local units, locating about 27,500 local units throughout northern Italy. We associated these local units with their corresponding four-digit ATECO codes (prevalent, primary, and secondary). By doing so, we constructed a dataset of approximately 54,700 local units with their corresponding geographic coordinates (Figure 1).

Figure 1 – Locations of the companies participating in FCA



Source: authors'elaboration

3. Methodology

In this section, we will introduce some measures of centrality that offer quantitative insight into the significance of a node within a network. Networks allow for the visualization of complex, multidimensional economic data and offer various statistical indices for interpreting the resulting graphs (Kolaczyk and Csárdi, 2020).

The structure we use to depict this is called a Bipartite graph, in which edges connect two entities. Only some connections in an extensive, complicated network are always equally significant. The most important connections are called the network's "backbone."

This 'backbone' structure is essential for figuring out meaningful connections and comprehending the network's fundamental structure.

The data on firms' geographic locations and ATECO code were converted into an undirected bipartite graph with 729 vertices and 9,827 edges. For our study, the main statistics computed on the bipartite graph examine centrality, and with the help of these, the backbone of the graph is computed. By removing some vertices, the structure of the graph G is simplified, resulting in a subgraph H that preserves the topology and most significant connections of the original graph, allowing for a simpler study.

In the context of graph theory and network analysis, the simplest measures to describe data are the degree, the betweenness and the eigenvector centralities.

Degree Centrality, namely, the number of edges incident on a vertex, represents the sum of the number of edges passing by a vertex. Since the graph is undirected, then the degree of a vertex v is simply the number of edges of value 1 that connect it to the other vertices u within the adjacency matrix A according to the formula:

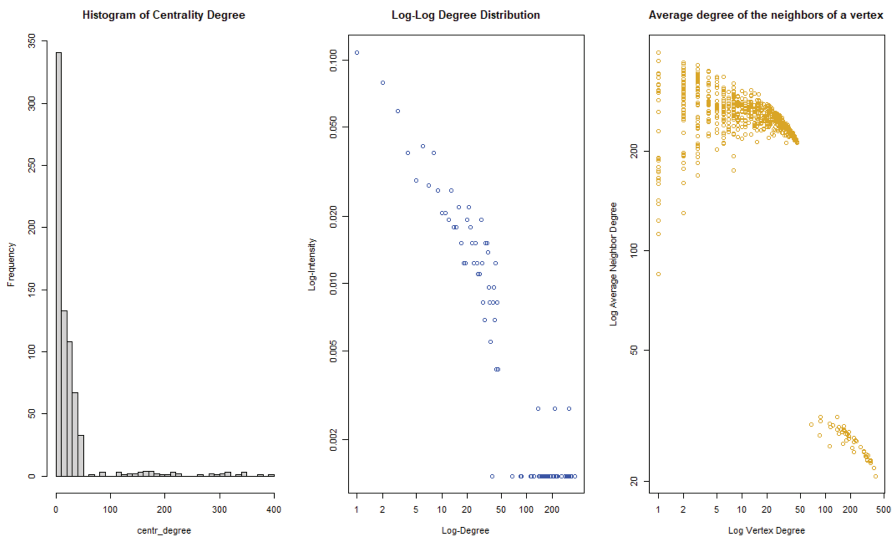
$$Degree(v) = \sum_{u \in V} A_{uv}$$

The higher the degree centrality of a node, the more connections the node has within the network. Nodes with a higher degree of centrality are more likely to be enabled for information and capabilities propagation.

The degree distribution of the vertices is shown below (Figure 2). The first graph shows the frequencies in a histogram, in the second the frequencies can be better appreciated thanks to a logarithmic scale, and in the third it is possible to observe the degree of neighboring vertices as a function of the degree of the vertex considered, i.e., vertices with low degree are connected to vertices with high degree and vice versa.

Higher-ranked nodes include the territories of Milano (398), Torino (374), Brescia (344), Udine (343) and Cuneo (342), while lower-ranked nodes include Lodi (68), Verbano Cusio Ossola (85) and Vicenza (86).

Figure 2 – The degree distribution



Source: authors' elaboration

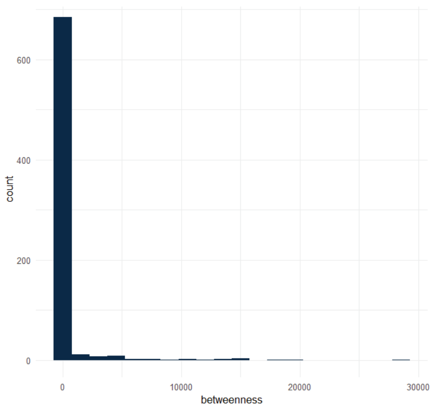
Betweenness centrality was also calculated, which measures how far a node v lies on the minimum paths between other pairs of vertex's s and t, an indication of how much a node acts as a "bridge" that is, how essential it is to the structure of the graph. Nodes with high Betweenness are critical vertices; if these are missing, the graph comes could split into multiple components by not transmitting information.

