

Re-inventing Willis

M.V. Simkin and V.P. Roychowdhury

Department of Electrical Engineering, University of California, Los Angeles, CA 90095-1594

Scientists often re-invent things which were long known. Here we review these activities as related to the mechanism of producing power law distributions, originally proposed in 1922 by Yule to explain experimental data on the sizes of biological genera, collected by Willis. We estimate that scientists are busy re-discovering America about 2/3 of time.

The book of Ecclesiastes says: “Is there any thing whereof it may be said, See, this is new? It hath been already of old time, which was before us.” The job of the scientists is to discover new things, thus they are the most affected. In this paper we report a case study involving the mechanism of generating power-law distributions known as Yule’s process, Simon’s model, Cumulative Advantage, or Preferential Attachment. We also discuss the relation of Yule’s process to Galton-Watson process and to Self-Organized Criticality.

Yule’s process

In 1922 Willis and Yule [1] analyzed the data on frequency distribution of the sizes of biological genera, collected by Willis [2]. They discovered that this distribution follows a power law. To explain this observation Yule [3] proposed the following model¹. Consider two types of mutations: specific (that is producing a new specie of the same genus), which occur in each specie at rate s , and generic (that is producing a new genus) which occur in each genus at rate g .

In this model the expectation value of the total number of genera grows with time as $\exp(g \times t)$. Therefore the probability distribution of the ages of genera is:

$$p(t) = g \exp(-gt). \quad (1)$$

The expectation number of species in a genus of age t is:

$$n(t) = \exp(s \times t). \quad (2)$$

Now suppose that chance can be ignored, that the number of species in a genus can be taken as a continuous variable, and that the above can be taken as absolute functional relations. The size of genus is then absolutely determined by its age, and we can find the number of genera of each size by eliminating t from Eq. (1) using Eq. (2):

$$t(n) = \frac{1}{s} \ln(n); \quad dt = \frac{dn}{s \times n}.$$

This leads to:

$$p(n) = \frac{g}{s} n^{-\gamma}; \quad \gamma = 1 + \frac{g}{s}. \quad (3)$$

The exact solution was also found by Yule [3]. The probability of a genus of an age t to be monotypic, i.e. to consist of just one specie, is, obviously, $p_1(t) = \exp(-st)$. For a genus to contain two species at time t , a single mutation must occur at some intermediate time t_1 , the original specie must not mutate until t_1 , and two resulting species must not mutate for the time $t - t_1$. The probability for a genus to contain two species is obtained by integrating over all possible values of t_1 :

$$p_2(t) = \int_0^t dt_1 p_1(t_1) s \exp(-2s(t-t_1)) = \\ \exp(-st)(1 - \exp(-st))$$

In general, it can be verified by induction that:

$$p_n(t) = \exp(-st)(1 - \exp(-st))^{n-1} \quad (4)$$

¹ A quarter of this paper consists of relevant passages copied from original papers. This is not plagiarism: the latter is copying from a single source. In contrast, copying from many sources is original in-depth research.

We see that the size distribution of genera of the same age is exponential. This result was rediscovered in 1992 by Günter *et al* [4] and in 2000 by Krapivsky and Redner [5], who used more complicated mathematical methods.

Combining the distribution of the number of species in genera of age t (given by Eq.(4)) with the distribution of the ages of genera (Eq.(1)) we obtain the overall probability distribution of genera with regard to the number of species:

$$\begin{aligned} p_n &= \int_0^\infty dt p_n(t) p(t) = \\ &\int_0^\infty dt \exp(-st)(1-\exp(-st))^{n-1} g \exp(-gt) = \\ &\frac{g}{s} \int_0^1 dx x^{\frac{g}{s}} (1-x)^{n-1} = \\ &\frac{g}{s} B\left(\frac{g}{s} + 1, n\right) = \frac{g}{s} \frac{\Gamma\left(\frac{g}{s} + 1\right) \Gamma(n)}{\Gamma\left(\frac{g}{s} + 1 + n\right)} \end{aligned} \quad (5)$$

Here B and Γ are Euler's Beta and Gamma functions.

The large- n asymptotic of Eq.(5) is

$$p_n \propto \frac{g}{s} \frac{\Gamma\left(\frac{g}{s} + 1\right)}{n^{\frac{g+1}{s}}}. \quad (6)$$

