XXIV SemeAd
Seminários em Administração

novembro de 2021
ISSN 2177-3866

The Influence of the Structural Aspects of the Network in the Innovation Activities: A Study in the Brazilian Wine Cluster of Serra Gaúcha

VITOR KLEIN SCHMIDT
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

AURORA CARNEIRO ZEN
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

BERNARDO SOARES FERNANDES
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

CRISTINA BOARI
UNIVERSITY OF BOLOGNA

The Influence of the Structural Aspects of the Network in the Innovation Activities: A Study in the Brazilian Wine Cluster of Serra Gaúcha

1. INTRODUCTION

Innovation plays a central role in the economic development of firms and regions. Innovation occurs through the recombination of different knowledge and can occur as a consequence of the firm's internal capabilities, but also as a result of the absorption of knowledge present in the environment in which the firm operates (Bathelt, Malmberg, & Maskell, 2004; Vicente, 2018). In this perspective, the competition and collaboration relations between the firms and the way that knowledge flows between them stand out (Porter, 1998). The term cluster is used in the business and geographic sciences to explain the role of the geographical agglomeration in the performance of firms (Porter, 1998). In this sense, the cluster approach highlights the impact of geographic proximity on the generation of regional competitive advantages, through the generation of positive externalities and increasing returns that emerge through the sharing of a specialized production structure in a given technological field (Holm & Østergaard, 2015).

The understanding of clusters as networks of firms and interconnected institutions has grown in recent years (Lazzeretti et al., 2019; Maghssudipour et al., 2020) and can be characterized as a set of nodes (firms) and connections (flow between firms) within a region and a technological field (Vicente, 2018). Network studies can be divided according to the relational and structural aspects of the network. While relational highlight how social values are incorporated into the exchange process and are reinforced by geographical proximity (Boschma, 2005; Calignano et al., 2018), structural aspects refer to the ties and positions taken by members within the network, highlighting the network architecture and its impact on the way of how knowledge is generated and transmitted (Balland et al., 2015; Crespo, Suire, et al., 2015; Tsouri & Pegoretti, 2020).

Knowledge networks formed by firms facilitate knowledge to be recombined and transmitted to other members of the network (Giuliani & Bell, 2005). In this sense, the collective innovative capacity of a cluster is directly related to its ability to encourage firms to disseminate and recombine different skills and knowledge (Crespo, Suire, et al., 2015). However, geographical proximity is an insufficient condition to ensure the emergence of benefits and externalities for firms (Boschma, 2005). In fact, both deliberate actions by central actors (Pinkse et al., 2018), as some network structures can constrain the dissemination of knowledge, impacting the collective survival capacity of clustered firms (Crespo et al., 2014; Expósito-Langa & Molina-Morales, 2010).

The different metrics of a network have sociological meanings and can assist in the understanding of the collective processes of knowledge diffusion, an essential aspect for the innovation and survival of firms. The knowledge that is regionally rooted in the firms routines does not spread automatically (Giuliani, 2005). For that, it is necessary that the firms build social bonds that stimulate the exchanges (Morrison & Rabellotti, 2009). In this sense, this study aims to study the relationship between innovation and network structure. More specifically, by studying the relationships and the position taken by wineries organized in clusters, we demonstrate the main network characteristics that influence the knowledge assimilation and transformation into innovation activity.

To achieve our objective, an exploratory quantitative analysis was carried out with the Wine Cluster of Serra Gaúcha (WCSG). We divided our study in three main stages. The first stage has as objective the formation of the Innovation Activity construct. For this analysis we conducted an Exploratory Factor Analysis. The second stage consists of mapping the knowledge exchange networks of wineries present in the WCSG. Such mapping took place through Social Network Analysis (SNA). Finally, the relationship between innovation activity and network metrics was established through different regression techniques and ensemble models.

2 CLUSTERS AS NETWORKS

The approach of clusters emerged from the administrative sciences, aiming to reconcile the importance of geographic agglomeration for the economic performance of firms (Porter, 1998). However, the understanding that a cluster could be treated as a network is relatively recent (Giuliani & Bell, 2005). The perspective of networks demonstrates that the cluster does not justify its existence only by market forces, but also through non-market interactions that can increase knowledge flow, resources, technologies and business opportunities (Crespo et al., 2014; Galaso & Miranda, 2020; Morrison & Rabellotti, 2009). Such relational aspects highlight relationships of trust, connections, exchanges, accumulation of social capital and cooperation (Calignano et al., 2018; Molina-Morales & Martínez-Fernández, 2010; Wal & Boschma, 2011). In this sense, the social bonds that emerge within the cluster allow the circulation of information and the increase of trust, facilitating the emergence of strong social ties (Boschma, 2005; Molina-Morales & Martínez-Fernández, 2010).

The clustered firms have easier access to the knowledge generated within the network due to geographical proximity and the embeddedness in a set of norms and values that are jointly shared (Morrison & Rabellotti, 2009). The term knowledge spillover is commonly used to describe knowledge that is disseminated within an economic system tacitly, through informal interactions between firms and through face-to-face interactions, thus influencing the innovative performance of clustered firms (Giuliani, 2005). Knowledge networks are networks intentionally formed by cluster members in order to seek effective knowledge to solve problems (Morrison & Rabellotti, 2009). Thus, the networking is essential for knowledge creation and transfer (Munari et al., 2012; Tsouri & Pegoretti, 2020). Intensive knowledge exchange networks facilitate not only knowledge transfer, but also inter-firm activities of cooperation and coordination, which may also reduce the risk of opportunistic behavior and allow the creation of new synergies between firms in a cluster (Eisingerich et al., 2012).

Despite the strong argument for the importance of geographical proximity for the diffusion of knowledge. The mere geographical agglomeration does not guarantee clustered firms greater access to new knowledge or any other type of competitive advantage (Boschma, 2005). Innovation occurs through the recombination of different knowledge that is present within the region (Giuliani, 2013). This knowledge can be created directly by a member of the network, or absorbed externally and integrated into the network later (Bathelt et al., 2004). However, knowledge is unevenly distributed within the network of a cluster (Morrison & Rabellotti, 2009). The decision to create and maintain a relationship within the network is linked to the cognitive distance between the actors (Boschma, 2005). In this sense, the increase in innovative capacities involves not only being located in a region, but being actively networked with other firms that have related and heterogeneous knowledge (Crespo, Vicente, et al., 2015).

