Simon's fundamental rich-get-richer model entails a dominant first-mover advantage

Peter Sheridan Dodds,[1,*] David Rushing Dewhurst,[1] Fletcher F. Hazlehurst,[1] Colin M. Van Oort,[1] Lewis Mitchell,[2]
Andrew J. Reagan,[1] Jake Ryland Williams,[3] and Christopher M. Danforth[1]

[1]*Vermont Complex Systems Center, Computational Story Lab, Vermont Advanced Computing Core, Department of Mathematics & Statistics,
University of Vermont, Burlington, Vermont 05401, USA*
[2]*School of Mathematical Sciences, North Terrace Campus, University of Adelaide, South Australia 5005, Australia*
[3]*Department of Information Science, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104, USA*

Herbert Simon's classic rich-get-richer model is one of the simplest empirically supported mechanisms capable of generating heavy-tail size distributions for complex systems. Simon argued analytically that a population of flavored elements growing by either adding a novel element or randomly replicating an existing one would afford a distribution of group sizes with a power-law tail. Here, we show that, in fact, Simon's model does not produce a simple power-law size distribution as the initial element has a dominant first-mover advantage, and will be overrepresented by a factor proportional to the inverse of the innovation probability. The first group's size discrepancy cannot be explained away as a transient of the model, and may therefore be many orders of magnitude greater than expected. We demonstrate how Simon's analysis was correct but incomplete, and expand our alternate analysis to quantify the variability of long term rankings for all groups. We find that the expected time for a first replication is infinite, and show how an incipient group must break the mechanism to improve their odds of success. We present an example of citation counts for a specific field that demonstrates a first-mover advantage consistent with our revised view of the rich-get-richer mechanism. Our findings call for a reexamination of preceding work invoking Simon's model and provide an expanded understanding going forward.

I. INTRODUCTION

Across the spectrum of natural and constructed phenomena, descriptions of the architecture and dynamical behavior of complex systems repeatedly involve heavy-tailed distributions. For systems involving components of variable size S, many bear size distributions with power-law decays of the form $P(S) \sim S^{-\gamma}$ [1,2]: word usage frequency in language [3–5], the number of species per genus [1,6], citation numbers for scientific papers [7,8], node degree in networks [9–12], firm sizes [13], and the extent of system failures such as forest fires [14,15]. These size distributions are often alternately cast in the form of a Zipf distribution [3] with components ordered by decreasing size and $S_r \sim r^{-\alpha}$ where r $(= 1,2, \ldots)$ is the size rank and $\alpha = 1/(\gamma - 1)$ [16].

Elucidating and understanding the most essential dynamical models leading to power-law size distributions is an essential task. While the mechanisms giving rise to such distributions are diverse, they generally involve growth and replication.

In his famous 1955 paper on skewed distributions [1], Simon built on classical urn model theory to show that a simple, single parameter, rich-get-richer mechanism could lead a growing population to produce a pure power-law size distribution of groups of elements of matching type [17]. Simon's model is governed by an innovation probability ρ which Simon argued controls the group size distribution exponent as $\gamma = 1 + 1/(1 - \rho)$ and, equivalently, the Zipf's law exponent as $\alpha = 1 - \rho$ (we rederive these results as part of our analysis in Sec. III).

Simon's model has endured because it is at once a boiled-down, easy-to-understand toy model representative of a large class of rich-get-richer mechanisms, and yet it is also a model that has a remarkable ability to capture the essential growth dynamics of disparate, real-world complex systems. While not without controversy, particularly for language [18–26], Simon's micro-to-macro link between the separately measurable innovation rate and power-law scaling for system component size distribution has been observed to roughly hold for word counts in books [1], citation counts in scientific literature [10,27,28], the early growth of the Web [11], and the development of software such as the Linux kernel [29].

Rich-get-richer models adjacent to Simon's model have been employed to characterize the essential features of many kinds of systems such as the emergence of novelties [30,31]. Arguably the most profound role of rich-get-richer mechanisms has been uncovered in complex networks. Simon's model is the explicit core of Price's cumulative advantage mechanism for the growth of citation networks in scientific literature [7,8]. A modified version of Simon's model is also at the heart of the independently discovered growing network model of preferential attachment due to the field-starting work of Barabási and Albert [9].

