A Mixture Model of Truncated Zeta Distributions with Applications to Scientific Collaboration Networks †

Hohyun Jung [1,2] and Frederick Kin Hing Phoa [1,*]

1   Institute of Statistical Science, Academia Sinica, Taipei City 11529, Taiwan; hhjung@stat.sinica.edu.tw
2   Department of Statistics, Sungshin Women's University, Seoul 02844, Korea
*   Correspondence: fredphoa@stat.sinica.edu.tw
†   This paper is an extended version of our paper published in Jung, H.; Phoa, F.K.H. Analysis of a Finite Mixture of Truncated Zeta Distributions for Degree Distribution. In Proceedings of the International Conference on Complex Networks and Their Applications, Madrid, Spain, 1–3 December 2020; pp. 497–507.

Abstract: The degree distribution has attracted considerable attention from network scientists in the last few decades to have knowledge of the topological structure of networks. It is widely acknowledged that many real networks have power-law degree distributions. However, the deviation from such a behavior often appears when the range of degrees is small. Even worse, the conventional employment of the continuous power-law distribution usually causes an inaccurate inference as the degree should be discrete-valued. To remedy these obstacles, we propose a finite mixture model of truncated zeta distributions for a broad range of degrees that disobeys a power-law behavior in the range of small degrees while maintaining the scale-free behavior. The maximum likelihood algorithm alongside the model selection method is presented to estimate model parameters and the number of mixture components. The validity of the suggested algorithm is evidenced by Monte Carlo simulations. We apply our method to five disciplines of scientific collaboration networks with remarkable interpretations. The proposed model outperforms the other alternatives in terms of the goodness-of-fit.

Keywords: truncated zeta distribution; power-law; collaboration network; degree distribution; mixture model

1. Introduction

Network science focuses on the study of complex networks such as telecommunication, computer, biological, cognitive, and social networks. A network consists of nodes and links. The topological structure, which explores how nodes are connected in the system, has been investigated with great interest [1–3]. Network researchers have examined foundational network topologies using various network-related quantities such as the degree distribution, clustering coefficient, and average path length. Most networks are dynamic, so accordingly, network-related quantities also change over time. Studying the evolution of these network quantities could provide insight into the behavior of individuals expressed by nodes and the change of topological properties of the network. The evolution of a network has been studied from various perspectives, e.g., the community [4,5], rich-get-richer [1], node heterogeneity [2,6], and link persistence [3].

The degree of a node is the number of links connected to a node. The degree distribution of a network, $P(k)$, tells us the probability of degree k that a randomly chosen node will have. One challenge in studying network science is to develop simplified measures that capture the network structure. The degree distribution is one of such measures that help to find influential nodes in the system. Many attempts have been made to study the degree distribution using Poisson, exponential, and power-law distributions. In particular, the analysis of the power-law degree distribution has been considered as one of the basic steps, and networks that have power-law degree distributions are often referred to as

scale-free networks. We can frequently observe power-law degree distributions in collaboration, World Wide Web, protein–protein interactions, and semantic networks [7–10]. The emergence of hubs (highly connected nodes) is a consequence of a scale-free property of networks. A rich-get-richer mechanism, also called a popularity effect [11], has been known to produce power-law degree distributions and hubs [1,6].

Many dynamic network models have been developed to explain the power-law degree distribution in real networks. The most widely known dynamic model for the power-law degree distribution is the rich-get-richer generative model in Barabasi and Albert (1999) [1], called the BA model. They employ the preferential attachment mechanism in which nodes with more neighbors tend to receive more links from other nodes. Specifically, the algorithm of the BA model has the following parts with the parameter m that controls the number of links over time:

- Growth: At each time point, a node enters the network. Then the node tries to connect with m nodes in the network.
- Preferential attachment: The newly entered node connects with node i with probability proportional to the degree of node i.

The BA model yields a power-law degree distribution with exponent 3, i.e., $P(k) \propto k^{-3}$. The exact form, given by Bollobas et al. in [12], is:

$$P(k|m) = \frac{2m(m+1)}{k(k+1)(k+2)}, \quad k = m, m+1, \cdots. \tag{1}$$

There are many variant models to cover a broad range of power-law exponents [13–15], degree correlations [16], accelerating growth [17–19], and node heterogeneity [6,20,21]. Another model for generating the power-law degree distribution is a copying network model presented in Kumar et al. (2000) [22]. In this model, newly entered nodes randomly select some existing nodes and copy some of the links.

The power-law distribution has the form of $P(k) \propto k^{-\alpha}$, which can be expressed as $\ln P(k) = -\alpha \ln k + constant$. Therefore, a straight line of $\ln P(k)$ plot on $\ln k$ (log-log plot) can be an indication of a power-law relationship, and its slope is a power-law exponent. In real networks, however, the degree distribution does not have a shape of a straight line in the entire range. There are many empirical distributions where the power-law behavior is not observed in the range of small degrees. Many variants of the power-law distribution are developed to address this issue, such as generalized power-law distributions [23,24], composite distributions with threshold [25,26], and power-law distributions with an exponential cutoff [27,28]. However, these methods do not consider the essential foundation of the power-law. According to the BA model and its variants, the power-law nature is an inherent property exhibited from the preferential attachment rule. The model presented in this paper preserves the power-law nature to avoid manual modifications of the power-law distribution function, given by $P(k) \propto k^{-\alpha}$.

Note that the BA model has a parameter m. Jordan (2006) [29] relieved the constant m condition that, at each time, the number M of connections can change over time according to the distribution of M. The degree distribution turned out to be

$$P(k) = \frac{2E[M(M+1)\mathbb{1}(M \le k)]}{k(k+1)(k+2)}, \quad k = 1, 2, \cdots. \tag{2}$$

We note the following statistical property.

Theorem 1. *Suppose that M has a finite support, $M = 1, \cdots, M_{max}$. Then the degree distribution of Jordan's model can be expressed as a mixture distribution, given by*

$$P(k) = \sum_{m=1}^{M_{max}} w_m P(k|m), \quad k = 1, 2, \cdots, \tag{3}$$

where $w_m = P(M = m)$ is a mixture weight corresponding to the mth mixture component $P(k|m)$ in Equation (1), $m = 1, \cdots, M_{max}$.

Proof. Note that we can rewrite $P(k|m)$ and $P(k)$ as

$$P(k|m) = \frac{2m(m+1)}{k(k+1)(k+2)} \mathbb{1}(k \geq m)$$

$$P(k) = \sum_{m=1}^{M_{max}} w_m \frac{2m(m+1)}{k(k+1)(k+2)} \mathbb{1}(m \leq k)$$

from Equations (1) and (2). Then, we have

$$P(k) = \sum_{m=1}^{M_{max}} w_m \frac{2m(m+1)}{k(k+1)(k+2)} \mathbb{1}(m \leq k)$$

$$= \sum_{m=1}^{M_{max}} w_m P(k|m)$$

by $\mathbb{1}(k \geq m) = \mathbb{1}(m \leq k)$. $\square$

Theorem 1 suggests that the degree distribution might be expressed as a mixture distribution. For example, we consider a network where a new node connects to one or two existing nodes, with probabilities $P(M = 1) = 2/3$ and $P(M = 2) = 1/3$. Then we have the following degree distribution in a mixture form of the two BA model's distributions with different m,

$$P(k) = \frac{2}{3} P(k|m = 1) + \frac{1}{3} P(k|m = 2).$$

Inspired by this property, we consider a mixture model as an explanation of the deviation from the power-law in the range of small degrees.

