

# Relational hyperevent models for the coevolution of coauthoring and citation networks

Jürgen Lerner<sup>1</sup> , Marian-Gabriel Hâncean<sup>2</sup> and Alessandro Lomi<sup>3</sup> 

<sup>1</sup>Department Computer and Information Science, University of Konstanz, Konstanz, Germany

<sup>2</sup>Department of Sociology, University of Bucharest, Bucharest, Romania

<sup>3</sup>Università della Svizzera italiana, Lugano, Switzerland

Address for correspondence: Jürgen Lerner, Department of Computer and Information Science, University of Konstanz, Universitätsstrasse 10, Konstanz 78464, Germany. Email: [juergen.lerner@uni-konstanz.de](mailto:juergen.lerner@uni-konstanz.de)

## Abstract

The development of appropriate statistical models has lagged behind the ambitions of empirical studies analysing large scientific networks—systems of publications connected by citations and authorship. Extant research typically focuses on either paper citation networks or author collaboration networks. However, these networks involve both direct relationships, as well as broader dependencies between references linked by multiple citation paths. In this work, we extend recently developed relational hyperevent models to analyse networks characterized by complex dependencies across multiple network modes. We introduce new covariates to represent theoretically relevant and empirically plausible mixed-mode network configurations. This model specification allows testing hypotheses that recognize the polyadic nature of publication data, while accounting for multiple dependencies linking authors and references of current and prior papers. We implement the model using open-source software to analyse publicly available data on a large scientific network. Our findings reveal a tendency for subsets of papers to be cocited, indicating that the impact of these papers may be partly due to endogenous network processes. More broadly, the analysis shows that models accounting for both the hyperedge structure of publication events and the interconnections between authors and references significantly enhance our understanding of the mechanisms driving scientific production and impact.

**Keywords:** citation networks, coauthoring networks, network dynamics, relational event models, hypergraphs, science of science

## 1 Introduction

Data on networks of scientific citations are reaching unprecedented levels of availability, scale, and scope (Lin et al., 2023). Parallel progress has been made on the analytical front where statistical models are now available to analyse increasingly large citation networks (Filippi-Mazzola & Wit, 2024). Stimulated by these developments, citation networks are now broadly recognized as the prime empirical basis for understanding the implications of the network structure and dynamics of scientific production, innovation, and impact (Barabási et al., 2002; Meng et al., 2024; Mukherjee et al., 2017; Newman, 2001; Radicchi et al., 2011).

As Liu et al. (2023, p. 1046) recently summarized: ‘The advent of large-scale datasets that trace the workings of science has encouraged researchers from many different disciplinary backgrounds to turn scientific methods into science itself’. The progressive development and expansion of the new interdisciplinary field of ‘science of science’ has considerably increased the interest in the

Received: February 5, 2024. Revised: July 2, 2024. Accepted: July 3, 2024

© The Royal Statistical Society 2024.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

network analysis of scientific networks—systems of scientific publications connected by bibliographic citations (Fortunato et al., 2018).

Science of science studies are increasingly present in prominent interdisciplinary science journals (Uzzi et al., 2013). For example, Guimera et al. (2005) explain performance of scientists working together in terms of team assembly mechanisms—or rules of partner selection. Uzzi et al. (2013) analyse how the atypicality of a paper's references affects the likelihood that a scientific publication will become a 'hit paper'. Park et al. (2023) found that papers are becoming less disruptive over time by showing a decline in the 'citation disruption index' (CDI)—a quantitative measure of the extent to which papers citing a focal paper do not cite its references. These results have been made possible by careful empirical analysis of scientific networks. While different in terms of methods and objectives, these works all frame the relation between scientific production and ultimate success of scientific ideas in terms of positions that scientific papers occupy in networks of paper citations (Sarigöl et al., 2014).

It is generally acknowledged that quantitative measures derived from scientific networks need to control for empirically observed features of these networks (Newman, 2001). For instance, Uzzi et al. (2013) normalize raw cocitation counts by expected values in randomized versions of the citation network, maintaining outdegree (i.e. the length of reference lists) and indegree (i.e. the number of citations that a paper receives). Park et al. (2023) adopt similar normalizations. However, randomization techniques for citation networks controlling for outdegrees and indegrees of papers are typically insufficient to provide *general* models for scientific networks and cannot test multiple competing hypotheses explaining scientific collaboration and impact. For instance, Petersen et al. (2024) claim that the CDI is confounded by changing citation behaviour caused, for example, by the increased availability of search engines for scientific publications, which alone can explain the increased tendency for 'triadic closure' in the citation network.

More generally, scientific networks emerging from patterns of paper citations involve two disjoint classes of entities that may be usefully represented as network nodes: scientific papers and their authors (Carusi & Bianchi, 2019). This implies the need to represent at least two network subsystems (Saavedra et al., 2009; Tumminello et al., 2011): authors connected to the papers they write, and papers connected to previous papers they cite (backward citations). In this view, citation networks involve a number of dual dependencies (Breiger, 1974; Lee & Martin, 2018), but extant research has typically restricted attention only to a subset of these dependence relations. For example, Ferligoj et al. (2015) model coauthoring networks with stochastic actor-oriented models (SAOMs) (Snijders, 2005). Fritz et al. (2023) adopt temporal exponential random graph models (TERGMs) (Krivitsky & Handcock, 2014; Lusher et al., 2013) to model collaboration among inventors that jointly file patents. Very much in the spirit of the current paper, Espinosa-Rada et al. (2024) specify and estimate multilevel SAOM to examine relations between papers, authors, and research institutions within a scientific community of astronomers from 2013 to 2015.

We premise the current study on what we consider three constitutive features of scientific networks emerging from time-ordered sequences of individual publication events. First, interaction among nodes (authors and papers) involves multiple parallel counting processes such as, for instance, the number of coauthored papers, or the number of citations to a paper. This property makes point-process models for network data (Perry & Wolfe, 2013), often denoted as relational event models (REMs) (Bianchi et al., 2024; Butts, 2008), a natural choice for modelling this kind of networks (Butts et al., 2023). Second, a single publication event does not connect only two nodes at a time, but it connects a varying and potentially very large number of authors and cited papers simultaneously, which involves 'polyadic' (one-to-many) interaction, or 'hyperevents' (Lerner & Hâncean, 2023; Lerner & Lomi, 2023). Treating these polyadic interactions as collections of dyadic interactions artificially inflates the number of observations, introduces dependence among observations by construction, and deprives analysts of the possibility to test for theoretically relevant and empirically plausible 'higher-order effects' (Lerner & Lomi, 2023). Third, publication events connect nodes of different types (authors and papers) with relations of different types: authors are connected to the papers they write (alone or in teams), and papers are connected to the multiple papers they cite. We argue that modelling either of these relations separately affords only a partial understanding of the dynamics of scientific networks.

Our main objective in this paper is to derive and test appropriate REMs that incorporate these core features of scientific networks and that support hypotheses testing about a variety of

empirically plausible and theoretically meaningful data generating mechanisms. We narrow the focus of our attention on REMs because, unlike better established exponential random graph models (ERGMs) (Amati et al., 2018), temporal ERGMs (Desmarais & Cranmer, 2012) or SAOMs (Snijders, 2017), REMs can deal with fine-grained time information thus removing the need to aggregate observations over arbitrary time intervals (Bianchi et al., 2024).

We contribute to—but also go beyond extant research in at least two ways. First, while current studies have recognized the analytical value of treating scientific networks as sequences of time-ordered sequences of relational hyperevents, we still know little about the underlying relational mechanisms that may be responsible for the observed structure of scientific networks. For example, Lerner and Hâncean (2023), model coauthoring networks via relational hyperevent model (RHEM) in an attempt to explain the ‘impact’ of papers, where impact is reconstructed in terms of a single numerical variable: the cumulative number of citations received at the time of data collection. However, the process through which papers accumulate citations was not modelled. In consequence, the model proposed by Lerner and Hâncean (2023) cannot be adopted to test or control for competing explanations of reference selection. In this paper, we take a more constructive approach by modelling directly the generating process underlying the observable network dynamics of paper citations.

Second, RHEMs are currently available for the analysis of processes defined on networks with a single class of nodes (one-mode, or unipartite networks) (Lerner & Lomi, 2023), or with two classes of nodes, e.g. participants and ‘labels’ of the event, where interaction is only between the node classes (two-mode, or bipartite networks) (Bright et al., 2024)—but not both simultaneously. In consequence, available RHEM cannot be specified to analyse mixed-mode networks (Snijders et al., 2013), where relational events can occur—and may be observed both between nodes in different classes, as well as nodes within of the same class. The value of such a model should be obvious in the analysis of scientific networks—the empirical setting of our study—where scientific collaboration produce direct relations between multiple authors, citations produce relations between multiple papers and their references, and publications produce associations between papers and their authors—and hence their institutional affiliation (Espinosa-Rada et al., 2024). In this paper, we extend the capacity of RHEM to analyse dynamic mixed-mode network processes. The corresponding effects that we specify and test are not only theoretically interesting, but also empirically important—as the results we report clearly demonstrate.

The empirical value of the new model we propose is tested using a large publicly available data on a scientific network containing more than a million papers and authors. We use this sample to test for the presence of higher-order dependencies and mixed two-mode dependencies and to compare the relative strength of the various effects, measured in terms of their individual contribution to the log likelihood.

We show that the magnitude of the new effects that we specify to capture salient aspects of mixed-mode network processes is comparable to some of the strongest effects documented by extant studies of paper citations based on one-mode network data. The exclusive modelling of citation networks—disregarding the interdependent coauthoring network—ignores that the choice of references cited in published papers is highly dependent on the authors of those papers. Plausible patterns in the coevolution of coauthoring and citation networks include the tendency of authors to: (i) cite their own prior papers (Seeber et al., 2019); (ii) cite papers of their past coauthors (Martin et al., 2013); (iii) adopt citations of their past coauthors (Goldberg et al., 2015); (iv) repeatedly cite the same papers (Merton, 1968; Price, 1976), and (v) repeatedly cite papers written by the same authors (White, 2001). Our analysis reveals a significant propensity of subsets of references to self-reproduce across publications. This result supports the conjecture that references come, at least to some extent, in discrete ‘packages’ of variable size. To the extent that the ‘impact’ of a scientific publication depends on the number of citations received (Radicchi et al., 2008), this results is important because it suggests that individual impact may be driven, in part, by nonindividual components.

We conclude our study inviting readers to reflect on the potential far-reaching implications, both pragmatic as well as conceptual, of the results afforded by the new models we present. These models can now be specified and estimated on very large datasets produced by scientific networks and connected systems such as, for example, patent citation systems (Filippi-Mazzola & Wit, 2024).

The results of the analysis that we report were obtained using a publicly available, open-source software<sup>1</sup> that interested readers can access freely. The combination of publicly available data and freely available software contributes to making the results of the analysis we present in this paper fully reproducible and potentially replicable on comparable datasets.

## 2 Models

The models considered in this paper are defined within the point-process class of models for relational event data proposed by [Perry and Wolfe \(2013\)](#). Theorems about the consistency and asymptotic normality of the maximum partial likelihood estimators presented in [Perry and Wolfe \(2013\)](#) also apply to the likelihood functions discussed below. Building directly on prior work ([Lerner & Hâncian, 2023](#); [Lerner & Lomi, 2023](#); [Perry & Wolfe, 2013](#)), we want to extend available models by developing novel hyperedge covariates, capturing the polyadic structure of co-authoring and citation and, more importantly, the multiple dependencies between these two kinds of networks.

We propose models for counting processes  $N_{(I,J)}(t)$ , where  $I$  is a *set* of authors, publishing papers that cite a *set* of references  $J$ . Thereby, we extend existing REM for citation networks ([Hunter et al., 2011](#)) in two directions. First, we do not focus exclusively on the cited papers, but consider also the interrelation between authors in  $I$  and references in  $J$ . Second, we take into account that coauthoring and citation networks cannot be completely represented in terms of purely dyadic interactions, connecting only two nodes at a time. Note that a single publication event entails a varying number of authors citing a varying—and arbitrarily large—number of references.

