



Untangling performance from success

Burcu Yucesoy¹ and Albert-László Barabási^{1,2,3,4*}

*Correspondence:

barabasi@gmail.com

¹Center for Complex Network Research and Department of Physics, Northeastern University, Boston, USA

²Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, USA

Full list of author information is available at the end of the article

Abstract

Fame, popularity and celebrity status, frequently used tokens of success, are often loosely related to, or even divorced from professional performance. This dichotomy is partly rooted in the difficulty to distinguish performance, an individual measure that captures the actions of a performer, from success, a collective measure that captures a community's reactions to these actions. Yet, finding the relationship between the two measures is essential for all areas that aim to objectively reward excellence, from science to business. Here we quantify the relationship between performance and success by focusing on tennis, an individual sport where the two quantities can be independently measured. We show that a predictive model, relying only on a tennis player's performance in tournaments, can accurately predict an athlete's popularity, both during a player's active years and after retirement. Hence the model establishes a direct link between performance and momentary popularity. The agreement between the performance-driven and observed popularity suggests that in most areas of human achievement exceptional visibility may be rooted in detectable performance measures.

Keywords: success; performance; popularity

1 Introduction

Performance, representing the totality of objectively measurable achievements in a certain domain of activity, like the publication record of a scientist or the winning record of an athlete or a team, captures the actions of an individual entity [1–5]. In contrast success, captured by fame, celebrity, popularity, impact or visibility, is a collective measure, representing a community's reaction to and acceptance of an individual entity's performance [6, 7]. The link between these two measures, while often taken for granted, is actually far from being understood and often controversial and lopsided. Indeed, even the most profound scientific discovery goes unnoticed if its importance is not acknowledged through discussions, talks and citations by the scientific community. The void between success and performance is well illustrated by the concepts of 'famesque', 'celebutante' or 'faminess', used to label an individual without tangible performance, but 'known for his well-knowingness' [8]. These often prompt us to see fame and success as only loosely related to [9–11] and often divorced [11–16] from performance. This dichotomy is illustrated by documented examples of scientists whose popular media visibility significantly exceeds their scientific credentials [17], or by countless celebrities, from the Kardashian sisters to athletes with no or only underwhelming accomplishments [18–20], as well as by high performers like David Beckham or Tiger Woods who are frequently featured in the media for reasons un-

related to their professional achievements [21, 22]. The source of this dichotomy is that in most areas of human achievement it is difficult to distinguish performance from success [23]. Indeed, while we can use citations, prizes and other measures to quantify the impact of a scientific discovery, we lack objective performance measures to capture the degree of innovation or talent characterizing a particular paper or a scientist.

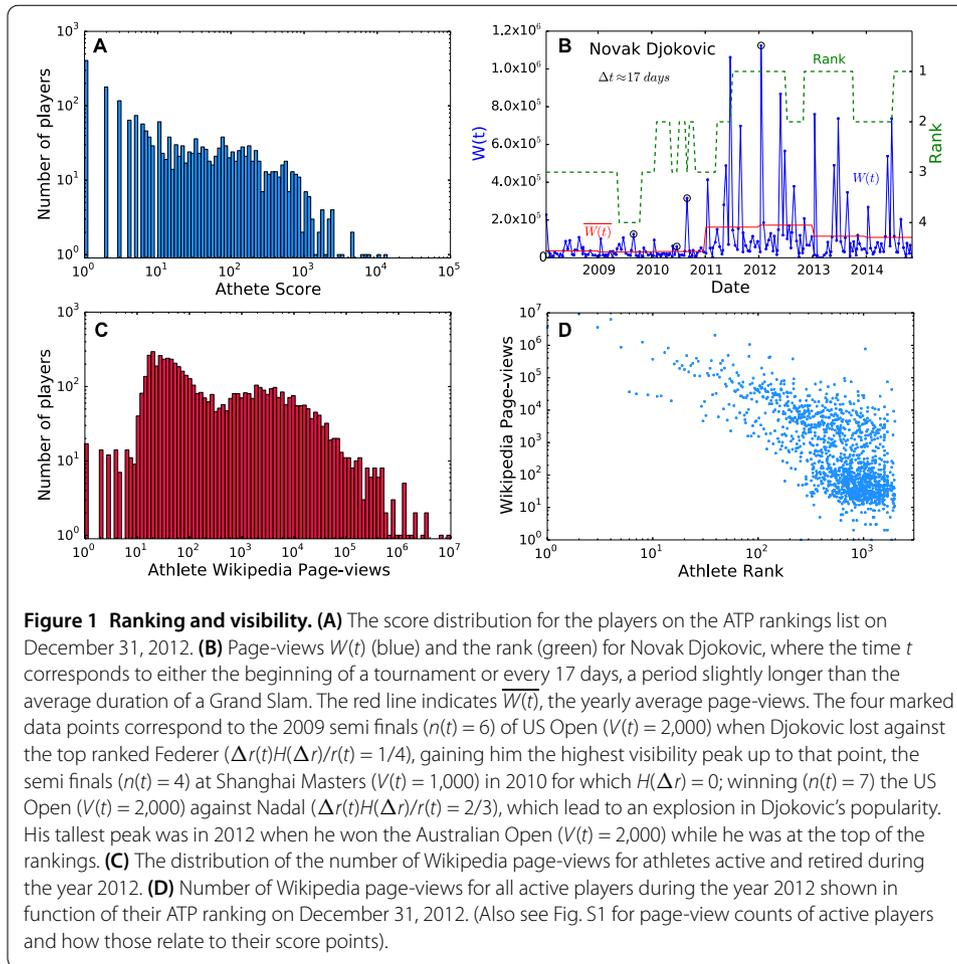
Our goal here is to explore in a quantitative manner the relationship between performance, an individual measure, and success, a collective measure capturing the societal acknowledgement of a given level of performance. Previous research, some using Google search results as a proxy for fame, have suggested that performance can indeed drive success [24–31]. Here, by exploring the time-lines of both achievement and its recognition, we uncover how performance affects success over the career of an individual. We do so through sports, an area where performance is accurately recorded in terms of number of wins, place in rankings or career records [32–38].