Modified Yule's model

We introduce this modification in order to make the similarity between Yule's and Simon's model (which we discuss later) more obvious. Let us assume that the rate of generic mutations is proportional to the number of species in a genus. In this case the number of genera grows with time as $\exp((s+g)\times t)$ and the age distribution of genera is:

$$p(t) = (s+g) \exp(-(s+g)t).$$

Substituting this into Eq.(5) we get:

$$\begin{aligned} p_n &= \int_0^\infty dt p_n(t) p(t) = \\ &\int_0^\infty dt \exp(-st)(1-\exp(-st))^{n-1} (s+g) \exp(-(s+g)t) \\ &= \frac{s+g}{s} \int_0^1 dx x^{\frac{s+g}{s}} (1-x)^{n-1} \\ &= \frac{s+g}{s} B\left(\frac{g}{s} + 2, n\right) = \\ &\frac{s+g}{s} \frac{\Gamma\left(\frac{g}{s} + 2\right) \Gamma(n)}{\Gamma\left(\frac{g}{s} + 2 + n\right)} \propto \frac{s+g}{s} \frac{\Gamma\left(\frac{g}{s} + 2\right)}{n^{\frac{g+2}{s}}} \end{aligned} \quad (7)$$

The major difference between the probability distribution, generated by the modified Yule's model (Eq.(7)) and the one generated by the standard Yule's model (Eq.(6)) is that Eq.(7) has the exponent of the power law equal to $\gamma = 2 + \frac{g}{s}$, while Eq.(6) has it

equal to $\gamma = 1 + \frac{g}{s}$. By changing parameters g and s , Eq.(6) can be tuned to have an exponent of any value greater than 1, while Eq.(7) has the exponent always greater than 2. Willis' data could be fitted with a power law with an exponent of around 1.5. Only standard Yule's model can account for it, while modified can't.

Champernowne's model of income distribution

To explain the power law in distribution of incomes, discovered by Pareto, Champernowne [6] invented the following model. He divided income recipients into ranges of equal proportionate width. That is, if I_{min} is the minimum income considered, then the first range contains persons with incomes between I_{min} and aI_{min} , the second range persons with incomes between aI_{min} and $a^2 I_{min}$, and so on. Next he introduces transition probabilities r_{nm} that a person who is in class m at time t will be in class n at time t

+ 1. He assumes that r_{nm} is a function of only $n-m$ (except for the small m , when it is modified to prohibit falling below minimum income).

To illustrate the Champernowne's model we will consider the simplest nontrivial transition function, the one which only allows transitions between adjacent classes:

$$r_{nm} = r(n-m) = \begin{cases} r_1 & \text{when } n-m=1 \\ r_0 & \text{when } n-m=0 \\ r_{-1} = 1 - r_1 - r_0 & \text{when } n-m=-1 \end{cases} \quad (8)$$

$$r_{n0} = \begin{cases} r_1 & \text{when } n=1 \\ r_0 + r_{-1} = 1 - r_1 & \text{when } n=0 \end{cases}$$

In equilibrium the occupation probabilities, p_n , of income ranges, n , should satisfy the following equations:

$$\begin{aligned} p_n &= r_0 p_n + r_1 p_{n-1} + r_{-1} p_{n+1} \\ p_0 &= (r_0 + r_{-1}) p_0 + r_{-1} p_1 \end{aligned},$$

which in the case $r_1 < r_{-1}$ have following solution:

$$p_n = (1 - r_1 / r_{-1}) \times (r_1 / r_{-1})^n. \quad (9)$$

The occupation probabilities decrease exponentially with income ranges range number, n . As p_n is the probability to have income between $a^{n-1} I_{min}$ and $a^n I_{min}$ - income exponentially grows with the range number. The situation is similar to what was in Yule's model, with time replaced by n . This leads to the income distribution of the form:

$$p(I) \sim I^{-\gamma}, \quad \gamma = 1 + \frac{\ln(r_{-1}/r_1)}{\ln(a)}. \quad (10)$$

A continuum version of Champernowne's model was re-discovered in 1996 by Levy and Solomon [7].

Simon's model

The distribution of words by the frequency of their use follows a power law. This fact was discovered sometime before 1916 by Estoup [8], re-discovered in 1928 by Condon [9], and once more in 1935 by Zipf [10]. Nowadays it is widely known as Zipf's law. To explain this observation Simon [11] proposed the following model. Consider a book that is being written, and that has reached a length of N words. With probability α the $(N+1)$ -st word is a new word – a word that has not occurred in the first N words. With probability $1-\alpha$ the $(N+1)$ -st word is one of the old words. The probability that it will be a particular old word is proportional to the number of its previous occurrences.

If certain word appeared K times among N words, the probability that $(N+1)$ -st word will be this word is $(1-\alpha) \times \frac{K}{N}$. The evolution of the word frequency distribution (here N_K denotes the number of words that appeared K times) is described by the following equations:

$$\begin{aligned} \frac{dN_1}{dN} &= \alpha - (1-\alpha) \times \frac{N_1}{N} \quad (11) \\ \frac{dN_K}{dN} &= (1-\alpha) \times \frac{(K-1) \times N_{K-1} - K \times N_K}{N} \quad (K > 1), \end{aligned}$$

Assuming that the distribution of words has reached its stationary state we can replace the derivatives with the ratios:

$$\frac{N_1}{N} = \alpha - (1-\alpha) \times \frac{N_1}{N} \quad (12)$$

$$\frac{N_K}{N} = (1-\alpha) \times \frac{(K-1) \times N_{K-1} - K \times N_K}{N} \quad (K > 1) .$$

The probability that the word occurs K times is equal to

$$P_K = \frac{N_K(N)}{D(N)}, \quad (13)$$

where $D(N) = \alpha N$ is the number of distinct words in the book.

