



The application of exponential random graph models to collaboration networks in biomedical and health sciences: a review

Fei Yu¹ · Helal M. El-Zaatari² · Michael R. Kosorok² · Andrea Carnegie^{3,4} · Gaurav Dave^{3,4}

Received: 6 July 2023 / Revised: 18 November 2023 / Accepted: 13 December 2023

© The Author(s) 2024

Abstract

Collaboration has become crucial in solving scientific problems in biomedical and health sciences. There is a growing interest in applying social network analysis to professional associations aiming to leverage expertise and resources for optimal synergy. As a set of computational and statistical methods for analyzing social networks, exponential random graph models (ERGMs) examine complex collaborative networks due to their uniqueness of allowing for non-independent variables in network modeling. This study took a review approach to collect and analyze ERGM applications in health sciences by following the protocol of a systematic review. We included a total of 30 studies. The bibliometric characteristics revealed significant authors, institutions, countries, funding agencies, and citation impact associated with the publications. In addition, we observed five types of ERGMs for network modeling (standard ERGM and its extensions—Bayesian ERGM, temporal ERGM, separable temporal ERGM, and multilevel ERGM). Most studies (80%) used the standard ERGM, which possesses only endogenous and exogenous variables examining either micro- (individual-based) or macro-level (organization-based) collaborations without exploring how the links between individuals and organizations contribute to the overall network structure. Our findings help researchers (a) understand the extant research landscape of ERGM applications in health sciences, (b) learn to control and predict connection occurrence in a collaborative network, and (c) better design ERGM-applied studies to examine complex relations and social system structure, which is native to professional collaborations.

Keywords Exponential random graph models (ERGMs) · Professional collaborations · Social network analysis · Review · Biomedical and health sciences

1 Introduction

In biomedical and health sciences, a collaborative effort has become crucial in addressing scientific challenges (Ho et al. 2021), improving health service effectiveness (Harris et al. 2012), reducing health disparities (Okamoto 2015), and

producing high-impact scientific publications (Smith et al., 2021). The synergy usually leads to innovation and acceleration of sciences and services by leveraging the perspectives, expertise, and resources of professionals from different disciplines, institutions, and countries (Bennett et al., 2010). Therefore, researchers examine collaboration networks to disclose patterns and characteristics of extant collaborative endeavors, measure the associations between factors and variables that can influence the network structure, and predict the probability of future connections between nodes (e.g., professionals and organizations). Previous studies have widely used network analysis in analyzing and visualizing co-authorship (Yu et al., 2020; 2020) and health policies (Provan et al., 2005; Luke et al. 2013).

Exponential random graph models (ERGMs), also known as p^* models, are a set of computational and statistical methods for analyzing social networks. They are predictive models of network structure, a new and rapidly evolving area of social network analysis. ERGMs assume that the presence or

✉ Fei Yu
feifei@unc.edu

¹ Health Sciences Library, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

² Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

³ North Carolina Translational and Clinical Sciences Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁴ School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

absence of other relationships or individual attributes influences the emergence of a connection/link. For example, “A,” “B,” and “C” are three nodes in a network. If “A” connects with “B” and “B” connects with “C,” the probability of the connection formation between “A” and “C” increases (i.e., transitivity).

Compared with other standard regression methods and inferential tests that researchers have adopted to assess the relationship between measures of variables in a collaborative social network (e.g., linear regressions (Fujimoto et al., 2009; Kesternich and Rank 2022), weighted least squares regression (Provan et al., 2009; Retrum et al. 2013), and ordinary least squares regression (Puro and Kelly 2022)), ERGMs stand out, because (1) practically, ERGMs can handle regression analyses and also simulate similar networks; and (2) theoretically, they are modified logistic regressions that allow for the probability of a connection/link to depend upon the presence or absence of other connections/links in the network (Stanford Human Evolutionary Ecology and Health; Yousefi-Nooraie et al. 2014; van der Pol 2018). ERGMs can model a network relation with both endogenous (dependent) and exogenous (independent) variables. While exogenous variables are attributes of a node and are not related to the network structure, endogenous variables are strictly a function of network modeling and the essence of a standard ERGM. Overall, ERGMs focus on the interaction between the network structures with endogenous variables (e.g., transitivity, reciprocity) and individual attributes of nodes with exogenous variables (e.g., age, gender, experience).

As a relatively new approach, ERGMs are considered “underutilized” for investigating collaboration networks (Harris et al. 2012). Although there are literature surveys of ERGMs on concepts (Drobyshevskiy and Turdakov 2020), new versions (Goodreau 2007; Robins et al. 2007), and statistical schemes and applications (Ghafouri and Khasteh 2020), it is unknown the extent of ERGM application to specific fields (Ghafouri and Khasteh 2020). Mainly, there is a lack of examining applications of ERGMs on collaborations in biomedical and health sciences. Therefore, this study fills in this gap by systematically searching and screening literature, extracting data, and synthesizing research evidence pertinent to the applications of ERGMs to collaboration networks within biomedical and health sciences. Random graph models are essential for building and analyzing complex networks. Thus, our findings help health sciences researchers understand ERGMs and learn to control and predict connection occurrence in a collaboration network (Drobyshevskiy and Turdakov 2020).

The remainder of this paper is structured as follows. First, in Sect. 2, titled “Method,” we describe the rigorous review methodology we adopted, including the literature search, study selection, and data extraction processes. Our primary

findings are presented in Sect. 3, “Results,” which includes the bibliometrics, study design, and network characteristics and features of the included studies. In Sect. 4, “Discussion,” we provide a detailed interpretation of the results, highlighting significant insights and addressing potential study limitations. Finally, Sect. 5, “Conclusion,” summarizes our main discoveries and their broader implications.

2 Methods

Since we aimed to produce research synthesis systematically, transparently, and timely, we chose a review approach. We followed the protocol of a systematic review (SR), a gold standard for knowledge synthesis. We reported results in descriptive summaries or data categorization instead of qualitative summaries or meta-analyses, typical in an SR. In addition, our synthesis process did not involve quality assessment.

