◎ EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

CrossMark

Generic temporal features of performance rankings in sports and games

José A Morales[1], Sergio Sánchez[1], Jorge Flores[2], Carlos Pineda[2,4*] iD, Carlos Gershenson[3,4,5,6,7],
Germinal Cocho[2], Jerónimo Zizumbo[1], Rosalío F Rodríguez[2] and Gerardo Iñiguez[3,8,9]

*Correspondence:
carlospgmat03@gmail.com
[2]Instituto de Física, Universidad
Nacional Autónoma de México,
México D.F., 01000, Mexico
[4]Centro de Ciencias de la
Complejidad, Universidad Nacional
Autónoma de México, México D.F.,
04510, Mexico
Full list of author information is
available at the end of the article

Abstract

Many complex phenomena, from trait selection in biological systems to hierarchy formation in social and economic entities, show signs of competition and heterogeneous performance in the temporal evolution of their components, which may eventually lead to stratified structures such as the worldwide wealth distribution. However, it is still unclear whether the road to hierarchical complexity is determined by the particularities of each phenomena, or if there are generic mechanisms of stratification common to many systems. Human sports and games, with their (varied but simple) rules of competition and measures of performance, serve as an ideal test-bed to look for universal features of hierarchy formation. With this goal in mind, we analyse here the behaviour of performance rankings over time of players and teams for several sports and games, and find statistical regularities in the dynamics of ranks. Specifically the rank diversity, a measure of the number of elements occupying a given rank over a length of time, has the same functional form in sports and games as in languages, another system where competition is determined by the use or disuse of grammatical structures. We use a Gaussian random walk model to reproduce the rank diversity of the studied sports and games. We also discuss the relation between rank diversity and the cumulative rank distribution. Our results support the notion that hierarchical phenomena may be driven by the same underlying mechanisms of rank formation, regardless of the nature of their components. Moreover, such regularities can in principle be used to predict lifetimes of rank occupancy, thus increasing our ability to forecast stratification in the presence of competition.

Keywords: complex systems; sports; data analysis; rank distribution; rank diversity

1 Introduction

Sports and games can be described as hierarchical complex systems due to the myriad of factors influencing the dynamics of competition and performance in them, including networked interactions, human and environmental heterogeneities, and other traits at the individual and group levels [1–4]. In particular, the performance of players and teams is influenced by a variety of causes: Economical, political and geographical conditions determine their rankings and may thus be used for predicting performance. Moreover, the (relatively) simple rules of competition and measures of performance associated with sports and games allow us to explore basic mechanisms of interaction leading to hierarchy for-

Springer

mation, which may be common to many systems driven by competition, not only leisure activities but other social, biological and economic systems. With this goal in mind, the availability of a large corpus of data related to sports, teams, and players allows researchers to perform multiple statistical analyses, in particular with respect to the structure and dynamics of performance rankings [5–7].

Data availability has made it possible not only to study the distribution of scores determining rankings, but also its time evolution [8]. In a recent paper, Deng *et al.* present a statistical analysis of 12 sports and report a universal scaling in rankings, despite the fact that the sports considered have very different ranking systems [9]. Here, we focus on the temporal trajectories of player and team performances, meaning the evolution of rank, with the objective of finding statistical regularities that indicate how competition shapes hierarchies of players and teams. In principle, rankings may be affected in time by events as apparently insignificant as a bad breakfast prior to an important event, or the weather during a competition [10]. Since these factors are inherently present for all activities, we would expect the evolution of rank to have generic features across sports and games.

We propose to quantify such evolution by means of a recently introduced measure, the *rank diversity*. With the help of the Google *n*-gram dataset [11], rank diversity has been used before to study how vocabulary changes in time [12]. That work shows that rank diversity has the same functional form for all languages studied, and is able to discriminate the size of the core of each language. Thus, here we concentrate on the temporal features of rank distributions corresponding to several sports and games with different ranking schemes. We consider data where an appropriate time resolution is available, and limit the analysis to six activities only: tennis, chess, golf, poker and football (both national teams and clubs). We find that all rank diversities have the same functional form as languages, despite having differences in their rank frequency distributions. Finally, we introduce a random walk model that, tuned by the parameter values of each dataset, reproduces qualitatively the diversity of all sports and games considered. Overall, our goal is to use rank diversity as a tool to understand rank dynamics in sports, games, and other hierarchical complex systems, thus enabling us to identify the dependence on rank of a change in the hierarchy of the system. By using this analysis, we may be able to estimate how well can a change in rank be predicted, regardless of the particularities of the phenomenon under study.

The article is organized as follows. In Section 2 we describe the datasets used. We then analyse ranking distributions in Section 3 and compare them with several models. In Section 4 we study the rank diversity for each sporting activity and compare it with a random walk model. The main conclusions of our analysis are included in Section 5. In Appendix A we discuss in detail the Kolmogorov-Smirnov index, which measures the goodness of fit for a given dataset. Finally, in Appendix B we describe the generic relation between rank diversity and the cumulative rank distribution in the random walk model.

2 Ranking data

We use ranking data on players and teams from six sports and games: (a) Tennis players (male), ranked by the Association of Tennis Professionals (ATP) [13]; (b) Chess players (male), ranked by the Fédération Internationale des Échecs (FIDE) [14]; (c) Golf players, ranked by the Official World Golf Ranking (OWGR) [15]; (d) Poker players, ranked by the Global Poker Index (GPI) [16]; (e) Football teams, ranked by the Football Club World

Table 1 Summary of ranking data for each sport and game considered in this study

Sport/game	Data source	Time period	Ranking resolution	#players/teams
Tennis players (male)	Association of Tennis Professionals (ATP) [13]	May 5 2003-Dec 27 2010	Weekly	1,600
Chess players (male)	Fédération Internationale des Échecs (FIDE) [14]	Jul 2012-Apr 2016	Monthly	13,500
Golf players	Official World Golf Ranking (OWGR) [15]	Sept 10 2000-Apr 19 2015	Weekly	1,000
Poker players	Global Poker Index (GPI) [16]	Jul 25 2012-Jun 10 2015	Weekly	1,799
Football teams	Football Club World Ranking (FCWR) [17]	Feb 1 2012-Dec 29 2014	Weekly	850
National football teams	Fédération Internationale de Football Association (FIFA) [18]	Jul 2010-Dec 2015	Monthly	150

Table listing the main properties of the ranking data used here (including data source, time period, ranking resolution, and number of players/teams). In order to have a homogeneous distribution of ranking snapshots and the same number of players/teams in each snapshot for a given activity, we disregard some data for the ATP, FIDE, OWGR, GPI, and FIFA datasets, as explained in the main text.

Ranking (FCWR) [17]; and (f) national football teams, ranked by the Fédération Internationale de Football Association (FIFA) [18].

