PageRank Under Interpolation Between Undirected- and Directed Networks - A Case Study

Florian Henning(✉), Remco van der Hofstad, and Nelly Litvak

Eindhoven University of Technology, 5600MB Eindhoven, The Netherlands
{f.b.henning,r.w.v.d.hofstad,n.v.litvak}@tue.nl

Abstract. Among centrality measures, PageRank is particularly famous due to its implementation in the Google search engine. We have recently shown that in general undirected networks, the graph-normalized PageRank of any node in the network is bounded from above by its degree. This general statement, however, is not true in directed networks, where, e.g., the directed version of the preferential attachment model exhibits *heavier* tails for the limiting PageRank distribution than for the limiting in-degree-distribution. In this note, we illustrate the general upper bound on PageRank in undirected networks by scatter plots of datasets from three scale-free real-world networks of different sizes. We observe and explain a concentration phenomenon within the scatter plots. Furthermore, to shed light on how the directed-ness of edges changes the relation between degrees and PageRank, we interpolate between undirected and directed graphs as follows: for each of the three networks, we construct a new - directed - network by randomly choosing a subset of the edges of prescribed size and for each of its elements deleting exactly one of the two possible directions. As a result of this procedure, some small-degree vertices will obtain a PageRank that is above their in-degree.

We illustrate and explain this phenomenon for the chosen datasets by comparing to what happens in the configuration model.

Keywords: Undirected network · Network centrality · PageRank · Power-law hypothesis · Power-law distributions

1 Introduction

1.1 Scale-Free Networks

A characteristic property of many real-world networks, e.g., online social networks, the web graph or router networks, is the absence of a typical scale for the degrees of nodes within the network. This means that there is a small, yet highly relevant, proportion of nodes with very high degrees (e.g., [10]). On the level of the degree distribution $(p_k)_{k\in\mathbb{N}}$, the absence of a typical scale in such a network corresponds to a power-law degree distribution, i.e.,

$$p_k \approx ak^{-\tau},$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
M. Bloznelis et al. (Eds.): WAW 2025, LNCS 15699, pp. 96–108, 2025.
https://doi.org/10.1007/978-3-031-92898-7_7

for both the network size n and k large, where $\approx$ denotes an unspecified approximation. In particular, the probability for a uniformly chosen node from the network to have a degree larger than k (i.e., the tail distribution at k) is approximated by

$$\mathbb{P}(d_{V_n} > k) \approx b k^{-(\tau-1)},$$

for some $b > 0$. In this note, we investigate the relation between power-law degree distributions and the tails of the PageRank distribution. Let us next define PageRank, and more generally, centrality measures in networks.

1.2 Network Centrality Measures

Network centrality measures are employed to establish a hierarchy between nodes in a (potentially large) network, and to identify the influential ones. Presumably, the most simple and intuitive network centrality measures are the (in-)degree and Google's PageRank. In the recent thesis [15], comparison techniques for centrality measures (in particular the Closeness Centrality Curve (CCC) comparing the overlap of the k most relevant nodes according to the considered centrality measures) are discussed and applied to the established ones such as in- and out-degree, closeness, betweeness, harmonic centrality, Katz centrality and the PageRank with different damping factors. This comparison is done on real-word network data sets, as well as on directed and undirected versions of the configuration model.

In view of the results of [15], we focus the present work on the comparison between PageRank and the (in-)degree, based on what is known as the power-law hypothesis.

1.3 PageRank and the Power-Law Hypothesis

PageRank. PageRank [6] is an influential centrality measure initially introduced and implemented by the Google search engine. The PageRank distribution is nothing but the unique stationary distribution of the *easily-bored surfer* Markov chain, which, in each time step and with probability $c \in [0,1]$, takes a simple random walk step from its current position, and with probability $1-c$ jumps to a node chosen uniformly from the entire network. Here, the *damping factor c interpolates* between being mainly local, or being more global instead. The PageRank vector is the solution to the linear system of equations

$$\pi(i) = c \sum_{j \to i} \frac{1}{d_j^{\text{out}}} \pi(j) + \frac{1-c}{n}.$$

To have PageRank on a similar scale as the (in-)degrees, it is convenient to go over to the graph-normalized PageRank, which is defined as

$$R(i) = n\pi(i),$$

where n is the graph size, so that the average PageRank R is equal to 1. Throughout this note, by PageRank we mean its graph-normalized version.

