X-distribution: retraceable power-law exponent of complex networks

Pandy, P.K.; Arya, A.; Saxena, A.

X-distribution: Retraceable Power-law Exponent of Complex Networks

PRADUMN KUMAR PANDEY and AIKTA ARYA, Indian Institute of Technology, Roorkee, India
AKRATI SAXENA, LIACS, Leiden University, The Netherlands

Network modeling has been explored extensively by means of theoretical analysis as well as numerical simulations for Network Reconstruction (NR). The network reconstruction problem requires the estimation of the power-law exponent (γ) of a given input network. Thus, the effectiveness of the NR solution depends on the accuracy of the calculation of γ. In this article, we re-examine the degree distribution-based estimation of γ, which is not very accurate due to approximations. We propose **X**-distribution, which is more accurate than degree distribution. Various state-of-the-art network models, including CPM, NRM, RefOrCite2, BA, CDPAM, and DMS, are considered for simulation purposes, and simulated results support the proposed claim. Further, we apply **X**-distribution over several real-world networks to calculate their power-law exponents, which differ from those calculated using respective degree distributions. It is observed that **X**-distributions exhibit more linearity (straight line) on the log-log scale than degree distributions. Thus, **X**-distribution is more suitable for the evaluation of power-law exponent using linear fitting (on the log-log scale). The MATLAB implementation of power-law exponent (γ) calculation using **X**-distribution for different network models and the real-world datasets used in our experiments are available at https://github.com/Aikta-Arya/X-distribution-Retraceable-Power-Law-Exponent-of-Complex-Networks.git.

CCS Concepts: • **Information systems** → **Data mining**; *Data analytics;*

Additional Key Words and Phrases: Degree distribution, **X**-distribution, power law, scale-free networks, network reconstruction, network modeling

ACM Reference Format:
Pradumn Kumar Pandey, Aikta Arya, and Akrati Saxena. 2024. X-distribution: Retraceable Power-law Exponent of Complex Networks. *ACM Trans. Knowl. Discov. Data.* 18, 5, Article 117 (February 2024), 12 pages.
https://doi.org/10.1145/3639413

1 INTRODUCTION

Networked systems are ubiquitous in nature, for example, transportation networks [6, 14], social networks [13], biological networks [12], and communication networks [21, 22], which are analyzed using graphs or networks to understand their complex dynamics. In the past two decades, the problem of *Structural Reconstruction* of real-world networks has received a lot of attention. The structural reconstruction of a real-world network is concerned with the reconstruction of a given

This work is supported by The Department of Science and Technology, Government of India, sponsored project under Grant No. DST-1401-CSE.

Authors' addresses: P. K. Pandey and A. Arya, Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, India- 247667; e-mails: {pradumn.pandey, a_arya}@cs.iitr.ac.in; A. Saxena, LIACS, Leiden University, The Netherlands; e-mail: a.saxena@liacs.leidenuniv.nl.

ACM 1556-4681/2024/02-ART117
https://doi.org/10.1145/3639413

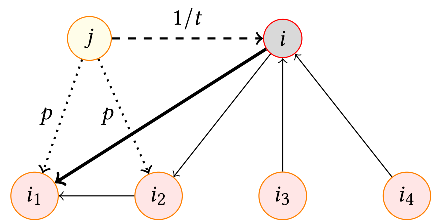


Fig. 1. Network evolution dynamics of CPM. A newly inserted node j at time t connects to an already existing node i with the given probability $1/t$ and establishes links with neighbors of node i with probability p.

network by using both a network model and limited information about the network [1]. The reconstruction means that the generated network should possess the same collective spectral and structural properties as the input real-world network. In literature, various network-generating models have been proposed to understand and study the evolution process of real-world networks, which exhibit various patterns and properties of real-world networks, such as degree distribution, clustering, triangle formation, and small-world phenomena [2, 7, 12, 18, 21]. These proposed models are used to generate synthetic networks that look alike real-world networks and are broadly used to understand network evolution and dynamic processes taking place on these networks, such as influence propagation, opinion formation, anomaly detection, and so on.

The first very-well-known network model in this direction is the **Barabási–Albert (BA)** model [3], in which each new node makes $\overline{k}$ connections with the existing nodes, and the probability of connecting with an existing node is directly proportional to its degree. This leads to the rich-get-richer phenomenon, and the degree distribution of the generated network follows a power law, i.e., approximated as $p(k) = c \cdot k^{-\gamma}$. After this, there have been proposed several models, including fitness model [5], triad-formation model [15], local-world model [19], mutual attraction model [28], copying model [16], **Network Reconstruction Model (NRM)** [24], RefOrCite2 Model [25], **Context Dependent Preferential Attachment Model (CDPAM)** [23], **Dorogovtsev, Mendes, and Samukhin (DMS)** model [10], and so on [26]. All these existing network-generation models primarily focus on the network's degree distribution so that the generated network follows the expected power-law degree distribution.

In the network reconstruction process of scale-free networks, estimating the power-law exponent of a given real-world network is required [24]. The novelty of network reconstruction solutions depends on the accuracy of power-law exponent calculation. Most of the state-of-the-art network models follow the power law if they use approximation, which may result in an error-prone estimation of the power-law exponent. The considered approximations in different models provide that model-generated networks follow power law in their tail only (high degree nodes) [9, 10, 23, 24].

