The Zipf-Poisson-Stopped-Sum Distribution with an Application for Modeling the Degree Sequence of Social Networks

Ariel Duarte-López[a,b,*], Marta Pérez-Casany[a,c], Jordi Valero[c]

[a]*Data Management Group (DAMA - UPC)*
[b]*Department of Computer Architecture. Technical University of Catalonia, Barcelona, Spain.*
[c]*Department of Statistics and OR. Technical University of Catalonia, Barcelona, Spain.*

Abstract

Under the Zipf Distribution, the frequency of a value is a power function of its size. Thus, when plotting frequencies versus size in log-log scale of data following that distribution, one obtains a straight line. The Zipf has been assumed to be appropriate for modeling highly skewed data from many different areas. Nevertheless, for many real data sets, the linearity is observed only in the tail; thus, the Zipf is fitted only for values larger than a given threshold and, consequently, there is a loss of information. The Zipf-Poisson-stopped-sum distribution is introduced as a more flexible alternative. It is proven that in log-log scale allows for top-concavity, maintaining the linearity in the tail. Consequently, the distribution fits properly many data sets in their entire range. To prove the suitability of our model 16 network degree sequences describing the interaction between members of a given platform have been fitted. The results have been compared with the fits obtained through other bi-parametric distributions.

Keywords: Degree sequence, Discrete compound distributions, Heavy-tail distributions, Poisson stopped sum, Power law, Zipf distribution

---

*Corresponding author. Address: Edifici H, Planta 6, Despatx 6-45, Av. Diagonal 647, 08028, Barcelona, Spain. Phone: +34 93 4016572. Fax: +34934016575

Email addresses: aduarte@ac.upc.edu (Ariel Duarte-López), marta.perez@upc.edu (Marta Pérez-Casany), jordi.valero@upc.edu (Jordi Valero)

1. Introduction

The Zipf distribution ([Zipf, 1949](#)) is also known as the discrete Pareto distribution and falls within the family of power law (PL) distributions. Its support comprises all the strictly positive integers, while other PLs are defined for val-
⁵ ues greater than a given threshold. Its most popular characteristic is the linear behavior observed when the probabilities are plotted in log-log scale. Many works have shown its suitability for fitting data sets corresponding to rank data or frequencies of frequency data. For example, [Newman (2005)](#) uses the PL distribution to fit the tail of data sets related to: the number of copies of books
¹⁰ sold in the US from 1895 to 1965; the populations of US cities; or earthquake magnitudes. Additional examples from computer science are presented in the review by [Mahanti et al. (2013)](#), where the authors use a PL distribution to adjust several Internet measures such as YouTube video popularity and web access, among others. In addition, the analysis by [Gomez-Lievano et al. (2012)](#)
¹⁵ suggests that the total number of homicides in Colombia, Mexico and Brazil can be described by a PL. A more recent work by [Wang et al. (2017)](#) on user authentication in cybersecurity shows that the "vulnerable portion of user-chosen passwords" can also be adjusted by this distribution. For a PL application to model the degree sequence of a network representation of a stock market, see [Bo-](#)
²⁰ [ginski et al. (2005)](#). Other practitioners have developed mechanisms that mimic scenarios where the distribution appears (see [Barabási & Albert, 1999](#); [Erling et al., 2015](#)). In general, the PL distribution is considered appropriate when one observes data distributed over orders of magnitudes and extreme values are not rare, which is quite common in nature.

²⁵ Most of the references cited above fit the PL distribution in the tail of the data, since it is not able to adjust the top-concave pattern drawn by the first values in the support of the variable. For [McKelvey et al. (2018)](#), the PL usually emerges for values greater than a given threshold that divide "the Gaussian and

2

Paretian worlds". This philosophy is also pointed out in Newman (2005) and
30 Clauset et al. (2009). The real existence of PL distributed data was analyzed in
the recent work by Broido & Clauset (2018), where the authors conclude that
only a small amount of the analyzed data sets are appropriately fitted by a PL
distribution.

In this work, we extend the Zipf distribution by means of the concept of
35 Poisson-stopped-sum (PSS), in order to obtain a distribution able to properly
fit the data in its whole range. The idea of stopped-sum was introduced by
Feller in 1943 under the term *generalizations* and renamed by Douglas in 1971
to *stopped sum*. Given two independent random variables (r.v.), N and X, where
N is Poisson distributed, the PSS is defined as the distribution of the sum of N
40 independently and identically distributed (i.i.d.) copies of X. Some examples
of distributions that belong to the family of PSS are: negative binomial, Pólya-
Aeppli and Neyman Type A. PSSs distributions are widely applied in dissimilar
fields. For example, Podur et al. (2010) use PSS to model the annual area
burned by forest fires in the Canadian province of Ontario; and the work by
45 Low et al. (2016) compares the performance of several PSSs used to model
citation data. For an application to insurance data, see Meng & Gao (2018).
The PSSs distributions appear naturally in many data generation processes, and
the parameter estimates provide important insights about the data generation
mechanism.

50 The new discrete family of probability distributions is denoted by *Zipf-
Poisson-stopped-sum* (Zipf-PSS). The Zipf-PSS is a bi-parametric distribution
that depends on the α parameter of the Zipf and the λ parameter of the Poisson.
In contrast to the perfect straight line drawn by the Zipf distribution in log-log
scale, the Zipf-PSS can adjust the top-concavity pattern shown by many real
55 data sets. This avoids having to look for a threshold and, what more impor-
tantly, allows consideration of all the collected data without loss of information.
It is important to comment that the Zipf-PSS includes the zero-value in its sup-
port, since it appears when N takes the zero value. Thus, in the example of
social network analysis, it will allow estimating the number of isolated nodes.

3

₆₀ These types of nodes could play an important role in the evolution of the network, since a change in the behavior of isolated nodes could induce considerable changes in the network dynamics (Bak et al., 1987; McKelvey et al., 2018). Some important properties related to the model are proved, for instance: the linearity of the tail. The Zipf-PSS distribution is implemented in the R-package

₆₅ *zipfextR* (Duarte-López & Pérez-Casany, 2018), which is available at CRAN. The proposed model is shown to be suitable through the analysis of a case study where the Zipf-PSS distribution is used to fit several degree sequences related to the independent annual networks of the MathOverflow community. The fit of the Zipf-PSS is compared with those achieved by the negative binomial (NB)

₇₀ and the discrete Weibull (DW) (Nakagawa & Osaki, 1975) distributions. The models are compared by means of the Akaïke Information Criterion (AIC).

The rest of the article is organized as follows. Section 2 introduces the Zipf distribution and the concept of PSS, then defines the notion of regularly varying function required to prove the tail linearity of our model. Section 3

₇₅ defines the Zipf-PSS distribution and analyzes its main properties. Section 4 presents the analysis of several real degree sequences. Finally, Section 5 gives some recommendations for using the Zipf-PSS distribution in random network generation.

2. Preliminaries

₈₀ This section reviews the Zipf model and also introduces the concept of a PSS distribution, both of which are used in defining the Zipf-PSS distribution. Finally, we introduce the notion of regularly varying functions, since they are required to prove some properties of the Zipf-PSS model.

2.1. The Zipf Distribution

₈₅ The Zipf distribution (Zipf, 1949) is the discrete version of the Pareto distribution and belongs to the family of PL distributions. It is also known as the *Zeta* distribution and, as proved in Zörnig & Altmann (1995), it is a particular

4

case of the Lerch distribution. The Zipf is a one-parametric distribution defined on the strictly positive integer numbers, where the probabilities change inversely to a power of the values. It is also a skewed distribution with a long right tail, which may be observed when plotting the probabilities in log-log scale. Most applications of the Zipf distribution are related to scenarios where practitioners deal with rank or frequencies of frequency data. Recent references to applications are, for instance: Mahanti et al. (2013) and Wang et al. (2017), who use the Zipf distribution in computer science; Gomez-Lievano et al. (2012), for PL application in Sociology; and its use by Ferrer-i Cancho & Vitevitch (2018) in linguistics.

It is said that r.v. X follows a Zipf distribution with parameter $\alpha > 1$ if, and only if, its probability mass function (PMF) is equal to:

$$P(X = x) = \frac{x^{-\alpha}}{\zeta(\alpha)}, x = 1, 2, ..., \alpha > 1, \tag{2.1}$$

where $\zeta(\alpha) = \sum_{i=1}^{+\infty} i^{-\alpha}$ is the Riemann Zeta function. Its probability generating function (PGF) is equal to:

$$G_X(z) = \frac{Li_\alpha(z)}{\zeta(\alpha)}, \alpha > 1, |z| \leq 1, \tag{2.2}$$

where $Li_\alpha(z) = \sum_{i=1}^{+\infty} z^i \, i^{-\alpha}$ is the polylogarithm function. Note that $Li_\alpha(z) = \zeta(\alpha)$.

Since the Riemann Zeta function converges only for $\alpha > 1$, the parameter space of the Zipf distribution is $(1, +\infty)$. Another consequence is that the k-th moment of the Zipf, $k \in \mathbb{Z}^+$, is finite if, and only if, $\alpha > k + 1$ and, in this case, it is equal to:

$$E[X^k] = \frac{\zeta(\alpha - k)}{\zeta(\alpha)}. \tag{2.3}$$

In particular, the first moment only exists if $\alpha > 2$ and it is equal to:

$$E[X] = \frac{\zeta(\alpha - 1)}{\zeta(\alpha)}, \alpha > 2. \tag{2.4}$$

Since the variance depends on the second moment of the distribution, it is finite if, and only if, $\alpha > 3$ and it is equal to:

$$Var[X] = \frac{\zeta(\alpha - 2)\,\zeta(\alpha) - \zeta(\alpha - 1)^2}{\zeta(\alpha)^2}, \alpha > 3. \tag{2.5}$$

5

₁₀₀ The maximum likelihood estimate (MLE) of the α parameter of the Zipf distribution is equal to the solution of the moment method which is applied to the logarithm of the variable, that exists for any $\alpha > 1$.

