The Science of Science

Dashun Wang and Albert-László Barabási

Appendix

This is the submitted version.
The printed version[1] is published by Cambridge University Press in 2021.

[1] Wang, Dashun, and Albert-László Barabási. 2021. *The Science of Science*. Cambridge University Press. doi:10.1017/9781108610834.

Appendix A1

Modeling team assembly

In any team, there are two types of members: (i) Newcomers, or rookies, who have limited experience and unseasoned skills, but who often bring a refreshing, brash approach to innovation, and (ii) incumbents, the veterans with proven track records, established reputations, and identifiable talents. If we categorize all scientists as either rookies or veterans, we can distinguish four different types of coauthorship links in a joint publication: (1) newcomer-newcomer, (2) newcomer-incumbent, (3) incumbent-incumbent, or, if both are incumbents who have worked together before, (4) repeat incumbent-incumbent.

Varying the proportion of these four types of links within a team allows researchers to develop a model capturing how teams are assembled, which then in turn helps us understand how certain coauthorship patterns impact the team's success [1]. The proportions of newcomer-newcomer, newcomer-incumbent, incumbent-incumbent, and repeat incumbent-incumbent links can be characterized by two parameters: The *incumbency parameter*, p, which represents the fraction of incumbents within a team, and the *diversity parameter*, q, which captures the degree to which veterans involve their former collaborators.

As illustrated in Fig A1.1, the model starts with a pool of newcomers (green circles) who haven't worked with anyone else before. Newcomers turn into incumbents (blue circles) when they are drafted onto a team for the first time. For simplicity, let's assume for now that all teams have the same size. To draft team members, with probability p we draw from the pool of incumbents, and with probability $1-p$, we resort to newcomers. If we decide to draw from the incumbents' pool and there is already another incumbent on the team (second panel), then we have another decision to make: If we need a veteran on the team, are we going to introduce a new one or bring in a past collaborator? This is determined by the diversity parameter q: (i) with probability q, the new member is randomly selected from past collaborators of a randomly selected incumbent already on the team, mimicking the tendency of existing team members to choose past collaborators; (ii) otherwise, the new member is selected at random among all incumbents (second panel).

Consider, for example, the creation of a three-person team (m = 3). At time zero, the collaboration network consists of five agents, all incumbents (blue circles). Along with the incumbents, there is a large pool of newcomers (green circles) eager to join new teams. As a concrete example, let us assume that incumbent 4 is selected as the first member in the new team (leftmost box). Let us also assume that the second agent is an incumbent, too (center-left box), which means we need to take the second step, to consider if we should choose from past collaborators or a new veteran. In this example, the second agent is a past collaborator of agent 4, specifically agent 3 (center-right box). Lastly, the third agent is selected from the pool of newcomers; this agent then becomes incumbent 6 (rightmost box).

