

Hermans, Frans

Article — Published Version

The contribution of statistical network models to the study of clusters and their evolution

Papers in Regional Science

Provided in Cooperation with:

Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale)

Suggested Citation: Hermans, Frans (2021) : The contribution of statistical network models to the study of clusters and their evolution, Papers in Regional Science, ISSN 1435-5957, Wiley, Oxford, Vol. 100, Iss. 2, pp. 379-403,
<https://doi.org/10.1111/pirs.12579> ,
<https://rsaiconnect.onlinelibrary.wiley.com/doi/10.1111/pirs.12579>

This Version is available at:

<https://hdl.handle.net/10419/232508>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

FULL ARTICLE

The contribution of statistical network models to the study of clusters and their evolution

Frans Hermans 

Leibniz Institute of Agricultural Development
in Transition Economies (IAMO), Halle,
Germany

Correspondence

Frans Hermans, Leibniz Institute of
Agricultural Development in Transition
Economies (IAMO), Theodor-Lieser-Strasse
2, 06120 Halle (Saale), Germany.
Email: hermans@iamo.de

Funding information

Bundesministerium für Bildung und
Forschung, Grant/Award Number: 031B0020;
The Role and Functions of Bioclusters in the
Transition to a Bioeconomy"

Abstract

This paper presents a systemic review of the contributions that stochastic actor-oriented models (SAOMs) and exponential random graph models (ERGMs) have made to the study of industrial clusters and agglomeration processes. Results show that ERGMs and SAOMs are especially popular to study network evolution, proximity dynamics and multiplexity. The paper concludes that although these models have advanced the field by enabling empirical testing of a number of theories, they often operationalize the same theory in completely different ways, making it difficult to draw conclusions that can be generalized beyond the particular case studies on which each paper is based. The paper ends with suggestions of ways to address this problem.

KEYWORDS

agglomeration, networks, Clusters, ERGM, SAOM

JEL CLASSIFICATION

C15; D85; O32; R12

1 | INTRODUCTION

Knowledge and information exchange is considered to be among the key drivers of regional and national economic development (Tödtling, Lehner, & Trippl, 2006). In order to promote such knowledge and information exchange and strengthen regional economic competitiveness and innovation, clusters have become a popular tool (Iammarino and McCann 2006; Porter, 1998). At the same time, interest has grown in the network characteristics

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Papers in Regional Science published by John Wiley & Sons Ltd on behalf of Regional Science Association International



of agglomerations and clusters and how these characteristics change over time (Glückler, 2007; Powell, White, Koput, & Owen-Smith, 2005). For instance, the effects of different types of proximity in restricted geographical areas has been a subject of intense interest in the regional science and business management literature (Boschma, 2005; Knoben & Oerlemans, 2006).

In general it can be said that the network characteristics of an industrial cluster are of great importance for both the performance of individual organizations within the cluster, as well as the cluster's overall functioning (Belussi & Sammarra, 2010; Breschi & Malerba, 2005; Karlsson, Johansson, & Stough, 2005). In their day-to-day operations, cluster actors engage in a number of different activities such as exchanging financial and material resources, generating and spreading information and collaborating (or competing) with other cluster actors. All these activities are essentially relational in nature: they require the creation and maintenance of a connection between two or more actors. Given the relational nature of many cluster activities, social network analysis has been a popular approach to the study of different aspects of cluster functioning (Bergman, 2009; Ter Wal & Boschma, 2009). In recent years new statistical network models have been developed that use the characteristics of the complete network as the basis for their analysis. Stochastic actor-oriented models (SAOMs) and exponential random graph models (ERGMs) are statistical inference models that are among the most popular and theoretically well-developed network models currently in use (Robins, Snijders, Wang, Handcock, & Pattison, 2007; Snijders, 2011). These statistical network models were deemed promising tools to investigate the drivers of economic agglomeration processes within regional studies (Broekel, Balland, Burger, & Van Oort, 2014; Maggioni & Uberti, 2011). A number of researchers have taken up these new tools and applied them to study different aspects of industrial clusters. However, critics of this network modelling approach in economic geography speak of "diminishing returns" because of the repetitive nature of studies that "take the same model and the same methodology to different databases" (Ferru & Rallet, 2016, p.112).

In this paper, I investigate this claim by assessing the different applications of ERGMs and SAOMs to see what kind of insights they have yielded and whether these statistical network models have fulfilled their promise. The aim of this review is to catalogue the contributions of SAOMs and ERGMs for the study of clusters and economic agglomerations and determine what types of questions have been answered using these new models and what lessons can be drawn from these applications. To answer these questions, I performed a systematic review of papers in which either an SAOM or an ERGM was applied on industrial clusters and agglomeration processes. I analysed how different theoretical concepts were operationalized in the models. This paper provides a methodological contribution by giving an overview of the different ways a number of theoretical concepts related to cluster and agglomeration studies have been operationalized and measured.

2 | COMPARISON OF STOCHASTIC ACTOR ORIENTED MODELS AND EXPONENTIAL RANDOM GRAPH MODELS IN CLUSTERS

In this section the general characteristics of ERGMs and SAOMs are described and the different types of micro-level configurations that they work with are introduced. Detailed comparisons between these methods can be found in Block, Stadtfeld, and Snijders (2019) and Broekel et al. (2014), with the latter also describing potential applications of both these methods within the field of economic geography.

A network can be depicted as a set of nodes connected by a set of ties. The nodes can represent different types of actors or organizations. Network ties can be directed, undirected and/or weighted to denote a range of relationships, such as similarities in positions and memberships, social interactions, and flows of money, information or resources (Borgatti, Mehra, Brass, & Labianca, 2009; Wasserman & Faust, 1994). Special cases are formed by multiplex networks, two-mode networks and multilevel networks, see Figure 1. ERGMs and SAOMs can be used to model any of these types of networks.

Within SAOMs, discrete choice models, continuous-time Markov chains and agent-based modelling are combined. Nodes representing actors can decide to create, maintain or dissolve links. They are assigned a rate




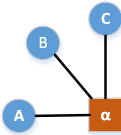
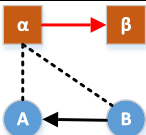
	Description and example
 <i>Multiplex network</i>	Two nodes (A and B) share more than one type of relationship ties (black and red arrows). For example, two managers may have a business relationship and friendship at the same time.
 <i>Two-mode network</i>	Two different types of nodes are combined in one network. These types are usually referred to as actors (A, B & C) and events (α). Typical two-mode networks include inventor-patent networks, citation networks (author–paper), and membership networks related to projects or executive boards.
 <i>Multilevel network</i>	Multilevel networks are a form of two-mode networks that allow different types of nodes and different types of relationships within the same network. For instance by separating the level of individual employees (A and B) from the firm level (α and β), the links between the individuals (black arrow), the links between the firms (red arrow) and the links between the individuals and the firms (dashed lines) can be investigated.

FIGURE 1 Overview of network types
Source: Adapted from Lusher, Koskinen, and Robins (2013)

function that determines how quickly they can change their connections and an objective function that determines what they value in making certain connections. Actors try to maximize this objective function. In SAOMs, the nodes are able to control their outgoing ties, therefore SAOMs are necessarily used for directed networks.¹ In contrast, ERGMs can be used for either directed or undirected networks as they are defined on the level of the complete network and ties or arrows are placed in such a way that they follow a number of pre-defined network characteristics (density, distribution and so on).

SAOMs have been used mostly for the analysis of temporal network evolution, whereas ERGMs have been used primarily for cross-case comparisons. However, as Block et al. (2019) point out: this distinction is slowly becoming obsolete as new types of ERGMs (longitudinal, separable temporal and temporal ERGMs) allow for the incorporation a temporal elements and SAOMs could, at least in theory, also be used for cross-case comparisons (Snijders & Steglich, 2015). In the following meta-analysis, these new types of temporal ERGMs have been excluded as they have not yet been applied on a sufficient number of industrial cluster studies to give a reliable overview.

The main commonality between SAOMs and ERGMs, which is the main reason that this paper includes them both, is the way in which they deal with local network dependencies between certain nodes. ERGMs and SAOMs have the advantage over regular regression models in that they explicitly account for a number of dependencies

¹SAOMs have also been used to model undirected ties, though this requires an additional assumption in the form of a mechanism through which outgoing ties are reciprocated, this is called the *unilateral initiative and reciprocal confirmation model* (Van De Bunt & Groenewegen, 2007). A tie is created if alter confirms the proposal made by ego; therefore, the decision depends on the expected utility of ego first and alter second.



between nodes that have been found to play important roles in social networks (Snijders, 2011). These network dependencies violate the assumption of “independency of observations” upon which regular logistic regression models rely: the existence of a particular tie within the network often predicts the existence of certain other ties within the network.