Motivation: We consider the **copying model (CPM)** [16], in which nodes appear in a sequence one by one. A newly appeared node j selects an older (existing) node i uniformly randomly, and then j connects neighbors (via outgoing edges) of node i with probability p; see Figure 1. The power-law exponent for CPM is $\gamma = 1/p$. By setting $p \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.12, 0.15\}$, we simulate networks of size $n = 10^5$. Using the degree distributions of simulated networks, the calculated values of γ are $\{4.9, 3.5, 2.8, 2.2, 2.9, 1.7, 1.5, 1.3, 1.3, 4.5, 4.1\}$ corresponding to the selected values of parameter p. But the expected values of γ for the selected values of p should be $\{10.0, 5.0, 3.3, 2.5, 2, 1.7, 1.4, 1.25, 1.1, 8.3, 6.7\}$. There is a significant deviation in the values of γ of simulated networks as compared to their expected values.

This motivates us to re-investigate the degree distribution for other models that are used for structural reconstruction. If degree distribution is not capable enough to be used for computation of the parameter γ, then another metric or variable, similar to degree, is required to calculate γ. Apart from that, in the literature, it is shown that the various network growth processes follow power-law degree distributions in the tail with the condition that the size of the network is very large [9, 10, 23, 24]. Thus, it is essential to define a new property to evaluate the value of power-law exponent γ of a given network more accurately, and it is expected to be more consistent with the change in the size of the networks. In this article, a variable X_i for node i is considered that is derived from the degree of the node and a constant. For various growing scale-free networks, X_i follows scale-free (power-law) distribution for $X_i > 0$; in the case of their respective degree distributions, it follows scale-free (power-law) distribution for higher values ($k_i >>> 1$). A novel method for more accurate power-law exponent computation is proposed based on the distribution of X in a growing scale-free network under a given model or growth dynamics. The proposed method is compared with the degree distribution-based power-law exponent computation method proposed in Reference [8].

Contributions: This article makes the following contributions:

— In this article, **X**-distribution (a derivative of degree) is defined, which is more accurate and consistent in calculating the power-law exponent of given networks.
— Extensive experimentation over different state-of-the-art network models, including CPM [16], NRM [24], RefOrCite2 Model [25], BA [3], CDPAM [23], and DMS model [10], exhibits novelty of **X**-distribution. We also apply our proposed algorithm successfully to calculate power-exponents of **X**-distribution for various real-world networks and compare with the degree distribution-based method.

The rest of the article is organized as follows: Section 2 is dedicated to discussing the limitation of degree distribution and the definition of **X**-distribution. An algorithm is proposed to calculate the power-law exponent γ for a given network. In Section 3, **X**-distribution and degree distribution are applied to retrace the microdynamics (γ) of the networks obtained under CPM, NRM, RefOrCite2, BA, CDPAM, and DMS models. The comparative analysis of degree distribution and **X**-distribution indicates the superiority of **X**-distribution in the estimation of γ more accurately and consistently. Finally, the work is concluded in Section 4.

2 X-DISTRIBUTION

Degree distribution to X-distribution: Here, we discuss the way we define **X**-distribution using the degree of nodes and its advantages over degree distribution.

We consider copying the model in References [4, 16] to explain **X**-distribution. Let us assume that $k_i^{\text{in}}(t)$, $k_i^{\text{out}}(t)$, and $k_i(t)$ ($= k_i^{\text{in}} + k_i^{\text{out}}$) be the in-degree, out-degree, and degree of node i, respectively, at time t. The growth in the degree of node i can happen in two ways: either a new coming node j gets attached with node i with probability $\frac{1}{t}$ directly (Figure 1), or node j first gets connected with one of the neighbors (nodes of incoming edges $\mathcal{N}_i$) of node i and then to node i with probability p ($i.e.\ p\frac{1}{t}$); see Figure 2. Thus,

$$\frac{dk_i(t+1)}{dt} = \frac{1}{t} + \left(1 - \frac{1}{t}\right) \sum_{l \in \mathcal{N}_i} p\frac{1}{t} = \frac{1 + pk_i}{t} - \frac{pk_i}{t^2}. \tag{1}$$

By mean-field approximation,

$$\frac{1}{p} \int \frac{dpk_i(t)}{1 + pk_i(t)} = \int \frac{dt}{t}.$$

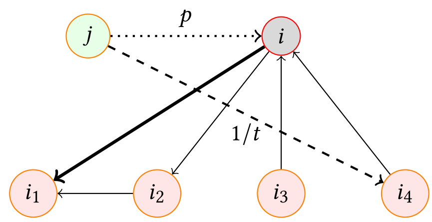


Fig. 2. Node i also gets new connection when j gets connected with neighbors (incoming) of node i under evolution dynamics of CPM. A node j newly introduced at time t connects to an older base node i_4 with probability $1/t$ and then gets connected with one of the first neighbors (out-links only) of node i_4 with probability p.

Asserting boundary condition $k_i(t_i) = k_i^{\text{out}}(t) = k_i^0$,

$$\ln \frac{k_i(t+1)p+1}{k_i^0 p+1} = p \ln \frac{t+1}{t_i},$$
$$\frac{k_i(t+1)+1/p}{k_i^0+1/p} = \left(\frac{t+1}{t_i}\right)^p.$$

For $k_i(t)$ to exceed k, we need

$$t_i < (t+1)(k+1/p)^{-1/p}(k_i^0+1/p)^{1/p}. \quad (2)$$

Since nodes arrive uniformly, we have

$$\Pr(k_i > k) \sim (k+1/p)^{-1/p}(k_i^0+1/p)^{1/p}, \quad (3)$$

where $\lim_{t\to\infty} k_i(t) \to k_i$.

Thus, the degree distribution closely follows a power law with a dependency on the initial degree, and this dependency leads to approximation and more error in curve fitting while retracing the model parameters. To work around the initial condition, we consider a variable

$$X_i = \frac{k_i(t+1)+1/p}{k_i^0+1/p} \quad (4)$$

instead of degree $k_i(t+1)$, the event $X_i > x$ corresponds to $(t/t_i)^p > x$, or $t_i < tx^{-1/p}$, implying that

$$\Pr(X_i > x) = x^{-1/p},$$

a perfect power law, and minimizes error in retracing the model parameters using curve fitting.