2.2. Poisson-stopped-sum Distributions

PSSs (Johnson et al., 2005) appear to be the distribution of an r.v. Y in relation to a branching process, where we initially assume to have N individuals or experimental units and that each individual gives rise to X_i new individuals in a given period of time, whereby the total number of first-generation individuals is equal to:

$$Y = X_1 + X_2 + ... + X_N, \tag{2.6}$$

where N is assumed to be a Poisson r.v. and X_i, $i = 1, \ldots, n$, are i.i.d r.v.'s with a given distribution that may be either continuous or discrete. It is assumed that when $N = 0$, $Y = 0$, meaning that if there are no initial individuals, then no new ones are generated. If X_i has a distribution with parameter vector θ and parameter space $\Theta \subseteq \mathbb{R}^n$, then the distribution of Y has parameter space $\{(\lambda, \theta) | \lambda \in (0, +\infty) \text{ and } \theta \in \Theta\}$. The r.v. X is obtained when $N = 1$, which takes place with probability $\lambda e^{-\lambda}$, and the Poisson distribution is obtained when X_i has a degenerate distribution at one. The distributions of N and X are called *primary* and *secondary* distributions, respectively. Denoting by $G_x(z)$ the PGF of X, one has that the PGF of Y defined as in (2.6) is equal to:

$$G_Y(z) = e^{\lambda (G_X(z)-1)}, \lambda > 0, \tag{2.7}$$

and it is defined at least for $|z| \leq 1$. Since the PGF of a PSS is the compo-
₁₀₅ sition of the PGF of the Poisson and the PGF of the distribution of X, these families are also know as *compound distributions*. A *discrete compound Poisson distribution* (DCP) is defined as a PSS with a discrete secondary distribution (Feller, 1971; Zhang & Li, 2016). The notation for an r.v. with a DCP distribution is $Y \sim DCP(\alpha_1 \lambda, \alpha_2 \lambda, \ldots)$, with λ being the Poisson parameter and
₁₁₀ $\alpha_i = P(X_i = i)$. Its approximation of order r, Y_r, is defined by approximating

6

the Taylor's expansion of $G_X(z)$ at $z = z_0$, by the first terms up to order r, and its distribution is denoted by: $DCP(\alpha_1 \lambda, \alpha_2 \lambda, \ldots, \alpha_r \lambda)$.

One of the most studied approximations is the second-order approximation, which is the Hermite distribution. Several works on the third and fourth ap-

115 proximations can be found in the literature under the name *sttutering Poisson distributions* (Patel, 1976) or *3rd and 4th order Hermite distribution* (Puig & Valero, 2007).

2.3. Regularly Varying Functions

A *regularly varying function* can roughly be defined as a function that per-

120 forms asymptotically as a power function. The formal definition below can be found in the book by Gulisashvili (2012, p. 201).

Definition 1. *Let f be a positive measurable function on $[a, +\infty)$ with $a > 0$, and let α be a real number. Then, the function f is a regularly varying function of index α at infinity, if for any $\lambda > 0$,*

$$\lim_{x \to \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\alpha. \tag{2.8}$$

The class of all regularly varying functions with index α is denoted by R_α.

Definition 2. *A slowly varying function at infinity is a function L which satisfies that for all $\lambda > 0$,*

$$\lim_{x \to \infty} \frac{L(\lambda x)}{L(x)} = 1. \tag{2.9}$$

The class of slowly varying at infinity functions is denoted by R_0. Observe that $f \in R_\alpha$ if, and only if, $f(x) = x^\alpha L(x)$, where the function L is slowly varying

125 at infinity.

The work by Jessen & Mikosch (2006) outlines several properties of the regularly varying functions and their relationship with random sums of non-negative r.v.'s. According to the authors, the tail of Y, defined as (2.6), is highly influenced by the tails of the primary and the secondary distributions. Understanding that $f(x) \sim g(x)$, as $x \to +\infty$ is equivalent to saying that $f(x)/g(x) \xrightarrow[x \to +\infty]{} 1$ if $g(x) \neq 0$, and it is equivalent to $f(x) = o(1)$ if $g(x) = 0$;

7

a *regularly varying random variable with index* $\alpha \geq 0$ is defined as r.v. X, such that

$$P(X > x) \sim p x^{-\alpha} L(x) \quad \text{and} \quad P(X \leq -x) \sim q x^{-\alpha} L(x),$$

where $p + q = 1$, and L is a slowly varying function.

Appendix A.3 establishes the relationship between the tails of r.v. X with a regularly varying function and r.v Y defined as (2.6).

3. The Zipf-Poisson-Stopped-Sum Distribution

₁₃₀   In this section, we define the Zipf-PSS distribution and state its main properties. The proofs of the results presented in this section appear in Appendix A.

3.1. Definition

The Zipf-PSS model is obtained by assuming that the distribution of the r.v.'s X_i in (2.6) is the Zipf(α) distribution. By substituting (2.2) in (2.7), the PGF of r.v. Y with a Zipf-PSS(α, λ) distribution is set to be equal to:

$$G_Y(z) = e^{\lambda \left(\frac{Li_\alpha(z)}{\zeta(\alpha)} - 1 \right)}, \ \lambda > 0, \ \alpha > 1, \ |z| \leq 1. \tag{3.1}$$

This family is a bi-parametric model with parameter space equal to $\{(\alpha, \lambda) \in (1, +\infty) \times (0, +\infty)\}$, and it pertains to the class of DCP distributions with

₁₃₅   parameters $(\frac{\lambda}{\zeta(\alpha)}, \frac{\lambda}{2^\alpha \zeta(\alpha)}, \frac{\lambda}{3^\alpha \zeta(\alpha)}, ...)$.

The probabilities of a Zipf-PSS(α, λ) distribution may be computed using the generalization of the Panjer recursion that appears in Panjer (1981, p. 24-25) and in Sundt & Jewell (1981, p. 38), which says that:

$$P(Y = 0) = \begin{cases} \frac{1 - a\, P(X=0)}{1-a}, & \text{if } a \neq 0 \\ \\ e^{-b\,[1 - P(X=0)]}, & \text{if } a = 0, \end{cases} \tag{3.2}$$

and, for $x = 1, 2, ...$

$$P(Y = x) = \frac{1}{1 - a\, P(Y=0)} \sum_{s=1}^{x} \left(a + \frac{b\,s}{x} \right) P(X = s)\, P(Y = x - s), \tag{3.3}$$

8

for given constants a and b. The work of Panjer (1981) also shows that if N has a Poisson distribution, then $a = 0$ and $b = \lambda$.

Thus, if $X \sim Zipf(\alpha)$ and $Y \sim$ Zipf-PSS(α, λ), from (3.2) one has that

$$P(Y = 0) = e^{-\lambda}, \tag{3.4}$$

and from (3.3) that

$$P(Y = x) = \frac{\lambda}{\zeta(\alpha)\,x} \sum_{s=1}^{x} s^{1-\alpha}\, P(Y = x - s),\, x \geq 1. \tag{3.5}$$

Fig. 1 shows the probabilities of the Zipf-PSS distribution for different values of α and λ. In (A) and (C), the plots are in normal scale while (B) and (D)
140 are shown in log-log scale. The probabilities obtained for $\alpha = 2.3$ and different values of λ can be observed in (A) and (B), while (C) and (D) contain the probabilities for $\lambda = 5$ and different values of α. We observe that the highest probabilities are obtained at the initial values when α and λ are small. Additionally, looking at (B), it is possible to compare the behavior of the Zipf-PSS
145 probabilities with those achieved by a Zipf distribution with the same α parameter. Note that even though it looks like a straight line for smaller λ values, the probabilities obtained are different from those of the Zipf. In comparison, the larger the value of λ becomes, more curvature is shown by the PMF. This brings us to consider the λ parameter as a measure of departure from the Zipf
150 distribution.

3.2. Moments and Index of Dispersion

It appears that in Satterthwaite (1942) and Johnson et al. (2005) the moments of any PSS are functions of the moments of the underlying secondary distribution, which in our case is the Zipf distribution. These results give rise
155 to the following propositions.

Proposition 1. *The k-th moment of the Zipf-PSS(α, λ) distribution is finite if, and only if, $\alpha > k + 1$.*

In what follows, we obtain the exact values for the mean and the variance.

9
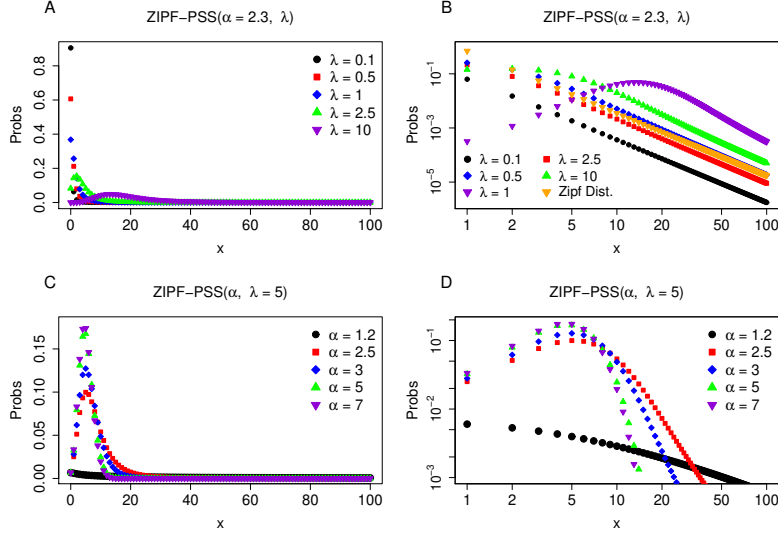
Fig 1. PMF of the Zipf-PSS(α, λ) distribution. For $\alpha = 2.3$ and $\lambda = 0.1, 0.5, 1, 2.5$ and 10, (A) in normal scale and (B) in log-log scale. For $\lambda = 5$ and $\alpha = 1.5, 2.5, 3, 5, 7$, (C) in normal scale and (D) in log-log scale. In normal scale $x = 0, \ldots, 100$, and $x = 1, \ldots, 100$ in log-log scale.

