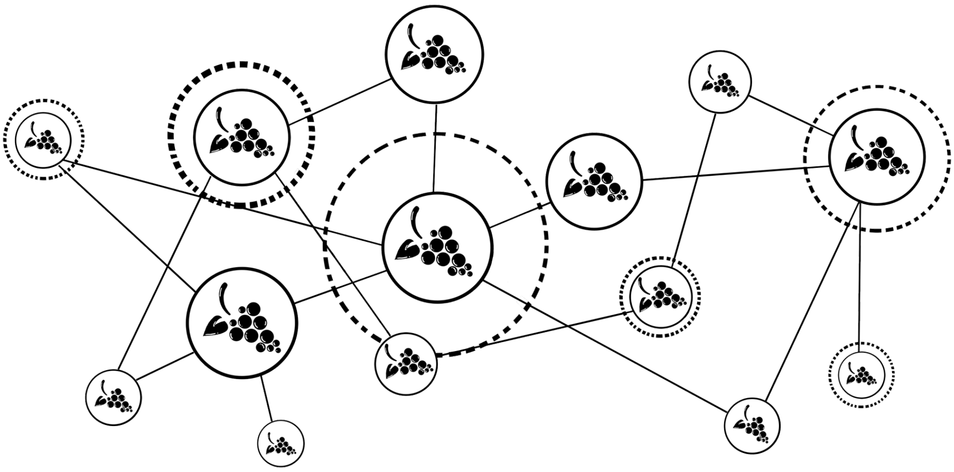
Copenhagen Business School

Local Knowledge, National Advantage?

The Relationship Between the Spatial Networks of German Vineyards and the Quality of Wine

Babette Sophia Bresser

Master Thesis
MSc. In Finance and Strategic Management



Supervisor: Steffen Blaschke
Department of Management, Society and Communication
Pages: 72
Characters: 154,477
Submission: September 4th, 2017

Abstract

In this study, the effect of the spatial network on quality is explored. Using the case of German wineries, the impact of their distances to each other as well as density and spatial cluster size is taken to assess if proximity increases the score these wineries attain on a national level. 60% percent of wine in Germany is sold through large discounters; small wineries are therefore increasingly exposed to the pressure of producing low-cost wine of decent quality. To escape these market dynamics, quality production plays a crucial role. As wine-making is very dependent on human expertise, it is intriguing to assess the factors influencing the quality of wine apart from geological conditions. A database consisting of 1,863 wineries, including their distances to each other and spatial factors, is developed and a network analysis is applied in order to assess if there is a relationship between the spatial network of wineries and quality. The results reveal some noteworthy dynamics: Wineries seem to thrive best in clusters with a limited number of close neighbors which suggests that proximity enforces the flow of expertise and thus quality. Interestingly however, more space and the presence of both close neighbors and actors further away seem to exhibit positive effects on quality. Thus, while the analysis finds support for the effect of proximity on quality, additional findings suggest that other dynamics play a role as well and offer potential for future research. The fact that spatial networks have an impact on quality indicates that local knowledge matters and offers individual wineries, as well as regions and national associations an incentive to promote communication in order to benefit from expertise embedded in local clusters.

Table of Contents

1. Introduction

In this study, the impact of spatial networks on success is tested taking the case of wine clusters in Germany. With more than 2,000 wineries spread across 13 wine regions, the German market can be considered very fragmented (Hanf et al. 2009). These mainly small businesses however, face a retail market dominated by a few large discounters, which control 60% of the market and thus exert considerable market power (CBI 2016). One way for small wineries to avoid the industry's low-cost pressures is to compete on quality instead of price; therefore, to know what impacts the quality of wine is crucial for firms in this field.

One dimension deemed to have a potential effect on quality is the geographical network of wineries, that is, the distance between them. The hypothesis is that wineries in denser networks are more likely to produce high quality, as the closer they are located to their neighbors, the more likely is the existence of social ties, trust and ultimately knowledge sharing, which in turn are argued to increase expertise and thereby the quality produced. While quality can be said to be a very subjective matter, awards received by wineries for their performance offer important indications for consumers and can therefore be used as a proxy for quality.

Thus, by building a network, taking 1863 wineries as nodes and the geographical distances between them as ties, winners of the "German Agricultural Society" award are identified to assess the impact of proximity on quality (DLG 2017a). The spatial network considered in this paper consists of 113 clusters, defined by zip code areas, of which intra-cluster distances, number of wineries and spatial dimensions are assessed by the use of statistical models.

1.1 Purpose and Contribution

It is the aim of this study to investigate if spatial proximity has an effect on the quality of wine as defined by the winning of awards in the industry. The specific case of German wine clusters is an example of an agricultural agglomeration of businesses defined as a "geographically proximate group of interconnected companies […] linked by commonalities and complementarities" (Porter 2000, p.4). Being close to other firms is argued to encourage valuable communication (Audretsch & Feldman 1996); spatial proximity and density of clusters are thus used as proxies for social networks. Knowledge sharing is further argued to impact learning and thereby performance, which in the wine industry can be considered to be dependent on the quality produced (Boschma 2005). By studying whether distance matters or not, this paper wants to offer new insights in terms of the relation between economic geography and social networks, as well as new knowledge regarding the German wine industry and the importance of spatial networks. If there is an effect, then, it can be hypothesized, that local knowledge in fact contributes to regional and national advantages for the wine industry and should be actively promoted.

This study is important as it helps to gain insights into the German wine sector, which in present literature has received little attention compared to countries such as France and Italy (see section 2.4, pp.20, Table 2.1). The results of this study can increase the understanding of success in the industry and how actors within it can utilize network links to improve performance. The sector becomes increasingly global, one example being "flying winemakers" (Giuliani 2007a, p.146) who export their knowledge to new wine countries such as Chile and Australia. However, while international collaboration increases, so does competition. New, less known wines from countries such as South Africa and Chile are increasingly getting a foothold in the German market, as do wines from Spain and France (Euromonitor 2017). Further, the national landscape is still highly fragmented with many small and local players torn between the country's established discount culture and the trend towards premium wine (Euromonitor 2017). Considering this competitive situation, both nationally and globally, an understanding of the sources of competitive advantages is crucial. If the study finds that spatial proximity enhances quality, then clusters and regions can make use of these findings by actively enforcing communication within and between clusters. On an international level, this could help the national industry to perform better. While this thesis takes the case of Germany, both the here developed method and potential conclusions can be extended to other countries and similar industries and thereby provide new insights on the link between spatial networks and performance.
More generally, wine is an interesting product category in terms of its relation to geography and performance. While its origin is spatially fixed, the final product is highly dependent on human activity (Gergaud & Ginsburgh 2008; Butler & Hansen 1991; Hira & Swartz 2014). As wineries are embedded in networks shaped by space and social coordination in clusters, their decisions depend on each other. This thesis thus wants to contribute new empirical evidence to this interdependence and offers insights into competitive advantages and possible explanations for the inner workings of wine clusters.

1.2 Scientific Contribution

This study adds value to existing research on multiple grounds. Considering the quantitative nature of the dataset, it contributes a spatial network theory approach to the field of research on wine. Founded in 2006, the American Association of Wine Economics publishes the "Journal of Wine Economics", a platform for peer-reviewed research on wine in order to promote the exchange of valuable ideas and research (AAWE 2017). While multiple network studies are published on this platform, such as Giuliani's research (2007a, 2010, 2013), the majority uses qualitative methodologies such as interviews, to assess the social relations of wineries. Quantitative analyses addressing the relationship between proximity and performance are scarce at best (Yang et al. 2012). Further, academic research on the German wine industry has so far lacked behind other wine producing countries. While this hypothetically might be due to the perceived insignificance of German wine, national industry differences and possible language barriers, this study adds to this scarce field of research focus.

In terms of spatial proximity and social networks, researchers such as Boschma (2005) and Oliver & Ebers (1998) criticize how little is known about the actual factors that influence performance in clusters. They call for empirical research on which dimensions and types of proximity impact clusters and how. This paper thus responds to this and contributes by providing a spatial network approach. It does so, by focusing on the geographic micro-level of clusters using actual quantitative data, that is, kilometers between the actors and density measures. While qualitative data provides a nuanced picture, it is also very subjective; a drawback that this thesis' database by nature does not exhibit.

This study also provides an assessment of clusters' competitive factors. As globalization has increased communication across borders, ultimately all industries have to address these immense opportunities but also the challenges that come along with it (Friedman 2005). In such a "flat" world (p.2) however, clusters such as Silicon Valley exist and thrive (Kenney 2000). The need of geographic proximity is thus not irrelevant and by assessing spatial networks' effect on quality, this paper gives new insights into the advantage of agglomerations. Lastly, the database was specifically created and thoroughly cleaned for this study and is believed to be the first combining German wineries, their geographical distances and awards. Apart from its originality, the database can be further extended for future research, both on the German market and other national settings.

1.3 Structure and Content

In this paper, the hypotheses to assess the effect of spatial networks on knowledge spillovers are derived based on existing literature, which includes the study of network theory, economic geography and organizational ecology. The case of this paper, the German wine sector, is an agricultural industry and based on traditions, climate and the national setting, exhibits specific dynamics. Thus, to provide a better understanding, the industry is explained in terms of history, structure and classification system. Following, the research design, including definitions, data selection and collection and the statistical model are explained. A coherent multiple regression is used with the aim to assess the hypotheses and answer the research question. More specifically, the regression aims at showing how different spatial dynamics, such as mean distances to others and density within a cluster impact the ranks as published by the German Agricultural Society (DLG). Besides the data on ranks, winery-specific characteristics provided by the German Wine Institute (DWI) are included in the dataset. Overall, the concept of networks is applied on a new, specifically for this study constructed, dataset with focus on how cluster-effects influence the score of individual wineries in Germany. Lastly, the study ends with a discussion and conclusion, including limitations and potential future research which can built upon the findings of this paper.

2. Literature Review

This thesis explores the connection between networks and geography and has the aim to assess whether the spatial network of clusters has a significant impact on the quality wineries produce. It is suggested that the closer they are to each other, the more they communicate and gain new knowledge that can be used in production. The research area is thus composed of both social and spatial concepts, which are laid out in this review. This section also shows that, while network theory and geography have been extensively studied by academia, this thesis adds value by using measurements which to the best of knowledge have yet not been applied to the wine industry in Germany.

While academia acknowledges that there exists a link between networks and geography, they nevertheless stem from two different and very diverse research areas. To structure these fragmented theoretical concepts, Oliver & Ebers' (1998) classification of organizational network literature serves as basis. In their paper, they in fact do a network analysis on network analyses to assess if there are certain common concepts that could structure this “messy” (p.549) accumulation of research. As illustrated below (Figure 2.1), they suggest that there are several paradigms that constitute the cornerstones of network theory (p.568) of which the two extremes are the social networks perspective and the governance perspective. The first focuses on the effects of actors' positions and the structure of networks. The latter is concerned with members' attributes and the form and content of relationships within the institutional environment.

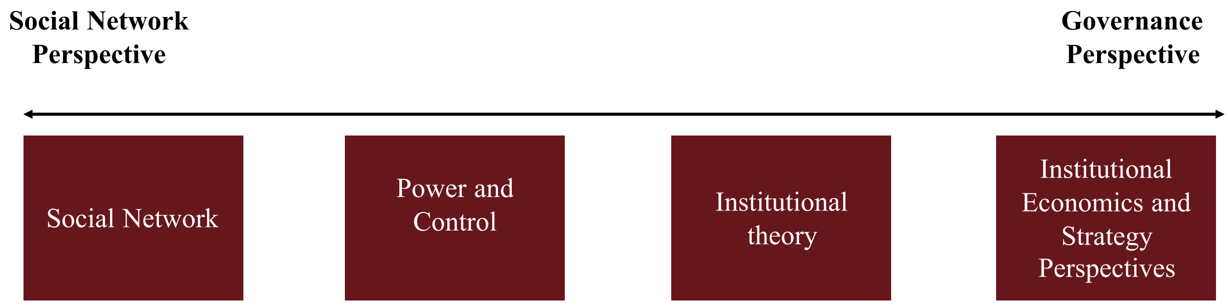


Figure 2.1**: Simplification of Oliver & Ebers (1998) Segmentation**

As their classification corresponds well to the organizational focus of this paper, previous research on network theory is explained within this context. To offer a well-round base for the theoretical understanding of this study, findings within the areas of network theory, economic geography and organizational ecology, including research specific to the wine industry are summarized. An overview of the relevant papers is provided at the end of the section (p.18;20;23).

2.1 Network Theory

Network theory, though often associated with social relations between people, spans far wider and exists in other contexts such as biology, physics and epidemiology (Barthélemy 2011). The common traits of all networks are their components: They are structures, containing nodes which are connected by ties (Borgatti & Halgin 2011). The nodes are the actors of the network while the ties are the connections or commonalities between them. As ties vary, networks exhibit specific structures in which actors take certain positions.

Taking this as central point, Borgatti & Halgin (2011) distinguish between "Network Theory" and the "Theory of Networks" (pp.1168). "Network Theory" aims at the effect of mechanisms of a network that results in certain outcomes. The "Theory of Networks" seeks to explain network structures, such as why people form ties with others or why not. While they focus on the social aspect of network theory, there is no consensus on whether networks are a theory or a methodology and it becomes increasingly hard to distinguish as the concept of networks is applied to a wide range of scientific areas (Borgatti & Halgin 2011). The study of networks is thus not only applied to dyadic relationships but also other entities such as organizations. As there is reliable proof that social networks influence economic outcomes, its study in this area has gained increasing attention in the last decades (Granovetter 2005).

Rooted in the field of sociology, network theory is built around central theories such as Granovetter's (1973) theory of “the strength of weak ties”, Burt’s (1992) “structural holes” and Watts’ (1999) definition of “small world networks” (Tichy & Fombrun 1979). Granovetter (1973), for instance, argues for the strength of weak ties is an important bridge between small-scale interactions and macro groups (p.1360). Based on the argument that people have strong ties with like-minded others, such as friends, weak ties, such as acquaintances, offer a broader range of new information. In other words, weak ties are a way to escape the problem of redundant information. The discussion of whether weak or strong ties are superior is still debated and empirical studies offer inconsistent results (Lin 1990; Watanabe 1988; Yanjie 1997). Ahuja (2000) makes the diplomatic suggestion that it ultimately depends on the objectives of the firm within the network.

The concept of “structural holes”, as developed by Burt (1992, pp.18) is based on Granovetter’s (1973) theory. These holes act as positions in networks which connect two clusters and thereby have influence on the knowledge transferred from one cluster to the other. For instance, in the context of wineries, firms with social bonds to other clusters would in Burt’s terms have a potentially favorable position as they can get diverse information from multiple clusters while controlling the flow of information between them.

Apart from ties and holes, worlds are also represented in network theory. Initiated by Watts (1999), “small world” networks describe very dense clusters, where the average path length between actors is short and the cluster coefficient is high. The average path length describes the average number of nodes one must cross to get to another node. In other words, the shorter the distance, the closer are the nodes. The cluster coefficient depicts how many of one node’s connections are connected to each other. Taking friendship networks as example, this would mean that one’s friends are most likely also friends with each other. The remarkable thing about these “small worlds” is their presence in very diverse settings. Researchers have found evidence for their existence in networks ranging from

Canadian investment banks (Baum et al. 2003) over power grids (Watts 1999) to American rap artists (Smith 2006). Therefore, network theory, though generally applied to social structures between people, firms and institutions, is a concept applicable to systems in general that consist of nodes and their connections, such as actual distances within geography.

2.2 Economic Geography

On the other side of Oliver & Ebers' (1998) theory scale is the governance perspective (Figure 2.1, p.8). This research area focuses on the formation of particularities of a network, in terms of actors' attributes and their relationships within the institutional environment (p.569). Regional conditions are such institutions (p.569) and these local environments are argued to be an important competitive factor of clusters which leads one to the study of economic geography (Maskell 2001). Parallel to research within network theory, the concept of economic geography has gained increasing attention. Its study and its relation to networks has been of academic interest since the introduction of the notion "industrial districts" in 1890 (Marshall, p.225). In the last three decades, research has extended this area to include relational factors and acknowledged the link between location and social dynamics (Maskell 2001; Bathelt & Glückler 2003; Markusen 1996).

The connection of economics and geography has to a large degree been coined by Marshall's "industrial districts" (1890, 1920). He differentiates in this context between reasons for formation and persistence of such agglomerations. According to him, the *foundation* of such districts is mainly based on physical conditions, such as soil and climate (Marshall 1890, p.223). Their *perseverance*, however, is due to the exchange of skills and knowledge which are "in the air" (1890, p.225) and learned "unconsciously" (1890, p.225) by the people within the district. While many physical conditions might not exhibit the same importance today as they have in Marshall's time, "industrial districts" can be considered to have set the stage for subsequent concepts on economies and networks. Porter's term of "clusters" (Porter 1990, 1998, 2000) is one such important keystone and also a famous application of the term to the wine industry (Porter 1990 p.86, 1998 pp.79). He argues that clusters' economic activities are "embedded" in relationships and that these informal linkages represent the "social glue" which ultimately makes knowledge and resources more accessible within the cluster (Porter 2000, p.19). The flow of knowledge has since then been the focus of several studies (Jaffe et al. 1993; Audretsch & Feldman 1996) and resulted in concepts such as "communities-of-practices" (Seely Brown & Duguid 1991), "buzz" (Storper & Venables 2003) and "epistemic communities" (Gittelman 2007). All these theories have in common that they see proximity and linkages themselves as an important source of competitive advantage for firms as they provide the mean to transfer knowledge.

Building on Marshall's and Porter's definitions, the effect of spatial proximity has subsequently been the focus of other research, recognizing the importance of regional institutions (Saxenian 1994; M. E. Porter 2000; Maskell 2001; Bell & Zaheer 2007), the effect of proximity on knowledge spillovers and performance (Jaffe et al. 1993; Audretsch & Feldman 1996; Balland 2012; Oerlemans et al. 2005; Arndt & Sternberg 2000), possible drawbacks such as lock-ins within the social communities (Asheim 1996; Asheim & Isaksen 2002; Boschma 2005; Giuliani & Bell 2005) and the special case of agriculture (Fanfani 1994; Murdoch 2000; Chiffoleau & Touzard 2014; Foster & Rosenzweig 1995).

Maskell (2001) for instance argues that institutions play a crucial role as they have evolved based on the requirements of the cluster. They are therefore very adapted to the specific environment and shape the way the cluster functions by encouraging certain activities and behavior. He suggests that there is an interdependence between the economic structure and the institutional environment.
Performance related, researchers such as Jaffe et al. (1993) prove that proximity has a positive effect on innovation in various industries. Balland (2012) adds to this research and finds that geographical, institutional and organizational proximity increases the likelihood of collaboration. Further, Audretsch & Feldman (1996) focus on the diffusion of new knowledge and find that innovation is more likely for firms in close spatial proximity to each other. Rosenthal & Strange (2003) find similar dynamics in urban space. Growth rates are also found to be affected by proximity, especially for smaller firms with up to 10 employees (Arndt & Sternberg 2000).

While many studies generally argue for the advantages of clusters, Asheim & Isaksen (2002) de-emphasize the importance of geographical proximity. They find, looking at four clusters in Norway, that linkages to firms outside the agglomoration have a significant impact on the innovation capabilities of the firms and observe this in very different clusters, from shipbuilding to technological agglomorations. Nevertheless, they also acknowledge that locally embedded contacts and knowledge play an important role, though they consider them not sufficient. Based on Granovetter's theory

(1985), Boschma (2005) follows this line of argument as he argues that clusters carry the risk of locking firms in an environment where they are provided with redundant information. Important, new knowlegde, he argues, comes from the outside. These findings are further supported by similar research, emphasizing the need of new information from inter-cluster relationships (Guerrieri & Pietrobelli 2004; Canina et al. 2005; Myles Shaver & Flyer 2000).

In the context of agriculture, the idea of applying Marshall's industrial districts has made its debut in Fanfani's (1994) study of Italian parmesan clusters. Agricultural food clusters, in his view, are characterized by a high degree of specialization and "artisan-like" processes (p.94). Furthermore, he finds that know-how, skills, tradition and cooperation are such local factors in these districts that they are hardly transferrable to other regions. In other words, the spatial factor is even more pronounced in rural districts than in Marshall's districts (1890). Foster & Rosenzweig (1995) find evidence in India where a farmer's proximity to experienced neighbors is found to increase his returns. The spatial factor of agricultural districts is also emphasized by Chiffoleau & Touzard (2014) and Murdoch (2000). The latter argues that it would be "naïve" (p.414) to think that new communication technologies make distance an obsolete factor for local network creation. Chiffoleau & Touzard (2014) emphasize that even though quality producers in agricultural clusters compete against each other, they are also in need of cooperation to deal with natural resources and potential complementary activities such as tourism. In general, the spatial factor of these clusters eventually forces actors to arrange themselves, because simply moving one's business is not a feasible option (Chiffoleau & Touzard 2014).

In traditional economic geography, firms located in clusters according to Porter's definition (1990; 1998; 2000) take an egocentric perspective with the general argument that agglomerations create a competitive advantage for the firms involved. From the central firms' perspective, other researchers have pointed out that advantages can stem from the co-location of actors as they produce common historical and geographical conventions, norms and institutions (Storper 1995; Barnes & Gertler 1999; Maskell 2001). Reduced transaction costs create accordingly the central competitive advantage as labor is more mobile and knowledge is more accessible (Maskell 2001). Economic geography, both for general clusters and rural agglomerations, is thus a field which spans multiple components of firms' environment within regional and national borders and emphasizes that location in fact does matter (Glückler 2007).

2.2.1 Relational Economic Geography

Assessing literature on network theory and economic geography, there is little doubt that firms are affected by both. In the last decades however, research on economic geography has begun to move from a more competition-based argument such as Porter's (1990; 1998; 2000) to a knowledge-based perspective.

In their paper "Toward a relational economic geography", Bathelt & Glückler (2003) argue that academia has undervalued the importance of relations in clusters by mainly focusing on economic equilibrium and trade advantages. They acknowledge that actors will choose the location which fits best their requirements, in other words location specificities make firms settle in an area. However, similar to Marshall's assessment of foundation and perseverance of clusters (1890), they emphasize that firms create their own environment by social engagement. The region becomes thus "socially constructed" (Bathelt and Glückler 2003, p.122). A follow-up study by Glückler (2007) supports this view and argues for an evolutionary network approach. He argues that path dependency and innovation in regions are dependent on the network of actors, both intra- and inter-regional. He again argues for a network-based approach within the area of economic geography, as networks are the construct through which norms, values and resources circulate in a region (p.13).

Informal, face-to-face communication is considered one of the characteristic components of clusters through which information can flow (Saxenian 1994; Storper & Venables 2003; Weterings & Boschma 2009; Duarte Alonso 2011; Rosenfeld 1997; Ganesan et al. 2005). Informal communication is theorized by Rosenfeld (1997) and Storper & Venables (2003) to play an important role for knowledge diffusion in clusters. Rosenfeld (1997) for instance, suggests that the highest performing clusters have the highest level of social capital. He backs his perspective by examples from various industries and calls for investments in face-to-face interactions. Storper & Venables (2003) built the theoretical argument that informal, personal meetings are especially important when knowledge is highly tacit.

According to relational economic geography, firms' advantages are hence embodied in the exchange of knowledge through relations in the cluster. *How* information flows has been the agenda of researchers in this field for the last decade. Instead of competitive advantages of individual firms, the information-sharing component thus comes into focus. Knowledge transfer within spatial dimensions has been found to be highly dependent on the industry, both rural (Yeung 1994; Amin & Roberts 2008) and urban (Jiaming et al. 2015; Rosenthal & Strange 2003), and on the size of firms involved (Sternberg & Arndt 2001; Freel 2003). For instance, Jiaming et al. (2015) and Rosenthal & Strange (2003) find that in urban areas, agglomerations are very dense in knowledge intensive industries and decrease after a few miles within a zip code area. They suggest that the need of knowledge spillovers is a reason for this density. Generally, Oerlemans et al. (2005) find that short distances favor interactions and exchange of tacit knowledge as spatial interaction literature suggests. He argues, referring to Marshall's industrial districts (1890), that informal relations, as well as intended meetings, create learning processes that need a certain degree of common values, norms and culture. These are created and evolve within regions and are hence locally embedded (Asheim 1994). A recent study by Stefano et al. (2016) tests why actors in networks voluntarily give valuable information to others and finds that in the context of Italian chefs, actors located close to other chefs would do so because they expect their counterpart to stick to the social norms of the area. The study therefore supports that spatial proximity facilitates knowledge flows via conformity pressures of the network.

Knowledge flows are however more complex than this (Breschi & Lissoni 2001). Multiple studies find that clusters not only facilitate information sharing, but keep it within certain boundaries. Asheim & Isaksen (2002) use the term "stickiness" (p.83) and judge it one of the reasons for the enduring importance of spatial proximity. "Stickiness" is considered the knowledge "partly embedded in local patterns of interactions" or "persons with first-hand experience" (p.86), in other words, it is knowledge which cannot be simply transmitted or communicated and presents "one of the few remaining genuinely localized phenomena" (Malmberg 1997 p.574). The exchange of this tacit knowledge is proven to be positively related to shorter distances (Torre & Gilly 1999); especially in industries where this type of knowledge is dominant (Cowan et al. 2004).

Agrawal et al. (2006) add a time variable to the discussion. Referring to the advantages of more distant network positions as laid out by Granovetter (1973) and Burt (1992), they find that knowledge flows between inventors increase by 50% when they have been co-located at some point. Their study thus draws an ambiguous picture: Proximity promotes social interactions, but once established, they survive without this proximity. Their explanation is that actors within the same industry have other options to communicate, such as trade fairs and conferences. Spatial proximity is thus not as important for intra-industrial linkages, they argue. The enduring social relations of individuals through time, such as inventors staying in touch, however can be considered a very specific case and not quite comparable to firms. Wal's study (2014) on bio-technology firms in Germany finds that spatial proximity is crucial for tie formation between inventors, but moves gradually towards triadic closure, that is, building of linkages with other inventors' partners, as knowledge needs move from a basic to a more specialized nature. This finding seems in line with Watt's (1999) theory of "small worlds".

Concluding, economic geography and its relational sub-category argue thus that proximity between firms is beneficial. The first however leans more towards whole clusters as unit of analysis, including the effect on performance and advantages of the firms involved. The latter takes a social network perspective, where the linkages between actors are considered the crucial edge that make clusters favorable systems for firms.

2.3 Organizational Ecology

Oliver & Ebers' (1998) study distinguishes thus between network theory and institutional environment within organizational network theory. Besides this classification, they also find that there are certain concepts which function as bridges between different theories. One such theory, which is relevant for this paper, is organizational ecology (p.558).

By definition, clusters can be said to be limited by geographical boundaries in which social networks play an important role. However, while network theory and economic geography deal with the interaction of actors and the emergence of clusters, this paper proposes that density also plays a crucial role. This falls in the sphere of organizational ecology, a concept which addresses the reasons for establishment and failures of populations of firms and their changes over time. The backbone of this concept is defined by Hannan & Freeman (1977) who make the important point that aggregates of firms are subjects to both selection and adaption processes, similar to the forces we see in biological ecosystems (pp.937). They explain the diversity of industry structures by the concept of niches and argue that firms in stable environments will be more likely to specialize as the probability of disruptive shifts is unlikely. Generalists however are not necessarily better off in uncertain environments. According to Hannan & Freeman (1977, pp.958), it depends on how long the "stable" periods in the environment are. If they are short, generalists would use too much time each period to adapt and specialists would thus also be favored in this environment. Glückler (2007) adds the concept of "network trajectory" (p.2) which states that networks are created due to evolutionary happenings, such as the formation and dissolution of ties. These ties in return are chosen based on selective competitive processes, where actors decide who to bond with as they cannot possibly be connected to everybody in a cluster. Firms therefore tend to build relations with firm's they judge to be of future importance for their own survival.

Hannan & Freeman (1977) and Glückler (2007) take hence the perspective of organizational ecology, which explains how populations gradually change by selection and adaption processes from within (Astley 1985). However, as highlighted in the literature review so far, firms are never isolated and interact with other actors. Similar to relational economic geography, a network-related view can be applied in this area. A stream of research building on this is community ecology. The concept can be considered a branch of organizational ecology but emphasizes that events shaping the ecosystem, are outside the scope of individual actors. The timing and shape of events are triggered by technical and institutional pressures and influence the ecological pressures on firms (Ruef 2000). Criticizing population ecology for being too slow and limited in their explanation of firms' ecosystems, community ecology takes populations themselves as units of change (Astley 1985).

Within this stream, researchers find that populations of firms are partially shaped by their social relations, as it limits resources that ultimately impact founding and failure of firms (Freeman & Audia 2006). Audia (2000) for instance finds that even though higher numbers of firms increase competition, clusters foster founding rates and sustain thereby the number of firms. He finds evidence in the American shoe industry and argues that entrepreneurs have access to social capital and knowledge and attain more confidence in clusters to start a business. Ingram & Inman (1996) also find that that increased competition is partially offset by institutions and a history of collective actions, as well as problems which firms face in a common location. More firms, they argue, do not mean that others are forced out of business. This is also in line with Baptista & Swann's (1998) findings in the British manufacturing industry. A more recent study by Diez-Vial (2011), taking the example of Iberian ham clusters, shows that the number of firms has a positive effect on other proximate firms. They reason that firms profit from skilled workers, knowledge spill-overs and lower transaction costs.

Baum & Mezias (1992) however, find opposing evidence in the Manhattan hotel industry. Their empirical findings show that the higher the density and the more alike actors are, the higher are the failure rates in the cluster. In terms of performance, Chung & Kalnins (2001) find similar evidence in the Texas lodging industry where agglomeration effects have a negative impact on firms of equal size in terms of performance.

Folta et al. (2006) bridge these opposing findings by arguing for an inverted u-shaped relationship between the performance of firms and their number in a cluster. They argue that the more firms, the more capital is available. However these benefits decrease at a certain point when competition effects take over. Fernhaber et al. (2007) find similar evidence in the ability of cluster firms with internationalization ventures. As research shows contrasting results, empirical evidence is needed to gain a better understanding of a specific industry.

The review on the theoretical concepts shows that networks, geography and density are interrelated based on their social nature. Oliver & Ebers (1998) lay out that network theory and institutional theory, which is argued to include economic geography, span this fragmented field of research. Networks take the social component as driver of relationships, while economic geography adds the institutional environment and attributes of actors to these networks. By combining organizational studies and spatially limited areas, networks and geography thus become evidently intertwined. By considering spatial borders however, density pressures are included in the dynamics of a cluster and population ecology therefore adds the necessary theoretical basis to understand these dynamics. Taking Hannan & Freeman's (1977) niche theory, a focus on high quality can be considered a crucial element for wineries as it relieves them of certain market pressures. Based on the literature, one should therefore expect that proximity has an influence on firms' performance, as does density within a cluster. In the following section, research on these dynamics within the wine industry is summarized.

