Papers in Evolutionary Economic Geography

# 07.07

Co-evolution of firms, industries and networks in space

Anne L.J. ter Wal & Ron A. Boschma
Co-evolution of firms, industries and networks in space

Anne L.J. ter Wal */ ** & Ron A. Boschma *

* Utrecht University, Department of Economic Geography, Faculty of GeoSciences
PO Box 80 115, 3508 TC Utrecht, The Netherlands

** Observatoire Français des Conjonctures Economiques
Department of Research on Innovation and Competition (OFCE-DRIC)
250, Rue Albert Einstein, 06560 Valbonne Sophia-Antipolis, France

e-mail: a.terwal @ geo.uu.nl
e-mail: r.boschma @ geo.uu.nl
homepage: http:// econ.geo.uu.nl/terwal/terwal.html
homepage: http:// econ.geo.uu.nl/boschma/boschma.html

Abstract
The cluster literature suffers from a number of shortcomings: (1) by and large, cluster studies do not take into account that firms in a cluster are heterogeneous in terms of capabilities; (2) cluster studies tend to overemphasize the importance of place and geographical proximity and underestimate the role of networks which are, by definition, a-spatial entities; (3) most, if not all cluster studies have a static nature, and do not address questions like the origins and evolution of clusters. Our aim is to overcome these shortcomings and propose a theoretical framework on the evolution of clusters. Bringing together bodies of literature on clusters, industrial dynamics, the evolutionary theory of the firm and network theory, we describe how clusters co-evolve with: (1) the industry they adhere to; (2) the (dynamic) capabilities of the firms they contain; and (3) the industry-wide knowledge network they are part of. Based on this framework, we believe the analysis of cluster evolution provides a promising research agenda in evolutionary economic geography for the years to come.

Key words: cluster evolution, network dynamics, industrial dynamics, co-evolution, evolutionary economic geography
1. INTRODUCTION
Since the 1980s, concepts like industrial districts, clusters, learning regions and regional innovation systems have conceived regions as drivers of innovation. Broadly speaking, the cluster literature claims that firms in a cluster benefit from knowledge externalities, because geographical proximity facilitates (tacit) knowledge-sharing, because cluster firms participate in extensive local networks, and because they belong to the same institutional environment. However, this way of conceptualising and analysing clusters has become subject to increased criticism. Until recently, economic geographers did not pay too much attention to the fact that firms in a cluster differ widely in terms of size, power and absorptive capacity. In addition, the role of geographical proximity in patterns of knowledge exchange tends to be overemphasized, whereas the effect of networks – by definition an a-spatial concept – tends to be underestimated (Boschma and ter Wal 2007). Finally, most studies analyse clusters from a static perspective, while questions like where clusters initially emerge, and why and how clusters and the advantages associated to them change over time are largely ignored.

While addressing these shortcomings, we propose an exploratory theoretical framework on the evolution of spatial clustering in an industry. This framework is grounded in an evolutionary economic geography approach that tackles questions in economic geography with theoretical insights and concepts derived from evolutionary economics (see e.g. Boschma and Lambooy (1999)). In this particular application to the evolution of spatial clustering, we argue that the evolution of patterns of clustering within an industry is part of wider co-evolutionary processes. These processes involve, beside the clusters themselves, the evolution of the industry’s constituent firms at the micro level, of the industry as a whole at the macro-level, and of the patterns of knowledge-based interaction, as expressed in the industry network. In sketching this framework, we link the geography literature on clusters to the evolutionary theory of the firm, the industrial dynamics literature and network theory.

In section 2, we give a short review of the literature on clusters. In section 3, we present insights from the evolutionary theory of the firm that explain how firms internally differ – in particular in terms of dynamic capabilities. Subsequently, in section 4, we combine the evolutionary theory of the firm with the literature on networks and explain how firm capabilities and their network positions are related through a bidirectional causality. We argue what implications differences in firm capabilities might have for the role firms play in cluster-based knowledge interaction and for the position firms have in the industry network. In section 5, we take a dynamic perspective and relate the evolution of networks to industry dynamics. We explain how networks evolve through the various life cycle stages of the industry and what role the bidirectional causality between firm heterogeneity and network position plays in this process. At the same time, we devote attention to the implications of the evolution of firm heterogeneity, networks and industries for the evolution of spatial clustering in an industry, making the final step towards an evolutionary economic geography approach to spatial clustering.

2. REVIEW OF THE LITERATURE ON CLUSTERS
When we refer to clusters, we have in mind the extensive literature on clusters, industrial districts, innovative milieux, regional innovation systems and learning regions published since the 1980s. While
we acknowledge that these concepts differ to some extent, they all stress the importance of local processes of collective learning, based on a high degree of embeddedness in clusters, in combination with the tacit nature of knowledge (Asheim 1996; Cooke 2001). The extensive literature on clusters has put emphasis on four mechanisms of inter-firm knowledge flows which contribute to their strongly localised character.

The first mechanism concerns a high level of informal interaction within cluster-based communities of entrepreneurs and technicians (Dahl and Pedersen 2004; Grabher and Ibert 2006). Clusters are characterised by a high level of embeddedness that is expressed in a cohesive and rather closed social environment in which entrepreneurs and employees exchange knowledge through informal social networks. In addition, due to the specialised nature of clusters, most of the relevant knowledge is highly tacit and therefore difficult to transfer over large distances. Hence, all firms in a cluster have access to more or less the same knowledge and hence can profit from that accordingly (Asheim and Gertler 2004). This knowledge is inaccessible to firms beyond the boundaries of the cluster: the social distance as well as the cognitive distance (that, in case of clusters, may coincide with geographical distance) make that the cluster’s knowledge does not reach firms outside the cluster or cannot be properly understood (Boschma and Lambooy 2002).