Moreover, many studies have considered the degree as a continuous variable using continuity assumptions [30,31]. This approach may mislead researchers since the degree is discrete-valued. Therefore, a discrete power-law distribution, called a truncated zeta distribution, is used in this paper.

In this study, we propose a mixture model of truncated zeta distributions for the analysis of degree distributions. The proposed model covers the entire range of the degree distribution through a mixture of truncated zeta distributions while maintaining the scale-free nature of a network. We can characterize the degree distribution more accurately through the discrete version of the power-law distribution. In addition, we present the maximum likelihood estimation algorithm along with a model selection method. A simulation study examines the validity of the proposed estimation procedure. In addition, real collaboration networks are investigated with the proposed model to describe the characteristics of the degree distribution.

We focus on analyzing actual scientific collaboration networks and have made significant advancements compared to the previous work in Jung and Phoa (2020) [32]. The major improvements are as follows:

- We detected some inconsistency in the scientific collaboration data. For example, "Smith, James," "Smith, John," and "Smith, Jacob" are all stated as "Smith, J" before 2007. Hence, we change the period of data to avoid the author name inconsistency for accurate inferences.
- Moreover, a more elaborate analysis of the real data is conducted with noteworthy interpretations.
- The validity of the presented algorithm is addressed with Monte Carlo simulations.

- Extensive comparison studies are performed to show the superiority of the proposed model. We compared the proposed model with generalized Pareto models as well as base models that lack discreteness or mixture nature.
- We provide more detailed explanations throughout the paper.

The rest of the paper is organized as follows. The continuous and discrete power-law distributions are defined in Section 2, and the proposed mixture model of truncated zeta distributions is defined in Section 3. Section 4 presents the estimation method of the mixture model, and the validity of the estimation procedure is demonstrated in Section 5. In Section 6, we analyze the scientific collaboration network by applying the proposed mixture model with interpretations. Section 7 concludes the paper.

2. The Power-Law Distribution

2.1. Continuous Power-Law Distribution

The probability density function (pdf) of the continuous power-law distribution parameterized by the power-law exponent $\alpha > 1$ and the minimum value $l > 0$, denoted by $PL(\alpha, l)$, is given by [7]

$$f(x|\alpha, l) = (\alpha - 1)l^{\alpha-1}x^{-\alpha}, \quad x \geq l. \tag{4}$$

Note that the support is real values greater than or equal to l.

The complementary cumulative distribution function is useful to describe the tail of the power-law distribution, expressed as

$$P(X \geq x) = \left(\frac{x}{l}\right)^{-\alpha+1}, \quad x \geq l.$$

Its moments are given by

$$E[X^m] = \frac{\alpha - 1}{\alpha - 1 - m}l^m,$$

provided $\alpha > m + 1$. We can deduce that $PL(\alpha, l)$ has the mean and variance given by

$$E[X] = \frac{\alpha - 1}{\alpha - 2}l, \quad Var[X] = \frac{\alpha - 1}{(\alpha - 3)(\alpha - 2)^2}l^2,$$

where the mean and variance are well-defined for $\alpha > 2$ and $\alpha > 3$, respectively.

The continuous power-law distribution has been widely used in the analysis of the degree distribution. To analyze the discrete-valued variable degree, we need an approximation method. Setting a constant c, $0 \leq c \leq 1$, for the correction of continuity, the degree distribution can be approximated by

$$f(k|\alpha, l) \approx \int_{k-c}^{k+1-c} f(x|\alpha, l)dx \tag{5}$$

for an integer value k.

One of the most common approaches is to round values to the nearest integer, which corresponds to $c = 0.5$ [7,33]. This *rounding* approach is acceptable when considering the tail part of the power-law distribution. However, if k is small, the constant c that satisfies the exact equation of (5) may be considerably less than 0.5, and it should be avoided. According to Clauset et al. (2009) in [7], the rounding approach is reasonable for $k > 6$.

Since the node degree is usually small, the approximation may lead to misunderstanding when performing statistical analysis on the degree distribution, such as generating node degrees and fitting the distribution to real data. We consider the *exact* version of the discrete power-law in the next subsection.

2.2. Truncated Zeta Distribution

The truncated zeta distribution, denoted by $TZ(\alpha, l)$, is a discrete form of the power-law distribution. Parameters are the same as the continuous power-law distribution in which $\alpha > 1$ is the power-law exponent and $l > 0$ is the minimum value. The probability mass function (pmf) of $TZ(\alpha, l)$ is given by [7]

$$g(k|\alpha, l) = \frac{1}{\zeta(\alpha, l)} k^{-\alpha}, \quad k = l, l+1, \cdots, \tag{6}$$

where $\zeta(\cdot, \cdot)$ is the Hurwitz zeta function

$$\zeta(\alpha, l) = \sum_{k=l}^{\infty} k^{-\alpha},$$

which can be regarded as the normalizing constant of the distribution. The Hurwitz zeta function in Equation (6) corresponds to the continuous counterpart $1/(\alpha - 1)l^{\alpha-1}$ in Equation (4) via the upper Riemann sum approximation, expressed as

$$\zeta(\alpha, l) \gtrsim \int_l^{\infty} x^{-\alpha} dx = 1/(\alpha - 1)l^{\alpha-1}.$$

The complementary cumulative distribution function is given by

$$P(K \geq k) = \frac{\zeta(\alpha, k)}{\zeta(\alpha, l)}, \quad k \geq l.$$

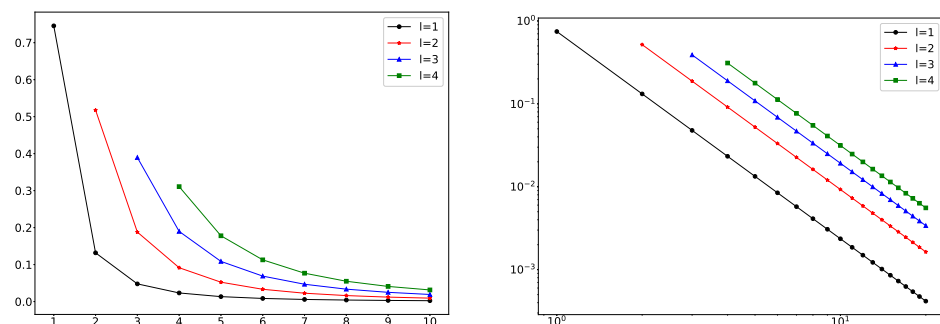Next, the moments of the truncated zeta distribution are expressed as

$$E[K^m] = \frac{\zeta(\alpha - m, l)}{\zeta(\alpha, l)}, \quad \alpha > m + 1.$$

We can straightforwardly derive the mean and variance as

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

provided $\alpha > 2$ for the mean and $\alpha > 3$ for the variance.

Figure 1 shows some pmfs of $TZ(\alpha, l)$ when $\alpha = 2.50$. We can check that the straight line of a log-log plot can be strong evidence of a power-law behavior.