The historical development of SAOMs and ERGMs can be viewed as increasingly sophisticated attempts to deal with these local dependencies. The latest advancement in SAOMs and ERGMs is represented in social circuit dependence models (Snijders, Pattison, Robins, & Handcock, 2006) in which the statistical dependence between ties is expressed in terms of local configurations: small connected substructures such as triangles or stars. Only network ties that do not share local configurations are assumed to be conditionally independent (Block et al., 2019; Harris, 2014).

These local configurations thus form the basic building blocks of both types of statistical network models and this makes it possible to compare the set-up of each type of models with the other. Lusher et al. (2013) provide a conceptual framework for processes of tie formation that categorizes these micro-level configurations according to three levels: the individual node, the dyad and the complete network. The first level is that of the individual node. At this level covariates are introduced to model the effects of certain characteristics of the individual nodes. For instance, certain attributes (age, work experience, or corporate size) can be used to explain the difference in the activity (outdegree) or popularity (indegree). At the second level, the dyadic level, the traits of both the sender and the receiver are considered and this is the level where the effects of homophily (“birds of a feather flock together”) and its inverse heterophily (“opposites attract”) are studied (McPherson, Smith-Lovin, & Cook, 2001). The third level of the framework contains the endogenous structural effects that are inherent to the internal process of the system of network ties. Lusher et al. (2013) refer to these effects as processes of network self-organization because they do not involve actor attributes or other exogenous influences to explain the formation of these local patterns. Figure 2 gives examples of these micro-level configurations of self-organization that include reciprocity (tendency to return an arrow), triadic closure (formation of different triangular patterns of ties), and degree differentials (centralization of the network) and their higher order effects. To account for more complex dependencies in observed networks, Snijders et al. (2006) introduced three higher order terms: alternating 2-paths, alternating k-triangles and alternating k-stars. Instead of these higher order effects being set to zero, they are constrained by means of a weighted sum with alternating signs. The inclusion of these indicators is often necessary to avoid model degeneracy and help to model the underlying network processes accurately, improving the model's goodness-of-fit. Later, Hunter and Handcock (2006) reworked these three parameters into a form that was easier to interpret. Geometrically weighted dyadwise shared partnerships (GWDSP) is a parametric summary measure of the tendency for two nodes in a network to have one or more partners in common, regardless of whether these two actors themselves are tied. Geometrically weighted edgewise partnerships (GWESP) is essentially the same measure for those actors that are, in fact, directly tied. For both parameters, the weights decrease geometrically as the number of shared partners increases (Robins et al., 2007).

3 | METHODOLOGY

The literature search was conducted in two steps and was based on the specific authors that have been involved in the development of ERGMs and SAOMs. At the moment, there are three popular software packages to perform either SAOM or ERGM based modelling. These packages are SIENA and RSIENA (Ripley, Snijders, & Prediado Lopez, 2011), the statnet package in R (Handcock, Hunter, Butts, Goodreau, & Morris, 2008) and the PNet program and its derivatives (Wang, Robins, & Pattison, 2009). The mathematical and statistical principles that these packages depend upon have been published in various peer-reviewed papers. The search parameters assumed that papers applying either ERGMs or SAOMs on clusters would cite at least one paper that the authors of these packages have written. Table 1 presents an overview of the authors that were included in the search.


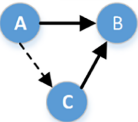
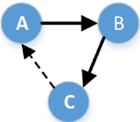
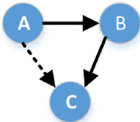


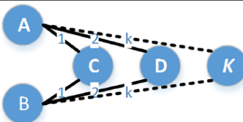
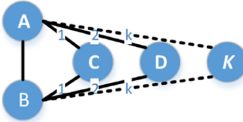
Effect name	Models the effects that the dashed lines or arrows will form, given the existence of the black ties or arrows
Reciprocity	
Triadic closure	<div></div> <div><i>Transitivity</i> <i>Cyclicity</i> <i>Balance</i></div>
Activity	 <i>k-outstar</i>
Popularity	 <i>k-instar</i>
Dyadwise shared partnerships, or k-2paths	
Edgewise shared partnerships, or k-triangle	

FIGURE 2 Overview of structural endogenous network configurations used in this paper
Source: adapted from Lusher et al. (2013)

TABLE 1 Overview of ERGM and SAOM software packages and their main developers

Software package	Names
SIENA and RSIENA	Tom A.B. Snijders; Christian E.G. Steglich
PNET, XPNET, MPNET	Philippa E. Pattison; Garry L. Robins, Peng Wang
Statnet	Steven M. Goodreau, Mark S. Handcock; David R. Hunter

Source: own elaboration.



A search in the Scopus database was performed on t27 August 2018. All papers from these authors, published since the year 2000 were identified, yielding a total of 379 papers. These 379 papers were cited in 12,613 studies. A similar procedure (performed on t23 August 2018) was used in Web of Science. Here, the initial search led to 11,500 studies that cited at least one paper by the authors listed in Table 1.

These initial sets of over 10,000 studies each covered a wide range of topics, not all of which were relevant for the study of clusters and agglomeration processes. In a multistep procedure these sets were further refined. The publication year was limited to 2010 or later. This was the year that SAOMs and ERGMs were first applied in the field of economic geography, as indicated by Maggioni and Uberti (2011). The results were limited to peer-reviewed papers, thereby excluding conference papers. The sets were then further refined by the use of keywords that are often used to describe different kinds of clusters and economic agglomerations: industrial district, technology parks, clusters, and regions. For an overview of the search terms and subject areas, see the Appendix.

This refinement procedure resulted in a list of 951 publications in Scopus and 388 publications in Web of Science. These publications were manually scanned for title and keyword relevance. Papers on topics or from fields not directly related to geography, knowledge exchange or innovation were manually removed. Bibliometric papers on citation networks were also removed. This remaining list of 66 papers was inspected in detail. Of this set, 19 papers were removed that cited one of the authors listed in Table 1 but did not apply an ERGM or SAOM. Another six papers were removed because they focused on topics other than clusters or industrial agglomerations (such as policy or environmental networks), and five papers dealing with intra-organizational networks were also removed. One paper and two book chapters were removed to avoid double counting as they described cases and approaches that were (almost) identical to other papers included in the final set. Finally, one paper could not be retrieved despite several emails to the authors. The 31 studies included in the final list were read and coded with respect to their research questions, method of data gathering, type of cluster analysed, setup and the operationalization of the statistical network model according to their endogenous structural effects, and the covariates and dyadic covariates employed in the model.

4 | RESULTS

4.1 | Overview of the application of statistical network papers

Table 2 provides an overview of the 31 papers selected. Since 2010, 17 papers have presented applications of SAOMs and 14 have presented applications of ERGMs. Many of these papers appeared in one of the top journals specializing in economic geography, business research, or innovation: the *Journal of Economic Geography* (4 papers), *Industry and Innovation* (3 papers), *Regional Studies* (2 papers), *Journal of Business Research* (2 papers), and *Organizational Studies* (2 papers). Table 3 shows the most frequently cited papers in the final set, further illustrating the general interest in these kinds of topics and the results of the applications of SAOMs and ERGMs.

Figure 3 shows the development over time of the most popular topics covered in these papers: network dynamics (15 papers), proximity (18 papers), and multiplexity (7 papers). These categories are closely linked. For instance, papers on network dynamics often apply Boschma's proximity framework (2005) to explain the underlying social mechanisms of network dynamics. Table 3 and Figure 3 show that the often-made empirical distinction between the application of ERGMs and SAOMs (cross case comparison versus longitudinal studies) can only be demonstrated to a certain extent. Although SAOMs are used to investigate questions of longitudinal changes in clusters, the papers that applied ERGMs do not use them to compare different clusters to one another, with the exception of one paper by Brennecke and Rank (2016). Instead, ERGMs are used to study single clusters at one point in time, or to study multiplexity: the effect that different types of ties can have within a cluster.