The second mechanism concerns direct inter-firm links in cooperation networks. Because of the high-level of embeddedness, direct cooperation links between firms are likely to be strongly localized within the boundaries of the cluster. The presence of a social community of engineers and entrepreneurs that is interlinked through an informal social network, does not only lead to implicit knowledge exchange on individual bases, it also leads to more explicit acts of collective learning taking place in local cooperative networks (Crevoisier 2004).

Thirdly, knowledge spills over from one firm to another through labour mobility. Next to knowledge flowing through formal and informal networks, appointing new employees is an important way to get access to external knowledge. This is especially relevant for acquiring knowledge in fields a firm is not already active in (Song et al. 2003). Since mobile labour is inclined to stay in their home region, these knowledge flows tend to be geographically localized (Almeida and Kogut 1999; Malmberg and Power 2005).

Finally, the creation of spin-offs can be considered a mechanism of knowledge transfer that tends to be strongly localized (Dahl et al. 2003). Spin-off firms inherit knowledge and experience from their parent company (Klepper and Sleeper 2005). Since spin-offs are strongly inclined to establish their firms in close geographical proximity to their mother company (Sorensen 2003), these knowledge flows tend to be geographically localized as well.

This broad literature on clusters is consistent in the view that inter-firm mobility of high-skilled workers and spin-offs, formal and informal forms of collaboration, and other forms of knowledge exchange are factors that have contributed to the success of clusters (Breschi and Malerba 2001). Because of the four inter-firm knowledge transfer mechanisms, the cluster literature puts a strong premium on the geographical concentration of knowledge flows between firms within clusters (and more generally regions). Consequently, clusters are put forward as key drivers of innovation.
(Malmberg and Maskell 2002). However, in stressing that flows of knowledge in clusters are highly concentrated in space, three misconceptions are made.

First, many cluster studies do not pay close attention to the fact that firms are highly heterogeneous in terms of capabilities, strategies and routines (Nelson and Winter 1982). Clusters mattered, and not so much firms. The performance of firms was largely attributed to their location in the cluster, because of the localized character of knowledge transfer in clusters. Cluster firms were supposed to outperform non-cluster firms, although that was hardly ever put to the test. But more important, the capabilities of firms are most likely to differ within clusters, with major consequences for their performance. Therefore, it would be wrong to treat cluster firms as being the same, and to relate their performance almost directly to their location, without controlling for firm-specific features.

The second shortcoming in most of the cluster literature is that the role of geographical proximity is overemphasised. When it comes to cooperation networks, the literature on clusters implicitly assumes that knowledge stemming from non-local sources is of inferior importance for firm competitiveness (Asheim and Isaksen 2002). However, there is increasing awareness that extra-cluster linkages might be crucial for cluster firms to avoid lock-in. Having said that, at the level of the cluster, there is little empirical evidence that clusters with strong local knowledge dynamics and a high degree of integration in global networks outperform other clusters in terms of growth (Krafft 2004). At the same time, social networks are assumed to be disclosed by the cluster boundaries and labour mobility flows to be essentially local. The degree to which these flows might cross over regional boundaries is often not addressed either in qualitative or quantitative studies on clusters. In other words, most of the cluster literature argues from the idea that it is the local environment of the cluster that affects the behaviour and performance of its constituent firms. If networks matter, their effects are believed to operate at the cluster level. However, networks are by definition a-spatial entities and, therefore, each of the four knowledge transfer mechanisms can be best conceptualized as mechanisms that are possibly, but not necessarily, of a local nature. Like Hendry and Brown (2006) showed in their study of German clusters, firms may take advantage from being connected to a network – irrespective of where their partners are located – rather than from being located in a cluster.

The third drawback of the cluster literature is that most studies are static, notable exceptions being Maggioni (2002), Brenner (2004) and Menzel and Fornhal (2007). This implies that the question as to where clusters come from and to why they emerged as they did received little or no attention from geographers. In addition, only limited attention has been paid to how clusters and inter-firm networks and clusters evolve, and whether the advantages that are associated to geographical clustering persist over time.

3. VARIETY ACROSS FIRMS IN CLUSTERS

Above, we argued that heterogeneity of firms within a cluster is largely neglected in many cluster studies. Evolutionary economic theory offers valuable concepts and ideas to enrich the cluster literature by paying more attention to how firms differ internally and how these differences matter for the roles and positions of cluster firms in knowledge networks. Highlighting the variety across of firms is the first step towards an evolutionary approach to geographical clusters.
The starting point here is the argument of Nelson and Winter (1982) that firms largely differ in their capabilities, strategies and routines. Differences in skills of individual organization members and firm strategies will, in turn, lead to the development of differences in routines and – at a more aggregate level – in firm capabilities. Routines and capabilities of firms are highly idiosyncratic and hence a source of competitive advantage. Important is the distinction between a firm’s capability to carry out highly frequent and strongly routinised daily tasks and its capability to change and develop these operational routines and capabilities (Dosi et al. 2000). According to this line of reasoning, a distinction should be made between substantive capabilities – defined as the ability to solve a problem – and higher order dynamic capabilities – constituting a firm’s ability to change the way a firm solves problems (Zahra et al. 2006). The latter – dynamic capabilities – can be considered the drivers behind the creation or continuation of long-term competitive advantage (Henderson and Cockburn 1994). A firm has to make strategic decisions how to allocate its scarce resources over the commercial exploitation of its existing knowledge on the one hand, and the exploration of alternatives on the other. Whereas for the first, the returns are more certain and immediate, the latter is accompanied by much more risk and uncertainty, but at the same time necessary in the long run to cope with future market and technology developments (March 1991).