X-distribution: Now, we define

$$X_i = \frac{k_i + C}{k_i^0 + C}, \quad (5)$$

where C is a constant. So, the distribution of the variable X_i is called **X**-distribution. If we compare Equations (4) and (5), constant C depends on model parameters. Due to the mean-field approximation made on Equation (1), $C = \gamma = 1/p$ as $t \longrightarrow \infty$. Thus, for the networks of limited sizes obtained using model (1), the value of C can differ from γ and $1/p$.

For different models that are working on the framework of the BA model, we can get constant C and γ using the following comparative analysis:

$$\frac{dk_i(t+1)}{dt} = \frac{1}{\gamma}\left(\frac{k_i(t)+C}{t}\right),$$
$$\Pr(X_i > x) = x^{-\gamma}. \tag{6}$$

If the growth equation of a model can be written in the form of Equation (6), then we can get the value of γ in terms of model parameters.

Equation (6) produces better approximation than Equation (3), thus an algorithm (Algorithm 1) is proposed to calculate γ more accurately using **X**-distribution. Algorithm 1 is divided into four blocks, namely, **B(I)**, **B(II)**, **B(III)**, and **B(IV)**. First block **B(I)** does the initialization of variable C, which varies from 0.001 to 50 in the interval of 0.01. For each value of C (**for** loop in line 5), values of X_i in block **B(II)** and the cumulative frequency of X_i ($Y1_i$) corresponding to unique values of X_i are calculated in **B(III)**, and finally, linear fitting on the log-log scale and error estimation is done in **B(IV)** using MATLAB functions polyfit[1] and polyval.[2] Meanwhile, γ is the negative slope of the linear fitting (line 18 in Algorithm 1). Algorithm 1 reports γ (in line 20) corresponding to the minimum error.

ALGORITHM 1: PLE: Power-law Exponent

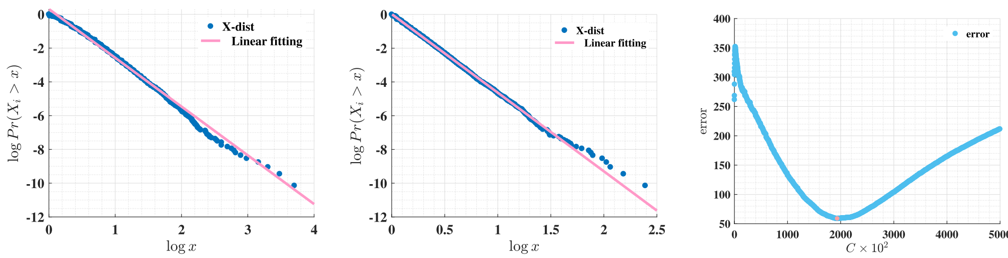
Input: Edge-list E_r of a given network $G_r = (V_r, E_r)$.

Output: Value of γ (power-law exponent).

```
1:  Procedure PLE: Fitting
2:  Let E_r be the edge-list of the network G_r.
3:  k_i^in and k_i^out are in-degree and out-degree of node i.
4:  C = 0.001 : .01 : 50; n = |V_r|;                          B(I)
5:  for all j = 1 : length(C) do
6:      for all i = 1 : n do
7:          X_i = (k_i + C_j) / (k_i^out + C_j);
8:      end for                                                B(II)
9:      X1 = sort(unique(X))
10:     for all i = 1 : length(X1) do
11:         Y1_i = length(find(X ≥ X1_i));
12:     end for                                                B(III)
13:     para = polyfit(log(X1), log(Y1), 1);
14:     Y = polyval(para, log(X1));
15:         If (sum(abs(Y-log(Y1))) ≤ error)
16:             error = sum(abs(Y-log(Y1)));
17:             tempC = C_j;
18:             γ = −para(1);                                  B(IV)
19: end for
20: return γ
```

Now, we consider a network dataset to understand the implementation of Algorithm 1. **Process:** In the first step, the algorithm does the calculation of X_i for all the nodes in the considered

[1] https://in.mathworks.com/help/matlab/ref/polyfit.html
[2] https://in.mathworks.com/help/matlab/ref/polyval.html



(a) **X**-distribution (b) **X**-distribution (c) Errors in fitting of **X**-distributions with respect to C.

Fig. 3. (a) For $C = 4.9610$, **X**-distribution is calculated. Its linear fitting (on log-log scale) is done to calculate the power-law exponent, $\gamma = 2.8885$. The error in fitting is 219.3640. (b) For $C = 19.7710$, **X**-distribution is calculated. Its linear fitting (on log-log scale) is done to calculate the power-law exponent, $\gamma = 4.68$. The error in fitting is 58.9. (c) Error in linear-fitting of **X**-distribution (on a log-log scale) of a real-world network (Supreme court) (in blue dots) is plotted for different considered values of $C = 0.001 : 0.01 : 50$. The minima of the pattern (pink dot) is identified to get the value of C, which is expected to produce the best linear fitting (on log-log scale) of **X**-distribution, which is shown in subfigure (b).

network for a value of constant C (Box **B(II)** in Algorithm 1), let $C = 4.9610$. Next, we calculate **X**-distribution (in Box **B(III)**), then we perform the linear fitting (on log-log scale) of the obtained **X**-distribution (in Box **B(IV)**, fitting is shown in Figure 3(a)) and calculate the error in the fitting. Repeat the explained process **Process** for different considered values of $C = 0.001 : 0.01 : 50$. In **B(IV)** (lines 15–18) stores the values of error and corresponding power-law exponent if the error corresponding to the current value of C is less than previously explored values of C. Finally, after the completion of the execution of Algorithm 1, we obtain the power-law exponent $\gamma = 4.68$ corresponding to the best linear fitting (on log-log scale) of **X**-distribution (shown in Figure 3(b)). The errors in the linear fitting of **X**-distributions corresponding to different values of C are plotted in Figure 3(c).

