(HOW) DO (FIRMS IN) CLUSTERS CREATE KNOWLEDGE?

Anders Malmberg and Dominic Power*
Department of Social and Economic Geography
CIND - Centre for Innovation and Industrial dynamics
Uppsala University
P.O. Box 513, S-751 20 Uppsala, Sweden
Telephone: +46-18-4710000
Fax: +46-18-4717418
dominic.power@kultgeog.uu.se
anders.malmberg@kultgeog.uu.se

Abstract
The literature on clusters and cluster building has been rapidly growing both in academic and policy-making circles. Central to this interest and body of work has been the assumption that clusters create or generate knowledge and that since knowledge is the basis of firm competitiveness clustered firms and industries will outperform others. Taking its starting point in a discussion of the cluster concept, this paper sets out to examine the evidence for proposition regarding the knowledge-enhancing qualities of clusters by reviewing and examining the literature with the expressed intention of testing whether such claims in fact rest upon rigorous and verifiable empirical findings. In order to do this we extract from the theoretical literature on clusters a series of foundational theoretical arguments. These are then grouped under three hypothetical arguments for the knowledge creating and competitiveness generating power of clusters: Knowledge in clusters is created through various forms of local inter-organizational collaborative interaction; Knowledge in clusters is created through increased competition and intensified rivalry; Knowledge in clusters is created through spill-over following from the local mobility and sociability of individuals. The paper goes on to assess and evaluate the number and rigour of empirical cases supporting these types of argument. The paper thus provides both a 'distilled' theory of if and how clusters create knowledge and an extensive review of empirical evidence for whether clusters do in fact create knowledge and competitiveness.

Keywords: Cluster; Knowledge

JEL - codes: L00; O00; O3; P00; R00
(HOW) DO (FIRMS IN) CLUSTERS CREATE KNOWLEDGE?

Introduction

How is the performance of firms affected by the characteristics of their place of location, and why do certain local milieus prosper while others don’t? Why do similar or related firms so often locate nearby each other, and how are such patterns of regional specialisation reproduced over time? The increased focus in recent years on the role of learning and innovation as key processes behind sustained industrial competitiveness has brought some of these core considerations of economic geography onto the research agenda in a wider community of economic scholars (Storper, 1995; Porter, 1998; Fujita et al., 1999). These questions are also being addressed in a new way. A contemporary view of these issues can be summarized in four broad assertions.

First, in today’s knowledge-based economy, the ability to innovate is more important than cost efficiency in determining the long-term ability of firms to prosper. Innovation is here defined broadly, as the ability to come up with new and better ways of organizing the production and marketing of new and better products (Porter, 1990; Lundvall, 1992; Nelson, 1993; Nonaka, 1994; Grant, 1996). This does not mean that cost considerations are unimportant, but simply that the combined forces of globalisation of markets and deepening divisions of labour make it increasingly difficult to base a competitive position on cost-advantage only.

Second, innovations predominantly occur as a result of interactions between various actors, rather than as a result of the solitary genius (Håkansson, 1987; Hippel, 1988; Lundvall, 1992). This fits with a Schumpeterian view of innovations as new combinations of already existing knowledge, ideas and artefacts (Schumpeter, 1934 (1959)). Most innovations are based on some form of problem solving. Someone perceives a problem and turns to someone else for help and advice. In an industrial context, these interactions often follow the value chain (Malmberg and Power, 2003 forthcoming). A firm facing a particular problem turns to a supplier, a customer, a competitor or some other related actor, to get help in specifying the problem and defining the terms for its solution. From this follows that the level of analysis for understanding the processes of industrial innovation and change is some notion of an industrial system or network of actors carrying out similar and related economic activity.

Third, and this is where ‘geography’ enters the picture, there are a number of reasons why interactive learning and innovation processes are not spaceless or universal, but on the contrary unfold in a way where geographical space plays an active role. Spatial proximity carries with it, among other things, the potential for intensified face-to-face interaction, short cognitive distance, common language, trustful relations between various actors, easy observation, and immediate comparison (Malmberg and Maskell, 2002). In short: spatial proximity seems to enhance processes of interactive learning and innovation, and therefore industrial systems should be assumed to have a distinctly localised component.