TABLE 2 Overview of papers included in the systematic review and the characteristics of their model set up

Authors (year)	Model (software)	Industry/cluster	Topics	Nodes
Ebbers and Wijnberg (2010)	SAOM (SIENA)	Film industry in the Netherlands	Reputation; network dynamics	Individuals
Pina-Stranger and Lazega (2011)	ERGM (PNET)	Biotech industry in France	Multiplexity	Individuals
Balland (2012)	SAOM (SIENA)	Global Navigation Satellite Systems (in Europe)	Proximity; network dynamics	Firms, research and non-profit organizations
Ferriani, Fonti, and Corrado (2012)	SAOM (SIENA)	Multi-media cluster in Italy	Multiplexity; network dynamics	Firms
Lomi and Pallotti (2012)	ERGM (unknown)	Public and private hospitals in Lazio, Italy (cluster)	Proximity	Hospitals
Balland, De Vaan, and Boschma (2013)	SAOM (SIENA)	Global video game industry	Proximity; network dynamics	Firms
Broekel and Hartog (2013b)	ERGM (statnet)	Netherlands Aerospace Industry	Proximity	Firms, research and non-profit organizations
De Stefano and Zaccarin (2013)	ERGMGM (XPNET)	Author-Innovator community in the area of Trieste, Italy (multiple industries)	Multiplexity; proximity	Individuals
Giuliani (2013)	SAOM (SIENA)	Colchagua Valley, Wine cluster in Chile	Network dynamics	Firms
Ter Wal (2013b)	SAOM (SIENA)	Biotech industry in Germany	Proximity	Individuals
Broekel and Hartog (2013a)	ERGM (statnet)	German chemical industry	Proximity	Regions
Hazir and Autant-Bernard (2014)	ERGM (PNET)	Life sciences, biotech, and biochemical industry in Europe	Proximity	Regions
Nicotra, Romano, and Del Giudice (2014)	SAOM (SIENA)	Ceramics and tile cluster in Sassuolo, Italy	Network dynamics	Firms
Rank (2014)	ERGM (PNET/ IPNET)	Biotech cluster in Germany	Network dynamics; firm survival	Firms and research
Molina-Morales, Belso-Martinez, Mas-Verdú, and Martínez-Cháfer (2015)	ERGM (unknown)	Xíxona chocolate and confectionary cluster, Spain	Proximity	Firms
Jiang, Gao, Chen, and Roco (2015)	ERGM (XPNET)	Nanotechnology in USA	Multiplexity; subsidies; diffusion	Individuals
Buchmann and Pyka (2015)	SAOM (RSIENA)	German automotive industry	Network dynamics; proximity; subsidies	Firms

(Continues)



TABLE 2 (Continued)

Authors (year)	Model (software)	Industry/cluster	Topics	Nodes
Balland, Belso-Martinez, and Morrison (2016)	SAOM (RSIENA)	Toy Valley, Valencia - Spain	Multiplexity; proximity	Firms
Lazzeretti and Capone (2016)	SAOM (RSIENA)	High Technology Cultural Goods cluster Tuscany, Italy	Proximity	Firms and research
Brennecke and Rank (2016)	ERGM	Metrology cluster and Photonics cluster in Germany.	Multiplexity	Multilevel: Individuals, firms and research
Pinheiro, Seródio, Pinho, and Lucas (2016)	ERGM	Research and industry related to the biological sciences in EU	Multiplexity; social capital	Firms, research and intermediaries
Belso-Martínez, Expósito-Langa, and Tomás-Miquel (2016)	ERGM (unknown)	Food stuffs cluster in Xixona, Spain	Network dynamics; absorptive capacity	Firms
Cao et al. (2017)	SAOM (RSIENA)	Building industry in Shanghai, China	Network dynamics	Individuals
Menzel, Feldman, and Broekel (2017)	SAOM (RSIENA)	ICT cluster in North Carolina, USA	Network dynamics; proximity	Multilevel: Individuals, firms and research
Juhász and Lengyel (2017)	SAOM (RSIENA)	Printing and paper product cluster in Hungary	Proximity	Firms
Roesler and Broekel (2017)	SAOM (SIENA)	Biotech industry in Germany	Network dynamics; subsidies; proximity	Firms, research and miscellaneous
Capone and Lazzeretti (2018)	ERGGM (statnet)	High Tech to Cultural Goods cluster in Tuscany, Italy	Proximity; multiplexity	Firms
Giuliani, Balland, and Matta (2018)	SAOM (RSIENA)	Consumer electronics cluster in Cordoba, Argentina	Network dynamics; proximity	Firms
Bauer, Hansen, and Hellsmark (2018)	SAOM (RSIENA)	Biorefinery industry in Sweden	Network dynamics	Firms, research and non-profit organizations
Schierjott, Brennecke, and Rank (2018)	ERGM (PNET)	Biotech cluster in Germany	Attitudes	Individuals
Tang, Wang, Li, and Cao (2018)	SAOM (RSIENA)	Chinese construction industry	Network dynamics; proximity	Firms

Notes:
^aMultiplex network with reciprocity in "other" network tie;
^bExcluded from final model to aid convergence; "sqrt" square root.



TABLE 2 (Continued)

Authors (year)	Ties	Type of ties	Reci-procity (Yes/No)	Transitivity	Degree distribution g
Ebbers and Wijnberg (2010)	Collaboration on same film	undirected	No	transitive triads	sqrt (degree of alter)
Pina-Stranger and Lazega (2011)	1) Knowledge 2) Friendship	directed	Yes	k-transitive triads; k-cyclic triads; k-2paths	mixed k-star
Balland (2012)	Collaborative projects	undirected	No	k-2paths	
Feriani, Fonti, and Corrado (2012)	1) Advice received (and indirectly given) 2) Economic supply relationship	directed	Yes	Transitive and balanced triplets	
Lomi and Pallotti (2012)	Flow of patients	directed	Yes	k-triads, k-2paths	mixed 2-star, k-instar, k-outstar
Balland, De Vaan, and Boschma (2013)	Collaboration on video game	undirected	No	Transitive triplets	sqrt (degree of alter) ^b
Broekel and Hartog (2013b)	Exchange of technical knowledge	undirected	No	GWESP and GWDSP ^b	gwdegree
De Stefano and Zaccarin (2013)	Collaboration on 1) patent; 2) paper writing	multiplex (undirected)	Yes ^a	triangles and k- triangles	2, 3 & k-star
Giuliani (2013)	Advice given and received	directed	Yes	transitive triplets	outdegree
Ter Wal (2013b)	Collaboration and knowledge creation	undirected	No	triadic closure	
Broekel and Hartog (2013a)	Collaboration	undirected	No	GWESP & GWDSP	gwdegree
Hazir and Autant-Bernard (2014)	Collaborative project network between regions	undirected	No	Triangles	2stars
Nicotra, Romano, and Del Giudice (2014)	Collaboration (patents)	directed	No	Transitivity	Sqrt (indegree popularity)

(Continues)



TABLE 2 (Continued)

Authors (year)	Ties	Type of ties	Reci-procity (Yes/No)	Transitivity	Degree distribution g
Rank (2014)	Exchange of task related advice	undirected	No	triadic closure	
Molina-Morales, Belso-Martínez, Mas-Verdú, and Martínez-Cháfer (2015)	Knowledge received	directed	Yes	triadic cycles; GWESP for cohesion	gwindegree; gwoutdegree
Jiang, Gao, Chen, and Roco (2015)	1) collaboration on patents 2) patent citation (knowledge transfer)	multiplex (directed and undirected)	Yes ^a	Transitivity in diff. Network ties	
Buchmann and Pyka (2015)	Collaboration	undirected	No	transitive triplets	
Balland, Belso-Martínez, and Morrison (2016)	1) Technical advice and 2) Business info received	directed	Yes	triadic closure and cyclicity	indegree popularity, outdegree activity
Lazzeretti and Capone (2016)	Shared project	undirected	No	1/nr of actors at geodesic distance 2	degree of alter
Brennecke and Rank (2016)	1) Collaborative network 2) Knowledge exchange network	multiplex (directed and undirected)	Yes	k-transitive triads; k cyclical triads	in-star and out2star; k-instars, k-outstars;
Pinheiro, Seródio, Pinho, and Lucas (2016)	Resource sharing, prior ties, trust and commitment	multiplex (directed and undirected)	Yes	cyclical triads; GWESP & GWDSP	gwoutdegree; gwindegree
Belso-Martínez, Expósito-Langa, and Tomás-Miquel (2016)	Advice given or received	directed	Yes	GWESP	
Cao et al. (2017)	Collaboration on same project	multiplex (directed and undirected)	No		indegree popularity
Menzel, Feldman, and Broekel (2017)	Patents	undirected		Transitivity; k-2stars; k-2paths	degree of alters & same degree of alters (assortativity)

(Continues)



TABLE 2 (Continued)

Authors (year)	Ties	Type of ties	Reci-procity (Yes/No)	Transitivity	Degree distribution g
Juhász and Lengyel (2017)	Advice/Knowledge transfer network	directed	Yes	Transitivity, cyclicity; triadic closure	
Roesler and Broekel (2017)	Collaborative projects	undirected			sqrt (degree of alters)
Capone and Lazzeretti (2018)	1) collaboration; 2) advice received; 3) friendship	directed		1/(nr. of actors at distance 2)	
Giuliani, Balland, and Matta (2018)	Advice given or received	directed	Yes	transitive triplets	
Bauer, Hansen, and Hellsmark (2018)	Shared projects	undirected		GWESP	sqrt (degree of alters)
Schierjott, Brennecke, and Rank (2018)	Received knowledge	directed	Yes	k-transitive & k-cyclical triads	k-outstar; k-instar
Tang, Wang, Li, and Cao (2018)	Shared projects	directed			sqrt (degree of indegree)

Source: own elaboration.