3 SIMULATION AND RESULTS

3.1 Data

Here, we consider the following network models to verify the superiority of **X**-distribution over degree distribution in calculating power-law exponent γ (Table 1): CPM [16], NRM [24], RefOrCite2 [25], BA [3], CDPAM [23], and DMS [10]. These are the state-of-the-art network models utilized for the structural reconstruction of real-world networks, in which at each time-step a new node appears and get attached to the older nodes according to the predefined rules of the respective models. The way to compute X_i for a growing network model is in Algorithm 1. We also consider various real-world networks, for example, Biomedical, Supreme court, ArxivTH, ArxivPH, Patent, and Facebook (refer to Table 2), and power-law exponents are calculated using **X**-distributions and respective degree distributions. The experimental computations are performed on Intel Xeon Gold 5120 dual CPU equipped with 128 GB RAM configuration system. Furthermore, Matlab implementations (using MATLAB R2022b software) of diverse network models are used to generate networks for experimental analysis.

3.2 Effectiveness of X-Distribution

This section discusses the experimental analysis of **X**-distribution through Algorithm 1 for various considered network models to show the effectiveness of the proposed **X**-distribution. For the

Table 1. Networks of Size 10^5 Nodes Under CPM, NRM, RefOrCite2, BA, CDPAM, and DMS Models Are Considered

<table>
<tr><th></th><th>γ</th><th colspan="2">$\gamma(X)$</th><th colspan="3">$\gamma(D)$</th></tr>
<tr><td rowspan="11">CPM</td><td>1.11</td><td colspan="2">1.108 ± 0.0335</td><td colspan="3">1.2533 ± 0.0264</td></tr>
<tr><td>1.25</td><td colspan="2">1.1918 ± 0.0111</td><td colspan="3">1.3406 ± 0.0264</td></tr>
<tr><td>1.43</td><td colspan="2">1.3633 ± 0.0181</td><td colspan="3">1.4938 ± 0.0295</td></tr>
<tr><td>1.67</td><td colspan="2">1.5739 ± 0.0253</td><td colspan="3">1.6930 ± 0.0200</td></tr>
<tr><td>2.00</td><td colspan="2">1.8918 ± 0.0418</td><td colspan="3">1.8559 ± 0.0309</td></tr>
<tr><td>2.50</td><td colspan="2">2.3679 ± 0.0771</td><td colspan="3">2.2452 ± 0.0959</td></tr>
<tr><td>3.30</td><td colspan="2">3.1549 ± 0.1362</td><td colspan="3">2.7707 ± 0.1593</td></tr>
<tr><td>5.00</td><td colspan="2">4.6118 ± 0.4223</td><td colspan="3">3.5281 ±0.2973</td></tr>
<tr><td>6.67</td><td colspan="2">6.349 ± 0.9251</td><td colspan="3">4.1296 ± 0.4066</td></tr>
<tr><td>8.33</td><td colspan="2">7.5686 ± 1.5666</td><td colspan="3">4.4651 ± 0.4232</td></tr>
<tr><td>10.00</td><td colspan="2">9.1658 ± 1.7703</td><td colspan="3">4.8878 ± 0.5612</td></tr>
<tr><td rowspan="6">NRM</td><td>1.33</td><td colspan="2">1.4103 ± 0.8030</td><td colspan="3">0.8707 ± 0.2851</td></tr>
<tr><td>1.72</td><td colspan="2">1.9124 ± 0.0588</td><td colspan="3">2.1771 ± 0.0499</td></tr>
<tr><td>2.27</td><td colspan="2">2.3436 ± 0.0746</td><td colspan="3">2.3249 ± 0.0895</td></tr>
<tr><td>3.57</td><td colspan="2">3.8684 ± 0.3082</td><td colspan="3">3.2931 ± 0.1313</td></tr>
<tr><td>5.26</td><td colspan="2">5.6594 ± 0.5490</td><td colspan="3">4.0311 ± 0.3263</td></tr>
<tr><td>10.26</td><td colspan="2">11.6199 ± 3.1616</td><td colspan="3">5.5043 ± 0.7103</td></tr>
<tr><td rowspan="11">RefOrCite2</td><td>1.11</td><td colspan="2">1.3226 ± 0.0024</td><td colspan="3">0.6909 ± 0.0492</td></tr>
<tr><td>1.25</td><td colspan="2">1.3762± 0.0124</td><td colspan="3">1.0242 ± 0.0638</td></tr>
<tr><td>1.43</td><td colspan="2">1.4581 ± 0.0103</td><td colspan="3">1.3097 ± 0.0637</td></tr>
<tr><td>1.67</td><td colspan="2">1.5833 ± 0.0165</td><td colspan="3">1.5627 ± 0.0645</td></tr>
<tr><td>2.00</td><td colspan="2">1.9729 ± 0.0321</td><td colspan="3">1.8303 ± 0.0855</td></tr>
<tr><td>2.50</td><td colspan="2">2.5095 ± 0.0303</td><td colspan="3">2.1749 ± 0.1282</td></tr>
<tr><td>3.30</td><td colspan="2">3.4217 ± 0.1190</td><td colspan="3">2.6272 ± 0.1875</td></tr>
<tr><td>5.00</td><td colspan="2">5.2578 ± 0.4249</td><td colspan="3">3.2503 ± 0.3531</td></tr>
<tr><td>6.67</td><td colspan="2">6.7698 ± 0.7115</td><td colspan="3">3.6699 ± 0.4457</td></tr>
<tr><td>8.33</td><td colspan="2">8.3429 ± 1.1467</td><td colspan="3">4.1082 ± 0.5815</td></tr>
<tr><td>10.00</td><td colspan="2">9.9375 ± 1.9204</td><td colspan="3">4.1476 ± 0.5886</td></tr>
<tr><td></td><td>γ</td><td>$\gamma(X, 10)$</td><td>$\gamma(D, 10)$</td><td colspan="2">$\gamma(X, 20)$</td><td>$\gamma(D, 20)$</td></tr>
<tr><td>BA</td><td>2.0</td><td>1.942</td><td>1.914 ± 0.0116</td><td colspan="2">1.9629</td><td>1.8985 ± 0.0033</td></tr>
<tr><td rowspan="6">CDPAM</td><td>1.00</td><td>1.4144 ± 0.0775</td><td>1.3601 ± 0.5103</td><td colspan="2">0.7845 ± 0.05</td><td>2.4621 ± 0.1468</td></tr>
<tr><td>1.33</td><td>1.0478 ± 0.0477</td><td>1.0417 ± 0.0774</td><td colspan="2">0.9 ± 0.005</td><td>0.8898 ± 0.0436</td></tr>
<tr><td>1.60</td><td>1.4927 ± 0.0250</td><td>1.8979 ± 0.0039</td><td colspan="2">1.5227 ± 0.015</td><td>2.2263 ± 0.0818</td></tr>
<tr><td>1.82</td><td>1.7552 ± 0.005</td><td>1.8143 ± 0.0158</td><td colspan="2">1.8059 ± 0.003</td><td>1.8665 ± 0.0073</td></tr>
<tr><td>1.91</td><td>1.8292 ± 0.005</td><td>1.8744 ± 0.0078</td><td colspan="2">1.8934 ± 0.005</td><td>1.9078 ± 0.0046</td></tr>
<tr><td>1.99</td><td>1.9186 ± 0.005</td><td>1.9329 ± 0.0108</td><td colspan="2">1.9579 ± 0.005</td><td>1.9229 ± 0.0050</td></tr>
<tr><td></td><td>γ</td><td>$\gamma(X, 10)$</td><td>$\gamma(D, 10)$</td><td>γ</td><td>$\gamma(X, 20)$</td><td>$\gamma(D, 20)$</td></tr>
<tr><td rowspan="4">DMS</td><td>2.01</td><td>1.9456 ± 0.0175</td><td>1.9405 ± 0.0114</td><td>2.01</td><td>1.9715 ± 0.012</td><td>2.2384 ± 0.0756</td></tr>
<tr><td>3.00</td><td>2.8710 ± 0.0175</td><td>2.2702 ± 0.0267</td><td>2.50</td><td>2.3641 ± 0.001</td><td>1.8979 ± 0.0031</td></tr>
<tr><td>4.00</td><td>3.9326 ± 0.0175</td><td>2.4381 ± 0.0288</td><td>3.00</td><td>2.87 ± 0.01</td><td>1.8996 ± 0.0032</td></tr>
<tr><td>12.00</td><td>7.65 ± 0.0175</td><td>2.8706 ± 0.0363</td><td>7.00</td><td>5.001 ± 0.010</td><td>1.9120 ± 0.0026</td></tr>
</table>