Fourth and finally, an implication is that there are reasons to believe that the knowledge-enhancing structures of a given geographical territory are more important than other characteristics (such as general factor supply, production costs...
etc) when it comes to determining where we should expect economic growth and prosperity in today’s world economy.

This brief paper is structured in three main sections. The next section is about the merits and shortcomings of one of the more influential approaches in economic geography in recent years – the cluster approach. The subsequent section posits three hypotheses on the role of clusters – and most specifically firms in clusters – in the creation of knowledge. Following from this an assessment of the evidence available supporting each of these hypotheses is given. The paper concludes by suggesting that at present empirical evidence for the role of clusters in knowledge creation is mixed, that some received hypotheses seem to constitute more promising avenues than others, and that the research field generally is in need of more (rigorous) empirical ‘testing’ in order to proceed further at the conceptual/theoretical level.

Clusters: functionally or spatially defined entities?

“Clusters are geographic concentrations of interconnected companies, specialized suppliers and service providers, firms in related industries, and associated institutions (e.g. universities, standard agencies, and trade associations) in particular fields that compete but also cooperate. Such clusters are a striking feature of virtually every economy, especially those of more economically advanced areas.” (Porter 2000, p. 253)

Persuasive arguments like the one cited above have become widely circulated in academic as well as in policy circles since the publication of Michael Porter’s book on The Competitive Advantage of Nations in 1990. In a recent paper, Martin & Sunley scrutinize the cluster concept and the broader “cluster trend” in economic geography and related disciplines and advance a number of more or less justified points of critique (Martin and Sunley, 2003). While providing a great review of a rapidly proliferation research field, their critique in our view overshoots the target in some important ways. In addition it fails to recognize some of the genuine advancements that “the cluster turn” has actually brought with it.

The cluster approach as presented by Porter (Porter, 1990) and subsequently developed by himself, his associates and others (Porter, 1994; Malmberg et al., 1996; Enright, 1998; Maskell et al., 1998; Porter, 1998; Porter, 2000; Malmberg and Maskell, 2002), brings some genuine contributions to the analysis of the key issues of economic geography. First, it provides a way to describe the systemic nature of an economy, i.e. how various types of industrial activity are related. Porter’s starting point here is the cluster chart (see Figure 1). Beginning with the firms in the industry where we find the main producers of the primary goods of the cluster (be they heavy trucks, telecom equipment or popular music), the chart proposes a way to analyse how these firms and industries are connected to supplier firms and industries providing various types of specialised input, technology and machinery and associated services, as well as to customer industries and more indirectly related industries

This way of approaching the ‘systemicness’ of economic activity has much in favour of it. It opens up a scope for analysing interactions and interdependencies between firms and industries across a wide spectrum of economic
activity. An additional advantage is that it contributes to the bridging of a number of more or less artificial and chaotic conceptual divides that characterize so much work in economic geography and related disciplines. These include, for example, manufacturing vs. services, high tech vs. low tech, large companies vs. SMEs, public and private activities etc. A single cluster, defined as a functional industrial system, may embrace firms and actors and activities on both sides of each of these divides (Dicken and Malmberg, 2001).

**Figure 1. The cluster chart: actors in an industrial cluster**

Furthermore, Porter’s model of the determinants of competitiveness in cluster, known under the “brand” of the diamond model, identifies a number of mechanisms proposed to foster industrial dynamism, innovations and long-term growth. Essentially, the model is built around four sets of intertwined forces related to factor conditions, demand conditions, related and supported industries, and firm structure, strategy and rivalry, respectively. The point here, which for example Martin and Sunley fail to acknowledge in their recent critique, is that the treatment of these factors include several points that are indeed novel (Martin and Sunley, 2001; Martin and Sunley, 2003). In particular this applies to the treatment of factor conditions (insisting that natural resources and general wage levels are subordinate to specialised infrastructures and skills, and the introduction of the notion of selective factor disadvantages), and the emphasis of the role of sophisticated local demand and local rivalry as drivers of innovation.