Let  $\mathcal{I}$  be a finite set of authors and  $\mathcal{J}$  be a finite set of papers. For a point in time  $t > 0$ , let  $\mathcal{I}_t \subseteq \mathcal{I}$  be those authors who could potentially publish a paper at  $t$  and let  $\mathcal{J}_t \subseteq \mathcal{J}$  be the set of those papers that could potentially be cited by a paper published at  $t$  (prior work). Models defined below specify distributions for sequences of publication events  $(t_1, j_1, I_1, J_1), \dots, (t_n, j_n, I_n, J_n)$ , where  $(t_m, j_m, I_m, J_m)$  indicates that authors  $I_m \subseteq \mathcal{I}_{t_m}$  publish paper  $j_m \in \mathcal{J}$  at time  $t_m$ , citing papers  $J_m \subseteq \mathcal{J}_{t_m}$  in their references. See [Figure 1](#) for the graphical illustration of a publication event, followed by a later publication whose reference list could partially be explained by the tendency of authors to repeatedly cite the same papers.

For any set of authors  $I \subseteq \mathcal{I}_t$  (i.e. a possible set of coauthors) and any set of papers  $J \subseteq \mathcal{J}_t$  (i.e. a possible set of references), let  $x_t(I, J)$  be a  $p$ -dimensional vector of covariates with associated unknown vector of parameters  $\beta_0$ .

### 2.1 The joint model

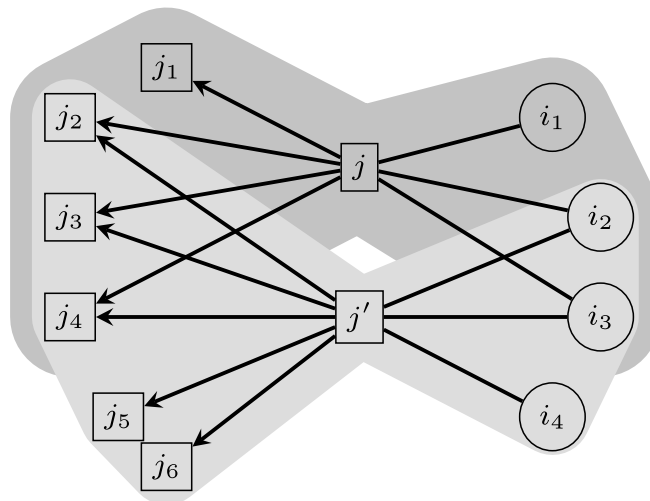
Consider the counting processes on  $\mathbb{R}_+ \times \mathcal{P}(\mathcal{I}) \times \mathcal{P}(\mathcal{J})$ , where the intensity on  $(I, J) \in \mathcal{P}(\mathcal{I}) \times \mathcal{P}(\mathcal{J})$  is specified as

$$\lambda_t(I, J) = \bar{\lambda}_t(|I|, |J|) \exp\{\beta_0^\top x_t(I, J)\} \mathbf{1}\{I \subseteq \mathcal{I}_t \wedge J \subseteq \mathcal{J}_t\}. \quad (1)$$

For a pair of positive integers  $K$  and  $L$ , giving the number of authors and the number of references, respectively,  $\bar{\lambda}_t(K, L)$  is the baseline intensity, and  $\mathbf{1}\{I \subseteq \mathcal{I}_t \wedge J \subseteq \mathcal{J}_t\}$  is the indicator whether the pair  $(I, J)$  is ‘at risk’ of experiencing an event at time  $t$ . For a pair  $(I, J)$  that is at risk, the intensity  $\lambda_t(I, J)$  is the product of the baseline intensity with the relative rate, or hazard ratio,  $\exp\{\beta_0^\top x_t(I, J)\}$ , whose estimation is of central interest for empirical applications. The hazard ratio indicates the factor by which a possible publication event at time  $t$  with authors  $I$  and references  $J$  is more or less likely than an alternative publication event with authors  $I'$  and references  $J'$ . The hyperedge covariates  $x_t(I, J)$  can be specified as functions of exogenous attributes of authors or papers and/or they can be functions of the history of the process, that is, of publication events at time  $t' < t$ . In the next section (Effects), we offer a representative list of possible hyperedge covariates.

Note that the specification in (1) includes the number of authors  $|I|$  and the number of references  $|J|$  in the baseline rate, implying that the hazard ratio reveals the factor by which observed events  $(I, J)$  are more or less likely than alternative events  $(I', J')$  with the same number of authors ( $|I'| = |I|$ ) and the same number of references ( $|J'| = |J|$ ). This decision is in line with—but

<sup>1</sup> <https://github.com/juergenlerner/eventnet>.



**Figure 1.** Publication event of paper  $j$ , authored by  $i_1, \dots, i_3$  and citing papers  $j_1, \dots, j_4$ , followed by the publication event of paper  $j'$ , authored by  $i_2, \dots, i_4$  and citing papers  $j_2, \dots, j_6$ . The two authors  $i_2, i_3$  repeatedly cite the three papers  $j_2, j_3, j_4$ . Mixed two-mode hyperedges, containing the published paper, its authors, and its references, are displayed as grey-shaded areas enclosing the participating nodes.

extends—the specification in [Perry and Wolfe \(2013\)](#) and [Lerner and Lomi \(2023\)](#) who include the number of receivers in the baseline rate (the number of senders is fixed to one in these papers). Ultimately, this decision can be motivated by the simple but crucial observation that the number of subsets of size  $k$  increases exponentially in  $k$ , as long as  $k$  is much smaller than the size of the entire set (i.e. the number of all authors who could publish or the number of all citable papers, respectively). For instance, the baseline publication rate on sets comprising 10 authors is tremendously lower than the baseline rate on sets comprising two authors—simply because there are so many more different author sets of size 10, than sets of size two. The decision to absorb the number of authors and references in the baseline rate ensures that effects of substantive interest, shaping the hazard ratio, do not get confounded by trivial but very strong variation in the number of subsets of given sizes.

The specification in (1) involves a stratified Cox proportional hazard model ([Aalen et al., 2008](#); [Cox, 1972](#)), where the strata are defined by the combination of the number of authors and the number of references. This simple observation is very useful since it implies that model parameters can be estimated by standard software for Cox regression, as it will be explained in more detail in the empirical section of this paper.

Let  $(t_1, j_1, I_1, J_1), \dots, (t_n, j_n, I_n, J_n)$  be the observed sequence of publication events. Adapting models proposed in [Perry and Wolfe \(2013\)](#) and [Lerner and Lomi \(2023\)](#) to the case of publication events, the model defined in (1) leads to the following partial likelihood at time  $t$ , evaluated at parameters  $\beta$ .

$$L_t(\beta) = \prod_{t_m \leq t} \frac{\exp\{\beta^T x_{t_m}(I_m, J_m)\}}{\sum_{(I,J) \in \binom{I_m}{|I_m|} \times \binom{J_m}{|J_m|}} \exp\{\beta^T x_{t_m}(I, J)\}} \quad (2)$$

(For any set  $M$  and integer  $k$ , we denote by  $|M|$  the cardinality of  $M$  and we denote by  $\binom{M}{k} = \{M' \subseteq M; |M'| = k\}$  the set of all subsets of size  $k$  of  $M$ .) Theorems from [Perry and Wolfe \(2013\)](#) imply that, under generally accepted assumptions, estimating parameters  $\hat{\beta}$  by maximizing the partial likelihood (2) produces consistent estimates.<sup>2</sup>

<sup>2</sup> To see that this claim holds, consider the comment by [Perry and Wolfe \(2013, p. 828\)](#): *the multicast case can be considered as a special case of the single-receiver case*. In our case, events have not only several ‘receivers’ (cited papers) but also several ‘senders’ (authors), which does not affect the applicability of their theorems.

Equation (2) is the partial likelihood of a stratified CoxPH model (see above) and standard maximum likelihood estimation of the parameters  $\beta$  relies on the assumption that events are conditionally independent, given the past events, and that the dependence on the past is completely captured by the vector of covariates  $x_t$ . We note that the validity of this assumption in any given empirical data is, in general, unprovable and it is unlikely to hold in the strict sense in complex, large-scale data, as the one analysed in this paper. Aalen et al. (2008) note that if not all dependence on the past is captured by the given covariates, then: *Often the estimation procedures are still valid, but the variance estimates from martingale theory are not correct (they will typically underestimate the true variance) and have to be substituted with sandwich type estimators* (Aalen et al., 2008, p. 303). Following this recommendation, we will estimate parameters with robust estimation in the empirical section of this paper.

For a large sample of authors and papers, the number of subsets in  $\binom{\mathcal{I}_{tm}}{|I_m|}$  and  $\binom{\mathcal{J}_{tm}}{|J_m|}$  grows exponentially in  $|I_m|$ , or  $|J_m|$ , respectively, leading to a computationally intractable likelihood for all but the smallest numbers of coauthors and references. We, therefore, resort to the approach of nested case-control sampling for CoxPH models (Borgan et al., 1995; Keogh & Cox, 2014), which has also been applied for the practical estimation of large REM and RHEM, for instance, in Lerner and Lomi (2020, 2023). For a positive integer  $q$ , representing the number of nonevents (or ‘controls’) per event (or ‘case’), let  $\tilde{\mathcal{R}}_t(\mathcal{I}_t, I, \mathcal{J}_t, J, q) \subseteq \binom{\mathcal{I}_t}{|I|} \times \binom{\mathcal{J}_t}{|J|}$  be a set of pairs of subsets that is sampled uniformly at random from

$$\left\{ \mathcal{R} \subseteq \binom{\mathcal{I}_t}{|I|} \times \binom{\mathcal{J}_t}{|J|}; \quad (I, J) \in \mathcal{R} \wedge |\mathcal{R}| = q + 1 \right\} \quad (3)$$

The sampled risk set  $\tilde{\mathcal{R}}_t(\mathcal{I}_t, I, \mathcal{J}_t, J, q)$  can be constructed by including  $(I, J)$  (the observed pair of authors and references, i.e. the ‘case’) plus  $q$  alternative pairs  $(I', J') \neq (I, J)$ , the ‘controls’, sampled without replacement uniformly and independently at random from  $\binom{\mathcal{I}_t}{|I|} \times \binom{\mathcal{J}_t}{|J|}$ .

We obtain a sampled partial likelihood  $\tilde{L}_t(\beta)$  by substituting the sampled risk set  $\tilde{\mathcal{R}}_{t_m}(\mathcal{I}_{t_m}, I_m, \mathcal{J}_{t_m}, J_m, q)$  for  $\binom{\mathcal{I}_{tm}}{|I_m|} \times \binom{\mathcal{J}_{tm}}{|J_m|}$  in the summation of (2):

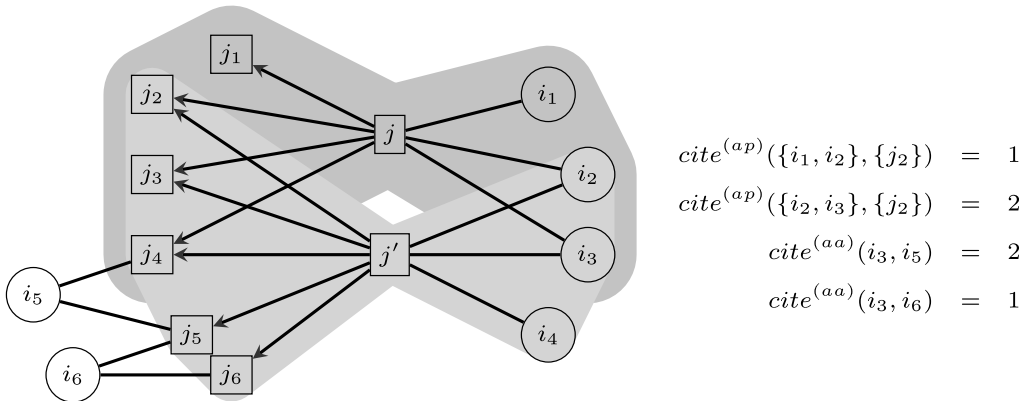
$$\tilde{L}_t(\beta) = \prod_{t_m \leq t} \frac{\exp\{\beta^T x_{t_m}(I_m, J_m)\}}{\sum_{(I, J) \in \tilde{\mathcal{R}}_{t_m}(\mathcal{I}_{t_m}, I_m, \mathcal{J}_{t_m}, J_m, q)} \exp\{\beta^T x_{t_m}(I, J)\}} \quad (4)$$

According to Borgan et al. (1995), estimating parameters  $\hat{\beta}$  by maximizing  $\tilde{L}_t(\beta)$  is based on an estimator which is consistent under reasonable conditions (also compare Lerner & Lomi, 2023). As noted above, parameter estimation can be done by standard software for Cox regression—provided that the software is given a table containing the values of the hyperedge covariates  $x_t(I, J)$  for all instances, that is, for all events and sampled controls. We provide details on this aspect in the empirical part of this paper.