Figure 1. The pmfs of $TZ(\alpha, l)$ on a standard scale (**left**) and a log-log scale (**right**) when $\alpha = 2.50$ and $l = 1, 2, 3, 4$.

3. Truncated Zeta Mixture Model

We consider the finite mixture model of truncated zeta distributions by fixing the power-law exponent α for mixture components while varying minimum values to produce a mixture of truncated zeta distributions. The probability mass function is represented as

$$p(k|\alpha, L, \mathbf{w}) = \sum_{l=1}^{L} w_l g(k|\alpha, l), \quad k = 1, 2, \cdots, \tag{7}$$

where $g(k|\alpha, l)$ is the pmf of $TZ(\alpha, l)$, and L is the number of mixture components. Mixture weights $\mathbf{w} = (w_1, w_2, \cdots, w_L)$

$$w_l \geq 0, \ l = 1, 2, \cdots, L-1, \ w_L > 0, \text{ and } \sum_{l=1}^{L} w_l = 1.$$

In this paper, we assume that the minimum value l is equal to 1, but it can be modified according to the data. The tail of most real networks follows the power-law distribution, and Equation (7) has the exact power-law behavior for sufficiently large degrees.

Theorem 2. *For k larger than or equal to L, the truncated zeta mixture distribution in Equation (7) has the exact power-law relationship, given by*

$$p(k|\alpha, L, \mathbf{w}) \propto k^{-\alpha}.$$

Proof. By using the pmf of $TZ(\alpha, l)$, we can write

$$p(k|\alpha, L, \mathbf{w}) = \left(\sum_{l=1}^{L} \frac{w_l}{\zeta(\alpha, l)} \right) k^{-\alpha}.$$

Since the term inside the bracket is independent with k, the pmf of the mixture is proportional to $k^{-\alpha}$. □

Mixture models may suffer from the non-identifiability issue even for finite mixtures. The following theorem proves that the proposed truncated zeta mixture model is identifiable.

Theorem 3. *The mixture distribution Equation (7) is identifiable with respect to α, L, and $\mathbf{w}$.*

Proof. Let $p(k|\alpha_1, L_1, \mathbf{w}_1)$ and $p(k|\alpha_2, L_2, \mathbf{w}_2)$ be the mixture distributions of α_1, L_1, $\mathbf{w}_1 = (w_{11}, w_{12}, \cdots, w_{1L})$ and α_2, L_2, $\mathbf{w}_2 = (w_{21}, w_{22}, \cdots, w_{2L})$, respectively. Suppose that $p(k|\alpha_1, L_1, \mathbf{w}_1)$ and $p(k|\alpha_2, L_2, \mathbf{w}_2)$ are identical, i.e., $p(k|\alpha_1, L_1, \mathbf{w}_1) = p(k|\alpha_2, L_2, \mathbf{w}_2)$ for all $k = 1, 2, \cdots$. Further, we define the slope function $s(k|\alpha, L, \mathbf{w})$ of the log-log degree distribution, given by

$$s(k|\alpha, L, \mathbf{w}) = \frac{\ln p(k+1|\alpha, L, \mathbf{w}) - \ln p(k|\alpha, L, \mathbf{w})}{\ln(k+1) - \ln(k)}, \quad k = 1, 2, \cdots.$$

The equality of the two mixture distributions give the identical slope function, $s(k|\alpha_1, L_1, \mathbf{w}_1) = s(k|\alpha_2, L_2, \mathbf{w}_2)$ for all $k = 1, 2, \cdots$. We then have $\alpha = s(k|\alpha, L, \mathbf{w})$ for the sufficiently large k ($\geq \max\{L_1, L_2\}$). Therefore, we obtain $\alpha_1 = \alpha_2$. The number of mixture components L is the largest integer k such that $s(k-1|\alpha, L, \mathbf{w}) \neq \alpha$, and we also have $L_1 = L_2$. Let $L (= L_1 = L_2)$ and $\alpha (= \alpha_1 = \alpha_2)$ be the common number of mixture components and the power-law exponent.

Using $p(k|\alpha_1, L_1, \mathbf{w}_1) = p(k|\alpha_2, L_2, \mathbf{w}_2)$ for $k = 1, 2, \cdots, L$, we have the following L equations:

$$w_{11}g(1|\alpha, 1) = w_{21}g(1|\alpha, 1),$$
$$w_{11}g(2|\alpha, 1) + w_{12}g(2|\alpha, 2) = w_{21}g(2|\alpha, 1) + w_{22}g(2|\alpha, 2),$$
$$\vdots$$
$$w_{11}g(L|\alpha, 1) + \cdots + w_{1L}g(L|\alpha, L) = w_{21}g(L|\alpha, 1) + \cdots + w_{2L}g(L|\alpha, L).$$

By solving these equations, we get $w_{1l} = w_{2l}$ for $l = 1, 2, \cdots, L$. Thus, the mixture of truncated zeta distributions is identifiable. $\quad\square$

In Figure 2, we depict some log-log plots of mixture distributions. We can see that this model can handle empirical degree distributions that do not follow the power-law distribution at small degrees.
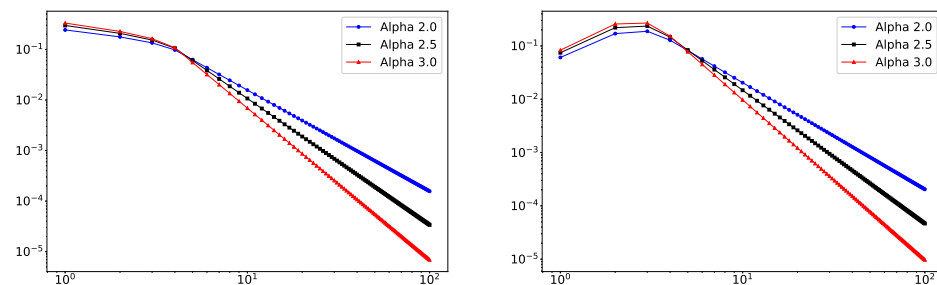
Figure 2. The pmfs of the mixture of truncated zeta distributions with $L = 4$ on a log-log scale. Mixture weights are $\mathbf{w} = (0.4, 0.3, 0.2, 0.1)$ (**left**) and $\mathbf{w} = (0.1, 0.4, 0.4, 0.1)$ (**right**). Power-law exponents are $\alpha = 2.0$ (blue), $\alpha = 2.5$ (black), and $\alpha = 3.0$ (red).

4. Estimation Algorithm

We use the Expectation-Maximization (EM) algorithm to estimate the exponent parameter α and mixture weights $\mathbf{w}$ for a given number of mixture components L. Let $\mathbf{k} = (k_1, k_2, \cdots, k_N)$ be the observed data, and z_n be the *membership* of k_n, where the membership z_n is assigned as $z_n = l$ if k_n is from the lth mixture component $TZ(\alpha, l)$. We consider the membership variable $\mathbf{z} = (z_1, z_2, \cdots, z_N)$ as missing. Let $\theta = (\alpha, \mathbf{w}) = (\alpha, w_1, \cdots, w_L)$ be the parameters of the mixture model.

