# Power Laws in Citation Distributions: Evidence from Scopus

**Article** *in* SSRN Electronic Journal · February 2014

DOI: 10.2139/ssrn.2397685 · Source: arXiv



**1 author:**

Michal Brzezinski
University of Warsaw
**51** PUBLICATIONS **282** CITATIONS

SEE PROFILE

# Measuring power laws in citation distributions: Evidence from Scopus

Michal Brzezinski

*Faculty of Economic Sciences, University of Warsaw, Poland*

**Abstract**

*Keywords:* power law, Lotka's law, citation distribution, Scopus, goodness of fit, model selection

## 1. Introduction

It is often assumed or derived in the literature belonging to informetrics and related disciplines, that distributions of some items (e.g., articles, citations) produced by some sources (e.g., authors, journals) follow power-law behaviour. These distributions are then said to conform to the Lotka's law, after Lotka (1926). Examples of such distributions include author productivity, occurrence of words, citations received by papers, nodes of social networks, number of authors per paper, scattering of scientific literature in journals, and many others; see Egghe 2005a, cha. 1.4, for a more complete list. In fact, power law models are widely used in many sciences as physics, biology, earth and planetary sciences, economics, finance, computer science, and others (see Newman, 2005; Clauset et al. 2009). Models equivalent to Lotkas law are known as Paretos law in economics (Gabaix, 2009) and as Zipfs law in linguistics (Baayen, 2001).

*Email address:* mbrzezinski@wne.uw.edu.pl (Michal Brzezinski)

## 2. Methods

### *2.1. Fitting power-law model to citation data*

We follow Clauset et al. (2009) in choosing methods for fitting power laws to citation distributions. These authors carefully show that, in general, the appropriate methods depend on whether the data are continuous or discrete. In our case, the latter is true as citations are non-negative integers. Let $x$ be the number of citations received by an article in a given field of science. The probability density function of discrete power-law model is defined as

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_0)}, \tag{1}$$

where $\zeta(\alpha, x_0)$ is the generalized or Hurwitz zeta function. The $\alpha$ is a shape parameter of the power-law distribution, known as the power-law exponent or scaling parameter. The power-law behaviour is usually found only for values greater than some minimum, denoted by $x_0$. In case of citation distributions, Albarrán et al. (2011a,b) show that the power-law behaviour can be found on average only in the top 2% of all articles published in a given scientific field.

what about the bulk of distribution - lower tail and the middle - other models - some citations -

The lower bound on the power-law behaviour, $x_0$, should be therefore estimated if we want to measure precisely in which part of the citation distributions the model applies. Moreover, we need an estimate for $x_0$ if we want to obtain an unbiased estimate of the power-law exponent, $\alpha$.

We estimate $\alpha$ using the maximum likelihood (ML) estimation. The log-likelihood function corresponding to (1) is

$$L(\alpha) = -n \ln \zeta(\alpha, x_0) - \alpha \sum_{i=1}^{n} \ln x_i \tag{2}$$

The ML estimate for $\alpha$ is found by numerical maximization of (2).[1]

[1] Clauset et al. (2009) provide also an approximate method of estimating $\alpha$ for discrete power-law model by assuming that continuous power-law distributed reals are rounded to the nearest integers. However, it this paper we use an exact approach based on maximizing (2).

Following Clauset et al. (2009), we use the following procedure to estimate the lower bound on the power-law behaviour, $x_0$. For each $x \geqslant x_{min}$, we calculate the ML estimate of the power-law exponent, $\hat{\alpha}$, and then we compute the well-known Kolmogorov-Smirnov (KS) statistic for the data and the fitted model. The KS statistic is defined as

$$KS = \max_{x \geqslant x_0} |S(x) - P(x; \hat{\alpha})| , \tag{3}$$

where $S(x)$ is the cumulative distribution function (cdf) for the observations with value at least $x_0$, and $P(x, \hat{\alpha})$ is the cdf for the fitted power-law model to observations for which $x \geqslant x_0$. The estimate $\hat{x}_0$ is then chosen as a value of $x_0$ for which the KS statistic is the smallest. The standard errors for both estimated parameters, $\hat{\alpha}$ and $\hat{x}_0$, are computed with standard bootstrap methods with 1,000 replications.

