OPEN ACCESS
Advances in Complex Systems
Vol. 26, No. 2 (2023) 2350006 (9 pages)
© The Author(s)
DOI: 10.1142/S0219525923500066

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

IS DISRUPTION DECREASING, OR IS IT ACCELERATING?

R. ALEXANDER BENTLEY*
College of Emerging and Collaborative Studies, University of Tennessee,
Knoxville, TN 37996, USA
rabentley@utk.edu

SERGI VALVERDE
Institute of Evolutionary Biology, Spanish National Research Council,
Barcelona 08003, Spain
s.valverde@csic.es

JOSHUA BORYCZ
Stevenson Science and Engineering Library,
Vanderbilt University,
Nashville, TN 37203, USA
joshua.d.borycz@vanderbilt.edu

BLAI VIDIELLA
Institute of Evolutionary Biology,
Spanish National Research Council,
Barcelona 08003, Spain
blai.vidiella@ibe.upf-csic.es

BENJAMIN D. HORNE
School of Information Sciences, University of Tennessee,
Knoxville, TN 37996, USA
bhorne6@utk.edu

SALVA DURAN-NEBREDA
Institute of Evolutionary Biology,
Spanish National Research Council,
Barcelona 08003, Spain
salva.duran@ibe.upf-csic.es

MICHAEL J. O'BRIEN
Department of History,
Texas A&M University, San Antonio, TX 78224, USA

*Corresponding author.
This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC BY) License which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Department of Life Sciences, Texas A&M University,
San Antonio, TX 78224, USA
Mike.obrien@tamusa.edu

Received 11 April 2023
Revised 17 July 2023
Accepted 10 August 2023
Published 13 September 2023

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

A recent highly publicized study [Park, M., Leahey, E. and Funk, R. J., Papers and patents are becoming less disruptive over time, *Nature* **613** (2023) 138–144] claiming that science has become less disruptive over recent decades represents an extraordinary achievement but with deceptive results. The measure of disruption, CD_5, in this study does not account for differences in citation amid decades of exponential growth in publication rate. In order to account for both the exponential growth as well as the differential impact of research works over time, here we apply a weighted disruption index to the same dataset. We find that, among research papers in the dataset, this weighted disruption index has been close to its expected neutral value over the last fifty years and has even increased modestly since 2000. We also show how the proportional decrease in unique words is expected in an exponentially growing corpus. Finding little evidence for recent decrease in disruption, we suggest that it is actually increasing. Future research should investigate improved definitions of disruption.

Keywords: Computational social science; bibliometrics; innovation; science of science.

1. Introduction

Sixty years ago, Kuhn [1] advanced the study of scientific paradigms as a punctuated leader–follower process. Becoming a paradigm in itself, this models science to be a social process, which does not entirely meet the ideal of objective hypothesis testing to surface the best ideas [2, 3]. As a social process of learning from others, scientific research is driven not only by necessity, validity, or originality of the research but also by the popularity of a publication [4] and the prestige of a journal and/or scholar [5, 6], for example. These human-related factors are then magnified by social learning between researchers through any combination of team dynamics, fashions, prestige bias, popular trends, and other social-learning strategies [7–13].

Like "paradigm", the related term "disruption" became fashionable in the early 1960s. Generally meaning radical change in an industry or market, disruption in science could be seen as the intensity or magnitude of a paradigm shift — the extent to which the old ideas are abandoned in favor of the new. As all science builds on previous ideas, a useful metric for disruption captures the extent to which followers of a new paradigm tend to cite only that "disruptive" paper and not the papers in its own reference list [14]. For it to be "disruptive", this tendency to ignore the references cited by a seminal paper should be beyond the normal decay of references being forgotten as they age [15, 16].