We thus have:

$$N_K(N) = P_K \alpha N \quad (14)$$

After substituting Eq. (14) in Eq. (12) we get:

$$\begin{aligned} P_1 &= \frac{1}{2-\alpha}, \\ \frac{P_K}{P_{K-1}} &= \frac{K-1}{K+1/(1-\alpha)} \quad (K>1). \end{aligned} \quad (15)$$

Iterating the above equation we get

$$\begin{aligned} P_K &= \frac{\Gamma(K)\Gamma(2+1/(1-\alpha))}{\Gamma(K+1+1/(1-\alpha))} \times \frac{1}{2-\alpha} = \\ &\frac{\Gamma(K)\Gamma(1+1/(1-\alpha))}{\Gamma(K+1+1/(1-\alpha))} \times \frac{1}{1-\alpha} \\ &\propto \frac{1}{1-\alpha} \times \frac{\Gamma(1+1/(1-\alpha))}{K^{1+1/(1-\alpha)}} \end{aligned} \quad (16)$$

In 1976 Price [12] used Simon's model to explain a power law distribution of citations to scientific papers, which he discovered in 1965 [13]². He proposed to call it "cumulative advantage process". Simon's model was re-discovered in 1992 by Günter *et al* [4] and 1999 by Barabasi and Albert [16]. In the latter case it acquired a new name: "preferential attachment".

Solution of the modified Yule's model by Simon's method

In Simon's model genus corresponds to a distinct word, and number of species in a genus corresponds to the number of occurrences of a word in the text. The probability of a new mutation to be a generic mutation corresponds to

the probability for the next word to be a new word. It is equal to:

$$\alpha = \frac{g}{s+g}$$

Substituting this in Eq. (16) we recover Eq. (7)

Solution of Yule's model by Simon's method

Original Yule's model is more difficult to solve by Simon's method. The problem is that the probability of a new mutation to be generic changes in time. This probability is given by the equation:

$$\alpha = \frac{gN_g}{sN_s + gN_g}, \quad (17)$$

where N_g is the total number of genera, and N_s the total number of species. Let us compute these numbers at time t . Suppose, that at time 0 there was one genus, consisting of single specie. The expectation number of genera at time t is, obviously,

$$N_g = \exp(g \times t). \quad (18)$$

The expectation number of species in the primal genus at time t is:

$$N_s^1 = \exp(s \times t)$$

The number of species in new genera is:

$$\begin{aligned} N_s^* &= \int_0^t dt_1 g \exp(gt_1) \exp(s(t-t_1)) = \\ &\frac{g}{g-s} (\exp(gt) - \exp(st)) \end{aligned}$$

The expectation number of all species at time t is:

$$N_s = N_s^1 + N_s^* = \frac{g}{g-s} \exp(gt) + \frac{s}{s-g} \exp(st)$$

The large t asymptotic of the above is:

² This was rediscovered in 1997 by Silagadze[14] and in 1998 by Redner [15].

$$N_s = \begin{cases} \frac{g}{g-s} \exp(gt) & \text{when } g > s \\ \frac{s}{s-g} \exp(st) & \text{when } s > g \end{cases} \quad (19)$$

Let us consider the case $g > s$. Substituting Eqs.(18) and (19) into Eq.(17) we get:

$$\alpha = \frac{ge^{gt}}{s - \frac{g}{g-s}e^{st} + ge^{gt}} = 1 - \frac{s}{g} \quad (20)$$

Substituting Eq.(20) into Eq.(14) we recover Eq.(5).

Let us now consider the case $s > g$. After substituting Eqs.(18) and (19) into Eq.(17) we get:

$$\alpha(t) = \frac{ge^{gt}}{s - \frac{s}{s-g}e^{st} + ge^{gt}} \propto g \frac{s-g}{s^2} e^{-(s-g)t} \quad (21)$$

After expressing t through N_s , using Eq.(19), and substituting the result into Eq.(21) we get:

$$\alpha(N_s) = CN_s^{-\gamma}; \gamma = 1 - \frac{g}{s}. \quad (22)$$

Here C is a function of g and s .

