Supporting Information

Stallings et al. 10.1073/pnas.1220184110

Formulation and Simulation of Scientific Impact

***A*-Index Derivation.** As described in the main text, there are n total coauthors on a publication who can be divided into m groups, and c_i coauthors in the ith group have the same credit x_i. Under *Axioms 1–3*, our problem is to compute not only the elemental mean $E(x_i)$ as a coauthor's credit share but also the corresponding SD $\sigma(x_i)(1 \le i \le m)$ for statistical testing.

Because $\sigma(x_i) = \sqrt{E(x_i^2) - E(x_i)^2}$, we need to find both $E(x_i)$ and $E(x_i^2)$ for general m and c_i. For notational convenience, let $R_{m,i} = E(x_i)$ and $S_{m,i} = E(x_i^2)$. From *Axioms 1* and *2*, the sample space of the above problem is

$$\Omega_m = \left\{ \boldsymbol{x}(m) = (x_2, \ldots, x_m) : \right.$$
$$\left. 0 \le x_m \le x_{m-1} \le \ldots \le x_2 \le \frac{1}{c_1}\left(1 - \sum_{i=2}^{m} c_i x_i\right) \right\}.$$

Let

$$M_m = \int_{\Omega_m} d\boldsymbol{x}(m)$$

$$E_{m,i} = \int_{\Omega_m} x_i \, d\boldsymbol{x}(m)$$

$$F_{m,i} = \int_{\Omega_m} x_i^2 \, d\boldsymbol{x}(m),$$

where $x_1 = \frac{1}{c_1}(1 - \sum_{i=2}^{m} c_i x_i)$. It follows that

$$R_{m,i} = \frac{E_{m,i}}{M_m}$$

$$S_{m,i} = \frac{F_{m,i}}{M_m}.$$

To determine $R_{m,i}$ and $S_{m,i}$ in a recursive fashion, we look at the restricted sample space whose utility will later become evident,

$$\Omega_m(a,b) = \left\{ \boldsymbol{x}(m) = (x_2, \ldots, x_m) : \right.$$
$$\left. b \le x_m \le x_{m-1} \le \ldots \le x_2 \le \frac{1}{c_1}\left(a - \sum_{i=2}^{m} c_i x_i\right) \right\},$$

where $x_1 = \frac{1}{c_1}(a - \sum_{i=2}^{m} c_i x_i)$, and a and b are constants with $a \ge b\sum_{i=1}^{m} c_i \ge 0$. Similarly, we can define $M_m(a,b)$, $E_{m,i}(a,b)$, and $F_{m,i}$ (a,b) where integration takes place over $\Omega_m(a,b)$, instead of Ω_m. We then have the following propositions.

Proposition 1.

$$M_m(a,b) = M_m\left(a - b\sum c_i, 0\right).$$

Proof: Transform x_i as $y_i = x_i - b$ for $2 \le i \le m$ and

$$M_m(a,b) = \int_{\Omega_m(a,b)} d\boldsymbol{x}(m) = \int_{\Omega_m(a-b\Sigma c_i, 0)} dy(m)$$
$$= M_m\left(a - b\sum c_i, 0\right).$$

□

Proposition 2.

$$M_m(a,0) = \frac{a^{m-1}}{(m-1)!\prod_{j=2}^{m}\left(\sum_{k=1}^{j} c_k\right)}.$$

Proof: We use proof by induction with respect to m. For $m = 2$, we have

$$M_2(a,0) = \int_{\Omega_2(a,0)} d\boldsymbol{x}(2) = \int_0^{\frac{a}{c_1+c_2}} dx_2 = \frac{a}{c_1+c_2},$$

so the proposition holds for this case. Now assume the case holds for $m-1$, and so

$$M_m(a,0) = \int_{\Omega_m(a,0)} d\boldsymbol{x}(m)$$
$$= \int_0^{\frac{a}{\Sigma c_i}} \left(\int_{\Omega_{m-1}(a - c_m x_m, x_m)} d\boldsymbol{x}(m-1) \right) dx_m$$
$$= \int_0^{\frac{a}{\Sigma c_i}} M_{m-1}\left(a - x_m \sum c_i, 0\right) dx_m$$
$$= \int_0^{\frac{a}{\Sigma c_i}} \frac{(a - x_m \sum c_i)^{m-2}}{(m-2)!\prod_{j=2}^{m-1}\left(\sum_{k=1}^{j} c_k\right)}$$
$$= \frac{a^{m-1}}{(m-1)!\prod_{j=2}^{m}\left(\sum_{k=1}^{j} c_k\right)}.$$