The concept of dynamic capabilities fulfils an important role in extending our explanation of divergent patterns of performance of cluster firms (Teece et al. (1997). Zollo and Winter (1999; 2002) perceive dynamic capabilities as a firm’s ability to replace or adapt a firm’s routinized activities of production by more effective operational routines. This implies that dynamic capabilities are a *structural* firm characteristic that conceals in a firm’s ability to introduce innovations in a relatively stable way over time. In other words, dynamic capabilities give way to replicable processes of change that are encapsulated in a firm’s routines. In our framework, dynamic capabilities perform three different functions in the evolution of firms.

The first – and most general – type of dynamic capability is absorptive capacity. The external environment of a firm provides stimuli for a firm to change its focus and to reconfigure its resource base in order to keep up competition. One potential and important way to do so is by use of the external knowledge a firm obtains through its network linkages. Such external knowledge might contain important information how to redirect the development of a firm away from its evolution along existing paths, causing a more path-breaking change in its development (Eisenhardt and Martin 2000). As a consequence, a firm’s absorptive capacity, defined as a firm’s ability to absorb, understand and exploit external knowledge (Cohen and Levinthal 1990) is a highly relevant dimension of firm heterogeneity that is especially relevant for the evolutionary analysis of cluster firms. Although there is still much debate on how the concepts of absorptive capacity and dynamic capabilities are related (Foss et al. 2006), we follow Zahra and George (2002) in recognizing absorptive capacity as a dynamic capability, with distinct cognitive and organizational dimensions.

The second role dynamic capabilities play in our framework of clustering can be considered a further specification of the first. Whereas absorptive capacity as a dynamic capability concerns the effective absorption and application of external knowledge in general, particularly firms that are able to change their network position potentially create a competitive advantage over other firms. As the
technology base of the industry evolves, a fixed position in the network might lead to a decay of newness of information and knowledge that reaches the firm (Grabher 1993). The heterogeneity – or variety across firms – decreases through long-lasting relationships among these firms (Nooteboom 2000). Eventually stability in network position carries the risk of cognitive lock-in among the group of interconnected firms. Then, a higher-order dynamic capability enables a firm to change its network position and hence to create new sources of external knowledge and hence a source of new variety across firms. This network change might in turn have implications for the pattern of spatial clustering of an industry. As to the extent that the new inter-firm relationships are local in comparison to the prior relationships, the need to be spatially proximate is likely to change along. If the new relationships are increasingly of a non-local nature, concentration of the industry in specific clusters might diminish, whereas an increase of local interaction might have the opposite effect.

A third dynamic capability, with special relevance to the cluster concept, concerns the ability of a firm to replicate its effective routines to new locations (Kogut and Zander 1992; Frenken and Boschma 2007). This can take place either when a firm moves entirely to a new location, when part of the firm’s activity is relocated or when it starts a subsidiary in another place, for instance to serve a new market. These acts of relocation directly affect the pattern of spatial clustering of an industry, either reinforcing or diminishing the extent to which an industry is clustered in space.

4. FIRM VARIETY, NETWORKS AND CLUSTERS

In this section, we set out which role networks play in clusters and critically assess the role of geographical proximity in the patterns of interaction in which cluster firms are involved. In doing so, we take a dynamic perspective, overcoming the predominantly static nature of most cluster studies. We argue that variety across firms in terms of capabilities drives the evolution of networks through time.

In cluster studies, the cluster environment and the spatially bounded knowledge dynamics are conceived as the most important forces making clusters and their constituent firms performing well. In order to be an innovative firm, it matters where you are located. However, we stated earlier that inter-firm interaction is not necessarily confined to the cluster area. Therefore, the local nature of knowledge exchange between firms in clusters – being the result of social networks, direct cooperation, labour mobility or spin-off relations – cannot be assumed beforehand. That is not to deny that each of these four types of inter-firm knowledge transfer mechanisms may have a certain bias to be local due to geographical proximity, but this will vary across regions, across industries and across time.

With respect to the time dimension, the literature on how networks emerge and evolve throughout the evolution of an industry is still weakly developed (Malerba 2006). To begin with, one needs to specify the determinants of matching in a network. Ahuja (2000), for instance, argues that the formation of strategic alliances depends on the interplay between inducements and opportunities of the firms involved. On the one hand, firms with superior capabilities to create new technology, products and processes and successfully commercialize them are attractive partners for other firms to start a strategic alliance with. On the other, these firms themselves may not have strong incentives to engage in alliances with firms with inferior capabilities. Whereas firms having a strong knowledge base – and superior dynamic capabilities – are attractive to be connected too, firms with lower dynamic
capabilities might not have any interesting knowledge to offer for others and, at the same time, might not be able to understand knowledge stemming from firms with high dynamic capabilities (Giuliani and Bell 2005). Related to this is Stuart’s (1998) argument that prestigious firms – firms that have built a good reputation through important technological advances – are desirable partners in collaboration. As a consequence, they come to be located centrally in knowledge networks, provided that they are willing to collaborate with less prestigious partners, for instance against attractive financial terms.