PageRank in Scale-Free Networks. Empirical data, particularly for the web graph, has shown that a power-law degree (in-)degree distribution often leads to a power-law for the PageRank distribution, with the same power-law exponent. The question of the generality of these empirical observations is coined the *power-law hypothesis*.

1.4 Heavier PageRank Tails in Directed Preferential Attachment Models

The (undirected) preferential attachment model (also called the *Barabási-Albert* model) [5] was introduced to describe the emergence of power laws in real-life networks. The preferential attachment model is a sequence of random graphs $(\mathrm{PA}_n^{m,\delta})_{n\in\mathbb{N}}$, whose size linearly grows with the time n. There are several versions of the preferential attachment model, and we focus on one. In each time step $n+1$, a new vertex v_{n+1} having a fixed number m half-edges is added to the existing graph. Each of the m half-edges is, one after the other, connected to one of the existing vertices in the graph by a linear update rule (see [9, (1.3.65)] for specific instances of the update rule and also the introduction of [2] for further variations of the model):

$$\mathbb{P}(v_{j+1,n+1}^{(m)} \text{ connects to } v_i^{(m)} \mid \mathrm{PA}_n^{m,\delta}) = \frac{D_i(n,j)+\delta}{2m(n-1)+j+\delta n},$$

where $D_i(n,j)$ is the degree of vertex i after the jth edge of vertex v_{n+1} has been added and $\delta > -1$ is a fixed parameter. In the directed version of the preferential attachment model, we regard each of the m half-edges as an outgoing edge, i.e., all edges are directed from younger to older vertices in the network, all vertices have a fixed out-degree m, but can have (as n grows) an arbitrarily large in-degree.

Surprisingly, as shown in [4], its directed version violates the power-law hypothesis in that the PageRank distribution at a uniformly chosen vertex converges weakly (with respect to the size n of the graph tending to infinity) to a power law that has *heavier* tails than the weak limit of the in-degree distribution. On the other hand, this heavier PageRank tails phenomenon does not occur in the directed configuration model [7] nor, more generally, in the entire class of random inhomogeneous digraphs [11].

1.5 General Upper Bound in Undirected Networks

The above-mentioned heavier tails of the PageRank distribution in the directed preferential attachment model raises the question whether such heavier tails can also occur in undirected networks.

In [8] we proved that this is *not* possible. Indeed, in undirected graphs of any size (in particular, not only asymptotically) the PageRank of '*every node* is bounded from above by the degree of the respective node. This result remains true (up to a scaling factor) in directed networks with a bounded ratio of in- and out-degrees.

For further results on the impact of this ratio in the context of the directed configuration model and inhomogeneous random digraphs, see, e.g., [14], and for results on personalized PageRanks on undirected networks with bounded ratio (w.h.p.) of maximal- and minimal degree, see [3].

2 Data Sets

We consider three different data sets that all describe undirected networks and are taken from the Stanford Large Networks Database. The largest of the three data sets is the *as-Skitter* [12] internet topology graph with 1,696,415 nodes and 11,095,298 edges. The second data set is the *musae-Twitch-de* [16] (**Mu**lti-**s**cale **A**ttributed Node **E**mbedding) friendship graph between German-language gamers on the Twitch platform consists of 9498 nodes and 153,138 edges. Finally, the *ego-Facebook* [13] data set on social circles, or friend lists, on Facebook consists of 4039 nodes and 88,234 edges.

3 Scatter Plots and Concentration Phenomena

To provide more insight into the relation between the graph-normalized PageRank and (in-)degrees beyond tail-distributions, we present our analysis by means of scatter plots shown in Fig. 1.

As a first visualization that Pagerank is upper bounded by degrees, in Fig. 1, we show scatter plots of the PageRank versus the degree for all vertices. We see that all points clearly lie under the identity line where the PageRank coincides with the degree.

