The Annals of Applied Probability
2023, Vol. 33, No. 6A, 4366–4394
https://doi.org/10.1214/22-AAP1921

RANKING-BASED RICH-GET-RICHER PROCESSES

BY PANTELIS P. ANALYTIS[1,a], ALEXANDROS GELASTOPOULOS[1,b] AND HRVOJE STOJIC[2,c]

[1]*Department of Business and Management, University of Southern Denmark,* [a]*pantelis@sam.sdu.dk,* [b]*alex@sam.sdu.dk*
[2]*Department of Economics and Business, Pompeu Fabra University,* [c]*hrvoje.stojic@upf.edu*

We study a discrete-time Markov process $X_n \in \mathbb{R}^d$ for which the distribution of the future increments depends only on the relative ranking of its components (descending order by value). We endow the process with a rich-get-richer assumption and show that, together with a finite second moments assumption, it is enough to guarantee almost sure convergence of X_n/n. We characterize the possible limits if one is free to choose the initial state and we give a condition under which the initial state is irrelevant. Finally, we show how our framework can account for ranking-based Pólya urns and can be used to study ranking algorithms for web interfaces.

1. Introduction. Wealthy individuals tend to become even wealthier [52], popular websites become even more popular [7], and highly cited papers overshadow less cited ones, gaining more future citations [54, 55]. Social and technological systems that preserve and amplify existing inequalities are said to be characterized by rich-get-richer dynamics [46, 54, 59]. In these systems, initial conditions and randomness early in time drastically affect the course of future events—advantages obtained by agents early on are conserved and reinforced [4, 16]. The above can result in socially objectionable outcomes, such as pervasive inequality in the distribution of wealth, and unfair outcomes where talented people or promising technologies cannot compete with already established ones [49].

In many systems, and increasingly in the online world, the rich-get-richer dynamics depend on the ranks of the various objects (people, options, institutions, etc.) in terms of some quantity of interest. For example, companies or academic institutions might receive job applications based on some status ranking, which in turn can help these institutions retain their status by employing qualified individuals [53]. Similarly, scientists might submit their work to journals by taking into account the journal's relative rank in terms of impact factor or some other metric, thus highly ranked journals are more likely to publish work of good quality and retain their position in the ranking [28, 40]. Last but not least, users of online interfaces are more likely to click on entries that appear at the top of the screen, hence making these entries appear more relevant to other users [33, 57]. In all of these cases, it is the ranking of the different entities that confers an advantage to the more successful ones and thus drives the rich-get-richer dynamics.

Although examples of systems characterized by ranking-based rich-get-richer dynamics abound, we still do not understand their dynamics and long-term behavior. There are a couple of lines of research that touch upon ranking-based processes. Hill et al. [27] study the case of a Pólya urn with balls of $d = 2$ colors, one ball added at a time, and allow the probability of adding a red ball to be a function of the proportion of red balls. In other words, there is some function $f : [0, 1] \to [0, 1]$, such that the probability of the next ball being red is $f(X_n/n)$, where X_n denotes the number of red balls at time n. If f is taken to be constant in

Received May 2021; revised August 2022.
MSC2020 subject classifications. Primary 60J05; secondary 60J20.
Key words and phrases. Ranking, rich-get-richer, Markov process, Pólya urn.

$[0, \frac{1}{2})$ and in $(\frac{1}{2}, 1]$, then we get a ranking-based urn. In [27] it is shown that X_n/n converges a.s., and then some results are given regarding the support of the limit (see also Section 3.2). Importantly, a subset of the results in [27] allows a nowhere dense set of discontinuities for f, so they apply to the ranking-based case. It is not obvious though how to generalize these results to Pólya urns with more colors or other types of processes.

The usual generalization to $d \in \mathbb{N}$ is to have the probability of adding a ball of color i be proportional to a function of the count (or proportion) of balls of that color only, thus not allowing comparison of the counts of balls of different colors (for recent examples see [10, 12, 41]—see also [51, 60] for surveys of results). A notable exception is the work of Arthur et al. [5], where the probabilities are allowed to depend on the whole vector of proportions of balls of each color. More precisely, there is an urn function

$$f : \Delta^{d-1} \to \Delta^{d-1} \quad \text{where } \Delta^{d-1} := \left\{ x \in [0, 1]^d, \sum_i x_i = 1 \right\}, \tag{1}$$

which takes as argument the vector of proportions of balls of each color, and its ith component gives the probability of adding a ball of color i. The authors generalize some of the results in [27] to any $d \in \mathbb{N}$. In particular they show that under mild conditions on f the process X_n/n (where X_n is now a vector) has positive probability of converging to any point $\theta \in \Delta^{d-1}$ that is a *stable* fixed point of f. According to the definition of stability used, in the ranking-based case all fixed points whose coordinates are all distinct are necessarily stable (see Section 3.2 for details). However, it is not claimed that the stable fixed points of f are the only possible limits for X_n/n. Also, convergence of X_n/n is shown only for certain special cases that do not cover ranking-based urns.

Even in the cases where the above results are applicable to ranking-based systems, their main limitation is that they are restricted to simple Pólya-type processes, that is, processes whose components increase one at a time and the increments are binary. But in many systems with ranking-dependent dynamics (e.g., journal impact factors, university rankings) the quantity of interest can take continuous values and the various components may change simultaneously. Thus, a more general setting is needed to model the dynamics of such systems.

In this work, we treat the problem in the context of (discrete-time) Markov processes, with the dynamics depending explicitly on the ranking. Specifically, we consider a nonhomogeneous random walk in $\mathbb{R}^d$, for which the distribution of the steps depends only on the ranking of its components (descending order of their values). Thus, in between changes of ranking, the process is an *ordered random walk* [18], which is a special case of a random walk in a cone [14, 15, 17, 23, 24]. The literature on this topic provides estimates on the exit time of a random walk from a ranking (cone) in the case that this time is almost surely finite, that is, the ranking is bound to eventually change. Here, instead of focusing on exit times, we find conditions under which the ranking eventually *stabilizes*, and we characterize the support of the limit ranking. Given that the applications that we are interested in are systems with rich-get-richer dynamics, the condition that we find for the ranking to stabilize is a ranking-based reinforcement condition (Assumption 2.5), and it is a weaker version of the following statement: conditioned on $X_n^i > X_n^j$, the difference $X_{n+1}^i - X_n^i$ has a larger mean than $X_{n+1}^j - X_n^j$.

Our results can be summarized as follows: under the abovementioned ranking-based reinforcement assumption and a finite second moments assumption, we show that in the limit $n \to \infty$ the ranking of the components of the process stops changing almost surely (Theorem 2.7). Independently of Assumption 2.5, we characterize the possible limits for the ranking (Theorem 2.10). By "possible limit" we mean that the probability of converging to this value is positive for *some* initial condition (distribution of X_0). Proposition 2.20 gives a condition under which the probability of converging to any of the possible limits is positive for any initial condition.

We also study the long-term behavior of X_n given that the ranking stabilizes. Conditioned on retaining the same ranking for all $n \geq n_0 \in \mathbb{N}$, in the limit $n \to \infty$ the process behaves like a regular random walk, satisfying the strong law of large numbers and central limit theorem that one would expect (Propositions 2.14 and 2.16). We note that, although intuitive, the validity of at least the central limit theorem is not trivial, since the finite-time conditional distributions of X_n differ from the (unconditional) ones for the corresponding regular random walk. Our results strengthen one of the few known results for random walks in cones with possibly infinite exit time [24], Proposition 5.1.

In Section 3 we consider applications to rankings in web interfaces; in Section 3.1 we show in detail how a commonly used recommendation algorithm gives rise to a ranking-based process, and in Section 3.2 we specialize our results to ranking-based Pólya urns and compare them to previously known results.

1.1. *Ranking-based processes without rich-get-richer dynamics.* Two more lines of research have studied ranking-based processes, but focus on regimes that exclude rich-get-richer dynamics.

Kemperman [39] studies the *oscillating random walk*, that is, a random walk in $\mathbb{R}^d$ whose increment distribution changes on either side of a hyperplane that passes through the origin. In the case of $d = 1$, this is equivalent to a ranking-based process with two components (by setting $Y = X^1 - X^2$). The author finds necessary and sufficient conditions for the random walk to be recurrent, which are nontrivial when some second moment does not exist. See also [37], Section 2, [38], and [45], Section 5.3. As the name "oscillating" suggests, this literature is concerned more with the recurrent case. This means that, if the first moments of the increment distributions on both sides of 0 are finite, they must point towards 0. This is quite the opposite of the rich-get-richer dynamics that we are interested in, since we want the drift to change (when the ranking changes) in the direction that the random walk moves. We are also not aware of any generalization of the oscillating random walk to $d > 1$ that can account for the ranking of multiple items.

In the continuous-time case, there is some related literature started by Banner et al. [6] (see also [29, 30, 32, 50] and references therein), which studies the case of Brownian particles with rank-dependent drifts and diffusion coefficients, that is, processes satisfying

$$dX_i(t) = \delta_{r_i(t)}\,dt + \sigma_{r_i(t)}\,dB_i(t), \tag{2}$$

where $r_i(t)$ is the rank of component i at time t, and the δ_i's and σ_i's are constants. This literature studies the limit behavior of the spacings $X^{(i)} - X^{(i+1)}$ of the rank statistics, under some assumption that guarantees that they do not escape to infinity. For example, in [6] the authors assume that

$$\frac{1}{k} \cdot \sum_{i=1}^{k} \delta_i < \frac{1}{d-k} \cdot \sum_{i=k+1}^{d} \delta_i, \tag{3}$$

that is, the average drift of the largest k components is smaller than that of the smallest $d-k$ components, for all k, which is a necessary and sufficient condition for the process $(X^{(1)} - X^{(2)}, \dots, X^{(d-1)} - X^{(d)})$ to be tight and converge to a stationary distribution [50]. As in the case of the oscillating random walk in the previous paragraph, this is in stark contrast with rich-get-richer dynamics, where being ranked higher should be beneficial. For the sake of comparison, the analogue of our Assumption 2.5 in the above setting would be to require $\delta_1 > \cdots > \delta_d$.

2. Main results. We begin by defining what we mean by ranking (Section 2.1) and ranking-based processes (Section 2.2). Sections 2.3 and 2.4 contain our two main results: convergence of ranking and characterization of terminal rankings. In Section 2.5 we look at the limit behavior of the process X_n itself and in Section 2.6 we consider the role of the initial condition.

2.1. *Rankings*. For a finite set S, we denote by $|S|$ its cardinality and by $[S] = \{1, \dots, |S|\}$ the set of the first $|S|$ positive integers.

DEFINITION 2.1. Let S be a finite set. A ranking of S is a function $r : S \to [S]$ with the property that for each $a \in S$,

$$\operatorname{card}\{b \in S : r(b) < r(a)\} = r(a) - 1. \tag{4}$$

We will say that a is ranked higher than b if $r(a) < r(b)$. Equation (4) requires that, for each $a \in S$, exactly $r(a) - 1$ elements are ranked higher than a. Thus, we will call $r(a)$ the *position* or *rank* of a in the ranking r. Note that two elements $a, b \in S$, $a \neq b$, can have the same position in r, that is, we may have $r(a) = r(b)$. In this case we will say that these elements are equally ranked by r. In Appendix A we show that rankings of a set S are equivalent to weak orderings on S.

Any bijection $r : S \to [S]$ satisfies Eq. (4), hence it is a ranking. Such rankings will be called *strict*. That is, strict rankings are such that no two elements of S are equally ranked.

Given a vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, we denote by $rk(x)$ the unique ranking r on the set $[d] = \{1, \dots, d\}$ that satisfies $r(i) < r(j)$ if and only if $X_i > X_j$, for any $i, j \in [d]$. It is easy to check that there is indeed a unique such ranking, given by $r(i) = \operatorname{card}\{j \in [d] : X_j > X_i\} + 1$. The folk name for this map is the *standard competition ranking*.

We will denote by $\mathcal{R}(d)$ the set of all rankings of the set $[d]$.

2.2. *Ranking-based processes*. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ a filtration on it. Let ν be a probability distribution on $\mathbb{R}^d$ with finite second moments, and for each $r \in \mathcal{R}(d)$ let μ^r be a probability distribution on $\mathbb{R}^d$, also with finite second moments. We consider a time-homogeneous Markov process $X_n \in \mathbb{R}^d$, adapted to $\{\mathcal{F}_n\}_n$, with initial distribution ν and with the law of its increments being μ^r, where r is the current ranking. More precisely, the transition kernel μ is given by

$$\mu(x, B) = \mu^{rk(x)}(B - x), \quad x \in \mathbb{R}^d, B \in \mathcal{B}(\mathbb{R}^d), \tag{5}$$

where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra of $\mathbb{R}^d$ and $B - x = \{y \in \mathbb{R}^d : y + x \in B\}$ denotes the translation of B by the vector $-x$. We will call such a process a (d-dimensional) *ranking-based process*.

Equation (5) implies that for any $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{P}(\Delta X_{n+1} \in B \mid \mathcal{F}_n) = \mu^{rk(X_n)}(B) \quad \text{a.s.}, \tag{6}$$

where $\Delta X_{n+1} = X_{n+1} - X_n$. In particular, the process is space-homogeneous within subsets of $\mathbb{R}^d$ that correspond to a fixed ranking, that is, subsets of the form $\{x \in \mathbb{R}^d : rk(x) = r\}$, for $r \in \mathcal{R}(d)$ (but it is not space-homogeneous in general). Equation (6) also implies that, conditioned on the ranking at time n, ΔX_{n+1} is independent of $\mathcal{F}_n$, that is,

$$\Delta X_{n+1} \underset{rk(X_n)}{\perp\!\!\!\perp} \mathcal{F}_n. \tag{7}$$

We will use the shorthand notation $\mu^r(x_i > x_j)$ to mean $\mu^r(\{x \in \mathbb{R}^d : x_i > x_j\})$ and similarly for other events that involve comparisons of components of x.

