A robust method for statistical testing of empirical power-law distributions

Davide Garbarino[1], Veronica Tozzo[1], Andrea Vian[2], and Annalisa Barla[1]

1 - DIBRIS, 2 - DAD, Universitá di Genova, Genova Italy
davide.garbarino@edu.unige.it
veronica.tozzo@dibris.unige.it
andrea.vian@unige.it
annalisa.barla@unige.it

Abstract. The World-Wide-Web is a complex system naturally represented by a directed network of documents (nodes) connected through hyperlinks (edges). In this work, we focus on one of the most relevant topological properties that characterize the network, *i.e.* being scale-free. A directed network is scale-free if its in-degree and out-degree distributions have an approximate and asymptotic power-law behavior. If we consider the Web as a whole, it presents empirical evidence of such property. On the other hand, when we restrict the study of the degree distributions to specific sub-categories of websites, there is no longer strong evidence for it. For this reason, many works questioned the almost universal ubiquity of the scale-free property. Moreover, existing statistical methods to test whether an empirical degree distribution follows a power law suffer when dealing with large sample size and/or noisy data.
In this paper, we propose an extension of a state-of-the-art method that overcomes such problems by applying a subsampling procedure on the graphs performing Random Walks (RW). We show on synthetic experiments that even small variations of true power-law distributed data causes the state-of-the-art method to reject the hypothesis, while the proposed method is more sound and stable under such variations.
Lastly, we perform a study on 3 websites showing that indeed, depending on the sub-categories of website we consider, some accept and some refuse the hypothesis of being power-law. We argue that our method could be used to further explore sub-categories of websites in order to better characterize their topological properties deriving from different generative principles: central or peripheral.

Keywords: Power-law Distribution, Random Walks, Statistical Test, World-Wide-Web, Network Analytics

1 Introduction

The World-Wide-Web (WWW) encodes associative links among a large amount of pages. Its structure has grown without any central control, thus make it approximable to the result of a random process, where pages link to each other

following local probabilistic rules.
Such probabilistic rules are defined through statistical properties of Web graph features. In particular, several investigations show that the WWW is scale-free [1, 5, 8] *i.e.*, both the distributions of incoming and outgoing links are well-approximated by a discrete power law [19]. This can be traced to the fact that the vast majority of documents in the Web have relatively few outgoing and incoming links, but few pages still have enormous number of links that skew the mean of the empirical distribution far above the median.

Nonetheless, when analyzing specific portions of the Web, *i.e.* websites, the scale-free property seems to be less evident especially for specific sub-categories of websites (*e.g.* university homepages) [21, 22]. Note that, differently from what is commonly done in literature [21], we consider websites as closed sub-systems of the Web whose temporal evolution is independent of the system they evolved into.

In this work, we are interested in developing a method able to assess if data from empirical observations follow a power-law. Indeed, testing power laws on empirical data is usually hard due to the large fluctuations that are present in the tail of the distribution.

One of the most commonly used methods for testing is the Kolmogorov-Smirnov test [9]. This method focuses on the center of the distribution, making it not suitable for testing heavy-tailed distributions. In [9] the authors make strong use of this test by performing a bootstrap procedure that is optimal in small sample size regimes. Indeed, as the sample size grows, the power of the statistical test increases thus leading to higher rate of rejections of the null hypothesis. Moreover, even in presence of small sample sizes, adding a low amount of noise may cause the test to reject.

As in real-world noisy or large samples are the common scenario, here, we propose an alternative testing pipeline that leverages on the Anderson Darling test [3] and Random Walks (RWs). Our pipeline is able to cope with the power of the test problem by reducing the sample size with random walks while maintaining the original degree distribution behavior.

We show synthetic experiments in which the state-of-the-art method fails under small variations or large sample sizes of input data. In all these cases, our method is proved to be more stable under variations and it can be shown that provides results with a better confidence. Lastly, we present case studies on 3 websites which present interesting results showing that indeed, closed sub-portion of the Web do not necessarily follow a power-law distribution.