2.1 Literature search

We systematically searched for literature in PubMed and Scopus in June 2022. PubMed is the optimal citation database for publications in biomedical and health sciences (Falagas et al. 2008), while Scopus is the largest citation database in the world with a broad journal indexing scope (Schotten et al. 2017). We constructed the search query with identified ERGM name variations (e.g., exponential family random graph models, temporal exponential random graph model, p^* model, etc.) and collaboration network terms (e.g., scientific network, team sciences, co-authorship, etc.). After testing and validating the query in PubMed, we translated it into a query for Scopus. The retrieved results were aggregated and deduplicated in Endnote. In addition, we conducted backward searching by manually reviewing relevant studies’ references and identified additional publications for screening (Fig. 1).

2.2 Study selection

Two reviewers (FY & HE) went through both title–abstract and full-text screening independently. We employed a set of inclusion and exclusion criteria for study selection. The exclusion criteria are (1) not published in the English language; (2) not in the domain of biomedical and health sciences; (3) not an original research article (e.g., commentary, editorial, erratum, abstract only, letter, note, review, etc.); (4) does not apply any ERGMs to the data set; (5) does not study collaboration networks; (6) focuses on ERGM applications in epidemiology and disease spreading (e.g., HIV infection, needle sharing, etc.); (7) focuses on a non-professional relationship like friendship, dating, or sexual relation, etc. The

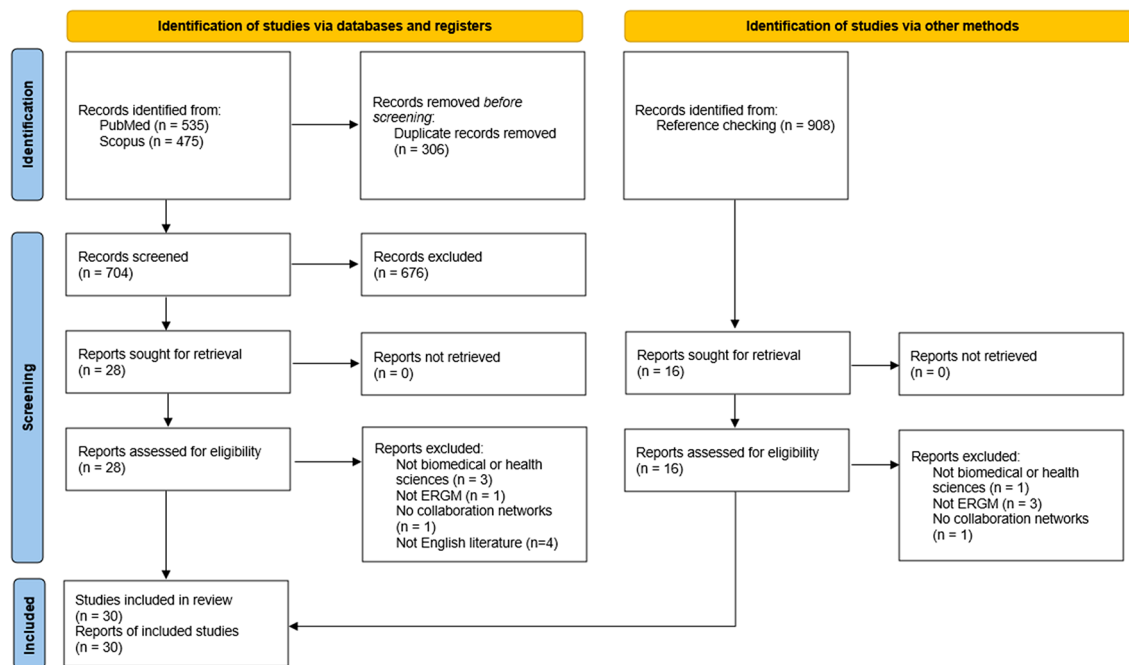


Fig. 1 Flow diagram of literature search and study selection (Page et al. 2021)

inter-rater reliability of the two reviewers was 74% (proportionate agreement) at the title/abstract level and 88% at the full-text level. Two reviewers met and discussed all conflicts until they agreed on the final inclusion.

2.3 Data extraction

Based on the research aim, we developed a data extraction schema (Table 1). Two reviewers (FY & HE) independently extracted data from each included study and recorded them in a Microsoft Excel Spreadsheet. The reviewers resolved discrepancies through meetings, discussions, and clarifications within the research team. The data extraction schema includes three categories: (1) bibliometrics characteristics depict the research landscape over time, significant authors, institutions, study sites, funding agencies, and citation impact of included studies.

If a publication does not explicitly specify the study site, which refers to the country, where the study implementation or the data source and results apply, we selected the first author's country as the study site. We measured citation impact using field-weighted citation impact (FWCI), a time- and field-normalized citation metric that Scopus provides, which is less biased than the traditional citation count (Elsevier; Yu et al. 2020). (2) Study design focuses on ERGM application to study collaboration networks regarding application purposes, ERGM type, and research methods (i.e., data collection and analysis). (3) Network modeling characteristics include specific descriptive network measures (e.g., number of networks, the average number of nodes per network, density, and centrality) and measures unique to ERGMs (e.g., endogenous and exogenous variables). We also included the methods of goodness of fit, which assesses how well ERGMs model networks.

Table 1 Data extraction schema

Category	Data field	Data source
Bibliometrics	author, affiliation, source title, funding source, citation count	Full-text article/Scopus
Study design	ERGM application, ERGM type, node type, data collection/source, data analysis software	Full-text article
Networks	# networks, average # nodes, network direction, # endogenous variables, # exogenous variables, density, centrality, and goodness of fit	Full-text article

3 Results

We include 30 studies that met our inclusion criteria (Fig. 1). The extracted data are presented in Table 2, 3, 4.

3.1 Bibliometrics

The included studies were published between 2010 and 2022, and five were published in 2018. About 80 researchers from 10 countries co-authored the studies. Harris J.K. published six studies, much more than the other authors.