For each $r \in \mathcal{R}(d)$, we denote by $Z^r = (Z_1^r, \dots, Z_d^r)$ a random variable with distribution μ^r. This will be especially useful when considering differences of the form $\Delta X_{n+1}^i - \Delta X_{n+1}^j$, whose distribution cannot be directly expressed via μ^r. Note that conditioned on $rk(X_n)$, $\Delta X_{n+1}^i - \Delta X_{n+1}^j$ has the distribution of $Z_i^{rk(X_n)} - Z_j^{rk(X_n)}$, that is, for each $B \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}\big(\Delta X_{n+1}^i - \Delta X_{n+1}^j \in B \mid rk(X_n)\big) = \mathbb{P}\big(Z_i^{rk(X_n)} - Z_j^{rk(X_n)} \in B\big). \tag{8}$$

We denote by q_i^r the mean of the ith component of the distribution μ^r, that is, $q_i^r = \mathbb{E}[Z_i^r]$. We will also use the vector notation $q^r = (q_1^r, \dots, q_d^r)$.

For the rest of the paper, we fix $d \in \mathbb{N}$ and a d-dimensional ranking based process X_n adapted to $\{\mathcal{F}_n\}_n$, with the associated μ^r's, μ, Z^r's, and q_i^r's. Strictly speaking, the Markov process X is described by the pair (ν, μ). However, we will often abuse terminology and talk about a single process X while allowing the initial distribution ν to vary. We will use the notation $\mathbb{P}_\nu$ for probabilities of events that depend on the initial distribution, but we will often omit the subscript ν otherwise (as in Eq. (8)). Both the distributions μ^r and initial distribution ν will always be assumed to have finite second moments.

We will also suppress the integer d in the notation for the set of all rankings of $[d]$ and write $\mathcal{R} = \mathcal{R}(d)$.

2.3. *Convergence of ranking.* As $n \to \infty$, the ranking of X_n may keep changing or it might converge to some particular ranking $r \in \mathcal{R}$ (where $\mathcal{R}$ is endowed with the discrete topology). We have the following definition.

DEFINITION 2.2. Let X be a ranking-based process. We say that $rk(X_n)$ converges to $r \in \mathcal{R}$ and write $rk(X_n) \to r$ (or $\lim_{n\to\infty} rk(X_n) = r$), if $rk(X_n) = r$ for all sufficiently large n, that is,

$$\{rk(X_n) \to r\} = \bigcup_{n_0 \in \mathbb{N}} \bigcap_{n=n_0}^{\infty} \{rk(X_n) = r\} = \liminf_n \{rk(X_n) = r\}. \tag{9}$$

We say that a ranking $r \in \mathcal{R}$ is terminal (for the transition kernel μ), if there exists some initial distribution ν, such that

$$\mathbb{P}_\nu\big(rk(X_n) \to r\big) > 0. \tag{10}$$

Otherwise, we say that r is transient.

Knowing that the ranking converges is useful because then we can predict the long-term behavior of the process (see Section 2.5). We will therefore seek conditions under which the ranking is guaranteed to converge.

As a first step, we ask the following question: if we know that $X_{n_0}^i > X_{n_0}^j$ occurs for some $n_0 \in \mathbb{N}$, is it likely that $X_n^i > X_n^j$ for all $n > n_0$? The following definition and proposition give a sufficient condition for the probability of this event to be positive and bounded away from zero.

DEFINITION 2.3. Let $\{X_n\}_{n\in\mathbb{N}}$ be a d-dimensional ranking-based process with the associated distributions μ^r and means $q_1^r, \dots, q_d^r$, and let $i, j \in [d]$.

- We say that i quasi-dominates j, if for any ranking r such that $r(i) < r(j)$ we have either $q_i^r > q_j^r$ or $\mu^r(x_i \neq x_j) = 0$.
- We say that i dominates j if we further have that for any ranking r such that $r(i) = r(j)$, either $\mu^r(x_i > x_j) > 0$ or $\mu^r(x_i \neq x_j) = 0$.

Note the relation between quasi-dominance and the (loosely defined) concept of rich-get-richer dynamics: if i quasi-dominates j, then X_n^i increases on average faster than X_n^j whenever X_n^i is already larger (or they vary in exactly the same way). The extra condition for dominance says that X_n^i has a nonzero probability of passing ahead after a tie (or, again, the two components vary in exactly the same way).

PROPOSITION 2.4. *Let $i, j \in [d]$.*

1. *If i quasi-dominates j, then there exists some $\epsilon > 0$ such that for any initial distribution ν and any $\mathcal{F}_n$-stopping time s, we have*

$$\mathbb{P}_\nu\left(\bigcap_{n=s}^{\infty}\{X_n^i > X_n^j\} \,\middle|\, \mathcal{F}_s\right) \geq \epsilon \quad \textit{a.s. on } \{s < \infty\} \cap \{X_s^i > X_s^j\}. \tag{11}$$

2. *If i dominates j, we further have*

$$\mathbb{P}_\nu\left(\bigcap_{n=s}^{\infty}\{X_n^i \geq X_n^j\} \,\middle|\, \mathcal{F}_s\right) \geq \epsilon \quad \textit{a.s. on } \{s < \infty\} \cap \{X_s^i \geq X_s^j\}. \tag{12}$$

For a concrete case, if we take the a.s. constant stopping time $s = n_0$, then Eq. (11) implies in particular that

$$\mathbb{P}_\nu\left(\bigcap_{n=n_0}^{\infty}\{X_n^i > X_n^j\} \,\middle|\, X_{n_0}^i > X_{n_0}^j\right) \geq \epsilon, \tag{13}$$

whenever the expression on the left hand side makes sense (i.e., whenever $\mathbb{P}_\nu(X_{n_0}^i > X_{n_0}^j) > 0$).

We postpone the proof in order to get to our main result for this section. For ease of reference we state the condition for that theorem separately:

ASSUMPTION 2.5 (Ranking-based reinforcement). For any pair of indices $i, j \in [d]$, either one of them dominates the other, or they quasi-dominate each other.

Note that it is possible for both i and j to dominate each other; the above assumption would still be satisfied. This means that the "dominance" relation does not have to be trichotomous. It does not have to be transitive either. However, a transitive trichotomous relation (i.e., a strict total order) on $[d]$ would satisfy Assumption 2.5.

EXAMPLE 2.6. Let $X_n = (X_n^1, \dots, X_n^d)$ give the number of balls of each of d colors in an urn. At each time step, a single ball is added, with probabilities for each color depending on the ranking. Note that in this case q_i^r is equal to the probability of adding a ball of color i when the ranking is r (see also the first paragraph of Section 3.2). These probabilities will be determined as follows: Each color has a propensity $a_i \geq 0$ to be chosen. Moreover, there are real numbers $\lambda_1 > \cdots > \lambda_d \geq 0$, with λ_i denoting an additive bonus to the propensity of the color(s) currently ranked ith. More specifically, the probability of adding a ball of color i, given that the current ranking is r, is

$$q_i^r = \frac{a_i + \lambda_{r(i)}}{\sum_{j=1}^d (a_j + \lambda_{r(j)})}. \tag{14}$$

We claim that this process satisfies Assumption 2.5. To see this, let $i, j \in [d]$ and suppose without loss of generality that $a_i \geq a_j$. We have the following cases:

- $a_i = a_j$: By Eq. (14) we have that $q_i^r > q_j^r$ whenever i is ranked higher than j and vice versa. That is, i and j quasi-dominate each other.
- $a_i > a_j$: We similarly get that color i quasi-dominates color j. Moreover, when i and j are ranked equally (i.e., $r(i) = r(j)$), Eq. (14) gives $q_i^r > q_j^r$, that is, it is more likely for color i to be chosen. This shows that i dominates j.

Thus our claim is proved.

We now state and prove our main theorem for this section.

THEOREM 2.7 (Convergence of ranking). *Let X_n be a ranking-based process satisfying Assumption 2.5. Then, $rk(X_n)$ converges a.s., for any initial distribution ν.*

PROOF. It is enough to show that for each pair of indices i, j, the relative ranking of X_n^i and X_n^j eventually stops changing with probability 1. So let $i \neq j$ and, without loss of generality, assume that i quasi-dominates j (see Assumption 2.5). Define $s_0 = 0$ and inductively $t_m = \inf\{n > s_{m-1} : X_n^i > X_n^j\}$ and $s_m = \inf\{n > t_m : X_n^i \leq X_n^j\}$. Notice that $\{s_m = \infty\} = \bigcap_{n=t_m}^{\infty}\{X_n^i > X_n^j\}$. Therefore, Proposition 2.4 applied to $s = t_m$ implies that there exists some ϵ, not depending on m, such that

$$\mathbb{P}_\nu(s_m = \infty \mid \mathcal{F}_{t_m}) \geq \epsilon > 0, \quad \text{a.s.} \tag{15}$$

on $\{t_m < \infty, X_{t_m}^i > X_{t_m}^j\} = \{t_m < \infty\}$. In particular, if $\mathbb{P}_\nu(t_m < \infty) > 0$, then

$$\mathbb{P}_\nu(s_m = \infty \mid t_m < \infty) \geq \epsilon \tag{16}$$

and

$$\begin{aligned} \mathbb{P}_\nu(s_m < \infty) &= \mathbb{P}_\nu\big(\{t_m < \infty\} \cap \{s_m < \infty\}\big) \\ &= \mathbb{P}_\nu(t_m < \infty) \cdot \mathbb{P}_\nu(s_m < \infty \mid t_m < \infty) \\ &\leq (1-\epsilon) \cdot \mathbb{P}_\nu(t_m < \infty) \\ &\leq (1-\epsilon) \cdot \mathbb{P}_\nu(s_{m-1} < \infty). \end{aligned} \tag{17}$$

Although we have assumed $\mathbb{P}_\nu(t_m < \infty) > 0$, Eq. (17) continues to hold even if $\mathbb{P}_\nu(t_m < \infty) = 0$, because then $\mathbb{P}_\nu(s_m < \infty) = 0$ as well.

By Eq. (17) and induction we have $\mathbb{P}_\nu(s_m < \infty) \leq (1-\epsilon)^m$, therefore

$$\mathbb{P}_\nu\Big(\bigcap_{m\in\mathbb{N}}\{s_m < \infty\}\Big) = 0. \tag{18}$$

Hence, with probability 1, either $X_n^i \leq X_n^j$ finitely often (henceforth abbreviated f.o.) or $X_n^i > X_n^j$ f.o. If $X_n^i \leq X_n^j$ f.o., then $X_n^i > X_n^j$ for all sufficiently large n, so we are done. Now assume that $X_n^i > X_n^j$ f.o. and separate two cases, according to Assumption 2.5:

- j also quasi-dominates i: We get similarly that either $X_n^i \geq X_n^j$ f.o. or $X_n^i < X_n^j$ f.o. As before, in the first case we are done. In the second case, we have both $X_n^i < X_n^j$ and $X_n^i > X_n^j$ f.o., so that $X_n^i = X_n^j$ for all sufficiently large n.
- i dominates j: Using the second part of Proposition 2.4 we get that either $X_n^i < X_n^j$ f.o. or $X_n^i \geq X_n^j$ f.o. The situation is identical to the first case. □

To give some more intuition on Assumption 2.5, let us demonstrate how it might fail. One way for this to happen is if a higher rank is harmful for the mean rate of increase. Then, it could be the case, for example, that X_n^1 increases more slowly than X_n^2 whenever $1 \succ_r 2$ and vice versa, hence neither of the indices 1 or 2 would quasi-dominate the other. That would describe a situation opposite to that of rich-get-richer dynamics, and more similar to the oscillating random walk discussed in Section 1.1.

But the above is not the only case in which Assumption 2.5 can fail. The next example describes a scenario in which being ranked higher is never harmful (everything else being equal), but Assumption 2.5 still fails.

EXAMPLE 2.8. Suppose that there are two types of people in a population, denoted by A and B, in proportions $p > 0.5$ and $1 - p$ respectively, and $d = 3$ types of items. People of type A always prefer items 1 and 2 over item 3, while people of type B always prefer item 3 over the other items. Between items 1 and 2, everyone always prefers the most popular one (and let's say they prefer item 1 in case of a tie). At every step, we randomly pick an individual from the population and ask them to choose among the three items. If we denote by X_n^i the number of times that item i has been chosen up to individual n, then X_n can be easily seen to be a ranking-based process. Since X_n^i increases by 1 whenever i is chosen, q_i^r is equal to the probability that item i is the most preferred one by a randomly picked individual, given that $rk(X_n) = r$.

In particular, if $X_n^3 > X_n^1 > X_n^2$, then there is a chance p that the randomly picked $(n+1)$th agent is of type A and they will choose item 1, while with probability $1 - p$ the agent will be of type B and they will choose item 3. Since $p > 1 - p$, for the ranking $3 \succ_r 1 \succ_r 2$ we have that $q_1^r > q_3^r$, implying that item 3 does not quasi-dominate item 1.

On the other hand, if $X_n^2 > X_n^1 > X_n^3$, then item 3 is again chosen with probability $1 - p$ (if the next agent is of type B), while item 1 cannot be chosen, because an agent of type A would prefer item 2. Therefore, for the ranking $2 \succ_r 1 \succ_r 3$, we have that $q_1^r < q_3^r$, showing that 1 does not quasi-dominate item 3 either. Since neither of the indices 1 or 3 quasi-dominates the other, Assumption 2.5 is not satisfied.

Given that there are natural examples of systems that do not satisfy Assumption 2.5, it is worth considering whether this assumption is not only sufficient for the a.s. convergence of $rk(X_n)$ for any initial distribution (Theorem 2.7), but also necessary, or whether a similar but weaker condition can be found that is both necessary and sufficient.

Let us consider the case $d = 2$ first and for simplicity let's assume that $\mu^r(x_1 > x_2) > 0$ and $\mu^r(x_2 > x_1) > 0$ for all rankings $r \in \mathcal{R}$. Then, Assumption 2.5 reduces to requiring that index 1 quasi-dominates index 2 or vice-versa. Denote by id and $-$id the rankings $1 \succ_{\text{id}} 2$ and $2 \succ_{\text{-id}} 1$, respectively. It is clear that, if $q_1^{\text{id}} \le q_2^{\text{id}}$ and $q_1^{\text{-id}} \ge q_2^{\text{-id}}$ both hold, then the process will oscillate between $X_n^1 > X_n^2$ and $X_n^1 < X_n^2$ (see also [39], Corollary 4.7). Therefore, $q_1^{\text{id}} > q_2^{\text{id}}$ or $q_1^{\text{-id}} < q_2^{\text{-id}}$ is necessary for convergence of $rk(X_n)$. But this is exactly the definition of quasi-dominance, hence in this case Assumption 2.5 is both necessary and sufficient for $rk(X_n)$ to converge, for any or all initial distributions.