## 2.2 Separation into coauthoring model and citation model

For some empirical applications, it may be useful to decompose the joint model into two submodels: a ‘coauthoring model’ explaining the propensity of sets of authors to coauthor papers and a ‘citation model’ explaining the conditional probability of the reference list of a paper published by a given set of authors. In some applications this decomposition may more faithfully reflect scientific collaboration and publication processes in which scientists first form teams and then write their common paper (which includes the decision about which papers to cite). The decomposition of the joint model into the two submodels reflects this separation and could potentially support a simpler interpretation of the different effects in the model. This decomposition is a possible option in modelling scientific networks, although it is not strictly necessary. We emphasize that this model decomposition, if adopted, would still maintain the interaction of coauthoring and citation networks. For instance, coauthoring probabilities can still be modelled dependent on past citations and the conditional probability of reference lists can still be modelled dependent on the authors of the current and past publications.





**Figure 2.** Publication events of papers  $j$  and  $j'$  with their authors  $i_1, \dots, i_4$  and references  $j_1, \dots, j_6$ . Additionally, authors ( $i_5, i_6$ ) of some of the cited papers are given. Values of network attributes in this example are computed without any decay over time. Values for  $\text{cite}^{(ap)}$  and  $\text{author}$  are binary and are given by the lines (edges) connecting papers to papers, or authors to papers, respectively. Values of other network attributes on selected nodes are given on the right-hand side.

Specifications of the coauthoring model and citation model are very similar to the joint model—except that the space of possibilities in these models are sets of authors (for the coauthoring model), or sets of papers (for the citation model), respectively. We provide details in the [Appendix A](#).

### 3 Effects

Effects in RHEM for scientific citation networks are operationalized by specifying hyperedge covariates  $x_t(I, J)$ . These covariates can be functions of exogenous attributes of nodes (authors or papers) or they can be functions of the history of the process, that is, functions of publication events  $(t', j', I', J')$  with  $t' < t$ . These history-dependent covariates are more demanding to understand, but also introduce empirically more interesting effects in the models. We will, therefore, focus on the definition of history-dependent covariates, which will proceed in two steps. First, we define a small number of time-dependent functions (denoted as ‘network attributes’) assigning values to nodes or sets of nodes, summarizing past interaction in prior events. These network attributes represent the state of the scientific network at any point in time. Second, we define the model covariates as functions of these network attributes. While this separation is not necessary, we argue that the network attributes are conceptually intuitive summaries of the interaction history, reduce the cognitive burden to understand the definition of covariates (in other words, they reduce the complexity of formulas), and represent a modular approach in which definitions of attributes might be modified without changing the definition of covariates as functions of attributes, and vice versa.

We illustrate the network attributes in [Figure 2](#) in an example extending [Figure 1](#). The definition of model covariates is further illustrated with small stylized examples and graphics in the [online supplementary material](#).

#### 3.1 Summarizing history through time-varying network attributes

Given a sequence of publication events  $E = \{(t_1, j_1, I_1, J_1), \dots, (t_n, j_n, I_n, J_n)\}$ , the value of network attributes at time  $t$  is a function of earlier events

$$E_{<t} = \{(t_m, j_m, I_m, J_m) \in E; t_m < t\},$$

where we let the effect of past events decay over time with a given half-life period  $T_{1/2} > 0$  ([Brandes et al., 2009](#); [Lerner et al., 2013](#)) using the following weight factor dependent on the elapsed time  $t - t_m$

$$w(t - t_m) = \exp\left(-[t - t_m] \frac{\log 2}{T_{1/2}}\right).$$

While using a decay is not theoretically necessary, it is empirically plausible that the effects of past collaborations or citations will fade over time. The inclusion of a decay parameter also often leads to a better model fit (Lerner & Hâncean, 2023).

The most important network attribute, which will be used as a basis for defining several covariates, encodes the extent to which a set of authors  $I'$  has cited a set of papers  $J'$  in joint publications. This attribute is a generalization of the ‘sender-specific hyperedge degree’, defined in Lerner and Lomi (2023), to the case of two-mode hyperevent networks where the number of ‘senders’ (authors in our case) is unconstrained. For a set of authors  $I' \subseteq \mathcal{I}$  and a set of papers  $J' \subseteq \mathcal{J}$ , past author-to-paper citations are encoded in the attribute

$$cite_t^{(ap)}(I', J') = \sum_{t_m < t} w(t - t_m) \cdot \mathbf{1}(I' \subseteq I_m \wedge J' \subseteq J_m).$$

The attribute adds up the contributions of all past publication events  $(t_m, j_m, I_m, J_m)$  satisfying that all members of the given set of authors  $I'$  are authors of  $j_m$  (i.e.  $I' \subseteq I_m$ ) and all members of the given set of papers  $J'$  are cited by  $j_m$  (i.e.  $J' \subseteq J_m$ ). As special cases, this attribute already encodes to what extent a set of authors has coauthored papers, irrespective of the citations of their papers, by using  $J' = \emptyset$  and it encodes the extent to which a set of papers have been cocited by other papers, regardless of the authors of these citing papers, by using  $I' = \emptyset$ .

For two papers  $j, j' \in \mathcal{J}$ , paper-to-paper citations are encoded in the attribute

$$cite_t^{(pp)}(j, j') = \sum_{t_m < t} w(t - t_m) \cdot \mathbf{1}(j = j_m \wedge j' \in J_m).$$

Note that at most one term of this sum is different from zero, since every paper is published only once, that is, appears only once as  $j_m$  in a publication event  $(t_m, j_m, I_m, J_m)$ . Seen from another angle, and first disregarding any decay, the values  $cite_t^{(pp)}(j, j')$  define the binary adjacency matrix of the citation network among all papers published before  $t$ ; the entry in row  $j$  and column  $j'$  is one if  $j$  cites  $j'$  and it is zero else. Considering the decay function, the impact of past citations decays exponentially in the elapsed time since the citing paper  $j$  has been published.

For two authors  $i, i' \in \mathcal{I}$ , author-to-author citations are encoded in the attribute

$$cite_t^{(aa)}(i, i') = \sum_{t_m < t} w(t - t_m) \cdot \mathbf{1}(i \in I_m \wedge \exists t_{m'} : j_{m'} \in J_m \wedge i' \in I_{m'}).$$

In words, the publication event of paper  $j_m$  is counted in the sum given above if and only if  $i$  is among the authors of  $j_m$  and  $j_m$  cites at least one paper  $j_{m'}$  that has  $i'$  among its authors. Thus, first disregarding any decay, the value  $cite_t^{(aa)}(i, i')$  is the number of papers  $j_m$  that are published before  $t$ , that have  $i$  among their authors, and that cite at least one paper having  $i'$  as an author. The entries  $cite_t^{(aa)}(i, i')$ , still disregarding any decay, therefore, define an integer-valued adjacency matrix of the author–cites–author network, whose edges  $(i, i')$  are weighted by the number of papers published by  $i$  before  $t$  and citing any paper published by  $i'$ . Considering the decay function, the impact of past citations decays exponentially in the elapsed time since the citing paper  $j_m$  has been published.

For an author  $i \in \mathcal{I}$  and a paper  $j \in \mathcal{J}$ , past author relations are encoded in the attribute

$$author_t(i, j) = \sum_{t_m < t} w(t - t_m) \cdot \mathbf{1}(j = j_m \wedge i \in I_m).$$

Again, at most one term of this sum is different from zero, since every paper is published only once. Disregarding any decay, the entries  $author_t(i, j)$  define binary adjacency matrix of the two-mode network connecting authors to the papers they publish (before  $t$ ). The value in row  $i$  and column  $j$  is one if  $i$  is an author of paper  $j$  and it is zero else.



For an author  $i \in \mathcal{I}$  the *author citation popularity* is the number of papers citing any paper published by  $i$  (downweighted by the elapsed time since the publication of the citing paper), defined in the node-level attribute

$$\text{citepop}_t^{(a)}(i) = \sum_{t_m < t} w(t - t_m) \cdot \mathbf{1}(\exists t_{m'} : j_{m'} \in J_m \wedge i \in I_{m'}).$$

The length of the reference list of a paper  $j \in \mathcal{J}$  is captured by the following node-level attribute

$$\text{outdeg}_t(j) = \sum_{t_m < t} w(t - t_m) \cdot \mathbf{1}(j = j_m) \cdot |J_m|.$$

We recall that some of the network attributes, which are auxiliary functions used in the definition of the hyperedge covariates in the following Section 3.2, are illustrated in Figure 2 on a small stylized example. We note that many different ways to scale or normalize these attributes are possible. It is not within the scope of this paper to explore different variants—or empirically to find the ‘optimal’ normalization.

### 3.2 Hyperedge covariates for network effects

A generic covariate that may be used as a base to define several different types of covariates is called *subset repetition of order  $(k, \ell)$*  (Lerner et al., 2019). In our case, subset repetition formalizes three variants of effects: (1) if a set of authors  $I'$  has already coauthored one or several papers (possibly together with others), they are likely to coauthor future papers (possibly together with other, potentially different authors); (2) if a set of papers  $J'$  has already been cocited (possibly within a larger list of references), they are likely to be cocited again (possibly within a different larger list of references); and (3) if a set of authors  $I'$  have already coauthored a paper citing a set of references  $J'$ , they are likely to coauthor another paper, also citing the set of references  $J'$ . The crucial pattern in all of these variants is that some—but not necessarily all—of the nodes (authors and/or references) participating in a past event, participate again in a future event—possibly within a varying set of further coparticipating nodes. Subset repetition is parameterized by the sizes of the subsets that are to be repeated. For two nonnegative integers  $k$  and  $\ell$  that are not both equal to zero, a set of authors  $I \in \mathcal{I}$ , and a set of papers  $J \in \mathcal{J}$ , subset repetition of order  $(k, \ell)$  is defined as

$$\text{subrep}_t^{(k, \ell)}(I, J) = \sum_{(I', J') \in \binom{I}{k} \times \binom{J}{\ell}} \frac{\text{cite}_t^{(ap)}(I', J')}{\binom{|I|}{k} \cdot \binom{|J|}{\ell}}.$$

The expression above iterates over all subsets of authors  $I' \subseteq I$  and references  $J' \subseteq J$  of the prescribed sizes and averages the past author-to-paper citation weight  $\text{cite}_t^{(ap)}(I', J')$ , defined above, over these combinations. Subset repetition can be specialized to the following three families of covariates. Recall that these and other covariates are further illustrated in the online supplementary material.

**Prior papers and prior joint papers.** For a set of authors  $I \subseteq \mathcal{I}$  the average number of prior papers (downweighted by the elapsed time) is captured by the covariate

$$\text{prior papers}_t(I) = \text{subrep}_t^{(1, 0)}(I, \emptyset)$$

and is a measure of the past publication activity of authors in  $I$ . Previous collaboration among pairs of authors is captured by

$$\text{prior joint papers}_t(I) = \text{subrep}_t^{(2, 0)}(I, \emptyset),$$

which averages the number of coauthored papers (downweighted by the elapsed time) over all unordered pairs of authors in  $I$ . It would be possible to measure past collaboration among larger sets

of authors via  $subrep^{(k,0)}(I, \emptyset)$  for  $k > 2$ . However, in our empirical analysis we found that the analysed scientific network is too sparse to allow estimation of these effects, so that we dropped them from our models.<sup>3</sup>

A related covariate captures the heterogeneity of past publication activity via

$$\text{diff prior papers}_t(I) = \sum_{\{i,i'\} \in \binom{I}{2}} \frac{|cite_t^{(ap)}(\{i\}, \emptyset) - cite_t^{(ap)}(\{i'\}, \emptyset)|}{\binom{|I|}{2}}.$$

A positive parameter associated with this covariate would suggest that teams of coauthors are typically composed of authors having a larger dispersion in the number of prior publications than expected by chance alone. A negative parameter would suggest homogeneous teams, whose prior publication activity tends to be more similar than expected by chance alone.

