REGULAR ARTICLE

Open Access

Success and luck in creative careers

Milán Janosov[1] , Federico Battiston[1] and Roberta Sinatra[2,3,4*]

*Correspondence: rsin@itu.dk
[2]Department of Computer Science,
IT University of Copenhagen,
Copenhagen, Denmark
[3]ISI Foundation, Torino, Italy
Full list of author information is
available at the end of the article

Abstract

Luck is considered a crucial ingredient to achieve impact in all creative domains, despite their diversity. For instance, in science, the movie industry, music, and art, the occurrence of the highest impact work and a hot streak within a creative career are very difficult to predict. Are there domains that are more prone to luck than others? Here, we provide new insights on the role of randomness in impact in creative careers in two ways: (i) we systematically untangle luck and individual ability to generate impact in the movie, music, and book industries, and in science, and compare the luck factor between these fields; (ii) we show the surprising presence of randomness in the relationship between collaboration networks and timing of career hits. Taken together, our analysis suggests that luck consistently affects career impact across all considered sectors and improves our understanding in pinpointing the key elements in driving success.

Keywords: Success; Dynamics of impact; Creative careers; Science of science

Research in developmental psychology has studied careers of prominent artists and scientists for decades, advocating the importance of chance for the successful unfolding of careers in various creative domains [1–4]. In recent years, the availability of big databases on scientific publications [5] and artistic records, from books to movies [6–8], has made it possible to test a number of previously suggested hypotheses on a large scale. For instance, in previous work [9, 10], the analysis of thousands of creative careers has shown that the biggest hit of an individual occurs randomly within an individual's career, a finding named equal-odds-rule [3] or random impact rule [9]. This rule explains the variability in the occurrence of creative individuals' best hits. Yet, career hits are not only the results of luck but also of other individual and team properties [11–17]. While previous literature suggests that luck and individual ability are both necessary to excel in art and science [18–21], a quantification of the role of luck across different creative domains is still lacking. In which creative fields are individuals more likely to go from rags to riches and vice-versa? How is the network position of an individual related to the occurrence of a career hit?

In this work, we quantify luck fluctuations in impact across creative careers from film, music, literature, and science, and create a framework to compare the broad observed differences in impact [7, 22]. Do these random fluctuations have the same magnitude across careers? To address this question, we build on the mathematical framework known as the Q-model proposed in Ref. [9] to untangle the impact into two components, one encoding

Springer

fluctuation that can be interpreted as luck, and another depending only on the individual. We show that this model is consistent with the classical test theory [23], also known as the true score theory [24], stating that the measured value of a certain measurable attribute consists of the sum of its true – error-free – score, and a stochastic error term. We find that the value of such *randomness* varies depending on the creative fields. By comparing this stochastic term to the typical impact score associated with each artist and scientist, we identify creative domains where the impact of single creative products are the most exposed to luck and fluctuate the most within individual careers. The pronounced role of randomness in achieving success in creative careers is confirmed by the unpredictable relation between the position of an individual in her collaboration network, captured by a number of network measures, and the timing of the hit of her career. To carry out these analyses, we rely on a large-scale data set covering more than four million individuals from c. 1902 up until 2017.

The outline of this paper is the following. First, we test the validity of the requirements of the *Q*-model proposed in Ref. [9]. Second, we use the *Q*-model impact decomposition method to factor impact in creative careers. Third, we apply the classical test theory to quantify the role of luck within each field and discuss the observed differences across fields. Finally, we construct the collaboration network within each domain and compare the time of the best hit of creative individuals to the time at which they reach their highest score in network centrality.
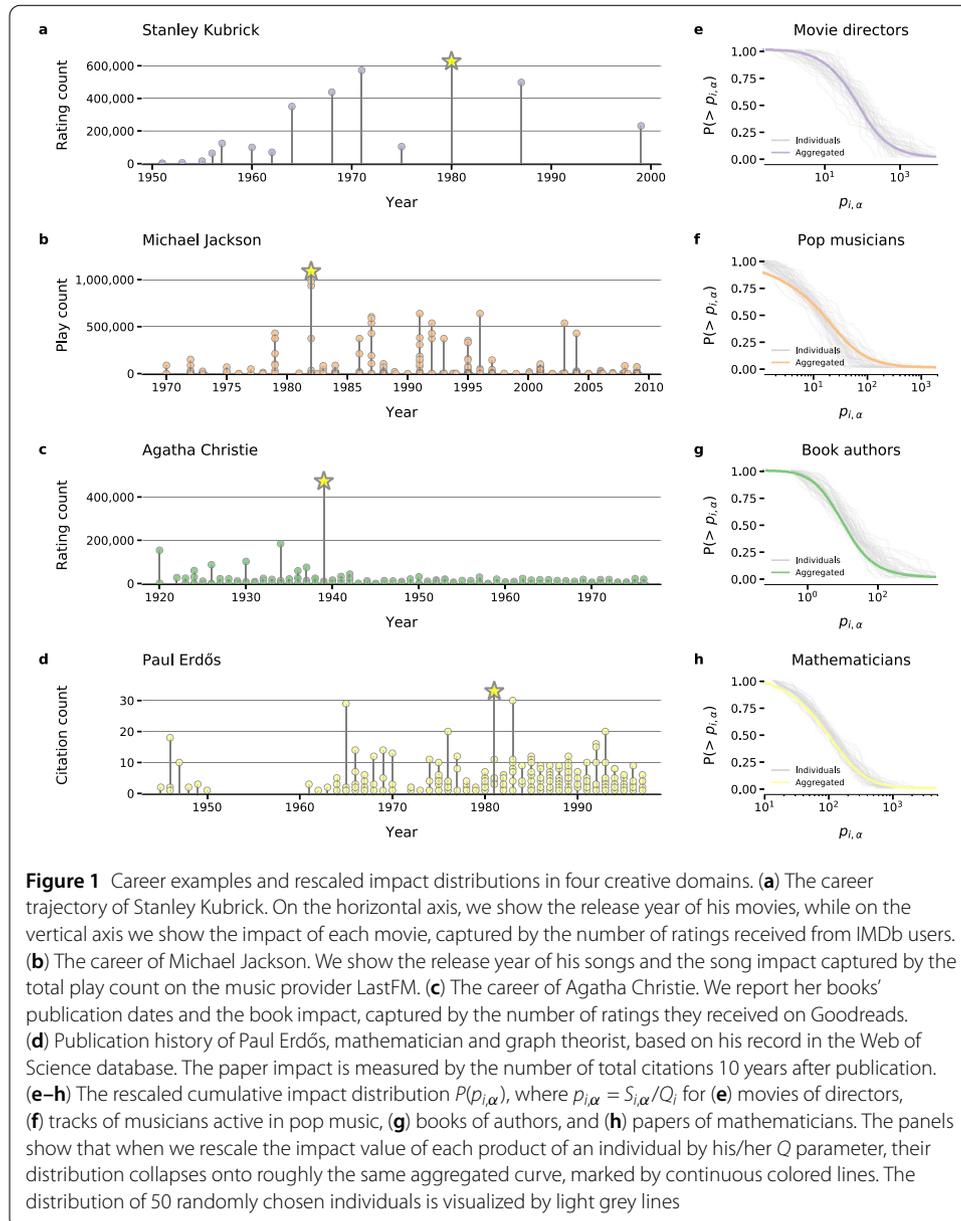
1 Data
We compiled four data sets of individual careers across the movie, music, and book industries, and across scientific fields, covering overall 28 different types of creative careers:

1. We mined the Internet Movie Database (IMDb [25]) and compiled a data set of 803,013 individuals in the movie industry working as movie directors, producers, art directors, soundtrack composers, and scriptwriters, altogether contributing to 1,297,275 movies.
2. By using the Discogs [26, 27] and LastFM [28] platforms, we constructed a database of 379,366 musicians released 31,841,981 songs in the genres of electronic, rock, pop, funk, folk, jazz, hip-hop, and classical music.
3. We extracted data from Goodreads [29] and built a data set containing information about 2,069,891 book authors and 6,604,144 books.
4. We used the Web of Science database [5] to reconstruct the scientific careers of 1,204,688 scientists from the fields of chemistry, mathematics, physics, applied physics, space science and astronomy, zoology, geology, agronomy, engineering, theoretical computer science, biology, environmental science, political science, and health science, altogether authoring approximately 87,4 million papers.

See further details about the data sets and the data collection in SI Section S1.1.

To measure the impact of movies, songs, books, and articles, we use their cumulated impact on large audiences, as captured by the rating counts for movies and books, the play counts for songs, and the number of citations received within the first ten years after publication for scientific papers [30] (SI Section S1.2). The existence of these cumulative impact measures in all data sets allows us to reconstruct individual careers consistently across domains by building the historical time series of each person. In Fig. 1a–d, we illustrate career examples in the four different databases: movie director Stanley Kubrick, pop singer Michael Jackson, writer Agatha Christie, and mathematician Paul Erdős. To en-

Figure 1 Career examples and rescaled impact distributions in four creative domains. (**a**) The career trajectory of Stanley Kubrick. On the horizontal axis, we show the release year of his movies, while on the vertical axis we show the impact of each movie, captured by the number of ratings received from IMDb users. (**b**) The career of Michael Jackson. We show the release year of his songs and the song impact captured by the total play count on the music provider LastFM. (**c**) The career of Agatha Christie. We report her books' publication dates and the book impact, captured by the number of ratings they received on Goodreads. (**d**) Publication history of Paul Erdős, mathematician and graph theorist, based on his record in the Web of Science database. The paper impact is measured by the number of total citations 10 years after publication. (**e**–**h**) The rescaled cumulative impact distribution $P(p_{i,\alpha})$, where $p_{i,\alpha} = S_{i,\alpha}/Q_i$ for (**e**) movies of directors, (**f**) tracks of musicians active in pop music, (**g**) books of authors, and (**h**) papers of mathematicians. The panels show that when we rescale the impact value of each product of an individual by his/her Q parameter, their distribution collapses onto roughly the same aggregated curve, marked by continuous colored lines. The distribution of 50 randomly chosen individuals is visualized by light grey lines

sure that impacts are comparable across fields, we used a previously introduced rescaling method [31] (SI Eq. (2)).

We also found that cumulative impact measures, like rating counts for movies, show high correlations with other cumulative measures, indicating that the impact patterns do not depend on the chosen cumulative measure. However, cumulative impact measures show low correlations to averaged measures, like the average movie rating, possibly due to the different nature of the social processes generating them. This finding might also reflect imbalances between popularity and quality, which have been found and discussed across several domains in the literature. For instance, recent work highlighted a pronounced difference between performance (following a normal distribution) and popularity (following heavy-tailed distributions) of tennis players [32]. Similarly, discrepancies between quality and popularity have been reported in controlled experiments in an artificial music market

[33], and in a social media platform where the quality of uploaded photos does not predict their popularity [34]. For further details, see SI Section S1.3.