*2.2. Goodness-of-fit and model selection tests*

The next step in measuring power laws involves testing goodness of fit. A positive result of such a test allows us to conclude that a power-law model is consistent with a given data set. Following Clauset et al. (2009) again, we use a test based on a semi-parametric bootstrap approach.[2] The procedure starts with fitting a power-law model to data using methods described in Section 2.1 and calculating a KS statistic for this fit, $k$. Next, we generate a large number of synthetic data sets that follow the originally fitted power-law model above the estimated $x_0$ and have the same non-power-law distribution as the original data set below $\hat{x}_0$. Then, a power-law model is fitted to each of the generated data sets using the same methods as for the original data set, and the KS statistics are calculated. The fraction of data sets for which their own KS statistic is larger than $k$ is the $p$-value of the test. It represents a probability that the KS statistics computed for data drawn from the power-law model fitted to the original data is at least as large as $k$. The power-law hypothesis is rejected if the $p$-value is smaller than some chosen threshold. Following Clauset et al. (2009), we rule out the power-law

[2]If our data were drawn from a given model, then we could use the KS statistic in testing goodness of fit, because the distribution of the $KS$ statistic is known in such a case. However, when the underlying model is not known or when its parameters are estimated from the data, which is our case, the distribution of the $KS$ statistic must be obtained by simulation.

model if the estimated $p$-value for this test is smaller than 0.1. In the present paper, we use 1,000 generated data sets.[3] If the goodness-of-fit test rejects the power-law hypothesis, we may conclude that the power law has not been found. However, if a data set is well fit by a power law, the question remains if there is an alternative distribution, which is an equally good or better fit to this data set. We need, therefore, to fit some rival distributions and evaluate which distribution gives a better fit. To this end, Clauset et al. (2009) use the likelihood ratio test proposed by Vuong (1989). The test computes the logarithm of the ratio of the likelihoods of the data under two competing distributions, $u$, which is negative or positive depending on which model fits data better. Vuong (1989) showed that in the case of non-nested models, the normalized log-likelihood ratio $v = n^{-1/2}u/\sigma$, where $\sigma$ is the estimated standard deviation of $u$, has a limit standard normal distribution. This result can be used to compute a $p$-value for the test discriminating between the competing models. In case of nested models, Vuong (1989) shows that $2u$ has a limit a chi-squared distribution.

We have followed Clauset et al. (2009) in choosing the following alternative discrete distributions: exponential, Weibull, log-normal, Poisson, Yule and the power law with exponential cut-off.[4] The definitions of these alternative distributions are given in Table 1.

## 3. Data

We use citation data from Scopus, a bibliographic database introduced in 2004 by Elsevier. Scopus is a major competitor to the most-widely used data source in informetric research – Web of Science (WoS) from Thomson Reuters. Scopus covers 29 million records with references going back to 1996 and 21 million pre-1996 records going back as far as 1823. An important limitation of the database is that it does not cover cited references for pre-1996

[3] In this procedure, some statistics other than the standard KS statistic could also be used. One could use, for example, a weighted KS statistic that accounts for the extreme tails of distributions, which was proposed recently by Chicheportiche & Bouchaud (2012).

[4] The power-law with exponential cut-off behaves like the pure power-law model for smaller values of $x$, $x \geqslant x_0$,while for larger values of $x$ it behaves like an expoential distribution. The pure power-law model is nested within the power-law with exponential cut-off, and for this reason the latter always provides a fit at least as good as the former. The Vuong's $u$ statistic for comparing these models will therefore be always negative or zero.