Proposition 2. *The mean and the variance of a Zipf-PSS(α, λ) distribution are respectively equal to:*

$$E[Y] = \lambda \, E[X] = \lambda \, \frac{\zeta(\alpha - 1)}{\zeta(\alpha)}, \alpha > 2, \tag{3.6}$$

and

$$Var[Y] = \lambda \, \frac{\zeta(\alpha - 2)}{\zeta(\alpha)}, \alpha > 3, \tag{3.7}$$

Taking into account (3.6), (3.7) and that $\zeta(\alpha)$ is a decreasing function of α, one has that if $\alpha > 2$, then

$$ID = \frac{Var[Y]}{E[Y]} = \frac{\zeta(\alpha - 2)}{\zeta(\alpha - 1)} > 1.$$

Consequently, the Zipf-PSS is over-dispersed compared to a Poisson distribution with the same mean, which is consistent with the relationship between the ID of Y and the ID of X and N, pointed out in Johnson et al. (2005, p. 386).

Fig. 2 shows the behavior of the mean of the Zipf-PSS distribution as a function of α, for $\lambda = 0.5, 1, 2.5$ and 3 (A); and, as a function of λ, for $\alpha =$

10

2.5, 3, 5 and 10 (B). Observe that the expected value of a Zipf-PSS is a decreasing

165 function of α for any value of λ. This makes sense, since $E[Y] = \lambda E[X]$ and the mean of a Zipf distribution is a decreasing function of α because increasing α leads to the probability concentrating in the first values. In (B), one can observe that the mean is an increasing function of λ for any α. This is a consequence of the fact that increasing λ increases the number of terms in (2.6) and, consequently, the value of Y.
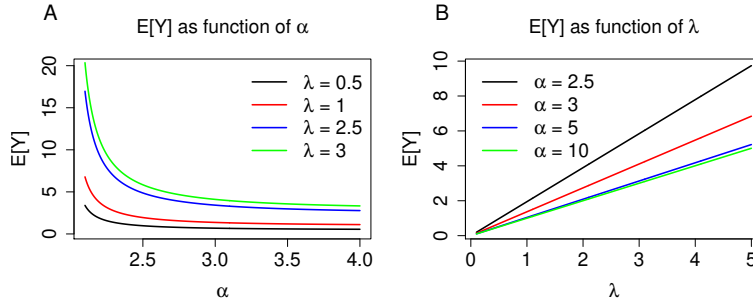

Fig 2. Expected values of a Zipf-PSS(α, λ) distribution as a function of α, for $\lambda = 0.5, 1, 2.5$ and 3 (A); and, as a function of λ, for $\alpha = 2.5, 3, 5$ and 10 (B).

170

A similar plot is presented in Fig. 3 in regard to the variance of the Zipf-PSS distribution as a function of α for several values of λ (A) and as a function of λ for several values of α (B). The variance of the distribution behaves quite similar to the mean. The variance is clearly a decreasing function of α, and it

175 decreases faster as the λ value becomes smaller. From Proposition 2, the mean and the variance are linear functions of λ with a slope that decreases when α increases, which is observed in Fig. 2 (B) and Fig. 3 (B), respectively.

In what follows, we provide a condition under which the variance of the Zipf-PSS is larger than the variance of a Zipf distribution with the same α parameter.

Proposition 3. *Let* $Y \sim Zipf\text{-}PSS(\alpha, \lambda)$ *and* $X \sim Zipf(\alpha)$. *If* $\alpha > 3$, *then one has that* $Var[Y] \geq Var[X]$ *if, and only if,*

$$\lambda \geq 1 - \frac{\zeta^2(\alpha - 1)}{\zeta(\alpha - 2)\,\zeta(\alpha)}.$$

11
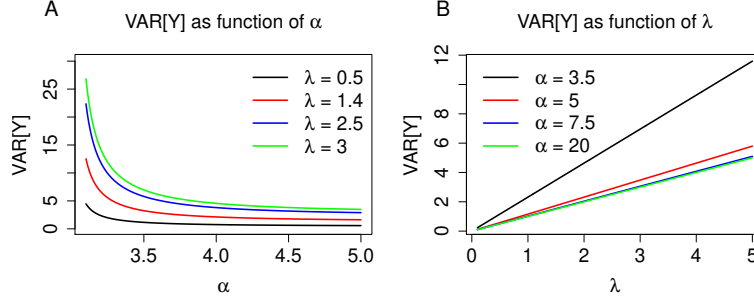
Fig 3. Variance values of a Zipf-PSS(α, λ) distribution as a function of α, for $\lambda = 0.5, 1.4, 2.5$ and 3 (A); and as a function of λ for $\alpha = 3.5, 5, 7.5$ and 20 (B).

3.3. Other Important Properties

This subsection first shows the relationship between the parameters of the distribution and its modality. Later, we compare the ratio of two consecutive probabilities of a Zipf-PSS(α, λ) with the one obtained from a Zipf distribution with the same α parameter.

Proposition 4. *If $Y \sim$ Zipf-PSS(α, λ) and denoting by $\lambda_0 = 2\zeta(\alpha)(1 - 2^{-\alpha})$, one has that:*

i) *if $\lambda \in (0, \zeta(\alpha)]$, then the distribution of Y is unimodal with a pseudo-mode at zero, and it is also log-concave;*

ii) *if $\lambda \in (\zeta(\alpha), \lambda_0]$, then the distribution of Y has a mode at one;*

iii) *if $\lambda \in (\lambda_0, +\infty)$, then the distribution of Y has a mode equal to or larger than two.*

Given that, $\zeta(\alpha) \to +\infty$ when $\alpha \to 1$, and that it tends to 1 when $\alpha \to +\infty$, one has that when $\alpha \to 1$, then $\lambda_0 \to +\infty$ and thus the parameter λ must be larger in order to have a mode larger than one. Also, when $\alpha \to +\infty, \lambda_0 \to 2$, and thus the distribution has a mode larger than one for any $\lambda \in (2, +\infty)$.

The next proposition establishes the relationship between the ratio of two consecutive probabilities of the Zipf-PSS(α, λ) and the Zipf(α) distributions.

12

Proposition 5. *The ratio of two consecutive probabilities of r.v. Y with a Zipf-PSS(α, λ) distribution is related to the same ratio of r.v. X with a Zipf(α) distribution by means of:*

$$\frac{P(Y = x + 1)}{P(Y = x)} = \frac{P(X = x + 1)}{P(X = x)} \left(\frac{x}{x+1} \right)^{1-\alpha} h(x; \alpha, \lambda),$$

where $h(x; \alpha, \lambda)$ is a ratio of two linear combinations of the probabilities $P(Y = i)$ for $i = 0, \ldots, x$.

Fig. 4 shows the behavior of this ratio for $\alpha = 2.3$ and two different values of the λ parameter. For comparison to the Zipf distribution, the ratio of the Zipf(α) probabilities is also included. When the λ value is close to zero, the ratio of the two consecutive probabilities of the Zipf-PSS(α, λ) quickly converges to the ratio of the same two consecutive Zipf probabilities. Otherwise, if the λ parameter is large, the ratio of the probabilities also converges to that of the Zipf, but the convergence takes longer. Moreover, both ratios tend to one when x tends to infinity. This implies that $h(x; \alpha, \lambda)$ also tends to one when x increases. In addition, this is another manner of observing the flexibility of the Zipf-PSS distribution, especially in the first integer values. Observe that increasing the value of x (i.e., considering those values in the tail of the distribution) leads to similar behavior as in the Zipf distribution, which is proved by Theorem 1 in the next subsection.
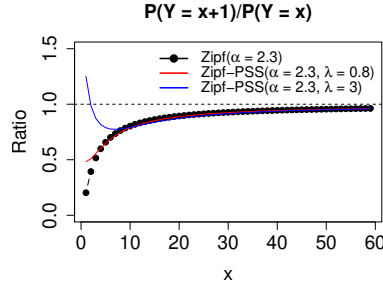
Fig 4. Ratio of two consecutive Zipf-PSS probabilities for $\alpha = 2.3$ and $\lambda = 0.8$ and 3. Jointly with the same ratio of probabilities of the Zipf for $\alpha = 2.3$.

13

3.4. Linear Tail Behavior

Section 2.3 introduces the concept of a regularly varying function, which is ²¹⁵ an essential tool for analyzing the tail behavior of the Zipf-PSS distribution. Particularly, we use Lemma 1, in Appendix A.3, to understand how the tails of the Poisson and the Zipf distributions influence the tail of the Zipf-PSS. To that aim, Lemmas 2, 3 and 4 of the same appendix are required.

The following theorem states that the tail of the Zipf-PSS is linear.