The situation is different, though, for $d > 2$. For example, in Example 2.8 $rk(X_n)$ converges a.s., no matter the initial distribution, even though Assumption 2.5 is not satisfied. For a more transparent case, consider the following example with $d = 3$:

- If component 3 is ranked third, then $\Delta X_{n+1} = (2, 1, 0)$ a.s.
- In any other case, $\Delta X_{n+1} = (1, 2, 0)$ a.s.

Here neither index 1 quasi-dominates index 2, nor vice versa. However, it is clear that, no matter the initial distribution, eventually component 3 will be ranked third, hence $\Delta X_{n+1} =$

$(2, 1, 0)$ for all sufficiently large n, which implies that $rk(X_n)$ converges a.s. to the ranking $1 \succ_r 2 \succ_r 3$. Both of these examples show that Assumption 2.5 is not necessary for a.s. convergence of $rk(X_n)$ for every initial distribution ν.

What these examples suggest is that, in determining the long-term behavior of $rk(X_n)$, the distributions μ^r for some rankings r are irrelevant, because these rankings cannot be accessed or they cannot be accessed infinitely often. However, the problem of figuring out which rankings cannot be accessed infinitely often is intertwined with the long-term behavior of $rk(X_n)$ itself, which we want to characterize. For this reason, the task of finding a necessary and sufficient condition for a.s. convergence of $rk(X_n)$ in terms of the distributions μ^r is nontrivial, and we leave it as an open question.

We now turn to the proof of Proposition 2.4. We will need the following lemma, which generalizes a property of biased random walks to the case when the transition probabilities are not constant, but vary in a finite set. Its proof is given in Appendix B. A related result is obtained in [45], Theorem 2.5.12, by different methods.

LEMMA 2.9. *Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space. Let S be a finite set and for each $r \in S$ let ν^r be a distribution on $\mathbb{R}$ such that it either has positive mean or $\nu^r(\{0\}) = 1$. Let $\{R_n\}_{n\in\mathbb{N}}$ be a sequence of random elements in S and $\{Y_n\}_{n\in\mathbb{N}}$ a sequence of random variables with $Y_0 = 0$. Suppose that ΔY_{n+1} is conditionally independent of $\{(Y_k, R_k)\}_{k\le n}$ conditioned on R_n, with distribution ν^{R_n}. In other words, for any $A \in \mathcal{B}(\mathbb{R})$, $n \in \mathbb{N}$,*

$$\mathbb{P}\big(\Delta Y_{n+1} \in A \,\big|\, \{(Y_k, R_k)\}_{k\le n}\big) = \nu^{R_n}(A) \quad a.s. \tag{19}$$

Then,

$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} \{Y_n \ge 0\}\Big) \ge \epsilon > 0, \tag{20}$$

where ϵ depends only on the distributions ν^r, $r \in S$.

We note that if $|S| = 1$, then Lemma 2.9 reduces to the well-known result that a biased one-dimensional random walk with positive mean has positive probability of never admitting negative values (see [35], Corollary 9.17).

PROOF OF PROPOSITION 2.4.

1. Let s and ν be given and define $\tau = \min\{n \ge s : X_n^i \le X_n^j\}$ and

$$Y_n = X^i_{\tau\wedge n} - X^j_{\tau\wedge n}. \tag{21}$$

Note that $Y_n > 0$ for all $n \ge s$ implies $X_n^i > X_n^j$ for all $n \ge s$. Therefore, it is enough to show that, for some $\epsilon > 0$ that does not depend on s or ν,

$$\mathbb{P}_\nu\Big(\bigcap_{n=s}^{\infty} \{Y_n > 0\} \,\Big|\, \mathcal{F}_s\Big) \ge \epsilon \quad \text{a.s. on } \{Y_s > 0\}. \tag{22}$$

We have

$$\Delta Y_{n+1} = \mathbf{1}_{\tau>n} \cdot \big(\Delta X^i_{n+1} - \Delta X^j_{n+1}\big), \tag{23}$$

where $\mathbf{1}_A$ denotes the indicator function of the set A. It follows that conditioned on $rk(X_n)$ and $\mathbf{1}_{\tau>n}$, ΔY_{n+1} is independent of $\mathcal{F}_n$ (see Eq. (6)). Moreover, its conditional distribution is equal to that of $Z_i^{rk(X_n)} - Z_j^{rk(X_n)}$ in the case $\tau > n$ (by Eq. (8)), while $\Delta Y_{n+1} = 0$ identically otherwise.

Let $F \in \mathcal{F}_s$ be any event with $\mathbb{P}(F) > 0$ and consider the probability measure $\mathbb{P}_{\nu,F}(\cdot) = \mathbb{P}_\nu(\cdot \mid F)$. We apply Lemma 2.9 for this measure and the sequence $\{Y_{n+s} - Y_s\}_{n\in\mathbb{N}}$, with $S = \mathcal{R} \sqcup \{\alpha\}$ (where α is an arbitrary new element) and

$$R_n = \begin{cases} rk(X_{n+s}) & \text{if } \tau > n+s, \\ \alpha & \text{otherwise.} \end{cases} \tag{24}$$

The distributions ν^r in Lemma 2.9 are equal to the distributions of $Z_i^r - Z_j^r$ for $r \in \mathcal{R}$, while ν^α is the singular probability measure satisfying $\nu^\alpha(\{0\}) = 1$. Lemma 2.9 thus gives

$$\mathbb{P}_\nu\left(\bigcap_{n\in\mathbb{N}} \{Y_{n+s} - Y_s \geq 0\} \,\middle|\, F\right) \geq \epsilon > 0, \tag{25}$$

where ϵ depends only on the μ^r's (distributions of Z^r's). Since F was arbitrary, we get

$$\mathbb{P}_\nu\left(\bigcap_{n\in\mathbb{N}} \{Y_{n+s} - Y_s \geq 0\} \,\middle|\, \mathcal{F}_s\right) \geq \epsilon \quad \text{a.s.,} \tag{26}$$

from where Equation (22) follows.

2. Let $\tau' = \inf\{n \geq s : X_n^i \neq X_n^j\}$. We may assume that $X_s^i = X_s^j$ and $\tau' < \infty$, since on $\{X_s^i > X_s^j\}$ part (a) applies, while on $\{\tau' = \infty\}$ the result holds trivially. On $\{X_s^i = X_s^j, \tau' < \infty\}$ we have $\bigcap_{n=s}^\infty \{X_n^i \geq X_n^j\} = \bigcap_{n=\tau'}^\infty \{X_n^i \geq X_n^j\}$ and $\tau' \geq s+1$; it is therefore enough to show that

$$\mathbb{P}_\nu\left(\bigcap_{n=\tau'}^{\infty} \{X_n^i > X_n^j\} \,\middle|\, \mathcal{F}_{\tau'-1}\right) \geq \epsilon \quad \text{a.s.} \tag{27}$$

or, by part (a),

$$\mathbb{P}_\nu\big(X_{\tau'}^i > X_{\tau'}^j \mid \mathcal{F}_{\tau'-1}\big) \geq \epsilon' \quad \text{a.s.,} \tag{28}$$

for some $\epsilon' > 0$ that does not depend on ν or s.

Let $R' = \{r \in \mathcal{R} : r(i) = r(j), \mu^r(x_i > x_j) > 0\}$ be the subset of rankings that rank i and j equally, but that give positive probability to i to pass ahead on the next step. Since by assumption all other rankings with $r(i) = r(j)$ satisfy $\mu^r(x_i \neq x_j) = 0$, $rk(X_n)$ must take a value in R' before we can have $X_n^i \neq X_n^j$. That is, $rk(X_{\tau'-1}) \in R'$ a.s., hence also

$$\begin{aligned} \mathbb{P}\big(X_{\tau'}^i > X_{\tau'}^j \mid \mathcal{F}_{\tau'-1}\big) &= \mathbb{P}\big(\Delta X_{\tau'}^i > \Delta X_{\tau'}^j \mid \mathcal{F}_{\tau'-1}\big) \\ &= \mu^{rk(X_{\tau'-1})}(x_i > x_j) \\ &\geq \min_{r\in R'} \mu^r(x_i > x_j) > 0 \quad \text{a.s.,} \end{aligned} \tag{29}$$

where the second equality follows from Eq. (6). □

2.4. *Terminal rankings*. Theorem 2.7 says that Assumption 2.5 guarantees convergence of $rk(X_n)$, but it doesn't say anything about the possible limits. In this section we deal with the question of what the possible limit rankings are. Recall that a ranking is terminal if $\mathbb{P}_\nu(rk(X_n) \to r) > 0$ for some probability distribution ν (Definition 2.2). Our main result in this section is the following:

THEOREM 2.10 (Terminal rankings). *Let X_n be a d-dimensional ranking-based process with the associated distributions μ^r and means $q_1^r, \ldots, q_d^r$. A ranking r is terminal if and only if, for any $i, j \in [d]$:*

- *If $r(i) = r(j)$ then $\mu^r(x_i \neq x_j) = 0$.*
- *If $r(i) < r(j)$ then either $q_i^r > q_j^r$ or $\mu^r(x_i \neq x_j) = 0$.*

Let us give some intuition behind Theorem 2.10. If $rk(X_n) \to r$, then there exists some $n_0 \in \mathbb{N}$ such that $rk(X_n) = r$ for all $n \geq n_0$, so ΔX_{n+1} is distributed according to μ^r for all $n \geq n_0$. In particular, for any $i \in [d]$, the ΔX_{n+1}^i's behave like i.i.d. random variables with mean q_i^r and finite variance, hence $X_n^i/n \to q_i^r$ (see also Proposition 2.14). Therefore, if r ranks i higher than j, for the ranking to remain equal to r, we must have that $q_i^r > q_j^r$. Note that in particular $q_i^r = q_j^r$ is not enough. An exception to the latter is if $\Delta X_{n+1}^i = \Delta X_{n+1}^j$ a.s. (equivalently $\mu^r(x_i \neq x_j) = 0$), so that the two components change in exactly the same way. On the other hand, if i and j are ranked equally, then we must necessarily have $\Delta X_{n+1}^i = \Delta X_{n+1}^j$ a.s. for the ranking not to change. The above theorem says that these conditions are not only necessary, but are also sufficient for the ranking to have a positive probability to remain the same for all $n \geq n_0$.

Theorem 2.10 characterizes all terminal rankings by an easy to check criterion. Note that it does not require Assumption 2.5. However, without that assumption $rk(X_n)$ is not guaranteed to converge (see Theorem 2.7). Also note that even if we know that r is terminal, we don't know whether $\mathbb{P}_\nu(rk(X_n) \to r) > 0$ for a *specific* initial distribution ν. This is the topic of Section 2.6 (see in particular Proposition 2.20).

If we can exclude the case $\mu^r(x_i \neq x_j) = 0$, then we get the following simplification of Theorem 2.10.

COROLLARY 2.11. *Suppose that $\mu^r(x_i \neq x_j) > 0$ for all $i, j \in [d]$ and all $r \in \mathcal{R}$. Then, a ranking r is terminal if and only if it is a strict ranking and*

$$q_{r^{-1}(1)}^r > q_{r^{-1}(2)}^r > \cdots > q_{r^{-1}(d)}^r, \tag{30}$$

where r^{-1} denotes the inverse of r.

PROOF. The case $\mu^r(x_i \neq x_j) = 0$ is excluded by assumption, so by Theorem 2.10 a ranking is terminal if and only if for any i, j with $r(i) < r(j)$ we have $q_i^r > q_j^r$, or equivalently, if for any $i < j$, $q_{r^{-1}(i)}^r > q_{r^{-1}(j)}^r$. □

For the proof of Theorem 2.10 we are going to need a construction that will be used again later on. Specifically, given a ranking-based process X_n and a ranking $r \in \mathcal{R}$, we construct another process Y_n that is identical to X_n up to some point $n_0 \in \mathbb{N}$, and it has i.i.d. increments afterwards with distribution μ^r. It has the additional property that it remains equal to X_n as long as their common ranking remains equal to r. The benefit of this is that we can work with the simpler process Y_n and then transfer results to X_n.

LEMMA 2.12. *For any $r \in \mathcal{R}$ and any $n_0 \in \mathbb{N}$, there exists a process $Y_n \in \mathbb{R}^d$ and a filtration $\mathcal{G}_n \supset \mathcal{F}_n$ such that*:

i. *$Y_n = X_n$ for all $n \leq n_0$.*

ii. *$\{\Delta Y_n\}_{n \geq n_0+1}$ is a sequence of i.i.d. random vectors with distribution μ^r. Moreover, $Y_n \in \mathcal{G}_n$ for each $n \in \mathbb{N}$, and $\Delta Y_{n+1} \perp\!\!\!\perp \mathcal{G}_n$ for each $n \geq n_0$.*

iii. *For any $n > n_0$, on both $\bigcap_{k=n_0}^{n-1}\{rk(X_k) = r\}$ and $\bigcap_{k=n_0}^{n-1}\{rk(Y_k) = r\}$ we have $Y_k = X_k$ a.s. for $k = 0, 1, \dots, n$. In particular, on both $\bigcap_{k=n_0}^{\infty}\{rk(X_k) = r\}$ and $\bigcap_{k=n_0}^{\infty}\{rk(Y_k) = r\}$ we have $Y_n = X_n$ a.s. for all $n \in \mathbb{N}$.*

A process Y_n that satisfies the above properties (for some filtration $\mathcal{G}_n$) will be said to (r, n_0)*-mimic* X_n.