Table 1: Definitions of alternative discrete distributions (discrete log-normal distribution is approximated by rounding the continuous log-normally distributed reals to the nearest integers).

| Distribution name | Probability distribution function |
|---|---|
| Exponential | $(1-\mathrm{e}^{-\lambda})\mathrm{e}^{\lambda x_0}\mathrm{e}^{-\lambda x}$ |
| Weibull | $q^{(x-1)^\beta} - q^{x^\beta}$ |
| Log-normal | $\sqrt{\frac{2}{\pi\sigma^2}}\left[\mathrm{erfc}(\frac{\ln x_0-\mu}{\sqrt{2}\sigma})\right]^{-1}\frac{1}{x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$ |
| Poisson | $\left[\mathrm{e}^{\mu}-\sum_{k=0}^{x_0-1}\frac{\mu^k}{k!}\right]^{-1}\frac{\mu^x}{x!}$ |
| Yule | $(\alpha-1)\frac{\Gamma(x_0+\alpha-1)}{\Gamma(x_0)}\frac{\Gamma(x)}{\Gamma(x+\alpha)}$ |
| Power law with exponential cut-off | $Cx^{-\alpha}\mathrm{e}^{-\lambda x}$ |

articles. Scopus contains 21,000 peer-reviewed journals from more than 5,000 international publishers. Scopus covers about 70% more sources compared to the WoS (López-Illescas et al., 2008), but a large part of the additional sources are low-impact journals. A recent literature review has found that the quite extensite literature that compares WoS and Scopus from the perspective of citation analysis produces mixed resuls (Aghaei Chadegani et al., 2013). However, most of the studies suggest that, at least for the period from 1996 on, the number of citations in both databases is either roughly similar or higher in Scopus than in WoS. Therefore, is seems that Scopus constitutes a useful alternative to WoS from the perspective of modeling citation distributions.

Journals in Scopus are classified under four main subject areas: life sciences (4,200 journals), health sciences (6,500 journals), physical sciences (7,100 journals) and social sciences including arts and humanities (7,000 journals). The four main subject areas are further divided into 27 major subject areas and more than 300 minor subject areas.[5] Journals may be classified under more than one subject area.

The analysis in this paper was performed on the level of 27 Scopus major subject areas of science. From the various document types contained in Scopus, we have selected only articles. For the purpose of comparability with Albarrán & Ruiz-Castillo (2011) and Albarrán et al. (2011a), only the articles published between 1998 and 2002 were considered. Following previous

[5]See Table 2 for a list of 27 .

literature, we have chosen a common 5-year citation window for all articles published in 1998-2002.[6] SeeAlbarrán & Ruiz-Castillo (2011) for a justification of choosing 5-year citation window common for all fields of science.

In order to measure the power-law behaviour of citations, we need data on the right tails of citation distributions. To this end, we have used the Scopus Citation Tracker to collect citations for $\min(100,000; x)$ of the highest cited articles, where $x$ is the actual number of articles published in a given field of science during 1998-2002. This analysis was performed separately for each of the 27 science fields categorized by Scopus.

Descriptive statistics for our data sets are presented in Table 2. In some cases, there was less than 100,000 articles published in a field of science during 1998-2002 and we were able to obtain complete or almost complete distributions of citations (see columns 2-4 of Table 2).[7] In other cases, we have obtained only some part of the relevant distribution. The smallest parts of citation distributions were obtained for Medicine (8.4% of total papers), Biochemistry, Genetics and Molecular Biology (15.7%) and Physics and Astronomy (18.4%). However, even in these cases it seems that the coverage of the right tails of citation distributions is satisfactory for our purposes. Using WoS data for 22 science categories, Albarrán & Ruiz-Castillo (2011) found that power laws account usually for less than 2% of the highest-cited articles.

[6]For example, for articles published in 1998, we have analyzed citations received during 1998-2002, while for articles published in 2002, those received during 2002-2006.

[7]For all fields of science analyzed, there were some articles with missing information on citations. These articles were removed from our samples. However, this has usually affected only about 0.1% of our samples.