$$Betweenness(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st(v)}}{\sigma_{st}}$$

In the graph (Figure 3), the absolute frequency for values of betweenness shows the vertices with higher values: Milano (28533), Torino (20202), Cuneo (17383), Bergamo (15700) and Venezia (14985). Finally, eigenvector centrality was calculated, which measures not only the number of edges touching each vertex, but also the importance of the vertices to which that vertex is connected; a node is considered important only if it is connected to other important nodes. The eigenvector centrality of a node v is an index between 0 and 1 defined as the *v-th* component of the principal vector of the equation $Ax = \lambda x$ where A is the adjacency matrix of the network (with $A_{ij} = 1$ if there is an arc between vertices i and j , otherwise with $A_{ij} = 0$), x is the vector of eigenvector centralities, where the *v-th* component x_v represents the eigenvector centrality of node v and λ is the largest eigenvalue of the matrix $\boldsymbol{A}$.

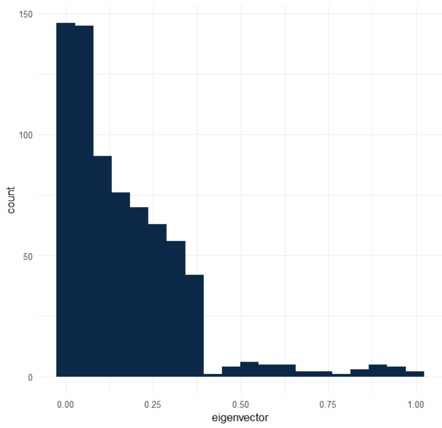
The greater the eigenvector centrality value of a node, the greater its importance within the network. The vertices with higher values represent the provinces of Milano (1.00), Torino (0.99), Udine (0.95), Brescia (0.95) and Cuneo (0.93) (Figure 4).

Figure 3 – Betweenness centrality distribution



Source: authors'elaboration

Figure 4 – Eigenvector centrality distribution



Source: authors'elaboration

4. The backbone model

Networks are valuable tools for modeling phenomena across various fields. While their ability to capture complexity is advantageous, simplifying a network to focus on its backbone can be beneficial in specific scenarios, such as computationally intensive analyses, dealing with noise, or improving visualization. For a complex network N, whether weighted or unweighted, its backbone N′ is a sparse, unweighted subgraph that aims to retain or highlight significant structural elements. Backbone models designed explicitly for bipartite projections are distinct because they operate directly on the bipartite network rather than on its projection. This approach allows these models to incorporate information inherent to the bipartite structure (Neal, 2021).

In a bipartite network, the vertices are separated into two partitions; each edge of the network connects the vertices of one partition to the vertices of the second partition, while there are no edges within the same partition. Therefore, the vertices V_1 of the first partition therefore are connected exclusively to the vertices V_2 of the second partition, and it is our interest to study the vertices of one partition pointing to the same vertices of the second partition, namely, the co-occurrences. Given the bipartite network $G_b = G\left(V_1, V_2, E_{1;2}\right)$, we can proceed by multiplying the adjacency matrix B by its transposed matrix $P = BB^T$, thus projecting the bipartite network into a unipartite network $G_u = G(V_1, E_{1;1})$. In the resulting graph, vertices V_1 are found to be interconnected if they were connected to the same vertex V_2 in the bipartite network. To calculate the significant edges in P, we used the fixed degree sequence model (FDSM) for extracting the backbone from bipartite projections, which we implemented using the *fdsm*() function of Neal's (2021) backbone library. This function requires estimating *p-*

values using Monte Carlo simulation and it employs efficient methods for simulating random networks under the FDSM null model. Nevertheless, many simulations are still necessary to estimate *p*-values with sufficient confidence.

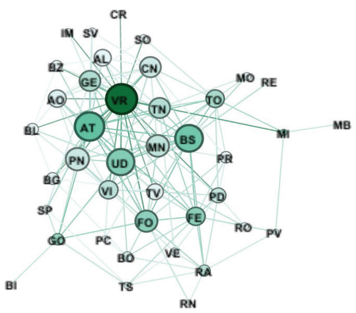
The FDSM is more appropriate than the hypergeometric model (HM) for most spatial data because it accounts for variations in rows and columns (Neal et al., 2022). FDSM is computationally intensive, but given the limited dimension of our dataset, it has been possible to use it proficiently. According to Neal (2021), other models exist, like SDSM, which is a close approximation of the FDSM.

Therefore, the factors guiding a choice between SDSM and FDSM backbones are not methodological, but practical (how large are the data?) and theoretical (how strict should the controls be?). When the data are small and/or strict control is desired, FDSM is more appropriate, while when the data are large and/or less strict control is suitable, SDSM is more appropriate**.** The resulting graph of the spatial backbone has been depicted in the following figures for different p-values (Figures 5-6). Particularly, vertices have size proportional to the eigenvector and colour intensity proportional to betweenness.

