Modeling citation network topology: A natural graph model and applications to citation theory

Lovro Šubelj[a,*], Dalibor Fiala[b], Slavko Žitnik[a], Marko Bajec[a]

[a] *University of Ljubljana, Faculty of Computer and Information Science, Tržaška 25, SI-1000 Ljubljana, Slovenia*
[b] *University of West Bohemia, Faculty of Applied Sciences, Univerzitní 8, CZ-30614 Plzeň, Czech Republic*

Abstract

TEXT

Keywords: graph model, citation networks, citation theory, infometrics, bibliometrics, scientometrics.

1. Introduction

TEXT

2. Related work

TEXT

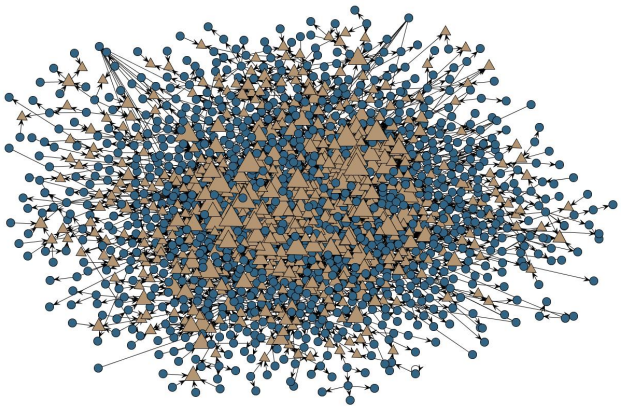


Figure 1: Largest connected component of citation network of WoS category *Information Science & Library Science* reduced to papers published before 1980 (see also text). Papers with more than one citation are shown as triangles, while node sizes are proportional to in-degrees. (Network layout was generated with Large Graph Layout [1].)

*Corresponding author. Tel.: +386 1 476 81 86.
Email addresses: lovro.subelj@fri.uni-lj.si (Lovro Šubelj), dalfia@kiv.zcu.cz (Dalibor Fiala), slavko.zitnik@fri.uni-lj.si (Slavko Žitnik), marko.bajec@fri.uni-lj.si (Marko Bajec)

3. Modeling citation networks

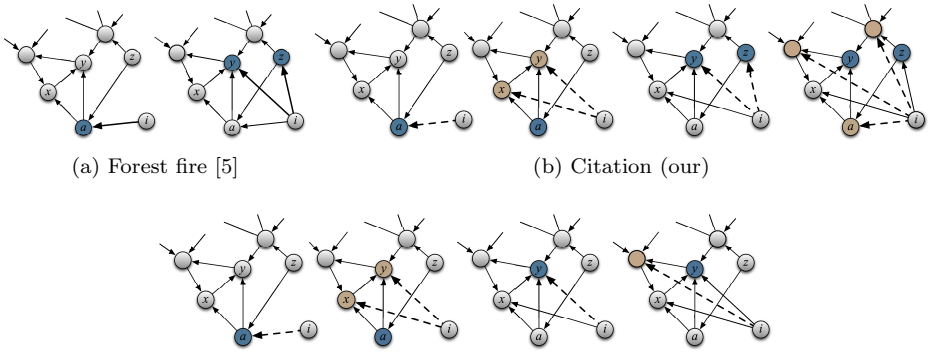
TEXT

3.1. Study of citation dynamics

TEXT

3.2. Graph model of citation networks

TEXT



(a) Forest fire [5] (b) Citation (our)

(c) Random walk [9]

Figure 2: Linking dynamics of random graph models of citation networks. (a) In Forest fire model [5], newly added node i chooses an ambassador a uniformly at random and links to it (solid lines). Next, some of its neighbors are taken as the new ambassadors by following the in-links and out-links (e.g., y and z). (b) In the proposed Citation model, i links to a only with a certain probability (dashed lines), while also to some of its neighbors by following the out-links (e.g., x and y). (c) In Random walk model [9, 6], i traverses the network by following the out-links, taking a single new ambassador on each step (e.g., y).

Initially, the graph consists of a single link, while the model repeats the following procedure $n-2$ times.

(1) Newly added node i chooses an ambassador node a uniformly at random (we say that i burns a).
(2) i links to a with probability q_a and randomly selects $x_l \sim G(q_l)$ neighbors of a that were not yet linked by following the out-links, and links to them.
(3) i randomly selects $x_f \sim G(p_f)$ neighbors of a that were not yet burned $a_1, \ldots, a_{x_f}$ by following the out-links and $x_b \sim G(p_b)$ neighbors that were not yet burned $a_{x_f+1}, \ldots, a_{x_f+x_b}$ by following the in-links.
(4) $a_1, \ldots, a_{x_f+x_b}$ are taken as the new ambassadors of i (step (2)).

When i cannot select enough neighbors of a, the remaining are selected uniformly at random from all nodes.

$$\langle s \rangle = \sum_{x=0}^{\infty} \left(\frac{p_f}{1-p_f} + \frac{p_b}{1-p_b} \right)^x = \left(1 - \frac{p_f}{1-p_f} - \frac{p_b}{1-p_b} \right)^{-1} \tag{1}$$

for $\frac{p_f}{1-p_f} + \frac{p_b}{1-p_b} < 1$. Hence,

$$m = n \langle s \rangle \left(q_a + \frac{q_l}{1-q_l} \right) \tag{2}$$

for $q_l < 1$. Thus,

$$q_a = \frac{m}{n \langle s \rangle} - \frac{q_l}{1-q_l} \tag{3}$$

and

$$q_l = \left(1 + \left(\frac{m}{n \langle s \rangle} - q_a \right)^{-1} \right)^{-1}. \tag{4}$$

Also,

$$n_+ = \frac{nq - nq\left(p - pq\right)^{\langle s \rangle}}{1 - p\left(1 - q\right)} \approx \frac{nq}{1 - p\left(1 - q\right)} \tag{5}$$

for $p\left(1 - q\right) < 1$, where $p = p_f + p_b - p_f p_b$ and $q = q_a + q_l - q_a q_l$. Thus,

$$q_l \approx 1 - \frac{2n_+ \left\langle s \right\rangle \left(1 - p\right)}{2m - n_+ \left\langle k \right\rangle p}. \tag{6}$$

3.3. Alternative citation models

TEXT

3.3.1. Poisson citation distribution

TEXT

Hence,

$$m = n \left\langle s \right\rangle \left(q_a + q_l\right). \tag{7}$$

Thus,

$$q_a = \frac{m}{n \left\langle s \right\rangle} - q_l \tag{8}$$

and

$$q_l = \frac{m}{n \left\langle s \right\rangle} - q_a. \tag{9}$$

Also,

$$n_+ = \frac{n\left(1 - q\right)\left(1 - \left(pq\right)^{\langle s \rangle}\right)}{1 - pq} \approx \frac{n\left(1 - q\right)}{1 - pq} \tag{10}$$

for $pq < 1$, where $p = p_f + p_b - p_f p_b$ and $q = \left(1 - q_a\right) e^{-q_l}$. Thus,

$$q_l \approx \frac{m}{n \left\langle s \right\rangle} - W\left(\exp\left(\frac{m}{n \left\langle s \right\rangle}\right)\left(n - n_+ p\right)^{-1}\right) - 1, \tag{11}$$

where W is the Lambert function.