2 The *Q*-model: decomposing luck and individual ability in impact

Kubrick's highest impact movie was released 30 years after his career start, while Michael Jackson had his biggest hit earlier in his career. These anecdotal examples suggest that a career's biggest hit can occur at any time. Indeed, a rigorous analysis of our data sets indicates that any work in a career has an equal chance to be the highest impact work, following the so-called *random-impact-rule*, consistently with what was previously found for large data sets of artists and scientists [9, 10] (SI Section S2.1 for a replication of this analysis). The magnitude of a career impact is not random though: individual impact distributions differ broadly from each other. These broad differences are reproduced and explained by several models, such as a cumulative advantage-based approach by Simkin et al. [35], and the so-called *Q*-model, a mechanistic stochastic model approach by Sinatra et al. [9] (SI Section S2.2). According to the *Q*-model, the impact $S_{i,\alpha}$ of a work α created by an individual i can be decomposed as the product of two independent factors $S_{i,\alpha} = Q_i p_{i,\alpha}$, where Q_i is an individual variable, depending only on the career history of individual i (and is robust throughout creative careers, as shown in SI Section S2.5), and $p_{i,\alpha}$ is a stochastic variable, independently drawn for every work from a field-specific distribution. The values of Q_i and $p_{i,\alpha}$ are obtained by maximizing a likelihood function which takes as input all the impact S of all products of all creative careers in a given field [9, 36].

Next, we assumed that the covariance σ_{QN}^2 between the distributions of the productivity N (number of creative products an individual has) and the parameter Q is negligible compared to the variance of the p and N distributions – an assumption we verify and validate in SI Sections S2.2–S2.3. Thanks to this assumption, we can write a simple approximated formula for Q_i:

$$Q_i = e^{\langle \log S_{i,\alpha} \rangle - \mu_p},\tag{1}$$

where μ_p is the mean of the p distribution within a given field. Equation (1) indicates that the exponent of Q_i is the average of the order of magnitude of the impact of i's works, minus a constant equal for all individuals in a field. To establish whether the *Q*-model reproduces the individual impact distributions in our data sets, we first check the hypothesis that both S and N follow log-normal distributions (SI Section S2.3). We then estimate the parameters associated with the distributions of p and Q, finding that within each creative domain Q_i and $p_{i,\alpha}$ are both log-normally distributed (SI Section S2.3.3).

The measured negligible covariances σ_{pN}^2 and σ_{pQ}^2 predict that the individual rescaled impact, $p_{i,\alpha} = S_{i,\alpha}/Q_i$, should follow a universal distribution, independent of Q_i. We use this prediction to validate the model in our data sets: we measure the distribution $p_{i,\alpha} = S_{i,\alpha}/Q_i$ and show that it collapses roughly on a single curve for different careers (Figs. 1e–h). Since this rescaled distribution is independent of any individual variables like N_i and Q_i, we can interpret p as a "luck factor" driving impact [9]. Finally, we compare the data with the scaling of the highest impact work with productivity as predicted by the *Q*-model, and show that the *Q*-model gives significantly better results than the random model (SI Section S2.4).

A single high impact work in a career is not sufficient to have a high Q_i; rather an individual needs to perform consistently well throughout her career. For instance, the movie director with the highest Q_i, Christopher Nolan, has a $Q_i = 1719.3$, due to his many high impact movies like "Inception" or "Interstellar". In contrast, one-hit wonders, who achieved fame with a single song or movie, and whose success was neither anticipated nor repeated throughout their career with many high impact works, are typically characterized by lower values of Q_i. An example is Michael Curtiz (1886–1962), director of the all-time classic Casablanca, who has only a modest $Q_i = 4.8$ as he did not direct any other movies with outstanding impact. In this case, the large impact of their career's biggest hit is explained by a lucky draw of a high p, rather than being due to the individual ability to consistently produce work of high impact, encoded in a high Q. Taken together, the Q-model reproduces well the career impact of individuals in our data sets.

3 From the *Q*-model to classical test theory to compare luck across different domains

Here we introduce a quantitative approach, based on the Q-model, to compare the fluctuations in luck and variations in the typical impact across different creative fields. Recalling the impact decomposition $S_{i,\alpha} = Q_i p_{i,\alpha}$ presented in Sect. 2, we can write:

$$\hat{S}_{i,\alpha} = \hat{Q}_i + \hat{p}_{i,\alpha}, \tag{2}$$

where $\hat{S}_{i,\alpha} = \log S_{i,\alpha}$, $\hat{Q}_i = \log Q_i$ and $\hat{p}_{i,\alpha} = \log p_{i,\alpha}$. Because p and Q are log-normally distributed (SI Section S2.3.3), $\hat{p}$ and $\hat{Q}$ are normally distributed. In addition, the covariance $\sigma_{pQ}^2 \approx 0$, then $\sigma_{\hat{p}\hat{Q}}^2 \approx 0$. Therefore, Eq. (2) takes the form proposed by classical test theory [24, 37–41] for decomposing the measured value of a certain quantity. Namely, according to this theory, the measurable value of an observed attribute, in this case $\hat{S}$, can be decomposed as the sum of two uncorrelated variables both following normal distributions. One of these two variables encodes the true score of the quantity, in this case, $\hat{Q}$, and the other variable encodes a random error term, $\hat{p}$ (Fig. 2a). The two normal distributions of the variables $\hat{Q}_i$ and $\hat{p}_{i,\alpha}$ are in line with previous studies, suggesting that individual based quantities, such as skill and ability [42, 43], and global ones such as luck, are typically normally distributed [20, 23, 40–42, 44].

