

Epidemiological Models of Topic Propagation in Co-authorship Networks: Degree-Block Models Versus Homogeneous Mixing

Ismael Robles-Martínez

irobles@cua.uam.mx

Universidad Autónoma Metropolitana

Jorge Cruz-Trani

Universidad Autónoma Metropolitana

Daniela Aguirre-Guerrero

Universidad Autónoma Metropolitana

Ricardo Marcelín-Jiménez

Universidad Autónoma Metropolitana

Research Article

Keywords: co-authorship networks, topic propagation, epidemiological models, diffusion of ideas, degree-block approximation, basic reproduction number, eigenvector localization, identifiability

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11254722/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Epidemiological Models of Topic Propagation in Co-authorship Networks: Degree-Block Models Versus Homogeneous Mixing

Ismael Robles-Martínez^{1*}, Jorge Cruz-Trani²,
Daniela Aguirre-Guerrero¹, Ricardo Marcelín-Jiménez²

¹*Department of Applied Mathematics and Systems, Universidad
Autónoma Metropolitana, Campus Cuajimalpa, Mexico City, Mexico.

²Department of Electrical Engineering, Universidad Autónoma
Metropolitana, Campus Iztapalapa, Mexico City, Mexico.

*Corresponding author(s). E-mail(s): irobles@cua.uam.mx;
Contributing authors: trani.jorge510@gmail.com;
daguirre@cua.uam.mx; calu@xanum.uam.mx;

Abstract

Epidemic models have described the spread of research topics among authors since the 1960s, mostly under homogeneous mixing. The conference version of this work fitted homogeneous Susceptible–Infected, Susceptible–Infected–Susceptible and Susceptible–Infected–Recovered models to the adoption of eight Computer Science topics in Scopus and attributed its largest residuals to the structure of co-authorship networks. We test that conjecture with degree-block models built from the empirical degree correlations of each topic co-authorship network, fitted to the same series as the homogeneous models with the same initial condition and number of parameters. By construction, this network contains only the eventual adopters of a topic, and its edges are created by the publications that define adoption. We report three results. First, the degree-block models fit better in only 3 of 24 comparisons. Second, between 64% and 95% of the authors enter a topic without co-signing with an already active author, so the network cannot be the substrate of the contagion, and its spectral radius, which enters the network reproduction number, can be set by the author list of a single article. Third, in three of the four growing topics the recovery rate, and with it the reproduction

number, follows the bound of its search interval in 11 of 12 fits, whereas growth rates, including the adoption growth rate that measures virality, do not. Testing contagion through collaboration requires a network that precedes adoption and includes non-adopters.

Keywords: co-authorship networks, topic propagation, epidemiological models, diffusion of ideas, degree-block approximation, basic reproduction number, eigenvector localization, identifiability

1 Introduction

Since Goffman and Newill (1964) described the transmission of ideas among scientists as an epidemic process, compartmental models have been fitted, under homogeneous mixing, to the number of authors who adopt an idea or enter a field (Goffman 1966; Bettencourt et al. 2006, 2008). Authors who have not yet published on a topic are susceptible, authors who publish on it are infected and authors who abandon it are recovered. The Susceptible–Infected (SI), Susceptible–Infected–Susceptible (SIS) and Susceptible–Infected–Recovered (SIR) models (Keeling and Rohani 2008) describe the dynamics through a transmission rate β and a recovery rate γ . They summarize the potential of a contagion to spread by the *basic reproduction number* R_0 , the average number of secondary cases generated by an infected individual in a completely susceptible population; the contagion can invade the population when $R_0 > 1$.

Scientific collaboration, however, takes place in co-authorship networks, whose nodes are authors and in which two authors are linked if they have published together. These networks have heavy-tailed degree distributions: most authors have few co-authors, while a few hubs collaborate with many (Newman 2004; Barabási 2016). In network epidemiology, degree heterogeneity and the correlations between the degrees of neighbouring nodes change the epidemic threshold (Pastor-Satorras and Vespignani 2001; Pastor-Satorras et al. 2015). The degree-block, or heterogeneous mean-field, approximation for correlated networks (Boguñá et al. 2003) captures both features

through a contact matrix whose spectral radius sets the threshold and enters the reproduction number.

Contact structure has been added to epidemic models of ideas in two ways. Kiss et al. (2010) fitted epidemic models on a network of citation flows between subject categories, defined independently of the topic, and obtained a better fit than with equal weights; Yue et al. (2019) formulated a degree-based model on the institutional collaboration network of the graphene field. A separate line of work measures, without epidemic models, whether collaboration precedes adoption: Venturini et al. (2024) showed that the probability of switching to a topic grows with the number of earlier collaborators already active in it. Our conference paper (Cruz-Trani et al. 2026), hereafter the conference version, fitted homogeneous models to eight topics, attributed the largest residuals of the models it selected, those of graph neural networks and tiny machine learning, to network structure and announced degree-block models as the next step. Here we take that step.

The question of this article is whether the topic co-authorship network, formed by the authors of a topic and their co-authorships on it, describes the adoption of the topic better than the homogeneous model when it enters the models through the degree-block approximation, as the conference version conjectured. Throughout, the *virality* of a topic is the speed at which it recruits new authors, a property of the topic rather than of its articles and distinct from the structural virality of diffusion cascades (Goel et al. 2016). The conference version measured it through β and R_0 ; we keep the former, as the adoption growth rate defined in Section 3.4, because a reproduction number cannot be inferred from a growth curve without further assumptions (Wallinga and Lipsitch 2007).

We make three contributions, presented in this order in Section 4. First, we fit degree-block SI, SIS and SIR models for correlated networks (Boguñá et al. 2003; Pastor-Satorras et al. 2015), built from the empirical degree correlations of each topic

co-authorship network, and compare them with the homogeneous models on the same observed series, with the same initial condition and the same number of parameters. Second, we show what the topic co-authorship network represents: it cannot be the substrate of the contagion, and its degree-block description is dominated by very small structures, such as the authors of a single article. Third, we show that the network reproduction number combines two quantities that these data cannot supply, the spectral radius and, in three topics, the recovery rate, which follows the bound of its search interval, whereas growth rates, including the adoption growth rate, do not.

This article extends the conference version, which fitted only homogeneous models and was presented at the Sixth International Symposium on Complex Systems (ISCS 2026). The new material comprises the documentation of the data and of their quality, including the completion of truncated author lists; the structural characterization of the eight topic co-authorship networks; the degree-block SI, SIS and SIR models on correlated networks and their reproduction number; the comparison of the two model families on the same series; the analysis of entry into a topic and of the localization of the leading eigenvector of the contact matrix; and the analysis of the identifiability of the recovery rate and of R_0 . Section 5 states how these results refine the conference version. The definitions of the epidemiological states and of the homogeneous models (Sections 3.1 and 3.2) reuse revised text of the conference version.

2 Data and topic co-authorship networks

2.1 Topics and records

We analyse the eight Computer Science topics of the conference version: adaptive neuro-fuzzy inference systems (ANFIS), the Business Process Execution Language (BPEL), distributed hash tables (DHT), online analytical processing (OLAP), Gaussian processes (GP), graph neural networks (GNN), tiny machine learning (TinyML) and quantum machine learning (QML). The conference version grouped the first four

as topics in saturation and the last four as trending topics; we keep these a priori groups only to order tables and figures, and no result depends on them. Each topic is named by a distinctive term or acronym, and much of the output of the field appears in conference proceedings indexed by Scopus, so yearly series capture fast changes; the topics range from 85 (TinyML) to 10066 (GP) records up to 2024. Two further topics, You Only Look Once object detection and self-supervised learning, were exported together with these eight but are not analysed here.