²²⁰ **Theorem 1.** *The tail of r.v. Y with a Zipf-PSS(α, λ) distribution is asymptotically equivalent to λ times the tail of r.v. X with a Zipf(α) distribution.*

Fig. 5 illustrates the result stated in Theorem 1. Observe that the larger the λ value becomes, the larger the x value must be to obtain the equivalence of the tails.
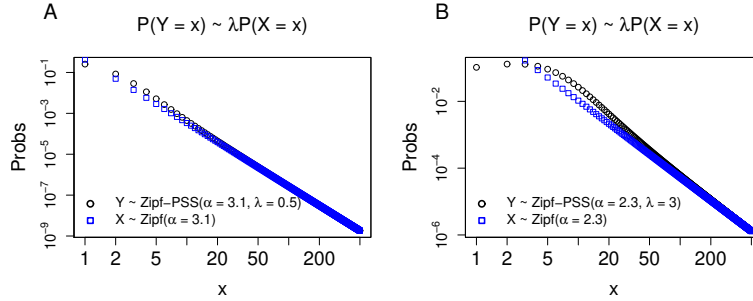
Fig 5. Probabilities of the Zipf and the Zipf-PSS distributions with the same α parameter for $\alpha = 3.1$ and $\lambda = 0.5$ (A); and for $\alpha = 2.3$ and $\lambda = 3$ (B).

²²⁵ 3.5. A Limit Distribution and Approximations

This subsection establishes the conditions under which the Zipf-PSS converges in distribution to the Poisson. It also studies their DCPs of order r approximations for some r values.

Proposition 6. *Let $Y_n \sim$ Zipf-PSS(α_n, λ), for $n \geq 1$, and Y_i is independent of Y_j. Also let $N \sim Po(\lambda)$, Y_i be independent of N $\forall i$. Then, if $\alpha_n \to +\infty$ when*

14

$n \to +\infty$,

$$Y_n \xrightarrow[n \to +\infty]{D} N.$$

In Subsection 3.1, we have introduced the approximation of order r of a DCP distribution. In what follows, we compute several approximations for the particular case of the Zipf-PSS.

Proposition 7. *The Zipf-PSS(α, λ) distribution has as:*

    a) first order approximation, a Poisson distribution, with parameter $\lambda/\zeta(\alpha)$;

    b) second order approximation, a Hermite distribution, with parameters $a_1 = \lambda/\zeta(\alpha)$ and $a_2 = a_1/2^\alpha$;

    c) third order approximation, a 3rd order Hermite distribution, with parameters $a_1 = \lambda/\zeta(\alpha)$, $a_2 = a_1/2^\alpha$ and $a_3 = a_1/3^\alpha$;

    d) fourth order approximation, a 4th order Hermite distribution, with parameters $a_1 = \lambda/\zeta(\alpha)$, $a_2 = a_1/2^\alpha$, $a_3 = a_1/3^\alpha$ and $a_4 = a_1/4^\alpha$.

Fig. 6 shows the behavior of the first four approximations. Observe that for small α values, a large λ value is required to properly approximate the distributions by a lower order approximation. In contrast, when α increases, the distribution is properly approximated by the first order approximation independently of the value of λ. Thus, the first order approximation is acceptable if α is large enough. For small values of α, it is better to consider the second or higher order approximations.

4. Application Examples

This section demonstrates the suitability of the Zipf-PSS distribution for fitting real data. We analyze the MathOverflow data set, which appears in a recent paper by Paranjape et al. (2017). This data set is available through the SNAP repository (Leskovec & Krevl, 2014) and contains interactions among the members of the MathOverflow platform, which was developed in 2009 by a

15
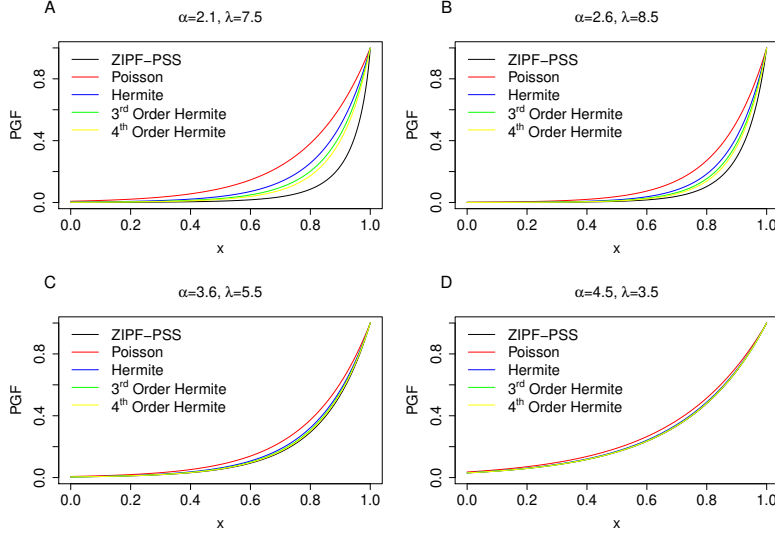
Fig 6. Behavior of the PGF of the higher order approximations of the Zipf-PSS for
$\alpha = 2.1$, $\lambda = 7.5$ (A), $\alpha = 2.6$, $\lambda = 8.5$ (B), $\alpha = 3.6$, $\lambda = 5.5$ (C) and $\alpha = 4.5$, $\lambda = 3.5$ (D).

group of PhD students and post-docs from Berkeley University. It is an on-line
question-and-answer (Q&A) forum for research mathematicians on the Stack
₂₅₅ Overflow network (Keller, 2010).

We have compared the performance of the Zipf-PSS to that of the NB (John-
son et al., 2005, p. 209) and the DW (Nakagawa & Osaki, 1975) distributions.
In addition, we also compared it with the PL fitted using the methodology de-
scribed by Clauset et al. (2009). The NB was used because it is also a PSS
₂₆₀ distribution, with the log-series(q) as primary distribution, and a Poisson with
$\lambda = -k \log(1 - q)$ as the secondary distribution. It is also the classical over-
dispersed two-parameter Poisson alternative. The DW was used since it is
proposed to model the failure time when life is measured by means of cycles,
blows or revolutions. Given that it makes sense to assume that a person is alive
₂₆₅ while it is active and creating new connections, and dead otherwise; the DW
seems to be a good alternative to use. The work by Koziol et al. (2013) uses
the DW in similar scenarios. To compare the models, we chose the AIC as the
goodness-of-fit criteria.

16

In the repository, the synergy among users is divided into three categories: answers to questions, comments to questions and comments to answers. For testing our model, we used the category, answers to questions, where an edge is created between users u and v when user u answers a question by user v. This gives rise to a directed network, since the origin and final nodes play different roles. The network comprises a total of 21688 nodes and 107581 edges collected from September 9, 2009 to March 6, 2016. As a consequence of including the component time during edge formation, it is possible to have multi-edges between different pairs of nodes. Moreover, due to the fact that a user can answer their own question, loops are also possible in the network.

Each annual network was extracted from the global data set. Because these are directed networks, the degree sequence was split into three sequences: the in-degree, the out-degree and the total-degree. The in-degree sequence corresponds to situations in which the node user's question receives a reply. The out-degree corresponds to situations where the node user answers a question. Finally, the total-degree corresponds to the sum of the in- and out-degrees, and it explains the total activity of a user. In this work, we analyze, for each one of the independent annual networks, the in- and out-degree sequences.

Table 1 summarizes the main statistics calculated for each network. It contains the number of nodes ($\mathbf{N}$), edges ($\mathbf{E}$), loops ($\mathbf{L}$) and multi-edges ($\mathbf{M}$), and – after splitting the degree sequence – the number of isolated nodes ($\mathbf{I}$), the mean ($\mathbf{Mean}$), the variance ($\mathbf{Var}$) and the skewness ($\mathbf{Skew}$) of the in- and out-degree sequences. Observe that a greater number of zeros occurs in the out-degree sequences, with the exception of the first year, which is when the community was created and only contains information covering approximately 4 months. One of the reasons for this may be that many people do not answer any questions while a few people answer many questions, which turns out to be an in-degree sequence with fewer nodes that have zero degrees. Observe that the mean values of the in- and out-degree sequences are equal for all years. This is because each edge involves a connection from the in- and the out-degree sequences, respectively. Moreover, this is one of the conditions of a graphical sequence, as

17

Table 1: For each of the considered degree sequences: *Year* of the sequence; number of nodes (*N*); number of edges (*E*); number of loops (*L*); number of multiple-edges (*M*); *type* of the sequence; number of isolated nodes (*I*) and its percentage; (*Mean*), variance (*Var*) and skewness (*Skew*) values.

Year	N	E	L	M	Type	I (%)	Mean	Var	Skew
2009	1278	7115	1363	188	in	526 (41.16)	5.57	297.69	9.39
					out	333 (26.06)	5.57	227.21	6.81
2010	4648	24799	3120	547	in	1468 (31.58)	5.34	224.15	8.69
					out	1839 (39.57)	5.34	335.19	8.10
2011	5358	18468	1712	529	in	1566 (29.23)	3.45	70.08	9.05
					out	2502 (46.7)	3.45	164.94	11.65
2012	5687	15954	1485	484	in	1530 (26.9)	2.81	46.55	13.98
					out	2910 (51.17)	2.81	141.50	13.00
2013	6101	14779	1245	500	in	1455 (23.85)	2.42	27.24	10.64
					out	3459 (56.7)	2.42	92.14	9.98
2014	5556	12574	1074	537	in	1376 (24.77)	2.26	25.45	15.11
					out	3121 (56.17)	2.26	75.74	10.17
2015	5409	11833	1166	566	in	1336 (24.7)	2.19	33.44	18.93
					out	2986 (55.2)	2.19	76.09	13.35
2016	1654	2059	83	92	in	548 (33.13)	1.24	3.27	5.17
					out	860 (52)	1.24	9.06	6.71

stated in the *Fulkerson-Ryser's theorem* (Kim et al., 2012). The mean values of the first (incomplete) and second year are clearly larger than the other years (the last year is also incomplete). This provides evidence indicating that the community was very active when it was created and the activity stabilized as the year passed. The variance is large for the out-degree sequences, meaning that some users are quite active in answering while others answer only few or no questions. The higher skewness value associated with each sequence leads us to assume that all the degree sequences will show a pronounced right tail.