Table 2 Bibliometrics

Study	Source title	Study site	Funding source	FWCI
Fattore 2010	BMC Health Service Research	Italy	N/A	1.09
Luke 2010	American Journal of Public Health	USA	N/A	1.95
Zappa 2011a	Quality and Quantity	Italy	N/A	2.31
Zappa 2011b	Procedia – Social & Behavior Science	Italy	N/A	1.89
Harris 2012	Implementation Science	USA	CDC	1.57
Lomi 2012	Social Networks	Italy/Switzerland	SNSF	4.79
Harris 2013	American Journal of Preventive Medicine	USA	N/A	1.55
Luke 2013	Health Education & Behavior	USA	CDC, NIA, NCI FIC, NIDCR	1.43
Uddin 2013	BMC Health Services Research	Australia	N/A	1.89
Wang 2013	Social Networks	Australia/France	N/A	14.64
Heaney 2014	Social Networks	USA	NIH	9.07
Yousefi-Nooraie 2014	Implementation Science	Canada	PHE	1.81
Shearer 2014	Implementation science	Canada	CIHR, IDRC	1.39
Bevc 2015a	American Journal of Public Health	USA	N/A	1.68
Bevc 2015b	International Journal of Environmental Research & Public Health	USA	N/A	0.58
Harris 2015	PLOS ONE	USA	NIDDKD	0.09
Okamoto 2015	Translational Behavioral Medicine	USA	N/A	0.65
Shoham 2016	Journal of Interprofessional Care	USA	N/A	0.75
Caimo 2017	Statistics in Medicine	Ireland/UK/Switzerland	SNSF	1
Azondekou 2018	AMIA Annual Symposium proceedings	Republic of Benin	N/A	0.13
Broekel 2019	Networks and Spatial Economics	Germany	BMBF, BMWi, BMU	0.89
Bunger 2018	Implementation Science	USA	NIH, NIMH, VA NCATS, QUERI, WUSTL	0.9
Marchand 2018	Alzheimer's and Dementia: Translational Research and Clinical Interventions	USA	NIH, NIGMS	0.49
Mascia 2018	Health Care Management Review	Italy/UK	SNSF, MIUR	0.85
McGlashan 2019	Scientific Reports	Australia	NHMRC, AGDH	0.4
Prusaczyk 2019	Journal of Rural Health	USA	NIMH, ACS	0.45
Prochnow 2020	Journal of Health Organization and Management	USA	N/A	0.37
Ho 2021	Journal of Clinical and Translational Science	USA	NIH, NIDA	No data
Bohnett 2022	Social Networks	USA/Italy	FIBER	No data
Kesternich 2022	Health Care Management Science	Germany	N/A	No data

N/A means the authors did not disclose the information in their acknowledgments

AGDH=Department of Health, Australian Government; BMBF=German Federal Ministry of Education and Research; BMWi=the Federal Ministry of Economics and Technology; BMUV=the Federal Ministry of the Environment; Nature Conservation and Nuclear Safety; CDC=Centers for Disease Control and Prevention (USA); CIHR=Canadian Institutes of Health Research (CIHR); FIBER=Florida Institute of Built Environment Resilience; FIC=Fogarty International Center; IDRC=International Development Research Centre; MIUR=Italian Ministry of Education, Research and University; NCATS=National Center for Advancing Translational Sciences; NCI=National Cancer Institute; NIDA=National Institute on Drug Abuse; NIDCR=National Institute of Dental and Craniofacial Research; NIDDKD=National Institute of Diabetes and Digestive and Kidney Diseases; NIGMS=National Institute of General Medical Sciences; NIH=National Institute of Health; NIMH=National Institute of Mental Health; NIA=National institute on Aging; NHMRC=National Health and Medical Research Council; PHE=Canadian Institutes of Health Research Partnerships for Health System Improvement; QUERI=Quality Enhancement Research Initiative; SNSF=Swiss national Science Foundation; VA=US Department of Veterans Affairs; WUSTL=Washington University in St. Louis

Table 3 Study design

Study	ERGM application	ERGM type	Node type	Data source	Analysis software
Fattore 2010	To examine the co-membership of physicians in the same collaboration initiative	STD	Physician	*	SIENA
Luke 2010	To predict the collaboration tie upon the organization's attribute	STD	Tobacco control organization	Survey	SPSS, R-statnet, Pajek
Zappa 2011a	To model the knowledge-sharing network and test exogenous variables	STD	Physician	Interview survey	N/A
Zappa 2011b	To examine social interaction among physicians for a newly launched drug	STD	Physician	*	SIENA
Harris 2012	To determine the association between the benefits and drawbacks of organizational collaboration and the presence of a collaboration tie between any two members	STD	Research organization	Survey	R-statnet
Lomi 2012	To specify the probability of patient referral network ties between organizations	STD	Hospital	* Survey interview	N/A
Harris 2013	To estimate the probability of a collaboration between health departments	STD	U.S. health department	*	R-statnet
Luke 2013	To examine attribute and structural predictors of collaboration and dissemination among public health organizations	STD	Public health agency	* interview	Pajek, R-statnet
Uddin 2013	To identify micro-level structural properties of a physician collaboration network	STD	Physician/ patient	Insurance claims	Organisation Risk Analyser
Wang 2013	To investigate the properties of a multilevel cancer researcher collaboration network	Mergm	Cancer researcher	*	N/A
Heaney 2014	To understand the influence of reputation on US health policy	STD	Interest group	*	R-statnet
Yousefi-Nooraie 2014	To predict the collaboration formation of information-seeking and expertise-recognition	STD	Public Health Department staff	* survey	SIENA StOCNET
Shearer 2014	To estimate the probability of the tie between two policy actors	STD	Policy actor	Survey interview	NVivo R-statnet
Bevc 2015a	To test for preferential tendencies in inter-organizational networks in the public health sector	STD	Public health organization	*	R-statnet

Table 3 (continued)