Notes:

^aMultiplex network with reciprocity in "other" network tie;

^bExcluded from final model to aid convergence; "sqrt" square root.



TABLE 3 Overview of most-cited papers in the final set

Year	Author(s)	Journal	Total citations	Citations per year
2012	Balland, P.A.	Regional Studies	170	21.25
2013	Balland, P.A., De Vaan, M., Boschma, R.	Journal of Economic Geography	95	13.57
2013	Ter Wall, A.L.J.	Journal of Economic Geography	76	10.86
2013	Giuliani, E.	Research Policy	86	12.29
2016	Balland, P.A., Belso-Martínez, J. A., Morrison, A.	Economic Geography	46	11.5
2013	Ferriani, S., Fonti, F., Corrado, R.	Strategic Organization	45	11.25
2012	Lomi, A., Pallotti, F.	Social Networks	44	5.5
2016	Lazzeretti, L., Capone, F.	Journal of Business Research	41	10.25
2015,	Molina-Morales, X., Belso-Martínez, J.A., Más-Verdú, F., Martínez-Cháfer, L.	Journal of Business Research	30	6
2013	Broekel, T., Hartog M.	Industry & Innovation	22	3,14

Source: Thompson Reuters' Web of Science [accessed July 2020].

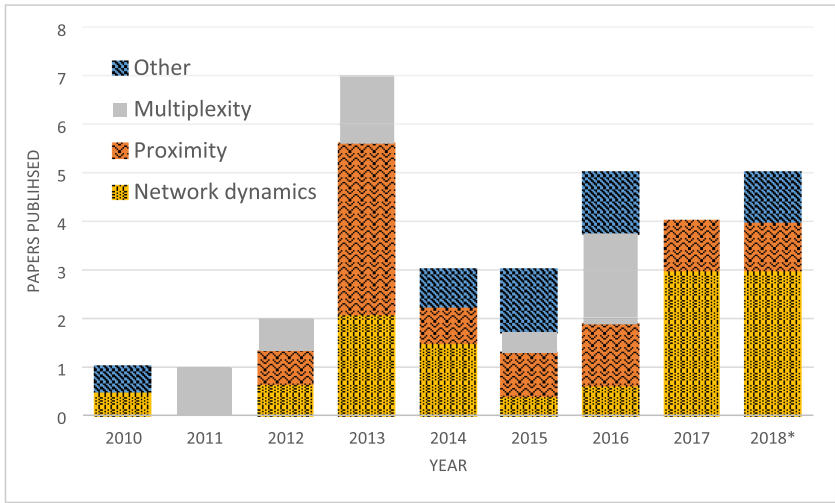


FIGURE 3 Overview of papers and main topics per year

Note: * First 8 months of 2018.

Source: Scopus and Thompson Reuters' Web of Science, accessed August 2018

4.2 | Characterization of cluster and networks

4.2.1 | Types of clusters and industries

A distinction can be made between the papers that study actual clusters (20 papers) and those that incorporate the effects of some form of economic agglomeration within a specific industrial sector (11 papers). The vast majority of the cluster studies were done on cases that are located within Europe (16 of 20). Locations of the studies not located



in Europe include the United States (1), China (1), Chile (1) and Argentina (1). The papers that focus on specific industries base their analyses at different geographical levels. For instance the video game industry is considered at the global level, European navigation and satellites systems, are analysed at the multi-national level of the European Union and the sectoral systems of innovation of the Dutch Aerospace sector, and the German biotech sector are considered at the national level.

4.2.2 | Network size

The sizes of the networks studied varies significantly, from the smallest network of 26 nodes (Juhász & Lengyel, 2017) ranging up to networks of approximately 600 to 700 nodes (Ter Wal, 2013b) and 1,575 nodes (Tang et al., 2018). The difference in size is related to the kind of data used in the different studies. Studies that make use of large databases of patents, scientific papers, or EU projects, end up with significantly larger networks than studies for which data were gathered by hand using online questionnaires or face-to-face interviews. This latter kind of primary data gathering is mostly associated with cluster studies, while secondary data gathering is mostly used for industrial sector agglomeration studies. As a result papers dealing with various forms of clusters are typically much smaller (average size—70 nodes) compared to industrial sector studies (average size—458 nodes).

4.2.3 | Nodes and composition

The most popular definition of a cluster comes from Porter (1990) who defines a cluster as “a geographically proximate group of interconnected companies and associated organizations (for example, universities, standards agencies, and trade associations) in a particular field, linked by commonalities and complementarities.” This definition explicitly includes many different types of organizations within a cluster. However, much of the cluster research within our set actually describes networks between a single type of actor: firms (12 papers), individuals (7 papers) or hospitals (1 paper). Only a minority of papers takes a broader view: two papers look at two types (firms and research), while five include the different links between three types of organizations, such as firms, research organizations and other organizations such as trade organizations or intermediaries. Most of the network studies thus take only a subsample of the networks in a cluster by limiting their focus to a single type of organization.

4.2.4 | Types of ties and relationships

The main type of relationship within the retrieved studies includes various types of knowledge exchange (which was one of the selection criteria for inclusion). However, how these knowledge networks are constructed differs from study to study with varying focus on knowledge creation, knowledge diffusion and knowledge exchange. Some studies that focus on knowledge creation view multi-actor patents as examples of successful collaboration and knowledge creation. For a second type of knowledge network, the emphasis is not on knowledge creation but on the diffusion of knowledge within the network through the directed flows of information: technical or business advice given or received, or a combination of these. Finally, there are the collaborative networks. These networks imply an exchange of information but, depending on the type of collaborative project, might involve both the exchange of technical information and the creation of new knowledge.

Although some studies also gathered data regarding the strength of particular ties, none of the studies ultimately uses this information in its model as the software available at the time was only suitable for dealing with binary graphs.



4.3 | Overview of model specifications

In this subsection I present how the different models have been operationalized according to the three different levels of nodes, dyads and networks, at which level the simple and higher-order structural effects are considered. ERGM and SAOM studies typically combine variables at these three levels in their models to test different hypotheses.

4.3.1 | Node level

There are three types of variables at the node level: categorical, continuous and network indicators of multiplexity. Categorical variables reflect different types or groups that need to be controlled for in the model, for instance different types of organizations. The second type of variables, quantitative variables, mostly contain simple statistics: firm age, experience, size, etc. However, in three studies, more elaborate status indicators are calculated. Balland (2012) constructs an indicator for the absorptive capacity of firms based on an aggregate of the scores of the OECD classification for sectoral R&D-intensity and the number of employees of a firm. Giuliani (2013) constructs a variable for the knowledge base of wineries based on a principal component analysis of indicators for the level of education of workers, their work experience in the wine industry and the experimentation intensity within the winery. This approach is also applied by Belso-Martínez et al. (2016).

Finally, there are the studies on multiplexity that use a network indicator for one network type as predictor for the nodes in the other network. For example, Jiang et al. (2015) construct two networks, one about patent co-authorships and one with patent citation information for the same set of nodes. At the node level, statistics from the citation network like indegree (continuous) and the existence of a structural holes (categorical) are used to set the nodal variables for the co-authorship model.

4.3.2 | Dyadic level

At the dyad level, the presented results are limited to the topic of proximity. See Table 4. At this level the proximity-related hypotheses are formulated based on the process of homophily. In their review of the different forms of proximity, Knoen and Oerlemans (2006) found that many forms of proximity can be defined both on the dyad and the network level. Due to the nature of SAOMs and ERGMs, most authors define the different forms of proximity on the dyad level as a form of homophily. The only exception is social proximity which some authors have chosen to define social proximity as a form of transitivity, which is a variable of network self-organization.

Table 4 presents those studies that operationalize at least two of the five types of proximity identified in Boschma's framework: (i) geographical proximity (similarity of physical location or physical distance between organizations); (ii) organizational proximity (similarities between, or sharing of organizational structures, for example, subsidiaries); (iii) institutional proximity (similarity of formal rules and regulations); and (iv) cognitive proximity (similarity of field or industry), and social proximity (similarity of social environment, e.g., friendship). The last column shows additional types of homophily that various authors have used in addition.

Geographical proximity is operationalized in two different ways. The first measures the physical distance between two nodes. Sometimes a mathematical operation is performed on this indicator: taking the square root or transforming the distance by some other measure that reflects propinquity instead of distance. The second type of operationalization uses a categorical variable of co-location as a measure for geographical proximity: if nodes are co-located in the same administrative unit such as a municipality or region, they are more likely to form ties. The idea of "temporary geographical proximity" (Torre, 2008), which states that knowledge exchange can be facilitated on a temporary basis by actors travelling to different locations, has not yet been taken up.



