Published in "Information Sciences 488(): 257–271, 2019"
which should be cited to refer to this work.

# The long-term impact of ranking algorithms in growing networks

Shilun Zhang[a], Matúš Medo[a,b,c], Linyuan Lü[a,d], Manuel Sebastian Mariani[a,e,*]

[a] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610051, PR China
[b] Department of Radiation Oncology, Inselspital, Bern University Hospital and University of Bern, Bern 3010, Switzerland
[c] Department of Physics, University of Fribourg, Fribourg 1700, Switzerland
[d] Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, PR China
[e] URPP Social Networks, Universität Zürich, Zürich 8050, Switzerland

*Keywords:*
Complex networks
Ranking
Popularity and quality
Popularity inequality
Algorithmic bias

When users search online for content, they are constantly exposed to rankings. For example, web search results are presented as a ranking of relevant websites, and online bookstores often show us lists of best-selling books. While popularity-based ranking algorithms (like Google's PageRank) have been extensively studied in previous works, we still lack a clear understanding of their potential systemic consequences. In this work, we fill this gap by introducing a new model of network growth that allows us to compare the properties of networks generated under the influence of different ranking algorithms. We show that by correcting for the omnipresent age bias of popularity-based ranking algorithms, the resulting networks exhibit a significantly larger agreement between the nodes' inherent quality and their long-term popularity, and a less concentrated popularity distribution. To further promote popularity diversity, we introduce and validate a perturbation of the original rankings where a small number of randomly-selected nodes are promoted to the top of the ranking. Our findings move the first steps toward a model-based understanding of the long-term impact of popularity-based ranking algorithms, and our novel framework could be used to design improved information filtering tools.

## 1. Introduction

Ranking algorithms allow us to efficiently sort massive amounts of online information, and quickly provide us with the relevant items that fit our needs. Due to the ubiquity of rankings, the implications of ranking-based information filtering tools such as search engines [3] and recommendation systems [21] for our society are widely debated [2,5,8,15]. Importantly, in social and information systems, the rankings that we are exposed to are influenced by social processes and, at the same time, influence social processes themselves [40]. To provide a few examples, rankings can heavily impact on the eventual popularity of movies and songs in cultural markets [39], affect the attention received by products in online e-commerce platforms [16,41,49], increase the sales of top-ranked dishes in restaurants [4], and even influence the choices of undecided

* Corresponding author at: Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, 610051 Chengdu, PR China; URPP Social Networks, Universität Zürich, 8050 Zürich, Switzerland.
*E-mail addresses:* linyuan.lv@uestc.edu.cn (L. Lü), manuel.mariani@business.uzh.ch (M.S. Mariani).

http://doc.rero.ch

http://doc.rero.ch

electors and, as a result, the outcome of political elections [9]. Therefore, understanding the potential systemic impact of ranking algorithms and correcting their potential flaws becomes a critical issue in diverse contexts.

So far, most studies on ranking in complex networks validated the algorithms in terms of their ability to find "important" nodes in the network. These studies assumed that there exists a set of target nodes to retrieve, and they assessed the performance of the ranking algorithms in terms of their accuracy in retrieving the target nodes. Relevant studies on the topic include papers on the identification of structural and dynamical influential nodes [20,29]; on the identification of expert-selected relevant nodes [19,25,26]; on the prediction of successful nodes [19,20]. Other studies [1] validated ranking algorithms with respect to their mathematical properties. We refer to the Related Works section (Section 2.1) for more details.

Yet, our understanding of the potential consequences of ranking algorithms for a given system remains elusive. An algorithm with high accuracy in identifying target nodes might lead, in the long term, to a highly-uneven system where only few nodes are able to gain visibility, making it harder for people to discover recent content [11,49]. Besides, both experimental [9] and theoretical [34] results suggest that opinion formation processes can be manipulated by means of information-filtering tools. Therefore, in our increasingly digitalized society, understanding the feedback mechanisms between individual behavior and ranking algorithms is of utmost importance.

The main goal of this article is to investigate the long-term implications of ranking algorithms from a theoretical perspective. To this end, we introduce a growing directed-network model where, at each step, a new node enters the system and links to a limited number of preexisting nodes. The links are chosen probabilistically based on two factors: (1) the preexisting nodes' ranking position as determined by a predefined, "adopted" ranking algorithm; (2) the nodes' quality. The relative importance of the two effects is determined by a homogeneous parameter that can be interpreted as the nodes' sensitivity to quality. In the model, quality is a node-level parameter [24] which can be interpreted as the success that the node would achieve in the absence of social influence mechanisms [39]. Crucially, different adopted ranking algorithms lead to different properties of the final network.

In this article, we aim to determine whether the adoption of a given ranking algorithm by a given system allows high-quality nodes to experience larger success than low-quality nodes. If this is not the case, we may conclude that the adopted ranking algorithm has a negative impact on the system, as it may prevent high-quality nodes from becoming popular and increase the popularity of low-quality nodes. Therefore, we use the model to address the following questions: Will a given algorithm facilitate or impede the success of high-quality nodes in the system? Is a given algorithm useful to discover high-quality nodes? Does it lead to uneven, highly-concentrated popularity distributions?

We postulate that a good ranking algorithm should lead to a network where: (1) the nodes' long-term popularity strongly correlates with their quality (*quality promotion*); (2) the nodes' score strongly correlates with their quality (*quality detection*); (3) popularity is not concentrated in a few nodes (*popularity diversity*). The quality promotion and detection properties favor the algorithms that help the system to improve the popularity-quality correlation (quality promotion) [6] and those that help the nodes to find high-quality nodes (quality detection) [24]. The popularity diversity property favors the algorithms that distribute popularity more evenly across the nodes, making it easier for the nodes to find high-quality nodes that are not among the most popular ones [50].

We find that in networks that adopt the ranking by cumulative popularity (as measured by the number of incoming links – node indegree [30]) as the ranking algorithm, the correlation between node popularity and quality strongly depends not only on the nodes' sensitivity to quality, but also on their willingness to select low-ranked nodes ("exploration cost" in [6]). Besides, the ranking by a popularity metric that is not biased by node age [25] (called rescaled indegree in [25]) leads to networks where both the nodes' final popularity and the nodes' score are significantly better correlated with node quality, and the final popularity distribution is significantly more diverse. Interestingly, when the exploration cost is large, networks that adopted a random ranking of the nodes exhibit even higher indegree-quality correlation than networks that adopted a popularity-based ranking: while popularity-based ranking algorithms are always useful for the nodes to discover high-quality content, they may accelerate the dissemination of low-quality content when individuals rely too heavily on them.

To further promote popularity diversity, we introduce and validate a ranking algorithm – the ranking by rescaled indegree with random promotion – where the original ranking by rescaled indegree is "perturbed" by promoting a small number of randomly-selected nodes to the top-10 or the top-20 of the ranking. Such a perturbation has a deterministic component (the number of nodes that are promoted is fixed) and a noisy one (the promoted nodes are chosen at random). It allows us to study the impact of a small amount of noise on systemic properties, in a similar spirit as previous studies that investigated the impact of noise on democratic consensus promotion in animal groups [7], dynamical influence detection [17], and the performance of human groups in coordination problems [42]. We find that with respect to the ranking by rescaled indegree, the ranking by rescaled indegree with random promotion generates networks with more even popularity distributions. Intriguingly, we find that the random promotion can have a marginal or a negative impact on the agreement between nodes' final popularity and quality, depending on the model parameters.

The main contribution of this article is to provide the first systematic comparison of network-based ranking algorithms with respect to their long-term systemic effects. Compared to existing works, our work shifts the focus from the performance of ranking algorithms in retrieving relevant nodes to the properties of the networks that emerge when the ranking algorithm influences the nodes' behavior. In addition, our work contributes to the literature on the relation between popularity and quality (see Section 2.2 for related works) by revealing that suppressing the bias by node age of popularity-based metrics is beneficial to quality promotion, detection, and popularity diversity.

The manuscript is organized as follows. Section 2 summarizes related works on the validation of ranking algorithms in time-evolving networks (Section 2.1) and the relation between popularity and quality in social and information systems (Section 2.2). Section 3 introduces a model of network growth together with the ranking algorithms considered here, and the ranking evaluation criteria. Section 4 presents the results of our numerical simulations both for the basic model (Sections 4.1–4.4) and for a variant of the model with node removal (Section 4.5), together with an application of our model to a real information network of scientific papers (Section 4.6). Section 5 is dedicated to a discussion of our results and their implications for algorithmic evaluation and the quality-popularity relation in information systems. Appendices A-B conclude the main text. The Supplementary Material (SM) file is available online – figures whose labels contain an "S" (e.g., Fig. S1) can be found in the SM file.

## 2. Related works

### *2.1. Validation of ranking algorithms in complex networks*

In fact, the synthetic networks generated with our ranking-based growth model can be interpreted as benchmark graphs for ranking algorithms. Our focus on quality promotion, quality detection, and diversity promotion makes our validation framework for ranking algorithm fundamentally different from existing validation methods. Indeed, network-based ranking algorithms are typically evaluated according to their ability to identify structural vital nodes [29], find those nodes that maximize the reach of a spreading process [20], single out expert-selected important nodes [25,26], or respect various sets of axioms [1,33]. In the following, we provide the basic ideas behind these studies.

Structural vital nodes can be defined as the nodes whose removal causes the largest damage to the network's connectedness [20,29]. The identification of such key nodes is referred to as the structural influence maximization problem [20]. Network-based ranking algorithms have been therefore compared with respect to their ability to identify these nodes. High accuracy can be obtained by means of optimal percolation theory [29].

Dynamic vital nodes can be defined as the nodes that maximize a given function that depends on both the topology of the network and a specific dynamical process on the network [20]. The identification of such influential nodes is referred to as the dynamic influence maximization problem [20]. The influential spreaders are a widely-studied example: They maximize the asymptotic reach of a given spreading process which unfolds on a networked population [20]. In line with this definition, a large number of studies have compared the accuracy of the rankings by different algorithms in identifying the optimal spreaders – see [20] for a review. In this context, a relevant (and still debated) question is whether ranking algorithms that only use local information [23] are able to outperform algorithms that take into account the whole network topology [22].