□

Proposition 3. *For all* $1 \le i \le m$,

$$E_{m,i}(a,b) = \left(a - b\sum c_i\right)^m F_{m,i} + \frac{b(a - b\sum c_i)^{m-1}}{(m-1)!\prod_{j=2}^{m}\left(\sum_{k=1}^{j} c_k\right)}$$

$$F_{m,i}(a,b) = \left(a - b\sum c_i\right)^{m+1} F_{m,i} + 2b\left(a - b\sum c_i\right)^m E_{m,i}$$
$$+ \frac{b^2(a - b\sum c_i)^{m-1}}{(m-1)!\prod_{j=2}^{m}\left(\sum_{k=1}^{j} c_k\right)}.$$

Proof:

$$E_{m,i}(a,b) = \int_{\Omega_m(a,b)} (x_i - b)d\boldsymbol{x}(m) + \int_{\Omega_m(a,b)} bd\boldsymbol{x}(m) = \int_{\Omega_m(a,b)} (x_i - b)d\boldsymbol{x}(m) + bM_m(a,b) \quad \textbf{[S1]}$$

$$F_{m,i}(a,b) = \int_{\Omega_m(a,b)} (x_i - b)^2 d\boldsymbol{x}(m) + \int_{\Omega_m(a,b)} 2b(x_i - b)d\boldsymbol{x}(m) + \int_{\Omega_m(a,b)} b^2 d\boldsymbol{x}(m) = \int_{\Omega_m(a,b)} (x_i - b)^2 d\boldsymbol{x}(m) + 2b\int_{\Omega_m(a,b)} (x_i - b)d\boldsymbol{x}(m) + b^2 M_m(a,b). \quad \textbf{[S2]}$$

Defining the new variable $y_i = \frac{x_i - b}{a - b\sum c_i}$, we have

$$\int_{\Omega_m(a,b)} (x_i - b)d\boldsymbol{x}(m) = \left(a - b\sum c_i\right)^m \int_{\Omega_m} y_i d\boldsymbol{y}(m) = \left(a - b\sum c_i\right)^m E_{m,i} \quad \textbf{[S3]}$$

$$\int_{\Omega_m(a,b)} (x_i - b)^2 d\boldsymbol{x}(m) = \left(a - b\sum c_i\right)^{m+1} \int_{\Omega_m} y_i^2 d\boldsymbol{y}(m) = \left(a - b\sum c_i\right)^{m+1} F_{m,i}. \quad \textbf{[S4]}$$

From *Propositions 1* and *2* we have

$$M_m(a,b) = \frac{(a - b\sum c_i)^{m-1}}{(m-1)!\prod_{j=2}^m \left(\sum_{k=1}^j c_k\right)}. \quad \textbf{[S5]}$$

Inserting Eqs. **S3**–**S5** into Eqs. **S1** and **S2**, we obtain the result. □

Proposition 4.

$$R_{2,1} = \frac{1}{c_1}(1 - c_2 R_{2,2})$$

$$S_{2,1} = \frac{1}{c_1^2}(1 - 2c_2 R_{2,2} + c_2^2 S_{2,2}).$$

Proof: For $m = 2$, we have $x_1 = \frac{1}{c_1}(1 - c_2 x_2)$, so

$$R_{2,1} = E(x_1) = \frac{1}{c_1}E(1 - c_2 x_2) = \frac{1}{c_1}(1 - c_2 R_{2,2})$$

$$S_{2,1} = E(x_1^2) = \frac{1}{c_1^2}E\left[(1 - c_2 x_2)^2\right] = \frac{1}{c_1^2}(1 - 2c_2 R_{2,2} + c_2^2 S_{2,2}).$$

□

Theorem 1.

$$R_{m,i} = \frac{1}{m}\sum_{j=i}^m \frac{1}{\sum_{k=1}^j c_k}.$$

Proof: By *Propositions 1* and *2*, we have

$$\begin{aligned} E_{m,m} &= \int_{\Omega_m} x_m d\boldsymbol{x}(m) \\ &= \int_0^{\frac{1}{\sum c_i}} x_m \left(\int_{\Omega_{m-1}(1 - c_m x_m, x_m)} d\boldsymbol{x}(m-1) \right) dx_m \\ &= \int_0^{\frac{1}{\sum c_i}} x_m M_{m-1}(1 - c_m x_m, x_m) dx_m \\ &= \int_0^{\frac{1}{\sum c_i}} x_m M_{m-1}\left(1 - x_m \sum c_i, 0\right) dx_m \\ &= \int_0^{\frac{1}{\sum c_i}} \frac{x_m (1 - x_m \sum c_i)^{m-2}}{(m-2)!\prod_{j=2}^{m-1}\left(\sum_{k=1}^j c_k\right)} dx_m \\ &= \frac{1}{m!(\sum c_i)\left[\prod_{j=2}^m \left(\sum_{k=1}^j c_k\right)\right]}. \end{aligned} \quad \textbf{[S6]}$$