The ranking procedure varies among sports. In ATP, for example, tennis players are ordered according to the number of points they have up to the date of publication of the ranking. The number of points depends on the tournaments players have participated in (and how well they have performed), but not all tournaments are taken into account. FIDE uses the Elo system [19] to rank players, which considers the number of matches, their results, and the opponent ranking. The FIFA ranking takes into account official matches between countries. The number of points depends on the confederation and classification of each team, as well as the importance and result of the match. Table 1 summarises the main properties of the ranking data considered in this study, including the time resolution used to measure rankings (*i.e.* the time difference between two snapshots of the ranking in a sport or game). In order to have a homogeneous distribution of ranking snapshots and the same number of players/teams in each snapshot for a given activity, we disregard some data for the ATP, FIDE, OWGR, GPI, and FIFA datasets: In all of these cases, the time elapsed between the publication of two rankings varies greatly (from less than a week to more than a month), and the number of players/teams across ranking snapshots may change as well. Therefore, for each dataset we choose a constant time resolution of rankings (weeks or months, as shown in Table 1) that maximises and keeps constant the number of ranked players/teams throughout time. All datasets, filtered as explained above, are included in Additional file 1.

3 Comparison with ranking models

Player or team performance is usually measured by a *score* that varies with time. This score results in a time-dependent *rank* with a rather complex behaviour, as we will explain below. We first focus on the distribution of scores versus ranks (*i.e.* a *rank distribution*) for a given time. Particularly, we are interested in seeing if this distribution can be reproduced by a single ranking model for all sports and games considered. We select five ranking models to fit the data, four of which are particular cases of

$$f(k) = \mathcal{N} \frac{(N+1-k)^q \exp(-bk)}{k^a},$$ (1)

where f is the score associated with rank k, a is an exponent that dominates most of the curve, b an exponent controlling its exponential decay, and q an algebraic decay that regulates a sharp drop of the curve for large k. Finally, N is the total number of elements (*i.e.* players or teams) in the system, and $\mathcal{N}$ is a normalization constant.

The first four models are

$$m_1(k) \propto \frac{1}{k^a}, \qquad m_2(k) \propto \frac{\exp(-bk)}{k^a},$$
$$m_3(k) \propto \frac{(N+1-k)^q}{k^a}, \qquad m_4(k) = f(k), \tag{2}$$

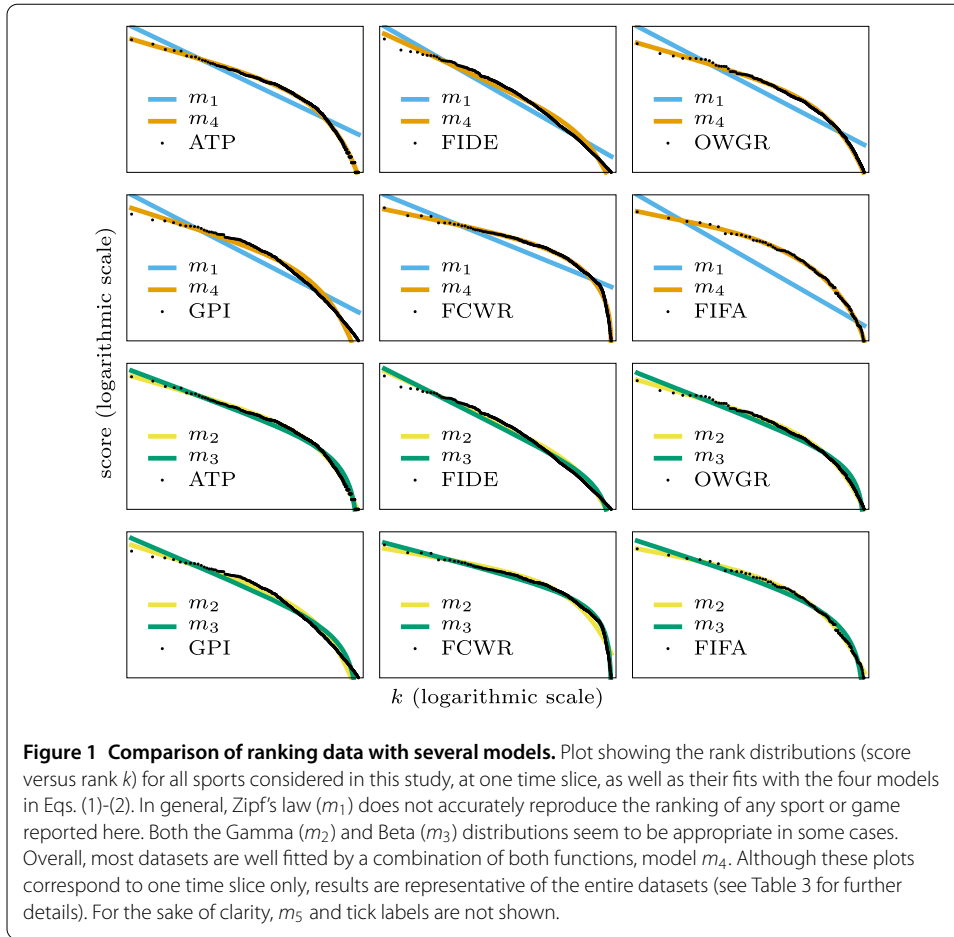whereas the fifth model is a double Zipf law [20],

$$m_5(k) = \mathcal{N} \begin{cases} \frac{1}{k^a}, & k \le k_c, \\ \frac{k_c^{a'-a}}{k^{a'}}, & k > k_c, \end{cases} \tag{3}$$

with a' an alternative exponent that regulates the behaviour of the curve after a critical rank k_c. Model m_1 is obtained by setting $q = b = 0$ in Eq. (1), and has been considered in a vast amount of studies, both in the realm of sports [21, 22] and in other studies of ranking behaviour [23], including the famous Zipf's law of languages where the particular case $a = 1$ has drawn a lot of attention (see, *e.g.*, [24] and references therein). The Gamma (m_2) and Beta (m_3) distributions have been useful in many disciplines for decades; a quick look at their Wikipedia entries provides numerous examples [25, 26]. Model m_4, being a more general expression than the previous ones, tends to provide a better fit at the expense of more parameters, and will serve as benchmark for the comparison between the rest of the models. Finally, model m_5 in Eq. (3) has been used with success in several contexts [20, 27], prompting us to test it in the area of sports and games.