Outline The remainder of the paper is organized as follows: Section 2 presents the state-of-the art algorithm for testing empirical power-law distribution; in Section 3 we present the limitations of such method with the related synthetic examples; in Section 4 we present our adaptation based on RWs to overcome the issue of power in empirical data; in Section 5 we present a set of synthetic experiments showing how our method is more stable; Section **??** shows results obtained on 3 websites; lastly, we conclude with Section 6 with some discussion on the obtained results and future research directions.

2 Discrete power-law distribution: definition, fit and statistical test

The discrete power-law distribution is defined as

$$\mathbb{P}(d_v = x) \approx \frac{1}{\xi(x_{min}, \alpha)} x^{-\alpha}, \tag{1}$$

where d_v is the random variable representing the degree of a node v, x_{min} is a fixed lower bound on the values x, α is a *scaling parameter*, and $\xi(x_{min}, \alpha) = \sum_{x=x_{min}}^{\infty} x^{-\alpha}$ is the Hurwitz-zeta function [13].

The parameter x_{min} is particularly important, as often the degree distribution of a network follows a power law only for degrees x greater than a lower bound. A network is said to be scale-free if the tail of its in-degree and out-degree distributions obeys to a discrete power law decay. In practice, this entails that we have a non-null probability to observe nodes with a degree much greater than average (hubs).

2.1 Maximum Likelihood Estimation

The parameters x_{min} and α of an empirical power-law distribution need to be estimated from data. Given as input a vector $\mathbf{x} \in \mathbb{N}^n$ representing the degrees of n nodes of a graph, we need to perform two different procedures to estimate these two parameters, as described by the pseudo-code in Algorithm 1.

Estimate of x_{min} add at least 25 remaining observations First, we pick $\hat{x}$ as the value that minimizes the difference between the empirical degree distribution and the fitted power-law model where $x_{min} = \hat{x}$ [10, 9]. Note that, if we select a $\hat{x} > x_{min}$, we are reducing the size of our training data, and our model will suffer from the statistical fluctuations in the tail of the empirical distribution. On the other hand, if $\hat{x} < x_{min}$, the maximum likelihood estimate of the scaling parameter $\hat{\alpha}$ may be severely biased.

In order to minimize the difference of the empirical and fitted distributions, we need to select a suitable distance. One of the most common is the Kolmogorov-Smirnov (KS) statistic, which is defined as the maximum distance between the cumulative distribution functions (CDFs) of the empirical data and the best-fit model [16]. Although the KS statistic is widely used, it presents some drawbacks in the detection of heavy-tailed distributions since, being based on the CDF, it mainly penalizes fluctuations in the center of the empirical distribution. A more reliable distance for the comparison of heavy-tailed distributions is the Anderson-Darling (AD) statistic as it puts more importance to the extreme values of the CDFs [3]. For this reason, we will recur to this statistic in the rest of the paper. The AD distance is defined as

$$A^2(\mathbf{x}, F_{x_{min}=x}) = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln F_{x_{min}=x}(x_i) + \ln(1 - F_{x_{min}=x}(x_{n+1-i})) \right], \tag{2}$$

Algorithm 1 Power-law fitting

Input: degrees vector of length n
distances = [ ]
for $x \in \{\min(\texttt{degrees}), \max(\texttt{degrees}) - 25\}$ **do**
    if len(degrees $> x$) too short **then**
        break
    $\alpha \leftarrow$ power_law_fit(degrees, $x_{min} = x$)
    $d \leftarrow$ Anderson-Darling(degrees, x, α)
    distances.append(d)
$\hat{x} \leftarrow \underset{x}{\text{argmin}}$ distances
$\hat{\alpha} \leftarrow$ power_law_fit(degrees, $x_{min} = \hat{x}$)
$\hat{d} \leftarrow$ Anderson-Darling(degrees, $\hat{x}, \hat{\alpha}$)
return $\hat{x}, \hat{\alpha}, \hat{d}$

where n is the sample size and $F_{x_{min}=x}$ is the power-law CDF.

The estimated lower bound $\hat{x}$ is then the observed degree that minimizes Equation (2).