Sports is characterized by an equally obsessive focus on popularity and fame, which strongly affects an athlete's market value [39], and previous research shows that fame can be an appropriate proxy for accomplishment [40]. Yet, in sports too, performance and success often follow different patterns, illustrated by the fact that only a small fraction of the earnings of a professional athlete is tied directly to his/her performance on the field, the vast majority coming from endorsements, determined more by the athlete's perceived success and popularity. For example, only \$4.2 million of Roger Federer's \$56.2 million reported 2014 income was from tournament prizes [41], the rest came from endorsements tied to the public recognition of the athlete. Yet, Novak Djokovic, who was better ranked than Federer during 2014, received over \$12.1 million prize money but only \$21 million via endorsements, about a third of Federer's purse. The fact that professional performance does not uniquely determine reward is further illustrated by Anna Kournikova, who in 2003 was the second best paid female tennis player, despite never reaching higher than the 8th place in rankings. Her popularity and consequential financial prowess is often attributed to her photogeneity and media-friendly personality, traits outside of her sports performance. This and many other well documented cases of 'famines' raise an important question: What performance factors affect popularity and how do they do so? In other words, can performance explain popularity and fame and if so, to what extent? These are fundamental issues in most areas that aim to fairly reward excellence, from science to education and business.

2 Performance and success in tennis

Professional tennis players accumulate score points based on their winning record during the previous year (Figure 1(A)). By ordering the players based on their score we obtain the official Association of Tennis Professionals (ATP) Rankings, $r(t)$, an accurate measure of a tennis player's performance relative to other players, having $r = 1$ for the top player while r is high for low-performing athletes.

We measure each player's time-resolved visibility through the number of hourly visits to its Wikipedia page [42–44], a sensitive time-dependent measure of the collective interest in a player (Figure 1(B)). We define a player's popularity or fame through his cumulative visibility, representing the average number of visits his Wikipedia page acquires during a year (red line in Figure 1(B)). Figure 1(C) indicates that top players can gather a total of 10^7 visits in a year, while those at the bottom of the rankings collect as few as 10, documenting popularity differences between players that span orders of magnitude. As Figure 1(D)