The values in $\gamma(X)$ column are calculated using the **X**-distribution and the values in $\gamma(D)$ column are calculated using the Degree distribution.

simulation purpose and to cover the wide range of γ, we set the parameter values of different models: parameter p in CPM is set to 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.15, 0.12, 0.1, (p, β) in NRM are set to (0.5, 0.5), (0.3, 0.4), (0.2, 0.3), (0.1, 0.2), (0.1, 0.1), (0.05, 0.05), in RefOrCite2, $p = 0.4$ and q is set to 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.15, 0.12, 0.1.

Table 2. Brief Descriptions of Datasets with Nodes and Edges

Networks	Description	Nodes	Edges	Ref
Biomedical	Consists of biomedical papers indexed in NCBI (2001–2008)	43,937	162,404	[25]
Supreme Court	US Supreme Court cases (1754–2002). Judgements refer to previous judgements	25,417	446,490	[11]
ArxivTH	High Energy Physics—Theory papers from arXiv.org (1992–2002)	27,770	352,807	[17]
ArxivPH	High Energy Physics—Phenomenology papers from arXiv.org (1992–2002)	34,546	421,578	[17]
Patent	Citation network among U.S. Patents	3,774,768	16,518,948	[17]
Facebook	Network of posts to other user's wall on Facebook	46,952	876,993	[27]

Table 3. Power-law Exponent (γ) of Different State-of-the-art Network Models

Model	CPM	NRM	RefOrCite2	BA	CDPAM	DMS
γ	$\frac{1}{p}$	$\frac{1}{\beta+(1-\beta)p}$	$\frac{1}{q}$	2	$\frac{2\beta}{\beta+0.5}$	$2+\frac{\beta}{\overline{k}}$

In BA, CDPAM, and DMS models, the average degree ($\overline{k}$) is also an input parameter. We set $\overline{k} \in \{10, 20\}$ for experimental simulations. $\gamma(X, 10)$ represents value of γ at $\overline{k} = 10$ for **X**-distribution. The CDPAM model set parameter β to 0.5, 1, 2, 5, 10, 1000 and $\overline{k} \in \{10, 20\}$, in DMS model parameter β set to 0.1, 10, 20, 100 and $\overline{k} \in \{10, 20\}$.

