State of the Art Review

Identifying Clusters – A Review of Methodological Approaches

Chris van Egeraat
Maynooth University
Chris.vanegeraat@MU.ie

SOTA Review No 9: October 2018

Although a seemingly intuitive concept, the identification of clusters involves a number of complexities linked to (1) the cluster definition (the industry boundaries) and (2) the spatial determination of clusters. Over the last 20 years, researchers have proposed a range of methods to address both aspects. The choice of cluster identification method will depend on the policy question, geographical context, type of industries under investigation and, importantly, data availability.

A combination of Delgado's et al. (2016) cluster definition method and Feser's et al. (2005) method for identifying geographical concentrations would usefully inform many policy contexts.

Background

The publication of *Competitive Advantage of Nations* by Michael Porter (1990) popularised the cluster concept as an industrial policy tool. Porter initially paid little attention to the spatial aspects of the cluster concept. In his book, he maintained a distinction between his concept of a cluster, as an inter-linked set of firms and institutions and a geographically proximate group of firms and institutions. The book has stimulated a substantial body of work seeking to develop methodologies suitable for the identification of clusters.

Porter (1998) defines clusters as geographic concentrations of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities. Although a seemingly intuitive concept, the identification of clusters involves a number of complexities linked to (1) the cluster definition (the industry boundaries) and (2) the spatial determination of clusters.

Over the last 20 years, researchers have made significant contributions on both aspects but, partly due to data limitations, the suggested methodologies may not be practical in all contexts. This review outlines the state-of-the art of cluster identification methodologies.

Progress with cluster identification methodologies

Cluster identification methodologies need to address (1) the cluster definition and (2) a methodology for identifying geographic concentrations.

In relation to the first issue, clusters involve firms that are related through skill, technology, supply, demand or other linkages. These clusters have been defined for single industries or for sets of related industries. Research shows that definitions based on single industries (e.g. three-digit NAISC classification or 3-digit SIC) perform poorly when trying to capture industry interdependencies. This is not surprising as the industry code system groups industries based on product similarity, without accounting for broader inter-industry inderdependencies (Delgado et al. 2016).

Porter (1990) focused on interactions across industries. But his method of grouping industries in a cluster chart was based on "best judgement of the researchers". More robust methodologies have since been proposed by Porter (2003) and Feser (2005). Both methods focus on input and output links as a means to capturing inter-industry relatedness. More recently, a more comprehensive methodology was proposed by members of the US Cluster Mapping Project (Delgado et al., 2016). Applying inter-industry measures of co-location patterns of employment and the number of establishments, input-output links and labour occupation links, the methodlogy groups 6-digit NAISC industries into 51 U.S. Benchmark Cluster Definitions. These Benchmark definitions are used in the latest *European Cluster Panorama* (Ketels and Protsiv, 2016).

Alcacer and Zhao (2016) make the case for a focus on firms' technical capabilities or knowlegde stocks to capture relatedness. Their proposed methodology employs patent data and may have particular application to research on technology clusters. However, apart from the problem with availability of patent data, if the goal is to capture multiple complementarities across industies, the U.S. Benchmark Cluster Definitions may be more useful.

In relation to the second issue, identifying discrete geograhic clusters requires a measure for geographic concentration and a determination of the appropriate geograpic unit. The literature presents a range of specialisaiton measures, including the location quotient (LQ), the local Gini coefficient (Krugman, 1991) and the localization coefficient of Ellison and Glaeser (1997). For an overview, see Prothero (2012). Industries with a high level of specialisation are often deemed to be geographically concentrated. The question remains, how high a LQ measure is required for one to consider it to be indicative of clustering. In addition, geographical concentration and regional specialisation should not be conflated. A high LQ does not necessarilly point to a substantial number of firms or workers. One partial way around this problem is to apply a minimum threshold value of employees (Lazzeretti et al., 2008; Ellison and Glaeser, 1997; Delgado et al., 2012).

In relation to the determination of the geographic unit, there are two established approaches. One involves using predetermined administrative or statistical units such as counties, Economic Areas and NUTS regions. This is common practice, driven by the obvious advantage of data availability. However, as industry concentrations may incorporate parts of different administrative or statistical units, restricting the analysis to such units is problematic. Applying units based on travel-to-work data (De Propris, 2005) only partially resolves this issue.

The other approach involves organically generating geographic units based on the industry data under investigation. Feser et al. (2005) make an important methodological contriution in this regard. Using employment data at county level, they employ a local *G* statistic to identify regions where counties posting comparatively high levels of employment for a given value chain are clustered together. The value of *G* for a county is based on value chain employment of both the county itself as well as neighbouting counties. In this way, *G* detects concentrations of cluster activity across county boundaries.

Other organic methods have been proposed by Alcacer (2016) and Van Egeraat et al. (2018). Both methods employ XY coordinates of economic activity. The cluster identification algorithms take full advantage of the information from the data and the output reflects the real spatial form and extent of the clusters. The application of these organic measures is limited by the availability of point data.

Conclusion

This review tracks the progress made in methodologies seeking to identify clusters. The core aspects of a selection of the reviewed approaches are summarised in Table 1. Although important progress has been made, even the most state-of-the-art methodologies have their drawbacks. The choice of cluster definition and method for identifying geographical concentrations will depend on the policy question, geographical context, type of industries under investigation and, importantly, data availability.