Other scholars have been interested in the ability of the ranking algorithms to identify important nodes whose value has been recognized by experts. For example, one can evaluate the algorithms by their ability to identify seminal papers selected by the editors of prestigious journals [25], scholars who have been awarded a prestigious prize [35], or patents that have been labeled as seminal by experts in the involved technological domains [26]. We refer to the review [19] for more examples.

Finally, scholars have characterized and validated network-based ranking algorithms based on their mathematical properties. In this approach, one postulates a set of axioms, and posits that a good ranking algorithm should fulfill all of them (or at least, most of them). For example, Boldi and Vigna [1] introduced three axioms (size axiom, density axiom, and score-monotonicity axioms) and used them to characterize the behavior of various centrality metrics. They found that the harmonic centrality (inspired by the popular closeness centrality) is the only metric that satisfies all of them. We refer to Boldi and Vigna [1] for more details and a summary of previous attempts to formalize the properties of network-based ranking algorithms through axioms.

### *2.2. The relation between popularity and quality*

Our work contributes to the literature on the relation between the popularity of an agent in a social system and its inherent quality. In this context, quality (also referred to as fitness [28], or talent [18]) refers to the popularity that the agent would experience in the absence of social influence mechanisms [39]. A robust finding in previous experiments of diverse nature [4,5,11,39] is that the current ranking position of an item (or its current popularity) heavily influences its eventual popularity or success. As a consequence, one of the key challenges is to assess whether in a given system, the final popularity of an item is a reliable proxy for its quality.

Cho [5] pointed out that ranking algorithms and search engines that favor already popular items create a strong "popularity bias" [11] – also dubbed as "search-engine bias" [5], and "googlearchy" [15] – such that only already-popular nodes can receive substantial attention in the future, whereas recent high-quality nodes will remain essentially unnoticed [5]. This bias can amplify initial differences between the items' popularity, leading to disproportionately high popularity of some items regardless of their quality. O'Madadhain et al. [32] emphasized that static ranking algorithms (like Google's PageRank [3]) are based on time-aggregate network representations and, for this reason, they do not respect the sequence of events that



http://doc.rero.ch

**Table 1**
The network growth model and ranking algorithms: Notation used in this article.

| | Variable | Description |
|---|---|---|
| Model properties | $\mathcal{A}$ | Adopted ranking algorithm |
| | $r_j^{(\mathcal{A})}(t)$ | Ranking position of node $i$ at time $t$ by algorithm $\mathcal{A}$ |
| | $q_i$ | Node quality |
| | $N$ | Final number of nodes |
| | $\alpha$ | Exploration cost parameter |
| | $\beta$ | Sensitivity to ranking parameter |
| Ranking algorithms | $k$ | Indegree |
| | $R(k)$ | Rescaled indegree |
| | $R(k)$+RP | Rescaled indegree with random promotion |

led to the formation of the network. To solve this issue, they introduced [32] and validated [33] a network-based ranking algorithm, called EventRank, that takes into account the detailed sequence of events, in a similar spirit as the recent literature on temporal networks [19].

Salganik et al. [39] found that in artificial cultural markets, showing the items' ranking by popularity to consumers significantly affects the items' final popularity. Recent studies [47] emphasized the individuals' limited attention as a determinant for the viral popularity of low-quality items. Other recent works focused on modeling the interplay between quality/fitness/talent and popularity for various types of information items or agents, including websites [18], scientific papers [28,46], researchers [27,43], and bestseller books [48]. Both model-based [18,28] and experimental results [39] indicate that in presence of social influence, the relation between popularity and quality is highly non-linear, meaning that small variations of quality lead to large variations in popularity.

## 3. Model and ranking algorithms

In this Section, we introduce the model of network growth (Section 3.1), the ranking algorithms considered in this paper (Section 3.2), and the metrics used to evaluate the long-term impact of ranking algorithms (Section 3.3). We refer to Table 1 for a summary of the notation used in this article.

### 3.1. The model of network growth

We focus here on monopartite directed networks. Our model is meant to represent a social or information network where the number of nodes grows with time. In the model, each node $i$ is endowed with a quality parameter $q_i$ that quantifies its attractiveness to new incoming connections in the absence of ranking influence. Before generating a network, we choose the ranking algorithm $\mathcal{A}$ that influences the growth: the nodes choose the targets of their links[1] based on the ranking of the nodes by $\mathcal{A}$. For the sake of brevity, we say that the network is $\mathcal{A}$-generated or that the system "has adopted" algorithm $\mathcal{A}$.

We introduce a model that features three essential elements:

1. *Growth.* At each time step, one new node enters the system. Nodes can thus be labeled directly by the time step in which they appeared. Each new node creates $m$ directed links to $m$ different preexisting nodes.[2]
2. *Ranking-driven attachment.* With probability $\beta$, node $t$ chooses $j$ as the target of a link with the probability

$$P^{(\mathcal{A})}(j,t) = \frac{(r_j^{(\mathcal{A})}(t))^{-\alpha}}{\sum_{s=1}^{t-1}(r_s^{(\mathcal{A})}(t))^{-\alpha}}, \tag{1}$$

where $r_j^{(\mathcal{A})}(t)$ is the ranking position of node $j$ at time $t$ according to $\mathcal{A}$; the real number $\alpha \geq 0$ is a tunable model parameter referred to as *exploration cost* by Ciampaglia et al. [6]: Large values of $\alpha$ imply that the nodes are only willing to connect to the top-nodes by the adopted ranking algorithm, whereas low values of $\alpha$ allow the nodes to also connect to low-ranked nodes.
3. *Quality-driven attachment.* With probability $1-\beta$, node $t$ chooses $j$ as the target of a link with the probability

$$P^{(q)}(j,t) = \frac{q_j}{\sum_{l=1}^{t-1} q_l}. \tag{2}$$

Our model reduces itself to the model by Fortunato et al. [10] in the special case $\beta = 1$. In other words, node quality and quality-driven attachment are novel elements with respect to Fortunato et al.'s model [10]. Differently from the recent popularity dynamics model by Ciampaglia et al. [6] that considers a cultural market composed of a fixed number $N$ of items,

[1] The network's directed links might be interpreted as friendship or follower relationships in online social networks, or as citations between documents in information networks.
[2] Self-loops and multiple links between a given pair of nodes are prohibited.



http://doc.rero.ch

our model represents a network that grows with time. Compared to static markets, growing networks better represent real information systems where new products, information items, songs, etc. continually appear with time. Differently from previous works [6,11,28], we aim to use our growing network model to compare the long-term properties of the networks generated by different ranking algorithms.

### 3.2. Ranking algorithms

Our main goal is to uncover the long-term implications of the temporal bias of static centrality metrics and the benefits from suppressing such bias. For this reason, we focus here on indegree (due to its simplicity and wide use [30]) and rescaled indegree (as it is a popularity metric that is not biased by temporal effects [25]). In addition, we introduce a random promotion mechanism which "perturbs" the ranking by rescaled indegree by promoting randomly-selected nodes to the top of the ranking, and we consider a random ranking of the node as a baseline. We provide below the details of these four ranking algorithms.

1. *Ranking by indegree, $k$.* The indegree[3] of a node is defined as the number of incoming connections received by that node [30]. We simply rank the nodes in order of decreasing indegree $k$, which is arguably the simplest way to rank the nodes in a directed network [30]. In growing networks, node indegree is strongly biased by node age [19,25], as confirmed by numerical simulations and analytic computations with our model (see Appendix B).
2. *Ranking by (age-)rescaled indegree, $R(k)$.* We rank the nodes in order of decreasing age-rescaled indegree [25] $R(k)$. The rescaled indegree is built on indegree by requiring that node score is not biased by node age. More specifically, for each node $i$, we consider a reference set $\mathcal{R}_i := \{i-\Delta/2, \ldots, i+\Delta/2\}$ of $\Delta+1$ nodes of similar age as node $i$ – we set $\Delta = 0.01\,N$. We compute the mean $\mu_i(k)$ and the standard deviation $\sigma_i(k)$ of node indegree within this reference set as

$$\begin{aligned}\mu_i(k) &= \tfrac{1}{\Delta+1}\sum_{j\in\mathcal{R}_i} k_j,\\ \sigma_i(k) &= \sqrt{\tfrac{1}{\Delta+1}\sum_{j\in\mathcal{R}_i}(k_j-\mu_i(k))^2}.\end{aligned}$$

   The rescaled indegree $R_i(k)$ of node $i$ is given by the $z$-score [25]

$$R_i(k) = \frac{k_i-\mu_i(k)}{\sigma_i(k)}. \tag{3}$$

   The rescaled indegree of a given node thus quantifies the difference between the node's indegree and the mean indegree of nodes of similar age, in units of standard deviations. Using the $z$-score to normalize static metrics of node importance is customary in scientometrics [44], where scholars aim to gauge the impact of a given scientific paper independently of its field and publication date [45]. Besides, in citation networks, the rescaled indegree allows us to early-identify seminal papers [25], movies [36] and patents [26] better than citation count.
3. *Ranking by age-rescaled indegree with Random Promotion, ($R(k)$+RP).* While the age-rescaled indegree suppresses the cumulative advantage of older nodes [25], it may still amplify the advantage of the nodes that, perhaps by chance, received quickly many more connections than nodes of similar age did. To reduce this potential problem and further increase diversity, we introduce the ranking by rescaled indegree with random promotion ($R(k)$+RP): We rank the nodes by age-rescaled indegree, $R(k)$, and then we "promote" $P$ randomly-selected nodes to the top-$T$ of the ranking.[4] Therefore, the fraction $\eta := P/T$ of nodes in the top-$T$ by the ranking is chosen at random. By placing some previously overlooked nodes at the top of the ranking, this mechanism gives these nodes enhanced visibility and, therefore, an additional opportunity to attract some links. In the following, we show results for $T = 10$ and $\eta = 0.5$; results for other values of $T$ and $\eta$ are shown in the Supplementary Material (Supplementary Figures S11– S13 and S19–S24).
4. *Random ranking.* The nodes are ranked at random. The resulting ranking is used as a baseline to understand for which parameter values does the final indegree-quality correlation benefit from the rankings by indegree and rescaled indegree.