2.1 Network properties and knowledge flow

Network-based analyzes combine the individual aspects of firms with the resulting network structure, in order to explain how different network architectures emerge and influence the actors' ability to create, access and disseminate knowledge (Crespo, Suire, et al., 2015), directly influencing the collective survival capacity of firms present in the cluster (Balland, Boschma, et al., 2015; Crespo, Suire, et al., 2015; Vicente, 2018). The interorganizational networks of a cluster can assume different natures (productive, commercial, cognitive and social) and of different structural formats. The heterogeneity of these elements indicates that the benefits of being in a network will not be the same for all actors, since there will be actors that will have a greater or lesser influence within the network (Giuliani, 2005). The influence of each actor is given by its relative position in the network, its rootedness, combinations of relationships with other actors, as well as its ability to bridge different actors that are not directly connected (Tsouri & Pegoretti, 2020). In this sense, such characteristics can be extracted and analyzed for the purpose of further understanding the role played by each actor within a network.

The architecture formed by the knowledge network has a direct impact on the transmission of knowledge within the cluster (Galaso & Miranda, 2020). If firms assume isolated strategies or strategies that restrict the diffusion of knowledge, the cluster may fragment and experience difficulties in getting new ideas to circulate within the network. In this sense, the connectedness of the network is measured through its density (Crespo et al., 2014; Morrison & Rabellotti, 2009). Maintaining multiple channels of knowledge circulation allows for greater recombination of knowledge, creating greater opportunities for innovation. The degree centrality of an actor can be measured by the number of ties it has. More central actors have an easier access to new knowledge (Morrison & Rabellotti, 2009). In this sense, Galaso and Miranda (2020) have shown a positive association between the centrality of the actor and his innovation activity.

The betweenness is a measure that assesses the number of shortest paths that pass through an actor in relation to the total possible pairs within the network. From the perspective of clusters, the betweenness assesses the degree of embeddedness of a firm in the cluster and assesses the ease with which important knowledge can be acquired by the firm based on its contacts within the network and the how easily important knowledge can be acquired by the firm based on its contacts within the network (Buenstorf & Costa, 2018). When connecting firms that were not directly related, a firm ends up assuming a position of broker of the network, a role that puts the firm in an advantageous position in the process of acquiring and disseminating knowledge (Fleming et al., 2007). However, the benefits in innovation by occupying a broker position in a network occur unevenly and depend on the absorptive capacity of the broker (Martínez-Cháfer et al., 2018).

The eigenvector centrality identifies that an individual's influence on the network is related to the influence of the actors with whom the actor is connected. In this sense, the centrality of a firm within the cluster is recursively related to the centrality of the actors with which it is related (Buenstorf & Costa, 2018). Kanno (2019) demonstrated the relationship between the eigenvector centrality and the credit contagion risk, indicating that the default risk of a financial institution is related to the risk associated with the financial institutions in which it maintains relation. By extending the analysis to the relations adjacent to the vertices, it is possible to understand the influence of being connected to central actors in innovation activities. By being connected with central actors in the cluster, a firm is able to access the knowledge generated by the network, without the need to maintain a large number of relations. Similar to the idea of the eigenvector centrality, power centrality determines that the power of an actor within the network is recursively related to the power of the actors with whom it is connected.

The cohesion is the term used to explain dense and overlapping relationships. Redundant networks allow the emergence of strong social relationships (Giuliani, 2013), which give rise to social capital within a region (Expósito-Langa & Molina-Morales, 2010) and allows actors to act collectively (Fleming et al., 2007). This aspect helps to explain the emergence of ties within a cluster, mutual monitoring and reduction of opportunistic behavior. (Boschma & Frenken, 2011). Cohesion also demonstrates whether all organizations are part of the same network, or whether there are multiple subnets with different organizations (Calignano et al., 2018). However, very cohesive networks may end up being involved in high transitivity. Transitivity is the property of a node being connected to its adjacent nodes. Transitivity is an important aspect for the growth of a network, because it facilitates the formation of new connections between similar agents. Such conformist behavior makes the central actors connect with greater intensity among themselves, making the peripheral actors poorly connected with the center, constraining knowledge dissemination (Crespo et al., 2014).

Eccentricity is the metric that defines the greatest geodesic distance of a vertex from the others within a network. Although not a widely used measure in regional studies, eccentricity can reveal the degree to which a firm is rooted within a network (Kanno, 2019). A high rootedness is represented by a low distance within the network, which indicates that the firm is densely connected to several other actors, being able to reach them quickly. However, a strong rootedness can also cause the lock-in effect, trapping the firm within the political, cognitive, and hierarchical structures (Schmidt et al., 2020).

Connections within the cluster tend not to occur at random, but through the preferential attachment mechanism (Crespo et al., 2014; Crespo, Suire, et al., 2015; Giuliani, 2013; Suire & Vicente, 2014; Wal & Boschma, 2011). The preferential attachment describes the process in by which a network grows as new nodes select one of the existing nodes in the network to connect (Crespo et al., 2014). The preferential attachment mechanism is an indicator of the existence of a center/periphery structure of a network, in which the firms in the center concentrate most of the links, while the firms in the periphery maintain few links between them (Suire & Vicente, 2014). In this way, central firms have better access to the flow of knowledge (Giuliani & Bell, 2005) and greater stability (Wal & Boschma, 2011) than peripheral firms. Clusters can also have different levels of structural homophily, which can be measured through the assortativity degree (Crespo et al., 2014; Crespo, Suire, et al., 2015). The structure of the relationships will be assortative when highly connected nodes tend to be, disproportionately, connected to other nodes that are also highly connected, or even when peripheral organizations have a great propensity to also be connected with other peripheral organizations. The disassortative structure, on the other hand, occurs when highly connected nodes tend to be disproportionately connected to other poorly connected nodes and vice versa (Crespo et al., 2014). In this sense, the homophily introduces the idea of opening of the network, demonstrating how central and peripheral firms connect and how knowledge circulates within the network (Crespo, Suire, et al., 2015).

Derived from the Web context, the hub and authority algorithms assess the importance of a node in directing information. The relationship between hub and authority occurs recursively: a hub firm is one that sends a large number of ties to other firms that already receive a large number of ties. In turn, a firm with a high authority value is one that receives a large number of ties from hub firms (Deguchi et al., 2014). Authorities are the firms that receive the most information, placing them in a privileged position to carry out cross-fertilization of knowledge (Holm & Østergaard, 2015). Hubs, on the other hand, are the information disseminators firms, playing an important role for knowledge to be transmitted. Munari et al. (2012) have shown that knowledge diffusion in a cluster tends to occur through a core set of firms. In this sense, firms can be divided between those that send a greater amount of information to more relevant sources and those that receive a greater amount of information from more relevant sources through their network connections.