Building on Eq. (2), on the properties of normal distributions and on the measured properties of the Q and p variables in our data sets, we can express the variances of $\hat{S}_{i,\alpha}$, $\sigma_{\hat{S}}^2$, as:

$$\sigma_{\hat{S}}^2 = \sigma_{\hat{Q}}^2 + \sigma_{\hat{p}}^2, \tag{3}$$

where $\sigma_{\hat{p}}^2$ and $\sigma_{\hat{Q}}^2$ are the variance of the distributions of $\hat{p}$ and $\hat{Q}$, respectively. This decomposition allows us to measure the relative importance of the luck component compared to the individual component in determining impact. Building on previous work [40, 41], we define the *randomness* index R capturing the share of luck in the overall impact variance as:

$$R = \frac{\sigma_{\hat{p}}^2}{\sigma_{\hat{Q}}^2 + \sigma_{\hat{p}}^2} = \frac{\sigma_{\hat{p}}^2}{\sigma_{\hat{S}}^2}. \tag{4}$$
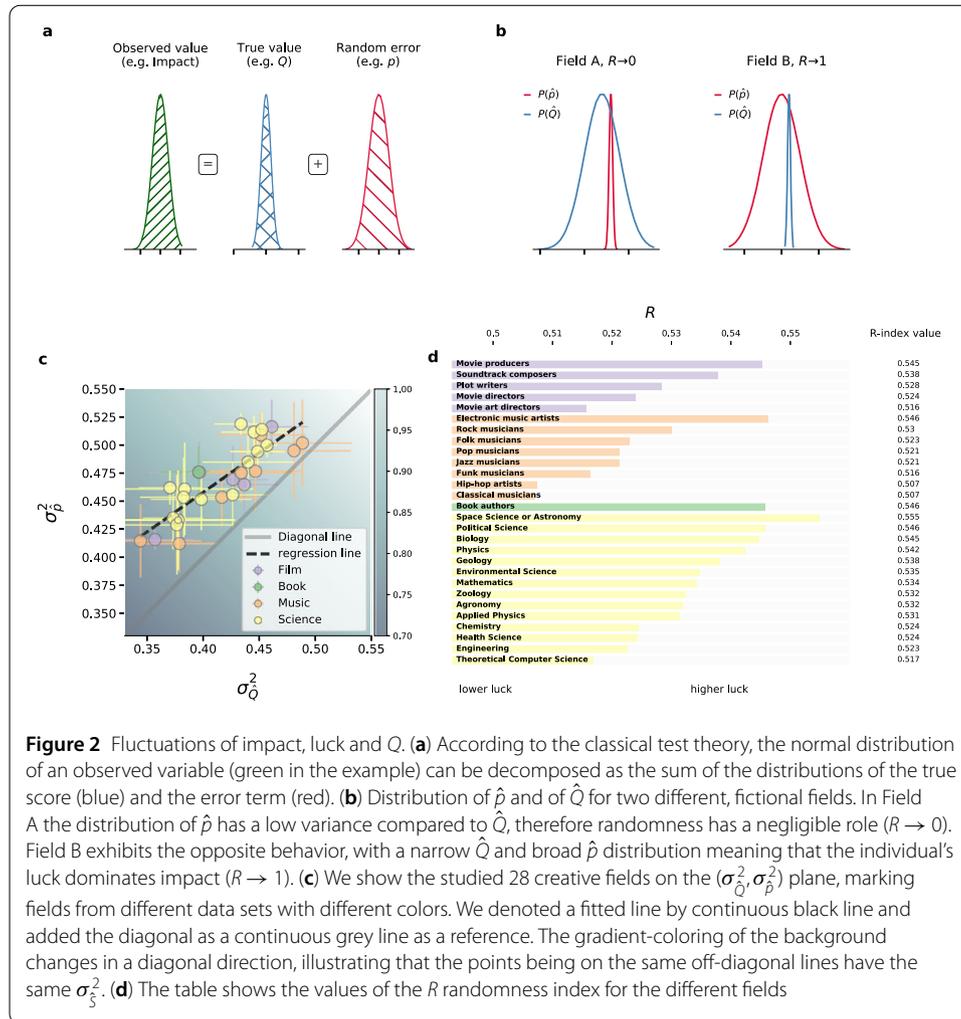
Figure 2 Fluctuations of impact, luck and Q. (**a**) According to the classical test theory, the normal distribution of an observed variable (green in the example) can be decomposed as the sum of the distributions of the true score (blue) and the error term (red). (**b**) Distribution of $\hat{p}$ and of $\hat{Q}$ for two different, fictional fields. In Field A the distribution of $\hat{p}$ has a low variance compared to $\hat{Q}$, therefore randomness has a negligible role ($R \rightarrow 0$). Field B exhibits the opposite behavior, with a narrow $\hat{Q}$ and broad $\hat{p}$ distribution meaning that the individual's luck dominates impact ($R \rightarrow 1$). (**c**) We show the studied 28 creative fields on the ($\sigma_{\hat{Q}}^2, \sigma_{\hat{p}}^2$) plane, marking fields from different data sets with different colors. We denoted a fitted line by continuous black line and added the diagonal as a continuous grey line as a reference. The gradient-coloring of the background changes in a diagonal direction, illustrating that the points being on the same off-diagonal lines have the same $\sigma_{\hat{S}}^2$. (**d**) The table shows the values of the R randomness index for the different fields

When individuals in a domain have a similar ability, captured by a narrow $\hat{Q}$ distribution, differences in impact are mainly driven by luck, and we have that $R \rightarrow 1$. In contrast, when $\hat{p}$ has a low variance compared to S, then $R \rightarrow 0$, and luck plays only a small role. This index allows us to compare the role of randomness across 28 different creative fields (Fig. 2b).