For the ranking algorithms considered above, if two or more nodes happen to have the same score, their relative order is determined at random.

Choosing the quality distribution in the model deserves some attention. A simple mean-field approximation shows that for $\beta = 0$, the expected final node popularity is proportional to node quality; simulation results show that for $\beta > 0$, node final popularity is a power-law function of node quality (see Appendix B for details). Motivated by this property, to mimic the broad popularity distributions typically observed in real data [30], we choose a Pareto distribution of the quality values $q$ (see Appendix A for details). This choice leads indeed to broad indegree distributions as shown in Fig. S1. We refer to Appendix A for further simulation details.

[3] In the following, we will use interchangeably "indegree", "popularity", and "cumulative popularity". This is because, in our simple setting, the incoming links received by a node are the only available information on its "popularity". The situation might be different in a real online system where, for example, the popularity of a video can be quantified by various means: the number of downloads, the number of views, the number of shares, etc.

[4] When a node whose original ranking position was $r_0 > T$ is promoted to the ranking position $r^* \leq T$ in the top-$T$ of the ranking, the ranking positions of the nodes that occupied the ranking positions $\{r^*, r^*+1, \ldots, r_0-1, r_0\}$ increase by one (i.e., these nodes lose one ranking position).



http://doc.rero.ch

**Table 2**
Evaluation metrics used in this article.

| Property | Symbol | Description |
|---|---|---|
| Quality promotion | $r^{\mathcal{A}}(k, q)$ | Pearson's linear correlation between node indegree and node quality $q$ in networks grown with algorithm $\mathcal{A}$ |
| | $P_{100}^{\mathcal{A}}(k, q)$ | Precision of indegree in identifying the top-100 nodes by quality, in networks grown with algorithm $\mathcal{A}$ |
| Quality detection | $r^{\mathcal{A}}(s, q)$ | Pearson's linear correlation between the score produced by algorithm $\mathcal{A}$ and node quality $q$ |
| | $P_{100}^{\mathcal{A}}(s, q)$ | Precision of algorithm $\mathcal{A}$ in identifying the top-100 nodes by quality, in networks grown with algorithm $\mathcal{A}$ |
| Diversity | $N_{eff}$ | The effective number of nodes, defined as the inverse of the Herfindahl index, $H(\mathbf{k})$, associated with node indegree |

We emphasize that the rescaling procedure described above is only one among the possible ways to design a time-aware ranking algorithm [19]. Already in 2005, O'Madadhain and Smyth [32] recognized that static centrality metrics are based on time-aggregate network representations and, for this reason, they do not respect the sequence of events that led to the formation of the network. To solve this issue, O'Madadhain and Smyth introduced [32] and validated [33] a ranking algorithm based on the detailed contact time-series, acting as precursors of the recent stream of literature on centrality in temporal networks based on time-preserving paths [19]. We refer to Liao et al. [19] for a review of time-dependent ranking algorithms in complex networks. Nevertheless, we focus on the age-rescaled indegree here because of its simplicity and effectiveness in suppressing the indegree's bias towards old nodes [19,25]. Testing alternative time-dependent ranking algorithms within our model-generated benchmark graphs is an interesting possibility for future research.

### 3.3. Evaluating the algorithms' long-term impact

To assess the long-term impact of different algorithms, we grow random networks based on the model described above for each ranking algorithm $\mathcal{A}$. The algorithm determines, at any time, the ranking of the nodes that, in turn, determines the probability that a node receives a new connection, according to Eq. (1). We refer the networks generated with algorithm $\mathcal{A}$ as *$\mathcal{A}$-generated networks*. Ideally, we would expect a good ranking algorithm $\mathcal{A}$ to exhibit the three main properties introduced above: (i) *Quality promotion*: The algorithm generates networks where the final popularity of the nodes strongly correlates with their quality; (ii) *Quality detection*: The algorithm is effective in identifying high-quality nodes; (iii) *Popularity diversity*: The algorithm generates networks where the popularity is not strongly concentrated in a few nodes. In the following, we introduce three groups of observables to quantify these three properties – see Table 2 for a summary.

#### 3.3.1. Quality promotion

For a given ranking algorithm $\mathcal{A}$, we evaluate how well the final popularity $k$ of the nodes reproduces the inherent quality values $q$ for $\mathcal{A}$-generated networks. We do so by calculating the Pearson's linear correlation $r^{\mathcal{A}}(k, q)$ between node popularity $k$ and node quality $q$. While this metric takes into account all nodes, we are also interested in the algorithm's ability to promote the top-quality nodes. To this end, we measure the precision $P_{100}^{\mathcal{A}}(k, q)$ defined as the fraction of nodes that are placed in the top-100 of both the ranking by $k$ and the ranking by $q$.

One caveat applies to the quality-promotion evaluation metrics, $r^{\mathcal{A}}(k, q)$ and $P_{100}^{\mathcal{A}}(k, q)$. Due to the always-present possibility in the model of choosing the target by node quality, even the random ranking yields positive correlation and precision. This correlation increases as the nodes' sensitivity to quality increases (i.e., as $\beta$ decreases) – see Fig. 2 and the related discussion below. As the random ranking does not contain any useful information, it is clear that it cannot provide any positive contribution to the quality promotion. To quantify this effect, we measure the algorithms' quality detection metrics (see next paragraph).

#### 3.3.2. Quality detection

For a given ranking algorithm $\mathcal{A}$, we evaluate how well do the node scores assigned by $\mathcal{A}$ reproduce the inherent node quality values $q$ for $\mathcal{A}$-generated networks. To this end, we measure the Pearson's linear correlation $r^{\mathcal{A}}(s, q)$ between the node-level scores $s$ produced by the algorithm $\mathcal{A}$ (measured at the end of the network growth) and node quality $q$ for $\mathcal{A}$-generated networks. In parallel, we also measure the precision [21] $P_{100}^{\mathcal{A}}(s, q)$ of the algorithm, defined as the fraction of nodes that are placed in the top-100 of both the ranking by $s$ and the ranking by $q$.

#### 3.3.3. Popularity diversity

To quantify the ability of the algorithms to evenly spread the popularity across the network's nodes, and thus avoid its disproportionate concentration in a small number of nodes, we measure the indegree's Herfindahl index [14] $H(\mathbf{k})$.

$$H(\mathbf{k}) = \sum_{i=1}^{N} \left(\frac{k_i}{L}\right)^2, \tag{4}$$

where $L$ is the total number of links. The index is proportional to the variance of the network's indegree distribution: the smaller $H(\mathbf{k})$, the less concentrated the indegree in a small group of nodes. More specifically, the index ranges between $H_{\min} = 1/N$ (egalitarian network where all the nodes have indegree equal to $L/N$) and $H_{\max} = 1$ (network where one node has indegree $L$, and all the other nodes have indegree zero). Consequently, $N_{eff}(\mathbf{k}) = 1/H(\mathbf{k})$ can be interpreted as the