In order to be able to communicate and exchange knowledge effectively, the technological or cognitive distance between partners should not be too great (Nootboom 2000). Likewise Lane and Lubatkin (1998) introduced the concept of relative absorptive capacity. They argue that it is the difference in absorptive capacity between related firms that determines the extent to which firms can learn from each other and hence the probability that a linkage between two firms is formed. Giuliani and Bell (2005) showed in their study of a Chilean wine cluster that knowledge diffusion in that cluster takes place mainly in a core group of firms with high absorptive capacities, whereas firms with inferior absorptive capacities remain isolated from the local knowledge network. Boschma and Ter Wal (2007) found evidence in their case study of the Barletta footwear cluster that absorptive capacity was positively related to the amount of non-local knowledge relationships.

The general argument that follows from the above is that a firm’s capabilities – relative to those of potential partners – are a crucial determinant for the formation of linkages. This implies that a firm’s capabilities – as for instance its absorptive capacity – are bidirectionally linked to firm performance (Malerba 2006). At the one hand, firms with a high absorptive capacity are attractive partners to be linked to in a network and hence are likely to be centrally connected in this network. At the other hand, a central network position is (to a certain threshold) argued to be positively related to performance and stimulates the further improvement of a firm’s capabilities. This, in turn, increases the attractiveness of partner, which might make them even more centrally located in the network. In other words, the bidirectional causality between firm capabilities and network position provokes a self-reinforcing and path-dependent process in which firm-internal capabilities and networks co-evolve throughout the evolution of an industry.

5. EVOLVING FIRMS, NETWORKS AND CLUSTERS ALONG THE INDUSTRY LIFE CYCLE

So, networks co-evolve with firm capabilities: the bidirectional causality between firm variety and network position spurs the evolution of networks and capabilities along the life cycle of an industry. In this section, we introduce the literature on industrial dynamics, putting the co-evolution of firms and networks within the wider evolution of the industry as a whole. Doing so, we devote particular attention to the geographical dimension of this co-evolutionary process, as reflected in the pattern of spatial clustering of the industry. The industry life cycle model, as originally developed by Abernathy and Utterback (1978) and further elaborated by Klepper (1997), serves as the basic framework through which the co-evolution of firms, networks and clustering is described. We distinguish between four stages of the life cycle of an industry, and we sketch how these affect the evolution of variety across firms in the industry, the network firms adhere to, and the pattern of spatial clustering accordingly.
First phase: the introductory stage

A new industry emerges when a number of pioneering firms – which can be either incumbent firms coming from a related industry or new start-up firms – introduce a radical innovation. At that time, the technological regime can be characterized by a high uncertainty with respect to the direction of technological development and the identification of the main players in the field (Nooteboom and Klein Woolthuis 2005). It is unclear which standards will become dominant in the emerging industry (Suarez and Utterback 1995). As a consequence, technological variety is high, and the pioneering firms will show considerable variety in their capabilities (Rigby and Essletzbichler 2006). Knowledge and technology are highly tacit and embodied in human capital in the introductory stage of its life cycle (Cowan et al. 2004). The technological regime, characterized by uncertainty and tacitness, is expected to result in instability and volatility at the network level and at the level of spatial clustering.

At the level of the network, the uncertainty associated to technological development makes firms eager to rely on inter-firm relationships. At the same time, however, the uncertainty and lack of knowledge about who are the main players in the field initially lead to a highly unstable network structure. Firms are likely to change links regularly by choosing new cooperation partners or attracting engineers originating from different companies because of this uncertainty. Thus, preferential attachment is not the main driver of network formation at this stage. The choice of partners in this process can be based on social networks (who do you know best) and chance events (accidental meetings with people who coincidentally happen to work on similar issues). Thus, we expect an unstable network in which the firms’ network positions tend to be normally distributed. This normal distribution is caused by the role of social networks and chance factors in partnering decisions.

The same line of reasoning holds for spatial clustering of firms in an emergent industry. The initial phase of industry development is characterized by instable clustering patterns. The forces towards clustering in later phases are not yet in place to exert their full influence. The initial pattern of an industry is mainly dependent on where the pioneers of a new industry emerge. Evolutionary entry models (e.g Arthur 1994) argue that new industries grow on the basis of spin-off dynamics and processes of imitation. Geography matters in this respect, since spin-off firms tend to start their activity in close geographical proximity to their parent company, and successful imitation is most likely to take place in close geographical proximity to the pioneering firms. Therefore, spinoffs and imitation behaviour may set in motion an initial process of spatial clustering. Nevertheless, the question in which locations the spin-off and imitation mechanisms result in industrial clustering is to a high degree dependent on chance events and, as a consequence, mainly the outcome of a probabilistic process (Lamboo and Boschma 2001). In the purest model of this kind, the role of geography in explaining spatial clustering of an industry is completely ruled out.