Let us consider a modified Simon's model where

$$\alpha = CN^{-\gamma}. \quad (23)$$

The number of distinct words as a function of total number of words will be

$$D(N) = \int_0^N dMCM^{-\gamma} = \frac{C}{1-\gamma} N^{1-\gamma} \quad (24)$$

After substituting Eq.(24) into Eq.(13) and the result together with Eq.(23) into Eq.(12) we get:

$$P_1 = \frac{1-\gamma}{2-\gamma}$$

$$\frac{P_K}{P_{K-1}} = \frac{K-1}{K+1-\gamma} \quad (25)$$

By iteration of Eq. (25) we get:

$$P_K = (1-\gamma) \frac{\Gamma(K)\Gamma(2-\gamma)}{\Gamma(K+2-\gamma)} \quad (26)$$

after substituting $\gamma = 1 - \frac{g}{s}$ into Eq.(26) we recover Eq.(5).

This model with N -dependent α was first suggested and solved by Simon [11]. It was rediscovered by Dorogovtsev and Mendes [17] in the context of science of networks.

This exercise shows that Yule's and Simon's models are two ways of looking at the same thing. In contrast, Champernowne's model is similar, but not identical.

Markov-Eggenberger-Polya Urn models

The urn contains one white and one black ball. We take out a ball at random, than put it back and add another ball of the same color. The procedure is repeated again and again. This model was considered in 1907 by Markov [18], and re-discovered in 1923 by Eggenberger and Polya (see Ref. [19] pp.176-177). The model is nowadays known as Polya's urn scheme. Obviously, the procedure has cumulative advantage feature of Yule-Simon process. This connection was, probably, first pointed out by Price [12] in 1975, and re-discovered in 2003 by Chung, Handjani, and Jungreis [20]. Markov's problem is easier to solve than Yule's. After one step in the urn there can be either one black and two white balls or one white and two black balls with equal probability. It is easy to show that after two steps the possible combinations are 3-1, 2-2, and 1-3, and all of them have the same probability. In general, after N steps there can be any number of white balls from 1 to $N+1$ all with equal probabilities $1/N$. This can be proved by induction. If this holds true for N steps, than the probability to have n white balls after $N+1$ steps is

$$p(n, N+1) = \frac{n-1}{N+1} p(n-1, N) + \frac{N+1-n}{N+1} p(n, N) = \frac{N}{N+1} \frac{1}{N} = \frac{1}{N+1}$$

This uniform distribution of balls reminds the zero-temperature distribution of particles obeying Bose-Einstein statistics, when the ground level is doubly degenerate. The relation of cumulative advantage process to Bose-Einstein statistics was pointed out in [21], [22], [23]. If we modify the urn scheme so that with probability α we add a ball of a new color – we recover the Yule-Simon process.

Spectrum of cosmic radiation

In 1949 Fermi [24] explained experimentally observed power-law spectrum of cosmic radiation as follows. The particles, like protons, are accelerated through collisions with wandering interstellar magnetic fields. An elementary estimate can be obtained by picturing the collisions against reflecting obstacles of very large mass, moving with random velocities averaging to V . Assuming this picture, one finds easily that the average gain in energy per collision is given as

order of magnitude by $dE = \left(\frac{V}{c}\right)^2 E$, where E

represents the energy of the particle inclusive of rest energy, and c is the speed of light. If we call τ the time between scattering collisions, the energy acquired by a particle of age t will

be $E(t) = Mc^2 \exp\left(\left(\frac{V}{c}\right)^2 \frac{t}{\tau}\right)$. During the process

of acceleration a proton may loose most of its energy by a nuclear collision. This absorption process can be considered to proceed according to an exponential law. We expect the age distribution to be $E(t) = \exp\left(-\frac{t}{T}\right)$, where T is the time

between absorption collisions. Combining relationships between age and energy with the probability distribution of age, we find the

probability distribution of the energy:

$$p(E) \propto E^{-1 - \frac{\tau \left(\frac{c}{V}\right)^2}{T}}$$

Renormalization group

Near critical temperature, T_c , of a second order phase transitions physical parameters of the system are power-law functions of the reduced temperature,

$t = \frac{T - T_c}{T_c}$. In 1971 Wilson [25] developed the

renormalization group (RG) method to explain this phenomenon. He studied how parameters of the system change after n successive re-scalings of the length by a factor l . He found that the reduced temperature grows exponentially with n :

$t^{(n)} = (l^{y_t})^n t$. At the same time correlation length decreases exponentially with n : $\xi^{(n)} = l^{-n} \xi$. As a result the correlation length scales with the reduced temperature as: $\xi(t) \propto t^{-1/y_t}$. Similar to what happened in Yule's model, the power law came out of two exponential dependencies.