3.3.2. Preferential attachment citation

TEXT

3.3.3. Random walk citation traversal

TEXT

4. Citation data and methodology

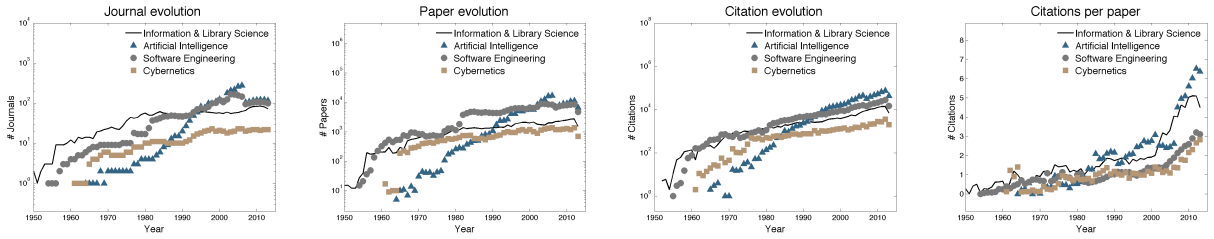
TEXT

4.1. Web of Science data

TEXT

Table 1: Basic statistics of *Web of Science* citation data analyzed in the study.

Data	WoS category	Period	# Journals	# Papers	# Citations
ILS	*Information Science & Library Science*[1]	1945–2013	177	71253	176240
AI	*Computer Science, Artificial Intelligence*	1964–2013	877	179510	639126
SE	*Computer Science, Software Engineering*	1954–2013	560	215745	323444
CY	*Computer Science, Cybernetics*	1961–2013	66	38097	45755



(a) (b) (c) (d)

Figure 3: Journal, paper and citation evolution of *Web of Science* data analyzed in the study (mind semilogarithmic scales).

4.2. Statistics of citation networks

TEXT

$$c_i = \frac{t_i}{\binom{k_i}{2}} \text{ and } d_i = \frac{t_i}{\omega_i}. \tag{12}$$

α is the scale-free exponent of the power-law degree distribution $P(k) \sim k^{-\alpha}$ (similarly for α_{in} and α_{out}). $r_{(\alpha,\beta)}$ is the Pearson correlation coefficient of α-degrees and β-degrees at links' source and target nodes, respectively, $\alpha, \beta \in \{in, out\}$.

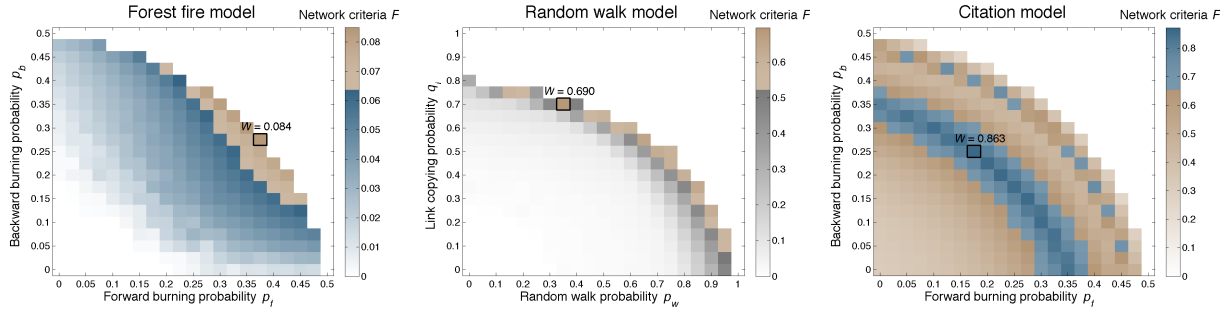
$$\mathcal{F} = \exp\left(-C\left(\frac{|m-m^*|}{m} + \frac{|\langle c\rangle - \langle c\rangle^*|}{\langle c\rangle} + \frac{|\langle d\rangle - \langle d\rangle^*|}{\langle d\rangle} + \frac{|\alpha-\alpha^*|}{\alpha} + \frac{|\alpha_{in}-\alpha^*_{in}|}{\alpha_{in}} + \frac{|\alpha_{out}-\alpha^*_{out}|}{\alpha_{out}} + \right.\right. \tag{13}$$
$$\left.\left. + |r_{(in,in)} - r^*_{(in,in)}| + |r_{(in,out)} - r^*_{(in,out)}| + |r_{(out,in)} - r^*_{(out,in)}| + |r_{(out,out)} - r^*_{(out,out)}|\right)\right),$$

where C is set to 0.4.

[1] Reduced to papers within WoS category *Computer Science* or its subcategories.

4.3. Parameters of citation models

TEXT



(a) (b) (c)

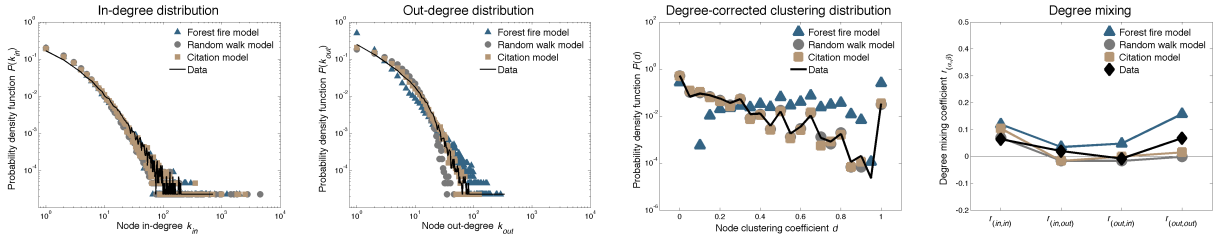
Figure 4: Random graph parameter estimation for citation network of WoS category *Information Science & Library Science* (note different color scales). (Results are estimates of the mean over 25 graph realizations with the global maximum highlighted.)