By *Proposition 3* and Eq. **S6**, for $1 \le i \le m-1$ we have

$$\begin{aligned} E_{m,i} &= \int_{\Omega_m} x_i d\boldsymbol{x}(m) = \int_0^{\frac{1}{\sum c_i}} \left(\int_{\Omega_{m-1}(1 - c_m x_m, x_m)} x_i d\boldsymbol{x}(m-1) \right) dx_m \\ &= \int_0^{\frac{1}{\sum c_i}} E_{m-1,k}(1 - c_m x_m, x_m) dx_m \\ &= \int_0^{\frac{1}{\sum c_i}} \left((1 - x_m \sum c_i)^{m-1} E_{m-1,k} + \frac{x_m (1 - x_m \sum c_i)^{m-2}}{(m-2)!\prod_{j=2}^{m-1}\sum_{k=1}^j c_k} \right) dx_m \\ &= E_{m,m} + \frac{E_{m-1,i}}{m\sum c_i}. \end{aligned}$$

By *Proposition 2*, we have

$$M_m = M_m(1,0) = \frac{1}{(m-1)!\prod_{j=2}^m \left(\sum_{k=1}^j c_k\right)}.$$

Hence we have

$$R_{m,m}=\frac{E_{m,m}}{M_m}=\frac{1}{m\sum c_i} \quad \textbf{[S7]}$$

and for $1\le i\le m-1$ we have

$$\begin{aligned}R_{m,i}&=\frac{E_{m,i}}{M_m}\\&=\frac{E_{m,m}}{M_m}+\frac{E_{m-1,i}}{mM_m\sum c_i}\\&=R_{m,m}+\frac{R_{m-1,i}M_{m-1}}{mM_m\sum c_i}\\&=R_{m,m}+\frac{m-1}{m}R_{m-1,i}.\end{aligned} \quad \textbf{[S8]}$$

For $2\le i\le m-1$, we repeatedly use Eq. **S8** and get

$$\begin{aligned}R_{m,i}&=R_{m,m}+\frac{m-1}{m}\left(R_{m-1,m-1}+\frac{m-2}{m-1}R_{m-2,i}\right)\\&=R_{m,m}+\frac{m-1}{m}R_{m-1,m-1}+\frac{m-2}{m}R_{m-2,i}\\&=R_{m,m}+\frac{1}{m}\sum_{j=1}^{m-i}(m-j)R_{m-j,m-j}\\&=\frac{1}{m}\sum_{j=i}^{m}\frac{1}{\sum_{k=1}^{j}c_k}.\end{aligned} \quad \textbf{[S9]}$$

For $i=1$, we also repeatedly use Eq. **S8** and get

$$\begin{aligned}R_{m,1}&=R_{m,m}+\left(\frac{1}{m}\sum_{j=1}^{m-3}(m-j)R_{m-j,m-j}\right)+\frac{2}{m}R_{2,1}\\&=\frac{1}{m}\sum_{j=3}^{m}\frac{1}{\sum_{k=1}^{j}c_k}+\frac{2}{mc_1}\left(1-\frac{c_2}{2(c_1+c_2)}\right)\\&=\frac{1}{m}\sum_{j=1}^{m}\frac{1}{\sum_{k=1}^{j}c_k}.\end{aligned} \quad \textbf{[S10]}$$

Combining Eqs. **S7**, **S9**, and **S10** we have

$$R_{m,i}=\frac{1}{m}\sum_{j=i}^{m}\frac{1}{\sum_{k=1}^{j}c_k}.$$

□

Theorem 2.

$$S_{m,i}=\frac{2}{m(m+1)}\sum_{i\le k\le j\le m}\frac{1}{(c_1+\ldots+c_j)(c_1+\ldots+c_k)}.$$