Focus	Researcher	Theory	Methodology	Country/ Industry	Findings	Related/ Criticism/ Response
Literature review	Oliver & Eber (1998)	Theoretical	Network Analysis		Theoretical structure of organizational network literature	Barthélemy (2011), Borgatti & Halgin (2011)
Social Network Analysis	Granovetter (1973,1985,2005)	Strength of Weak Ties	Theoretical		Weak ties offer a way to attain new information and are of high value	Lin (1990), Watanabe (1988), Yanjie (1997), Tichy & Fombrun (1979), Ahuja (2000)
	Burt (1992)	Structural Holes	Literature-based		Conceptualization of structural holes as link between network clusters	
	Watts (1999)	Small Worlds		Social & biological, networks	Evidence of small world networks with short separation and high clustering factors	Baum et al. (2003), Smith (2006)
Organizational Ecology	Hannan & Freeman (1977)	Organizational Ecology			Density pressures regulate the number of firms within a spatial limit or industry	Astley (1985), Baum & Mezias (1992), Chung & Kalnins (2001), Canina et al. (2005), Audia (2000), Staber (1998), Audia & Freeman (2006)
	Folta (2006)	Organizational Ecology	Quantitative	U.S. Bio-tech. firms	There are increasing returns to cluster size but diseconomies of agglomeration increase by number of firms in a cluster	Fernhaber et al (2007) Jiaming et al. (2015)
	Ingram & Inman (1996)	Community Ecology	Theoretical	U.S. tourism industry	Competition enhances collective action	Babtista & Swann (1998), Arthur (1990), Diez-Vial (2011)
Industrial District	Marshall (1890)	Macro Economics	Equilibrium Analysis		Definition of Industrial Districts	Markusen (1996)
Competitive advantage of clusters	Porter (1998, 2000)	Economic Geography	Case studies	Wine, Leather	Conceptualization of clusters	
Knowledge transfer	Jaffe et al. (1993)	Economic Geography	Patent citation and geography	Academia and firms	Geographical proximity facilitates knowledge flows	Audretsch & Feldmann (1996), Agrawal et al. (2006), Wal (2014), Ruef (2000), Rosenthal & Strange (2003), Breschi & Lissoni (2011)
	Gittelmann (2000)	Epistemic Communities	Co-authorship and geography	Biotech. firms	Epistemic communities reach beyond cluster boundaries	
	Ganesan et al. (2005)	Social Network Theory	Survey	U.S. optics industry	Product innovation might be attributable to close relational ties instead of spatial proximity	
	Seely Brown & Duguid (1991)	Communities-of-practice	Literature-based		Informal communication is an important driver of learning and innovation within clusters	Amin & Roberts (2008)
Institutional focusdfg	Bell & Zaheer (2007)	Institutional Theory	Questionnaire and board membership	Canadian mutual funds	Institutional proximity is only significant when spatial proximity is given	
	Saxenian (1994)	Institutional Theory	Case study	Technolog y	Importance of social, institutional factors in terms of a cluster's competitive advantage	Barnes & Gertler (1999)
	Maskell (2001)	Institutional Theory	Theoretical		Institutions shape the growth and the inner workings of clusters	

Table 2.1: **Literature Overview.**

Focus	Researcher	Theory	Methodology	Country/ Industry	Findings	Related/ Criticism/Response
Informal relations	Rosenfeld (1997)	Economic Geography	Case studies	U.S., Italy	Clusters are local production systems embedded in social systems	Guerrieri & Pietrobelli (2004)
	Storper & Venables (2003)	Game theory	Theoretical		Face-to-face communication is important when knowledge not easily codified	Cowan et al. (2004)
Agriculture	Chiffoleau & Touzard (2014)	Network Theory	Survey	France	Advice networks are drivers of innovation	
	Foster & Rosenzweig (1995)	Economic Geography	Target - input model	Indian village	Having knowledgeable neighbors lead to higher profits and performance	
	Fanfani (1994)	Economic Geography	Case study	Parmesan industry	Conceptualization of agri-food clusters	Murdoch (2000)
Network position	Asheim (1994) Asheim & Isaksen (2002)	Economic Geography	Case studies	Norway	Innovative capabilities are enhanced by linkages to external firms	Malmberg (1997)
	Boschma (2005)	Network Theory	Literature-based		Proximity can create lock-ins, different types of proximity are needed in clusters	Myles Shaver & Flyer (2000)
Relational Economic Geography	Sternberg & Arndt (2001)	Resource-based Theory	Survey	European SMEs	Inter-firm relations affect innovation more than regional characteristics	
	Freel (2003)	Economic Geography, Network Theory	Survey	British SMEs	The importance of inter- and intra-relations of firms depend on size and innovation level	
	Oerlemans et al. (2005)	Resource-based Theory, Network Theory	Questionnaire	Dutch manufacturing firms	Inter- and intra-regional innovative ties impact firm performance	
	Stefano et al. (2016)	Economic Geography, Network Theory	Interviews	Italy	Compliance with norms is affected by spatial proximity	
	Torre & Gilly (1999)	Economic Geography, Network Theory	Survey		Both social, organizational proximity and spatial distance are important for firms	Balland (2012)
	Yeung (1994)	Network Theory	Literature-based		Compared to regulational and fordist perspectives, social networks contribute to a better understanding of cluster dynamics	
	Glückler (2007) Bathelt & Glückler (2003)	Relational Economic Geography	Literature-based		Conceptualization of relational theory within economic geography	

2.4 On Wine Clusters

As an agricultural good, geological conditions, such as soil and climate are literally, by nature, part of the quality of wine. While generally the environment thus plays a role, researchers such as Butler & Hansen (1991), Gergaud & Ginsburgh (2008) and Hira & Swartz (2014) argue that performance of wine clusters is also based on the social component. The latter two find evidence which suggests that the term terroir, the assessment of quality based on regional origin, only explains a small part of the perceived quality.

Taking this perspective, research on winery networks is therefore mainly focused on social connections between actors. The concept of clusters and its application to wineries is first famously assessed by Porter (1998; 2000) who establishes that wine agglomerations fit his definition of a "geographically proximate group of interconnected companies […] linked by commonalities and complementarities" (2000, p.4). Cluster formations have since then been found to be a very characteristic trait of the wine industry across different countries such as the U.S. (Butler & Hansen 1991; Taplin 1999; Benjamin et al. 1999; Tor Guthey 2008; Yang et al. 2012; Hira & Swartz 2014), Italy (Morrison & Rabellotti 2005), Britain (Turner 2010), Australia (Lockshin et al. 2005; Aylward 2006) and Chile (Giuliani & Bell 2005; Giuliani 2007a). A thorough search for similar and elaborate studies on the German wine market did not yield considerable findings. Countries differ in the structures of their respective wine industries and the industry is generally divided between new world and old world wines, where the latter stems from mainly European countries while new world wines are from South American and African countries, as well as Australia and the U.S (Migone & Howlett 2010). That wine grows in clusters can be considered a fact. But the inner workings of networks in these clusters is complex and thus the subject of various papers. Migone & Howlett (2010) compare wine clusters in different countries according to the nature of their networks, such as informal, organized and innovative social structures. Old world wines, such as France and Italy they argue, are characterized by a high level of cooperation which is partially rooted in the long tradition and history of the wine clusters. According to them, "clusters coalesce geographically, while networks coalesce ideationally" (p.2). They thus stress that networks and clusters are two different things where the first is embedded in the latter and that networks in wine clusters are crucial to success.

Content-wise, a row of social network studies looks at the dynamics shaping these clusters, such as collaboration (Tor Guthey 2008; Hira & Swartz 2014; Rasch 2008; Rasch & Gretzel 2008; Butler & Hansen 1991; Migone & Howlett 2010; Taplin 1999), informal ties (Brito 2006; Tor Guthey 2008;

Duarte Alonso 2011), knowledge flows (Morrison & Rabellotti 2009; Giuliani & Bell 2005), status (Benjamin et al. 1999) and quality perception (Benjamin et al. 1999; Yang et al. 2012) (see Table 2.2, pp.24).

Cluster performance is shown to be influenced by their collaborative networks, as comparative studies in various regions have shown. Butler & Hansen (1991) compare Washington state wineries in 1988 to other regions with the same geological conditions. They explain the region's thriving wine industry by the pre-existing network of institutions and entrepreneurs; other clusters in the study lacked this vital component and therefore never thrived as wine regions. A study by Hira & Swartz (2014) uses quantitative county data to assess if the social aspect can explain why Napa Valley in California performs better than other regions with the same geological zone. By taking the number of associations which require close collective actions of the wineries within the cluster, they find that Napa Valley outperforms other regions by far. Similar studies by Rasch (2008) and Rasch & Gretzel (2008) find the same dynamics in their study. Taking the case of wine tourism in Texas, they suggest that the region's weak performance is based on the lack of collaboration between the wineries which is partially due to the very large distances between wineries. While acknowledging that nature, political institutions and climate play a role, collaboration within a region seems thus to have a considerable effect on price and quality.

Studies on the social aspects of clusters are to a great number of qualitative nature such as Duarte Alonso's (2011) study of Southern U.S. wine clusters. By conducting interviews, he finds that collaboration in wine clusters is not only seen as economic means to an end, but that actors see their neighbors as friends. Similar studies in California and Portugal find the same dynamics, where social objectives seem to influence actors' actions to a large degree (Brito 2006; Tor Guthey 2008). While numerous papers consider intra-cluster networks crucial for wine clusters (Migone & Howlett 2010; Duarte Alonso 2011; Brito 2006), Turner's (2010) study points at the contrary. Conducting a qualitative study of English wineries, he finds that their performance is enhanced by contractual relationships with wineries *outside* the cluster. This research is in line with the mentioned finding by Asheim & Isaksen's (2002) that inter-cluster relationships are important providers of new information for clusters. However, high-quality viticulture has only existed in the UK since the early 1990's (Skelton 2008), which might add different dynamics compared to old wine countries such as Germany. An industry having "grown up" in the age of new communication technologies might exhibit other strategies and might lack the historical heritage of traditional wine countries. Nevertheless, the findings reflect that the relation between geography and networks in wine clusters is not clear-cut.

Taking knowledge flows within networks as research subject, Giuliani finds in multiple studies that wine clusters exhibit “small world” traits (Giuliani & Bell 2005; Giuliani 2007a; Giuliani & Matta 2013; Giuliani & Arza 2009; Giuliani 2013; Watts 1999). For instance, taking Italian and Chilean wine clusters as subjects, she examines their knowledge networks and finds that linkages are firstly, very dense and secondly chosen on competition based considerations (2007, 2013). Results suggests that these ties are based on how heterogeneous firms’ knowledge is perceived to be by others. Morrison & Rabellotti (2005) find similar evidence by conducting a qualitative study in Italian wine clusters. According to them, knowledge is a “club good” (p.999) exchanged between members in a core group which pushes other firms to the network periphery of the cluster. This seems also in line with Glückler's (2007) argument, that firms form ties with others based on strategic considerations. It follows thus that firstly, clusters can be very fragmented in terms of network ties, and secondly, that this fragmentation exists because some wineries are considered to produce higher quality than others within the same cluster.

Research finds thus evidence for the importance of networks. Nevertheless, regions play a role as they are marketing devices for producers and quality indications for consumers. Related to this is Benjamin et al.'s (1999) research on Californian wineries and their affiliations to certain regions over a 10-year period. They find that associations to certain regions, which are visible on the bottles, give wineries a certain status and have a significant effect on the quality ratings they get. This does not diminish the impact of vintners’ expertise but it suggests that simple affiliation can influence the perceived quality. The status of regions, at least in the American market, can be of importance for the wineries involved.

Another important factor is that wine is an experience good, which means that to assess its quality, consumers must taste it. As this is not easily feasible in normal supermarkets, labels on the bottle function as indications for the potential consumer and therefore play an important role in the wine industry (Lockshin et al. 2005). That such attributes and especially awards matter is based on empirical findings (Lockshin et al. 2005; Orth & Krska 2002; Sáenz-Navajas et al. 2013). Studying the behavior of Burgundy wine consumers, Sáenz-Navajas et al. (2013) for instance, prove that

awards are important indicators of high quality. Lockshin et al. (2005) differentiate this relationship even more as they find that awards are especially important for low and medium involvment consumers. Recognition is thus to a large degree based on the information consumers attain on the bottles, and as sales can be said to drive the return for wineries, these are highly important.
Considering the large amount of social network studies on wine, Yang et al.'s (2012) paper seems to be one of the few studies relating the actual intra-cluster distances to quality perception. Measuring the price effect of wineries in relation to their proximity to others in California and Washington state, they find that neighbors to high-reputation wineries achieve higher prices. They suggest that this clustering effect can stem from shared terroir or from knowledge spillovers.

Overall studies on wine clusters have mostly focused on the social networks within the clusters, establishing that they exist and play a crucial role. Both networks and institutional aspects shape hence the dynamics within a cluster. Research on the actual distances and population ecology is however very scarce or not the focus of any paper. This thesis therefore builds on existing literature and wishes to contribute a quantitative study of intra-cluster distances and their effect on quality.

Focus	Researcher	Theory	Methodology	Country	Findings
Informal networks & collaboration	Migone & Howlett (2010)	Network Analysis	Literature-based	Global	Across countries, certain cluster types are recurring
	Butler & Hansen (1991)	Network Analysis	Interviews	U.S.	Both broad social and inter-organizational networks are crucial for a cluster's success
	Hira & Swartz (2014)	Network Analysis	Quantitative	U.S.	Social linkages and entrepreneurship were significant for Napa Valley's success
	Rasch (2008), Rasch & Gretzel (2008)	Network Analysis	Interviews	U.S.	Weak performance is based on lack of collaborative spirit
	Turner (2010)	Economic Geography, Network Analysis	Interviews	United Kingdom	Informal linkages are replaced over time by more contractual forms of relationships
	Butler & Hansen (1991)	Network Analysis	Interviews	U.S.	Both broad social and inter-organizational networks are crucial for a cluster's success
	Duarte Alonso (2011)	Network Analysis	Interviews	U.S.	Social capital, especially friendship is of value to vintners
	Brito (2006)	Network Analysis	Interviews	Portugal	Non-economic driver play a crucial role, such as social and political factors
	Tor Guthey (2008)	Network Analysis	Interviews	U.S.	Wine clusters competiveness depends on local social network
Knowledge transfer	Giuliani (2007,2010, 2013), Co-author: Bell (2005), Arza (2009), Matta (2013)	Network Analysis	Interviews, Case Studies	Italy, Chile	Knowledge networks impact cluster performance, sub-clusters of knowledge exist
	Morrison & Rabellotti (2005)	Network Analysis	Interviews	Italy	Knowledge is a club good, only available to selected wineries
	Taplin (1999)	Network Analysis	Interviews	U.S.	Cooperative activity is time sensitive
Spatial proximity, quality & performance	Gergaud & Ginsburgh (2008)		Quantitative	France	Terroir has only a small effect, technological actions by the vintner affect quality
	Yang et al. (2012)	Economic Geography	Quantitative	U.S.	Wineries located near high performers can charge higher prices
	Benjamin et al. (1999)	Economic Geography, Network Analysis	Quantitative	U.S.	Actors' status position and their affiliations impact the quality they produce
	Lockshin et al. (2005)	Consumer Behavior	Quantitative	Australia	Extrinsic cues' value on bottles depends on the involvement level of the consumers
	Aylward (2006)	Organizational Ecology	Qualitative	Australia	Australian wineries are torn between the image of "Australian" wine and regional marketing
	Orth & Krska (2002)	Consumer Behavior	Qualitative	Czech Republic	Wine exhibition awards are crucial and affect consumer preferences
	Sáenz-Navajas et al. (2013)	Consumer Behavior	Qualitative	France	Trade-off in quality perception exist depending on different labels such as origin, labels, identity of producer and the presence of awards

Table 2.2: **Literature Overview on Wine**

3. Theory and Hypothesis Development

3.1 Research Question and Hypotheses

Based on findings summarized in in the literature section (see section 2, pp.7), space matters as firms have historically clustered to benefit from the accumulation of knowledge in a spatially limited area. Studies have established that knowledge-spillovers exist and findings support the argument of clusters' advantages. But papers on the effect of intra-cluster proximity of firms in terms of performance is scarce; even more so in terms of agriculture and wine. Yang et al.'s (2012) study is one of the few relating performance, in their case price, to distances within wine clusters. They state that their study is the first to apply actual spatial distances to gain insights into the effect on performance of wineries and they call for further studies on this "largely unexplored area of research" (p.675). This paper thus follows their call with the objective to contribute new data with focus on the German wine industry. The research question and hypotheses are laid out in this section, followed by the theoretical argument based on the literature review. The overall argument of this paper is that the geographic network matters for knowledge transfer and thereby affects performance in terms of quality. The network is not of any social nature, but takes spatial proximity as a proxy for possible social connections between wineries. Neither does the analysis provide information on the nature of knowledge exchanged. Spatial proximity acts as a pipeline which can contain both informal and formal information. To sum up, the study wants to answer the following question by assessing three hypotheses.

Is there a relationship between the spatial networks of wineries in a cluster and the quality of wine?

Hypothesis 1

The closer wineries are geographically, the more likely is that knowledge flows between them and, on average, the higher is the score these wineries receive for the quality of wine they produce.

Hypothesis 2

The average score of wineries will, up to a certain point, be positively related to the number of wineries in a cluster and will thereafter exhibit a negative relationship.

Hypothesis 3

The score of wineries will be negatively related to the spatial size of the cluster.

3.2 Theoretical Argument

This section presents the theoretical argument of the paper, which is based on previous studies highlighted in the literature section (see section 2, pp.7). The analysis is conducted using a database specifically created for this paper by merging information on German wineries, their rankings and spatial proximity to each other. The database offers analyses on two levels: vertices, which are the wineries, and the cluster, defined by zip code areas. With the ambition to offer a comprehensive analysis, the argument is divided into three hypotheses, specifying how geographical distance influences knowledge transfer taking proximity between nodes, clustering and spatial network size into account. A thorough search of relevant literature yielded no results in terms of the actual spatial distance between wine businesses and their effect on quality in Germany. This thesis therefore aims at providing a first step towards understanding this fundamental relationship.

As mentioned in the literature review (see section 2.2, pp.10), networks, spatial proximity and clusters have been researched by numerous studies taking both formal and informal networks, inter- and extra-cluster relations and their effects on innovation into account (Table 2.1). One key term within the sphere of geographical proximity is Porter's (2000) notion of clusters as "geographic concentrations of interconnected companies […] in a particular field" (p.15). Firms in clusters are therefore by definition spatially proximate to each other. Another concept is Marshall's "industrial district" (1890, pp.225), a network of small, co-located firms whose formation is the result of historical events and which share a mutual understanding and culture. Taking these classic definitions, the literature review has laid out the findings of different aspects of clusters and this will be used to argue that spatial distance matters.

3.2.1 Intra-cluster Proximity and Knowledge Transfer

As highlighted in the literature review, geographical proximity affects knowledge flows and performance (see section 2, pp.7). Studies by Jaffe et al. (1993), Audretsch & Feldmann (1996) and Balland (2012) show that the closer firms are located geographically, the more knowledge flows between them and eventually the more likely are these firms to innovate. Further, Oerlemans et al. (2005) find a positive relationship between spatial proximity and performance.

Having established that spatial proximity, knowledge diffusion and performance are positively related, other studies have specified that geographical networks' importance is very contextual in terms of size of the firms (Freel 2003; Karlsson & Olsson 1998) and type of knowledge (Antonio

Belso-Martínez & Morrison 2015). For instance, size related, Freel (2003) discovers that smaller firms in mature industries are more locally embedded due to resource constraints. How knowledge flows within clusters has been further considered by concepts such as “local buzz” (Storper & Venables 2003) and “communities of practices” (Seely Brown & Duguid 1991). As mentioned in the literature review, both terms refer to clusters as being characterized by face-to-face interactions, shared culture, knowledge and values, which thus form the competitive advantage of clusters (e.g. Storper & Venables 2003; Duarte Alonso 2011, see section 2, pp.7). Common problems and mutual experiences only reinforce these values and increase the expertise of the actors. To conclude, under certain conditions, spatial proximity has been found to influence knowledge dispersion, innovation and performance.

Research also finds considerable differences in terms of the type of industry (Bell & Zaheer 2007; Ganesan et al. 2005; Sternberg & Arndt 2001) and it therefore makes sense to take an industry specific perspective. First, wine is an agricultural good, hence it is by definition limited to the location where the grapes are grown. Contrary to technological clusters, such as the Silicon Valley, wine clusters developed primarily due to the right climatic condition for the grapes. One would therefore expect quality of wine and geological conditions to be positively related (Migone & Howlett 2010; Hira & Swartz 2014). However, based on findings by Gergaud & Ginsburgh (2008), Butler & Hansen (1991) and Hira & Swartz (2014) as discussed in the review section, these factors do not tell the whole story. They find that social networks and the embedded capabilities are crucial traits of wine clusters. This seems in line with Marshall’s (1890) view that clusters are founded due to geographical reasons, but survive based on the advantages that firms create for each other. Further, wineries are generally small establishments in a mature industry and based on Freel’s (2003) findings, this suggests that localities are essential. Taking these results into account, the following logic can be applied: If competing wineries are naturally located in the same climatic zones and if soil does not alone account for differences in quality, then the human factor must be of importance.

In fact, winemaking is a highly innovative and increasingly technological business. Grapes require treatments throughout the year, such as trimming, help for self-pollination in spring and the pressing of grapes after harvest. Several chemical steps are then taken whose specific timing is the vintners’ decision and ultimately have an impact on the quality of the final product (DWI 2017e). Hira & Swartz (2014) find that the available knowledge flows between wineries are responsible for the development and continuous success as reflected in higher scores at wine tastings. This is in line with

Sternberg & Arndt's (2001) finding that small- to medium-sized knowledge-intensive firms cooperate more locally. Supporting these findings, other studies have established that information-sharing, informal relations and the individual knowledge of winemakers within clusters is of big importance for the overall competitiveness of a region (Yeung 1994; Giuliani 2010; Taplin 1999; Tor Guthey 2008). Thus, as emphasized in the literature review, wineries' embeddedness in their regional clusters can be considered to constitute communities of practices, where owners exchange knowledge and act according to a shared set of values, rules and culture (Seely Brown & Duguid 1991; Giuliani & Bell 2005; Morrison & Rabellotti 2009, see section 2, pp.8).

Linking spatial proximity and performance, Yang et al. (2012) find, taking ratings from the Wine Spectator Magazine, that wineries in California can charge higher prices if they are neighbors to high-qualities producers. Possible explanations, they suggest, are knowledge and reputational spillovers (p.683). Foster & Rosenzweig (1995) prove the same in the context of farmers. The need to be close to each other is also discussed by Turner (2010), who applies Amin & Roberts' (2008) definition to wine, labeling it a craft- and task-based industry which exhibits the need for co-location and face-to-face communication. Rasch (2008) and Rasch & Gretzel (2008) suggest that the long distances between Texas wineries contribute to the region's weak performance in the wine tourism industry. All these studies thus establish a connection between geographical distance and performance.

To conclude, previous studies find that wineries' competitive advantages are a composition of their geographical embeddedness together with the knowledge sharing within the cluster. Hence, if wineries are embedded in geographical clusters and knowledge diffusion happens to a large degree within these spatially limited areas and this in turn influences the quality of the wine, then the logical consequence is that shorter distances of wineries will lead to more knowledge sharing and ultimately to better wine.

While this proposition considers spatial proximity an advantage for quality production of wine, there is a line of argument which challenges this perspective. The main argument is based on Burt's (1992) concept of structural holes. In terms of wine, this states that a winery, which is linked to several other clusters, might have access to diverse information which gives it superior knowledge compared to wineries with only intra-cluster relations. As stated in the literature review, research related to wine, such as Boschma's (2005) and Turner's (2010), finds evidence in favor of this theory (see section 2.4, pp.20). Because many of these studies refer to knowledge networks, it is intriguing to consider how longer distances in spatial networks might affect quality. Considering the connection of proximity and communication, being geographical more isolated might be favorable for a winery as it has the freedom to choose to be part of multiple networks without being locked-in by a community.

This argument raises a valid point, but the traditional and historical character of the wine industry is important to consider (Migone & Howlett 2010). 70% of wineries worldwide are still family-owned and managed (Woodfield 2012). While its connection to spatial proximity can only be hypothesized, the role of historical patterns can be argued to matter. Therefore, it is sensible to assume that wineries in their long history had to be embedded in local clusters to gain superior knowledge, especially considering that communication over long distances was more challenging and local support crucial. It thus seems more likely that path-dependency plays a role in the development of wine clusters and high quality wineries would therefore be expected to be in dense clusters.

To summarize, wineries are located in clusters and proximity is found to facilitate communication. Communication in turn, is crucial for information-sharing and knowledge exchange. This again enhances the expertise of wineries and increases product quality as the latter is to a large degree based on human know-how. Thus, being closer leads to more knowledge exchange, more knowledgeable production methods, and consequently to better quality. Based on this assumption, the following can be hypothesized:

> *H_1: The closer wineries are geographically, the more likely is it that knowledge flows between them and, on average, the higher is the score these wineries receive for the quality of wine they produce.*

In other words, wineries with shorter distances to others in a cluster are expected to produce higher quality compared to wineries located further away from others. The mean distance of each individual winery to all other wineries within the cluster is taken as proxy to assess the effect of spatial relationships between the nodes.

Related to economic geography, another interesting factor is the relation between quality and cluster size. There are two ways to look at it: Size can either be the number of nodes within a cluster or the actual spatial dimension of it. Taking each of these perspectives, would one expect award winners' presence in clusters to decrease with increasing cluster size or are high quality wineries on average in large clusters? The following two hypotheses assess each of these dimensions.

3.2.2 Clustering

Considering the first, the number of firms is a result of their entries, growth and exits. This mechanism is addressed in the field of organizational ecology, as pioneered by Hannan & Freeman (1977) who argue for the adaptive and selective forces of firms in an ecosystem, as observed in nature. As explained in the organizational ecology review (pp.15), literature offers a contradictory view as some studies have found that the growth in number of firms has a negative impact on firm performance due to unintentional knowledge spillovers and competition (Myles Shaver & Flyer 2000; Pouder & John 1996; Staber 1998; Wai & Yeung 2005; Baum & Mezias 1992) while others argue the opposite (Baptista & Swann 1998; Arthur 1990; Diez-Vial 2011; Freeman & Audia 2006; Stefano et al. 2016). Stefano et al. (2016) for instance argue in the case of rural clusters, that although competition exists, fixed natural and institutional factors make cooperation necessary. Considering the contradicting findings, it is questionable how clustering affects rankings. Basically, if knowledge transfer is important for quality, then the question is how knowledge flows are affected by the agglomeration of wineries.

This thesis bases its hypothesis on Folta et al.'s (2006) and Fernhaber et al.'s (2007) findings and argues for an inverted U-shaped relationship of density and performance. The more firms, the more knowledge is available, hence it could be argued that wineries have a larger pool to extract advice from. Further, based on basic mathematics, the higher the number of wineries, the higher the probability of having award-winners amongst them. This results in a positive relationship between award winners and number of wineries in a cluster. At some point however, with increasing density this advantage erodes as congestion costs increase. First, competition costs rise. Heightened demand of skilled labor will increase the prices of such. Second, coordination will be harder to implement. For instance, with more actors, decision-making within the local community will be more difficult and some wineries might become very passive in the cluster while others take leadership roles. This might lead to wrong decisions or social tensions and can affect important knowledge exchange. Third, density might affect the space of the individual slopes and smaller wineries might lack the capabilities or resources to enter competitions. An increase in the number of wineries in a cluster might therefore lead to lower rankings. The argument is thus that the number of wineries has a positive effect on ranking in the cluster up to a certain point where cooperation and competition costs lead to a negative influence on knowledge exchange and thereby to decreased quality. While death and birth rates of wineries probably happen at a slow rate compared to technological clusters, the number of actors nevertheless is considered to be shaped by ecological dynamics. The second hypothesis is thus:

> *H_2: The average score of wineries will, up to a certain point, be positively related to the number of wineries in a cluster and will thereafter exhibit a negative relationship.*

The number of firms per zip code area accounts for density pressures in the cluster together with an interaction term considering the spatial dimension of the cluster.

3.2.3 Geographical Cluster Size

The third factor to account for is the size of the cluster. Naturally, a larger region will also have space for more wineries which in turn can exhibit the same dynamics as stated by H_2. When accounting for the number of wineries in a cluster, it is intriguing to see if actual distances affect quality.

Compared to the clustering of actors, actual research focusing on the spatial size of rural clusters is scarce but has been studied in the context of urban areas. As mentioned in the literature review (see section 2.2.1, pp.13), Jiaming et al. (2015) find that spatial patterns in the Beijing area vary greatly by industry, where knowledge- and labor-intensive industries are more likely to be closer to each other than large manufacturing firms. Rosenthal & Strange (2003) also find that localization effects, in their case employment within an industry, decreases rapidly within the first miles of a zip code center, but very slowly thereafter. Acknowledging that both studies use different industries, agglomerations and methodologies, they have in common that they emphasize the effect of geographical distance on firms. As highlighted in the literature review, proximity between actors affects performance and that has been proven in various industries (Jaffe et al. 1993; Audretsch & Feldman 1996; Balland 2012; Oerlemans et al. 2005) as well as in wine clusters (Giuliani 2007a; Giuliani 2013, see section 2, pp.5). Further, from an intuitive point of view, clusters' characteristic trait is their proximity within a limited spatial region. Thus, an increasing size of an area eventually moves away from this definition of cluster.

Wineries, being in agricultural clusters, need naturally more space than urban industries, but accounting for the number of actors in a cluster, it can be argued that longer distances in a cluster make it more difficult for actors to meet and less likely to have knowledge exchanges. The following can therefore be hypothesized:

> *H_3: The score of wineries will be negatively related to the spatial size of the cluster.*

The spatial cluster size is calculated by taking the sum of all the distances between the wineries in a cluster. As this is also correlated with the number of members, the analysis will account for the interaction between the size of the cluster and the number of wineries.

4. Background

The German wine industry is used as case to assess the effect of spatial networks on quality. This highly fragmented market with more than 2000 wineries in the 13 official wine regions is characterized by a long history and traditions (Hanf et al. 2009). This long history however has also contributed to the less well-known image of German quality wine. This section provides an overview of the industry and relates it back to the research question as to why geographical networks in this industry should be expected to have an effect on quality.

4.1 Historical Overview

The first traces of viticulture in Germany date back 2000 years, to the time when the Romans conquered the region with their heirs (DWI 2017c). In the Middle Ages, monasteries took the lead as wineries and became responsible for advances in German wine production and quality (Pfalz 2017). Europe flourished during that period due to a medieval warm period and the population, and with it the need of monasteries and churches, increased considerably (Wein-Plus 2017c). There was thus high demand for wine which additionally was a safer alternative to, often contaminated, water (Seidel et al. 1978; Wein-Plus 2017c). In order to meet this need, farmers started to grow wine in regions less suitable for wine such as the Baltic Sea and this eventually resulted in a decrease in quality of wine produced during that time (Seidel et al. 1978; Wein-Plus 2017c; German Mission in the U.S. 2017).

The 17th century then brought changes to Europe and its wine industry. First, the Thirty Years' War left the region in a disastrous state (German Mission in the U.S. 2017). Combined with increased imports of wine from France, the rise of beer as popular drink and colder climate, the wine industry experienced a significant decline in both production and quality (Seidel et al. 1978; Wein-Plus 2017c; DWI 2017c). Following the political changes triggered by the French Revolution, the feudal system ceased to exist and many lords as well as the church lost their possessions, including vineyards (Wein-Plus 2017b). Quality saw two opposing trends at this point: While some wineries were transferred to noblemen with focus on high quality wine, many smaller vineyards were now in the hands of farmers who lacked the financial means to strive for such quality (Seidel et al. 1978; Wein-Plus 2017c). The latter thus began to establish the first local cooperatives to facilitate the production of wine for the farmers (Wein-Plus 2017d).