More realistic models include geographical factors. For instance, regions may have to fulfill generic conditions like infrastructure, human capital and the like, in order to be a potential candidate for the new industry (Storper and Walker 1989). Another geographical factor that may affect the location of a new industry is that it can be either created through related variety or through Jacobs-externalities in space. Related variety concerns the emergence of a new sector that grows out of an existing, related industry. An example is the British automobile sector that initially emerged mainly on
the prior industrial structure of bicycle and coach making firms and the spin-offs they generated. The new industry came to be concentrated in regions that used to be specialized in those related industries (Boschma and Wenting 2007). Hence, who will be the early players of the new industry, and in which locations they will concentrate might be partially dependent on the geographical pattern of prior regional specializations. However, which related industries will provoke the emergence of a new industry remains unpredictable beforehand. New industries may also emerge out of Jacobs’ externalities. Starting from the Schumpeterian idea that innovation basically is a recombination of knowledge and ideas, it is argued that regions with a diversified industry structure, as opposed to a specialized but related structure, are most conducive to breed new industries by means of exploiting so-called Jacobs externalities (Jacobs 1969). Some regions will turn out to be better equipped in terms of a diversified structure than others. Again, it is unpredictable ex ante which recombination will lead to the emergence of a new industry and hence which regions exhibit the right mix of prior industrial activity. Due to this uncertainty – and due to the fact that a new industry can emerge either from related variety or Jacobs’ externalities – many regions are a potential candidate to get pioneering entrepreneurial activity within their boundaries. The windows of locational opportunity concerning the emergence of a new industry are open for many regions, as long as some generic conditions are fulfilled (Storper and Walker 1989).

In conclusion, in the initial phase of industry evolution, chance factors and unpredictable outcomes related to the pioneer’s social networks and the region’s industrial structure produce unstable and volatile patterns of interaction and firm location. The subsequent growth phase of the industry, however, is characterized by forces towards stability in the industry knowledge network as well as in the pattern of spatial concentration of the industry.

Second phase: the growth stage
In the second phase of the industry life cycle, a dominant technological design emerges and the market for products in the new industry expands. As a result, the number of active firms in the industry grows rapidly, mainly through imitation behaviour and the formation of spin-off firms attracted by the high rents in an expanding market (Utterback 1994). The increase in the number of firms through spin-offs and imitation as well as the development of a dominant design result in forces towards stability both at the network level and at the level of spatial clustering.

At the level of the network, a tendency toward the formation of stable core-periphery pattern can be observed, starting from the growth stages of the evolution of an industry. For instance, Orsenigo et al. (1998) showed that during the life cycle of the industry the network of strategic alliances in biotechnology was characterized by a highly stable core-periphery profile. There are several forces that lead to the establishment of this stable pattern. As new firms enter the industry, the network will grow. This growth will mainly take place through the mechanism of preferential attachment. Preferential attachment describes a process of network growth in which new nodes select one of the existing nodes in the network to connect to. The probability of node to be selected is proportional to the number of links this node already has. As a consequence, firms that are centrally
located in the network initially are likely to become even more central (see Barabási and Albert 1999). The preferential attachment process is nurtured by the following forces.

First, there is a strong first-mover advantage. The preferential attachment process is nurtured by the bidirectional causality between capabilities and network position, as explained in the previous section. Since firms with ‘cutting-edge’ technology are attractive partners to be linked to, new entrants are inclined to link themselves to central nodes in the network. As a consequence, a fit-get-richer process in the network can be observed. Gay and Dousset (2005) found evidence that the firms that are continuously found in the core of the network are firms that hold the key patents within the industry. Early players in an industry tend to establish themselves centrally in the network and are likely to retain this position throughout the evolution of the industry. Continuous flows of entry in the industry and, hence, in the network do not result in major deformations of this network structure: entry into the core of the network becomes increasingly difficult for new entrants as the network continues to grow (Orsenigo et al. 1998; Gay and Dousset 2005). As a result, the variety in firm capabilities between central and peripheral firms is growing. In addition, the positive effect of being an early entrant on firm survival (e.g. Klepper 1997) might be partly attributed to the fact that those firms can establish themselves early in the network and get a central position through preferential attachment. An exception to the rule is possibly formed by spin-off firms. New entrants in the industry might be better able to get a more central position in a network when they are a spin-off of an existing (core) firm. These firms have inherited successful routines from their parents (Klepper 1997), and they might have the opportunity to take over part of the network linkages of their parent company.

A second force leading to stable core-periphery patterns can be found at the exit side of industrial dynamics. Firms with inferior network positions are more likely to end their business and to exit the industry (Mitchell and Singh 1996). Conversely, centrally positioned firms will survive and the core-periphery pattern in the network will be reinforced. In order to empirically validate this hypothetical relationship, a firm’s network position should be included in models that aim to explain firm survival probability and industry dynamics. Beside time of entry in the industry and entrepreneurial experience (e.g. as a spin-off) (Klepper 1997, 2002), the (evolving) position of a firm in a knowledge network might act as an additional explanatory variable for the survival probability of a firm. At the same time, the possibility that spin-off firms might take over relationships from their parent company might partly account for the higher survival probability that typically characterizes spin-off firms.

The stability of the network structure is further fed by the fact that the formation of new alliances is largely based on a network of prior alliances (Gulati 1995; Gulati and Gargiulo 1999). Prior direct alliances are likely to have led to the formation of trust and effective routines of cooperation. Additionally, the network of prior indirect alliances acts as a channel of information on opportunities for future cooperation and as reputational circuits concerning the reliability of potential future partners.