Bradford's law

Bradford Law states that the distribution of scientific journals by the number of articles, they contain, follows a power law. In 1970 Naranan [26] considered the model where both the number of journals and the size of each journal increase exponentially with time. He obtained an equation identical to Eq. (3) with g and s being growth rates of the number of journals and number of articles in a journal respectively. This very mechanism was rediscovered in 1999 by Huberman and Adamic [28] in the context of distribution of websites by the amount of webpages they contain. Both aforementioned reinventions ([26], [27]) appeared in the same journal (*Nature*), were the original paper by Willis and Yule [1] was published

In his first paper on this topic [26] Naranan mentioned the 1949 paper by Fermi [24] as the source of the idea. Later he found [27] that that the original idea was older and that it was several times

re-invented. He gave a list of related papers, from which we benefited in writing this article.

Monkey-typed text

In 1957 Miller [29] had explained Zipf's law in word frequencies as follows. A monkey, sitting at a typewriter, is randomly hitting its keys. The resulting discourse is a random sequence of independent symbols: the letters and the space that mark the boundaries between words. There are M letters. The probabilities of all letters and of space are equal $1/(M+1)$. If a word contains m letters, the probability for it to be printed by a monkey is $M^m/(M+1)^{m+1}$. The number of m -letter words is M^m . Now m plays again the role of t in Yule's problem. Thus we get a power law distribution with exponent $\gamma = 1 + \frac{\ln(M/(M+1))}{\ln(M)} = 2 - \frac{\ln(M+1)}{\ln(M)}$.

This explanation was rediscovered in 1992 by Li [30]. By the way, Li credited Miller for the idea, citing his Introduction to the book by Zipf [10], where Miller wrote that typing monkeys will produce Zipf's law, but did not give a mathematical proof. Apparently, Li decided that Miller never produced that proof.

A more interesting thing is that Miller's explanation of Zipf's law was a re-invention of the explanation proposed by Mandelbrot back in 1953 [31]. Mandelbrot derived Zipf's law from optimization criteria. He considered a word-by-word coding of a text using q elementary symbols, one of which is used as a separator between words. He asked a question what distribution of word frequencies will lead to maxim Shannon information per symbol of encoded text. After some algebra he found that this distribution should be Zipfian. Now note that maximum Shannon entropy is achieved when the sequence of symbols is completely random. We see that Miller just solved the Mandelbrot's problem by another method (a simpler one). Mandelbrot understood that these two explanations of Zipf's Law are identical (see [32]), but some authors, who wrote on the issue lately, don't.

Psychophysical law

Perceived intensity varies with physical intensity as a power law [33]. For example, for brightness the exponent is about 1/3. This behavior is puzzling since frequency of nerve impulses from sensory receptors is related logarithmically to intensity of stimulations. In 1963 MacKay [34] proposed the matched response model, which can resolve apparent paradox. In his model perception is adaptive response to stimulation. Perceived intensity of a stimulus reflects not the frequency of impulses from receptor organ, but the magnitude of internal activity evoked to match or counterbalance that frequency. MacKay proposed that frequency of matching impulses is logarithmically related to the magnitude of internal activity. In such case both perceived intensity and physical intensity depend exponentially on the frequency of nerve impulses and we get a power law relation between them, similarly how we get it in Yule's model.

The Science of Networks

In 1999 in order to explain the power-law distribution of connectivity in the World Wide Web [35] and other networks Barabasi and Albert [16] proposed the following model. Starting with a small number of nodes, at every time step we add a new node and link it to one of the nodes already present in the system. When choosing the nodes to which the new node connects, we assume that the probability p_i that a new node will be connected to node i is proportional to its degree k_i . With the substitution $node = distinct\ word$ and $degree = number\ of\ occurrences$ Barabasi-Albert model reduces to Simon's model with $\alpha=1/2$. This was pointed out in [36].

Galton-Watson processes

In the 19th century some British gentlemen had noticed that many families that had occupied conspicuous positions in the past became extinct. At first they concluded that increase in intellectual capacity is accompanied by a decrease in fertility. However, in 1875 Watson and Galton [37] developed the theory of branching processes, which showed that a large proportion of families (or surnames) should become extinct by the ordinary law of chances. The

mathematical model considered was that in each generation $p(0)$ per cent of the adult males have no sons, $p(1)$ have one son and so on. The model can be solved using method of generating functions and apart from the probability of extinction one can derive the exact formula for the probability distribution of total offspring (i.e. the sum of the numbers of sons, grandsons, great-grandsons and so on). In the case when the average number of sons per adult male equals one, this turns out to be a power law with exponent 3/2. One can also compute an approximate expression for the probability distribution of surname frequencies in a given generation. This turns out to be a power law with exponent 1. Branching processes were re-discovered in genetics (replace surname with gene), chemical chain reactions, nuclear chain reactions, cosmic rays (see the book by Harris [38]) and most recently in sandpiles [39].

