

# How Famous is a Scientist? — Famous to Those Who Know Us.

James P. Bagrow,<sup>1</sup> Hernán D. Rozenfeld,<sup>1</sup> Erik M. Boltt,<sup>2,1</sup> and Daniel ben-Avraham<sup>1</sup>

<sup>1</sup>*Department of Physics, Clarkson University, Potsdam NY 13699-5820*

<sup>2</sup>*Department of Math and Computer Science, Clarkson University, Potsdam, NY 13699-5805*

(Dated: June 4, 2004)

Following a recent idea, to measure fame by the number of *Google* hits found in a search on the WWW, we study the relation between fame (*Google* hits) and merit (number of papers posted on an electronic archive) for a random group of scientists in condensed matter and statistical physics. Our findings show that fame and merit in science are linearly related, and that the probability distribution for a certain level of fame falls off exponentially. This is in sharp contrast with the original findings about WW II ace pilots, for which fame is exponentially related to merit (number of downed planes), and the probability of fame decays in power-law fashion. Other groups in our study show similar patterns of fame as for ace pilots.

PACS numbers: 01.75.+m, 02.50.-r, 87.23.Ge, 89.75.Hc

The concept of fame is socially and economically important to many people, and the organizations to which they belong. However, it is not a well defined concept, since each person has their own idea of what it means to be famous, including perhaps: recognizable to the “common” person on the street (but how do we define the common person?), being on television, appearing frequently in newspapers and in other media. Recently, researchers at UCLA, Simkin and Roychowdhury [1] performed an empirical study in which they catalogued the fame of World War I pilot “Aces.” For reasons of specificity in measurement, they chose an interesting definition of fame: *the number of hits a search for a person’s name garners in the Google search engine*. In this view, our fame is taken to be how well linked we are in what has quickly become a most popular medium — the World Wide Web — and is related to the number of webpages that mention us, as measured by the PageRank system that is behind *Google*’s popular search engine [2]. This is an ingenious idea, in that it provides an inexpensive measurement of social impact, by enlisting powerful computer resources, freely available to all, to perform what would otherwise be an expensive social study. However, we mention that in practice even this measurement is hard to make completely precise without a great deal of effort, due to difficulty in separating coincidences in popular names (doubling and tripling a person’s fame), and also possibly missing a person’s full fame due to too restrictive a search. Nonetheless, in this study, it is precisely this *Google*-hits measure that we adapt, as carefully specified below.

The purpose of this communication is to explore whether there is a difference between relative *fame* and *achievement* (merit) in science, as compared to the findings for ace pilots. In the UCLA study of fighter pilots [1], a pilot’s achievement was measured by how many enemy planes the pilot had downed; was he an ace? It was found that fame increases exponentially with achievement, while the distribution of fame falls off algebraically, nearly as  $(\text{fame})^{-2}$ . A model mechanism behind these findings was presented, describing the social context of fame within a random graph. We have catalogued similar measurements of fame and achievement for scientists working in the area

of condensed matter or statistical physics. We find dramatically different behavior, with fame increasing *linearly* with achievement, and its probability falling off *exponentially*. A simple argument to account for these facts suggests that in difference to ace pilots, that enjoy public renown, scientist are well known mostly within their own community and do not truly reach real fame. Preliminary studies of other groups of people reveal similar fame patterns as for ace pilots.

We begin with a brief summary of the findings in [1]. The distribution of fame  $F$  (as measured by *Google* hits) falls off roughly as an inverse square,

$$P(F) \sim F^{-\gamma}; \quad \gamma \approx 2, \quad (1)$$

while it rises exponentially with achievement  $A$  (number of downed planes),

$$F(A) \sim e^{\beta A}. \quad (2)$$

The two relations imply that achievement ought to be exponentially distributed:

$$P(A) \sim e^{-\alpha A}; \quad \alpha = \beta(\gamma - 1) \approx \beta. \quad (3)$$

This is indeed confirmed from independent measurements.

Our findings for scientists — researchers in the area of condensed matter and statistical physics — are dramatically different. We have examined a list of 449 researchers, drawn randomly from among those who post articles on the web-based electronic board <http://www.arxiv.org/archive/cond-mat>. As a measure of fame, we used the UCLA *Google*-hit criterion, with search lexicon: “Author’s name” AND “condensed matter” OR “statistical physics” OR “statistical mechanics”. The distribution of fame, as measured by the number of hits, decays exponentially, rather than in power-law fashion (Fig. 1):

$$P(F) \sim e^{-\eta F}; \quad \eta = 0.00102 \pm 0.00006. \quad (4)$$

The figure results, as well as the computed correlation for the two possibilities ( $R^2 = 0.977$  for the exponential, vs.