Proof: By *Propositions 1* and *2*, we have

$$\begin{aligned}F_{m,m}&=\int_{\Omega_m}x_m^2d\boldsymbol{x}(m)\\&=\int_0^{\frac{1}{\sum c_i}}x_m^2\left(\int_{\Omega_{m-1}(1-c_mx_m,x_m)}d\boldsymbol{x}(m-1)\right)dx_m\\&=\int_0^{\frac{1}{\sum c_i}}x_m^2M_{m-1}(1-c_mx_m,x_m)dx_m\\&=\int_0^{\frac{1}{\sum c_i}}x_m^2M_{m-1}\left(1-x_m\sum c_i,0\right)dx_m\\&=\int_0^{\frac{1}{\sum c_i}}\frac{x_m^2(1-x_m\sum c_i)^{m-2}}{(m-2)!\prod_{j=2}^{m-1}\left(\sum_{k=1}^{j}c_k\right)}dx_m\\&=\frac{2}{(\sum c_i)^2(m+1)!\prod_{j=2}^{m-1}\left(\sum_{k=1}^{j}c_k\right)}.\end{aligned} \quad \textbf{[S11]}$$

For $1\le i\le m-1$, using *Proposition 3* and Eq. **S11**, we have

$$\begin{aligned}F_{m,i}&=\int_{\Omega_m}x_i^2d\boldsymbol{x}(m)\\&=\int_0^{\frac{1}{\sum c_i}}\left(\int_{\Omega_{m-1}(1-c_mx_m,x_m)}x_i^2d\boldsymbol{x}(m-1)\right)dx_m\\&=\int_0^{\frac{1}{\sum c_i}}F_{m-1,i}(1-c_mx_m,x_m)dx_m\\&=\int_0^{\frac{1}{\sum c_i}}\Big(\left(1-x_m\sum c_i\right)^mF_{m-1,i}\\&\qquad+2x_m(1-x_m\sum c_i)^{m-1}E_{m-1,i}\Big)dx_m\\&\qquad+\int_0^{\frac{1}{\sum c_i}}\frac{x_m^2(1-x_m\sum c_i)^{m-2}}{(m-2)!\prod_{j=2}^{m-1}\left(\sum_{k=1}^{j}c_k\right)}dx_m\\&=F_{m,m}+\frac{F_{m-1,i}}{(m+1)\sum c_i}+\frac{2E_{m-1,i}}{m(m+1)(\sum c_i)^2}.\end{aligned}$$

Therefore, we have

$$S_{m,m}=\frac{F_{m,m}}{M_m}=\frac{2}{m(m+1)(\sum c_i)^2},$$

and for $1 \le i \le m-1$ we have

$$S_{m,i} = \frac{F_{m,m}}{M_m} + \frac{F_{m-1,i}}{(m+1)(\sum c_i)M_m} + \frac{2E_{m-1,i}}{m(m+1)(\sum c_i)^2 M_m}$$
$$= S_{m,m} + \frac{m-1}{m+1} S_{m-1,i} + \frac{2(m-1)}{m(m+1)\sum c_i} R_{m-1,i}.$$

For $2 \le i \le m-1$, repeatedly using Eq. **S12** we have

$$\begin{aligned}
S_{m,i} &= S_{m,m} + \frac{2(m-1)}{m(m+1)\sum c_i} R_{m-1,i} + \frac{m-1}{m+1} S_{m-1,i} \\
&= S_{m,m} + \frac{2(m-1)}{m(m+1)\sum c_i} R_{m-1,i} \\
&\quad + \frac{m-1}{m+1}\left(S_{m-1,m-1} + \frac{2(m-2)}{m(m-1)\sum_{k=1}^{m-1} c_k} R_{m-2,i} + \frac{m-2}{m} S_{m-2,i} \right) \\
&= S_{m,m} + \frac{m-1}{m+1} S_{m-1,m-1} + \frac{2(m-1)}{m(m+1)\sum c_i} R_{m-1,i} \\
&\quad + \frac{2(m-2)}{m(m+1)\sum_{k=1}^{m-1} c_k} R_{m-2,i} + \frac{(m-1)(m-2)}{m(m+1)} S_{m-2,i} \\
&= S_{m,m} + \frac{m-1}{m+1} S_{m-1,m-1} + \frac{(m-1)(m-2)}{(m+1)m} S_{m-2,m-2} \\
&\quad + \ldots + \frac{(i+1)i}{(m+1)m} S_{i,i} + \frac{2(m-1)}{m(m+1)\sum c_i} R_{m-1,i} \\
&\quad + \frac{2(m-2)}{m(m+1)\sum_{k=1}^{m-1} c_k} R_{m-2,i} + \ldots + \frac{2i}{m(m+1)\sum_{k=1}^{i+1} c_k} R_{i,i} \\
&= \frac{2}{m(m+1)} \left(\sum_{j=i}^{m} \frac{1}{\left(\sum_{k=1}^{j} c_k\right)^2} \right. \\
&\quad \left. + \sum_{i \le k < j \le m} \frac{1}{(c_1 + \ldots + c_j)(c_1 + \ldots + c_k)} \right).
\end{aligned} \tag{S12}$$