Here, we show analytically and through simulations that Simon's analysis, for all its successes, was strikingly incomplete: The initial group enjoys a profound "first-mover advantage" on the order of the inverse of the innovation probability, $1/\rho$. This is not a small correction to a long established theory. As the innovation probability is typically less than 0.1 and often much closer to 0 [2,3,5,11,29,32], the initial group's size may be orders of magnitude greater than would be consistent with a simple power law. Nor, as we will show, can the first group be dismissed as a transient or as a kind of null group and not part of the system. Indeed, we provide evidence from scientific citation data that a first-mover advantage manifested by Simon's model is a real phenomenon.
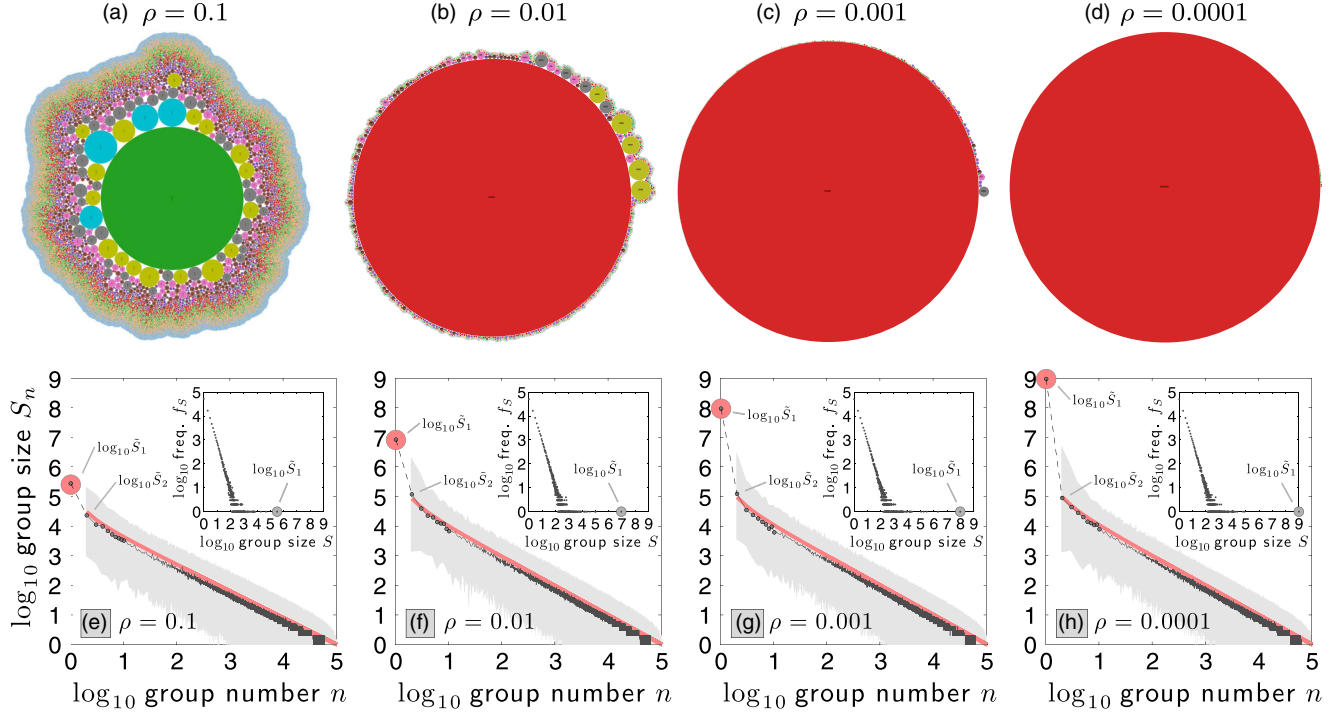
_____
*peter.dodds@uvm.edu

FIG. 1. Results from simulations of Simon's model showing an inherent dominant first-mover advantage. (a)–(d) Visualizations of group sizes after approximately $t = 10^6$ time steps for $\rho = 0.1$, 0.01, 0.001, and 0.0001. Group sizes are proportional to disk area. The colors are a function of each simulation's specific history but do match across groups of the same size within each system. (e)–(h) Main plots: Zipf-like distributions for an ensemble of 200 simulations in gray and theory in red [see Eq. (5)]. Differing from the simulations for the visualizations, these were run until $N^{(g)} = 10^5$, approximately $10^5/\rho$ time steps. The dark gray curve indicates the median size of the nth arriving group and the light gray bounds 2.5 to 97.5 percentiles. For clarity, the median sizes of the first and second groups are highlighted as $\log_{10} \tilde{S}_1$ and $\log_{10} \tilde{S}_2$. Insets: For each value of ρ, we show the (raw) size frequency distribution f_S for a single simulation, again indicating $\log_{10} \tilde{S}_1$. The exponents for the Zipf distribution and size distributions are connected as $\gamma = 1 + 1/\alpha$. We provide interactive simulations for Simon's model as part of the paper's online appendices [33].

In what follows, we first describe Simon's model and present results from simulations (Sec. II); analytically determine the first mover's advantage (Sec. III); explore the detailed long term dynamics of all groups (Sec. IV); self-referentially compare the model's output with citation data concerning scale-free networks [9] (Sec. V); and, finally, consider broader implications for understanding real-world, rich-get-richer systems (Sec. VI).

II. EVIDENCE OF A FIRST-MOVER ADVANTAGE FROM SIMULATIONS

The algorithm for Simon's classic rich-get-richer model is simple. We are concerned with the growth of a population of elements where each element has a type, and elements of the same type form a group. In modeling real systems, types may represent a city, a word, or the destination of a link in a network [1,29]. Beginning with a single element at time $t = 1$, an element is added to the population at each discrete time step $t \geqslant 2$. Representing innovation, the arriving element has, with probability ρ, a new, previously unseen type. Alternately, with probability $1 - \rho$, the arriving element is a replication taking on the type of a randomly chosen existing element. As one of a number of generalizations, we will, as Simon did himself in [1], later consider a dynamic innovation probability ρ_t in Sec. VI.