The results of the fitting process between data and Eqs. (1)-(3) are shown in Figure 1, while Table 2 summarises the parameter values obtained. Data corresponds to a single time snapshot for all sports and games: Dec 27 2010 (ATP); Sept 2014 (FIDE); Mar 18 2015 (GPI); Apr 19 2015 (OWGR); Dec 29 2014 or Week 53 2014 (FCWR); and Dec 18 2014 (FIFA). The following results are, however, representative of all time snapshots (see Table 3 and the text below for further details). Both models and data show variation in their goodness of fit. From Figure 1 it is clear that Zipf's law (m_1) is not adequate. On the other hand, the Gamma distribution (m_2) fits some datasets rather well, particularly those that do not show an abrupt fall of score as a function of rank. Datasets with an abrupt decay of frequency are well fitted by the Beta distribution (m_3) instead. However, most sports and games seem to be an intermediate case where both functions capture global behaviour accurately, and thus the fit is considerably better for a combination of both models, *i.e.* m_4. We also see that the double Zipf law (m_5) is a good fit for FIDE and GPI, as seen from Table 3.

In order to objectively compare goodness of fit between models, we consider several measures: The coefficient of determination R^2 [28], the maximum deviation between theory and observation D, and the Kolmogorov-Smirnov index p [28, 29]. The coefficient R^2 is calculated from the 2-norm with respect to the data coming from a single time snapshot,

$$R^2 = 1 - \frac{\sum_k [m_i(k) - y_k]^2}{\sum_k [y_k - \langle y \rangle]^2}, \tag{4}$$

Figure 1 Comparison of ranking data with several models. Plot showing the rank distributions (score versus rank k) for all sports considered in this study, at one time slice, as well as their fits with the four models in Eqs. (1)-(2). In general, Zipf's law (m_1) does not accurately reproduce the ranking of any sport or game reported here. Both the Gamma (m_2) and Beta (m_3) distributions seem to be appropriate in some cases. Overall, most datasets are well fitted by a combination of both functions, model m_4. Although these plots correspond to one time slice only, results are representative of the entire datasets (see Table 3 for further details). For the sake of clarity, m_5 and tick labels are not shown.

for a given model $m_i(k)$, $i = 1, \ldots, 5$, and data y_k, where $\langle y \rangle$ is the expectation value of y_k. The closer R^2 is to one, the better the fit is. To calculate D, we consider the cumulative of both the proposed distribution $m_i(k)$ and a dataset with N data points (the equivalence between an empirical rank-value distribution and the empirical cumulative distribution corresponding to scores is discussed in [28]). There it is also shown that for a given theoretical rank distribution $m_i(k)$, the cumulative is simply $M_i(m_i) = [N + 1 - k(m_i)]/[N + 1]$, where $k = k(m_i)$ is the inverse function of m_i that implicitly depends on scores. Whereas for a dataset, $M_{\text{data}}(s) = (1/N) \sum_j \theta(s - s_j)$, with θ a step function and $\{s_i, \ldots, s_N\}$ the set of scores in the data [28]. We then define D (the so-called Kolmogorov statistics) as the maximum vertical difference between the two curves, $D = \sup_s |M_i(m_i) - M_{\text{data}}(s)|$. The calculation of the Kolmogorov-Smirnov index p is more involved so we discuss it in Appendix A. The measure p allows us to consider that a small dataset will have some noise due to poor statistics. Thus, if a model is consistent with a dataset, but we have poor statistics, we might still have a good (large) p. Usually, a 'good' fit is required to have $p > 0.1$, see *e.g.* [30].

Table 3 shows the mean values $\langle R^2 \rangle$ and $\langle D \rangle$ (and their associated standard deviations σ_D and σ_{R^2}), averaged over all time slices available, for the fitting process between the six datasets and five models m_i used here. We also include values of p for the single time slice of Figure 1. Higher $\langle R^2 \rangle$ and lower $\langle D \rangle$ imply better fits. Since σ_D and σ_{R^2} are small, the fits shown in Figure 1 are representative of the entire datasets. We observe that none

Table 2 Parameter values for fitting process between sports data and ranking models

	Model m_1		Model m_2			Model m_3			Model m_4				Model m_5			
	$\log \mathcal{N}$	a	$\log \mathcal{N}$	a	b	$\log \mathcal{N}$	a	q	$\log \mathcal{N}$	a	b	q	$\log \mathcal{N}$	a	a'	k_c
ATP	4.51	1.04	4.11	0.626	3.18×10^{-3}	-1.46	0.816	1.79	4.01	0.628	3.13×10^{-3}	3.12×10^{-2}	4.2	0.746	3.004	3.19×10^{2}
FIDE	3.46	0.0252	3.46	2.01×10^{-2}	6.25×10^{-6}	3.32	2.21×10^{-2}	3.28×10^{-2}	3.46	0.0202	6.25×10^{-6}	5.05×10^{-9}	3.45	0.016	0.036	1.97×10^{2}
OWGR	1.35	0.702	1.05	0.383	2.68×10^{-3}	-0.928	0.53	0.703	1.023	0.385	2.65×10^{-3}	1.18×10^{-2}	1.09	0.452	1.73	2.03×10^{2}
GPI	3.75	0.234	3.66	0.144	6.63×10^{-4}	2.54	0.193	0.358	3.66	0.144	6.63×10^{-4}	3.64×10^{-9}	3.65	0.133	0.437	1.08×10^{2}
FCWR	4.52	0.529	4.24	0.218	3.06×10^{-3}	2.19	0.341	0.732	2.93	0.269	1.39×10^{-3}	0.458	4.35	0.371	3.47	4.26×10^{2}
FIFA	3.37	0.473	3.24	0.142	1.02×10^{-2}	2.30	0.262	0.456	3.15	0.148	9.58×10^{-3}	4.16×10^{-2}	3.26	0.227	0.8622	33.86

Table listing parameter values for all models in Eqs. (1)-(3), obtained in the fitting process with empirical data. These values correspond to the model curves in Figure 1 (model m_5 not shown there), but are representative of the entire datasets (see Table 3 for further details).
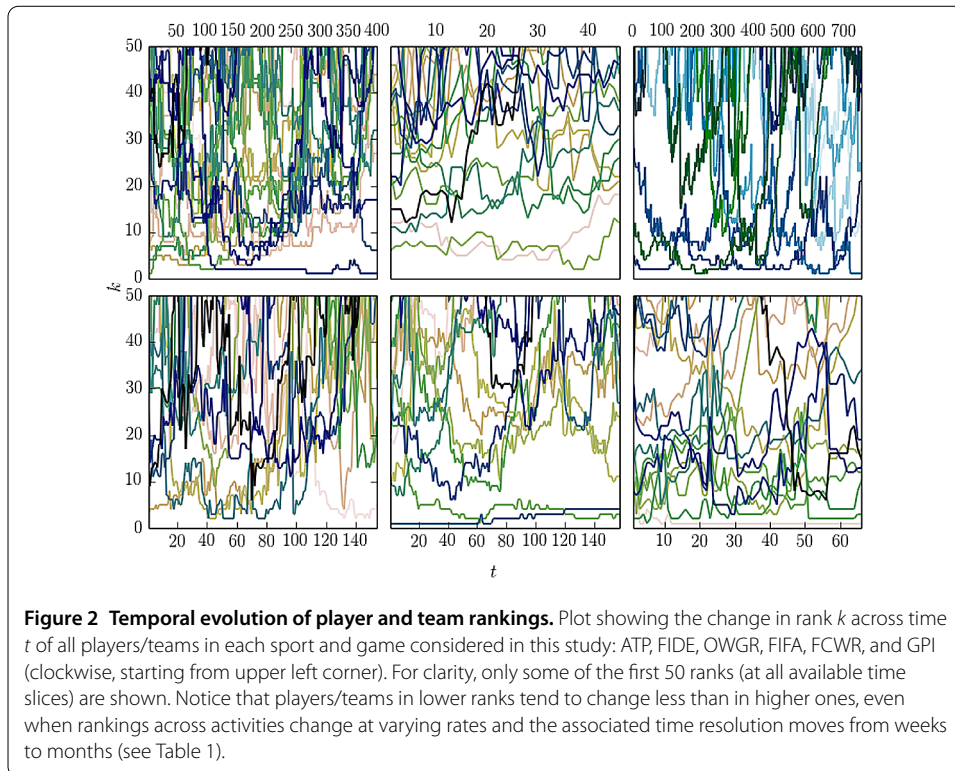
Table 3 Goodness of fit measures