The complete-data likelihood function is given by

$$\mathcal{L}(\theta|\mathbf{k}, \mathbf{z}) = \prod_{n=1}^{N} p(z_n|\mathbf{w})g(k_n|\alpha, z_n)$$
$$= \prod_{n=1}^{N} w_{z_n} \frac{1}{\zeta(\alpha, z_n)} k_n^{-\alpha}, \quad k_n = z_n, z_n + 1, \cdots.$$

We proceed by taking the logarithms of both sides to have the complete-data log-likelihood function:

$$\ln \mathcal{L}(\theta|\mathbf{k}, \mathbf{z}) = \sum_{n=1}^{N} (\ln w_{z_n} - \ln \zeta(\alpha, z_n) - \alpha \ln k_n).$$

We define $Q(\theta|\theta^*)$ as the expected value of the log-likelihood given the observed data $\mathbf{k}$ and the current parameter estimate $\theta^* = (\alpha^*, \mathbf{w}^*)$, which can be expressed as

$$Q(\theta|\theta^*) = E[\ln \mathcal{L}(\theta|\mathbf{k}, \mathbf{Z})|\mathbf{k}, \theta^*]$$
$$= \sum_{n=1}^{N} \sum_{l=1}^{L} \gamma(l, n, \theta^*)(\ln w_l - \ln \zeta(\alpha, l) - \alpha \ln k_n),$$

where $\gamma(l, n, \theta^*)$ is the *membership responsibility* of the nth observation k_n corresponding to the lth mixture component $TZ(\alpha, l)$. They are defined by the posterior probabilities of mixture component memberships for each observation,

$$
\begin{aligned}
\gamma(l, n, \theta^*) &= P(z_n = l | k_n, \theta^*) \\
&= \frac{w_l^* g(k_n | \alpha^*, l)}{\sum_{l'=1}^{L} w_{l'}^* g(k_n | \alpha^*, l')}, \quad l = 1, 2, \cdots, L, \ n = 1, 2, \cdots, N.
\end{aligned}
\tag{8}
$$

The E-step computes membership responsibilities.

In the M-step, a new parameter estimate $\check{\theta} = \text{argmax}_\theta \, Q(\theta | \theta^*)$ is computed using the quantity previously computed in the E-step. To find $\check{\mathbf{w}}$, we need to solve the following optimization problem:

$$
\text{Maximize} \quad \sum_{n=1}^{N} \sum_{l=1}^{L} \gamma(l, n, \theta^*) \ln w_l,
$$

$$
\text{Subject to} \quad \sum_{l=1}^{L} w_l = 1, \quad \text{and} \quad w_l \geq 0, \ l = 1, 2, \cdots, L.
$$

The Lagrange multiplier method yields

$$
\check{w}_l = \frac{1}{N} \sum_{n=1}^{N} \gamma(l, n, \theta^*), \quad l = 1, 2, \cdots, L.
\tag{9}
$$

Next, $\check{\alpha}$ can be found in the partial derivative of Q with respect to α, given by

$$
\frac{\partial Q(\theta | \theta^*)}{\partial \alpha} = -\sum_{n=1}^{N} \sum_{l=1}^{L} \gamma(l, n, \theta^*) \left(\frac{\zeta_\alpha(\alpha, l)}{\zeta(\alpha, l)} + \ln k_n \right).
$$

Here, $\zeta_\alpha(\alpha, l)$ is the partial derivative of the Hurwitz zeta function with respect to α, given by

$$
\zeta_\alpha(\alpha, l) = \frac{\partial \zeta(\alpha, l)}{\partial \alpha} = -\sum_{k=l}^{\infty} (\ln k) k^{-\alpha}.
$$

Then, the equation

$$
\frac{\partial Q(\theta | \theta^*)}{\partial \alpha} = 0
\tag{10}
$$

gives the desired $\check{\alpha}$. Unfortunately, a closed-form solution does not exist in general. In this paper, we employ Brent's method [34] to solve Equation (10) with respect to α.

The two steps are necessarily repeated until the convergence obtains the final parameter estimate $\hat{\theta}$. The process is summarized in Algorithm 1.

In order to select the number of mixture components L, we employ the Bayesian information criterion (BIC) considering the trade-off between the goodness-of-fit and the complexity of the model. BIC is given by the following formula:

$$
BIC = L \ln N - 2 \sum_{n=1}^{N} \ln p(k_n | \alpha, L, \mathbf{w}),
$$

where α and $\mathbf{w}$ are the obtained estimated parameters given the number of mixture components L. We choose L giving the smallest BIC, where the candidates of L are the integers from 1 to the minimum of the two values: 100 and the nearest integer to $0.90 \times$ (maximum degree).

Algorithm 1: EM Algorithm

---

input : Observed data $\mathbf{k} = (k_1, \cdots, k_N)$, the number of components L, and initial parameter estimate $\hat{\theta}^{(0)}$.

1   Initialize: $s = 0$, *converged = False*.

2   **while** *not converged* **do**

3   (E-step) Compute $\gamma(l, n, \hat{\theta}^{(s)})$ in Equation (8) for $l = 1, \cdots, L, n = 1, \cdots, N$.

4   (M-step) Find $\hat{\theta}^{(s+1)} = \text{argmax}_\theta \, \hat{Q}\left(\theta | \hat{\theta}^{(s)}\right)$ using (9) and (10).

5   **if** *convergence criteria is satisfied* **then**

6   *converged* ← *True*.

7         $\hat{\theta} \leftarrow \hat{\theta}^{(s+1)}$.

8   **end**

9      $s \leftarrow s + 1$.

10   **end**

     output: Converged parameter estimate $\hat{\theta}$.

---

5. Monte Carlo Simulation

We study the validity of the presented estimation methods using the synthetic data. We consider three values of true power-law exponents $\alpha = 2.50, 3.00, 3.50$ and also three values of the number of mixture components $L = 2, 4, 6$. For each pair of α and L, we generate 1000 samples from the finite mixture of truncated zeta distributions in Equation (7). Each process is repeated 30 times. Mixture weights $w_l, l = 1, 2, \cdots, L$ are set to $1/L$ for each dataset.

We assume that the true number of components is given. Algorithm 1 is applied to each dataset to obtain the parameter estimate $\hat{\theta} = (\hat{\alpha}, \hat{\mathbf{w}})$. Table 1 presents the parameter estimation result. We find that the average values of estimated parameters are very close to the true parameters, implying that the EM algorithm works well in estimating the exponent parameter α and weights $\mathbf{w}$.

Table 1. Estimates of θ of the EM algorithm applied to the synthetic data with the true number of mixture components. We present the average (standard deviation in parentheses) values of estimated parameters $\hat{\theta} = (\hat{\alpha}, \hat{\mathbf{w}})$ of 30 datasets for each case.