http://doc.rero.ch

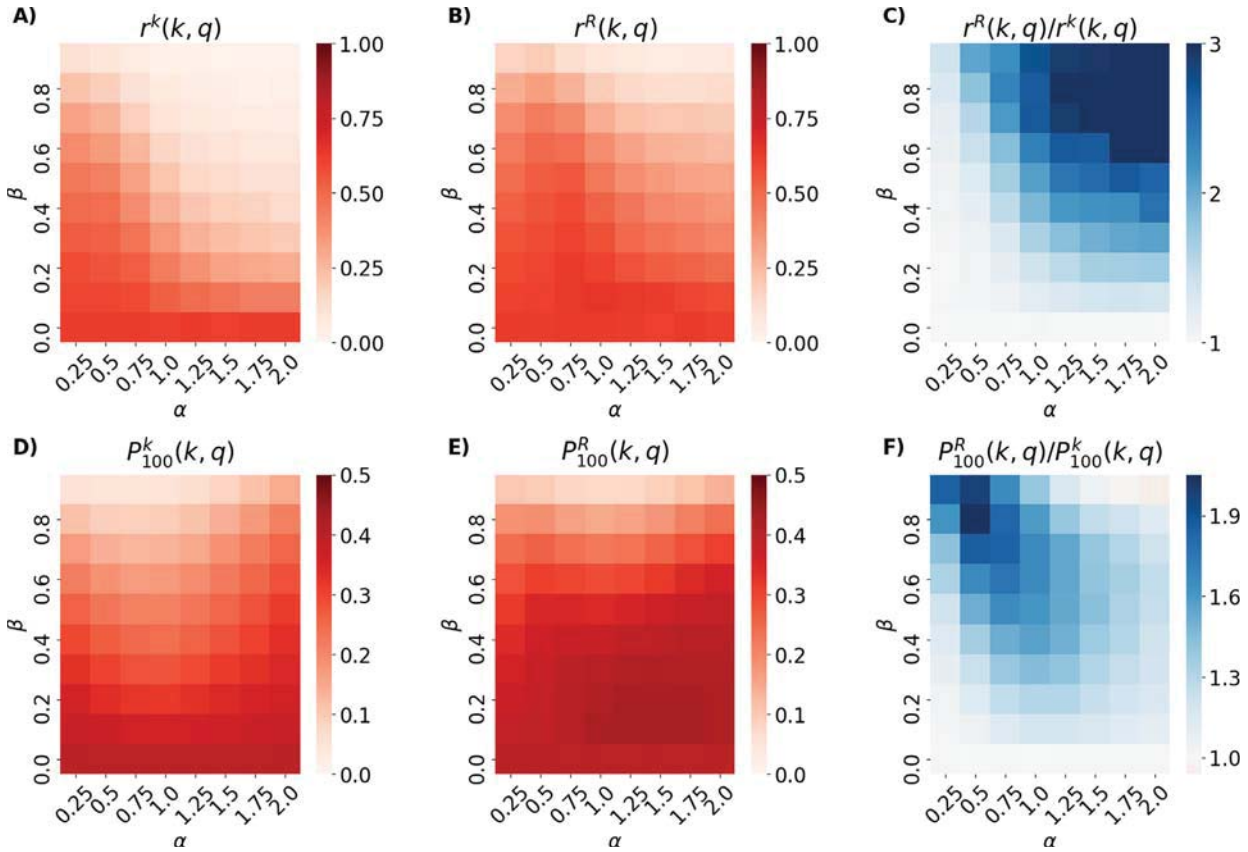


**Fig. 1.** Quality promotion as measured by $r(k, q)$ (the Pearson's linear correlation between node indegree $k$ and node quality $q$ – *top* panels), and $P_{100}(k, q)$ (the precision of node indegree $k$ in identifying the top-100 nodes by quality $q$ – *bottom* panels): a comparison between indegree-generated and $R(k)$-generated networks. (A-B): $r(k, q)$ for indegree-generated ($r^k(k, q)$, panel A) and $R(k)$-generated ($r^R(k, q)$, panel B) networks, as a function of the model parameters $\alpha$ (exploration cost) and $\beta$ (reliance on ranking). (C): Ratio $r^R(k, q)/r^k(k, q)$ as a function of the model parameters. (D-E): $P_{100}(k, q)$ for indegree-generated ($P^k_{100}(k, q)$, panel D) and $R(k)$-generated ($P^R_{100}(k, q)$, panel E) networks, as a function of the model parameters. (F): Ratio $P^R_{100}(k, q)/P^k_{100}(k, q)$ as a function of the model parameters. Results are averaged over 500 realizations.

"effective number of nodes" that received incoming links: $N_{eff}$ is equal to 1 if one single node received all the incoming links, and it is equal to $N$ if all the nodes received the same number of links. We posit that a good ranking algorithm should not produce too concentrated networks, and its generated networks should therefore exhibit relatively large values of $N_{eff}$.

## 4. Results

We grow networks of $N = 10,000$ nodes according to the model described above; we refer to Appendix A for all the simulation details. Here, we show the results for node outdegree $m = 6$; the results for $m = 3$ are commented below and shown in the SM (Figs. S2– (S13). Our goal is to compare the properties of networks generated by the four ranking algorithms defined in Section 3, according to the observables described above.

### *4.1. Quality promotion: impact of the age bias suppression*

Fig. 1A shows the indegree-quality Pearson's linear correlation $r(k, q)$ in indegree-generated networks, as a function of the model parameters $\alpha$ and $\beta$. The correlation between final popularity and quality is sensitive to both $\alpha$ and $\beta$. As $\alpha$ grows, it becomes harder for low-ranked high-quality nodes to acquire new incoming connections, which results in lower indegree-quality correlation. The (approximately) monotonous dependence of $r^k(k, q)$ on $\alpha$ was not found for the model of a static market[5] by Ciampaglia et al. [6], which indicates that it is a consequence of the network's growth. As $\beta$ grows, the nodes become less sensitive to quality and, as a direct consequence, the indegree-quality correlations deteriorates.

[5] By static market, we mean a collection of a fixed number of nodes. This is different from our growing network model where, at each time step, a new node enters the system and connects to the preexisting nodes.



http://doc.rero.ch

http://doc.rero.ch

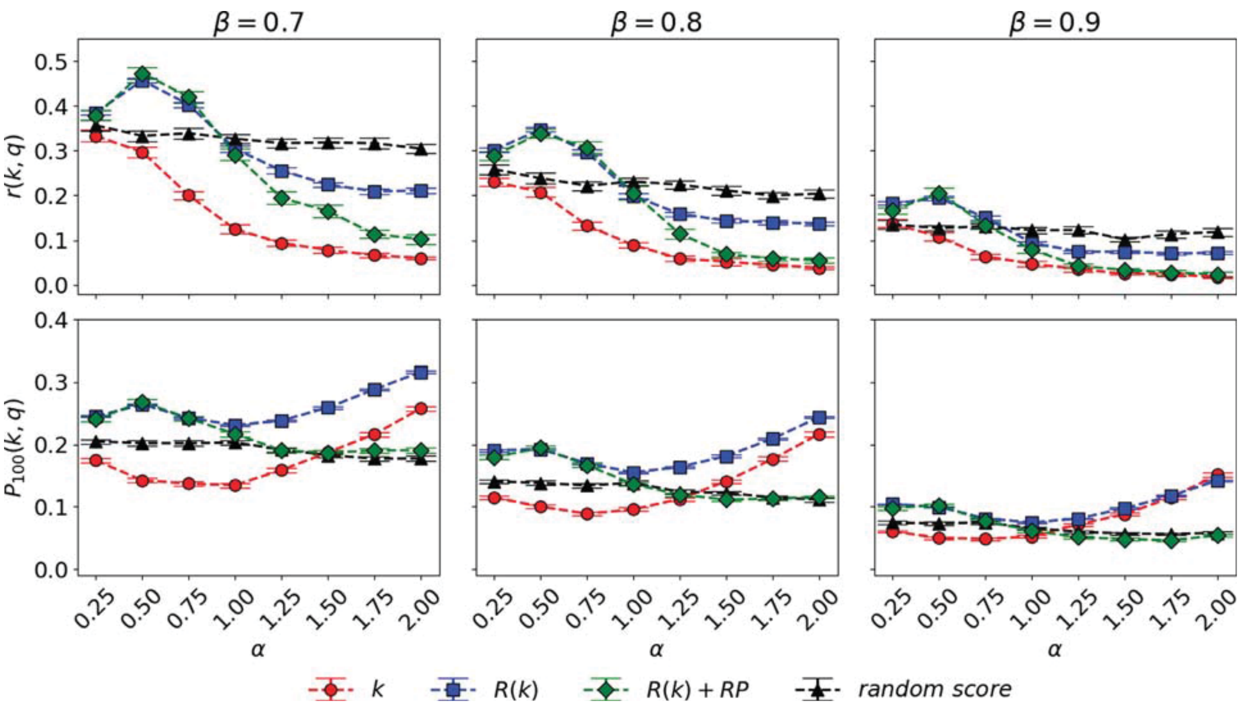


**Fig. 2.** Quality promotion as measured by $r(k, q)$ (the Pearson's linear correlation between node indegree $k$ and node quality $q$ – *top* panels), and $P_{100}(k, q)$ (the precision of node indegree $k$ in identifying the top-100 nodes by quality $q$ – *bottom* panels): a comparison between indegree-generated (red circles), $R(k)$-generated (blue squares), $R(k) + RP$-generated (green rhombuses), and random-generated networks (triangles). The three columns correspond, from left to right, to $\beta = 0.7, 0.8, 0.9$, respectively. Results are averaged over 100 realizations; the error bars represent the standard error of the mean.

Fig. 1B shows the correlation $r^R(k, q)$ between node indegree $k$ and node quality in $R(k)$-generated networks. The figure shows that by adopting the age-rescaled metric $R(k)$, the indegree-quality correlation stays large for a broader parameter region. For example, we observe values of $r^R(k, q)$ as high as 0.35 when $(\alpha, \beta) = (1, 0.7)$, which corresponds to a scenario where the nodes are driven by quality only three times out of ten.[6]

To visually appreciate the parameter regions where the indegree-quality correlations significantly differ between the two classes of networks, we represent the heatmap of the ratio $r^R(k, q)/r^k(k, q)$ (Fig. 1C). When $\beta = 0$, nodes are only sensitive to quality and the plotted ratio is one by definition (on average) because node ranking has no influence. When $\beta > 0$, the nodes become driven both by quality and by ranking, which makes it possible to observe differences between the networks grown with different algorithms. We find that the rescaled indegree produces networks with a higher indegree-quality correlation for all the parameter space; we observe the largest advantage of the rescaled indegree in terms of quality promotion for the region where both $\alpha$ and $\beta$ are relatively large – i.e., in the region where the nodes are unwilling to choose low-ranked nodes and, at the same time, are highly sensitive to ranking. Analogous heatmaps for the precision metrics (Fig. 1D–F) show that the indegree's precision in $R(k)$-generated networks is systematically larger than that in indegree-generated networks. Differently from Fig. 1C, Fig. 1F shows that the largest gaps between the precision in $R(k)$- and indegree-generated networks occur in the small $\alpha$, large $\beta$ region. We discuss the reasons behind the different trend for correlation and precision in Section 4.2.

### *4.2. Quality promotion: comparing the four ranking algorithms*

Fig. 1 indicates that adopting the rescaled indegree promotes node quality better than indegree for a broad range of model parameters. At the same time, it is important to compare the performance achieved by adopting rescaled indegree with that achieved by adopting rescaled indegree with random promotion and random ranking, respectively. As pointed out above, for $\beta < 1$, the nodes have a non-zero probability to choose their targets based on quality, which results in a positive indegree-quality correlation even for networks generated by adopting a random ranking. We focus on three values of $\beta$ ($\beta = 0.7, 0.8, 0.9$) which correspond to populations of nodes that are mostly driven by ranking when selecting their targets; analogous results for smaller values of $\beta$ are shown in Supplementary Figs. S16– S18 for $m = 6$ and S8–S10 for $m = 3$.