In Fig. 1, we provide visualizations along with Zipf-like distributions and size distributions for Simon's model for $\rho = 0.1$, 0.01, 0.001, and 0.0001. In Figs. 1(a)–1(d), we show results for four sample simulations corresponding to these values of ρ after 10^6 time steps. Each disk represents a group with area proportional to group size. Interactive simulations from which these images are drawn are available in the paper's online appendices [33].

To the eye, for $\rho = 0.1$, the first group appears to be somewhat outsized but perhaps not inconsistent. However, for the next three (decreasing) values of ρ, the first group is evidently different, increasingly accruing the bulk of all new elements.

In Figs. 1(e)–1(h), we make clear the first-mover advantage through Zipf-like distributions for group size (main plots) and size distributions (inset plots) for the same ordering of ρ values. The group sizes in the main plots are a function of group arrival number rather than rank according to decreasing size (hence Zipf-like), and come from an ensemble of simulations (median in dark gray, 2.5 to 97.5 percentile range in light gray; see caption for details) and overlaid theory (red). We examine the variation in group size with group arrival order later on. As per Simon's analysis, median group size $\tilde{S}_n$ behaves as $n^{-\alpha}$ for $n \geqslant 10$ where $\alpha \, (= 1 - \rho)$ is the Zipf exponent. However, as we demonstrate theoretically below, the dominant group $\tilde{S}_1$ is larger than would be expected by a factor of $1/\rho$.

Again for single example simulations, the inset size frequency distributions (raw counts; notation: f_S) in Figs. 1(e) and 1(f) show the same disparity with the largest element isolated from the main power-law-obeying size distribution. We can now see why Simon's analysis, while technically correct, fell short: The size distribution he derived fit all but one point which, if not observed and handled appropriately, vanishes in contribution in the infinite system size limit.

III. ANALYTIC DETERMINATION OF THE FIRST-MOVER ADVANTAGE

To understand this first-mover advantage, we carry out a reanalysis of Simon's model. New groups are initiated stochastically with the nth group first appearing on average at time t_n^{init}. We write the number of elements in the nth group at time $t \geqslant t_n^{\text{init}}$ as $S_{n,t}$, and each group starts with a single element: $S_{n,t_n^{\text{init}}} = 1$ (we examine this choice later). As Simon did, we assume an initial condition of a population of 1 element at time $t = 1$ [we consider a general initial condition later in Sec. VI, Eq. (8)]. If at time t, we have $N_t^{(\text{g})}$ distinct groups, the probability that a randomly drawn element belongs to the nth group is then

$$P_{n,t} = \frac{S_{n,t}}{\sum_{n'=1}^{N_t^{(\text{g})}} S_{n',t}} = \frac{S_{n,t}}{t}. \tag{1}$$

We construct an evolution equation for the size of the nth arriving group, $S_{n,t}$. At time t, for the nth group to increase in number by 1, replication must be chosen (occurring with probability $1 - \rho$), and then an element in the nth group must be replicated, leading to the probabilistic statement

$$\langle S_{n,t+1} - S_{n,t} \rangle = (1 - \rho_t) \frac{S_{n,t}}{t} (+1). \tag{2}$$

In the case of Simon's fundamental model with a fixed innovation probability, $\rho_t = \rho$, we proceed with a difference equation calculation. Our primary analysis is approximate as we drop the expectation on the left hand side of Eq. (2).

For fixed $\rho_t = \rho$ and shifting from t to $t - 1$, Eq. (2) gives the approximation, for $t \geqslant t_n^{\text{init}}$,

$$S_{n,t} = \left[1 + \frac{(1 - \rho)}{t - 1} \right] S_{n,t-1}. \tag{3}$$

Again given that $S_{n,t_n^{\text{init}}} = 1$, we have

$$\begin{aligned} S_{n,t} &= \left[1 + \frac{(1 - \rho)}{t - 1} \right] \left[1 + \frac{(1 - \rho)}{t - 2} \right] \cdots \left[1 + \frac{(1 - \rho)}{t_n^{\text{init}}} \right] 1 \\ &= \left[\frac{t - \rho}{t - 1} \right] \left[\frac{t - 1 - \rho}{t - 2} \right] \cdots \left[\frac{t_n^{\text{init}} + 1 - \rho}{t_n^{\text{init}}} \right] \\ &= \frac{\Gamma(t + 1 - \rho)\Gamma(t_n^{\text{init}})}{\Gamma(t_n^{\text{init}} + 1 - \rho)\Gamma(t)} = \frac{B(t_n^{\text{init}}, 1 - \rho)}{B(t, 1 - \rho)}, \end{aligned} \tag{4}$$

where B is the beta function. Our analysis thus far has been heading towards the same conclusion as Simon's which took a different route following the evolution of the system's size distribution. Now, however, we find a distinction that renders the initial group special. For constant $\rho \ll 1$, the nth group first arrives, on average, at $t_n^{\text{init}} \simeq [\frac{n-1}{\rho}]$ where $[\cdot]$ is the rounding

operator. But this is only valid for $n \geqslant 2$ because $t_1^{\text{init}} = 1$. For large t, we find the size of the nth arriving group to be

$$S_{n,t} \sim \begin{cases} \frac{1}{\Gamma(2-\rho)} \left[\frac{1}{t} \right]^{-(1-\rho)} & \text{for } n = 1, \\ \rho^{1-\rho} \left[\frac{n-1}{t} \right]^{-(1-\rho)} & \text{for } n \geqslant 2, \end{cases} \tag{5}$$

where we have used the asymptotic scaling $B(x,y) \sim \Gamma(y)x^{-y}$ for large x, and replaced $t + 1 - \rho$ with t.