α	L	w_l	$\hat{\alpha}$	$\hat{\mathbf{w}}$
	2	0.50	$2.51_{(0.05)}$	$[0.49_{(0.02)}, 0.51_{(0.02)}]$
2.50	4	0.25	$2.51_{(0.06)}$	$[0.25_{(0.01)}, 0.25_{(0.02)}, 0.25_{(0.03)}, 0.25_{(0.02)}]$
	6	0.17	$2.50_{(0.07)}$	$[0.17_{(0.02)}, 0.16_{(0.02)}, 0.17_{(0.03)}, 0.18_{(0.04)}, 0.16_{(0.04)}, 0.16_{(0.04)}]$
	2	0.50	$3.01_{(0.09)}$	$[0.49_{(0.02)}, 0.51_{(0.02)}]$
3.00	4	0.25	$3.03_{(0.10)}$	$[0.25_{(0.01)}, 0.24_{(0.02)}, 0.25_{(0.03)}, 0.26_{(0.03)}]$
	6	0.17	$3.03_{(0.11)}$	$[0.17_{(0.01)}, 0.17_{(0.02)}, 0.17_{(0.03)}, 0.17_{(0.03)}, 0.16_{(0.03)}, 0.17_{(0.03)}]$
	2	0.50	$3.50_{(0.14)}$	$[0.50_{(0.02)}, 0.50_{(0.02)}]$
3.50	4	0.25	$3.51_{(0.13)}$	$[0.25_{(0.02)}, 0.25_{(0.02)}, 0.26_{(0.02)}, 0.25_{(0.02)}]$
	6	0.17	$3.50_{(0.13)}$	$[0.17_{(0.01)}, 0.17_{(0.01)}, 0.17_{(0.02)}, 0.17_{(0.02)}, 0.15_{(0.03)}, 0.17_{(0.03)}]$

Next, we assume that the number L of mixture components is not given. We will check whether the presented algorithm can find the appropriate number of mixture components. Note that BIC is used as a model selection criterion for the generated datasets. Table 2 shows the result of determining L. The result indicates that the maximum likelihood method yields a reasonable estimation result in finding L.

Table 2. The estimation result of L applied to the maximum likelihood method. $\hat{L}$ is selected with the smallest BIC for each dataset. The average, standard deviation, and count of $\hat{L}$ over 30 datasets are presented.

α	L	Average $\hat{L}$	St.Dev. $\hat{L}$	Count $\hat{L}$
	2	2.03	0.18	{2: 29, 3: 1}
2.50	4	4.00	0.00	{4: 30}
	6	5.87	0.34	{6: 26, 5: 4}
	2	2.00	0.00	{2: 30}
3.00	4	4.00	0.00	{4: 30}
	6	6.00	0.00	{6: 30}
	2	2.00	0.00	{2: 30}
3.50	4	4.00	0.00	{4: 30}
	6	5.97	0.18	{6: 29, 5: 1}

6. Application: Collaboration Networks
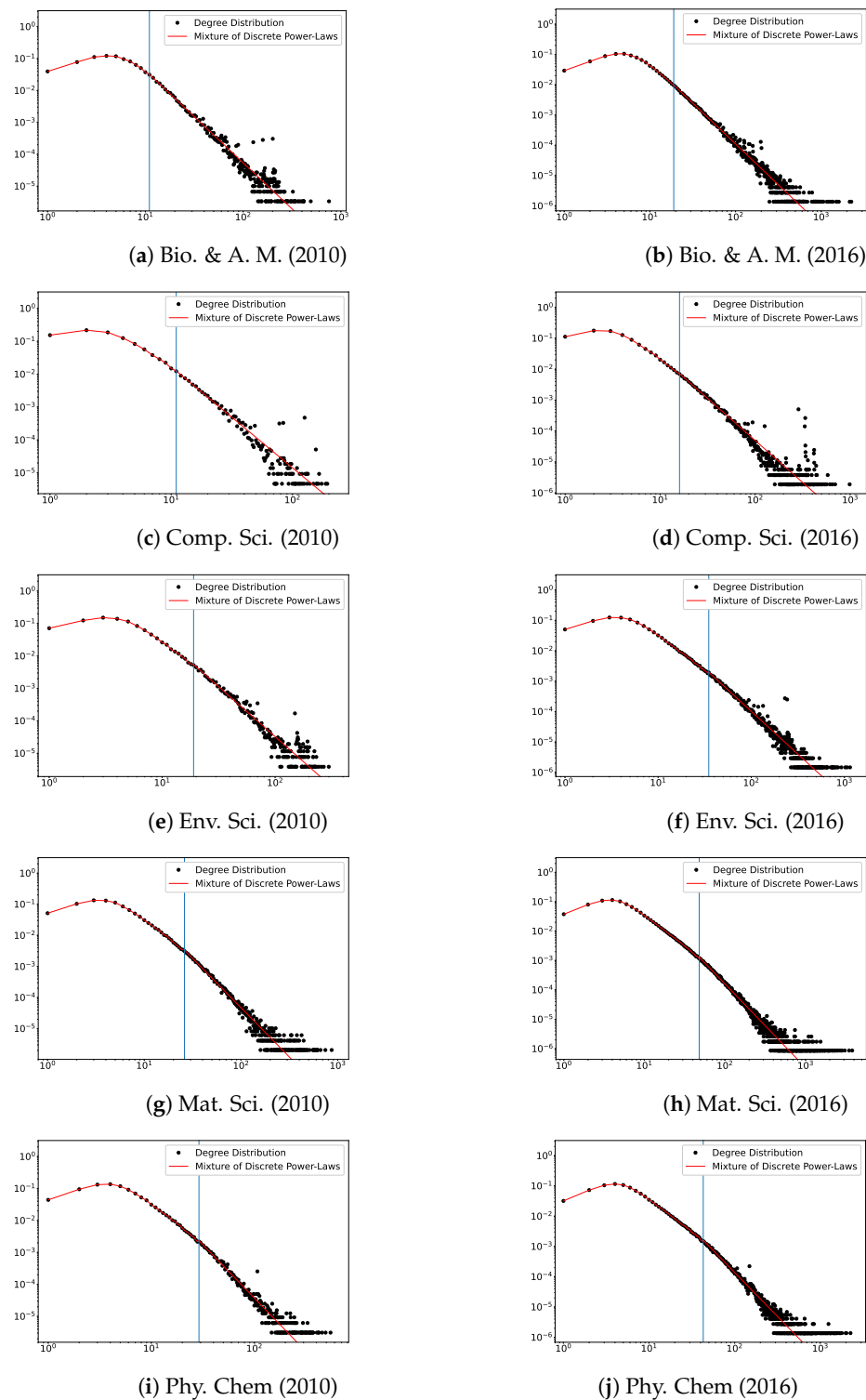
6.1. The Data

We study the scientific collaboration network obtained from the Web of Science, where a large-scale database collects the information of all published scientific articles in the world. Among all 275 disciplines, we randomly choose five, which are Biotechnology and Applied Microbiology, Computer Science, Environmental Science, Materials Science, and Physical Chemistry, for demonstration. We reorganize the Web of Science database into a network data structure such that the nodes of networks are authors, and two authors are connected by an undirected link if there is at least one paper co-authored by them. We employ the data from 2007 to 2016 as the inconsistency of author's names is observed before the year 2007. The data from 2007 to 2009 are used to accumulate data so that the dependence of node degrees can be ignorable. We analyze the degree distribution of the data from 2010 to 2016. Table 3 shows the summary statistics in the year 2016.

Table 3. The summary statistics of collaboration networks in the year 2016. For each field, we present the number of authors (nodes) and links between them. The mean, median, standard deviation, and maximum of degrees are also presented.

