Supplementary Information

Measuring Originality in Science

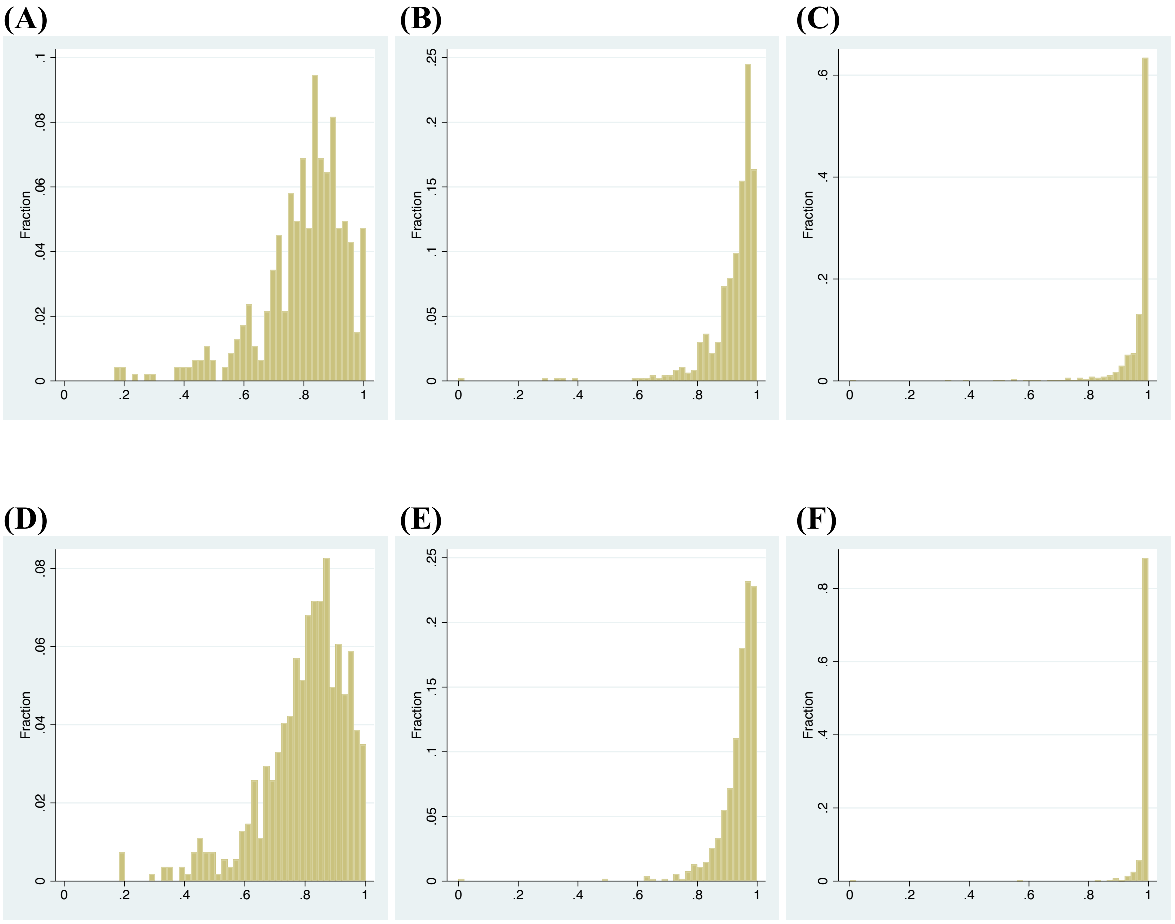
Sotaro Shibayama[1,*] and Jian Wang[2]

[1] School of Economics and Management, Lund University, Lund, 22363, Sweden

[2] Leiden University, Leiden, 2333 CA, Netherlands

[*]Corresponding author:
Email: sotaro.shibayama@fek.lu.se.
Phone: +46 (0)46 2227812