**Paper citation popularity and cocitation.** For a set of papers  $J \subseteq \mathcal{J}$  the average number of past citations (downweighted by the elapsed time) is captured by the covariate

$$\text{paper citation popularity}_t(J) = subrep_t^{(0,1)}(\emptyset, J).$$

A positive parameter associated with this covariate would indicate a ‘rich get richer’ effect in the sense that papers with more past citations are also cited at a higher rate in the future. Previous cocitations of pairs and triples of papers are captured by

$$\begin{aligned} \text{paper-pair cocitation}_t(J) &= subrep_t^{(0,2)}(\emptyset, J) \\ \text{paper-triple cocitation}_t(J) &= subrep_t^{(0,3)}(\emptyset, J). \end{aligned}$$

Positive parameters for these covariates would suggest that pairs (or triples) of papers that jointly appeared in reference lists are more likely to be cocited again in future publications.

**Author citation repetition.** The hypothetical tendency of (groups of) authors to repeatedly cite the same papers (compare Figure 1) is captured by the following covariates combining the history of a set of authors  $I \subseteq \mathcal{I}$  and a set of papers (potential references)  $J \subseteq \mathcal{J}$

$$\begin{aligned} \text{author citation repetition}_t(I, J) &= subrep_t^{(1,1)}(I, J) \\ \text{author-pair citation repetition}_t(I, J) &= subrep_t^{(2,1)}(I, J) \\ \text{author-triple citation repetition}_t(I, J) &= subrep_t^{(3,1)}(I, J). \end{aligned}$$

**Citing a paper and its references.** The tendency to adopt (some of) the references of a cited paper—or, conversely, to cite a paper and some of its citing papers—is captured by the covariate measuring the past citation density within a set of papers  $J \subseteq \mathcal{J}$  (representing a possible list of references):

$$\text{cite paper and its refs}_t(J) = \sum_{\{j,j'\} \in \binom{J}{2}} \frac{cite_t^{(pp)}(j, j') + cite_t^{(pp)}(j', j)}{\binom{|J|}{2}}.$$

We note that this covariate is likely to have an influence on the CDI (Park et al., 2023) since a positive parameter would suggest a lower expected CDI.

<sup>3</sup> However, see Lerner and Hâncian (2023) who estimated subset repetition among coauthors up to order 10 in a smaller and denser coauthoring network.

**Paper outdegree popularity.** We define the average length of reference lists of a set of papers  $J \subseteq \mathcal{J}$  via the covariate

$$\text{paper outdegree popularity}_t(J) = \sum_{j \in J} \frac{\text{outdeg}_t(j)}{|J|}.$$

A positive parameter associated with this covariate would suggest a tendency to cite papers with long reference lists. This covariate can be seen as a precondition of the covariate ‘citing a paper and its references’ which should be included to control for varying length of reference lists.

**Collaborate with citing author.** The tendency of scientists to coauthor papers with those who cited their previous work is captured by the covariate measuring the past citation density within a set of authors  $I \subseteq \mathcal{I}$ :

$$\text{collab with citing author}_t(I) = \sum_{\{i, i'\} \in \binom{I}{2}} \frac{\text{cite}_t^{(aa)}(i, i') + \text{cite}_t^{(aa)}(i', i)}{\binom{|I|}{2}}.$$

**Author citation popularity.** We define the average citation popularity of a set of authors  $I \subseteq \mathcal{I}$  via the covariate

$$\text{author citation popularity}_t(I) = \sum_{i \in I} \frac{\text{citepop}_t^{(a)}(i)}{|I|},$$

and the heterogeneity in author citation popularity (similar to the heterogeneity in the past publication activity) via the covariate

$$\text{diff auth cite pop}_t(I) = \sum_{\{i, i'\} \in \binom{I}{2}} \frac{|\text{citepop}_t^{(a)}(i) - \text{citepop}_t^{(a)}(i')|}{\binom{|I|}{2}}.$$

A positive parameter associated with the author citation popularity would suggest that authors whose papers have been cited more often in the past are more likely to publish future papers. A positive parameter associated with the difference in authors’ citation popularity would suggest that teams of coauthors are typically composed of authors with varying citation popularity, e.g. senior scientists coauthoring together with junior scientists.

**Author self-citation.** Capturing the tendency of authors to cite their own past work, we define the covariate *self-citation* measuring the density of the two-mode subgraph connecting a set of authors  $I \subseteq \mathcal{I}$  and a set of papers  $J \subseteq \mathcal{J}$  with respect to the *author* relation

$$\text{self-citation}_t(I, J) = \sum_{i \in I, j \in J} \frac{\text{author}_t(i, j)}{|I| \cdot |J|}.$$

**Triadic closure (author pairs).** The remaining list of covariates are variants of triadic closure in the sense that pairs of nodes (which can be authors or papers) are characterized by their past relations to ‘third’ nodes (which, again, can be authors or papers) in various types of relations, defined as network attributes above. The first four covariates in this family establish indirect relations among pairs of authors, that may have an influence on the probability that these authors collaborate in the future.

**Coauthor closure.** The covariate *common coauthor* measures to what extent authors in  $I \subseteq \mathcal{I}$  have coauthored papers with the same ‘third’ authors. To abbreviate notation, we write  $coauth_t(i, i') = cite_t^{(ap)}(\{i, i'\}, \emptyset)$ .

$$\text{common coauthor}_t(I) = \sum_{\{i, i'\} \in \binom{\mathcal{I}}{2} \wedge i'' \neq i, i'} \frac{\min[coauth_t(i, i''), coauth_t(i', i'')]}{\binom{|I|}{2}}.$$

In the summation above, the ‘third’ author index  $i''$  runs over all authors  $i'' \in \mathcal{I}$ , different from the two focal authors  $i$  and  $i'$ .

**Citing common paper.** A related covariate *citing common paper* measures to what extent authors in  $I \subseteq \mathcal{I}$  have cited the same papers in the past.

$$\text{cite common paper}_t(I) = \sum_{\{i, i'\} \in \binom{\mathcal{I}}{2} \wedge j \in \mathcal{J}} \frac{\min[cite_t^{(ap)}(\{i\}, \{j\}), cite_t^{(ap)}(\{i'\}, \{j\})]}{\binom{|I|}{2}}.$$

**Citing common author.** The covariate *citing common author* measures to what extent authors in  $I \subseteq \mathcal{I}$  have cited (papers published by) the same authors in the past.

$$\text{cite common auth}_t(I) = \sum_{\{i, i'\} \in \binom{\mathcal{I}}{2} \wedge i'' \neq i, i'} \frac{\min[cite_t^{(aa)}(i, i''), cite_t^{(aa)}(i', i'')]}{\binom{|I|}{2}}.$$

**Cited by common author.** Finally, the covariate *cited by common author* reverses these relations and measures to what extent (papers published by) authors in  $I \subseteq \mathcal{I}$  have been cited by the same authors in the past.

$$\text{cited by common auth}_t(I) = \sum_{\{i, i'\} \in \binom{\mathcal{I}}{2} \wedge i'' \neq i, i'} \frac{\min[cite_t^{(aa)}(i'', i), cite_t^{(aa)}(i'', i')]}{\binom{|I|}{2}}.$$

**Triadic closure (author–paper pairs).** The remaining four ‘triadic closure’ covariates establish indirect relations between authors and papers via various types of relations.

**Adopt citations of coauthors.** A tendency of authors to cite the papers that have been cited in the past by their past coauthors can be assessed by the following covariate:

$$\text{adopt cite coauth}_t(I, J) = \sum_{i \in I \wedge j \in J \wedge i' \neq i} \frac{\min[coauth_t(i, i'), cite_t^{(ap)}(\{i'\}, \{j\})]}{|I| \cdot |J|}.$$

**Citing papers of coauthors.** Similarly, a tendency of authors to cite the papers that have been published by their past coauthors can be assessed by the following covariate:

$$\text{cite coauthor's paper}_t(I, J) = \sum_{i \in I \wedge j \in J \wedge i' \neq i} \frac{\min[coauth_t(i, i'), author_t(i', j)]}{|I| \cdot |J|}.$$

**Citation-repetition on the author level.** We define a covariate *author–author citation repetition* that can test the tendency of authors to repeatedly cite the papers of the same authors

$$\text{auth–auth cite repet}_t(I, J) = \sum_{i \in I \wedge j \in J \wedge i' \neq i} \frac{\min[cite_t^{(aa)}(i, i'), author_t(i', j)]}{|I| \cdot |J|}.$$

Note that this covariate is structurally different from ‘author citation repetition’ since the latter tests whether the same author tends to repeatedly cite the same paper, while the new ‘author–author citation repetition’ tests whether the same author  $i$  tends to repeatedly cite potentially different papers written by the same other author  $i'$ .

**Citation reciprocation on the author level.** Reversing a relation in the previously defined author–author citation repetition, we define a covariate *author–author citation reciprocation* testing the tendency of authors to cite the papers of those other authors who have previously cited their papers

$$\text{auth–auth cite recipr}_i(I, J) = \sum_{i \in I \wedge j \in J \wedge i' \neq i} \frac{\min[\text{cite}_i^{(aa)}(i', i), \text{author}_i(i', j)]}{|I| \cdot |J|}.$$

We emphasize that many different ways to scale or normalize these covariates are possible and there are also many possibilities to define yet other structurally different covariates. The objective of this paper is to define a representative set of covariates, illustrating structurally different ways how past publication events may shape the distribution of future publications. It is outside the scope of this paper to provide a ‘complete’ list of covariates.

### 3.3 Hyperedge covariates dependent on exogenous factors

Another family of hyperedge covariates may be expressed in terms of numeric or categorical exogenous attributes of authors or papers. Examples of such exogenous attributes for papers include subject classifiers, keywords, and numeric summaries of the text of abstracts or whole papers. For the case of authors, exogenous attributes include gender, seniority, or affiliation. Types of hyperedge covariates based on such exogenous attributes include summary statistics of attribute values over authors, papers, or of the interaction between author attributes and paper attributes (e.g. interaction between research interest of authors and subject classifiers of papers cited by them). Simple examples for the functional form of such covariates are provided by the two effects ‘author citation popularity’ and ‘diff author citation popularity’, defined above. Although these covariates are defined as functions of the history-dependent node-level attribute  $\text{citepop}_i^{(a)}(i')$ , their functional form could be reused to define similar covariates dependent on exogenous attributes (e.g. seniority or gender).

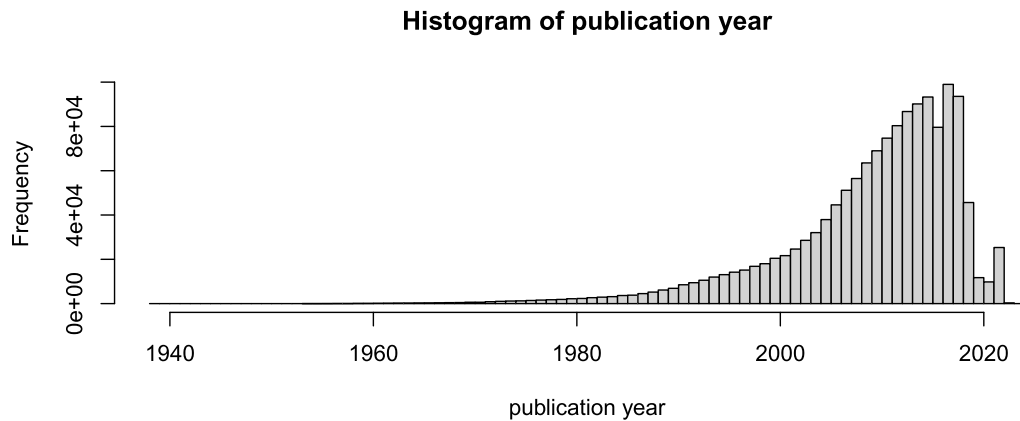
Because we do not use any exogenous attributes in our empirical example below—and given that their definition may be considered relatively straightforward compared to some of the more complex history-dependent attributes—we do not explicitly define covariates depending on exogenous attributes. We further refer the reader to [Lerner and Lomi \(2023\)](#) for explicit examples of such covariates.