We would expect disruption, in the sense of radical change, to initiate a new paradigm of some size, such that disruptive papers would be highly-cited. Without this aspect, an obscure Paper A, even if cited only once by Paper B, could be defined technically as highly disruptive, if Paper B did not cite any of the references in Paper

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

A. If the disruptiveness of a focal paper is defined as the tendency for papers that cite a focal paper to ignore the papers in the focal-paper's reference list, then it makes sense to factor in the number of times the focal paper has been cited overall. In other words, rather than record simply the proportion of "disruptive" papers — equally weighted, whether cited once, twice or thousands of times — it would be better to measure the proportion of citations that point to disruptive papers.

Using the equally weighted measure, among millions of published papers as well as patents, scientific research appears to have become less disruptive [17] in the citation patterns. This result is based on the observed decrease in both a metric for disruption and the relative fraction of unique words over the last 70 years [17].

Scientific metrics provide evidence about a certain element of reality but may not always reflect trends accurately. The reported decline in the disruption index, CD_5, since 1950 [17] is largely due to the diluting effect of exponential growth in quantity of scientific papers and patent applications, as well as their respective citations, over those decades.

In these results [17], CD_5 measures the proportion of citations that a focal work receives relative to the works cited by the focal work. This index does not account for exponential growth of publications and citations. It also treats all papers equally in the annual average, such that a research paper cited by, say, just two other papers with a CD_5 score of -1 would offset a paper from the same year cited 1000 times with a CD_5 score of $+1$. Averaged across an exponentially increasing number of papers each year, this annually averaged CD_5 has decreased in concave-up fashion [17]. This does not necessarily signal a decrease in disruption as defined by how much subsequent work is affected at an accelerating pace.

This highlights the need to account for decades of growth in numbers of papers (and patents) as well as a corpus of expanding citations. For example, the number of patents in the United States has grown exponentially since the 19th century (www.uspto.gov/patents/milestones). In academic publications, this growth rate averaged about 4% per year for more than a century and about 5% per year since 1950 [18]. Accelerated growth appears to be responsible for the steady decline in disruption as defined in the study [17] because the number of citations is increasing in a similar way (Fig. 1). We confirmed that the annual numbers of research papers in the dataset increased exponentially for over eight decades since 1930 (Fig. 1, blue).

2. A Citation-Weighted Measure of Disruption

Using the term "paper" for simplicity, the disruption metric used in the original study [17] is defined as follows:

$$\mathrm{CD}_5(t) = \frac{1}{n_t} \sum_{i}^{n_t} -2 f_{\mathrm{it}} b_{\mathrm{it}} + f_{\mathrm{it}}, \tag{1}$$

where f_{it} is a binary variable (1 or 0) for including the future n_t papers within the 5-year window that cite ($f = 1$) the focal paper i in year t or not ($f = 0$). Similarly, b_{it} is a binary variable indicating whether the future paper cites a predecessor of

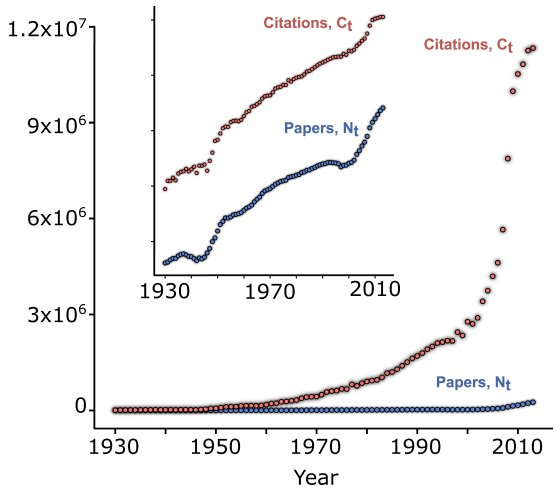

Fig. 1. Decades of exponential growth in scientific publications and citations: C_t and N_t are annual numbers of citations and publications, respectively. Inset shows logarithmic scale. Source data from [17].

the focal paper. Each term in the sum varies from 1 to +1, and each paper published has equal weight in the calculation of the average $\mathrm{CD}_5(t)$ for year t. Note how equation 1 is an average across all papers in year t, such that a paper cited a few times counts as much toward the annual disruption score CD_5 as a paper cited thousands of times.