At the same time, there is indeed a good deal of fuzziness surrounding the cluster concept (Markusen, 1999; Martin and Sunley, 2001; Martin and Sunley, 2003). In our view, the really disturbing lack of clarity exists at the most basic level: what is meant by the terms cluster and clustering? This seemingly trivial question is causing continuing and increasing problems. We are not thinking here of subtle definitional issues relating to the scales, boundaries of and criteria for identification of
clusters. Instead, in our view, the main confusion is related to whether clusters and clustering are primarily functional or indeed spatial phenomena.

It seems obvious the cluster concept has come to denote (at least) three rather different phenomena. The first is a functionally defined industrial system, composed by all the actors, resources and activities that come together to develop, produce and market various types of goods and services (as captured in Figure 1 above). The second is the spatial agglomeration of similar and related economic activity. The third is a more policy-driven concept where a cluster has come to refer to a specific policy initiative. In our view, industry cluster, localised cluster and policy cluster could work as relatively straightforward notion for these respective phenomena. In the existing literature, it is often hard to tell whether the use of the cluster concept refers to either of the three phenomena or indeed some imagined ‘ideal case’ where all three coincide.

This is more than a question of terminology. It seems obvious that (functional) industry clusters will normally not be confined to, or contained within, any narrowly defined and spatially bounded scale (Malmberg and Power, 2003 forthcoming). On the contrary, most industry clusters will have widespread global connections and if we would be able to identify their boundaries in spatial terms, the spatial scale would in most cases certainly not be an urban region. By making spatial configuration (i.e. degree of agglomeration) an attribute of an industry cluster, rather than part of its definition, one could better establish a platform for more fruitful analyses of how ‘geography’ comes into play in the overall process of industrial competitiveness, growth and transformation. In other words, rather than trying to squeeze ‘cluster charts’ into narrowly defined regions (where they rarely will fit in), we should research the hypotheses found in the ‘diamond’ regarding the role of proximity and local milieu on the proposed mechanisms leading to competitiveness.

When it comes to spatial agglomerations of similar and related economic activity, i.e. localised clusters in the terminology proposed here, there are also reasons believe that firms in such settings are less interrelated than Porter and others have led us to believe. This issue will be further developed in the following section.

Clusters, localisation and knowledge creation: three received hypotheses

A preoccupation with spatial readings of the cluster concept has contributed to sidetracking empirical research on clustering. The introduction of the cluster concept could have triggered lots of research on the fruitful issue of how industrial transformation occurs as a result of interactions within and across industrial systems (i.e. clusters defined in the functional sense) and the role of geographical proximity (concentration or agglomeration, i.e. clustering in the spatial sense) in such processes. Instead, we would argue, there has been far too much focus on interaction between firms within geographically defined spaces and numerous rather pointless attempts of trying to assess to what degree there is actual interaction going on locally and thus whether a specific region can indeed be said to contain a “fully-fledged” or “true” cluster or not (Martin and Sunley, 2001; Martin and Sunley, 2003).
Thus, the empirical validation of the propositions advanced in the cluster literature leaves a lot to be desired. This is due to the fact that there has been a general reluctance to spell out the theoretical propositions made in a form that would make it possible to subject them to systematic empirical validation. Nonetheless it is possible to identify certain broad areas of agreement in the literature that could be seen as basic underlying hypotheses driving current research. Consider the following three hypothetical propositions, for example:

1. Knowledge in clusters is created through various forms of local inter-organizational collaborative interaction. This hypothesis is grounded in the proposal that firms that collaborate more on technology with firms and other actors (e.g. universities) in the local milieu will innovate more, and in the idea that firms that meet sophisticated demand from demanding customers in the local milieu will be forced to innovate at a higher pace than other firms.

2. Knowledge in clusters is created through increased competition and intensified rivalry. The claim here is that rivalry between similar firms in a local milieu will be more intense, almost emotional, and this will create a pressure to innovate in order to outsmart the local rival. In part, this is related to the fact that co-located firms are more visible to each other, and thus that observation, monitoring and benchmarking thereby is easier and more efficient. Therefore, firms with nearby rivals will be more innovative than firms who have their main competitors located elsewhere.