Orsenigo et al. (2001) argued that the stable core-periphery pattern is also nurtured through the path-dependent nature of technology development. The fact that the core of firms in a network might continue the development of technology along a certain technological path might strongly diminish the probability that competing technologies will establish themselves. Consequently, firms developing these technologies find difficulty to connect themselves to the industry network and
eventually might fail to survive. In other words, the emergence of collaborations in the early growth stages of the development of a new technology might lead to dominant standards. During subsequent stages of more incremental change, the early developers of new standards are likely to position themselves in the core of the industry network (Suarez and Utterback 1995; Soh and Roberts 2003).

With respect to the spatial level, comparable forces towards stability are likely to be observed concerning the industry’s spatial pattern. In contrast to the first phase, in which no clear-cut pattern of spatial clustering is established, the growth stage of the industry is characterized by forces towards stability. Several forces that lead to the concentration of firms in clusters can be distinguished.

The first force is closely associated to the growth of the number of firms that characterizes the second stage of the industry life cycle. As explained before, in the introductory stage, it is quite unpredictable where visionary entrepreneurs emerge and where the first successful firms generate other spin-off companies or provoke the strongest imitation behaviour. But as soon as they start to develop somewhere, these forces towards clustering are complemented by another force based on agglomeration advantages (Arthur 1994). As soon as clustering occurs somewhere, various types of Marshallian externalities come into being: new infrastructure is built to cope with increasing demand, relevant knowledge spillovers become increasingly available, the labour market becomes more specialized, specialized suppliers emerge after some time, supportive institutions come into being, etc. (Boschma and Lambooy 1999). These agglomeration advantages make it increasingly attractive for new entrants to be located in the emerging cluster and hence further stimulate the evolution towards a stable pattern of geographical clustering (Brenner 2004). As a consequence, industrial concentration selectively takes place in a number of regions only. The more an industry gets clustered, the more difficult it becomes for other regions to localize part of the emerging industry within its boundaries. In other words, as clustering proceeds, the ‘windows of locational opportunity’ close for the regions not taking part in the clustering of the new industry (Storper and Walker 1989).

The process of network growth through preferential attachment that generates a stable core-periphery network has also a distinct geographical component. During the growth stage, many firms enter the industry and want to connect to the industry network. The bidirectional relationship between capabilities and network position gives way to a process of network growth through preferential attachment, in which firms with superior capabilities come to be centrally located in the network. The new links that are added to the network might have a relatively strong tendency to be local, to be concentrated in a cluster. Because uncertainty is still high and the nature of knowledge remains considerably tacit, geographical proximity is especially relevant for the knowledge exchange between firms (Audretsch and Feldman 1996; Cowan et al. 2004). Tacit knowledge flows most easily through the mobility of people, which is likely to take place locally, or through repeated interaction among people, which is eased by geographical proximity as well. In addition to this direct effect, an indirect effect of geographical proximity may stimulate local clustering as well. The uncertainty that is associated to the emergence of a new industry can be partly compensated for through social proximity – and the associated presence of trust – which are likely to coincide to a considerable degree with geographical proximity.
In conclusion, the growth stage of the industry life cycle coincides with stabilizing patterns of interaction in the industry network as well as stabilizing patterns of spatial clustering. This does certainly not imply that the evolution of networks and the evolution of clusters automatically and completely coincide. Although mechanisms of geographical proximity cause a bias of network links to be locally concentrated in clusters, dense and stable parts of the network need not show overlap with established clusters. As a consequence, in addition to clusters characterized by a dense local network structure, there might occur clusters without strong local knowledge-based interaction, as well as stable and dense parts of a network that are dispersed over various geographical locations.

**Third phase: the maturity stage**

The growth of an industry is not infinite. At some point, the industry will show symptoms of maturity. Market size ceases expanding, the number of new entrants will decline rapidly, and the technological potential for further innovation decreases (Klepper 1997). Furthermore, the maturity stage of the industry is characterized by a shake-out process. That is to say, there is a massive wave of firm exits, because the size of firms matter more, and the nature of competition shifts from an emphasis on technology and product innovation to an emphasis on price and cost reduction (Utterback 1994).

At the level of the industry, the variety across firms declines through a massive shake-out. As stated previously, network position might positively impact on firm survival. If the (core of the) knowledge network coincides with the main geographical clusters of the industry, it is very well possible that, on average, firms in these clusters outperform those outside the clusters. For instance, Krafft (2004) demonstrated that during the recent shake-out in the ICT industry, firms in the ICT-business park of Sophia-Antipolis, unlike comparable firms outside the cluster, continued to survive. The park as a whole even continued to grow, though at a lower speed than before. Krafft suggests that strong local knowledge dynamics could have been responsible for the fact that a shake-out did not occur in Sophia-Antipolis. Building on these ideas, we could hypothesise that clusters that are characterized by strong local knowledge dynamics and a high degree of integration in global networks outperform other clusters in terms of growth, especially in the shake-out phase.

However, being peripherally positioned in a network or being located outside a cluster is not necessarily disadvantageous for a firm. It is certainly true that the more stable patterns of interaction among firms that emerged during the growth stage of industry lead to trust-building and provide opportunities for following the lengthy trajectories that are needed to develop innovations. However, the tendency of stability at the level of networks and clusters – that do not necessarily coincide – might get some strong downsides as the industry life cycle proceeds towards maturity.

First, lengthy interaction among firms in stable networks tends to decrease variety in capabilities across firms. Cowan et al. (2006) noted that through collaboration, firms’ competences will become more similar and the technological distance between the two will decrease. This will in turn diminish the opportunities for future learning. As a consequence, firms might decide to break up the redundant network linkages, which will result in a declining network. In line with this hypothesis, Darr and Talmud (2003) found in the electronics industry that the technological dialogue between sellers and buyers was substantially more intense in a sub-sector with emergent technologies than in a more
mature branch of the industry. However, even if relationships among firms endure, the information and knowledge that flow through them gets less valuable through time because firms become more similar in what they know and in what technologies they possess.