As we assume instead that highly cited papers contribute more to disruption, here we re-analyze the trend described in [17] using their dataset of millions of scientific publications. For each focal paper, we calculate disruption via a citation-weighted measure, as follows:

$$m\mathrm{CD}_5(t) = \frac{c_{it}}{n_t} \sum_{i}^{n} -2f_{it}b_{it} + f_{it}, \tag{2}$$

where c_{it}, the average number of citations of the focal paper per year, captures the impact of the focal paper [14]. When we sum the $m\mathrm{CD}_5(t)$ for all papers in year t from Eq. (2), we then need to divide by the sum of citations across all papers in year t, a number that also increased exponentially over the eight decades (Fig. 1, red).

In order to calculate this from the data available in the original study [17], we use the $\mathrm{CD}_5(t)$ metric and the average number of times c_{it} each paper is cited each year:

$$m\mathrm{CD}_5(t) = \frac{1}{C_t} \sum_{i}^{N_t} c_{it}\mathrm{CD}_5(i), \tag{3}$$

where C_t is the sum of all the citation counts c_{it} of the N_t papers published in year t. Normalizing in this way helps this weighted measure to account for the exponential growth in the total number of papers and citations over time. Note that, since the number of citations is also growing exponentially, $m\mathrm{CD}_5(t)$ also has an exponentially

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

growing denominator, just as $CD_5(t)$ does. With this normalization, and because the disruption score in the sum ranges from 1 to +1, our null expectation is that Eq. (3) should be close to zero in most years.

This exponential increase in information obscures the overall pattern of disruption. By incorporating citation rate into the disruption score for each publication, we might better estimate the disruptive impact. Here we show how an impact-based measure of disruption, weighted by overall citation count, affects the overall picture. Measured in the dataset from the study [17], the new citation-weighted measure (2) has been close to zero for the 50 years after about 1970 (Fig. 2). Zooming in, we see a modest but consistent growth in weighted disruption since 2000 (Fig. 2, inset), which coincides closely with the digital age of publishing [15], suggesting that technology is a main force behind disruption.

The other effects observed in the study — the decrease in unique words, as well as changes in word frequency — appear to be mostly due to the accelerated growth in the number of publications. This is seen across modern information in general [19, 20], as shown by the word count in English-language books over time [21, 22] in Fig. 3(a).

Although we do not fully understand why things grow exponentially, many technological advances can accelerate cultural changes [23] and promote the need for novel words [20] (Fig. 3(a), inset). In an expanding corpus, the ratio of unique terms to total words naturally declines as information grows. An exponentially growing corpus in books, for example, corresponds to the decline in the ratio of unique to total words per year [24].

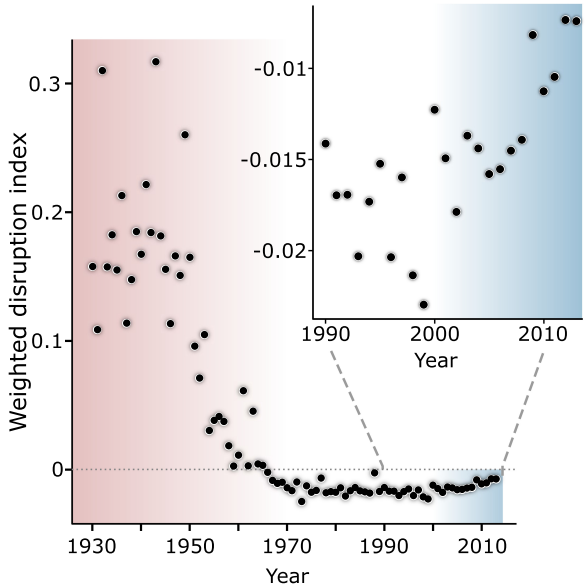

Fig. 2. The citation-weighted disruption index has been close to zero since around 1970. Inset shows a modest increase since 2000 (blue area). Source data from [17].