Estimate of α Given the lower bound x_{min}, we estimate the scaling parameter α by means of maximum likelihood, which provides consistent estimates in the limit of large sample sizes [11].
In the discrete case, a good approximation of the true scaling parameter can be reached mostly in the $x_{min} \geq 6$ regime [9]. And it can be computed as:

$$\hat{\alpha} \approx 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min} - \frac{1}{2}} \right]^{-1}. \tag{3}$$

2.2 Goodness-of-fit test

Once $\hat{\alpha}$ and $\hat{x}$ have been estimated, we need to assess if observed data are plausibly sampled from the related power-law distribution. To such extent, we perform a goodness-of-fit (GoF) test procedure [17].

A goodness of fit test measures how well a statistical model fits into a set of observations. Typically, a GoF makes use of a so-called test statistic in order to evaluate the discrepancy between the observed values and the values expected under the tested statistical model. GOF statistics are functions which do not depend on the parameters of the statistical model. The goodness-of-fit test produces a p-value corresponding to the probability that the test statistic is greater than the realization of the same statistic on the observed data.
The pseudo-code of the test we perform is presented in Algorithm 2. In particular, we recur to a semi-parametric bootstrap approach where we fixed as statistic the Anderson-Darling distance. The bootstrap phase is needed as, when estimating parameters from data, we do not know the distribution of the test statistic under the null hypothesis [23, 9]. Given n samples, we indicate with n_{tail} the

Algorithm 2 Power-law testing

Input: `degrees` vector of length $n, \hat{x}, \hat{\alpha}, \hat{d}$
`distances` = [ ]
for $i = 1, \ldots, M$ **do**
  n_{tail} = count(`degrees` $> \hat{x}$)
  for $j = 1, \ldots, n$ **do**:
    $b \leftarrow$ bernoulli_sample(n_{tail}/n)
    if b is 1 **then**
      $s_i[k]$ = power_law_sample($\hat{x}, \hat{\alpha}$)
    else
      $s_i[j] \leftarrow$ uniform_sample(`degrees` $< \hat{x}$)
  $\alpha_i, x_i \leftarrow$ power_law_fit(s)
  $d \leftarrow$ Anderson-Darling(s, x_i, α_i)
  `distances`.append(d)
`p-value` = count(`distances` $> \hat{d}$)$/M$
return `p-value`

amount of samples that are greater than $\hat{x}$. Bootstrap is then performed by simulating n_{tail} examples from a power law with parameters $\hat{\alpha}$ and $\hat{x}$, and for the remaining sample size $n - n_{tail}$ we sample degrees from the empirical data that are smaller than $\hat{x}$. We repeat this procedure M times. The value of M depends on the desired significance of the p-value. Typically, if we want a p-value that approximates its true value with an error smaller than ϵ, then $M = \frac{1}{4\epsilon^2}$.

Given the M simulated data sets, we fit to each of them its own power-law model and compute the AD distance. This provides the empirical distribution of the AD statistic that we use to compute the associated p-value, defined as the fraction of synthetic distances larger than the observed one.

If p is large (relatively to a fixed significance level, *e.g.* 0.1), we cannot reject the null hypothesis, then possibly the difference between the empirical and theoretical distributions may be attributed to statistical fluctuations. Differently, if p is smaller than the significance level, we say that the empirical data are not power law.

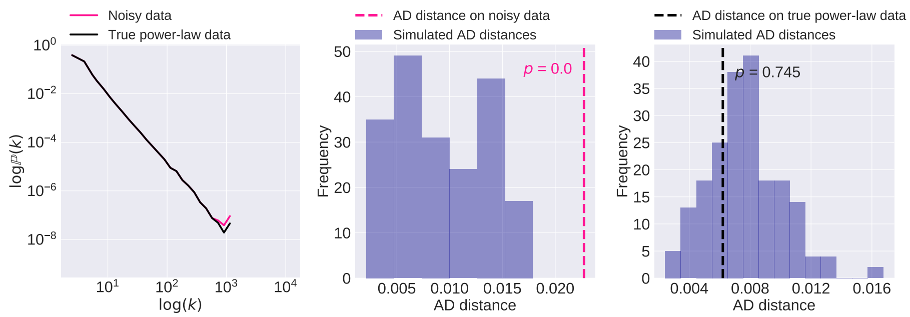
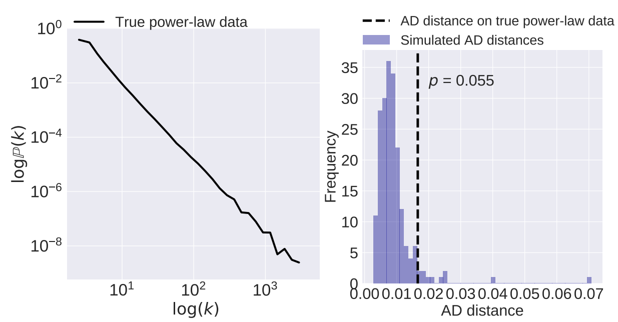