TABLE 4 Types of homophily to operationalize proximity

Authors (year)	Type of proximity				Other forms of homophily	
	Geographical	Organizational	Cognitive	Social	Institutional	
Balland (2012)	Same region	Same corporate group	Similar field/domain	1/Geodesic distance 2 ^b	Same organizational type	
Lomi and Pallotti (2012) ^a	1) Distance	Same administrative unit <i>Spatial competition:</i> 3) patient pool overlap	1) Same field/domain [secondary or tertiary care]	Same organizational type	Same organizational type	1) Same size [employees]; 2) Same occupancy rate (%); 3) Same comparative Performance Index
	2) Same		2) Index based on available clinical specialties for complementarity			
			3) Case Mix Index CMI for complexity of organization (absorptive capacity)			
Balland et al. (2013)	Distance	Same corporate group	Shared genres index	Project history [collaborations in last 5 yrs]	Same country	Profile similarity [publisher or developer]
Broekel and Hartog (2013b)	Distance		Similar field/domain	Shared history at same firm [y/n]	Same organizational type	1) Dedicated to aviation [y/n]; 2) Openness: use of external knowledge vs. internal knowledge
Broekel and Hartog (2013a)	1) Distance, 2) Same region				Same region	both have university

(Continues)



TABLE 4 (Continued)

Authors (year)	Type of proximity			Other forms of homophily		
	Geographical	Organizational	Cognitive	Social	Institutional	
Hazir and Autant-Bernard (2014)	Contiguity [y/n]		Tech. distance: 1) Difference in biotech patenting 2) Difference in R&D human resources; Absorptive capacity: 3) Sum of biotech patenting, and 4) Sum of R&D human resources		Same country	
Buchmann and Pyka (2015)	Distance		Tech. distance: [diff. in vector of patent classifications]			
Molina-Morales et al. (2015)	Distance	Same corporate group	Same NACE code	Shared family history [y/n]	Same organizational type	
Balland et al. (2016)	Distance	Same corporate group	Nr. of digits similarity in NACE codes	Shared family history [y/n]	Same organizational type	Perceived similarity
Lazzeretti and Capone (2016)	Same region	Years of collaboration between a pair;	Similar field/domain	1/Geodesic distance 2 ^b	Same organizational type	1) Funds similarity; 2) Leadership similarity
Juhász and Lengyel (2017)	Distance		Nr. of digits similarity in NACE codes			
Menzel et al. (2017)		Same corporate group	Cosine of two inventors patent vectors (1 digit patent class)			

(Continues)



TABLE 4 (Continued)

Authors (year)	Type of proximity				Other forms of homophily
	Geographical	Organizational	Cognitive	Social	
Roesler and Broekel (2017)	1) Distance	Same corporate group	Nr. of co-occurrence in research area codes in last 10 yrs	Shared project history [project in last 10 years]	Same organizational type
	2) Same region				
Capone and Lazzeretti (2018)	Same region		Similar field/domain;	1/Geodesic distance 2 ^b	Same organizational type
Giuliani et al. (2018)	Distance (in km)	Membership of local business board [y/n]			Industrial/sectoral proximity = same sub-sector
Tang et al. (2018) ^a	Same region			Same organizational type [state owned, non-state owned]	Size/scale similarity

Source: own elaboration.

Notes: Bold entries are continuous variables;

^aThese papers do not use an explicit proximity framework. Their homophily terms have been categorized by the author into different proximity categories;

^bTransitivity and the number of actors at distance 2 are not dyadic variables. In this paper, they are categorized on the level of network self-organization. These indicators have been included for completeness sake.



Organizational proximity refers to rules and procedures that organizations share. When organizations share the same parent organization or belong to the same corporate group, they share the same framework and their degree of autonomy within these organizational arrangements is low. Table 3 shows that six out of eight studies that consider organizational proximity use a categorical indicator to evaluate whether two organizations belong to the same type of group of firms.

Cognitive proximity depends on similarities in the technological field, domain or industrial sector an organization is associated with. The category of cognitive proximity therefore also includes studies that refer to cognitive proximity as technical proximity. If organizations are active within the same field or domain, they are considered to have high cognitive proximity. Continuous variables measure the overlap in between fields or domains that two organizations share. Indices based on the Statistical Classification of Economic Activities in the European Community, commonly referred to as NACE codes or patent vectors are used to indicate firms in the same economic sub-sector. A minimum level of absorptive capacity is thought to be an important element of cognitive proximity (Cohen & Levinthal, 1990), and two studies include separate indices for absorptive capacity to define cognitive proximity.

Institutional proximity reflects the similarity of formal and informal rules that organizations encounter. Organizations that fall under the same type of regulatory framework are considered to enjoy a closer proximity and therefore can more easily cooperate. Two different operationalizations of institutional proximity were found. Some authors argue that different types of organizations (research, commercial or governmental) have different organizational aims and incentive structures and therefore differ in their institutional environments as well. This type of institutional proximity looks both at the formal and informal “rules of the game”. Other authors only look at the formal regulatory framework (laws) and argue that within the same administrative unit (country or region), the rules are basically the same for all types of organizations. In this case, the operationalization of institutional proximity is basically the same measure as the co-location variable that is sometimes used to operationalize geographical proximity.

Social proximity has two types of operationalizations within ERGMs and SAOMs. The first group of authors define social proximity based on a shared history between two nodes. The nature of this shared history can differ: a personal relationship through family or friendship ties or a professional working relationship such as a shared former employer or the number of previous projects that have been done together. For the second type of social proximity, the social environment is defined based on the position of the node within the social network. Thus, some indicators for social proximity are not defined as processes of homophily, but are defined as belonging to the class of endogenous, structural indicators, like transitivity, that are also used to model network self-organization.

4.3.3 | Network level: structural effects

The last three columns of Table 2 contain an overview of the endogenous structural effects like reciprocity, triadic closure, degree differentials and their higher order configurations. A practical disadvantage of including higher order effects like GWESP, GWDSP and alternating triangles, is that they make running the computer simulations very time consuming (several hours up to days). Some studies therefore ultimately drop the higher order parameters in order to facilitate model convergence.

A general problem with these endogenous structural parameters is that the theoretical underpinnings of various endogenous network developments in clusters are still largely unknown (see Abbasiharofteh, 2020 for an exception). Network evolution is therefore often viewed as a mix of endogenous and exogenous effects, but differentiating between these two remains difficult. There are two exceptions. The first is the negative density coefficient that is included in all the models that indicates that establishing or maintaining a connection costs effort or energy, and thus actors do not connect with everyone, but make conscious decisions that lead to sparse networks. The second exception is the set of indicators for transitivity, discussed earlier, that have been used as operationalizations of social proximity. Here triadic closure is not only viewed as an endogenous structural effect, but also has a theoretical



meaning that has been used by a number of different papers. However, the theoretical implications for most other endogenous structural effects are more or less formulated on an *ad hoc* basis. In most other cases, these indicators are used only as control variables.

This problem is further complicated by the fact that it is sometimes unclear how these endogenous network effects influence the other properties of the network. For instance, Juhász and Lengyel (2017) suggest that endogenous network effects and proximity effects are not independent from each other in network evolution because of homophily effects that leads to the establishment of cohesive groups of similar individuals. They conclude that triadic closure and cognitive proximity counteract each other because firms look for new varieties of knowledge within the cluster.

Preferential attachment is another well-known structural network effect that many authors include in their models (using the *gwdegree* or alternating *k-stars*). However as Giuliani (2013 p. 1417) writes: “It seems that, so far, scholars have confused preferential attachment with the strengthening of one or a few prominent actors that have become outstanding not through the number of their ties, but based on their agency, their resources and skills, and their capacity to be leading firms.” Balland et al. (2016) make a distinction between network status as a degree-related concept and industrial status that is derived from the reputation, expertise and visibility a firm builds over time. Their results show that the type of knowledge network is an important factor, because highly skewed degree distributions are the result of both industrial and network status for business advice, while these same two variables do not play a significant role in the degree distribution of technical advice networks.

5 | DISCUSSION AND RECOMMENDATIONS

This paper a systematic review of two statistical network tools that have enjoyed growing popularity in recent years: ERGMs and SAOMs. The results show some of the versatility of these new statistical tools in operationalizing and empirically testing some of the theoretical concepts related to the network evolution paradigm in economic geography (Glückler, 2007; Powell et al., 2005). As could be expected when reviewing the application of a statistical tool, the limits to the application of ERGMs and SAOMs depend to a large extent, on the creativity of the individual researcher and his or her particular interests. However, based on this review it is also clear that there is often a lack of reflection about the effects that choices in research design may have on the outcomes of the study. In this section I make recommendations to overcome this problem.