Moreover, the scatter plots in Fig. 1 indicate that the pairs of degrees and PageRanks seem to concentrate around the graph of the linear function $x \mapsto x/\alpha$, where α is the average degree of the nodes. Moreover, the PageRanks are (for sufficiently high degrees) above the graph of the linear function $x \mapsto x/\beta$, where β is four times the average degree of the nodes. We will provide an explanation for these phenomena on the basis of the configuration model. While the networks that we analyze do not look like typical realizations of a configuration model (our networks generally have more triangles than in the configuration model), the configuration model is an established model in network analysis and our theoretical considerations below match to the concentration phenomena that become visible in the plots. Here, the *undirected configuration model* $\mathrm{CM}_n(\boldsymbol{d})$ is a model for a random graph of size n which has a prescribed degree vector $\boldsymbol{d} \in \mathbb{N}_0^n$ as parameter. Each node i is assigned d_i *half-edges*. Let $|\boldsymbol{d}| = \sum_{i\in[n]} d_i$ be the total degree. Then, a random graph is sampled by choosing any permutation $\varphi\colon [|\boldsymbol{d}|] \equiv \{1, 2, \ldots, |\boldsymbol{d}|\} \to [|\boldsymbol{d}|]$ uniformly at random and afterwards connecting the jth half-edge to the $\varphi(j)$th half-edge, where j runs from 1 to $|\boldsymbol{d}|$, and we impose the condition that $\varphi(j) \neq j$ for all $j \in [|\boldsymbol{d}|]$.

Conditions on Degree Sequence. Let $D_n = d_{\phi_n}$ denote the degree of a vertex in G_n where ϕ_n is a uniformly chosen node from G_n. We impose the following two assumptions on the degree sequence:

- **Condition 1.** $D_n \overset{n\to\infty}{\longrightarrow} D$ in distribution; **and**
- **Condition 2.** $\mathbb{E}[D_n] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[D]$.

Under these two conditions, the configuration model converges locally in probability towards the *unimodular branching process tree* [9], whose root-degree is distributed as D, and all other vertices have i.i.d. offspring-distribution $\tilde{D}$ according to the size-biased distribution minus one, i.e.,

$$\mathbb{P}(\tilde{D}=k) = \frac{(k+1)\mathbb{P}(D=k+1)}{\mathbb{E}[D]}. \tag{1}$$

The total degree (including the edge to the parent) is given by the sized-biased distribution $D^* = \tilde{D}+1$, i.e.,

$$\mathbb{P}(D^* = k) = \frac{k\mathbb{P}(D=k)}{\mathbb{E}[D]}. \tag{2}$$

The following asymptotic lower bound on PageRank at a uniformly chosen node ϕ_n holds for the undirected configuration model:

Theorem 1 (Theorem 2.8 in [8]). *Consider the undirected configuration model $G_n = \mathrm{CM}_n(\boldsymbol{d}^{(G_n)})$ where the degree distribution satisfies Conditions 1 and 2. Let $c \in [0,1]$ be a constant (the damping factor for PageRank). Then the PageRank vector $\boldsymbol{R}^{(G_n)}$ satisfies that, for all n, k,*

$$\mathbb{P}(R_{\phi_n}^{(G_n)} > k) \leq \mathbb{P}(D_n > k), \tag{3}$$

while further, for any $\beta > \frac{4\mathbb{E}[D]}{c(1-c)}$,

$$\liminf_{k\to\infty} \liminf_{n\to\infty} \frac{\mathbb{P}(R_{\phi_n}^{(G_n)} > k/\beta)}{\mathbb{P}(D_n > k)} \geq 1. \tag{4}$$

While [8, Theorem 2.8] states that the PageRank *distribution* has thinner tails than the degree in undirected graphs, the result proved is in fact pointwise, in that the proof shows that in *any undirected graph* G, and all v in the vertex set of G,

$$R_v^{(G)} \leq d_v^{(G)}. \tag{5}$$

It is this relation that we will investigate empirically in this paper, and, in particular, how it is changed in graphs that become more directed using an interpolation scheme.

The proof of the lower bound (4) involves a first-order approximation of the PageRank. As the local limit of $\left(\mathrm{CM}_n(\boldsymbol{d}^{(G_n)})\right)_{n\in\mathbb{N}}$ is a branching-process tree with i.i.d.-offspring distribution for all vertices except the root we conjecture the following improvement with a potential proof being based on taking into account all terms of the series expansion and employing independence of degrees:

Conjecture 1. In the lower bound (4) in Theorem 1 we conjecture that the scaling factor β can be decreased by a factor $c(1-c)$. More explicitly, we conjecture that for every $\tilde{\beta} > 4\mathbb{E}[D]$ it holds

$$\liminf_{k\to\infty} \liminf_{n\to\infty} \frac{\mathbb{P}(R_{\phi_n}^{(G_n)} > k/\tilde{\beta})}{\mathbb{P}(D_n > k)} \geq 1. \tag{6}$$