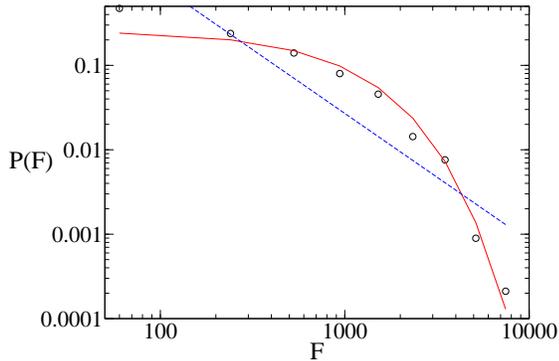


FIG. 1: Distribution of fame in science. The distribution is clearly better fitted by an exponential,  $P(F) \sim 0.26e^{-0.001F}$  (curve,  $R^2 = 0.98$ ), than by a power-law (straight line,  $R^2 = 0.82$ )

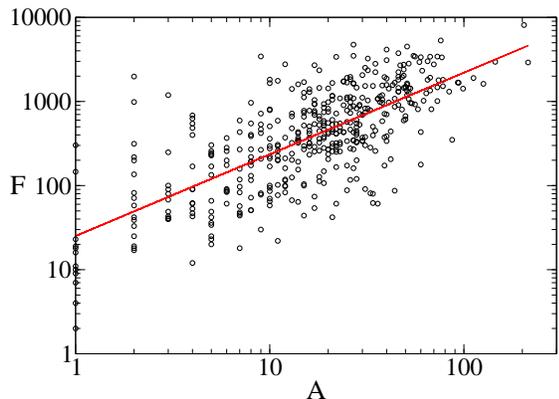


FIG. 2: Fame of individual scientists (*Google* hits) versus achievement (number of papers posted in the /cond-mat e-archive). The data ( $\circ$ ) is better fitted by an almost linear relation,  $F \sim cA$  (straight line), than by the exponential dependence found in [1].

0.82 for a power-law), leave little doubt as to which is the better fit.

Next, we consider the relation between fame and achievement in science. As a simple-minded measure of achievement we take the total number of publications by an author on the cond-mat board, perhaps thus paying too much heed to the popular dictum “publish or perish.” (The cond-mat board has been active since 1991.) Our findings are summarized in Fig. 2. The wide scatter of data points resembles that found for ace pilots by the UCLA team, but our best fit indicates a power-law (almost linear) increase of fame with achievement, rather than the exponential dependence found there. Indeed, we find  $R^2 = 0.513$  for a power-law fit, vs. 0.328 for an exponential: quite the reverse from [1], which (for 393 ace pilots) cites  $R^2 = 0.72$  for the exponential, vs. 0.48 for the linear fit. Our findings suggest the linear relation:

$$F(A) \sim cA^\xi; \quad \xi = 0.97 \pm 0.04 \approx 1. \quad (5)$$

If Eqs. (4), (5) are right (with  $\xi = 1$ ), it follows that

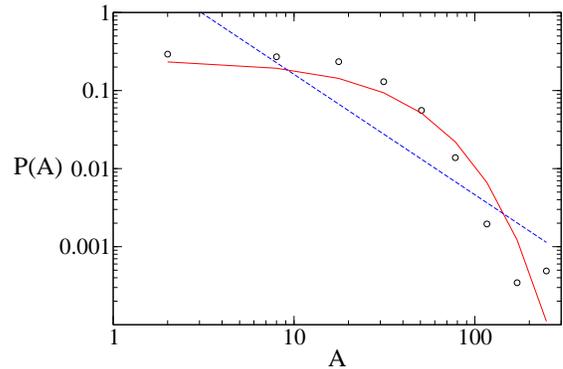


FIG. 3: Distribution of achievement in science, as measured by the number of papers posted by each author on the web archive <http://arxiv.org/archive/cond-mat>. The data ( $\circ$ ) is better fitted by an exponential,  $P(A) = 0.25e^{-0.031A}$  (curve,  $R^2 = 0.90$ ), than by a power-law (straight line,  $R^2 = 0.76$ ).

achievement in science is distributed exponentially:

$$P(A) \sim e^{-\nu A}; \quad \nu = c\eta. \quad (6)$$

Indeed, independent measurements of the probability distribution of achievement do support this prediction (Fig. 3). Moreover, the value found from a best fit for  $\nu = 0.031 \pm 0.004$  is consistent with  $c = 25 \pm 1$ ,  $\eta = 0.00102 \pm 0.00006$ , and  $c\eta = 0.0255 \pm 0.0025$  found from the two previous plots. We note that the observed exponential decay of the probability of achievement in science is the only feature that seems to be shared with the fame-achievement question in the case of ace pilots.