The records were exported from the web interface of the Scopus database (Elsevier), restricted to journal articles and conference papers. Each record contains the publication year, the title, the abstract, the author keywords, the index keywords and the Scopus identifiers of the authors, which we use to identify authors; the BPEL export contains no abstracts or keywords. The exact query strings and the export dates were not preserved. The counts for 2025 are incomplete, so we analyse the records published up to 2024. The exports of GP, GNN and TinyML listed at most ten authors per record; the full author lists of the affected records up to 2024 (36 in GP and 15 in GNN; the two TinyML records with ten authors are from 2025) were retrieved by record identifier through the Scopus Search application programming interface on 3 October 2026. That interface returns at most 100 authors per record, so the author list of one GP record, a 2022 consortium article with more than 100 authors, remains incomplete.

The queries used the terms `anfis`, `bpel`, `distributed hash table`, `olap`, `gaussian process`, `gnn`, `tinyml` and `quantum machine learning`. In the exports of ANFIS, OLAP, GP, GNN and TinyML, the term appears in the keywords of at least 99.9% of the records but in the title or abstract of only 71.6% to 89.0%, so these exports are consistent with queries restricted to the keyword fields. The authors of a topic are thus the authors of the records labelled with its term. The BPEL export cannot be checked. The QML and GNN queries were not restricted to the topic phrase:

231 53.6% of the QML records and 18.4% of the GNN records do not contain it in any field,
232 and their series start in 1999 and 1983. These two exports therefore contain records
233 outside the topic. In DHT, 4.8% of the records do not contain the term in any field.
234
235

237 2.2 Network construction

238
239 For each topic, we build an undirected co-authorship network whose nodes are the
240 authors of its records and in which two authors are linked if they have co-authored
241 at least one record of the topic up to 2024. Each record is thus projected onto a
242 clique of its authors. The topology is static: it accumulates all co-authorships of the
243 period, and what evolves year after year is the epidemiological state of each author
244 (Section 3.1). The models use the unweighted topology, so the degree k of an author
245 is their number of distinct co-authors on the topic; authors with only single-authored
246 records do not appear in the network. The conference version described a yearly,
247 weighted network; here the network is static and unweighted because the degree-
248 block formulation requires a single degree sequence. The networks are built with the
249 NetworkX library. By construction, the topic co-authorship network contains only the
250 eventual adopters of the topic, its edges are created by the publications that define
251 adoption, and the degree of an author includes collaborations that occur after their
252 adoption.
253
254
255

264 2.3 Structural characterization

265
266 Let N be the number of nodes, N_k the number of nodes of degree k and $P(k) = N_k/N$
267 the degree distribution, with moments $\langle k^n \rangle = \sum_k k^n P(k)$. Degree correlations are
268 described by the conditional probability $P(k'|k)$ that an edge leaving a node of degree
269 k reaches a node of degree k' . If $E_{kk'}$ denotes the number of edges between the blocks
270
271
272
273
274
275
276

of nodes of degree k and k' (counted twice when $k = k'$), then

$$P(k'|k) = \frac{E_{kk'}}{kN_k}, \quad \sum_{k'} P(k'|k) = 1. \quad (1)$$

In an uncorrelated network $P(k'|k)$ does not depend on k and reduces to $k'P(k')/\langle k \rangle$ (Dorogovtsev and Mendes 2022; Newman 2018). We define the *contact matrix* $\mathbf{C}$ with entries

$$C_{kk'} = k P(k'|k), \quad (2)$$

the expected number of neighbours of degree k' of a node of degree k , and its spectral radius $\rho(\mathbf{C})$, which controls the epidemic threshold of the degree-block models (Section 3.3).

Table 1 summarizes the eight networks and Fig. 1 shows their degree distributions. The networks range from 291 (TinyML) to 22443 (GP) authors, with mean degrees between 3.5 and 5.3 and maximum degrees between 10 and 107. All degree distributions are right-skewed: most authors have a few co-authors, about the size of a typical author list, and a few hubs have many more. The spectral radius $\rho(\mathbf{C})$ exceeds its uncorrelated value $\langle k^2 \rangle / \langle k \rangle$ in every network, by factors between 1.21 (BPEL) and 5.85 (GP). The networks are fragmented: they have between 69 (TinyML) and 3451 (GP) connected components, and the largest connected component (LCC) contains between 2.3% (GNN) and 26.6% (GP) of the authors. In ANFIS, BPEL and DHT the LCC holds at most 5.1% of the authors, unlike the co-authorship networks of the fields studied by Bettencourt et al. (2009), which develop a giant component as the fields mature.

Table 1 Structure of the eight topic co-authorship networks and entry into each topic

Topic	Since	Records	N	M	$\langle k \rangle$	$k_{\max}$	n_b	$\frac{\langle k^2 \rangle}{\langle k \rangle}$	$\rho(\mathbf{C})$	LCC (%)	Entry (%)
ANFIS	1998	3878	9798	17241	3.52	64	30	5.39	9.71	3.98	22.2
BPEL	2004	1079	2212	3925	3.55	59	20	5.45	6.60	5.06	26.4
DHT	1996	1805	3950	7273	3.68	35	24	5.98	23.09	4.43	25.2
OLAP	1983	2014	3907	8626	4.42	68	39	9.27	20.85	19.32	31.9
GP	1970	10066	22443	55717	4.97	107	59	16.75	98.00	26.58	36.3
GNN	1983	1131	4410	10128	4.59	26	22	6.40	22.00	2.34	9.4
TinyML	2020	85	291	519	3.57	10	9	4.29	5.31	3.78	5.2
QML	1999	2856	8474	22531	5.32	49	41	8.84	18.91	17.10	25.4

Since: year of the first record; Records: articles and conference papers up to 2024; N and M : numbers of nodes (authors) and edges; $\langle k \rangle$ and $k_{\max}$: mean and maximum degree; n_b : number of degree blocks; $\rho(\mathbf{C})$: spectral radius of the contact matrix, equal to $\langle k^2 \rangle / \langle k \rangle$ in an uncorrelated network; LCC: share of authors in the largest connected component; Entry: share of the authors entering the topic after its first year who, in their year of entry, co-sign with an author already active in it (Section 4.2).

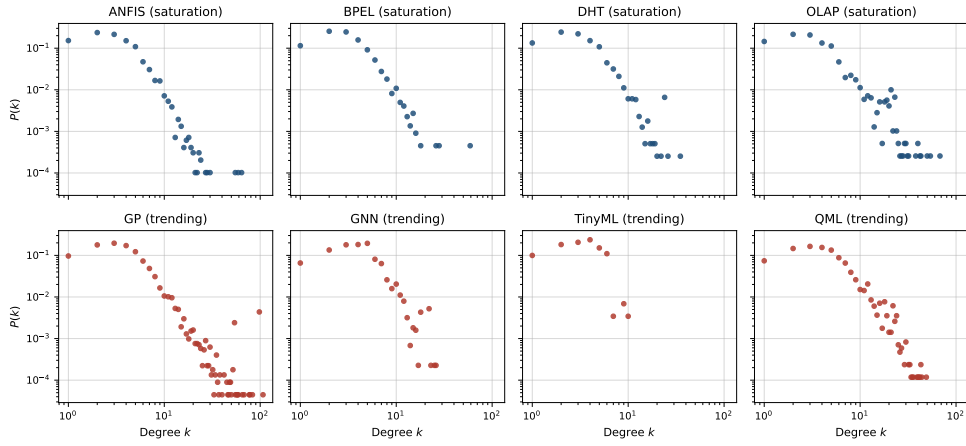


Fig. 1 Degree distributions of the eight topic co-authorship networks. Log-log plots of $P(k)$ for ANFIS, BPEL, DHT and OLAP (top row) and GP, GNN, TinyML and QML (bottom row). Panel titles and colours mark the a priori groups of the conference version (Section 2.1).