PROOF. Let $\{U_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ with distribution μ^r, independent of $\mathcal{F}_\infty$, and let $\mathcal{G}_n = \sigma(U_1, \dots, U_n, \mathcal{F}_n)$ and $\tau = \min\{n \geq n_0 : rk(X_n) \neq r\}$. Define

$$Y_n = X_{\tau\wedge n} + \sum_{m=\tau+1}^{n} U_m \tag{31}$$

with the convention that the sum is 0 if $\tau + 1 > n$. Property (i) follows from the fact that $\tau \geq n_0$. For property (ii), note that since $\{\tau \leq n\}$ is $\mathcal{F}_n$-measurable, we get $Y_n \in \mathcal{G}_n$. Moreover,

$$\Delta Y_{n+1} = \mathbf{1}_{\tau>n} \cdot \Delta X_{n+1} + \mathbf{1}_{\tau\leq n} \cdot U_{n+1}. \tag{32}$$

In particular, for any $n \geq n_0$ and any $S \in \mathcal{B}(\mathbb{R}^d)$, on $\{\tau > n\}$ we have

$$\mathbb{P}_\nu(\Delta Y_{n+1} \in S \mid \mathcal{G}_n) = \mathbb{P}_\nu(\Delta X_{n+1} \in S \mid \mathcal{G}_n) = \mu^{rk(X_n)}(S) = \mu^r(S) \quad \text{a.s.}, \tag{33}$$

where the second equality follows from Eq. (6). Also, on $\{\tau \leq n\}$ we have

$$\mathbb{P}_\nu(\Delta Y_{n+1} \in S \mid \mathcal{G}_n) = \mathbb{P}_\nu(U_{n+1} \in S \mid \mathcal{G}_n) = \mathbb{P}_\nu(U_{n+1} \in S) = \mu^r(S) \quad \text{a.s.} \tag{34}$$

Combining the last two equations we get

$$\mathbb{P}_\nu(\Delta Y_{n+1} \in S \mid \mathcal{G}_n) = \mu^r(S) \quad \text{a.s.}, n \geq n_0, S \in \mathcal{B}(\mathbb{R}^d). \tag{35}$$

Therefore, the sequence $\{\Delta Y_{n+1}\}_{n\geq n_0}$ is i.i.d. and, for each $n \geq n_0$, ΔY_{n+1} has distribution μ^r and is independent of $\mathcal{G}_n$, which completes the proof of (ii).

For property (iii), let $m \leq n$ and note that on the set $\{Y_m \neq X_m\}$ we have $\tau < m < \infty$ and by definition $\tau \geq n_0$, $Y_\tau = X_\tau$, and $rk(Y_\tau) = rk(X_\tau) \neq r$. Therefore, the intersection of $\{Y_m \neq X_m\}$ with both $\bigcap_{k=n_0}^{n-1}\{rk(X_k) = r\}$ and $\bigcap_{k=n_0}^{n-1}\{rk(Y_k) = r\}$ is empty a.s. □

PROOF OF NECESSITY FOR THEOREM 2.10. Let r be a terminal ranking. Then, there exists some initial distribution ν and some $n_0 \in \mathbb{N}$ such that $\mathbb{P}_\nu(A) > 0$, where

$$A = \bigcap_{n=n_0}^{\infty} \{rk(X_n) = r\}. \tag{36}$$

Let Y_n (r, n_0)-mimic X_n. By Lemma 2.12(iii) we have

$$\bigcap_{n=n_0}^{\infty} \{rk(Y_n) = r\} = \bigcap_{n=n_0}^{\infty} \{rk(X_n) = r\} = A. \tag{37}$$

Fix some $i, j \in [d]$ and note that the sequence $d_n = Y_n^i - Y_n^j$, $n \geq n_0$, performs a random walk, starting at $d_{n_0} = Y_{n_0}^i - Y_{n_0}^j = X_{n_0}^i - X_{n_0}^j$, and with the step $\Delta d_{n+1} = d_{n+1} - d_n$ having the same distribution as $Z_i^r - Z_j^r$ (see Eq. (8)). In particular, for any $n \geq n_0$,

$$\mathbb{P}_\nu(\Delta d_{n+1} \neq 0) = \mathbb{P}(Z_i^r \neq Z_j^r) = \mu^r(x_i \neq x_j) \tag{38}$$

and

$$\mathbb{E}_\nu[\Delta d_{n+1}] = \mathbb{E}[Z_i^r - Z_j^r] = q_i^r - q_j^r. \tag{39}$$

If $\mu^r(x_i \neq x_j) \neq 0$, then the random walk is nontrivial, and in particular $\bigcap_{n=n_0}^{\infty}\{Y_n^i = Y_n^j\} = \bigcap_{n=n_0}^{\infty}\{d_n = 0\}$ has probability 0. If $r(i) = r(j)$, this means that $\bigcap_{n=n_0}^{\infty}\{rk(Y_n) = r\}$

has probability 0, contradicting the fact that $\mathbb{P}_\nu(A) > 0$. We conclude that if $r(i) = r(j)$, then $\mu^r(x_i \neq x_j) = 0$.

For the second assertion, assume that in addition to $\mu^r(x_i \neq x_j) \neq 0$, we also have $q_i^r \leq q_j^r$. This means that either $d_n \to -\infty$ or the random walk is recurrent. In either case, $\mathbb{P}_\nu(\bigcap_{n=n_0}^\infty \{Y_n^i > Y_n^j\}) = \mathbb{P}_\nu(\bigcap_{n=n_0}^\infty \{d_n > 0\}) = 0$. Therefore, if $r(i) < r(j)$, then $\mathbb{P}_\nu(\bigcap_{n=n_0}^\infty \{rk(Y_n) = r\}) = 0$, again contradicting the fact that $\mathbb{P}_\nu(A) > 0$. We conclude that if $r(i) < r(j)$, then either $\mu^r(x_i \neq x_j) = 0$ or $q_i^r > q_j^r$. □

For the sufficiency part of Theorem 2.10, we are going to prove the following more general result.

LEMMA 2.13 (Terminal rankings sufficient condition). *Let $r \in \mathcal{R}$ and define $A = \{(i,j) \in [d] \times [d] : r(i) < r(j)\}$ and $A' = \{(i,j) \in [d] \times [d] : r(i) = r(j)\}$. Assume that for any $(i,j) \in A$, either $q_i^r > q_j^r$ or $\mu^r(x_i \neq x_j) = 0$, and that for any $(i,j) \in A'$, $\mu^r(x_i \neq x_j) = 0$. Then, there exists some $M > 0$, such that for any initial distribution ν and any $n_0 \in \mathbb{N}$ that satisfy*

$$\mathbb{P}_\nu\Bigg(\bigcap_{(i,j)\in A} \{X_{n_0}^i > X_{n_0}^j + M\}, \bigcap_{(i,j)\in A'} \{X_{n_0}^i = X_{n_0}^j\}\Bigg) > 0, \tag{40}$$

we have

$$\mathbb{P}_\nu\Bigg(\bigcap_{n=n_0}^{\infty} \{rk(X_n) = r\}\Bigg) > 0. \tag{41}$$

PROOF. Consider the collection of random variables $\{U_n^i\}_{n\in\mathbb{N}}^{i\in[d]}$, independent of $\mathcal{F}_\infty$, such that for each i, $\{U_n^i\}_n$ are i.i.d. with distribution equal to that of Z_i^r. For any pair $(i,j) \in A$, $U_n^i - U_n^j$ is either identically zero (if $\mu^r(x_i \neq x_j) = 0$) or it has positive mean and finite variance (if $q_i^r - q_j^r > 0$). In the latter case, by the strong law of large numbers, $\sum_{m=1}^n (U_m^i - U_m^j) \to \infty$ a.s. as $n \to \infty$. Therefore, in both cases, $\sum_{m=1}^n (U_m^i - U_m^j)$ is bounded below a.s. Hence, there exists some $M > 0$, such that for any pair $(i,j) \in A$,

$$\mathbb{P}\Bigg(\min_{n\in\mathbb{N}} \sum_{m=1}^{n} (U_m^i - U_m^j) \leq -M\Bigg) < \frac{1}{d^2}. \tag{42}$$

Now let the initial distribution ν and $n_0 \in \mathbb{N}$ satisfy Eq. (40) for the value of M specified in Eq. (42), that is, $\mathbb{P}_\nu(D) > 0$, where

$$D = \bigcap_{(i,j)\in A} \{X_{n_0}^i > X_{n_0}^j + M\} \cap \bigcap_{(i,j)\in A'} \{X_{n_0}^i = X_{n_0}^j\} \tag{43}$$

We want to show that $\mathbb{P}_\nu(\bigcap_{n=n_0}^\infty rk(X_n) = r) > 0$. Let Y_n be a process that (r, n_0)-mimics X_n (see Lemma 2.12) and note that Eq. (43) implies

$$D = \bigcap_{(i,j)\in A} \{Y_{n_0}^i > Y_{n_0}^j + M\} \cap \bigcap_{(i,j)\in A'} \{Y_{n_0}^i = Y_{n_0}^j\}. \tag{44}$$

For any $(i,j) \in A'$, $n \geq n_0$, we have $\mathbb{P}_\nu(\Delta Y_{n+1}^i \neq \Delta Y_{n+1}^j) = \mu^r(x_i \neq x_j) = 0$ by Lemma 2.12(ii) and by assumption, hence on the set D we have

$$Y_n^i = Y_n^j \quad \text{for all } (i,j) \in A', n \geq n_0. \tag{45}$$

We further define

$$
\begin{aligned}
B_{(i,j)} &= \bigcap_{n\ge n_0} \{(Y_n^i - Y_n^j) - (Y_{n_0}^i - Y_{n_0}^j) > -M\}, \quad (i,j)\in A,\\
B &= \bigcap_{(i,j)\in A} B_{(i,j)}.
\end{aligned} \tag{46}
$$

Note that on the set $D\cap B_{(i,j)}$ we have $Y_n^i > Y_n^j$ for all $n\ge n_0$ and any $(i,j)\in A$. Combining this with Eq. (45), we get that on the set $D\cap B$, it holds that $rk(Y_n)=r$ for all $n\ge n_0$, which implies $rk(X_n)=r$ for all $n\ge n_0$ (Lemma 2.12(iii)). It is therefore enough to show that $\mathbb{P}_\nu(D\cap B)>0$.

Note that for each $(i,j)\in A$, $\{(Y_{n+n_0}^i - Y_{n+n_0}^j)-(Y_{n_0}^i - Y_{n_0}^j)\}_{n\in\mathbb{N}}$ has the same distribution as $\{\sum_{m=1}^n (U_m^i - U_m^j)\}_{n\in\mathbb{N}}$, therefore Eq. (42) implies that $\mathbb{P}_\nu(B_{(i,j)}) > 1-1/d^2$, and since $\mathrm{card}(A)<d^2$, we get $\mathbb{P}_\nu(B)>0$. By assumption we also have $\mathbb{P}_\nu(D)>0$. Finally observe that by Lemma 2.12(ii), $D\in\mathcal{G}_{n_0}$ and $B\perp\!\!\!\perp \mathcal{G}_{n_0}$, hence $\mathbb{P}_\nu(D\cap B)=\mathbb{P}_\nu(D)\cdot\mathbb{P}_\nu(B)>0$, which completes the proof. □

PROOF OF SUFFICIENCY FOR THEOREM 2.10. By assumption r satisfies the conditions of Lemma 2.13. Let M be as in that lemma and define the initial distribution ν as follows: $X_0^i=(d-r(i))\cdot(M+1)$ a.s. Then, $r(i)=r(j)$ implies $X_0^i = X_0^j$ a.s., while $r(i)<r(j)$ implies $X_0^i > X_0^j + M$ a.s. That is, ν satisfies Eq. (40) with $n_0=0$, hence $\mathbb{P}_\nu(\bigcap_{n=0}^\infty\{rk(X_n)=r\})>0$, in particular r is terminal. □

2.5. *Limit theorems for X_n.* In this section we prove a number of results regarding the long-term behavior of the process X_n, if we know that $rk(X_n)$ eventually stabilizes. Note that as long as $rk(X_n)$ remains constant and equal to r, X_n behaves like a regular random walk with increment distribution μ^r. We thus expect the standard limit laws to apply. One has to be careful, however, because the distribution of the paths of X_n changes when conditioned on the event $\bigcap_{n=n_0}^\infty\{rk(X_n)=r\}$. Still, it turns out that in the limit X_n behaves as expected.

PROPOSITION 2.14 (Strong law of large numbers). *For any $r\in\mathcal{R}$,*

$$
\lim_{n\to\infty}\frac{X_n}{n} = q^r \text{ a.s. on the set } \Big\{\lim_{n\to\infty} rk(X_n)=r\Big\}. \tag{47}
$$

PROOF. Since by definition $\{rk(X_k)\to\infty\}=\bigcup_{n_0\in\mathbb{N}}\bigcap_{k=n_0}\{rk(X_k)=r\}$, it is enough to show that

$$
\lim_{n\to\infty}\frac{X_n}{n} = q^r \text{ a.s. on the set } \bigcap_{k=n_0}\{rk(X_k)=r\} \tag{48}
$$

for any $n_0\in\mathbb{N}$ that satisfies $\mathbb{P}_\nu(\bigcap_{k=n_0}\{rk(X_k)=r\})>0$. Fix such an n_0 and let Y_n be a process that (r,n_0)-mimics X_n (see Lemma 2.12). Since $\{\Delta Y_n\}_{n\ge n_0+1}$ is an i.i.d. sequence with mean q^r, we have by the strong law of large numbers that $Y_n/n\to q^r$ a.s. Hence, Equation (48) follows from Lemma 2.12(iii). □

We also have the following partial converse.

PROPOSITION 2.15. *If $X_n/n\to x\in\mathbb{R}^d$ and the components of x are all distinct, then $rk(X_n)\to rk(x)$ and $x=q^{rk(x)}$.*

PROOF. Let $i, j \in [d]$ and assume without loss of generality that $x_i > x_j$. Then, for large enough n, $X_n^i > X_n^j$, so $rk(X_n)$ ranks i higher than j. Since this is true for all pairs i, j, we get that for large enough n, $rk(X_n) = rk(x)$, hence $rk(X_n) \to rk(x)$. By Proposition 2.14, $X_n/n \to q^{rk(x)}$. □

Next we state some central limit theorem-type results. We denote by ξ^r a centered multivariate normal distribution on $\mathbb{R}^d$ with covariance matrix equal to that of μ^r.

PROPOSITION 2.16 (Central limit theorem - 1). *Let ν be some initial distribution, $r \in \mathcal{R}$, and $n_0 \in \mathbb{N}$, such that $\mathbb{P}_\nu(\bigcap_{k=n_0}^\infty \{rk(X_k) = r\}) > 0$. Then, for any $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\lim_{n\to\infty} \mathbb{P}_\nu\left(\frac{X_n - n\cdot q^r}{\sqrt{n}} \in A \;\middle|\; \bigcap_{k=n_0}^{\infty} \{rk(X_k) = r\}\right) = \xi^r(A). \tag{49}$$

In particular, if $\mathbb{P}_\nu(\lim_{k\to\infty} rk(X_k) = r) > 0$, then

$$\lim_{n\to\infty} \mathbb{P}_\nu\left(\frac{X_n - n\cdot q^r}{\sqrt{n}} \in A \;\middle|\; \lim_{k\to\infty} rk(X_k) = r\right) = \xi^r(A). \tag{50}$$

We postpone the proof in order to state the following corollary, which strengthens Proposition 5.1 in [24].