Fig. 2: On the left, the empirical probability density functions of true power-law data (black line) and noisy power-law data (pink). On the right, the Anderson-Darling test on both samples. Little variations from an exact power-law sample lead to reject the null hypothesis.

3 Problems of goodness-of-fit on empirical data



On the left, the empirical probability density function of true power law data. On the right, the Anderson-Darling test. Large sample size (5×10^5) leads to reject the null hypothesis.

Testing whether empirical data are power-law distributed is a hard task. This is due to the following reasons: a) the probability of rejecting the null hypothesis grows with sample size; and, as a consequence b) the procedure is too sensitive to even minimal amount of noise. Little attention has been put on these issues, but we argue that they are crucial as they heavily affect the final response of the statistical test.

In particular, both problems can be addressed by considering the *power of the test*, which, fixed a significance level, is defined as the probability of correctly rejecting the null hypothesis. Such probability increases accordingly to the sample size, hence, when the number of nodes n is large, we tend to reject the null hypothesis even in cases of true power-law distributed data (as the power of the test is very close to 1). Indeed, by performing bootstrap, we simulate nearly exact power-law samples, which induce the Anderson-Darling test to be very sensitive to even minimal fluctuations in the observed distribution.

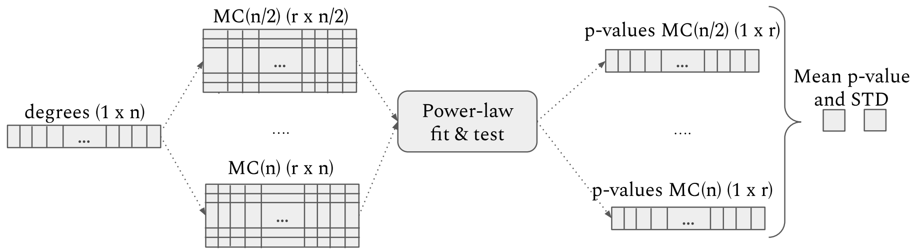


Fig. 3: Schematic representation of the proposed pipeline.

In Figure 1 and 3, we show two synthetic experiments where such test fails, in particular:

(a) we generated $n = 10^5$ samples from a discrete power-law distribution with parameters $x_{min} = 7$ and $\alpha = 2.7$. We perturbed the data by adding one occurrence to the last 13 degrees in the extreme tail (see Figure 1 left panel for the true and perturbed data);
(b) we generated $n = 5 \times 10^5$ samples from a discrete power-law distribution with parameters $x_{min} = 2$ and $\alpha = 2.7$.

We applied the procedure in Section 2 on both datasets, with $M = 200$ and significance level set to 0.1. Results are shown on the right side of Figure 1 and 3. In Figure 1, the empirical probability density functions of the two samples are indistinguishable from each other except in the extreme tail, where little divergences can be traced. Thus, it becomes evident that for large sample sizes the test is very sensitive even to little fluctuations in the observed sample. Also, with example (b) we show that even perfect power-law samples induce the test to fail when the sample size is too large (Figure 3).
Both examples show that the high power of the Anderson-Darling test in large sample size regimes constitutes a drawback of the previously introduced method [9]. Since it is never the case that an observed degree distribution is exactly drawn from a discrete power law, we propose a variation of the method in Section 2 that aims at testing the goodness of fit of heavy tail distributions.