For the given settings of parameter values of different models, 100 networks are simulated of the size of 10^5 nodes. Then, power-law exponents $\gamma(X)$ and $\gamma(D)$ are calculated using **X**-distributions and degree distributions, respectively. The mean values with standard deviation are reported in Table 1. In Table 1, γ is the theoretical value corresponding to input parameters. The mathematical formulation for computing γ is given in Table 3. Furthermore, the numerically simulated values ($\gamma(X)$ or $\gamma(D)$) close to theoretical γ correspond to a more accurate estimation of the power-law exponent. Values written in bold are closer to theoretical γ. From Table 1, it is observed that **X**-distribution outperforms in most of the cases. The improvement is marginal whenever degree distribution exhibits improved results, but significant improvements are noticed in the case of **X**-distribution.

For pictorial verification, degree distributions and **X**-distributions of considered models are plotted in Figure 4 on log-log scale. **X**-distributions plots (in dark green hexagon) are more linear than respective degree distributions (plots in pink color squares). The extensive experimental results support the claim that **X**-distribution can estimate power-law exponent more accurately compared to respective degree distribution.

We also calculate ***goodness of fit*** **(GoF)** using cost function **Mean-squared Error (MSE)**, and ***Two-sample Kolmogorov-Smirnov Test*** **(*K2*)** [20] to evaluate the quality of fitting of **X**-distribution and degree distribution of networks obtained under different network models (considered in Figure 4) and real-world networks (in Table 2), and noted in Table 4. Lower values (values in bold in Table 4) of GoF and $K2$ signify better curve fitting. From Table 4, it is observed that **X**-distribution of networks exhibit better curve fitting with lower values of GoF and $K2$ than corresponding degree distributions.

3.3 Consistency of X-distribution

In this section, we discuss the stability and consistency of the proposed algorithm for calculating γ. It has already been mentioned that most of the models follow the power law in their tail (for higher values of degree). Thus, the size of the network plays a critical role in the estimation of γ. Here, two networks of different sizes (having 10^5 and 10^6 nodes) are generated using CPM, NRM, RefOrCite2, BA, CDPAM, and DMS models, and their **X**-distribution and degree distributions are plotted in Figure 5. The degree distribution of a model network follows the power law in its tail,

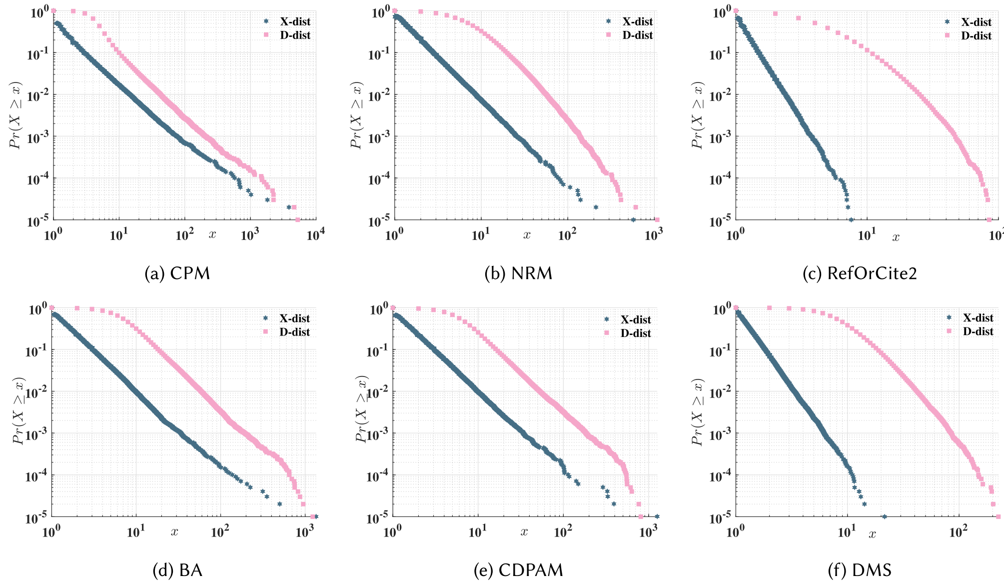


Fig. 4. (Best viewed in color.) **X**-distributions (**X-dist**) and Degree distributions (**D-dist**) for (a) CPM ($p = 0.7$), (b) NRM ($p = 0.4, \beta = 0.2$), (c) RefOrCite2 ($p = 0.4, \ q = 0.2$), (d) BA ($p = 5$), (e) CDPAM ($\beta = 5, \overline{k} = 5$), and (f) DMS ($\beta = 10, \overline{k} = 5$) models.

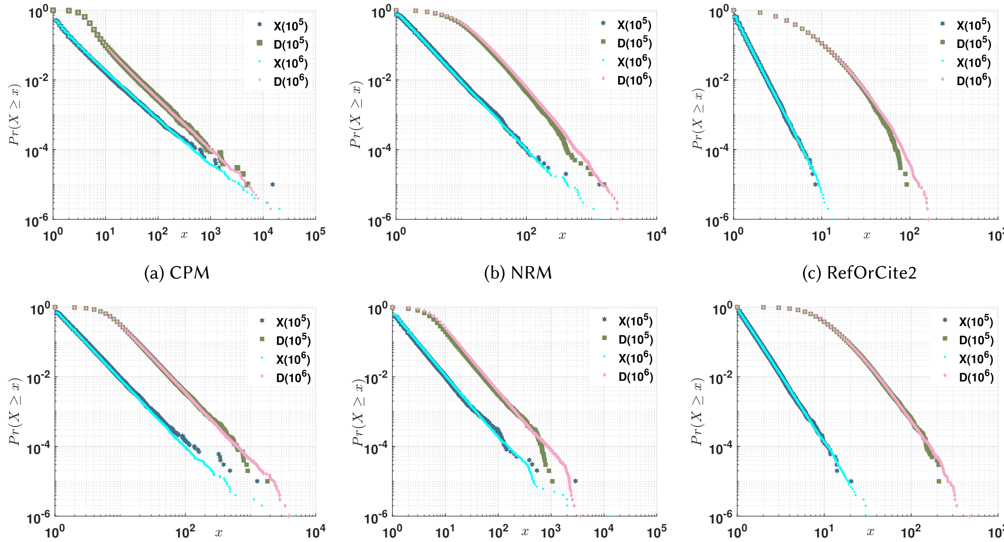
Table 4. Goodness of Fit Values[3](GoF) Using MSE (Mean-squared Error) Cost Function, and $K2$ Represents Test Statistic of *Two-sample Kolmogorov-Smirnov Test* [4][20] that Measures the Maximum Absolute Difference between the Cdfs of Two Input Distributions