On average, we expect the groups to be ranked by decreasing size according to their arrival number n. For large n, Zipf's law appears with $S_{n,t} \sim \rho^{1-\rho} [\frac{n}{t}]^{-\alpha}$ with exponent $\alpha = 1 - \rho$. As we move down toward $n = 2$, a small correction factor arises as $[\frac{n-1}{n}]^{-\alpha}$, maximally $2^{1-\rho} < 2$ at $n = 2$.

We now see that the initial and largest group manifestly does not conform because of the absence of the $\rho^{(1-\rho)}$ term. Moreover, as $\rho \to 0$, the size of the dominant group departs rapidly from Zipf's law with a factor of $1\rho^{-1}$ greater than what would be internally consistent with $S_{n,t}$ for $n \gg 2$.

Shifting to the size frequency distribution [16], we have $P(S) \propto S^{-\gamma} = S^{-(1+\frac{1}{\alpha})} = S^{-(1+\frac{1}{1-\rho})}$, and again for values of ρ consistent with real-world data [1,2], the first group is an outlier clearly separated from the power-law size distribution.

Simon's analysis, like many subsequent treatments [10,27,28,34–38], was a mixture of discrete and continuous pieces, obtaining the tail of the group size distribution through an asymptotic expansion. Such a continuum approximation washed out the discrete nature of the system rendering the first-mover anomaly hidden. Further, because the first mover is only one of an infinite number of groups for $t \to \infty$, its weight in the distribution tends to 0 and will thus be unobserved.

While it may be tempting to interpret the first group as a so-called "dragon king" [39,40], we see that it is an endogenously generated product of an elementary rich-get-richer model rather than an exogenous mechanism singularly affecting the largest events within a system.

The preferential attachment models developed in [34] show evidence of a first-mover advantage and may profit from an analysis similar to the one we have laid out here.

IV. LONG TERM VARIABILITY OF GROUP SUCCESS

We now delve into how arrival order of groups (n) relates to final rank (r), an issue arising with the Zipf-like distributions we presented in Fig. 1 and elided in our preceding analysis.

While it is evident that Simon's model must produce a degree of indeterminacy in ranking, our goal here is largely to use simulations and some specific analysis to characterize the nature and extent of rank variability for growing network models based on Simon's model [8,9]. Krapivsky and Redner [10] and Newman [27,28] have analytically determined the distribution of citations that the nth arriving group will receive as a function of time. Further, Newman was able to show good agreement with citation data. Our focus on the original Simon model and different treatment means we arrive at some complementary findings.

In the main plot of Fig. 2(a), we show results for a single simulation with $\rho = 0.01$, and 10^7 total elements (equivalent to run time), and $N^{(\text{g})} \simeq 10^5$ distinct groups. The dark points indicate the actual size of the nth group while the blue curve is the resulting, properly ranked Zipf distribution. While the
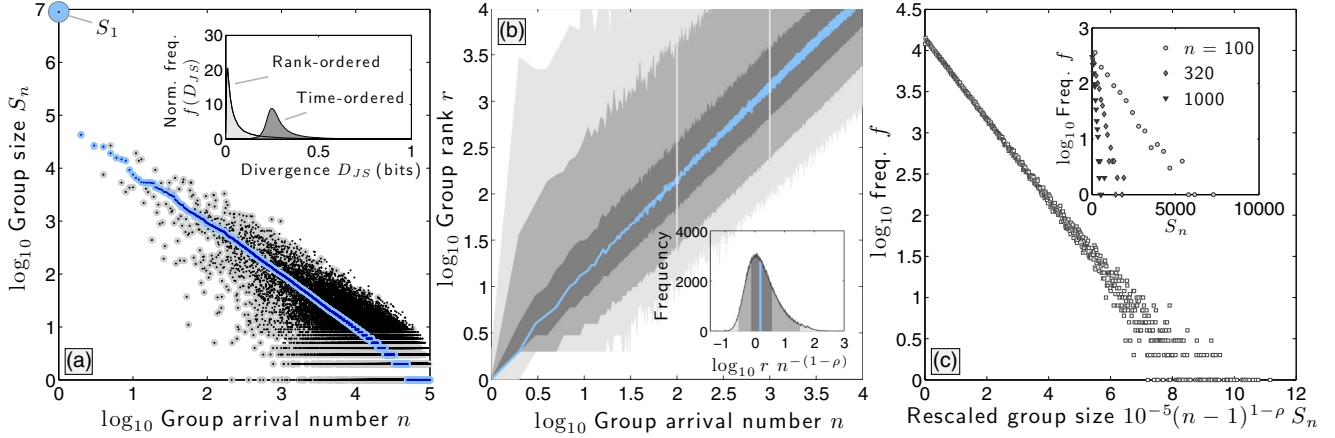
FIG. 2. Simulation results for a numerical investigation of the variability of final rank r as a function of group arrival number n. All panels are derived from an ensemble of 1000 systems with $\rho = 0.01$ evolved for 10^7 time steps. (a) Main panel: For a single example system, group size as a function of group arrival number n (black) with ranked sizes overlaid (blue). Inset: Comparison of Jenson-Shannon divergences for all pairs of group sizes as a function of time of arrival (time-ordered) and all pairs of groups sizes when ranked by size (rank-ordered). (b) Main panel: Group rank r as a function of group arrival number n with median in light blue and the increasingly gray regions delimiting minimum-maximum, 2.5 to 97.5 percentiles, and 25 to 75 percentiles. Inset: Rescaled distribution of group ranks for $10^2 \leqslant n \leqslant 10^3$ (white lines in main plot). (c) Main panel: Rescaled size distributions for $10^2 \leqslant n \leqslant 10^3$ indicating an exponential form. Inset: Raw size distributions for three example arrival group numbers, $n = 100$, 320, and 1000.