We find (Fig. 2, top panels) that the indegree-quality correlation observed in $R(k)$-generated networks is not always larger than the indegree-quality correlation observed in random-generated networks: as the exploration cost $\alpha$ grows, the

[6] As opposed to $r^k(k, q) = 0.12$ observed for indegree-generated networks for the same pair of $(\alpha, \beta)$ values.

indegree-quality correlation in $R(k)$-generated networks dwindles; when $\alpha$ is larger than one, $R(k)$-generated networks exhibit a smaller indegree-quality correlation than random-generated networks. While a large exploration cost is harmful for the overall indegree-quality correlation in both indegree-generated and $R(k)$-generated networks, for $\alpha \geq 1$, indegree's precision in promoting the top-quality nodes (Fig. 2, bottom panels) tends to grow with the exploration cost. Remarkably, $R(k)$-generated networks exhibit the largest precision values for all values of $\alpha$.

The behavior of the ($R(k)$+RP)-generated networks (i.e, the networks generated with the ranking by rescaled indegree with random promotion, $R(k)$+RP) is non-trivial. For small values of $\alpha$, the indegree-quality correlation and indegree's precision of these networks are comparable with those of $R(k)$-generated networks, which indicates that randomly promoting a few nodes to the top of the ranking is not harmful to quality promotion. On the other hand, for $\alpha > 1$, the indegree-quality correlation (or indegree's precision) of $R(k)$-generated networks becomes larger than that of ($R(k)$+RP)-generated networks. We conclude that for large values of $\beta$ ($\beta = 0.7, 0.8, 0.9$), the random promotion mechanism has a marginal impact on quality promotion for sparse networks when $\alpha < 1$, whereas it is harmful to quality promotion when the exploration cost is large, i.e., for $\alpha > 1$. For smaller values of $\beta$ ($\beta = 0.1, 0.3, 0.5$), the ($R(k)$+RP)-generated networks and the $R(k)$-generated networks exhibit comparable values of indegree-quality correlation and precision (Figs. S8– (S16).

The qualitative difference between the top and the bottom panels of Fig. 2 is explained by the different indegree distributions of the networks generated with different values of $\alpha$. When $\alpha$ is large, the incoming links concentrate on few top items (as the effective number of nodes shows, see Fig. 4 below and the related discussion) and, at the same time, the low-quality items remain unnoticed. The small number of incoming links received by low-quality items do not allow the metrics to discriminate their quality, which results in small indegree-quality correlation values. By contrast, high-quality nodes receive a large number of incoming links, and it is possible for the metrics to rank them at the top, which results in relatively large precision values.

### 4.3. Quality detection

In indegree-generated and $R(k)$-generated networks, node indegree and age-rescaled indegree, respectively, are the scores that are used for the nodes' ranking. Our ability to detect quality in such networks is determined by the strength of the relation between node score $s$ and $q$ (as measured by both the Pearson's linear correlation $r^s(s, q)$ and the precision $P^s_{100}(s, q)$). Remarkably, the correlation $r^R(R(k), q)$ is larger than the correlation $r^k(k, q)$ for all the parameter values (Fig. 3, top panels, and Fig. S14). The precision of age-rescaled indegree in identifying the top-quality nodes is also larger than indegree's precision for all the parameter values (Fig. 3, bottom panels). In qualitative agreement with the results obtained for quality promotion (Fig. 1), for $m = 6$ and $\beta = 0.7, 0.8, 0.9$, the random promotion mechanism turns out to have a marginal impact for $\alpha < 1$, whereas it is harmful for $\alpha > 1$. For $m = 3, 6$ and $\beta = 0.1, 0.3, 0.5$, the ($R(k)$+RP)-generated and $R(k)$-generated networks exhibit similar levels of score-quality correlation (see Figs. S9 and S17). By being completely insensitive to node popularity, the random score achieves on average zero precision in identifying the top-quality nodes. As expected, while the random score can still generate networks with non-zero indegree-quality correlation (Fig. 2), the rankings it produces have no practical utility.

### 4.4. Diversity

For both indegree- and $R(k)$-generated networks, the effective number of nodes $N_{eff}$ depends on both $\alpha$ and $\beta$ (Figs. 4). Unsurprisingly, the random ranking produces the most egalitarian networks (Fig. 4), with $N_{eff}$ values above $3000 = 0.3\,N$. In qualitative agreement with previous findings [39], for all the studied metrics, a larger sensitivity to ranking position (i.e., a larger $\alpha$ value) leads to a more unequal popularity distribution—this manifests itself in the decrease of $N_{eff}$ as $\alpha$ increases. Importantly, the popularity distribution is more even for $R(k)$-generated networks than for indegree-generated networks (Figs. 4 and S15), and the ($R(k)$+RP) algorithm further enhances popularity diversity. In summary, the age normalization procedure not only improves the indegree-quality and the score-quality correlation, but it also decreases the popularity inequality in the system; as expected, the random promotion mechanism further decreases the popularity inequality. At the same time, both the $R(k)$-generated and the ($R(k)$+RP)-generated networks exhibit $N_{eff}$ values significantly smaller than the $N_{eff}$ achieved by the random ranking. It remains open to design ranking algorithms that lead to more egalitarian networks than those generated with rescaled indegree and ($R(k)$+RP), yet maintaining a similar level of quality promotion and quality detection.

### 4.5. Network growth model with node removal

In real social and information systems, not only can new nodes enter the system, but also existing nodes can disappear or lose relevance. The members of an online community, for example, may lose interest in the platform and deactivate their accounts, which may eventually lead to the "death" of the platform [12]. In the WWW, many webpages are deleted every day, which makes it essential to incorporate node deletion into network growth models [18]. To take into account this situation, we introduce a variant of the model introduced in Section 3.1 by assuming that the nodes are in one of two possible states: "active" or "removed". At every time step $t$, each active node becomes removed with probability $\mu$, and one new active node enters the system and chooses its targets among the existing active nodes according to the rules described



http://doc.rero.ch

http://doc.rero.ch

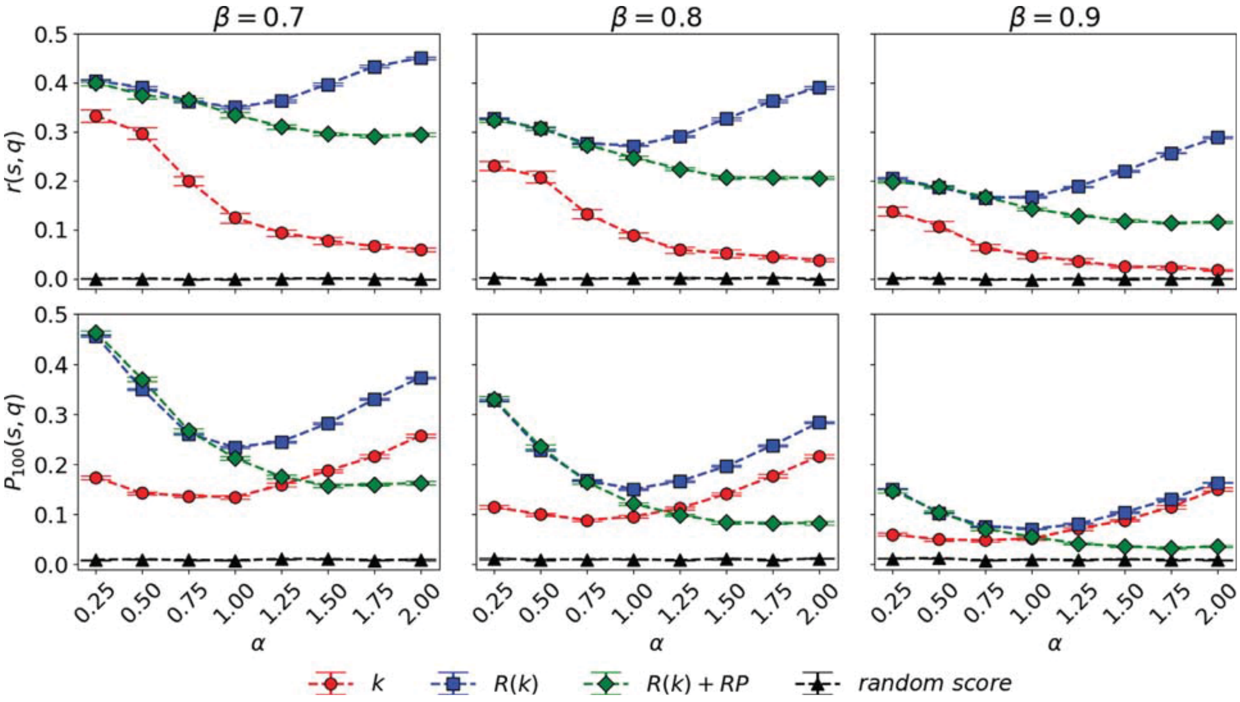


**Fig. 3.** Quality detection as measured by $r(s, q)$ (the Pearson's linear correlation between node score $s$ and node quality $q$ – *top* panels), and $P_{100}(s, q)$ (the precision of node score $s$ in identifying the top-100 nodes by quality $q$ – *bottom* panels): a comparison between indegree-generated ($s = k$, red circles), $R(k)$-generated ($s = R(k)$, blue squares), $(R(k) + RP)$-generated ($s = R(k)$, green diamonds), and random-generated networks ($s$ is given by a random score, black triangles). The three columns correspond, from left to right, to $\beta = 0.7, 0.8, 0.9$, respectively. The dots represent averages over 100 realizations; the error bars represent the standard error of the mean.

in Section 3.1. In the continuum approximation, the number of active nodes, $A(t)$, follows the equation $\dot{A}(t) = 1 - \mu A(t)$ whose solution (with the initial condition $A(1) = 1$) is

$$A(t) = \mu^{-1} + (1 - \mu^{-1}) \exp(-\mu (t - 1)). \tag{5}$$

After an initial linear growth ($A(t) \approx t$ for $t \ll \mu^{-1}$), the number of active nodes eventually equilibrates at $A^* = 1/\mu$. The validity of Eq. (5) is confirmed by our numerical simulations (see Fig. 5F). Note that when $\mu = 0$ (i.e., without node removal), $A(t) = t$ and $N = t^*$, where $t^*$ denotes the total number of performed simulation steps. In the following, we set $t^* = 10^4$; due to the node removal mechanism, the number $A(10^4)$ of active nodes at the end of the simulation is substantially smaller than $10^4$, as expected based on Eq. (5) (see Fig. 5F).