Second, the necessity for explicit forms of inter-firm interaction decreases, because knowledge may become more codified in the maturity stage. As the industry evolves, its technological regime changes along (Dosi 1988; Malerba and Orsenigo 1996). Whereas technology and knowledge tend to be highly tacit and embodied in human capital during the first stage of the industry, they get more codified during the growth and particularly the maturity stage (Cowan et al. 2004). At the same time, uncertainty about how technology will develop decreases (Robertson and Langlois 1995). As a result, geographical proximity might be less needed, while congestion costs or high rents in the cluster might make cluster firms decide to move to cheaper locations (Audretsch and Feldman 1996).

Third, when networks are stable over a long period of time, firms might get locked in established lines of thinking (Grabher 1993). It is unlikely this situation of cognitive lock-in will be perturbed, because due to a lack of network change, virtually no new external knowledge – from outside the rigid core structure of the network – comes in. As a consequence, the mature character of an industry in terms of a decline of innovative activity is not merely due to exhaustion of the technological opportunities for further innovation, but does also relate to inertness in patterns of interaction among firms within the industry.

In such a situation of decreasing variety across firms and cognitive lock-in – being the result of the shakeout and the fixed patterns of interaction in dense parts of the industry network or within local clusters –, firms might need their dynamic capabilities to survive in the long run. These dynamic capabilities can be exercised in two ways. First, firms can decide to delocalize (part of) their activity to other (cheaper) locations in order to avoid congestion costs and high land prices in the cluster. In order to effectively replicate their successful routines to the new location, firms need dynamic capabilities. The relocation decision of firms directly affects the spatial clustering of the industry, leading either to a more dispersed spatial pattern, or to the emergence of new clusters. Second, firms need their dynamic capabilities when they want to change their network position radically. A new position in the industry network, for instance, connecting to a group of firms that are devoted to more up-to-date technology, might enable a firm to break through the situation of cognitive lock-in (Glückler 2007). In order to do so, a firm might even decide to relocate to another cluster within the industry that does not suffer yet from the negative spiral of cognitive lock-in.

In short, the maturity stage of the industry life cycle is characterized by a massive shake-out which is highly selective as to which firms exit the industry. Firms with a peripheral network position are more likely to exit the industry than firms in more central network positions and places. The result is that the variety across firms in the industry decreases. A tendency of cognitive lock-in is likely to emerge due to fixed patterns of interaction. This can take place either in dense and stable parts of the industry network, or in clusters with a dense interaction structure. Firms need dynamic capabilities to overcome such a situation of cognitive lock-in, changing either their network position or their location. Cognitively locked-in firms that are not able to do so are likely to be part of the industry’s shake-out.
Fourth phase: industry decline or the start of a new cycle

The maturity phase of the industry life cycle coincides with a shake-out process among the population of the industry and with increasing negative effects of the relatively stable core-periphery profile of the industry network. In the fourth phase, two different scenarios are possible.

First, if no radically new technologies are introduced, the industry will eventually decline. The market demand for the industry’s products might decrease rapidly, and the innovative potential of the industry may become completely exhausted (Utterback 1994). Eventually, the survivors of the industry are forced to exit the industry when they are not able to diversify to new industrial activities by exercising their dynamic capabilities. For individual firms, a situation of lock-in can also be perturbed through ‘relocating’ themselves in other more vibrant parts of the network, or in more ‘up-to-date’ geographical locations by means of their dynamic capabilities. However, this might not be sufficient for breaking the inertness of the network and the industry at a more aggregate level. These stable patterns can be disturbed only through exogenous shocks such as the implementation of new basic technologies (Buckhardt and Brass 1990; Orsenigo et al. 2001).

Second, in case there is an exogenous shock like the development of radical technological breakthrough, a new cycle of industry evolution and an associated evolution of networks can be provoked. Successive waves of new technologies might radically reshape the structure of an industry network (Gay and Dousset 2005). When such a breakthrough is developed by firms that are peripherally located in the network, this shock is an opportunity for them to structurally improve their network position. Experienced firms, on the other hand, might react slowly on new challenges in the industry, for instance because of inferior dynamic capabilities or cognitive lock-in. As a consequence they might have to pass leadership to new pioneers and new entrants (Dosi et al. 2000), and lose their central network position. A radical reshuffling of the structure of network might be the result. By contrast, when radically new technologies are invented by established firms, the existing structure of the network tends to be further reinforced (Soh and Roberts 2003). In line with this, Madhavan et al. (1998) distinguish structure-reinforcing and structure-loosening exogenous shocks.