5. Results and applications

TEXT

5.1. Validation of citation model

TEXT



(a) (b) (c) (d)

Figure 5: Degree and clustering distributions and degree mixing in citation network of WoS category *Information Science & Library Science* and graphs generated with random models (see also Figs. B.9 and B.10).

Table 2: Degree and clustering distributions and degree mixing in citation network of WoS category *Information Science & Library Science* and graphs generated with random models (see also Tables B.8 and B.9). (Results are estimates of the mean over 25 graph realizations.)

Model	$\langle k \rangle$	$\langle c \rangle$	$\langle d \rangle$	α_{in}	α_{out}	$r_{(in,in)}$	$r_{(in,out)}$	$r_{(out,in)}$	$r_{(out,out)}$	$\mathcal{F}$
Data	8.05	0.11	0.14	2.47	3.35	0.06	0.02	−0.01	0.07	*1.000*
Citation (our)	7.97	0.11	0.13	2.58	3.31	0.12	−0.01	0.01	0.01	*0.863*
Random walk [9]	7.91	0.11	0.13	2.61	4.68	0.07	−0.01	−0.02	0.00	*0.695*
Random-citing [7]	7.48	0.21	0.26	2.36	3.26	0.16	−0.02	0.08	0.15	*0.400*
Forest fire [5]	7.88	0.40	0.52	2.51	2.45	0.13	0.04	0.06	0.17	*0.087*
Poisson citation	8.03	0.11	0.14	2.63	2.17	0.12	−0.02	0.00	0.01	*0.749*
Yule dynamics	8.06	0.11	0.14	2.14	2.16	0.12	−0.04	0.01	0.03	*0.697*
Yule burning	8.02	0.11	0.13	2.53	3.28	0.10	−0.01	0.01	0.02	*0.882*
Yule citing	8.01	0.11	0.14	2.59	2.17	0.10	−0.02	−0.01	0.02	*0.759*

5.2. Assessment of citation dynamics

TEXT

Recall that $\langle s \rangle$ in Eq. (1) is the number of papers actually read by an author, while q_a is the probability of citing such paper. The total number of cited papers is $\langle k \rangle /2$, while $\langle s \rangle q_l/(1-q_l)$ is the number of copied citations. The ratio between the papers discovered by citations and services is $p_f/(1-p_f) : p_b/(1-p_b)$.

Table 3: Paper citation, study and discovery dynamics of WoS category *Information Science & Library Science* as predicted by random graph models (see also Table C.10). (Results are estimates of the mean over 25 graph realizations.)

	Paper citation		Paper study		Paper discovery by		
Model	# Cite	% Copy	# Read	% Cite	% Citation	% Service	% Other
Citation (our)	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
Random walk [9]	3.95	90.4%	1.54	29.2%	83.7%	-	16.3%
Random-citing [7]	3.73	65.4%	1.29	100.0%	-	-	-
Forest fire [5]	3.95	-	3.95	100.0%	62.3%	37.7%	-

5.3. Applications to citation theory

TEXT

Table 4: Paper citation, study and discovery dynamics of WoS categories *Information Science & Library Science* (ILS), *Computer Science, Artificial Intelligence* (AI), *Computer Science, Software Engineering* (SE) and *Computer Science, Cybernetics* (CY) as predicted by citation model (see also Table C.10). (Results are estimates of the mean over 25 graph realizations.)

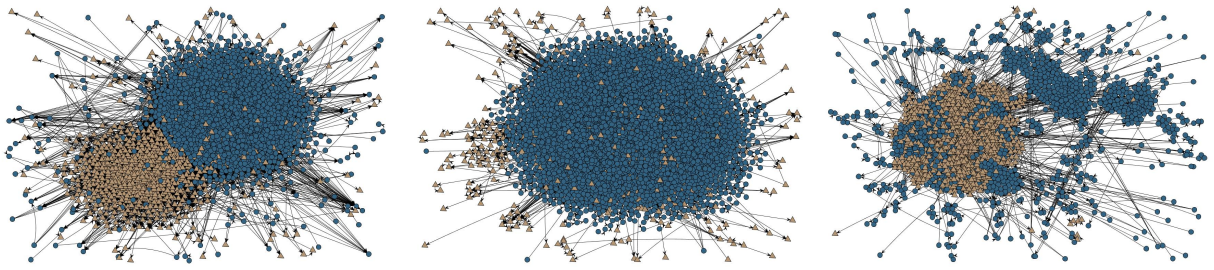
	Paper citation		Paper study		Paper discovery by		
Data	# Cite	% Copy	# Read	% Cite	% Citation	% Service	% Other
ILS	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
AI	4.52	87.3%	1.47	40.9%	25.8%	47.6%	26.6%
SE	2.78	81.5%	1.58	36.4%	68.8%	2.0%	29.2%
CY	2.18	69.6%	1.59	43.2%	24.5%	37.8%	37.6%

Table 5: Paper citation, study and discovery evolution of WoS category *Information Science & Library Science* as predicted by citation model (see also Table D.11). (Results are estimates of the mean over 25 graph realizations.)

	Paper citation		Paper study		Paper discovery by		
Period	# Cite	% Copy	# Read	% Cite	% Citation	% Service	% Other
1945–2013	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
1970–1980	2.23	52.1%	3.39	33.5%	41.4%	0.0%	58.5%
1980–1990	2.62	65.1%	2.96	33.0%	48.3%	1.1%	50.6%
1990–2000	3.42	81.6%	2.38	29.0%	40.3%	23.2%	36.5%
2000–2010	5.06	83.6%	2.90	32.2%	40.7%	27.5%	31.7%

Table 6: Paper citation, study and discovery evolution of WoS journals *Scientometrics* (SCI), *Expert Systems with Applications* (ESWA), *Communications of the ACM* (CACM) and *IEEE Transactions on Systems, Man, and Cybernetics* (TSMC) as predicted by citation model (see also text). (Results are estimates of the mean over 25 graph realizations.)