4 Randomness in creative careers

In which creative domains are inequalities driven more by luck than by individual ability? Using the Q-model, we measure $\sigma_{\hat{Q}}^2$ and $\sigma_{\hat{p}}^2$ for 28 types of creative careers in the movie, music, and book industries, and in science (Fig. 2c). We also report the linear regression between $\sigma_{\hat{p}}^2$ and $\sigma_{\hat{Q}}^2$ (black dashed line on Fig. 2c). This figure offers a number of findings. First, we observe that all the fields are placed above the diagonal line ($\sigma_{\hat{p}}^2 > \sigma_{\hat{Q}}^2$), indicating that within each domain fluctuations in luck are broader than those in the typical career impact. Second, we do not observe any domain-specific clustering on the ($\sigma_{\hat{Q}}^2, \sigma_{\hat{p}}^2$) plane, which suggests that the studied four domains do not differ from each other in the magnitude of the effects of random fluctuations. Third, we report that the linear regression has a slope ($s \approx 0.7$) lower than one; therefore, it intercepts the diagonal for high $\sigma_{\hat{Q}}^2$. Also, the regression slope due to simple geometric reasons is equal to the ratio $\sigma_{\hat{p}}^2 / \sigma_{\hat{Q}}^2$. Consequently, the values of $\sigma_{\hat{p}}^2$ increase slower as the values of $\sigma_{\hat{Q}}^2$ (symmetric increase is illustrated by

the shading on Fig. 2c). Hence large fluctuations in impact are dominated by larger fluctuations in the individual ability, captured by Q, in comparison to fluctuations in luck.

Next, we measure the randomness index R of Eq. (4) to compare the characteristics of career success across domains (Fig. 2d). The values of R span from a minimum of 0.5 for classical music, indicating that luck and Q have the same variance, to a maximum of 0.55 for space science and astronomy, indicating that the luck variance is slightly higher than the variance of Q. To ensure that these small differences in R are statistically significant, we performed a Mann–Whitney U test on the Q and p distributions of all possible field pairs (see S2.3.3 and Fig. S7) and found that the differences in R, even if tiny, are significant.

When we observe the R index values within and across fields, we find that on the one hand, within the movie industry, producers' careers are the most driven by luck, followed by composers. On the other hand, being an art director is associated with the lowest R index, suggesting that achieving high impact as an art director happens less by chance than in other professions within the movie industry. It is also interesting to compare the randomness index of scriptwriters ($R = 0.528$) and book authors ($R = 0.546$), due to the apparent similar nature of these two creative careers. The value of the indices show that writing for the movie industry is less driven by luck than in the book industry. In music, classical and hip-hop are the most robust against luck fluctuations with the lowest randomness index of our data set, $R = 0.507$. This could be explained by classical music being more dependent on skills, experience, and musical training. Regarding hip-hop music, we could speculate that being largely an underground genre, it is less exposed to the rich-gets-richer effect and leaves more space for rising junior individuals. In contrast, the most popular genres, namely electronic music ($R = 0.546$) and rock music ($R = 0.530$) are on the other side of the range with the highest R. These two genres contain the largest number of one-hit-wonder careers; therefore impact has more pronounced fluctuations. Regarding science, we find a wider range of randomness, with space science and astronomy ($R = 0.555$) and political science ($R = 0.546$) being the highest R-index fields, and theoretical computer science ($R = 0.517$) and engineering ($R = 0.523$) being among the least influenced fields by luck fluctuations.

5 The role of collaboration networks

In the previous sections, we have analyzed the randomness and magnitude of impact focusing on individual careers. However, a movie, a song or a paper is rarely the result of the work of only one individual. Therefore next we ask: Can collaborations between individuals improve our ability to predict the timing of career hits? Previous research suggests that scientific career impact and network position can be connected, for instance, according to regressive models on predicting success by using network features. [15, 45–49].