For $i = 1$, repeatedly using Eq. **S12** we have

$$\begin{aligned}
S_{m,1} &= S_{m,m} + \frac{m-1}{m+1} S_{m-1,m-1} + \frac{(m-1)(m-2)}{(m+1)m} S_{m-2,m-2} \\
&\quad + \ldots + \frac{12}{(m+1)m} S_{3,3} + \frac{6}{(m+1)m} S_{2,1} + \frac{2(m-1)}{(m+1)m\sum c_i} R_{m-1,1} \\
&\quad + \frac{2(m-2)}{(m+1)m\sum_{k=1}^{m-1} c_k} R_{m-2,1} + \ldots + \frac{4}{(m+1)m(c_1+c_2+c_3)} R_{2,1} \\
&= \frac{2}{(m+1)m} \left(\sum_{j=3}^{m} \frac{1}{\left(\sum_{k=1}^{j} c_k\right)^2} \right. \\
&\quad \left. + \sum_{1 \le k < j \le m} \frac{1}{(c_1 + \ldots + c_j)(c_1 + \ldots + c_k)} + 3S_{2,1} - \frac{1}{c_1(c_1+c_2)} \right).
\end{aligned} \tag{S13}$$

Because

$$\begin{aligned}
S_{2,1} &= \frac{1}{c_1^2}\left(1 - 2c_2 R_{2,2} + c_2^2 S_{2,2}\right) \\
&= \frac{1}{c_1^2}\left(1 - \frac{2c_2}{2(c_1+c_2)} + \frac{2c_2^2}{6(c_1+c_2)^2}\right) \\
&= \frac{1}{c_1^2} - \frac{c_2}{c_1^2(c_1+c_2)} + \frac{c_2^2}{3c_1^2(c_1+c_2)^2} \\
&= \frac{c_1 + c_2 - c_2}{c_1^2(c_1+c_2)} + \frac{c_2^2}{3c_1^2(c_1+c_2)^2} \\
&= \frac{1}{c_1(c_1+c_2)} + \frac{c_2^2}{3c_1^2(c_1+c_2)^2},
\end{aligned}$$

we have

$$3S_{2,1} - \frac{1}{c_1(c_1+c_2)} = \frac{1}{c_1^2} + \frac{1}{(c_1+c_2)^2}. \tag{S14}$$

Inserting Eq. **S14** into Eq. **S13** we have

$$S_{m,1} = \frac{2}{(m+1)m}\left(\sum_{j=1}^{m} \frac{1}{\left(\sum_{k=1}^{j} c_k\right)^2} + \sum_{1 \le k < j \le m} \frac{1}{(c_1 + \ldots + c_j)(c_1 + \ldots + c_k)} \right).$$

Combining Eqs. **S11**, **S12**, and **S15**, we obtain

$$S_{m,i} = \frac{2}{(m+1)m} \sum_{i \le k \le j \le m} \frac{1}{(c_1 + \ldots + c_j)(c_1 + \ldots + c_k)}.$$

□

It follows that $\sigma(x_i) = \sqrt{S_{m,i} - R_{m,i}^2}$.

Simulation Study. The goal of the simulation was to compare the performance of the four indexes in a variety of situations. The simulation was based on classifying researchers on the basis of the following four categories:

1) Number of publications (*N*-index)

2) Journal impact factor (JIF)

3) Number of coauthors

4) Coauthor's rank

Each category had two different distributions labeled high and low, where the high distributions assigned larger probability to values that would increase a simulated researcher's *P*-index. In this case, the high distribution of number of coauthors means that the researcher collaborated with few other researchers, because fewer coauthors increases a researcher's potential *A*-index and hence the *P*-index. Hence, there were $2^4 = 16$ different types of researchers considered. We made strong assumptions about these distributions so that we could simulate publication data to show how the *A*- and *P*-indexes perform in a variety of situations.

Methods. For each of the 16 combinations, we created 200 virtual researchers. For each of these researchers, we generated publication data over 5 y. Publications for each researcher and year were treated as independent. The parameters of the distributions were primarily determined on the basis of what was observed for the biomedical engineering (BME) researchers in the original analysis. A graphic representation of the distributions is in Fig. S1.

For the number of publications per year, or the *N*-index, we generated data from a specified Gamma distribution and then rounded the number to the nearest integer, because the *N*-index must be an integer. We chose the Gamma distribution because it is nonnegative, is right skewed, and can take on many different forms. The high distribution had an expected value of seven publications and a SD of 4, whereas the low distribution had an expected value of two publications with a SD of 2. The larger variability and greater skewness for the high level were justified because it is difficult for any researcher to consistently publish a large number of publications.