4 Monte Carlo approach

Our proposal is based on the idea of performing iterative Monte Carlo (MC) sub-samplings of different length on the original degree sequence. We argue that with this sub-sampling scheme we can reduce the sample size without modifying the trend of the original degree distribution and possibly obtain a more reliable test.
The global scheme of the procedure is provided in Figure 2. In particular, we define a set of lengths, $\{l_1, \ldots, l_{max}\}$, for each length we perform r corresponding MC samplings.(see Section 4.1 for a more detailed RWs description). For each sample we fit a power-law distribution and assess its

plausibility following Algorithm 1 and Algorithm 2. Thus obtaining r p-values of the Anderson-Darling test. We consider, as final output of the procedure, the mean of all p-values for all different lengths and the related standard deviation. To reject the null hypothesis, we fix the significance level at 0.1. This is a conservative choice implying that the power law hypothesis is ruled out if there is a probability of 1 in 10 or less that we obtain data that agree with the model as the data we have.

To the best of our knowledge, it is not usual to exploit RWs to test for power-law decay in the degree distribution. In fact, performing RWs does not allow to exactly estimate the parameters of the power-law distribution, indeed, to each RW may correspond a different set of parameters. Nonetheless, we do not use RWs as a fitting method but rather to say if a network is plausible to asymptotically satisfying the scale-free property. We argue that using RWs as a way to obtain suitable sub-samples of smaller sample size would provide better understanding of the degree sequence behavior while overcoming the drawbacks induced by large sample sizes.

The algorithm to obtain a RW given a graph and a starting node is presented in Algorithm **??**. Given a node, we iteratively select one of its neighbors at random and move to it. We repeat the procedure until we reach the desired length sequence of nodes. The collected sequence is defined as a RW on the graph.

length of monte carlo resampling The problem of selecting adequate lengths for the random walks is not trivial. We now want to provide a lower and an upper bound for the lengths given the following considerations: on the one hand, a too short random walk would lead to very different degree sequences due to the large fluctuations present in the original network, while, on the other hand, lengths close to the original degree sequence would not provide a smaller sample, leading to higher rates of rejection of the power-law hypothesis.

5 Experimental results

In order to evaluate the performance of the proposed pipeline, we perform four experiments and compare the results with the state-of-the-art method. In the results and in the rest of the narration we will refer to the state-of-the-art method as Bootstrap and to our method as Monte Carlo + Bootstrap.

5.1 Validation of the proposed method on different graph models

In the first expweriment we aim at verifying if the prposed method is comparable to the SoA when considering two cases at varyiong sample size:

1. Erdős-Renyi models of different sample sizes, $\{75 \times 10^3, 15 \times 10^4, 3 \times 10^5\}$, we expect both ours and state-of-the-art method to refuse the null hypothesis as the degree distribution of this model is known to follow a binomial distribution [12]. Thus, we use this as base test to assess the probability of correctly rejecting the power-law hypothesis.

2. Barabasi-Albert models of different sample sizes, $\{75\times10^3, 15\times10^4, 3\times10^5\}$, we expect both methods to have high p-values as the degree distribution follows a power law [6]. We use this experiment to provide proof of the soundness of the method in presence of true power-law data.

Each experiment listed above is repeated 10 times to estimate the mean and standard deviation of performances. Results are rported in Table **??** Results are reported in Table **??** and Figure 3. In particular, in the table we report the mean p-values and standard deviations obtained when analysing two extreme cases: true power-law data and completely non power-law data. Results show that we always reject the null hypothesis in the latter case, while in the former we always provide p-values with a smaller variance than the bootstrap approach.

5.2 Robustness to noise

Configuration models, where we provide an input distribution that follows a power law and we want to show that our method is more robust under increasing noise in the input distribution[20]. We simulated from a discrete power law, with parameters $\alpha = 2.3$ and $x_{min} = 1$, a sample of size $n = 4 \times 10^5$. For different levels of noise in the set $\bar{n} \in \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$, we perturbed the power law observation by adding $\bar{n}$ values uniformly sampled from the original observation. For each perturbed distribution we generate the related configuration model. For our pipeline, we fixed 5 lengths of RWs within the lower and upper bounds defined in Section **??**. For each length, we perform 5 different RWs. All the simulations are performed in Python. We used the package **powerlaw** [2] for fitting power-law distributions to empirical data and compute the AD distances. We provide all the notebooks used for the experiments of this paper in a GitHub repository[1]. Figure 3 shows that the proposed methods is ... Figure 3 shows similar results, in mean our approach is almost always better than the simple bootstrap approach while also providing a smaller variance. Also, it never reject the null-hypothesis in cases in which the noise is small while sometimes it rejects it in presence of high amount of noise (100 added observations). Differently from the boostrap approach that, depending on the simulated sample, sometimes rejects it even in presence of zero noise.