3 Models

3.1 Epidemiological states

As in the conference version, the state of each author in each year is obtained from the years in which they publish on the topic. An author is susceptible until their first record. In the SI model, an author is infected from the year of their first record onwards; in the SIS model, an author is infected in the years in which they publish on

the topic and susceptible otherwise; and in the SIR model, an author is infected from their first to their last record up to 2024 and recovered afterwards, the definition of [Goffman \(1966, 1971\)](#). Each model is therefore fitted to a different observed series: the cumulative number of adopters (SI), the number of authors active in each year (SIS) and the number of authors between their first and last record (SIR). The same three series are used for the homogeneous and the degree-block models.

3.2 Homogeneous models

Let N be the number of authors and $s(t)$, $i(t)$ and $r(t)$ the proportions of susceptible, infected and recovered authors. Under homogeneous mixing ([Keeling and Rohani 2008](#)), the dynamics are

$$\text{SI: } \frac{di}{dt} = \beta(1-i)i, \quad (3)$$

$$\text{SIS: } \frac{di}{dt} = \beta(1-i)i - \gamma i, \quad (4)$$

$$\text{SIR: } \frac{ds}{dt} = -\beta si, \quad \frac{di}{dt} = \beta si - \gamma i, \quad \frac{dr}{dt} = \gamma i, \quad (5)$$

where β is the transmission rate (in year^{-1}), which combines the rate of contacts and the probability that a contact transmits the topic, and γ is the recovery, or abandonment, rate. For the SIS and SIR models $R_0 = \beta/\gamma$; the SI model has no recovery and no finite R_0 .

3.3 Degree-block models on correlated networks

To incorporate the network structure, we adopt the degree-block approximation ([Pastor-Satorras and Vespignani 2001](#); [Boguñá et al. 2003](#); [Barrat et al. 2008](#); [Pastor-Satorras et al. 2015](#)), which partitions the nodes into blocks of nodes with the same degree k and assumes that nodes in the same block are statistically equivalent. Let $s_k(t)$, $i_k(t)$ and $r_k(t)$ be the proportions of susceptible, infected and recovered nodes

415 in the block of degree k . The population quantities are recovered as

416

417

418

419

420

421

422 and analogously for $s(t)$ and $r(t)$. The infection pressure on a susceptible node of

423

424

425 infected node:

426

427

428

429

430

431

432

433 The excess factor $e_{k'}$ accounts for the fact that, in models without reinfection, one

434

435 of the k' edges of an infected node is the one through which it was infected. Since

436

437

438 $P(k'|k)$ is computed from the observed edges through Eq. (1), the formulation retains

439

440 the degree correlations of each network. Replacing the homogeneous contact term $s i$

441

442 by $s_k \theta_k$ and the average number of contacts by k , the degree-block models read, for

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

where β is now a transmission rate per edge. Linearizing the SIS and SIR models

around the disease-free state shows that the infection grows if and only if $\beta\rho > \gamma$, so

$$i(t) = \sum_k P(k) i_k(t), \quad I(t) = N i(t), \quad (6)$$

$$\theta_k(t) = \sum_{k'} P(k'|k) e_{k'} i_{k'}(t), \quad e_{k'} = \begin{cases} (k' - 1)/k' & \text{for SI and SIR,} \\ 1 & \text{for SIS.} \end{cases} \quad (7)$$

$$\text{SI: } \frac{di_k}{dt} = \beta k (1 - i_k) \theta_k, \quad (8)$$

$$\text{SIS: } \frac{di_k}{dt} = \beta k (1 - i_k) \theta_k - \gamma i_k, \quad (9)$$

$$\text{SIR: } \frac{ds_k}{dt} = -\beta k s_k \theta_k, \quad \frac{di_k}{dt} = \beta k s_k \theta_k - \gamma i_k, \quad \frac{dr_k}{dt} = \gamma i_k, \quad (10)$$

that the basic reproduction number is (Boguñá et al. 2003; Pastor-Satorras et al. 2015)

$$R_0 = \frac{\beta}{\gamma} \rho, \quad \rho = \begin{cases} \rho(\mathbf{C}) & \text{for SIS,} \\ \rho(\mathbf{C}_{\text{SIR}}), \quad \mathbf{C}_{\text{SIR}} = \mathbf{C} \text{diag}\left(\frac{k'-1}{k'}\right) & \text{for SIR.} \end{cases} \quad (11)$$

The network enters R_0 only through a spectral property of the contact matrix. For uncorrelated networks $\rho(\mathbf{C}) = \langle k^2 \rangle / \langle k \rangle$ and $\rho(\mathbf{C}_{\text{SIR}}) = (\langle k^2 \rangle - \langle k \rangle) / \langle k \rangle$, the classical heterogeneous mean-field thresholds (Pastor-Satorras and Vespignani 2001; Pastor-Satorras et al. 2015); the factor $\langle k^2 \rangle / \langle k \rangle$ proposed in the conference version is this uncorrelated case, which ρ generalizes.

3.4 Parameter estimation

For each topic and model, we estimate β and, for the SIS and SIR models, γ by non-linear least squares (Boyd and Vandenberghe 2018),

$$\hat{\Theta} = \arg \min_{\Theta} \sum_{j=1}^m \left[I_{\text{obs}}(t_j) - N i(t_j; \Theta) \right]^2, \quad (12)$$

where Θ denotes the fitted parameters, $I_{\text{obs}}(t_j)$ is the observed series of Section 3.1 in year t_j , m the number of observed years, and N the number of distinct authors of the topic for the homogeneous models and the number of network nodes $\sum_k N_k$ for the degree-block models, which excludes only the authors with no co-authored record (for instance, 3907 against 4010 authors in OLAP). The problem is solved with the Trust Region Reflective algorithm, with $\beta \in [10^{-10}, 10]$ and $\gamma \in [10^{-2}, 10]$ (for the degree-block SI model, $\beta \in [0, 10]$). The homogeneous models are integrated with the fourth-order Runge–Kutta (RK4) method (Ascher and Petzold 1998) and a step of one year. The degree-block models are integrated with RK4 and a step of 0.1 years and then aggregated through Eq. (6); observations at non-consecutive years are compared with

the trajectories by linear interpolation. Both families start from the observed initial condition, the proportion of infected authors in the first year of the series, which the degree-block models assign uniformly to all degree blocks. An earlier implementation estimated this proportion as a free parameter of the degree-block models; the optimizer could then drive it to the equivalent of a small fraction of an author, consistent with a fast mode localized on a few blocks that grows from an infinitesimal seed and can mimic a late take-off of the series. Fixing it at its observed value gives both families the same initial condition and the same number of parameters; besides the contact structure, they differ only in the integration step and in the population, by 0.7% to 2.6%. The fits use information up to 2024, including the population of the SI model, so they describe the observed curves.