The JIF distributions were similarly defined, although they were not rounded because they do not need to be integers. The high level followed a Gamma distribution with mean and SD of 5, whereas the low level was also Gamma with a mean of 2 and a SD of $\sqrt{2}$.

Although the data analysis of BME researchers gave evidence that researchers who had many publications tended to collaborate more, we ignored this relationship in the simulation. The number of coauthors for each paper was generated using a rounded Gamma distribution truncated at 1, because there must be at least one author of each paper. This was accomplished by randomly generating a value from the proposed Gamma distribution, rounding it to the closest integer, and then adding 1. Recall that the high distribution means fewer collaborators on average. Hence, we used a Gamma distribution with mean and SD of 2. The low distribution had a mean of 5 and a SD of 4.

Ranking of the virtual researcher on the paper corresponded to credit, not necessarily the order of the authors in the publication. For example, usually the last author is listed as the corresponding author, who is often as equally important as the first author. The distribution of author rank clearly depended on the number of coauthors for the publication. For simplicity, we assumed that every author assumed unequal credit, so given the m total authors we had m ranks. To assign a rank for a given publication and number of coauthors, we used a multinomial distribution where each category corresponds to a rank, and the probability of having a certain rank depends on the high or low level. For a high-rank author, the probability vector was the $m \times 1$ vector $\boldsymbol{p}_H = c^{-1}(m, m-1, \ldots, 1)'$, where $c = \frac{m(m+1)}{2}$, so that the elements summed to one. For example, the probability for a virtual researcher who had a high-rank distribution being the first author was

$$\frac{m}{1+2+\ldots+m} = \frac{m}{\frac{m(m+1)}{2}} = \frac{2}{m+1},$$

and the probability of being the last author was

$$\frac{1}{1+2+\ldots+m} = \frac{1}{\frac{m(m+1)}{2}} = \frac{2}{m(m+1)}.$$

The probability vector for a low-rank author was $\boldsymbol{p}_L = c^{-1}(1, 2, \ldots, m)'$, which is $\boldsymbol{p}_H$ with the elements reversed. With the probability vectors defined, we took one random draw from that multinomial distribution and let that be the assigned rank for the virtual researcher.

For each publication, we calculated the *A*-index assuming an unequal contribution. Recall that the *A*-index comes from an expected credit vector that was derived from the three axioms specified in the main text. To see how well the *P*-index performed, a "true" credit vector was also randomly generated for each publication from the distribution specified by the axioms, assuming an unequal contribution. This allowed us to calculate the true *P*-index.

For each virtual researcher and year, we first generated the *N*-index, say N_0. For each of the N_0 publications, we generated the JIFs and number of coauthors. Once the number of coauthors was known, we generated the researcher's rank on that publication. Finally, the *A*-index was calculated and a true credit vector was produced. After the simulated publication data were produced, each year's *P*-index and a "pseudo"-5-y *H*-index, based on the JIF instead of citations, were calculated.

We summarized the 16 different types of virtual researchers by looking at averages and SDs of the four metrics: *P*-, *C*-, and *N*-indexes and pseudo-5-y *H*-index. We ranked the combinations by the *P*-index to see which virtual researchers had the highest *P*-index and compared their corresponding *N*- and *H*-indexes, which do not take collaboration into account. We also looked at consistency of the *P*-index from one year to the next by looking at the proportion of virtual researchers in a given year that were classified as high impact (*P*-index ≥ 5). The mean *P*-index for each group was also compared with their mean true *P*-index. Finally, we investigated what situations caused virtual researchers in low-mean *P*-index groups to have a high *P*-index for a year.

Results and Discussion. Behavior of the indexes across all researchers and years for each of the 16 combinations is summarized in Table S1, sorted by mean *P*-index. Table S1 is further broken up on the basis of the *P*-index cutoff of 5. Of the 16 combinations, 7 had a mean *P*-index greater than 5, meaning that we expect these virtual researchers to be high impact for a given year. Of these 7, only one researcher group had a low *N*-index distribution, and only one had both low coauthor and low-rank distributions.