### 3.4 Hyperedge covariates dependent on the published paper

We note that none of the covariates defined above takes the focal paper  $j_m$  of a publication event  $(t_m, j_m, I_m, J_m)$  into account. Indeed, since we defined only covariates dependent on the history of the process, considering the paper  $j_m$  would not make any difference, since a paper has an empty history at its publication time. If, however, we considered covariates dependent on exogenous attributes, it would be possible in a straightforward manner to also take attributes of  $j_m$  into account. For instance, overlap in keywords of  $j_m$  with the papers in  $J_m$ , related subject classifiers, or some measure of text similarity, could make it more likely that  $j_m$  has  $J_m$  as its reference list.

## 4 Empirical example

We test the empirical value of the new RHEM we have described in an analysis of a large scientific networks of bibliographic citations. The analysis focuses on the estimation and interpretation of the mixed-mode effects, and on the assessment of their relative importance vis-à-vis the set of more common ‘network effects’ routinely incorporated in empirical specification of statistical models for citation networks ([Newman, 2004](#)) to capture indegrees and outdegrees and clustering of authors and papers.



**Figure 3.** Histogram of publication year of papers used in the empirical analysis.

#### 4.1 Data

To construct our data, we take the Aminer Citation Network Dataset (<https://www.aminer.org/citation>) ‘DBLP-Citation-network V14’ (Tang et al., 2008), which we restrict to journal publications (indicated by the field value `doc_type=“Journal”`). We also retain only those references that appear in this restricted data (i.e. only citations from and to journal articles). If by this process, the list of references of some paper  $j$  becomes empty, we drop the respective publication event from our analysis (however,  $j$  could still be cited by another paper in our data). The resulting 1,416,353 papers (i.e. publication events) are written by 1,286,941 unique authors. The number of authors per paper ranges from 1 to 134 with mean equal to 2.8 and the number of references (included in the analysis) ranges from 1 to 2,268 with the mean number of references equal to 8.6. The year of publication ranges from 1939 to 2023, where the bulk of papers has been published after the year 2000; see the distribution of publication years in Figure 3.

#### 4.2 Preprocessing, estimation, and reproducibility

Explanatory variables (i.e. hyperedge covariates) are computed with the open-source software `eventnet` (Lerner & Lomi, 2020; Lerner et al., 2021), available at <https://github.com/juergenlerner/eventnet>. The `eventnet` software samples  $q$  nonevent hyperedges associated with each observed hyperedge (i.e. publication event) and computes all covariates of observed events and sampled nonevents. For defining the set of authors  $\mathcal{I}_t$  that could publish a paper in year  $t$  and the set of papers  $\mathcal{J}_t$  that could be cited in year  $t$ , we use all authors, respectively papers, that appear in any publication event at  $t$  or earlier. In particular, we do never remove authors from  $\mathcal{I}_t$  since we do not have any information about when an author cannot (or decides not to) publish anymore. (Given such information, `eventnet` would allow to remove nodes from the risk set.) Note, however, that covariates capturing author activity (e.g. ‘prior papers’) have a decay and approach zero after longer periods of inactivity. Thereby, the models could learn implicitly that authors are unlikely to publish again after longer periods of inactivity.

We set the number of nonevents per event to  $q = 5$ . All network attributes are defined with the half-life period  $T_{1/2}$  set to 3 years. We apply to all covariates the square-root transformation  $x \mapsto \sqrt{x}$ , which consistently results in a better fit for all models. We note that a general search for optimal  $q$ , half-life periods, or variable transformation is out of scope of this paper. However, we comment more on the sample size and choice of the half-life parameter after having sketched the estimation method.

For each model, we estimate parameters with the `coxph` function in the R package `survival` (Therneau, 2015; Therneau & Grambsch, 2013), using robust estimation. As noted when defining the intensity (1) and the (sampled) likelihood function, (2) and (4), our models are stratified CoxPH models where the strata are defined by the combination of the number of authors and the number of references. However, for reasons lying in the implementation of the `coxph`



function, we must explicitly provide the information that only controls and events having the same event time (i.e. publication time in our case) are put in the same fraction of the sampled likelihood (4). This has for instance been noted by Aalen et al. (2008), who write when discussing the Cox model with nested case-control sampling: *For computing, one may use standard software for Cox regression and other relative risk regression models, formally treating the label of the sampled risk sets as a stratification variable in the model [...]* (Aalen et al., 2008, p. 197). Thus, in our case, the `strata` argument of the `coxph` function is a variable defined by the concatenation of publication time (year), number of authors, and number of references.

We observe that robust estimation via `coxph` leads for some of the effects to considerably different estimates for standard errors than default estimation. However, this rarely affects significance at conventional levels ( $p < .01$  or  $p < .001$ ), most likely due to the size of our sample. We nevertheless recommend robust estimation in studies where it is reasonable to expect the presence of residual dependencies among the observations.

Since, in general, many papers are published in a given year, we have a dataset with ‘tied event times’ (Breslow, 1974; Efron, 1977; Hertz-Picciotto & Rockhill, 1997; Kalbfleisch & Prentice, 1973). In the `coxph` function that we adopt tied event times are handled by default through the so-called ‘Efron approximation’ (Efron, 1977; Hertz-Picciotto & Rockhill, 1997). We further note that, even in the presence of tied event times, the assumption that the values of hyperedge covariates at time  $t$  are only functions of prior publication events at time  $t' < t$  still holds. Our models assume that simultaneous events are conditionally independent of each other, given the set of previous events (Lerner et al., 2013). In practice, this implies that, for instance, authors do not base their decision about the selection of coauthors or references on other publications appearing in the same year. While testing the implications of this assumption is outside of scope of this paper, we do not think that this assumption could be considered implausible or unnecessarily restrictive. We note that, for instance, patent collaboration and citation networks typically come with a much finer time granularity, often given by the day. Scientific papers published in the last few years often provide more fine-grained information on (online) publication dates given by the day. When available, our model is able to make efficient use of these pieces of finer-grained information about the timing of events.

To assess the reliability of parameter estimates with the given chosen value of  $q$ , we generate 100 bootstrap samples (Efron, 1992)—a possibility mentioned by Lerner and Lomi (2020)—and reestimate parameters of the ‘joint model’ from Table 1 on each of these samples. We find that for each of the 20 significant effects, all 100 parameters estimated by bootstrap sampling have the same sign. For the single nonsignificant effect (citing common author), the parameters estimated by bootstrap sampling are rather centred around zero—confirming the finding that this parameter is not significantly different from zero. Thus, judging by bootstrap sampling, our sample size seems to be large enough to support reliable estimates.

We argue—on theoretical grounds—that the choice of the half-life period is likely to impact findings. Changing the half-life period implies that we test different effects: long-term effects of, e.g. past collaboration or citation versus their short-term effects. It could be expected that long-term effects are different from short-term effects, at least for some of the structural patterns. For instance, there might be an effect if two scientists have ever coauthored a paper, even if this happened decades ago; but this long-term effect might be different from a short-term effect of having coauthored a paper few years ago. To probe this issue further, we also fitted the three models from Table 1 where all covariates were computed without any decay (‘infinite’ half-life). Indeed findings are qualitatively different. One notable difference is that in the no-decay model, the parameter of ‘prior papers’ (average number of prior papers of the authors included in a hyperedge) has a significantly negative sign, whereas it is positive in the models reported in Table 1 (with half-life set to 3 years). This could be explained by the fact that we are analysing a period of more than 60 years. Authors who have been very active in the (distant) past may simply be no longer active scientists and therefore do no longer publish at all. The model without decay might reflect this possibility by producing a negative effect of the number of prior papers (e.g. once scientists have accumulated many citations, they are closer to retirement age and stop publishing). More generally, models without decay are not able to express that scientists may be active in some periods but later become less active, that collaborations among scientists might dissolve, that once popular papers might lose their popularity (and hence get cited less), and so on. While we discard the model

**Table 1.** Estimated parameters and standard errors (in brackets) obtained by robust estimation

	Coauthors	Citations	Joint
Prior papers	0.215 (0.008)***		0.129 (0.004)***
Diff prior papers	0.181 (0.004)***		0.185 (0.003)***
Prior joint papers	1.729 (0.018)***		0.823 (0.009)***
Author citation popularity	−0.032 (0.002)***		−0.017 (0.001)***
Diff auth cite pop	0.027 (0.002)***		−0.007 (0.001)***
Collab with citing author	0.503 (0.013)***		0.318 (0.006)***
Common coauthor	−0.080 (0.004)***		−0.014 (0.003)***
Citing common paper	−0.008 (0.006)		−0.112 (0.004)***
Citing common author	−0.056 (0.003)***		0.001 (0.002)
Cited by common author	−0.059 (0.002)***		0.007 (0.002)***
Paper citation popularity		0.188 (0.001)***	0.170 (0.001)***
Paper-pair cocitation		0.020 (0.011)	0.032 (0.012)**
Paper-triple cocitation		−0.208 (0.021)***	−0.170 (0.021)***
Author citation repetition		0.829 (0.018)***	0.643 (0.017)**
Author-pair citation repet		−0.524 (0.016)***	
Author-triple citation repet		−0.195 (0.020)***	
Paper outdegree popularity		0.192 (0.005)***	0.195 (0.006)***
Cite paper and its refs		7.746 (0.023)***	7.360 (0.024)***
Adopt citation of coauthor		−0.038 (0.006)***	−0.264 (0.007)***
Self-citation		1.745 (0.028)***	1.799 (0.025)***
Cite coauthor's paper		0.497 (0.017)***	0.291 (0.016)***
Author–author citation repet		−0.042 (0.017)*	−0.381 (0.015)***
Author–author citation recip		−0.258 (0.014)***	−0.343 (0.014)***
Akaike Information Criterion (AIC)	20602304.582	19168404.629	18927741.318
Num. events	1, 416, 353	1, 416, 353	1, 416, 353
Num. obs.	8, 493, 351	8, 493, 344	8, 493, 353

Note. \*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$ .

without decay on theoretical grounds we argue that it is hard to find optimal half-life parameters—at least for a network of this size. The optimal value might vary by effect and models might even be specified by adding the same structural effect several times with varying half-life parameters (Amati et al., 2019).

The analysis we present next is fully reproducible. We provide all software and scripts (preprocessing, eventnet configuration files, and R scripts), as well as explanations on how to use them, on a dedicated page in the eventnet wiki, available at: [https://github.com/juergenlerner/eventnet/wiki/Coevolution-of-collaboration-and-references-to-prior-work-\(tutorial\)](https://github.com/juergenlerner/eventnet/wiki/Coevolution-of-collaboration-and-references-to-prior-work-(tutorial)). The open-source software eventnet is published under the GPL 3.0 license, allowing inspection, modification, and republishing of code and providing a reference implementation for computing hyperedge covariates.

### 4.3 Results

**Model parameters.** Table 1 provides estimated parameters of the coauthor model, citation model, and joint model. Note that, while the observed events are identical for the three models, the sampled nonevents are different and are indeed sampled from different spaces: sets of authors, sets of papers, and pairs of author sets and paper sets, respectively.

Parameters typically—but not always—have the same sign and significance levels in the joint model and in one of the two submodels. An exception is the difference in the author citation popularity (diff auth cite pop) and the effect assessing whether two scientists are likely to become co-authors if they have been ‘cited by [a] common author’. Below, we discuss findings related to some of the strongest, or most interesting and interpretable effects.