first mover dominates as expected, and the second arriving group obtains the second ranking in this instance, we observe substantial and increasing decoupling between arrival order and final rank as the system grows.

To further explore this decoherence, we generate an ensemble of 1000 simulations of the same system. The effect on final group ordering brought about by decreasing ρ saturates quickly and all of the following results regarding arrival time and rank are essentially the same for $\rho \leqslant 0.01$ with little quantitative change. We first measure the variability of the resulting distributions using the Jenson-Shannon divergence (JSD), time-ordered against time-ordered and rank-ordered against rank-ordered, a total of $\binom{1000}{2}$ such comparisons for each. In the inset to Fig. 2(a), we show the distributions of JSD for each ordering, finding a typical disparity for the time-ordered size distributions of 0.25 bits.

In the main plot of Fig. 2(b), we present overall group rank r as a function of group arrival number n for our ensemble of 1000 simulations. The pale blue line indicates the median, and the surrounding gray regions mark the 25 to 75 percentile range, the 2.5 to 97.5 percentile range, and the minimum to maximum range. In the inset, we rescale and collapse the final rank distribution for arrival group numbers 10^2 to 10^3 based on Eq. (5). We see that from around just the 10th arriving group on, 95% of a group's final rank spans a remarkable two orders of magnitude around the median $r = n$, skewed towards higher values as shown in the inset. Thus, while by equating rank and arrival number, our analysis is effective for the median size of groups, the system's specific dynamics are considerably more complex.

Lastly, in Fig. 2(c), we examine the distribution of possible group sizes as a function of group arrival number n. Rather than possessing a single maximum that increases with n, we find that an exponential distribution is a good approximation.

The inset gives three example distributions and the main plot shows an appropriate rescaling of all $100 \leqslant n \leqslant 1000$.

For all but a small initial collection of groups, the mode group size is thus close to 1, consistent with Simon's asymptotic result that $1/(2 - \rho) \simeq 1/2$ of all groups contain only one element (hapax legomena for texts). To understand how half of all initiated groups never grow in size, we determine the probability of a group gaining a new member as a function of its arrival time and current size. The probability that the nth arriving group, which is of, say, size $S_{n,t} = k$ at time t, fails to replicate for all of times t through $t + \tau - 1$ before replicating at time $t + \tau$ is

$$P(S_{n,t+\tau} = k + 1 \mid S_{n,t+i} = k \text{ for } i = 0, \ldots, \tau - 1)$$

$$= \prod_{i=0}^{\tau-1} \left[1 - (1 - \rho)\frac{k}{t + i} \right] (1 - \rho)\frac{k}{t + \tau}$$

$$= k \frac{B(\tau, t)}{B(\tau, t - (1 - \rho))} \frac{1 - \rho}{t + \tau} \propto \frac{\tau^{-(1-\rho)k}}{t + \tau}, \qquad (6)$$