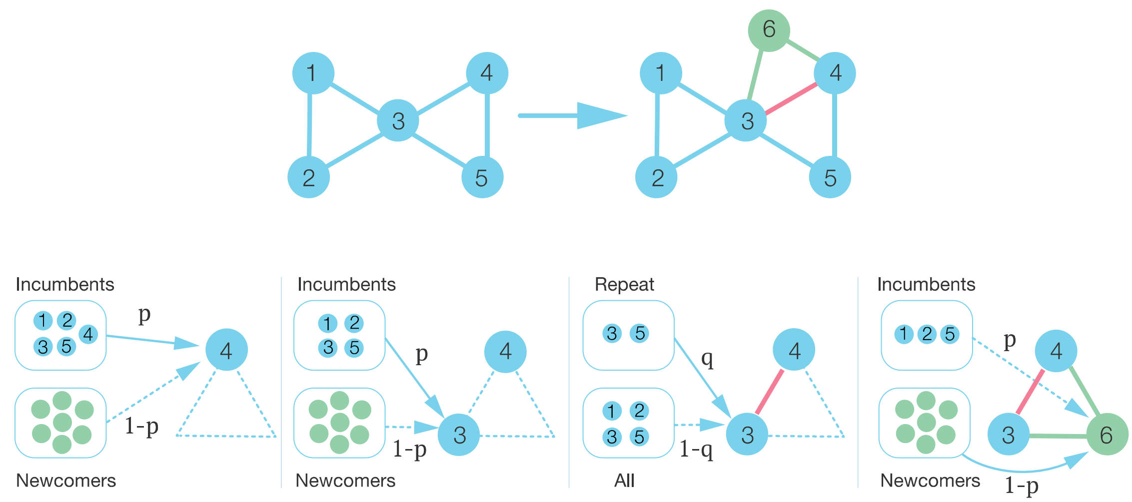

Figure A1.1: **Assembling teams in science.** The model starts with a pool of newcomers (green circles) and incumbents (blue circles). To draft team members, with a probability p we draw from the pool of incumbents, and with a probability 1-p, we resort to newcomers. If we decide to draw from the incumbents' pool, the diversity parameter q determines the likelihood of involving past collaborators: (i) with probability q, the new member is randomly selected from past collaborators of a randomly selected incumbent already on the team; (ii) otherwise, the new member is selected at random among all incumbents (center-left box). After Guimera *et al.* [1]

The model predicts two distinct outcomes for the coauthorship network. The precise shape of the network is largely determined by the incumbency parameter p: When p is small, choosing experienced veterans is not a priority, offering ample opportunities for newcomers to enter the field. Yet frequently choosing rookies who have not worked with others before means that the coauthorship network will be fragmented into many small teams with little overlap. Increasing p increases the likelihood of having veterans on the team, thus increasing the chance that a team connects with other teams in the same network. Indeed, because veterans are on multiple teams, they are the source of overlaps throughout the network.

Therefore, as *p* increases, the formerly fragmented teams start to form a larger cohesive cluster within the coauthorship network.

Interestingly, comparing the incumbency and diversity parameters (*p* and *q*) and a journal's impact factor, which serves as a proxy for the overall quality of the team's output, researchers find that the impact factor is positively correlated with the incumbency parameter *p*, but negatively correlated with the diversity parameter, *q*. This means that teams publishing in high-impact journals often have a higher fraction of incumbents. On the other hand, the negative correlation between the journal impact factor and diversity parameter, *q*, implies that teams suffer when they are composed of incumbents who mainly choose to work with prior collaborators. While these kinds of team alignments may breed familiarity, they do not breed the ingenuity that new team members can offer.

Appendix A2

Modeling Citations

A2.1 The Price model

The Price model can explain the citation disparity among scientific publications, and the universal, field-independent nature of citation distributions. To formalize the citation process, let's have *m* represent the number of citations on a given paper's reference list. When a scientist cites a paper, he/she does not choose it at random. Rather, the probability that the new paper cites paper *i* depends on how many citations *i* has already received, c_i :

$$\Pi_i = \frac{c_i}{\Sigma_i c_i}, \qquad \text{(A2.1.1)}$$

an expression known in the network science literature as preferential attachment [2].This means that when a scientist is choosing between two papers to cite, if one has twice as many citations as the other, she is about twice as likely to pick the more cited paper.

As written, preferential attachment (A2.1.1) leads to a catch-22: If a paper has no citations yet ($c_i = 0$), it can not attract new citations. We can resolve this by acknowledging that each new paper has a finite initial probability of being cited for the first time, called the initial attractiveness of a paper, c_0. Hence the probability that a paper *i* is cited can be modified as

$$\Pi_i = \frac{c_i + c_0}{\Sigma_i (c_i + c_0)}. \qquad \text{(A2.1.2)}$$

The model described above captures two key aspects of citations.

1. **The growth of the scientific literature.** New papers are continuously published, each of which cite *m* previous papers.

2. **Preferential attachment.** The probability that an author chooses a particular paper to cite is not uniform, but proportional to how many citations the paper already has.