COROLLARY 2.17 (Central limit theorem - 2). *For any initial distribution ν, $A \in \mathcal{B}(\mathbb{R}^d)$, $n_0 \in \mathbb{N}$, and $r \in \mathcal{R}$,*

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{P}_\nu\left(\frac{X_n - n\cdot q^r}{\sqrt{n}} \in A, \bigcap_{k=n_0}^{n} \{rk(X_k) = r\}\right)\\ &\quad = \mathbb{P}_\nu\left(\bigcap_{k=n_0}^{\infty} \{rk(X_k) = r\}\right) \cdot \xi^r(A).\end{aligned} \tag{51}$$

For the sake of comparison, Proposition 5.1 in [24] gives only an exponential rate for the quantity $\mathbb{P}_\nu(\frac{|X_n - n\cdot q^r|}{\sqrt{n}} \in A, \bigcap_{k=0}^n \{rk(X_k) = r\})$, for sets A of a certain type, and under a nondegeneracy assumption on μ^r. See also [22], Corollary, for a related result.

PROOF. Since $\bigcap_{k=n_0}^n \{rk(X_k) = r\} \to \bigcap_{k=n_0}^\infty \{rk(X_k) = r\}$, we have

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{P}_\nu\left(\frac{X_n - n\cdot q^r}{\sqrt{n}} \in A, \bigcap_{k=n_0}^{n} \{rk(X_k) = r\}\right)\\ &\quad = \lim_{n\to\infty} \mathbb{P}_\nu\left(\frac{X_n - n\cdot q^r}{\sqrt{n}} \in A, \bigcap_{k=n_0}^{\infty} \{rk(X_k) = r\}\right).\end{aligned} \tag{52}$$

The result now follows from Eq. (49). □

For the proof of Proposition 2.16 we are going to need a couple of lemmas whose proofs are given in Appendix B.

LEMMA 2.18. *Let A_n, $n \in \mathbb{N}$, and A be measurable sets in a probability space, each with positive probability, and suppose that $A_n \to A$ a.s. (i.e., $\mathbb{P}((A_n \setminus A) \cup (A \setminus A_n)) \to 0$). Then, $\mathbb{P}(S \mid A_n) \to \mathbb{P}(S \mid A)$ uniformly in $S \in \mathcal{F}$.*

LEMMA 2.19. *Let $a_{m,n} \in \mathbb{R}$, $m, n \in \mathbb{N}$, and suppose that* $\lim_{m\to\infty} a_{m,n} = a_n \in \mathbb{R}$ *uniformly in* n, *and* $\lim_{n\to\infty} a_{m,n} = a \in \mathbb{R}$ *for all* $m \in \mathbb{N}$. *Then*, $\lim_{n\to\infty} a_n = a$.

PROOF OF PROPOSITION 2.16. Let Y_n be a process that (r, n_0)-mimics X_n (see Lemma 2.12). Since $\{\Delta Y_n\}_{n\geq n_0+1}$ is a sequence of i.i.d. random vectors with distribution μ^r, we have by the central limit theorem,

$$\mathbb{P}_\nu\left(\frac{Y_n - n\cdot q^r}{\sqrt{n}} \in A\right) \to \xi^r(A). \tag{53}$$

In fact, Eq. (53) can be strengthened: since $\Delta Y_{n+1} \perp\!\!\!\perp \mathcal{F}_m$ for any $n \geq m \geq n_0$, we have that for any $m \geq n_0$ and any $x \in \mathbb{R}$,

$$\mathbb{P}_\nu\left(\frac{Y_n - n\cdot q^r}{\sqrt{n}} \in A \,\middle|\, \bigcap_{k=n_0}^{m}\{rk(X_k) = r\}\right) \overset{n\to\infty}{\to} \xi^r(A). \tag{54}$$

Furthermore, since $\bigcap_{k=n_0}^m \{rk(X_k) = r\} \overset{m\to\infty}{\longrightarrow} \bigcap_{k=n_0}^\infty \{rk(X_k) = r\}$, Lemma 2.18 implies that

$$\begin{aligned} &\mathbb{P}_\nu\left(\frac{Y_n - n\cdot q^r}{\sqrt{n}} \in A \,\middle|\, \bigcap_{k=n_0}^{m}\{rk(X_k) = r\right) \\ &\overset{m\to\infty}{\to} \mathbb{P}_\nu\left(\frac{Y_n - n\cdot q^r}{\sqrt{n}} \in A \,\middle|\, \bigcap_{k=n_0}^{\infty}\{rk(X_k) = r\}\right) \end{aligned} \tag{55}$$

uniformly in n. Combining this with Eq. (54) and Lemma 2.19 we get

$$\mathbb{P}_\nu\left(\frac{Y_n - n\cdot q^r}{\sqrt{n}} \in A \,\middle|\, \bigcap_{k=n_0}^{\infty}\{rk(X_k) = r\}\right) \to \xi^r(A), \tag{56}$$

as $n \to \infty$. By Lemma 2.12(iii), this is equivalent to Eq. (49). Equation (50) now follows from Eq. (49) and Definition 2.2. □

2.6. *Terminal rankings and initial distributions*. Although Theorem 2.10 gives the possible limits of the ranking for a ranking-based process in principle, it doesn't say for which pairs of initial distributions ν and terminal rankings r we have $\mathbb{P}_\nu(rk(X_n) \to r) > 0$. To see that for the same terminal ranking r it is possible to have $\mathbb{P}_\nu(rk(X_n) \to r) > 0$ for some initial distributions ν and not for others, consider a deterministic system with $d = 2$ and such that

$$\begin{aligned} &\mathbb{P}\big(\Delta X_{n+1} = (1, 0) \mid rk(X_n) = \mathrm{id}_2\big) = 1 \quad \text{and} \\ &\mathbb{P}\big(\Delta X_{n+1} = (0, 1) \mid rk(X_n) \neq \mathrm{id}_2\big) = 1, \end{aligned} \tag{57}$$

where id_2 is the identity function on the set $\{1, 2\}$. In words, if $X_n^1 > X_n^2$, then X_n^1 increases by 1 and X_n^2 remains constant. If $X_n^1 \leq X_n^2$, then X_n^2 increases by 1 and X_n^1 remains constant. Clearly, if we start at $X_0 = (0, 0)$, $rk(X_n) \to r$ a.s., where $r(1) = 2$, $r(2) = 1$, while if we start at $X_0 = (1, 0)$, $rk(X_n) \to \mathrm{id}_2$ a.s.

From the above example it might seem that the only reason that a strict ranking r satisfying $q^r_{r^{-1}(1)} > q^r_{r^{-1}(2)} > \cdots > q^r_{r^{-1}(d)}$ might fail to satisfy $\mathbb{P}_\nu(rk(X_n) \to r) > 0$ is that it is not reachable from the given initial distribution, in the sense that $\mathbb{P}_\nu(\bigcup_{n=1}^\infty \{rk(X_n) = r\}) = 0$. However, this is not the only case. For example, let $d = 3$, and suppose that

$$\begin{aligned} &\mathbb{P}\big(\Delta X_{n+1} = (5, -2, 0) \mid rk(X_n) = \mathrm{id}_3\big) = 1/2, \\ &\mathbb{P}\big(\Delta X_{n+1} = (-3, 3, 0) \mid rk(X_n) = \mathrm{id}_3\big) = 1/2 \quad \text{and} \\ &\mathbb{P}\big(\Delta X_{n+1} = (0, 0, 1) \mid rk(X_n) \neq \mathrm{id}_3\big) = 1. \end{aligned} \tag{58}$$

In words, whenever $X_n^1 > X_n^2 > X_n^3$, with probability $1/2$ the first component will increase by 5 and the second will decrease by 2, and also with probability $1/2$ the first component will decrease by 3 and the second will increase by 3, while the last component remains constant a.s. For any other ranking, the third component increases by 1 and the rest remain constant a.s.

Now suppose we begin at $X_0 = (2, 1, 0)$ a.s., so that $rk(X_0) = \mathrm{id}_3$ a.s. Clearly, after the first step the ranking will necessarily change and after that $\Delta X_n = (0, 0, 1)$ deterministically, so that for large n we will have either $X_n^3 > X_n^1 > X_n^2$ or $X_n^3 > X_n^2 > X_n^1$. We see that despite the fact that $rk(X_0) = \mathrm{id}_3$ and $q_1^{\mathrm{id}_3} > q_2^{\mathrm{id}_3} > q_3^{\mathrm{id}_3}$, for the specific initial distribution ν we get $\mathbb{P}_\nu(rk(X_n) \to \mathrm{id}_3) = 0$.

The above examples might seem discouraging. We have the following positive result, which states that such situations do not arise if a certain condition is satisfied. The condition roughly says that, no matter the ranking, there is some positive probability for any component to increase faster than the rest, and for the increments of the rest to follow any given nonstrict order.

PROPOSITION 2.20. *Suppose that for any permutation σ of $[d]$ and any $r' \in \mathcal{R}$,*

$$\mu^{r'}(x_{\sigma_1} > x_{\sigma_2} \geq x_{\sigma_3} \geq \cdots \geq x_{\sigma_d}) > 0. \tag{59}$$

Then, for any initial distribution ν and any terminal ranking r,

$$\mathbb{P}_\nu\big(rk(X_n) \to r\big) > 0. \tag{60}$$

REMARK 2.21. The condition of Proposition 2.20 implies that $\mu^r(x_i \neq x_j) > 0$ for all $i, j \in [d]$ and $r \in \mathcal{R}$, which in particular implies the condition of Corollary 2.11. Consequently, under the condition of Proposition 2.20, only strict rankings may be terminal.

EXAMPLE 2.22. In a ranking-based Pólya urn, with probability one, exactly one of the components of ΔX_{n+1} is 1 and the rest are 0 (see also Section 3.2). Therefore, Eq. (59) is satisfied if and only if for any ranking there is positive probability of adding a ball of any given color. In Example 2.6, this is equivalent to either $\lambda_d > 0$ or $a_i > 0$ for all $i \in [d]$.

More generally, for processes that change one component at a time, Eq. (59) is satisfied if and only if, for any ranking, every component has nonzero probability of increasing.

PROOF OF PROPOSITION 2.20. By Remark 2.21 we may assume that r is a strict ranking. Also, by renaming the indices, we may assume that r is the identity map on $[d]$, that is, $r(i) = i$ for all $i \in [d]$. Let $M > 0$ be as in Lemma 2.13 and define

$$\begin{aligned} C_n^j &= \{X_n^j > X_n^{j+1} + M\}, \quad j \in [d-1], n \in \mathbb{N}, \\ B_n^i &= \bigcap_{j=i}^{d-1} C_n^j, \quad i \in [d-1], n \in \mathbb{N}, \end{aligned} \tag{61}$$

and $B_n^d = \Omega$, $n \in \mathbb{N}$. By Lemma 2.13, it is enough to show that $\mathbb{P}_\nu(\bigcup_{n\in\mathbb{N}} B_n^1) > 0$. We will use (backwards) induction on i to show that $\mathbb{P}_\nu(\bigcup_{n\in\mathbb{N}} B_n^i) > 0$ for all $i \leq d$, with the base case $i = d$ being trivially true. Suppose then that $\mathbb{P}_\nu(\bigcup_{n\in\mathbb{N}} B_n^{i+1}) > 0$ or, equivalently, that there exists some $n \in \mathbb{N}$ such that $\mathbb{P}_\nu(B_n^{i+1}) > 0$. Fix such an n. From Eq. (59) and continuity, there exists some $\epsilon > 0$ such that $\mu^{r'}(A_i) > 0$ for all $r' \in \mathcal{R}$, where

$$A_i = \{x \in \mathbb{R}^d : x_i - \epsilon \geq x_{i+1} \geq x_{i+2} \geq \cdots \geq x_d\}. \tag{62}$$

For any $j \in [d-1]$ and $k \in \mathbb{N}$, define

$$D^j_{m,k} = \{X^j_{m+k} - X^j_m \geq X^{j+1}_{m+k} - X^{j+1}_m\} \tag{63}$$

and

$$D^j_{m,k}(\epsilon) = \{X^j_{m+k} - X^j_m - \epsilon \geq X^{j+1}_{m+k} - X^{j+1}_m\}. \tag{64}$$

In particular, $D^j_{m,1} = \{\Delta X^j_{m+1} \geq \Delta X^{j+1}_{m+1}\}$, and similarly for $D^j_{m,1}(\epsilon)$. Therefore, from Eq. (6) we get that for any $m \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}_\nu\left(D^i_{m,1}(\epsilon), \bigcap_{j=i+1}^{d-1} D^j_{m,1} \,\middle|\, \mathcal{F}_m\right) &= \mu^{rk(X_m)}(A_i)\\ &\geq \min_{r' \in \mathcal{R}} \mu^{r'}(A_i) > 0 \quad \text{a.s.}\end{aligned} \tag{65}$$

Let $K \in \mathbb{N}$ be such that

$$\mathbb{P}_\nu(F_K \cap B^{i+1}_n) > 0, \tag{66}$$

where

$$F_K = \{X^i_n - X^{i+1}_n > M - K\epsilon\}. \tag{67}$$

This is always possible, since $\bigcup_{K\in\mathbb{N}} F_K = \Omega$ and $\mathbb{P}_\nu(B^{i+1}_n) > 0$ by assumption. Applying Eq. (65) for $m = n, n+1, \dots, n+(K-1)$ and using $D^j_{m,1} \in \mathcal{F}_{m+1}$, it easily follows that

$$\mathbb{P}_\nu\left(D^i_{n,K}(K\epsilon), \bigcap_{j=i+1}^{d-1} D^j_{n,K} \,\middle|\, \mathcal{F}_n\right) > 0 \quad \text{a.s.} \tag{68}$$

Observe that

$$\begin{aligned} C^j_n \cap D^j_{n,K} &\subset C^j_{n+K}, \quad j = i+1, \dots, d-1\\ F_K \cap D^i_{n,K}(K\epsilon) &\subset C^i_{n+K}.\end{aligned} \tag{69}$$

Combining these two relations and the definition of B^i_n we get

$$\begin{aligned}\mathbb{P}_\nu(B^i_{n+K}) &= \mathbb{P}_\nu\left(\bigcap_{j=i}^{d-1} C^j_{n+K}\right)\\ &\geq \mathbb{P}_\nu\left(\bigcap_{j=i+1}^{d-1}(C^j_n \cap D^j_{n,K}), F_K, D^i_{n,K}(K\epsilon)\right)\\ &= \mathbb{P}_\nu\left(F_K, B^{i+1}_n, D^i_{n,K}(K\epsilon), \bigcap_{j=i+1}^{d-1} D^j_{n,K}\right)\\ &= \mathbb{P}_\nu(F_K, B^{i+1}_n) \cdot \mathbb{P}_\nu\left(D^i_{n,K}(K\epsilon), \bigcap_{j=i+1}^{d-1} D^j_{n,K} \,\middle|\, F_K, B^{i+1}_n\right)\\ &> 0,\end{aligned} \tag{70}$$

with the last line following from Eqs. (66) and (68). This concludes the inductive proof. □

3. Applications. A crucial setting where we expect the ranking-based reinforcement assumption (Assumption 2.5) to hold is when modeling the dynamic interaction between people's choices and algorithms in online interfaces such as search engines, online marketplaces, newspapers, and discussion forums. The algorithms implemented in these venues rank the available content for the users to facilitate their access to information [42]. People, in return, pay more attention to and interact more (download, click, buy) with content that appears higher on ranked lists [25, 33]. In the past two decades, several behavioral models have been proposed in economics, management, marketing, and computer science to capture the effect of rank on people's choices [3, 9]. In our first application (Section 3.1), we describe the dynamics produced by a commonly used ranking algorithm and a standard behavioral model in computer science. In a separate paper [3] we show how a similar framework can be used to study dynamics under a variety of behavioral models from the economics literature.