Study	ERGM application	ERGM type	Node type	Data source	Analysis software
Bevc 2015b	To examine the partnership between public health organizations	STD	Public health organization	*	R-statnet
Harris 2015	To estimate the probability of a collaborative tie among diabetes translational scientists	STD	Diabetes researcher	Survey	N/A
Okamoto 2015	To model the scientific collaborations of key investigators of ten research centers	STD	Researcher	Survey	R-statnet
Shoham 2016	To examine the patterns of ties in the interprofessional and intra-professional communication network at a burn unit	STD	Hospital professionals	Survey	R-statnet
Caimo 2017	To characterize interhospital patient referral networks	Bayesian	Hospital	*	R-bergm
Azondekon 2018	To test the co-authorship network hypotheses of Malaria researchers in Belin	STD & Tergm	Malaria researcher	Web of Science	Tethne NetworkX R-mixer R-btergm R-igraph
Broekel 2019	To demonstrate how to use Stergm to investigate an R&D collaboration network	Stergm	R&D organization	*	N/A
Bunger 2018	To test hypotheses about advice-seeking ties of physicians	STD	Mental health physician	Survey	N/A
Marchand 2018	To model co-authorship of researchers in a research center	STD	Neuroscience researcher	Citations survey	R-statnet
Mascia 2018	To test hypotheses about advice-seeking ties between pediatricians	STD	Pediatrician	Survey	N/A
McGlashan 2019	To model the structure of community-based collaboration networks for obesity prevention	Mergm	Steering Committee member	Survey	MPNet
Prusaczyk 2019	To understand the referral networks of rural clinics and health centers	STD	Clinic/ health center	*	R-statnet R-igraph
Prochnow 2020	To determine significant factors that may influence collaborative ties between organizations	STD	Health organization	Survey	R-statnet
Ho 2021	To investigate factors that contributed to the formation and persistence of collaboration networks	Stergm	Researcher	Survey	R-tergm
Bohnett 2022	To examine how resource contributions may influence collaborative network	STD	Health organization	*	R-statnet

Table 3 (continued)

Study	ERGM application	ERGM type	Node type	Data source	Analysis software
Kesternich 2022	To explore open and closed triadic substructures in patient-sharing networks among physicians	STD	Physician	Medical claims	N/A

N/A means that the authors did not disclose the information in their publications

STD = Standard ERGM; *Bergm* = Bayesian ERGM; *Tergm* = Temporal ERGM; *Stergm* = Separable temporal ERGM; *Mergm* = Multilevel ERGM

*Secondary data which were collected by the authors previously, other researchers, or government projects.

Table 4 Network characteristics

Study	# Networks	Ave. Nodes	Direction	# Endo. Var	# Exo. Var	Dens	Centr	GOF
Fattore 2010	1	223	U	0	4	N	N	Y
Luke 2010	5	34.6	U	4	6	Y	Y	Y
Zappa 2011a	1	784	U	4	11	Y	Y	Y
Zappa 2011b	2	891	D	2	8	Y	Y	Y
Harris 2012	3	18	U	4	17	Y	Y	Y
Lomi 2012	1	91	D	10	15	Y	Y	Y
Harris 2013	1	2107	D	4	10	Y	Y	Y
Luke 2013	8	18.88	U	2	5	Y	Y	Y
Uddin 2013	85	26.22	U	5	0	Y	Y	Y
Wang 2013	1	97,82	D	8	0	Y	Y	Y
Heaney 2014	4	168	U (<i>n</i> =2) D (<i>n</i> =2)	5	9	Y	Y	Y
Yousefi-Nooraie 2014	2	28	D	3	10	N	Y	Y
Shearer 2014	6	69	D	4	5	Y	Y	Y
Bevc 2015a	162	18	D	1	1	Y	Y	N
Bevc 2015b	162	27.77	D	1	1	N	N	Y
Harris 2015	4	49	U	3	8	Y	Y	Y
Okamoto 2015	2	167	D	4	15	N	Y	Y
Shoham 2016	1	71	U	4	5	Y	Y	Y
Caimo 2017	1	110	D	9	10	Y	Y	Y
Azondekou 2018	1	1792	U	4	6	Y	Y	Y
Broekel 2019	1	910	U	3	12	Y	Y	N
Bunger 2018	2	131	D	7	12	N	N	Y
Marchand 2018	4	168	U	1	2	Y	Y	Y
Mascia 2018	2	48.5	D	9	10	Y	Y	Y
McGlashan 2019	2	19	D	13, 21	2	Y	Y	Y
Prusaczyk 2019	1	69	U	1	2	N	N	N
Prochnow 2020	1	18	U	2	6	Y	Y	N
Ho 2021	3	9.33	D	4	6	Y	Y	Y
Bohnett 2022	42	19.4	U	2	10	N	Y	Y
Kesternich 2022	2	113	U	9	5	Y	Y	Y

U = Undirected; D = Directed; *Endo.* = Endogenous; *Exo.* = Exogenous; *Var.* = Variable; *Dens.* = Density; *Centr.* = Centrality; *GOF* = Goodness of Fit; Y = Yes; N = No.

Luke D.A. and Pallotti F. each published three studies. Luke is Harris' colleague and research collaborator at the Brown School of Social Work at Washington University in St. Louis

(WUSL). At the same time, Pallotti F. was affiliated with the Center for Business Network Analysis at the University of Greenwich in the UK. WUSL, the University of Colorado

Denver, and the University of North Carolina at Chapel Hill were the top three author affiliations. In addition, more than half of the studies were conducted in the United States, followed by Italy (6 studies) and Australia (3). However, studies in European countries involved more international collaborations than US-based studies. For example, five studies had multiple study sites, where either co-authors or data were from different countries, such as Italy and Switzerland (Lomi and Pallotti 2012), Italy and the UK (Mascia et al. 2018), Ireland, UK, and Switzerland (Caimo et al. 2017), or Australia and France (Wang et al. 2013), USA and Italy (Bohnett et al. 2022). Based on the available funding disclosure in the publications, the federal government agencies in the US supported at least one-third of the included studies, including the National Institutes of Health and CDC. In contrast, the Swiss government funded three studies through the Swiss national Science Foundation (SNSF), the most in European countries.

The included studies are published across 20 journals and one conference proceeding. *Implementation Science* and *Social Networks* each published four studies, followed by the *American Journal of Public Health* (2 studies), *BMC Health Services Research* (2 studies), and several other health sciences journals (e.g., *American Journal of Preventive Medicine*, *Health Care Management Science*) or multi-discipline journals (*PLOS One*, *Scientific Reports*). Out of the 27 studies with FWCI scores, 15 of them have FWCI scores above 1, indicating that they are cited more frequently than expected based on the global average. Particularly, Wang et al.'s study (Wang et al. 2013) of a multilevel collaboration network of French cancer research elites has the highest FWCI score, 14.64. The second highest FWCI (9.07) is Heaney's "multiplex network" modeling of interest group connections/disconnection in the health policy process (Heaney 2014).