The model with Eq. (A2.1.2) was first proposed by de Solla Price in 1976, hence it is sometimes called the *Price model* [3]. It allows us to analytically calculate the distribution of the number of citations received by papers, yielding:

$$p_c \sim (c + c_0)^{-\gamma}, \tag{A2.1.3}$$

where the citation exponent γ follows

$$\gamma = 2 + \frac{c_0}{m}. \tag{A2.1.4}$$

For $c \gg c_0$, (A2.1.3) becomes $p_c \sim c^{-\gamma}$, predicting a power-law citation distribution. Equation (A2.1.3) is in remarkable agreement with the citation distribution observed by Price in 1965 [4], as well as later measurements [5-9]. It predicts that the citation exponent (A2.1.4) is strictly greater than two. Many empirical measurements put the exponent around $\gamma = 3$, consistent with the case of $c_0 = m$.

Box A2.1: Do the Rich Really Get Richer?

Preferential attachment relies on the assumption that new papers tend to cite highly cited papers. But how do we know that preferential attachment is actually present when it comes to citations? We can answer this by measuring the citation rate of a paper (citations per year, for example) as a function of its existing citation count [20]. If preferential attachment is active, then a paper's citation rate must be linearly proportional to its total citations. Measurements have shown that this is indeed the case [10-12], offering direct empirical support for the presence of preferential attachment.

When does the rich-get-richer effect start to kick in? The answer lies in the initial attractiveness parameter, c_0, introduced in Eq. (A1.2). According to it, when a paper has very few citations ($c < c_0$), its chance of getting cited is determined mainly by the initial attractiveness, c_0. To see when a paper begins to benefit from the rich-get-richer effect, we can compare the predictions of Price's model with the Barabasi-Albert model, which ignores the initial attractiveness [13], finding that the tipping point for preferential attachment is around $c_0 \approx 7$. That is, before a paper acquires 7 citations, its citations accumulate as though preferential attachment does not apply. Preferential attachment only turns on once the paper gets past this 7-citation threshold (Fig. A2.1.1).

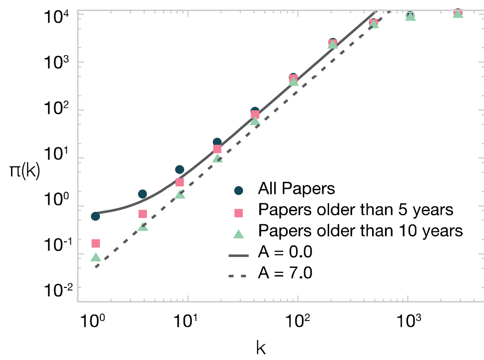

Figure A2.1.1 Empirical Validation of Initial Attractiveness and Preferential Attachment. The solid line captures the case of initial attractiveness $c_0 = 7$ citations. The dashed line corresponds to the case without initial attractiveness ($c_0 = 0$). After Eom and Fortunato [9]

A2.2 The origins of preferential attachment

The rich-get-richer effect might seem to suggest that each scientist meticulously keeps track of the citation counts of every paper, so that she can cite the more cited ones. This is obviously not the case. So where does preferential attachment come from? We can answer this question by inspecting the way we encounter and cite new papers. A common way to discover the research relevant to one's work is by using the papers we have read to find other papers related to our topic of interest. That is, when we read a paper, we tend to inspect its reference list, ocassionally choosing to read some of the papers referenced therein. When we later cite a paper that we discovered by reading another paper, we are effectively "copying" a citation from the earlier paper.

This copying process can help explain the origins of preferential attachment [14-21]. More specifically, imagine a scientist who is deciding which papers to cite in his next paper. He could pick a random paper from the literature, something he encountered by searching for the appropriate topic or keywords. If he only chooses papers in this random way, the resulting citation distribution would follow a Poisson distribution, meaning that every citation is a random, independent event. Imagine, however, that the scientist occasionally chooses to "copy" one of the references of the paper that he randomly selected and cite that reference instead. As simple as it is, this act of copying naturally leads to preferential attachment. This is because a paper with a large number of citations will inevitably be included in multiple papers' reference lists. So, the more citations a paper has, the more likely it is that it will show up on the reference list of the paper we chose, and so the more likely it is that we will cite it.

The beauty of the copy model is that it does not require us to keep track of the citation counts of the published papers. Rather, preferential attachment arises naturally and necessarily from the model, if individuals only rely on local information (i.e. the reference lists of the papers they read previously) to find additional papers to cite. The copy model is not merely a theory—the fingerprints of this process can be detected directly in the data (Box A2.2).