where we have again used that $B(x, y) \sim \Gamma(y)x^{-y}$ for large x. We observe a power-law decay with two scaling regimes. For $\tau \gg t$, the probability behaves as $\tau^{-1-(1-\rho)k}$. As the nth arriving group starts with $S_{n,t_n^{\text{init}}} = 1$ element, the exponent is $-(2 - \rho) > -2$ and the expected time for replication is infinite. Once a group has replicated, the expected time becomes finite. A newly arriving group is therefore greatly advantaged if it can step out of the mechanism and begin with even just two elements. Another variation allowing for later success is for elements to have variable inherent qualities as has been done, for example, for the Barabási-Albert model [9,41,42].
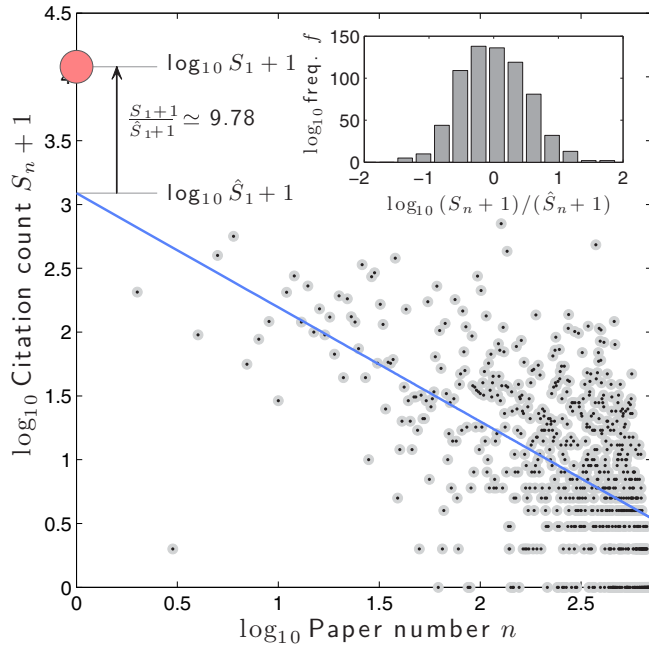
FIG. 3. Citation counts for (1) the original paper on scale-free networks by Barabási and Albert [9] (larger pale red disk) and (2) 692 subsequent papers with "scale-free networks" in their title ordered by publication date. We obtained citation counts from the Web of Science [44] on October 16, 2016. The blue line is the line of best fit using standard linear regression for all articles excluding the initiating article. A slope of -0.89 ± 0.09 gives an estimate of $\rho \simeq 0.11 \pm 0.09$. The first mover is a factor of 9.78 greater than would be consistent with the line of best fit, approximately the expected ratio of $1/\rho = 1/0.11 = 9.46$. The inset shows the binned residuals span two orders of magnitude, in keeping with the results shown in the inset of Fig. 2(b).

V. EMPIRICAL EVIDENCE OF A FIRST-MOVER ADVANTAGE OBEYING SIMON'S MODEL

While we do not expect systems with truly pure Simon mechanisms operating from a single initial element to be widespread (more below), one possible application lies in citation data for papers central to the development of a well-defined area of research. In Fig. 3, we show citation counts for papers with titles matching "scale-free networks" (gray dots) [43] along with citation counts for the incipient work of Barabási and Albert [9] (large red disk) (in the manner of [8], we add 1 to all citation counts). We used Web of Science [44] to find these papers and their citation counts obtaining a total of 692 on October 16, 2016. We note that although Barabási and Albert introduced the term in [9], they did not elevate it to being part of their own title. Of course, our model is not a perfect fit structurally as the citations are from all sources and each new paper would likely cite those coming before, especially [9]; we are also not attempting to capture the growth of a citation network.

Because our results show that a pure Simon model will produce a Zipf distribution with exponent $\alpha = 1 - \rho$ and a first-mover factor of $1/\rho$, we have two distinct measurable quantities that estimate ρ. Regressing through all but the initial paper's citation counts (blue line), we obtain a Zipf exponent

$\alpha = 0.89 \pm 0.09$ which in turn gives the estimate $\hat{\rho} = 1 - \alpha = 0.11 \pm 0.09$. The ratio of the number of citations for the first paper relative to that indicated by the regression is 9.78 which gives a very comparable value of $\hat{\rho} = 1/9.78 \simeq 0.10$, one that is well within standard error of the first estimate.

The inset for Fig. 3 shows the binned residuals for citation counts around the line of best fit for Zipf's law. The two orders of magnitude variation in citation counts conforms well with results from simulations displayed in the inset of Fig. 2(b).

Presuming all 692 papers cite the original Barabási and Albert paper, then a fraction $692/11982 \simeq 0.058$ of the first paper's citations come from papers centrally occupied with scale-free networks. Removing these and all citations internal to the set of 693 papers would not greatly alter the observed congruence between the first mover's outsized citation count and Zipf's law.

Improving and extending this analysis of citation data would be natural if potentially difficult to automate at a large scale. To be a potential fit for the pure Simon's model with the first-mover advantage considered a feature, a new research area would need to be started cleanly with a single paper. The paper would have to introduce a singular and lasting catchphrase for the focal topic, one that subsequent researchers developing the area would see fit to be included in a paper title. In order to build a sufficiently large data set, the analysis may also be successfully expanded to include papers with matches for the catchphrase in their abstracts rather than titles. We also note that we have ignored real time in our present analysis and that some modification of the model will be needed to properly accommodate the probability that the nth relevant paper appears at time t.

VI. CONCLUDING REMARKS

We close with some thoughts on how the dominant first-mover advantage of Simon's model and the variability of ordering may confound empirical analyses of real-world systems, and offer some potential resolutions.

As an idealized process, Simon's model evidently and purposefully fails to involve many aspects and details of real-world systems. Nevertheless, we must address the issue that Simon's rich-get-richer model has performed extremely well in analyses of real systems that do not exhibit a dominant group. Particular successes [45] have been found in measures of the innovation probability ρ and the Zipf exponent α with the expectation $\alpha = 1 - \rho$, as well as in the fraction of groups with one or two members [1,8,11,29].