Table 2: Descriptive statistics for citation distributions, Scopus, 1998–2002

| | Total number of papers | No. of papers in the sample | % of total papers | Mean no. of citations | Std. Dev. of citations | Max. no. of citations |
|---|---|---|---|---|---|---|
| Agricultural and Biological Sciences | 372575 | 99804 | 26.8 | 15.17 | 14.36 | 628 |
| Arts and Humanities | 47191 | 47074 | 99.8 | 1.256 | 3.357 | 91 |
| Biochemistry, Genetics and Molecular Biology | 636421 | 99819 | 15.7 | 49.09 | 46.29 | 3118 |
| Business, Management and Accounting | 61211 | 61156 | 99.9 | 3.452 | 7.273 | 287 |
| Chemical Engineering | 158673 | 98989 | 62.4 | 7.232 | 9.236 | 344 |
| Chemistry | 416660 | 99398 | 23.9 | 21.07 | 21.17 | 1065 |
| Computer Science | 134179 | 99933 | 74.5 | 6.44 | 18.13 | 2737 |
| Decision Sciences | 27409 | 27393 | 99.9 | 3.467 | 5.496 | 143 |
| Earth and Planetary Sciences | 228197 | 99788 | 43.7 | 14.1 | 17.03 | 1195 |
| Economics, Econometrics and Finance | 49645 | 49559 | 99.8 | 4.652 | 8.653 | 287 |
| Energy | 67076 | 66378 | 99.0 | 2.553 | 5.596 | 334 |
| Engineering | 439719 | 99765 | 22.7 | 11.77 | 15.83 | 971 |
| Environmental Science | 186898 | 99847 | 53.4 | 10.72 | 11.27 | 730 |
| Immunology and Microbiology | 195339 | 99858 | 51.1 | 22.11 | 25.11 | 926 |
| Materials Science | 331310 | 99591 | 30.1 | 12.48 | 14.49 | 697 |
| Mathematics | 193740 | 99922 | 51.6 | 6.912 | 11.38 | 929 |
| Medicine | 1191154 | 99823 | 8.4 | 48.55 | 60.14 | 4365 |
| Neuroscience | 445181 | 99886 | 22.4 | 18.97 | 20.39 | 771 |
| Nursing | 51283 | 50464 | 98.4 | 5.274 | 12.07 | 518 |
| Pharmacology, Toxicology and Pharmacautics | 179427 | 99757 | 55.6 | 12.19 | 12.28 | 347 |
| Physics and Astronomy | 541328 | 99817 | 18.4 | 24.75 | 31.64 | 3118 |
| Psychology | 104449 | 99736 | 95.5 | 7.446 | 11.55 | 377 |
| Social Sciences | 215410 | 99890 | 46.4 | 6.148 | 8.055 | 519 |
| Veterinary | 53203 | 53117 | 99.8 | 3.637 | 5.843 | 128 |
| Dentistry | 27470 | 27437 | 99.9 | 4.943 | 6.736 | 115 |
| Health Professions | 75491 | 75414 | 99.9 | 7.272 | 11.49 | 348 |
| Multidisciplinary | 50287 | 50226 | 99.9 | 30.38 | 76.08 | 5187 |
| All Sciences | 6480926 | 2203841 | 34.0 | 14.92 | 27.74 | 5187 |

## 4. Empirical results and discussion

Table 3: Power-law fits to citation distributions, Scopus, 19..