Fitting these data with the Zipf-PSS and the NB distributions allows insights into the network's data generation mechanism as a direct consequence of the

₃₁₀ fact that both are PSS. Thus, in both cases, $\widehat{E[N]}$ is interpreted as the average number of times that a user is active on the network. Each time that a user is active, she or he receives (in-degree) or answers (out-degree) an expected $\widehat{E[X_i]}$ number of questions. In contrast, the DW has no straightforward parameter interpretation.

₃₁₅   The total number of degree sequences fitted is 16, which is obtained by multiplying 8 (years) by 2 (type of sequences). Parameter estimates for the Zipf-PSS are computed by means of the MLE using the R-package *zipfextR* (Duarte-López & Pérez-Casany, 2018). Since the MLE of the Zipf-PSS model needs to be found numerically, initial values for each of the parameters are ₃₂₀ required. These values are computed as follows:

1) The initial value for α is set to be equal to the α estimate obtained for the Zipf distribution proposed by Güney et al. (2017).

2) The initial value for λ is set to be equal to 1, because when the mean is $\lambda = 1$, we have just one term in the summation (2.6).

₃₂₅ Tables 2 and 3, in Appendix B, summarize the results obtained for the periods 2009-2012 and 2013-2016, respectively. The main conclusions are:

a) In 93.75% of the sequences (15 out of 16), the Zipf-PSS gives the best fit. In the remaining case, the DW is the best model.

b) The behavior of the NB and the DW is quite similar, independently of the ₃₃₀ type of sequence, but the DW is always the best of the two.

c) The values of $\hat{\lambda}$ for the out-degree sequences are always smaller than one, while they are slightly larger than one for the in-degree sequences. It is expected that, over the course of one year, less than one person is activated to answer the question of a given node, while the node is active more than ₃₃₅ once over the same period to answer someone else's question(s).

Fig. 7 shows the behavior of the parameter estimates over time. Since the data set contains only partial information related to the years 2009 and 2016, these

19
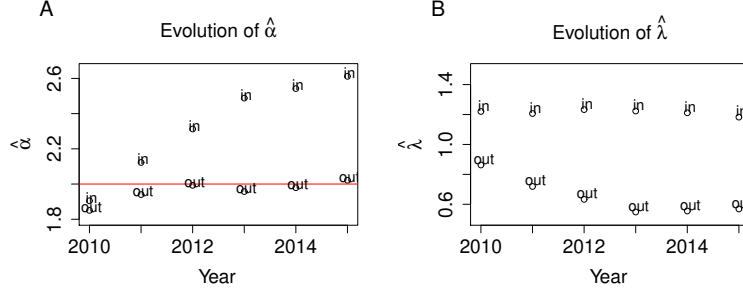
Fig 7. Evolution of the $\hat{\alpha}$ (A) and the $\hat{\lambda}$ (B), MLE of the parameters of the Zipf-PSS distribution over the period 2010-2015 for the in-degree (in) and the out-degree (out) sequences.

years have been excluded from the figure. In Fig. 7 (A), we can observe the evolution of the system in terms of the $\hat{\alpha}$, which increases significantly in the

in-degree sequence and slightly in that of the out-degree. The work of McKelvey et al. (2018), states that exponents of the PL that are lower than 2 appear in nascent systems, which agrees with our in-degree results, since a value smaller than two occurs only for the year 2010. The differences between the in- and out-degree sequence of the Q&A network are shown. While the $\hat{\alpha}$ of the in-degree

sequences quickly departs from the threshold established in the literature, the $\hat{\alpha}$ for the out-degree stabilized at around 2.0, and thus time does not change the number of questions answered by a member of the community each time that she or he is active. With respect to the $\hat{\lambda}$ behavior (see, Fig. 7 (B)), it seems to remain constant for the in-degree from nearly the beginning while it takes more

time to stabilize for the out-degree. From this, it seems that, in all the years analyzed, the users are active in the networks approximately the same number of times.

Fig. 8 shows the fits obtained by the models studied using the network's degree sequences for the year 2015. The plots also incorporate the fit obtained

from the PL distribution by using the methodology proposed by Clauset et al. (2009), which initially determines a value x_{min} (cut-off), from which the PL is fitted. Establishing a cut-off point equal to 6 in the in-degree sequence generates

20
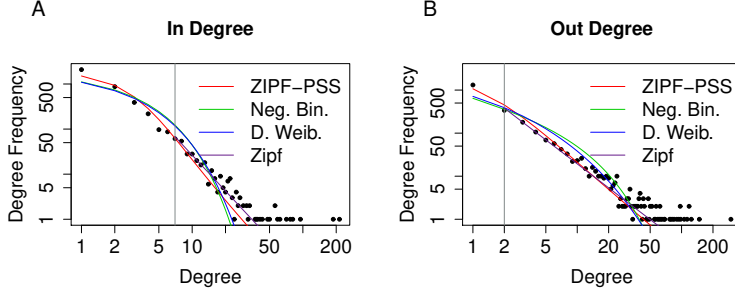
Fig 8. In-degree (A) and out-degree (B) of the 2015 network, jointly with the fit obtained by each of the considered models. It also includes the estimate of the PL and the cut-off point achieved using the methodology proposed by Clauset et al. (2009).

a loss in observations of 92.3 %. In the case of the out-degree sequence, the cut-off is taken to be equal to 2, which implies that those users with less than two replies are not considered in the analysis, which accounts for 79.4% of the nodes in the sequence for the year 2015. The Zipf-PSS, the NB and the DW, avoid lost information by covering the whole range of the degree sequence. Note that the Zipf-PSS is the only bi-parametric one that can maintain linearity in the tail. Comparing the fits of the considered models shows that the NB and DW have a more pronounced curvature than the Zipf-PSS; which, together with the earlier decay of their probabilities, demonstrates a clear deviation from the pattern shown by the data.

Since the Zipf-PSS and the NB are PSS, their estimated parameters capture some insights into the behavior of the community members. For instance, for 2015 and the in-degree sequence, the parameter estimates of the Zipf-PSS and the NB are equal to: $\hat{\alpha} = 2.6118, \hat{\lambda} = 1.1825$ and $\hat{k} = 0.8251, \hat{q} = 0.7262$ (see, Table 3) respectively. From this, one has that the expected number of active users answering someone's questions is approximately equal to one ($E[N] = \hat{\lambda} = 1.1828$ for the Zipf-PSS and $E[N] = -\hat{k} \, log(1 - \hat{q}) = 1.0689$ for the NB). Moreover, given that according to (2.4) $E[X] = 1.732$ for the Zipf-PSS, and $E[X] = -\hat{q}/(log(1 - \hat{q})(1 - \hat{q})) = 2.05$ for the NB, the expected number of answers provided by the active user is around 2.

21

On the other hand, for 2015 and the out-degree sequence, the parameter estimates of the Zipf-PSS and the NB are equal to: $\hat{\alpha} = 2.0214, \hat{\lambda} = 0.5688$

and $\hat{k} = 0.223, \hat{q} = 0.9075$ respectively. Thus, the expected number of times that a particular user is active for answering questions is approximately 0.5 ($E[N] = 0.5686$ for the Zipf-PSS and $E[N] = 0.5309$ for the NB). Once the user becomes active, the expected number of answered questions is $E[X] = 28.844$ for the Zipf-PSS and $E[X] = 4.12$ for the NB.

Note that the models agree in their estimates for the in-degree activity, but they provide quite different results for the number of answers in the out-degree. However, even though the Zipf-PSS is the model that gives the best fit to the data, the estimated number of a user's answers (out-degree) may be highly influenced by the large variability observed in the sample variance of the out-degree, which is larger than that obtained for the in-degree sequence. In addition, that result might also be a consequence of using the Riemann zeta function with a parameter close to the lower boundary of its parameter space; according to (2.4), its numerator is $\zeta(\hat{\alpha} - 1) = \zeta(1.0214)$, which may lead to a less robust result.

5. The Zipf-PSS in Synthetic Data Generation

Many researchers these days must face the problem of generating synthetic networks that mimic the characteristics of the real ones. This is because, for either economic or privacy reasons, the desired number of real networks is not available for a given particular type.

Given that one of the main characteristics of the networks is the degree sequence, we propose several steps for using a synthetic network generator to produce networks with a degree sequence similar to a given real one. The algorithms used for this purpose are conditioned to different inputs. However, a degree sequence close to the one observed in the real network is essential for several of them, such as, for instance, in Seshadhri et al. (2012) and Erling et al. (2015).

Assuming that the degree sequence observed in the real network follows

22

the Zipf-PSS distribution, the next steps will generate networks with a degree sequence with that has the same distribution as the real one. If n networks must be generated,

410   1. Estimate the parameters of the Zipf-PSS by means of maximum likelihood, denoted by $(\hat{\alpha}, \hat{\lambda})$, from the degree sequence of the real network.

   2. Generate n random samples from a Zipf-PSS$(\hat{\alpha}, \hat{\lambda})$ using the R-package *zipfextR*.

   3. For each sample, check if it is a *graphic* degree sequence (Sierksma & Hoogeveen, 1991). If it is not, suppress this sample and generate a new one.