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

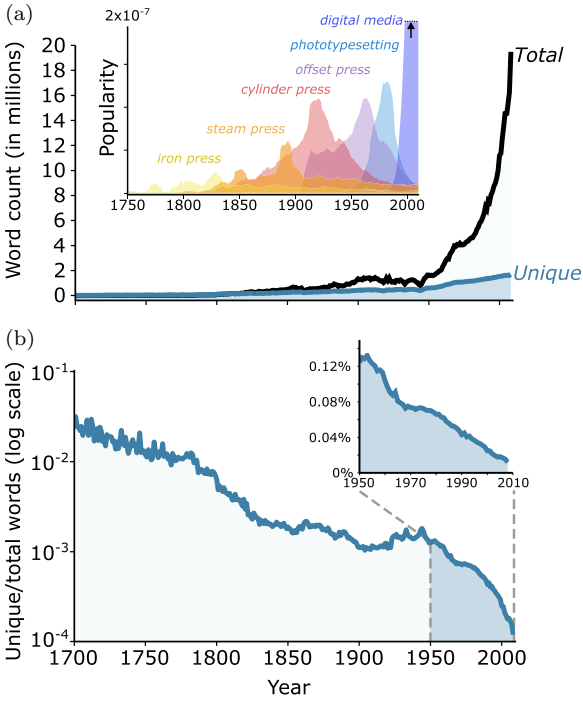

Fig. 3. Corpus growth, fashion cycles, and dilution. (a) Annual count of words (black curve), in billions, and vocabulary (unique words, blue curve), in tens of millions, in English-language books. Inset: popularity cycles of selected technological terms (the max of "digital media" lies above the plot). (b) Unique words (appeared once in the year) as a fraction of total word count, with logarithmic y-axis (inset: 1950–2008, not logarithmic). All data are from the 2012 Google books database (https://books.googlecom/ngrams/info).

In this case, the ratio of unique terms to total words used declines as information grows (Fig. 3(b)). Interestingly, the 80% decrease in this ratio in scientific publications between 1950 and 2010 [17, Fig. 2] matches that in millions of books during that period [26]. This trend has also been reported in scientific publications and in books published in different languages across very different domains [24]. Similarly, the most-used verbs in science and patents have changed over time, reflecting common English language usage, and this pattern can be easily predicted by simple models of word choice [25–27]. For example, the use of the word "produce" has declined over time both in patents and in English books in general (see [17, Fig. 3]).

3. Discussion and Conclusions

Despite the exponentially growing denominator in $m\mathrm{CD}_5$, this citation-weighted measure of disruption has either held steady or increased over the last half century. In other words, not only is the absolute number of disruptive works relatively

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

constant, even the fraction of citations that point to disruptive works has been either constant or increasing. Similarly, Petersen *et al.* [29] show that the decrease in CD_5 reported by Park *et al.* [17] since the mid-20th century is an artifact of how it was defined combined with "citation inflation" — the exponential growth in the number of published papers and citations, as well as the increase in mean number of references per paper over time. The decline in CD_5 is expected due to the increase in the denominator in the expression amid an exponentially expanding corpus [29, Eqs. (1) and (2)]. If, for instance, a limit were imposed by publishers on the number of paper references, the decline in CD_5 would stop (see [29, Fig. 4]).

In summary, we see little evidence that disruption is decreasing. The unweighted disruption metric is going down, but the citation-weighted measure is still going up. With each paper equally weighted regardless of impact, the measure of disruption with number of papers in the denominator (Eq. (1)) is expected to decrease over time amid an exponentially expanding corpus, as should the fraction of unique words. Instead, when we use the citation-weighted measure (Eq. (3)), we find that the disruption measure moves close to its expected value of zero over time, but also increases modestly in the last 20 years. Given this is amidst an exponentially growing corpus, we might even conclude that disruption is increasing in the current age of digital publishing.

Disruption means that the ideas that help form a new idea are disproportionately left behind. Darwin's *Origin of Species* was built upon ideas in geology, geography, and biology at the time [30]. In the 1990s, particular publications on preferential attachment [31] or small-world networks [32] became enormously disruptive, in the sense that they each came to be cited tens of thousands of times — orders of magnitude more than similar contemporary works [33] — despite building on similar ideas of proportional advantage [25, 34] from decades before.