		m_1	m_2	m_3	m_4	m_5
ATP	$\langle R^2 \rangle$	0.222	**0.982**	0.879	**0.982**	0.964
	$\langle D \rangle$	0.433	0.044	0.08	**0.038**	0.077
	σ_{R2}	0.0969	0.01652	0.009	0.0124	0.0288
	σ_D	0.211	0.0126	0.0672	0.0128	0.0287
	p	0.01	**0.17**	0.0	0.12	0.0
FIDE	$\langle R^2 \rangle$	0.777	0.936	0.657	0.936	**0.991**
	$\langle D \rangle$	0.477	0.2	0.188	0.2	**0.141**
	σ_{R2}	0.0071	0.0053	0.0028	0.0054	0.0035
	σ_D	0.0072	0.0048	0.0166	0.0048	0.0005
	p	0.0	0.0	0.0	0.0	0.0
OWGR	$\langle R^2 \rangle$	0.631	0.981	0.943	**0.982**	0.97
	$\langle D \rangle$	0.316	0.046	0.088	**0.043**	0.088
	σ_{R2}	0.0264	0.0388	0.0138	0.0381	0.0391
	σ_D	0.1292	0.0165	0.0192	0.0152	0.0104
	p	0.0	**0.92**	0.0	0.89	0.0
GPI	$\langle R^2 \rangle$	0.791	0.978	0.937	0.978	**0.985**
	$\langle D \rangle$	0.531	0.201	**0.149**	0.201	0.202
	σ_{R2}	0.01029	0.0115	0.0044	0.0115	0.0459
	σ_D	0.01612	0.0039	0.0048	0.0039	0.00533
	p	0.0	0.0	0.0	0.0	0.0
FCWR	$\langle R^2 \rangle$	0.727	0.986	0.981	**0.997**	0.947
	$\langle D \rangle$	0.295	0.115	0.057	**0.055**	0.172
	σ_{R2}	0.0186	0.0183	0.0098	0.0112	0.0268
	σ_D	0.02833	0.0046	0.0052	0.00128	0.0104
	p	0.0	0.0	0.0	0.0	0.0
FIFA	$\langle R^2 \rangle$	0.833	0.993	0.981	**0.996**	0.979
	$\langle D \rangle$	0.387	0.076	0.071	**0.041**	0.155
	σ_{R2}	0.0277	0.0324	0.0135	0.0114	0.0413
	σ_D	0.02888	0.004	0.007	0.002	0.0147
	p	0.0	**0.99**	0.0	**0.99**	0.02

Table listing mean values $\langle R^2 \rangle$ and $\langle D \rangle$ (and their associated standard deviations σ_D and σ_{R2}), averaged over all time slices available, for the fitting process between the six sports and five theoretical rank distributions used here. We also include values of the Kolmogorov-Smirnov index p for the single time slice of Figure 1. Higher $\langle R^2 \rangle$ and lower $\langle D \rangle$ imply better fits. Since σ_D and σ_{R2} are small, the fits shown in Figure 1 are representative of the entire datasets. The best fits for each sport are shown in bold.

of the models are a good fit for all sports and games, although m_4 and m_5 are the most appropriate (in terms of R^2). However, in three cases (FIDE, GPI, and FCWR) we have $p = 0$ for model m_4, and no model fits well, meaning that the theoretical distribution is not followed by the data. We stress again that Zipf's law (m_1) is the worst fit among all considered, except for FIDE. It is interesting to notice that R^2 and p lead to different criteria of what is a 'good' or a 'bad' model. This is due to both the amount of available data and the number of parameters in the model. The larger the data, the easier it is to distinguish the best available model from a good (but not accurate enough) approximation. On the other hand, the more parameters the model has, the easier it is to fit any data. Both of these aspects are taken into account in the definition of p, but not in R^2.

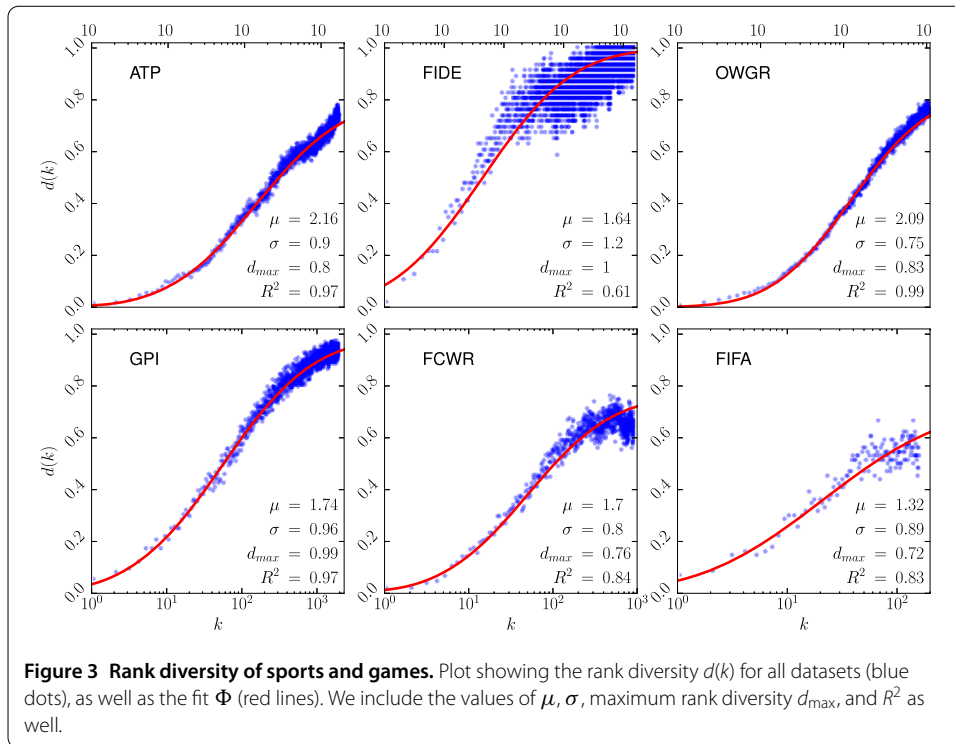## 4 Rank diversity in sports and games

The previous analysis of the functional form of the rank distribution in several sporting activities (even when the goodness of fit has been averaged over time) is restricted by the fact that the rank distribution is inherently an *instantaneous* measure, in the sense that it captures ranking at a given point in time and does not take into account the dynamics of