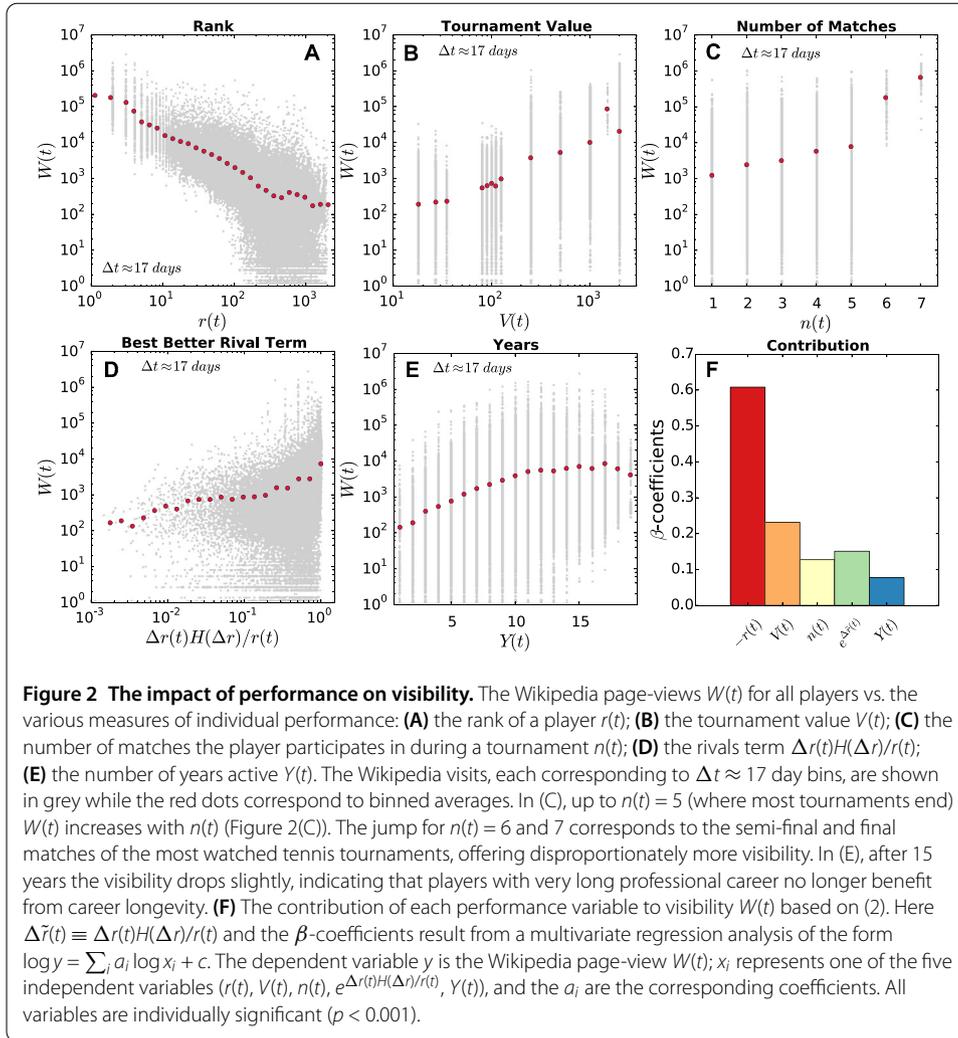


indicates, a player's rank (performance) and Wikipedia visits (popularity) are correlated: The lower the ranking, the higher are the Wikipedia page-views. Yet, we also observe a significant scattering: Athletes ranked around $r = 1,000$ can gather anywhere from 10 to 10^5 visits per year, wide differences that support the impression that popularity is often divorced from performance.

To address the degree to which performance determines popularity, we reconstructed the number of Wikipedia visits $W(t)$ for all ranked professional male tennis players between 2008 and 2015 (Figure 1(B)), together with their time-dependent history of achievements on the field (supplementary material S3 in Additional file 1). The two datasets allow us to identify several performance-related factors that influence a player's visibility, and eventually his popularity:

(i) *Rank $r(t)$* : Figure 2(A) shows the measured Wikipedia page-views $W(t)$ vs. a player's momentary rank $r(t)$, indicating that the number of visits rapidly drop with the increasing rank of a player. It also shows that the variations in $W(t)$ is higher for players with lower performance (i.e. higher rank) (Figure 1(B)).

(ii) *Tournament value $V(t)$* : The more points $V(t)$ a tournament offers to the winner, a measure of the tournament prestige, the more visibility it confers to its players, win or lose (Figure 2(B)). For example, the early peaks in Djokovic's visibility correspond to his participation in the US and Australian Opens, two high value tournaments (Figure 1(B)).



(iii) *Number of matches $n(t)$* : The more matches an athlete plays within a tournament, reflecting his advance within the competition, the more exposure he receives (Figure 2(C)). As the losing player is eliminated, $n(t)$ also determines the amount of points the player collects within a tournament. Up to $n(t) = 5$ (where most tournaments end) we see a steady increase of $W(t)$ with $n(t)$. The jump for $n(t) = 6$ and 7 exist only for the semi-final and final matches in the most watched tennis tournaments, offering disproportional visibility.

(iv) *Rank of the best opponent*: A match against a better ranked rival generates additional interest in a player. To capture this effect we measure the relative rank difference for each match $\Delta r(t)H(\Delta r)/r(t)$, where $\Delta r(t) = r(t) - r_{BR}(t)$ is the difference between the rank of the considered player and his best rival in the tournament; the Heaviside step function $H(\Delta r)$ is one when the opponent has a better rank and zero otherwise. The increase of the average page-views with $\Delta r(t)H(\Delta r)/r(t)$ quantifies the boost in visibility from playing against a better athlete (Figure 2(D)), like the visibility peak of Djokovic when he played against the then #1 Federer in the 2009 US Open (Figure 1(B)) (see S9 for the detailed opponent statistics).

(v) *Career length $Y(t)$* : The longer the player has been an active professional player, the more Wikipedia visitations he collects (Figure 2(E)).