3. Knowledge in clusters is created through spill-over effects following from the local mobility and sociability of individuals. This hypothesis is based on the idea that knowledge diffusion will be more rapid among local firms than among globally dispersed firms, due to the intensity of informal interaction in the local milieu as well as through flows of people in the local labour market.

These, we would argue, are all interesting and researchable hypotheses that could be deduced from the cluster literature, based on the underlying argument that the forces that enhance the dynamism of an industry cluster are strengthened by geographical proximity, via a series of mechanisms. We will in the following section present reviews of the existing evidence for each of the three hypotheses presented below. It is hoped that in doing so we can point out some key findings and some key lacunas present in the empirical evidence and provoke constructive debate on where empirical efforts may best be targeted.

The status of the empirical evidence

*Knowledge in clusters is created through various forms of local inter-organizational collaborative interaction*

Throughout the cluster literature it is often said that knowledge creation is enhanced by various forms of inter-organisational collaborative action. Such a hypothesis is most often said to be exemplified by two localised collaborative phenomena: inter-
firm cooperative transactions; and cooperation links or partnerships with nearby R&D institutes, universities and the like.

What we found in relation to the literature on inter-firm transactions (such as buyer-supplier relations, etc.) and inter-firm collaboration is an extremely mixed picture. Certainly there is some consistent evidence that intensified input-output relations and supplier-buyer relations characterise many industry clusters. However, there exist a number of studies pointing out that what local inter-firm transactions that do occur are often exchanges of basic inputs and services rather than R&D or knowledge (Brown, 2000). In addition to there being limited local inter-firm transactions, there is also limited evidence supporting the widespread existence – and therein benefits or not – of local organised inter-firm collaboration.

Nevertheless there does exist a number of industries and studies, where transactions and collaboration were seen to be enhanced by a firm's location within a localised industrial cluster. In particular, there is a rapidly growing set of empirical results suggesting that local business-university links and localised tacit knowledge might be important in some industries. What is not clear however is that such collaborations, etc. with R&D institutes and universities are important to clusters' knowledge and innovations systems in the long term. Audretsch & Feldman’s study of US clusters suggests that while the generation of new economic knowledge tends to result in a greater propensity for innovative activity to cluster during the early stages of the industry life cycle, innovative activity tends to be more highly dispersed during the mature and declining stages of the life cycle (Audretsch and Feldman, 1996). They argue that this may suggest that the positive agglomeration effects seen during the early stages of the industry life cycle get replaced in the latter stages by congestion effects and that local collaboration with universities, etc. is more important in early phases of a 'cluster life cycle'.

In assessing the role of clusters in knowledge creation it is important also to note that many studies point out that extra-local collaboration is often more important than collaboration which happens locally. Hendry et al.’s study of the opto-electronics industry found that national and international relationships were much stronger than local ones and that this was a function of customer and supplier markets. Indeed this is interesting in that it draws attention to the fact that commercialisation processes have an enormously important role in knowledge creation and innovation processes and one cannot assume that firms will be located near commercialisation sites (Hendry et al., 2000). Staying with the electronics sector, Kearns and Gorg’s study of the Irish electronics industry found that though there was evidence of cluster effects the leading firms (most often foreign) undertook the majority of their research abroad and that there was little or no spillover between the larger global firms and indigenous ones (Kearns and Gorg, 2002).

In summary, we found that there is evidence available showing that there are generally limited transactions going on between firms in the local milieu. When asked questions about where the most important suppliers or customers are located, most firms report fairly spatially extended networks and that global connections tend to dominate over the local (Larsson and Lundmark, 1991; Angel and Engström, 1995; Larsson, 1998; Markgren, 2001). Secondly, there seems to be limited formal collaboration between firms (or between firms and other organizations) going on locally. Such collaborations tend to follow the value chain and therefore tend to be
fairly globally extended (Larsson, 1998; Fuellhart, 1999; Zeller, 2001; Owen-Smith and Powell, 2002; MacKinnon et al., 2003).