In Section 3.2 we focus on a class of models that can be described as ranking-based Pólya urns and that lie on the intersection of our framework and that of [5, 27]. After specializing our results to this type of models, we compare them to the previously known results.

3.1. *Ranking items in online interfaces.* One of the most fundamental and commonly employed ranking algorithms places the options on the screen according to their popularity, that is the number of clicks, sales, citations, or upvotes that different options have obtained so far. The rank-by-popularity algorithm is very simple to implement, and many popular websites have relied on it in the past or use some version of it at present. (For example, Reddit used to order comments by the number of upvotes, Google scholar used to order articles by the number of citations—and still offers that possibility when looking at a profile—, Amazon offers the possibility to order options by the number of reviews, Goodreads orders user comments by the number of likes, etc.) We will use a staple computer science model for the probability of clicking on a link, called the position-based model [9], p. 10. We note that although we will refer to clicks, the model can also be used to describe downloads and citations of papers, purchases of products, likes of comments, etc.

In the position-based model, a link is first examined by the user and then clicked if its content is considered to be relevant. This can be stated as

$$C_n^i = E_n^i \cap D_n^i, \tag{71}$$

where

$$\begin{aligned} E_n^i &= \{n\text{-th user examines link } i\}, \\ D_n^i &= \{\text{link } i \text{ is relevant to the } n\text{th user}\}, \\ C_n^i &= \{n\text{th user clicks on link } i\}. \end{aligned} \tag{72}$$

We are interested in the vector $X_n = (X_n^1, \dots, X_n^d)$, where X_n^i is the number of users that have clicked on link i, up to the nth user. Clearly, we have $\Delta X_{n+1}^i = 1$ if C_{n+1}^i occurs, and $\Delta X_{n+1}^i = 0$ otherwise.

The probability that a link is examined depends only on the position where it appears, and typically decreases for later positions. Assuming that results appear according to the rank-by-popularity algorithm, that is, by descending number of clicks so far (and randomly breaking ties), this factor depends only on (a) the current rank of result i with respect to the number of clicks and (b) the number of links that are ranked equally with it. For our purposes, we may allow the probability that a link is examined to depend on the full ranking (i.e., how all of the links are ranked), so we will denote

$$a_i^r = \mathbb{P}(E_{n+1}^i \mid rk(X_n) = r). \tag{73}$$

The expression on the right hand side makes sense whenever $\{rk(X_n) = r\}$ has positive probability. We will be making this assumption below whenever similar expressions appear, without further mention.

We also assume that links that appear higher are more likely to be examined, that is, if $r(i) < r(j)$, then $a_i^r > a_j^r$. Finally, we assume that $a_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$, so that there is always positive probability of clicking on any of the links.

The probability of link i being relevant to the user depends only on the link itself, that is, D_{n+1}^i is independent of $\{E_{n+1}^j\}_{j\in[d]}, \{D_{n+1}^j\}_{j\neq i}$, and $rk(X_n)$. We denote

$$u_i = \mathbb{P}(D_{n+1}^i) \tag{74}$$

and assume that $u_i \in (0, 1)$.

The number u_i can be considered a measure of objective quality of the link (not necessarily known to the ranking algorithm). Combining Eqs. (71), (73), and (74) we get

$$\begin{aligned} q_i^r &= \mathbb{P}(\Delta X_{n+1}^i = 1 \mid rk(X_n) = r) \\ &= \mathbb{P}(C_{n+1}^i \mid rk(X_n) = r) \\ &= \mathbb{P}(E_{n+1}^i \cap D_{n+1}^i \mid rk(X_n) = r) \\ &= \mathbb{P}(E_{n+1}^i \mid rk(X_n) = r) \cdot \mathbb{P}(D_{n+1}^i \mid E_{n+1}^i, rk(X_n) = r) \\ &= \mathbb{P}(E_{n+1}^i \mid rk(X_n) = r) \cdot \mathbb{P}(D_{n+1}^i) \\ &= a_i^r \cdot u_i. \end{aligned} \tag{75}$$

Since we are assuming that $a_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$, and $u_i \in (0, 1)$ for all $i \in [d]$, we also have $q_i^r \in (0, 1)$ for all $i \in [d], r \in \mathcal{R}$. Moreover, using the fact that the D_{n+1}^j's are independent of everything else and $\mathbb{P}(D_{n+1}^j) < 1$ for all $j \in [d]$, we get

$$\begin{aligned} &\mathbb{P}(\Delta X_{n+1}^i = 1, \Delta X_{n+1}^j = 0 \text{ for all } j \neq i \mid rk(X_n) = r) \\ &\quad \geq q_i^r \cdot \prod_{j\neq i} \mathbb{P}((D_{n+1}^j)^c) > 0, \end{aligned} \tag{76}$$

for any $i \in d, r \in \mathcal{R}$.

Now let $i \neq j$ and suppose (without loss of generality) that $u_i \geq u_j$. Recall that, by assumption, for any ranking r that ranks i higher than j, we have $a_i^r > a_j^r$, hence Eq. (75) gives $q_i^r > q_j^r$. That is, i quasi-dominates j. By Eq. (76), $\mu^r(x_i \neq x_j) > 0$ for all $r \in \mathcal{R}$, therefore i actually dominates j, hence Assumption 2.5 is satisfied.

Theorem 2.7 now says that $rk(X_n)$ converges a.s. Equation (76) also implies that the conditions of Corollary 2.11 and Proposition 2.20 are satisfied, therefore the possible limits for $rk(X_n)$ are those strict rankings r for which $q_{r^{-1}(1)}^r > \cdots > q_{r^{-1}(d)}^r$. Note that in general there will be more than one ranking r satisfying this condition, especially if the effect of the position is strong (a_i^r decreases quickly with the position of i in the ranking r). Thus, it is likely that links of smaller objective quality u_i will end up being ranked higher in the long-term (thus getting more clicks) than links of higher quality. This is an important consequence, because it implies that in general people will be directed towards links that are less likely to be relevant to them, and it reveals an inherent drawback of algorithms that rank results by popularity.

Our framework can be generalized to other models of user behavior. For example, we could allow the probability a_i^r of examining a link to depend on the ranking in an arbitrary way (subject to Assumption 2.5 being satisfied). In particular the model applies to cases where the position of other links also affects the probability of examining a link at a certain position, such as in the cascade model in computer science [13] or satisficing models in economics [8]. More generally, the assumption that links are first examined and then independently judged to be relevant or not can be discarded altogether; it is enough to require that the links possess some objective quality u_i, and whenever link i is ranked higher than j and $u_i > u_j$, it is more likely for i to be clicked (i.e., $q_i^r > q_j^r$). For example, the q_i^r's can be described by a multi-attribute utility model [36], where the link position is one of the attributes and u_i is a summary of the rest of the attributes.

In a similar vein, we can relax assumptions related to the ranking algorithm. For instance, more sophisticated ranking algorithms may not rank the links based on their number of clicks only, but according to some calculated score that takes into account several other features [42, 48]. The conceptual framework we developed in this section still applies, as long as the popularity ranking is taken into account in calculating the score. Further, recent algorithmic approaches estimate the objective utility or relevance u_i of different items by debiasing the number of clicks from attention imbalances [1, 34]. Even for these algorithms, however, ranking-based rich-get-richer dynamics can be at play if a link's actual or perceived utility for the users depends on the object's popularity [4, 47]. For example, when ranking social networking applications, the rank may convey information about their utility, therefore some form of advantage may persist even when correcting for attention disparities.

3.2. *Ranking-based Pólya urns and urn functions*. Several of the models that have been used to study the dynamics of popularity rankings in online interfaces can be expressed as ranking-based Pólya urns [11, 21, 25], that is ranking-based processes $X_n \in \mathbb{R}^d$ where $\Delta X_n \in \{0,1\}^d$ and $\sum_{i=1}^d \Delta X_n^i = 1$ a.s. Note that in this case we have

$$q_i^{rk(X_n)} = \mathbb{E}[\Delta X_{n+1}^i \mid rk(X_n)] = \mathbb{P}(\Delta X_{n+1}^i = 1 \mid rk(X_n)), \tag{77}$$

that is, q_i^r is the probability of adding a ball of color i, when the ranking is r. Such models are special cases of both our framework and that of [5, 27]. In this section we specialize our results to this class of models and compare them to the previously known results.

In references [5, 27], the results are stated in terms of the fixed points of the urn function. The urn function $f : \Delta^{d-1} \to \Delta^{d-1}$, where

$$\Delta^{d-1} := \left\{ x \in [0,1]^d, \sum_i x_i = 1 \right\} \tag{78}$$

is the standard $(d-1)$-dimensional simplex, takes as argument the vector of proportions of balls of each color, and its ith component f_i gives the probability of the next ball being of color i. For a ranking-based urn, $f(x)$ must be constant in regions of constant ranking, that is, its value may only depend on $rk(x)$. With our notation we have

$$f_i(x) = \mathbb{P}(\Delta X_n^i = 1 \mid rk(X_n) = rk(x)) = q_i^{rk(x)}. \tag{79}$$

The next proposition uses our results from Section 2 to relate the fixed points of f with the limiting behavior of X_n.

PROPOSITION 3.1. *Consider a ranking-based Pólya urn with urn function f and let A be the set of fixed points of f whose coordinates are all distinct, that is*,

$$A = \{x \in \Delta^{d-1} : f(x) = x, x_i \neq x_j \text{ for all } i \neq j\}. \tag{80}$$

Then:

1. *For any* $x \in A$, *there is some* ν *such that* $\mathbb{P}_\nu(X_n/n \to x) > 0$. *If* $q_i^r > 0$ *for all* $i \in [d], r \in \mathcal{R}$, *then* $\mathbb{P}_\nu(X_n/n \to x) > 0$ *holds for all* ν.

2. *If* $q_i^r > 0$ *for all* $i \in [d], r \in \mathcal{R}$ *and furthermore Assumption* 2.5 *is satisfied*, *then for any initial distribution* ν, $\lim_{n\to\infty} X_n/n \in A$ *a.s.* (*in particular* X_n/n *converges a.s.*).

3. *Conditioned on* $\lim_{n\to\infty} X_n/n = x \in A$, $\frac{X_n - n\cdot x}{\sqrt{n}}$ *converges to a centered multivariate normal distribution* ξ_σ *with covariance matrix* (σ_{ij}), *where* $\sigma_{ii} = x_i(1-x_i)$ *and* $\sigma_{ij} = -x_i x_j$, $i \neq j$. *More precisely*, *for any initial distribution* ν, $x \in A$, *and* $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{P}_\nu\left(\frac{X_n - n\cdot x}{\sqrt{n}} \in B \,\middle|\, \lim_{k\to\infty} \frac{X_k}{k} = x\right) \to \xi_\sigma(B) \tag{81}$$

whenever $\mathbb{P}_\nu(\lim_{k\to\infty} \frac{X_k}{k} = x) > 0$.

PROOF.

1. Let $x = (x_1, \dots, x_d) \in A$ and denote $r = rk(x)$, so that $x_{r^{-1}(1)} > \cdots > x_{r^{-1}(d)}$. Then, $f(x) = q^r$ (Eq. (79)). Since x is a fixed point of f, we get $q^r = x$, hence also $q^r_{r^{-1}(1)} > \cdots > q^r_{r^{-1}(d)}$. By Theorem 2.10 r is terminal, so $\mathbb{P}_\nu(rk(X_n) \to r) > 0$ for some initial distribution ν. By Proposition 2.14, $\mathbb{P}_\nu(X_n/n \to x) = \mathbb{P}_\nu(X_n/n \to q^r) \geq \mathbb{P}_\nu(rk(X_n) \to r) > 0$. If $q_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$, then the condition of Proposition 2.20 is satisfied, therefore r being terminal implies $\mathbb{P}_\nu(rk(X_n) \to r) > 0$ for any initial distribution ν.

2. By Theorem 2.7 $rk(X_n)$ converges a.s. and by Corollary 2.11 the limit R has to be a strict ranking, in particular $q_i^R \neq q_j^R$ for all $i \neq j$ a.s. By Proposition 2.14 $X_n^i/n \to q^R$ and by Proposition 2.15 $q^R = q^{rk(q^R)}$, which is a fixed point of f by Eq. (79), thus $q^R \in A$.

3. Denote $r = rk(x)$. By Propositions 2.14 and 2.15, $x = q^r$ and

$$\left\{\lim_{k\to\infty} X_k/k = x\right\} = \left\{\lim_{k\to\infty} rk(X_k) = r\right\} \quad \text{a.s.} \tag{82}$$

Hence, by Proposition 2.16,

$$\mathbb{P}_\nu\left(\frac{X_n - n\cdot x}{\sqrt{n}} \in B \,\middle|\, \lim_{k\to\infty} \frac{X_k}{k} = x\right) \to \xi^r(B) \quad \text{a.s.}, \tag{83}$$

where ξ^r is a centered multivariate normal distribution with covariance matrix equal to that of μ^r. The result follows once we recall that each component of μ^r describes a Bernoulli random variable with mean $q_i^r = x_i$ and that no two of these components can be positive simultaneously. □

We now compare our results to the ones that appear in [5, 27]. We are going to restrict ourselves to ranking-based Pólya urns with the urn function being constant in $n \in \mathbb{N}$ (in [5] the urn function is allowed to be a function of n).