We now want to get insight into the behavior of $\mathbb{E}\left[\frac{R_\phi}{D_\phi}\right]$ on the unimodular branching process tree with root ϕ. We can offer an interesting heuristic insight by assuming that the fractions $X(v) = \frac{R_v}{D_v}$ of the limiting PageRank and the respective degree along the vertices of the tree are independent of the rest of the degree sequence. While there is a subtle dependence in reality, this simplification helps us to understand how PageRank dissipates through the layers of the tree. We will use $\hat{R}$ to denote this independent version. We formalize the analysis in the next proposition. While this is an obvious simplification, to the best of our knowledge, these are the first calculations for the PageRank of an undirected configuration model.

Proposition 1. *Let $(D_\phi, (D_{ij})_{i,j\in\mathbb{N}})$ be an independent family of $\mathbb{N}$-valued random, where D_ϕ has distribution D and D_{ij}'s have distribution D^* that has a size-biased distribution of D as defined in (2). Let the random variables $(X_\phi, (X_{ij})_{i,j\in\mathbb{N}})$ satisfy the stochastic recursion*

$$X_\phi \stackrel{d}{=} (1-c)\frac{1}{D_\phi} + c\frac{1}{D_\phi}\sum_{j=1}^{D_\phi} X_{1,j}, \tag{7}$$

$$X_{i,j} \stackrel{d}{=} (1-c)\frac{1}{D_{i,j}} + c\frac{1}{D_{i,j}}\sum_{j=1}^{D_{i,j}-1} X_{i+1,j} + c\frac{1}{D_{i,j}} X_{i-1,1}, 0 > i \tag{8}$$

For $i = \phi, 1, 2, \ldots$, denote $x_i = \mathbb{E}[X_{i,j}]$, which is equal for all $j = 1, 2, \ldots$. Then the column vector $\mathbf{x} = (x_\phi, x_1, x_2, \ldots)^T$ is defined by

$$\mathbf{x} = [I - cB]^{-1}\mathbf{b}. \tag{9}$$

Here, the matrix B in (9) is the probability transition matrix of a simple random walk on the non-negative integers reflected at $0 = \phi$; this random walk transitions from $i > 0$ to $i-1$ with probability

$$b = \mathbb{E}\left[\frac{1}{D_{i,j}}\right] = \sum_{k=1}^{\infty} \frac{1}{k}\frac{k\mathbb{P}(D=k)}{\mathbb{E}[D]} = \frac{1}{\mathbb{E}[D]};$$

from $i > 0$ to $i+1$ with probability $1-b$, and from ϕ to 1 with probability 1. The vector $\mathbf{b}$ is given by $\mathbf{b} = (1-c)\cdot(a, b, b, \ldots)^T$, where $a = \mathbb{E}\left[\frac{1}{D}\right]$.

Proof. Using the tower rule (7), we get

$$x_\phi = \mathbb{E}\left[\mathbb{E}\left[(1-c)\frac{1}{D_\phi} + c\frac{1}{D_\phi}\sum_{j=1}^{D_\phi} X_{1,j}|D_\phi\right]\right] = (1-c)a + cx_1. \quad (10)$$

Similarly, from (8), we get

$$x_i = \mathbb{E}\left[\mathbb{E}\left[(1-c)\frac{1}{D_{i,j}} + c\frac{1}{D_{i,j}}\sum_{j=1}^{D_{i,j}-1} X_{i+1,j} + c\frac{1}{D_{i,j}}X_{i-1,1}|D_{i,j}\right]\right]$$
$$= (1-c)b + c(1-b)x_{i+1} + cbx_{i-1},\ i \geq 0. \quad (11)$$

Denote $\mathbf{x} = (x_\phi, x_1, x_2, \ldots)^T$. Then system of linear equations (10), (11) can be written in the form

$$[I - cB]\mathbf{x} = \mathbf{b}.$$

Since $I - cB$ is invertible, this proves (9). □

Note that $a > b$, and this makes x_ϕ larger than x_i's. However, the difference between a and b is typically not that large. If we had $b = a$ then all the x_j's would be the same and equal to $b = 1/\mathbb{E}[D]$. This heuristic derivation intuitively explains the observed slope close to $1/\mathbb{E}[D]$ in the scatter plots.