We investigate whether the node removal mechanism substantially alters the results described above. More specifically, we perform numerical simulations with values of $\mu$ ranging from $10^{-4}$ to $5 \cdot 10^{-4}$, which corresponds to expected values of the asymptotic number of active nodes, $A^* = 1/\mu$, ranging from 2,000 to 10,000. We find that through the whole range of $\mu$ values, the results are qualitatively similar to those obtained for the networks without node removal (see Fig. 5A-E for the results for $\alpha = 0.5$, $\beta = 0.7$, and Figs. S25 for $\alpha = 1$, $\beta = 0.7$). All the considered observables display remarkable stability with respect to $\mu$, which indicates that node removal has a marginal impact on the algorithms' quality promotion, quality detection, and popularity diversity. It remains open to determine the largest removal probability $\mu$ tolerated by the system before the results change qualitatively.

### *4.6. A case study: growing a network of scientific papers based on different metrics*

So far, we have considered model-generated networks. How are our model and our results on ranking algorithms relevant to real systems? If our model provides a plausible description of the growth of a given system of interest, and we are able to infer the $(\alpha, \beta)$ parameters from the available data, we should be able to quantify the potential impact of various ranking algorithms on the future properties of the system. Stimulated by this observation, in this Section, we analyze real information networks to fit our model's parameters to the empirical data, and use the resulting parameters to grow again the network based on different ranking algorithms.

http://doc.rero.ch

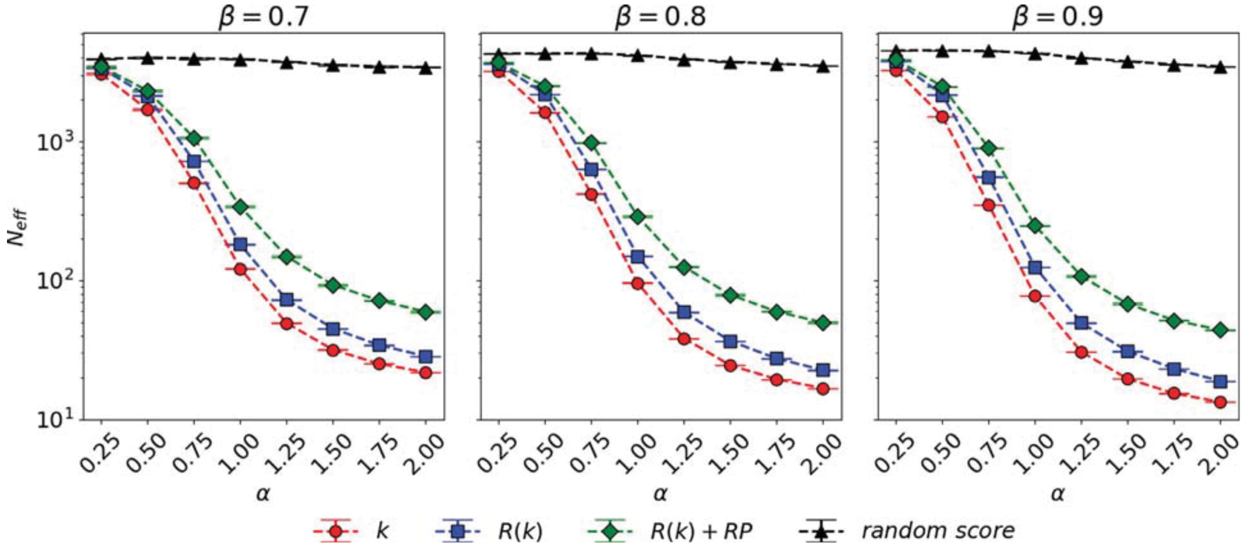


**Fig. 4.** Diversity as measured by $N_{eff}$(the larger the value, the more egalitarian the indegree distribution): a comparison between indegree-generated (red circles), $R(k)$-generated (blue squares), $(R(k) + RP)$-generated (green diamonds) and random-generated (black triangles) networks. The three columns correspond, from left to right, to $\beta = 0.7, 0.8, 0.9$, respectively. The dots represent averages over 100 realizations; the error bars represent the standard error of the mean.

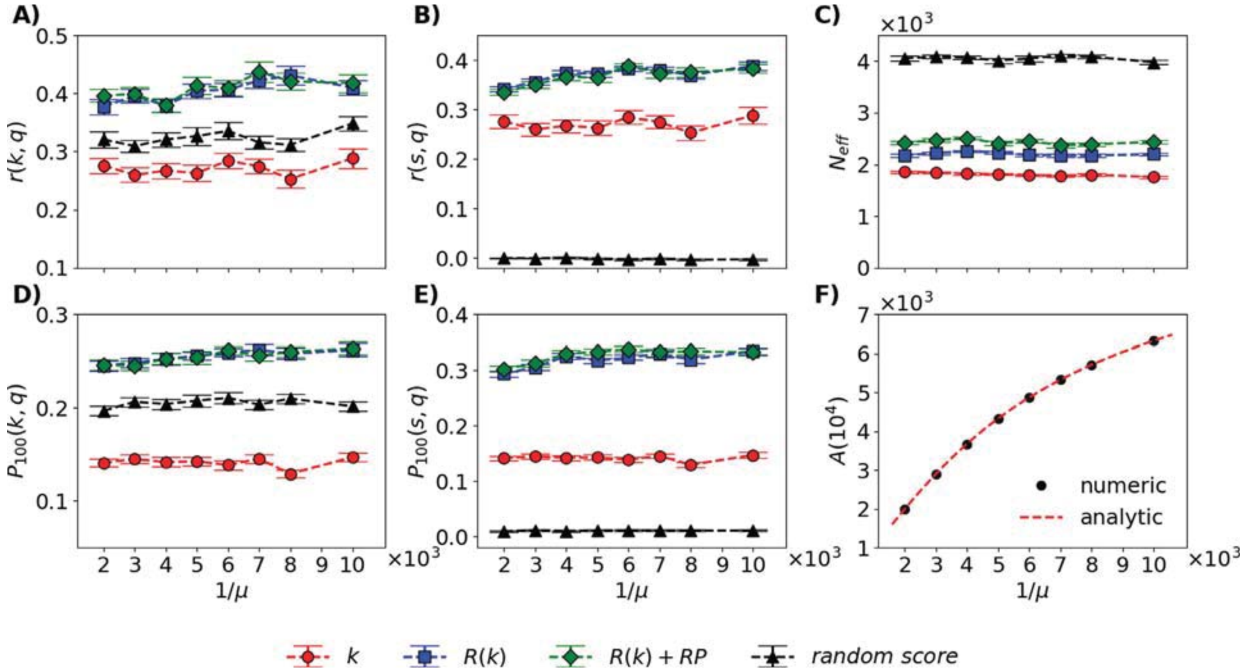


**Fig. 5.** The impact of node removal on quality promotion, quality detection, and popularity diversity. We grow networks generated with the model described in Section 4.5 with $\alpha = 0.5$, $\beta = 0.7$, $T = 10^{-4}$. We show five ranking evaluation metrics as a function of the inverse of the removal probability, $A^* = \mu^{-1}$: (A) the Pearson's correlation between node indegree and node quality; (B) the Pearson's correlation between node score and node quality (see Fig. 3's caption for the definition of node score); (C) the effective number of nodes, $N_{eff}$; (D) the indegree's precision in identifying the top-100 nodes by quality; (E) the score's precision in identifying the top-100 nodes by quality. As in the previous figures, different lines correspond to the networks generated with different algorithms. The symbols represent averages over 50 realizations; the error bars represent the standard error of the mean. Panel F shows the number of active nodes at time $t = 10^4$, i.e., at the time when we halt the simulations, as a function of $\mu^{-1}$; the values computed analytically through Eq. (5) well match the values observed in the simulations.

The information networks analyzed here are subsets of the American Physical Society (APS) citation network of scientific papers.[7] The citation dataset provided by the APS contains all 539974 papers published by the APS from 1893 to 2013 together with their 5992897 references to other papers published by APS journals and their publication dates. For our analysis, we extract two subsets of papers that include the papers with the PACS number[8] 89.75.Hc ("Networks and genealogical trees"; the subset includes $N = 1615$ papers and $L = 8222$ citations) and 03.67.Lx ("Quantum computation architectures and implementations"; the subset includes $N = 3876$ papers and $L = 24213$ citations), respectively.

To quantify the potential impact of different ranking algorithms, the first step is to infer the optimal pair $(\hat{\alpha}, \hat{\beta})$ of model parameters together with the optimal algorithm that together lead to the best agreement between the $\mathcal{A}$-generated model networks and the observed real network. The real data determine the final number of nodes in the network, their order of appearance, and their outdegree values; in an $\mathcal{A}$-generated model network, the nodes choose their references based on the rules of the model described in Section 3.1, where the ranking position of the nodes is determined by the chosen algorithm $\mathcal{A}$.

Quality is not accessible in real data as opposed to synthetic networks where it is a well-defined node-level variable. To overcome this limitation, as rescaled indegree is the best-performing metric in quantifying node quality in synthetic networks (Fig. 3), we use the papers' final rescaled indegree score $R(k)$ as a proxy for their quality if $R(k) \geq 0$, whereas we set $q = 0$ for papers whose $R(k)$ is negative.