In the following centuries, the German wine industry experienced multiple setbacks, which to a large part shape the industry until today. At the end of the 19th century, a parasite imported from America, destroyed large parts of the country's viticulture and many indigenous varieties became extinct in the course of the plague (DWI 2017c; Pfalz 2017). Thus until today, only wine on American phylloxera-resistant rootstocks is allowed to be cultivated in Germany (DWI 2017c). In the 20th century, after two world wars, recessions and an increasing influence of the United States, the national wine industry was marked by cheap, fordist-inspired mass-production and quality decreased constantly, taking the infamous "Liebfraumilch" as an example (Rössel & Beckert 2012; Seidel et al. 1978 p.14; Pfanner 2012). In the post-war period, especially with the emergence of the EU, the German viticulture became more structured. One such development was the German Wine Law of 1971 which had the objective to standardize the labeling of wine to facilitate the purchase for consumers (DWI 2017c). By this law, certain rootstocks and indications of origin had to be labeled (BGBl 1971). Furthermore, and often criticized, it collected smaller, individual vineyards to larger appellations (*Grosslagen*) with the goal to simplify marketing (BGBl 1971). Until today, high-quality wineries consider this law disastrous as it lets lower quality wineries free-ride on their efforts and reputation (VDP 2017b). While Germany follows EU legislations, it is still allowed to grade its wine by must weight, which takes the ripeness as indicator of quality, not the origin as is the EU's initially aim (Seidel et al. 1978). As recent as 2009 however, a new EU law moved the whole system towards a terroir-based classification (DWI 2009).

Historically, distance and thus the local social network can be considered to have been crucial elements of wineries' daily operations. Harder, long-distance travels made regions more dominant on a social level as close neighbors most likely represented the knowledge pool that one could extract expertise from. It is thus sensible to say that a certain dependence of wineries in their close spatial environment is rooted in the industry's history.

Quality-wise, societal changes, wars and influences from abroad made regions and origins to bad indicators of quality, as the country experienced periods where quality was not an essential part of wine consumption. This might also explain the distinctive quality system the country has developed, as will be explained in the classification section (see section 4.4, pp.37).

4.2 Structure & Industry

Today, there are approximately 35,000 wine businesses in Germany, of which 2,000 are wineries cultivating more than 10 hectares (Hanf et al. 2009). German viticulture is one of the most Nordic wine growing regions in the world and as such, wine is grown in concentrated areas of the country (Wein-Plus 2017c).

Figure *4.1* illustrates the geographical structure of German wineries. On the broadest level, Germany consists of 13 wine regions, which are mostly located along the Rhine and the Mosel rivers. After the reunification of Germany in 1990, two new regions were added, Sachsen and Saale-Unstrut (Wein-Plus 2017c). With some exceptions due to size, 10 out of 13 regions are further divided into "Districts" (*Bereiche*) which are made up of "Collective Sites" (*Grosslagen*). These sites are groupings of several smaller individual wineries. The smallest entity are "Individual Sites" (*Einzellagen*) which can span from 1 ha to 200 ha (Wein-Plus 2017c).

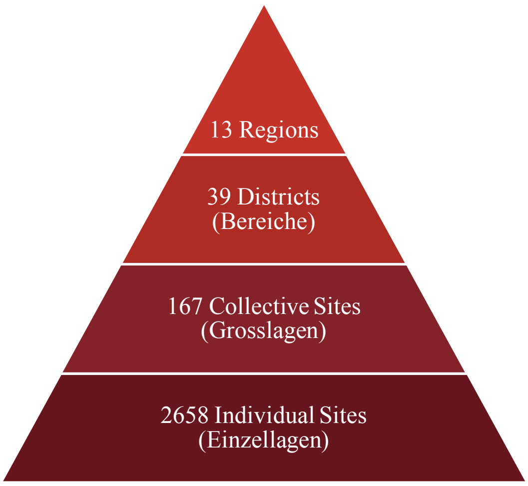


Figure 4.1: **Geographical Division of German wineries**. Source: DWI (2017)

Considering the historical image of German wine, national interest groups have formed, each promoting their own definition of quality. One such group is the "Association of German Prädikat Wine Estates" (hereafter VDP), which was founded in 1910 with the goal to promote German wine of high-quality (VDP 2017b). With the introduction of the German Wine Law of 1971, which clustered smaller vineyards into larger "Collective Sites", the VDP has isolated itself from the national classification system (VDP 2017b). According to the association, a consolidation of both

low- and high-quality wineries into larger “Collective Sites” would happen at their expense and undermine the incentive to produce good wine (VDP 2017b).
The VDP consists currently of 197 wineries which produce 3% of the German wine harvest (VDP 2017c). The members do not follow the German Wine Law but define their quality according to the French notion of terroir (VDP 2017a). While the VDP claims not to operate under the Law, it is important to note, that members still must pass the sensorial and chemical test, as all other German wines, in order to be accepted for sale. However, they choose to not promote their rank as indicated by the official system. Their classification can hence be considered an alternative addition to quality standards (Brunke et al. 2016). While the VDP’s influence was rather limited in the post-war period, a new, more individualized focus in the consumer market has given the association more legitimization in recent years, most likely also because of the generally high-status perception of the French system (Rössel & Beckert 2012).

On the other side of the producer scale are cooperatives of vineyards, whose members are owners at the same time (Wein 2017). They function as democratic “self-help organizations” where each member has one vote, regardless the size of the vineyard (Hanf & Iselborn 2014, p.6). Cooperatives combine their grapes at commonly owned wine-making facilities and market their wines under the name of the cooperative to cut costs. Today, the 169 German cooperatives produce one third of the country’s wine (Deutsche Winzergenossenschaften 2017). Despite their importance for sales in Germany, wine cooperatives suffer under a reputation of producing medium- to low-quality wine (Hanf & Iselborn 2014). Partially, this perception is based on structural problems within the organization: Cooperatives’ members use common facilities to produce the wine and in many cases, these facilities are obliged to accept all the grapes of the members. However, members, being principals at the same time, can often decide to keep the best grapes to themselves and sell them privately (Hanf & Iselborn 2014). Further, as the wine is marketed under the same name, there is no competition which would necessarily lead to higher quality (Hanf & Iselborn 2014). These factors thus result in a typical principal-agent problem and diminish the quality of the wine that consumers associate with cooperatives.

On the wholesaler side, the German Market shows some striking characteristics that seem to follow the historical lack of quality focus. Unlike other wine-producing countries, discounters play a big role in the German market (CBI 2016). Aldi, Lidl, Edeka and a few other discounters account for more than 60% of all wine sales when one excludes the gastronomy sector (CBI 2016). While this might also be contributed to the country's "money for value" culture, discounters are now also offering wines of higher quality, which can account for the strong market position (Euromonitor 2017). However, these discounters are known for harsh bargaining and the Ministry of Foreign Affairs advises potential foreign players to offer competitive prices in order to gain a foothold in the German market (CBI 2016). Cooperatives are thus under increased pressure of delivering larger amounts of acceptable quality (Hanf et al. 2009). To accommodate these demands, two central wine cooperatives exist in Germany which collect the wines from multiple cooperatives and produce them in two standardized facility. By centralizing the production of bulk wine, cooperatives can thus better meet the requirements of large discounters and secure their survival on the mature market (Hanf et al. 2009).

The German wine market can hence be considered to face very fragmenting dynamics. On one side, there are chains such as Aldi and Lidl, which exhibit huge market power and demand standardized quality of wine at a low price. Cooperatives and individual wineries must comply with them if they want to make it in the market. On the other side, the VDP officially rejects any association with the current classification system and takes the niche segment by marketing its wineries as the top-quality producers of the country. To provide an overview, Figure 4.2 shows a simplification of the market players excluding import, export and middlemen.

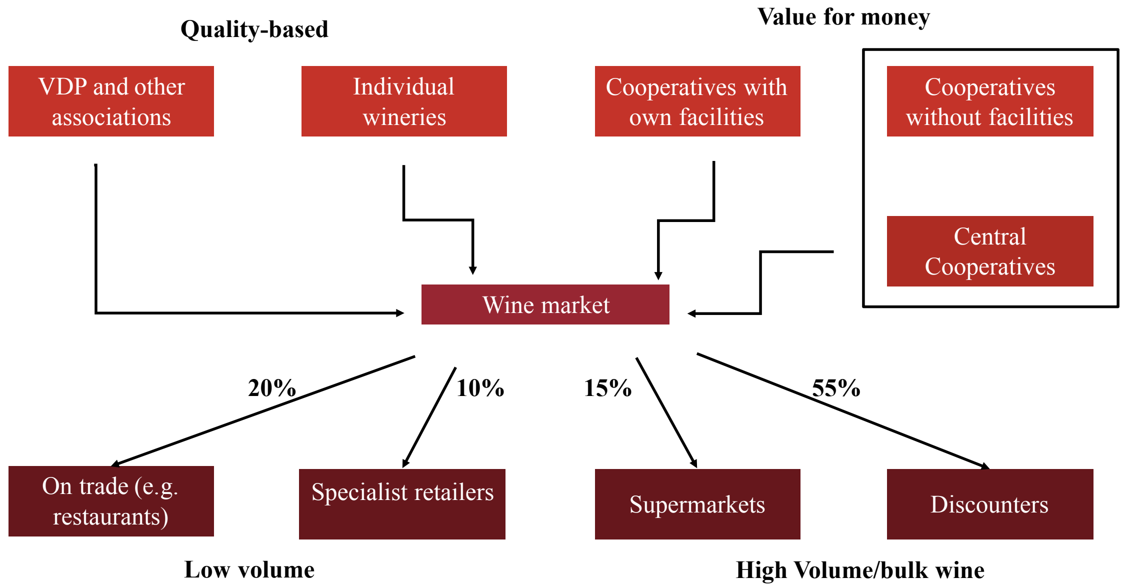


Figure 4.2: **Market flows in the German Wine industry.** The graph is a simplification of Hanf et al.'s (2004) illustration of the German wine market. The updated percentages stem from the CBI market report 2016 (CBI 2016).

4.3 Outlook

In terms of the industry's development, Euromonitor (2017) expects that Germany will see an increase in demand of high-quality wine even though the strong market power of discounters is expected to keep the price increase at a moderate level. The rise in popularity of high-end wines benefits regional wines in particular, as consumers trust the domestic quality standards more than foreign ones and look for more local wines. Whereas the same argument applies to domestic medium-quality wines, their strong sales can also be explained by their increased presence in supermarkets where every second bottle in Germany is currently bought (Euromonitor 2017). Thus, the continued requirement of cheap, but relatively high quality wine, suggest that the market will continue to be highly competitive. Based on consumer preferences and EU legislation, regions gain importance and this suggests that wineries, both individual sites, the VDP and cooperatives, will move towards a more local focus in terms of production and marketing.

4.4 German Quality Standards

Quality is a very vague concept. Firstly, because every individual can be considered to taste differently. Secondly, because quality is judged depending on the context, such as nation or product category. Nevertheless, quality assessments are important, especially in the wine industry, as many consumers judge the wine by labels on the bottle (Schamel & Anderson 2001). This section introduces the German classification system as it is unique compared to other old wine countries, such as France (Seidel et al. 1978). Further, as the analysis uses the DLG award with its specific definition of quality, an overview of the national definitions and awards are provided for a better understanding.

The classification system in Germany contains several levels and is therefore rather complicated. For once, as wine is an agricultural good, it falls under the "Common Agricultural Policy" as governed by EU law (USDA 2017; EC 2017). However, as long as wineries adhere to these rules, there are different classifications they can attain depending on country and even state. While France follows the system of terroir, so origin, Germany generally classifies its wine according to ripeness (Rössel & Beckert 2012). The scale of ripeness is also called the "must weight" and represents the natural amount of sugar in the grape juice (DWI 2017d). The more sunlight the grapes receive, the higher the amount of sugar and this again affects the final amount of alcohol, which is an important part of the classification in Germany (Brunke et al. 2016). Depending on the region, different "must weights" apply as more northern regions will have fewer sun hours and will accordingly have less sugar in the grapes juice (DWI 2017b). Figure 4.3 illustrates the current classification system which has been in place since the new EU regulation passed in 2009.

Figure 4.3: **Current German Classification System.** Source: DWI (2017)

The upper categories, "Quality Wine of Guaranteed Origin" and "Wine with Special Attributes" must be from one of the 13 regions and must have passed chemical tests as well as a certain must weight to be considered (DWI 2017d). "Wine with Special Attributes" is then further divided into six categories, which, true to the German model, are based on the grapes' ripeness (Rössel & Beckert 2012).

4.4.1 Quality in the Glass vs. Terroir

Germany has two distinctive classification systems, "Terroir" and "Quality in the Glass" (hereafter QiG) (DWI 2017a). Terroir is a term known from the French system which points at the origin of the wine and bases its classification more on geological factors such as soil and climate (Brunke et al. 2016). QiG, on the other hand, is a German concept and was incorporated in the German Wine Law since 1971. The classification is solely based on the chemical and sensorial composition of the wine. The must weight is part of this examination and measured in *Oechlse* (DWI 2017a). From a historical point of view, the QiG's perseverance is reasonable. After two World Wars and recessions, Germany embraced the mass-producing culture and the focus concerning wine was that it had to be drinkable (Rössel & Beckert 2012). The fact that this system survived can therefore be considered a child of its time. With the consolidation of smaller vineyards in 1971, origin became even less important and just reinforced the system in place, against the interest of the VDP (DWI 2017a; Rössel & Beckert 2012).

So far, Germany has successfully resisted EU legislation by only changing the terms used, but not the methods and QiG is thus still in place (Brunke et al. 2016).

4.4.2 Awards

As there are to systems in Germany, there are also awards which base their principles on one or the other, of which the two most important ones are explained in the section. The first is the award given by the "German Agricultural Society" (hereafter DLG), a non-profit organization whose main purpose is to "promote technical and scientific progress" (DLG 2017a). The DLG gives awards in several food and beverage categories on a yearly basis, where the wine category consists of rankings including "The Best Collections", "The Best Young Vintners" and "The Top Producers" (DLG 2017c). As testing and classification is a state responsibility in Germany, the DLG is the only nation-wide competition (Brunke et al. 2016; DLG 2017b)**.** Winners are selected on their current and former performance in several DLG quality tests. DLG quality tests are based on the association's five-point scale consisting of a blind tasting for chemical testing and a quality assessment by 150 wine experts (DLG 2017e). The winners of these tests are then awarded medals in bronze, silver and gold which they can advertise on their products (DLG 2017d). By using sensorial and chemical testing, the DLG thus follows the philosophy of QiG and largely ignores origin as a factor for quality.
In recent years, the DLG, and the wine award industry in general, have been criticized for not being selective enough (Bock 2016). In the state of Hessen for instance, only 9% of the wines in the competition did not receive an award in 2016. Based on a survey by the School of Geisenheim, many vintners find that wine competitions have become "inflationary" in the way they give out prizes (Bock 2016, p.1). However, based on a survey of 2,000 participants, DLG is still the most recognized label for wine by 60%, while other awards in Germany are considerably less well-known (Bock 2016).

Another influential indicator of quality in Germany is the Gault& Millau Guide. Originally, a guide to the high-end restaurants, the Gault& Millau today publishes its guide for multiple countries and has expanded to include wines (Wein-Plus 2017a). The Gault& Millau uses a 100-point scale to classify wines and the final guide offers short descriptions of approximately 1000 wines. Wineries are tested by being compared to others from the region, both blind tastings and open tests are used under the procedure. The competition then moves one level higher and the best collections from the 13 regions are compared (Gault Millau 2017). Contrary to the DLG award, the Gault& Millau bases its quality philosophy on terroir (Rössel & Beckert 2012). The guide thus has a strong affiliation with the VDP, as is also indicated by the fact that one of its members was a leading editor of the guide (Rössel & Beckert 2012).

4.5 Summary

As the history and the current structure of the German wine industry make clear, the country can be said to be divided when it comes to the definition of what good wine is. For once, rooted in history and culture, consumers are not very quality-minded which makes the two-classification system possible in the first place. Second, for each definition of quality, there are large support bodies in place which exhibit market power. QiG is still widely used by vintners and the influential DLG. "Must weight" is a known term in Germany and the fact that regions have historically not been indicators of quality, this classification method stands on a stable basis. On the other hand, terroir is used by the VDP and marketed as the sophisticated niche of German Wines, which again is supported by the Gault& Millau, promoting itself in a similar manner as a gourmet guide. While EU Legislation wants to move the classification system towards the French model, the QiG system can be said to be part of the nation's DNA and has therefore a strong foothold in the industry. As both EU law and the consumer preferences show, regional wine gains increasingly attention which suggests that local differences embedded in traditions and specific know-how will be an important advantage in this competitive market. In order to consider the country-specific system and the German consumer market, the DLG award is judged a suitable proxy for quality and its rating given to wineries is thus the dependent variable used in the analysis of this paper.

5. Methodology

The previous sections have explained the research question and its importance as well as relevant background knowledge regarding the case. After having explained the *what* and *why* of this study, this section wants to explain the *how*, in other words, the research design used to answer the research question. It is the aim of this study to assess if the spatial network of wineries has an effect on the quality they produce taking the rankings of wineries as a proxy. As emphasized in the literature review, proximity fosters communication and innovation and thereby performance (Storper & Venables 2003; Rosenfeld 1997; Foster & Rosenzweig 1995; Jaffe et al. 1993, see section 2, pp.8). As cluster firms are naturally close to each other but research on detailed intra-cluster distances and their effect is scarce, this thesis aims at assessing if wineries in dense spatial networks on average produce wine of higher quality. To analyze this relationship, the model and variables chosen will address the network dynamics of the hypotheses as stated in section 3.1 (pp.25).

Firstly, the proximity between the wineries is reflected using the mean distances of each winery to the others in a cluster. Secondly, the clustering effect is accounted for by including the number of wineries. Thirdly, the sum of distances is taken to consider the spatial size of the cluster. Lastly, standard deviation, varieties, region and awards per cluster are added to consider possible other effects on score. This section provides important definitions and a justification of the data selected by considering possible other network linkages. The data collection is then described, followed by an explanation of the statistical model and the variables.

5.1 Definitions

5.1.1 Wineries

While the terms "vineyard" and "winery" are generally used in the same context, "winery" is intentionally used for this thesis. The reason for this is the research design of the paper. According to Merriam-Webster's dictionary, a winery is a "wine-making establishment", while a vineyard is defined as the "sphere of activity" where grapes are grown (Merriam-Webster 2017). Often, these terms can be said to be synonymously, as a wine manufacturer often grows the grapes and makes the wine on the same grounds. However, as the award winners in the research sample also include cooperatives, the terms have different meanings. Members of cooperatives grow their own grapes on their respective vineyards but the actual making of wine happens in a common wine-making facility, in this context the winery. As awards attained by cooperatives are given to the winery and not individual vineyards, and as this thesis focuses on the value created by human hand and not the soil, the term winery is used.

5.1.2 Quality

Taste is subjective and any quality definition can therefore only serve as a proxy. As described in section 4.4 (pp.38), EU legislation provides the basis of quality testing, after which it is the country's responsibility to classify the wine (Rössel & Beckert 2012). A definition of quality can thus be approached from two sides. On the one side is the consumers' perspective, which can be said to be represented by sales numbers. As their perception ultimately affects sales, it can be argued to drive the ambition of producers. Considering however that every second wine in Germany is bought in supermarkets, it is questionable how well this measure would assess quality (Euromonitor 2017). To single out more sophisticated consumers, poses however problems in terms of reliable data. On the other side of the supply chain is the producers' perception of quality, visible by labels and appellations on the bottle. These labels however are subjective, diverse and often dependent on grape ripeness (DWI 2017d). They are evidently biased; hence this measure is not optimal to assess quality.

Quality perception is therefore widely dispersed. The common ground however are guides and awards which serve as aspiration for producers and orientation for consumers. In Germany, there are several guides, such as the DLG, Gault& Millau and Wein Plus which each their distinctive classification system and philosophy (DLG 2017b; Gault Millau 2017; Wein-Plus 2017c). For this thesis, quality is thus judged to be well represented by the DLG awards based on multiple factors. Firstly, every winner of the competition has passed the national minimum standards as well as a good performance in other competitions compared to other wineries (Rössel & Beckert 2012). This serves as a legitimate status for vintners. Secondly, the DLG competition is well known and based on a classification system unique to Germany (Rössel & Beckert 2012; Bock 2016, see section 4.4.3, pp.32). It therefore represents well the common perception of quality in the country. Quality is thus defined as the rating received in the annual DLG award competition based on a winery's current and past performance in other competitions.

5.1.3 Cluster

One of the more dominant definitions of clusters is Porter's take on them as "geographic concentrations of interconnected companies and institutions in a particular field" (Porter 1998, p.78). Taking a historical and industry focused approach, his definition suggests that the official wine regions are suitable clusters as they span spatial areas tied to the wine industry. However, these regions are too large to assume that they offer a common basis for communication. For instance, wineries within the region of Rheinhessen, will naturally identify themselves with it, but with an area of 26,600 hectares, active and frequent knowledge exchange is unlikely (DWI 2017a). Regions are therefore not a good proxy for clusters in this context. Official districts (*Bereiche,* see Figure *4.1*) which are a sub-category of wine regions, would offer another cluster definition. However, here too, they are considered too large. Some wine regions, such as Franken, are so small that they do not have any district, which disqualifies the latter for the same reason as regions (DWI 2017a). Regions and districts are therefore on an industry basis good cluster definitions, but not on a historical and practical basis as they are too large to have fostered a tight network, especially a 100 years ago when distances were harder to overcome. To account for communication and a denser network based on face-to-face meetings, zip code areas are judged more suitable. The German five-digit zip code system has been in place since 1993, of which the first digit represents the zone and the second digit the region, together they define the geographic location, the zip code area (*Zielregion)*, as illustrated in Figure 5.1 (Deutsche Post AG 2017).

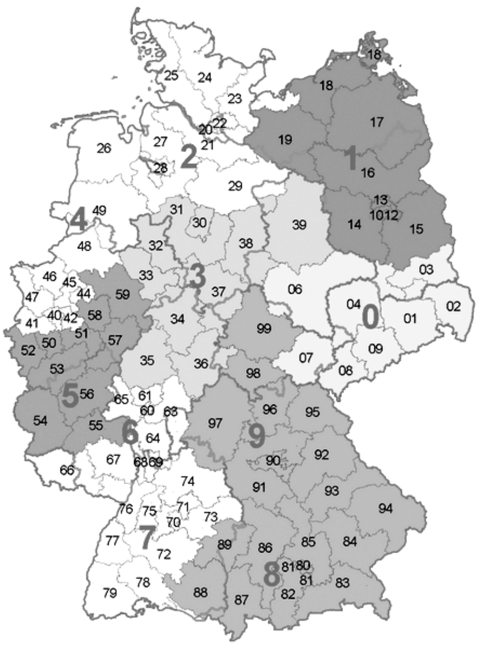


Figure 5.1: **The Division of German Zip Code Areas.** Source**:** Porto-Info (2017)

The last three digits were partially set according to the specific location of the receivers, if they have a post box or a real address and whether they are individuals or larger entities (Deutsche Post AG 2017). The zip code area is therefore suitable for the cluster definition as they are geographically bound and owners of wineries are likely to be present on their grounds due to the nature of the product. Apart from the practical argument for using this cluster definition, it has also been applied in academia. Grubesic (2008) for instance describes how zip codes are increasingly used by network studies, because the zones represent "like-minded consumers, of similar demographics and socioeconomic status" (p.129). A similar argument can be made for wineries in terms of commonalities. Adams (2002) and Herrera-Yagüe et al. (2015) define spatial networks and knowledge exchange by the use of zip codes. As highlighted in the literature review, regional institutions are found to be of importance for the evolution and economic structure of clusters (Saxenian 1994; M. E. Porter 2000; Maskell 2001; Bell & Zaheer 2007, see section 2.2, pp.7). Zip code areas can hence be considered to include these establishments, which eventually are part of wineries' daily life. By being in the same area, they are likely to share the same cultural background, the same agricultural and climatic conditions, vintners' children are more likely to go the same schools and they are more inclined to visit the same town hall meetings and fairs. In the case of larger wine regions or districts, this is less likely to be the case. Based on the options available, a cluster definition is thus a tradeoff, as larger clusters will exhibit less biases but more variability and finer grained ones the opposite dynamics. For historical and practical reasons, zip code areas are however used as cluster definitions for this study since communication within them is more likely.

5.2 Data Selection

This paper considers geographical distances as linkages between wineries. As grapes are an agricultural good, they are locally bound and related businesses are therefore not as mobile as firms in urban space. Based on the climatic requirements, wine is thus grown in clusters. As these are spatially fixed, it can be argued that geographical distance might be a key factor as it is a hurdle that these businesses proactively have to overcome to communicate and which is a constant factor of the environment, as moving one's vineyard is not possible. A spatial network of wineries is therefore likely to affect the social linkages of the clusters.

However, as mentioned in the literature review, the concept of networks encompasses not only social but also biological, informational and technological networks (Borgatti & Halgin 2011, see section 2.1, pp.9). The fundamental trait of networks is that they consist of items which are connected by ties as explained in the literature review (see section 2.1, pp.8). In regard to clusters, Boschma (2005) and Balland (2012) stress that proximity can be defined by other variables such as organizational, institutional and cognitive proximity. Hence, firms can be connected by having the same structure, being part of the same industry or share the same knowledge base (Balland 2012). Having this in mind, multiple linkages have been considered for this paper. The aim of this section is to illustrate why certain linkages have *not* been chosen for the analysis, although they could provide interesting results, and why the spatial network is the subject of analysis.

5.2.1 Events

As Giuliani (2007; 2013) argues, spatial proximity offers actors the opportunity to connect on social events to form a business network. Specifically, for wineries, local social events, such as trade fairs, markets (Brunori & Rossi 2000) and wine festivals (Bruwer 2003) are of importance. As wineries are embedded in local clusters and the concept of terroir aims at promoting a whole region, increased knowledge flows can be assumed to exist between wineries that meet each other at these occasions. It would be intriguing to see if participation of the same wineries at several events, national and international, would increase the likelihood that they attain awards. However, taking events as nodes presents practical problems. On the local scale, these events are often not advertised on the internet and participation of specific wineries is rarely observable. Also, many advertisements are deleted or updated after the event, which consequently makes it very difficult to investigate the network links through this channel. On the international scale, another consideration emerges. Large fairs, such as the annual ProWein conference in Düsseldorf, had more than 6,300 exhibitors in 2017, of which 1,003 were German (ProWein 2017). It is questionable, how valuable knowledge transfer at these conferences is, especially because these trade fairs are mainly focused on self-promotion on an international scale. This type of network has therefore primarily been excluded due to practical reasons.

5.2.2 Membership

Another interesting network are affiliation networks. These networks are characterized by actors' collaboration within a group which form links based on common membership (Newman 2003). One such entity are cooperatives (see section 4.2, pp.34), representing approximately a third of the German wine market (Hanf & Schweickert 2014). While cooperatives exist to represent their members' interest, internal conflicts and governance issues are a commonly known obstacles within these groups (Schweickert & Hanf 2007; 2014; Hanf et al. 2009) Due to this active and personal relationship structure within cooperatives, their effect on knowledge transfer is likely to be high. However, as mentioned in the background section (see section 4.2, pp.32), cooperatives have the reputation of producing low quality wine and member lists are therefore rarely published (Sáenz-Navajas et al. 2013, 2014; Hanf & Iselborn 2014). Consequently, this makes a network analysis impractical.

Another considered affiliation network is the membership of regional wine routes. This form of attracting tourists has increasingly gained attention in several countries (Hall et al. 1997; Rasch 2008; Koch et al. 2013; Brunori & Rossi 2000). As Brunori & Rossi (2000) argue, being part of a wine route demands collective action from the wineries involved as these have to follow a common set of rules such as keeping their cellars open for certain hours each day. However, like cooperative memberships, many wine routes do not mention the specific wineries on the trails. Further, wineries are commonly part of multiple routes as some regions are located close to each other, for example along the Rhine river. Membership, despite the interesting information it would offer, was therefore excluded.

5.2.3 Varieties

While varieties of grapes are included as variable in terms of number of white and red varieties, they have also been considered as tie between nodes. Reasons in favor of this approach are possible commonalities of cultivation procedures and knowledge. In other terms, wineries with common varieties would channel their advice seeking towards specific wineries and thereby establish more intensive knowledge flows. Further, while the importance of soil for quality is argued to be less significant than generally believed, grapes can still be considered a very spatially ground factor (Mueller & Sumner 2006). Nevertheless, taking grapes as linkage has been excluded as they are often mixed and the number of varieties is too limited to infer geographical communication.

5.2.4 Distance as Proxy

Excluding events, membership and varieties as possible network ties, geographical distance is chosen for the analysis. By excluding certain commonalities, a researcher can be said to face the "boundary specification problem" as derived by Laumann et al. (1983). According to them, making the choice of ties to study leads to the drawback that important nodes might be excluded or irrelevant nodes might be added to the network. Borgatti & Halgin (2011) argue however that, ultimately, it is the research question that dictates the ties chosen. By choosing a network, the researcher does not claim that other ties are irrelevant, but rather aims at answering the specific research question. Laumann et al. (1983) state that every research question generates its own network and in this paper, the interest in the effect of spatial networks on wine quality creates such a geographic network of wineries in Germany.

While taking pure interest into account, the choice of spatial proximity as nevertheless concrete reasons: Firstly, as emphasized in the literature section (see Table 2.1, pp.18), most studies on social links in clusters are based on qualitative studies, such as interviews. While this approach is straightforward, it is the aim of the paper to offer new knowledge as to what facilitates these social networks and the quality produced within a cluster. Further, considering the general perception that globalization has made spatial closeness obsolete, this kind of spatial network analysis and its effect on quality can contribute to the discussion whether this statement is true or not.

Secondly, a very simple reason for the choice of network is that space matters. As Barthélemy (2011) explains, spatial networks are crucial in multiple areas of life, ranging from infrastructure over brain neurons to communication. Based on evidence taking social networks and distance into account[1], friends are found to be most likely in one's neighborhood. Space therefore matters in social networks and as social networks matter in firm clusters, space is argued have an influence on quality.

A third reason is literally the nature of the industry. Wineries, being an agricultural industry, are affected by distance, both historically and practically. Historically, distance can be said to have shaped old social ties in these clusters, as in such a traditional industry, vintners had to overcome distances which were far more pronounced a 100 years ago than they are today. Considering that 70% of wineries worldwide are still family-owned and managed, distance was therefore a crucial part of shaping social networks which still exist today (Woodfield 2012). Further, wineries are very inflexible. For instance, firms in urban spaces can cluster. They can move into the same district or building and they might do this actively or not. However, wineries in the same cluster are located due to the requirements of their good. As they are immobile and in need of more space than non-agricultural firms, space becomes a very static factor which has to be overcome every time relations are made or nurtured. Distance can thus be said to be of importance for social networks in agricultural space, in the past and today.

Lastly, the data chosen for this analysis is also based on the novelty of the dataset and convenience. To the best of knowledge, no study has so far taken the spatial network of German wineries and connected it to awards. As the data is available on the internet and considering the scope of the paper as well as the absence of previous research, the spatial network analysis is deemed the best choice available.

The methodology thus takes distance as a proxy for social networks and measures thereby its effect on quality. It is important to note the limits of this methodology as well as the advantages. First, by taking distance as facilitator of social relations, the network does not give any account on the amount of information, nor the content of information transmitted. This can be considered the goal of the numerous qualitative studies as listed in the literature review (see Table 2.1, pp.18). Secondly, quality is a very vague concept, especially with experience goods such as wine. The methodology, focusing on the link between distance, communication and quality, is therefore prone to ignore other factors, such as other awards or distributive factors in Germany.