3.2 Study design

We observed five types of ERGMs for network modeling: 1) standard ERGM (STD) that consists of only endogenous and exogenous variables and assumes that the network structure is fixed and does not change over time; 2) Bayesian ERGM (Bergm) that combines the principles of Bayesian statistics and ERGM for modelling complex networks that exhibit various patterns of connectivity such as clustering, reciprocity, and transitivity; 3) temporal ERGM (Tergm) that accounts for the inter-temporal dependence in longitudinal collaborative network modeling; 4) separable temporal ERGM (Stergm) that extends Tergm to account for the temporal dependencies among the network structures over time and can handle both within-period component (i.e., within each time period) and between-period component (i.e., across time); and 5) multilevel ERGM (Mergm) that

consists micro-, meso-, and macro-level network modelling accounting for the effects of both individual-level and group-level covariates on the network structure.

Most studies (80%) used STD. As an extension to STD, Bergm, Tergm, Stergm, and Mergm were adopted to either address the limitations in STD or add additional parameters for network modeling. For example, only one study (Caimo et al. 2017) applied Bayesian computational methods to address the limitations of ERGM diffusion complexity. Tergm and Stergm predict the probability of maintaining a connection between two nodes concerning time. Incorporating this time-dependent parameter, researchers gained insight into the processes that drive connection formation and change (Azondekon 2018; Broekel and Bednarz 2019; Ho et al. 2021). Both Wang et al. (2013) and McGlashan et al. (2019) adopted a Mergm network to map and test complex interdependencies in a collaboration network. Wang was also a co-author in McGlashan et al.'s study. While Wang et al. focused on the French cancer researcher collaboration network at laboratory and advice-seeking levels, McGlashan et al. explored the complex collaborative system, where steering committee members connected for community-based public health interventions.

3.3 Networks

The networks reviewed in this study were modest in size, with only two studies (Harris 2013; Azondekon 2018) that applied ERGMs to model a network with more than 1000 nodes. In addition, density and centrality (e.g., degree, betweenness, closeness) were typical network measures in 80% of the included studies. Furthermore, researchers predominantly used simulation for goodness-of-fit (GOF) evaluation.

Notably, researchers used several endogenous variables to model network transitivity and centrality. For example, 22 studies modeled transitivity operating edgewise shared partners, dyadwise shared partners, their geometrically weighted counterparts, triangles and alternating triangles (Luke et al. 2010, 2013; Lomi and Pallotti 2012; Harris et al. 2012, 2015; Wang et al. 2013; Harris 2013; Uddin et al. 2013; Heaney 2014; Shearer et al. 2014; Okamoto 2015; Shoham et al. 2016; Caimo et al. 2017; Mascia et al. 2018; Bunker et al. 2018; Azondekon 2018; Broekel and Bednarz 2019; McGlashan et al. 2019; Prochnow et al. 2020; Ho et al. 2021; Kesternich and Rank 2022; Bohnett et al. 2022). Similarly, 21 studies adopted k-stars and geometrically weighted degrees to model network centrality (Fattore and Salvatore 2010; Luke et al. 2010, 2013; Zappa and Mariani 2011; Zappa 2011; Lomi and Pallotti 2012; Harris et al. 2012, 2015; Wang et al. 2013; Harris 2013; Uddin et al. 2013; Yousefi-Nooraie et al. 2014; Okamoto 2015; Bevc et al. 2015a, b; Shoham et al. 2016; Caimo et al. 2017;

Mascia et al. 2018; Broekel and Bednarz 2019; McGlashan et al. 2019; Kesternich and Rank 2022) (Table 4).

Most studies (93%) conducted GOF diagnostics for their ERGMs, in which two diagnostic measures were adopted. The first type is local by comparing the final ERGM with a null ERGM, while the second type is global by reaching the final ERGM with a network regression model constructed via a quadratic assignment procedure. The majority of authors relied on simulation for the GOF diagnostic.

4 Discussion

4.1 Bibliometrics

Overall, the bibliometrics characteristics of the included studies confirmed that ERGM is a relatively new approach to collaborative network analysis (Harris et al. 2012), and most biomedical and health sciences applications emerged only in the most recent decade. However, researchers in several countries have attempted to apply ERGMs to model professional collaborative networks in various health domains. The contributing authors have background and expertise in social sciences (e.g., social work, economics, management, political science, education), biomedical and health sciences (i.e., public health, medicine, nursing), or both. Government funding is the primary support for the studies, demonstrating the increasing force and determination to bring people, expertise, and resources together for optimal effectiveness.

4.2 Study design

There are two types of nodes in the reviewed ERGM networks: individuals and groups/organizations. Collaboration networks modeled upon individuals are at the micro-level. This review found that the micro-level network studies are primarily about physician collaborations (7 studies) for knowledge sharing or advice seeking (Zappa and Mariani 2011; Zappa 2011; Mascia et al. 2018; Bunger et al. 2018), patient care (e.g., Uddin et al. 2013; Kesternich and Rank 2022), or organization membership (Fattore and Salvatore 2010). The other micro-level collaborative network is researcher scientific collaborations (Wang et al. 2013; Okamoto 2015; Harris et al. 2015; Azondekon 2018; Ho et al. 2021). Collaboration networks modeled upon organizations are at the macro-level, involving different types of institutes, such as public health organizations (Luke et al. 2010, 2013) (Bevc et al. 2015a, b), hospitals/healthcare organizations (Lomi and Pallotti 2012; Caimo et al. 2017; Prusaczyk et al. 2019; Bohnett et al. 2022), R&D organizations in the industry (Broekel and Bednarz 2019), government public health departments (Harris 2013; Yousefi-Nooraie et al. 2014), and health policy interest groups (Heaney 2014). In addition,

we observed only two studies incorporating a multilevel network analysis approach, which characterized micro-, macro-, and meso-level networks and their interactions. The meso-level network models the network between individuals and organizations. Since networks rarely exist in isolation, the multilevel approach provides a complete picture of the mechanisms vital to the overall network structure, which shall be investigated in future research.