3. METHOD

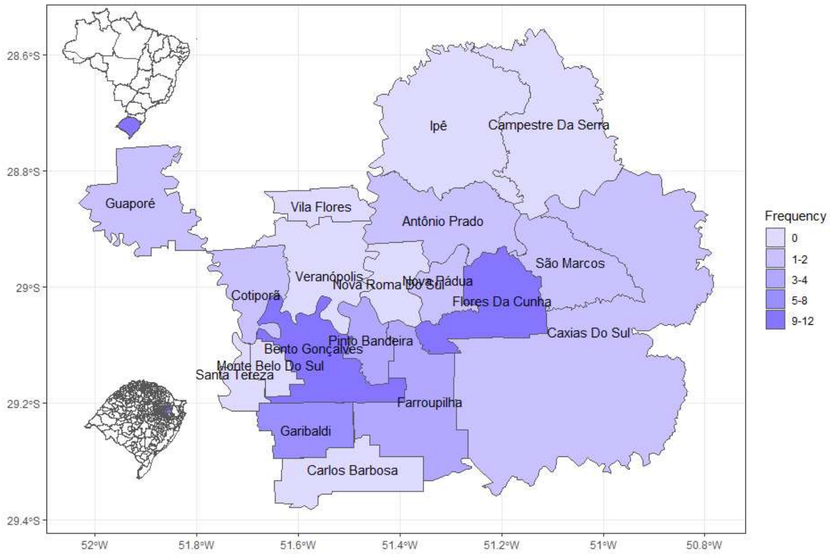
Despite the quantitative character, this research is characterized as exploratory, since the data were obtained in a cross-sectional way and the idiosyncrasies of the WCSG limit the generalization of the results found here. For analysis, this study combined Social Network Analysis with psychometric and machine learning (ML). SNA is a methodological tool that serves to assess interactions and knowledge exchange in regions and clusters (Tsouri & Pegoretti, 2020) and can be combined with ML techniques in order to achieve better results and analyze possible relationships between network variables and variables related to the performance of firms (Galaso & Miranda, 2020). Ties between firms can be characterized by trade, workers, knowledge, or any type of exchange that two or more firms can maintain. However, within the evolutionary literature, greater emphasis is placed on interactions and the flow of knowledge (Boschma & Frenken, 2011; Vicente, 2018), being the unit of analysis normally used in studies within the literature of clusters (Crespo et al., 2014; Giuliani & Bell, 2005), since a firm's ability to innovate depends not only on its internal capabilities, but also on the quality of the links it has with other organizations (Giuliani et al., 2019; Vicente, 2018). In this sense, firms collaborate with each other in order to acquire complementary sets of knowledge and develop innovations that make them more competitive (Crespo, Suire, et al., 2015).

The cluster that served as the object of analysis is the Wine Cluster of Serra Gaúcha. The WCSG is located in the Brazilian state of Rio Grande do Sul and originated from Italian immigration in the late 19th century. The cluster employs 2,554 employees in 154 wineries of which the vast majority are micro and small (RAIS, 2020). For most of its history, the cluster has focused on producing table wines, white wines, vinegar and cognac, products that, in general, have a low added value. However, since the early 2000s, wineries have increasingly invested in improvements in wine production, modernizing equipment, new products, new business models related to tourism and new technologies aimed at the production of wines and grape products.

As performed by Giuliani e Bell (2005), this research established some criteria, in order to operationalize the data collection. Three criteria were used in order to determine the sample population. The criteria were: (I) having produced at least 1,000 liters of fine wines in 2018; (ii) having won at least one award in the last three years (2016-2018); (iii) be present at the Serra Gaúcha Wine Cluster. Once the criteria were determined, data was sought from the Brazilian Institute of Wine and Vine in order to determine the population size for this study. Such data were made available and a total of 56 wineries were reached.

Once the wineries were determined, telephone contacts were made with them, in order to schedule the questionnaire. The questionnaire was applied in person, in order to minimize problems of non-participation, missing values or answer errors. The application of the questionnaire had an average time of 40 minutes, which application started in late 2018 and was completed in early 2019. In total, of the 56 selected wineries, 46 accepted to participate in the study (82.1%). The respondents were all owners or high-level managers involved in innovation activities in the wineries. As a data collection instrument, a questionnaire was used that covered questions related to innovation activities and questions related to knowledge exchange. Figure 1 presents the cluster map with the frequency distribution by municipality of the survey respondents.

Figure 1 – Wine Cluster of Serra Gaúcha Map



All statistical analyzes were performed with the aid of the R software, through its packages. For this research, three main analyzes were carried out: Exploratory Factor Analysis for forming the Innovation Activity construct, SNA for measuring network metrics and different statistical models for inferential statistics. The dependent variable of this study is the construct "Innovation Activity", which was represented by five questions measured using a 5-point Likert-Type scale, in which the wineries positioned themselves. In addition to the Likert-Type scale, wineries were also asked whether or not they introduced different activities related to viticulture in the last three years. Such questions allowed the elaboration of a factorial map of the innovation activities, through the Multiple Correspondence Analysis.

For the SNA, we asked the wineries to indicate, based on a complete list of the other wineries that belonged to the original sample, and we asked them to name all the other wineries that they received technical information in the last three years (Morrison & Rabellotti, 2009). This procedure starts the formation of a network for exchanging technical information, information that is essential for the development of innovations (Giuliani & Bell, 2005; Maghssudipour et al., 2020). From the extraction of network statistics, the data were concatenated in a new dataset and were used as independent variables for the models. For this study, we used the ego network metrics used are: Local Transitivity, All-Degree, In-Degree, Out-Degree, Closeness, Hub, Authority, Betweenness, Eccentricity, Ego Size, Eigen Centrality e Power Centrality. The choice for these metrics was due to their recurrent use in several studies that propose to analyze the relationship between the knowledge network of firms in clusters and innovation (Galaso & Miranda, 2020; Giuliani, 2013; Giuliani et al., 2019; Giuliani & Bell, 2005; Morrison & Rabellotti, 2009; Tsouri & Pegoretti, 2020). These metrics translate firms' tendency to receive or send information, their embeddedness, prestige, and power within the network. In addition to the ego network metrics, we also analyzed the structure of the network as a whole, identifying the patterns of connections between members and the division between center and periphery generated by the network. This aspect is important, as it can influence the transmission of knowledge within the network (Crespo et al., 2014; Crespo, Suire, et al., 2015; Suire & Vicente, 2014).

The inference between the network metrics and the innovation activity occurred through different regression techniques and ensemble models. Machine Learning (ML) models are increasing in popularity due to the fact that their results tend to be superior to traditional econometric and psychometric models. ML algorithms are designed to maximize the performance of independent variables over dependent variables. This result is possible due to the fact that an ML algorithm follows the patterns revealed, inductively, by the data set itself (Choudhury et al., 2021).