To study network effects, we first reconstruct the temporal aggregated network of movie directors, pop musicians, and mathematicians to quantify the relationship between their network positions and impact. We use a yearly time resolution. In these weighted undirected networks, each individual is represented by a node. To compute the weight of each edge we took the set of products ($P_i(T)$), e.g. publications or movies, each individual i contributed to up until year T. Then, we defined the weight of the connection between nodes i and j at year T as $w_{ij}(T)$, the Jaccard-index of the sets of works of the two individuals i
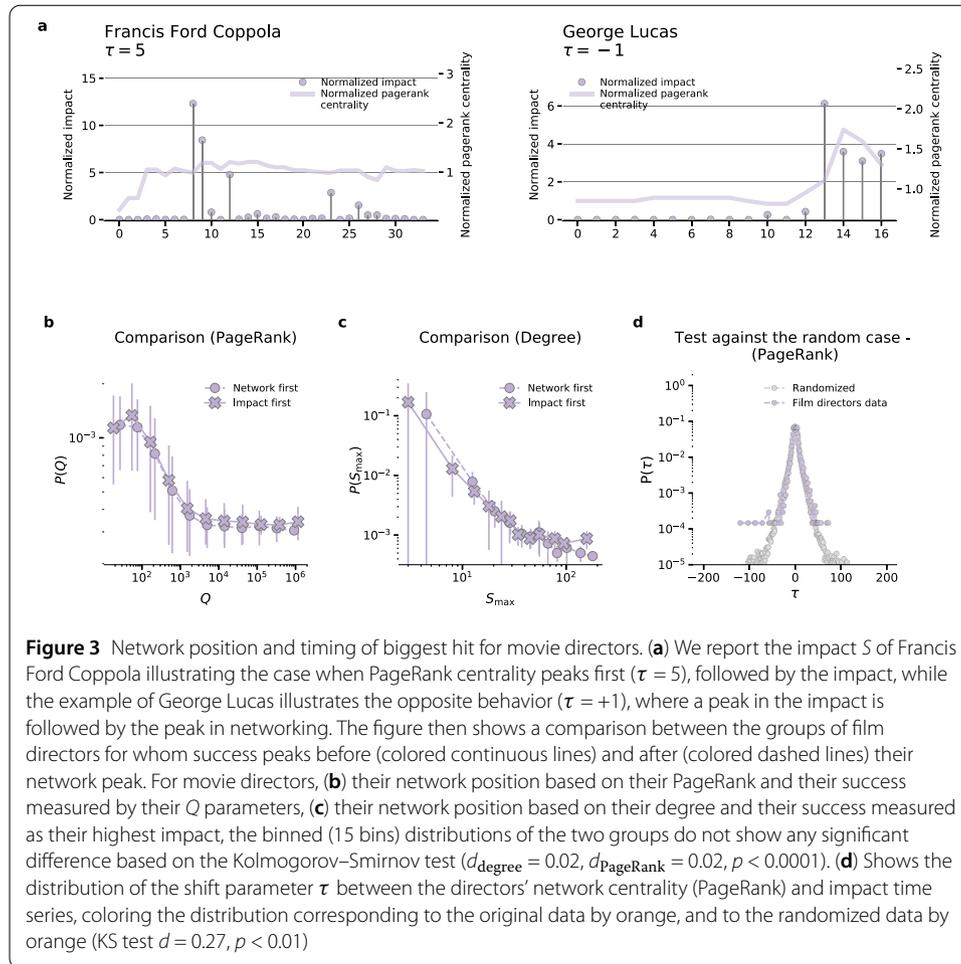
and j:

$$w_{ij}(T) = \frac{|P_i(T) \cap P_j(T)|}{|P_i(T) \cup P_j(T)|}, \tag{5}$$

that is the number of works both individuals collaborated on, divided by the total number of works they contributed to until year T. Based on this definition, the final aggregated collaboration network of movie directors consists of 8,091,208 links between 184,220 people active between 1927–2017 (giant connected component only). In the pop music network, we have 52,366 musicians active between 1926–2017 connected by 8,232,349 links, while the network of mathematicians consists of 94,755 links between 27,401 mathematicians during 1944–2016.

For each individual, we measure her degree centrality, PageRank centrality, clustering coefficient, node strength, betweenness centrality, closeness centrality, network constraint [50], and coreness centrality [51] in the aggregated network at the time she has produced her different pieces of works. We note that while some of these measures are highly correlated, some of them are not, encoding different angels of the individuals' network positions (details showed in SI Section S3 and SI Figure S14). We then create individual time-series for each of these network measures, where time points correspond to the works in the individual careers. Finally we study these network-based time-series together with the evolution of individual the impact over a career. Our hypothesis is that the dynamics of the network position and the dynamics of impact are correlated over time, however with a delay of τ. We measure τ by shifting the network time-series with respect to the impact time-series, and choose the value for which we obtain the maximum correlation between the time-series. For further information on the network analysis see SI Section S3.

By analyzing the time-series of movie directors, pop musicians, and mathematicians, we find that there are two groups of individuals: those for whom the network measures peak before the highest impact work occurs, and those for whom the peak occurs after. For example, the director Francis Ford Coppola ($\tau = 5$) belongs to the first category, while George Lucas ($\tau = -1$) to the second (Fig. 3a). However, there are no significant differences between these two groups when we consider impact: the two groups have similar distributions of the Q-parameter (Fig. 3b) and of the magnitude of the highest success withing a career (Fig. 3c). Further details on these findings can be found in SI Figures S11–S12 and SI Tables S6–S7.

Given the indistinguishable nature of impact in these two groups, we ask whether the observed shift τ is different from that obtained from reshuffled time-series, where time correlations are canceled. We measure the distribution of the delay parameter τ and compare it to the distribution of a randomized data set in which the time-series are randomly reshuffled. The two distributions are closely overlapping, confirmed by the double-sided Kolmogorov–Smirnov test (Fig. 3d, details about the KS test in SI Section S3 and Table S8). Taken together, the collaboration network among individuals does not improve our ability to predict the timing of the biggest hit since it is similarly likely to happen before or after the network peak, suggesting that chance has much higher importance than the collaboration network to determine the timing of the biggest hit within a career. Our results on the timing of career hits and network position complement previous work where collaborations between authors were positively associated with scientific impact [48, 52].