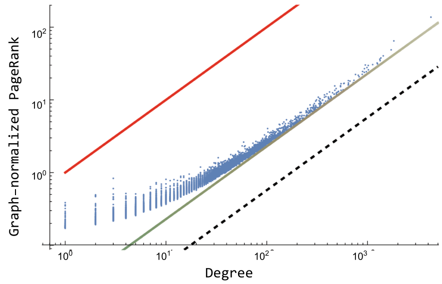
4 Interpolation Between Undirectedness and Directedness in Real-World Networks

As discussed above, the behavior of PageRank on undirected networks can be quite different to that on directed networks. By means of a case study, we perform an oriented percolation very similar to [1]. Here, we at first identify the undirected network with a directed one with bi-directed edges and then uniformly choose a subset of prescribed size from the edges. This subset contains those edges which we will make single-directional. After sampling this subset, for any of its elements, we uniformly choose one of the two possible directions and delete it. Depending on the proportion p of edges that we make single-directional, the procedure will significantly alter the ratio of in- and out-degrees.

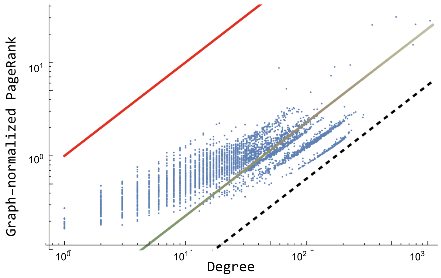
Interpolation Between Undirected and Directed Networks
Here, we first describe our randomized algorithm that interpolates between undirected and directed graphs:

Input: List of m undirected edges $E = \{\{i_1, j_1\}, ..\{i_m, j_m\}\}$, proportion $p \in [0,1]$.

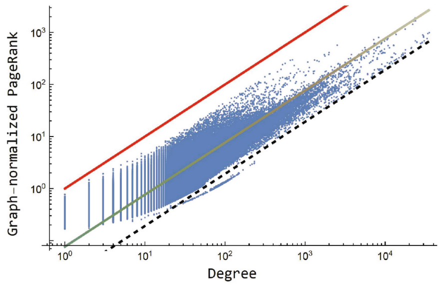
1. Create two lists E^- and E^+ of directed edges from LoE as follows:
   - For all $\{a,b\} \in E$ do $(a,b) \in E^-$ and $(b,a) \in E^+$.
2. Uniformly choose a subset $RaInd$ of $\{1,2,..,m\}$ of size $round(pm)$, where we recall that m denotes the total number of edges given as input for the algorithm.



(a) musae-twitch-de [16]



(b) ego-Facebook [13]



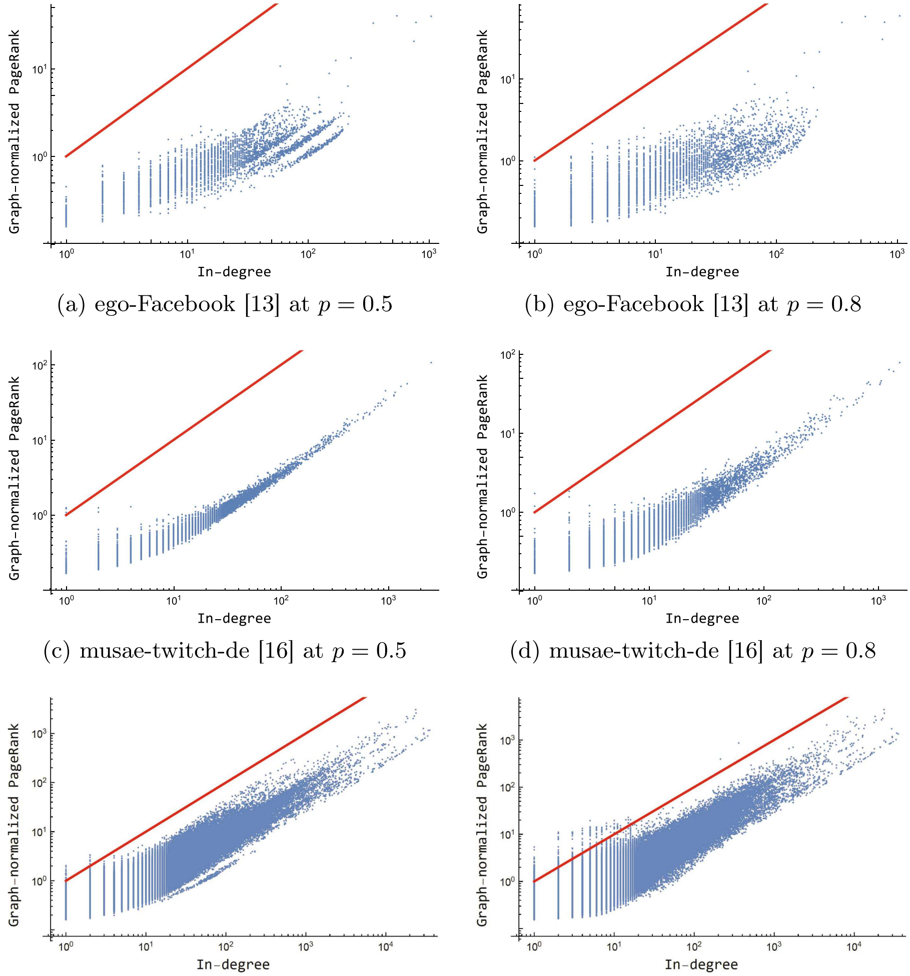
(c) as-Skitter [12]