415   4. For each sample, generate the random network using the chosen random network generator.

An application of this procedure can be found in the work of Duarte-López et al. (2015), where the authors use the Marshal-Olkin Extended Zipf (Pérez-Casany & Casellas, 2013) distribution for generating random copies of real networks.

420   **Acknowledgments**

425   **Appendix A. Lemmas and Proofs**

Appendix A.1. Proofs of Subsection 3.2

Proof of Proposition 1: Let $M(t), t \in \mathbb{R}$ be the moment generating function (MGF) of the Zipf-PSS distribution. According to equation (3.1), it is equal to:

$$M(t) = G_Y(e^t) = e^{\lambda \left(\frac{Li_\alpha(e^t)}{\zeta(\alpha)} - 1 \right)}. \tag{A.1}$$

In general $E[Y^k] = M^{(k)}(t)|_{t=0}$, which depends only on the derivatives of the PGF of the secondary distribution. Satterthwaite (1942) establishes the relationship between the moments of any PSS distribution and those of the under-

430   lying secondary distribution. Since the existence of Zipf-PSS moments depends

23

on the existence of Zipf moments, the moment of order k of the Zipf-PSS exists if, and only if, $\alpha > 1$. $\qquad\square$

Proof of Proposition 2: The expectation of the Zipf-PSS, as well as its variance expression, can be derived either by means of the derivatives of the moment function or, what is more straightforward, from the *Law of Total Expectation* and the *Law of Total Variance*, which state that:

$$E[Y] = E[N]\,E[X], \text{ and}$$

$$Var[Y] = E[N]\,Var[X] + E^2[X]\,Var[N]. \qquad (A.2)$$

Let us take into account that if N is Poisson distributed, $E[N] = Var[N] = \lambda$, and that $E[X]$ and $Var[X]$ appear in (2.4) and (2.5), respectively. Then, in a simple substitution, we have:

$$Var[Y] = \lambda\,\frac{\zeta(\alpha-2)\,\zeta(\alpha) - \zeta(\alpha-1)^2}{\zeta(\alpha)^2} + \left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}\right)^2 \lambda = \lambda\,\frac{\zeta(\alpha-2)}{\zeta(\alpha)}. \quad \square$$

Proof of Proposition 3: From (A.2), and given that $E[N] = Var[N] = \lambda$, one has that:

$$Var[Y] = \lambda\,(Var[X] + E^2[X]).$$

Thus,

$$Var[Y] \geq Var[X] \Leftrightarrow \lambda\,(Var[X] + E^2[X]) \geq Var[X]$$

$$\Leftrightarrow \lambda \geq \frac{Var[X]}{Var[X] + E^2[X]} = \frac{1}{1 + \frac{E^2[X]}{Var[X]}}.$$

Taking into account (2.4) and (2.5),

$$1 + \frac{E^2[X]}{Var[X]} = \frac{\zeta(\alpha-2)\,\zeta(\alpha)}{\zeta(\alpha-2)\,\zeta(\alpha) - \zeta^2(\alpha-1)},$$

and thus,

$$Var[Y] \geq Var[X] \Leftrightarrow \lambda \geq \frac{\zeta(\alpha-2)\,\zeta(\alpha) - \zeta^2(\alpha-1)}{\zeta(\alpha-2)\,\zeta(\alpha)} = 1 - \frac{\zeta^2(\alpha-1)}{\zeta(\alpha-2)\,\zeta(\alpha)}.$$

Given that the Riemann Zeta function is always positive, then: if λ is larger than one (which is usually the case when fitting real data), the condition is always satisfied and the Zipf-PSS has a larger variance than the corresponding Zipf distribution. $\qquad\square$

24

Appendix A.2. Proof of Subsection 3.3

Proof of Proposition 4: i) It is necessary to see that $P(Y = 0) \geq P(Y = 1)$

$$P(Y = 0) \geq P(Y = 1) \Leftrightarrow e^{-\lambda} \geq \frac{\lambda e^{-\lambda}}{\zeta(\alpha)} \Leftrightarrow \zeta(\alpha) \geq \lambda.$$

Hence, point i) holds when $\lambda \in (0, \zeta(\alpha)]$.

ii) Applying (3.5) for $x = 1$ and $x = 2$, one has that:

$$P(Y = 2) \geq P(Y = 1) \Leftrightarrow \frac{\lambda e^{-\lambda}}{2\,\zeta(\alpha)} \left[\frac{\lambda}{\zeta(\alpha)} + 2^{1-\alpha} \right] \geq \frac{\lambda e^{-\lambda}}{\zeta(\alpha)} \tag{A.3}$$

$$\Leftrightarrow \frac{1}{2}\left(\frac{\lambda}{\zeta(\alpha)} + 2^{1-\alpha} \right) \geq 1 \Leftrightarrow \lambda \geq 2\,\zeta(\alpha)\,(1 - 2^{-\alpha}). \tag{A.4}$$

To prove point ii), it is only necessary to see that for $\lambda \in (\zeta(\alpha), 2\,\zeta(\alpha)\,(1 - 2^{-\alpha}))$, $P(Y = 1) \geq P(Y = 0)$ and $P(Y = 1) \leq P(Y = 2)$. Taking into account that:

$$P(Y = 0) \leq P(Y = 1) \Leftrightarrow e^{-\lambda} \leq \frac{\lambda e^{-\lambda}}{\zeta(\alpha)} \Leftrightarrow \zeta(\alpha) \geq \lambda, \tag{A.5}$$

and based on (A.3), one has that, if $\lambda \in (\zeta(\alpha), 2\,\zeta(\alpha)\,(1 - 2^{-\alpha}))$, there is a mode at one.

iii) This is a straightforward consequence of points i) and ii). $\square$

Proof of Proposition 5: From (3.5) one has that for $x = 1, 2, \ldots$

$$P(Y = x + 1) = \frac{\lambda}{\zeta(\alpha)\,(x + 1)} \sum_{s=1}^{x+1} s^{1-\alpha}\, P(Y = x + 1 - s).$$

Dividing this expression by $P(Y = x)$, as in (3.5), one has that:

$$\frac{P(Y = x + 1)}{P(Y = x)} = \frac{\frac{1}{x+1}}{\frac{1}{x}} \frac{\sum_{s=1}^{x+1} s^{1-\alpha}\, P(Y = x + 1 - s)}{\sum_{s=1}^{x} s^{1-\alpha}\, P(Y = x - s)}. \tag{A.6}$$

Thus, denoting by $h(x; \alpha, \lambda)$ the second ratio on the right hand side of the equation one has that:

$$\frac{P(Y = x + 1)}{P(Y = x)} = \frac{x}{x + 1}\, h(x; \alpha, \lambda). \tag{A.7}$$

Taking into account that the ratio of two consecutive probabilities of the Zipf(α) distribution is equal to $(\frac{x}{x+1})^{\alpha}$, (A.7) is equivalent to:

$$\frac{P(Y = x + 1)}{P(Y = x)} = \frac{P(X = x + 1)}{P(X = x)} \left(\frac{x}{x + 1} \right)^{1-\alpha} h(x; \alpha, \lambda). \qquad \square$$

25

Appendix A.3. Proofs of Subsection 3.4

From Jessen & Mikosch (2006, p. 177), the following results are obtained:

Lemma 1. *Assuming that X is a regularly varying r.v. with index $\alpha > 0$, that $E[N] < +\infty$ and that $P(N > x) = o(P(X > x))$, then, as $x \to +\infty$*

$$P(Y > x) \sim E[N] \, P(X > x). \tag{A.8}$$

The next lemmas are required, since they prove that the hypothesis for Lemma 1 is satisfied.

Lemma 2. *An r.v. X with a Zipf distribution is a regularly varying r.v.*

Proof. According to Gulisashvili (2012, p. 220), Pareto-type distributions (i.e., the Zipf distribution) are camouflaged versions of regularly varying functions. Basically, a function f belongs to the class of Pareto-type distribution if it is asymptotically equivalent to a regularly varying function. This implies that f is also a regularly varying function. □

Lemma 3. *Let $f(x)$ be defined as $f(x) = \frac{\lambda^x \, x^\alpha}{x!}$, then it verifies that: a) it is a decreasing function of x for x large enough; and b) $\lim_{x \to +\infty} f(x) = 0$.*

Proof. Given that,

$$\frac{f(x+1)}{f(x)} = \frac{\lambda^{(x+1)} \, (x+1)^\alpha}{(x+1)!} \, \frac{x!}{\lambda^x \, x^\alpha} = \frac{\lambda}{(x+1)} \left(1 + \frac{1}{x} \right)^\alpha,$$

one has that $\lim_{x \to +\infty} f(x+1)/f(x) = 0$ and, thus for large values of x, $f(x + 1) \le f(x)$, which proves a).