Data/Model	**X**-distribution		Degree distribution	
	GoF	$K2$	GoF	$K2$
Biomedical	**0.0084**	**0.0452**	0.0235	0.0480
Supreme Court	**0.0026**	**0.0166**	0.1212	0.0651
ArxivTH	**0.0029**	**0.0273**	0.0210	0.0336
ArxivPH	**0.0085**	0.0466	0.0554	**0.0362**
Patent	**0.0066**	0.0614	0.0110	**0.0375**
Facebook	**0.0018**	**0.0307**	0.1579	0.0791
CPM	0.0227	**0.0183**	**0.0060**	0.0240
NRM	**0.0028**	0.0603	0.0184	**0.0339**
RefOrCite2	**0.0036**	**0.0250**	0.0617	0.0500
BA	**0.0118**	**0.0193**	0.0237	0.0433
CDPAM	0.0485	0.0697	**0.0159**	**0.0254**
DMS	**0.0034**	**0.0125**	0.0302	0.0368

Values in bold (lower values of GoF or $K2$) represent better performance.

and maximum deviation is observed in the tail. It may result in inaccurate computation of γ. From the figure, it is observed that **X**-distribution is more stable and consistent with the growth of the network until the model changes its parameters.

[3]https://in.mathworks.com/help/ident/ref/goodnessoffit.html
[4]https://in.mathworks.com/help/stats/kstest2.html#btno0gd-ks2stat



(a) CPM (b) NRM (c) RefOrCite2

(d) BA (e) CDPAM (f) DMS

Fig. 5. Consistency of Degree distribution and **X**-distribution with the growth of networks generated using (a) CPM ($p = 0.7$), (b) NRM ($p = 0.4, \beta = 0.2$), (c) RefOrCite2 ($p = 0.4,\ q = 0.2$), (d) BA ($\overline{k} = 5$), (e) CDPAM ($\beta = 5, \overline{k} = 5$), and (f) DMS ($\beta = 10, \overline{k} = 5$) network reconstruction models.

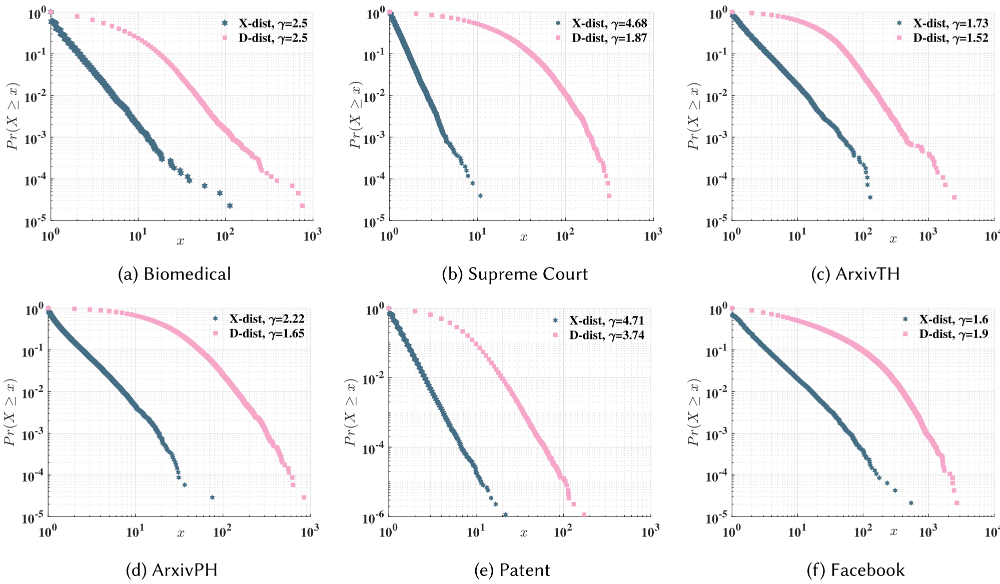
3.4 Verification on Real-world Networks

We also applied Algorithm 1 over several real-world networks, for example, Biomedical, Supreme court, ArxivTH, ArxivPH, Patent, and Facebook real-world networks (for descriptions refer to Table 2) and reported power-law exponents evaluated using **X**-distribution and degree distribution. The plots of distributions are available in Figure 6. We can clearly observe that the linearity (low values of GoF and $K2$ in Table 4) in the plots of **X**-distributions is better than the linearity observed in degree distribution plots. From Table 4, it is observed that **X**-distributions exhibit more linearity (low values of GoF and $K2$) on the log-log scale than degree distributions in most of the cases. Thus, **X**-distribution is more suitable for the evaluation of power-law exponent using linear fitting (on the log-log scale).

4 CONCLUSION

In this article, the problem of retraceability of microdynamics of a growing network is considered, which has importance in network reconstruction. We propose the **X**-distribution to compute the model parameters more accurately compared to the networks' associated degree distribution. Retracing the parameter values using **X**-distribution is successfully applied over the networks obtained under the BA, CP, NRM, RefOrCite2, CDPAM, and DMS models. The experimental results show that the **X**-distribution is more consistent with the growth of a network than the degree distribution. We also verified the effectiveness of **X**-distribution over degree distribution on various real-world networks.