Initially, the data were scaled using the Z-score. After scaling, the data were divided into two subsets: training set and test set. The separation of the data was carried out through the same random seed, in order to guarantee the equality of the sets and the reproducibility of the analyzes. The training set made up 70% of the sample (34) and was used to train the models, whereas the test set made up 30% (12) of the sample, being used to test the results and the generation of error metrics. As a cross-validation technique, Leave-One-Out Cross-Validation (LOOCV) was used for all analyzes. Lasso Regression and Elastic Regression were used to perform the inference as linear techniques and ensemble techniques based on trees as nonlinear techniques.

Lasso's Regression introduces a penalty on the estimated coefficients, reducing to zero those that do not assist in the prediction, thus removing them from the model. Elastic Regression maintains Lasso's characteristic and combines it with Ridge regularization, which reduces the chances of overfitting the data and eliminates problems with multicollinearity between the predictor variables. In this sense, both techniques aim to reduce the number of variables that are not important for the model, being suitable for the evaluation of the importance of coefficients in a linear model (James et al., 2013). Non-linear models are more flexible and thus better able to capture the variability and complexity of the relationships between variables (James et al., 2013). Thus, we also used nonlinear models to analyze the dataset. The nonlinear models were chosen based on maintaining an ensemble logic. The ensemble models add a large amount of individual learning results to create more accurate predictive models. We used models based on decision trees, which adopt an independent voting system of weak classifiers. The models used were: Random Forest, Bootstrap Aggregation (Bagging), Gradient Boosting Machine (GBM) and Extreme Gradiente Boosting.

The results are then evaluated based on the error metrics. The model with the best results is then saved and used for the analyses. The focus of this research is to evaluate the individual contribution of each variable in the model. However, while linear models produce easily interpretable coefficients associated with a T-statistic value, ensemble models do not produce such coefficients (Choudhury et al., 2021). In this sense, in order to be able to compare the importance of each variable in each model, we proceeded with the method of evaluating the Variable Importance. The use of different techniques allows the comparison of the performance of the influence of each network metric on the Innovation Activity, thus reducing biases that some technique may offer on the results. The script used can be made available on request.

4 RESULTS

In order to reduce the five items of the scale of innovation activity into a single factor, we conducted an EFA. Despite the small sample size, both the KMO and Bartlett tests demonstrated sufficient sample adequacy for the performance of EFA, reaching 0.73 in the KMO test and a significance level of less than 5% in the Bartlett test. ($\rho < 0,001$; $x^2 = 50.6$). For the retention of factors, both the Kaiser and the parallel lines criterion signed only one factor with an eigenvector above 1. Also, the choice of a single factor proved to be superior in the analysis of Simple Structure Criterion, Wayne Velicer's Minimum Average Partial, Root Mean Square Error of Approximation (RMSEA) and Bayesian Information Criterion (BIC). In this way, we proceeded the EFA by extracting only 1 factor. As the extraction method, we used Principal Axis Factoring Promax orthogonal rotation. The model converged into a stable solution with adequate adjustment rates (RMSR = 0.08, TLI = 0.908, RMSEA = 0.086, CFI = 0.955, CFI = 0.955, Cronbach Alfa Standardized = 0.75). The resulting factor explains 50.06% of the total variation of the questions. In addition, all the factorial loads obtained were above |0,4|. The scores were extracted and concatenated in the database according to the respective observation, being used later as a dependent variable. Table 1 shows the list of questions used and the factorial loads obtained.

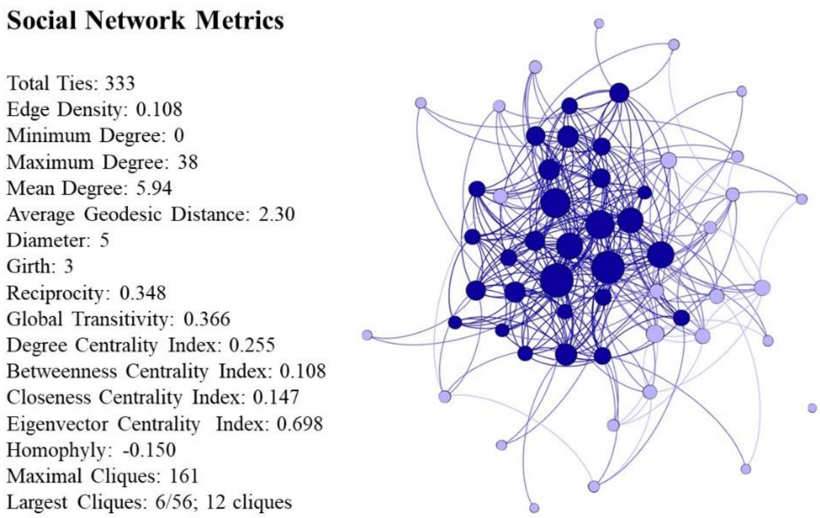
Once the factor scores were obtained, we continued with SNA. The formed network has 333 ties, having a directed character. The network metrics extracted and used as independent variables. In addition to the ego network metrics, we also analyzed metrics related to the network as a whole. All global measurements were extracted using the format of directed networks (with the exception of the analysis of clicks and global transitivity, which need to be calculated in a non-directed format. In these cases, the network was transformed into non-directed). The general density of the network is 0.108, that is, 10.8% of all possible connections are present, resulting in a network with a medium-low density. The average of total connections was close to 6 ties for each winery. Still, few relationships are in fact reciprocal (34.8%). In other words, the winery that provides information to a winery does not necessarily ask or receive for information in return. This type of behavior occurs within environments marked by opportunism, in which the firms do not disseminate their proprietary knowledge, fearing that such action may increase competition (Giuliani, 2013). The average geodesic distance is 2.30, indicating that wineries are capable to access each other relatively easily, which facilitates knowledge to circulate more freely within the network.

The global transitivity, assesses the probability that two actors are connected. When many actors connect, they tend to create communities within the network. This does not seem to be the case for the network formed by the wine cluster, since the overall coefficient remained mediocre (0.366). The great difference in the number of relationships indicates the existence of an unequal relationship in the network, in which some wineries concentrate a good part of the relationships (center) and few wineries concentrate a smaller number of ties (periphery). The biggest clicks were made up of 6 wineries, occurring this event 12 times. The vast majority of wineries that belong to these distinct social groups are large wineries that are connected to medium-sized wineries and that have a central position in the network.

