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, Universitá di Genova, Genova, Italy
davide.garbarino@edu.unige.it, veronica.tozzo@dibris.unige.it,
annalisa.barla@unige.it
[2] DAD, Universitá di Genova, Genova, Italy
andrea.vian@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 from large sample sizes 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 Monte Carlo sub-sampling procedure on the graphs. 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 their category, some accept and some refuse the hypothesis of being power-law. We argue that our method could be used to better characterize topological properties deriving from different generative principles: central or peripheral.

Keywords: Power-law distribution · Monte Carlo · 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 [1,5,10] *i.e.*, both the distributions of incoming and outgoing links are well-approximated by a discrete power law [21]. 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 categories (*e.g.* university homepages) [22,24]. Note that, differently from what is commonly done in literature [22], 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 statistical test is the Kolmogorov-Smirnov [11]. This method focuses on the center of the distribution, making it not suitable for testing heavy-tailed distributions. In [11] 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 Monte Carlo sub-sampling. Our pipeline is able to cope with the power of the test problem by reducing the sample size 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 representative of different generative processes. These case studies present interesting results showing that indeed, closed sub-portion of the Web do not necessarily follow a power-law distribution. And, they seem to point in the direction that the more the generative process is centralized the less the degree distribution can be associated to a power law decay.

Outline. The remainder of the paper is organized as follows: Sect. 2 presents the state-of-the art algorithm for testing empirical power-law distribution; in Sect. 3 we present the limitations of such method with the related synthetic examples; in Sect. 4 we present our adaptation based on Monte Carlo sub-sampling to overcome the issue of power in empirical data; in Sect. 5 we present a large variety of experiments showing how our method is more stable and the case

studies; lastly, we conclude with Sect. 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 [15].

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}. First, we pick $\hat{x}$ as the value that minimizes the difference between the empirical degree distribution and the fitted power-law model where $x_{min} = \hat{x}$ [11,12].

In order to minimize such difference, we need to select a suitable distance. One of the most common is the Kolmogorov-Smirnov (KS) statistic, which is defined as the supremum norm of the difference between two distribution functions (CDFs) of the empirical data and the best-fit model [18]. 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

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

Algorithm 1. Power-law fitting

1: **Input**: degrees vector of length n
2: distances $= [\,]$
3: **for** $x \in \{\min(\text{degrees}), \ldots, \max(\text{degrees})\}$ **do**
4:     **if** len(unique(degrees) $> x$) <25 **then**
5:         **break**
6:     $\alpha \leftarrow$ power_law_fit(degrees, $x_{min} = x$)
7:     $d \leftarrow$ Anderson-Darling(degrees, x, α)
8:     distances.append(d)
9: $\hat{x} \leftarrow \underset{x}{\text{argmin}}$ distances
10: $\hat{\alpha} \leftarrow$ power_law_fit(degrees, $x_{min} = \hat{x}$)
11: $\hat{d} \leftarrow$ Anderson-Darling(degrees, $\hat{x}, \hat{\alpha}$)
12: **return** $\hat{x}, \hat{\alpha}, \hat{d}$

---

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

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.

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 [13].

In the discrete case, a good approximation of the true scaling parameter can be reached mostly in the $x_{min} \geq 6$ regime [11]. And it can be computed as:

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

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 [19].

A goodness-of-fit test measures how well a statistical model fits into a set of observations. Given the statistical model under testing, a GoF makes use of a statistic that evaluates the discrepancy between the observed values and the expected value of the model. By definition, a statistic is a function which does not depend on the parameters of the model. The output of the GoF procedure is a p-value corresponding to the probability that the statistic is greater than its realization on the observed data.

Note that, since we estimate the model parameters from data we do not know the distribution of the statistic. Thus, we perform a semi-parametric bootstrap approach to estimate such distribution empirically [11,25].

Algorithm 2. Power-law testing

```
 1: Input: degrees vector of length n, x̂, α̂, d̂
 2: distances = [ ]
 3: for i = 1, . . . , M do
 4:     n_tail = count(degrees > x̂)
 5:     for j = 1, . . . , n do:
 6:         b ← bernoulli_sample(n_tail/n)
 7:         if b is 1 then
 8:             s_i[k] = power_law_sample(x̂, α̂)
 9:         else
10:             s_i[j] ← uniform_sample(degrees < x̂)
11:     α_i, x_i ← power_law_fit(s)
12:     d ← Anderson-Darling(s, x_i, α_i)
13:     distances.append(d)
14: p-value = count(distances > d̂)/M
15: return p-value
```

In particular, we fixed as statistic the Anderson-Darling distance and we perform a procedure described in Algorithm 2. Given n samples, we indicate with n_{tail} the 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.