The agreement between the $\mathcal{A}$-generated networks and the original network is quantified by the model's mean error per node, $e^{(\mathcal{A},\alpha,\beta)}$. For each $\mathcal{A}$-generated network[9] $\mathcal{G}(\mathcal{A}, \alpha, \beta)$, we first define the network's mean error per node, $e(\mathcal{G}(\mathcal{A}, \alpha, \beta))$, as

$$e(\mathcal{G}(\mathcal{A}, \alpha, \beta)) = \frac{1}{N} \sum_{i=1}^{N} \left| k_i(\mathcal{G}(\mathcal{A}, \alpha, \beta)) - k_i^* \right|, \tag{6}$$

where $k_i(\mathcal{G})$ denotes paper $i$'s indegree in network $\mathcal{G}$, whereas $k_i^*$ denotes the observed indegree of paper $i$. The model's mean error per node, $e^{(\mathcal{A},\alpha,\beta)}$ is defined as the average of $e(\mathcal{G})$ over a sample of 100 $\mathcal{A}$-generated networks. Essentially, the model's mean error per node quantifies the average difference between node indegree in real and model networks. The lower $e^{(\mathcal{A},\alpha,\beta)}$, the better the agreement between the $\mathcal{A}$-generated model networks and the real network.

Finally, we compare the properties of the networks generated by different ranking algorithms for $(\alpha, \beta) = (\hat{\alpha}, \hat{\beta})$. The rationale behind such a comparison is that the parameter values $(\hat{\alpha}, \hat{\beta})$ that yield the best agreement represent our best estimations of the nodesâ exploration cost $\alpha$ and sensitivity to popularity $\beta$. Based on this assumption, we estimate the long-term implications of various algorithms by simply changing the algorithm that is used to grow the network, whilst keeping $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$ fixed.

For the subset that corresponds to the PACS code 89.75.Hc, we find that indegree-generated networks with $\hat{\alpha} = 0.5$ and $\hat{\beta} = 0.5$ exhibit the best agreement with the real network (mean error per node $e^{(k,0.5,0.5)} = 3.1$, see Fig. 6A. Therefore, we fix $\alpha = 0.5$ and $\beta = 0.5$, and we compare the networks generated by three different algorithms: indegree, age-rescaled indegree, and the ranking by rescaled indegree with random promotion. In qualitative agreement with our previous results, we find that $R(k)$-generated networks and ($R(k)$+RP)-generated networks exhibit substantially larger popularity diversity than $k$-generated networks and the original network (see Fig. 6B). Qualitatively similar results are obtained for the subset that corresponds to the PACS code 03.67.Lx (see Fig. 6C-D). These results confirm that in a real information system, suppressing the age bias of popularity metrics can increase popularity diversity. Beyond numerical simulations, we envision that future studies might further validate this conclusion by means of field experiments where subjects in different groups can select various information items based on the items' ranking position by different ranking algorithms.

## 5. Discussion

To summarize, we find that age-rescaled indegree allows us not only to fairly compare old and recent nodes [25], but also to generate networks that exhibit a larger indegree-quality correlation and a more even popularity distribution than the networks generated with indegree. Examples of widely-used cumulative popularity metrics include the number of views or downloads for online content, the number of received citations for scientific papers, among others. Our results indicate that despite the widespread use of cumulative popularity metrics, age-rescaled metrics may better help both users to find high-quality content, and high-quality content to experience larger success than low-quality content. The random promotion mechanism introduced here consistently improves popularity diversity, whereas its impact on quality promotion and detection ranges from marginal to substantial depending on the exploration cost parameter $\alpha$.

The main message of our work is that network-based growth models can help us to understand the impact of network growth mechanisms on the rankings by a given algorithm [19,24,27], as well as to estimate the impact of the adoption of different ranking algorithms by a given system. In other words, we can investigate not only how the past evolution of

[7] The dataset has been already analyzed in [25,26,28], and it can be downloaded here: https://journals.aps.org/datasets.

[8] The PACS (Physics and Astronomy Classification Scheme) codes refer to a hierarchical classification of research areas in physics, astronomy, and related sciences. We refer to https://journals.aps.org/PACS for further information.

[9] An $\mathcal{A}$-generated network is a realization of the stochastic process that generates $\mathcal{A}$-generated networks.

http://doc.rero.ch

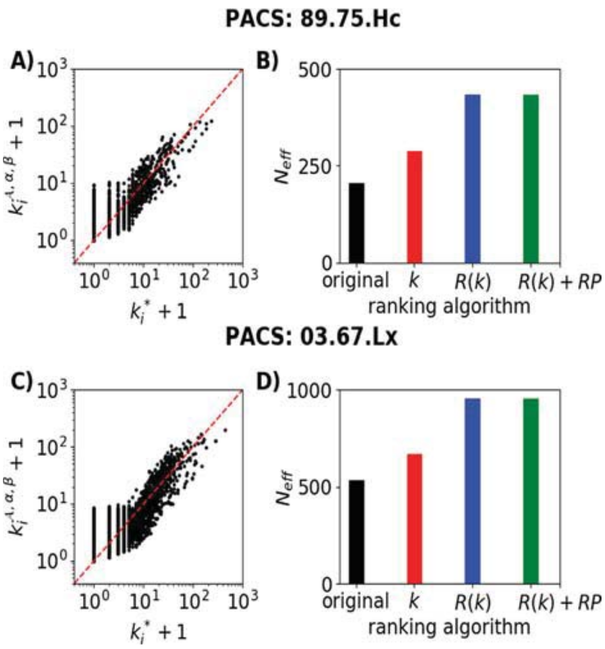


**Fig. 6.** A comparison of two real networks (subsets of the APS citation network) with the respective synthetic networks generated with the model described in Section 3.1, using different ranking algorithms. Panels (A,C) show the scatter plots between the nodes' original indegree $k^*$ and the nodes' average indegree $\bar{k}$ in indegree-generated networks ($\alpha = \hat{\alpha},\ \ \beta = \hat{\beta}$) generated as described in the main text, for the APS network's subsets that correspond to the PACS codes 89.75.Hc and 03.67.Lx, respectively. Panels (B,D) show the averages of $N_{eff}$ over 100 realizations of the model networks generated by different ranking algorithms. The ($R(k)$+RP)-generated ($\eta = 0.5, T = 10$) and $R(k)$-generated networks exhibit significantly larger values of $N_{eff}$ than the original network and the indegree-generated network. This indicates that in a real information system, suppressing the age-bias of popularity-based metrics can substantially improve popularity diversity.

http://doc.rero.ch

the system influenced the current rankings, but also how the adopted rankings may influence the future evolution of the system. As a result, our ranking-driven and quality-driven growing network model can be interpreted as a generative model for benchmark graphs to evaluate the performance of ranking algorithms in terms of quality promotion, quality detection, and popularity diversity.

We stress that the ranking algorithms considered here are simple in the sense of being based on a single criterion that is furthermore readily quantified (e.g., the ranking by node indegree). In a general scenario where multiple criteria are to be considered, especially when they are contradicting such as node popularity and novelty, the existing vast body of literature on multiple-criteria decision analysis [13] – with techniques such as the analytic hierarchy process [38] and outranking [37] – becomes relevant. This goes beyond directly combining node scores by various metrics [31], yet it can become relevant when designing an information filtering system for real users with their heterogeneous, and often contradictory, needs and preferences.

Our work sheds light on the long-studied relation between quality/talent and success [39,48]: Do the high-quality nodes experience larger success than the low-quality nodes? Why nodes of similar worthiness experience widely different success? In real systems, addressing these questions is challenging as defining “node quality” in an unbiased and objective way is often not possible. Our model-based approach bypasses this obstacle by defining node quality as an intrinsic node property, and by building multiple independent realizations of an artificial system where the nodes choose their connections based on both the other nodes' ranking and their quality. At the same time, while the model studied here is arguably one of the simplest models which feature all the elements of interest in our analysis (network growth, and the joint influence of ranking and quality on network growth), it can only provide a stylized description of the growth of real networks.

Finally, the application of our growing network model to the American Physical Society citation network of scientific papers constitutes an attempt to estimate the potential consequences of various ranking algorithms on a real information system. We envision that more sophisticated models together with suitable field experiments will improve the reliability of model-based predictions of the effects of ranking algorithms, providing us with a robust basis for more informed choices of ranking algorithms for real-world applications. To draw a parallel, in a similar way as high-resolution models of epidemic spreading have led to accurate predictions of the properties of disease outbreaks [30], detailed models of network evolution may lead to the accurate quantification of the consequences of the adoption of a given metric in a given system.

http://doc.rero.ch

**Author contributions statement**

M.S.M. and M.M. conceived the idea, M.S.M. and L.L. designed research, S.Z. performed the numerical simulations, S.Z. and M.S.M. performed the analytic computations, all authors analyzed and discussed the results. S.Z. and M.S.M. wrote the manuscript. All authors reviewed the manuscript.

**Acknowledgments**

We thank Yi-Cheng Zhang for many enlightening discussions on the topic. This work has been supported by the National Natural Science Foundation of China (Grants Nos. 11622538, 61673150), the Science Strength Promotion Program of the University of Electronic Science and Technology of China (Grant No. Y030190261010020), and the Natural Science Foundation of Zhejiang Province of China (Grant no. LR16A050001). MSM acknowledges the University of Zürich for support through the URPP Social Networks.

**Appendix A. Details on the numerical simulations**

We focus on networks composed of $10^4$ nodes, and study the following model parameters: $\alpha$ from 0.25 to 2.0 with step 0.25 and $\beta$ from 0 to 1 with step 0.1. Node quality values are drawn from the Pareto distribution $P(q) \sim q^{-3}$ where $q \in [1, \infty)$. The network is initialized with a network of $m$ nodes, each of them with one outgoing and incoming link. At each time step $t > m+1$, a new node $t$ is added to the system and $m$ nodes (only results for $m = 6$ are shown in the main text, whereas results for both $m = 3$ and $m = 6$ are shown in the Supplementary Material) are chosen as targets to establish $m$ new links. With probability $\beta$, the probability that a given node is chosen is given by Eq. (1), with probability $1 - \beta$, it is given by Eq. (2). To save computational time, for times $t \leq 10^2$, we re-compute and update the rankings at each time step, whereas for times $t \geq 10^2$, the newly introduced nodes are placed at the bottom of the node ranking, and we re-compute and update the rankings every 10 time steps. To make our results insensitive to random fluctuations, for each parameter pair $(\alpha, \beta)$, all the results shown here represent averages over a sufficiently large number of realizations of the network growth process.