Field	Number of		Degree Distribution			
	Nodes	Links	Mean	Median	St.Dev.	Max.
Biotechnology & A. M.	729,478	3,977,919	10.91	7.00	19.52	2250
Computer Science	528,267	1,765,283	6.68	4.00	15.19	989
Environmental Science	680,924	3,291,780	9.67	5.00	17.42	1149
Materials Science	1,154,908	6,861,189	11.88	6.00	28.47	3801
Physical Chemistry	727,213	4,150,183	11.41	6.00	23.42	2260

### 6.2. Application of the Truncated Zeta Mixture Model

We apply the presented estimation method to degree distributions from 2010 to 2016.

In Figure 3, we plot degree distributions and the estimated fitting lines of the proposed model in the years 2001 and 2016. The plots suggest that the truncated zeta mixture shows a good performance in fitting degree distributions. The beginning points (see the vertical lines in Figure 3) of power-law behaviors are reasonably estimated via $\hat{L}$. This result shows the usefulness of the proposed model for the degree distribution that deviates from the power-law distribution on a small degree part.

(**a**) Bio. & A. M. (2010)

(**b**) Bio. & A. M. (2016)

(**c**) Comp. Sci. (2010)

(**d**) Comp. Sci. (2016)

(**e**) Env. Sci. (2010)

(**f**) Env. Sci. (2016)

(**g**) Mat. Sci. (2010)

(**h**) Mat. Sci. (2016)

(**i**) Phy. Chem (2010)

(**j**) Phy. Chem (2016)

Figure 3. Degree distributions and fitted curves for five fields of scientific collaboration in the years 2010 and 2016. The estimated L values are depicted in blue vertical lines.

Figure 4 shows the temporal changes of $\hat{L}$ and $\hat{\alpha}$. The estimated number L of mixture components exhibits a rising tendency, and $\hat{\alpha}$ shows an opposite trend. They constantly move towards large L and small α regions for all fields, indicating that the stabilization of the degree distribution has not yet been achieved. Many network scientists [1,13,14] have concentrated on the stationary or the converged degree distribution. The result implies that, however, non-stationary network models are of great importance, such as

acceleration models [15]. Moreover, many works in real data analysis have focused on the power-law exponent of a snapshot of a network. However, the temporal variation of $\hat{\alpha}$ tells us that the significance of temporal models can account for the temporal change of the power-law exponent.
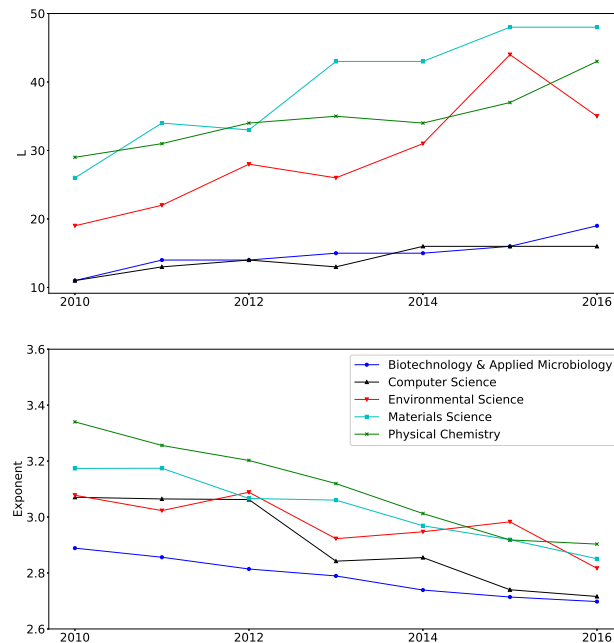
Figure 4. The change of the estimated number of mixture components L (**top**) and power-law exponents α (**bottom**).

It should be noted that $\hat{L}$ tends to approach different values across fields. According to Equation (3) and the relevant interpretation in Section 1, the number of mixture components L and mixture weights **w** are closely related to the distribution of the number of links in the system. Therefore, $\hat{L}$ could be a measure to the distribution of the number of links. On the other hand, $\hat{\alpha}$ seems to converge a value near 2.80 for all fields, suggesting that α could be a network-specific quantity instead of a field-specific quantity.

We now focus on the field of Computer Science. Table 4 presents the result of applying the model to the field of Computer Science in more detail. We can observe an interesting pattern in the estimated mixture weights, where $\hat{w}_1$ and $\hat{w}_2$ show decreasing trends whilst $\hat{w}_3$, $\hat{w}_4$, and $\hat{w}_5$ show the opposite. According to Jordan's model [29] and Equation (3), mixture weights have much to do with the number of newly made links. The decreasing trend of w_1 and w_2 and the increasing trend of $w_3, w_4, \cdots$ indicates the increasing number of links over time. As shown in Table 4, the number of created links tends to increase over time. Since there are many new links compared to new nodes (authors), the average degree gradually increased from 4.70 in 2010 to 6.68 in 2016. The increase in the average degree also explains that the state of this network is still evolving. With the rapid advancement in technology and science, many publications have been produced by researchers. In particular, Computer Science is making greater progress due to recently emerging areas: artificial intelligence and big data. The proposed model tries to explain the increasing mean degree in two ways: decreasing power-law exponents α and an appropriate change in the mixture weight, suggesting that the proposed model is helpful to describe the change of the network pattern.

Table 4. The results of $\hat{L}$, $\hat{\alpha}$, $\hat{\mathbf{w}}$, and BIC using the proposed model for the discipline of Computer Science. We also present the number of new links and nodes as well as the mean degree over time.

Year	Discrete Mixture								Number of New		Mean Deg.
	$\hat{L}$	$\hat{\alpha}$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	$\hat{w}_4$	$\hat{w}_5$	BIC	Nodes	Links	
2010	11	3.07	0.18	0.32	0.25	0.12	0.06	1,010,549	-	-	4.70
2011	13	3.06	0.17	0.30	0.25	0.12	0.06	1,241,555	43,237	132,560	4.94
2012	14	3.06	0.16	0.30	0.25	0.13	0.07	1,486,634	45,004	144,305	5.16
2013	13	2.84	0.16	0.30	0.26	0.13	0.07	1,813,101	56,392	256,646	5.78
2014	16	2.86	0.15	0.29	0.26	0.14	0.07	2,137,666	56,091	215,236	6.03
2015	16	2.74	0.15	0.29	0.26	0.14	0.07	2,481,780	57,317	277,983	6.48
2016	16	2.72	0.14	0.28	0.26	0.14	0.07	2,810,563	55,800	234,714	6.68

6.3. Comparison to Other Models

Our model is developed to deal with two essential characteristics of the degree distribution: the non-power-law behavior at small degrees and the discreteness. We study the superiority of the proposed model to base models that lack one of these characteristics.