The goodness of fit is measured by the normalized root mean square error (NRMSE), the root mean square of the residuals divided by the range of the observed series. The two families are compared with the Akaike information criterion for least squares, $AIC = m \ln(RSS/m) + 2n_p$, where RSS is the residual sum of squares and n_p the number of fitted parameters. The difference $\Delta AIC = AIC_{\text{block}} - AIC_{\text{hom}}$ is used descriptively. With equal numbers of parameters it equals $m \ln(RSS_{\text{block}}/RSS_{\text{hom}})$. The yearly series are serially correlated, and cumulative for the SI model, which inflates the magnitude of ΔAIC and understates uncertainty (King et al. 2015): the residuals of the homogeneous SI fits have lag-1 autocorrelations between 0.81 and 0.97 in the seven topics other than TinyML. The 24 comparisons, from eight purposively chosen topics and three series built from the same records, are not independent. We attach no significance to the counts and read only the sign and the order of magnitude of the differences, taking $|\Delta AIC| \leq 2$ as not discriminating. The approximate 95% confidence intervals (CIs) of R_0 in Appendix A, from the Jacobian of the residuals and the delta method, are descriptive for the same reasons.

We measure the virality of a topic by its *adoption growth rate*, the transmission rate β of the homogeneous SI model of Eq. (3) fitted, as in the conference version, to the cumulative number of adopters (homogeneous SI rows of Appendix A). While adopters are few, β is their per-capita growth rate. It is a whole-series rate that averages over long periods of sparse early records, since 1970 in GP and 1983 in GNN, so it depends on the start of the series, and we do not compare it across topics. All tables, figures and numbers of this article are produced by an open-source pipeline written in Python (Robles-Martínez et al. 2026).

4 Results

4.1 Degree-block versus homogeneous models

Figs. 2 and 3 show the homogeneous and degree-block fits of the three models, and Appendix A lists all estimated parameters. Each column of the figures is a different observed series (Section 3.1), so errors are compared only within a column, between the two families fitted to the same series. In BPEL, DHT and OLAP the yearly number of new adopters has fallen since its maximum and the activity series peaked between 2009 and 2012, whereas ANFIS still recruited between 722 and 879 new authors per year in 2019–2024. The SIR model, which allows a permanent abandonment of the topic, follows the rise and decline of its series, whereas the solutions of the SIS model are monotone: the homogeneous SIS fits of BPEL and DHT degenerate, with β at the upper bound of its search interval and NRMSE 0.2852 and 0.2154. In GP, GNN, TinyML and QML, the adoption and activity series are still growing in 2024.

Table 2 compares the two families on each series. Both start from the observed initial condition and have the same number of fitted parameters, the entries of $P(k'|k)$ being computed from the network rather than fitted, so ΔAIC reflects only the ratio of their residual sums of squares. The degree-block model built from the empirical $P(k'|k)$ of the topic co-authorship network fits better in 3 of the 24 comparisons, all in ANFIS

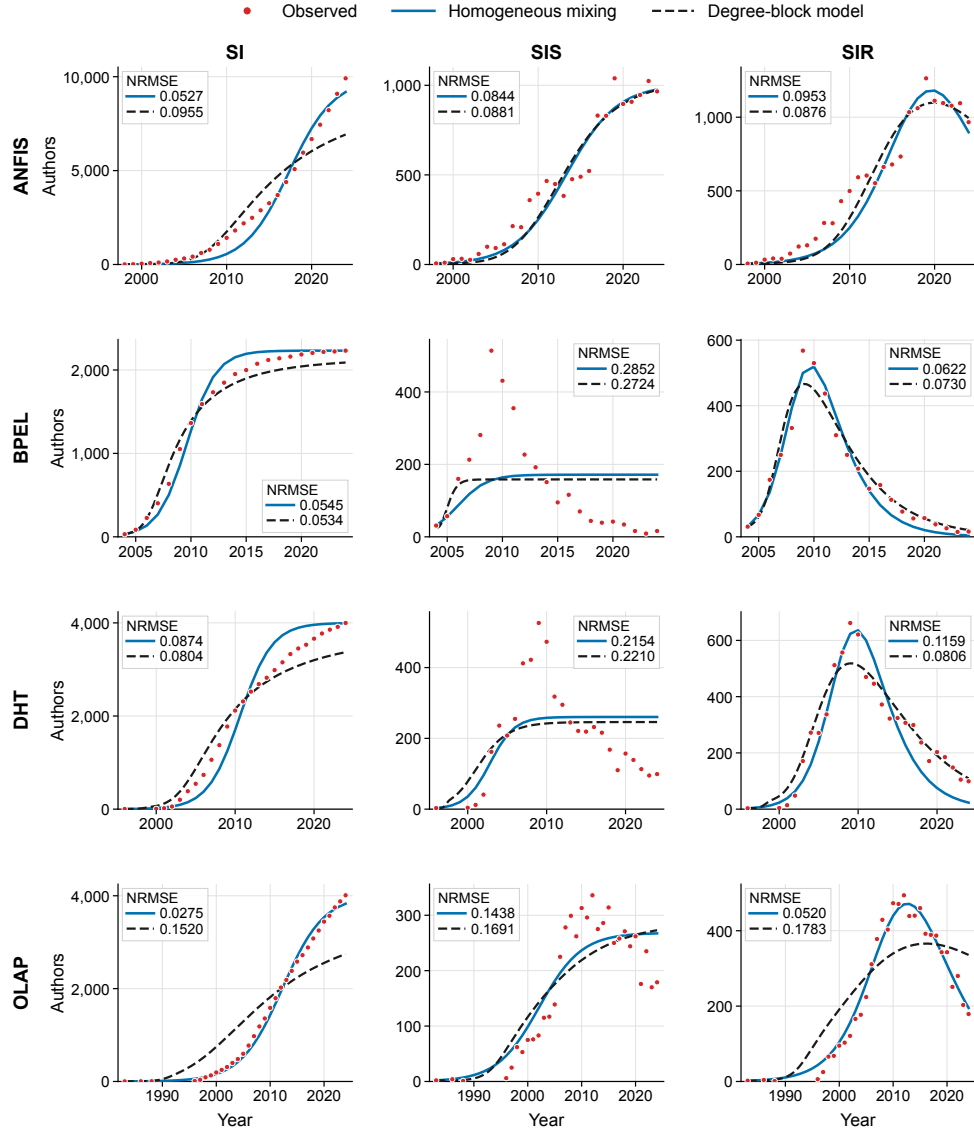


Fig. 2 Homogeneous and degree-block fits for ANFIS, BPEL, DHT and OLAP. Columns: SI, SIS and SIR models. Red dots are the observed series: cumulative adopters (SI), authors active in each year (SIS) and authors between their first and last record (SIR). Solid lines are the homogeneous fits, which reproduce those of the conference version, and dashed lines the degree-block fits aggregated with Eq. (6). Boxes give the NRMSE of both fits.