**Appendix B. The relation between popularity and quality in indegree-generated networks**

We start by considering networks where the nodes cannot perceive the other nodes ranking, and are completely driven by quality ($\beta = 1$). In this scenario, the average indegree of node $i$ at time $N$ is given by

$$\overline{k_i}(N) = \sum_{t=i+1}^{N} m \frac{q_i}{\sum_{j=1}^{t-1} q_j}. \tag{B.1}$$

In the thermodynamic limit $N \gg i$, by using a similar mean-field approximation as in [10], we obtain

$$\overline{k_i}(N) \simeq m \frac{q_i}{\overline{q}} \log\left(\frac{N-1}{i-1}\right). \tag{B.2}$$

There is a good agreement between Eq. (B.2) and the results of numerical simulations (see Supplementary Fig. S26). Such linear relation between indegree and quality does not hold for $\beta > 0$, where the analytic calculation is made difficult by the fact that the ranking position of a given node at a given time is influenced by both its quality and the previous dynamics of the system. Nevertheless, we find that the relation $\overline{k_i}(N) = C q_i^{\delta}$ fits reasonably well the simulation results, and the dependence of the fitted exponent $\delta$ on node age is relatively weak (see Fig. S27 and Table S1 for details).

**Supplementary material**

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ins.2019.03.021.

**References**

[1] P. Boldi, S. Vigna, Axioms for centrality, Internet Math. 10 (3–4) (2014) 222–262.
[2] E. Bozdag, Bias in algorithmic filtering and personalization, Ethics Inf. Technol. 15 (3) (2013) 209–227.
[3] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1) (1998) 107–117.
[4] H. Cai, Y. Chen, H. Fang, Observational learning: evidence from a randomized natural field experiment, Am. Econ. Rev. 99 (3) (2009) 864–882.
[5] J. Cho, S. Roy, Impact of search engines on page popularity, in: Proceedings of the 13th International Conference on World Wide Web, ACM, 2004, pp. 20–29.
[6] G.L. Ciampaglia, A. Nematzadeh, F. Menczer, A. Flammini, How algorithmic popularity bias hinders or promotes quality, Sci. Rep. 8 (2018).
[7] I.D. Couzin, C.C. Ioannou, G. Demirel, T. Gross, C.J. Torney, A. Hartnett, L. Conradt, S.A. Levin, N.E. Leonard, Uninformed individuals promote democratic consensus in animal groups, Science 334 (6062) (2011) 1578–1580.
[8] A. Diaz, Through the Google Goggles: Sociopolitical Bias in Search Engine Design, Web Search, 2008, pp. 11–34.
[9] R. Epstein, R.E. Robertson, The search engine manipulation effect (seme) and its possible impact on the outcomes of elections, Proc. Natl. Acad. Sci. 112 (2015) E4512–E4521.

[10] S. Fortunato, A. Flammini, F. Menczer, Scale-free network growth by ranking, Phys. Rev. Lett. 96 (21) (2006) 218701.
[11] S. Fortunato, A. Flammini, F. Menczer, A. Vespignani, Topical interests and the mitigation of search engine bias, Proc. Natl. Acad. Sci. 103 (2006) 12684–12689.
[12] D. Garcia, P. Mavrodiev, F. Schweitzer, Social resilience in online communities: the autopsy of friendster, in: Proceedings of the First ACM Conference on Online Social Networks, ACM, 2013, pp. 39–50.
[13] S. Greco, J. Figueira, M. Ehrgott, Multiple Criteria Decision Analysis, Springer, 2016.
[14] O.C. Herfindahl, Copper Costs and Prices: 1870–1957, 1959, Baltimore, Published, P, 1959. Published for Resources for the Future by Johns Hopkins Press.
[15] M. Hindman, K. Tsioutsiouliklis, J.A. Johnson, Googlearchy: how a few heavily-linked sites dominate politics on the web, Annual Meeting of the Midwest Political Science Association, 2003.
[16] D. Jannach, K. Hegelich, A case study on the effectiveness of recommendations in the mobile internet, in: Proceedings of the Third ACM Conference on Recommender Systems, ACM, 2009, pp. 205–208.
[17] J.-J. Jiang, Z.-G. Huang, L. Huang, H. Liu, Y.-C. Lai, Directed dynamical influence is more detectable with noise, Sci. Rep. 6 (2016) 24088.
[18] J.S. Kong, N. Sarshar, V.P. Roychowdhury, Experience versus talent shapes the structure of the web, Proc. Natl. Acad. Sci. 105 (2008) 13724–13729.
[19] H. Liao, M.S. Mariani, M. Medo, Y.-C. Zhang, M.-Y. Zhou, Ranking in evolving complex networks, Phys. Rep. 689 (2017) 1–54.
[20] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, T. Zhou, Vital nodes identification in complex networks, Phys. Rep. 650 (2016) 1–63.
[21] L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender systems, Phys. Rep. 519 (1) (2012) 1–49.
[22] L. Lü, Y.C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS ONE 6 (6) (2011) e21202.
[23] L. Lü, T. Zhou, Q.M. Zhang, H.E. Stanley, The h-index of a network node and its relation to degree and coreness, Nat. Commun. 7 (2016) 10168.
[24] M.S. Mariani, M. Medo, Y.-C. Zhang, Ranking nodes in growing networks: when PageRank fails, Sci. Rep. 5 (2015).
[25] M.S. Mariani, M. Medo, Y.-C. Zhang, Identification of milestone papers through time-balanced network centrality, J. Informetr. 10 (4) (2016) 1207–1223.
[26] M.S. Mariani, M. Medo, F. Lafond, Early identification of important patents: design and validation of citation network metrics, Technol. Forecast. Soc. Change (2018).
[27] M. Medo, G. Cimini, Model-based evaluation of scientific impact indicators, Phys. Rev. E 94 (3) (2016) 032312.
[28] M. Medo, G. Cimini, S. Gualdi, Temporal effects in the growth of networks, Phys. Rev. Lett. 107 (23) (2011) 238701.
[29] F. Morone, H.A. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (7563) (2015) 65.
[30] M. Newman, Networks: An Introduction, Oxford University Press, 2010.
[31] K. Okamoto, W. Chen, X.-Y. Li, Ranking of closeness centrality for large-scale social networks, in: International Workshop on Frontiers in Algorithmics, Springer, 2008, pp. 186–195.
[32] J. O'Madadhain, P. Smyth, Eventrank: a framework for ranking time-varying networks, in: Proceedings of the 3rd International Workshop on Link Discovery, ACM, 2005, pp. 9–16.
[33] J. O'Madadhain, J. Hutchins, P. Smyth, Prediction and ranking algorithms for event-based network data, ACM SIGKDD Explor. Newslett. 7 (2) (2005) 23–30.
[34] N. Perra, L.E.C. Rocha, Modelling opinion dynamics in the age of algorithmic personalisation, 2018, arXiv:1811.03341.
[35] F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Diffusion of scientific credits and the ranking of scientists, Phys. Rev. E 80 (5) (2009) 056103.
[36] Z.-M. Ren, M.S. Mariani, Y.-C. Zhang, M. Medo, Randomizing growing networks with a time-respecting null model, Phys. Rev. E 97 (5) (2018) 052311.
[37] B. Roy, The outranking approach and the foundations of electre methods, in: Readings in Multiple Criteria Decision Aid, Springer, 1990, pp. 155–183.
[38] T.L. Saaty, Analytic hierarchy process, in: Encyclopedia of Operations Research and Management Science, Springer, 2013, pp. 52–64.
[39] M.J. Salganik, P.S. Dodds, D.J. Watts, Experimental study of inequality and unpredictability in an artificial cultural market, Science 311 (5762) (2006) 854–856.
[40] I. Scholtes, R. Pfitzner, F. Schweitzer, The social dimension of information ranking: a discussion of research challenges and approaches, in: Socioinformatics-The Social Impact of Interactions between Humans and IT, Springer, 2014, pp. 45–61.
[41] A. Sharma, J.M. Hofman, D.J. Watts, Estimating the causal impact of recommendation systems from observational data, in: Proceedings of the Sixteenth ACM Conference on Economics and Computation, ACM, 2015, pp. 453–470.
[42] H. Shirado, N.A. Christakis, Locally noisy autonomous agents improve global human coordination in network experiments, Nature 545 (7654) (2017) 370.
[43] R. Sinatra, D. Wang, P. Deville, C. Song, A.-L. Barabási, Quantifying the evolution of individual scientific impact, Science 354 (6312) (2016) aaf5239.
[44] G. Vaccario, M. Medo, N. Wider, M.S. Mariani, Quantifying and suppressing ranking bias in a large citation network, J. Informetr. 11 (3) (2017) 766–782.
[45] L. Waltman, A review of the literature on citation impact indicators, J. Informetr. 10 (2) (2016) 365–391.
[46] D. Wang, C. Song, A.-L. Barabási, Quantifying long-term scientific impact, Science 342 (6154) (2013) 127–132.
[47] L. Weng, A. Flammini, A. Vespignani, F. Menczer, Competition among memes in a world with limited attention, Sci. Rep. 2 (2012) 335.
[48] B. Yucesoy, X. Wang, J. Huang, A.-L. Barabási, Success in books: a big data approach to bestsellers, EPJ Data Sci. 7 (1) (2018) 7.
[49] A. Zeng, C.H. Yeung, M. Medo, Y.-C. Zhang, Modeling mutual feedback between users and recommender systems, J. Stat. Mech. (7) (2015) P07020.
[50] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J.R. Wakeling, Y.-C. Zhang, Solving the apparent diversity-accuracy dilemma of recommender systems, Proc. Natl. Acad. Sci. 107 (2010) 4511–4515.

http://doc.rero.ch