Galton-Watson process is similar to Yule's process. If we apply Yule's model to find distribution of species by number of individuals than "specie" will be replaced with "individual" and "genus" with "specie". Specific mutation will correspond to a birth without mutation and generic mutation to specific mutation. These are only verbal changes, which do not affect math. Now remember that, unlike species, individuals are mortal. We will have to adjust Yule's model for that, and it will transform it into Galton-Watson model.

The scientists will appreciate the following illustration of the connection between Yule-Simon and Galton-Watson processes. A model where, when scientist is writing a manuscript, he picks up few random papers, cites them, and also copies a fraction of their references leads to Yule-Simon process. A modification of that model: when scientist is writing a manuscript, he picks up few random *recent* papers, cites them, and also copies a fraction of their references leads to Galton-Watson process. The only difference between the two models is the word *recent*. See Ref. [40] for details.

The Science of Self-Organized Criticality

In its mean-field version Self Organized Criticality (SOC) can be described as a Galton-Watson branching process [41], [42]. Here the sand grains, which are moved after the original toppling, are equivalent to sons. This moved grains can cause further toppling, resulting in the motion of more grains, which are equivalent to grandsons, and so on. The total number of displaced grains is the size of the avalanche and is equivalent to total offspring in the case of a branching process. Size distribution of offspring is equivalent to distribution of avalanches in SOC. To be fair SOC is not completely reduced to a critical branching process: it has a built in dissipative mechanism (sand grains fall from the edges), which tunes the branching process into a critical state.

Genetic model of Moran

A model of the evolution of a gene pool can be formulated by replacing "surname" with "gene" in Galton-Watson model. Unfortunately, this model is not exactly soluble. However, a modification of the model, suggested in 1962 by Moran [43] and further studied by Karlin and McGregor [44], is exactly soluble. In this model genes die *one by one* at random and are replaced with new genes.

The exact model we consider is as follows. There is a gene pool of fixed but large size. At each step one selected at random gene dies. We select from the pool a random gene and add to the pool a gene of the same type. With the probability α the added gene can mutate.

Moran's model can be reformulated in verbal terms. Consider a string of words of large but fixed length N . At each step we delete one randomly selected word and add one word according to the rules of Simon's model. (With probability α it is a new word. With probability $1-\alpha$ it is one of the old words. The probability that it is a particular old word is proportional to the number of its previous occurrences.)

Equilibrium word frequencies should satisfy the following equations (here N_K denotes the number of words that appeared K times):

$$-\frac{N_1}{N} + \frac{2N_2}{N} - (1-\alpha) \times \frac{N_1}{N} + \alpha = 0 \quad (27a)$$

$$\begin{aligned} & -\frac{K \times N_K}{N} + \frac{(K+1) \times N_{K+1}}{N} - \\ & (1-\alpha) \times \frac{K \times N_K}{N} + \quad \quad \quad (K > 1) \quad (27b) \\ & (1-\alpha) \times \frac{(K-1) \times N_{K-1}}{N} = 0 \end{aligned}$$

To find N_1 note that the rate of creation of new words is α . At the same time the rate of loss of distinct words is $\frac{N_1}{N}$: clearly, a word gets extinct when a word which occurs only once is selected for deletion. It follows that $\frac{N_1}{N} = \alpha$. Substituting this in Eq.(27a), we get $\frac{N_2}{N} = \frac{\alpha(1-\alpha)}{2}$, and, similarly, $\frac{N_3}{N} = \frac{\alpha(1-\alpha)^2}{3}$. In general, it can be verified by induction that

$$\frac{N_K}{N} = \frac{\alpha(1-\alpha)^{K-1}}{K}. \quad (28)$$

The probability that the word occurs K times is equal to,

$$P_K = \frac{N_K}{D}, \quad (29)$$

where D is the number of distinct words in the string. The latter can be computed as

$$D = \sum_{K=1}^{\infty} N_K \quad (30)$$

After substituting Eq.(28) in Eq.(30), and the result into Eq.(29) we get:

$$P_K = \frac{(1-\alpha)^K}{\ln(1/\alpha)K}. \quad (31)$$

The word frequency distribution follows a hyperbolic law with an exponential cut-off.

A small modification of Moran's model makes it very similar to Simon's model. The rules of word addition remain the same. The rule of deletion will be as follows. The deletion happens with

probability α . We select a random *distinct* word (the probability is equal for all distinct words in the sequence and does not depend on the number of occurrence of the word). Then we delete all occurrences of this word from the sequence. With these new rules the equilibrium equations (27) will change only slightly. Third and fourth terms will not change, as the rules of addition didn't change. Second term will disappear, because if a word is deleted than all its occurrences are deleted, it does not just move from N_K to N_{K-1} category, like it was in previous model. The first term will become $-\alpha \frac{N_K}{D}$, where D is the number of distinct words in the sequence. To compute D note that the average number of occurrences of a word is N/D . In equilibrium the rate of addition must be equal to the rate of deletion. Therefore $\alpha N/D = 1$, or $D = \alpha N$. The first term becomes $\frac{N_K}{N}$, and Eq.(27) transform into:

$$\begin{aligned} & -\frac{N_1}{N} - (1-\alpha) \times \frac{N_1}{N} + \alpha = 0, \\ & -\frac{N_K}{N} + (1-\alpha) \times \frac{(K-1) \times N_{K-1}}{N} - \\ & (1-\alpha) \times \frac{K \times N_K}{N} = 0 \quad (K > 1) \quad (32) \end{aligned}$$