Figures 2(A)-(E) document clear correlations between the performance measures (i)-(v) and the momentary visibility of a player. Yet, a linear fit $y(t) = Ax(t) + C$ of each individual performance measure to the observed Wikipedia page-views results in $R^2 < 0.1$, except for $1/r(t)$, for which $R^2 = 0.29$. Therefore, no individual performance measure can fully explain visibility, indicating that performance drives popularity through a combination of performance measures (see supplementary materials S4 in Additional file 1 for variable interdependencies). We therefore explored the predictive power of the sum of these variables with multipliers obtained via an ordinary least squares (OLS) fitting [45] process, resulting in $R^2 = 0.31$, only slightly better than $1/r(t)$ alone. We find, however, that a multiplicative process offers a much better predictive power, an OLS fitting leading to the formula

$$W_M(t) = A \frac{1}{r(t)} V(t)n(t)e^{\frac{\Delta r(t)H(\Delta r)}{r(t)}} Y(t) + C, \quad (1)$$

which yields $R^2 = 0.57$. This indicates that the influence of the performance measures (i)-(v) do not add up, but amplify each other in a multiplicative manner, allowing for the emergence of the extreme fluctuations in visibility, as observed in Figure 1(B) and Figures 2(A)-(E). By taking the logarithm of (1) and calculating the standardized β -coefficients for each term, we can evaluate how strongly each performance variable influences $W(t)$ in units of standard deviation. Figure 2(F) shows the obtained standardized β -coefficients, indicating that rank is the strongest driving force of visibility, followed by the value (prestige) of the tournaments, the rank of the opponents and the number of matches the player participated in a tournament. While career length contributes to a lesser degree, all terms are significant ($p < 0.001$).

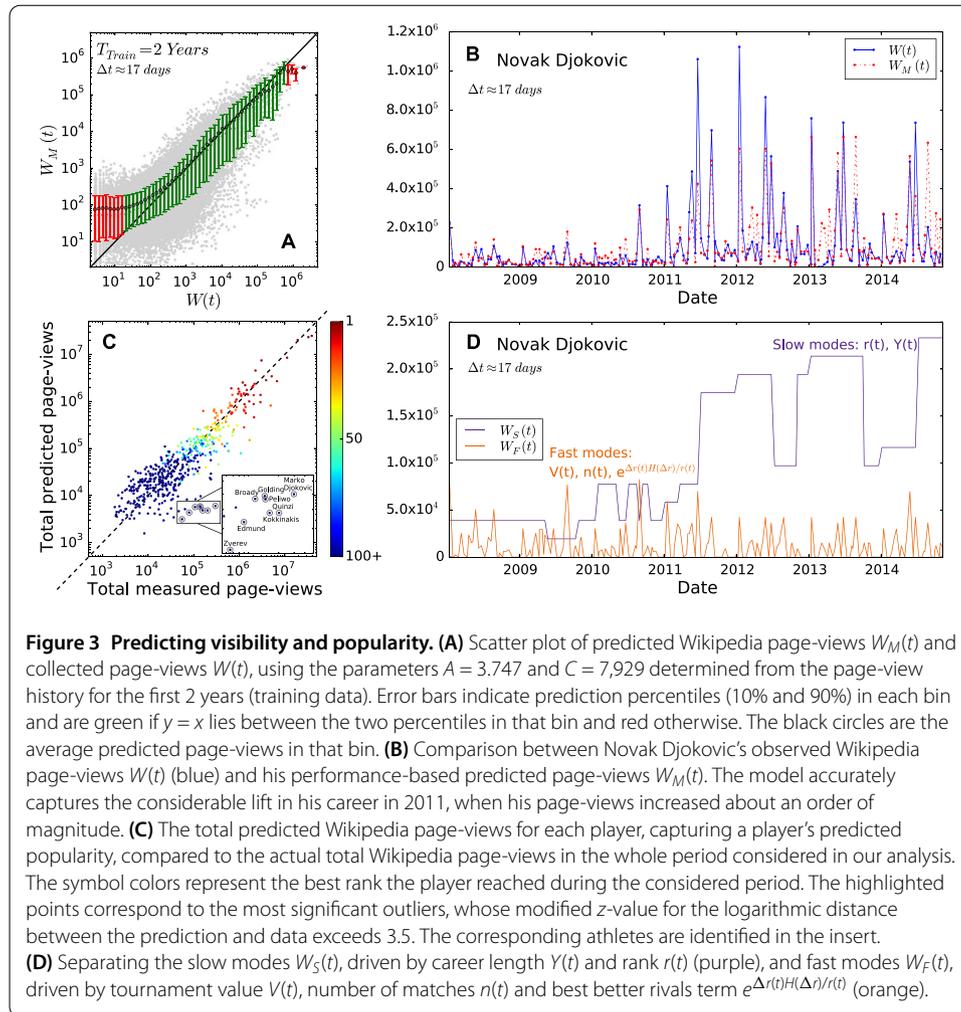
These results lead to a popularity model (PROMO), which predicts the time dependent visibility of a player $W_M(t)$ using the athlete's performance as an input,

$$W_M(t) = A \frac{Y(t)}{r(t)} V(t)n(t)e^{\Delta r(t)H(\Delta r)/r(t)} + C \frac{Y(t)}{r(t)}. \quad (2)$$