---

[1] For a complete literature list, please see Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1), 1-101

The approach offers however certain advantages. It is straightforward in its result: Proximity has a significant effect or it has not; this kind of analysis can be considered an advantage compared to qualitative social network analyses which are subject to human biases and subjective statements given in interviews. Further, as emphasized in the literature review (see section 2, pp.7), proximity has been found to be relevant (Jaffe et al. 1993; Audretsch & Feldman 1996; Wal 2014; Stefano et al. 2016), especially in wine clusters (Tor Guthey 2008; Hira & Swartz 2014; Yang et al. 2012). Proximity is thus chosen as network ties, based on its relevance proven by other researchers, its simplicity, its novelty in the wine industry and its convenience in terms of data collection.

5.3 Data Collection

The data used for the analysis consists of two merged lists: The list of wineries in Germany provided by the German Wine Institute and a dataset containing the winners of the German Agricultural Society award. The raw data is rich on information but the lists contain multiple outdated or wrong entries, which were manually cleaned and with help of the statistical program RStudio©, if the missing information would have constrained the analysis. The edges of the network dataset, that is the distances between the wineries, were collected by a query using Bing Maps© (Bing 2017).

5.3.1 Winery Dataset

The wineries dataset consists of 2,128 wineries, which have been extracted from the website www.deutscheweine.de, the online platform of the German Wine Institute (hereafter DWI) (DWI 2017a). The DWI is responsible for marketing and promotion of the German wine industry, and provides an extensive list of wineries in the 13 wine-growing regions. 265 wineries had to be removed from the dataset as they were either not wineries, no longer in business or counted twice in the data set. Table 5.1 and Map 1 provide an overview and illustration of the number and location of wineries per regions. “Borderliners” are the wineries which are located on the border of multiple areas and therefore officially grow wine in more than one region.

Region	# of wineries
Mosel	375
Rheinhessen	341
Pfalz	310
Franken	190
Baden	160
Württemberg	127
Nahe	138
Rheingau	109
Mittelrhein	37
Saale-Unstrut	25
Ahr	25
Hessische Bergstraße	12
Sachsen	10
"Borderliners"	4
Total	**1863**

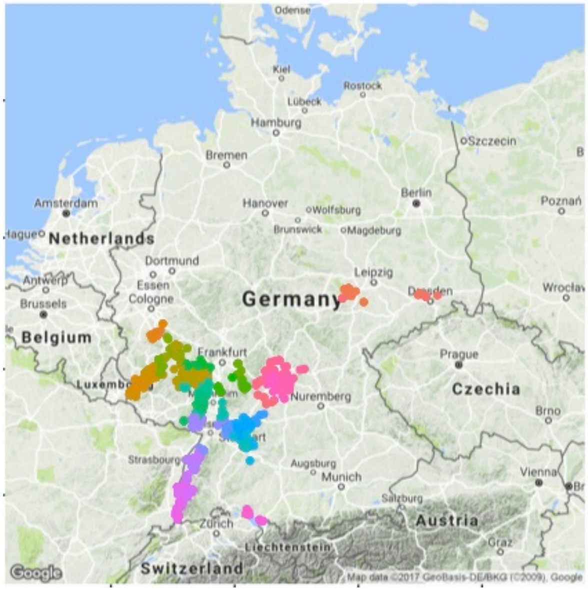


Table 5.1 and Map 1: **Distribution of individual wineries per wine region (for codes, see Appendix A)**

The original list of wineries includes the wineries' names, region, street address, city, website and the different grape varieties of white and red grapes grown by the winery. As this thesis uses distance as proxy for the likelihood of communication, spatial limitations are needed. Thus, as described in the data selection part (pp.44), zip code areas are deemed the best option to establish the link between communication and spatial distance. Using this division, the 1,863 wineries have been divided into 113 zip code regions, according to the first three-digits of their local zip codes.

5.3.2 Awards Dataset

The list of award winners was also extracted from the internet, only focusing on the Top 100 wineries in Germany, which are announced every year by the German Agricultural Society (DLG 2017b). The data consists of the ranks from eight years, that is, the winners from 2008 to 2016. Although the DLG award has existed since 2004, it was only possible to retrieve rankings from 2008 onwards. The data included in these lists consists of the name of the wineries, the rank, street address, region, five-digit zip code and website of the winners. For all the years, except for 2009, all ranks were obtained. Nine wineries in 2009 did not have any rank but were indicated as "new" in the ranking compared to 2008. These were deleted in the cleaning process.

5.3.3 Data Cleaning and Network Construction

As mentioned, two separate lists were merged to create the final database for the analysis. The statistical program RStudio© was used to prepare and analyze the data. The wineries provided from the DWI were then checked in terms of duplicates, outdated entries and other specificities, such as German "umlaute". During this process, 37 addresses were missing and added manually. Following, the first three digits of the zip codes were then selected to create the clusters.

Using RStudio©'s functions igraph and ggplot2, a network was created by placing the wineries as vertices and the edges as the distances between the nodes (Wickham 2016; Csardi & Nepusz 2006; R Core Team 2017). A Bing Maps© query was used to extract the distances within each zip code, which resulted in 141,620 edges in total (Bing 2017). The data collection had to be done in a stepwise manner, as the search engine only allows for a certain amount of data to be extracted. Following, due to faults in the original database, some addresses were not found and required manual rechecking and correcting of the dataset. Clusters with one winery were omitted as other variables, such as mean distance and sum of distance, could not be calculated from them. After cleaning the network dataset, the variables were calculated and merged with the cleaned award data. While merging the datasets, the year variable was not included, as the focus of the paper is which spatial network traits increase the score and thereby quality. A time variable would have distorted this focus as many wineries would not have been included and thereby potential valuable information would have been lost. Lastly, after constructing the final dataset, descriptive statistics and regressions are run to assess the research question of the thesis.

5.4 The Statistical Model

In this section, the statistical model and its basic assumptions are explained. To assess the research question, a sound statistical model has to be used. With this intention in mind, a multiple linear regression model is deemed to be the best approach for the analysis. The advantage of this model is its relatively easy implementation – by help of a statistical software such as RStudio© – and its capacity to answer a variety of research questions (R Core Team 2017). The model aims at explaining the variation in the dependent variable, y - the score of wineries - with help of a range of different explanatory variables. These consist of multiple numeric and one categorical explanatory variable, x_j , which are mean distance, sum of distances, standard deviation of the distances, number of wineries and its square root, number of grapes, awards per cluster and region. The model has thus the following specifications:

$$Y_{score} = \beta_0 + \beta_1 mean.dist + \beta_3 sd.dist + \beta_2\, sum.dist + \beta_4 n.wineries + \beta_5 n.wineries^2$$
$$+ \beta_6 n.whiteGrapes + \beta_7 n.redGrapes + \beta_8 awa.by.reg + \beta_9 region + \varepsilon$$

Y_{score} is the predicted score, the constant β_0 is the predicted value of Y_{score} when all the other variables are set to zero and β_j are the parameters associated with their respective variables. The regression aims at showing the expected change of Y_{score} per one unit change of x_j while holding the other explanatory variables constant (Stock & Watson 2015). The parameters β_j are calculated by the ordinary least squares (OLS) estimation procedure. The OLS method calculates the parameters by minimizing the sums of squares of deviation between the observed data and expected values (Wooldridge 2013).

For an OLS regression to be coherent, certain assumptions should be met. For this analysis, only the relevant assumptions are discussed[2]. The first crucial criterion for an OLS analysis is the absence of perfect multi-collinearity amongst the different explanatory variables (Wooldridge 2013, pp.80). Even if variables are not perfectly correlated, high correlations can cause biased estimates and thus result in an inconsistent analysis. A high correlation between explanatory variables for instance increases the standard error and can lead to some variables becoming insignificant even though they have a significant effect (Wooldridge 2013, p.81). Ideally, each explanatory variable is desired to contribute with some new information to explain changes in the response variable and this entails that correlation between the independent variables in the model would at best be low. Collinearity is acceptable up to a certain degree, where as a rule of thumb a correlation of 0.6 between two different variables is said to pose a problem. To identify possible multicollinearity, the regression is run by sub-sequentially adding variables to the model to see if, and how, the coefficients change the effect on score.

As the correlation table (Table 5.2) shows, the highest correlation is 0.945, between the number of wineries and the square root of number of wineries as the latter is derived from the first and serves the purpose to account for the hypothesized non-linear relationship. The sum of kilometers in a cluster and the number of wineries have a correlation of 0.873. This comes from the fact that a larger area

2 For a detailed discussion of all six Markov assumptions and their implications please see (Wooldridge, 2013 pp.80)

will probably also have more wineries and while the correlation is high it does not invalidate the model. The numbers of red and white varieties have a correlation of 0.817. Two potential explanations are suggested: Firstly, the ambition or size of the winery account for this. A high number of white varieties for instance could be argued to need a large area. A winery of this size might as well cultivate many red varieties. In other words, wineries managing large numbers of varieties have also more space to cultivate different grapes in terms of size. They might also be more ambitious, and experiment with certain varieties. The other reason for the high correlation could be related to the data collection. Not all wineries in the DWI dataset provide their varieties. Those that have published their varieties might be more thorough in terms of their listing while others have not listed theirs at all. Thus, the dataset might primarily consist of two extremes, those stating their grapes and those who do not. Taking the variables related to the hypothesis into account, the correlation matrix seems in line with the rule of thumb, as the values which are above 0.6 can be reasonably explained.

Correlation Matrix

	Score	Mean of Distances	Sum of Distances	St. dev. of Distances	Number of Wineries	Squared Number of Wineries	Number of White Varieties	Number of Red Varieties	Awards by Cluster
Score	1								
Mean of Distances	-0.157	1							
Sum of Distances	-0.148	0.485	1						
Standard Deviation of Distances	-0.031	0.540	0.290	1					
Number of Wineries	-0.203	0.381	**0.873**	0.218	1				
Squared Number of Wineries	-0.159	0.374	**0.945**	0.195	**0.952**	1			
Number of White Varieties	0.148	-0.046	-0.011	-0.020	0.027	0.048	1		
Number of Red Varieties	0.134	-0.079	-0.008	-0.052	0.020	0.053	**0.817**	1	
Awards by Cluster	0.253	0.007	0.358	-0.035	0.425	0.445	0.254	0.181	1

Table 5.2: **Correlation Matrix.**

A second assumption for an OLS regression to be valid is the absence of heteroscedasticity, that is, when the size of the error term is different depending on the value of a variable. For instance, the scatterplot Figure 5.2 shows that for a low number of wineries in a cluster, the number of awards is also low, however the observations become more spread out the more wineries are in a cluster, as some clusters win, and others don't despite the presence of more wineries. When a variable's value changes, it contradicts the assumption that the variance of the residuals is independent of the value of the explanatory variables.

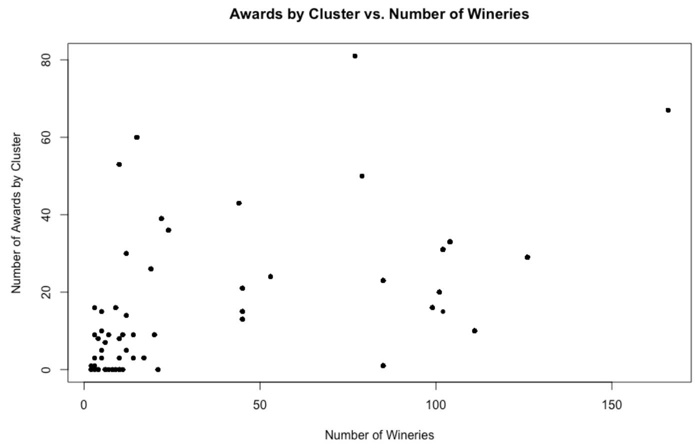


Figure 5.2: **Scatterplot showing the Number of Awards per Cluster vs. Number of Wineries.**

The Breusch-Pagan chi-square tests, where the residuals are regressed with the explanatory variables, is a way to quantitatively test for heteroscedasticity (Wooldridge 2013, pp.267). If a significant relationship is found, heteroscedasticity is present and the regression analysis should account for this by, for instance, using robust standard errors. For this dataset, the test shows a p-value of < 2e-16 (see Table *5.3*) so heteroscedasticity is present. The model accounts for it by using heteroscedastic robust standard errors in the analysis.

Breusch Pagan Test			
Regression Model		score ~ mean.dist + sd.dist + sum.dist + n_mi_win + n_mi_win2 + sum.dist*n_mi_win + n.whiteGrapes + n.redGrapes+ awa.by.reg + factor(region),	
BP	360.65		
Df	25	**p-value**	**< 2.2e-16**

Table 5.3: **Result of Breusch Pagan Test**.

An additional factor that influences the standard errors of the analysis is the fact that the wineries are clustered in zip code areas. Wineries within the same cluster will most likely have certain traits in common, such as bureaucratic requirements, weather or certain events. These are factors not included in the model, but they nevertheless create a relation between the wineries within a cluster. Ignoring this correlation can lead to small standard errors and very narrow confidence intervals and thereby obscure the actual relationship of the explanatory and response variable.

The statistical model should therefore allow for correlations between observations in a cluster while assuming independence across clusters. Each cluster has therefore an assigned ID and cluster robust standard errors have been included in the computation of the regression, which also treats the mentioned heteroscedasticity (see section 5.4, pp.50). For a comparison between the use of normal standard errors and cluster robust standard errors, please see Appendix F (pp.100)

5.5 Variables

Below, the variables of the model are explained. With the data available, they were chosen to reflect the measurement that are considered best to assess the research question, specifically the proximity and the clustering in the wineries' environments.

Score

As highlighted in the quality section (see section 4.4.2, pp.39), awards play a crucial role in the wine industry as they serve as indicators of quality for a good that consumers usually cannot assess before purchase (Lockshin et al. 2005; Orth & Krska 2002; Sáenz-Navajas et al. 2013). Thus, the DLG award ranking serves as scale to assess quality. The dependent variable ranges from 0 to 100, where 0 stands for the wineries never having received an award, and the score 1 to 100 representing the actual rank of the winery. For convenience purposes, the original ranking has been inverted, such that the score ranges between 1, being the lowest score, and 100, the top score. This ensures a straightforward analysis and interpretable results.

Mean Distances

The mean distance is the average taken from all the outgoing ties of a central winery. Figure 5.3 shows the spatial distances taken from a cluster in the database and illustrates how the average of the ties that connect a winery to all the other wineries in the cluster is taken. As highlighted in the literature section (see section 2, pp.7), previous studies have found a positive effect of proximity and performance (Jaffe et al. 1993; Audretsch & Feldman 1996; Balland 2012; Oerlemans et al. 2005). One should therefore expect a shorter mean distance to have a positive effect on score.

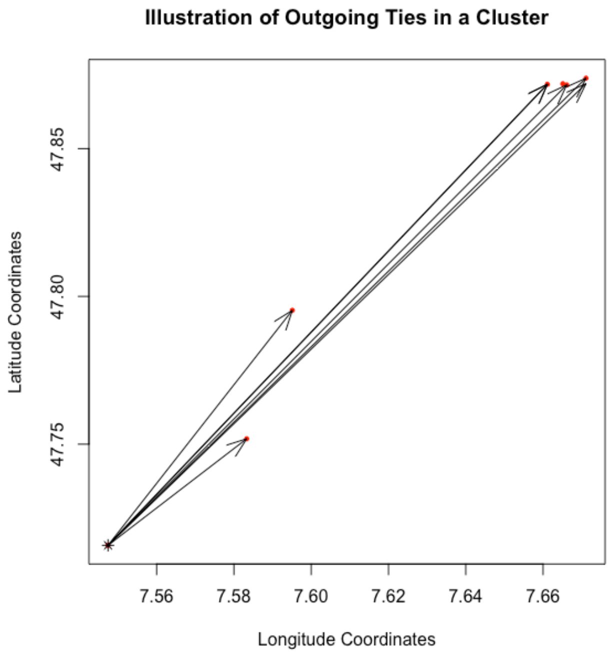
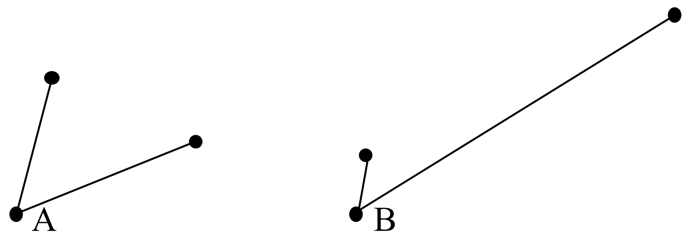


Figure 5.3: **Example of Cluster Positions and Distances.** The figure illustrates a real cluster (zip code area 794) of which the mean distance for the central winery (star-shaped) is taken by calculating the average of the distances to all other wineries in the cluster.

Standard Deviation of Distances

The standard deviation of the distances is taken to account for the clustering effect, which is not shown by the mean distance. To illustrate, winery A has two medium large distances to two wineries, while winery B has one very large distance and one very short distance to two other wineries (Figure 5.4). The means of the distances for winery A and B would therefore be similar even though they exhibit very different clustering effects. The standard deviation is hence included in the model to account for the effect of deviations from the mean of distances. A larger standard deviation means that a winery has some wineries very close and others very far located from it. Based on the theoretical argument that closer distances facilitate knowledge exchange (Saxenian 1994; Storper & Venables 2003; Weterings & Boschma 2009; Duarte Alonso 2011; Rosenfeld 1997; Ganesan et al. 2005), we should expect that wineries with higher standard deviations from the mean have some wineries closer to them and therefore on average higher scores.



$$\bar{x}_A = \bar{x}_B$$
$$\sigma_A \neq \sigma_B$$

Figure 5.4: **Example of Cluster Positions with the Same Mean Distances but Different Standard deviations**

Sum of Distances

As illustrated by Figure 5.5, the variable of sum of distances is taken by adding all the distances in km between the nodes in a cluster to attain a cumulative distance. While this serves as proxy for the spatial size of a cluster, it has however the weakness that a cluster with a large sum of distances can be caused by a few wineries far away from each or because a lot of wineries are present. To address this issue, one of the regression models further includes an interaction term between the sum of distance and number of wineries. This also enables to pin-point the specific effect of sum of distances in large versus small clusters. Referring to the literature (see section 2, pp.7), we should expect the sum of distances to have a negative effect on the score of wineries.

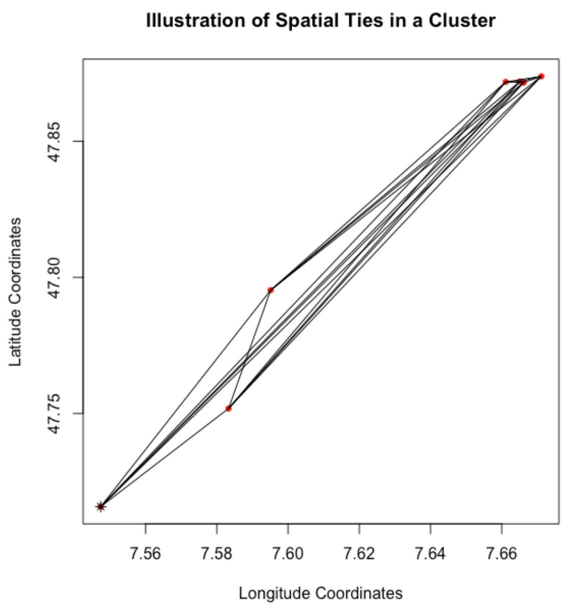


Figure 5.5: **Example of Cluster Ties (zip code area 794) used to Calculate the Sum of Distances**

Number of Wineries

The number of wineries per cluster is the sum of all its wineries subtracted by one. The reason for the subtraction is of statistical nature. There are multiple clusters with only one winery and therefore no mean distance, standard deviation and sum of distance. For statistical reasons, these wineries have to be either deleted or altered. The latter has been done in order to keep them in the database. "The Number of Wineries – 1" represents the number of all the *other* possible wineries the respective firm can form relationships with in its cluster. Furthermore, by subtracting one from the total number of wineries, the clusters with only one winery automatically become part of the baseline. This alteration will however have no impact on the outcome of the regression analysis. Based on the principles of organizational ecology, we should expect there to be a relationship in the shape of an inverted u-shape to reflect the gains from increased knowledge but also the disadvantages of high competition.

Square Root of Number of Wineries

The square root of the number of wineries is included to test for a non-linear relationship explained in the theoretical argument section (pp.26). By making the number of wineries to a quadratic variable, the regression can account for any non-linearity as hypothesis H_2 suggests. It therefore accounts for the assumption that an increasing number of wineries has a positive influence on score up to a certain point after which a negative effect is expected.

Number of Varieties

The varieties of grapes in the database is divided in two variables, white grapes and red grapes. In the regression model, the number of varieties has been included to account for possible effects. One assumption could be that a lower number of varieties could have a positive effect on score, as a winery specializes more. On the other hand, more varieties could mean more communalities with other wineries and therefore more possible partners for knowledge exchange about the right cultivation of the varieties. While this variable is not a focus point, it is included to account for possible effects. One consideration concerning this variable was to include the total number of grapes instead, that is, the sum of the number of white and red grapes. This has however been excluded due to two reasons: Statistically seen, for a regression, it would have contradicted the assumption of non-collinearity, as the sum would naturally be perfectly correlated to the distinctive numbers of white and red varieties.

Further, a finer division has the potential of showing more, and therefore varieties have been separated.

Region

The variable region, taking the value of one of the 13 official regions, is included as a control variable (Wooldridge 2013, pp.470). The dummy variable of region introduces possible factors which cannot be observed such as weather or political conditions. The primary reason for using region as a fixed effect variable is the assumption that some factors within region might impact the winning of scores. As each region is expected to have some individual characteristics, their error terms and the constant should not be correlated with the other variables.

Awards by Cluster

Awards by cluster is calculated as all the sum of all the awards a cluster has won in the last nine years, and is constant for all the wineries within the cluster. By including this variable, it can be tested if on a cluster level, it has an effect for a cluster when wineries within it receive awards. One possible explanation could be that wineries within a cluster pay more attention to the winner and begin collaboration with him to gain knowledge they can use themselves. We should therefore expect to see a positive effect in score.

Zip Area Code

Due to the mentioned heteroscedasticity and the fact that the dataset must account for intra-cluster correlations, the zip area codes are included, though not as variable but in the regression analysis to define the robust standard error.

	Variable	Regression code	Type	Description
Independent variable	Score	*score*	Ordinal	Attained score, 0 = no award, 1(lowest) – 100(highest) ranking
On individual winery level	Mean of distances	*mean.dist*	numerical	Mean of all the distances going from a winery to all others within the same cluster
	Standard deviation of mean	*sd.dist*	numerical	Standard deviation taken from the mean of all the distances going from a winery to all the others within the same cluster
	Number of white varieties	*n.whiteGrapes*	numerical	Total number of white grape varieties by winery
	Number of red Varieties	*n.redGrapes*	numerical	Total number of red grape varieties by winery
On cluster level	Sum of distances	*sum.dist*	numerical	Sum taken from all the distances within a cluster
	Number of wineries	*n_mi_win*	numerical	Total number of wineries within a cluster minus one
	Square of number of wineries	*n_mi_win2*	numerical	Square root taken from the variable "number of wineries" to account for non-linearity
	Number of wineries * Sum of distances	*n_mi_win*sum.dist*	numerical	Interaction term to account for cluster-effects
	Region	*region*	dummy	One of the 13 official wine regions or "Borderlines"
	Attained awards per cluster	*awa.by.reg*	numerical	Number of attained awards of a given cluster in the last 9 years

Table 5.4: **Regression Variables.** The columns show the name of the variable, the code used in the statistics software R, the type and a short description of the variable including its purpose. The variables are grouped according to their level of clusters, where the smallest entity is the individual firm level. The cluster level represents all variables which have the same value for the whole cluster.

6. Regression Analysis

Having laid out the research design including the statistical model and the variables, this section provides a short summary of the descriptive statistics followed by the regression analysis and the results.

6.1 Summary Statistics

As mentioned in the methodology section (see section 5, pp.40), the raw dataset contains 2,128 individual wineries, of which 265 had to be omitted due to mistakes in the dataset, double entries or outdated information. 1,863 wineries are thus the subjects of the final dataset. One outlier was removed, as its mean distance fell considerably outside the range of the other data points and would thus have caused a substantial bias in the regression (see Appendix B, pp.97). Table 6.1 provides the descriptive statistics of the wineries in the dataset for the numeric variables. The mean distance between wineries is approximately 15 km, with a minimum of 38 meters and a maximum of 102 km. The sum of distances depicts twice the actual distance per cluster for the following reason: Taking the distance between A and B using Bing Maps is not necessarily the same as the length as B to A. One-way streets can be one reason for this. The data query thus counts each edge, that is A to B and B to A which results in the distance being approximately twice as long in the analysis than in reality. However, as this is the case for every cluster in the database, it does not affect the interpretation of the results. Further, for a practical representation, the sums of distances are divided by 1,000, and hence denoted in thousand km. Again, this does not have an effect on the interpretation of the final results. The average sum of distances, as extracted from the analysis, is thus 140,000 km, where the cluster with the lowest sum spans 1 km and the largest 514,000 km. In terms of density, there are on average 75 wineries per cluster, the smallest consisting of two wineries, and the largest of 166. On average, a winery has four white varieties and approximately two red varieties. This might be explained by the fact that Germany traditionally has grown white grapes which today represent 64% of the grapes grown in the country (DWI 2017a). Lastly, the average number of awards won by a cluster in the last nine years is approximately 27; the standard deviation is quite large as there are multiple wineries in the sample without any awards, while others have won in consecutive years.

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Median	Min	Max
Score	1,863	3.51	14.56	0	0	100
Mean of Distances	1,863	14.19	8.98	12.88	0.38	101.92
Sum of Distances (in tsd)	1,863	139.78	166.12	104.32	0.01	514.03
Standard Deviation	1,863	10.02	11.12	7.69	0.00	166.23
Number of Wineries	1,863	75.35	48.23	85	2	166
Squared Number of Wineries	1,863	8,002	7,826	7,225	4	27,556
Number of White Varieties	1,863	3.96	3.39	4	0	14
Number of Red Varieties	1,863	2.40	2.40	2	0	12
Awards by Cluster	1,863	26.57	22.25	23	0	81

Table 6.1: **Descriptive Statistics for the spatial network of wineries in Germany.** The table summarizes the number of unique wineries with the respective mean, standard deviation, median and minimum as well as maximum values of the variable. The sum of distances has been divided by 1000 to make the values more comprehensible.

Figure 6.1 shows the distribution of the variables necessary to assess the hypotheses stated in the theoretical argument section (pp.26). The mean distance shows a right-skewed distribution with most wineries being in the range of 0 to 20 km. Any wineries with longer distances are very low in numbers. The distribution of the sum of distances is very right-skewed. Most clusters exhibit shorter total distances and there are only a few clusters within the 400 to 500 range. While these could be considered outliers, an omitted version of the regression did not show any change whether these were included or not (see Appendix C, pp.97). The observations were thus kept in the database for completeness sake. The number of wineries per cluster spans a wide range of values, showing an accumulation in the range of 10 to 25 wineries, though there are also clusters with 100 and more present. The distribution of awards per cluster exhibits a decreasing trend with most observations being close to or precisely 0.

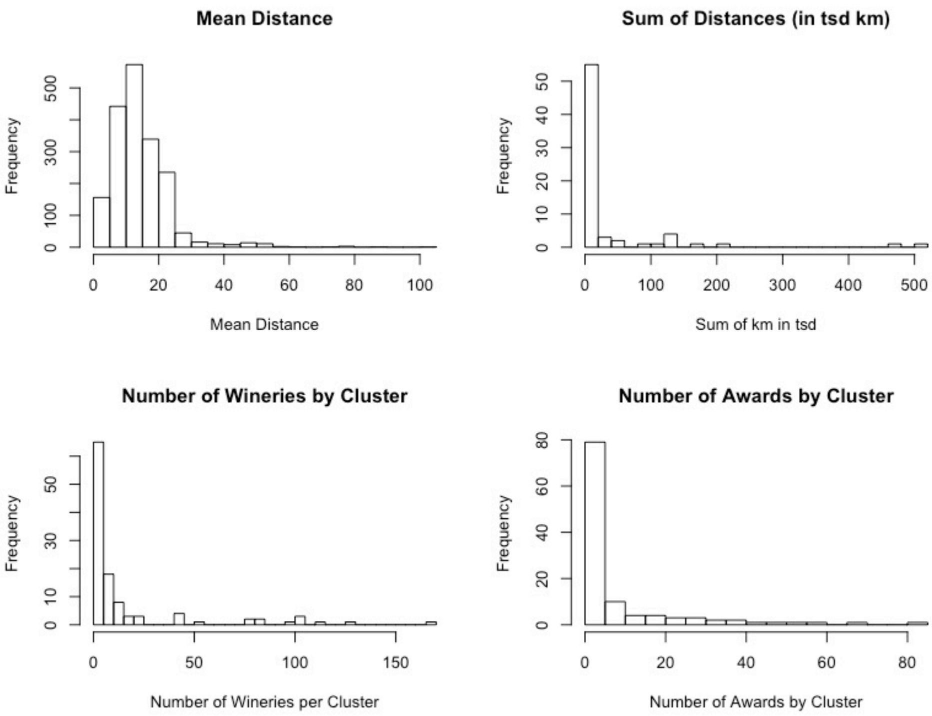


Figure 6.1: **Histograms of the Distributions of four Variables.** The upper left histogram shows the individual wineries of mean distance. The other histograms show the distribution of the individual clusters in terms of sum of distances, number of wineries by cluster and awards by cluster. For histograms showing the distribution of all the observations see Appendix E.

6.2 Results

In order to conduct a coherent analysis, each type of variable is added gradually. This assures that possible biases and changes of the effect on score and R squared become apparent. The regression analysis therefore results in nine models, of which the last one includes all the variables as described in the statistical model in section 5.5 (pp.54).

The analysis explains the effect and changes of each variable throughout the different models considering t-statistics, p-value and R squared. The t- and p-values address the question of how likely it is that the true coefficient of a variable is zero (Teetor 2011). As the t-value translates into the p-value, both indicate the significance of a variable on the score. Following the general practice, a p-value equal to or less than 5% is the threshold to assess significance. R squared indicates how much of the variation of the independent variable is explained by the model (Wooldridge 2013). The test models (1) – (3) address H_1 and H_3 by covering the effect of geographical distances between wineries and spatial cluster size. Model (4) – (5) introduce the density component as stated in H_2 (pp.30). The remaining models (6) – (9) introduce possible fixed effects such as varieties, awards by cluster, regions and an interaction term to account for the correlation between number of wineries and sum of distances.

Table 6.2 below shows the results of the regression analysis.