Box A2.2: Evidence of citation copying

With more than 220,000 citations on Google Scholar, John P. Perdew, the pioneer of density functional theory, is one of the world's most cited physicists. His 1992 *Physical Review B* paper [22], coauthored with Y. Wang, has alone collected over 20,000 citations. However, Perdew himself has noted that thousands of those citations were likely misplaced, as many of those authors apparently intended to cite a completely different paper. Perdew and Wang had coauthored another lesser-known paper just a year before their breakthrough—but in some popular references, the paper was mistakenly listed as the more cited 1992 paper.

Analyzing such *citation misprints* can offer direct empirical evidence for citation copying [23]. Occasionally a reference to a certain paper will include a typo; for example, one digit of a paper's 4-digit page number may be misprinted. If the same misprinted number shows up repeatedly in many reference lists, this suggests that citations were simply copied from earlier publications. Indeed, the chance that multiple researchers make the same mistake independently is very low (10^{-4} in this example). Yet when researchers [23] traced a particular citation misprint back to a relatively well-known paper [24], they found that subsequent citations were disproportionately likely to carry the exact same typo. Although different authors cited the article 196 different ways, the citation with the typo was observed 78 times, suggesting that those who cited this paper must have simply copied the reference from some other paper. These repeated misprints indicate that citation "copying" is not just metaphorical, but can be quite literal.

A2.3 The Fit Get Richer

Price's model assumes that the growth rate of a paper's citations is determined solely by its current number of citations. To build upon this basic model, let's assume that citation rate is driven by both preferential attachment and a paper's fitness. This is called the *fitness model* or the Bianconi-Barabási model [25, 26], which incorporates the following two assumptions:

- **Growth**: In each time step, a new paper i with m references and fitness η_i is published, where η_i is a random number chosen from a distribution $p(\eta)$. Once assigned, the paper's fitness does not change over time.
- **Preferential Attachment**: The probability that the new paper cites an existing paper i is proportional to the product of paper i's previous citations and its fitness η_i,

$$\Pi_i = \frac{\eta_i c_i}{\sum_j \eta_j c_j} \quad \text{(A2.3.1).}$$

In (A2.3.1) the probability's dependence on c_i captures the preferential attachment mechanism we have discussed earlier. Its dependence on η_i indicates that between two papers with the same number of citations (c_i), the one with higher fitness will attract citations at a higher rate. Hence, (A2.3.1) assures that even a relatively new paper, with a few citations initially, can acquire citations rapidly if it has greater fitness than other papers.

A2.4 Minimal citation model for individual papers.

In Part 3, we discussed four different mechanisms that are shown to affect the impact of a paper [13]: the *exponential growth* of science (Ch. **3.1**), *preferential attachment* (Ch. **3.3**), *fitness* (Ch. **3.3**), and *aging* (Ch. **3.5**). Combining these four mechanisms allows us to build a minimal citation model that captures the time evolution of the citations a paper receives [13]. To do so, we write the probability that paper i is cited at time t after publication as

$$\Pi_i(t) \sim \eta_i c_i^t P_i(t)\ , \quad \text{(A2.4.1)}$$

In (A2.4.1), η_i captures the paper's fitness, which is a collective measure capturing the community's response to the work and c_i measures preferential attachment, indicating that the paper's probability of being cited is proportional to the total number of citations it has received previously. Lastly, the long-term decay in the paper's citations is well approximated by a log-normal survival probability function

$$P_i(t) = \frac{1}{\sqrt{2\pi}\sigma_i t} exp\left(-\frac{(ln\, t-\mu_i)^2}{2\sigma_i^2}\right), \quad \text{(A2.4.2)}$$

where t is time elapsed since publication; μ captures the immediacy of impact, governing the time required for a paper to reach its citation peak; and σ is longevity, capturing the decay rate of citations.