What could be the reason for the different fame vs. merit patterns found for scientists and ace pilots? One likely difference is in the set of people who author webpages that refer to ace pilots as opposed to those who write webpages about scientists. In [1] the authors explained their results using a “rich-get-richer” scheme [3], whereby individuals that are already popular attract people to generate new webpages at a rate proportional to their current popularity. Implicit in this mechanism is the assumption that there is an inexhaustible (or at least very large) pool of people that may author webpages on a popular subject. Such an assumption might perhaps be justified for a subject that enjoys wide notoriety within the public at large — in other words, for truly famous subjects. We maintain that, for whatever reason, scientists are simply not known to the general public, thereby curtailing the option of a rich-get-richer growth. Instead, it is mostly scientist that write webpages about other scientists in their own discipline. In fact, a simple explanation to the linear increase of fame (number of *Google* hits) and achievement (number of papers published) in science is that scientists get cited on other scientists’ webpages in relation to their published work. If each published work typically generates citations in  $c$  webpages, it follows that  $A$  publications would connect a scientist with  $F = cA$  webpages, on average.

It is also worth noting that it is not merely the relation between fame and merit that is different for scientists and ace pilots, but also the distribution of fame itself. This

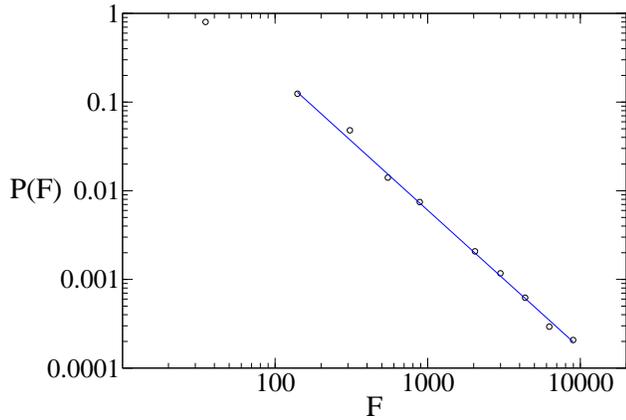


FIG. 4: Fame distribution for 291 athletes in various events: 100 m, 200 m, 1,500 m, and 10,000 m running, and 10,000 m walking. The data ( $\circ$ ) is well-fitted by a power-law (straight line) with exponent 1.55.

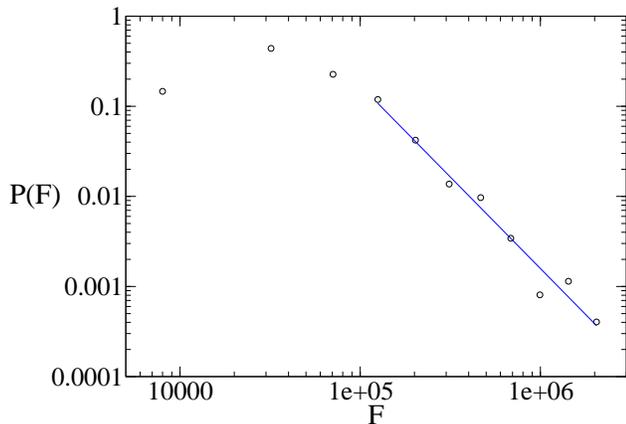


FIG. 5: Fame distribution among 263 movies and television actors. The data ( $\circ$ ) is well fitted by a power-law (straight line) with exponent  $2.03 \pm 0.14$ , in line with the prediction in [1].

gives us a valuable simpler way to explore fame among various groups of people, for while achievement is not usually easy to define, one can always use the *Google*-hit number as a universal measure of fame. We expect that truly famous groups would exhibit a power-law distribution of fame, as for ace pilots, while those that are famous only within their own community might display a fast decaying or exponential distribution, as for scientists. We have tested this idea against a group of 291 olympic runners (all distances) and a group of 263 television and movie actors (randomly drawn from an alphabetical list). The data, presented in Figs. 4 and 5 show that in both cases the probability of a certain level of fame drops off as a power-law, similar to Eq. (1), implying that “true” fame has been achieved — fame within the public at large. It is interesting, though, that the exponent  $\gamma \approx 1.55$  found for runners is quite different from  $\gamma = 2$  predicted by the theory advanced for ace pilots [1].