We anticipated that researchers with high *N*-index and JIF distributions would have the highest *P*-indexes regardless of their collaborative behaviors, because a high *N*-index and high average JIF can neutralize overcollaboration. The four researcher groups with high *N*-index and JIF distributions were all in the high-impact group and had similar *H*-indexes around 7.00. The researcher group with all high distributions had the highest mean *P*-index, 19.43, as well as the highest SD, 14.48. The high SD was caused by the larger variance for all high distributions. Researchers with high *N*-index and JIF distributions, but low coauthor and rank distributions had a mean *P*-index of 7.06 and a SD of 6.50, whereas their mean *N*- and *H*-indexes were among the highest. Even though these researchers were involved in multiple, high-impact publications, their mean *P*-index was much lower than that of the other researcher groups with high *N*-index and JIF distributions. By including collaborative behavior, the *P*-index was able to distinguish certain researchers that would otherwise be identical relative to the *N*- and *H*-indexes.

The fourth-highest researcher class had a high distribution in every category except JIF and had a mean *P*-index of 7.92 and a SD of 5.64. However, their mean 5-y *H*-index was 3.37, whereas the mean *H*-indexes for researchers with high *N*-index and JIF distributions were above 7.00. The low JIF virtual researcher published frequently in low-cited journals, but worked in small groups and tended to be the most significant author. Because he/she did not publish in highly cited journals, his/her pseudo-5-y *H*-index was limited. This case shows that a researcher can achieve a relatively high *P*-index without publishing in high-impact journals, as long as he/she is vital to the publications and works diligently. Young, prolific researchers may fit into this category, and the *P*-index would quickly identify them, whereas traditional metrics could ignore them.

We expected researchers with the low *N*-index and JIF distributions to have the lowest average *P*-index, along with the other indexes. The mean *P*-indexes for these groups were 2.04 or less whereas the mean *H*-indexes were only slightly higher than 2.00. In this case, collaborative behaviors of these researchers were difficult to detect because they published infrequently.

Next we looked at how consistently researchers were classified as having high impact across the 5 y. We calculated the proportion of researchers in each group that were high impact for all 5 y and then for at least 4 y, etc. (Table S2). Although it was possible to classify a normal researcher as high impact for a given year, very few were high impact for multiple years. In fact, none of them were classified as high impact for all 5 y. Also, of the researchers who tend to have high *N*-index and JIF distributions, but low coauthor and rank distributions, only 6% were high impact for all 5 y. By looking at the *P*-index for many years, it may become clearer who is the driving force behind the scientific impact.

The ability of the *P*-index to measure the true productivity was also of interest. Fig. S2 shows that the *P*-index does an excellent job of approximating the true *P*-index, because the points follow the reference line closely. This was assuming that the true credit vectors followed the uniform distribution in *Axiom 3*. In the future, it would be interesting to see how robust the *P*-index is when the true credit vectors follow some other distribution.

For the bottom two types of researchers (LLLH and LLLL in Tables S1 and S2), we investigated what caused them to appear as high impact for a given year and found it was mainly caused by either a high number of publications or a few publications with a high JIF. These situations will come up from time to time but we should not expect it to happen repeatedly from one year to the next. In reality, there is the possibility of a researcher briefly working with someone who is of truly high impact, potentially causing a short time period of the researcher being classified as high impact. By raising the standard *P*-index by which we classify high-impact researchers, or considering it alongside the *C*-index, these cases could be discovered quickly.

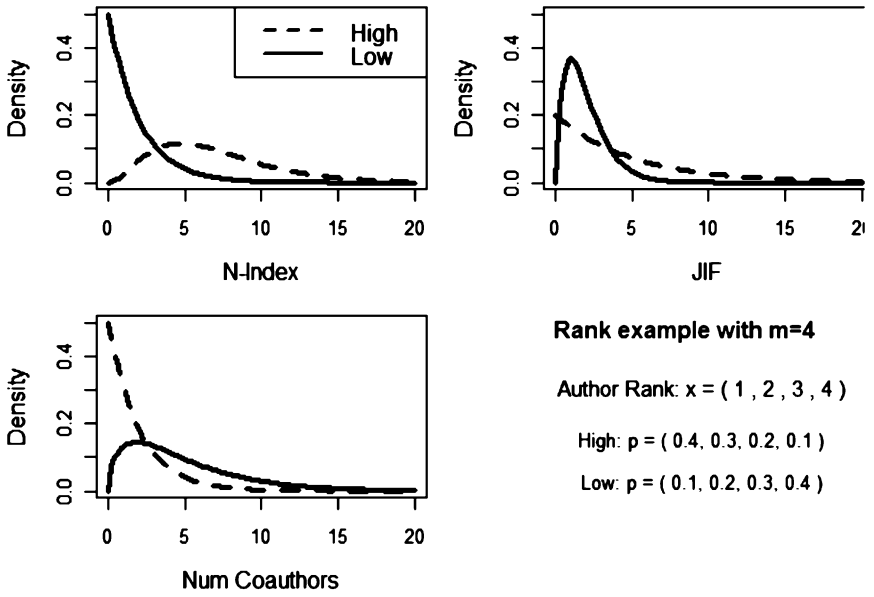

Fig. S1. Category distributions.