Findings explaining the formation of coauthor teams include a positive effect of the number of prior papers (i. e., scientists publishing more in the recent past, tend to publish at a higher rate in the future) and a positive effect of the number of prior joint papers (suggesting a tendency to repeat scientific collaboration with the same coauthors). These self-reinforcing process are consistent with established hypotheses on the cumulative advantage in science (Price, 1976), and social inertia in scientific collaboration (Ramasco & Morris, 2006), respectively,

The difference in the number of prior papers has a positive effect, suggesting that coauthor teams are typically composed of scientists with diverse publication histories such as, for example, senior scientists publishing papers with their PhD students. We find that scientists have a tendency to collaborate with those who cited their papers. The four triadic effects explaining coauthoring, from ‘common coauthor’ to ‘cited by common author’, are mostly negative in the coauthoring model, with the effect of having cited common papers not being significant. This result is in line with previous findings on negative closure in coauthor networks (Lerner & Hăncean, 2023) and hyperevent networks of meeting events (Lerner & Lomi, 2022; Lerner et al., 2021). Following the interpretation offered in prior work, negative closure suggests the existence of various dense groups of actors that do not merge over time—or, alternatively, the existence of actors occupying stable broker positions, surrounded by ‘structural holes’, or open triads (Burt, 1992).

Strong and positive effects explaining the selection of references include the number of times that these references have been cited before (paper citation popularity), the tendency of authors to repeatedly cite the same papers (author citation repetition), to cite their own papers (self-citation), and to cite papers published by their past coauthors (cite coauthor’s paper). Another strong positive effect is the tendency to cite papers together with some of their references. This is consistent with the finding by Park et al. (2023) that the citation disruption index is lower than expected in a randomized citation network (although it does not say anything about their main claim stating that the CDI is declining over time).

**Contributions of covariates in the joint model.** Table 2 presents contributions of the individual covariates to the log likelihood of the joint model. The first numeric column gives the differences in log likelihoods of the respective single-parameter model (specified with only one covariate at a time) over the null model (no parameters). The second numeric column gives the differences between the ‘full model’ (specified with all 21 covariates) and the models with 20 parameters obtained by dropping the respective effect from the full model. The last column gives the type of effect: ‘A’ for pure authorship effects, that is, covariates that could still be computed if we discarded the set of cited papers  $J_m$  from all publication events  $(t_m, j_m, I_m, J_m)$ , ‘C’ for pure citation network effects, that is, covariates that could still be computed if we discarded the set of authors  $I_m$  from all publication events, and ‘M’ for mixed effects that depend on authors and citations. The log likelihood of the null model is  $-11,063,051$  and the difference in log likelihood between the full model and the null model is  $1,599,202$ .

We find that the strongest effect, by far, is the tendency of authors to cite a paper together with some of the references of that cited paper. The implied pattern in citation networks emerges if, for instance, paper  $j_1$  is cited by  $j_2$  and later paper  $j_3$  cites  $j_1$  and  $j_2$ . Different explanations could lead to this pattern: authors may copy part of the references of a paper they cite—or authors may search for citations to a paper they want to cite, and include also some of these citing papers in their reference list. Besides these purely mechanistic explanations, the papers’ content (similarity) or relevance for a specific new paper could as well explain this pattern. If  $j_2$  cites  $j_1$ , it is a signal that these two papers are likely to be related; if  $j_3$  cites  $j_2$  they are related as well; by transitivity  $j_1$  and  $j_3$  are related, which may increase the probability that  $j_3$  cites  $j_1$ . With the given data we cannot distinguish between such different explanations for the same structural effect.

The second-strongest effect is the paper citation popularity—an effect linked to the so-called ‘Matthew effect’ in science (Merton, 1968)—indicating that papers that have been cited more often in the past accumulate citations at a higher rate in the future. These two covariates are

**Table 2.** Contributions to the log likelihood of individual covariates in the joint model

	To null	In full	Type
Cite paper and its refs	1,150,680	237,136	C
Paper citation popularity	629,911	115,535	C
Author citation repetition	614,921	3,495	M
Self-citation	537,228	9,967	M
Num prior joint papers	508,556	21,586	A
Num prior papers	508,359	1,230	A
Author–author citation repetition	441,856	1,545	M
Collab with citing author	415,377	4,324	M
Diff num prior papers	402,533	5,537	A
Paper-pair cocitation	400,152	278	C
Author–author citation reciprocation	359,472	1,331	M
Cite coauthor’s paper	355,529	849	M
Adopt citation of coauthor	329,170	4,649	M
Author citation popularity	251,639	198	M
Diff auth cite pop	217,229	55	M
Paper-triple cocitation	212,057	1,123	C
Citing common paper	202,071	1,682	M
Common coauthor	189,043	67	A
Paper outdegree popularity	168,659	12,428	C
Citing common author	160,407	0	M
Cited by common author	112,673	66	M

*Note.* First numeric column: improvements over the null model; second numeric column: contributions in the full model. Last column: type of effect (‘A’ for pure coauthorship effects, ‘C’ for pure citation effects, ‘M’ for mixed effects).

pure citation network effects and they have the strongest individual contributions over the null model and in the full model.

The next effects, ordered by contributions over the null model, are mixed effects: the tendency of authors to repeatedly cite the same papers and the tendency of authors to cite their own prior work. For these effects we observe that the contributions over the null model can lead to a different relative ordering than the contributions in the full model. A possible explanation is that for some effects we have included more related effects than for others—which implies that the focal effect has a smaller additional contribution in the full model. For instance, part of the contributions of ‘author citation repetition’ can presumably be explained by the papers’ citation popularity and/or by the authors’ publication activity (prior papers). Conversely, the ‘paper outdegree popularity’ has a relatively weak effect over the null model but the fourth strongest effect in the full model. We note that the outdegree popularity is one out of few effects that reverse the roles of citing and cited papers: papers that cite many others (high outdegree) are more likely to be cited. The only other effect that also reverses the roles of ‘citing’ and ‘being cited’ is the ‘author–author citation reciprocation’—although in this effect the nodes are authors and not papers.

The effects at Positions 5 and 6 are the pure authorship effects ‘prior joint papers’ (i.e. repeatedly coauthoring papers with the same others) and ‘prior papers’ (i.e. past publication activity measured by the number of prior papers, downscaled by the elapsed time). Some of the weakest effects are the four triadic closure effects that relate two authors through past relations to common cited papers, common coauthors, common cited authors, or by having been cited by common authors. The difference in contributions from one effect to the next, ordered by contributions over the null model, are associated with relatively small jumps in the log-likelihood function—with the notable exception of the unexpectedly strong contribution of the strongest effect (partial copying of the reference lists of cited papers).

## 5 Discussion and conclusions

We proposed, implemented, and tested new hyperedge covariates to extend RHEM to the analysis of scientific networks—dynamic mixed two-mode networks whose nodes are authors and papers, where authors are connected to the papers they write, and papers are connected to other papers they cite. The models presented in this paper extend the analytical possibilities in the study of scientific citation networks by affording consideration of a broader set of dynamic network dependencies. Existing models considering only dependencies in either the coauthoring network, or in the citation network, respectively, can only take into account a subset of possible dependencies in complex systems of scientific citation networks.

We argued and demonstrated that the network dynamics of the two components of this evolving network system (authors-by-papers, and papers-by-references) are heavily interdependent. On the one hand, the selection of reference lists depends on the authors of the current and past papers. On the other, the formation of teams of coauthors also depends on their and others past citations. We argued, further, that interaction in scientific networks is inherently polyadic, since a single publication event connects a varying and theoretically unbounded number of authors and cited papers. Relational hyperevent model are a family of statistical models that can deal with the polyadic (or ‘multicast’) nature of scientific networks. The main contribution of this paper was to adapt RHEM to mixed two-mode networks of time-stamped observations in the spirit of [Snijders et al. \(2013\)](#).

We adapted RHEM to networks of publication events by defining a list of representative covariates that implement structurally distinctive effects. We conducted an empirical analysis of a large scientific network to demonstrate the empirical value of the model. We found that mixed-mode effects—explaining scientific publications through a combination of authoring and citation relations—are generally strong and can be used to test theoretically relevant and empirically meaningful effects explaining scientific collaboration and research orientation. Likewise, we found significant polyadic effects—acknowledging that cocitation probabilities of two or more papers are not determined by individual citation probabilities of single papers. Most notably, we found a significant tendency to cite a paper together with subsets of its references. The corresponding effects are among the strongest in our models. We think that this result is particularly important because it suggests that the impact of an individual paper depends, at least in part, on the propensity of that paper to become part of a ‘package’ of other papers that are frequently cited together. This seems to be a particularly important conjecture in the light of the standard practice to consider the number of citations received as a general measure of scientific impact ([Radicchi et al., 2008](#)).

We emphasize that the contribution of our paper is neither in the general form of the likelihood function, nor the estimation procedure as both should be considered well established features of relational event and hyperevent models ([Bianchi et al., 2024](#); [Butts, 2008](#); [Lerner & Lomi, 2023](#); [Perry & Wolfe, 2013](#)). Our emphasis on dependence among publications determined by intersecting references is also not new. The germane concept of bibliographic coupling as an approach to measure similarity among scientific papers has been well established within bibliometry at least since the work of [Kessler \(1963\)](#). Rather, the novelty of the present work should be evaluated in terms of the new effects (hyperedge covariates) for publication events that can be tested after inclusion in the partial likelihood (2) and (4). These effects allow us to go well beyond aggregate measures of similarity among publications that may be derived from bibliographic coupling ([Liu, 2017](#)). The new effects that we have presented in this paper allow to specify and test the possible network micro-mechanisms underlying observed macrolevel patterns of bibliographic coupling. The empirically validated conclusion that some of these new effects make significant and very strong contributions to our understanding of the structure of scientific networks and their constructive relational mechanisms may be interpreted as evidence that models not accounting for the polyadic nature of publication events, and/or not accounting for the interdependence between authoring and citation networks might miss important dependencies of current publications on prior publications. Thus, our paper makes a small but significant step forward in the quest for statistical models for citation networks that are both analytically rigorous, as well as empirically plausible.

Much work remains to do be done to alleviate some of the limitations of the present study. For example, a comprehensive approach to evaluating the goodness of fit of the models we have presented like the one recently established by [Amati et al. \(2024\)](#) is still not available at a scale

comparable to that of the sample we have analysed. Also, model estimation procedures based on the sampled partial likelihood that we have implemented building on prior studies (Perry & Wolfe, 2013) assumes that events are conditionally independent, given the earlier events, and that this dependence is fully reflected in the model specification. We have not tested for potential violations of this assumption. We simply followed best recommended practice (Aalen et al., 2008) and controlled for generic sources of violation by adopting robust estimation techniques. A final issue worth mentioning is more of an empirical issue than a hard-wired limitation of the model. In studies that, like ours, are based on case-control sampling it is common for a particular covariate (or set of covariates) to be particularly sparse in the risk set but not infrequent in the subset of observed cases. Lerner and Lomi (2020) discuss this problem of sparsity and the convergence issues that it may generate in unfavourable empirical circumstances.

Despite these limitations, we think our study provides reliable evidence that citations linked to scientific papers are not independent: citations reported in a paper are typically overlapping with the citations reported in cited papers. The fact that a paper citing another often also cites some of its references may not be surprising in light of what we now know about the practice of scientific production (Meng et al., 2024; Pandey et al., 2020; Simkin & Roychowdhury, 2002). In fact, we would expect that technological progress in information retrieval and the increased availability of large digital paper repositories is likely to accentuate and accelerate these trends in the practice of scientific research. What may be surprising is that, before the current paper, no model was available to adequately incorporate these predictable empirical features in the analysis of scientific citation networks.

We think that the results presented—while only suggestive at the current stage—have potentially far-reaching implications for the evaluation of scientific productivity, science policy and, ultimately, for the evaluation of scientists' academic career. If citations are not independent, but—as our study suggests—‘come in packages’, then it should be obvious that bibliometric indexes computed under the assumption of independence will be imperfect, and possibly inflated by the tendency of specific references to be included in subsets of references that are frequently cited together by the same authors, their coauthors or simply by other papers connected by direct citation links. Understandably perhaps, individual success measured by number of citations remains the dominant metric for assessing scientific impact. Our work suggests that individual success may be affected by nonindividual components generated by complex network dependencies.