Fig. 1. The graph-normalized PageRank distribution with damping factor $c = 0.85$ (vertical axis) and the degrees (horizontal axis) with a $\log_{10} - \log_{10}$-scaling of the axes. All points lie below the red line corresponding to the graph of the identity function, which exemplifies the general upper bound for the PageRank in undirected networks. Furthermore, the point sets are approximately centered around the khaki-colored line that is vertically shifted by $-\log_{10} \alpha$ with α being the average degree of the respective graph. The black dashed line is vertically shifted by $-\log_{10} \tilde{\beta}$ with $\tilde{\beta}$ being four times the average degree as given in Conjecture 1.

3. For each $l \in RaInd$ uniformly choose among the two options: either delete the lth entry of the list E^-, or delete the lth entry of the list E^+ of reversed edges.
   Let $E^{'-} \subset E^-$ and $E^{'+} \subset E^+$ denote the so-obtained lists of directed edges with deletions.
4. Generate a directed graph $G_{\text{dir}} = (V, E^{'-} \cup E^{'+})$ from the lists with deletions. Here, V is the set of all vertices appearing in E.
   Generate graph $G_{\text{ref}} = (V, E^- \cup E^+)$ (the reference graph) from the list without deletions that contains each edge in both directions.

Note that in [1], each directed edge is removed with probability p, while in our case each edge can be removed with probability $p/2$, and we guarantee to keep an edge in at least one of the directions.

In Fig. 2 we see that as the proportion p of single-directed edges increases, some of the nodes will get a PageRank that is higher than the in-degree.



(a) ego-Facebook [13] at $p = 0.5$ (b) ego-Facebook [13] at $p = 0.8$

(c) musae-twitch-de [16] at $p = 0.5$ (d) musae-twitch-de [16] at $p = 0.8$

(e) as-Skitter [12] at $p = 0.5$ (f) as-Skitter [12] at $p = 0.8$

Fig. 2. Six realizations of the above algorithm applied to the three data sets and different values of the proportion p. The respective data plots for the initial undirected graphs are shown in Fig. 1. For each realization of the algorithm, the graph-normalized PageRank with damping factor $c = 0.85$ on the vertical axis are plotted against the in-degrees on the horizontal axis. As the proportion p of edges that have been made single-directional increases, some of the low-degree nodes get shifted above the red line, i.e., they obtain a graph-normalized PageRank that is higher than their in-degree.

The Parents of Nodes Whose PageRank Overshoots the In-Degree. To determine the main effects that contribute to this phenomenon, we further specifically look at those nodes in the truncated graph G_{dir} that have in-degree 1, but a PageRank that is larger than 1. Heuristically, from the definition of

PageRank, the unique parent of each of these nodes must have a large PageRank. For each of these nodes, we compare the in- and out-degree and PageRank of its unique ancestor with the respective values of these quantities for the (undirected) reference graphs.

For the two smaller datasets, musae-twitch-de and ego-Facebook, this reveals that the phenomenon of increasing PageRank seems to be mostly explained by a superposition of two effects on the parent of those nodes (see Tables 1a and 1b):

- The out-degree of the parent is reduced. Thus, the remaining children get a higher share of the total degree weight.
- The PageRank of the parent increases.