	Paper citation		Paper study		Paper discovery by		
Journal	# Cite	% Copy	# Read	% Cite	% Citation	% Service	% Other
SCI	3.83	87.6%	1.40	38.5%	54.3%	23.4%	22.3%
ESWA	2.12	81.0%	1.67	25.9%	34.7%	27.8%	37.6%
CACM	1.32	46.5%	1.70	42.0%	8.7%	38.5%	52.8%
TSMC	1.47	37.1%	1.66	56.8%	13.6%	39.9%	46.4%



(a) (b) (c)

Figure 6: Largest connected components of citation networks of WoS journals (a) *Scientometrics*, (b) *Expert Systems with Applications* and (c) *Communications of the ACM* (see also text). Papers published in the first or second half of the journal's lifespan are shown as triangles and circles, respectively. (Network layouts were generated with Large Graph Layout [1].)

6. Final remarks

TEXT

Acknowledgments

Authors would like to thank Vladimir Batagelj for insightful comments and discussions, and Thomson Reuters for providing the access to WoS citation data. This work has been supported in part by the Slovenian Research Agency Program No. P2-0359, by the Slovenian Ministry of Education, Science and Sport Grant No. 430-168/2013/91, and by the European Union, European Social Fund.

Appendix A. Parameters of graph models

TEXT

Table A.7: Random graph model parameters for citation networks of WoS categories *Information Science & Library Science* (ILS), *Computer Science, Artificial Intelligence* (AI), *Computer Science, Software Engineering* (SE) and *Computer Science, Cybernetics* (CY).

		Burning probability		Linking probability	
Data	Model	p_f (p_w)	p_b	q_a	q_l
ILS	Citation (our)	0.181	0.237	0.279	0.615
	Random walk [9]	0.349	-	0.292	0.699
	Forest fire [5]	0.384	0.274	-	-
AI	Citation (our)	0.102	0.173	0.409	0.728
	Random walk [9]	0.261	-	0.417	0.747
	Forest fire [5]	0.423	0.209	-	-
SE	Citation (our)	0.264	0.010	0.364	0.588
	Random walk [9]	0.351	-	0.333	0.602
	Forest fire [5]	0.444	0.128	-	-
CY	Citation (our)	0.128	0.184	0.432	0.488
	Random walk [9]	0.255	-	0.427	0.549
	Forest fire [5]	0.343	0.191	-	-

		Kronecker graph initiator $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$			
Data	Model	a	b	c	d
ILS	Kronecker [4]	0.503	0.398	0.429	0.898
AI		0.450	0.506	0.505	0.701
SE		0.859	0.509	0.434	0.337
CY		0.536	0.356	0.371	0.908

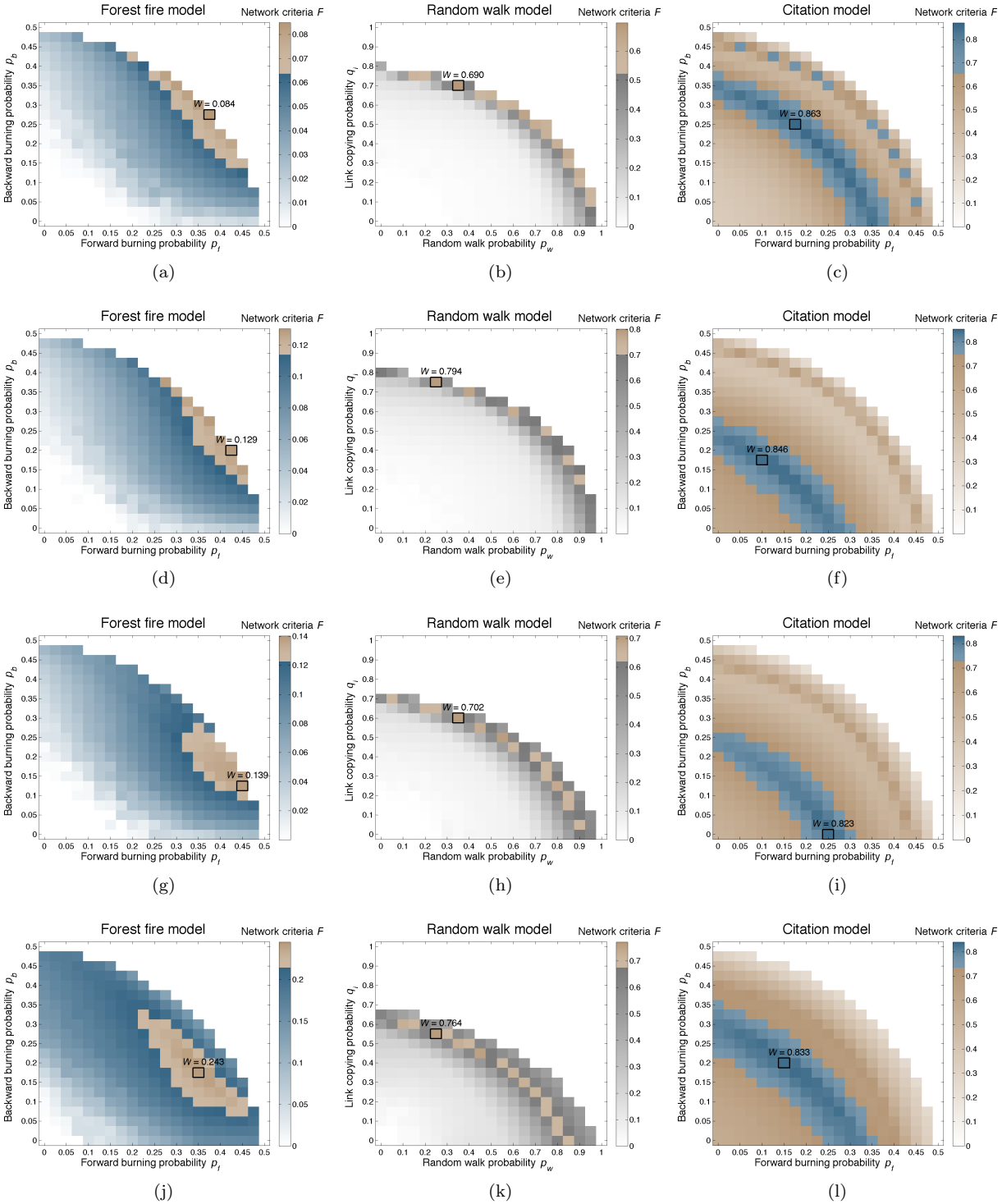


Figure A.7: Random graph parameter estimation for citation networks of WoS categories (a)–(c) *Information Science & Library Science*, (d)–(f) *Computer Science, Artificial Intelligence*, (g)–(i) *Computer Science, Software Engineering* and (j)–(l) *Computer Science, Cybernetics* (note different color scales). (Results are estimates of the mean over 25 graph realizations with the global maximum highlighted.)