3   Problems of Goodness-of-Fit on Empirical Data

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.
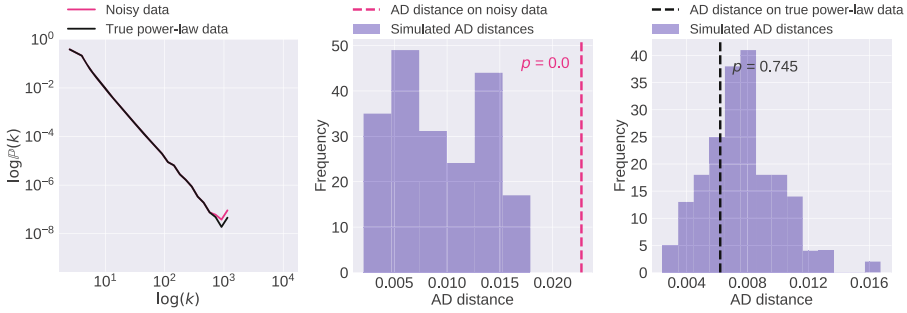
Fig. 1. On the left, the empirical probability density functions of true power-law data (black line) and noisy power-law data (pink). On the right, the Anderson-Darling test on both samples. Little variations from an exact power-law sample lead to reject the null hypothesis. (Color figure online)
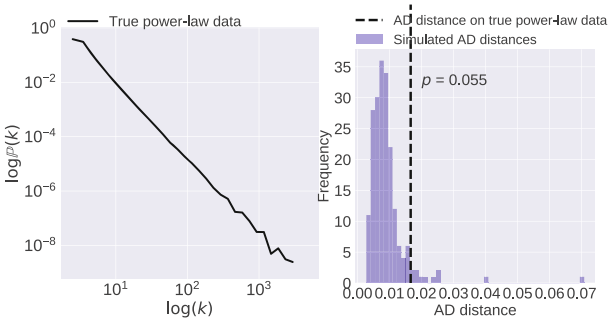

Fig. 2. On the left, the empirical probability density function of true power law data. On the right, the Anderson-Darling test. Large sample size (5×10^5) leads to reject the null hypothesis.

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.

In Fig. 1 and 2, 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° in the extreme tail (see Fig. 1 left panel for the true and perturbed data);
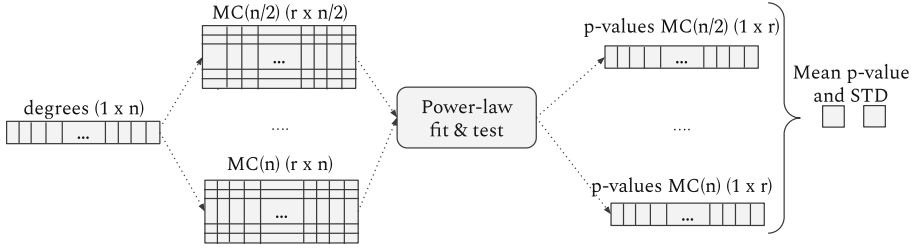
Fig. 3. Schematic representation of the proposed pipeline.

(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 Sect. 2 on both datasets, with $M = 200$ and significance level set to 0.1. Results are shown on the right side of Fig. 1 and 2. In Fig. 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 (Fig. 2).

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 [11]. 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 Sect. 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 Fig. 3. In particular, we define a set of lengths, $\{l_1, \ldots, l_{max}\}$, for each length we perform r corresponding MC samplings. For each sample, we fit a power-law distribution and assess its plausibility exploiting Algorithm 1 and Algorithm 2 and, thus, obtaining a sequence of p-values of the Anderson-Darling test of length r. We consider, as final output of the procedure, the mean of all p-values sequences for all different lengths and the related standard deviation.

To the best of our knowledge, it is not usual to exploit MC sub-sampling to test for power-law decay in the degree distribution. In fact, performing MC

does not allow to exactly estimate the parameters of the power-law distribution, indeed, to each sub-sample may correspond a different set of parameters. Nonetheless, we do not use MC as a fitting method but rather to say if a network is plausible to asymptotically satisfying the scale-free property. We argue that using MC 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.

4.1   Instantiation of Parameters

In order to apply the Monte Carlo approach we need to fix different values, specifically l_1, l_{max}, r and the significance level.