Figure 2 Temporal evolution of player and team rankings. Plot showing the change in rank k across time t of all players/teams in each sport and game considered in this study: ATP, FIDE, OWGR, FIFA, FCWR, and GPI (clockwise, starting from upper left corner). For clarity, only some of the first 50 ranks (at all available time slices) are shown. Notice that players/teams in lower ranks tend to change less than in higher ones, even when rankings across activities change at varying rates and the associated time resolution moves from weeks to months (see Table 1).

players and teams changing rank as time goes by. In order to overcome this issue, here we contribute to the analysis of ranking in sports and games by computing the rank diversity, a measure of the number of elements occupying a given rank over a length of time. From previous [12] and current work, it appears that rank diversity has the same functional form, not only for sports but also for other complex systems, such as countries classified by their economic complexity, the 500 leading enterprises ranked by *Fortune* magazine, or a set of millions of words in six Indo-European languages.
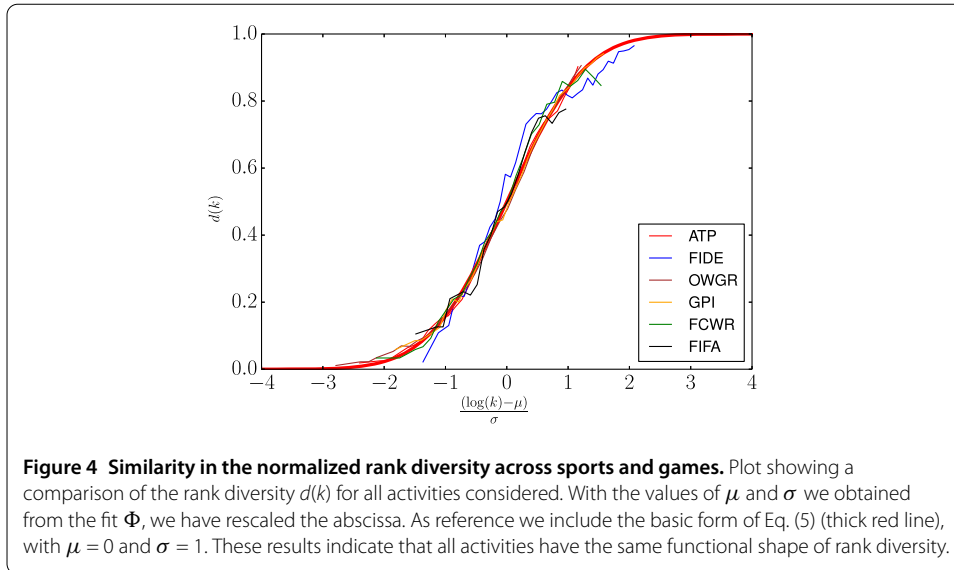
The rank diversity $d(k)$ is defined as the number of distinct elements in a complex system that occupy the rank k at some point during a given length of time. In other words, we choose to focus on the time dependence of ranks, rather than on the static (*i.e.* defined for a single time) rank distribution $f(k)$. An example of the change of ranks in time for the sports and games studied here can be seen in Figure 2. These so-called 'spaghetti' curves show how elements - individuals or teams - change their rank in time. The rank diversity $d(k)$ is simply the normalized number of different elements (curves) that spend at least one time interval at a given rank k. The rank diversity for the various sports and games considered here is shown in Figure 3.

We should stress that $d(k)$ and $f(k)$ measure different aspects of the hierarchical structure of a complex system. First of all, the rank diversity includes information on how elements change rank throughout time in a single function, while the rank distribution captures the hierarchy in the system for a single time interval. Secondly, the rank diversity disregards any information on the scores of elements beyond their order, and thus the same $d(k)$ may be obtained for several shapes of $f(k)$ (power-law, Gamma, Beta, etc.). As an example, consider any transformation in time of the scores of elements in the system, such that their ranking order stays the same; then $f(k)$ could interpolate between differ-

Figure 3 Rank diversity of sports and games. Plot showing the rank diversity $d(k)$ for all datasets (blue dots), as well as the fit Φ (red lines). We include the values of μ, σ, maximum rank diversity d_{max}, and R^2 as well.

ent functional shapes as time goes on, while $d(k)$ would stay constant. The inverse case is also possible, and any rank distribution may produce a wide variety of rank diversities. For example, we could construct several dynamics of scores that keep the number of elements with a given score constant, but that change the amount of time an element holds certain score, thus keeping $f(k)$ fixed and changing $d(k)$. Overall, both $d(k)$ and $f(k)$ measure some aspects of the structure and dynamics of hierarchy in a complex system, but only the rank diversity captures the way elements change their positions in the hierarchy, beyond minor changes in scores that could be attributed, for example, to different ways of measuring performance.

From Figure 3 we see that the empirical curves for rank diversity are (roughly) monotonic and have a single shoulder. The cumulative of a square-integrable function with a single bump would have these properties, and a Gaussian is arguably the simplest choice. Moreover, an analytical argument (see Appendix B) suggests that this may be an appropriate ansatz under very general conditions, at least qualitatively. In a large variety of physical systems composed of alike elements with similar interactions between them, the macroscopic response of the system is usually determined by general laws such as equations of state. However, in different empirical realisations of the same dynamics there may be differences associated to the law of the large numbers or the central limit theorem. These differences across realisations follow a normal Gaussian distribution, according to the Gaussian theory of errors. However, for complex systems with competitive dynamics, there may be generic features described by the Gamma (m_2) and Beta (m_3) distributions [31, 32], and there may also be differences across realisations that follow a multiplicative dynamics. This is indeed the case for several Indo-European languages [12] and for the games and sports datasets considered here (See Figure 1). In Appendix B we introduce the non-trivial idea that there are two different dynamics associated with so-called generic and contingent fea-

Figure 4 Similarity in the normalized rank diversity across sports and games. Plot showing a comparison of the rank diversity $d(k)$ for all activities considered. With the values of μ and σ we obtained from the fit Φ, we have rescaled the abscissa. As reference we include the basic form of Eq. (5) (thick red line), with $\mu = 0$ and $\sigma = 1$. These results indicate that all activities have the same functional shape of rank diversity.

tures, which may be described in terms of a one-step Markovian, Gaussian process. This allows us to establish an explicit relation between the diversity $d(k)$ and the cumulative of the rank distribution, $S(t)$.