The last term accounts for periods when the player is not playing (between tournaments, periods of injury, etc.). An OLS fitting process of this two-parameter PROMO results in $R^2 = 0.60$, offering the best predictive accuracy of the models we tested (an analysis of all sub-models is provided in the supplementary materials S5 in Additional file 1). Hence (2) represents our main result, linking an athlete's visibility ($W(t)$) to his performance captured by $r(t)$, $V(t)$, $n(t)$, $\Delta r(t)H(\Delta r)/r(t)$ and $Y(t)$. While it is possible to improve the accuracy further by implementing additional measures such as including a multiplier to $n(t)$ for when $n(t) \geq 6$, the increase in accuracy is not sufficient to justify the added complexity.

We designed several tests to validate PROMO's predictive power:

(i) In a prospective study we fit the r.h.s. of (2) to $W(t)$ for 2008 and 2009, obtaining the coefficients $A = 3.747$ and $C = 7,929$, the same for all athletes (see supplementary materials S3 in Additional file 1). We then use these parameters to predict the momentary visitations $W_M(t)$ for the subsequent 5 years (Figure 3(A)). We find that the model accurately predicts the bulk of the real data (Pearson correlation coefficient is $r = 0.77$ resulting in an R^2 of 0.6). The deviation for very low Wikipedia page-views has a simple technical reason: beginning athletes often lack a dedicated Wikipedia page, hence their page-views are



counted only indirectly and incompletely on Wikipedia (see supplementary materials S3 and S7 in Additional file 1). Once a player gets a Wikipedia page, his visibility approaches the model-predicted values.

(ii) Figure 3(B) compares the time dependent visibility $W_M(t)$ to the real visitation $W(t)$ for Novak Djokovic, indicating that the model (2) accurately captures not only the explosion of his overall popularity in 2011 following his exceptional performance on the field, but also the peaks in the visitation patterns. Comparing Figure 3(B) to Figure 1(B) helps clarify the lack of significant increase in popularity when Djokovic first reached #2 in the rankings: Rankings alone cannot explain visibility; we need the full predictive power of PROMO to unveil the underlying dynamics.

(iii) The model (2) allows us to separate the role of the different parameters: Rank and career length act as *slow modes*, driving the *popularity* of an athlete, capturing his overall fame or celebrity (Figure 2(D)). In contrast the tournament value $V(t)$, the number of matches played within a tournament $n(t)$, and the rank of the best opponent represent *fast modes* that drive the *momentary visibility*, like the timing and the height of the individual visibility peaks (Figure 3(D)). As the slow and fast modes get multiplied, they together can account for the major visibility peaks on a slowly varying background, which in turn determines an athlete's overall fame.

(iv) To assess our ability to predict a player's popularity or fame from his performance, we use the total page-views a player received across several years. Figure 3(C) shows a comparison between the predicted and observed popularity of each active player between 2008 and 2015, indicating that the observed popularity $W(t)$ closely follows the performance-driven predicted popularity $W_M(t)$. A color-coding by the player's peak rank reveals that for players that reached top rankings, the accuracy of the prediction is remarkable; scattering is only seen for lower ranked players.

To understand if popularity in tennis can be induced by factors unrelated to performance, we inspected the outliers, athletes whose observed popularity is significantly higher than their performance-based popularity. We find that the outliers highlighted in Figure 3(C) (also listed in Table S3) are young players at the bottom of the rankings, who participated in only a few tournaments. An inspection of their career reveals that their added popularity is also performance driven, routed in outstanding results in junior or doubles tournaments, performance factors not considered in (2). For example Quinzi, Peliwo, Zverev and Golding reached number one or two in junior rankings, and Broady and Edmund had considerable successes in doubles. Only Marco Djokovic's visibility could not be explained by his performance. His higher than expected fame is most likely related to the attention (and potential confusion) he earns as the brother of Novak Djokovic, one of the best active tennis players.

(v) Finally, we find that the Wikipedia pages of retired players continue to attract visitors (Figure 1(C)), prompting us to ask if a player's enduring popularity can be explained by his past performance. Given that for retired players $r(t)$, $V(t)$, $n(t)$ and $\Delta r(t)H(\Delta r)/r(t)$ are not recorded, their visibility can be determined only by the second term in (2). By using the median for $r(t)$, reflecting a player's overall performance during his active career, and Y_T for the number of active years $Y(t)$, we predict an athlete's popularity during retirement as

$$W_M^{\text{inact}}(t) = C \frac{Y_T}{r_{\text{med}}}, \quad (3)$$

using the same fitting parameter we had before ($C = 7,929$). As Figure 4(A) shows, the predicted popularity of inactive players is in excellent agreement with the measured popularity, indicating that past performance is the main source of their enduring fame, at least for several years following the player's retirement (see supplementary materials S8 in Additional file 1 for outliers).