Appendix B. Comparison of graph models

TEXT

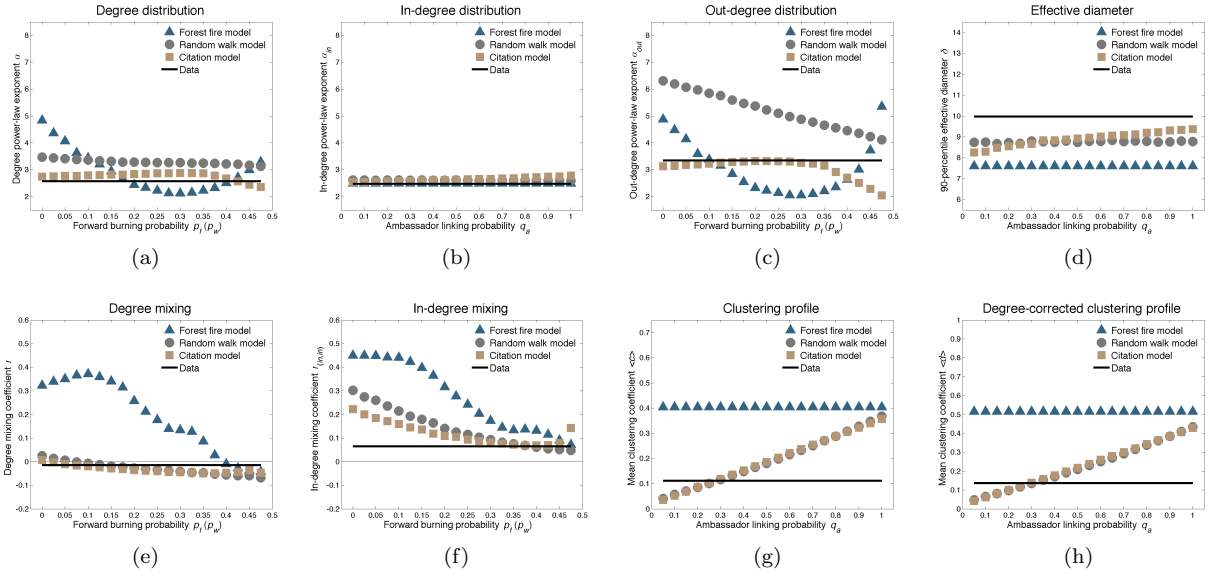


Figure B.8: Degree and clustering distributions and degree mixing in citation network of WoS category *Information Science & Library Science* and graphs generated by varying parameters of random models.

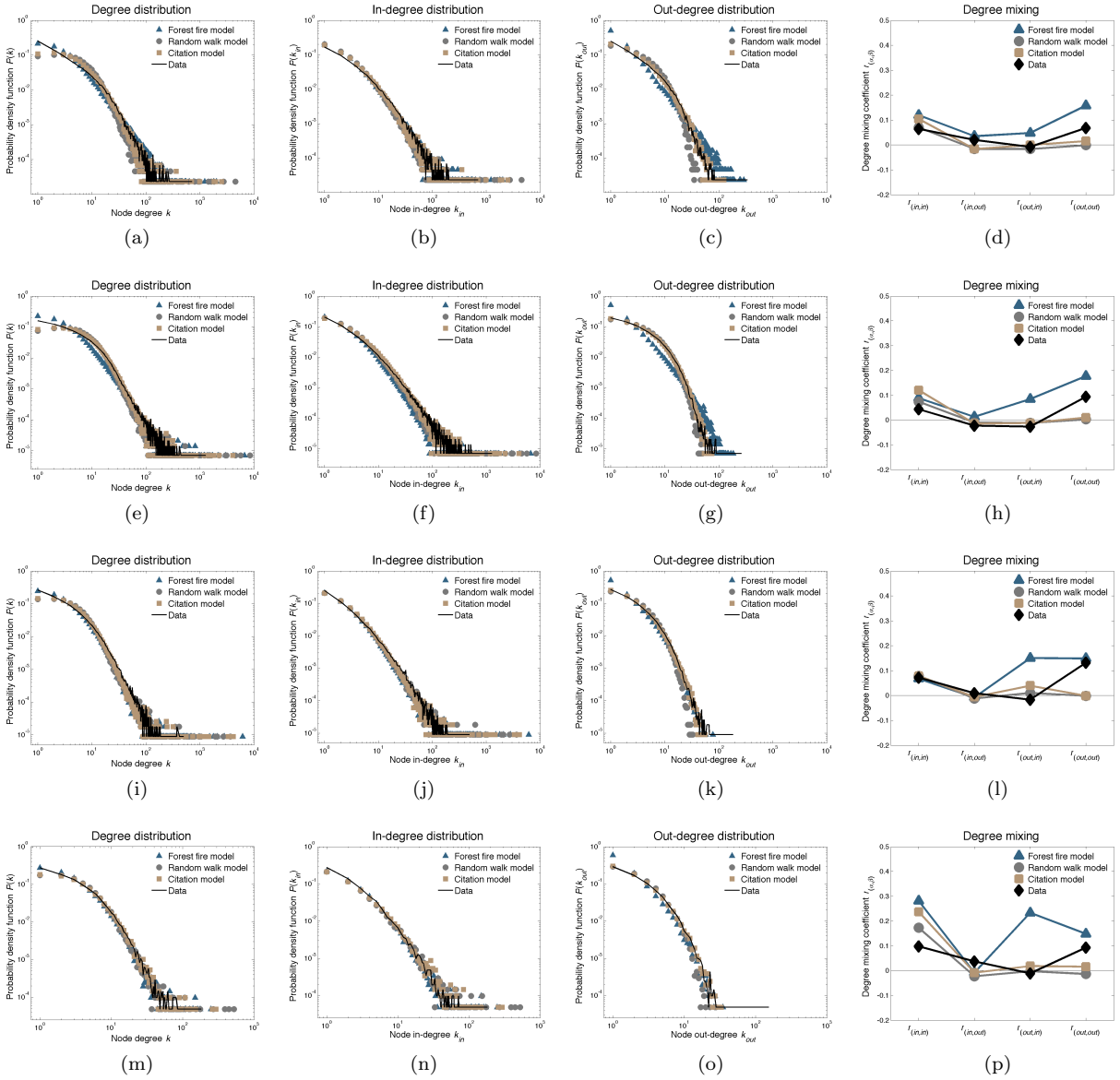


Figure B.9: Degree distributions and mixing in citation networks of WoS categories (a)–(d) *Information Science & Library Science*, (e)–(h) *Computer Science, Artificial Intelligence*, (i)–(l) *Computer Science, Software Engineering* and (m)–(p) *Computer Science, Cybernetics*, and graphs generated with random models (see also Table B.9).