In moving away from a pure Simon model, we consider three variant conceptions. A first possibility is to treat the first group as being of a different kind to those that come after by declaring the first element to be a kind of unobservable null element—its selection and replication at time t represents a failure of the system to produce any visible element. Given our analysis above, Zipf's law would shift with $n \to n - 1$ and then be closely approximated by a power-law decay [45]. However, this is a problematic mechanism as it requires the null element to be hidden from an observer of the system, while at the same time visible and equal in nature to all other elements from the point of view of the replication mechanism [46]. This mechanism is not the same as the one formed by

adding an overall master update probability to Simon's model (i.e., at each time step, engage the element-adding mechanism according to some fixed probability ρ_{add}). Such a modification would only serve to slow the dynamics. Further, starting the original model with no element at time $t = 1$ would also fail for as soon as the first element appears, the model would then act in the same way. The first-mover advantage of Simon's model cannot be dismissed by any meaningful reinterpretation of the mechanism.

A second modification would be to allow the innovation probability to vary with time (something Simon considered in [1]; see also [5,47]). Such a dynamic ρ_t is plausible for real-world systems [48], and is exemplified by Heap's law, the observation that "word birth" rate decays as a function of text or corpus length [49–52]. For example, the innovation rate may be initially high with new groups appearing rapidly, while in the long run, the system may stabilize in its expansion rate with the innovation probability ρ_t dropping and tending towards a constant. For such a dynamic, we would obtain a power-law tail for Zipf's law, but early on, a high innovation probability would suppress the first group, and smooth out the overall distribution. While in principle ρ_t could be estimated from data, great care would have to be taken given the stochastic evolution of a single run of a pure Simon model that we have demonstrated above and observed in real-world systems [11,29].

Finally, and perhaps more realistically, we may have a system that is initially configured by an entirely different growth mechanism up until some time t_0 at which a pure Simon model takes over. Building on our earlier analysis, we can instantiate a simple version of such a system with an initial $t = 1$ condition of S^{init} elements spread over n^{init} groups ranked by size ($S_{n,1} \geqslant S_{n+1,1}$ for $1 \leqslant n \leqslant n^{\text{init}}$). Allowing the basic rich-get-richer mechanism to then go into effect, the same approach of Eqs. (2)–(5) returns

$$S_{n,t} = \frac{B(t_n^{\text{init}}, 1 - \rho)}{B(S^{\text{init}} + t - 1, 1 - \rho)} S_{n,t_n^{\text{init}}}. \tag{7}$$

For large t, we have the approximate result

$$S_{n,t} \sim \begin{cases} \frac{S_{n,1}}{\Gamma(2-\rho)} \left[\frac{1}{S^{\text{init}} + t - 1} \right]^{-(1-\rho)} & \text{for } 1 \leqslant n \leqslant n^{\text{init}}, \\ \rho^{1-\rho} \left[\frac{n - n^{\text{init}}}{S^{\text{init}} + t - 1} \right]^{-(1-\rho)} & \text{for } n > n^{\text{init}}, \end{cases} \tag{8}$$

which reduces to Eq. (5) if $n^{\text{init}} = S^{\text{init}} = S_{1,1} = 1$. Equation (8) shows how the first-mover advantage is distributed across the groups present when Simon's model is introduced, and quantifies how the likelihood for newly arriving groups to replicate diminishes as the population size of the initial elements increases. For real-world systems with Zipf's laws that have clean power-law tails leading out from difficult-to-characterize forms for the largest groups, Eq. (8) offers a possible fit that would also estimate a time of transition from an establishing mechanism to a rich-get-richer one.

A parallel finding is that the initial network structure also matters for the Barabási-Albert growing network model [35,36], offering the possibility that the initial conditions and onset of the preferential attachment mechanism may be estimated from an observed network. For even relatively small seed networks with average degree 10, the asymptotic behavior of the degree distribution will not match the classic k^{-3} form found in [9]. A Simon-like first-mover advantage does not arise for the Barabási-Albert model with a very small seed network as a new node appears in each time step.

In sum, we have shown through simulations and analysis that Simon's fundamental rich-get-richer model—now 60 years old—carries an intrinsic first-mover advantage. The first group's size in the idealized model is outsized by a factor of $1/\rho$, potentially several orders of magnitude. Any attempt to attribute Simon's mechanism to the growth of real-world systems must take into account this potentially dominant feature of the model, along with the complications of the variability of final rankings for later arriving groups [10,27,28,35,36], and past work must come under a new scrutiny.

[1] H. A. Simon, Biometrika **42**, 425 (1955).

[2] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).

[3] G. K. Zipf, *Human Behaviour and the Principle of Least-Effort* (Addison-Wesley, Cambridge, MA, 1949).

[4] R. Ferrer i Cancho and R. V. Solé, J. Quant. Linguist. **8**, 165 (2001).

[5] J. R. Williams, J. P. Bagrow, C. M. Danforth, and P. S. Dodds, Phys. Rev. E **91**, 052811 (2015).

[6] G. U. Yule, Philos. Trans. B **213**, 21 (1925).

[7] D. J. de Solla Price, Science **149**, 510 (1965).

[8] D. J. de Solla Price, J. Am. Soc. Inform. Sci. **27**, 292 (1976).

[9] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[10] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[11] S. Bornholdt and H. Ebel, Phys. Rev. E **64**, 035104(R) (2001).

[12] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[13] R. Axtell, Science **293**, 1818 (2001).

[14] J. Doyle and J. M. Carlson, Phys. Rev. Lett. **84**, 5656 (2000).