While the first effect might be not surprising with respect to the definition of and what is known about PageRank, the second effect might be more surprising. We conjecture that the second effect is explained by the first effect affecting the nodes which remain a directed edge pointing towards the previously considered node (the parent's parents).

A Positive Proportion of Vertices with In-Degree 1 and Large PageRank. We next explain, based on the first of the two effects described above, how a positive proportion of vertices with in-degree 1 and large PageRank can arise.

Proposition 2. *Consider any undirected graph that has a local limit with a positive proportion* $q_{\ell,k}$ *of vertices that have degree equal to some fixed small* ℓ *(for example* $\ell = 1$*), and that have a neighbor of degree k for k large. Then, for k sufficiently large the following holds true:*
with a positive probability, the graph obtained from applying the randomized algorithm that makes a proportion p of edges single-directed will contain at least one vertex with a PageRank larger than the in-degree.

Proof. Consider a pair (v_l, w_k) of neighboring vertices in the local limit such that v_l has degree ℓ and w_k has degree k. After applying the above randomized algorithm to make a proportion p of edges directed, the probability that vertex v_ℓ will have in-degree 1, getting the directed edge from w_k, equals $p^\ell \cdot (1/2)^\ell$, while the probability that w_k directs all its remaining $k-1$ edges which are not attached to v_ℓ to in-edges for itself is $p^{k-1} \cdot (1/2)^{k-1}$. This implies that the degree-k neighbor w_k sends all of its PageRank to the vertex of interest. Since $R_u \geq 1-c$ for every vertex u, this means that the PageRank of the vertex v_ℓ will be at least $kc(1-c)$, which can be made arbitrarily large by making k large. In particular, we can make this PageRank larger than 1. This shows that with a positive probability, after directing the edges, there will be a positive (though possibly quite small) proportion of vertices with degree 1 and PageRank larger than 1. □

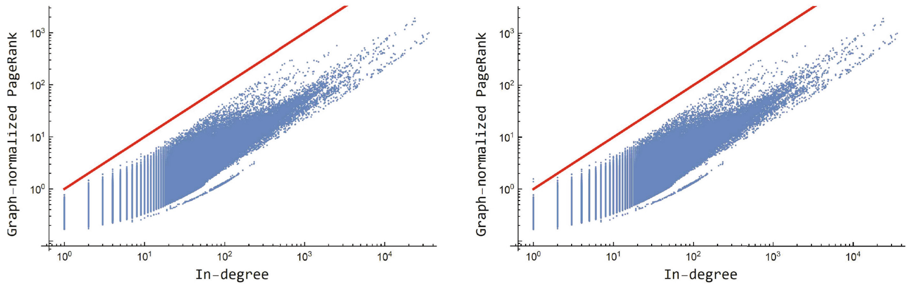
Table 1. Parameters of the unique parents of those nodes in Figs. 2b and 2d that get shifted above the line and have in-degree 1. The node numbers correspond to the ones used in the data sources.

(a) In Figure 2b, nodes 3984, 4010, 4024, 4035 and 4036 get shifted above the line and have in-degree one. They all have the common unique parent 3980.

Data for unique parent 3980	before removal	after removal
In-degree	59	58
Out-degree	59	11
PageRank	8.71032	12.2927

(b) In Figure 2d, nodes 5918, 6150 and 1462 get shifted above the line and have in-degree one. The corresponding parents are 18, 967 and 8114

Data for the parents	before removal	after removal
In-degrees	2;16;3	2;16;3
Out-degrees	2;16;3	1;1;1
PageRanks	0.470515;0.588953;0.433619	1.83977;1.00621;1.21459



(a) $p = 0.001$ (b) $p = 0.01$

Fig. 3. Two realizations of the above algorithm for smaller values of p and the as-Skitter [12] dataset, again with damping factor $c = 0.85$. We see that already at $p = 0.01$ some small-degree nodes obtain a higher PageRank than their in-degree.

Large PageRanks at Low Proportion p of Single-Directed Edges. In the two smaller networks, a large value (0.5 or higher) for the proportion p of single-directed edges is needed to evoke the effect of PageRanks becoming larger than the corresponding in-degrees (cf. Figs. 2a to 2d).