The growth rate (A2.4.1) helps us calculate the rate at which paper i acquires new citations at time t after its publication,

$$\frac{dc_i^t}{dN} = \frac{\Pi_i}{\sum_{i=1}^{N} \Pi_i}, \qquad \text{(A2.4.3)}$$

Here N represents the total number of papers, with $N(t) \approx \exp(\beta t)$, where β characterizes the rate of science's exponential growth (Ch 1.1). The rate equation (A2.4.3) tells us that with the publication of each new paper, paper i has a smaller and smaller chance of acquiring an additional citation. The analytical solution of the master equation (A2.4.3) leads to the closed-form solution, (3.6.1), predicting the cumulative number of citations acquired by paper i at time t after publication.

Table A1 The average citation counts $< c >$ up to 2012 for all subject categories in 2004 [27]. The Table also lists the number of publications, N, in each category.

SUBJECT CATEGORIES	$< c >$	N
PHYSICS, ACOUSTICS	8.78	3361
AGRICULTURE, AGRICULTURAL ECONOMICS & POLICY	7.87	592
AGRICULTURE, DAIRY & ANIMAL SCIENCE	8.99	3868
AGRICULTURE, MULTIDISCIPLINARY	12.07	2803
AGRICULTURE, AGRONOMY	10.0	4767
MEDICINE, ALLERGY	18.97	1617
MEDICINE, ANATOMY & MORPHOLOGY	11.79	1022
MEDICINE, ANDROLOGY	11.46	248
MEDICINE, ANESTHESIOLOGY	10.06	4122
PHYSICS, ASTRONOMY & ASTROPHYSICS	21.41	13392
ENGINEERING, AUTOMATION & CONTROL SYSTEMS	12.91	3449
BEHAVIORAL SCIENCES	16.95	3426
BIOLOGY, BIOCHEMICAL RESEARCH METHODS	20.54	9674
BIOLOGY, BIOCHEMISTRY & MOLECULAR BIOLOGY	26.18	43556
BIOLOGY, BIODIVERSITY CONSERVATION	14.03	2117
BIOLOGY	16.13	5302
BIOPHYSICS	19.48	9609
BIOLOGY, BIOTECHNOLOGY & APPLIED MICROBIOLOGY	19.63	13899
MEDICINE, CARDIAC & CARDIOVASCULAR SYSTEMS	20.21	12472
BIOLOGY, CELL & TISSUE ENGINEERING	31.4	322
BIOLOGY, CELL BIOLOGY	32.72	17610
CHEMISTRY, ANALYTICAL	15.04	14446
CHEMISTRY, APPLIED	11.76	7542
CHEMISTRY, INORGANIC & NUCLEAR	12.33	10219
CHEMISTRY, MEDICINAL	14.62	6444
CHEMISTRY, MULTIDISCIPLINARY	21.38	23501
CHEMISTRY, ORGANIC	14.56	16878

CHEMISTRY, PHYSICAL	18.52	29735
MEDICINE, CLINICAL NEUROLOGY	16.95	15563
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	16.77	4690
COMPUTER SCIENCE, CYBERNETICS	10.42	1068
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE	9.83	2890
COMPUTER SCIENCE, INFORMATION SYSTEMS	11.58	4633
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	10.25	5761
COMPUTER SCIENCE, SOFTWARE ENGINEERING	8.85	4718
COMPUTER SCIENCE, THEORY & METHODS	9.59	3918
ENGINEERING, CONSTRUCTION & BUILDING TECHNOLOGY	7.36	2302
MEDICINE, CRITICAL CARE MEDICINE	18.15	3116
CHEMISTRY, CRYSTALLOGRAPHY	8.1	7032
MEDICINE, ORAL SURGERY & MEDICINE	11.11	5040
MEDICINE, DERMATOLOGY	10.08	4808
BIOLOGY, DEVELOPMENTAL BIOLOGY	31.01	3289
ECOLOGY	18.02	9860
EDUCATION, SCIENTIFIC DISCIPLINES	6.32	1930
CHEMISTRY, ELECTROCHEMISTRY	17.34	5539
MEDICINE, EMERGENCY MEDICINE	7.59	1661
ENDOCRINOLOGY & METABOLISM	21.68	11259
ENGINEERING, ENERGY & FUELS	11.9	5977
ENGINEERING, AEROSPACE	4.7	1902
ENGINEERING, BIOMEDICAL	18.82	4717
ENGINEERING, CHEMICAL	10.78	13612
ENGINEERING, CIVIL	7.0	5972
ENGINEERING, ELECTRICAL & ELECTRONIC	11.32	26432
ENGINEERING, ENVIRONMENTAL	16.39	4850
ENGINEERING, GEOLOGICAL	7.08	1406
ENGINEERING, INDUSTRIAL	8.28	3109
ENGINEERING, MANUFACTURING	8.28	3385
ENGINEERING, MARINE	1.06	489
ENGINEERING, MECHANICAL	7.54	8503