In fact, studying $d(k)$ for six Indo-European languages [12], we found that the observed rank diversity closely follows the cumulative of a Gaussian (*i.e.* a sigmoid)

$$\Phi_{\mu,\sigma}(\log k) = \frac{\max_i d(k_i)}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log k} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy. \tag{5}$$

The mean value μ is set as the smallest k_0 for which $d(k_0) = \frac{\max_i d(k_i)}{2}$, while the width σ is fitted and gives the scale for which $d(k)$ gets close to its extreme values. If $k_\pm$ are given by $\log_{10} k_\pm = \mu \pm 2\sigma$, the bulk of the changes in the values of diversity lies between k_- and k_+. In Figure 3 we show the fit Φ for all sports and games considered here (R^2 values for the Φ curves are shown there as well). We do not consider neither D nor p, since these measures are only meaningful for distributions, which $d(k)$ is not. To compare different rank diversity curves, their rank can be normalised to $\frac{\log(k)-\mu}{\sigma}$, as shown in Figure 4. Since all the cases considered can be fitted with the sigmoid curve of Eq. (5), we argue that the rank diversity of sports seems to have a generic shape.

4.1 A random walk model

From Figure 2 and Figure 4 we see that players and teams with low ranks change very slowly or not at all, while those with higher k have a larger rank variation in time. This intuition is clear from recent experience in sports like tennis and football: According to the analysed datasets, Hewitt, Nadal, Roddick, Ferrero, Agassi and Federer have been the only number one tennis players from May 2003 till December 2010. The same holds for football clubs: Real Madrid, Atlético Madrid, Barcelona, and Bayern München have been the best-ranked teams from January 2012 till December 2014. In other words, players and teams with small k tend to have a small rank diversity.

In what follows we propose a simple model [12] that captures such intuition (*i.e.* a variation approximately proportional to the current rank), and whose rank diversity resembles

Figure 5 Comparison between empirical and simulated rank diversities. Plot showing the rank diversity $d(k)$ of the empirical data (blue shaded area) and of simulated data from our random walk model (yellow shaded area), as well as the fit Φ (red/green lines for empirical/simulated data, respectively). We also include the values of μ, σ, and maximum rank diversity d_{max} for the simulated data. The random walk model manages to qualitatively reproduce the rank diversity seen in all sports and games considered here, although clear differences reveal that the model is insufficient for a full quantitative agreement.

the data presented here. We call this model a scale-invariant random Gaussian walk, since a member with rank k_t, at the discrete time t, is converted to rank k_{t+1} according to the following procedure: We define an auxiliary variable l_{t+1}, which we call pre-rank, at time $t + 1$ by the relation

$$l_{t+1} = k_t + G(k_t\hat{\sigma}), \tag{6}$$

where $G(k_t\hat{\sigma})$ is a Gaussian-distributed random number with standard deviation $k_t\hat{\sigma}$ and mean 0. This means that the random variable l_{t+1} has a width distribution proportional to k_t, and thus will, for small k_t, have small changes as well. Once the values of the pre-ranks l_{t+1} for all members are obtained, we order them according to their magnitude. This new order gives new rankings, *i.e.* the k values at time $t + 1$. The only parameter left in the model is the relative width $\hat{\sigma}$, which we fit by using a least-squares method over a smoothed version of the empirical rank diversity. In Figure 5 we show the rank diversity for systems with the same number of elements as those of Figure 3, but generated with the random model. We see that these two sets of plots are qualitative similar, although clear differences reveal that the model is insufficient for a full quantitative agreement. The fact that both the empirical and simulated rank diversities have a sigmoid shape suggests that rank changes in real systems may be the result of a large number of multiplicative processes. We discuss some analytical ideas supporting this insight in Appendix B. However, the mismatch between model and data seen in Figure 5 shows that not all characterizing features of the empirical process are captured by our model, and further investigation is needed.

5 Discussion and conclusions

Competition and heterogeneous performance are characteristic of the elements of many complex systems in biological, social and economic settings. Despite the fact that these systems show a large variation in the definitions of their constituents and in the relevant interactions between them, it remains to be seen whether the emergence of hierarchical structure is mostly determined by the particularities of each phenomenon, or if there are mechanisms of stratification common to the temporal evolution of many systems. We have explored this notion by considering a set of relatively controlled and simplified systems driven by competition: Human sports and games, where the rules of engagement and measures of performance are well defined, in contrast to, say, the ranking of physicists (the question of whom is the 'best' physicist would have an ambiguous answer, to say the least). This allows us to characterise the emergence of hierarchical heterogeneity by comparing the temporal features of rankings of individuals and teams across activities in a clear way. Explicitly, we analysed the statistical properties of rank distributions in six sports and games, each with different number of members and rules for calculating scores (and, therefore, ranks). By comparing rank distributions with several ranking models, we find that the Zipf law (model m_1) does not provide a suitable fit for the empirical data. Even if the more generic ranking model m_4 (a combination of the Gamma and Beta distributions) tends to offer good fits, it is not always the best.

Furthermore, we studied the temporal features of rankings explicitly by calculating the rank diversity $d(k)$, a measure of the number of individuals or teams occupying a given rank over a length of time. We found that $d(k)$ has the same sigmoid-like functional form, even for relatively small systems like FIFA (with only 150 elements per time slice). Coupled to the fact that a sigmoid rank diversity has also been found in the way vocabulary changes in time [12], our results suggest that the emergence of hierarchical complexity - as measured by $d(k)$ - may have traits common to many systems. This claim is underlined by the fact that a simple model (the scale-invariant random Gaussian walk) can reproduce the diversity of the sports and games studied here, and also of languages [12]. One could initially suspect that rank changes depend on the intrinsic strength or qualities of players and teams. However, given the fact that our random walk model reproduces relatively well the rank dynamics of several sports and games, it seems that rank change can instead be characterised as a random process. This does not imply that rank change is random, but that the specific mechanisms associated with each activity and ranking system are irrelevant for the calculation of rank diversity.

A natural direction to follow in the near future is to study the behaviour of rank diversity in other competitive phenomena beyond sporting activities and language, such as physical, social and economic processes of stratification. If indeed a certain universality in the temporal features of rankings is present in other complex settings, it would indicate that hierarchical phenomena may be driven by the same underlying mechanisms of rank formation, regardless of the nature of their components. Potentially, we may exploit such regularities to predict lifetimes of rank occupancy, thus increasing our ability to forecast stratification in the presence of competition.