The problem of selecting adequate lengths for the MC sub-samples is not trivial. On the one hand, a too small sub-sample 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 lead to higher rates of rejection of the power-law hypothesis. Then, we arbitrarily decided to set l_1 at $n/2$ which is half the length of the observed data. As for l_{max}, we fix it to n as in case of true power-law samples we want to being able to obtain a high p-value, while in case of noisy data, considering one length equal to the original size does not particularly affect the resulting mean p-value.

The value of r affects the robustness of the final result, the more repetitions the better approximation of the true p-value. Nonetheless, its value depends on constraints deriving from computational power. Thus, we leave the definition of such value to the user.

We fixed the significance level at 0.1 for the rejection of the null hypothesis. 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 data sampled from the true model agree with the model as the empirical data.

Lastly, we fixed the maximal possible x_{min} to be least 25 observations less than the maximal observed degree. This is due to limit the chances of fitting a power-law distribution on too few observations.

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 rest of the narration we will refer to the state-of-the-art method as Bootstrap and to our method as Monte Carlo + Bootstrap.

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]. For all experiments, we fixed 30 lengths of Monte Carlo re-sampling in the interval $[\frac{n}{2}, n]$ and for each of this length we get $r = 10$ re-samplings.

---

[1] https://github.com/DaviGarba/netanalytics.

Table 1. Results to assess the goodness of the proposed testing pipeline in cases of scale-free graphs (Barabasi-Albert) or not (Erdős-Renyi), in terms of mean p-value and standard deviation on 10 repetitions of the test for different sample sizes.

Test type	Erdős-Renyi		Barabasi-Albert		
	75000	150000	75000	150000	300000
Bootstrap	0.00 ± 0.00	0.00 ± 0.00	0.85 ± 0.24	0.75 ± 0.27	0.78 ± 0.18
MC + Boostrap	0.00 ± 0.00	0.00 ± 0.00	0.85 ± 0.06	0.69 ± 0.16	0.71 ± 0.15

5.1 Validation of the Proposed Method on Different Graph Models

In the first experiment we aim at verifying if Monte Carlo + Bootstrap is comparable to just Bootstrap when considering two cases at varying sample sizes:

1. Erdős-Renyi models of size $\{75 \times 10^3, 15 \times 10^4\}$, we expect both methods to refuse the null hypothesis as the degree distribution of this model is known to follow a binomial distribution [14]. Thus, we use this as base test to assess the probability of correctly rejecting the power-law hypothesis.
2. Barabasi-Albert models of size $\{75 \times 10^3, 15 \times 10^4, 3 \times 10^5\}$, 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 p-values. Results are reported in Table 1 where we observe that our approach (Monte Carlo + Bootstrap) always reject the null hypothesis in the Erdős-Renyi case as the Bootstrap method, while in the Barabasi-Albert case we always provide p-values with a smaller variance.

5.2 Robustness to Noise

We now want to assess that our method is indeed more robust under increasing noise in the input empirical distribution. We simulated from a discrete power law with parameters $\alpha = 2.3$ and $x_{min} = 1$, a sample of size $n = 10^5$. For different levels of noise in the set $\bar{n} \in \{10, 40, 70, 100\}$, we perturbed the power law observation by adding $\bar{n}$ values uniformly sampled from the original observation.

Figure 4 shows that the proposed methods is in mean 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 Bootstrap approach that, depending on the simulated sample, sometimes rejects it even in presence of zero noise.
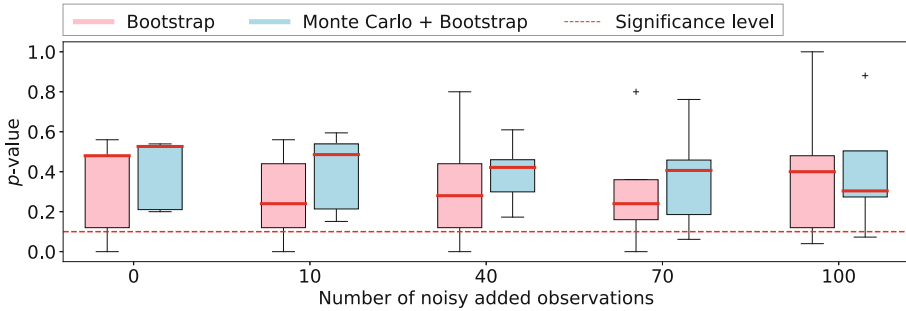
Fig. 4. Results in terms of p-values for the two testing pipelines as the input data present an increasing level of noisy observations.