We first consider the standard discrete power-law distribution, $TZ(\alpha, 1)$. The estimates of the power-law exponent α are obtained by fitting the data into $TZ(\alpha, 1)$, and the result is presented in Table 5. The $\hat{\alpha}$ estimates of the standard discrete power-law distribution are smaller than those of the mixture distribution. As we can observe in Figure 5, the small-degree data that deviates from the power-law makes the exponent small. The estimated fitting lines indicate that the standard discrete power-law distribution is inappropriate to describe the degree distribution.
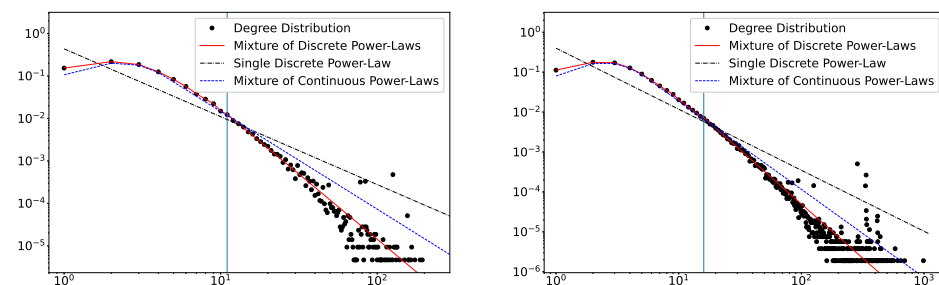
Figure 5. The degree distributions of the collaboration network data of the Computer Science field in the years 2010 (**left**) and 2016 (**right**). The estimated mixture of truncated zeta distributions (red solid), standard discrete power-law distributions (black dash-dot), and the mixture of continuous power-law distributions (blue dashed) are presented. Vertical lines refer to the estimated number $\hat{L}$ of mixture components.

Next, the degree distribution is applied to the mixture of the continuous power-law distribution using the rounding method with the constant $c = 0.5$ of the continuity correction. The continuous version of the mixture of truncated zeta distributions can be constructed by the substitution of $TZ(\alpha, l)$ to $PL(\alpha, l - c)$. In addition, we imitate the method in Section 4 for the continuous mixture to compare the performance of the continuous mixture with the discrete mixture. The estimation procedure is identical to the case of the discrete mixture model. Table 5 presents the estimation result of α and mixture weights $\mathbf{w}$. The estimates $\hat{L}$ and $\hat{\alpha}$ are much smaller in continuous mixture models due to the non-exactness of the continuous model. In addition, estimated mixture weights $\hat{\mathbf{w}}$ are considerably different from the discrete mixture. According to Equation (3), mixture weights are involved in the distribution of the number of links. Therefore, we should use the discrete version of the power-law to determine mixture weights correctly. Figure 5 shows that the fitting lines of the continuous mixture seem to deviate from empirical distributions.

As we can see in Tables 4 and 5, the smallest BIC values are achieved in the proposed model as well. Thus, we can conclude that the proposed model outperforms the continuous model as well as the standard discrete power-law model.

Table 5. The results of $\hat{L}$, $\hat{\alpha}$, $\hat{\mathbf{w}}$, and BIC using the continuous mixture model for the discipline of Computer Science. Estimated α and BIC for the standard discrete power-law distribution $TZ(\alpha, 1)$ are also presented.

| Year	Continuous Mixture						Zeta	
	$\hat{L}$	$\hat{\alpha}$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	BIC	$\hat{\alpha}$	BIC
2010	4	2.30	0.20	0.40	0.30	1,053,289	1.60	1,170,929
2011	4	2.29	0.19	0.39	0.30	1,290,015	1.59	1,436,579
2012	5	2.29	0.18	0.38	0.30	1,540,812	1.57	1,719,325
2013	5	2.26	0.17	0.37	0.30	1,871,783	1.56	2,090,369
2014	5	2.24	0.16	0.36	0.30	2,201,834	1.55	2,462,841
2015	5	2.22	0.16	0.35	0.31	2,550,130	1.54	2,852,831
2016	5	2.21	0.15	0.34	0.30	2,884,064	1.53	3,229,651

Next, we compare the proposed model with generalized Pareto distributions in which all degree ranges are covered. We use both continuous and discrete types of generalized Pareto distributions. The complementary cumulative density function of the continuous generalized Pareto distribution $GPD(\mu = 0.5, \sigma, \xi)$ is given by

$$P(X \geq x) = \left(1 + \frac{\xi(x - \mu)}{\sigma}\right)^{-1/\xi}, \quad x \geq \mu.$$

There are few studies concerning the discrete version of the generalized Pareto distribution. We here use the model in Prieto et al. (2013) [35], expressed as $DGP(\alpha, \lambda, \mu = 1)$. This distribution has advantages over the continuous distribution since the node degree is discrete. Table 6 shows the results of discrete and continuous generalized Pareto distributions applied to the field of Computer Science. In Figure 6, we plot fitting results in the years 2010 and 2016.

Table 6. The results of discrete and continuous generalized Pareto distributions applied to the field of Computer Science.

Year	Continuous Generalized Pareto			Discrete Generalized Pareto		
	$\hat{\sigma}$	$\hat{\xi}$	BIC	$\hat{\alpha}$ (10^4)	$\hat{\lambda}$ (10^{-6})	BIC
2010	3.54	0.14	1,032,317	2.51	9.53	1,043,227
2011	3.68	0.16	1,267,627	5.42	4.17	1,282,398
2012	3.83	0.16	1,517,145	5.65	3.81	1,535,652
2013	3.93	0.21	1,852,799	5.69	3.34	1,911,678
2014	4.09	0.22	2,183,341	1.40	12.92	2,249,684
2015	4.20	0.25	2,535,774	5.63	2.98	2,633,601
2016	4.31	0.26	2,871,026	3.49	4.65	2,980,178
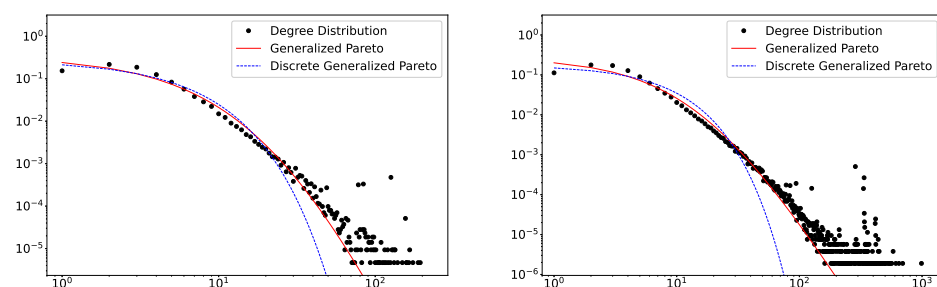
Figure 6. The degree distributions of the collaboration network data of the Computer Science field in the years 2010 (**left**) and 2016 (**right**). The estimated fitting lines of continuous (red solid) and discrete (blue dashed) generalized Pareto distributions are presented.

Interestingly, Figure 6 suggests that the continuous version has better fitting results. The discrete version has difficulty in explaining a range of large degrees. We can see that our model performs better than the two generalized Pareto distributions, as well indicated by the BIC values (see Tables 4 and 6).

7. Concluding Remark

Inspired by Jordan's model [29], a novel mixture model for the degree distribution is proposed to describe the entire range of degrees while maintaining the power-law or the scale-free property of a network. The truncated zeta distribution enables us to analyze discrete distributions for accuracy purposes. The parameter estimation procedure is presented along with the model selection criterion for determining the number of mixture components. A simulation study shows the validity of the suggested estimation procedure. The practical performance of the model is studied through the comparison analysis with the other techniques.