Equally we found a number of studies bemoaning the fact that little or no significant local collaboration and transactions could be found in their target of study (Barrera, 2002; Watts et al., 2003). The tone of some of these studies is disappointment grounded in a clear normative framework that assumes localised collaboration and transactions are, of course, key to knowledge creation (and a whole host of other positive competences). The underlying problem, though, is the expectation at the outset that there should be a high magnitude of local inter-firm relations for a cluster to be said to exist – and indeed for a cluster to be dynamic and prosper. In that case, the disappointment is largely a result of an initial misconception. If we accept that industry clusters are normally not confined to local milieus, then we should not expect them to be primarily locally integrated. There would presumably be much to gain from dropping the underlying assumption that ‘the more localised interaction, the better’.

**Knowledge in clusters is created through increased competition and intensified rivalry**

Porter himself has long said that rivalry is an essential part of cluster dynamics. Throughout his work on clusters he has consistently argued that perhaps the most important dimension of localisation is that it acts to boost domestic rivalry, which in turn he regards as the most important element behind a nation’s competitive advantage. Given that from the start rivalry and competition were corner stones of the cluster literature it is perhaps rather surprising that rivalry has been almost totally ignored by empirical studies. Rather than looking for rivalry and competition much of the empirical work has, as pointed out above, favoured the search for cooperation and almost altruistic behaviour among firms within clusters. Nevertheless a small body of work is beginning to take rivalry seriously and show that in some cases rivalry seems to ‘drive competitiveness’.

One recent example of a study that attempts to account for knowledge creation in terms of the impact of rivalry is Boari et al’s interesting and careful study of the role of rivalry in the Italian packaging machinery industry (Boari et al., 2003 Forthcoming). They introduce the paper by suggesting that rivalry is a cognitive and social dimension underlying action in competitive business environments. They make the point then that rivalry is very different and much more intense than simple competition. Competition is something more general whereas rivalry is to do with individual firms’ and actors’ behaviour towards others operating in the same market and as such can be a very powerful force. The empirical part of their study showed that rivalry was in fact a key driving force in entrepreneurs’ efforts and a useful tool, or focus, they used to better understand the competitive environment.

Another example of a study that shows that localised rivalry can drive the competitiveness of a cluster or industrial district is Power and Hallencreutz’s paper on creativity in the Jamaican music industry (Power and Hallencreutz, 2002). In this case it is clear that personal rivalries combined with (and partly driven by) resource scarcity have helped to drive localised actors there to breakneck production speeds and extremely rapid innovation and knowledge transfer (most often through imitation and indeed theft). The paradox here is that although rivalry has created one of the world’s most productive and innovative music industry clusters the extreme
nature of that rivalry has undermined firm and industry structures and made an effective intellectual property regime impossible.

Despite such indicatory evidence it is of course plausible that firms can be creative, innovative, create knowledge and do very well even in the absence of any local rivalry. Larsson’s work on the Swedish machinery industry makes the interesting point that there are plenty of companies that because of the unique, or tailored, nature of the products they produce believe they have no rivals. Indeed several studies have reported that many firms believe that they have few direct competitors worldwide, and that those few rival firms are located elsewhere or abroad (Glaeser et al., 1992; Audretsch and Feldman, 1996; Baptista and Swann, 1996; Baptista and Swann, 1998; Malmberg et al., 2000). In such cases there is perhaps little need for knowledge transference or learning at the level of a cluster or industrial system.

In summary there is growing evidence that there is indeed sometimes intense local rivalry. However despite being a basic hypothesis worthy of scrutiny the issue of rivalry in knowledge creation has been relatively underserved by the research community.

**Knowledge in clusters is created through spill-over following from the local mobility and sociability of individuals.**

Another promising avenue of research follows from a focus on the role of individuals’ work and social lives. By focusing more attention on individuals and actors in and around clusters and firms there is growing evidence that: important knowledge exchanges do seem to take place in informal interactions locally; and that the transfer of labour between firms/organisations in localised clusters does seem to boost knowledge creation.