Appendix A: Explicit calculation of Kolmogorov-Smirnov p-value

The Kolmogorov-Smirnov p-value is a way to quantify the goodness of fit of some theoretical distribution to the empirical distribution of a dataset. For a given dataset $\{s_1, s_2, \ldots, s_N\}$,

the corresponding empirical distribution is defined as

$$M_{\text{data}}(s) = \frac{1}{N} \sum_j \theta(s - s_j), \tag{7}$$

where θ is a step function and the variable s represents scores. The goodness of fit is obtained via such empirical distribution and a theoretical cumulative distribution (CCD). Thus, we need to define the rank distribution in terms of a CCD in order to use this criterion. Ref. [28] shows that there is an equivalence between an empirical rank-value distribution and the empirical cumulative distribution of scores (or frequencies) available from the data. The formula that relates these two functions is

$$M_i(m_i) = \frac{N + 1 - k(m_i)}{N + 1}, \tag{8}$$

where m_i is the value of the theoretical rank distribution, and k the rank related to score s. So, to obtain the corresponding CCD of a rank distribution m_i, it is enough to apply Eq. (8). Note that $k = k(m_i)$, *i.e.* k is the inverse function of m_i. The p-value will then measure how good $M_i(m_i)$ fits the empirical distribution of scores. Indirectly, we are obtaining a measure of goodness of fit of m_i to the empirical rank-value distribution, due to the equivalence stated in Eq. (8). In our case, the theoretical m_i is given by Eq. (2) and Eq. (3).

Next we define the Kolmogorov statistic D as the maximum distance between the empirical distribution of scores and the theoretical cumulative distribution,

$$D = \sup_s \left| M_i(m_i) - M_{\text{data}}(s) \right|. \tag{9}$$

We stress that when we talk about m_i, value means a score in the system.

Finally, we describe the process used to calculate the p-value:

1. Compute the parameters of fit m_i for the empirical rank-value distribution (scores).
2. Obtain the empirical distribution of scores and the $M_i(s)$ with Eq. (7) and Eq. (8).
3. Calculate the Kolmogorov statistic D between M_i and M_{data}.
4. Generate (*e.g.* 2,500) artificial datasets of scores, distributed according to the fitted M_i. For each of them, fit to an artificial $M_{i,\text{art}}$ in order to obtain a value D_{art}.
5. Count how many of the 2,500 D_{art} values are larger than the D value of the real dataset and divide it by 2,500. The result is the p-value.

Appendix B: Diversity and cumulative distribution

In previous work we have shown that, under very general conditions in which dynamic competition exists between positive and negative mechanisms, like birth and death processes, the rank distribution is given by the ratio of two power laws [32]. In this Appendix we analyse the difference between the data associated with different realisations of such competitive dynamics and the adjustments to real data in terms of stochastic models such as $m_2(k)$, $m_3(k)$, and $m_4(k)$ given by Eq. (2). Specifically, we adopt the more general point of view that the data (obtained for Indo-European languages [12] and several sports and games) may be represented by a one-step Markovian stochastic process for the allocation of ranks.

The difference between the data associated with several realisations of the competitive dynamics and the adjustments to the real data may be analysed by treating k as a continuous variable. In this case, the time evolution of the probability density distribution of ranks $P(k, t)$ is described by a Fokker-Planck equation (FPE),

$$\frac{\partial}{\partial t} P(k, t) = -\frac{\partial}{\partial k}\big[A(k)P\big] + \frac{\partial^2}{\partial k^2}\big[B(k)P\big], \tag{10}$$

where $A(k)$ and $B(k)$ are rank-dependent drift and diffusion coefficients, respectively.

Note that in Figures 1, 3 and 4 the abscissa is not the rank k, but $x = \log k$. In other words, the systems exhibit a simpler behaviour in terms of the variable x, a fact that suggests a multiplicative behaviour and, in turn, a log-normal process. This process is the statistical realisation of the multiplicative product of many independent positive random variables, a feature that is justified by considering the central limit theorem in the logarithmic domain, and thus obeys the log-normal distribution. As a consequence, $P(k, t)$ can be expressed in the general form

$$P(x, t) = P^{\text{st}}(x) + P_1(x, t), \tag{11}$$

with $x = \log k$. The explicit form of the stationary distribution $P^{\text{st}}(x)$ is well known [33, 34], and the time dependent solution $P_1(x, t)$ may be determined as follows. We first note that Eq. (10) may be rewritten as

$$\frac{\partial}{\partial t} P(x, t) = \frac{\partial}{\partial x}\big[B(x)P_x\big] + \alpha P_x + \beta P, \tag{12}$$

where $\alpha = -A + B_x$, $\beta = -A_x + B_{xx}$, and each subscript $\bullet_x$ denotes a partial derivative with respect to x. This equation can be further simplified by introducing the variable $v(x, t) \equiv B(x)P(x, t)$. Moreover, in order to simplify the discussion and the resulting equations, we consider the particular case where the drift and diffusion coefficients $A(x)$ and $B(x)$ are proportional to the same function $g(x)$, i.e., $A(x) = \lambda_A g(x)$ and $B(x) = \lambda_B g(x)$. If $\tau \equiv B(x)t$, then Eq. (12) reduces to

$$\frac{\partial}{\partial \tau} v(x, \tau) = -\Lambda \frac{\partial v}{\partial x} + \frac{\partial^2 v}{\partial x^2}, \tag{13}$$

with $\Lambda \equiv \lambda_A/\lambda_B$. Let us now introduce the multiplicative character mentioned above by introducing $u(x, \tau)$ through the following change of variables,

$$\log \frac{v(x, \tau)}{u(x, \tau)} = \Lambda x - \frac{\Lambda^2}{4}\tau. \tag{14}$$

As a result Eq. (13) reduces to the diffusion equation

$$\frac{\partial}{\partial \tau} u(x, \tau) = \frac{\partial^2 u}{\partial x^2}, \tag{15}$$

whose formal solution is a Gaussian,

$$u(x, \tau) = \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{+\infty} e^{-(x-x')^2/4\tau} u(x', 0)\, dx'. \tag{16}$$

Starting from some initial state x_0, the distribution of the amount of time required for a stochastic process to encounter a threshold for the first time is known as the first passage time distribution (FPTD). We will now exhibit the relation between the diversity and the diffusion equation (15). To this end and to simplify the notation, in what follows we shall again use the symbol t to denote τ.