One-third of the included studies used secondary data from previous studies, projects, or existing data sets from databases, among which three studies supplemented the secondary data with additional data collected by interviews (Harris 2013; Luke et al. 2013) or interviews with a survey (Lomi and Pallotti 2012). Among all the data sources, The PARTNER database (Program to Analyze, Record, and Track Networks to Enhance Relationships) was adopted in three studies (Bevc et al. 2015a, b; Bohnett et al. 2022). As one of the most extensive network databases, PARTNER captures the connection of health organizations. It has been primarily used by practitioners engaged in inter-organizational collaborations in their communities for outcome evaluation.

For cross-sectional studies, surveys were used in ten studies to collect network-related data. In addition, medical or health insurance claim data were used in two studies (Uddin et al. 2013; Kesternich and Rank 2022). Interestingly, two studies used citation data from literature database searching or researchers' curriculum vitae (Marchand et al. 2018; Azondekon 2018). The co-authorship from the citations was helpful for data modeling upon homophily, cluster assignment, and temporal dependencies.

Twenty-two of 30 studies disclosed the software they adopted for network analysis and modeling. The ERGM package in the R statnet is the most frequently used software for ERGM fitting, especially for the studies which adopted the standard ERGM. Other R packages include R tergm, R btergm, R igraph, and R-mixer. The R libraries for ERGMs like statnet are crucial for network modeling. However, we have not identified a similar library in Python, a widely used data analytics platform. Researchers are calling for an ERGM library in Python so that users who are more familiar with this environment can also conduct social network analysis (Ghafouri and Khasteh 2020).

4.3 Networks

Networks modeled upon individuals (i.e., researchers or physicians) are always nested within a more extensive organizational network, such as academic institutions or hospitals. However, we found only two studies utilized this hierarchical approach by applying Mergm (Wang et al. 2013; McGlashan et al. 2019). All other studies modeled upon individuals at the micro-level or organizations at the macro-level without

exploring how the links between individuals and organizations contribute to the overall network structure. Similarly, only three studies modeled a network with time as a parameter, while most studies investigated collaboration networks at a single or fixed timepoint. Since social networks are dynamic, constantly changing, and evolving over time, future studies are warranted in applying ERGMs to multilevel (micro-, macro-, and meso-levels) and longitudinal collaboration networks.

In an ERGM, inherent network dependencies are modeled via endogenous variables. Geometrically weighted edgewise shared partner (GWESP) triadic closures and alternating k-stars were used in multiple studies (Harris 2013; Uddin et al. 2013; Caimo et al. 2017). Despite leaving out the meso-level network, several researchers acknowledged this hierarchical structure by including organizational membership as an exogenous variable (Shearer et al. 2014; Okamoto 2015; Bevc et al. 2015b).

To select a high-quality ERGM, all included studies commonly adopted the Akaike/Bayesian information criterion (AIC/BIC). After the ERGM was fitted, model diagnostics for that fitted ERGM were performed via simulation. Key network statistics summarizing the observed network's characteristics, such as the number of triangles and the number of edgewise-shared partners, were recorded. The simulated network statistics values did not differ significantly from the observed network in all studies reviewed. However, only four studies evaluated whether an ERGM was appropriate (Fattore and Salvatore 2010; Uddin et al. 2013; Heaney 2014; Okamoto 2015). They used a multiple regression model to achieve this goal with multiple regression quadratic assignment procedure as a popular choice.

The increased computational power of computers enables researchers to fit complex and computationally intensive models, such as ERGMs (Supplementary material 1). The leading utility of ERGMs lies in the specification of endogenous variables. Proper identification of network dependencies via the inclusion of endogenous variables can shed light on the characteristics and nature of a network, helping researchers answer complex questions regarding connection formation, change, and prediction. However, it is hard to discern a priori which endogenous variables to include as there are hundreds of possibilities used to model the same network dependence structure, such as transitivity and centrality. This challenge may explain the underutilization of ERGMs in practice. Although we identified the most popular endogenous variables from the included studies were geometrically weighted edgewise shared partners and k-stars, it is unknown why such endogenous variables were selected. Since the correct specification of endogenous variables is vital to GOF, for ERGMs to reach their full potential, a systematic and accessible guide to uncovering network dependencies via endogenous variables is needed. Furthermore, for

a network consisting of two types of nodes, organizational and individual level, researchers should consider using an Mergm. In addition, a Tergm or Stergm is better for modeling dynamic networks that change significantly over time. Finally, researchers who wish to leverage domain-specific expertise should consider a Bergm.

4.4 Limitations

Although we followed an SR protocol by systematically collecting, extracting, and summarizing research evidence pertinent to ERGM applications to collaborations in biomedical and health sciences, we may miss studies that are not indexed in PubMed or Scopus. In addition, studies published in a language other than English were not included. Furthermore, due to the heterogeneity of the studies, we did not conduct a quality assessment in this review. Finally, we only focused on common network characteristics (nodes, density, centrality) for network data extraction and did not compare or discuss ERGM statistical formulas that researchers applied to their studies.

ERGMs are powerful tools in social network analysis, adept at capturing intricate network interactions. However, they are not without challenges. Their inherent complexity often makes parameter estimation and interpretation daunting, especially in expansive networks (Stivala et al. 2020; Alberty et al. 2021). A significant concern of ERGMs is the proper selection of endogenous variables, which, if not accurately chosen, can lead to model misfits (An 2016). Moreover, the intractable normalizing constant in ERGMs adds to computational difficulties (Huang and Butts 2024). Another prevalent issue is model degeneracy, where ERGMs might yield unrepresentative or non-convergent results (Kei et al. 2023). While ERGMs are promising for network studies, researchers must tackle these challenges for their broader and more effective application.

5 Conclusion

This review demonstrates the research landscape of how researchers have applied ERGMs to collaboration network analysis in biomedical and health sciences. The bibliometric, study design and network modeling characteristics of included studies provide insights and guidance for future research. Professional collaboration networks are essential to solving complex scientific problems and optimizing healthcare delivery and health policy. Our systematic synthesis contributes to the design, development, and utilization of ERGMs in studies examining complex relationships and system structures native to professional collaboration networks. Furthermore, we advocate novel methodologies tailored for selecting endogenous variables and fitting ERGMs,

addressing the pronounced challenges that currently impede the optimal use of ERGMs.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13721-023-00439-w>.