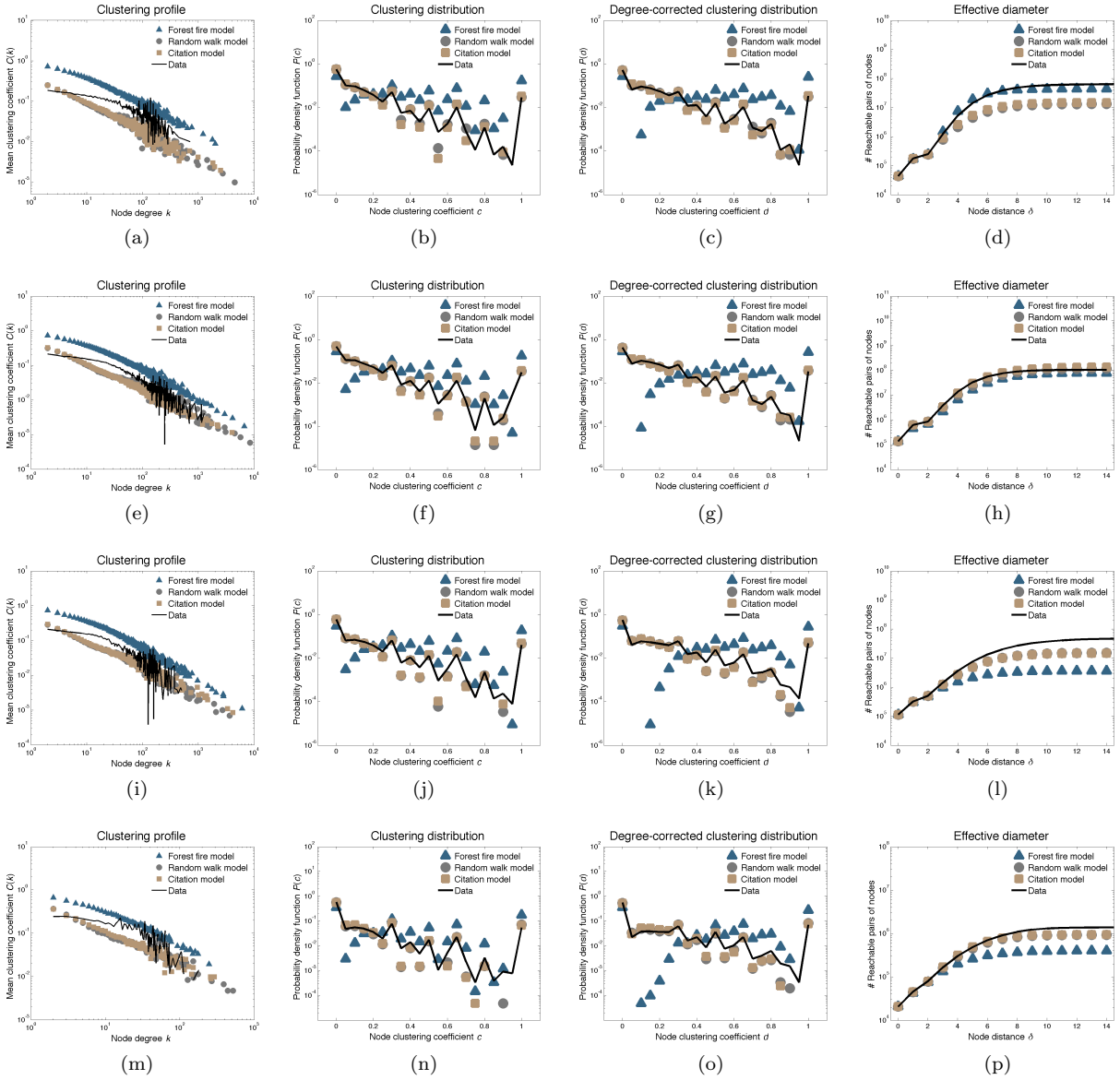


Figure B.10: Clustering profile and distributions in citation networks of WoS categories (a)–(d) *Information Science & Library Science*, (e)–(h) *Computer Science, Artificial Intelligence*, (i)–(l) *Computer Science, Software Engineering* and (m)–(p) *Computer Science, Cybernetics*, and graphs generated with random models (see also Table B.8).

Table B.8: Standard statistics and clustering in citation networks of WoS categories *Information Science & Library Science* (ILS), *Computer Science, Artificial Intelligence* (AI), *Computer Science, Software Engineering* (SE) and *Computer Science, Cybernetics* (CY), and graphs generated with random models. (Results are estimates of the mean over 25 graph realizations.)