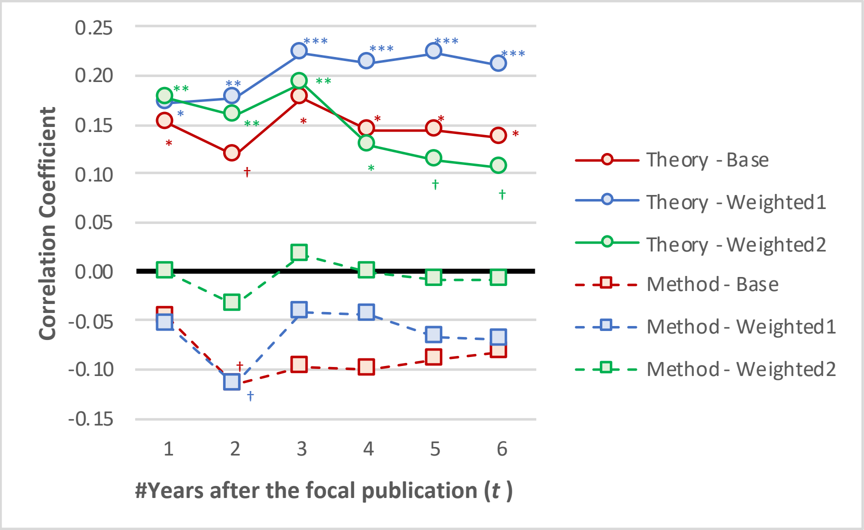
Figure S1. Distribution of proposed originality measures



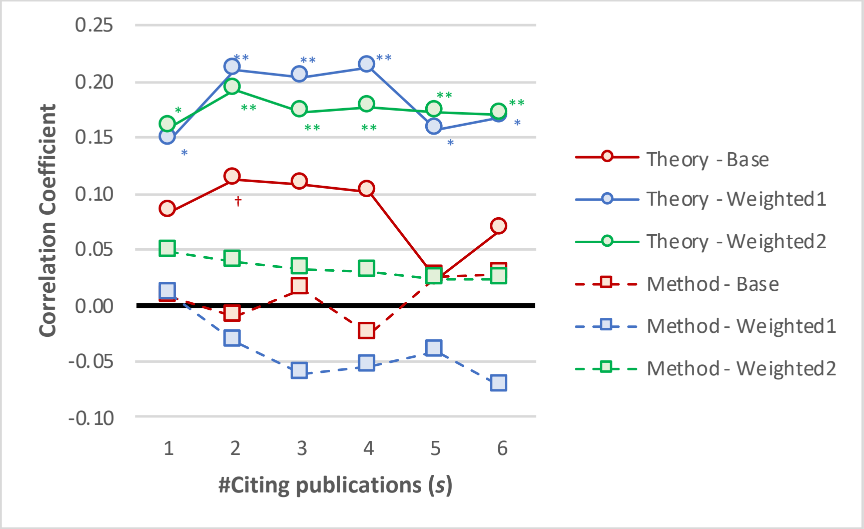
Note. (A) Base measure computed with first-year citing publications. *N* = 465. (B) Weighted measure 1 computed with first-year citing publications. *L* = 14.4. *N* = 465. (C) Weighted measure 2 computed with first-year citing publications. *L* = 57.6. *N* = 465. (D) Base measure computed with the first three citing publications. *N* = 544. (E) Weighted measure 1 computed with the first three citing publications. *L* = 12.0. *N* = 544. (F) Weighted measure 2 computed with the first three citing publications. *L* = 14.0. *N* = 544.

Figure S2. Correlation between the proposed originality measures (mean at the scientist level) and self-assessed originality

(A) Originality measured with citing publications in the first *t* years (*t* = 1, …, 6)



(B) Originality measured with the first s citing publications (*s* = 1, …, 6)



Notes. †p<0.1. *p<0.05. **p<0.01. ***p<0.001. (A) The sample size ranges from 219 to 245. (B) The sample size ranges from 197 to 244.