Regression Results

	Score								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Constant	24.20***	24.16***	24.07***	32.93***	27.79***	27.73***	18.64***	20.77*	23.99**
	t = 6.40	t = 6.59	t = 6.70	t = 5.91	t = 5.46	t = 5.16	t = 4.38	t = 1.70	t = 2.00
Mean of Distances	-0.54***	-0.39***	-0.53***	-0.49***	-0.50***	-0.50***	-0.35***	-0.45***	**-0.44*****
	t = -3.62	t = -2.73	t = -2.94	t = -2.92	t = -3.18	t = -3.09	t = -2.63	t = -3.05	t = -3.03
Sum of Distances		-0.02	-0.02	-0.005	0.01	0.01	0.04**	0.03	**0.15****
		t = -1.25	t = -1.19	t = -0.34	t = 0.96	t = 1.02	t = 2.43	t = 1.54	t = 2.09
Standard Deviation of Distances			0.21*	0.24**	0.22**	0.22**	0.26***	0.30***	**0.29*****
			t = 1.66	t = 2.05	t = 2.09	t = 2.11	t = 2.81	t = 3.10	t = 3.03
Number of Wineries				-0.32***	-0.29***	-0.29**	-0.28***	-0.16**	**-0.49****
				t = -2.77	t = -2.69	t = -2.50	t = -2.93	t = -2.29	t = -2.06
Squared Number of Wineries				0.002**	0.001	0.001	-0.0002	-0.001	0.002
				t = 2.43	t = 1.64	t = 1.50	t = -0.31	t = -1.47	t = 0.94
Number of White Varieties					1.17***	1.06	0.05	-0.07	-0.11
					t = 2.76	t = 1.02	t = 0.06	t = -0.09	t = -0.13
Number of Red Varieties						0.17	0.73	1.21	1.25
						t = 0.13	t = 0.66	t = 1.16	t = 1.20
Awards by Cluster							0.50***	0.56***	**0.60*****

							t = 3.75	t = 4.35	t =4.20
Ahr and Mittelrhein								-17.67 t = -1.44	-17.12 t = -1.44
Baden								-0.20 t = -0.02	-1.01 t = -0.08
Baden and Hess. Bergstrasse								-8.53 t = -0.65	-9.74 t = -0.76
Franken								-8.30 t = -0.64	-8.00 t = -0.64
Hess. Bergstrasse								-4.33 t = -0.33	-5.74 t = -0.45
Hess. Bergstrasse and Rheingau								-5.54 t = -0.42	-5.65 t = -0.45
Mittelrhein								-6.34 t = -0.46	-7.38 t = -0.55
Mittelrhein and Rheingau								-6.02 t = -0.46	-7.36 t = -0.59
Mosel								-4.37 t = -0.35	-3.84 t = -0.31
Nahe								-2.41 t = -0.18	-3.97 t = -0.31
Pfalz								-16.68 t = -1.29	-15.69 t = -1.24
Rheingau								-5.30 t = -0.40	-5.46 t = -0.43
Rheinhessen								-3.86 t = -0.30	-4.24 t = -0.34
Saale-Unstrut								-16.59 t = -1.33	-15.30 t = -1.25
Sachsen								-16.91 t = -1.31	-18.62 t = -1.47
Wuerttemberg								-3.01 t = -0.23	-3.70 t = -0.30
Sum of Dist. x Nr of Wineries									-0.001 t = -1.59
Observations	2,605	2,605	2,605	2,605	2,605	2,605	2,605	2,605	2,605
R^2	0.02	0.03	0.04	0.07	0.09	0.09	0.20	0.23	0.23
Wald Test (alternative F-Test)	13.12^{***} (df = 1; 2603)	6.21^{***} (df = 2; 2602)	4.27^{***} (df = 3; 2601)	4.52^{***} (df = 5; 2599)	5.80^{***} (df = 6; 2598)	6.82^{***} (df = 7; 2597)	9.80^{***} (df = 8; 2596)	10.11^{***} (df = 24; 2580)	10.35^{***} (df = 25; 2579)
Note:									*p<0.1; **p<0.05; ***p<0.01

Table 6.2: **Regression Analyses on Spatial Network Factors.** The columns labeled by numbers represent the incremental regression models as described in the regression analysis section. The variables are listed in the first column. For every model, each variable's coefficient and t-value are stated. The stars next to the coefficient indicate the significance by p-value, where * equals a p-value less than 10%, ** equals a p-value less than 5% and *** represents a p-value less than 1%. The last rows in the table represent the number of observations, R squared and Wald - statistics.

The first variable, the mean of distances per winery, is highly significant. The variable's effect on score is very robust as there are no significant changes of the coefficient when different control variables are added. Taking the t- and p-value, the variable is statistically highly significant on a 1% level. As the coefficient is negative, it can be concluded that the closer on average a winery is to others in its cluster, the higher is the score.

The sum of distances has a less stable effect as can be seen by the changing sign of the coefficient and varying significance level. Model (8) shows no significance for this variable, however by adding the interaction term of number of wineries and sum of distances in model (9), it becomes significant on a 5% level with a positive effect on score. Overall, the positive effect on score implies that the longer the distances in a cluster, the higher, on average, the score. This could potentially indicate that lower competition contributes to higher performance.

The third variable, standard deviation of the distances, is also highly significant on a 1% level. The effect is generally robust as the coefficient does not change considerably when adding control variables. The p-value of the standard deviation points towards a positive effect on the score when a winery has a few neighbors very close by and others very far away. This result offers various potential interpretations and is further treated in the discussion part.

Model (4) introduces the effect of number of wineries in a cluster. The variable is not very robust, as the coefficient changes throughout the models; especially when the interaction term is added in model (9). The fact that the coefficient changes from -0.16 to -0.49 implies the presence of an omitted variable bias. The variable is however significant on a 5% level. As the squared variable is not significant, it suggests that there is a linear relationship, thus more wineries in a cluster imply a lower score on average, which could be based on negative competition effects when many wineries are present.

Model (5) - (6) introduce the numbers of white and red varieties to the regression analysis, of which none is significant. Number of grape varieties does thus not seem to have an effect on score.

The number of awards by cluster however, is highly significant with a robust and high coefficient. On average, every award won by a cluster increases the score by 0.6, which indicates a high spill-over effect.

Model (8) contains the 17 official regions to account for fixed effects. R automatically excludes one region to prevent perfect multicollinearity, in this case the region Ahr. Overall, regions do not seem to have a considerable impact on the dependent variable.

Model (9) includes the interaction term of sum of distances and number of wineries which has no significant effect on score. It is however still important to include it as a control variable, in order to ensure no omitted variable biases (see Model (4)).
In terms of the R squared, the adding of variables throughout the analysis improves the explanatory value. Especially when the number of awards by cluster variable is added in model (7), R squared increases considerably namely from 0.09 to 0.2. The final model (9) explains 23% of the variation of the dependent variable. As some other potential correlations exist between the variables, interaction terms are tested (see Table 5.2, pp.52). Standard deviation is combined with the number of wineries and mean distances to form two interactions as are the number of wineries and the awards won by cluster. Due to the scope of the paper the regression results are in appendix D (pp.98). Only one interaction term however was significant which is treated in the discussion section (see section 7, pp.65). The Wald statistic is used as robustness check as the regression takes cluster robust standard errors (Wooldridge 2013, pp.579). The values are equivalent to F statistics and show throughout the analysis high significance.

To conclude, the analysis shows that mean distance between wineries has a significant effect and that shorter mean distances on average imply higher scores. Number of wineries in a cluster is found to be significant and exhibits a negative effect on score. The opposite is true for the spatial dimension: the larger the cumulative distances in a cluster, even controlled for the number of wineries, the higher the score. Other variables such as grapes and the fixed effects variable of regions do not show any significant relationship. However, the variation of the distances to other wineries in a cluster is highly significant; a combination of close neighbors and actors further away thus has a positive effect. Lastly, awards won by a cluster in the last nine years are found to be highly significant, thus the number of past awards increases scores on average. The following discussion treats these findings in relation to the research question and hypotheses.

7. Discussion

The significance of the variables in the analysis supports that spatial proximity matters. In short, it has an effect on the score and thereby the quality of wine produced. However, some dynamics of distance seem to work in unexpected ways. This section revisits all the hypotheses in light of the findings and discusses other potential interpretations of the results. Table 7.1 summarizes the hypotheses, including the expected and actual results.

7.1 Hypothesis 1

The analysis finds that a shorter mean distance to other wineries in a cluster has a significant positive effect on score. The hypothesis that proximity to other cluster members improves quality of the product is thus supported. As argued, knowledge is more likely to flow over short spatial distances; and hence the result adds to the stream of findings that firms within clusters benefit from knowledge spillovers as innovation increases (Jaffe et al. 1993; Audretsch & Feldman 1996; Balland 2012) and performance, such as growth rates and innovative results (Arndt & Sternberg 2000; Oerlemans et al. 2005). This conclusion is also in line with Foster & Rosenzweig's (1995) finding that immediate neighbors in agricultural clusters have an effect on one's performance through exchange of expertise. Industry-specific, these results also add a quality argument to Yang et al.'s (2012) finding that wineries with successful neighbors can charge higher prices. While they find a positive relation between spatial proximity and price, this study finds that simple proximity to neighbors enhances quality. While the mentioned studies define clusters by cities, regions or official agglomerations (Jaffe et al. 1993; Audretsch & Feldman 1996; Balland 2012; Arndt & Sternberg 2000; Oerlemans et al. 2005), this study reaches for the micro-level as defined by individual actors in the network and how their distances in kilometers, affect quality. It thereby offers a more detailed approach towards how spatial networks affect quality. The regression indicates that shorter distances to one's neighbors provide a more stable social network through which important knowledge can flow. Vintners thus attain more expertise and apply this on the production of wine. As the fixed effect of regions is not significant, geological conditions on that level do not seem to affect the quality. Nevertheless, a limitation of the variables is that finer divisions of space, such as districts, could show differences between clusters in terms of natural endowments (see section 5.1.3, pp.42).

7.2 Hypothesis 2

The number of wineries in a cluster has a negative effect on the score. Hypothesis 2 is thus not supported by the findings. Taking an organizational ecology perspective, density was expected to have a positive effect due to the increase of potential partners and, after a certain threshold, a negative impact due to competition. The results however suggest that only the latter holds true. Further the insignificance of the interaction term shows that density's effect is not dependent on cluster size. Thus, more wineries have a negative effect regardless of the sum of distances in a cluster. This linear negative relation is in line with studies by Staber (1998), Pouder & John (1996), Baum & Mezias

(1992) and Myles Shaver & Flyer (2000). Unintentional knowledge spillovers (Staber 1998) and free-riding by low performers (Myles Shaver & Flyer 2000) are thus factors that could potentially create incentives for high performers to locate in less dense environments. Although wineries are spatially less mobile, based on evolutionary dynamics, it can be suggested that high performers are more likely to prefer clusters with fewer wineries. The result thus does not support the hypothesis of an inverted u-shaped relationship as argued by Folta et al. (2006) and Fernhaber et al. (2007). However, it adds an agricultural perspective to existing literature. The finding suggests that multiple dynamics are at work: Wineries, requiring specialized labor, are possibly competing for skilled workers. Thus, the more wineries are in a cluster, the higher the prices for labor and thereby competition. As clusters are defined as three-digit zip code areas, the space is nevertheless limited and the sum of distances might not completely accurately reflect a cluster size. So, the denser the network, the smaller a winery's area might be. This could impact its resources to experiment, improve or even enter competitions. As generally known from urban settings, growing numbers of actors might also contribute to more anonymity, which could decrease communication within the cluster. While empirical studies are needed to determine these dynamics, the result offers insights into the ecological dynamics of an agricultural setting where more possible ties inhibit the production of high quality wine.

7.3 Hypothesis 3

One interesting finding of the regression analysis is that the sums of distances per cluster are positively related to score. Wineries located in a cluster with large total sums of distances, regardless of the number of wineries, have therefore higher scores on average. This rejects hypothesis 3 that more space in a cluster will lead to less communication and thus lower quality. This result is interesting as it is not in line with findings such as Jiaming et al. (2015) and Rosenthal & Strange (2003). It also contradicts established studies by, for instance, Jaffe et al. (1993) and Giuliani (2007). Intuitively, expertise through communication should be fostered in small clusters not large ones. However, an attempt to explain this finding leads to the following suggestions: While hypothesis 1 is supported, thus wineries are found to benefit from close distances *between each other*, hypothesis 3 indicates that total distances in a cluster, that is distances between *all* wineries, increase score because of lower competition. Thus, hypothesis 1 shows that on the individual node level, firms need proximity, but on cluster level, space might be important. In these clusters, wineries might grow larger areas and have thus more means to invest in facilities. The more space is available, the more possibilities to produce different wines, which vintners can compete with in competitions. Logically, the more wines are handed-in, the higher the chances to win awards. On the other hand, specialization on a few varieties would intuitively contribute more to quality than having a broad range. The relationship is thus not clear-cut and future research should assess the dynamics more in depth. One important thing to note is that the variable itself might have a technical weakness. As stated in the analysis section (pp.59), the coefficient is not robust, which makes an interpretation less certain. Further, the sum of distances does not reflect a real spatial unit such as km^2. Therefore, while the variable is significant, these results should be interpreted with care.

Hypotheses	Expected Effect	Regression Result
Hypothesis 1: Closer mean distances between wineries, increase the knowledge flow and lead, on average, to higher scores.	Significant, positive effect	Supported
Hypothesis 2: The average score of wineries will, up to a certain point, be positively related to the number of wineries in a cluster and will thereafter exhibit a negative relationship.	Significant, non-linear relation to score	Partially. Significant, but a negative linear relation to score
Hypothesis 3: The score of wineries will be negatively related to the spatial size of the cluster	Significant, negative effect	Partially. Significant, but a positive relation to score

Table 7.1: **Overview of hypotheses and final results**

7.4 Other Findings

The regression also contains other interesting results. First, wineries with more variation in their distances to others have on average higher scores. This suggests that a winery benefits from a few close neighbors as a higher potential for face-to-face interaction increases possible knowledge spillovers (Jaffe et al. 1993) and trust between actors (Arndt & Sternberg 2000). These dynamics have also been found to exist in the wine industry (Yang et al. 2012) and support the evidence found by H_1. While the advantages of a few close exchange partners can be considered the main effect, longer distances in a cluster might be beneficial as well: As highlighted in the literature review (see section 2, pp.7), research related to wine such as Boschma's (2005) and Turner's (2010) papers, finds evidence for the importance of intra- *and* inter-cluster ties. This argument relates back to Granovetter's (1973) theory of the advantages of weak ties, as having exclusively close neighbors brings the danger of redundant information within clusters but weak ties to others can potentially provide new knowledge from other sources outside the region.

While the analysis of this paper lacks data on social inter-cluster relationships, the effect of distance within the cluster might exhibit similar dynamics. Short distances to one's neighbors, as found by hypothesis 1, have a positive effect on score, which is argued by a stronger social network. But a winery with *both,* that is an actor with a few close neighbors and other cluster members further away, might gain some new information by having firms further away but still in the same cluster. The reasoning is that wineries, regardless their distances, still have important commonalities as they live in the same cluster defined by zip code areas. They live and work in the same institutional environment which is argued to be of importance (Saxenian 1994; M. E. Porter 2000; Maskell 2001; Bell & Zaheer 2007). Thus, higher variation of distances might affect score due to the presence of both close and far distances in a cluster. Further research is however needed to prove the validity of this theory.

Another interesting finding is that performance on a cluster level also seems to improve scores. Thus, wineries appear to benefit when their respective cluster has won many awards. This indicates that knowledge circulates in these agglomerations. Acknowledging that a possible bias of the jury could also increase the score of a wine when it comes from a famous region, the test conditions make this less likely as the rating is conducted by a blind tasting (DLG 2017b). When accounting for the number of actors within a cluster (see Appendix D, pp.98), the results suggest that achievements are more dominant for clusters with fewer members. In large clusters, the effect probably becomes diffused while with fewer actors, ties are closer, less selective and mutual learning is more likely.

To conclude, while not all variables establish a significant effect as expected, the analysis finds empirical evidence that the spatial network matters when considering its effect on wine rankings and thus quality. To revisit the research question:

"*Is there a relationship between the spatial networks of wineries in a cluster and the quality of wine*",

we can hence conclude that the answer is yes. On the individual node level, average proximity between wineries is found to affect scores as do wineries' numbers and the spatial dimension on the cluster level.

The findings of this thesis thus support other studies such as Fanfani's (1994) research on parmesan clusters as the here found results indicate that know-how and cooperation seem to matter in agricultural clusters. The thesis' results also back Murdoch's (2000) opinion that despite new communication technologies, local networks are still of importance. Overall, wineries seem to benefit from a selective but dense and close spatial network in which they can maintain tight social relations to a few others. Thus, local knowledge is a crucial factor, which is facilitated by shorter distances between wineries, lower numbers of competitors in a cluster and previous achievements in national competitions. While the spatial dimension should be interpreted with caution, the finding that wineries with short *and* long distances within a cluster have higher scores leaves room for future research.

7.5 Limitations and Future Research

While the study uses a self-built database, this approach presents limitations, broadly regarding the definition of the spatial network and quality. In terms of clusters, Borgatti & Halgin (2011) state rightly that clusters do not have natural boundaries, they are defined by the researcher. The subjects of this study are thus grouped according to zip code areas. While this ensures a more detailed division of space than wine regions, it ignores that wineries, which are close to borders between zip code areas, will not have any relation in the database. As the database only considers intra-cluster variables, inter-cluster relations and possibly interesting dynamics stay unobserved. For instance, taking Burt's (1992) theory of structural holes, it would be intriguing to see if the individual wineries, or clusters for this matter, take such positions between networks to gain certain advantages.

On another note, it can be argued that wineries close to each other not only share social bonds, but also geological conditions such as sunny slopes, which could affect the wine quality. This very detailed information is difficult to obtain but could contribute to the final product and should be considered for future research. Lastly, due to the scope of the paper, only official wine regions are included to account for fixed effects, however, a more detailed division, such as districts, could be added to account for certain industry-specific factors.

In terms of quality, the DLG award only takes one of Germany's classification systems, that is, QiG, into account. Other awards in Germany, such as the Gault& Millau Guide, are also influential and could add to the findings if included in the dataset. Furthermore, the DLG award chosen for this analysis might represent the German consumer market very well, but considering the country's strong international focus, wineries selling their wine abroad might not be represented by the DLG award, nor the database of the DWI.

The dataset offers potential for further studies as it can be extended by variables to create a better understanding of the effect of spatial networks on organization, specifically for the wine industry.

For instance, more detailed information on wineries could be added, such as size of firms, which is found to have an impact (Diez-Vial 2011) or organizational form, such as private wineries and cooperatives (Schamel 2015; 2014). As Staber (1998) states, research within economic geography should also add idiosyncratic factors, such as labor regulations, traditions and institutions. Cities close to or within the clusters could also give new insights, as wineries might profit from the attracted pool of labor. In terms of tourism, the database could offer a starting point to see if the spatial network has an effect and this might be considered when regions implement tourism routes. Further, due to the scope of the paper spatial cluster size is reflected by the sum of distances. As this does not ideally reflect space, further research might add official information such as square meters. Lastly, while the country focus of this study is Germany, the methodology can be easily applied to other countries as well. Comparative studies could therefore gain new insights when implemented in other national settings.

8. Conclusion

This study is believed to be one of the first to assess spatial networks' effect on quality in the German wine industry. By using actual geographical distances between 1863 wineries located in one of the 13 official wine regions, award winners were identified and a network analysis was applied to assess impact of spatial networks on success. The results of the analysis suggest that spatial networks have an effect on quality. Specifically, wineries seem to thrive when they are located in clusters with a few actors and short distances to their neighbors. These findings indicate that trust and knowledge flow best through close ties to neighbors and ultimately increase expertise which results in better products. The analysis also provides other interesting findings for future research: On the individual actor level, wineries are found to produce better quality when they have both short and long distances to others in their cluster. This suggests that close neighbors foster social ties and thus communication. Further, long distances within the same cluster provide potentially diverse knowledge. On the cluster level, space, seems to matter. Thus, more space is found to increase quality. While this potentially indicates that lower competition is beneficial, future research should assess the role of space in agricultural clusters in more detail. Further, performance of a cluster in general is found to increase the average score of the wineries within it. This strongly suggests that knowledge sharing exists within these agglomerations.

The findings of this paper support previous research but also provide new insights in terms of the relation between economic geography, social networks and performance, as well as new knowledge regarding the German wine industry. While this paper does not reject that soil, climate and a winery's individual resources are factors for quality, it nevertheless argues that spatial proximity enhances communication and knowledge sharing and that wine production is strongly dependent on human expertise by such. The results provide support in favor of this argument as proximity is found to have a positive effect on quality, while regions do not. In an increasingly competitive market, local knowledge seems to matter; and wineries, as well as regional and national associations, can make use of these findings by actively enforcing communication, across neighborhoods, clusters and regions. By identifying where competitive advantages lie, and by acting upon it, local and national advantages can emerge. The results also call for more research on spatial proximity within clusters and its effect on performance in various industries and countries. Competition is global but the findings of this thesis show that advantages are embedded in localities. To explore these dynamics in more detail should be the subject of future studies.

9. References

AAWE, 2017. The Journal of Wine Economics | American Association of Wine Economists. *American Association of Wine Economics*. Available at: http://www.wine-economics.org/journal/ [Accessed July 25, 2017].

Adams, J.D., 2002. Comparative localization of academic and industrial spillovers. *Journal of Economic Geography*, 2, pp.253–278. Available at: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/joeg/2/3/10.1093/jeg/2.3.253/2/020253.pdf?Expires=1490349810&Signature=VFD03seRBPZB5fs9s8BVy7~FByeF443ClSsguzfL2itREz52Yc7cWrFwVwfvQJdpHDN5WAb~dcw6kChNFCpsPXHxeorV6jCSn~oVpks58~Pw0shbHW6l [Accessed March 20, 2017].

Agrawal, A., Cockburn, I. & Mchale, J., 2006. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography Advance Access published on*, 6, pp.571–591. Available at: https://static1.squarespace.com/static/578cf5ace58c62ac649ec9ce/t/5820d39bd482e9a9a704f3e7/1478546332970/10.1.1.507.2595.pdf [Accessed April 15, 2017].

Ahuja, G., 2000. Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quartely*, 45(3), pp.425–455. Available at: https://utexas-ir.tdl.org/bitstream/handle/2152/29302/CollaborationNetworks.pdf?sequence=1&isAllowed=y [Accessed May 16, 2017].

Amin, A. & Roberts, J., 2008. Knowing in action: Beyond communities of practice. *Research Policy*.

Antonio Belso-Martínez, J. & Morrison, A., 2015. The Dynamics of Technical and Business Knowledge Networks in Industrial Clusters: Embeddedness, Status, or Proximity? *Economic Geography*, 92(1), pp.35–60. Available at: https://www.researchgate.net/profile/Andrea_Morrison/publication/287119689_The_Dynamics_of_Technical_and_Business_Knowledge_Networks_in_Industrial_Clusters_Embeddedness_Status_or_Proximity/links/57055e9d08ae44d70ee34253.pdf [Accessed March 20, 2017].

Arndt, O. & Sternberg, R., 2000. Do Manufacturing Firms Profit from Intraregional Innovation Linkages? An Empirical Based Answer. *European Planning Studies*. Available at: http://www.tandfonline.com/action/journalInformation?journalCode=ceps20 [Accessed March 5, 2017].

Arthur, W.B., 1990. "Silicon Valley" Locational Clusters: When Do Increasing Returns Imply Monopoly? *Mathematical Social Sciences*, 19(3), pp.235–273. Available at: www.santafe.edu [Accessed April 22, 2017].

Asheim, B., 1996. *Industrial districts as "learning regions". A condition for prosperity?*, Seoul. Available at: https://brage.bibsys.no/xmlui/bitstream/handle/11250/226797/STEPrapport3-1995.pdf?sequence=1&isAllowed=y [Accessed April 14, 2017].

Asheim, B.T., 1994. *Industrial districts, inter-firm cooperation and endogenous technological development: the experience of developed countries*

Asheim, B.T. & Isaksen, A., 2002. Regional Innovation Systems: The Integration of Local "Sticky" and Global "Ubiquitous" Knowledge. *Journal of Technology Transfer*, 27, pp.77–86. Available at: http://download.springer.com/static/pdf/262/art%253A10.1023%252FA%253A1013100704794.pdf?ori ginUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1023%2FA%3A1013100704794&token2=exp=1492095526~acl=%2Fstatic%2Fpdf%2F262%2Fart%25253A10.1023%25252FA%25253A1013 [Accessed April 13, 2017].

Astley, W.G., 1985. The Two Ecologies: Population and Community Perspectives on Organizational Evolution The Two Ecologies: Population and Com- munity Perspectives on Organizational Evolution. *Source: Administrative Science Quarterly*, 30(2), pp.224–241. Available at: http://www.jstor.org/stable/2393106 [Accessed July 8, 2017].

Audia, P.G., 2000. The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States. *American Journal of Sociology*, 106(2), pp.424–462. Available at: http://www.jstor.org/stable/10.1086/316962 [Accessed July 8, 2017].

Audretsch, D.B. & Feldman, M.P., 1996. R&D Spillovers and the Geography of Innovation and Production R&D Spillovers and the Geography of Innovation and Production. *Source: The American Economic Review*, 86(3), pp.630–640. Available at: http://www.jstor.org/stable/2118216 [Accessed March 25, 2017].

Aylward, D.K., 2006. Innovation lock-in: unlocking research and development path dependency in the Australian wine industry. *Strategic Change*, 15, pp.7–8. Available at: http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1237&context=commpapers [Accessed March 28, 2017].

Balland, P.-A., 2012. Proximity and the Evolution of Collaboration of Networks: Evidence from research and Development Projects within the Global Navigation Satellite System Industry. *Regional Studies*, 46(6), pp.741–756. Available at: http://www.tandfonline.com/doi/pdf/10.1080/00343404.2010.529121?needAccess=true [Accessed March 28, 2017].

Baptista, R. & Swann, P., 1998. Do firms in clusters innovate more? *Research Policy*, 27, pp.525–540. Available at: http://in3.dem.ist.utl.pt/mscdesign/03ed/files/other_8_03.pdf [Accessed April 25, 2017].

Barnes, T. & Gertler, M.S., 1999. *The New Industrial Geography: Regions, Regulation and Institutions - Google Books*, London. Available at: https://books.google.dk/books?hl=en&lr=&id=2baBAgAAQBAJ&oi=fnd&pg=PA1&dq=Industrial+geography,+institutional+economics+barnes&ots=eSs7ExvxEC&sig=px7SpgAKfYrk_3OhdVUYyDqK30Q&redir_esc=y#v=snippet&q=geographical sensibility&f=false [Accessed April 14, 2017].

Barthélemy, M., 2011. Spatial networks. *Physics Reports*, 499, pp.1–101. Available at: http://ac.els-cdn.com.esc-web.lib.cbs.dk/S037015731000308X/1-s2.0-S037015731000308X-main.pdf?_tid=7531d3ec-6e4b-11e7-bc09-00000aacb360&acdnat=1500665782_2530358a556d25eb8327214f258c5440 [Accessed July 21, 2017].

Bathelt, H. & Glückler, J., 2003. Toward a relational economic geography. *Journal of Economic Geography*, 3, pp.117–144. Available at: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/joeg/3/2/10.1093/jeg/3.2.117/3/030117.pdf?Expires=1492613062&Signature=LwsNE3E16KIBcYOAnxJ791Iw4p7sDGOec7YxaRnQhN8Oz~t7HfSTOlIi7g0UD~79D9-DQyUvz~ljAGOCB~k3tUgOfI8tBn4asU2YhFccyarwtBhv1nH3 [Accessed April 15, 2017].

Baum, J.A.C., Shipilov, A. V & Rowley, T.J., 2003. Where do small worlds come from? *Industrial and Corporate Change2*, 12(4). Available at: https://www.researchgate.net/profile/Tim_Rowley2/publication/5212449_Where_do_small_worlds_co me_from/links/00b7d5303cae7eb29b000000.pdf [Accessed May 16, 2017].

Bell, G.G. & Zaheer, A., 2007. Geography, Networks, and Knowledge Flow. *Institute for Operations Research and the Managment Sciences*, 18(6), pp.955–972. Available at: http://www.jstor.org/stable/25146152 [Accessed March 28, 2017].

Benjamin, B.A. et al., 1999. Status, Quality, and Social Order in the California Wine Industry. *Source: Administrative Science Quarterly*, 44225166(3), pp.563–589. Available at: http://www.jstor.org/stable/2666962 [Accessed March 23, 2017].

BGBl, 1971. *Bundesgesetzblatt*, Available at: https://www.bgbl.de/xaver/bgbl/start.xav?start=//*%5B@attr_id=%27bgbl171063.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl171063.pdf%27%5D__1499067099347 [Accessed July 3, 2017].

Bing, 2017. Bing Maps - Directions, trip planning, traffic cameras & more. Available at: http://www.bing.com/maps [Accessed August 9, 2017].

Bock, O., 2016. Kaum ein Wein ohne Auszeichnung. *Frankfurter Allgemeine Zeitung*, p.430.

Borgatti, S.P. & Halgin, D.S., 2011. On Network Theory. *Organization Science*, 22(5), pp.1168–1181. Available at: http://dx.doi.org/10.1287/orsc.1100.0641 [Accessed July 23, 2017].

Boschma, R., 2005. Proximity and Innovation: A Critical Assessment. *Regional Studies*, 39(391), pp.61–74. Available at: http://www.informaworld.com/smpp/title~content=t713393953 [Accessed March 28, 2017].

Breschi, S. & Lissoni, F., 2001. Knowledge spillovers and local innovation systems: A critical survey. *Industrial and Corporate Change*, 10(4). Available at: http://people.unica.it/stefanousai/files/2015/10/breschi-lissoni-icc-2001.pdf [Accessed March 25, 2017].

Brito, C., 2006. A network perspective of the port wine sector. *IJWM International Journal of Wine Marketing*, 18(2), pp.124–138. Available at: https://www.researchgate.net/profile/Carlos_Brito5/publication/247623657_A_network_perspective_of_the_port_wine_sector/links/00b4953263734315850000000.pdf [Accessed March 25, 2017].

Brunke, H. et al., 2016. *Tested Quality in the Glass:*, Available at: http://www.wine-economics.org/aawe/wp-content/uploads/2016/03/AAWE_WP193.pdf [Accessed June 27, 2017].

Brunori, G. & Rossi, A., 2000. Synergy and Coherence through Collective Action: Some Insights from Wine Routes in Tuscany. *Sociologia Ruralis*, 40(4).

Bruwer, J., 2003. South African wine routes: some perspectives on the wine tourism industry's structural dimensions and wine tourism product. *Tourism Management*, 24, pp.423–435.

Burt, R.S., 1992. *Structural holes : the social structure of competition*, Harvard University Press. Available at: https://books.google.dk/books?hl=en&lr=&id=FAhiz9FWDzMC&oi=fnd&pg=PR5&dq=burt+structural+holes&ots=vISYct8PMJ&sig=sE65lFrpIR7-GkiF7ZA2FgBHZ6A&redir_esc=y#v=onepage&q=burt structural holes&f=false [Accessed April 3, 2017].

Butler, J. & Hansen, G., 1991. Network evolution, entrepreneurial success, and regional development. *Entrepreneurship and Regional Development*, 3, pp.1–16. Available at: http://www-tandfonline-com.esc-web.lib.cbs.dk/doi/pdf/10.1080/08985629100000001?needAccess=true [Accessed March 5, 2017].

C Baum, J.A. & Mezias, S.J., 1992. Localized Competition and Organizational Failure in the Manhattan Hotel Industry, 1898- Localized Competition and Organizational Failure in the Manhattan. *Source: Administrative Science Quarterly*, 37(4), pp.580–604. Available at: http://www.jstor.org/stable/2393473 [Accessed July 8, 2017].

Canina, L., Enz, C.A. & Harrison, J.S., 2005. AGGLOMERATION EFECTS AND STRATEGIC ORIENTATIONS: EVIDENCE FROM THE U.S. LODGING INDUSTRY. *Academy of Management Journal*, 48(4), pp.565–581. Available at: http://amj.aom.org/cgi/doi/10.5465/AMJ.2005.17843938 [Accessed April 25, 2017].

CBI, 2016. *CBI Market channels and segments: Wine in Germany*, Available at: https://www.cbi.eu/sites/default/files/market_information/researches/channels-segments-germany-wine-2016.pdf [Accessed March 3, 2017].