Based on the idea of clicks and communities, the Cohesive Blocking technique was used to determine the belonging of the vertices in the central or peripheral region of the network. In total, 7 substructures were identified. Wineries that belonged to the maximum number of substructures were labeled "central" and the others as "periphery". In this way, it was possible to identify the wineries that belong to the center (29) and those that belong to the periphery (27) of the network. Finally, the analysis of the degree correlation indicates a weak disassortative structure (-0.15). This fact indicates that there is no tendency for central and/or peripheral firms to be strongly connected with each other. This aspect is essential to ensure that new knowledge generated, especially in the periphery, circulates to the center.

Figure 2 presents a representation of the graph formed by the network. The color of the vertices was defined based on their central (dark blue) or peripheral (light blue) position. In addition, the size of the vertices was weighted based on the total number of relationships that each node has (all-degree). In this sense, it is possible to identify, through the Mann-Whitney test that, as a rule, the nodes belonging to the center are also those that concentrate most of the relationships ($W = 19.5$; $p < 0.001$; $r = -0.787$).

Figure 2 – Graph and metrics extracted from the network



In addition to the Likert-Type scale used to measure innovation activity, the innovation activity was also measured in a dichotomous manner, asking whether the winery performs that particular activity (yes/no) and can be seen in Table 2. Such dichotomous variables were used as active categories for the construction of a multiple correspondence model (MCA). MCA is an exploratory multivariate dimension reduction technique for categorical variables and graphical analysis. As a result of the analysis, a Euclidean representation of the data is obtained in a few dimensions. The results can be interpreted based on the relative positions of the points and their distribution along the axes. Each point represents a category of a variable and its distance on the plane measures the correspondence between them: the closer they are on the plane, the more similar the relations are. The labels for the variables used are shown in Table 1. In addition to the variables of interest, the position of the winery in the network (center/periphery), the existence of tourism activities, production size and whether the winery has production with Geographical Indication were used as supplementary variables[i].

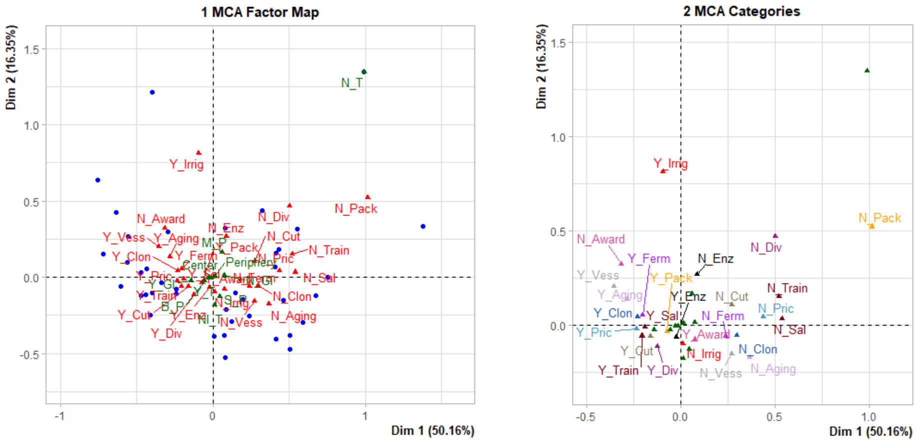
Table 1 – Categorized Innovation Activities

In the past three years, which of these practices have been introduced in your winery	**No**		**Yes**	
	Frequency	MCA Label	Frequency	MCA Label
1. Introduction of different clones or varieties in the terroir of the vineyard	20	N_Clon	26	Y_Clon
2. Irrigation system	41	N_Irrig	5	Y_Irrig
3. Viticulture training systems	13	N_Train	33	Y_Train
4. New fermentation techniques	21	N_Ferm	25	Y_Ferm
5. New enzyme and/or yeast	9	N_Enz	37	Y_Enz
6. Aging period	20	N_Aging	26	Y_Aging
7. Use of different vessels for wine aging	26	N_Vess	20	Y_Vess
8. New varietal cuts	17	N_Cut	29	Y_Cut
9. Product packaging	3	N_Pack	43	Y_Pack
10. New channels to promote products	9	N_Div	37	Y_Div
11. New channels for product sales	12	N_Sal	34	Y_Sal
12. New pricing strategies	16	N_Pric	30	Y_Pric
13. Participation in new national and/or international awards or contests	9	N_Award	37	Y_Award

Source: Authors

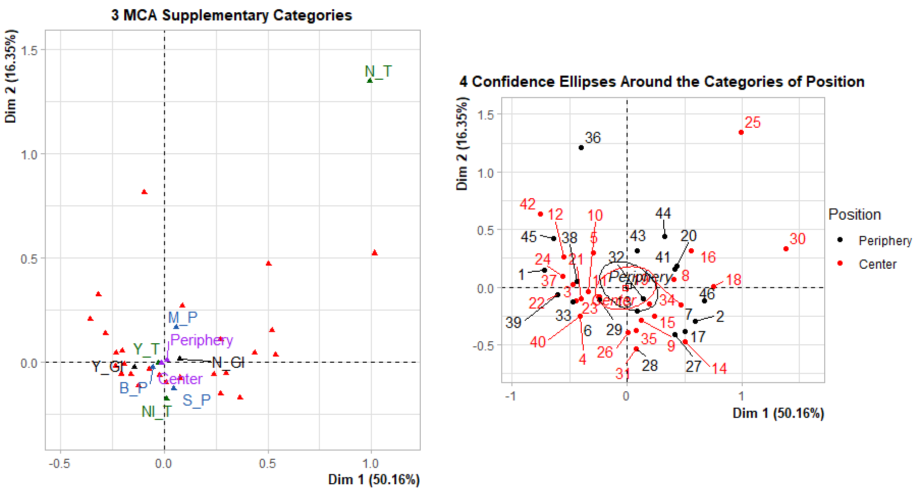
Due to a question of graphic parsimony, the graphs were divided between the graphs of the active categories and the supplementary categories. Figure 3 shows the factorial maps generated by the MCA analysis, with emphasis on the active categories. Map 1 presents a condensate of all active categories (red), supplementary (green) and individuals (blue). While the individuals are more evenly distributed across the graph, it is possible to verify the existence of a concentration of active and supplementary categories around the average. The distribution of active categories is best explored on map 2. To facilitate visualization, a color is assigned to each set of categories associated with a question, keeping the color green for supplementary categories. From Map 2, it is possible to verify that most innovation activities are carried out in the quadrants located on the left, while the negation of that particular activity is located on the right.