The above equations are identical to Eq. (12) and therefore have the same solution, Eq.(15). Although the above model has the same solution as standard Simon's model, it is not identical to it. In the latter words are added but not deleted and the number of words in the sequence is growing. In the above model words are both added and deleted and the number of words in the sequence is constant. This model was formulated and solved by Simon in [11].

Cumulative advantage and Bose-Einstein statistics

In Simon's model the probability that a new word is a word that has already appeared exactly N times is proportional to N . This resembles the transition rules for particles obeying Bose-Einstein statistics. For such particles the probability to move to certain

quantum level is proportional to the number of particles already sitting on that level. This was first pointed out in 1974 by Hill [21] and further discussed by Ijiri and Simon (see the Chapter "Some distributions associated with Bose-Einstein statistics" in Ref. [22]). This link was rediscovered in 2001 by Bianconi and Barabasi [23].

Re-inventing Willis

We summarized the re-inventions, described in this paper, in Table 1. We treat Simon's and Yule's models as different things, because they

use different mathematical approaches (alternative ways to America). We count it as a re-discovery when the same America is discovered in the same way. Even with this restriction almost everything appears to be re-discovered twice.

We conclude that re-discovering America is the most common scientific occupation, both in this country and abroad: the scientists are busy with it two thirds of time.

Table 1. Re-inventing Willis.

Phenomenon	Discovered	Re-discovered	
Yule's process	Yule (1925)	Fermi (1949)	Huberman and Adamic (1999)
Simon's process	Simon (1955)	Günter et al (1992)	Barabasi and Albert (1999)
Champernowne's process	Champernowne (1953)	Levy and Solomon (1996)	
Power law of word frequencies	Estoup (before 1916)	Condon (1928)	Zipf (1935)
Power law of scientific citing	Price (1965)	Silagadze (1997)	Redner (1998)

1. J.C. Willis and G.U. Yule, "Some statistics of evolution and geographical distribution in plants and animals, and their significance", *Nature*, **109**, 177 (1922).
2. J.C. Willis, *Age and area: a study in geographical distribution and origin of species* (University press, Cambridge, 1922).
3. G.U. Yule, "A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.", *Philosophical Transactions of the Royal Society of London B*, **213**, 21 (1925).
4. R. Günter, B. Schapiro, and P. Wagner, "Physical complexity and Zipf's law", *International Journal of Theoretical Physics* **31**, 524 (1992); R. Günter, L. Levitin, B. Schapiro, and P. Wagner, "Zipf's law and the effect of ranking on probability distributions", *International Journal of Theoretical Physics* **35**, 395 (1996).
5. P. L. Krapivsky and S. Redner, "Organization of growing random networks", *Phys. Rev. E*, **63**, 066123 (2001).
6. D.G. Champernowne, "A model of income distribution", *Economic Journal* **63**, 318 (1953).
7. M. Levy and S. Solomon, "Dynamical Explanation for the Emergence of Power Law in a Stock Market Model", *Int. J. Mod. Phys. C*, **7**, 65 (1996); M. Levy and S. Solomon, "Power laws are logarithmic Boltzmann laws", *Int. J. Mod. Phys. C*, **7**, 595 (1996), <http://www.arxiv.org/abs/adap-org/9607001>
8. Estoup, J.B., *Gammes sténographique*, 1916 (as cited in Ref. [32]). See also: <http://torvald.aksis.uib.no/corpora/2002-2/0070.html>
9. E.U. Condon, "Statistics of vocabulary", *Science*, **67**, 300 (1928).
10. G.K. Zipf, *The psycho-biology of language* (MIT Press, Cambridge, 1965). This is a re-issue of the book originally published in 1935.
11. H.A. Simon, "On a class of skew distribution functions", *Biometrika*, **42**, 425 (1955).
12. D. de S. Price, "A general theory of bibliometric and other cumulative advantage process", *Journal of American Society for Information Science*, **27**, 292 (1976).
13. D. de S. Price, "Networks of Scientific Papers", *Science*, **149**, 510 (1965).
14. Silagadze, Z.K., "Citations and Zipf-Mandelbrot law", *Complex Syst.* **11**, 487 (1997).
15. Redner, S., "How popular is your paper? An empirical study of citation distribution", *cond-mat/9804163; Eur. Phys. J. B* **4**, 131 (1998).
16. Barabasi, A.-L. and R. Albert, "Emergence of scaling in random networks", *Science*, **286**, 509 (1999).
17. S.N. Dorogovtsev and J.F.F. Mendes, "Evolution of networks", *Advances in Physics*, **51**, 1079 (2002). (see the chapter "accelerated growth of networks")
18. A.A. Markov "Extension of the law of large numbers to dependent variables" in A.A. Markov, *Selected Works*, (Izdatel'stvo Akademii Nauk SSSR, Moscow, 1951) (in Russian). (The original paper was published in 1907 in an obscure journal.) The ball problem is discussed on pp.351-354.
19. N.L. Johnson and S. Kotz, *Urn Models and their Application*, (John Wiley, New York, 1977).
20. F. Chung, S. Handjani, and D. Jungreis, "Generalizations of Polya's urn Problem", *Annals of Combinatorics* **7** (2003) 141-153.
21. B.M. Hill "Zipf's law and prior distributions for the composition of a population", *Journal of American Statistical Association*, **65**, 1220 (1970); B.M. Hill "The rank-frequency form of Zipf's law", *Journal of American Statistical Association*, **69**, 1017 (1974).
22. Y. Ijiri and H.A. Simon "Skew distributions and the size of business firms" (North-Holland, Amsterdam, 1977)
23. G. Bianconi and A.-L. Barabási, "Bose-Einstein condensation in complex networks", *Phys. Rev. Lett.* **86**, 5632 (2001).
24. E. Fermi, "On the origin of the cosmic radiation", *Phys. Rev.* **75**, 1169 (1949).
25. K. G. Wilson, *Phys. Rev. B* **4**, 3174 (1971) (for simpler explanation see Chapt. 9.4 of N.D.

- Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, 1992)).
26. S. Naranan, “Bradfords’s law of bibliography of science: an interpretation”, *Nature*, **227**, 631 (1970).
 27. S. Naranan, “Power law relations in science bibliography – a self-consistent interpretation”, *Journal of Documentation*, **27**, 83 (1971).
 28. B.A. Huberman and L. Adamic, “Growth dynamics of the World-Wide Web”, *Nature*, **401**, 131 (1999).
 29. G. A Miller. “Some effects of intermittent silence”, *American Journal of Psychology*, **70**:311-314, 1957.
 30. W. Li, “Random texts exhibit Zipf’s-law-like word frequency distribution”, *IEEE Transactions on Information Theory*, **38** (1992) 1842.
 31. B. Mandelbrot, “An informational theory of the statistical structure of languages”, in *Communication Theory*, edited by W. Jackson (Academic Press, New York, 1953).
 32. B. Mandelbrot, “Information theory and Psycholinguistic: a theory of word frequencies” in *Readings in mathematical social science*, ed. P.F. Lazarsfeld and N.W. Henry (Science Research Associates, Chicago, 1966)
 33. S.S. Stevens, “Neural events and the psychophysical law”, *Science*, vol. 170, pp 1043-1055 (1970).
 34. D.M. MacKay, “Psychophysics of perceived intensity: a theoretical basis for Fechner’s and Stevens’ laws”, *Science*, vol. 139, pp1213-1216 (1963).
 35. R. Albert, H. Jeong, and A.-L. Barabasi, “Diameter of the world-wide web”, *Nature*, **401**, 130 (1999).
 36. S. Bornholdt and H. Ebel, “World Wide Web scaling exponent from Simon’s 1955 model”, *Phys. Rev. E* **64**, 035104 (2001).
 37. H. W. Watson and F. Galton, “On the probability of the extinction of families”, *Journal of the Anthropological Institute* **4**, 138 (1875). Available online at: <http://www.mugu.com/galton/essays/1870-1879/galton-1874-jaigi-family-extinction.pdf>
 38. T.E. Harris, “The theory of branching processes” (Springer, Berlin, 1963)
 39. P. Bak, *How Nature Works: the Science of Self-Organized Criticality* (Copernicus, New York, 1999)
 40. M.V. Simkin and V.P. Roychowdhury, “A mathematical theory of citing”, <http://arxiv.org/abs/physics/0504094>
 41. P. Alstrøm “Mean-field exponents for self-organized critical phenomena”, *Phys. Rev. A* **38**, 4905–4906 (1988)
 42. K. B. Lauritsen, S. Zapperi, and H. E. Stanley, “Self-organized branching processes: Avalanche models with dissipation”, *Phys. Rev. E* **54**, 2483–2488 (1996).
 43. P.A.P. Moran, *The statistical processes of evolutionary theory*, (Clarendon, Oxford, 1962)
 44. S. Karlin and J. McGregor, “The number of mutant forms maintained in a population”, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol.4, pp 415-438 (University of California Press, Berkeley, 1967)