| Science Category | $\hat{x}_0$ | $\hat{\alpha}$ | No. of power-law papers | % of total papers | $p$–value |
|---|---|---|---|---|---|
| Agricultural and Biological Sciences | 92 (15.1) | 4.19(0.25) | 488 | 0.1 | 0.566 |
| Arts and Humanities | 14 (5.4) | 3.46(0.47) | 655 | 1.4 | 0.005 |
| Biochemistry, Genetics and Molecular Biology | 148 (28.0) | 3.72(0.13) | 2813 | 0.4 | 0.175 |
| Business, Management and Accounting | 24 (10.1) | 3.4(0.38) | 1339 | 2.2 | 0.000 |
| Chemical Engineering | 38(6.7) | 4.01(0.19) | 1418 | 0.9 | 0.099 |
| Chemistry | 41(7.1) | 3.4(0.05) | 8193 | 2.0 | 0.110 |
| Computer Science | 26(10.6) | 2.78(0.11) | 3989 | 3.0 | 0.000 |
| Decision Sciences | 12(4.0) | 3.36(0.24) | 1596 | 5.8 | 0.000 |
| Earth and Planetary Sciences | 36(8.9) | 3.37(0.09) | 5834 | 2.6 | 0.000 |
| Economics, Econometrics and Finance | 21(10.2) | 3.13(0.36) | 1995 | 4.0 | 0.000 |
| Energy | 32(5.4) | 3.91(0.22) | 356 | 0.5 | 0.825 |
| Engineering | 26(9.4) | 3.14(0.09) | 7986 | 1.8 | 0.000 |
| Environmental Science | 63(10.3) | 4.33(0.22) | 624 | 0.3 | 0.506 |
| Immunology and Microbiology | 78(13.6) | 3.48(0.10) | 2713 | 1.4 | 0.049 |
| Materials Science | 43(8.9) | 3.47(0.11) | 2687 | 0.8 | 0.193 |
| Mathematics | 24(4.0) | 3.11(0.06) | 4152 | 2.1 | 0.012 |
| Medicine | 59(16.3) | 3.07(0.04) | 20163 | 1.7 | 0.000 |
| Neuroscience | 135(28.4) | 4.69(0.41) | 423 | 0.1 | 0.896 |
| Nursing | 60(15.7) | 3.68(0.40) | 439 | 0.9 | 0.256 |
| Pharmacology, Toxicology and Pharmacautics | 56(6.8) | 4.1(0.12) | 1215 | 0.7 | 0.865 |
| Physics and Astronomy | 61(6.5) | 3.35(0.04) | 5034 | 0.9 | 0.797 |
| Psychology | 52(8.8) | 3.9(0.17) | 1060 | 1.0 | 0.812 |
| Social Sciences | 24(6.4) | 3.56(0.15) | 2963 | 1.4 | 0.007 |
| Veterinary | 23(4.0) | 4.09(0.27) | 858 | 1.6 | 0.017 |
| Dentistry | 20(2.4) | 3.89(0.18) | 1012 | 3.7 | 0.011 |
| Health Professions | 49(10.2) | 3.85(0.24) | 942 | 1.2 | 0.352 |
| Multidisciplinary | 209(40.4) | 3.24(0.14) | 1147 | 2.8 | 0.100 |
| All Sciences | 186(46.3) | 3.45(0.10) | 6364 | 0.2 | 0.076 |

Table 4: Model selection tests for citation distributions, Scopus, 19..