A combinaiton of Delgado's et al. (2016) cluster definition method and Feber's et al. (2005) method for identifying geographical concentrations would usefully inform many policy contexts. The Delgado's Benchmark Cluster Definitions are capable of detecting multiple complementarities across industries. Idealy the method should be applied to other countries using their specific data. However, the existing 51 cluster definitions are a good strating point, especially for economies that lack the required data. These definitions will be especially useful for countries with an industry code schema similar to the one in the USA. It can also be applied with higher level of aggregation to many other countries through matching the U.S. NAISC code to the U.N. ISIC coding system. The Benchmark definitions have allready been employed in the latest *European Cluster Panorama* (Ketels and Protsiv, 2016) to identify clusters in EU NUTS 2 Regions.

Where possible, research should employ organic approaches to identifying geographic units. The approach developed by Feser et al. (2005) is directly transferable to other international contexts and relatively feasible to apply given current data availability in most countries. Other organic approaches reviewed may be preferable, but their feasibility will depend on the availability of point data while, the methodology proposed by Van Egeraat et al. (2018) may not be directly transferable to other contexts with different urban settlement and sectoral structures.

Finally, it is important to point out that the methodologies reviewed here measure concentrations of nominally linked industries. However, the identified concentrations are only suggestive of relations between actors. The existence of a concentration does not guarantee that beneficial cluster advantages and processes are in operation. Whether individual concentrations should be a target for industrial policy or whether such processes could be stimulated always requires more detailed case-study investigation.

Table 1: Selected cluster identification methods

Study	Context and Data	Cluster Definition	Methodology for identifying geographic concentrations
Porter (1990)	Case studies of OECD countries; UN SITC industry data amongst others.	Groupings of UN SITC industries organised in cluster charts on the basis of best judgement of researchers.	Not applicable.
Feser et al. (2005)	U.S all industries; county-level industry employment data (SIC) from Bureau of Labour Statistics.	Groupings of SIC industries based on input-output analysis.	Local *G* statistic which measures specialisation in groups of neighbouring counties. (Organic approach).
Delgado et al. (2016)	U.S. all industries; Employment data (6-digit NAICS), input-output data and labour occupation data at Economic Area level.	Groupings of 6-digit NAICS industries based on inter-industry measures of co-location patterns of employment and the number of establishments, input-output links and labour occupation links.	Based on employment specialisation and share. (Predetermined statistical units).
Alcacer and Zhao (2016)	Global semiconductor industry; USPTO patent data.	Single industry.	Based on densities of patents in given locations using latitude and longitude data. (Organic approach).
Van Egeraat et al. (2018)	Republic of Ireland; Data on firms supported by enterprise agencies. 2-digit NACE.	Individual 2-digit NACE industries.	Using point-level data and travel-to work data to construct industry-specific discrete geographical units. (Organic approach).

Sources

Alcacer, J. and Zhao, M. (2016) Zooming in: a practical manual for identifying geographic clusters. Strategic Management Journal 37: 10-21.

Delgado, M., Porter, M. and Stern, S. (2016) Defining clusters of related industries. *Journal of Economic Geography* 16:1-38.

Delgado, M. Porter, M. and Stern, S. (2012) Clusters, convergence, and economic performance, *NBER Working Paper* 18250.

De Propris, L. (2005) Mapping local production systems in the UK: Methodolgy and application. Regional Studies 39: 197-211.

Ellison, G. and Glaeser, E. (1997) Geographic concentration in US manufacturing industries: A dartboard approach. Journal of Political Economy 105: 889-927.

Feser, E. (2005) *Benchmark Value Chain Industry Clusters for Applied Regional Research*. Regional Economics Application Laboratory, University of Illinois at Urbana Champaign.

Feser, E., Sweeney, S. and Renski, H. (2005) A descriptive analysis of discrete U.S. industrial complexes. *Journal of Regional Science* 45(2):395-419.

Ketels, C. and Protsiv, S. (2016) *European Cluster Panorama 2016*. Brussels: European Commission.

Krugman, P.(1991) *Geography and Trade*. Leuven University Press/The MIT Press, Leuven/Cambridge, MA.

Lazzeretti, L. Boix, R. and Capone, F. (2008) Do creative industries cluster? Mapping creative local production systems in Italy and Spain. *Industry and Innovation* 15:549-567.

Porter, M. (2003) The economic performance of regions. *Regional Studies* 37:549-578

Porter, M. (1990) *The Competitive Advantage of Nations*. Basingstoke: Parlgrave.

Porter, M. (2008) Clusters and competition: new agendas for companies, governments, and institutions. In Porter M (ed.), *On Competition*. Harvard: Harvard Business Press

Prothero, R. (2012) Industrial Specialisation in Local Areas. London and Newport: Office for National Statistics.

Van Egeraat, C., Morgenroth, E., Kroes, R., Curran, D. and Gleeson, J. (2018) A measure for identifying substantial geographic concentrations. *Papers in Regional Science* 97(2):281-300.

About the author

Chris van Egeraat is Senior Lecturer at the Department of Geography, Maynooth University. He has 25 years experience of researching issues related to clusters and regional development policy in Ireland and internationally and has published widely in both areas. Chris regularly acts as a consultant for national and regional economic development organisations on issues related to industrial development policy. He can be contacted at chris.vanegeraat@mu.ie.