Table S1. Correlation between the proposed originality measures and self-assessed originality

(A) Originality measured with citing publications in the first *t* years and self-assessed theoretical originality

t	*N*	Theory - Base	Theory - Weighted1	Theory - Weighted2
1	461	0.130 *	0.136 ***	0.142 ***
2	516	0.099 †	0.147 **	0.138 **
3	540	0.147 **	0.174 ***	0.143 ***
4	549	0.120 *	0.163 ***	0.087
5	551	0.120 *	0.160 ***	0.074
6	553	0.113 *	0.145 ***	0.066

(B) Originality measured with citing publications in the first *t* years and self-assessed methodological originality

t	*N*	Method - Base	Method - Weighted1	Method - Weighted2
1	455	-0.040	-0.044	-0.002
2	510	-0.096 †	-0.097 *	-0.028
3	534	-0.082	-0.033	0.012
4	543	-0.083	-0.034	0.000
5	545	-0.076	-0.047	-0.006
6	547	-0.070	-0.048	-0.005

(C) Originality measured with the first s citing publications and self-assessed theoretical originality

s	*N*	Theory - Base	Theory - Weighted1	Theory - Weighted2
1	540	0.067	0.110 **	0.107 ***
2	502	0.093 †	0.169 ***	0.130 *
3	468	0.091	0.174 ***	0.149 **
4	430	0.090	0.185 **	0.171 **
5	390	0.022	0.144 *	0.144 *
6	359	0.060	0.156 *	0.182 ***

(D) Originality measured with the first s citing publications and self-assessed methodological originality

s	*N*	Method - Base	Method - Weighted1	Method - Weighted2
1	534	0.005	0.007	0.032
2	496	-0.008	-0.027	-0.023
3	463	0.012	-0.053	-0.015
4	425	-0.022	-0.048	-0.010
5	385	0.023	-0.038	0.008
6	354	0.025	-0.067	0.020

Note. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Since our respondents can have multiple papers, we introduced a weight (the reciprocal of the paper count) into the computation of correlation coefficients.

Table S2. Prediction of Citation Rank

	Model 1		Model 2		Model 3	
$\boldsymbol{ln\ N}$	1.433**	(0.553)	1.399*	(0.578)	1.043†	(0.566)
$\boldsymbol{Orig_{base}}$	1.149	(1.553)				
$\boldsymbol{Orig_{weighted1}}$			2.780	(3.795)		
$\boldsymbol{Orig_{weighted2}}$					107.669***	(31.184)
Chi-squared stat	11.093**		11.698**		23.464***	
Log likelihood	-153.715		-153.729		-148.852	
N	439		439		439	
#Scientist	205		205		205	

	Model 4		Model 5		Model 6	
$\boldsymbol{ln\ N}$	1.175*	(0.571)	1.096†	(0.587)	0.838	(0.583)
$\boldsymbol{Orig_{base}}$	4.442*	(2.116)				
$\boldsymbol{Orig_{weighted1}}$			10.910*	(5.144)		
$\boldsymbol{Orig_{weighted2}}$					162.262***	(45.624)
Chi-squared stat	12.389**		15.526***		20.038***	
Log likelihood	-149.057		-149.585		-144.919	
N	410		410		410	
#Scientist	194		194		194	

	Model 7		Model 8		Model 9	
$\boldsymbol{ln\ N}$	0.930†	(0.562)	0.796	(0.572)	0.627	(0.568)
$\boldsymbol{Orig_{base}}$	5.313*	(2.349)				
$\boldsymbol{Orig_{weighted1}}$			14.686**	(5.266)		
$\boldsymbol{Orig_{weighted2}}$					166.365***	(48.996)
Chi-squared stat	10.804**		16.115***		16.911***	
Log likelihood	-144.937		-144.796		-140.905	
N	375		375		375	
#Scientist	185		185		185	

Notes. Logistic regressions with errors clustered in scientists. Unstandardised coefficients (robust errors in parentheses). Two-tailed test. $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Only focal articles published in 2008 or before are included in the analysis to allow a 10-year citation window. Originality measured with *s* citing publications: *s* = 2 (Models 1-3), *s* = 3 (Models 4-6), and *s* = 4 (Models 7-9).