ENGINEERING, MULTIDISCIPLINARY	7.3	4443
ENGINEERING, OCEAN	7.5	874
ENGINEERING, PETROLEUM	2.14	1613
BIOLOGY, ENTOMOLOGY	7.76	4371
ENVIRONMENTAL SCIENCES	14.88	16938
BIOLOGY, EVOLUTIONARY BIOLOGY	22.88	3170
AGRICULTURE, FISHERIES	10.79	3495
FOOD SCIENCE & TECHNOLOGY	11.97	9457
FORESTRY	11.42	2811
MEDICINE, GASTROENTEROLOGY & HEPATOLOGY	19.03	7518
BIOLOGY, GENETICS & HEREDITY	25.56	12947
GEOGRAPHY, GEOCHEMISTRY & GEOPHYSICS	15.79	5777
GEOGRAPHY, PHYSICAL	14.6	2230
GEOGRAPHY, GEOLOGY	12.42	1604
GEOGRAPHY, MULTIDISCIPLINARY	11.71	10683
GERIATRICS & GERONTOLOGY	15.1	2387
MEDICINE, HEALTH CARE SCIENCES & SERVICES	11.78	3577
MEDICINE, HEMATOLOGY	25.88	9875
HISTORY & PHILOSOPHY OF SCIENCE	4.18	919
IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY	16.28	1136
MEDICINE, IMMUNOLOGY	22.17	17048
MEDICINE, INFECTIOUS DISEASES	18.47	7727
ENGINEERING, INSTRUMENTS & INSTRUMENTATION	8.28	8599
MEDICINE, INTEGRATIVE & COMPLEMENTARY MEDICINE	10.45	885
LIMNOLOGY	13.27	1208
BIOLOGY, MARINE & FRESHWATER BIOLOGY	12.33	6939
MATERIALS SCIENCE, BIOMATERIALS	23.02	2082
MATERIALS SCIENCE, CERAMICS	7.87	3443
MATERIALS SCIENCE, CHARACTERIZATION & TESTING	4.59	1293
MATERIALS SCIENCE, COATINGS & FILMS	11.03	4993
MATERIALS SCIENCE, COMPOSITES	9.44	1539
MATERIALS SCIENCE, MULTIDISCIPLINARY	13.68	34391

MATERIALS SCIENCE, PAPER & WOOD	4.67	1048
MATERIALS SCIENCE, TEXTILES	5.72	949
BIOLOGY, MATHEMATICAL & COMPUTATIONAL BIOLOGY	20.05	2304
MATHEMATICS	4.61	13390
MATHEMATICS, APPLIED	6.65	11863
MATHEMATICS, INTERDISCIPLINARY APPLICATIONS	10.37	4370
MECHANICS	9.49	10165
MEDICAL ETHICS	6.27	443
MEDICINE, MEDICAL INFORMATICS	11.66	1196
MEDICINE, MEDICAL LABORATORY TECHNOLOGY	11.13	2210
MEDICINE, GENERAL & INTERNAL	19.97	14814
MEDICINE, LEGAL	7.02	993
MEDICINE, RESEARCH & EXPERIMENTAL	20.29	8861
METALLURGY & METALLURGICAL ENGINEERING	8.12	8077
PHYSICS, METEOROLOGY & ATMOSPHERIC SCIENCES	15.86	6720
BIOLOGY, MICROBIOLOGY	19.84	13224
MICROSCOPY	9.9	674
ENGINEERING, MINERALOGY	10.95	1724
ENGINEERING, MINING & MINERAL PROCESSING	6.01	1553
MULTIDISCIPLINARY SCIENCES	48.85	10909
BIOLOGY, MYCOLOGY	10.45	1019
NANOSCIENCE & NANOTECHNOLOGY	20.63	7183
BIOLOGY, NEUROIMAGING	25.95	1430
BIOLOGY, NEUROSCIENCES	23.48	23796
ENGINEERING, NUCLEAR SCIENCE & TECHNOLOGY	6.03	7589
MEDICINE, NURSING	7.86	2365
MEDICINE, NUTRITION & DIETETICS	18.44	4767
MEDICINE, OBSTETRICS & GYNECOLOGY	11.78	7384
GEOGRAPHY, OCEANOGRAPHY	13.66	4159
MEDICINE, ONCOLOGY	23.44	19647
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	10.69	3902
MEDICINE, OPHTHALMOLOGY	11.97	6359