5.1 | The need for reflection on internal and external cluster characteristics

The main characteristics of the clusters and industries covered in the studies and how these characteristics will likely affect the selection of relevant network types within the cluster or industry is often missing, or not discussed systematically. For instance, there is often no clear justification for the choice to consider cluster networks composed of only a single type of organizations (firms) versus networks that also include research institutes and/or intermediate organizations. There is a similar lack of clarity around the selection of type of knowledge network: formal, tacit or collaborative. Advice networks are organized hierarchically, but collaborative networks could potentially also show some cyclical knowledge flows. What complicates the analysis is that the research on multiplexity shows that knowledge can follow complementary routes in addition to overlapping ones (Maghssudipour, Lazzeretti, & Capone, 2020).

Likewise, the studied clusters include different types of industries, ranging from the relatively low-tech industries of toys, chocolates, and wine, to the more high-tech clusters like ICT, photonics and biotech. All these characteristics are likely to influence the outcomes of the studies and more reflection is needed about the various internal cluster characteristics that are relevant for issues of cluster evolution, proximity and multiplexity.



In addition, more consideration should be given to the regional characteristics of the innovation system in which the cluster is embedded (Hermans, 2018). SAOMs could potentially incorporate more external influences on knowledge evolution, for instance with regards to the regional characteristics relevant to a particular cluster. This would also be in line with Trippl, Grillitsch, Isaksen, and Sinosis (2015) and Isaksen and Trippl (2014), who distinguish different types of regional innovation systems based on their organizational thickness. They argue that these different types of regional innovation systems also support different forms of cluster development pathways. Giuliani et al. (2018) have begun to work in this direction with a study on network evolution in a peripheral cluster, but more case studies are still needed.

5.2 | Cross-case comparisons

The theories that are tested most frequently in the selected studies, are those related to different forms of proximity. Even critics of the network approach to proximity have acknowledged that the empirical operationalization of the proximity framework has been a major contribution to economic geography (Ferru & Rallet, 2016). Statistical network models have contributed to providing the operationalizations and subsequent measurements of the theoretical concepts behind the different proximity categories. Even though all studies included in this review refer to the same proximity theory developed by Boschma (2005), the operationalization of the framework within the five different categories shows some inconsistencies across studies. The variable co-location is used to measure geographical proximity by some authors and institutional proximity by others. Similarly, the variable measuring the years of collaboration between a pair of nodes has been used both for organizational proximity and social proximity. These differences in operationalization might make sense within the context of the individual studies; however, such differences make it difficult to extent their conclusions beyond the single case study investigated. Many of the ambiguities related to the application of different proximity categories that Knoben and Oerlemans (2006) identified, are still alive and kicking.

In order for the ERGMs and SAOMs to deliver on their promise to help build and test a common framework for proximity research, investigations must move beyond single case studies. One explanation for the lack of cross-case comparisons is no doubt the elaborate nature of data collection and the need for very high response rates that makes manual data collection in clusters both time and resource intensive. As a result, the potential of these tools, and especially ERGMs, to support cross-case comparison of different networks between different types of clusters has not yet been effectively explored. One of the main advantages of ERGMs is their ability to compare networks of different sizes, something that is notoriously difficult to do with traditional social network analysis tools (Anderson, Butts, & Carley, 1999). However, with the exception of Brennecke and Rank (2016) and Pinheiro et al. (2016) all the papers that use ERGMs deal with only a single cluster. An open access repository for network data that accompany these papers would be very beneficial for the field as a whole.

5.3 | Distinction between slow and fast variables in network evolution studies

The second important topic studied, especially with SAOMs, is related to cluster evolution. These studies have so far ignored questions on the frequency or the time scale of network changes and the choice of time frames is often based on the availability of data. Papers on cluster evolution present this as a steady process of change that can be explained by dynamics within the cluster or the dominant industrial sector it depends on, and no distinction is made between the influence of various fast and slow processes (with the notable exception of Menzel et al, 2017, who identify a period before and after the “dot-com” bubble in the ICT industry). Papers that construct long-term networks that span decades based on patent data, often smooth out high fluctuations in the network from year to year by applying a five year moving window to their data (Buchmann & Pyka, 2015; Ter Wal, 2013a). Although this practice can be defended as necessary, given that the SAOM software makes it difficult modelling high frequency changes in networks, this characteristic of SAOMs makes it difficult to identify types of changes other than linear,



more or less incremental ones. The implicit assumption is therefore that cluster dynamics will follow the standard cluster life-cycle pattern of emergence, growth, maturity and eventual decline (Menzel & Fornahl, 2010). However other cluster development pathways, such as those based on a punctuated equilibrium (Rosenkopf & Tushman, 1998), or on the adaptive cycle (Martin & Sunley, 2011) where long phases of relative stability alternate with short phases of major change could be difficult to identify. Consideration for the theoretical implications of cluster phases on network characteristics is growing, but more empirical evidence is needed.

6 | CONCLUSIONS

This paper presents a systematic review of the scientific literature on the applications of ERGMs and SAOMs with a focus on cluster studies and agglomeration processes. Results show that these two methods have been especially popular in economic geography to investigate network dynamics, proximity and the effects of multiplexity and multilevel networks in clusters. Although these models have advanced the field through enabling empirically testing of a number of theories, they are often used to operationalize the same theory in completely different ways, making it difficult to draw conclusions that can be generalized beyond the particular case study on which each paper is based. In order to remedy these problems, a more systematic description of the development of a cluster, the characteristics of its industry and the type of knowledge networks should be offered and these factors should be the subject of more explicit reflection. In addition, more cross-case comparisons, reflection about different types of slow and fast changes occurring in clusters, a broadening the focus of cluster studies beyond knowledge networks and the incorporation of financial or material flows within the networks are needed as well. This might benefit both cluster studies in general and the further development of ERGMs and SAOMs specifically.

ACKNOWLEDGEMENTS

This study was financed through the TRAFORBIT project "The Role and Functions of Bioclusters in the Transition to a Bioeconomy" (031B0020) under the call "Bioökonomie als gesellschaftlicher Wandel" of the German Federal Ministry of Education and Research (BMBF). The author wants to thank Milad Abbasiharofteh, Tom Broekel and an anonymous reviewer for their useful comments on a draft version of this paper.

ORCID

Frans Hermans  <https://orcid.org/0000-0003-3394-9012>

REFERENCES

- Abbasiharofteh, M. (2020). Endogenous effects and cluster transition: A conceptual framework for cluster policy. *European Planning Studies*, 30, 1–24. <https://doi.org/10.1080/09654313.2020.1724266>
- Anderson, B. S., Butts, C. T., & Carley, K. (1999). The interaction of size and density with graph-level indices. *Social Networks*, 21, 239–267. [https://doi.org/10.1016/S0378-8733\(99\)00011-8](https://doi.org/10.1016/S0378-8733(99)00011-8)
- Balland, P.-A. (2012). Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies*, 46, 741–756.
- Balland, P.-A., Belso-Martinez, J. A., & Morrison, A. (2016). The dynamics of technical and business knowledge networks in industrial clusters: Embeddedness, status, or proximity? *Economic Geography*, 92, 35–60.
- Balland, P.-A., De Vaan, M., & Boschma, R. (2013). The dynamics of interfirm networks along the industry life cycle: The case of the global video game industry, 1987–2007. *Journal of Economic Geography*, 13, 741–765.
- Bauer, F., Hansen, T., & Hellsmark, H. (2018). Innovation in the bioeconomy: Dynamics of biorefinery innovation networks. *Technology Analysis & Strategic Management*, 30, 935–947.
- Belso-Martinez, J.-A., Expósito-Langa, M., & Tomás-Miquel, J.-V. (2016). Knowledge network dynamics in clusters: Past performance and absorptive capacity. *Baltic Journal of Management*, 11, 310–327. <https://doi.org/10.1108/BJM-02-2015-0044>