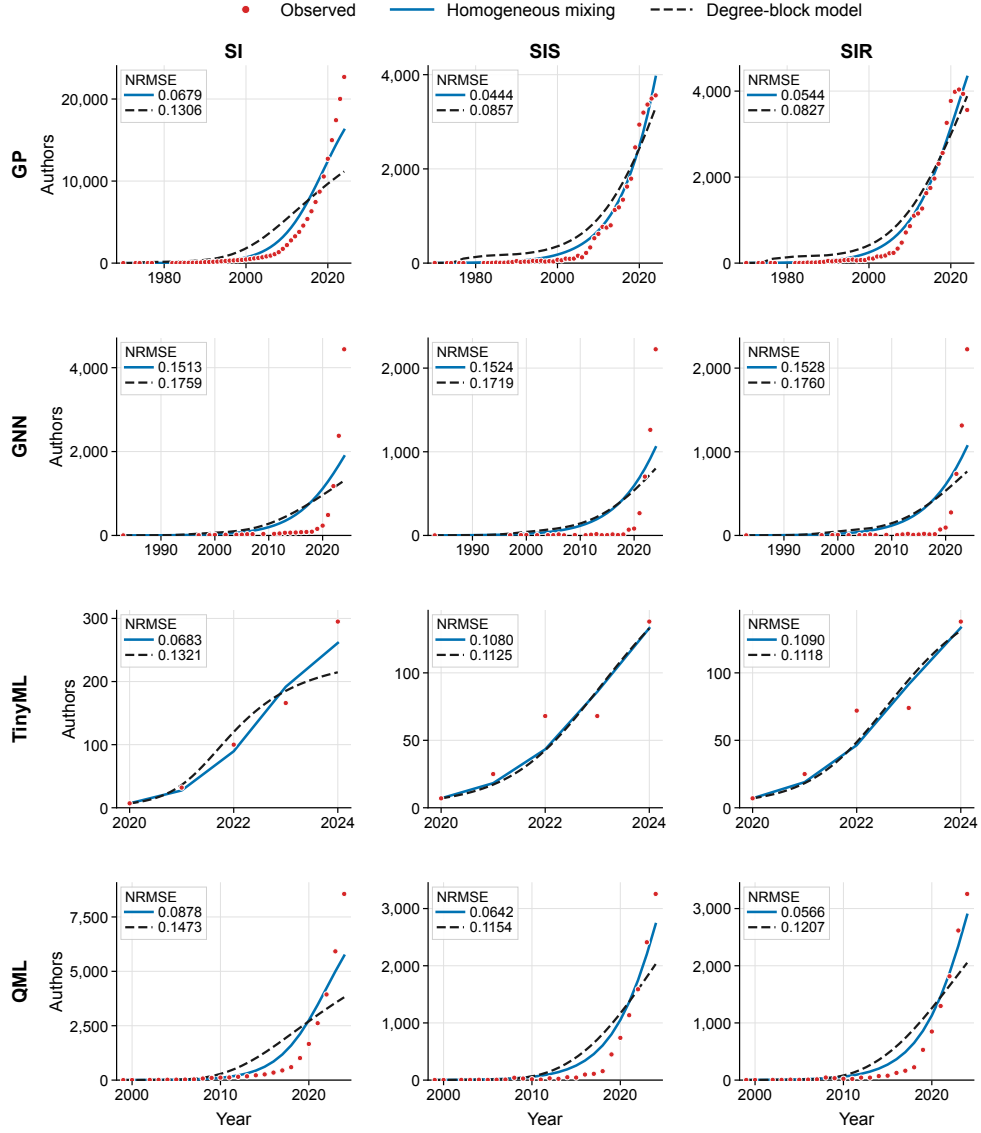


Fig. 3 Homogeneous and degree-block fits for GP, GNN, TinyML and QML. Columns, observed series, line styles and boxes as in Fig. 2; the homogeneous fits of GP and GNN differ slightly from those of the conference version because of the completed author lists (Section 2.1). TinyML has only five yearly observations (2020–2024).

and DHT, and by more than 10 points in one (DHT, SIR); the homogeneous model fits better in 16, markedly so in OLAP, GP and QML. Four of the five comparisons within 2 points are uninformative: TinyML has five observations and the homogeneous SIS fits of BPEL and DHT are degenerate. Excluding QML and GNN, whose exports contain records outside the topic, the counts are 3 and 10 of 18. The degree-block models do not reduce the largest residuals of the conference version: they fit worse in the three series of GNN and, in TinyML, worse in the SI series, while the other two comparisons do not discriminate.

Table 2 Homogeneous versus degree-block fits on the same observed series

Topic	SI (cumulative adopters)			SIS (active authors)			SIR (first to last record)		
	hom.	block	ΔAIC	hom.	block	ΔAIC	hom.	block	ΔAIC
ANFIS	0.0527	0.0955	+32.1	0.0844	0.0881	+2.3	0.0953	0.0876	-4.5
BPEL	0.0545	0.0534	-0.8	0.2852	0.2724	-1.9	0.0622	0.0730	+6.7
DHT	0.0874	0.0804	-4.3	0.2154	0.2210	+1.3	0.1159	0.0806	-18.9
OLAP	0.0275	0.1520	+109.5	0.1438	0.1691	+10.4	0.0520	0.1783	+78.9
GP	0.0679	0.1306	+62.8	0.0444	0.0857	+63.2	0.0544	0.0827	+40.2
GNN	0.1513	0.1759	+7.2	0.1524	0.1719	+5.8	0.1528	0.1760	+6.8
TinyML	0.0683	0.1321	+6.6	0.1080	0.1125	+0.4	0.1090	0.1118	+0.3
QML	0.0878	0.1473	+25.9	0.0642	0.1154	+29.4	0.0566	0.1207	+37.9

NRMSE of the homogeneous (hom.) and degree-block (block) fits and $\Delta\text{AIC} = \text{AIC}_{\text{block}} - \text{AIC}_{\text{hom}}$, descriptive (Section 3.4); values below -2 favour the degree-block model and values above +2 the homogeneous model. Both families start from the observed initial condition and have the same number of fitted parameters. TinyML has only five observations.

4.2 What the topic co-authorship network represents

The edges of a topic co-authorship network are created by the publications that define adoption and include collaborations that occur after an author has adopted the topic (Section 2.2). We draw two separate consequences.

First, the network cannot be the substrate of the contagion. The networks are fragmented (Section 2.3), and most authors do not enter a topic through them. We classify each author who publishes on a topic for the first time in year t as entering *with an active co-author* if at least one of their co-authors in year t had already

published on the topic in an earlier year; authors who enter in the first year of a topic are excluded. Between 63.7% (GP) and 94.8% (TinyML) of the entering authors enter without an active co-author (Table 1 gives the complementary shares). The measure only sees co-signatures on records of the topic in the year of entry: earlier collaborations on other subjects, collaborations on records not labelled with the term of the topic and teams that enter a topic together count as entry without an active co-author. Moreover, keyword-restricted queries retrieve only part of the literature of a topic, so the networks are thinned samples; their fragmentation and the share of entry without an active co-author are upper bounds. Neither observation bears directly on the comparison of Section 4.1, because the degree-block approximation is a mean-field description that does not see components or paths.

Second, the degree-block description of each network is dominated by very small structures, often the authors of a single article. Since $\mathbf{C}$ is non-negative, $\rho(\mathbf{C}) \geq \max_k kP(k|k)$: an article with n authors that forms an isolated clique of degree $n - 1$ forces $\rho(\mathbf{C}) \geq n - 1$ when no other author has degree $n - 1$. In DHT, the 26 authors of degree 24 (25 of them from one article) have $P(24|24) = 0.96$, so that $\rho(\mathbf{C}) \approx 24 \times 0.96 \approx 23$, and their block carries 99% of the weight of the leading eigenvector of $\mathbf{C}$, measured by its squared normalized components. In GP, the 99 distinct authors retrieved for the consortium article, 98 of degree 98 and one of degree 107, carry practically all the weight and set $\rho(\mathbf{C}) = 98.00$. In GNN, the 23 authors of a 2024 article form an isolated clique, with $P(22|22) = 1$, that carries all the weight and sets $\rho(\mathbf{C}) = 22.00$, although even in the model it cannot transmit the topic to the rest of the network. In all eight networks, at most two degree blocks carry at least half of the weight of the leading eigenvector. This parallels the eigenvector localization described for heterogeneous networks, on a hub or on a small dense subgraph (Goltsev et al. 2012; Pastor-Satorras and Castellano 2016); here the dense subgraphs are author lists. The localization is a property of projecting each article, without weights, onto a clique

of its authors, so that a single article with many authors can set $\rho(\mathbf{C})$, which enters the network R_0 ; such projections can misrepresent the dynamics on the underlying author–article network (Lee et al. 2021). That the degree-block models seldom improve on the homogeneous ones (Section 4.1) is consistent with this: their threshold and fastest mode are set by a few blocks, often the authors of one article, a negligible fraction of the adopters whose curve is fitted.