Data	Model	n	m	WCC	$\langle k \rangle$	$\langle c \rangle$	$\langle d \rangle$	δ_e	$\mathcal{F}$
ILS	Data	43808	176219	94.7%	8.05	0.11	0.14	9.6	*1.000*
	Citation (our)	43808	174467	92.5%	7.97	0.11	0.13	8.6	*0.863*
	Random walk [9]	43808	173241	94.7%	7.91	0.11	0.13	8.8	*0.695*
	Random-citing [7]	43808	163743	77.3%	7.48	0.21	0.26	5.6	*0.400*
	Forest fire [5]	43808	172510	100.0%	7.88	0.40	0.52	7.7	*0.087*
	Kronecker [4]	32768	165822	98.6%	10.12	0.00	0.00	7.7	*0.209*
	Small-world [10]	43808	175228	100.0%	8.00	0.12	0.14	11.5	*0.000*
	Scale-free [2]	43808	175226	100.0%	8.00	0.00	0.00	7.3	*0.329*
	Random [3]	43808	176219	100.0%	8.05	0.00	0.00	9.8	*0.001*
AI	Data	140362	639110	97.0%	9.11	0.14	0.16	8.8	*1.000*
	Citation (our)	140362	635007	95.8%	9.05	0.14	0.16	9.7	*0.849*
	Random walk [9]	140362	629652	96.4%	8.97	0.14	0.16	9.9	*0.798*
	Random-citing [7]	140362	590798	76.9%	8.42	0.22	0.26	5.7	*0.486*
	Forest fire [5]	140362	464120	100.0%	6.61	0.41	0.53	9.6	*0.129*
	Kronecker [4]	131072	487535	99.0%	7.44	0.00	0.00	10.0	*0.073*
	Small-world [10]	140362	561445	100.0%	8.00	0.14	0.16	12.7	*0.000*
	Scale-free [2]	140362	561442	100.0%	8.00	0.00	0.00	7.8	*0.309*
	Random [3]	140362	639110	100.0%	9.11	0.00	0.00	9.8	*0.003*
SE	Data	113907	323406	90.5%	5.68	0.13	0.16	12.3	*1.000*
	Citation (our)	113907	316471	91.1%	5.56	0.13	0.16	9.3	*0.841*
	Random walk [9]	113907	316495	91.5%	5.56	0.13	0.16	9.1	*0.718*
	Random-citing [7]	113907	314972	70.9%	5.53	0.21	0.26	5.0	*0.475*
	Forest fire [5]	113907	295989	100.0%	5.20	0.41	0.53	7.5	*0.139*
	Kronecker [4]	131072	410583	88.8%	6.26	0.00	0.00	8.7	*0.253*
	Small-world [10]	113907	227813	74.0%	4.00	0.00	0.00	-	*0.000*
	Scale-free [2]	113907	227812	100.0%	4.00	0.00	0.00	16.1	*0.288*
	Random [3]	113907	323406	99.7%	5.68	0.00	0.00	13.7	*0.000*
CY	Data	20706	45743	83.4%	4.42	0.16	0.20	9.4	*1.000*
	Citation (our)	20706	45196	88.2%	4.37	0.16	0.20	7.9	*0.849*
	Random walk [9]	20706	44731	89.5%	4.32	0.16	0.19	7.6	*0.770*
	Random-citing [7]	20706	44892	72.1%	4.34	0.19	0.24	5.4	*0.608*
	Forest fire [5]	20706	42883	100.0%	4.14	0.37	0.50	7.3	*0.246*
	Kronecker [4]	16384	51143	96.2%	6.24	0.00	0.00	9.2	*0.284*
	Small-world [10]	20706	41412	68.6%	4.00	0.00	0.00	-	*0.000*
	Scale-free [2]	20706	41410	100.0%	4.00	0.00	0.00	14.0	*0.311*
	Random [3]	20706	45743	98.7%	4.42	0.00	0.00	15.4	*0.031*

Table B.9: Degree distributions and mixing in citation networks of WoS categories *Information Science & Library Science* (ILS), *Computer Science, Artificial Intelligence* (AI), *Computer Science, Software Engineering* (SE) and *Computer Science, Cybernetics* (CY), and graphs generated with random models. (Results are estimates of the mean over 25 graph realizations.)

Data	Model	α	α_{in}	α_{out}	$r_{(in,in)}$	$r_{(in,out)}$	$r_{(out,in)}$	$r_{(out,out)}$	$\mathcal{F}$
ILS	Data	2.57	2.47	3.35	0.06	0.02	−0.01	0.07	*1.000*
	Citation (our)	2.83	2.58	3.31	0.12	−0.01	0.01	0.01	*0.863*
	Random walk [9]	3.24	2.61	4.68	0.07	−0.01	−0.02	0.00	*0.695*
	Random-citing [7]	2.72	2.36	3.26	0.16	−0.02	0.08	0.15	*0.400*
	Forest fire [5]	2.39	2.51	2.45	0.13	0.04	0.06	0.17	*0.087*
	Kronecker [4]	3.13	4.75	4.12	0.16	0.16	0.16	0.16	*0.209*
	Small-world [10]	15.51	-	-	−0.07	−0.07	−0.07	−0.07	*0.000*
	Scale-free [2]	3.13	3.06	3.06	−0.03	−0.03	−0.03	−0.03	*0.329*
	Random [3]	8.58	22.30	22.34	0.00	0.00	0.00	0.00	*0.001*
AI	Data	2.74	2.39	3.88	0.04	−0.02	−0.03	0.09	*1.000*
	Citation (our)	2.94	2.60	3.84	0.13	−0.01	−0.01	0.01	*0.849*
	Random walk [9]	3.12	2.61	4.36	0.07	−0.01	−0.01	0.00	*0.798*
	Random-citing [7]	2.57	2.30	2.97	0.11	−0.02	0.08	0.15	*0.486*
	Forest fire [5]	2.67	2.63	2.91	0.08	0.01	0.09	0.17	*0.129*
	Kronecker [4]	4.70	7.88	8.15	0.02	0.02	0.03	0.02	*0.073*
	Small-world [10]	16.52	-	-	−0.07	−0.07	−0.07	−0.07	*0.000*
	Scale-free [2]	3.13	3.06	3.06	−0.02	−0.02	−0.02	−0.02	*0.309*
	Random [3]	7.12	19.03	18.93	0.00	0.00	0.00	0.00	*0.003*
SE	Data	3.01	2.61	4.46	0.07	0.01	−0.02	0.13	*1.000*
	Citation (our)	3.32	2.73	4.47	0.08	−0.01	0.03	0.00	*0.841*
	Random walk [9]	3.50	2.69	6.17	0.06	−0.01	0.01	0.00	*0.718*
	Random-citing [7]	2.95	2.50	3.63	0.15	−0.01	0.11	0.19	*0.475*
	Forest fire [5]	3.19	2.68	4.34	0.07	−0.01	0.16	0.15	*0.139*
	Kronecker [4]	2.96	4.35	3.26	0.04	0.04	0.04	0.04	*0.253*
	Small-world [10]	61.87	-	-	−0.14	−0.14	−0.14	−0.14	*0.000*
	Scale-free [2]	3.13	3.05	3.05	−0.02	−0.02	−0.02	−0.02	*0.288*
	Random [3]	14.19	37.46	37.74	0.00	0.00	0.00	0.00	*0.000*
CY	Data	3.03	3.00	4.05	0.10	0.04	−0.01	0.09	*1.000*
	Citation (our)	3.05	2.81	4.09	0.23	−0.01	0.01	0.02	*0.849*
	Random walk [9]	3.25	2.78	4.98	0.15	−0.01	0.00	0.00	*0.770*
	Random-citing [7]	2.74	2.67	3.26	0.26	−0.02	0.10	0.18	*0.608*
	Forest fire [5]	3.14	2.96	4.06	0.23	−0.00	0.21	0.17	*0.246*
	Kronecker [4]	2.91	4.39	4.16	0.18	0.18	0.18	0.19	*0.284*
	Small-world [10]	5.31	73.77	73.77	−0.13	−0.13	−0.13	−0.13	*0.000*
	Scale-free [2]	2.58	3.10	3.10	−0.05	−0.05	−0.05	−0.05	*0.311*
	Random [3]	5.85	12.91	12.94	0.00	0.00	0.00	0.00	*0.031*