5.3 Benchmark on University of Notre Dame website

In literature, we find a widely studied example of empirical data that is assumed to follow a power-law distribution [1, 4, 18], *i.e.* the web graph of the University of Notre Dame website. This graph, in 1999, has been studied in order to obtain information regarding the topology of the WWW. In [1], the authors found that the in-degree and out-degree distributions of the graph

[1] LINKTO

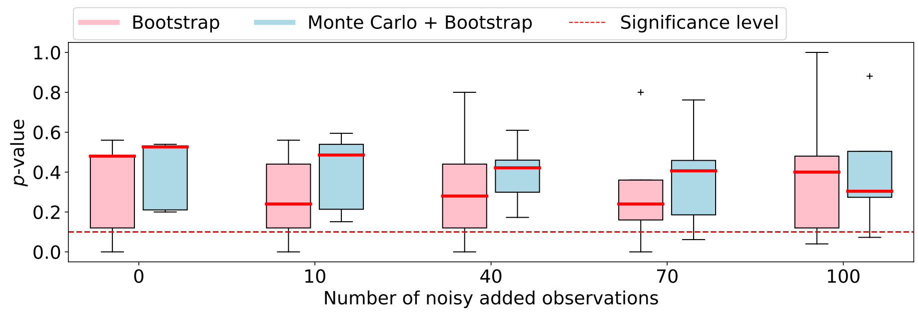


Fig. 4: Results in terms of p-values for the two testing pipelines as the input data present an increasing level of noisy observations.

underlying the hyperlink structure of the domain nd.edu were well approximated by power-law distributions with scaling parameters 2.7 and 2.1 respectively. We downloaded the hyperlink graph from http://snap.stanford.edu/ [14]; the crawl consists of 325729 documents and 1497134 links. We tested the proposed pipeline on the in-degree distribution of the network. For this purpose we performed 5 MC samplings for different sizes equally spaced between 162864 and the number of nodes. We obtained a mean p-value equal to 0.15, meaning that there is no strong evidence against the power-law hypothesis for the in-degree distribution.
Differently, when applying the state-of-the-art method we observed a p-value equal to 0.00, which would lead us to reject the null hypothesis.
This allows us to conclude that....

5.4 Websites analysis

In order to assess the proposed method on real scenarios, we considered three different websites that we deemed representative of different strategies of content creation: e-commerce, academic and public forum. The first category is typically characterized by a strong central control in the design and evolution of the information architecture and content generation. Conversely, the last category is completely user-guided and its evolution is, thus, likely to be random. We argue that the academic category, as well as most complex institutions, should be a trade-off between the two, as usually many contributors have access to writing and adding content with a mild central control. We crawled the following websites using the open source framework Scrapy[2]:

1. Goop, the website of a wellness and lifestyle company; during the crawl we restricted to the domains `goop.com` and `shop.goop.com`;

[2] https://scrapy.org/.

2. Stanford, the website of Stanford University. We downloaded a crawl performed in 2002 available at http://snap.stanford.edu/;
3. InsideHoops, a basketball news website in which users can also exchange views in a forum section. We crawled starting from `insidehoops.com/forum` limiting the depth to 5 edges, in order to retrieve only content within the public forum.

Table **??** describes the characteristics of the three considered websites, in terms on nodes, edges, and type and reports the results of hypothesis testing on the in-degree distribution with the proposed method.