Part 1 of Proposition 3.1, in particular the case $q_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$, agrees with Theorem 5.1 in [5]. In that theorem, the authors show that X_n/n has positive probability of converging to any point $\theta \in \Delta^{d-1}$ that is a stable fixed point of f, in the sense that $f(\theta) = \theta$ and there is a neighborhood U of θ and a positive-definite matrix C such that

$$\langle C(x - f(x)), x - \theta)\rangle > 0 \quad \text{for all } x \in \Delta^{d-1} \cap U, x \neq \theta. \tag{84}$$

Note that in the ranking-based case, where f is piecewise constant, any fixed point θ with all coordinates being distinct (i.e., $\theta \in A$) is always stable, since then $f(x) = \theta$ identically in a neighborhood of θ, so the above condition is satisfied if we take C to be the identity matrix. The result in [5] is more general than part 1 of Proposition 3.1, because it also applies to fixed points whose coordinates are not distinct. On the other hand, there are no analogues of parts 2 and 3 of our Proposition 3.1 in [5] that apply to the ranking-based case (but Theorem 3.1 in that reference is an analogue of part 2 for continuous urn functions f).

As mentioned in the Introduction, in [27] the case $d = 2$ is studied and it is shown that X_n/n converges a.s. Note that we have shown this only if $q_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$, and Assumption 2.5 is satisfied. In [27] no such assumption is made. However, the proof there relies on properties of the real line (when $d = 2$, the process is described by X_n^1 alone, because $X_n^2 = n - X_n^1$), thus it is not obvious how to generalize to $d \in \mathbb{N}$.

Regarding the support of the limit, Theorem 4.1 in [27] is similar to part 2 of our Proposition 3.1: assuming that $d = 2$ and $q_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$, if A contains a single point, then the two results coincide. Part 2 of Proposition 3.1 also applies when A contains more than one (i.e., two) points, while Theorem 4.1 in [27] does not. On the other hand, if A is empty, which (in the case $d = 2$ with $q_i^r > 0$ for all $i \in [d], r \in \mathcal{R}$) is equivalent to Assumption 2.5 not being satisfied, part 2 of Proposition 3.1 does not apply, while Theorem 4.1 in [27] implies that $X_n/n \to 1/2$.

We emphasize that the above is a comparison of results in the special case of ranking-based Pólya urns (and in the case of [27], when $d = 2$). However, both our results and those in [5, 27] apply to more general settings: our results apply to more general (ranking-based) processes than Pólya urns, while those in [5, 27] apply to nonranking-based Pólya urns.

4. Discussion. We have developed a mathematical framework for describing systems characterized by ranking-based rich-get-richer dynamics. Specifically, we defined a ranking-based process as a discrete-time Markov process in $\mathbb{R}^d$ whose increment distributions depend only on the current ranking of the components of the process. Under a ranking-based reinforcement assumption (Assumption 2.5), we showed that the ranking converges (Theorem 2.7) and proved a strong law of large numbers and central limit theorem-type results for the process itself (Propositions 2.14 and 2.16). We also found conditions in terms of the Markov transition kernel that can be used to check whether a particular ranking is a possible limit ranking (Theorem 2.10). In some cases we were able to characterize the support of the limit of the ranking independently of the initial distribution (Proposition 2.20). We also translated our results in terms of urn functions for the special case of ranking-based Pólya urns, in order to compare them with previous results with which they partially overlap (Section 3.2). Finally, we described an application to rank-ordered web interfaces (Section 3.1).

Models of systems with rich-get-richer dynamics have been commonplace in the social, behavioral, and computer sciences, and they have been used to describe the observed dynamics in a wide variety of settings. So far, there have been two main families of such models. The first family goes back to Gibrat's law [26], which states that firms grow proportionally to their current size, and independently of the performance of their competitors. Variations of the notion of proportional growth have been applied across disciplines, for example in models of citation growth [2] and city growth [19, 20]. Models based on Gibrat's law are inherently unsuitable for capturing ranking-based dynamics, because of their assumption that growth is independent of any competitors.

The second family of rich-get-richer models builds on the notion of preferential attachment [58], which assumes that entities grow when new units "attach" to them, but these new units are more likely to attach to entities that are already larger. Such models are usually described mathematically as Pólya urns or one of their many generalizations [43, 51]. What is common in almost all of these generalized Pólya urns, and relevant to us, is the fact that the number of added balls of a given color is chosen from a finite set, with probabilities that are each a *continuous* function of the proportion of balls of a *single* color, except that they are normalized to sum to one. Although this allows for some form of competition among colors, it precludes direct comparison of the proportions of balls of different colors, so it does not allow for the modeling of systems in which growth rates depend on the differences between the sizes of different entities, let alone their ranking. Two exceptions are the works of Arthur et al. [5] and of Hill et al. [27], which allow arbitrary comparisons of proportions of balls of different colors, but they only treat the simplest type of Pólya urn processes. These works do not specifically focus on ranking-based competition, but they partially cover them as extreme cases, with a subset of their results applying to them. See the Introduction and Section 3.2 for details.

Compared to these existing approaches, our work differs in two main ways. First, our approach is at a more abstract level; the literature related to preferential attachment and Gibrat's law usually starts with a specific model, with the goal of reproducing some empirically observed phenomena, such as outcome unpredictability and skewed popularity distributions. Our approach in contrast is model-independent; we have identified conditions that are sufficient to lead to certain rich-get-richer phenomena, that is, conditions that when satisfied by *any* model, regardless of the exact assumptions made, lead to the stated results. This is illustrated in Section 3.1, where we point out that ranking-based rich-get-richer dynamics could be set in motion under a wide array of behavioral or algorithmic assumptions, as long as Assumption 2.5 is satisfied. In this respect, our work is similar in spirit to the work of Arthur et al. [5]. Such a model-free approach is more suitable for explaining the universality of certain properties of real-world rankings [31].

The second and perhaps more distinctive difference of our work, is the fact that it covers the opposite end of the spectrum of rich-get-richer dynamics. The distributions of the increments of the various components, instead of depending (continuously) on the current level of each of the components separately, they are piecewise constant with respect to the current levels, with discontinuities occurring when the ranking of the components changes. In other words, we focus explicitly on the role of ranking-based competition. However, our framework does not incorporate other types of competition, nor does it allow for any explicit dependence of the increments on the current level of the process, other than through the ranking.

The above delineates a promising future research direction: one could envisage a general mathematical theory of Markov rich-get-richer processes that encompasses all of the above cases, by allowing for an arbitrary dependence of the increments' distribution on the current level of the whole vector of the process, subject to the minimal conditions for rich-get-richer dynamics. The work of Arthur et al. [5] is in this direction for the case of simple Pólya urn processes, but no such framework currently exists for more general processes.

APPENDIX A: RANKINGS ARE EQUIVALENT TO WEAK ORDERINGS

The following proposition says that rankings are equivalent to weak orderings. A weak ordering on a set is like a total ordering, except that it allows for "ties". More precisely, a weak ordering "$\succeq$" on S is a binary relation that is transitive and strongly complete, that is, that for any two elements $a, b \in S$, at least one of the relations $a \succeq b$ or $b \succeq a$ holds [56]. Recall that we would get a *total* order if we further required that $a \succeq b$ and $b \succeq a$ implies $a = b$.

PROPOSITION A.1. *There is a bijection between rankings of a finite set S and weak orderings on S, given by $r \mapsto \succeq_r$, where*

$$a \succeq_r b \quad \text{whenever } r(a) \le r(b). \tag{85}$$

The above map satisfies

$$r(a) = \operatorname{card}\{b \in S : a \not\succeq_r b\} + 1. \tag{86}$$

The ranking r is strict if and only if $\succeq_r$ is a total order on S.

PROOF. It is easy to check that $\succeq_r$, as defined by Eq. (85), is a weak ordering on S. Using Eq. (85), Eq. (86) can be rewritten as

$$r(a) = \operatorname{card}\{b \in S : r(b) < r(a)\} + 1, \tag{87}$$

which is equivalent to Eq. (4), so it holds by definition. By Eq. (86), r is uniquely determined by $\succeq_r$, so the map $r \mapsto \succeq_r$ is one-to-one. To show that it is onto, let "$\succeq$" be a weak ordering on S and define $r : S \to [S]$ by

$$r(a) = \operatorname{card}\{b \in S : a \not\succeq b\} + 1. \tag{88}$$

We claim that $r(b) \le r(a)$ is equivalent to $b \succeq a$. First note that if $b \succeq a$, then by transitivity $\{c \in S : b \not\succeq c\}$ is a subset of $\{c \in S : a \not\succeq c\}$, hence $r(b) \le r(a)$. For the converse, assume that $b \not\succeq a$. Then we must have $a \succeq b$, and we get as above that $\{c \in S : a \not\succeq c\}$ is a subset of $\{c \in S : b \not\succeq c\}$, but this time it is a proper subset, because a belongs to the latter. Therefore $r(a) < r(b)$, which completes the proof of our claim. Hence, by Eq. (85), $\succeq$ is the same relation as $\succeq_r$, which shows that the mapping $r \mapsto \succeq_r$ is onto.

The last assertion follows from the fact that $a \succeq_r b$ and $b \succeq_r a$ hold simultaneously if and only if $r(a) = r(b)$. □

APPENDIX B: SUPPORTING PROOFS

Here we give the proofs of Lemmas 2.9, 2.18, and 2.19. For ease of reference, we repeat each statement before the proof.

LEMMA 2.9. *Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space. Let S be a finite set and for each $r \in S$, ν^r a distribution on $\mathbb{R}$ such that it either has positive mean or $\nu^r(\{0\}) = 1$. Let $\{R_n\}_{n\in\mathbb{N}}$ be a sequence of random elements in S and $\{Y_n\}_{n\in\mathbb{N}}$ a sequence of random variables with $Y_0 = 0$. Suppose that ΔY_{n+1} is conditionally independent of $\{(Y_k, R_k)\}_{k\le n}$ conditioned on R_n, with distribution ν^{R_n}. In other words, for any $A \in \mathcal{B}(\mathbb{R})$, $n \in \mathbb{N}$,*

$$\mathbb{P}\big(\Delta Y_{n+1} \in A \mid \{(Y_k, R_k)\}_{k\le n}\big) = \nu^{R_n}(A) \quad a.s. \tag{89}$$

Then,

$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} \{Y_n \ge 0\}\Big) \ge \epsilon > 0, \tag{90}$$

where ϵ depends only on the distributions ν^r, $r \in S$.

PROOF. Let $\{U_n^r\}_{n\in\mathbb{N}}^{r\in S}$ be a collection of independent random variables, independent of $\{(Y_n, R_n)\}_{n\in\mathbb{N}}$, and such that $U_n^r \sim \nu^r$ for all $r \in S$, $n \in \mathbb{N}$, where the relation $\sim$ means equality in distribution. Define $Y_0' = Y_0 = 0$ and for each $n \in \mathbb{N}$,

$$Y_{n+1}' = Y_n' + U_n^{R_n}. \tag{91}$$

Clearly, for any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{P}\big(\Delta Y_{n+1}' \in A \mid \{(Y_k', R_k)\}_{k\le n}\big) = \mathbb{P}\big(U_n^{R_n} \in A \mid R_n\big) = \nu^{R_n}(A), \tag{92}$$

therefore $\{Y_n'\}_{n\in\mathbb{N}} \sim \{Y_n\}_{n\in\mathbb{N}}$. It is hence enough to show that Eq. (90) holds for the sequence Y_n' instead of Y_n.

For each $r \in S$, define $\tau_0^r = -1$ and inductively $\tau_n^r = \inf\{k > \tau_{n-1}^r : R_k = r\}$. Note that each Y_n is a sum of terms of the form $U_{\tau_k}^r$, for $k = 1, \dots, m_r$, where $m_r \in \mathbb{N}$, $r \in S$. Therefore,

$$\bigcap_{n\in\mathbb{N}} \{Y_n' \ge 0\} \supset \bigcap_{r\in S} \bigcap_{n\in\mathbb{N}} \left\{\sum_{k=1}^{n} U_{\tau_k^r}^r \ge 0\right\}, \tag{93}$$

with the convention $U_{\tau_k^r}^r = 0$ when $\tau_k^r = \infty$.

Since $\tau_k^r \perp\!\!\!\perp \{U_n^r\}_{n\in\mathbb{N}}$, if the τ_k^r's were all finite a.s., it would easily follow that $U_{\tau_k^r}^r$ has the same distribution as U_1^r, $k \in \mathbb{N}$, and since τ_k^r is strictly increasing in k we would even get that $\{U_{\tau_k^r}^r\}_{k\in\mathbb{N}}$ is i.i.d. To deal with the case $\tau_n^r = \infty$, we define the random times σ_n^r as follows: Let $v^r = \sup\{n \in \mathbb{N} : \tau_n^r < \infty\}$ and

$$\sigma_n^r = \tau_n^r \cdot \mathbf{1}_{n\le v^r} + \left(\tau_{v^r}^r + n - v_r\right) \cdot \mathbf{1}_{n>v^r}. \tag{94}$$

The σ_n^r's are almost surely finite and distinct for fixed $r \in S$, and $\{\sigma_n^r\}_{n\in\mathbb{N}}^{r\in S} \perp\!\!\!\perp \{U_n^r\}_{n\in\mathbb{N}}^{r\in S}$. Therefore, by [44], Theorem 2.1, we get that $\{U_{\sigma_n^r}^r\}_{n\in\mathbb{N}}^{r\in S} \sim \{U_n^r\}_{n\in\mathbb{N}}^{r\in S}$. (In [44] it is assumed that the σ_n^r's are all distinct a.s., even for different r's, but this assumption can be substituted by the fact that the sequences $\{U_n^r\}_{n\in\mathbb{N}}$ are independent for different r's and the proof goes through.)