- Belussi, F., & Sammarra, A. (2010). *Business networks in clusters and industrial districts: The governance of the global value chain*. London: Routledge.
- Bergman, E. M. (2009). Embedding network analysis in spatial studies of innovation. *The Annals of Regional Science*, 43, 559–565.
- Block, P., Stadtfeld, C., & Snijders, T. A. B. (2019). Forms of dependence: Comparing SAOMs and ERGMs from basic principles. *Sociological Methods & Research*, 48(1), 202–239.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323, 892–895.
- Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39, 61–74.
- Brennecke, J., & Rank, O. N. (2016). Multilevel network analysis for the social sciences: Theory, methods and applications. In E. Lazega & T. A. B. Snijders (Eds.), *Multilevel network analysis for the social sciences* (pp. 273–293). Cham; Heidelberg; New York; Dordrecht; London: Springer.
- Breschi, S., & Malerba, F. (2005). *Clusters, networks and innovation*. Oxford: Oxford University Press.
- Broekel, T., Balland, P.-A., Burger, M., & Van Oort, F. (2014). Modeling knowledge networks in economic geography: A discussion of four methods. *The Annals of Regional Science*, 53, 423–452.
- Broekel, T., & Hartog, M. (2013a). Determinants of cross-regional R&D collaboration networks: An application of exponential random graph models. In T. Scherngell (Ed.), *The geography of networks and R&D collaborations* (pp. 49–70). Cham: Springer.
- Broekel, T., & Hartog, M. (2013b). Explaining the structure of inter-organizational networks using exponential random graph models. *Industry and Innovation*, 20, 277–295.
- Buchmann, T., & Pyka, A. (2015). The evolution of innovation networks: The case of a publicly funded German automotive network. *Economics of Innovation and New Technology*, 24, 114–139.
- Cao, D., Li, H., Wang, G., Luo, X., Yang, X., & Tan, D. (2017). Dynamics of project-based collaborative networks for BIM implementation: Analysis based on stochastic actor-oriented models. *Journal of Management in Engineering*, 33, 04016055.
- Capone, F., & Lazzeretti, L. (2018). The different roles of proximity in multiple informal network relationships: Evidence from the cluster of high technology applied to cultural goods in Tuscany. *Industry and Innovation*, 25(9), 1–21.
- Cohen, W., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128–152.
- De Stefano, D., & Zaccarin, S. (2013). Modelling multiple interactions in science and technology networks. *Industry and Innovation*, 20, 221–240. <https://doi.org/10.1080/13662716.2013.791130>
- Ebbers, J. J., & Wijnberg, N. M. (2010). Disentangling the effects of reputation and network position on the evolution of alliance networks. *Strategic Organization*, 8, 255–275.
- Elsevier's Scopus database. Available online: <https://www.scopus.com/> [accessed August 2020].
- Ferriani, S., Fonti, F., & Corrado, R. (2012). The social and economic bases of network multiplexity: Exploring the emergence of multiplex ties. *Strategic Organization*, 11, 7–34.
- Ferru, M., & Rallet, A. (2016). Proximity dynamics and the geography of innovation: Diminishing returns or renewal? In R. Shearmu, C. Carrincazeaux, & D. Doloreux (Eds.), *Handbook on the geographies of innovation* (pp. 100–122). Cheltenham: Edward Elgar.
- Giuliani, E. (2013). Network dynamics in regional clusters: Evidence from Chile. *Research Policy*, 42, 1406–1419.
- Giuliani, E., Balland, P.-A., & Matta, A. (2018). Straining but not thriving: Understanding network dynamics in underperforming industrial clusters. *Journal of Economic Geography*, 30, 147–172.
- Glückler, J. (2007). Economic geography and the evolution of networks. *Journal of Economic Geography*, 7, 619–634.
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24, 1–11.
- Harris, J. K. (2014) *An introduction to exponential random graph modeling (quantitative applications in the social sciences)*. Los Angeles CA, Sage Publications
- Hazir, C. S., & Autant-Bernard, C. (2014). Determinants of cross-regional R&D collaboration: Some empirical evidence from Europe in biotechnology. *The Annals of Regional Science*, 53, 369–393.
- Hermans, F. (2018). The potential contribution of transition theory to the analysis of bioclusters and their role in the transition to a bioeconomy. *Biofuels, Bioproducts and Biorefining*, 12, 265–276.
- Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15, 565–583.
- Iammarino, S., & Mccann, P. (2006). The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. *Research Policy*, 35, 1018–1036. <https://doi.org/10.1016/j.respol.2006.05.004>
- Izaksen, A., & Trippl, M. (2014). *Regional industrial path development in different regional innovation systems: A conceptual analysis*. *Papers in innovation studies*. Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE). Lund, Sweden: Lund University.



- Jiang, S., Gao, Q., Chen, H., & Roco, M. C. (2015). The roles of sharing, transfer, and public funding in nanotechnology knowledge-diffusion networks. *Journal of the Association for Information Science and Technology*, 66, 1017–1029.
- Juhász, S., & Lengyel, B. (2017). Creation and persistence of ties in cluster knowledge networks. *Journal of Economic Geography*, 121, 1203–1226.
- Karlsson, C., Johansson, B., & Stough, R. R. (2005). *Industrial clusters and inter-firm networks (new horizons in regional science)*. Cheltenham: Edward Elgar.
- Knoben, J., & Oerlemans, L. A. G. (2006). Proximity and inter-organizational collaboration. *International Journal of Management Reviews*, 8, 71–89.
- Lazzeretti, L., & Capone, F. (2016). How proximity matters in innovation networks dynamics along the cluster evolution. A study of the high technology applied to cultural goods. *Journal of Business Research*, 69, 5855–5865.
- Lomi, A., & Pallotti, F. (2012). Relational collaboration among spatial multipoint competitors. *Social Networks*, 34, 101–111.
- Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods and applications*. New York: Cambridge University Press.
- Maggioni, M. A., & Uberti, T. E. (2011). Networks and geography in the economics of knowledge flows. *Quality and Quantity*, 45, 1053–1057.
- Maghssudipour, A., Lazzeretti, L., & Capone, F. (2020). The role of multiple ties in knowledge networks: Complementarity in the Montefalco wine cluster. *Industrial Marketing Management*, 90, 667–678. <https://doi.org/10.1016/j.indmarman.2020.03.021>
- Martin, R., & Sunley, P. (2011). Conceptualizing cluster evolution: Beyond the life cycle model? *Regional Studies*, 45, 1299–1318.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, 415–444.
- Menzel, M.-P., Feldman, M. P., & Broekel, T. (2017). Institutional change and network evolution: Explorative and exploitative tie formations of co-inventors during the dot-com bubble in the Research Triangle region. *Regional Studies*, 51, 1179–1191.
- Menzel, M.-P., & Fornahl, D. (2010). Cluster life cycles—dimensions and rationales of cluster evolution. *Industrial and Corporate Change*, 19, 205–238.
- Molina-Morales, X. F., Belso-Martinez, J. A., Mas-Verdú, F., & Martínez-Cháfer, L. (2015). Formation and dissolution of inter-firm linkages in lengthy and stable networks in clusters. *Journal of Business Research*, 68, 1557–1562.
- Nicotra, M., Romano, M., & Del Giudice, M. (2014). The evolution dynamic of a cluster knowledge network: The role of firms' absorptive capacity. *Journal of the Knowledge Economy*, 5, 240–264.
- Pina-Stranger, A., & Lazega, E. (2011). Bringing personalized ties back in: Their added value for biotech entrepreneurs and venture capitalists interorganizational networks. *The Sociological Quarterly*, 52, 268–292.
- Pinheiro, M. L., Seródio, P., Pinho, J. C., & Lucas, C. (2016). The role of social capital towards resource sharing in collaborative R&D projects: Evidences from the 7th Framework Programme. *International Journal of Project Management*, 34, 1519–1536.
- Porter, M. E. (1990). *The competitive advantage of nations*. New York: The Free Press.
- Porter, M. E. (1998). *Clusters and the new economics of competition*, No. 6. Boston, MA: Harvard Business School Press.
- Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *The American Journal of Sociology*, 110, 1132–1205.
- Rank, O. N. (2014). The effect of structural embeddedness on start-up survival: A case study in the German biotech industry. *Journal of Small Business & Entrepreneurship*, 27, 275–299.
- Ripley, R. M., Snijders, T. A. B., Prediado Lopez, P. (2011) Manual for RSiena. University of Oxford, Department of Statistics, Nuffield College,
- Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29, 192–215.
- Roesler, C., & Broekel, T. (2017). The role of universities in a network of subsidized R&D collaboration: The case of the biotechnology-industry in Germany. *Review of Regional Research*, 37, 135–160.
- Rosenkopf, L., & Tushman, M. (1998). The co-evolution of community networks and technology: Lessons from the flight simulation industry. *Industrial and Corporate Change*, 7, 311–346.
- Schierjott, I., Brennecke, J., & Rank, O. N. (2018). Entrepreneurial attitudes as drivers of managers' boundary-spanning knowledge ties in the context of high-tech clusters. *Journal of Small Business Management*, 56, 108–131.
- Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, 37, 131–153.
- Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36, 99–153.
- Snijders, T. A. B., & Steglich, C. E. G. (2015). Representing micro-macro linkages by actor-based dynamic network models. *Sociological Methods & Research*, 44, 222–271. <https://doi.org/10.1177/0049124113494573>