| Science category | $p$-value | Exponential | | Weibull | | Log-normal | | Poisson | | Yule | | PL with cut-off | | Support |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LR | $p$ | LR | $p$ | LR | $p$ | LR | $p$ | LR | $p$ | LR | $p$ | for power law |
| Agricultural and Biological Sciences | 0.566 | 20.740 | 0.009 | 0.338 | 0.779 | -0.096 | 0.782 | 3664.8 | 0.000 | -0.011 | 0.858 | -0.268 | 0.464 | |
| Arts and Humanities | 0.005 | 6.287 | 0.457 | -6.93 | 0.023 | -6.56 | 0.025 | 742.25 | 0.000 | -1.38 | 0.000 | -7.37 | 0.000 | |
| Biochemistry, Genetics and Molecular Biology | 0.175 | 204.5 | 0.000 | 1.22 | 0.758 | -1.12 | 0.473 | 67812.8 | 0.000 | -0.155 | 0.108 | -0.567 | 0.287 | |
| Business, Management and Accounting | 0.000 | 34.390 | 0.034 | -9.60 | 0.013 | -9.24 | 0.013 | 3790.4 | 0.000 | -1.39 | 0.000 | -9.98 | 0.000 | |
| Chemical Engineering | 0.099 | 69.480 | 0.001 | -0.021 | 0.994 | -0.972 | 0.480 | 4978.3 | 0.000 | -0.358 | 0.187 | -0.78 | 0.211 | |
| Chemistry | 0.110 | 736.0 | 0.000 | 7.48 | 0.262 | -2.67 | 0.204 | 69748.9 | 0.000 | -0.999 | 0.060 | -3.31 | 0.010 | |
| Computer Science | 0.000 | 609.4 | 0.000 | -7.05 | 0.248 | -8.80 | 0.035 | 59824.4 | 0.000 | -2.00 | 0.000 | -5.23 | 0.001 | |
| Decision Sciences | 0.000 | 77.730 | 0.001 | -6.71 | 0.046 | -6.81 | 0.048 | 2580.4 | 0.000 | -2.66 | 0.000 | -5.91 | 0.001 | |
| Earth and Planetary Sciences | 0.000 | 459.7 | 0.000 | -4.69 | 0.451 | -7.52 | 0.045 | 42409.3 | 0.000 | -1.95 | 0.000 | -5.69 | 0.001 | |
| Economics, Econometrics and Finance | 0.000 | 45.080 | 0.021 | -21.6 | 0.000 | -20.4 | 0.000 | 6385.7 | 0.000 | -2.68 | 0.000 | -22.9 | 0.000 | |
| Energy | 0.825 | 20.630 | 0.065 | 0.357 | 0.789 | -0.072 | 0.838 | 1176.5 | 0.002 | -0.023 | 0.884 | -0.119 | 0.625 | |
| Engineering | 0.000 | 825.5 | 0.000 | - | - | -7.98 | 0.032 | 58523.9 | 0.000 | -2.71 | 0.000 | -7.52 | 0.000 | |
| Environmental Science | 0.506 | 26.730 | 0.104 | 0.003 | 0.999 | -0.422 | 0.685 | 3034.7 | 0.001 | -0.114 | 0.334 | -0.18 | 0.547 | |
| Immunology and Microbiology | 0.049 | 170.3 | 0.000 | -1.85 | 0.539 | -2.48 | 0.176 | 35217.4 | 0.000 | -0.268 | 0.076 | -3.98 | 0.005 | |
| Materials Science | 0.193 | 233.4 | 0.000 | 2.02 | 0.610 | -1.02 | 0.460 | 22545.5 | 0.000 | -0.412 | 0.178 | -0.850 | 0.192 | |
| Mathematics | 0.012 | 414.8 | 0.000 | -1.54 | 0.784 | -4.97 | 0.083 | 27866.8 | 0.000 | -1.56 | 0.007 | -5.19 | 0.001 | |
| Medicine | 0.000 | 2740.0 | 0.000 | - | - | -7.78 | 0.043 | 468152.0 | 0.000 | -2.03 | 0.000 | -5.62 | 0.001 | |
| Neuroscience | 0.896 | 11.920 | 0.072 | -0.018 | 0.987 | -0.178 | 0.726 | 3098.1 | 0.000 | -0.020 | 0.637 | -0.285 | 0.451 | |
| Nursing | 0.256 | 21.520 | 0.012 | -0.284 | 0.803 | -0.372 | 0.580 | 3193.4 | 0.000 | -0.045 | 0.565 | -0.733 | 0.226 | |
| Pharmacology, Toxicology and Pharmacautics | 0.865 | 47.520 | 0.000 | -0.361 | 0.844 | -0.747 | 0.449 | 5413.9 | 0.000 | -0.1480 | 0.337 | -1.24 | 0.115 | |
| Physics and Astronomy | 0.797 | 706.2 | 0.000 | 19.5 | 0.006 | 0.048 | 0.646 | 97142.8 | 0.000 | 0.091 | 0.771 | 0.000 | 1.000 | |
| Psychology | 0.812 | 53.220 | 0.000 | 0.186 | 0.920 | -0.460 | 0.562 | 5597.1 | 0.000 | -0.112 | 0.475 | -0.791 | 0.208 | |
| Social Sciences | 0.007 | 173.3 | 0.000 | -3.56 | 0.366 | -4.27 | 0.114 | 9621.9 | 0.000 | -1.43 | 0.007 | -4.21 | 0.004 | |
| Veterinary | 0.017 | 38.090 | 0.000 | 0.841 | 0.598 | -0.183 | 0.677 | 1402.2 | 0.000 | -0.047 | 0.874 | -0.542 | 0.298 | |
| Dentistry | 0.011 | 11.830 | 0.200 | -6.60 | 0.025 | -6.26 | 0.028 | 1173.6 | 0.000 | -1.28 | 0.000 | -7.14 | 0.000 | |
| Health Professions | 0.352 | 38.620 | 0.001 | -0.944 | 0.599 | -1.10 | 0.352 | 4470.9 | 0.000 | -0.192 | 0.189 | -1.63 | 0.071 | |
| Multidisciplinary | 0.100 | 98.560 | 0.001 | -1.37 | 0.595 | -1.67 | 0.339 | 64131.9 | 0.000 | -0.067 | 0.069 | -1.44 | 0.090 | |
| All_Sciences | 0.076 | 672.3 | 0.000 | 18.30 | 0.009 | -0.125 | 0.797 | 289249.0 | 0.000 | -0.054 | 0.625 | -0.240 | 0.488 | |