PHYSICS, OPTICS	12.16	12693
BIOLOGY, ORNITHOLOGY	8.36	928
MEDICINE, ORTHOPEDICS	13.0	5607
MEDICINE, OTORHINOLARYNGOLOGY	9.01	3235
BIOLOGY, PALEONTOLOGY	9.7	1559
PARASITOLOGY	11.06	2239
MEDICINE, PATHOLOGY	14.42	5694
MEDICINE, PEDIATRICS	10.69	9553
MEDICINE, PERIPHERAL VASCULAR DISEASE	25.41	8353
MEDICINE, PHARMACOLOGY & PHARMACY	14.65	20991
PHYSICS, APPLIED	14.24	28999
PHYSICS, ATOMIC, MOLECULAR & CHEMICAL	14.32	12973
PHYSICS, CONDENSED MATTER	13.39	22654
PHYSICS, FLUIDS & PLASMAS	12.31	5648
PHYSICS, MATHEMATICAL	11.38	6624
PHYSICS, MULTIDISCIPLINARY	16.17	15438
PHYSICS, NUCLEAR	10.24	4987
PHYSICS, PARTICLES & FIELDS	14.33	8759
PHYSIOLOGY	18.73	7846
BIOLOGY, PLANT SCIENCES	15.45	12844
MATERIALS SCIENCE, POLYMER SCIENCE	14.16	11170
MEDICINE, PRIMARY HEALTH CARE	6.93	1140
MEDICINE, PSYCHIATRY	20.88	9108
MEDICINE, PSYCHOLOGY	17.93	2942
MEDICINE, PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH	15.31	10171
MEDICINE, RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING	15.78	12165
REHABILITATION	11.45	1863
ENGINEERING, REMOTE SENSING	14.94	1301
BIOLOGY, REPRODUCTIVE BIOLOGY	15.26	3710
MEDICINE, RESPIRATORY SYSTEM	16.34	6259

MEDICINE, RHEUMATOLOGY	18.28	3058
ENGINEERING, ROBOTICS	11.17	497
SOIL SCIENCE	11.08	2766
MATERIALS SCIENCE, SPECTROSCOPY	9.56	6648
SPORT SCIENCES	13.44	4701
MATHEMATICS, STATISTICS & PROBABILITY	9.04	4922
SUBSTANCE ABUSE	16.99	1049
MEDICINE, SURGERY	12.25	22687
ENGINEERING, TELECOMMUNICATIONS	9.97	5196
PHYSICS, THERMODYNAMICS	9.76	3809
MEDICINE, TOXICOLOGY	13.82	6214
MEDICINE, TRANSPLANTATION	13.21	4665
MEDICINE, TRANSPORTATION SCIENCE & TECHNOLOGY	5.44	1562
MEDICINE, TROPICAL MEDICINE	11.39	1298
MEDICINE, UROLOGY & NEPHROLOGY	15.58	7784
MEDICINE, VETERINARY SCIENCES	7.64	7967
MEDICINE, VIROLOGY	21.66	4713
WATER RESOURCES	10.86	5490
BIOLOGY, ZOOLOGY	10.07	6684