Finally we can apply the tools and insights developed above to explore the emergence of fame. For this we group players into ten categories based on their peak ranking during their career. The change in the average ranking of each group over time (Figures 4(B)-(C)) indicates that players who reach the best ranks distinguish themselves in their first 20 tournaments, i.e. they climb in rank very fast early in their careers. This effect is particularly clear in Figure 4(C), which shows the rate of change in rankings during the first 20 tournaments, indicating that players that eventually reach the top rise much faster early on than the rest. Therefore, top ranks are not reached by a slow improvement in skill, but instead young players come in with a given skill set, some remarkable, others less so, and rapidly reach the vicinity of their skill-determined ranking level, where they fluctuate for most of their career (Figure 4(B)).

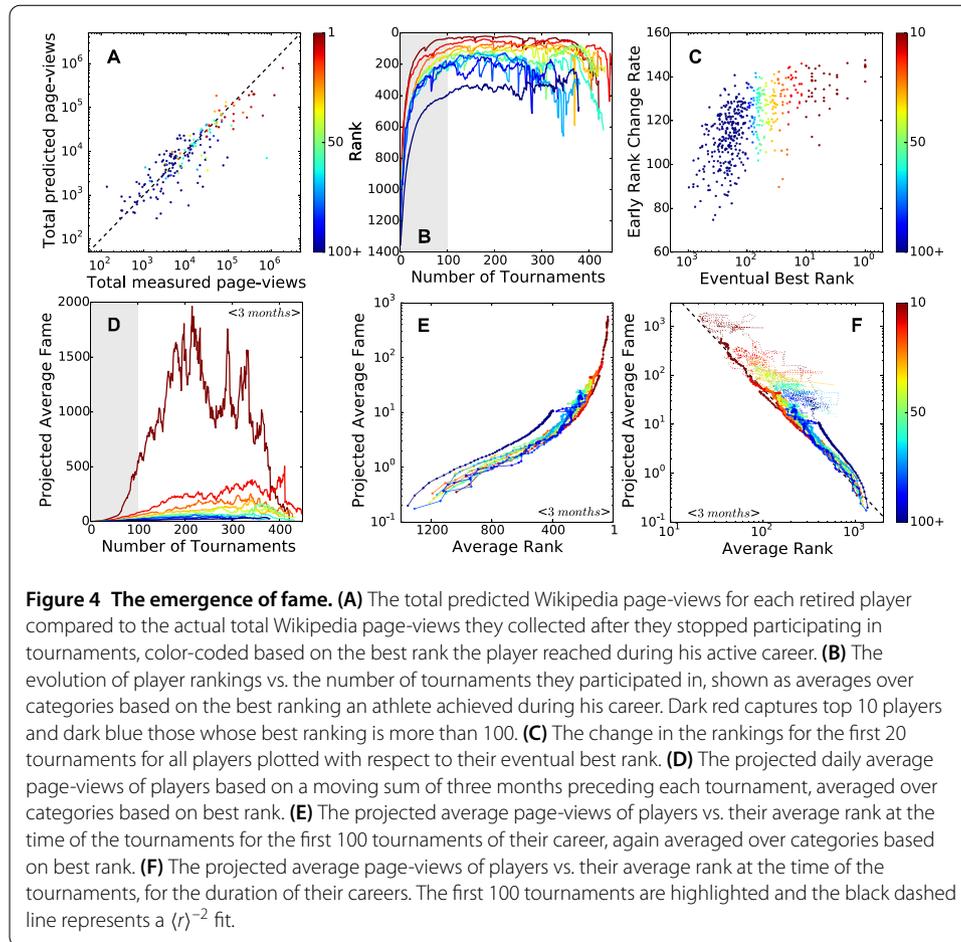


Figure 4(D) shows the projected daily average Wikipedia visitation of players grouped based on their best rank (supplementary materials S9 in Additional file 1). It allows us to uncover a highly nonlinear relationship between rank and popularity (Figure 4(E)): popularity raises fast immediately following a player's entry into the professional field, but this growth rate slows down between ranks 1,000 and 400. This is followed by an exploding popularity for the elite players, those that reach ranking 200 and below. In other words, elite players benefit from a disproportional popularity bonus, not accessible for other ranked players. Overall, we find a robust relationship between a player's average popularity and rank: At the beginning of a player's career popularity increases as $1/\langle r \rangle^\alpha$, with $\alpha \approx 2.0 \pm 0.02$, indicating that in this early stages of an athlete's career rank is the most important determinant of popularity (Figure 4(F)). After the first 100 tournaments, however, the influence of rank on popularity is less pronounced and average popularity fluctuates in the vicinity of its top value.