Now observe that $\sigma_k^r = \tau_k^r$ on $\{\tau_k^r < \infty\}$, therefore, by Eq. (93),

$$\bigcap_{n\in\mathbb{N}} \{Y_n' \ge 0\} \supset \bigcap_{r\in S} \bigcap_{n\in\mathbb{N}} \left\{ \sum_{k=1}^n U_{\sigma_k^r}^r \ge 0 \right\}. \tag{95}$$

Consequently,

$$\begin{aligned} \mathbb{P}\Bigl(\bigcap_{n\in\mathbb{N}} \{Y_n' \ge 0\}\Bigr) &\ge \mathbb{P}\Bigl(\bigcap_{r\in S} \bigcap_{n\in\mathbb{N}} \Bigl\{ \sum_{k=1}^n U_{\sigma_k^r}^r \ge 0 \Bigr\}\Bigr) \\ &= \prod_{r\in S} \mathbb{P}\Bigl(\bigcap_{n\in\mathbb{N}} \Bigl\{ \sum_{k=1}^n U_k^r \ge 0 \Bigr\}\Bigr) > 0, \end{aligned} \tag{96}$$

because $\{U_k^r\}_{k\in\mathbb{N}}$ is an i.i.d. sequence of random variables that are either identically 0 or they have a positive mean. □

LEMMA 2.18. *Let A_n, $n \in \mathbb{N}$, and A be measurable sets in a probability space, each with positive probability, and suppose that $A_n \to A$ a.s.* (*i.e.*, $\mathbb{P}((A_n\backslash A) \cup (A\backslash A_n)) \to 0$). *Then*, $\mathbb{P}(S \mid A_n) \to \mathbb{P}(S \mid A)$ *uniformly in $S \in \mathcal{F}$.*

PROOF. We have

$$\begin{aligned} &\left|\mathbb{P}(S \mid A) - \mathbb{P}(S \mid A_n)\right| \\ &\quad= \left| \frac{\mathbb{P}(S \cap A)}{\mathbb{P}(A)} - \frac{\mathbb{P}(S \cap A_n)}{\mathbb{P}(A_n)} \right| \\ &\quad= \left| \frac{\mathbb{P}(S \cap A_n) + \mathbb{P}(S \cap A\backslash A_n) - \mathbb{P}(S \cap A_n\backslash A)}{\mathbb{P}(A)} - \frac{\mathbb{P}(S \cap A_n)}{\mathbb{P}(A_n)} \right| \\ &\quad\le \mathbb{P}(S \cap A_n) \cdot \left| \frac{1}{\mathbb{P}(A)} - \frac{1}{\mathbb{P}(A_n)} \right| + \frac{|\mathbb{P}(S \cap A\backslash A_n) - \mathbb{P}(S \cap A_n\backslash A)|}{\mathbb{P}(A)} \\ &\quad\le \left| \frac{1}{\mathbb{P}(A)} - \frac{1}{\mathbb{P}(A_n)} \right| + \frac{\mathbb{P}(A\backslash A_n) + \mathbb{P}(A_n\backslash A)}{\mathbb{P}(A)}. \end{aligned} \tag{97}$$

The quantity in the last line does not depend on S and, by assumption, it converges to 0 as $n \to \infty$. □

LEMMA 2.19. *Let $a_{m,n} \in \mathbb{R}$, $m, n \in \mathbb{N}$, and suppose that* $\lim_{m\to\infty} a_{m,n} = a_n \in \mathbb{R}$ *uniformly in n, and* $\lim_{n\to\infty} a_{m,n} = a \in \mathbb{R}$ *for all $m \in \mathbb{N}$. Then*, $\lim_{n\to\infty} a_n = a$.

PROOF. Let $\epsilon > 0$ and let $m_0 \in \mathbb{N}$ be such that $|a_{m_0,n} - a_n| < \epsilon$ for all $n \in \mathbb{N}$. Now let $n_0 \in \mathbb{N}$ be such that $|a_{m_0,n} - a| < \epsilon$ for all $n \ge n_0$. It follows that $|a_n - a| < 2\epsilon$ for all $n \ge n_0$. □

Acknowledgments. We would like to thank Thorsten Joachims, Gabor Lugosi, and Murad Taqqu for their remarks in previous versions of this manuscript.

Funding. Pantelis P. Analytis was supported in part through NSF Award IIS-1513692 granted to Thorsten Joachims.

REFERENCES

[1] AGARWAL, A., TAKATSU, K., ZAITSEV, I. and JOACHIMS, T. (2019). A general framework for counterfactual learning-to-rank. In *Proceedings of the* 42*nd International ACM SIGIR Conference on Research and Development in Information Retrieval* 5–14.

[2] ALLISON, P. D., LONG, J. S. and KRAUZE, T. K. (1982). Cumulative advantage and inequality in science. *Am. Sociol. Rev.* 615–625.

[3] ANALYTIS, P. P., CERIGIONI, F., GELASTOPOULOS, A., STOJIC, H. et al. (2022). Sequential choice and self-reinforcing rankings. Technical Report, Universitat Pompeu Fabra, Department of Economics and Business.

[4] ARTHUR, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Econ. J.* **99** 116–131.

[5] ARTHUR, W. B., ERMOLIEV, Y. M. and KANIOVSKI, Y. M. (1986). Strong laws for a class of path-dependent stochastic processes with applications. In *Stochastic Optimization* (*Kiev*, 1984). *Lect. Notes Control Inf. Sci.* **81** 287–300. Springer, Berlin. MR0890992 https://doi.org/10.1007/BFb0007105

[6] BANNER, A. D., FERNHOLZ, R. and KARATZAS, I. (2005). Atlas models of equity markets. *Ann. Appl. Probab.* **15** 2296–2330. MR2187296 https://doi.org/10.1214/105051605000000449

[7] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

[8] CAPLIN, A., DEAN, M. and MARTIN, D. (2011). Search and satisficing. *Amer. Econ. Rev.* **101** 2899–2922.

[9] CHUKLIN, A., MARKOV, I. and DE RIJKE, M. (2015). Click models for web search. *Synthesis Lectures on Information Concepts*, *Retrieval*, *and Services* **7** 1–115.

[10] CHUNG, F., HANDJANI, S. and JUNGREIS, D. (2003). Generalizations of Polya's urn problem. *Ann. Comb.* **7** 141–153. MR1994572 https://doi.org/10.1007/s00026-003-0178-y

[11] CIAMPAGLIA, G. L., NEMATZADEH, A., MENCZER, F. and FLAMMINI, A. (2018). How algorithmic popularity bias hinders or promotes quality. *Sci. Rep.* **8** 1–7.

[12] COLLEVECCHIO, A., COTAR, C. and LICALZI, M. (2013). On a preferential attachment and generalized Pólya's urn model. *Ann. Appl. Probab.* **23** 1219–1253. MR3076683 https://doi.org/10.1214/12-AAP869

[13] CRASWELL, N., ZOETER, O., TAYLOR, M. and RAMSEY, B. (2008). An experimental comparison of click position-bias models. In *Proceedings of the* 2008 *International Conference on Web Search and Data Mining* 87–94.

[14] DENISOV, D. and WACHTEL, V. (2010). Conditional limit theorems for ordered random walks. *Electron. J. Probab.* **15** 292–322. MR2609589 https://doi.org/10.1214/EJP.v15-752

[15] DENISOV, D. and WACHTEL, V. (2015). Random walks in cones. *Ann. Probab.* **43** 992–1044. MR3342657 https://doi.org/10.1214/13-AOP867

[16] DENRELL, J., FANG, C. and LIU, C. (2014). Chance explanations in the management sciences. *Organization Science* **26** 923–940.

[17] DURAJ, J. (2014). Random walks in cones: The case of nonzero drift. *Stochastic Process. Appl.* **124** 1503–1518. MR3163211 https://doi.org/10.1016/j.spa.2013.12.003

[18] EICHELSBACHER, P. and KÖNIG, W. (2008). Ordered random walks. *Electron. J. Probab.* **13** 1307–1336. MR2430709 https://doi.org/10.1214/EJP.v13-539

[19] GABAIX, X. (1999). Zipf's law for cities: An explanation. *Q. J. Econ.* **114** 739–767.

[20] GABAIX, X. (2008). Power laws in economics and finance. Technical Report, National Bureau of Economic Research.

[21] GAETA, R., GARETTO, M., RUFFO, G. and FLAMMINI, A. (2023). Reconciling the quality vs popularity dichotomy in online cultural markets. *ACM Trans. Inf. Syst.* **41** 1–34.

[22] GARBIT, R. (2007). Temps de sortie d'un cône pour une marche aléatoire centrée. *C. R. Math. Acad. Sci. Paris* **345** 587–591. MR2374468 https://doi.org/10.1016/j.crma.2007.10.016

[23] GARBIT, R. (2018). On the exit time from an orthant for badly oriented random walks. *Braz. J. Probab. Stat.* **32** 117–146. MR3770866 https://doi.org/10.1214/16-BJPS334

[24] GARBIT, R. and RASCHEL, K. (2016). On the exit time from a cone for random walks with drift. *Rev. Mat. Iberoam.* **32** 511–532. MR3512425 https://doi.org/10.4171/RMI/893

[25] GERMANO, F., GÓMEZ, V. and LE MENS, G. (2019). The few-get-richer: A surprising consequence of popularity-based rankings?. In *Proceedings of the World Wide Web Conference* 2764–2770.

[26] GIBRAT, R. (1931). Les inégalits économiques. *Sirey*.

[27] HILL, B. M., LANE, D. and SUDDERTH, W. (1980). A strong law for some generalized urn processes. *Ann. Probab.* **8** 214–226. MR0566589

[28] HUDSON, J. (2013). Ranking journals. *Econ. J.* **123** F202–F222.

[29] ICHIBA, T., PAL, S. and SHKOLNIKOV, M. (2013). Convergence rates for rank-based models with applications to portfolio theory. *Probab. Theory Related Fields* **156** 415–448. MR3055264 https://doi.org/10.1007/s00440-012-0432-5

[30] ICHIBA, T., PAPATHANAKOS, V., BANNER, A., KARATZAS, I. and FERNHOLZ, R. (2011). Hybrid atlas models. *Ann. Appl. Probab.* **21** 609–644. MR2807968 https://doi.org/10.1214/10-AAP706

[31] IÑIGUEZ, G., PINEDA, C., GERSHENSON, C. and BARABÁSI, A.-L. (2022). Dynamics of ranking. *Nat. Commun.* **13** 1–7.

[32] ITKIN, D. and LARSSON, M. (2021). On a class of rank-based continuous semimartingales. ArXiv preprint. Available at arXiv:2104.04396.

[33] JOACHIMS, T., GRANKA, L. A., PAN, B., HEMBROOKE, H. and GAY, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th International ACM SIGIR Conference on Research and Development in Information Retrieval* **5** 154–161.

[34] JOACHIMS, T., SWAMINATHAN, A. and SCHNABEL, T. (2017). Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* 781–789.

[35] KALLENBERG, O. (2002). *Foundations of Modern Probability* (*2nd ed*). Springer.

[36] KEENEY, R. L. and RAIFFA, H. (1976). *Decisions with Multiple Objectives*: *Preferences and Value Tradeoffs*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0449476

[37] KEILSON, J. and SERVI, L. D. (1986). Oscillating random walk models for $GI/G/1$ vacation systems with Bernoulli schedules. *J. Appl. Probab.* **23** 790–802. MR0855384 https://doi.org/10.2307/3214016

[38] KEILSON, J. and WELLNER, J. A. (1978). Oscillating Brownian motion. *J. Appl. Probab.* **15** 300–310. MR0474526 https://doi.org/10.2307/3213403

[39] KEMPERMAN, J. H. B. (1974). The oscillating random walk. *Stochastic Process. Appl.* **2** 1–29. MR0362500 https://doi.org/10.1016/0304-4149(74)90010-6

[40] LABAND, D. N. (2013). On the use and abuse of economics journal rankings. *Econ. J.* **123** F223–F254.

[41] LARUELLE, S. and PAGÈS, G. (2019). Nonlinear randomized urn models: A stochastic approximation viewpoint. *Electron. J. Probab.* **24** Paper No. 98. MR4017116 https://doi.org/10.1214/19-ejp312

[42] LIU, T.-Y. et al. (2009). Learning to rank for information retrieval. *Found. Trends Inf. Retr.* **3** 225–331.

[43] MAHMOUD, H. M. (2009). *Pólya Urn Models*. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR2435823

[44] MELFI, V. F. and PAGE, C. (2000). Estimation after adaptive allocation. *J. Statist. Plann. Inference* **87** 353–363. MR1771124 https://doi.org/10.1016/S0378-3758(99)00198-6

[45] MENSHIKOV, M., POPOV, S. and WADE, A. (2017). *Non-homogeneous Random Walks*: *Lyapunov Function Methods for Near-Critical Stochastic Systems*. *Cambridge Tracts in Mathematics* **209**. Cambridge Univ. Press, Cambridge. MR3587911 https://doi.org/10.1017/9781139208468

[46] MERTON, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science* **159** 56–63.

[47] MUCHNIK, L., ARAL, S. and TAYLOR, S. J. (2013). Social influence bias: A randomized experiment. *Science* **341** 647–651.

[48] PAGE, L., BRIN, S., MOTWANI, R. and WINOGRAD, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report, Stanford InfoLab.

[49] PAGE, S. E. (2006). Path dependence. *Q. J. Polit. Sci.* **1** 87–115.

[50] PAL, S. and PITMAN, J. (2008). One-dimensional Brownian particle systems with rank-dependent drifts. *Ann. Appl. Probab.* **18** 2179–2207. MR2473654 https://doi.org/10.1214/08-AAP516

[51] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79. MR2282181 https://doi.org/10.1214/07-PS094

[52] PIKETTY, T. (2015). About capital in the twenty-first century. *Amer. Econ. Rev.* **105** 48–53.

[53] PODOLNY, J. M. (1993). A status-based model of market competition. *Am. J. Sociol.* **98** 829–872.

[54] PRICE, D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inf. Sci.* **27** 292–306.

[55] REDNER, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* **4** 131–134.

[56] ROBERTS, F. S. (1979). *Measurement Theory*. *Advanced Book Program*. Addison-Wesley Co., Reading, MA. MR0551364

[57] SALGANIK, M. J., DODDS, P. S. and WATTS, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311** 854–856.

[58] SIMON, H. A. (1955). On a class of skew distribution functions. *Biometrika* **42** 425–440. MR0073085 https://doi.org/10.1093/biomet/42.3-4.425

[59] YULE, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. *Philos. Trans. R. Soc. Lond., B Contain. Pap. Biol. Character* **213** 21–87.

[60] ZHU, T. (2009). Nonlinear Pólya urn models and self-organizing processes. Unpublished dissertation, Univ. Pennsylvania, Philadelphia.