Taking into account that the Poisson distribution has moments of any order, and denoting by $\llcorner \alpha \lrcorner$ the integer part of α, we have that for any $\alpha \in \mathbb{R}$, and $\alpha > 1$,

$$0 \le \sum_{x=0}^{+\infty} \frac{\lambda^x \, x^\alpha}{x!} \le \sum_{x=0}^{+\infty} \frac{\lambda^x \, x^{\llcorner \alpha \lrcorner + 1}}{x!} < +\infty,$$

and consequently, $\lim_{x \to +\infty} f(x) = 0$. □

Lemma 4. *Let $N \sim Po(\lambda)$ and $X \sim Zipf(\alpha)$, N and X being independent r.v.'s. Then*

$$P(N > x) = o(P(X > x)).$$

26

Proof. We need to prove that $P(N > x)/P(X > x)$ has limit zero at infinity. However, taking into account point a) of Lemma 3, one has that for x large enough,

$$0 \le \frac{P(N > x)}{P(X > x)} = \frac{\sum_{i=x+1}^{+\infty} \frac{e^{-\lambda}\lambda^i}{i!}}{\sum_{i=x+1}^{+\infty} \frac{i^{-\alpha}}{\zeta(\alpha)}} = e^{-\lambda}\zeta(\alpha)\frac{\sum_{i=x+1}^{+\infty} \frac{\lambda^i i^\alpha i^{-\alpha}}{i!}}{\sum_{i=x+1}^{+\infty} i^{-\alpha}}$$

$$\le e^{-\lambda}\zeta(\alpha)\frac{\lambda^x x^\alpha}{x!}\frac{\sum_{i=x+1}^{+\infty} i^{-\alpha}}{\sum_{i=x+1}^{+\infty} i^{-\alpha}} \le e^{-\lambda}\zeta(\alpha)\frac{\lambda^x x^\alpha}{x!}.$$

Taking limits in the inequality, and as a consequence of point b) of Lemma 3, one has that:

$$0 \le \lim_{x \to +\infty} \frac{P(N > x)}{P(X > x)} \le e^{-\lambda}\zeta(\alpha)\lim_{x \to +\infty}\frac{\lambda^x x^\alpha}{x!} = 0,$$

which proves the proposition. $\qquad\square$

Proof of Theorem 1: The proof is a consequence of Lemma 1, where $X \sim \mathrm{Zipf}(\alpha)$ and $N \sim \mathrm{Po}(\lambda)$, and X and N are independent r.v.'s Thus, we have that $P(Y > x) \sim E[N]\,P(X > x)$, and given that

$$P(Y = x) = P(Y > x - 1) - P(Y > x),$$

one has that

$$P(Y = x) \sim \lambda\left[P(X > x - 1) - P(X > x)\right] \Leftrightarrow P(Y = x) \sim \lambda\,P(X = x). \quad \square$$

Appendix A.4. Proof of Subsection 3.5

Proof of Proposition 6: Given that, when $\alpha \to +\infty$, the Zipf distribution tends towards the degenerate distribution at one, we have that, for large values of α, the Zipf-PSS(α, λ) is equal to the sum of N ones, which is equal to N. These results can be observed in the top right-hand side of Figure 1, where for $\alpha = 2.3$ and $\lambda = 10$, the PMF of the Zipf-PSS looks like the PMF of a Poisson. $\qquad\square$

Proof of Proposition 7: Taking into account that,

$$\frac{Li_\alpha(z) - \zeta(\alpha)}{\zeta(\alpha)} = \frac{1}{\zeta(\alpha)}\left[\sum_{k=1}^{+\infty}\frac{z^k}{k^\alpha} - \sum_{k=1}^{+\infty}\frac{1}{k^\alpha}\right] = \frac{1}{\zeta(\alpha)}\sum_{k=1}^{+\infty}\frac{(z^k - 1)}{k^\alpha}$$

$$= \frac{(z-1)}{\zeta(\alpha)} + \frac{(z^2 - 1)}{2^\alpha\,\zeta(\alpha)} + \frac{(z^3 - 1)}{3^\alpha\,\zeta(\alpha)} + \dots,$$

27

from (3.1), the first to fourth order approximations of the PGF of r.v. Y with
a Zipf-PSS(α, λ) distribution are those mentioned in points a) to d). $\qquad \square$

Appendix B. Tables Associated with the Fitted Degree Sequences

Table 2: For each of the degree sequences, one has: in the first column, the names of the models fitted; from the second to the fourth columns, the maximum likelihood parameter estimates together with their corresponding confidence intervals (**CI**); and the values of the AIC in the last column.

Year	Type	Model	$param_1$	CI_{param_1}	$param_2$	CI_{param_2}	AIC
2009	in	Zipf-PSS	$\hat{\alpha} = 1.7636$	(1.7111, 1.816)	$\hat{\lambda} = 0.9752$	(0.906, 1.0444)	6216.0239
		NB	$\hat{k} = 0.2734$	(0.2484, 0.2984)	$\hat{q} = 0.9532$	(0.9469, 0.9595)	6241.2867
		DW	$\hat{q} = 0.5717$	(0.5465, 0.597)	$\hat{\beta} = 0.5239$	(0.4967, 0.5511)	**6170.7334**
	out	**Zipf-PSS**	$\hat{\alpha} = 1.9202$	(1.8633, 1.977)	$\hat{\lambda} = 1.1738$	(1.1009, 1.2467)	**6246.7526**
		NB	$\hat{k} = 0.3761$	(0.3448, 0.4074)	$\hat{q} = 0.9367$	(0.9294, 0.9441)	6567.5277
		DW	$\hat{q} = 0.6386$	(0.6157, 0.6616)	$\hat{\beta} = 0.6112$	(0.5841, 0.6383)	6479.5210
2010	in	**Zipf-PSS**	$\hat{\alpha} = 1.9048$	(1.8744, 1.9352)	$\hat{\lambda} = 1.2202$	(1.179, 1.2614)	**23120.8875**
		NB	$\hat{k} = 0.3736$	(0.3564, 0.3908)	$\hat{q} = 0.9346$	(0.9305, 0.9387)	23707.2106
		DW	$\hat{q} = 0.6361$	(0.6238, 0.6485)	$\hat{\beta} = 0.6118$	(0.5968, 0.6267)	23397.4229
	out	**Zipf-PSS**	$\hat{\alpha} = 1.8503$	(1.8194, 1.8813)	$\hat{\lambda} = 0.8625$	(0.8311, 0.894)	**20353.4265**
		NB	$\hat{k} = 0.2493$	(0.2378, 0.2608)	$\hat{q} = 0.9553$	(0.9522, 0.9585)	21433.7524
		DW	$\hat{q} = 0.5340$	(0.521, 0.547)	$\hat{\beta} = 0.5066$	(0.4936, 0.5197)	21040.4257
2011	in	**Zipf-PSS**	$\hat{\alpha} = 2.1226$	(2.0864, 2.1588)	$\hat{\lambda} = 1.2067$	(1.17, 1.2434)	**23703.7434**
		NB	$\hat{k} = 0.5101$	(0.4867, 0.5336)	$\hat{q} = 0.8711$	(0.8642, 0.8779)	24555.5682
		DW	$\hat{q} = 0.6393$	(0.628, 0.6507)	$\hat{\beta} = 0.7229$	(0.7069, 0.7388)	24322.7655
	out	**Zipf-PSS**	$\hat{\alpha} = 1.9400$	(1.9051, 1.975)	$\hat{\lambda} = 0.7191$	(0.6931, 0.7451)	**20165.4236**
		NB	$\hat{k} = 0.2407$	(0.2294, 0.2521)	$\hat{q} = 0.9347$	(0.9302, 0.9392)	21277.6210
		DW	$\hat{q} = 0.4753$	(0.463, 0.4876)	$\hat{\beta} = 0.5180$	(0.5048, 0.5313)	20869.6246
2012	in	**Zipf-PSS**	$\hat{\alpha} = 2.3127$	(2.2705, 2.355)	$\hat{\lambda} = 1.2328$	(1.1976, 1.2681)	**23527.8020**
		NB	$\hat{k} = 0.6456$	(0.6151, 0.6761)	$\hat{q} = 0.8130$	(0.804, 0.822)	24603.2748
		DW	$\hat{q} = 0.6473$	(0.6366, 0.6581)	$\hat{\beta} = 0.8046$	(0.788, 0.8211)	24415.4935
	out	**Zipf-PSS**	$\hat{\alpha} = 1.9926$	(1.9545, 2.0306)	$\hat{\lambda} = 0.6330$	(0.6098, 0.6563)	**19413.6498**
		NB	$\hat{k} = 0.2244$	(0.2136, 0.2352)	$\hat{q} = 0.9259$	(0.9208, 0.9311)	20617.3471
		DW	$\hat{q} = 0.4352$	(0.4233, 0.4472)	$\hat{\beta} = 0.5146$	(0.5014, 0.5278)	20159.4326

Table 3: For each of the degree sequences, one has: in the first column, the names of the models fitted; from the second to the fourth columns, the maximum likelihood parameter estimates together with their corresponding confidence intervals (**CI**); and the values of the AIC in the last column.