Author contributions All authors contributed to the study conceptualization, manuscript review and revisions. FY conducted systematic literature search, study screening and data extraction, and produced the original draft of the manuscript; HE screened the retrieved publications and extracted data independently as well as participating in writing the original draft; MK provided guidance on results interpretation; AC and GD provided the context for this study and acquired financial support. All authors reviewed, discussed, and agreed to their individual contributions before the manuscript submission.

Funding The project described was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) through Grant Award Number UL1TR002489. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data availability The data sets extracted during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing financial interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albery GF, Kirkpatrick L, Firth JA, Bansal S (2021) Unifying spatial and social network analysis in disease ecology. *J Anim Ecol* 90:45–61. <https://doi.org/10.1111/1365-2656.13356>
- An W (2016) Fitting ERGMs on big networks. *Soc Sci Res* 59:107–119. <https://doi.org/10.1016/j.ssresearch.2016.04.019>
- Azondekon R (2018) Modeling the Complexity and Dynamics of the Malaria Research Collaboration Network in Benin, West Africa: papers indexed in the Web Of Science (1996–2016). *AMIA Annu Symp Proc* 2018:195–204
- Bennett LM, Gadlin H, Levine-Finley S (2010) Collaboration and Team Science: A Field Guide. National Institute of Health
- Bevc CA, Retrum JH, Varda DM (2015a) New perspectives on the “silo effect”: initial comparisons of network structures across public health collaboratives. *Am J Public Health* 105(Suppl 2):S230–S235. <https://doi.org/10.2105/AJPH.2014.302256>
- Bevc CA, Retrum JH, Varda DM (2015b) Patterns in PARTNERing across Public Health Collaboratives. *Int J Environ Res Public Health* 12:12412–12425. <https://doi.org/10.3390/ijerph121012412>
- Bohnett E, Vacca R, Hu Y et al (2022) Resilience and fragmentation in healthcare coalitions: The link between resource contributions and centrality in health-related interorganizational networks. *Soc Netw* 71:87–95. <https://doi.org/10.1016/j.socnet.2022.07.004>
- Broekel T, Bednarz M (2019) Disentangling link formation and dissolution in spatial networks: An Application of a Two-Mode STERGM to a Project-Based R&D Network in the German Biotechnology Industry. *Netw Spat Econ*. <https://doi.org/10.1007/s11067-018-9430-1>
- Bunger AC, Doogan N, Hanson RF, Birken SA (2018) Advice-seeking during implementation: a network study of clinicians participating in a learning collaborative. *Implement Sci* 13:101. <https://doi.org/10.1186/s13012-018-0797-7>
- Caimo A, Pallotti F, Lomi A (2017) Bayesian exponential random graph modelling of interhospital patient referral networks. *Stat Med* 36:2902–2920. <https://doi.org/10.1002/sim.7301>
- Drobyshevskiy M, Turdakov D (2020) Random Graph Modeling. *ACM Comput Surv* 52:1–36. <https://doi.org/10.1145/3369782>
- Elsevier Research Metrics Guidebook. <https://www.elsevier.com/research-intelligence/resource-library/research-metrics-guidebook>. Accessed 19 Jul 2021
- Falagas ME, Pitsouni EI, Malietzis GA, Pappas G (2008) Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 22:338–342. <https://doi.org/10.1096/fj.07-9492LSF>
- Fattore G, Salvatore D (2010) Network organizations of general practitioners: antecedents of formation and consequences of participation. *BMC Health Serv Res* 10:118. <https://doi.org/10.1186/1472-6963-10-118>
- Fujimoto K, Volente TW, Pentz MA (2009) Network structural influences on the adoption of evidence-based prevention in communities. *J Community Psychol* 37:830–845. <https://doi.org/10.1002/jcop.20333>
- Ghafouri S, Khasteh SH (2020) A survey on exponential random graph models: an application perspective. *PeerJ Comput Sci* 6:e269. <https://doi.org/10.7717/peerj-cs.269>
- Goodreau SM (2007) Advances in exponential random graph (p*) models applied to a large social network. *Soc Netw* 29:231–248. <https://doi.org/10.1016/j.socnet.2006.08.001>
- Harris JK (2013) Communication ties across the national network of local health departments. *Am J Prev Med* 44:247–253. <https://doi.org/10.1016/j.amepre.2012.10.028>
- Harris JK, Provan KG, Johnson KJ, Leischow SJ (2012) Drawbacks and benefits associated with inter-organizational collaboration along the discovery-development-delivery continuum: a cancer research network case study. *Implement Sci* 7:69. <https://doi.org/10.1186/1748-5908-7-69>
- Harris JK, Wong R, Thompson K et al (2015) Networks of Collaboration among Scientists in a Center for Diabetes Translation Research. *PLoS ONE* 10:e0136457. <https://doi.org/10.1371/journal.pone.0136457>
- Heaney MT (2014) Multiplex networks and interest group influence reputation: An exponential random graph model. *Soc Netw* 36:66–81. <https://doi.org/10.1016/j.socnet.2012.11.003>
- Ho E, Jeon M, Lee M et al (2021) Fostering interdisciplinary collaboration: A longitudinal social network analysis of the NIH mHealth Training Institutes. *J Clin Transl Sci* 5:e191. <https://doi.org/10.1017/cts.2021.859>
- Huang P, Butts CT (2024) Parameter estimation procedures for exponential-family random graph models on count-valued networks: A comparative simulation study. *Soc Netw* 76:51–67. <https://doi.org/10.1016/j.socnet.2023.07.001>
- Kei YL, Chen Y, Madrid Padilla OH (2023) A partially separable model for dynamic valued networks. *Comput Stat Data Anal* 187:107811. <https://doi.org/10.1016/j.csda.2023.107811>