3 Conclusions

While we would like to believe that fame, visibility and popularity are uniquely determined by performance, representing well-deserved recognition for some sustained or singular achievement, a significant body of media research indicates otherwise, suggesting that fame follows patterns on its own, divorced from talent or performance [8–22]. Here we aimed to quantify the relationship between performance and popularity in an area where

these two quantities can be individually measured. We did so by constructing a model to predict a tennis player's visibility captured by his Wikipedia page-views, a proxy of the athlete's popularity and fame. Taken together, we find that in tennis a player's popularity and momentary visibility are uniquely determined by his performance on the court. The agreement is especially good for elite players. This indicates that for athletes exceptional performance offers exceptional visibility, a level that is hard to modulate by exogenous events. For less accomplished players we observe deviations from the performance-predicted popularity, suggesting that in this case visibility can be manipulated by exogenous events or personal attributes outside of the scope of sports performance. It is comforting, however, that for most outliers the extra visibility can be explained by performance factors not considered in our model, like achievements in doubles or junior tournaments. In short, the better the performance of an athlete, the more accurately it determines his popularity, and the lesser the role of exogenous factors. Finally, we find that the fame of retired players is also determined by their past performance, indicating that exceptional past performance can lead to a commensurate lasting legacy.

We expect that our methodology can be readily generalized to other areas where performance and visibility can be independently measured. The generalization to sports like chess, table tennis, golf or car racing should be straightforward as the ranking systems of these sports are similar to tennis. With careful adjustments it may also be appropriate for team sports, allowing us to systematically explore how the performance of a professional athlete is tied to his/her or the team's collective success.

We would like to believe that in most areas of human achievement fame and visibility are determined by some underlying performance indicators. The scientific challenge, however, is to systematically separate performance from success, a fundamental goal from science to management. The excellent agreement between the performance-based and the observed popularity documented here makes us wonder to what degree fame is real, suggesting that outstanding fame and popularity may be rooted in performance measures that are perhaps not yet accessible to us.

Additional material

Additional file 1: Supplementary materials. (pdf)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

BY and A-LB designed the research. BY analyzed the data and prepared the figures. BY and A-LB prepared the manuscript.

Author details

¹Center for Complex Network Research and Department of Physics, Northeastern University, Boston, USA. ²Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, USA. ³Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA. ⁴Center for Network Science, Central European University, Budapest, Hungary.

Acknowledgements

We wish to thank Tamás Hámori for his guidance toward a deeper understanding of the tennis world, Kim Albrecht for helpful visualizations and other colleagues at the CCNR, especially those in the success group, for valuable discussions and comments. This research was supported by Air Force Office of Scientific Research (AFOSR) under agreements FA9550-15-1-0077 and FA9550-15-1-0364.

Received: 18 January 2016 Accepted: 20 April 2016 Published online: 29 April 2016