[15] J. M. Carlson and J. Doyle, Proc. Natl. Acad. Sci. USA **99**, 2538 (2002).

[16] We connect a power-law size distribution $P(S) \sim S^{-\gamma}$ to the corresponding Zipf's law $S_r \sim r^{-\alpha}$ through the complementary cumulative distribution $P_{\geqslant}(S) = \int_{s=S}^{\infty} ds\, P(s) \sim S^{-\gamma+1}$. For a finite system with N elements, $N P_{\geqslant}(S)$ estimates the number of groups with size at least S. We see that this must also be

the rank r of the group of size S_r, and we have $r \simeq N P_{\geqslant}(S_r) \sim N S_r^{-\gamma+1}$. Because $S_r \sim r^{-\alpha}$, we therefore have $\gamma = 1 + 1/\alpha$ and equivalently $\alpha = 1/(\gamma - 1)$.

[17] Wikipedia, Yule-Simon distribution (January 2, 2017), https://en.wikipedia.org/w/index.php?title=YuleSimon_distribution&oldid=757940425.

[18] B. B. Mandelbrot, in *Communication Theory*, edited by W. Jackson (Butterworth, Woburn, MA, 1953), pp. 486–502.

[19] B. B. Mandelbrot, Information and Control **2**, 90 (1959).

[20] H. A. Simon, Information and Control **3**, 80 (1960).

[21] B. B. Mandelbrot, Information and Control **4**, 198 (1961).

[22] H. A. Simon, Information and Control **4**, 217 (1961).

[23] B. B. Mandelbrot, Information and Control **4**, 300 (1961).

[24] H. A. Simon, Information and Control **4**, 305 (1961).

[25] G. A. Miller, Introduction to reprint of G. K. Zipf's *The Psycho-Biology of Language* (MIT Press, Cambridge, MA, 1965).

[26] R. Ferrer-i-Cancho and B. Elvevåg, PLoS ONE **5**, e9411 (2010).

[27] M. E. J. Newman, Europhys. Lett. **86**, 68001 (2009).

[28] M. E. J. Newman, Europhys. Lett. **105**, 28002 (2014).

[29] T. Maillart, D. Sornette, S. Spaeth, and G. von Krogh, Phys. Rev. Lett. **101**, 218701 (2008).

[30] F. Tria, V. Loreto, V. D. P. Servedio, and S. H. Strogatz, Sci. Rep. **4**, 5890 (2014).

[31] V. Loreto, V. D. Servedio, S. H. Strogatz, and F. Tria, *Creativity and Universality in Language* (Springer, Berlin, Germany, 2016), pp. 59–83.

[32] J. R. Williams, P. R. Lessard, S. Desu, E. M. Clark, J. P. Bagrow, C. M. Danforth, and P. S. Dodds, Sci. Rep. **5**, 12209 (2015).

[33] Online appendices can be found at http://compstorylab.org/share/papers/dodds2016b/.

[34] L. Hébert-Dufresne, A. Allard, V. Marceau, P.-A. Noël, and L. J. Dubé, Phys. Rev. E **85**, 026108 (2012).

[35] Y. Berset and M. Medo, Eur. Phys. J. B **86**, 1 (2013).

[36] B. Fotouhi and M. G. Rabbat, Phys. Rev. E **88**, 062801 (2013).

[37] J. P. Gleeson, J. A. Ward, K. P. O'Sullivan, and W. T. Lee, Phys. Rev. Lett. **112**, 048701 (2014).

[38] J. P. Gleeson, K. P. O'Sullivan, R. A. Baños, and Y. Moreno, Phys. Rev. X **6**, 021019 (2016).

[39] D. Sornette and G. Ouillon, Eur. Phys. J.: Spec. Top. **205**, 1 (2012).

[40] Wikipedia, King effect (May 7, 2016), https://en.wikipedia.org/w/index.php?title=King_effect&oldid=719041025.

[41] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).

[42] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).

[43] We have chosen the restriction to such titles so as to limit ourselves to publications that are primarily concerned with the topic of scale-free networks; we acknowledge the crude nature of this approach and that better and far more exhaustive analyses would be available for future work.

[44] Web of Science, http://webofknowledge.com/ (accessed on October 16, 2016).

[45] J. R. Williams, J. P. Bagrow, A. J. Reagan, S. E. Alajajian, C. M. Danforth, and P. S. Dodds, Zipf's law is a consequence of coherent language production, arXiv:1601.07969.

[46] One possible exception may be language evolution, with whitespace framed as a kind of null element [45].

[47] M. Gerlach and E. G. Altmann, Phys. Rev. X **3**, 021006 (2013).

[48] W. J. Baumol, Am. Econ. Rev. **76**, 1072 (1986).

[49] H. S. Heaps, *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Inc., Orlando, FL, 1978).

[50] G. Herdan, *Type-Token Mathematics*, Vol. 4 (Mouton, The Hague, Netherlands, 1960).

[51] A. M. Petersen, J. N. Tenenbaum, S. Havlin, H. E. Stanley, and M. Perc, Sci. Rep. **2**, 943 (2012).

[52] E. A. Pechenick, C. M. Danforth, and P. S. Dodds, arXiv:1503.03512.