Figure 3 – MCA of Active Categories



The supplementary variables do not directly affect the geometric space of the map, but assist in the interpretation of the results. Factor map 3, shown in Figure 4, shows the layout of the supplementary variables and, just like the active variables (those marked in red), each set of categories received a different color. Based on the map, it is possible to verify that neither the quantity of wines produced nor the production by geographical indication have significant differences in terms of the distribution of the groups, a fact confirmed by the confidence intervals of the ellipses. Map 4 shows the confidence interval of the ellipses for the position of the network. It is possible to verify the overlapping of the ellipses, indicating that there are no differences between the wineries that are in the center or on the periphery of the network for the development of innovation activities, a fact still confirmed by the Student t statistic between the position of the network and the Innovation Activities construct ($t[44] = -1.239$; $p > 0.05$; $d = -0.368$).

Figure 4 – MCA of supplementary variables



In order to make the inference about which network metrics are capable of explaining the variation in the Innovation Activities construct, different linear and non-linear models were applied and the models were chosen through their adequacy to the data and the objectives of this study. The models followed the pre-processing process described in the method. Each model was trained only using the training data and its performance occurred through the calculation of the error metrics for the test data. The calculated error metrics were Root Mean Square Error (RMSE), Determination Coefficient (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE). As shown in Table 3, both linear models did not obtain satisfactory results in the prediction of the test data. This is due to two main reasons: the low amount of available data and the non-linearity of the relationship between the network metrics and the Innovation Activities construct. In order to deal with the non-linearity of the relationship, we used non-linear ensemble models.

The algorithms of Variable Importance order the contribution of each predictor variable within the model. Through permutations, different models are constructed by removing one variable at a time. After removal, the corresponding reduction in the predictive capacity of the model is calculated. The importance of the variable is then calculated by increasing the of the error metrics. A variable is considered important when its removal increases the error and decreases the predictive capacity of the model. In order to facilitate the interpretation of the individual importance of each variable, all measures are normalized and multiplied by 100, the greater the importance of the variable, the higher the score obtained. Table 3 presents the scores obtained for each network metric in each model and the error metrics.

Table 3 – Models results

Network Metric	Lasso Regression	Elastic Regression	Ranger: Random Forest	Bagging	GBM	XGBoost
Authority	100	100	78.70	100	100	100
Betweenness	.	34.80	1.87	53.29	16.01	68.06
Closeness	.	21.53	16.96	62.40	31.49	57.55
Eccentricity	.	.	.	.	.	.
Ego Size	.	.	18.64	31.13	7.11	29.73
Eigen Centrality	55.32	90.23	100	71.70	81.77	17.66
In-Degree	.	.	71.24	87.08	31.59	9.88
Out-Degree	66.59	39.98	53.81	21.51	22.61	7.62
All-Degree	.	.	42.58	37.07	13.72	1.41
Hub	8.96	49.25	39.36	53.69	99.27	17.90
Power Centrality	8.25	29.13	11.98	23.83	56.35	24.73
Local Transitivity	.	.	11.35	31.58	35.05	19.15
Error Metrics Based on The Test Data						
RMSE	1.011	1.024	0.957	0.966	0.955	0.954
R^2	0.226	0.225	0.342	0.415	0.472	0.509
MAE	0.812	0.824	0.739	0.691	0.696	0.688
MAPE	1.565	1.697	1.292	1.125	1.116	1.003

Source: Authors

Based on the results obtained, it is possible to verify that the use of non-linear models resulted in a significant improvement in the coefficient of determination and a reduction in the error statistics. On average, the variables that were most important in the different models were: Authority, Eingen Centrality, Hub and Closeness. The most important metrics highlight the importance of the quality of the ties more than the number of ties that a winery has within the network. In this sense, receiving a large amount of information or information from privileged and more relevant sources within the network seems to have a greater impact on the innovative activity of wineries. Conversely, the variables that make the least contribution to the models were: Eccentricity, Ego Size and Local Transitivity. Such metrics refer to the connections that a given vertex has with closer and similar vertices. In this sense, a weak embedded in the network, or the maintaining ties only within their respective clicks and neighborhoods can limit the acquisition of new knowledge and, consequently, restrict the activity of innovation.

5 DISCUSSIONS

Clustered firms are rooted in a large social and economic structure and their performance is influenced by social, institutional and cultural aspects that are transmitted locally, through the exchange networks between firms (Boschma, 2005). The dynamics existing within a network influences how knowledge is disseminated and recombined (Vicente, 2018), limiting diffusion and leading the cluster to decline (Schmidt et al., 2020), or facilitate knowledge to be disseminated and promote renewal and growth (Grillitsch et al., 2018). In this perspective, the network literature has placed a lot of emphasis on the disseminating knowledge within the network: the more dynamic and open the network, the more easily the knowledge spreads, collectively improving the routines present in the cluster. Such process occurs through the dissemination of knowledge and other externalities (Holm & Østergaard, 2015). This research demonstrates that the innovation activity of a clustered firm is not necessarily related to the absolute number of connections, or to the hierarchical position in the network.

The evolutionary perspective highlights the fact that knowledge is heterogeneously concentrated among firms. In this sense, not all ties should be considered to have the same weight, since knowledge is spread unevenly within a cluster (Maghssudipour et al., 2020). The network formed by the cluster has different hierarchical levels and with the domain of some brokers in its structure. Both aspects influence the associated power of each firm within the network. However, none of these elements proved to be relevant in this study. This aspect can be explained due to the ease with which knowledge spreads through the network, represented by the geodesic distance, as well as due to the disassortative character of the network (Crespo, Suire, et al., 2015).

For the acquisition of new knowledge, the position of a firm in the center or on the periphery of the network has a secondary role (Morrison & Rabellotti, 2009). Even firms inserted in the periphery can also have access to the knowledge generated by the network, if they are connected with highly connected firms. This argument favors the analysis of the structural homophily of the network: disassortative networks favor the dissemination of knowledge and the renewal of the cluster, through the maintenance of relations between firms in the center and in the periphery (Crespo, Suire, et al., 2015). Another explanation for the fact that the central actors do not have a greater innovative activity is that these actors are precisely the firms that have the smallest knowledge bases (Martínez-Cháfer et al., 2018). A smaller knowledge base implies the need for the firm to collaborate more intensively in the local knowledge network, since such actors would not have sufficient absorptive capacity and resources to reach the high international standards. In this way, strong roots in the local network would be one of the few alternatives to access new knowledge (Morrison & Rabellotti, 2009).

The network metrics that showed a better predictive capacity were those related to the qualitative aspects of the relationships. The in-degree metric points to the importance of receiving information within the network; this information is received and recombined to the routines that already exist, allowing the cross-fertilization of knowledge and the development of innovations (Galaso & Miranda, 2020). In fact, the authority metric was the one that obtained the best performance in most models, indicating the importance of receiving information within the big diffuser's nodes, the more information the winery receives from the network, the greater the possibilities for recombination and innovation. The importance of the eigenvector centrality reveals that wineries with a better position in the network is more relevant than the absolute number of relationships. This aspect resides in the idea that maintaining relationships includes both costs and benefits (Crespo, Vicente, et al., 2015; Giuliani, 2007). In this sense, maintaining relationships only with more prestigious members within the network reduces the costs associated with maintaining a relationship, while it maximizes information gains through indirect contacts (Morrison & Rabellotti, 2009).