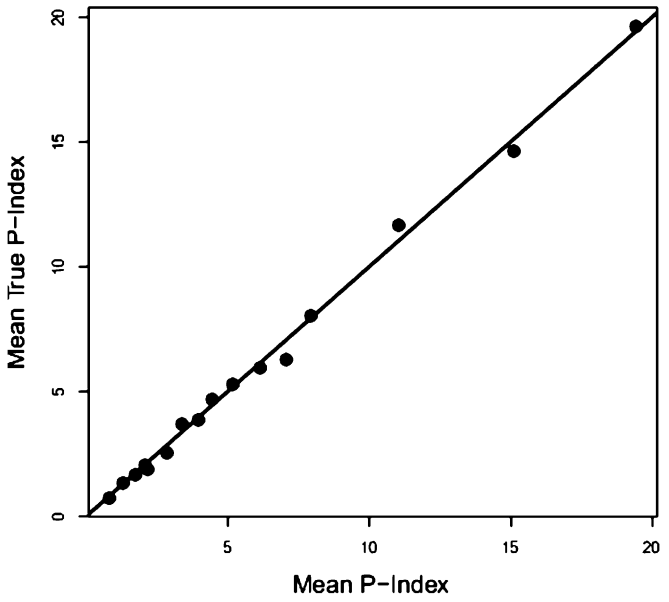

Fig. S2. Mean true *P*-index vs. mean *P*-index.

Table S1. Index means and SDs across all researchers and years

Combinations				*P*-index		*C*-index		*N*-index		*H*-index	
N	JIF	Coauthors	Rank	Mean	SD	Mean	SD	Mean	SD	Mean	SD
H	H	H	H	19.43	14.48	3.88	2.39	7.08	4.07	7.29	1.17
H	H	H	L	15.07	12.93	3.06	2.05	7.06	4.11	7.20	1.37
H	H	L	H	11.00	9.07	2.15	1.46	6.75	4.07	7.17	1.24
H	L	H	H	7.92	5.64	3.95	2.51	7.20	4.29	3.37	0.61
H	H	L	L	7.06	6.50	1.38	0.99	7.04	4.06	7.30	1.21
H	L	H	L	6.11	4.56	3.04	1.99	7.00	4.07	3.37	0.64
L	H	H	H	5.16	6.87	1.05	1.13	1.95	2.01	3.83	1.23
H	L	L	H	4.41	3.30	2.19	1.49	6.93	4.08	3.32	0.58
L	H	H	L	3.94	5.78	0.84	0.97	1.94	1.97	3.77	1.22
L	H	L	H	3.36	4.77	0.64	0.76	2.00	2.17	3.90	1.35
H	L	L	L	2.85	2.29	1.41	0.99	7.12	3.96	3.37	0.59
L	H	L	L	2.16	3.50	0.43	0.57	2.20	2.20	4.07	1.39
L	L	H	H	2.04	2.36	1.09	1.18	1.96	1.97	2.13	0.62
L	L	H	L	1.73	2.36	0.86	1.02	2.00	2.06	2.21	0.72
L	L	L	H	1.26	1.67	0.64	0.75	2.03	2.06	2.14	0.76
L	L	L	L	0.81	1.27	0.40	0.54	2.00	2.04	2.22	0.70

Table S2. Proportion of consistent high-impact researchers (blanks are zeros)

N	JIF	Coauthors	Rank	All 5 y	≥4 y	≥3 y	≥2 y	≥1 y
H	H	H	H	0.52	0.85	0.99	1.00	1.00
H	H	H	L	0.29	0.72	0.92	0.99	1.00
H	H	L	H	0.14	0.52	0.87	0.98	1.00
H	L	H	H	0.12	0.42	0.75	0.91	1.00
H	H	L	L	0.06	0.30	0.66	0.92	0.99
H	L	H	L	0.05	0.17	0.50	0.85	0.97
L	H	H	H		0.03	0.19	0.56	0.88
H	L	L	H		0.06	0.23	0.56	0.85
L	H	H	L		0.01	0.09	0.41	0.76
L	H	L	H			0.07	0.38	0.74
H	L	L	L			0.04	0.21	0.54
L	H	L	L		0.01	0.02	0.14	0.50
L	L	H	H				0.09	0.45
L	L	H	L				0.04	0.42
L	L	L	H				0.01	0.17
L	L	L	L					0.08