4.3 Identifiability of R_0

Because the network enters the reproduction number only through ρ , the product $R_0 = (\beta/\gamma)\rho$ inherits two quantities that these data cannot supply: ρ , which a single large group of co-authors can set (Section 4.2), and γ , which the search bound sets in GP, GNN and QML. What the homogeneous fits do identify is a growth rate.

The recovery rate reaches the lower bound of its search interval, $\gamma = 0.01$, in 11 of the 12 SIS and SIR fits of GP, GNN and QML, homogeneous and degree-block, and in no other fit; the exception is the homogeneous SIR fit of GP ($R_0 = 2.71$). When the bound is lowered to 10^{-3} and 10^{-4} , γ follows it in all 11 fits and, at 10^{-4} , R_0 is multiplied by 93 to 97, while the NRMSE decreases by less than 3% in the homogeneous fits and by less than 5% in the degree-block fits, and the growth rate $\beta - \gamma$ of the homogeneous fits is stable (Table 3). This follows from the model: the homogeneous SIS equation is a logistic equation with rate $\beta - \gamma$ and capacity $N(1 - \gamma/\beta)$, so only these two quantities are identified. When the series shows no plateau below N , the fit pushes γ towards zero and $R_0 = 1 + (\beta - \gamma)/\gamma$ towards infinity while $\beta - \gamma$ stays fixed. A plausible reason for the absence of a plateau is that the models assume a closed population, whereas growing topics recruit authors from a pool that keeps growing; a closed model can reproduce sustained growth only by suppressing recovery. That a growth rate is identified while a reproduction number requires further assumptions is well known in epidemiology (Wallinga and Lipsitch 2007), and Bettencourt et al.

(2006) found that their data constrained only the sum of the recovery and exit rates. Where no parameter lies at a bound, R_0 depends on the model: in DHT it ranges from 2.03 (homogeneous SIR) to 13.18 (degree-block SIR; Appendix A). The adoption growth rate, in contrast, involves neither γ nor ρ .

Table 3 Sensitivity of the growth rate and of R_0 to the lower bound of γ

Topic	Model	$\beta - \gamma$ (year ⁻¹)			R_0			NRMSE
		10 ⁻²	10 ⁻³	10 ⁻⁴	10 ⁻²	10 ⁻³	10 ⁻⁴	
GP	SIS	0.137	0.137	0.137	14.7	138	1367	0.0443–0.0444
GNN	SIS	0.177	0.177	0.177	18.7	178	1767	0.1518–0.1524
GNN	SIR	0.178	0.177	0.177	18.8	178	1774	0.1515–0.1528
QML	SIS	0.305	0.304	0.304	31.5	305	3045	0.0634–0.0642
QML	SIR	0.309	0.308	0.308	31.9	309	3082	0.0549–0.0566

Homogeneous SIS and SIR fits whose recovery rate reaches the lower bound of the search interval, refitted with lower bounds 10⁻², 10⁻³ and 10⁻⁴. In all cases the estimate of γ equals the bound. $\beta - \gamma$: growth rate; NRMSE: range over the three bounds.

5 Discussion

Relation to the conference version. The homogeneous fits of the conference version (Cruz-Trani et al. 2026) are reproduced on the conference data; completing truncated author lists changes those of GP and GNN only slightly (Appendix A). Four of its statements are superseded by the present analysis. First, the SI, SIS and SIR models are fitted to different observed series, so we compare models only within a series and no longer select one model per topic. Second, the SI model fixes the population at the last observed value, so a good SI fit does not by itself indicate saturation. Third, the conference version noted that the recovery rate of GP, GNN and QML reached the lower bound of the search interval; Section 4.3 shows that the corresponding values of R_0 (14.7 to 31.5) are therefore not identified. Finally, the conference version attributed its largest residuals, those of GNN and TinyML, to network structure; the degree-block models do not reduce them (Table 2). In this sense the present analysis

refines the conference version: the adoption growth rate used here is the transmission rate of its homogeneous SI model.

What the results do not show. The results do not show that the homogeneous model describes the adoption of topics better than degree-block models in general, nor that collaboration does not influence adoption: they concern the degree-block model built from the empirical $P(k'|k)$ of the topic co-authorship network, whose nodes are all adopters and whose edges are created by adoption. Answering either question requires a network that precedes adoption and includes non-adopters, such as a prior-collaboration network of all the authors of a field, built year by year. [Venturini et al. \(2024\)](#) built such a network and found that the probability of switching to a topic grows with the number of earlier collaborators already active in it. Their measure and ours are compatible: we measure the share of adopters with an active co-author on the records of the topic in the year of entry, whereas they measure the probability of adoption given the number of active collaborators in the whole network. What remains open is whether a degree-block model built on a prior-collaboration network would improve on the homogeneous model. Any unweighted projection of articles onto cliques would inherit the spectral bound of Section 4.2, so such a test would also need to weight the contribution of long author lists or model the author–article network directly. We leave this test for future work.

Limitations. The eight topics were chosen by hand from a single field, so the results may not extend to other topics or fields. Their records come from keyword queries whose strings were not preserved, and the QML and GNN exports contain records outside the topic; Section 4.1 gives the comparison without them. The parameters are estimated by least squares on cumulative or autocorrelated series, so ΔAIC and the CIs are descriptive (Section 3.4), and TinyML has only five observations. The models assume a closed population with constant rates (Section 4.3). The conference version presented the framework as a predictive tool; here the fits are retrospective

descriptions, and we do not evaluate forecasts. Finally, the framework contains no
variable describing individual articles and cannot predict the virality of an article.

6 Conclusions

The conference version conjectured that the structure of co-authorship networks explained the largest residuals of its homogeneous epidemic models. Taking the step it announced, we fitted degree-block SI, SIS and SIR models built from eight topic co-authorship networks. By construction, each network contains only the eventual adopters of a topic, and its edges are created by the publications that define adoption. The degree-block models fitted better in only 3 of 24 comparisons; between 64% and 95% of the authors entered a topic without an active co-author; and the spectral radius that enters the network reproduction number was set, in DHT, GP and GNN, by the authors of a single article. In GP, GNN and QML the recovery rate, and with it the reproduction number, followed the bound of its search interval in 11 of 12 fits, whereas growth rates, including the adoption growth rate, did not. Testing contagion through collaboration requires a prior-collaboration network that includes non-adopters.

Abbreviations.