We perform the real data analysis on five disciplines of the scientific collaboration data obtained from the Web of Science. We observe the increasing tendency in the number of mixture components and the decreasing tendency in the power-law exponent. In addition, mixture weights change over time. It can be suggested from these results that the analyzed networks are still in an evolving state, highlighting the practical importance of non-stationary temporal network models. The non-convergence of the degree distribution might be due to the short-term analysis performed. Determining whether the collaboration network will stabilize the equilibrium remains as future work.

We can observe power-law distributions not only in the degree distribution but also in sandpile avalanches, species extinctions, city sizes, and so on. The proposed model could be useful when (i) the distribution does not follow the power-law only in small values while the power-law is suitable for large values, (ii) the background knowledge does not support the manual modification of the power-law relationship, or (iii) a mixture distribution can be regarded as reasonable for describing data.

References

1. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef] [PubMed]
2. Jung, H.; Lee, J.G.; Lee, N.; Kim, S.H. Comparison of fitness and popularity: Fitness-popularity dynamic network model. *J. Stat. Mech. Theory Exp.* **2018**, *2018*, 123403. [CrossRef]
3. Mazzarisi, P.; Barucca, P.; Lillo, F.; Tantari, D. A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. *Eur. J. Oper. Res.* **2020**, *281*, 50–65. [CrossRef]

4. Xu, K.S.; Hero, A.O. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 552–562. [CrossRef]

5. Peixoto, T.P.; Rosvall, M. Modelling sequences and temporal networks with dynamic community structures. *Nat. Commun.* **2017**, *8*, 582. [CrossRef]

6. Bianconi, G.; Barabási, A.L. Competition and multiscaling in evolving networks. *Europhys. Lett.* **2001**, *54*, 436. [CrossRef]

7. Clauset, A.; Shalizi, C.R.; Newman, M.E. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]

8. Newman, M.E. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]

9. Newman, M.E. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351. [CrossRef]

10. Albert, R.; Jeong, H.; Barabási, A.L. Diameter of the world-wide web. *Nature* **1999**, *401*, 130–131. [CrossRef]

11. Jung, H.; Lee, J.G.; Lee, N.; Kim, S.H. PTEM: A popularity-based topical expertise model for community question answering. *Ann. Appl. Stat.* **2020**, *14*, 1304–1325. [CrossRef]

12. Bollobás, B.E.; Riordan, O.; Spencer, J.; Tusnády, G. The degree sequence of a scale-free random graph process. *Random Struct. Algorithms* **2001**, *18*, 279–290. [CrossRef]

13. Dorogovtsev, S.N.; Mendes, J.F.F.; Samukhin, A.N. Structure of growing networks with preferential linking. *Phys. Rev. Lett.* **2000**, *85*, 4633. [CrossRef] [PubMed]

14. Krapivsky, P.L.; Redner, S. Organization of growing random networks. *Phys. Rev. E* **2001**, *63*, 066123. [CrossRef] [PubMed]

15. Dorogovtsev, S.N.; Mendes, J.F.F. Effect of the accelerating growth of communications networks on their structure. *Phys. Rev. E* **2001**, *63*, 025101. [CrossRef]

16. Fotouhi, B.; Rabbat, M.G. Degree correlation in scale-free graphs. *Eur. Phys. J. B* **2013**, *86*, 510. [CrossRef]

17. Yu, X.; Li, Z.; Zhang, D.; Liang, F.; Wang, X.Y.; Wu, X. The topology of an accelerated growth network. *J. Phys. A Math. Gen.* **2006**, *39*, 14343. [CrossRef]

18. Jung, S.; Kim, S.; Kahng, B. Geometric fractal growth model for scale-free networks. *Phys. Rev. E* **2002**, *65*, 056101. [CrossRef] [PubMed]

19. Dorogovtsev, S.N.; Mendes, J.F. Accelerated growth of networks. *arXiv* **2002**, arXiv:cond-mat/0204102.

20. Pham, T.; Sheridan, P.; Shimodaira, H. Joint estimation of preferential attachment and node fitness in growing complex networks. *Sci. Rep.* **2016**, *6*, 32558. [CrossRef] [PubMed]

21. Jung, H.; Lee, J.G.; Kim, S.H. On the analysis of fitness change: Fitness-popularity dynamic network model with varying fitness. *J. Stat. Mech. Theory Exp.* **2020**, *2020*, 043407. [CrossRef]

22. Kumar, R.; Raghavan, P.; Rajagopalan, S.; Sivakumar, D.; Tomkins, A.; Upfal, E. Stochastic models for the web graph. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, 12–14 November 2000; pp. 57–65.

23. Prieto, F.; Sarabia, J.M. A generalization of the power law distribution with nonlinear exponent. *Commun. Nonlinear Sci. Numer. Simul.* **2017**, *42*, 215–228. [CrossRef]

24. Arnold, B.C. *Pareto Distributions*; Chapman and Hall/CRC: New York, NY USA, 2015.

25. Meibom, A.; Balslev, I. Composite power laws in shock fragmentation. *Phys. Rev. Lett.* **1996**, *76*, 2492. [CrossRef] [PubMed]

26. Garcıa, F.; Garcıa, R.; Padrino, J.; Mata, C.; Trallero, J.; Joseph, D. Power law and composite power law friction factor correlations for laminar and turbulent gas–liquid flow in horizontal pipelines. *Int. J. Multiph. Flow* **2003**, *29*, 1605–1624. [CrossRef]

27. Fenner, T.; Levene, M.; Loizou, G. A model for collaboration networks giving rise to a power-law distribution with an exponential cutoff. *Soc. Netw.* **2007**, *29*, 70–80. [CrossRef]

28. Mossa, S.; Barthelemy, M.; Stanley, H.E.; Amaral, L.A.N. Truncation of power law behavior in "scale-free" network models due to information filtering. *Phys. Rev. Lett.* **2002**, *88*, 138701. [CrossRef]

29. Jordan, J. The degree sequences and spectra of scale-free random graphs. *Random Struct. Algorithms* **2006**, *29*, 226–242. [CrossRef]

30. Barabási, A.L.; Albert, R.; Jeong, H. Mean-field theory for scale-free random networks. *Phys. A* **1999**, *272*, 173–187. [CrossRef]

31. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47. [CrossRef]

32. Jung, H.; Phoa, F.K.H. Analysis of a Finite Mixture of Truncated Zeta Distributions for Degree Distribution. In *Studies in Computational Intelligence, Proceedings of the International Conference on Complex Networks and Their Applications, Madrid, Spain, 1–3 December 2020*; Springer: Cham, Switzerland, 2020; pp. 497–507.

33. Gillespie, C. Fitting Heavy Tailed Distributions: The poweRlaw Package. *J. Stat. Softw.* **2015**, *64*, 1–16. [CrossRef]

34. Brent, R.P. *Algorithms for Minimization without Derivatives*; Dover: New York, NY USA, 2013.

35. Prieto, F.; Gómez-Déniz, E.; Sarabia, J.M. Modelling road accident blackspots data with the discrete generalized Pareto distribution. *Accid. Anal. Prev.* **2014**, *71*, 38–49. [CrossRef] [PubMed]