Figure 3 Network position and timing of biggest hit for movie directors. (**a**) We report the impact S of Francis Ford Coppola illustrating the case when PageRank centrality peaks first ($\tau = 5$), followed by the impact, while the example of George Lucas illustrates the opposite behavior ($\tau = +1$), where a peak in the impact is followed by the peak in networking. The figure then shows a comparison between the groups of film directors for whom success peaks before (colored continuous lines) and after (colored dashed lines) their network peak. For movie directors, (**b**) their network position based on their PageRank and their success measured by their Q parameters, (**c**) their network position based on their degree and their success measured as their highest impact, the binned (15 bins) distributions of the two groups do not show any significant difference based on the Kolmogorov–Smirnov test ($d_{degree} = 0.02$, $d_{PageRank} = 0.02$, $p < 0.0001$). (**d**) Shows the distribution of the shift parameter τ between the directors' network centrality (PageRank) and impact time series, coloring the distribution corresponding to the original data by orange, and to the randomized data by orange (KS test $d = 0.27$, $p < 0.01$)

6 Conclusion

In this work, we provided a framework to understand and quantify the role of randomness in the success of creative fields across different domains. To understand the emergence of high-impact creative works, we built large-scale data sets and investigated thousands of careers from the movie, book, and music industries, and science. We built on an existing model, known as Q-model, to decompose the impact of the individual creative works into two independent components, one expressing the ability of an individual to have consistently high or low typical impact, captured by the *Q-parameter*, and one associated to random fluctuations, capturing the role of *luck*. We also cast the model into the framework of classical test theory, which aims to disentangle the true score of a variable from noisy fluctuations.

Using this framework, we found that on average, fluctuations in the impact of single creative works are more influenced by luck than by individual ability, as all the fields in Fig. 2c are placed above the diagonal. However, we conclude a change in trends: the fluctuations in the individual-based parameter are more pronounced for fields with large fluctuations in impact. The extrapolated linear trend between fluctuations in individual parameters and luck predicts that when impact fluctuations become large ($\sigma_Q^2 \approx 0.7$), the fluctuations in individual parameter become larger than the random ones. In this fictional case, the fluctuations in impact would be mainly due to individual differences. Moreover, we found

that sub-disciplines within different do not show spatial clustering on the studied variance plane of Fig. 2c. This absence of segregation of the studied four creative domains suggests the magnitude of luck is not a distinctive feature of domains.

We introduced a synthetic randomness index, defined as the relative ratio of the variance of the random component to that of success, and investigated its values across different domains. We found that the randomness index varies in a relatively narrow range, despite the differences in typical impact within creative professions. This further confirms the lack of distinct typical scales of random fluctuations associated with the four different domains investigated in the paper. Finally, in this narrow range of randomness, we found that the careers with the highest values of luck are those of movie producers, electronic music artists, book authors, and scientists working in the fields of space science, and political science. On the other hand, randomness has the lowest influence on hip-hop and classical music, theoretical computer science, and movie art directors.

Finally, we also studied the temporal relationship between success and centrality in the collaboration network for movie directors, pop musicians, and mathematicians as a case study. For each individual, we compared the temporal evolution of their network centrality to the evolution of their impact. We found that these two are correlated, yet with a delay. We computed these delay parameters and found two distinct classes of creative careers regardless of their creative domain. Individuals belonging to the first group produce their big hit first, and become well-connected in the network only after the occurrence of the hit, while people falling into the second category first build favorable connections, and produce their big hit afterward. However, we found no correlation between individual impact and the social environment the individual belongs to. We also showed that the delay between the impact and the network time-series follows the same distribution as randomized data. In conclusion, our analysis revealed that the evolution of the individual position in the network is random in respect to the timing of the career hits, regardless of the particular choice of network measures.

Future studies could further untangle the individual Q-parameter and pinpoint what Q means, for example, in terms of access to resources or early career steps. Also, the variable p, interpreted here as luck, could contain more information than just randomness if further data is incorporated in the analysis. Nevertheless, its universal distribution across careers suggests that this information is homogeneously distributed among individuals.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-020-00227-w.

Additional file 1. Supplementary information. (PDF 1.3 MB)

Abbreviations
IMDb, Internet Movie Database; SI, Supplementary Information.