Consider the absorbing boundary $u(x_c, t) = 0$, where the subscript c identifies the absorption point x_c, and let $u(x, t; x_0, x_c)$ denote the probability density satisfying this boundary condition for $x < x_c$. The survival probability $S(t, x_c)$ that the particle has remained at a position $x < x_c$ for all times up to t, is given by

$$S(t, x_c) \equiv \int_{-\infty}^{x_c} u(x, t; x_0, x_c) \, dx, \qquad (17)$$

which is also the cumulative distribution of x at time t. Let the probability that a particle has reached the absorption point between times t and $t + dt$ be $h(t) \, dt = S(t) - S(t + dt)$. If we use a first order Taylor approximation, the first passage time distribution $h(t)$ is then given by

$$h(t) = -\frac{\partial S(t)}{\partial t}, \qquad (18)$$

and the relation between the cumulative distribution $S(t)$ and the FPTD (between two arbitrary times t_1 and t_2) is [35, 36]

$$S(t_1) - S(t_2) = \int_{t_1}^{t_2} h(t') \, dt'. \qquad (19)$$

Clearly, as shown in Figure 3, the diversity $d(k)$ (that counts events having achieved rank k in a fixed time window) may be identified with the right hand side of Eq. (19). This equation shows, firstly, the relation between diversity and the diffusion equation (15). Secondly, since there is a relation between the solutions of the diffusion equation and random walks, there is also one between $d(k)$ and the random walk model given by Eq. (6). We have already studied a particular case of these models in [12].

Additional material

Additional file 1: The datasets supporting the conclusions of this article. (zip)

Author details
[1]Facultad de Ciencias, Universidad Nacional Autónoma de México, México D.F., 01000, Mexico. [2]Instituto de Física, Universidad Nacional Autónoma de México, México D.F., 01000, Mexico. [3]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México D.F., 01000, Mexico. [4]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, México D.F., 04510, Mexico. [5]SENSEable City Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]MoBS Lab, Network Science Institute, Northeastern

University, Boston, MA 02115, USA. [7]ITMO University, St. Petersburg, 199034, Russian Federation. [8]Centro de Investigación y Docencia Económicas, Consejo Nacional de Ciencia y Tecnología, México D.F., 01210, Mexico. [9]Department of Computer Science, Aalto University School of Science, Aalto, 00076, Finland.

References
1. Duch J, Waitzman JS, Amaral LAN (2010) Quantifying the performance of individual players in a team activity. PLoS ONE 5(6):10937
2. Ben-Naim E, Vazquez F, Redner S (2007) What is the most competitive sport? J Korean Phys Soc 50:124
3. Merritt S, Clauset A (2013) Environmental structure and competitive scoring advantages in team competitions. Sci Rep 3:3067
4. Merritt S, Clauset A (2013) Social network dynamics in a massive online game: network turnover, non-densification, and team engagement in halo reach. Eprint. arXiv:1306.4363
5. Albert J, Bennett J, Cochran JJ (2005) Anthology of statistics in sports, vol 16. SIAM, Philadelphia
6. Radicchi F (2011) Who is the best player ever? A complex network analysis of the history of professional tennis. PLoS ONE 6(2):17249
7. Yucesoy B, Barabási A-L (2016) Untangling performance from success. EPJ Data Sci 5:17
8. Merritt S, Clauset A (2014) Scoring dynamics across professional team sports: tempo, balance and predictability. EPJ Data Sci 3:4
9. Deng W, Li W, Cai X, Bulou A, Wang QA (2012) Universal scaling in sports ranking. New J Phys 14(9):093038
10. Klaassen FJ, Magnus JR (2001) Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. J Am Stat Assoc 96(454):500-509
11. Michel J-B, Shen YK, Aiden AP, Veres A, Gray MK, Team TGB, Pickett JP, Hoiberg D, Clancy D, Norvig P, Orwant J, Pinker S, Nowak MA, Aiden EL (2011) Quantitative analysis of culture using millions of digitized books. Science 331(6014):176-182
12. Cocho G, Flores J, Gershenson C, Pineda C, Sánchez S (2015) Rank diversity of languages: generic behavior in computational linguistics. PLoS ONE 10(4):0121898
13. ATP World Tour. http://www.atpworldtour.com/. Accessed 4 April 2016
14. World Chess Federation. http://ratings.fide.com/. Accessed 6 April 2016
15. Official World Golf Ranking. http://www.owgr.com/. Accessed 6 April 2016
16. Global Poker Index. http://www.globalpokerindex.com/. Accessed 6 April 2016
17. Football Club World Ranking. http://www.clubworldranking.com/ranking-clubs.aspx. Accessed 6 April 2016
18. Fédération Internationale de Football Association. http://www.fifa.com/. Accessed 6 April 2016
19. Elo AE (1978) The rating of chessplayers, past and present. Arco Pub., London
20. Gerlach M, Altmann EG (2013) Stochastic model for the vocabulary growth in natural languages. Phys Rev X 3:021006
21. Katz JS, Katz L (1999) Power laws and athletic performance. J Sports Sci 17(6):467-476
22. Alvarez-Ramirez J, Rodriguez E (2006) Scaling properties of marathon races. Physica A 365(2):509-520
23. Visser M (2013) Zipf's law, power laws and maximum entropy. New J Phys 15(4):043021
24. Baek SK, Bernhardsson S, Minnhagen P (2011) Zipf's law unzipped. New J Phys 13(4):043004
25. Wikipedia: Gamma distribution. https://en.wikipedia.org/wiki/Gamma_distribution. Accessed 1 March 2016
26. Wikipedia: Beta distribution. https://en.wikipedia.org/wiki/Beta_distribution. Accessed 1 March 2016
27. Jóhannesson G, Björnsson G, Gudmundsson EH (2006) Afterglow light curves and broken power laws: a statistical study. Astrophys J Lett 640(1):L5-L8
28. Li W, Miramontes P, Cocho G (2010) Fitting ranked linguistic data with two-parameter functions. Entropy 12(7):1743
29. Kolmogorov AN (1933) Sulla determinazione empirica di una legge di distribuzione. G Ist Ital Attuari 4(1):83-91
30. Clauset A, Shalizi CR, Newman ME (2009) Power-law distributions in empirical data. SIAM Rev 51(4):661-703
31. Alvarez-Martinez R, Cocho G, Rodríguez RF, Martínez-Mekler G (2014) Birth and death master equation for the evolution of complex networks. Physica A 402:198-208
32. Martínez-Mekler G, Martínez RA, del Río MB, Mansilla R, Miramontes P, Cocho G (2009) Universality of rank-ordering distributions in the arts and sciences. PLoS ONE 4(3):4791
33. Van Kampen NG (2007) Stochastic processes in physics and chemistry. North Holland, Amsterdam
34. Wheeler JC, Gordon RG, Baker GA, Gammel JL (1970) The Padé approximant in theoretical physics. Academic, New York
35. Perline R (1996) Zipf's law, the central limit theorem, and the random division of the unit interval. Phys Rev E 54(1):220
36. Perline R, Perline R (2016) Two universality properties associated with the monkey model of Zipf's law. Entropy 18(3):89