The fame-merit behavior found for ace pilots is well

documented at least in one other case, and independently from the *Google*-hit measure of fame. Consider a network of social contacts, where two individuals (nodes) are connected if they know each other on a first name basis. In this case, the number of links  $k$  coming out from a node constitutes an obvious measure of fame,  $F = k$ . As a measure of achievement, we may take the distance of a node from an individual at the top of the social ladder (the president of the USA, say). A node is distance  $l$  from the president if the shortest path connecting between the two, along the links of the net, consists of  $l$  links. Thus, people immediately connected to the president boast  $l = 1$ , while those connected to them are (at least) distance  $l = 2$  away, etc. We shall deem being *closer* to the president to represent a higher level of social achievement, that is  $A = -l$ . It is well known that social nets of contacts are scale-free, possessing a degree distribution  $P(k) \sim k^{-\gamma}$ . Hence,

$$P(F) \sim F^{-\gamma},$$

as in Eq. (1). Meanwhile, since the number of nodes at distance  $l$  increases as  $\langle k \rangle^l$  (where  $\langle k \rangle$  is the average degree of a node), we have

$$P(A) \sim e^{-\alpha A}, \quad \alpha = \ln \langle k \rangle.$$

The two relations imply

$$F(A) \sim e^{\beta A}, \quad \beta = \alpha / (\gamma - 1),$$

exactly as found for ace pilots. Note however that in this example the fame exponent  $\gamma$  need not equal 2. On the other hand, we know from recent studies on growth of scale-free networks that some kind of rich-get-richer mechanism is invariably involved, whether implicitly or explicitly [4].

To summarize, the relations for fame and merit found for ace pilots are probably universal among groups of people that enjoy true fame, within the general public. Behind these relations there is a possible rich-get-richer mechanism for the acquisition of fame, though the details might vary from one group to the next. Most importantly, the nature of fame can be studied on its own with the cheap and easy measure of number of *Google* hits: power-law distributions indicate true fame.

For scientists, we find completely different relations, including an exponentially decaying distribution of fame. This suggests that scientists (at least as measured by *Google* hits) are primarily famous within their own science community, and do not enjoy public fame as other groups in the study do. Needless to say, our study excluded the limited number of truly famous scientists, such as Albert Einstein (1,660,000 hits), Isaac Newton (902,000), Galileo Galilei (245,000), Richard Feynman (112,000), and perhaps a handful of others. Their popularity pales in comparison with other public figures, such as the performers Michael Jackson (5,570,000), Janet Jackson (3,190,000), and Barry Williams [5] (2,400,000) .

Our findings for scientists might of course be affected by our choice of the *Google*-hit measure. Had we measured fame in any other way, we may have found the telltale

power-law distribution characteristic of other groups. A reasonable argument is that scientists make heavy use of the WWW for propagation of their work, and as a result the number of webpages dictated by their professional activities overwhelms the number of webpages written about

them by the general public. In this respect, there is comfort in our poor showing of fame, for at the same time it is an indication of a productive level of research [6].

We thank the NSF for partial funding of this research.

- 
- [1] M.V. Simkin, V.P. Roychowdhury, "Theory of Aces: Fame by chance or merit?" (preprint, [arxiv.org/abs/cond-mat/0310049](http://arxiv.org/abs/cond-mat/0310049), 2003).
- [2] The idea of measuring renown by number of *Google* hits has been also suggested recently by: V. Krebs, "What's Your Google Number?" *International Association for Human Resource Information Management Journal*, Vol VII No. 2, 40–42 (2003).
- [3] G. Bianconi and A.-L. Barabási, "Competition and multiscaling in evolving networks," *cond-mat/0011029*; *Europhys. Lett.* **54**, 436 (2001).
- [4] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002); P. L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001); H. Rozenfeld and D. ben-Avraham, "Designer nets from local strategies," *cond-mat/0401196*.
- [5] "Greg Brady," in popular 1970's television show.
- [6] Indeed, a brief *Google* search for undergraduate students at Clarkson University (lexicon: "Student's Name" AND "Clarkson University") — another non-famous group — shows that most students garner zero hits, and at most a few dozen hits, well below the least famous scientists in our study.