Author details
[1]Department of Network and Data Science, Central European University, Budapest, Hungary. [2]Department of Computer
Science, IT University of Copenhagen, Copenhagen, Denmark. [3]ISI Foundation, Torino, Italy. [4]Complexity Science Hub
Vienna, Vienna, Austria.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Lehman HC (1953) Age and achievement. Princeton University Press, Princeton
2. Campbell DT (1960) Blind variation and selective retentions in creative thought as in other knowledge processes. Psychol Rev 67(6):380
3. Simonton DK (1984) Creative productivity and age: a mathematical model based on a two-step cognitive process. Dev Rev 4(1):77–111
4. Simonton DK (1988) Age and outstanding achievement: what do we know after a century of research? Psychol Bull 104(2):251
5. https://webofknowledge.com. Web of science. Date accessed: 2018.11.06
6. Spitz A, Horvát E (2014) Measuring long-term impact based on network centrality: unraveling cinematic citations. PLoS ONE 9(10):e108857
7. Yucesoy B, Wang X, Huang J, Barabási A-L (2018) Success in books: a big data approach to bestsellers. EPJ Data Sci 7(1):7
8. Williams OE, Lacasa L, Latora V (2019) Quantifying and predicting success in show business. Nat Commun 10(1):1–8
9. Sinatra R, Wang D, Deville P, Song C, Barabási A-L (2016) Quantifying the evolution of individual scientific impact. Science 354(6312):aaf5239
10. Liu L, Wang Y, Sinatra R, Giles CL, Song C, Wang D (2018) Hot streaks in artistic, cultural, and scientific careers. Nature 559(7714):396
11. Guimera R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. Science 308(5722):697–702
12. Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. Science 342(6157):468–472
13. Wang D, Song C, Barabási A-L (2013) Quantifying long-term scientific impact. Science 342(6154):127–132
14. Lee Y-N, Walsh JP, Wang J (2015) Creativity in scientific teams: unpacking novelty and impact. Res Policy 44(3):684–697
15. Zagovora O, Weller K, Janosov M, Wagner C, Peters I (2018) What increases (social) media attention: research impact, author prominence or title attractiveness? In: Proceedings of the 23rd international conference on science and technology indicators, pp 1182–1190
16. Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B et al (2018) Science of science. Science 359(6379):eaao0185
17. Jadidi M, Karimi F, Lietz H, Wagner C (2018) Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. Adv Complex Syst 21(03n04):1750011
18. Flugel JC, West DJ (1964) A hundred years of psychology
19. Petersen AM, Jung W-S, Yang J-S, Stanley HE (2011) Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. Proc Natl Acad Sci 108(1):18–23
20. Pluchino A, Biondo AE, Rapisarda A (2018) Talent vs Luck: the role of randomness in success and failure. arXiv preprint arXiv:1802.07068
21. Pluchino A, Burgio G, Rapisarda A, Biondo AE, Pulvirenti A, Ferro A, Giorgino T (2019) Exploring the role of interdisciplinarity in physics: success, talent and luck. PLoS ONE 14(6):e0218793
22. Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: toward an objective measure of scientific impact. Proc Natl Acad Sci 105(45):17268–17272
23. Crocker L, Algina J (1986) Introduction to classical and modern test theory. ERIC, U.S. Department of Education
24. Lord FM (1965) A strong true-score theory, with applications. Psychometrika 30(3):239–270
25. www.imdb.com. Internet movie database. Date accessed: 2017.02.04
26. www.discogs.com. Discogs music release database. Date accessed: 2017.02.04
27. Hartnett J (2015) Discogs.com. Charlest Advis 16(4):26–33
28. www.last.fm. LastFM. Date accessed: 2017.02.06
29. www.goodreads.com. Goodreads book database. Date accessed: 2017.02.04
30. Garfield E, Merton RK (1979) Citation indexing: its theory and application in science, technology, and humanities, vol 8. Wiley, New York

31. Radicchi F, Castellano C (2011) Rescaling citations of publications in physics. Phys Rev E 83(4):046116
32. Yucesoy B, Barabási A-L (2016) Untangling performance from success. EPJ Data Sci 5(1):17
33. Salganik MJ, Dodds P, Sheridan P, Watts DJ, (2006) Experimental study of inequality and unpredictability in an artificial cultural market. Science 311(5762):854–856
34. Aiello LM, Schifanella R, Redi M, Svetlichnaya S, Liu F, Osindero S (2017) Beautiful and damned. Combined effect of content quality and social ties on user engagement. IEEE Trans Knowl Data Eng 29(12):2682–2695
35. Simkin MV, Roychowdhury VP (2007) A mathematical theory of citing. J Am Soc Inf Sci Technol 58(11):1661–1673
36. Vásárhelyi G, Virágh C, Somorjai G, Nepusz T, Eiben AE, Vicsek T (2018) Optimized flocking of autonomous drones in confined environments. Sci Robot 3(20):eaat3536
37. Kristof W (1974) Estimation of reliability and true score variance from a split of a test into three arbitrary parts. Psychometrika 39(4):491–499
38. Kline T (2005) Psychological testing: a practical approach to design and evaluation. Sage, Thousand Oaks
39. Kean J, Reilly J (2014) Item response theory. In: Handbook for clinical research: design, statistics and implementation, pp 195–198
40. Mauboussin MJ (2010) Untangling skill and luck: how to think about outcomes—past, present, and future. Legg Mason Capital Management
41. Mauboussin MJ (2012) The success equation: untangling skill and luck in business, sports, and investing. Harvard Business Press, Brighton
42. Stewart J (1983) The distribution of talent. Marilyn Zurmuehlen Work Pap Art Educ 2(1):21–22
43. Galton F (1869) Hereditary genius
44. Allen MJ, Yen WM (2001) Introduction to measurement theory. Waveland Press, Mountain View
45. Figg WD, Dunn L, Liewehr DJ, Steinberg SM, Thurman PW, Barrett JC, Birkinshaw J (2006) Scientific collaboration results in higher citation rates of published articles. Pharmacother J Hum Pharmacol Drug Ther 26(6):759–767
46. Hsu J-W, Huang D-W (2011) Correlation between impact and collaboration. Scientometrics 86(2):317–324
47. Radicchi F (2012) In science "there is no bad publicity": papers criticized in comments have high scientific impact. Sci Rep 2:815
48. Sarigöl E, Pfitzner R, Scholtes I, Garas A, Schweitzer F (2014) Predicting scientific success based on coauthorship networks. EPJ Data Sci 3(1):9
49. Janosov M, Musciotto F, Battiston F, Iñiguez G (2020) Elites, communities and the limited benefits of mentorship in electronic music. Sci Rep 10(1):1–8
50. Burt RS (2004) Structural holes and good ideas. Am J Sociol 110(2):349–399
51. Seidman SB (1983) Network structure and minimum degree. Soc Netw 5(3):269–287
52. Petersen AM (2015) Quantifying the impact of weak, strong, and super ties in scientific careers. Proc Natl Acad Sci 112(34):E4671–E4680
53. Galton F (1889) Natural inheritance