The large amount of information diffusers (hub) also indicated to be relevant. The explanation behind this aspect may be due to the logic of increasing returns associated with the establishment of a dominant design (Crespo et al., 2014). Within a network, some firms concentrate most of the links and information, through the preferential attachment mechanism (Wal & Boschma, 2011). These central firms start to encourage the circulation of knowledge, compliance and establish social norms among network members (Suire & Vicente, 2014), concentrating most of the knowledge diffusion (Munari et al., 2012). More than knowledge, the diffusers also disseminate a design among the firms, establishing greater stability and adaptation to the network (Tsouri & Pegoretti, 2020). Thus, as more firms adopt a disseminated routine, the returns from that routine tend to increase, generating a positive feedback cycle that benefits the diffusers.

The most irrelevant metrics were Eccentricity, Ego Size and Local Transitivity. In this sense, the irrelevance of eccentricity indicates that having a large geodetic distance does not influence the acquisition of new knowledge, especially within a scenario of small words, as in the case of clusters (Crespo, Suire, et al., 2015). Likewise, the size of the neighborhood of an ego size firm is also not relevant. Both aspects reinforce the idea that central firms, which have a shorter geodesic distance and a large neighborhood, do not necessarily have a superior innovative activity. The low importance of local transitivity points to the idea that communicating with firms that are present within the same substructures makes it difficult to introduce new knowledge. Despite the advantages associated with the social capital that emerges in cohesive and densely connected groups (Expósito-Langa & Molina-Morales, 2010), maintaining ties only with firms that have a similar relational structure might turn the network inflexible (Eisingerich et al., 2012), leading to a conformist and isomorphic behavior (Crespo et al., 2014), which can hinder the diffusion of new knowledge (Balland et al., 2015), increasing the chances of the emergence of a lock-in effect within the network (Schmidt et al., 2020).

6 CONCLUSIONS

In this paper, we explore the role of the network metrics in innovation activities. More specifically, we examined the relation between network analysis metrics in the innovation activities for the Wine Cluster of Serra Gaúcha. Clusters can be analyzed as economic structures formed by firms connected through intense exchange network, which are vital for maintaining an innovative and vibrant cluster (Bathelt et al., 2004; Giuliani, 2005). When analyzing the process of acquiring knowledge within a network, some actors end up concentrating a greater amount of relationships within the network, becoming important hubs or authorities in the process of disseminating or acquiring knowledge (Crespo, Vicente, et al., 2015). Such aspects are important for the emergence of a dominant design and the generation of positive externalities that guarantee the competitive advantages for the clustered firms (Porter, 1998).

Exploring the role of network in the innovation activities, this article has contributed to the current knowledge network literature. Maintaining an innovative cluster requires firms to be able to recombine their knowledge for new possibilities for economic exploitation (Grillitsch et al., 2018). Therefore, both individual capacities and interorganizational relations influence the survival of the cluster. (Crespo et al., 2014). In this study, we demonstrate that it is not the absolute number of ties or the hierarchical position of the winery that influences innovation. Instead, what influences innovation in clusters is being connected with firms that are highly connected within the network. Receiving information from other actors who have a large number of ties means that the winery is capable to receive knowledge generated within the network without the burden of having to maintain a large number of relationships. In this sense, even a peripheral winery, but which has relationships with more prestigious firms, manages to access the knowledge disseminated within the network. To improve the knowledge spillover, the network must have a disassortative pattern (Crespo et al., 2014).

Most of the network studies highlight the fact that the actors get closer due to their similarities (Expósito-Langa & Molina-Morales, 2010; Fleming et al., 2007; Molina-Morales & Martínez-Fernández, 2010). However, there is little to learn from those who are cognitively close, or from those who shares the same beliefs systems and knowledge bases (Boschma, 2005). In fact, maintaining the same routines and interactions can lead the cluster to lock-in, one of the main explanations for the decline of clusters (Schmidt et al., 2020). In this sense, it is not surprising that the eccentricity, ego size and transitivity, which are metrics associated with the conformity, are not relevant. In this sense, the lack of the relation could be potentially explained by the idea that wineries that are connected only with neighboring wineries which tend to belong to the same social group, tend to have a greater difficulty in accessing heterogeneous knowledge sources.

Despite the efforts to measure the effects of the structural aspects of the network on the innovation activity with different techniques, this research has a bottleneck related to some wineries' non-response bias: Social Network Analysis is very sensitive to missing data. In this sense, in order to validate the results found, future studies may replicate the techniques used here in larger and complete networks.

References

Balland, P. A., Boschma, R., & Frenken, K. (2015). Proximity and Innovation: From Statics to Dynamics. *Regional Studies*, *49*(6), 907–920. https://doi.org/10.1080/00343404.2014.883598

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31–56. https://doi.org/10.1191/0309132504ph469oa

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Boschma, R., & Frenken, K. (2011). The emerging empirics of evolutionary economic geography. *Journal of Economic Geography*, *11*(2), 295–307. https://doi.org/10.1093/jeg/lbq053

Buenstorf, G., & Costa, C. (2018). Drivers of spin-off performance in industry clusters: Embodied knowledge or embedded firms? *Research Policy*, *47*(3), 663–673. https://doi.org/10.1016/j.respol.2018.01.015

Calignano, G., Fitjar, R. D., & Kogler, D. F. (2018). The core in the periphery? The cluster organization as the central node in the Apulian aerospace district. *Regional Studies*, *52*(11), 1490–1501. https://doi.org/10.1080/00343404.2017.1420155

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30–57. https://doi.org/10.1002/smj.3215

Crespo, J., Suire, R., & Vicente, J. (2014). Lock-in or lock-out? How structural properties of knowledge networks affect regional resilience. *Journal of Economic Geography*, *14*(1), 199–219. https://doi.org/10.1093/jeg/lbt006

Crespo, J., Suire, R., & Vicente, J. (2015). Network structural properties for cluster long-run dynamics : evidence from collaborative R&D networks in the European mobile phone industry. *Industrial and Corporate Change*, *25*(2), 1–22. https://doi.org/10.1093/icc/dtv032

Crespo, J., Vicente, J., & Amblard, F. (2015). Micro-behaviors and structural properties of knowledge networks : toward a ' one size fits one ' cluster policy. *Economics of Innovation and New Technology ISSN:*, *8599*(October), 1–20. https://doi.org/10.1080/10438599.2015.1076199