In the near future, machine learning and natural language-processing tools may substantially revise this history of disruption [28]. As academic publications are increasingly read by, and perhaps written by, machines and artificial intelligence, the nature of scientific paradigms and disruption will change, potentially dramatically. Machines can read all digitized papers, no matter how obscure, whereas human researchers are all too prone to follow popular ideas or prestigious thought leaders (e.g. citation count). That could mean that the dynamics of disruption become different, and machines may even discover lost revolutionary ideas that become disruptive long after human learning biases have forgotten them. Academic creative innovation tends to occur in bursts, which lead to paradigm shifts but also founder effects, in which one paper becomes highly cited and other papers of that productive time, with equally original and potentially disruptive ideas, being left behind in the inherently human process of following prestigious leaders and/or popular ideas. With machines reading and summarizing large numbers of obscure papers, however, a measure of disruption per paper, independent of subsequent citation by humans, would become more important.

Acknowledgments

S.V. is supported by the Spanish Ministry of Science and Innovation (MCIN) AEI grant PID2020-117822GB-I00/AEI/10.13039/501100011033. B.V. and S.V. are further supported through the BiodivRestore ERA-NET Cofund (GA No. 101003777) MPA4Sustainability Project, as well as MCIN grant PCI2022-132936 and PIE-202120E047-Conexiones-Life. S.D.-N. is supported by the Beatriu de Pins Programme and the Ministry of Research and Universities of the Government of Catalonia (2019 BP 00206) and the Marie Sklodowska-Curie COFUND (BP3 Contract No. 801370) of the H2020 programme.

ORCID

Alexander Bentley https://orcid.org/0000-0001-9086-2197
Sergi Valverde https://orcid.org/0000-0002-2150-9610
Joshua Borycz https://orcid.org/0000-0002-1505-148X
Blai Vidiella https://orcid.org/0000-0002-4819-7047
Benjamin D. Horne https://orcid.org/0000-0002-5779-3019
Salva Duran-Nebreda https://orcid.org/0000-0002-2539-3539
Michael J. O'Brien https://orcid.org/0000-0001-6433-8295

References

[1] Kuhn, T. S., *The Structure of Scientific Revolutions* (University of Chicago Press, 1962).
[2] Kitcher, P., *The Advancement of Science* (Oxford University Press, 1993).
[3] Hull, D. L., *Science and Selection* (Cambridge University Press, 2001).
[4] Borycz, J. D., Horne, B. D. and Bentley, R. A., Forking paths of cultural complexity: A case study of research paradigms, *Adv. Complex Syst.* **25** (2022) 2240004.
[5] Malmgren, R. D., Ottino, J. M. and Amaral, L. A. N., The role of mentorship in protégé performance, *Nature* **465** (2010) 622–626.
[6] Ma, Y., Mukherjee, S. and Uzzi, B., Mentorship and protégé success in STEM fields, *Proc. Natl. Acad. Sci. USA* **117** (2020) 14077–14083.
[7] Zeng, A., Fan, Y., Di, Z., Wang, Y. and Havlin, S., Fresh teams are associated with original and multi-disciplinary research, *Nat. Hum. Behav.* **5** (2021) 1314–1322.
[8] Evans, J. A. and Foster, J. G., Metaknowledge, *Science* **331** (2011) 721–725.
[9] Brock, W. A. and Durlauf, S. N., A formal model of theory choice in science, *Econ. Theory* **14** (1999) 113–130.
[10] Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., Spring, B., Stokols, D., Trochim, W. and Uzzi B., A multi-level systems perspective for the science of team science, *Sci. Transl. Med.* **2** (2010) 49cm24.
[11] Rivera, M., Soderstrom, S. and Uzzi, B., Dynamics of dyads in social networks, *Ann. Rev. Sociol.* **36** (2010) 91–115.
[12] Guimerà R., Uzzi, B., Spiro, J. and Amaral, L. A. N., Team assembly mechanisms determine collaboration network structure and team performance, *Science* **308** (2005) 697–702.