The fixed degree sequence odel application returned a p-value of 0.05 and a p-value of 0.10 for the networks with more components. These results are crucial as they guide our decision to retain vertices with degrees greater than zero, thereby shaping further analysis.

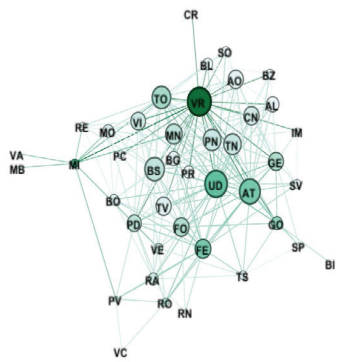
The graph with a p-value of 0.05 shows fewer focal nodes, suggesting a more selective network by retaining only the most significant connections; on the contrary, the graph with a p-value of 0.10 has more relevant nodes, indicating greater inclusiveness of connections, even the least significant ones; as a result, the eigenvector and betweenness values also vary slightly. The difference in *p*-values allows us to observe how the significance level affects density (0.146 applying a p-value = 0.05 versus 0.198 if p-value = 0.10) and network structure (average eigenvector of 0.338 using a p-value = 0.05 versus 0.387 if p-value = 0.10)

Figure 5 – Spatial network using backbone model – *p*-value =0.05



Source: authors'elaboration

Figure 6 – Spatial network using backbone model – *p*-value =0.10



Source: authors'elaboration

The analysis, conducted using the FDSM model, makes it possible to infer the spatial network among a set of locations (provinces) from data on their shared characteristics, i.e., the technological content of economic activities. This confirms the added value in terms of strategic alliance of geographic proximity among firms that have joined at least one network contract.

5. Conclusions

The present study explored the connections between ATECO codes and northern Italian provinces using the backbone of a graph with the FDSM model. The bipartite graph projection analysis revealed that some neighboring provinces show significant connections, suggesting a relevant correlation between geographic proximity and the distribution of economic activities represented by ATECO codes. These results offer new insights into district and regional economic structure and the dynamics of interaction and centrality among provinces. Future studies could further explore these connections, examining the impact of specific ATECO subcategories and their evolution over time to understand better the economic and industrial dynamics of Italian provinces and businesses. The goal is to extend the study to the entire Italian territory to obtain a more complete view of economic connections at the national level.

In this work, we decided to limit our analysis scope to Northern Italy to reduce as much as possible issues related to spatial inhomogeneity.

References

Burlina, C. (2020). Networking policy and firm performance. *Growth and Change*, **51**(1), pp. 161–17.

Cisi, M., Devicienti, F., Manello, A., Vannoni, D. (2020). The advantages of formalizing networks: New evidence from Italian SMEs. *Small Business Economics*, **54**(4), pp. 1183–1200.

Cohen,W. M., Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, **35**(1), pp. 128–152.

Dickson, M. M., Espa, G., Gabriele, R., Mazzitelli, A. (2021). Small businesses and the effects on the growth of formal collaboration agreements: Additional insights and policy implications. *Applied Economics,* **53**(46), pp. 5397–5414.

Gulati, R., Nohria, N., Zaheer, A. (2000). Strategic networks. *Strategic Management Journal*, **21**(3), pp. 203–215.

Huggins, R. (2001). Inter-firm network policies and firm performance: Evaluating the impact of initiatives in the United Kingdom. *Research Policy*, **30**(3), pp. 443–458.

Kolaczyk, E.D., Csárdi, G. (2020). *Statistical Analysis of Network Data with R*, II edition, Springer

Latham,W.R., Le Bas, C. (2006). *The Economics of Persistent Innovation: An Evolutionary View*. Springer.

Laurell, H., Achtenhagen, L., Andersson, S. (2017). The changing role of network ties and critical capabilities in an international new venture's early development. *International Entrepreneurship and Management Journal,* 13(1), pp. 113–140.

Malecki, E. J., Veldhoen, M. E. (1993). Network activities, information and competitiveness in small firms. *Geografiska Annaler*: *Series B, Human Geography*, **75**(3), pp. 131–147.

Neal, Z.P. (2021) backbone: An R package to extract network backbones. *PLoS ONE* **17**(5), pp. 1-24.

Neal, Z.P., Domagalski, R., Sagan, B. (2022) Analysis of spatial networks from bipartite projections using R backbone package. *Geographical Analysis*, **54**(3), pp. 623-647.

Rosenfeld, S. A. (1996). Does cooperation enhance competitiveness? Assessing the impacts of inter-firm collaboration*. Research Policy*, **25**(2), pp. 247–263.

Rubino, M., Vitolla, F. (2016). The effects of network agreements on firms' performance. *Journal of Corporate Governance, Insurance and Risk Management,* **3**(1), pp. 17– 26.

Schoonjans, B., Van Cauwenberge, P., Bauwhede, H. (2013). Formal business networking and SME growth. *Small Business Economics*, **41**, pp. 169–181.