In future, **X**-distribution would be applied to real-world data to analyze the universality of power law and reconstruction of real-world networks. One can further explore to propose better network reconstruction models using **X**-distribution. To retrace their microdynamics, we will further study the applicability of **X**-distribution for other real-world networks, including weighted networks,



(a) Biomedical (b) Supreme Court (c) ArxivTH

(d) ArxivPH (e) Patent (f) Facebook

Fig. 6. (Best viewed in color.) **X**-distributions and Degree distributions for (a) Biomedical, (b) Supreme Court, (c) ArxivTH, (d) ArxivPH, (e) Patent, and (f) Facebook real-world datasets.

signed networks, and multi-layer networks. Additionally, we will investigate effectiveness of X-distribution in temporal and dynamic environment where node degree and connections change over time. Such investigations can provide insights of structural dynamics of real-world networks. Furthermore, the relation among X-distribution and community structure of the network is still left unexplored. Hence, we also keep this for our future work.

REFERENCES

[1] A. Arya and P. K. Pandey. 2022. Structural reconstruction of signed social networks. In *IEEE Transactions on Computational Social Systems* 10, 5 (2022), 2599–2612. Oct. 2023, DOI: 10.1109/TCSS.2022.3186017
[2] A. Arya, P. K. Pandey, and A. Saxena. 2023. Balanced and unbalanced triangle count in signed networks. In *IEEE Transactions on Knowledge and Data Engineering* 35, 12 (2023), 12491–12496. DOI: 10.1109/TKDE.2023.3272657
[3] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286, 5439 (1999), 509–512.
[4] U. Bhat, P. L. Krapivsky, R. Lambiotte, and S. Redner. 2016. Densification and structural transitions in networks that grow by node copying. *Phys. Rev. E* 94, 6 (2016), 062302.
[5] Ginestra Bianconi and A.-L. Barabási. 2001. Competition and multiscaling in evolving networks. *Europhys. Lett.* 54, 4 (2001), 436.
[6] Steffen Bohn and Marcelo O. Magnasco. 2007. Structure, scaling, and phase transition in the optimal transport network. *Phys. Rev. Lett.* 98, 8 (2007), 088702.
[7] Deepayan Chakrabarti and Christos Faloutsos. 2006. Graph mining: Laws, generators, and algorithms. *ACM Comput. Surveys* 38, 1 (2006), 2.
[8] Aaron Clauset, Cosma Rohilla Shalizi, and Mark E. J. Newman. 2009. Power-law distributions in empirical data. *SIAM Rev.* 51, 4 (2009), 661–703.
[9] Reuven Cohen and Shlomo Havlin. 2003. Scale-free networks are ultrasmall. *Phys. Rev. Lett.* 90, 5 (2003), 058701.
[10] Sergey N. Dorogovtsev, José Fernando F. Mendes, and Alexander N. Samukhin. 2000. Structure of growing networks with preferential linking. *Phys. Rev. Lett.* 85, 21 (2000), 4633.
[11] James H. Fowler and Sangick Jeon. 2008. The authority of Supreme Court precedent. *Soc. Netw.* 30, 1 (2008), 16–30.
[12] Michelle Girvan and Mark E. J. Newman. 2002. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12 (2002), 7821–7826.

[13] Marta C. González, Pedro G. Lind, and Hans J. Herrmann. 2006. System of mobile agents to model social networks. *Phys. Rev. Lett.* 96, 8 (2006), 088702.
[14] Roger Guimera, Stefano Mossa, Adrian Turtschi, and L. A. Nunes Amaral. 2005. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 22 (2005), 7794–7799.
[15] Petter Holme and Beom Jun Kim. 2002. Growing scale-free networks with tunable clustering. *Phys. Rev. E* 65, 2 (2002), 026107.
[16] Pavel L. Krapivsky and Sidney Redner. 2005. Network growth by copying. *Phys. Rev. E* 71, 3 (2005), 036118.
[17] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2005. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. 177–187.
[18] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2007. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data* 1, 1 (2007), 2.
[19] Xiang Li and Guanrong Chen. 2003. A local-world evolving network model. *Physica A: Stat. Mech. Appl.* 328, 1 (2003), 274–286.
[20] Frank J. Massey Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Stat. Assoc.* (1951), 68–78.
[21] Mark Newman. 2018. *Networks*. Oxford University Press.
[22] J.-P. Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A.-L. Barabási. 2007. Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18 (2007), 7332–7336.
[23] Pradumn Kumar Pandey and Bibhas Adhikari. 2015. Context dependent preferential attachment model for complex networks. *Physica A: Stat. Mech. Appl.* 436 (2015), 499–508.
[24] Pradumn Kumar Pandey and Bibhas Adhikari. 2017. A parametric model approach for structural reconstruction of scale-free networks. *IEEE Trans. Knowl. Data Eng.* 29, 10 (2017), 2072–2085.
[25] Pradumn Kumar Pandey, Mayank Singh, Pawan Goyal, Animesh Mukherjee, and Soumen Chakrabarti. 2020. Analysis of reference and citation copying in evolving bibliographic networks. *J. Informetr.* 14, 1 (2020), 101003.
[26] Akrati Saxena. 2022. Evolving models for dynamic weighted complex networks. *Principles of Social Networking: The New Horizon and Emerging Challenges* (2022), 177–208.
[27] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. 2009. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*. 37–42.
[28] Wen-Xu Wang, Bo Hu, Bing-Hong Wang, and Gang Yan. 2006. Mutual attraction model for both assortative and disassortative weighted networks. *Phys. Rev. E* 73, 1 (2006), 016133.

Received 9 October 2023; accepted 15 December 2023