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

Advs. Complex Syst. 2023.26. Downloaded from www.worldscientific.com
by 161.111.10.231 on 02/26/24. Re-use and distribution is strictly not permitted, except for Open Access articles.

[13] Wuchty, S., Jones, B. F. and Uzzi, B., The increasing dominance of teams in production of knowledge, *Science* **316** (2007) 1036–1039.

[14] Funk, R. J. and Owen-Smith, J., A dynamic network measure of technological change, *Manag. Sci.* **63** (2017) 791–817.

[15] Evans, J. A., Electronic publication and the narrowing of science and scholarship, *Science* **321** (2008) 395–399.

[16] Wu, F. and Huberman, B. A., Novelty and collective attention, *Proc. Natl. Acad. Sci. USA* **104** (2007) 17599–17601.

[17] Park, M., Leahey, E. and Funk, R. J., Papers and patents are becoming less disruptive over time, *Nature* **613** (2023) 138–144.

[18] Bornmann, L., Mutz, R. and Haunschild, R., Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases, *Humanit. Soc. Sci. Commun.* **8** (2020) 115.

[19] Hidalgo, C. A., *Why Information Grows: The Evolution of Order, from Atoms to Economies* (Basic Books, New York, NY, 2015).

[20] Duran-Nebreda, S., O'Brien, M. J., Bentley, R. A. and Valverde, S., Dilution of expertise in the rise and fall of collective innovation, *Humanit. Soc. Sci. Commun.* **9**(1) (2022) 110.

[21] Michel, J.-B., Kui Shen, Y., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A. and Aiden, E. L., Quantitative analysis of culture using millions of digitized books, *Science* **331** (2010) 176–182.

[22] Perc, M., Evolution of the most common English words and phrases over the centuries, *J. Roy. Soc. Interface* **9** (2012) 3323–3328.

[23] Bentley, R. A. and O'Brien, M. J., *The Acceleration of Cultural Change* (M.I.T. Press, Cambridge, MA, 2017).

[24] Petersen, A. M., Tenenbaum, J. N., Havlin, S., Stanley, H. E. and Perc, M., Languages cool as they expand: Allometric scaling and the decreasing need for new words, *Sci. Rep.* **2** (2012) 943.

[25] Simon, H. A., On a class of skew distribution functions, *Biometrika* **42** (1955) 425–440.

[26] Ruck, D. J., Bentley, R. A., Acerbi, A., Garnett, P. and Hruschka, D. J., Role of neutral evolution in word turnover during centuries of English word popularity, *Adv. Complex Syst.* **20** (2017) 1750012.

[27] Vidiella, B., Carrignon, S., Bentley, R. A., O'Brien, M.J. and Valverde, S., A cultural evolutionary theory that explains both gradual and punctuated change, *J. Roy. Soc. Interface* **19** (2022) 20220570.

[28] Bentley, R. A., Borycz, J., Carrignon, S., Ruck, D. J. and O'Brien, M. J., Machine learning for rediscovering revolutionary ideas of the past, *Adaptive Behav.* **30** (2021) 279–286.

[29] Petersen, A. M., Arroyave, F. and Pammolli, F., The disruption index is biased by citation inflation, preprint (2023), arXiv:2306.01949.

[30] Davies, R., How Charles Darwin received Wallace's Ternate paper 15 days earlier than he claimed, *Biol. J. Linn. Soc.* **105** (2012) 472–477.

[31] Barabsi, A.-L. and Albert, R., Emergence of scaling in random networks, *Science* **286** (1999) 509–512.

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

[33] Huberman, B. A. and Adamic, L. A., Growth dynamics of the World-Wide Web, *Nature* **401** (1999) 131.

[34] Yule, G. U., A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, *Philos. Trans. Roy. Soc. London B* **213** (1925) 21–87.