Deguchi, T., Takahashi, K., Takayasu, H., & Takayasu, M. (2014). Hubs and authorities in the world trade network using a weighted HITS algorithm. *PLoS ONE*, *9*(7), 1–16. https://doi.org/10.1371/journal.pone.0100338

Eisingerich, A., Falck, O., Heblich, S., & Kretschmer, T. (2012). Firm Innovativeness across Cluster Types. *Industry and Innovation*, *19*(3), 233–248. https://doi.org/10.1080/13662716.2012.669619

Expósito-Langa, M., & Molina-Morales, F. X. (2010). How relational dimensions affect knowledge redundancy in industrial clusters. *European Planning Studies*, *18*(12), 1975–1992. https://doi.org/10.1080/09654313.2010.515817

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, *52*(3), 443–475. https://doi.org/10.2189/asqu.52.3.443

Galaso, P., & Miranda, A. R. (2020). The leading role of support organisations in cluster networks of developing countries. *Industry and Innovation*, *00*(00), 1–30. https://doi.org/10.1080/13662716.2020.1856046

Giuliani, E. (2005). Cluster absorptive capacity: Why do some clusters forge ahead and others lag behind? *European Urban and Regional Studies*, *12*(3), 269–288. https://doi.org/10.1177/0969776405056593

Giuliani, E. (2007). The selective nature of knowledge networks in clusters: Evidence from the wine industry. *Journal of Economic Geography*, *7*(2), 139–168. https://doi.org/10.1093/jeg/lbl014

Giuliani, E. (2013). Network dynamics in regional clusters: Evidence from Chile. *Research Policy*, *42*(8), 1406–1419. https://doi.org/10.1016/j.respol.2013.04.002

Giuliani, E., Balland, P. A., & Matta, A. (2019). Straining but not thriving: Understanding network dynamics in underperforming industrial clusters. *Journal of Economic Geography*, *19*(1), 147–172. https://doi.org/10.1093/jeg/lbx046

Giuliani, E., & Bell, M. (2005). The micro-determinants of meso-level learning and innovation: Evidence from a Chilean wine cluster. *Research Policy*, *34*(1), 47–68. https://doi.org/10.1016/j.respol.2004.10.008

Grillitsch, M., Asheim, B., & Trippl, M. (2018). Unrelated knowledge combinations: The unexplored potential for regional industrial path development. *Cambridge Journal of Regions, Economy and Society*, *11*(2), 257–274. https://doi.org/10.1093/cjres/rsy012

Holm, J. R., & Østergaard, C. R. (2015). Regional Employment Growth, Shocks and Regional Industrial Resilience: A Quantitative Analysis of the Danish ICT Sector. *Regional Studies*, *49*(1), 95–112. https://doi.org/10.1080/00343404.2013.787159

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning - with Applications in R*. Springer. https://doi.org/10.1007/978-1-4614-7138-7

Kanno, M. (2019). Network structures and credit risk in cross-shareholdings among listed Japanese companies. *Japan and the World Economy*, *49*(September 2018), 17–31. https://doi.org/10.1016/j.japwor.2018.09.003

Lazzeretti, L., Capone, F., Caloffi, A., & Sedita, S. R. (2019). Rethinking clusters. Towards a new research agenda for cluster research. *European Planning Studies*, *27*(10), 1879–1903. https://doi.org/10.1080/09654313.2019.1650899

Maghssudipour, A., Lazzeretti, L., & Capone, F. (2020). The role of multiple ties in knowledge networks: Complementarity in the Montefalco wine cluster. *Industrial Marketing Management*, *April 2019*. https://doi.org/10.1016/j.indmarman.2020.03.021

Martínez-Cháfer, L., Molina-Morales, F. X., & Peiró-Palomino, J. (2018). The cluster is not flat. Uneven impacts of brokerage roles on the innovative performance of firms. *BRQ Business Research Quarterly*, *21*(1), 11–25. https://doi.org/10.1016/j.brq.2017.12.002

Molina-Morales, F. X., & Martínez-Fernández, M. T. (2010). Social network effect on innovation. *Journal of Small Business Management*, *48*(2), 258–279.

Morrison, A., & Rabellotti, R. (2009). Knowledge and information networks in an Italian wine cluster. *European Planning Studies*, *17*(7), 983–1006. https://doi.org/10.1080/09654310902949265

Munari, F., Sobrero, M., & Malipiero, A. (2012). Absorptive capacity and localized spillovers: Focal firms as technological gatekeepers in industrial districts. *Industrial and Corporate Change*, *21*(2), 429–462. https://doi.org/10.1093/icc/dtr053

Pinkse, J., Vernay, A. L., & D'Ippolito, B. (2018). An organisational perspective on the cluster paradox: Exploring how members of a cluster manage the tension between continuity and renewal. *Research Policy*, *47*(3), 674–685. https://doi.org/10.1016/j.respol.2018.02.002

Porter, M. E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, *76*(6), 77–90. https://doi.org/10.1042/BJ20111451

RAIS. (2020). *Relação Anual de Informações Sociais*. http://www.rais.gov.br/sitio/index.jsf

Schmidt, V. K., Santos, D. A. G. dos, Zen, A. C., Bittencourt, B. A., & Belussi, F. (2020). Trajectory Dependence , Lock-In Effect , and Cluster Decline : A Case Study of the Footwear Cluster in Sinos-Paranhana Valley. *Latin American Business Review*, *21*(4), 1–21. https://doi.org/10.1080/10978526.2020.1770607

Suire, R., & Vicente, J. (2014). Clusters for life or life cycles of clusters. *Entrepreneurship & Regional Development*, *26*(1–2), 142–164.

Tsouri, M., & Pegoretti, G. (2020). Structure and resilience of local knowledge networks: the case of the ICT network in Trentino. *Industry and Innovation*, *00*(00), 1–20. https://doi.org/10.1080/13662716.2020.1775070

Vicente, J. (2018). *Economics of Clusters: A Brief History of Cluster Theories and Policy*. Palgrave macmillan. https://doi.org/10.1007%2F978-3-319-78870-8

Wal, A. L. J., & Boschma, R. A. (2011). Co-evolution of firms, industries and networks in space. *Regional Studies*, *45*(7), 919–933.

---

[i] Labels in the MCA for supplementary variables: production (S_P = Small Production; M_P = Medium Production; B_P = Big producion), tourism (N_T = No Tourism; NI_T = No but Intend to have Tourism activities soon; Y_T = Yes Tourism), geographical indication (N_GI = No Geographical Indication; Y_GI = Yes Geographical Indication).