Year	Type	Model	$param_1$	CI_{param_1}	$param_2$	CI_{param_2}	AIC
2013	in	**Zipf-PSS**	$\hat{\alpha} = 2.4882$	(2.4412, 2.5353)	$\hat{\lambda} = 1.2234$	(1.1905, 1.2562)	**23726.1601**
		NB	$\hat{k} = 0.7861$	(0.7487, 0.8235)	$\hat{q} = 0.7548$	(0.7442, 0.7655)	25134.4768
		DW	$\hat{q} = 0.6520$	(0.6418, 0.6622)	$\hat{\beta} = 0.8721$	(0.8553, 0.8888)	25011.8353
	out	**Zipf-PSS**	$\hat{\alpha} = 1.9572$	(1.9196, 1.9949)	$\hat{\lambda} = 0.5484$	(0.5277, 0.5691)	**19446.4311**
		NB	$\hat{k} = 0.1977$	(0.188, 0.2075)	$\hat{q} = 0.9245$	(0.9192, 0.9299)	20389.9913
		DW	$\hat{q} = 0.3951$	(0.3835, 0.4066)	$\hat{\beta} = 0.4896$	(0.4765, 0.5028)	20006.2583
2014	in	**Zipf-PSS**	$\hat{\alpha} = 2.5422$	(2.4898, 2.5947)	$\hat{\lambda} = 1.2115$	(1.1772, 1.2458)	**21150.4685**
		NB	$\hat{k} = 0.8385$	(0.7949, 0.8821)	$\hat{q} = 0.7297$	(0.7175, 0.7419)	22304.9266
		DW	$\hat{q} = 0.6485$	(0.6378, 0.6592)	$\hat{\beta} = 0.8943$	(0.8763, 0.9123)	22217.8324
	out	**Zipf-PSS**	$\hat{\alpha} = 1.9792$	(1.9391, 2.0194)	$\hat{\lambda} = 0.5551$	(0.5333, 0.577)	**17633.8811**
		NB	$\hat{k} = 0.2108$	(0.1998, 0.2218)	$\hat{q} = 0.9148$	(0.9086, 0.921)	18450.5365
		DW	$\hat{q} = 0.4000$	(0.3879, 0.4121)	$\hat{\beta} = 0.5064$	(0.4923, 0.5206)	18127.9951
2015	in	**Zipf-PSS**	$\hat{\alpha} = 2.6118$	(2.5554, 2.6683)	$\hat{\lambda} = 1.1825$	(1.1487, 1.2164)	**20006.8297**
		NB	$\hat{k} = 0.8251$	(0.782, 0.8683)	$\hat{q} = 0.7262$	(0.7137, 0.7386)	21403.2895
		DW	$\hat{q} = 0.6352$	(0.6244, 0.6461)	$\hat{\beta} = 0.8830$	(0.8655, 0.9004)	21282.5344
	out	**Zipf-PSS**	$\hat{\alpha} = 2.0214$	(1.9791, 2.0637)	$\hat{\lambda} = 0.5688$	(0.5464, 0.5912)	**17084.7695**
		NB	$\hat{k} = 0.2230$	(0.2112, 0.2348)	$\hat{q} = 0.9075$	(0.9008, 0.9142)	17969.2179
		DW	$\hat{q} = 0.4058$	(0.3936, 0.4181)	$\hat{\beta} = 0.5222$	(0.5077, 0.5367)	17633.6200
2016	in	**Zipf-PSS**	$\hat{\alpha} = 3.3093$	(3.1162, 3.5024)	$\hat{\lambda} = 0.9708$	(0.9178, 1.0238)	**4897.3457**
		NB	$\hat{k} = 1.7500$	(1.4651, 2.0348)	$\hat{q} = 0.4156$	(0.3738, 0.4575)	5053.1047
		DW	$\hat{q} = 0.5992$	(0.5782, 0.6203)	$\hat{\beta} = 1.1396$	(1.094, 1.1852)	5070.1318
	out	**Zipf-PSS**	$\hat{\alpha} = 2.3938$	(2.2856, 2.502)	$\hat{\lambda} = 0.6210$	(0.5787, 0.6633)	**4760.6672**
		NB	$\hat{k} = 0.4404$	(0.3903, 0.4905)	$\hat{q} = 0.7387$	(0.7113, 0.766)	4921.4060
		DW	$\hat{q} = 0.4397$	(0.4173, 0.4621)	$\hat{\beta} = 0.7250$	(0.6896, 0.7604)	4888.3861

References

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, *59*, 381.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512. URL: https://doi.org/10.1126/science.286.5439.509. doi:10.1126/science.286.5439.509.

Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of financial networks. *Computational statistics & data analysis*, *48*, 431–443.

Broido, A. D., & Clauset, A. (2018). Scale-free networks are rare. *arXiv preprint arXiv:1801.03400*, .

Ferrer-i Cancho, R., & Vitevitch, M. S. (2018). The origins of Zipf's meaning-frequency law. *Journal of the Association for Information Science and Technology*, *69*, 1369–1379.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.*, *51*, 661–703. URL: https://doi.org/10.1137/070710111. doi:10.1137/070710111.

Duarte-López, A., & Pérez-Casany, M. (2018). *zipfextR: Zipf Extended Distributions*. URL: https://CRAN.R-project.org/package=zipfextR r package version 1.0.0.

Duarte-López, A., Prat-Pérez, A., & Pérez-Casany, M. (2015). Using the Marshall-Olkin Extended Zipf Distribution in Graph Generation. In *European Conference on Parallel Processing* (pp. 493–502). Springer.

Erling, O., Averbuch, A., Larriba-Pey, J., Chafi, H., Gubichev, A., Prat, A., Pham, M.-D., & Boncz, P. (2015). The LDBC social network benchmark: Interactive workload. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (pp. 619–630). ACM.

Feller, W. (1971). An introduction to probability theory and its applications. *Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1971, 3rd ed.*, .

Gomez-Lievano, A., Youn, H., & Bettencourt, L. M. (2012). The statistics of urban scaling and their connection to Zipf's law. *PloS one*, *7*, e40393.

Gulisashvili, A. (2012). *Analytically tractable stochastic stock price models*. Springer Finance. Springer, Heidelberg. URL: https://doi.org/10.1007/978-3-642-31214-4. doi:10.1007/978-3-642-31214-4.

30

Güney, Y., Tuaç, Y., & Arslan, O. (2017). Marshall-Olkin distribution: parameter estimation and application to cancer data. *J. Appl. Stat.*, *44*, 2238–2250. URL: https://doi.org/10.1080/02664763.2016.1252730. doi:10.1080/02664763.2016.1252730.

Jessen, H. A., & Mikosch, T. (2006). Regularly varying functions. *Publications de l'Institut Mathematique*, *80*, 171–192.

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions*. Wiley Series in Probability and Statistics (3rd ed.). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. URL: https://doi.org/10.1002/0471715816. doi:10.1002/0471715816.

Keller, J. (2010). Beyond Facebook: How the World's Mathematicians Organize Online. https://www.theatlantic.com/technology/archive/2010/09/beyond-facebook-how-the-worlds-mathematicians-organize-online/63422/. (Accessed on 10/01/2019).

Kim, H., Del Genio, C. I., Bassler, K. E., & Toroczkai, Z. (2012). Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics*, *14*, 023012.

Koziol, J., Griffin, N., Long, F., Li, Y., Latterich, M., & Schnitzer, J. (2013). On protein abundance distributions in complex mixtures. *Proteome science*, *11*, 5.

Leskovec, J., & Krevl, A. (2014). SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

Low, W. J., Wilson, P., & Thelwall, M. (2016). Stopped sum models and proposed variants for citation data. *Scientometrics*, *107*, 369–384.

Mahanti, A., Carlsson, N., Mahanti, A., Arlitt, M., & Williamson, C. (2013). A tale of the tails: Power-laws in internet measurements. *IEEE Network*, *27*, 59–64.

McKelvey, B. et al. (2018). Using maximum likelihood estimation methods and complexity science concepts to research power law-distributed phenomena. In *Handbook of Research Methods in Complexity Science*. Edward Elgar Publishing.

Meng, S., & Gao, G. (2018). Compound Poisson Claims Reserving Models: Extensions and Inference. *ASTIN Bulletin: The Journal of the IAA*, (pp. 1–20).

Nakagawa, T., & Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, *24*, 300–301.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, *46*, 323–351.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *Astin Bull.*, *12*, 22–26. URL: https://doi.org/10.1017/S0515036100006796. doi:10.1017/S0515036100006796.

Paranjape, A., Benson, A. R., & Leskovec, J. (2017). Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 601–610). ACM.

Patel, Y. C. (1976). Estimation of the parameters of the triple and quadruple stuttering-Poisson distributions. *Technometrics*, *18*, 67–73. URL: https://doi.org/10.2307/1267918. doi:10.2307/1267918.

Pérez-Casany, M., & Casellas, A. (2013). Marshall-Olkin Extended Zipf Distribution. *arXiv preprint arXiv:1304.4540*, .

Podur, J. J., Martell, D. L., & Stanford, D. (2010). A compound Poisson model for the annual area burned by forest fires in the province of Ontario. *Environmetrics*, *21*, 457–469. URL: https://doi.org/10.1002/env.996. doi:10.1002/env.996.

Puig, P., & Valero, J. (2007). Characterization of count data distributions

₅₅₅ involving additivity and binomial subsampling. *Bernoulli*, *13*, 544–555. URL: https://doi.org/10.3150/07-BEJ6021. doi:10.3150/07-BEJ6021.

Satterthwaite, F. (1942). Generalized Poisson distribution. *The Annals of Mathematical Statistics*, *13*, 410–417.

Seshadhri, C., Kolda, T. G., & Pinar, A. (2012). Community structure and

₅₆₀ scale-free collections of erdős-rényi graphs. *Physical Review E*, *85*, 056109.

Sierksma, G., & Hoogeveen, H. (1991). Seven criteria for integer sequences being graphic. *Journal of Graph Theory*, *15*, 223–231.

Sundt, B. r., & Jewell, W. S. (1981). Further results on recursive evaluation of compound distributions. *Astin Bull.*, *12*, 27–39. URL: https://doi.org/10.1017/S0515036100006802. doi:10.1017/S0515036100006802.

₅₆₅

Wang, D., Cheng, H., Wang, P., Huang, X., & Jian, G. (2017). Zipf's law in passwords. *IEEE Transactions on Information Forensics and Security*, *12*, 2776–2791.

Zhang, H., & Li, B. (2016). Characterizations of discrete compound

₅₇₀ Poisson distributions. *Comm. Statist. Theory Methods*, *45*, 6789–6802. URL: https://doi.org/10.1080/03610926.2014.901375. doi:10.1080/03610926.2014.901375.

Zipf, G. K. (1949). Human Behaviour and the Principle of Least-Effort. Cambridge MA edn.

₅₇₅ Zörnig, P., & Altmann, G. (1995). Unified representation of Zipf distributions. *Computational Statistics & Data Analysis*, *19*, 461–473.