<i>AIC</i> Akaike Information Criterion	
<i>ANFIS</i> Adaptive Neuro-Fuzzy Inference Systems	
<i>BPEL</i> Business Process Execution Language	
<i>CI</i> Confidence Interval	
<i>DHT</i> Distributed Hash Tables	
<i>GNN</i> Graph Neural Networks	
<i>GP</i> Gaussian Processes	
<i>ISCS</i> International Symposium on Complex Systems	
<i>LCC</i> Largest Connected Component	
<i>NRMSE</i> Normalized Root Mean Square Error	

967 *OLAP* Online Analytical Processing
 968
 969 *QML* Quantum Machine Learning
 970
 971 *RK4* Fourth-order Runge–Kutta
 972
 973 *RSS* Residual Sum of Squares
 974
 975 *SI* Susceptible–Infected
 976
 977 *SIR* Susceptible–Infected–Recovered
 978
 979 *SIS* Susceptible–Infected–Susceptible
 980
 981 *TinyML* Tiny Machine Learning
 982

983 **Declarations**

984 **Ethics approval and consent to participate.** Not applicable.
 985

986 **Consent for publication.** Not applicable.
 987

988 **Availability of data and materials.** The bibliographic records analysed in this
 989 study were obtained from Scopus (Elsevier) under an institutional subscription; restric-
 990 tions apply to the availability of these data, which were used under license for the
 991 current study, and so are not publicly available. The aggregated data derived from
 992 them and the code that generates every table and figure of this article are avail-
 993 able in the Zenodo repository (Robles-Martínez et al. 2026), [https://doi.org/10.5281/](https://doi.org/10.5281/zenodo.23128970)
 994 [zenodo.23128970](https://doi.org/10.5281/zenodo.23128970). Record-level tables are available from the corresponding author on
 995 reasonable request and with permission of Elsevier.
 996
 997
 998
 999

1000 **Competing interests.** The authors declare that they have no competing interests.
 1001

1002 **Funding.** This work was supported by SECIHTI (Secretaría de Ciencia,
 1003 Humanidades, Tecnología e Innovación, Mexico) under grant CBF-2025-G-1812 (Basic
 1004 and Frontier Science programme).
 1005
 1006
 1007
 1008
 1009
 1010
 1011
 1012

Authors' contributions. IRM designed the study. JCT collected the data. All authors contributed equally to the analysis, the interpretation of the results and the writing of the manuscript, and read and approved the final manuscript.

Acknowledgements. Not applicable.

Appendix A Estimated parameters of all fits

Tables A1 and A2 list the parameters, errors, AIC and reproduction numbers of the 48 fits (8 topics, 3 models, 2 families).

Table A1 Parameters of all fits for ANFIS, BPEL, DHT and OLAP

Topic	Family	Model	β	γ	p_0	NRMSE	AIC	R_0 [95% CI]
ANFIS	hom.	SI	0.3824	—	—	0.0527	339.9	—
ANFIS	hom.	SIS	3.2969	2.9628	—	0.0844	245.2	1.11 [1.10, 1.13]
ANFIS	hom.	SIR	0.7196	0.3984	—	0.0953	262.5	1.81 [1.76, 1.86]
ANFIS	block	SI	0.1040	—	6.12e-04	0.0955	372.1	—
ANFIS	block	SIS	0.1481	0.6681	6.12e-04	0.0881	247.6	2.15 [2.06, 2.24]
ANFIS	block	SIR	0.0995	0.1810	6.12e-04	0.0876	257.9	4.95 [4.72, 5.19]
BPEL	hom.	SI	0.7579	—	—	0.0545	203.0	—
BPEL	hom.	SIS	10.0000*	9.2315	—	0.2852	212.7	n.i.
BPEL	hom.	SIR	1.3510	0.5475	—	0.0622	152.6	2.47 [2.35, 2.59]
BPEL	block	SI	0.2750	—	1.40e-02	0.0534	202.2	—
BPEL	block	SIS	1.6241	8.0931	1.40e-02	0.2724	210.8	1.32 [1.18, 1.47]
BPEL	block	SIR	0.3177	0.3772	1.40e-02	0.0730	159.3	4.81 [4.44, 5.19]
DHT	hom.	SI	0.4924	—	—	0.0874	306.5	—
DHT	hom.	SIS	10.0000*	9.3487	—	0.2154	249.7	n.i.
DHT	hom.	SIR	1.0166	0.5000	—	0.1159	229.4	2.03 [1.94, 2.13]
DHT	block	SI	0.1337	—	7.59e-04	0.0804	302.1	—
DHT	block	SIS	0.3153	1.6848	7.59e-04	0.2210	251.0	4.32 [4.04, 4.60]
DHT	block	SIR	0.1547	0.2597	7.59e-04	0.0806	210.6	13.18 [12.55, 13.81]
OLAP	hom.	SI	0.2598	—	—	0.0275	302.8	—
OLAP	hom.	SIS	3.8247	3.5688	—	0.1438	252.0	1.07 [1.06, 1.08]
OLAP	hom.	SIR	0.5424	0.3019	—	0.0520	211.6	1.80 [1.78, 1.82]
OLAP	block	SI	0.0477	—	5.12e-04	0.1520	412.4	—
OLAP	block	SIS	0.0421	0.2970	5.12e-04	0.1691	262.4	2.95 [2.68, 3.23]
OLAP	block	SIR	0.0392	0.0969	5.12e-04	0.1783	290.5	8.07 [7.42, 8.72]

hom.: homogeneous model; block: degree-block model; p_0 : initial proportion of infected nodes in every degree block, fixed at the observed value of the first year. Rates in year⁻¹; β is a per-edge rate in the degree-block models. R_0 : basic reproduction number with its descriptive CI, treating ρ of Eq. (11) as fixed; n.i.: not identified because a parameter lies at a bound.

*Estimate at a bound of the search interval.

1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104

Table A2 Parameters of all fits for GP, GNN, TinyML and QML

Topic	Family	Model	β	γ	p_0	NRMSE	AIC	R_0 [95% CI]
GP	hom.	SI	0.1825	—	—	0.0679	706.6	—
GP	hom.	SIS	0.1469	0.0100*	—	0.0444	490.0	n.i.
GP	hom.	SIR	0.2347	0.0865	—	0.0544	521.6	2.71 [2.20, 3.23]
GP	block	SI	0.0256	—	1.34e-04	0.1306	769.4	—
GP	block	SIS	0.0161	0.0100*	1.34e-04	0.0857	553.2	n.i.
GP	block	SIR	0.0188	0.0100*	1.34e-04	0.0827	561.8	n.i.
GNN	hom.	SI	0.1975	—	—	0.1513	314.5	—
GNN	hom.	SIS	0.1868	0.0100*	—	0.1524	283.6	n.i.
GNN	hom.	SIR	0.1876	0.0100*	—	0.1528	283.8	n.i.
GNN	block	SI	0.0351	—	2.27e-04	0.1759	321.7	—
GNN	block	SIS	0.0276	0.0100*	2.27e-04	0.1719	289.4	n.i.
GNN	block	SIR	0.0317	0.0100*	2.27e-04	0.1760	290.5	n.i.
TinyML	hom.	SI	1.4549	—	—	0.0683	31.8	—
TinyML	hom.	SIS	1.5675	0.5444	—	0.1080	30.5	2.88 [−5.52, 11.28]
TinyML	hom.	SIR	1.3055	0.2450	—	0.1090	30.6	5.33 [−4.17, 14.83]
TinyML	block	SI	0.5736	—	2.41e-02	0.1321	38.4	—
TinyML	block	SIS	0.2737	0.1181	2.41e-02	0.1125	30.9	12.31 [−134.57, 159.18]
TinyML	block	SIR	0.3721	0.0904	2.41e-02	0.1118	30.8	18.13 [−62.63, 98.90]
QML	hom.	SI	0.3625	—	—	0.0878	333.1	—
QML	hom.	SIS	0.3147	0.0100*	—	0.0642	271.1	n.i.
QML	hom.	SIR	0.3190	0.0100*	—	0.0566	264.8	n.i.
QML	block	SI	0.0453	—	2.36e-04	0.1473	359.0	—
QML	block	SIS	0.0321	0.0100*	2.36e-04	0.1154	300.4	n.i.
QML	block	SIR	0.0357	0.0100*	2.36e-04	0.1207	302.7	n.i.

As in Table A1. TinyML has only five observations, so its CIs are uninformative.