Chiffoleau, Y. & Touzard, J.-M., 2014. Understanding local agri-food systems through advice network analysis. *Agri Hum Values*, 31, pp.19–32. Available at: http://download.springer.com/static/pdf/761/art%253A10.1007%252Fs10460-013-9446-6.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs10460-013-9446-6&token2=exp=1494417872~acl=%2Fstatic%2Fpdf%2F761%2Fart%25253A10.1007%25252Fs10460-013-944 [Accessed May 10, 2017].

Chung, W. & Kalnins, A., 2001. Agglomeration Effects and Performance: A Test of the Texas Lodging Industry. *Strategic Management Journal Strat. Mgmt. J*, 22, pp.969–988. Available at: http://www.jstor.org/stable/pdf/3094379.pdf [Accessed April 25, 2017].

Cowan, R., Jonard, N. & Özman, M., 2004. Knowledge dynamics in a network industry. *Technological Forecasting & Social Change*, 71, pp.469–484. Available at: http://ac.els-cdn.com/S0040162503000453/1-s2.0-S0040162503000453-main.pdf?_tid=3e10ef7a-1b7f-11e7-83ac-00000aab0f6b&acdnat=1491562077_a8bc484c361f599513b83f930853b43d [Accessed April 7, 2017].

Csardi, G. & Nepusz, T., 2006. The igraph software package for complex network research. , p.1695. Available at: http://igraph.org.

Deutsche Post AG, 2017. Die Geschichte der PLZ. Available at: https://www.deutschepost.de/de/p/plz-buch/plz_finden/die_geschichte_der.html.

Deutsche Winzergenossenschaften, 2017. Deutsche Winzergenossenschaften. Available at: http://www.deutsche-winzergenossenschaften.de/start.html [Accessed July 3, 2017].

Diez-Vial, I., 2011. Geographical cluster and performance: The case of Iberian ham. *Food Policy*, 36, pp.517–525. Available at: http://ac.els-cdn.com/S0306919211000613/1-s2.0-S0306919211000613-main.pdf?_tid=880018dc-29ac-11e7-86d4-00000aab0f6b&acdnat=1493120845_4382f1f4b3a92963a750b137e4402536 [Accessed April 25, 2017].

DLG, 2017a. DLG e.V. - About DLG. Available at: http://www.dlg.org/about.html [Accessed July 4, 2017].

DLG, 2017b. DLG e.V. - Die DLG testet Wein und Sekt. Available at: http://www.dlg.org/wein.html

[Accessed June 2, 2017].

DLG, 2017c. DLG e.V. - TOP 100 Weinerzeuger. Available at: http://www.dlg.org/top100_weinerzeuger.html [Accessed July 4, 2017].

DLG, 2017d. Was bedeutet DLG-prämiert? Available at: http://www.dlg-verbraucher.info/de/dlg-testzentrum/dlgpraemiert.html [Accessed July 5, 2017].

DLG, 2017e. Wein & Sekt. Available at: http://www.dlg-verbraucher.info/de/testergebnisse/wein-sekt.html [Accessed July 5, 2017].

Duarte Alonso, A., 2011. Standing Alone You Can't Win Anything: The Importance of Collaborative Relationships for Wineries Producing Muscadine Wines. *Journal of Wine Research*, 22(1), pp.43–55. Available at: http://www-tandfonline-com.esc-web.lib.cbs.dk/doi/pdf/10.1080/09571264.2011.550761?needAccess=true [Accessed April 14, 2017].

DWI, 2009. Aktuelles Weinrecht. Available at: http://www.deutscheweine.de/fileadmin/user_upload/Website/Intern/Dozentenportal/Aktuelles_Weinrecht.pdf [Accessed July 3, 2017].

DWI, 2017a. DWI. Available at: http://www.germanwines.de/about-us/who-we-are/ [Accessed June 2, 2017].

DWI, 2017b. Güteklassen. Available at: http://www.deutscheweine.de/wissen/qualitaet-standards/gueteklassen/ [Accessed July 3, 2017].

DWI, 2017c. History. Available at: http://www.germanwines.de/knowledge/viticulture-winemaking/history/ [Accessed June 28, 2017].

DWI, 2017d. Must weights. Available at: http://www.germanwines.de/knowledge/quality-standards/must-weights/ [Accessed June 28, 2017].

DWI, 2017e. Work in the Wineyard & Cellar. *The German Wine Institute*. Available at: http://www.germanwines.de/knowledge/viticulture-winemaking/work-in-the-wineyard-cellar/ [Accessed August 5, 2017].

EC, 2017. *EU Wine Market Data Portal | Agriculture and rural development*, Available at: https://ec.europa.eu/agriculture/wine/statistics_en [Accessed June 28, 2017].

Euromonitor, 2017. *Euromonitor International - Analysis*, Available at: http://www.portal.euromonitor.com.esc-web.lib.cbs.dk/portal/analysis/tab [Accessed July 3, 2017].

Fanfani, R., 1994. Agricultural restructuring and rural change in Europe. Available at: http://edepot.wur.nl/318085#page=87 [Accessed April 2, 2017].

Fernhaber, S., Gilbert, B.A. & McDougall, P., 2007. International entrepreneurship and geographic location: an empirical examination of new venture internationalization. *Journal of International Business Studies*, 39(2), pp.267–290. Available at: http://www.palgrave-journals.com/doifinder/10.1057/palgrave.jibs.8400342.

Folta, T.B., Cooper, A.C. & Baik, Y.-S., 2006. Geographic cluster size and firm performance. Available at: http://koasas.kaist.ac.kr/bitstream/10203/6370/1/2006-111.pdf [Accessed April 18, 2017].

Foster, A.D. & Rosenzweig, M.R., 1995. Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Source Journal of Political Economy*, 10369(6), pp.1176–1209. Available at: http://www.jstor.org [Accessed April 3, 2017].

Freel, M.S., 2003. Sectoral patterns of small firm innovation, networking and proximity. *Research Policy*, 32, pp.751–770. Available at: http://ac.els-cdn.com/S0048733302000847/1-s2.0-S0048733302000847-main.pdf?_tid=32714fc4-0186-11e7-a028-00000aacb35f&acdnat=1488706334_d1e199063098af86f1ee0277727cc93f [Accessed March 5, 2017].

Freeman, J.H. & Audia, P.G., 2006. COMMUNITY ECOLOGY AND THE SOCIOLOGY OF ORGANIZATIONS. *Annu. Rev. Sociol*, 32, pp.145–69. Available at: https://www.researchgate.net/profile/Pino_Audia/publication/228215874_Community_Ecology_and_the_Sociology_of_Organizations/links/00b4952ef972d9d018000000.pdf [Accessed July 8, 2017].

Friedman, T.L., 2005. It's a Flat World, After All. *The New York Times*. Available at: http://www.wright.edu/~tdung/Friedman_excerp.pdf [Accessed June 5, 2017].

Ganesan, S., Malter, A.J. & Rindfleisch, A., 2005. Does Distance Still Matter? Geographic Proximity and New Product Development. *Journal of Marketing*, 69(4), pp.44–60. Available at: http://www.jstor.org/stable/30166551 [Accessed March 28, 2017].

Gault Millau, 2017. Weinguide | GaultMillau Deutschland - Guide für Gourmets. Available at: http://gaultmillau.de/weinguide [Accessed July 3, 2017].

Gergaud, O. & Ginsburgh, V., 2008. Natural Endowments, Production Technologies and the Quality of Wines in Bordeaux. Does Terroir Matter? *The Economic Journal*, 118(529), pp.F142–F157. Available at: http://doi.wiley.com/10.1111/j.1468-0297.2008.02146.x [Accessed March 10, 2017].

German Mission in the U.S., 2017. German Missions in the United States - History & Facts. Available at: http://www.germany.info/Vertretung/usa/en/04__W__t__G/04/04/01/Feature__1.html [Accessed July 7, 2017].

Gittelman, M., 2007. Does Geography Matter for Science-Based Firms? *Geography of Research and Patenting in Biotechnology Author*, 18(4), pp.724–741. Available at: http://www.jstor.org/stable/25146132 [Accessed April 19, 2017].

Giuliani, E., 2010. Clusters, Networks and Economic development: an evolutionary economics perspective. In *The Handbook of Evolutionary Economic Geography*. pp. 261–280. Available at: http://www.univpgri-palembang.ac.id/perpus-fkip/Perpustakaan/East Phylosopy/Economic Geography/THE HANDBOOK OF EVOLUTIONARY ECONOMIC.pdf#page=131 [Accessed March 7, 2017].

Giuliani, E., 2013. Network dynamics in regional clusters: Evidence from Chile. *Research Policy*, 42(8), pp.1406–1419. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733313000796 [Accessed April 14, 2017].

Giuliani, E., 2007a. The selective nature of knowledge networks in clusters: evidence from the wine industry. *Journal of Economic Geography*, 7, pp.139–168. Available at: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/joeg/7/2/10.1093/jeg/lbl014/2/lbl014.pdf?Expires=1494692468&Signature=GnVU0Delv-dQ2bGdBGujRRVZURuIaqcbLe2p~6npBLkXVDpUebW9q6IOed65z6EVz8G9ejpCOR5eDzWlvGe~n2kBurBVfUspkWpVthcXOU4QaxGQmzuJwU8WGIpv9aOvSoWr6SG4WBKOMrFPqq4FGmAxZf5lVrx3iiMVlKeAlZD-

b7u8HM9LWOfiSA6iJjJPVDaa2NfXiBrYHmQgvs9lKlXg0xuSvmmkK6W3XNop6GsCH4uFtP98ml8RaBgEbwufDcuhoiXvKu7B6S8zcvBQdBtZ5oNnUKYl~LcoQapUSUE-LIqNK-Uz25taonCsE34JNp1eFyDEKofe2mOXZq-VqA__&Key-Pair-Id=APKAIUCZBIA4LVPAVW3Q [Accessed May 12, 2017].

Giuliani, E. & Arza, V., 2009. What drives the formation of "valuable" university–industry linkages?: Insights from the wine industry. *Research Policy*, 38(6), pp.906–921.

Giuliani, E. & Bell, M., 2005. The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster. *Research Policy*, 34, pp.47–68. Available at: http://ac.els-cdn.com.esc-web.lib.cbs.dk/S0048733304001441/1-s2.0-S0048733304001441-main.pdf?_tid=cdcce8ca-1077-11e7-89b4-00000aacb360&acdnat=1490349419_0b8558788c4380ccd319215112134121 [Accessed March 24, 2017].

Giuliani, E. & Matta, A., 2013. Impact evaluation with social network analysis methods, program for supply chain development in the province of Cordoba. *ATN*. Available at: http://www.fomin.org/Portals/0/Impact Evaluation/MIF2013_Impact_SupplyChainsCordoba_Giuliani.pdf [Accessed March 29, 2017].

Glückler, J., 2007. Economic geography and the evolution of networks. *Journal of Economic Geography*, pp.1–16. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.819.595&rep=rep1&type=pdf [Accessed March 2, 2017].

Granovetter, M., 2005. The Impact of Social Structure on Economic Outcomes. *Journal of Economic Perspectives*, 19(1), pp.33–50. Available at: http://pubs.aeaweb.org/doi/pdfplus/10.1257/0895330053147958 [Accessed July 12, 2017].

Granovetter, M., 1985. The Problem of Embeddedness. *American Journal of Sociology*, 91(3), pp.481–510. Available at: http://www.jstor.org/stable/2780199 [Accessed July 13, 2017].

Granovetter, M.S., 1973. The Strength of Weak Ties. *American Journal of Sociology*, 78(6), pp.1360–1380. Available at: https://sociology.stanford.edu/sites/default/files/publications/the_strength_of_weak_ties_and_exch_w-gans.pdf [Accessed April 13, 2017].

Grubesic, T.H., 2008. Zip codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, 42, pp.129–149. Available at: http://ac.els-cdn.com/S0038012106000516/1-s2.0-S0038012106000516-main.pdf?_tid=3bc263c0-0d5b-11e7-9afe-00000aab0f01&acdnat=1490007295_830bcc16017b1c20074ef9372e50712f [Accessed March 20, 2017].

Guerrieri, P. & Pietrobelli, C., 2004. Industrial districts' evolution and technological regimes: Italy and Taiwan. *Technovation*, 24, pp.899–914. Available at: www.elsevier.com/locate/technovation [Accessed March 30, 2017].

Hall, C.M. et al., 1997. Wine Tourism and Network Development in Australia and New Zealand: Review, Establishment and Prospects. *International Journal of Wine Marketing*, 9, pp.5–31.

Hanf, J., Schweickert, E. & Belaya, V., 2009. *Power asymmetries in the wine chain Implications for the management of the chain*, Leibnitz. Available at: https://pdfs.semanticscholar.org/e974/7b34865cd5f7eab2d9727bd5e24e6fe1f71e.pdf [Accessed April 27, 2017].

Hanf, J.H., Gagalyuk, T. & Schweickert, E., 2004. Success of Co-operatives: Thoughts on German Wine Co-operatives. Available at: http://www.embassyconferences.co.za/media/file_icons/Success of Co-operatives Thoughts on German Wine Co-operatives.pdf [Accessed July 3, 2017].

Hanf, J.H. & Iselborn, M., 2014. How to deal with quality problems of German wine cooperatives -A double principal-agent approach, Geisenheim. Available at: http://academyofwinebusiness.com/wp-content/uploads/2014/07/BM05_Iselborn_Maximilian.pdf [Accessed April 27, 2017].

Hanf, J.H. & Schweickert, E., 2014. Cooperatives in the balance between retail and member interests: the challenges of the German cooperative sector. *Journal of Wine Research*, 25(1), pp.32–44. Available at: http://www.tandfonline.com/doi/abs/10.1080/09571264.2014.871122 [Accessed March 4, 2017].

Hannan, M.T. & Freeman, J., 1977. The Population Ecology of Organizations. *American Journal of Sociology*, 82(5), pp.929–964. Available at: http://www.jstor.org/stable/2777807 [Accessed March 5, 2017].

Herrera-Yagüe, C. et al., 2015. The anatomy of urban social networks and its implications in the searchability problem. *Scientific reports*, 5, p.10265. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26035529 [Accessed March 20, 2017].

Hira, A. & Swartz, T., 2014. What makes Napa Napa? The roots of success in the wine industry. *Wine Economics and Policy*, 3, pp.37–53. Available at: www.sciencedirect.com [Accessed March 29, 2017].

Ingram, P. & Inman, C., 1996. Institutions, Intergroup Competition, and the Evolution of Hotel Populations around Niagara Falls. *Sage Publications*, 41(4), pp.629–658. Available at: http://www.jstor.org.esc-web.lib.cbs.dk/stable/pdf/2393870.pdf?refreqid=excelsior%3Aa97e8a6ae3faef0e0dfc0f64c81fff48 [Accessed July 8, 2017].

Jaffe, A.B., Trajtenberg, M. & Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Source: The Quarterly Journal of Economics*, 108(3), pp.577–598. Available at: http://www.jstor.org/stable/2118401 [Accessed March 29, 2017].

Jiaming, L. et al., 2015. Industrial Spatial Agglomeration Using Distance-based Approach in Beijing, China. *Chinese Geographical Science*, 25(256), pp.698–712. Available at: www.springerlink.com/content/1002-0063 [Accessed May 2, 2017].

Karlsson, C. & Olsson, O., 1998. Product Innovation in Small and Large Enterprises. *Source: Small Business Economics*, 10(1), pp.31–46. Available at: http://www.jstor.org/stable/40228568 [Accessed March 25, 2017].

Kenney, M., 2000. *Understanding Silicon Valley: The Anatomy of an Entrepreneurial Region*, Stanford University Press. Available at: https://books.google.dk/books?hl=en&lr=&id=CLxzUW4V_2cC&oi=fnd&pg=PR9&dq=silicon+valley&ots=HYP1KmeBXw&sig=61Oyzx5YTs3pGRU3hB_Kpn7XPf8&redir_esc=y#v=snippet&q=cluster&f=false [Accessed June 5, 2017].

Koch, J., Martin, A. & Nash, R., 2013. Overview of perceptions of German wine tourism from the winery perspective. *International Journal of Wine Business Research Iss International Journal of Contemporary Hospitality Management*, 25(2), pp.50–74. Available at: http://dx.doi.org/10.1108/17511061311317309 [Accessed February 25, 2017].

Laumann, E.O., Marsden, P. V. & Prensky, D., 1983. *The boundary specification problem in network analysis*, Beverly Hills: Sage.

Lin, N., 1990. Social resources and social mobility: A structural theory of status attainment. In *Social Mobility and Social Structure*. New York, pp. 247–271.

Lockshin, L. et al., 2005. Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice. *Food Quality and Preference*, 17, pp.166–178. Available at: http://ac.els-cdn.com/S0950329305000339/1-s2.0-S0950329305000339-main.pdf?_tid=97b8a746-03f3-11e7-a13e-00000aacb361&acdnat=1488973221_3dcbaabe59daf911319d98b555de5d8a [Accessed March 8, 2017].

Malmberg, A., 1997. Industrial geography: location and learning. *Progress in Human Geography*, 21(4), pp.573–582. Available at: http://journals.sagepub.com.esc-web.lib.cbs.dk/doi/pdf/10.1191/030913297666600949 [Accessed July 8, 2017].

Markusen, A., 1996. Sticky Places in Slippery Space: A Typology of Industrial Districts. *Economic Geography*, 72(3), pp.293–313. Available at: ftp://131.252.97.79/Transfer/ES_Pubs/ESVal/econ_geog/markusen_1996_industrialTypology_v72_3_p293.pdf [Accessed April 2, 2017].

Marshall, A., 1920. *Industry and Trade: A Study of industrial technique and businnes organization; and of their influences on the condition of various classes and nations*, Available at: http://socserv2.mcmaster.ca/~econ/ugcm/3ll3/marshall/Industry%26Trade.pdf [Accessed April 7, 2017].

Marshall, A., 1890. *The Principles of Economics* 8th ed., Available at: http://download.springer.com.esc-web.lib.cbs.dk/static/pdf/690/bok%253A978-1-137-37526-1.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Fbook%2F10.1057%2F9781137375261&token2=exp=1492025680~acl=%2Fstatic%2Fpdf%2F690%2Fbok%25253A978-1-137-37526-1.pdf%3Forig [Accessed April 12, 2017].

Maskell, P., 2001. Towards a Knowledge-based Theory of the Geographical Cluster. *Industrial and Corporate Change*, 10(4). Available at: https://myweb.rollins.edu/tlairson/pek/knowclusters.pdf [Accessed April 14, 2017].

Merriam-Webster, 2017. Winery | Definition of Winery by Merriam-Webster. Available at: https://www.merriam-webster.com/dictionary/winery [Accessed June 5, 2017].

Migone, A. & Howlett, M., 2010. *Comparative Networks and Clusters in the Wine Industry*, Available at: http://www.wine-economics.org/aawe/wp-content/uploads/2012/10/AAWE_WP62.pdf [Accessed March 5, 2017].

Morrison, A. & Rabellotti, R., 2009. Knowledge and Information Networks in an Italian Wine Cluster. *European Planning Studies*, 17(7), pp.983–1006. Available at: http://www.tandfonline.com/doi/pdf/10.1080/09654310902949265?needAccess=true [Accessed March 25, 2017].

Mueller, R.A.E. & Sumner, D.A., 2006. *Clusters of Grapes and Wines*, Available at: http://www.agmrc.org/media/cms/Wine_Clusters2_84CD1EE476398.pdf [Accessed April 7, 2017].

Murdoch, J., 2000. Networks - a new paradigm of rural development ? *Journal of Rural Studies*, 16, pp.407–419.

Myles Shaver, J. & Flyer, F., 2000. Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic Management Journal*, 21(12), pp.1175–1193. Available at: http://doi.wiley.com/10.1002/1097-0266%28200012%2921%3A12%3C1175%3A%3AAID-SMJ139%3E3.0.CO%3B2-Q [Accessed April 25, 2017].

Newman, M.E.J., 2003. The structure and function of complex networks. Available at: http://www-personal.umich.edu/~mejn/courses/2004/cscs535/review.pdf [Accessed April 27, 2017].

Oerlemans, L.;, Meeus, M. & Meeus, M.T.H., 2005. Do organizational and spatial proximity impact of firm performance? Do organisational and spatial proximity impact on firm performance? *Regional Studies*, 39(1), pp.89–104. Available at: https://pure.uvt.nl/portal/files/611431/Doorganpostprint.pdf [Accessed March 25, 2017].

Oliver, A.L. & Ebers, M., 1998a. Networking Network Studies: An Analysis of Conceptual Configurations in the Study of Inter-organizational Relationships. *Organization Studies*, 19(4), pp.549–583. Available at: http://journals.sagepub.com/doi/10.1177/017084069801900402.

Orth, U.R. & Krska, P., 2002. Quality signals in wine marketing: the role of exhibition awards. *International Food and Agribusiness Management Review*, 4, pp.385–397. Available at: https://www.ifama.org/resources/Documents/v4i4/Orth-Krska.pdf [Accessed March 5, 2017].

Pfalz, 2017. Weingeschichte | PFALZ.de. Available at: https://www.pfalz.de/wein-und-genuss/weingeschichte# [Accessed June 30, 2017].

Pfanner, E., 2012. After the Debacle Called Liebfraumilch. *The New York Times*. Available at: http://www.nytimes.com/2012/10/13/dining/13iht-wine13.html [Accessed June 28, 2017].

Porter, M.E., 2000. Clusters and Competition. *Oxford handbook of economic geography*, p.197. Available at: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=50ABCB5C96060AC1E5E0B8339D338C22?doi=10.1.1.199.4104&rep=rep1&type=pdf [Accessed April 7, 2017].

Porter, M.E., 1998. Clusters and the New Economics of Competition. *Harvard Business Review*. Available at: http://clustermapping.us/sites/default/files/files/resource/Clusters_and_the_New_Economics_of_Competition.pdf [Accessed March 6, 2017].

Porter, M.E., 2000. Location, Competition, and Economic Development: Local Clusters in a Global Economy. *Economic Development Quarterly*, 14(1), pp.15–34. Available at: http://edq.sagepub.com/cgi/doi/10.1177/089124240001400105 [Accessed March 6, 2017].

Porter, M.E., 1990. The Competitive Advantage of Nations Harvard Business Review. *Harvard Business Review*. Available at: http://www.clustermapping.us/sites/default/files/files/resource/The Competitive Advantage of Nations HBR.pdf [Accessed April 19, 2017].

Porto-Info, 2017. Porto-Info.de. Available at: http://www.porto-info.de/ [Accessed August 13, 2017].

Pouder, R. & John, C.H.S., 1996. Hot Spots and Blind Spots: Geographical Clusters of Firms and Innovation. *Source: The Academy of Management Review*, 21(4), pp.1192–1225. Available at: http://www.jstor.org/stable/259168 [Accessed April 25, 2017].

ProWein, 2017. ProWein Trade Fair. Available at: http://www.prowein.com/cgi-bin/md_prowein/lib/pub/tt.cgi/Exhibitors_by_country.html?oid=32214&lang=2&ticket=g_u_e_s_t [Accessed April 27, 2017].

Rasch, L.A., 2008. *Online Wine Tourism Marketing Efforts of Wineries and Convention and Visitors*. Texas A&M. Available at: http://oaktrust.library.tamu.edu/bitstream/handle/1969.1/85745/Rasch Thesis.pdf [Accessed April 14, 2017].

Rasch, L. & Gretzel, U., 2008. Wineries' Involvement in Promoting Tourism Online: The Case of Texas. *Pasos*, 6(2), pp.317–326. Available at: www.pasosonline.org [Accessed July 9, 2017].

Rosenfeld, S.A., 1997. Bringing Business Clusters into the Mainstream of Economic Development. *European Planning Studies*, 5(1). Available at: https://www.researchgate.net/profile/Stuart_Rosenfeld/publication/237446231_Bringing_business_clusters_into_the_mainstream_of_economic_development/links/544138b40cf2e6f0c0f60511/Bringing-business-clusters-into-the-mainstream-of-economic-development.pdf [Accessed April 15, 2017].

Rosenthal, S.S. & Strange, W.C., 2003. GEOGRAPHY, INDUSTRIAL ORGANIZATION, AND AGGLOMERATION. *The Review of Economics and Statistics*, 85(2), pp.377–393. Available at: http://www.jstor.org/stable/pdf/3211587.pdf [Accessed March 2, 2017].

Rössel, J. & Beckert, J., 2012. Quality Classifications in Competition: Price Formation in the German Wine Market. *Max Planck Institut*. Available at: http://www.mpifg.de/pu/mpifg_dp/dp12-3.pdf [Accessed March 8, 2017].

Ruef, M., 2000. The Emergence of Organizational Forms: A Community Ecology Approach The Emergence of Organizational Forms: A Community Ecology Approach 1. *Source: American Journal of Sociology*, 106(3), pp.658–714. Available at: http://www.jstor.org/stable/10.1086/318963 [Accessed July 8, 2017].

Sáenz-Navajas, M.-P. et al., 2014. Extrinsic attributes responsible for red wine quality perception: A cross-cultural study between France and Spain. *Food Quality and Preference*, 35, pp.70–85. Available at: http://ac.els-cdn.com/S0950329314000238/1-s2.0-S0950329314000238-main.pdf?_tid=5c1c8d14-03f9-11e7-b951-00000aab0f6c&acdnat=1488975698_285ac158bcaefe6887cf3be38f0a14ee [Accessed March 8, 2017].

Sáenz-Navajas, M.-P. et al., 2013. Perception of wine quality according to extrinsic cues: The case of Burgundy wine consumers. *Food Quality and Preference*, 27, pp.44–53. Available at: http://ac.els-cdn.com/S0950329312001085/1-s2.0-S0950329312001085-main.pdf?_tid=a29c2f02-03f3-11e7-8469-00000aacb361&acdnat=1488973239_92e792dad700c8cbe960f143c9f8215c [Accessed March 8, 2017].

Saxenian, A., 1994. Inside-Out: Regional Networks and Industrial Adaptation in Silicon Valley and Route 128. *Cityscape*, 2(2), pp.41–60.

Schamel, G. & Anderson, K., 2001. Wine Quality and Varietal, Regional and Winery Reputations: Hedonic Prices for Australia and New Zealand. Available at: www.vdqs.fed-eco.org [Accessed March 8, 2017].

Schamel, G.H., 2015. Can German wine cooperatives compete on quality? *BIO Web of Conferences*, 5(3003). Available at: https://oiv.edpsciences.org/articles/bioconf/pdf/2015/02/bioconf_oiv2015_03003.pdf [Accessed August 4, 2017].

Schamel, G.H., 2014. Wine quality, reputation, denominations: How cooperatives and private wineries compete? *BIO Web of Conferences*, 3(3008). Available at: https://oiv.edpsciences.org/articles/bioconf/pdf/2014/02/bioconf_oiv2014_03008.pdf [Accessed August 4, 2017].

Schweickert, E. & Hanf, J.H., 2007. *Changes in the wine chain – Managerial challenges and threats for German wine co-ops*.

Seely Brown, J. & Duguid, P., 1991. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science March*, 2(1), pp.40–57. Available at: http://links.jstor.org/sici?sici=1047-7039%281991%292%3A1%3C40%3AOLACTA%3E2.0.CO%3B2-3 [Accessed March 26, 2017].

Seidel, A., Jäckel, M. & Hahn, A., 1978. Deutscher Wein ein Luxusgut? Eine Positionierungsanalyse deutschen Weins im Hochpreissegment in Trier. Available at: http://ubt.opus.hbz-nrw.de/volltexte/2007/432/pdf/Gesamtdokument_24.07.07.pdf [Accessed June 30, 2017].

Skelton, S., 2008. *UK Vineyards Guide 2008*, SP Skelton Ltd.

Smith, R.D., 2006. The network of collaboration among rappers and its community structure. *J. Stat. Mech*. Available at: http://iopscience.iop.org/1742-5468/2006/02/P02006 [Accessed May 16, 2017].

Staber, U., 1998. Organizational Survival in Small-Firm Clusters. *Academy of Management Proceedings*, 1998(1), pp.A1–A7. Available at: http://proceedings.aom.org/cgi/doi/10.5465/APBPP.1998.27643355 [Accessed April 25, 2017].

Stefano, G. Di, King, A.A. & Verona, G., 2016. Too Many Cooks Spoil the Broth? Geographic Concentration, Social Norms, and Knowledge Transfer. *Advances in Strategic Management, Geography, Location, and Strategy*, 36. Available at: http://www.giadadistefano.com/wp-content/uploads/2016/10/Di-Stefano-King-Verona-2016-AiSM.pdf [Accessed April 15, 2017].

Sternberg, R. & Arndt, O., 2001. The Firm or the Region: What Determines the Innovation Behavior of European Firms? *Economic Geography*, 77(4), pp.364–382. Available at: http://www.jstor.org/stable/pdf/3594106.pdf [Accessed March 29, 2017].

Stock, J.H. & Watson, M.W., 2015. *Introduction to Econometrics*, Pearson Education Limited.

Storper, M., 1995. The regional world: Territorial development in a global economy. *Revue Région & Développement*, 1. Available at: http://region-developpement.univ-tln.fr/en/pdf/R1/R1_Storper.pdf [Accessed April 14, 2017].

Storper, M. & Venables, A.J., 2003. Buzz: Face-To-Face Contact and the Urban Economy. *Centre for Economic Performance*. Available at: http://eprints.lse.ac.uk/20008/1/Buzz_Face-to-Face_Contact_and_the_Urban_Economy.pdf [Accessed March 31, 2017].

Taplin, I.M., 1999. Network structure and knowledge transfer in cluster evolution The transformation of the Napa Valley wine region. *International Journal of Wine Marketing*, 19(3), pp.127–145. Available at: http://dx.doi.org/10.1108/19348831111135074 [Accessed March 28, 2017].

Team, R.C., 2017. R: A language and environment for statistical computing.

Teetor, P., 2011. *R Cookbook: Proven recipes for data analysis, statistics, and graphics*, O'Reilly Media.

Tichy, N. & Fombrun, C., 1979. Network Analysis in Organizational Settings. *Human Relations*, 32(11), pp.923–965. Available at: http://journals.sagepub.com/doi/10.1177/001872677903201103 [Accessed February 23, 2017].

Tor Guthey, G., 2008. Agro-industrial conventions: some evidence from northern California's wine industry. *The Geographical Journal*, 174225121(2), pp.138–148. Available at: http://www.jstor.org/stable/40205209 [Accessed March 28, 2017].

Torre, A. & Gilly, J.-P., 1999. On the analytical dimension of proximity dynamics. *Regional Studies*, 34(2), pp.169–180. Available at: http://andre-torre.com/pdf/PDFpub31N1.pdf [Accessed April 13, 2017].

Turner, S., 2010. Networks of learning within the English wine industry. *Journal of Economic Geography*, 10, pp.685–715. Available at: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/joeg/10/5/10.1093/jeg/lbp053/2/lbp053.pdf?Expires=1490889744&Signature=Une93VEaGvjRAYjouL3ILvhYBz9mEDu8EUY57uoLsyr9BjC8NC3UFzHHA7Z2WLAyIpat~eAYm1hoeYRKjzFLLTQlE2Ds9M1p62xT-LigUXWWAYK4~orF [Accessed March 26, 2017].

USDA, 2017. USDA ERS - Common Agricultural Policy. Available at: https://www.ers.usda.gov/topics/international-markets-trade/countries-regions/european-union/common-agricultural-policy.aspx [Accessed June 28, 2017].

VDP, 2017a. Der Traubenadler schaut ... das Video: The VDP.Classification. Available at: http://www.vdp.de/en/endorsements/details/artikel/der-traubenadler-schaut-das-video-the-vdpclassification-8404/ [Accessed July 7, 2017].