Some of the effects that we have defined in this paper will help derive and test specific hypotheses about the mechanisms that shape the size and composition of ‘citation packages’, the frequency with which they recur across papers, and the tendency of authors and coauthors to include citations in packages. In the light of the results presented in this paper, we believe that future research in the science of science should reexamine the implicit assumption found in many studies that the number of citations received represents a direct measure of scientific impact at the individual, or even team-level (Fortunato et al., 2018).

In closing, it is worth emphasizing that the dynamic structure of scientific citation networks also arises in other settings. For instance, the data structure typical of scientific networks is commonly found in studies of patent citation networks connecting teams of inventors citing related patents (Acemoglu et al., 2016). Similarly, artistic productions where multiple artists are involved in collaborative productions make explicit or implicit references to other productions or artists (Burgdorf et al., 2024; Lena, 2004; Schuster et al., 2019). Many of the statistical effects that we have derived, implemented and tested in this paper in the context of scientific networks seem to arise naturally from repeated collaboration in artistic productions (Faulkner, 1983; Uzzi & Spiro, 2005), and music more specifically (Gleiser & Danon, 2003). These different interactive contexts tend to create communities with permeable boundaries whose members reassemble over time in different combinations around different productions. We offer the conjecture that collaborative production of ideas tends to produce networks with very similar structural features almost regardless of contextual details (Saavedra et al., 2009).

Avenues for future research include both empirical, as well as methodological developments. Examples of empirical work that may benefit from the results reported in this paper include reproducibility studies of influential empirical findings in research on the science of science concerning, for example, the effect of citation atypicality on the impact of scientific papers (Uzzi et al., 2013), or the apparent decline of disruptive scientific innovations (Park et al., 2023). Using the more



general models discussed in this paper, it may be possible to define new ways to normalize quantitative measures in citation networks and test the robustness of existing results across normalization strategies. Other possible empirical studies employing models proposed in this work could test theories of social selection (e.g. scientists start collaboration due to common research interest), social influence (e.g. scientists adopt research orientation of their current or past collaborators), assembly of scientific teams, or determinants of scientific productivity, innovation, and impact.

Future progress on the modelling front might come from developing a more exhaustive list of covariates for mixed two-mode networks, techniques to find optimal scaling or normalization of network attributes and covariates, and techniques to find optimal half-life periods for the decaying influence of past events. Model estimation could be made more robust, for instance, through Markov chain simulation to more efficiently explore the large space of (plausible) controls. Unobserved heterogeneity of authors and/or papers may be accounted for by adapting ideas from multilevel REM (e.g. [Uzaheta et al., 2023](#)) or latent-space REM (e.g. [Mulder & Hoff, 2021](#)) to RHEM that we discussed in this paper.

*Conflicts of interest:* None declared.

## Funding

Financial support for this work was provided by: the Deutsche Forschungsgemeinschaft (DFG)—through grant number 321869138 to J.L.; The Executive Agency for Higher Education, Research, and Innovation Funding (UEFISCDI) grant, code PN-III-P4-ID-PCE-2020-2828) to M.-G.H., and the Schweizerischer Nationalfonds (SNF) through grant number 100013\_192549 to A.L.

## Data availability

Links to data and code for preprocessing and analysis, as well as instructions on how to use this code, are provided at: [https://github.com/juergenlerner/eventnet/wiki/Coevolution-of-collaboration-and-references-to-prior-work-\(tutorial\)](https://github.com/juergenlerner/eventnet/wiki/Coevolution-of-collaboration-and-references-to-prior-work-(tutorial))

## Author contributions statement

J.L., M.-G.H., and A.L. conceived the model and the experiments, J.L. implemented and conducted the experiments, J.L. and A.L. analysed and discussed the results, J.L. and A.L. wrote the manuscript, J.L., M.-G.H., and A.L. reviewed and revised the manuscript.

## Supplementary material

[Supplementary material](#) is available online at *Journal of the Royal Statistical Society: Series A*.

## Appendix A. Separation into coauthoring and citation model

### A.1 The coauthoring model

We define the coauthoring model by focusing on the differences to the joint model defined above. We adapt the notation by using the superscript <sup>(a)</sup> whenever necessary (where ‘a’ stands for ‘authors’). The coauthoring model is for counting processes on  $\mathbb{R}_+ \times \mathcal{P}(\mathcal{I})$ , where the intensity on a set of authors  $I \in \mathcal{P}(\mathcal{I})$  is specified as

$$\lambda_t^{(a)}(I) = \bar{\lambda}_t^{(a)}(|I|) \exp\left\{\beta_0^{(a)\top} \mathbf{x}_t^{(a)}(I)\right\} \mathbf{1}\{I \subseteq \mathcal{I}_t\}. \quad (\text{A1})$$

Note that, although covariates  $\mathbf{x}_t^{(a)}(I)$  are only functions of the set of authors  $I$ , this does not preclude that these covariates can consider *past* citations. For example,  $\mathbf{x}_t^{(a)}(I)$  could count, for pairs of two different authors  $i, i' \in I$ , the number of papers that have been cited by both,  $i$  and  $i'$  in the past (this could indicate similarity of their research focus and could, therefore, increase the probability of a future collaboration); compare the precise definition of such a covariate in Section 3. All covariates  $\mathbf{x}_t^{(a)}(I)$  of the coauthoring model can be used as covariates  $\mathbf{x}_t(I, J)$  in the joint model in a

straightforward way by setting  $x_t(I, J) = x_t^{(a)}(I)$ , that is, by ignoring the second argument  $J$ .

For an observed sequence of publication events  $(t_1, j_1, I_1, J_1), \dots, (t_n, j_n, I_n, J_n)$ , the sampled partial likelihood  $\tilde{L}_t^{(a)}(\beta)$  of the coauthoring model is very similar to the one of the joint model, given in (4). The only remarkable difference is that the summation is over the sampled risk set  $\tilde{\mathcal{R}}_{t_m}^{(a)}(\mathcal{I}_{t_m}, I_m, q)$ , sampled uniformly at random from

$$\left\{ \mathcal{R} \subseteq \binom{\mathcal{I}_t}{|I|}; I \in \mathcal{R} \wedge |\mathcal{R}| = q + 1 \right\}, \quad (\text{A2})$$

instead of being sampled from (3).

## A.2 The citation model

As before, we define the citation model by focusing on the differences to the joint model defined above, where we use the superscript  $^{(c)}$  whenever necessary (where ‘c’ stands for ‘citations’). Same as for the joint model, the citation model is for counting processes on  $\mathbb{R}_+ \times \mathcal{P}(\mathcal{I}) \times \mathcal{P}(\mathcal{J})$ . The first remarkable difference to the joint model is that the intensity on a pair  $(I, J) \in \mathcal{P}(\mathcal{I}) \times \mathcal{P}(\mathcal{J})$  absorbs the set of authors  $I$  in the baseline intensity  $\tilde{\lambda}_t^{(c)}$ :

$$\lambda_t^{(c)}(I, J) = \tilde{\lambda}_t^{(c)}(I, |J|) \exp\left\{\beta_0^{(c)\top} x_t^{(c)}(I, J)\right\} \mathbf{1}\{I \subseteq \mathcal{I}_t \wedge J \subseteq \mathcal{J}_t\}. \quad (\text{A3})$$

This difference implies that the hazard ratio  $\exp\{\beta_0^{(c)\top} x_t^{(c)}(I, J)\}$  explains by which factor a possible set of references  $J$  is more or less likely than an alternative set of references  $J'$ , for a given publication by authors  $I$ . In other words, the citation model conditions on the observed set of authors of a publication event and explains which references are likely to be cited by them. The other implication is that any covariate  $x_t^{(c)}(I, J)$  that is only a function of the authors  $I$  (but ignores the references  $J$ ) would lead to a nonidentifiable parameter  $\beta_0^{(c)}$  in the citation model, since its effect is absorbed in the baseline intensity. (From another perspective, the effect of such a covariate would cancel out in the resulting likelihood function.) For instance, a covariate  $x_t^{(c)}(I, J)$  counting the number of previous papers of authors  $i \in I$  would lead to a nonidentifiable parameter, since it satisfies  $x_t^{(c)}(I, J) = x_t^{(c)}(I, J')$  for all sets of references  $J, J'$ . On the other hand, a covariate  $x_t^{(c)}(I, J)$  counting the number of those previous papers of authors  $i \in I$  that cite papers in  $J$  could be included in the citation model. All covariates of the citation model can also be included in the joint model.

The sampled partial likelihood  $\tilde{L}_t^{(c)}(\beta)$  of the citation model is very similar to the one of the joint model, given in (4). The only remarkable difference is that it uses in the summation the sampled risk set  $\tilde{\mathcal{R}}_t^{(c)}(I, \mathcal{J}_t, J, q) \subseteq \{I\} \times \binom{\mathcal{J}_t}{|J|}$  which is sampled uniformly at random from

$$\left\{ \mathcal{R} \subseteq \{I\} \times \binom{\mathcal{J}_t}{|J|}; (I, J) \in \mathcal{R} \wedge |\mathcal{R}| = q + 1 \right\}, \quad (\text{A4})$$

instead of being sampled from (3). Note that the first component of the elements in  $\tilde{\mathcal{R}}_t^{(c)}(I, \mathcal{J}_t, J, q)$  is fixed to  $I$  and only the second component is random.

## References

- Aalen O., Borgan O., & Gjessing H. (2008). *Survival and event history analysis: A process point of view*. Springer Science & Business Media.
- Acemoglu D., Akcigit U., & Kerr W. R. (2016). Innovation network. *Proceedings of the National Academy of Sciences of the United States of America*, 113(41), 11483–11488. <https://doi.org/10.1073/pnas.1613559113>
- Amati V., Lomi A., & Mascia D. (2019). Some days are better than others: Examining time-specific variation in the structuring of interorganizational relations. *Social Networks*, 57, 18–33. <https://doi.org/10.1016/j.socnet.2018.10.001>
- Amati V., Lomi A., & Mira A. (2018). Social network modeling. *Annual Review of Statistics and Its Application*, 5(1), 343–369. <https://doi.org/10.1146/statistics.2018.5.issue-1>