References

1. Lehmann S, Jackson AD, Lautrup BE (2006) Measures for measures. *Nature* 444(7122):1003-1004
2. Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: toward an objective measure of scientific impact. *Proc Natl Acad Sci USA* 105(45):17268-17272
3. Grubb HJ (1998) Models for comparing athletic performances. *J R Stat Soc, Ser D, Stat* 47(3):509-521
4. Radicchi F (2011) Who is the best player ever? A complex network analysis of the history of professional tennis. *PLoS ONE* 6(2):e17249
5. Radicchi F (2012) Universality, limits and predictability of gold-medal performances at the Olympic Games. *PLoS ONE* 7(7):e40335
6. Uzzi B (2008) A social network's changing statistical properties and the quality of human innovation. *J Phys A, Math Theor* 41(22):224023
7. Ke Q, Ferrara E, Radicchi F, Flammini A (2015) Defining and identifying Sleeping Beauties in science. *Proc Natl Acad Sci USA* 112(24):7426-7431
8. Boorstin DJ (2012) *The image: a guide to pseudo-events in America*. Vintage Books, New York
9. Van de Rijt A, Shor E, Ward C, Skiena S (2013) Only 15 minutes? The social stratification of fame in printed media. *Am Sociol Rev* 78(2):266-289
10. Gabler N (2001) *Toward a new definition of celebrity*. The Norman Lear Center, USC Annenberg
11. Andrews DL, Jackson SJ (2001) *Sport stars: the cultural politics of sporting celebrity*. Routledge, London
12. Argetsinger A (2013) *Famesque: Amy Argetsinger on celebrities famous for being famous*. Washington Post
13. Whannel G (2002) *Media sport stars: masculinities and moralities*. Routledge, London
14. Turner G (2004) *Understanding celebrity*. Sage, London
15. Evans J, Hesmondhalgh D (2005) *Understanding media: inside celebrity*. Open University Press, Milton Keynes
16. Ivaldi A, O'Neill SA (2008) Adolescents' musical role models: whom do they admire and why? *Psychol Music* 36(4):395-415
17. Hall N (2014) The Kardashian index: a measure of discrepant social media profile for scientists. *Genome Biol* 15(7):424
18. McClain AS (2013) The Kardashian phenomenon: news interpretation. *Media Report to Women* 41(2)
19. Female athletes only famous for their looks. <http://www.rantsports.com/clubhouse/2014/04/04/15-hot-female-athletes-who-are-only-famous-for-their-looks/>. Accessed 9 Dec 2014
20. Celebrities who are famous for doing nothing. <http://www.ajc.com/gallery/entertainment/celebrities-who-are-famous-doing-nothing/gCDry/>. Accessed 9 Dec 2014
21. Vincent J, Hill JS, Lee JW (2009) The multiple brand personalities of David Beckham: a case study of the Beckham brand. *Sport Mark Q* 18(3):173-180
22. Davie WR, King CR, Leonard DJ (2010) A media look at Tiger Woods - two views. *J Sports Media* 5(2):107-116
23. Murray C (2003) *Human accomplishment: the pursuit of excellence in the arts and sciences, 800 BC to 1950*. Harper Collins, New York
24. Rosen S (1981) The economics of superstars. *Am Econ Rev* 71(5):845-858
25. Hamlen WA Jr (1991) Superstardom in popular music: empirical evidence. *Rev Econ Stat* 73(4):729-733
26. Bagrow JP, Rozenfeld HD, Bollt EM, ben-Avraham D (2004) How famous is a scientist? - Famous to those who know us. *Europhys Lett* 67(4):511
27. Bagrow JP, ben-Avraham D (2005) On the Google-fame of scientists and other populations. In: *Modeling cooperative behavior in the social sciences*. AIP conference proceedings, vol 779, pp 81-89
28. Adler M (2006) Stardom and talent. In: Ginsburg VA, Throsby D (eds) *Handbook of the economics of art and culture*, vol 1. Elsevier, Amsterdam, pp 895-906
29. Simkin M, Roychowdhury V (2011) Von Richthofen, Einstein and the AGA. *Significance* 8(1):22-26
30. Simkin MV, Roychowdhury VP (2013) A mathematical theory of fame. *J Stat Phys* 151(1-2):319-328
31. Simkin MV, Roychowdhury VP (2015) Chess players' fame versus their merit. *Appl Econ Lett* 22(18):1499-1504
32. Sire C, Redner S (2009) Understanding baseball team standings and streaks. *Eur Phys J B* 67(3):473-481
33. Merritt S, Clauset A (2014) Scoring dynamics across professional team sports: tempo, balance and predictability. *EPJ Data Sci* 3(1):1
34. Clauset A, Kogan M, Redner S (2015) Safe leads and lead changes in competitive team sports. *Phys Rev E* 91:062815
35. Lago-Ballesteros J, Lago-Peñas C (2010) Performance in team sports: identifying the keys to success in soccer. *J Human Kinet* 25:85-91
36. 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 USA* 108(1):18-23
37. Motegi S, Masuda N (2012) A network-based dynamical ranking system for competitive sports. *Sci Rep* 2:904
38. Vaz de Melo POS, Almeida VAF, Loureiro AAF, Faloutsos C (2012) Forecasting in the NBA and other team sports: network effects in action. *ACM Trans Knowl Discov Data* 6(3):13
39. Herm S, Callsen-Bracker H-M, Kreis H (2014) When the crowd evaluates soccer players' market values: accuracy and evaluation attributes of an online community. *Sport Manag Rev* 17(4):484-492
40. Yu AZ, Ronen S, Hu K, Lu T, Hidalgo CA (2016) Pantheon 1.0, a manually verified dataset of globally famous biographies. *Sci Data* 3:150075
41. The world's 100 highest-paid athletes 2013. <http://www.forbes.com/pictures/mli45mmlg/the-worlds-100-highest-paid-athletes-4/>. Accessed 1 Dec 2014
42. Spoerri A (2007) What is popular on Wikipedia and why? *First Monday* 12(4)
43. Mestyán M, Yasseri T, Kertész J (2013) Early prediction of movie box office success based on Wikipedia activity big data. *PLoS ONE* 8(8):e71226
44. Keegan B, Gergle D, Contractor N (2013) Hot off the Wiki: Structures and dynamics of Wikipedia's coverage of breaking news events. *Am Behav Sci* 57(5):595-622
45. Cohen J, Cohen P, West SG, Aiken LS (2003) *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Mahwah