VDP, 2017b. Der VDP - VDP. Available at: http://www.vdp.de/de/vdp/ [Accessed July 3, 2017].

VDP, 2017c. Zahlen, Daten, Fakten - VDP. Available at: http://www.vdp.de/de/vdp/zahlen-daten-fakten/ [Accessed July 7, 2017].

Wai, H. & Yeung, -Chung, 2005. Rethinking Relational Economic Geography. *Transactions of the Institute of British Geographers*, 30(1), pp.37–51. Available at: http://www.jstor.org/stable/3804528 [Accessed May 17, 2017].

Wal, A.L.J. Ter, 2014. The dynamics of the inventor network in german biotechnology: Geographic proximity versus triadic closure. *Journal of Economic Geography*, 14, pp.589–620. Available at: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/joeg/14/3/10.1093/jeg/lbs063/2/lbs063.pdf?Expires=1499597260&Signature=ges~OOW-OQa7hbEC0bqWnuv6zMgec2HlfhqaGuJ4DLknZyNLPmJNDyAPRj0f~n9yHy4716Vwy8iDh9L8kkeDBj2HcYAL5Jxj07vi2g7urM03G2OMMi1L [Accessed July 8, 2017].

Watanabe, S., 1988. Job-searching: A comparative study of male employment relations in the United States and Japan. *Business Administration General*, 7, pp.2469–2469.

Watts, D.J., 1999. Networks, Dynamics and the Small-World Phenomenon. *American Journal of Sociology*, 105(2), pp.493–527. Available at: http://www.jstor.org/stable/10.1086/210318 [Accessed April 15, 2017].

Wein, 2017. Gemeinsam stark: Winzergenossenschaften — Wein.de. Available at: https://www.wein.de/de/deutsche-weine/weinkauf-pflege/gemeinsam-stark-winzergenossenschaften/ [Accessed July 7, 2017].

Wein-Plus, 2017a. Gault Millau | Wein-Plus. Available at: https://glossar.wein-plus.eu/gault-millau [Accessed July 5, 2017].

Wein-Plus, 2017b. Napoleon Bonaparte | Wein-Plus. Available at: https://glossar.wein-plus.eu/napoleon-bonaparte [Accessed June 30, 2017].

Wein-Plus, 2017c. Weine von Deutschland | Wein-Plus. *Wein-Plus*. Available at: https://weinfuehrer.wein-plus.eu/weinregion/deutschland [Accessed June 30, 2017].

Wein-Plus, 2017d. Winzergenossenschaft | Wein-Plus. *Wein-Plus*. Available at: https://glossar.wein-plus.eu/winzergenossenschaft [Accessed June 30, 2017].

Weterings, A. & Boschma, R., 2009. Does spatial proximity to customers matter for innovative performance? Evidence from the Dutch software sector. , 38(5), pp.746–755. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.669.5844&rep=rep1&type=pdf [Accessed April 15, 2017].

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis.

Woodfield, P.J., 2012. *Intergenerational entrepreneurship in family business: conceptualising ways entrepreneurial family businesses can be sustained across generations*. The University of Auckland. Available at: https://researchspace.auckland.ac.nz/bitstream/handle/2292/19446/whole.pdf?sequence=2 [Accessed May 3, 2017].

Wooldridge, J.M., 2013. *Introductory Econometrics - A modern approach* 5th ed.

Yang, N., Mccluskey, J.J. & Brady, M.P., 2012. The Value of Good Neighbors: A Spatial Analysis of the California and Washington State Wine Industries. *Land Economics*, 88(4), pp.674–684. Available at: https://muse.jhu.edu/article/484305 [Accessed March 29, 2017].

Yanjie, B., 1997. Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China. *American Sociological Review*, 62(3), pp.366–385. Available at: https://www.researchgate.net/profile/Yanjie_Bian/publication/234021344_Bringing_Strong_Ties_Back_in_Indirect_Ties_Network_Bridges_and_Job_Searches_in_China/links/57f5624308ae8da3ce55268b/Bringing-Strong-Ties-Back-in-Indirect-Ties-Network-Bridges-and-Job-S [Accessed July 13, 2017].

Yeung, H.W., 1994. Critical reviews of geographical perspectives on business organizations and the organization of production: towards a network approach. *Progress in Human Geography*, 18(4), pp.460–490.

10. Appendices

Appendix A

Codes used for the cleaning, assembling and analysis of the data

c) 2017 Babette Sophia Bresser
(c) 2017 Steffen Blaschke

# load packages
# library(lubridate)
# library(stringr)
library(dplyr)
# library(reshape2)
library(igraph)
# library(ggplot2)
# library(tm)
# library(XML)
library(ggmap)

# set locale character conversion to
Sys.setlocale(locale="C")
# Sys.setlocale(locale="de_DE")

# set working directory
# test:
setwd("/Users/Steffen/Documents/Projects/Networks/UN/data/test")
setwd("/Users/Steffen/Documents/Projects/Networks/wineries/data")

# wineries
load("Merged_5.RData")

# clean up
rm(awards_fixed, awards_fixed_list, wineries_whole_list_fixed, merged_4, Merged)
wineries <- Merged_5
rm(Merged_5)
names(wineries) <- c("name", "region", "street", "city", "zip", "website", "whiteGrapes", "redGrapes", "rank", "year")

Encoding(wineries$name) <- "bytes"
wineries$name <- gsub("\\xf4", "ue", wineries$name)
wineries$name <- gsub("\\x8a__\\x92\\x99", "c", wineries$name)

Encoding(wineries$street) <- "bytes"
wineries$street <- gsub("\\xf4", "ue", wineries$street)

Encoding(wineries$city) <- "bytes"
wineries$city <- gsub("\\xf4", "ue", wineries$city)
wineries$city <- gsub("\\x99", "ueh", wineries$city)

manual corrections
wineries$street <- gsub("Schillstrasse", "Schillerstr.", wineries$street)
wineries$street <- gsub("stuedter", "staedter", wineries$street)
wineries <- wineries[!(wineries$name == "schafer-reichart"),] # delete Schaefer-Reichart winery because it's a P.O. Box of an American company

built networks from first three digits of zip code # 113
wineries$zip.network <- NA # initialize empty zip.network column
wineries[nchar(wineries$zip) == 4,]$zip.network <- substr(wineries[nchar(wineries$zip) == 4,]$zip, 1, 2) # first two digits from four-digit zips
wineries[nchar(wineries$zip) == 5,]$zip.network <- substr(wineries[nchar(wineries$zip) == 5,]$zip, 1, 3) # first three digits from five-digit zips

Missing wineries

wineries.missing.data <- read.csv("missing_streets_excel.csv", sep = ";", stringsAsFactors = FALSE)
wineries.missing.data <- left_join(wineries[is.na(wineries$street),], wineries.missing.data[, 1:2], by = "name")
wineries.missing.data <- select(wineries.missing.data, name, region, street = street.y, city, zip, website, whiteGrapes, redGrapes, rank, year, zip.network)

wineries <- rbind(wineries[!is.na(wineries$street),], wineries.missing.data)
rm(wineries.missing.data)

Single network with three-digit zip networks

vertices <- unique (select(wineries, name, street, city, zip, zip.network))
g <- make_empty_graph() # initialize empty graph
g <- add_vertices(g, nrow(vertices), name = vertices$name, street = vertices$street, city = vertices$city, zip = vertices$zip, zip.network = vertices$zip.network) # add vertices
for(i in unique(wineries$zip.network)) {g[V(g)$zip.network == i, V(g)$zip.network == i] <- TRUE # add all possible edges for each three-digit zip network}
g <- simplify(g) # remove loops
E(g)$dist <- NA # initialize distance on edges
edgelist <- get.edgelist(g) # get edgelist

from <- data.frame(name = edgelist[, 1], stringsAsFactors = FALSE) # from
from <- left_join(from, vertices) # join edgelist with vertices
from <- paste0(from$street, ", ", from$zip, " ", from$city, ", ", "Germany", sep = "") # from query
from <- split(from, 1:58) # split data frame into list of 58 x 2475

to <- data.frame(name = edgelist[, 2], stringsAsFactors = FALSE) # to
to <- left_join(to, vertices) # join edgelist with vertices
to <- paste0(to$street, ", ", to$zip, " ", to$city, ", ", "Germany", sep = "") # to query
to <- split(to, 1:58) # split data frame into list of 58 x 2475

Bing Query

```
missing <- edgelist[which(is.na(distance)), ]
from <- V(g)[missing[, 1]]
to <- V(g)[missing[, 2]]

#building the URI + 83 missing distances:
missing.dist <- numeric()

uri <-
paste0("http://dev.virtualearth.net/REST/v1/Routes?wayPoint
=",gsub(" ", "%20", from),"&wayPoint.2=",gsub(" ", "%20",
to),"&maxSolutions=1&optimize=distance&routePathOutput
=Points&distanceUnit=km&travMode=Driving","&key=",bi
ng.api.key)

bing.api.key <- "whateverthekeyis"

for(i in uri) {missing.dist <- c(missing.dist,
tryCatch({fromJSON(i)$resourceSets$resources[[1]]$routeL
egs[[1]]$routeSubLegs[[1]]["travelDistance"]}, error =
function(e){print(e)return(data.frame(travelDistance =
NA))})
Sys.sleep(sample(1:5,1)) # sleep for 1 seconds in order not to
overload server
print(paste0("Processing ", i)}}
```

Cleaning of Data Set

```
#####extract the missing distances

>
distance$travelDistance[which(is.na(distance$travelDistance)
)] <- missing.dist$travelDistance

###test###
> distance[123870,]

> which(is.na(distance))
integer(0)

E(g)$dist <- distance[, 1]

#change 77 to 777, as importing it from Excel resulted in the
dataset mistaking some of the 777 zip network wineries for
77
View(wineries)
V(g)[V(g)$name == "weingut schwoerer, josef
rohrer"]$zip.network

V(g)[V(g)$name == "weingut schwoerer, josef
rohrer"]$zip.network <- 777

#Saale-Ustrut in Capital Letters
wineries[2627,3] <- "saale-unstrut"

#there is a stdv error = 0 somewhere (77), rerun therefore
sum of distances, mean and everything else

V(g)[V(g)$name == "weingut schwoerer, josef
rohrer"]$mean.dist    #for comparison and check
V(g)[V(g)$name == "weingut karl friedrich aust"]$mean.dist
```

Calculation of Variables

```
# mean geographical distance for each winery in its network
pb <- txtProgressBar(0, vcount(g), char = "*", style = 3) #
initialize progress bar
for(i in 1:vcount(g)) {V(g)[i]$sd.dist <-
sd(as.numeric(E(g)[unlist(incident_edges(g, V(g)[i]))]$dist))
setTxtProgressBar(pb, i) # update progress bar}

# sum of geographical distances for each subnetwork
for(i in unique(V(g)$zip.network)) {V(g)[V(g)$zip.network
== i]$sum.dist <- sum(as.numeric(E(induced_subgraph(g,
V(g)[V(g)$zip.network == i]))$dist))

# standard deviation of geographical distance for each winery
in its network

V(g)[V(g)$name == "weingut schwoerer, josef
rohrer"]$sd.dist

pb <- txtProgressBar(0, vcount(g), char = "*", style = 3)
for(i in 1:vcount(g)) {V(g)[i]$sd.dist <-
sd(as.numeric(E(g)[unlist(incident_edges(g,
V(g)[i]))]$dist))setTxtProgressBar(pb, i)}

#number of wineries
  for(i in 1:vcount(g)) {V(g)[V(g)$zip.network ==
i]$n.wineries <-vcount((induced_subgraph(g,
V(g)$zip.network == i)))}

#merge it to data frame

df <- data.frame(name = V(g)$name, zip.network =
V(g)$zip.network, city = V(g)$city,  mean.dist =
V(g)$mean.dist, sum.dist =V(g)$sum.dist, sd.dist =
V(g)$sd.dist, n.wineries = V(g)$n.wineries, stringsAsFactors
= TRUE)
df$city <-as.character(df$city)
df$name<-as.character(df$name)
library(dplyr)
wineries_analysis <- left_join(wineries,df, by = c("name",
"city"))

######Varieties###

# change factor to character for whiteGrapes
wineries.analysis$whiteGrapes <-
as.character(wineries.analysis$whiteGrapes)
wineries.analysis[is.na(wineries.analysis$whiteGrapes),
]$whiteGrapes <- "character(0)"
wineries.analysis[wineries.analysis$whiteGrapes ==
"character(0)", ]$whiteGrapes <- NA
# number of white grapes
wineries.analysis$n.whiteGrapes <- 0
for(i in 1:nrow(wineries.analysis))
{if(is.na(wineries.analysis[i, ]$whiteGrapes)) next
wineries.analysis[i, ]$n.whiteGrapes <-
length(unlist(strsplit(as.character(wineries.analysis[i,
]$whiteGrapes), "\\+")))}

# change factor to character for redGrapes
wineries.analysis$redGrapes <-
as.character(wineries.analysis$redGrapes)
```

```
wineries.analysis[is.na(wineries.analysis$redGrapes),
]$redGrapes <- "character(0)"
wineries.analysis[wineries.analysis$redGrapes ==
"character(0)", ]$redGrapes <- NA
# number of white grapes
wineries.analysis$n.redGrapes <- 0
for(i in 1:nrow(wineries.analysis))
{if(is.na(wineries.analysis[i, ]$redGrapes)) next
wineries.analysis[i, ]$n.redGrapes <-
length(unlist(strsplit(as.character(wineries.analysis[i,
]$redGrapes), "\\+")))}

wineries.analysis$n.grapes <-
wineries.analysis$n.whiteGrapes +
wineries.analysis$n.redGrapes


# number of awards per cluster in the last 9 years

wins.per.zip <-
as.data.frame(aggregate(ana_work_2$winner,by =
list(ana_work_2$zip.network.y), FUN = sum))
```

Cleaning after merging

```
#join and delete redundant columns
clean_analysis_with_wins <-
left_join(cl_wineries_analyzed,wins.per.zip,by =
"zip.network.y")
clean_analysis_with_wins<-clean_analysis_with_wins[,-1]

names(clean_analysis_with_wins)[names(clean_analysis_wit
h_wins) =="x"] <- "awards by region"

# wineries
load("wineries_clean.RData")

# read csv
awards <- read.csv("dlg_awards.csv")

# clean up wineries
wineries <- wineries_clean # copy clean data frame to
wineries
wineries$name <- trimws(tolower(wineries$name)) # lower
case, trim whitespace
wineries$name <- gsub("\\s+", " ", wineries$name) #
substitute one or more whitespaces
wineries$name <- gsub(" eg", "", wineries$name)
wineries$name <- gsub(" e\\.g\\.", "", wineries$name)
wineries$name <- gsub(" gbr", "", wineries$name)
wineries$name <- gsub(" gdbr", "", wineries$name)
wineries$name <- gsub(" gmbh.*$", "", wineries$name)

wineries$region <- trimws(tolower(wineries$region)) # lower
case, trim whitespace

# clean up awards
awards$name <- trimws(tolower(awards$name)) # lower
case, trim whitespace
awards$name <- gsub("\\s+", " ", awards$name) # substitute
one or more whitespaces
awards$name <- gsub(" eg", "", awards$name)
awards$name <- gsub(" e\\.g\\.", "", awards$name)
awards$name <- gsub(" gbr", "", awards$name)
awards$name <- gsub(" gdbr", "", awards$name)
awards$name <- gsub(" gmbh.*$", "", awards$name)

awards$region <- trimws(tolower(awards$region))
awards$region <- sub("hess.+? bergstra.+?e", "hessische
bergstrasse", awards$region)

####delete remaining duplicates

#####NAMES####

> which(duplicated(no_dup$name))
 [1]  37  38  72 124 250 269 349 350 454 584 751 790
838  888 1010 1511 1861 1894 2008 2081
> View(no_dup)
> no_dup<-no_dup[-73,]
> which(duplicated(no_dup[,1:2]))
[1]  72  123  249  268  789  887 1510 1860 2007
> View(no_dup)
> View(no_dup)
> which(duplicated(no_dup[,1&2&4]))
integer(0)
> View(no_dup)
> which(duplicated(no_dup[,1&4]))
integer(0)
> which(duplicated(no_dup$name))
 [1]  37  38  72 123 249 268 348 349 453 583 750 789
837  887 1009 1510 1860 1893 2007 2080
> which(duplicated(no_dup[,1&3]))
integer(0)
> View(no_dup)
> View(no_dup)
> save.image("~/Documents/Documents 2017/Master
thesis/R/No_dups set.RData")

> wineries_no_dups <- wineries[-
c(5,39,73,125,145,204,223,272,286,332,339,362,368,397,424
,464,596,1467,649,662,702,733,765,816,818,905,982,1028,1
193,1194,1531,1773,1819,1855,1922,2032,2050,2052,2077),
]
> View(wineries_no_dups)
> which(duplicated(wineries_no_dups[,1&3]))
integer(0)

#####Addresses####

#extracted to Excel and looked through the missing adresses
> View(wineries_no_dups_addedadress)
#check for duplicates
> dups_2 <-
wineries_no_dups_addedadress[duplicated(wineries_no_dups
_addedadress[,2]),2]
> dups_2
 [1] "Weingut Fischer"         "Weingut Hildegardishof"
"Weingut Jung"
 [4] "Weingut Jung"            "Weingut Michel"           "Weingut
Schmitt"
 [7] "Weingut Sonnenberg"      "Weingut Neumer"
"Weingut Michael Schafer"
[10] "Weingut Salwey"          "Weingut Antony"
"Weingut Bauer"
[13] "Weingut Escher"
> wineries_no_dups_2 <- wineries_no_dups_addedadress[-
c(1526,1301,1437,1738), ]
```

```
#after deleting duplicates = 1068 observations
#Finding the values where ZIP=NA
> which(is.na(wineries_no_dups_2[,6]))
 [1]  116  314  436  695  858  962  975 1062 1066 1068 1084
1196 1199 1200 1255 1267 1281 1291 1332
[20] 1338 1428 1479 1492 1497 1524 1538 1539 1551 1565
1598 1612 1620 1628 1639 1735 1944 2034
> View(wineries_no_dups_2)
> View(wineries_no_dups_addedadress)
#export and import the dataset to change one adress I
overlooked, new name of dataset =wineries_no_dups_3
> which(is.na(Wineries_no_dups_3[,5]))
 [1] 314  436  695  858  962  975 1062 1066 1068 1084 1196
1199 1200 1255 1267 1281 1291 1332 1338
[20] 1428 1479 1492 1497 1524 1538 1539 1551 1565 1598
1612 1620 1628 1639 1735 1944 2034
> wineries_no_zipNA<-Wineries_no_dups_3[-
which(is.na(Wineries_no_dups_3[,5]))]
>Wineries_noZipNA<-Wineries_no_dups_3[-
c(314,436,695,858,962,975,1062,1066,1068,1084,1196,1199,
1200,1255,1267,1281,1291,1332,1338,1428,1479,1492,1497,
1524,1538,1539,1551,1565,1598,1612,1620,1628,1639,1735,
1944,2034),]
> Wineries_cleaned<-Wineries_noZip_cleaned[-
c(1129,1130,1304),]
> View(Wineries_cleaned) Wineries_cleaned<-
Wineries_noZip_cleaned[-
c(589,1404,1403,1234,1319,1423,1464,1517,1786,299,371,6
71,768,840,1007,1215,1150,1161,1736,1421,1398,1239,1801
,1747,39,990,677,1592,795,789,1596,444,1520,212,69,600,5
93,1399,413,1123,1435,82,978,1164,1991,227,1939,692,202
7,1786,1632,1126,1627,1569,1594,1742,1014,1548,1526,151
7,1452,1376,449,1793,1310,1134,1092,150,139,1394,1989,1
957,237,229,667,668,574,1535,1964,1119,1714,1030),]
> View(Wineries_cleaned)
```

Cluster extraction

```
#added leading 0 using sprintf, as some zip codes start with 0
wineries_clean[,6]<-sprintf("%05d",wineries_clean$zip)
#separated the first 3 digits to identify the leitregion
> wineries_V6<-separate(wineries_V4,zip,into =
c("Leitbereiche","last_digits"),sep = 3)
> View(wineries_V6)
> count(wineries_V6,"Leitbereiche")
```

Final cleaning in Excel

```
# 2008 there is no 99
# 2009 8 had to be deleted because the rank was not visible
# 2012, two double entries were fixed
# turn ranks into scale from 0 to 100
```

Descriptive Statistics and Plots

```
summary(wineries_analysis)
stargazer(as.data.frame(wineries_analysis_complete_V5,type
),type = "text")

stargazer(data.frame(wineries_analysis),title = "Descriptive
Statistics" , digits = 2,summary.stat =
c("n","mean","sd","median","min","max") ,  type = "html",
covariate.labels = c("Score", "Mean of Distances","Sum of
Distances","Standard Deviation", "Number of Wineries",
"Squared Number of Wineries", "Number of wh. Varieties",
"Number of red Varieties","Awards by Region"), notes.align
= "c", out = "Descr_1.htm")

################Plots##################

par(mfrow=c(1,1))

#all scatterplots
plot(wineries_ana_cor_div)
plot(wineries_analysis$n_mi_win,
wineries_analysis$awa.by.reg,pch = 20, main = "Awards per
Cluster vs. Number of Wineries per Cluster", xlab =
"Number of Wineries per Cluster", ylab = "Number of
Awards per Cluster")
plot(wineries_analysis$sum.dist,
wineries_analysis$awa.by.reg,pch = 20, main = "Sum of
Distances vs. Number of Awards per Cluster", xlab = "Sum
of Distances per Cluster", ylab = "Number of Awards per
Cluster")
plot(wineries_analysis$mean.dist,
wineries_analysis$sd.dist,pch = 20, main = "Mean Distances
vs. Standard Deviation", xlab = "Mean of Distances", ylab =
"Standard Deviation")
plot(wineries_analysis$awa.by.reg,
wineries_analysis$sd.dist,pch = 20, main = "Awards per
Cluster vs. Standard Deviation", xlab = "Awards per
Cluster", ylab = "Standard Deviation")
plot(wineries_analysis$mean.dist,
wineries_analysis$score,pch = 20, main = "Mean Distance
vs. Score", xlab = "Mean Distances", ylab = "Score")

#score

plot(score ~ mean.dist+ sum.dist +sd.dist + awa.by.reg +
sum.dist*n_mi_win, data = wineries_analysis, pch = 20)

#histograms

par(mfrow=c(2,2))
hist(wineries_analysis$mean.dist, main = "Distribution of
Mean Distances",xlab = "Mean Distance", breaks = 30)
hist(wineries_analysis$sum.dist, main = "Distribution of Sum
of Distances (in tsd km)", xlab = "Sum of km in tsd", breaks
= 30)
hist(wineries_analysis$n_mi_win, main = "Distribution of
Number of Wineries per Cluster",freq = TRUE, xlab =
"Number of Wineries per Cluster",  breaks = 25)
hist(wineries_analysis$awa.by.reg, main = "Distribution of
Number of Awards per Cluster", xlab = "Number of Awards
by Cluster",breaks = 25)


#REMOVE OUTLIERS
 write.csv(wineries_analysis, "no_outliers.csv")
 no_outliers <- read_csv("~/Documents/Documents
2017/Master
thesis/Data/Regression/20170729/no_outliers.csv")

hist(no_outliers$mean.dist, main = "Distribution of Mean
Distances",xlab = "Mean Distance", breaks = 30)
```

```
 hist(no_outliers$sum.dist, main = "Distribution of Sum of Distances (in tsd km)", xlab = "Sum of km in tsd", breaks = 30)
 hist(no_outliers$n_mi_win, main = "Distribution of Number of Wineries per Cluster",freq = TRUE, xlab = "Number of Wineries per Cluster",  breaks = 25)
 hist(no_outliers$awa.by.reg, main = "Distribution of Number of Awards per Cluster", xlab = "Number of Awards by Cluster",breaks = 25)

 plot(no_outliers$n_mi_win, no_outliers$awa.by.reg,pch = 20, main = "Awards per Cluster vs. Number of Wineries per Cluster", xlab = "Number of Wineries per Cluster", ylab = "Number of Awards per Cluster")
 plot(no_outliers$sum.dist, no_outliers$awa.by.reg,pch = 20, main = "Sum of Distances per Cluster vs. Number of Awards per Cluster", xlab = "Sum of Distances per Cluster", ylab = "Number of Awards per Cluster")
 plot(no_outliers$mean.dist, no_outliers$sd.dist,pch = 20, main = "Mean Distances vs. Standard Deviation", xlab = "Mean of Distances", ylab = "Standard Deviation")
 plot(no_outliers$awa.by.reg, no_outliers$sd.dist,pch = 20, main = "Awards per Cluster vs. Standard Deviation", xlab = "Awards per Cluster", ylab = "Standard Deviation")



 ########CORRELATION

cor_table <- cor(wineries_ana_cor_div)
stargazer(cor_table, title = "Correlation Matrix", type = "html", out = "correlation_V2.htm")
stargazer(cor_table, title = "Correlation Matrix", type = "html",covariate.labels = c("Score", "Mean of Distances", "Sum of Distances", "Standard Deviation of Distances", "Number of Wineries","Squared Number of Wineries", "Number of White Varieties", "Number of Red Varieties", "Awards by Cluster"),out = "Correlation_V1.htm")

#get the histogram without counting the multiple winners double

uni_n_by_id <- subset(N_by_id, !duplicated(wineries_ana_stat.id))

#Number of wineries histo

hist(uni_n_by_id$wineries_ana_stat.n_mi_win, main = "Number of Wineries by Cluster",freq = TRUE, xlab = "Number of Wineries per Cluster",  breaks = 25)
length(uni_n_by_id$wineries_ana_stat.n_mi_win)
summary(uni_n_by_id)

#award by cluster histogram

Awa_by_id <- data.frame(wineries_ana_stat$id,wineries_ana_stat$awa.by.reg )
uni_awa_by_id <- subset(Awa_by_id, !duplicated(wineries_ana_stat.id))
hist(uni_awa_by_id$wineries_ana_stat.awa.by.reg, main = "Number of Awards by Cluster", xlab = "Number of Awards by Cluster",breaks = 25)

#Overall four histograms for appendix
par(mfrow=c(2,2))
hist(unique(wineries_analysis$mean.dist), main = "Mean Distance",xlab = "Mean Distance", breaks = 30)
hist(unique(wineries_analysis$sum.dist), main = "Sum of Distances (in tsd km)", xlab = "Sum of km in tsd", breaks = 30)
hist(uni_n_by_id$wineries_ana_stat.n_mi_win, main = "Number of Wineries by Cluster",freq = TRUE, xlab = "Number of Wineries per Cluster",  breaks = 25)
hist(uni_awa_by_id$wineries_ana_stat.awa.by.reg, main = "Number of Awards by Cluster", xlab = "Number of Awards by Cluster",breaks = 25)
##################################
```

Regression

```
# regressions, wald, stargazer with the right outputs
library(lmtest)
library(multiwayvcov)
library(stargazer)
library(sandwich)
require(sandwich)
require(lmtest)
require(multiwayvcov)
require(stargazer)
```

Added mean of distances

```
reg_1 <-lm(score ~ mean.dist, data = wineries_analysis)
summary(reg_1)

vcov_reg_1 <- cluster.vcov(reg_1, wineries_analysis$id)
new_1 <- coeftest(reg_1, vcov_reg_1)

cl.robust.se.1 <- sqrt(diag(vcov_reg_1))
cl.wald1 <- waldtest(reg_1, vcov = vcov_reg_1)
cl.wald1

stargazer(reg_1,reg_1,title = "Results", align = TRUE, type = "html",column.labels=c("default","robust"), no.space = TRUE, se = list(NULL,cl.robust.se.1))
```

Added sum of distances

```
reg_2 <-lm(score ~ mean.dist + sum.dist, data = wineries_analysis)
summary(reg_2)
plot(reg_2)

vcov_reg_2 <- cluster.vcov(reg_2, wineries_analysis$id)
new_2 <-coeftest(reg_2, vcov_reg_2)
new_2

cl.robust.se.2 <- sqrt(diag(vcov_reg_2))
cl.wald2 <- waldtest(reg_2, vcov = vcov_reg_2)
cl.wald2

stargazer(new_1, new_2,title = "Results", align = TRUE, type = "html",report = "vc*t", no.space = TRUE, se = list(cl.robust.se.1,cl.robust.se.2))

##test if there is an improvement in terms of using robust se
```

```
stargazer(reg_2,new_2,title = "Results", align = TRUE, type
= "html",column.labels=c("default","robust"),report = "vc*t"
,no.space = TRUE, se = list(NULL,cl.robust.se.2))
```

Added standard deviation of distances

```
reg_3 <-lm(score ~ mean.dist + sum.dist + sd.dist, data =
wineries_analysis)
summary(reg_3)

vcov_reg_3 <- cluster.vcov(reg_3, wineries_analysis$id)
new_3 <- coeftest(reg_3, vcov_reg_3)
cl.robust.se.3 <- sqrt(diag(vcov_reg_3))



stargazer(reg_3,new_3,title = "Results", align = TRUE, type
= "html",column.labels=c("default","robust"), no.space =
TRUE, se = list(NULL,cl.robust.se.3))


cl.robust.se.3 <- sqrt(diag(vcov_reg_3))
cl.wald3 <- waldtest(reg_3, vcov = vcov_reg_3)
cl.wald3

stargazer(new_1,new_2,new_3,title = "Results",report =
"vc*t" ,align = TRUE, type =
"html",column.labels=c("m1","m2","m3"), no.space =
TRUE, se = list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3))

waldtest(new_2, new_3)
```

Added number of wineries (m_4) and square root (reg_5)

```
#######
m_4 <-lm(score ~ mean.dist + sum.dist + sd.dist + n_mi_win,
data = wineries_ana_stat)
#######

reg_5 <-lm(score ~ mean.dist + sum.dist + sd.dist +
n_mi_win + n_mi_win2, data = wineries_analysis)
summary(reg_5)
plot(reg_5)


vcov_reg_5 <- cluster.vcov(reg_5, wineries_analysis$id)
new_5<-coeftest(reg_5, vcov_reg_5)
cl.robust.se.5 <- sqrt(diag(vcov_reg_5))

cl.wald5 <- waldtest(reg_5, vcov = vcov_reg_5)
cl.wald5


stargazer(new_1,new_2,new_3,new_5,title = "Results",report
= "vc*t" ,align = TRUE, type =
"html",column.labels=c("m1","m2","m3","m4"), no.space =
TRUE, se = list(cl.robust.se.1,cl.robust.se.2,
cl.robust.se.3,cl.robust.se.5))

stargazer(reg_5,reg_5,title = "Results", align = TRUE, type =
"html",column.labels=c("default","robust"), no.space =
TRUE, se = list(NULL,cl.robust.se.5))
```

Added white varieties

```
reg_6 <-lm(score ~ mean.dist + sum.dist + sd.dist +
n_mi_win + n_mi_win2 + n.whiteGrapes, data =
wineries_analysis)

summary(reg_6)
plot(reg_6)


vcov_reg_6 <- cluster.vcov(reg_6, wineries_analysis$id)
new_6<-coeftest(reg_6, vcov_reg_6)
cl.robust.se.6 <- sqrt(diag(vcov_reg_6))

cl.robust.se.6 <- sqrt(diag(vcov_reg_6))
cl.wald6 <- waldtest(reg_5, vcov = vcov_reg_6)
cl.wald6

waldtest(reg_5, reg_6)

stargazer(reg_1,reg_2,reg_3,reg_5,reg_6,title = "Results",
align = TRUE, type =
"html",column.labels=c("m1","m2","m3","m4","m5"),
no.space = TRUE, se = list(cl.robust.se.1,cl.robust.se.2,
cl.robust.se.3,cl.robust.se.5, cl.robust.se.6))

stargazer(reg_5,reg_5,title = "Results", align = TRUE, type =
"html",column.labels=c("default","robust"), no.space =
TRUE, se = list(NULL,cl.robust.se.5))
```