Bibliography

1. Guimera, R., et al., *Team assembly mechanisms determine collaboration network structure and team performance.* Science, 2005. **308**(5722): p. 697-702.
2. Barabási, A.-L. and R. Albert, *Emergence of scaling in random networks.* science, 1999. **286**(5439): p. 509-512.
3. Price, D.d.S., *A general theory of bibliometric and other cumulative advantage processes.* Journal of the American society for Information science, 1976. **27**(5): p. 292-306.
4. de Solla Price, D.J., *Networks of Scientific Papers.* Science, 1965. **149**(3683): p. 510-515.
5. Redner, S., *How popular is your paper? An empirical study of the citation distribution.* The European Physical Journal B-Condensed Matter and Complex Systems, 1998. **4**(2): p. 131-134.
6. Seglen, P.O., *The skewness of science.* Journal of the American Society for Information Science, 1992. **43**(9): p. 628-638.
7. Vázquez, A., *Statistics of citation networks.* arXiv preprint cond-mat/0105031, 2001.
8. Bommarito, I.I., J. Michael, and D.M. Katz, *Properties of the United States code citation network.* arXiv preprint arXiv:0911.1751, 2009.
9. Eom, Y.-H. and S. Fortunato, *Characterizing and modeling citation dynamics.* PloS one, 2011. **6**(9): p. e24926.
10. Jeong, H., Z. Néda, and A.-L. Barabási, *Measuring preferential attachment in evolving networks.* EPL (Europhysics Letters), 2003. **61**(4): p. 567.
11. Newman, M.E., *Clustering and preferential attachment in growing networks.* Physical Review E, 2001. **64**(2): p. 025102.
12. Redner, S., *Citation Statistics from 110 Years of Physical Review.* Physics Today, 2005. **58**: p. 49.
13. Wang, D., C. Song, and A.-L. Barabási, *Quantifying long-term scientific impact.* Science, 2013. **342**(6154): p. 127-132.
14. Krapivsky, P.L. and S. Redner, *Organization of growing random networks.* Physical Review E, 2001. **63**(6): p. 066123.
15. Peterson, G.J., S. Pressé, and K.A. Dill, *Nonuniversal power law scaling in the probability distribution of scientific citations.* Proceedings of the National Academy of Sciences, 2010. **107**(37): p. 16023-16027.
16. Simkin, M.V. and V.P. Roychowdhury, *Copied citations create renowned papers?* arXiv preprint cond-mat/0305150, 2003.
17. Simkin, M.V. and V.P. Roychowdhury, *A mathematical theory of citing.* Journal of the American Society for Information Science and Technology, 2007. **58**(11): p. 1661-1673.
18. Bentley, R.A., M.W. Hahn, and S.J. Shennan, *Random drift and culture change.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2004. **271**(1547): p. 1443-1450.
19. Vazquez, A., *Knowing a network by walking on it: emergence of scaling.* arXiv preprint cond-mat/0006132, 2000.
20. Kleinberg, J.M., et al., *The web as a graph: measurements, models, and methods*, in *Computing and combinatorics*. 1999, Springer. p. 1-17.
21. Simon, H.A., *On a class of skew distribution functions.* Biometrika, 1955: p. 425-440.
22. Perdew, J.P. and Y. Wang, *Accurate and simple analytic representation of the electron-gas correlation energy.* Physical Review B, 1992. **45**(23): p. 13244.
23. Simkin, M.V. and V.P. Roychowdhury, *Read before you cite!* arXiv preprint cond-mat/0212043, 2002.
24. Kosterlitz, J.M. and D.J. Thouless, *Ordering, metastability and phase transitions in two-dimensional systems.* Journal of Physics C: Solid State Physics, 1973. **6**(7): p. 1181.
25. Bianconi, G. and A.-L. Barabási, *Competition and multiscaling in evolving networks.* EPL (Europhysics Letters), 2001. **54**(4): p. 436.
26. Bianconi, G. and A.-L. Barabási, *Bose-Einstein condensation in complex networks.* Physical Review Letters, 2001. **86**(24): p. 5632.
27. Radicchi, F. and C. Castellano, *A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions.* PLoS One, 2012. **7**(3): p. e33833.