- Amati V., Lomi A., & Snijders T. A. (2024). A goodness of fit framework for relational event models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 187(4), 967–988. <https://doi.org/10.1093/jrssa/qnae016>
- Barabási, A.-L., Jeong, H., Nda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4), 590–614. <https://doi.org/10.1016/S0378-43710200736-7>
- Bianchi, F., Filippi-Mazzola, E., Lomi, A., & Wit, E. C. (2024). Relational event modeling. *Annual Review of Statistics and its Application*, 11(1), 297–319. <https://doi.org/10.1146/annurev-statistics-040722-060248>
- Borgan Ø., Goldstein L., & Langholz B. (1995). Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *The Annals of Statistics*, 23(5), 1749–1778. <https://doi.org/10.1214/aos/1176324322>
- Brandes U., Lerner J., & Snijders T. A. (2009). Networks evolving step by step: Statistical analysis of dyadic event data. In *Proceedings of the international conference on advances in social network analysis and mining (ASONAM)* (pp. 200–205). IEEE Computer Society.
- Breiger R. L. (1974). The duality of persons and groups. *Social Forces*, 53(2), 181–190. <https://doi.org/10.2307/2576011>
- Breslow N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30(1), 89–99. <https://doi.org/10.2307/2529620>
- Bright D., Lerner J., Sadewo G. R. P., & Whelan C. (2024). Offence versatility among co-offenders: A dynamic network analysis. *Social Networks*, 78, 1–11. <https://doi.org/10.1016/j.socnet.2023.10.003>
- Burgdorf, K., Wittek, M., & Lerner, J. (2024). Communities of style: Artistic transformation and social cohesion in Hollywood, 1930–1999. *Socius*, 10. <https://doi.org/10.1177/23780231241257334>. forthcoming.
- Burt R. S. (1992). *Structural holes: The social structure of competition*. Harvard University Press.
- Butts C. T. (2008). A relational event framework for social action. *Sociological Methodology*, 38(1), 155–200. <https://doi.org/10.1111/j.1467-9531.2008.00203.x>
- Butts C. T., Lomi A., Snijders T. A., & Stadtfeld C. (2023). Relational event models in network science. *Network Science*, 11(2), 175–183. <https://doi.org/10.1017/nws.2023.9>
- Carusi C., & Bianchi G. (2019). Scientific community detection via bipartite scholar/journal graph co-clustering. *Journal of Informetrics*, 13(1), 354–386. <https://doi.org/10.1016/j.joi.2019.01.004>
- Cox D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- Desmarais B. A., & Cranmer S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, 391(4), 1865–1876. <https://doi.org/10.1016/j.physa.2011.10.018>
- Efron B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359), 557–565. <https://doi.org/10.1080/01621459.1977.10480613>
- Efron, B. (1992). Bootstrap methods: another look at the jackknife. In S. Kotz, N.L. Johnson (Eds.), *Breakthroughs in statistics* (pp. 569–563). Springer Science & Business Media.
- Espinosa-Rada A., Bellotti E., Everett M. G., & Stadtfeld C. (2024). Co-evolution of a socio-cognitive scientific network: A case study of citation dynamics among astronomers. *Social Networks*, 78, 92–108. <https://doi.org/10.1016/j.socnet.2023.11.008>
- Faulkner R. R. (1983). *Music on demand*. Transaction Publishers.
- Ferligoj A., Kronegger L., Mali F., Snijders T. A., & Doreian P. (2015). Scientific collaboration dynamics in a national scientific system. *Scientometrics*, 104(3), 985–1012. <https://doi.org/10.1007/s11192-015-1585-7>
- Filippi-Mazzola E., & Wit E. C. (2024). A stochastic gradient relational event additive model for modelling us patent citations from 1976 to 2022. *Journal of the Royal Statistical Society Series C: Applied Statistics*, qlae023. <https://doi.org/10.1093/jrssc/qlae023>
- Fortunato S., Bergstrom C. T., Börner K., Evans J. A., Helbing D., Milojević S., Petersen A. M., Radicchi F., Sinatra R., Uzzi B., Vespignani A., Waltman L., Wang D., & Barabási A.-L. (2018). Science of science. *Science*, 359(6379), eaao0185. <https://doi.org/10.1126/science.aao0185>
- Fritz, C., De Nicola, G., Kevork, S., Harhoff, D., & Kauermann, G. (2023). Modelling the large and dynamically growing bipartite network of German patents and inventors. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186, 557–576. <https://doi.org/10.1093/jrssa/qnad009>
- Gleiser P. M., & Danon L. (2003). Community structure in jazz. *Advances in Complex Systems*, 6(04), 565–573. <https://doi.org/10.1142/S0219525903001067>
- Goldberg S. R., Anthony H., & Evans T. S. (2015). Modelling citation networks. *Scientometrics*, 105(3), 1577–1604. <https://doi.org/10.1007/s11192-015-1737-9>
- Guimera R., Uzzi B., Spiro J., & Amaral L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722), 697–702. <https://doi.org/10.1126/science.1106340>
- Hertz-Picciotto I., & Rockhill B. (1997). Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics*, 53(3), 1151–1156. <https://doi.org/10.2307/2533573>

- Hunter, D., Smyth, P., Vu, D. Q., & Asuncion, A. U. (2011). Dynamic egocentric models for citation networks. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (pp. 857–864). ACM.
- Kalbfleisch J. D., & Prentice R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60(2), 267–278. <https://doi.org/10.1093/biomet/60.2.267>
- Keogh, R. H., & Cox, D. R. (2014). *Case-control studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139094757>
- Kessler M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25. <https://doi.org/10.1002/asi.v14:1>
- Krivitsky P. N., & Handcock M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 29–46. <https://doi.org/10.1111/rssb.12014>
- Lee M., & Martin J. L. (2018). Doorway to the dharma of duality. *Poetics*, 68, 18–30. <https://doi.org/10.1016/j.poetic.2018.01.001>
- Lena J. C. (2004). Meaning and membership: Samples in rap music, 1979–1995. *Poetics*, 32(3-4), 297–310. <https://doi.org/10.1016/j.poetic.2004.05.006>
- Lerner J., Bussmann M., Snijders T. A., & Brandes U. (2013). Modeling frequency and type of interaction in event networks. *Corvinus Journal of Sociology and Social Policy*, 4(1), 3–32. <https://doi.org/10.14267/issn.2062-087X>
- Lerner J., & Hâncean M.-G. (2023). Micro-level network dynamics of scientific collaboration and impact: Relational hyperevent models for the analysis of coauthor networks. *Network Science*, 11(1), 5–35. <https://doi.org/10.1017/nws.2022.29>
- Lerner J., Indlekofer N., Nick B., & Brandes U. (2013). Conditional independence in dynamic networks. *Journal of Mathematical Psychology*, 57(6), 275–283. <https://doi.org/10.1016/j.jmp.2012.03.002>
- Lerner J., & Lomi A. (2020). Reliability of relational event model estimates under sampling: How to fit a relational event model to 360 million dyadic events. *Network Science*, 8(1), 97–135. <https://doi.org/10.1017/nws.2019.57>
- Lerner J., & Lomi A. (2022). A dynamic model for the mutual constitution of individuals and events. *Journal of Complex Networks*, 10(2), cnac004. <https://doi.org/10.1093/comnet/cnac004>
- Lerner, J., & Lomi, A. (2023). Relational hyperevent models for polyadic interaction networks. *Journal of the Royal Statistical Society: Series A*, 186, 577–600. <https://doi.org/10.1093/rjsssa/qnac012>
- Lerner J., Lomi A., Mowbray J., Rollings N., & Tranmer M. (2021). Dynamic network analysis of contact diaries. *Social Networks*, 66, 224–236. <https://doi.org/10.1016/j.socnet.2021.04.001>
- Lerner, J., Tranmer, M., Mowbray, J., & Hancean, M.-G. (2019). REM beyond dyads: Relational hyperevent models for multi-actor interaction networks arXiv arXiv:1912.07403 preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.1912.07403> Version: v1
- Lin Z., Yin Y., Liu L., & Wang D. (2023). SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data*, 10(1), 315. <https://doi.org/10.1038/s41597-023-02198-9>
- Liu L., Jones B. F., Uzzi B., & Wang D. (2023). Data, measurement and empirical methods in the science of science. *Nature Human Behaviour*, 7(7), 1046–1058. <https://doi.org/10.1038/s41562-023-01562-4>
- Liu R.-L. (2017). A new bibliographic coupling measure with descriptive capability. *Scientometrics*, 110(2), 915–935. <https://doi.org/10.1007/s11192-016-2196-7>
- Lusher D., Koskinen J., & Robins G., (Eds.) (2013). *Exponential random graph models for social networks*. Cambridge University Press.
- Martin T., Ball B., Karrer B., & Newman M. (2013). Coauthorship and citation patterns in the physical review. *Physical Review E*, 88(1), 012814. <https://doi.org/10.1103/PhysRevE.88.012814>
- Meng X., Varol O., & Barabási A.-L. (2024). Hidden citations obscure true impact in science. *PNAS Nexus*, 3(5), pgae155. <https://doi.org/10.1093/pnasnexus/pgae155>
- Merton R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810), 56–63. <https://doi.org/10.1126/science.159.3810.56>
- Mukherjee S., Romero D. M., Jones B., & Uzzi B. (2017). The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances*, 3(4), e1601315. <https://doi.org/10.1126/sciadv.1601315>
- Mulder, J., & Hoff, P. D. (2021). A latent variable model for relational events with multiple receivers arXiv arXiv:2101.05135 preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.2101.05135>. Version v3.
- Newman M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 404–409. <https://doi.org/10.1073/pnas.98.2.404>
- Newman M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(suppl\_1), 5200–5205. <https://doi.org/10.1073/pnas.0307545100>
- Pandey P. K., Singh M., Goyal P., Mukherjee A., & Chakrabarti S. (2020). Analysis of reference and citation copying in evolving bibliographic networks. *Journal of Informetrics*, 14(1), 101003. <https://doi.org/10.1016/j.joi.2019.101003>

- Park M., Leahey E., & Funk R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613(7942), 138–144. <https://doi.org/10.1038/s41586-022-05543-x>
- Perry P. O., & Wolfe P. J. (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, 75(5), 821–849. <https://doi.org/10.1111/rssb.12013>
- Petersen, A. M., Arroyave, F., & Pammolli, F. (2024). The disruption index suffers from citation inflation and is confounded by shifts in scholarly citation practice. arXiv preprint arXiv:2406.15311 non peer reviewed <https://doi.org/10.48550/arXiv.2406.15311> Version v1
- Price D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292–306. <https://doi.org/10.1002/asi.v27:5>
- Radicchi F., Fortunato S., & Castellano C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45), 17268–17272. <https://doi.org/10.1073/pnas.0806977105>
- Radicchi, F., Fortunato, S., & Vespignani, A. (2011). Citation networks. In A. Scharnhorst, K. Börner, & P. Besselaar (Eds.), *Models of science dynamics: Encounters between complexity theory and information sciences* (pp. 233–257). Springer.
- Ramasco J. J., & Morris S. A. (2006). Social inertia in collaboration networks. *Physical Review E*, 73(1), 016122. <https://doi.org/10.1103/PhysRevE.73.016122>
- Saavedra S., Reed-Tsochas F., & Uzzi B. (2009). A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457(7228), 463–466. <https://doi.org/10.1038/nature07532>
- Sarigöl E., Pfitzner R., Scholtes I., Garas A., & Schweitzer F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, 3(1), 1–16. <https://doi.org/10.1140/epjds/s13688-014-0009-x>
- Schuster M., Mitchell D., & Brown K. (2019). Sampling increases music sales: An empirical copyright study. *American Business Law Journal*, 56(1), 177–229. <https://doi.org/10.1111/ablj.v56.1>
- Seeber M., Cattaneo M., Meoli M., & Malighetti P. (2019). Self-citations as strategic response to the use of metrics for career decisions. *Research Policy*, 48(2), 478–491. <https://doi.org/10.1016/j.respol.2017.12.004>
- Simkin, M. V., & Roychowdhury, V. P. (2002). Read before you cite arXiv arXiv:cond-mat/0212043 preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.cond-mat/0212043>. Version v1.
- Snijders T. A. (2005). Models for longitudinal network data. In P. J. Carrington, J. Scott, & S. Wasserman, (Eds.), *Models and methods in social network analysis*. Cambridge University Press.
- Snijders T. A. (2017). Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and its Application*, 4(1), 343–363. <https://doi.org/10.1146/statistics.2017.4.issue-1>
- Snijders T. A., Lomi A., & Torló V. J. (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks*, 35(2), 265–276. <https://doi.org/10.1016/j.socnet.2012.05.005>
- Tang J., Zhang J., Yao L., Li J., Zhang L., & Su Z. (2008). Arnetminer: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 990–998).
- Therneau T. M. (2015). *A package for survival analysis in S*. <https://CRAN.R-project.org/package=survival>. Version no. 2.38.
- Therneau T. M., & Grambsch P. M. (2013). *Modeling survival data: Extending the Cox model*. Springer Science & Business Media.
- Tumminello M., Micciche S., Lillo F., Piilo J., & Mantegna R. N. (2011). Statistically validated networks in bipartite complex systems. *PLoS One*, 6(3), e17994. <https://doi.org/10.1371/journal.pone.0017994>
- Uzaheta A., Amati V., & Stadtfeld C. (2023). Random effects in dynamic network actor models. *Network Science*, 11(2), 249–266. <https://doi.org/10.1017/nws.2022.37>
- Uzzi B., Mukherjee S., Stringer M., & Jones B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157), 468–472. <https://doi.org/10.1126/science.1240474>
- Uzzi B., & Spiro J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, 111(2), 447–504. <https://doi.org/10.1086/432782>
- White H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2), 87–108. [https://doi.org/10.1002/\(ISSN\)1532-2890](https://doi.org/10.1002/(ISSN)1532-2890)