Added red varieties

```
reg_7 <-lm(score ~ mean.dist + sum.dist + sd.dist +
n_mi_win + n_mi_win2 + n.whiteGrapes + n.redGrapes, data
= wineries_analysis)

summary(reg_7)
plot(reg_7)

vcov_reg_7 <- cluster.vcov(reg_7, wineries_analysis$id)
new_7 <-coeftest(reg_7, vcov_reg_7)
cl.robust.se.7 <- sqrt(diag(vcov_reg_7))

waldtest(reg_7, new_7, test = "F")
waldtest(reg_7, vcov_reg_7)

cl.robust.se.7 <- sqrt(diag(vcov_reg_7))
cl.wald7 <- waldtest(reg_7, vcov = vcov_reg_7)
cl.wald7

waldtest(new_6, new_7)

stargazer(reg_1,reg_2,reg_3,reg_5,reg_6,reg_7,title =
"Results", align = TRUE, type =
"html",column.labels=c("m1","m2","m3","m4","m5","m6"),
no.space = TRUE, se = list(cl.robust.se.1,cl.robust.se.2,
cl.robust.se.3,cl.robust.se.5, cl.robust.se.6,cl.robust.se.7))

stargazer(reg_7,reg_7,title = "Results", align = TRUE, type =
"html",column.labels=c("default","robust"), no.space =
TRUE, se = list(NULL,cl.robust.se.7))
```

Added awards by cluster

```
reg_8 <-lm(score ~ mean.dist + sum.dist + sd.dist +
n_mi_win + n_mi_win2 + n.whiteGrapes + n.redGrapes+
awa.by.reg, data = wineries_analysis)
summary(reg_8)
plot(reg_8)

vcov_reg_8 <- cluster.vcov(reg_8, wineries_analysis$id)
new_8<-coeftest(reg_8, vcov_reg_8)
cl.robust.se.8 <- sqrt(diag(vcov_reg_8))

cl.robust.se.8 <- sqrt(diag(vcov_reg_8))
waldtest(reg_8, vcov_reg_8)
stargazer(reg_1,reg_2,reg_3,reg_5,reg_6,reg_7,reg_8,title =
"Results", align = TRUE, type =
"html",column.labels=c("m1","m2","m3","m4","m5","m6","
m7"), no.space = TRUE, se = list(cl.robust.se.1,cl.robust.se.2,
cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8))

stargazer(reg_7,reg_7,title = "Results", align = TRUE, type =
"html",column.labels=c("default","robust"), no.space =
TRUE, se = list(NULL,cl.robust.se.7))
```

Added regions

```
reg_9 <-lm(score ~ mean.dist + sum.dist + sd.dist +
n_mi_win + n_mi_win2 + n.whiteGrapes + n.redGrapes+
awa.by.reg + factor(region), data = wineries_analysis)

cl.wald9 <- waldtest(reg_9, new_9)
cl.wald9
summary(reg_9)
plot(reg_9)
waldtest(new_8,new_9)

vcov_reg_9 <- cluster.vcov(reg_9, wineries_analysis$id)
new_9 <-coeftest(reg_9, vcov_reg_9)
cl.robust.se.9 <- sqrt(diag(vcov_reg_9))

cl.robust.se.9 <- sqrt(diag(vcov_reg_9))
cl.wald9 <- waldtest(new_9,new_10)
cl.wald9
cl.wald9 <- waldtest(reg_9, vcov = vcov_reg_9)

stargazer(reg_1,reg_2,reg_3,reg_5,reg_6,reg_7,reg_8,reg_9,ti
tle = "Results", align = TRUE, type =
"html",column.labels=c("m1","m2","m3","m4","m5","m6","
m7","m9"), no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9))

stargazer(reg_9,reg_9,title = "Results", align = TRUE, type =
"html",column.labels=c("default","robust"), no.space =
TRUE, se = list(NULL,cl.robust.se.9))
```

Interaction term sum of distances * number of wineries

```
reg_10 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win + n_mi_win2 + sum.dist*n_mi_win+
n.whiteGrapes + n.redGrapes+ awa.by.reg + factor(region),
data = wineries_analysis)
plot(reg_10)
summary(reg_10)

vcov_reg_10 <- cluster.vcov(reg_10, wineries_analysis$id)
new_10 <- coeftest(reg_10, vcov_reg_10)
new_10
plot(new_10)
cl.robust.se.10 <- sqrt(diag(vcov_reg_10))
```

Compare normal standard errors vs. robust standard errors

```
stargazer(reg_10,new_10,title = "Results", align = TRUE,
type =
"html",column.labels=c("default","robust"),covariate.labels =
c("Mean of Distances", "Sum of Distances", "Standard
Deviation of Distances", "Number of Wineries","Squared
Number of Wineries", "Number of White Varieties",
"Number of Red Varieties", "Awards by Region", "Ahr and
Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Sum of Distanes x Number of Wineries"), no.space =
TRUE, se = list(NULL,cl.robust.se.10), out =
"reg_10_comp.htm")
```

Compare all models

```
stargazer(new_1,new_2,new_3,new_5,new_6,new_7,new_8,
new_9,new_10,title = "Regression Results",covariate.labels =
c("Mean of Distances", "Sum of Distances", "Standard
Deviation of Distances", "Number of Wineries","Squared
Number of Wineries", "Number of White Varieties",
"Number of Red Varieties", "Awards by Region", "Ahr and
Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Sum of Distanes x Number of Wineries"),digits =
2,dep.var.labels = "Score",dep.var.caption = "",report =
"vc*t", align = TRUE, type = "html", no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9,
cl.robust.se.10),out = "ref_final_8.htm" )


waldtest(reg_10, vcov = firm_c_vcov, test = "F")

cl.wald10 <- waldtest(new_10, vcov = vcov_reg_10)
cl.wald10

coeftest_10 <- coeftest(reg_10, vcov_reg_10)
```

Exclude region

```
reg_12 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win + n_mi_win2 + sum.dist*n_mi_win+
n.whiteGrapes + n.redGrapes+ awa.by.reg,data =
wineries_analysis)
summary(reg_12)

vcov_reg_12 <- cluster.vcov(reg_12, wineries_analysis$id)
new_12 <- coeftest(reg_12, vcov_reg_12)
new_12
```

```
cl.robust.se.12 <- sqrt(diag(vcov_reg_12))
cl.wald12 <- waldtest(reg_12, vcov = vcov_reg_12)

stargazer(new_1,new_2,new_3,new_5,new_6,new_7,new_8,
new_9,new_10,reg_12, title = "Regression
Results",covariate.labels = c("Mean of Distances", "Sum of
Distances", "Standard Deviation of Distances", "Number of
Wineries","Squared Number of Wineries", "Number of
White Varieties", "Number of Red Varieties", "Awards by
Region", "Ahr and Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Sum of Distanes x Number of Wineries"),digits =
2,dep.var.labels = "Score",dep.var.caption = "",report =
"vc*t", align = TRUE, type = "html", no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9,
cl.robust.se.10, cl.robust.se.12),out = "ref_final_8.htm" )
```

Interaction: term mean of distances * standard deviation

```
reg_13 <-lm(score ~ mean.dist + sd.dist +sum.dist +
mean.dist*sd.dist+  n_mi_win + n_mi_win2 +
sum.dist*n_mi_win+  n.whiteGrapes + n.redGrapes+
awa.by.reg,data =  wineries_analysis)
summary(reg_13)

vcov_reg_13 <- cluster.vcov(reg_13, wineries_analysis$id)
new_13 <- coeftest(reg_13, vcov_reg_13)
new_13
cl.robust.se.13 <- sqrt(diag(vcov_reg_13))
cl.wald13 <- waldtest(reg_13, vcov = vcov_reg_13)

stargazer(new_1,new_2,new_3,new_5,new_6,new_7,new_8,
new_9,new_10,new_12,new_13, title = "Regression
Results",covariate.labels = c("Mean of Distances", "Sum of
Distances", "Standard Deviation of Distances", "Number of
Wineries","Squared Number of Wineries", "Number of
White Varieties", "Number of Red Varieties", "Awards by
Region", "Ahr and Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Mean od Distances x St. Deviation of Distances","Sum of
Distanes x Number of Wineries"),digits = 2,dep.var.labels =
"Score",dep.var.caption = "",report = "vc*t", align = TRUE,
type = "html", no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9,
cl.robust.se.10, cl.robust.se.12, cl.robust.se.13),out =
"ref_final_9.htm" )
```

Add regions --> not significant

```
reg_14 <-lm(score ~ mean.dist + sd.dist +sum.dist +
mean.dist*sd.dist+  n_mi_win + n_mi_win2 +
n.whiteGrapes + n.redGrapes+ awa.by.reg +
factor(region),data =  wineries_analysis)
summary(reg_14)

vcov_reg_14 <- cluster.vcov(reg_14, wineries_analysis$id)
new_14 <- coeftest(reg_14, vcov_reg_14)
new_14
cl.robust.se.14 <- sqrt(diag(vcov_reg_14))
cl.wald14 <- waldtest(reg_14, vcov = vcov_reg_14)

stargazer(new_1,new_2,new_3,new_5,new_6,new_7,new_8,
new_9,new_10,new_12,new_13, title = "Regression
Results",covariate.labels = c("Mean of Distances", "Sum of
Distances", "Standard Deviation of Distances", "Number of
Wineries","Squared Number of Wineries", "Number of
White Varieties", "Number of Red Varieties", "Awards by
Region", "Ahr and Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Mean od Distances x St. Deviation of Distances","Sum of
Distanes x Number of Wineries"),digits = 2,dep.var.labels =
"Score",dep.var.caption = "",report = "vc*t", align = TRUE,
type = "html", no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9,
cl.robust.se.10, cl.robust.se.12, cl.robust.se.13),out =
"ref_final_9.htm" )
```

Interaction: number of wineries and st.dv (without region)

```
reg_15 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win*sd.dist+  n_mi_win + n_mi_win2 +
n.whiteGrapes + n.redGrapes+ awa.by.reg,data =
wineries_analysis)
summary(reg_15)

vcov_reg_15 <- cluster.vcov(reg_15, wineries_analysis$id)
new_15 <- coeftest(reg_15, vcov_reg_15)
new_15
cl.robust.se.15 <- sqrt(diag(vcov_reg_15))
cl.wald15 <- waldtest(reg_15, vcov = vcov_reg_15)

stargazer(new_1,new_2,new_3,new_5,new_6,new_7,new_8,
new_9,new_10,new_12,new_13,new_15 title = "Regression
Results",covariate.labels = c("Mean of Distances", "Sum of
Distances", "Standard Deviation of Distances", "Number of
Wineries","Squared Number of Wineries", "Number of
White Varieties", "Number of Red Varieties", "Awards by
Region", "Ahr and Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Mean od Distances x St. Deviation of Distances","Sum of
Distanes x Number of Wineries"),digits = 2,dep.var.labels =
"Score",dep.var.caption = "",report = "vc*t", align = TRUE,
type = "html", no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9,
cl.robust.se.10, cl.robust.se.12, cl.robust.se.13),out =
"ref_final_9.htm" )
```

Adding regions

```
reg_16 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win*sd.dist+  n_mi_win + n_mi_win2 +
sum.dist*n_mi_win+  n.whiteGrapes + n.redGrapes+
awa.by.reg + factor(region),data =  wineries_analysis)
```

```
vcov_reg_16 <- cluster.vcov(reg_16, wineries_analysis$id)
new_16 <- coeftest(reg_16, vcov_reg_16)
new_16
cl.robust.se.16 <- sqrt(diag(vcov_reg_16))
cl.wald16 <- waldtest(reg_16, vcov = vcov_reg_16)
stargazer(new_1,new_2,new_3,new_5,new_6,new_7,new_8,
new_9,new_10,new_12,new_13,new_15,new_16 title =
"Regression Results",covariate.labels = c("Mean of
Distances", "Sum of Distances", "Standard Deviation of
Distances", "Number of Wineries","Squared Number of
Wineries", "Number of White Varieties", "Number of Red
Varieties", "Awards by Region", "Ahr and
Mittelrhein","Baden", "Baden and Hess.
Bergstrasse","Franken","Hess. Bergstrasse", "Hess.
Bergstrasse and Rheingau", "Mittelrhein","Mittelrhein and
Rheingau", "Mosel","Nahe", "Pfalz", "Rheingau",
"Rheinhessen", "Saale-Unstrut", "Sachsen", "Wuerttemberg",
"Mean od Distances x St. Deviation of Distances","Sum of
Distanes x Number of Wineries"),digits = 2,dep.var.labels =
"Score",dep.var.caption = "",report = "vc*t", align = TRUE,
type = "html", no.space = TRUE, se =
list(cl.robust.se.1,cl.robust.se.2, cl.robust.se.3,cl.robust.se.5,
cl.robust.se.6,cl.robust.se.7,cl.robust.se.8,cl.robust.se.9,
cl.robust.se.10, cl.robust.se.12, cl.robust.se.13),out =
"ref_final_9.htm" )
```

Interaction: n_mi_win*awa by region

```
reg_17 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win*sd.dist+  n_mi_win + n_mi_win2 +
sum.dist*n_mi_win+ n_mi_win*awa.by.reg +
n.whiteGrapes + n.redGrapes+ awa.by.reg +
factor(region),data =  wineries_analysis)
summary(reg_17)

vcov_reg_17 <- cluster.vcov(reg_17, wineries_analysis$id)
new_17 <- coeftest(reg_17, vcov_reg_17)
new_17
cl.robust.se.17 <- sqrt(diag(vcov_reg_17))
cl.wald17 <- waldtest(reg_17, vcov = vcov_reg_17)
cl.wald17
```

Interaction: sd.dist * sum.dist

```
reg_19 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win + n_mi_win2  +  n.whiteGrapes + n.redGrapes+
awa.by.reg + factor(region)+ sd.dist*sum.dist ,data =
wineries_analysis)
summary(reg_19)

vcov_reg_19 <- cluster.vcov(reg_19, wineries_analysis$id)
new_19 <- coeftest(reg_19, vcov_reg_19)
```

Interaction: mean dist * number

```
reg_22 <-lm(score ~ mean.dist + sd.dist +sum.dist +
n_mi_win + n_mi_win2  +  n.whiteGrapes + n.redGrapes+
awa.by.reg + factor(region)+ mean.dist*n_mi_win ,data =
wineries_analysis)
summary(reg_22)

vcov_reg_22 <- cluster.vcov(reg_22, wineries_analysis$id)
new_22 <- coeftest(reg_22, vcov_reg_22)
```

Map of Germany

```
# having extracted the zip codes and found the corresponding
longitude and latitude in excel, re-import  the zip codes +
longitude + latitude

> zips_map_2 <- read_csv("~/Documents/Documents
2017/Master thesis/Data/zips_map_2.csv",
                    +     col_types = cols(zip = col_number()))

#joining
> wineries_map <-
left_join(analysis_060617_withqu,zips_map_2)
Joining, by = "zip"
> View(wineries_map)
> save.image("~/Documents/Documents 2017/Master
thesis/Data/0806_dataset for mapping.RData")

> names(merged_zip)[names(merged_zip)=="Lat"]<-"y"
> names(merged_zip)[names(merged_zip)=="longitude"]<-
"x"

> is.numeric(wineries_map$zip)
[1] TRUE
> merged_zip$zip<-as.numeric(merged_zip$zip)
> is.numeric(merged_zip$zip)
[1] TRUE

> save.image("~/0806_dataset with zips.RData") #just to be
sure

zips_map_3 <- read_csv("~/Documents/Documents
2017/Master thesis/Data/zips_map_3.csv",
                  +     col_types = cols(longitude =
col_number()))
> View(zips_map_3)
> is.numeric(zips_map_3$longitude)
[1] TRUE
> zips_map_3<-zips_map_3[,-1]
> left_join(zips_map_2,zips_map_3)

Joining, by = "zip"
# A tibble: 418 <U+00D7> 3
Lat   zip longitude
<dbl> <chr>     <dbl>
  1  50.51840 53508   7.01857
2  49.07394 74232   9.30241
3  49.74729 97318  10.13598
4  50.04079 65347   8.04118
5  49.94400 56841   7.11890
6  49.70047 67593   8.24427
7  48.69414 77815   8.14726
8  49.86564 97332  10.23802
9  49.27120 76835   8.05209
10 48.48648 77770   8.02852

# ... with 408 more rows
> merged_zip <- left_join(zips_map_2,zips_map_3)

#get the map

Germany <- get_map(location = "Germany",zoom = 6)
```

Map from URL :
http://maps.googleapis.com/maps/api/staticmap?center=Germany&zoom=6&size=640x640&scale=2&maptype=terrain&language=en-EN&sensor=false
Information from URL :
http://maps.googleapis.com/maps/api/geocode/json?address=Germany&sensor=false

```
> plot(Germany)

map <- ggmap(Germany) + geom_point(aes(x=x,y=y), data =
wineries_map,colour="red")
> plot(map)
> save.image("~/Documents/Documents 2017/Master
thesis/Data/080617_dataset ITWORKED.RData")

#WITH COLOURS

map_zip <- ggmap(Germany) +
  geom_point(aes(x=x,y=y,colour=as.factor(zip.network.x)),
data = wineries_map)+
  ggtitle("German Wineries by Zip Area")
> plot(map_zip)
```

Table of Wineries per Region

```
#table count per region
library(dplyr)


df_1<- wineries_analysis %>%
  group_by(region) %>%
  summarise(count = n_distinct(mean.dist))

df_1 <- as.data.frame(df_1)

library(knitr)
kable(df_1, format = "rst", caption = "Wineries by Region" )
```

Example of Cluster Graph

```
#Sum of Distances
plot(Baden$Long, Baden$Lat, pch=19, col="red", cex=0.5,
main = "Illustration of Spatial Ties in a Cluster", ylab =
"Latitude Coordinates", xlab = "Longitude Coordinates")
lines(Baden$Long, Baden$Lat, pch=19, col="red", cex=0.5)
apply(combn(seq_len(nrow(Baden)), 2), 2,
      function(x) lines(Baden[x, ]$Long, Baden[x, ]$Lat))

#Mean Distances
plot(Baden$Long, Baden$Lat, pch=19, col="red", cex=0.5,
main = "Illustration of Outgoing Ties in a Cluster", ylab =
"Latitude Coordinates", xlab = "Longitude Coordinates")
lines(arrows(7.547420,47.71578,x1 =
7.671100,47.87383,length = 0.2, angle = 20))
lines(arrows(7.547420,47.71578,x1 = 7.666133,
47.87156,length = 0.2, angle = 20))
lines(arrows(7.547420,47.71578,x1 = 7.671100,
47.87195,length = 0.2, angle = 20))
lines(arrows(7.547420,47.71578,x1 = 7.661123,
47.87178,length = 0.2, angle = 20))
lines(arrows(7.547420,47.71578,x1 = 7.661123,
47.87178,length = 0.2, angle = 20))
lines(arrows(7.547420,47.71578,x1 = 7.583305 ,
47.75183,length = 0.2, angle = 20))
lines(arrows(7.547420,47.71578,x1 = 7.595120 ,
47.79529,length = 0.2, angle = 20))
```

Citations

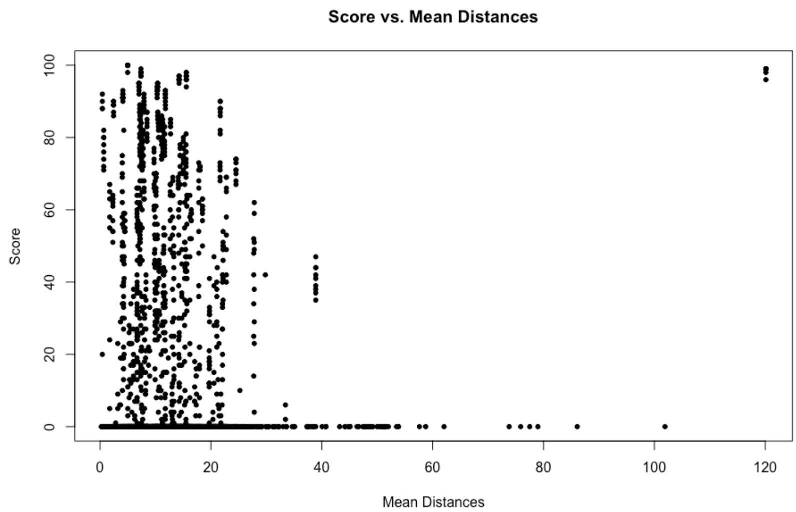
```
> citation()

> citation(package = "igraph")

> citation(package = "ggplot2")
```

Appendix B

Illustration of outliers which were removed



Appendix C

Regression results without values in Sum of Distances which could be considered outliers

Regression Results

	Score
Mean of Distances	-0.49***
	t = -2.61
Standard Deviation of Distances	0.34***
	t = 3.55
Sum of Distances	0.39***
	t = 2.91
Number of Wineries	-0.72***
	t = -3.36
Squared Number of Wineries	0.001
	t = 0.21
Number of White Varieties	-0.11
	t = -0.14
Number of Red Varieties	0.77
	t = 0.76
Awards by Region	0.67***
	t = 5.60
Sum of Distanes x Number of Wineries	-0.001
	t = -0.94
Constant	21.94***
	t = 4.95
Note:	*p**p***p<0.01

Appendix D

Regression with interaction terms

Regression Results with Interaction Terms

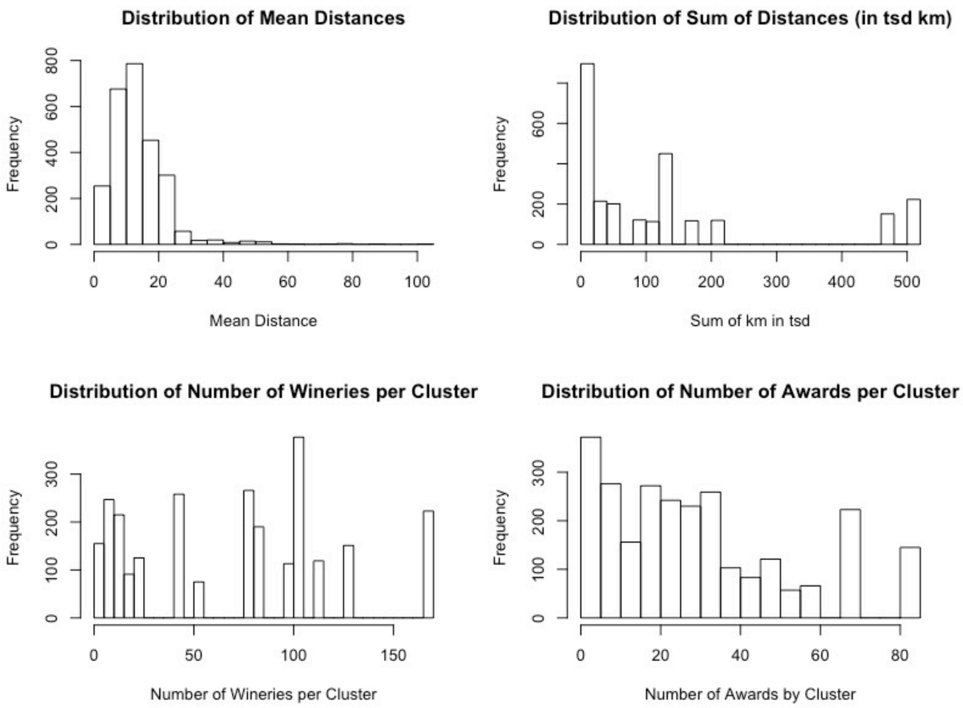
	Score				
	(1)	(2)	(3)	(4)	(5)
Mean of Distances	-0.44^{***}	-0.30^{**}	-0.39^{***}	-0.42^{***}	-0.43
	t = -3.03	t = -2.05	t = -2.61	t = -2.64	t = -1.56
Standard Deviation of Distances	0.29^{***}	0.55^{***}	0.35^{**}	0.21	0.21
	t = 3.03	t = 2.90	t = 2.24	t = 0.77	t = 0.77
Sum of Distances	0.0001^{**}	0.0001^{**}	-0.0001^{**}	-0.0001	-0.0001
	t = 2.09	t = 2.01	t = -2.18	t = -1.44	t = -1.48
Number of Wineries	-0.49^{**}	-0.51^{**}	-0.39^{**}	-0.40^{**}	-0.40^{**}
	t = -2.06	t = -2.12	t = -2.45	t = -2.52	t = -2.52
Squared Number of Wineries	0.002	0.002	0.004^{***}	0.004^{***}	0.004^{***}
	t = 0.94	t = 1.00	t = 3.53	t = 3.46	t = 3.44
Number of White Varieties	-0.11	-0.10	-0.23	-0.24	-0.24
	t = -0.13	t = -0.12	t = -0.27	t = -0.28	t = -0.28
Number of red Varieties	1.25	1.22	1.40	1.40	1.40
	t = 1.20	t = 1.18	t = 1.42	t = 1.42	t = 1.42
Awards by Cluster	0.60^{***}	0.61^{***}	1.23^{***}	1.24^{***}	1.24^{***}
	t = 4.20	t = 4.42	t = 19.04	t = 18.26	t = 17.90
Ahr and Mittelrhein	-17.12	-18.07	-15.61	-15.33	-15.30
	t = -1.44	t = -1.51	t = -1.43	t = -1.41	t = -1.39
Baden	-1.01	-2.60	-11.15	-10.98	-10.96
	t = -0.08	t = -0.21	t = -0.98	t = -0.96	t = -0.96
Baden and Hess. Bergstrasse	-9.74	-10.15	-14.50	-14.49	-14.48
	t = -0.76	t = -0.79	t = -1.25	t = -1.26	t = -1.25
Franken	-8.00	-9.54	-4.52	-4.17	-4.16
	t = -0.64	t = -0.76	t = -0.39	t = -0.36	t = -0.36
Hess. Berstrasse	-5.74	-5.83	-8.91	-9.04	-9.03
	t = -0.45	t = -0.45	t = -0.79	t = -0.80	t = -0.80
Hess. Bergstrasse and Rheingau	-5.65	-6.32	-17.50	-16.23	-16.23
	t = -0.45	t = -0.51	t = -1.53	t = -1.41	t = -1.41
Mittelrhein	-7.38	-8.33	-7.56	-7.47	-7.46
	t = -0.55	t = -0.61	t = -0.60	t = -0.60	t = -0.59
Mittelrhein and Rheingau	-7.36	-9.09	-12.56	-12.26	-12.27
	t = -0.59	t = -0.73	t = -1.12	t = -1.09	t = -1.10
Mosel	-3.84	-5.19	-4.85	-4.63	-4.62
	t = -0.31	t = -0.42	t = -0.43	t = -0.41	t = -0.41
Nahe	-3.97	-5.12	-14.52	-14.20	-14.19
	t = -0.31	t = -0.40	t = -1.23	t = -1.20	t = -1.20
Pfalz	-15.69	-16.50	-16.51	-15.85	-15.86
	t = -1.24	t = -1.31	t = -1.48	t = -1.42	t = -1.42
Rheingau	-5.46	-6.10	-15.74	-14.64	-14.63
	t = -0.43	t = -0.49	t = -1.39	t = -1.28	t = -1.28
Rheinhessen	-4.24	-5.49	-9.27	-8.66	-8.66
	t = -0.34	t = -0.44	t = -0.82	t = -0.77	t = -0.77
Saale-Unstrut	-15.30	-18.60	-11.84	-10.64	-10.63
	t = -1.25	t = -1.49	t = -1.05	t = -0.92	t = -0.92
Sachsen	-18.62	-19.85	-15.23	-14.70	-14.70
	t = -1.47	t = -1.57	t = -1.32	t = -1.26	t = -1.27
Wuerttemberg	-3.70	-4.84	-8.59	-8.37	-8.35

	t = -0.30	t = -0.39	t = -0.77	t = -0.75	t = -0.75
Mean of Distance x Number of Wineries					0.0001
					t = 0.03
Standard Deviation of Distances x Sum of Distances				-0.0000	-0.0000
				t = -0.99	t = -1.02
Awards per Cluster x Number of Wineries			**-0.01*****	**-0.01*****	**-0.01*****
			t = -8.06	t = -8.01	t = -7.98
Standard Deviation of Distances x Number of Wineries			0.0004	0.003	0.003
			t = 0.26	t = 1.09	t = 1.02
Mean of Distances x Standard Deviation of Distances		-0.01*	-0.004	-0.001	-0.001
		t = -1.85	t = -1.22	t = -0.20	t = -0.22
Sum of Distances x Number of Wineries	-0.0000	-0.0000	0.0000	0.0000	0.0000
	t = -1.59	t = -1.58	t = 1.41	t = 1.36	t = 1.37
Constant	23.99**	23.08*	18.89*	19.40*	19.40*
	t = 2.00	t = 1.92	t = 1.72	t = 1.78	t = 1.78

Note: *p**p***p<0.01

Appendix E

Histograms of the distribution of the mean distance of wineries and the sum of distances in German wine clusters.



Appendix F

Comparison of regression with normal standard errors and robust standard errors

Results

	score	
	default	robust
	(1)	(2)
Mean of Distances	-0.438***	-0.438***
	(0.085)	(0.145)
Sum of Distances	0.289***	0.289***
	(0.057)	(0.096)
Standard Deviation of Distances	0.147***	0.147**
	(0.044)	(0.070)
Number of Wineries	-0.486***	-0.486**
	(0.125)	(0.236)
Squared Number of Wineries	0.002*	0.002
	(0.001)	(0.002)
Number of White Varieties	-0.113	-0.113
	(0.277)	(0.847)
Number of Red Varieties	1.246***	1.246
	(0.383)	(1.034)
Awards by Region	0.605***	0.605***
	(0.037)	(0.144)
Ahr and Mittelrhein	-17.116	-17.116
	(25.874)	(11.889)
Baden	-1.009	-1.009
	(4.835)	(12.462)
Baden and Hess. Bergstrasse	-9.741	-9.741
	(10.655)	(12.821)
Franken	-7.998*	-7.998
	(4.829)	(12.515)
Hess. Bergstrasse	-5.744	-5.744
	(7.224)	(12.852)
Hess. Bergstrasse and Rheingau	-5.653	-5.653
	(25.966)	(12.567)
Mittelrhein	-7.384	-7.384
	(5.996)	(13.311)
Mittelrhein and Rheingau	-7.363	-7.363
	(25.952)	(12.440)
Mosel	-3.840	-3.840
	(4.883)	(12.233)
Nahe	-3.974	-3.974
	(5.217)	(12.745)
Pfalz	-15.686***	-15.686
	(4.792)	(12.613)
Rheingau	-5.455	-5.455
	(5.363)	(12.606)
Rheinhessen	-4.241	-4.241

	(5.117)	(12.545)
Saale-Unstrut	-15.304**	-15.304
	(6.869)	(12.220)
Sachsen	-18.620**	-18.620
	(8.660)	(12.639)
Wuerttemberg	-3.695	-3.695
	(4.696)	(12.515)
Sum of Distanes x Number of Wineries	-0.001***	-0.001
	(0.0004)	(0.001)
Constant	23.989***	23.989**
	(4.587)	(11.984)
Observations	2,605	
R^2	0.230	
Adjusted R^2	0.223	
Residual Std. Error	25.492 (df = 2579)	
F Statistic	30.866*** (df = 25; 2579)	
Note:		*p**p***p<0.01