- Tang, Y., Wang, G., Li, H., & Cao, D. (2018). Dynamics of collaborative networks between contractors and subcontractors in the construction industry: Evidence from national quality award projects in China. *Journal of Construction Engineering and Management*, 144, 05018009.
- Ter Wal, A. L. J. (2013a). Cluster emergence and network evolution: A longitudinal analysis of the inventor network in Sophia-Antipolis. *Regional Studies*, 47, 651–668.
- Ter Wal, A. L. J. (2013b). The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. *Journal of Economic Geography*, 14, 589–620.
- Ter Wal, A. L. J., & Boschma, R. A. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, 43, 739–756.
- Thompson Reuters' Web of Science. <https://apps.webofknowledge.com/> [accessed July 2020].
- Tödtling, F., Lehner, P., & Tripl, M. (2006). Innovation in knowledge intensive industries: The nature and geography of knowledge links. *European Planning Studies*, 14, 1035–1058.
- Torre, A. (2008). On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies*, 42, 869–889.
- Tripl, M., Grillitsch, M., Izaksen, A., & Sinosis, T. (2015). Perspectives on cluster evolution: Critical review and future research issues. *European Planning Studies*, 23, 2028–2044. <https://doi.org/10.1080/09654313.2014.999450>
- Van De Bunt, G. G., & Groenewegen, P. (2007). An actor-oriented dynamic network approach: The case of inter-organizational network evolution. *Organizational Research Methods*, 10, 463–482.
- Wang, P., Robins, G., Pattison, P. (2009) PNet: Program for the simulation and estimation of exponential random graph models. University of Melbourne, Melbourne School of Psychological Sciences, Melbourne
- Wasserman, S., & Faust, K. (1994). *Social network analysis (structural analysis in the social sciences)*. Cambridge: Cambridge University Press.

How to cite this article: Hermans F. The contribution of statistical network models to the study of clusters and their evolution. *Pap Reg Sci.* 2021;100:379–403. <https://doi.org/10.1111/pirs.12579>

APPENDIX

Search terms used

Web of Science (23th of August 2018)

In Web of Science 10 different searches were combined. The name “Peng Wang” yielded too many results, so for this author the three papers included in Web of Science that he (co-) authored were included (searches 8, 9 & 10).

Indexes used for all searches:

SCI-EXPANDED, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED, IC Timespan = All years

1. **CITED AUTHOR:** (SNIJDERS T OR SNIJDERS T A OR SNIJDERS T A B OR SNIJDERS TAB OR SNIJDERS TOM OR SNIJDERS TOM A B OR SNIJDERS TOM AB) **AND CITED YEAR:** (2000–2018) -- > [3,733 results]
2. **CITED AUTHOR:** (STEGLICH C OR STEGLICH C E G OR STEGLICH CEG OR STEGLICH CHRISTIAN OR STEGLICH CHRISTIAN E G) **AND CITED YEAR:** (2000–2018) [2,531 results].
3. **CITED AUTHOR:** (PATTISON P E OR PATTISON PE OR PATTISON PHILIPPA OR PATTISON PHILIPPA E OR PATTISON PHILIPPA ELEANOR OR PATTISON PHILLIPA OR PATTISON PIP) **AND CITED YEAR:** (2000–2018) [1,885 results]
4. **CITED AUTHOR:** (ROBINS GARRY OR ROBINS GARRY L OR ROBINS G L) **AND CITED YEAR:** (2000–2018) [1,086 results].
5. **CITED AUTHOR:** (GOODREAU S M OR GOODREAU SM OR GOODREAU STEVEN OR GOODREAU STEVEN M OR GOODREAU STEVEN MICHAEL) **AND CITED YEAR:** (2000–2018) [1,536 results]



6. **CITED AUTHOR:** (HANDCOCK MARK OR HANDCOCK MARK S OR HANDCOCK M S OR HANDCOCK MS) **AND CITED YEAR:** (2000–2018) [2,368 results].
7. **CITED AUTHOR:** (HUNTER D R OR HUNTER D ROBERT OR HUNTER DAVID R OR HUNTER DR) **AND CITED YEAR:** (2000–2018) [2,053 results]
8. **CITED DOI:** (10.5751/ES-06880-190,423) **AND CITED YEAR:** (2000–2018) [22 results].
9. **CITED DOI:** (10.1016/j.jmp.2013.05.004) **AND CITED YEAR:** (2000–2018) [12 results]
10. **CITED DOI:** (10.1371/journal.pone.0142181) **AND CITED YEAR:** (2000–2018) [2 results]
11. #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1 [11,500 results]
12. (#10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1) AND TS = (cluster* OR district* OR industr* OR agglom* OR region* NOT clustering) AND PY = (2010–2019)

Refined by: WEB OF SCIENCE CATEGORIES: (MULTIDISCIPLINARY SCIENCES OR GREEN SUSTAINABLE SCIENCE TECHNOLOGY OR GEOSCIENCES MULTIDISCIPLINARY OR MANAGEMENT OR BIOTECHNOLOGY APPLIED MICROBIOLOGY OR INFORMATION SCIENCE LIBRARY SCIENCE OR SOCIAL SCIENCES BIOMEDICAL OR SOCIAL SCIENCES INTERDISCIPLINARY OR GEOGRAPHY PHYSICAL OR BUSINESS OR FOOD SCIENCE TECHNOLOGY OR ECONOMICS OR TRANSPORTATION OR MATHEMATICS INTERDISCIPLINARY APPLICATIONS OR SOCIAL SCIENCES MATHEMATICAL METHODS OR REMOTE SENSING OR SOCIOLOGY OR OPERATIONS RESEARCH MANAGEMENT SCIENCE) AND [excluding] **DOCUMENT TYPES:** (PROCEEDINGS PAPER) **Timespan:** All years. **Indexes:** SCI-EXPANDED, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED, IC.

Scopus (27th of August 2018)

- Step 1: (AU-ID("Wang, Peng" 55,790,445,600) OR AU-ID("Snijders, Tom A.B." 7,004,565,512) OR AU-ID("Steglich, Christian" 24,367,743,300) OR AU-ID("Steglich, Christian" 57,200,919,132) OR AU-ID("Pattison, Philippa Eleanor" 7,005,357,222) OR AU-ID("PATTISON, PHILIPPA E." 56,431,403,100) OR AU-ID("Robins, Garry L." 7,006,457,420) OR AU-ID("Goodreau, Steven Michael" 8,207,010,100) OR AU-ID("Handcock, Mark Stephen" 6,603,858,417) OR AU-ID("Hunter, David R." 7,402,537,048)) AND (PUBYEAR AFT 1999)
- Step 2: Refined search: ((cluster* OR industr* OR region OR region* OR district* OR agglom*)) AND (proximity OR collaboration OR multiplexity OR cooperation OR innovation) AND (LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010)) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (SUBJAREA, "SOCI") OR LIMIT-TO (SUBJAREA, "BUSI") OR LIMIT-TO (SUBJAREA, "DECI") OR LIMIT-TO (SUBJAREA, "ECON") OR LIMIT-TO (SUBJAREA, "MULT") OR LIMIT-TO (SUBJAREA, "ENER"))



Resumen. Este artículo presenta un examen sistémico de las contribuciones de los modelos estocásticos orientados a los actores (SAOM, en inglés) y los modelos de gráficos aleatorizados exponenciales (ERGM, en inglés) al estudio de los conglomerados industriales y los procesos de conglomeración. Los resultados muestran que los ERGM y los SAOM son especialmente populares para estudiar la evolución de las redes, las dinámicas de proximidad y la multiplexidad. En el artículo se llega a la conclusión de que, si bien esos modelos han hecho avanzar este campo de estudio al permitir realizar estudios empíricos de varias teorías, a menudo ponen en práctica la misma teoría mediante maneras completamente diferentes, lo que hace difícil sacar conclusiones que puedan generalizarse más allá de los estudios de caso particulares en que se basa cada artículo. El artículo concluye con sugerencias de formas de abordar este problema.

抄録: 本稿では、確率的アクター指向モデル (Stochastic Actor-Oriented Model: SAOM)と指数ランダムグラフモデル (Exponential Random Graph Model: ERGM)による産業クラスターと集積の過程の研究への寄与の系統的レビューを行う。結果から、ERGMとSOAMは、特にネットワークの進化、近似性の変化、多重性の研究によく用いられていることが示される。本稿では、これらのモデルは、多くの理論の実証的な検証を可能にし、その研究分野を進歩させてきたが、同じ理論を全く異なる方法で扱うことにより、各研究の基となっている特定のケーススタディを超えて一般化できる結論を得ることが困難になっていると結論する。最後にこの問題に対処する方法を提案する。