6 Discussion

1. we proposed a method for hypothesis testing of power-law distributions in empirical data to overcome issues related to power and sample size and noise. When the sample size is big, we rely on iterative Monte-Carlo subsamplings
2. we verified that the proposed method retains the ability of assessing if a distribution is a power-low with different sample sizes
3. we observe that the method is indeed more reliable than the state-of-the art in sythetic data when dealing with noisy data
4. we confirm that the UND follows a power-law
5. we start exploring how different content generation strategies for websites may induce a different connectivity structure of the hyperlink graph
6. future research direction may involve considering RWs instead of simple Monte Carlo Random Walks represent a sub-sampling technique on graphs aimed at evaluating empirical characteristics on suitable sub-networks [7, 15].

References

1. Albert, R., Jeong, H., Barabási, A.L.: Diameter of the world-wide web. nature **401**(6749), 130–131 (1999)
2. Alstott, J., Bullmore, D.P.: powerlaw: a python package for analysis of heavy-tailed distributions. PloS one **9**(1) (2014)
3. Anderson, T.W., Darling, D.A.: A test of goodness of fit. Journal of the American statistical association **49**(268), 765–769 (1954)
4. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. science **286**(5439), 509–512 (1999)
5. Barabási, A.L., Albert, R., Jeong, H.: Scale-free characteristics of random networks: the topology of the world-wide web. Physica A: statistical mechanics and its applications **281**(1-4), 69–77 (2000)
6. Barabási, A.L., et al.: Network science. Cambridge university press (2016)
7. Basirian, S., Jung, A.: Random walk sampling for big data over networks. In: 2017 International Conference on Sampling Theory and Applications (SampTA). pp. 427–431. IEEE (2017)

8. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., Wiener, J.: Graph structure in the web. Computer networks **33**(1-6), 309–320 (2000)
9. Clauset, A., Shalizi, C.R., Newman, M.E.: Power-law distributions in empirical data. SIAM review **51**(4), 661–703 (2009)
10. Clauset, A., Young, M., Gleditsch, K.S.: On the frequency of severe terrorist events. Journal of Conflict Resolution **51**(1), 58–87 (2007)
11. Daniels, H.: The asymptotic efficiency of a maximum likelihood estimator. In: Fourth Berkeley Symposium on Mathematical Statistics and Probability. vol. 1, pp. 151–163. University of California Press Berkeley (1961)
12. Erdös, P., et al.: On random graphs
13. Hardy, M.: Paretos law. The Mathematical Intelligencer **32**(3), 38–43 (2010)
14. Leskovec, J., Sosič, R.: Snap: A general-purpose network analysis and graph-mining library. ACM Transactions on Intelligent Systems and Technology (TIST) **8**(1), 1 (2016)
15. Lovász, L., et al.: Random walks on graphs: A survey. Combinatorics, Paul erdos is eighty **2**(1), 1–46 (1993)
16. Massey Jr, F.J.: The kolmogorov-smirnov test for goodness of fit. Journal of the American statistical Association **46**(253), 68–78 (1951)
17. Maydeu-Olivares, A., Garcia-Forero, C.: Goodness-of-fit testing. International encyclopedia of education **7**(1), 190–196 (2010)
18. Mossa, S., Barthélémy, M., Eugene Stanley, H., Nunes Amaral, L.A.: Truncation of power law behavior in "scale-free" network models due to information filtering. Phys. Rev. Lett. **88**, 138701 (Mar 2002). https://doi.org/10.1103/PhysRevLett.88.138701, https://link.aps.org/doi/10.1103/PhysRevLett.88.138701
19. Newman, M.E.: Power laws, pareto distributions and zipf's law. Contemporary physics **46**(5), 323–351 (2005)
20. Newman, M.: Networks: An introduction. 2010 oxford
21. Pennock, D.M., Flake, G.W., Lawrence, S., Glover, E.J., Giles, C.L.: Winners don't take all: Characterizing the competition for links on the web. Proceedings of the national academy of sciences **99**(8), 5207–5211 (2002)
22. Stumpf, M.P., Porter, M.A.: Critical truths about power laws. Science **335**(6069), 665–666 (2012)
23. Stute, W., Manteiga, W.G., Quindimil, M.P.: Bootstrap based goodness-of-fit-tests. Metrika **40**(1), 243–256 (1993)