## 5. Conclusions

**Acknowledgements:** I would like to acknowledge gratefully the use of Matlab and R software written by Aaron Clauset and Cosma R. Shalizi, which implements empirical methods used in this paper. The software can be obtained from http://tuvalu.santafe.edu/∼aaronc/powerlaws/. Any remaining errors are my responsibility.

## References

Aghaei Chadegani, A., Salehi, H., Md Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science*, *9*, 18–26.

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011a). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, *88*, 385–397.

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011b). *The skewness of science in 219 sub-fields and a number of aggregates*. Working Paper 11-09 Universidad Carlos III.

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology*, *62*, 40–49.

Chicheportiche, R., & Bouchaud, J.-P. (2012). Weighted kolmogorov-smirnov test: Accounting for the tails. *Physical Review E*, *86*, 041115.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*, 661–703.

López-Illescas, C., de Moya-Anegón, F., & Moed, H. F. (2008). Coverage and citation impact of oncological journals in the¡ i¿ web of science¡/i¿ and¡ i¿ scopus¡/i¿. *Journal of Informetrics*, *2*, 304–316.

Lotka, A. (1926). The frequency distribution of scientific productivity. *Journal of Washington Academy Sciences*, *16*, 317–323.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*, 307–333.

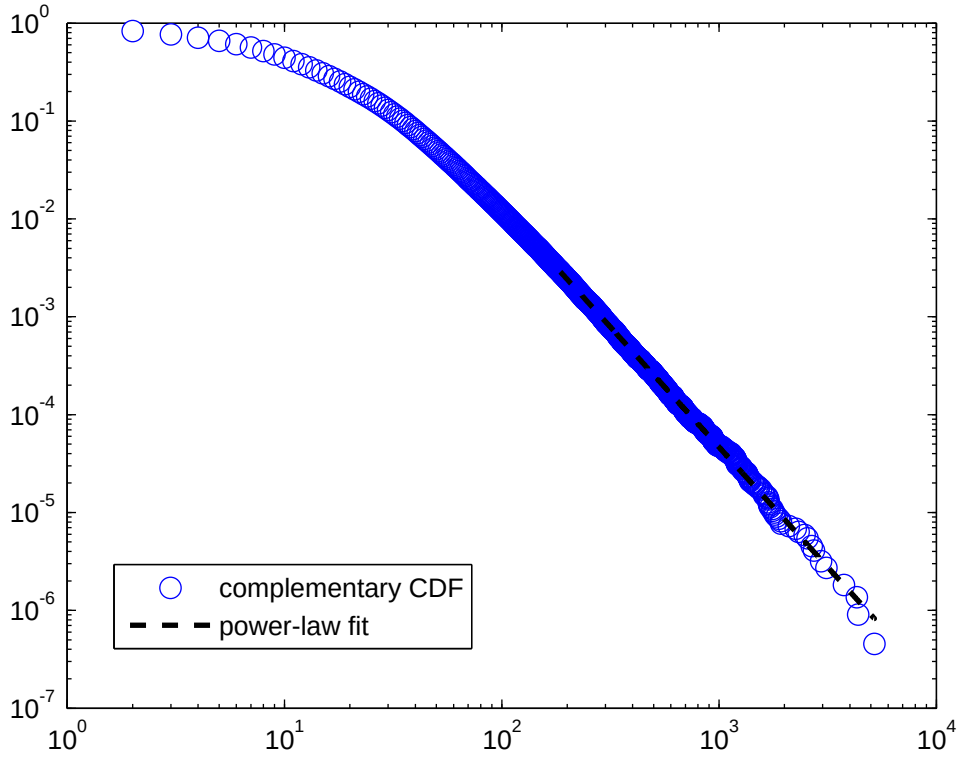


Figure 1: Power-law fit to citation data for all sciences, Scopus, 19...

