Physica A 428 (2015) 173–177

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Phase transitions in the web of science

J.C. Phillips

*Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, 08854, United States*

## HIGHLIGHTS

- Numbers of scientific papers published annually are growing superlinearly.
- Searching the Web of Science with selected key words reveals discontinuous shifts.
- Shifts are first- or second-order, and resemble phase transitions.
- The transitions occurred in 1990 (1st order) and 2000 (2nd order).
- Specific historical events explain the abruptness of these shifts.

## ARTICLE INFO

*Article history:*
Received 7 August 2014
Available online 11 February 2015

*Keywords:*
Key words
Interdisciplinary
Phase transition
Innovation
Russian jump
Globalization

## ABSTRACT

The Internet age is changing the structure of science, and affecting interdisciplinary interactions. Publication profiles connecting mathematics with molecular biology and condensed matter physics over the last 40 years exhibit common phase transitions indicative of the critical role played by specific interdisciplinary interactions. The strengths of the phase transitions quantify the importance of interdisciplinary interactions.



## 1. Introduction

Mathematical and scientific literature, as measured by numbers of papers published each year, is growing super-linearly, with the number of specialized journals proliferating in the 21st century. Citations are widely used as a measure of the impact of research efforts, if not the absolute importance of their content. This growth has stimulated more than 700 papers analysing the literature scientometrically. The most spectacular scientometric study so far surveyed nearly all the 20th century literature, consisting of 25 million papers and 600 million citations, with $10^6$–$10^7$ more entries than typical data bases and unique in the history of epistemology [1], see also arXiv:0810.1426. This study found that stretched exponentials (SE) give better fits than popular power laws to citation chain distributions for both low and intermediate citation levels $n < n_1$, where $n_1$ is the "crossover to fame". Here $n_1 \sim 40$ earlier in the century, and $n_1 \sim 200$ later. The middle SE region with $n < n_1$ accounts for 95% of the citations, and it, not the 5% scale-free or power law high end $n > n_1$, probably objectively represents the essential features of working research citation patterns.

A physical interpretation of this result is that the self-similarity of the "5%" power laws is an example of the "rich grow richer"; famous papers with $n > n_1$ are often cited unread by many citers. The "95%" stretched exponential is used to describe foraging, and its fractal exponent $\beta$ can be explained topologically. In the topological "diffusion to traps" model, $\beta$ is

*E-mail address:* jcphillips8@comcast.net.

http://dx.doi.org/10.1016/j.physa.2015.02.013
0378-4371/

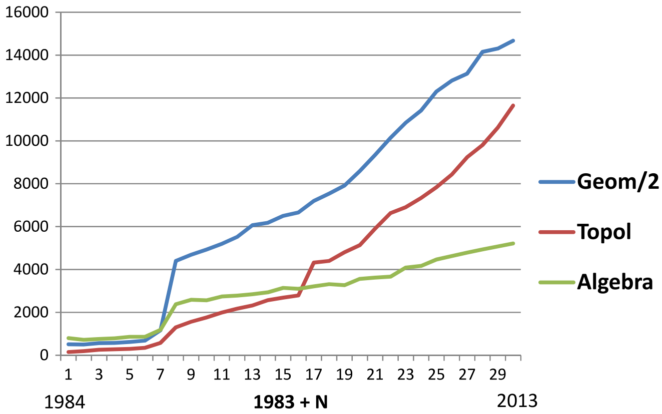


**Fig. 1.** (Colour online) The annual numbers of papers on each of three mathematical disciplines, with the geometry number divided by 2. In the version of the Web of Science Core Collection used in this survey, the entire geometry profile covers 330,000 papers. Other versions may cover different numbers, but the general trends should be similar for the two phase transitions around 1990 and 2000.

either $d/(d+2)$, or $d/(d+4)$, with $d = 3\,(=2+1$, dynamical surface interactions). The former corresponds to short-range interactions only, and the latter to a maximum entropy dissipation rate mixture of short-range and long-range (Lévy flights) actions. The observed crossover from $\beta = 3/5$ to $\beta = 3/7$ in 1960 [1] is then explained as the result of globalization [2]. Several factors contributed to fixing this 1960 crossover, including the appearance of jet travel, and international scientific competition (Sputnik). Physical examples of SE are found in the relaxation of homogeneously interconnected network glasses, where both values of $\beta$ occur, depending on the ranges of interactions and the concentrations of randomly distributed defects [2,3].

While citations have attracted much attention [4,5], an equally or even more important issue is the interdisciplinary impact of scientific research. Here we extend citation analyses to discuss the evolution of several mathematical fields associated with unambiguous keywords, and compare their evolution to two broad scientific fields, condensed matter physics, and molecular biology. The latter are each divided into two subtopics, where phase transitions (thresholds in numbers of papers published each year, all Web of Science data bases) are unmistakable. One of these phase transitions is readily anticipated, and is associated with the Internet, beginning in 2000. The other, which is much larger, may come as a surprise. We attribute it to the Gorbachev emigration of a large number of Russian scientists after 1988.

## 2. Results

### 2.1. Mathematics

We begin with three mathematical disciplines, in the period 1984–2013, shown in Fig. 1. Geometry is the oldest branch of quantitative knowledge, dating originally to prehistoric land surveying at least 6000 years ago. It still produces the largest numbers of papers, which are divided by 2 in Fig. 1. Algebra can be traced back 3800 years (first studies of quadratic equations), and it held second place until recently. Apart from a few isolated examples, topology emerged only in the late 19th century in the work of Cantor and Poincare. As Fig. 1 shows, topology overtook algebra in 2000, with a jump associated with popular interest in the Internet. Both geometry and topology accelerated after 2000, while algebra increased at nearly the same rate as before. The jumps resemble step functions and allowing for small time publication delays, can be described as first-order phase transitions, as the step amplitudes are far above noise level.

Another jump can be seen around 1990: what does it mean? A semilog plot in Fig. 2 makes it easier to discuss the 1990 jump, which is larger for geometry and topology. These are subjects natural to computers, as their internal networks have both geometrical and topological aspects. The rapid advances in computer hardware (Moore's law, about 60%/year) have never been matched by software (much slower, informally estimated at about 20%/year), so by 1990 a large backlog had accumulated in unanalyzed data. At the same time, post-Gorbachev emigration included a large number of not only expert but also resourceful Russian computer scientists, who had had to utilize inferior tools. One of their first actions after emigrating was to publish papers in Western journals, and this appears to be a natural explanation for the 1990 jumps.

### 2.2. Molecular biology

The largest backlogs in unanalyzed scientific data were in the biological sciences, and by 1990 the Russians were providing many "new hands" for analysing emerging complex data on protein sequences and structures. In Fig. 3 we see several interesting features, which are much stronger in analysing gene and amino acid sequences than in discussing structures. The latter exhibits the usual 1990 Russian Jump, already seen in mathematical research in Figs. 1 and 2. The Russian influx contributed to protein structure studies by providing additional engineering support for building new synchrotron sources,