- Kesternich E, Rank O (2022) Beyond patient-sharing: Comparing physician- and patient-induced networks. *Health Care Manag Sci*. <https://doi.org/10.1007/s10729-022-09595-3>
- Lomi A, Pallotti F (2012) Relational collaboration among spatial multipoint competitors. *Soc Netw* 34:101–111. <https://doi.org/10.1016/j.socnet.2010.10.005>
- Luke DA, Harris JK, Shelton S et al (2010) Systems analysis of collaboration in 5 national tobacco control networks. *Am J Public Health* 100:1290–1297. <https://doi.org/10.2105/AJPH.2009.184358>
- Luke DA, Wald LM, Carothers BJ et al (2013) Network influences on dissemination of evidence-based guidelines in state tobacco control programs. *Health Educ Behav* 40:33S–42S. <https://doi.org/10.1177/1090198113492760>
- Marchand GC, Hilpert JC, Bragg KM, Cummings J (2018) Network-based assessment of collaborative research in neuroscience. *Alzheimers Dement (n y)* 4:433–443. <https://doi.org/10.1016/j.trci.2018.08.006>
- Mascia D, Pallotti F, Dandi R (2018) Determinants of knowledge-sharing networks in primary care. *Health Care Manage Rev* 43:104–114. <https://doi.org/10.1097/HMR.0000000000000139>
- McGlashan J, de la Haye K, Wang P, Allender S (2019) Collaboration in Complex Systems: Multilevel Network Analysis for Community-Based Obesity Prevention Interventions. *Sci Rep* 9:12599. <https://doi.org/10.1038/s41598-019-47759-4>
- Okamoto J (2015) Scientific collaboration and team science: a social network analysis of the centers for population health and health disparities. *Transl Behav Med* 5:12–23. <https://doi.org/10.1007/s13142-014-0280-1>
- Page McKenzie, Bossuyt, et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. <https://doi.org/10.1136/bmj.n71>
- Prochnow T, Patterson MS, Sharkey J, Umstattd Meyer MR (2020) Health coalition collaboration network, perceived satisfaction and success. *J Health Organ Manag* 34:885–897. <https://doi.org/10.1108/JHOM-04-2020-0120>
- Provan KG, Harvey J, de Zapien JG (2005) Network structure and attitudes toward collaboration in a community partnership for diabetes control on the US-Mexican border. *J Health Organ Manag* 19:504–518. <https://doi.org/10.1108/14777260510629706>
- Provan KG, Huang K, Milward HB (2009) The evolution of structural embeddedness and organizational social outcomes in a centrally governed health and human services network. *J Public Adm Res Theory* 19:873–893. <https://doi.org/10.1093/jopart/mun036>
- Prusaczyk B, Maki J, Luke DA, Lobb R (2019) Rural health networks: how network analysis can inform patient care and organizational collaboration in a rural breast cancer screening network. *J Rural Health* 35:222–228. <https://doi.org/10.1111/jrh.12302>
- Puro N, Kelly RJ (2022) Community social capital or health needs: What is driving hospital-community partnerships to address social determinants of health? *SSM Popul Health* 18:101129. <https://doi.org/10.1016/j.ssmph.2022.101129>
- Retrum JH, Chapman CL, Varda DM (2013) Implications of network structure on public health collaboratives. *Health Educ Behav* 40:13S–23S. <https://doi.org/10.1177/1090198113492759>
- Robins G, Snijders T, Wang P et al (2007) Recent developments in exponential random graph (p*) models for social networks. *Soc Netw* 29:192–215. <https://doi.org/10.1016/j.socnet.2006.08.003>
- Schotten M, el Aisati M, Meester WJN et al (2017) A brief history of scopus: the world's largest abstract and citation database of scientific literature. In: Cantú-Ortiz FJ (ed) *Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics*. Taylor & Francis, Auerbach Publications, Boca Raton, FL, pp 31–58
- Shearer JC, Dion M, Lavis JN (2014) Exchanging and using research evidence in health policy networks: a statistical network analysis. *Implement Sci* 9:126. <https://doi.org/10.1186/s13012-014-0126-8>
- Shoham DA, Harris JK, Mundt M, McGaghie W (2016) A network model of communication in an interprofessional team of health-care professionals: A cross-sectional study of a burn unit. *J Interprof Care* 30:661–667. <https://doi.org/10.1080/13561820.2016.1203296>
- Smith TB, Vacca R, Krenz T, McCarty C (2021) Great minds think alike, or do they often differ? Research topic overlap and the formation of scientific teams. *J Informetr*. <https://doi.org/10.1016/j.joi.2020.101104>
- Stanford Human Evolutionary Ecology and Health Social Network Analysis for Anthropologists. <https://eehh-stanford.github.io/SNA-workshop/ergm-intro.html#what-is-a-random-graph?> Accessed 5 Sep 2022
- Stivala A, Robins G, Lomi A (2020) Exponential random graph model parameter estimation for very large directed networks. *PLoS ONE* 15:e0227804. <https://doi.org/10.1371/journal.pone.0227804>
- Uddin S, Hossain L, Hamra J, Alam A (2013) A study of physician collaborations through social network and exponential random graph. *BMC Health Serv Res* 13:234. <https://doi.org/10.1186/1472-6963-13-234>
- Yu F, Hayes B (2018) Applying data analytics and visualization to assessing the research impact of the cancer cell biology (CCB) program at the University of North Carolina at Chapel Hill. *J Esience Librariansh* 7:e1123. <https://doi.org/10.7191/jeslib.2018.1123>
- van der Pol J (2018) Introduction to Network Modeling Using Exponential Random Graph Models (ERGM): Theory and an Application Using R-Project. *Comput Econ*. <https://doi.org/10.1007/s10614-018-9853-2>
- Wang P, Robins G, Pattison P, Lazega E (2013) Exponential random graph models for multilevel networks. *Soc Netw* 35:96–115. <https://doi.org/10.1016/j.socnet.2013.01.004>
- Yousefi-Nooraie R, Dobbins M, Marin A (2014) Social and organizational factors affecting implementation of evidence-informed practice in a public health department in Ontario: a network modelling approach. *Implement Sci* 9:29. <https://doi.org/10.1186/1748-5908-9-29>
- Yu F, Van AA, Patel T et al (2020) Bibliometrics approach to evaluating the research impact of CTSAs: A pilot study. *J Clin Transl Sci* 4:336–344. <https://doi.org/10.1017/cts.2020.29>
- Zappa P (2011a) The network structure of knowledge sharing among physicians. *Qual Quant* 45:1109–1126. <https://doi.org/10.1007/s11135-011-9494-1>
- Zappa P, Mariani P (2011b) The interplay of social interaction, individual characteristics and external influence in diffusion of innovation processes: An empirical test in medical settings. *Procedia-Social Behav Sci* 10:140–147. <https://doi.org/10.1016/j.sbspro.2011.01.017>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.