Increasingly research has shown that informal social networks are crucial to the functioning of many industries, innovation processes, and cluster cohesiveness. Grabher’s work on the advertising industry points very clearly to the importance of social networks and also makes the interesting point that for temporary collaborations or project-based working (an increasing trend in many service and cultural industries) social networks play a vital role (Grabher, 2001; Grabher, 2001; Grabher, 2002; Grabher, 2002; Grabher, 2002). Indeed the role of communities and networks is often highlighted as being important in the creation of knowledge. Lissoni has written about mechanical firms in Brescia, Italy and the way in which certain groups of mechanical engineers within a cluster can form “epistemic communities” that is more crucial to knowledge and innovation than the firms themselves (Lissoni, 2001).

Much of the research on actors within clusters has concentrated on how existing networks of actors “negotiate” (Thrift and Leyshon, 1994) discourses, cultures, milieus, practices and tacit or untraded knowledges. Other researchers are starting from slightly earlier in the individual’s engagement with firms within clusters: i.e. how labour market flows, entry, and exit act as the key determinant of who, where, and when potential knowledge creators are. Thus rather than focusing on often diffuse and vague notions that knowledge and innovation reside ‘in the air’ or in the ‘buzz’ of urban life emerging research is focused on the role of flows of workers in the labour market and within firms themselves.
Studies using mainly quantitative data on labour markets and workers’ movements confirm the idea that labour mobility is a key to knowledge and innovation creation and transference. Dahl and Peterson’s (Dahl, 2002; Dahl and Pedersen, 2003) work on engineers in Denmark, and Almeida and Kogut’s (Almeida and Kogut, 1999) on the mobility of engineers and patent holders in the US semiconductor industry give more than adequate empirical backing to the idea that labour market movements are an essential area of study. Using the extremely comprehensive Swedish taxation and population records Power and Lundmark (Power and Lundmark, 2003) found clear evidence that the most successful industrial clusters in Stockholm were the ones with the highest rates of inter-firm labour mobility and also that the most successful individuals were the ones with the highest rates of labour mobility. Thus there is evidence to say that there tends to be labour mobility between firms locally and that this characterises successful clusters. However, it is also worth noting that many firms see this as a problem more than as an advantage (Almeida and Kogut, 1999; Lawson, 1999; Dahl, 2002).

Conclusions

In this paper we have posited three received hypotheses and used these as a guide to assessing what the current status of empirical knowledge on clusters and knowledge creation is. Despite being a somewhat bold and stylised overview we believe that the summary of empirical results presented above is fairly accurate.

A crude summary of the above is that there tend to be modest commercial relations between firms within spatial clusters, that other types of collaboration are more common locally and temporally but that also such relations normally extend well beyond the borders of narrowly defined regions, that the degree of local rivalry varies, that informal knowledge exchanges do occur, and that local labour mobility is presumably an important factor. Equally crudely and generally, the evidence to date suggests to us that rivalry, labour mobility and knowledge spillovers following from informal types of social interaction are more likely to be important advantages of spatial clustering than organised inter-firm transactions and collaborations. In short, there seems little evidence that organised inter-firm transactions and cooperation characterises successful firms, however evidence is building that rivalry, labour market dynamics, and knowledge spillovers can play important roles in firms’ and clusters’ competitiveness.

From the perspective of two geographers, a particularly interesting theme apparent in the empirical findings is that localised clusters of similar economic activity are normally not ‘locally defined industrial systems’. Much of the evidence seems to point to a reality in which industrial sectors are a lot less spatially localised than geographers and regional scientists might like to wish. This is most likely due to the fact that in the area of innovation and knowledge creation it seems especially important for firms and clusters to seek out the best possible partners or to challenge those they value as being their closest rivals. Whilst occasionally this happens locally it is not surprising perhaps that in the modern business world this does not necessarily happen locally (and perhaps should not?).

In conclusion we feel the research field is in need of more (rigorous) empirical ‘testing’ in order to proceed further at the conceptual/theoretical level. It is
important that research takes the basic received hypotheses on the role of clusters in knowledge creation and try to empirically valid them. Our study of the available empirical evidence shows that there are many studies out there right now. However, the bad news is that there is still very little evidence of the actual effects of clustering and moreover the evidence that does exist does not seem to show what we want them to show.
References


betydelse. [Is proximity a geographical question in business relationships].
Uppsala, Uppsala University, Department of Business Studies.