References	1105
	1106
Ascher UM, Petzold LR (1998) Computer Methods for Ordinary Differential Equations	1107
and Differential-Algebraic Equations. Society for Industrial and Applied Mathemat-	1108
ics (SIAM)	1109
	1110
	1111
	1112
Barabási AL (2016) Network Science. Cambridge University Press, Cambridge, United	1113
Kingdom	1114
	1115
	1116
Barrat A, Barthélemy M, Vespignani A (2008) Dynamical Processes on Complex	1117
Networks. Cambridge University Press, Cambridge, UK	1118
	1119
	1120
	1121
Bettencourt LMA, Cintrón-Arias A, Kaiser DI, et al (2006) The power of a good idea:	1122
Quantitative modeling of the spread of ideas from epidemiological models. Physica	1123
A: Statistical Mechanics and its Applications 364:513–536. https://doi.org/10.1016/	1124
j.physa.2005.08.083	1125
	1126
	1127
	1128
Bettencourt LMA, Kaiser DI, Kaur J, et al (2008) Population modeling of the emer-	1129
gence and development of scientific fields. Scientometrics 75(3):495–518. https:	1130
//doi.org/10.1007/s11192-007-1888-4	1131
	1132
	1133
	1134
Bettencourt LMA, Kaiser DI, Kaur J (2009) Scientific discovery and topological	1135
transitions in collaboration networks. Journal of Informetrics 3(3):210–221. https:	1136
//doi.org/10.1016/j.joi.2009.03.001	1137
	1138
	1139
	1140
Boguñá M, Pastor-Satorras R, Vespignani A (2003) Absence of epidemic threshold in	1141
scale-free networks with degree correlations. Physical Review Letters 90(2):028701.	1142
https://doi.org/10.1103/PhysRevLett.90.028701	1143
	1144
	1145
	1146
Boyd S, Vandenberghe L (2018) Introduction to Applied Linear Algebra: Vectors,	1147
Matrices, and Least Squares. Cambridge University Press	1148
	1149
	1150

1151 Cruz-Trani J, Robles-Martínez I, Aguirre-Guerrero D, et al (2026) Dynamic modeling
 1152 of topic propagation in coauthorship networks using epidemiological models. In:
 1153 Cherifi H, Guillaume JL, Melançon G (eds) Proceedings of The Sixth International
 1154 Symposium on Complex Systems (ISCS 2026), Studies in Systems, Decision and
 1155 Control, vol 709. Springer, Cham, in press
 1156
 1157
 1158
 1159
 1160 Dorogovtsev SN, Mendes JFF (2022) The Nature of Complex Networks. Oxford Uni-
 1161 versity Press, Oxford, UK, <https://doi.org/10.1093/oso/9780199695119.001.0001>
 1162
 1163
 1164 Goel S, Anderson A, Hofman J, et al (2016) The structural virality of online diffusion.
 1165 Management Science 62(1):180–196. <https://doi.org/10.1287/mnsc.2015.2158>
 1166
 1167
 1168 Goffman W (1966) Mathematical approach to the spread of scientific ideas—the
 1169 history of mast cell research. Nature 212(5061):449–452. [https://doi.org/10.1038/](https://doi.org/10.1038/212449a0)
 1170
 1171 [212449a0](https://doi.org/10.1038/212449a0)
 1172
 1173
 1174 Goffman W (1971) A mathematical method for analyzing the growth of a scien-
 1175 tific discipline. Journal of the ACM 18(2):173–185. [https://doi.org/10.1145/321637.](https://doi.org/10.1145/321637.321640)
 1176
 1177 [321640](https://doi.org/10.1145/321637.321640)
 1178
 1179
 1180 Goffman W, Newill VA (1964) Generalization of epidemic theory: An application to the
 1181 transmission of ideas. Nature 204(4955):225–228. <https://doi.org/10.1038/204225a0>
 1182
 1183
 1184 Goltsev AV, Dorogovtsev SN, Oliveira JG, et al (2012) Localization and spreading
 1185 of diseases in complex networks. Physical Review Letters 109(12):128702. [https:](https://doi.org/10.1103/PhysRevLett.109.128702)
 1186
 1187 [//doi.org/10.1103/PhysRevLett.109.128702](https://doi.org/10.1103/PhysRevLett.109.128702)
 1188
 1189
 1190 Keeling MJ, Rohani P (2008) Modeling Infectious Diseases in Humans and Animals.
 1191 Princeton University Press, <https://doi.org/10.2307/j.ctvcn4gk0>
 1192
 1193
 1194
 1195
 1196

King AA, Domenech de Cellès M, Magpantay FMG, et al (2015) Avoidable errors in	1197
the modelling of outbreaks of emerging pathogens, with special reference to Ebola.	1198
Proceedings of the Royal Society B: Biological Sciences 282(1806):20150347. https://doi.org/10.1098/rspb.2015.0347	1199
	1200
	1201
	1202
	1203
Kiss IZ, Broom M, Craze PG, et al (2010) Can epidemic models describe the diffusion	1204
of topics across disciplines? Journal of Informetrics 4(1):74–82. https://doi.org/10.1016/j.joi.2009.08.002	1205
	1206
	1207
	1208
	1209
Lee HA, Alves LGA, Amaral LAN (2021) Spreader events and the limitations of pro-	1210
jected networks for capturing dynamics on multipartite networks. Physical Review	1211
E 103(2):022320. https://doi.org/10.1103/PhysRevE.103.022320	1212
	1213
	1214
	1215
Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration.	1216
Proceedings of the National Academy of Sciences 101(Suppl 1):5200–5205. https://doi.org/10.1073/pnas.0307545100	1217
	1218
	1219
	1220
	1221
Newman MEJ (2018) Networks, 2nd edn. Oxford University Press, Oxford, UK, https://doi.org/10.1093/oso/9780198805090.001.0001	1222
	1223
	1224
	1225
Pastor-Satorras R, Castellano C (2016) Distinct types of eigenvector localization in	1226
networks. Scientific Reports 6:18847. https://doi.org/10.1038/srep18847	1227
	1228
	1229
Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks.	1230
Physical Review Letters 86(14):3200–3203. https://doi.org/10.1103/PhysRevLett.86.3200	1231
	1232
	1233
	1234
	1235
Pastor-Satorras R, Castellano C, Van Mieghem P, et al (2015) Epidemic processes in	1236
complex networks. Reviews of Modern Physics 87(3):925–979. https://doi.org/10.1103/RevModPhys.87.925	1237
	1238
	1239
	1240
	1241
	1242

1243 Robles-Martínez I, Cruz-Trani J, Aguirre-Guerrero D, et al (2026) epidemic-mean-
 1244 field-topics: code and aggregated data for topic propagation in co-authorship
 1245 networks. Zenodo, <https://doi.org/10.5281/zenodo.23128970>
 1246
 1247
 1248
 1249 Venturini S, Sikdar S, Rinaldi F, et al (2024) Collaboration and topic switches in
 1250 science. Scientific Reports 14:1258. <https://doi.org/10.1038/s41598-024-51606-6>
 1251
 1252
 1253 Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between
 1254 growth rates and reproductive numbers. Proceedings of the Royal Society B:
 1255 Biological Sciences 274(1609):599–604. <https://doi.org/10.1098/rspb.2006.3754>
 1256
 1257
 1258
 1259 Yue Z, Xu H, Yuan G, et al (2019) Modeling study of knowledge diffusion in scientific
 1260 collaboration networks based on differential dynamics: A case study in graphene
 1261 field. Physica A: Statistical Mechanics and its Applications 524:375–391. <https://doi.org/10.1016/j.physa.2019.04.201>
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276
 1277
 1278
 1279
 1280
 1281
 1282
 1283
 1284
 1285
 1286
 1287
 1288