The firms causing the exogenous shock are not necessarily located in the existing clusters of the industry. Where this new activity emerges, is largely dependent on chance factors, as in the first phase. Since the pioneering firms bringing the new technology are likely to be located outside the current clusters, they might not only reshuffle the industry network, but also its spatial pattern. New clusters of firms with path-breaking technology can emerge outside the traditional core clusters (Storper and Walker 1989): new pioneering firms might emerge at the technological frontier, the core of the industry network will redirect itself around the new core of pioneers, and the new firms may set in motion clustering dynamics in new regions. As explained before, where a new industry emerges is not completely random. Viewing radical innovation as recombination, new industries can emerge from Jacobs’ externalities or related variety. That is to say, either regions with a diversified economic structure or regions with a structure of related industries might have a higher probability to function as seedbeds for new industries. Due to the unpredictable nature of innovation, the question which regions exhibit the right mix can be answered only ex post.
When a new technological breakthrough is introduced, a new cycle of co-evolution of firms, networks, industries and clusters might start. Dependent on the extent to which the ‘new’ industry has its roots in the previous one, the new cycle will involve new players and new clusters. Firms from the old technology that had superior dynamic capabilities might have been able to survive and to leap successfully to the new industry. By contrast, firms with inferior dynamic capabilities might eventually die, in particular when the new technology completely substitutes the prior one.

6. CONCLUSION

In this paper, we argued that most cluster studies suffer from a number of shortcomings. First, they often neglect that firms in a cluster differ in terms of internal capabilities. In the context of cluster firms, we have claimed that absorptive capacity – conceptualized as a dynamic capability that captures the cognitive and organizational dimensions of absorbing external knowledge effectively – is an important dimension of this heterogeneity. Second, these studies tend to overemphasize the role of geographical proximity in patterns of inter-firm knowledge flows. As a consequence, the role of networks is often underestimated. Finally, the majority of cluster studies is static and does not address questions concerning the origins and evolution of clusters. In providing an evolutionary approach to spatial clustering, we made an attempt to overcome these shortcomings, setting up a theoretical framework on how clusters co-evolve with the industry they adhere to, with the (variety of) capabilities of firms in that industry, and with the industry-wide knowledge network they are part of.

The central idea in the framework we proposed is that the pattern of spatial clustering in an industry co-evolves with three entities: with the firm at the micro level, with the industry and its technological properties at the macro level, and with the network that describes the patterns of interaction among firms of the industry. We made a distinction between various phases of the industry life cycle: the introductory, the growth and the maturity phase. These phases are either followed by structural decline of the industry, or a ‘regenerative’ phase in which breakthroughs provoke the start of a new cycle. The hypothesized outcomes of this co-evolutionary process are summarized in Table 1.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Firm Variety</th>
<th>Industry Variety</th>
<th>Network Variety</th>
<th>Clustering Variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introductory stage</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>2. Growth stage</td>
<td>Increasing</td>
<td>Increasing</td>
<td>High, but decreasing</td>
<td>High, but decreasing</td>
</tr>
<tr>
<td>3. Maturity stage</td>
<td>Decreasing</td>
<td>Decreasing (shake-out)</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>4A Decline</td>
<td>Decreasing</td>
<td>Decreasing</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>4B Start of a new cycle</td>
<td>Increasing</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
At the level of the firm, the heterogeneity in capabilities is responsible for divergent patterns of firm network position and hence firm performance. At the same time, the evolution of networks and clusters affects the heterogeneity among firms by increasing or decreasing variety in capabilities. Furthermore, individual firms need dynamic capabilities in later stages of the industry life cycle, characterized by a considerable risk of cognitive lock-in, in order change their network position or to relocate and replicate their routines to new – more vibrant – locations.

At the level of the industry, entry and exit dynamics might be selective in the extent to which they concern firms in clusters or not. This selectiveness directly affects the pattern of spatial clustering. In addition, the changing characteristics of the industry’s technological regime throughout the evolution of an industry result in changes in the necessity and hence the tendency for firms to cluster in space. The negative effects of clustering might even come to prevail over the positive effects as the industry evolves towards maturity.

At the level of the network, networks and clusters experience a similar pattern of evolution throughout the various stages of industry evolution. After initial clustering induced by spin-off and imitation dynamics, clusters and networks may become interlinked through the working of a bias in network growth towards the formation of local linkages. Among other things, this bias is based on the tacit character of knowledge and the high level of uncertainty during the growth phase. Both factors make knowledge-based interaction among firms easier in case of geographical proximity. As a consequence, parts of the industry network tend to become localized in spatial clusters. However, since this is a probabilistic process, dense parts of the industry network do not necessarily show complete overlap with the pattern of spatial clustering.

It is important to note here that our exploratory evolutionary approach to clusters needs further development and refinement from a theoretical perspective. In particular, we are in need for empirical validation of the ideas we suggested. Therefore, our contribution should be considered mainly as a research agenda, inviting researchers to tackle the numerous theoretical and empirical challenges.

Further refinement of our theoretical framework is particularly necessary with respect to the role of institutions. In order to streamline our approach, we did hardly pay attention to the role of institutions, although we acknowledge institutions play a crucial role in clustering and network formation over time (Mumman 2003). Many research challenges remain in how an institutional set-up – at the level of cities, regions or nations – develops over time as new industries emerge and others decline. Maskell and Malmberg (2007) suggest that institutions in a region develop path-dependently, in response to the special requirements of the region’s dominating industry. As industries evolve and new ones emerge, this path-dependency may turn into inertness, closing the way for alternative paths of development associated to the emergence of new industries.

The mechanisms underlying our framework on co-evolution of firms, industries, networks and clusters need thorough empirical testing. Although our framework is based to a certain extent on prior empirical research, a key challenge remains to validate the consistency of the framework as a whole, as well as several mechanisms of co-evolution that underlie it, by means of extensive empirical research across industries. Doing so, we believe the analysis of cluster evolution provides a promising and challenging research agenda in evolutionary economic geography for the years to come.
References