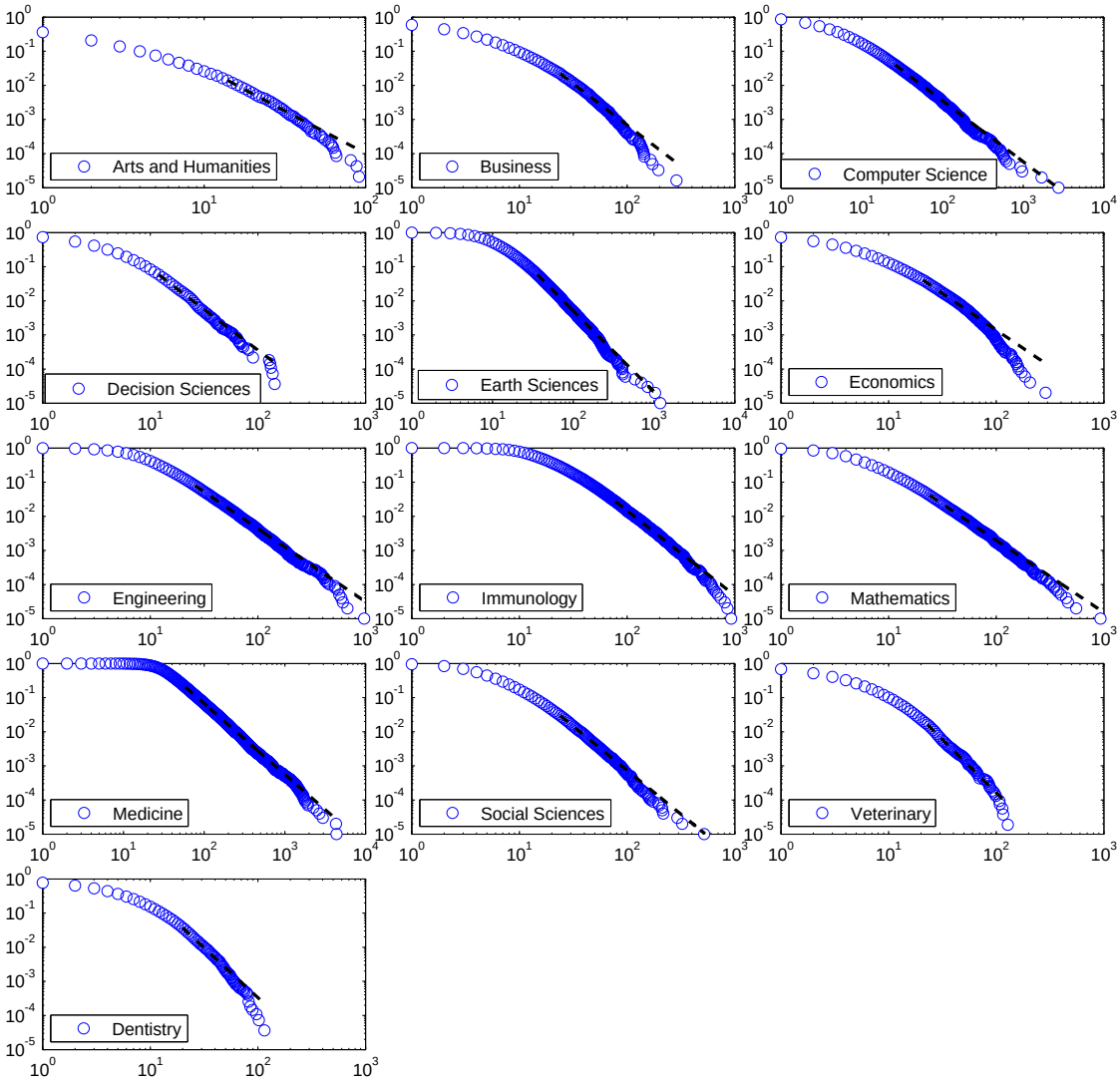


Figure 2: Bad fits of power-law model to citation distributions, Scopus, 19...