Appendix C. WoS citation dynamics

TEXT

Table C.10: Paper citation, study and discovery dynamics of WoS categories *Information Science & Library Science* (ILS), *Computer Science, Artificial Intelligence* (AI), *Computer Science, Software Engineering* (SE) and *Computer Science, Cybernetics* (CY) as predicted by random graph models. (Results are estimates of the mean over 25 graph realizations.)

		Paper citation		Paper study		Paper discovery by		
Data	Model	# Cite	% Copy	# Read	% Cite	% Citation	% Service	% Other
ILS	Citation (our)	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
	Random walk [9]	3.95	90.4%	1.54	29.2%	83.7%	-	16.3%
	Random-citing [7]	3.73	65.4%	1.29	100.0%	-	-	-
	Forest fire [5]	3.95	-	3.95	100.0%	62.3%	37.7%	-
AI	Citation (our)	4.52	87.3%	1.47	40.9%	25.8%	47.6%	26.6%
	Random walk [9]	4.49	88.9%	1.35	41.7%	88.1%	-	11.9%
	Random-citing [7]	4.21	69.7%	1.28	100.0%	-	-	-
	Forest fire [5]	3.31	-	3.31	100.0%	73.5%	26.5%	-
SE	Citation (our)	2.78	81.5%	1.58	36.4%	68.8%	2.0%	29.2%
	Random walk [9]	2.78	83.7%	1.54	33.3%	80.2%	-	19.8%
	Random-citing [7]	2.75	61.0%	1.07	100.0%	-	-	-
	Forest fire [5]	2.59	-	2.59	100.0%	84.4%	15.6%	-
CY	Citation (our)	2.18	69.6%	1.59	43.2%	24.5%	37.8%	37.6%
	Random walk [9]	2.16	75.7%	1.34	42.7%	79.0%	-	21.0%
	Random-citing [7]	2.17	49.1%	1.10	100.0%	-	-	-
	Forest fire [5]	2.07	-	2.07	100.0%	68.9%	31.1%	-

Appendix D. WoS citation evolution

TEXT

Table D.11: Paper citation, study and discovery evolution of WoS categories *Information Science & Library Science* (ILS), *Computer Science, Artificial Intelligence* (AI), *Computer Science, Software Engineering* (SE) and *Computer Science, Cybernetics* (CY) as predicted by citation model. (Results are estimates of the mean over 25 graph realizations.)

		Paper citation		Paper study		Paper discovery by		
Data	Period	# Cite	% Copy	# Read	% Cite	% Citation	% Service	% Other
ILS	1945–2013	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
	1970–1980	2.23	52.1%	3.39	33.5%	41.4%	0.0%	58.5%
	1980–1990	2.62	65.1%	2.96	33.0%	48.3%	1.1%	50.6%
	1990–2000	3.42	81.6%	2.38	29.0%	40.3%	23.2%	36.5%
	2000–2010	5.06	83.6%	2.90	32.2%	40.7%	27.5%	31.7%
AI	1964–2013	4.52	87.3%	1.47	40.9%	25.8%	47.6%	26.6%
	1970–1980	1.23	38.3%	2.11	36.8%	28.9%	19.5%	51.6%
	1980–1990	2.84	78.1%	1.56	42.1%	25.2%	39.8%	35.0%
	1990–2000	3.38	79.7%	1.70	43.4%	36.0%	34.9%	29.1%
	2000–2010	4.27	85.5%	1.53	44.7%	58.9%	18.4%	22.7%
SE	1954–2013	2.78	81.5%	1.58	36.4%	68.8%	2.0%	29.2%
	1970–1980	1.89	25.1%	2.69	54.5%	33.9%	23.3%	42.7%
	1980–1990	2.22	65.0%	2.01	40.9%	40.7%	21.0%	38.3%
	1990–2000	2.54	76.2%	1.70	37.9%	50.7%	14.1%	35.2%
	2000–2010	2.90	83.0%	1.49	36.1%	63.1%	5.8%	31.1%
CY	1961–2013	2.18	69.6%	1.59	43.2%	24.5%	37.8%	37.6%
	1970–1980	1.62	27.6%	1.81	66.9%	27.5%	34.1%	38.4%
	1980–1990	2.06	53.5%	1.76	56.6%	35.8%	25.8%	38.4%
	1990–2000	2.07	67.9%	1.49	45.8%	16.1%	43.4%	40.5%
	2000–2010	2.19	72.3%	1.65	40.0%	43.9%	21.4%	34.7%

References

[1] Adai, A.T., Date, S.V., Wieland, S., Marcotte, E.M., 2004. LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks. J. Mol. Biol. 340, 179–190.

[2] Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

[3] Erdős, P., Rényi, A., 1959. On random graphs i. Publ. Math. Debrecen 6, 290–297.

[4] Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., Ghahramani, Z., 2010. Kronecker graphs: An approach to modeling networks. J. Mach. Learn. Res. 11, 985–1042.

[5] Leskovec, J., Kleinberg, J., Faloutsos, C., 2007. Graph evolution: Densification and shrinking diameters. ACM Trans. Knowl. Discov. Data 1, 1–41.

[6] McGlohon, M., Akoglu, L., Faloutsos, C., 2008. Weighted graphs and disconnected components: Patterns and a generator, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA. p. 524–532.

[7] Simkin, M.V., Roychowdhury, V.P., 2005. Copied citations create renowned papers? Annals Improb. Res. 11, 24–27.

[8] Šubelj, L., Bajec, M., 2013. Model of complex networks based on citation dynamics, in: Proceedings of the WWW Workshop on Large Scale Network Analysis, Rio de Janeiro, Brazil. pp. 527–530.

[9] Vázquez, A., 2003. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev. E 67, 056104.

[10] Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.