5 Outlook and Discussion

In this paper, we have investigated how PageRank can become larger than the in-degree in directed networks, while PageRank is bounded by the degree in undirected networks. We have done this by interpolating between the undirected version of the graphs, and the randomly truncated directed versions of the real-world networks. We have studied which nodes are such that the PageRank is larger than the in-degree. This occurs for vertices that have low in-degree, and have a unique high-degree parent for whom relatively many edges have been changed to in-edges.

Our work raises the following follow-up questions. It remains to formally prove the concentration results that become visible in Fig. 1 on the level of random graph models.

6 Data Availability

The plots shown in Figs. 1, 2 and 3 have been generated using Wolfram Mathematica and using the datasets mentioned in the respective captions. The underlying realizations of the above randomized algorithm presented in Figs. 2 and 3 and Tables 1b and 1a have been obtained with Wolfram Mathematica. The lists of edges for the resulting directed graphs have been saved as .csv-files and are available upon request.

Acknowledgments. The authors thank the three anonymous referees for their insightful comments on the first version of this paper.

The work of RvdH and NL is supported by the Netherlands Organisation for Scientific Research (NWO) through the Gravitation NETWORKS grant 024.002.003, and by the National Science Foundation under Grant No. DMS-1928930 while the authors were in residence at the Simons Laufer Mathematical Sciences Institute in Berkeley, California, during the Spring semester.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alimohammadi, Y., Borgs, C., Saberi, A.: Locality of random digraphs on expanders. Ann. Probab. **51**(4), 1249–1297 (2023). https://doi.org/10.1214/22-AOP1618
2. Antunes, N., Banerjee, S., Bhamidi, S., Pipiras, V.: Attribute network models, stochastic approximation, and network sampling and ranking algorithms. Preprint (2023)
3. Avrachenkov, K., Kadavankandy, A., Ostroumova Prokhorenkova, L., Raigorodskii, A.: PageRank in undirected random graphs. In: Gleich, D.F., Komjáthy, J., Litvak, N. (eds.) WAW 2015. LNCS, vol. 9479, pp. 151–163. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26784-5_12
4. Banerjee, S., Olvera-Cravioto, M.: PageRank asymptotics on directed preferential attachment networks. Ann. Appl. Probab. **32**(4), 3060–3084 (2022). https://doi.org/10.1214/21-aap1757
5. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999). https://doi.org/10.1126/science.286.5439.509

6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. **30**(1), 107–117 (1998). https://doi.org/10.1016/S0169-7552(98)00110-X. proceedings of the Seventh International World Wide Web Conference
7. Chen, N., Litvak, N., Olvera-Cravioto, M.: Generalized PageRank on directed configuration networks. Random Struct. Algorithms **51**(2), 237–274 (2017). https://doi.org/10.1002/rsa.20700
8. Henning, F., van der Hofstand, R., Litvak, N.: Power-law hypothesis for PageRank on undirected graphs. Preprint (2024)
9. van der Hofstad, R.: Random Graphs and Complex Networks, vol. 2. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press (2024). https://doi.org/10.1017/9781316795552
10. Lawrence, S., Giles, C.L.: Searching the world wide web. Science **280**(5360), 98–100 (1998). http://www.jstor.org/stable/2895232
11. Lee, J., Olvera-Cravioto, M.: Pagerank on inhomogeneous random digraphs. Stochast. Processes Appl. **130**(4), 2312–2348 (2020). https://doi.org/10.1016/j.spa.2019.07.002
12. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 177–187 (2005). https://www.cs.cmu.edu/~jure/pubs/powergrowth-kdd05.pdf
13. Leskovec, J., Mcauley, J.: Learning to discover social circles in ego networks. In: Advances in Neural Information Processing Systems, vol. 25 (2012). http://i.stanford.edu/~julian/pdfs/nips2012.pdf
14. Olvera-Cravioto, M.: PageRank's behavior under degree correlations. Ann. Appl. Probab. **31**(3), 1403–1442 (2021). https://doi.org/10.1214/20-aap1623
15. Pandey, M.: Centrality Measures and Connectivity Properties in Large Networks: Who is the most influential in a network? Ph.d. thesis 1 (research tu/e / graduation tu/e), Mathematics and Computer Science (2025). https://pure.tue.nl/ws/portalfiles/portal/345650626/20250108_Pandey_hf.pdf, proefschrift
16. Rozemberczki, B., Allen, C., Sarkar, R.: Multi-scale attributed node embedding. J. Complex Netw. **9**(2), cnab014 (2021)