5.3 Benchmark on University of Notre Dame Website

We exploit a widely studied example of empirical data that is assumed to follow a power-law distribution [1,4,20], *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 Web. In [1], the authors found that the in-degree and out-degree distributions of the graph 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/ [16]; the crawl consists of 325729 documents and 1497134 links. We tested the Monte Carlo + Bootstrap approach against the Bootstrap approach as the empirical data are noisy and we want to provide further validation of our testing procedure on the in-degree distribution of the network.

We performed $r = 5$ MC re-samplings for different sizes equally spaced in the interval [162864, 325729]. Monte Carlo + Bootstrap results in a mean p-value of 0.15, meaning that there is no strong evidence against the power-law hypothesis for the in-degree distribution. Differently, when applying the Bootstrap method we observed a p-value equal to 0.00, which would lead us to reject the null hypothesis.

As in literature many have argued the power-law nature of this graph, this allows us to conclude that our testing procedure is more robust and thus can be applied on real-world data with higher reliability.

5.4 Websites Analysis

We now want to exploit our procedure in real scenarios to seek for evidence of differences in the degree distributions deriving from different generative processes. We considered three different websites that we deemed representative of different strategies of content creation: e-commerce, academic and free encyclopedia. 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 other website of 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 (Fig. 5).
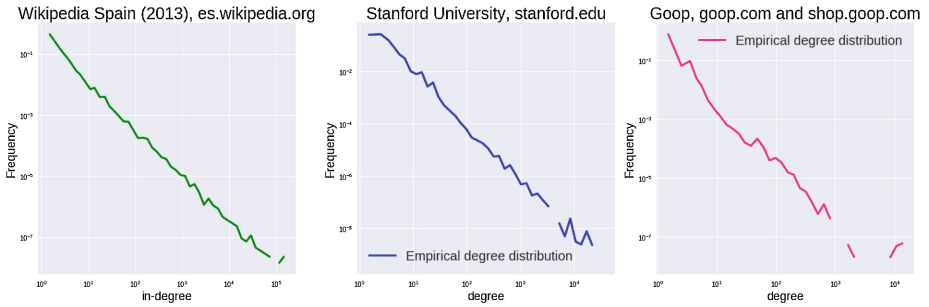

Fig. 5. Log-log plots of the empirical distributions of the considered case studies.

Table 2. Analyzed websites with the related information about number of nodes, number of edges and category.

Name	Url	Website type	No. nodes	No. edges	p-value
Goop	goop.com	E-commerce	100.482	731.259	0.00 ± 0.00
Stanford	stanford.edu	Academic	281.903	2.312.497	0.01 ± 0.00
Wikipedia (ES)	es.wikipedia.org	Encyclopedia	972.933	23.041.488	0.74 ± 0.21

We consider the following websites:

1. Goop, the website of a wellness and lifestyle company; we crawled the entire website using the open source framework Scrapy[2], during the crawl we restricted to the domains `goop.com` and `shop.goop.com`;
2. Stanford, the website of Stanford University. We downloaded a crawl performed in 2002 available at http://snap.stanford.edu/;
3. Wikipedia (ES), the website of the free spanish encyclopedia. We downloaded a crawl of 2013 at http://law.di.unimi.it/index.php [8,9].

Table 2 describes the characteristics of the three considered websites, in terms of category, number of nodes and number of edges. Table 2 also reports the mean p-values obtained with Monte Carlo + Bootstrap on the in-degree distributions. Results seems to validate our hypothesis about an inverse correlation between the centrality of the content generative process and the scale-free property.

---

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

6   Discussion

In this paper, we proposed a method for hypothesis testing of power-law distributions in empirical data that overcomes issues related to the power of the test. In particular, our method mediates the effect of possibly noisy data through Monte-Carlo sub-samplings of the empirical distribution. We verified that the proposed method retains the ability of assessing if observations are indeed plausibly sampled from a power law, under different sample sizes and level of noise. Indeed, the method is more reliable than the state-of-the art on synthetic data. To further assess the reliability of our approach we also provide a real-world example, specifically the University of Notre Dame website, which is a well studied dataset and it is considered to be scale-free. Our method does indeed provide a p-value higher than the significance level, differently from the state-of-the-art method that rejects the null hypothesis.

This allowed us to use our method to test different websites corresponding to different content generative processes. From a first insights, we observed that different content generation strategies may induce a different connectivity structure of the hyperlink graph.

For future research we intend to increase the number of real networks studied and consider current websites related to different generative processes to provide a more comprehensive understanding of specific sub-categories of the Web.

Future research directions may also involve the use of random walks instead of Monte Carlo as a sub-sampling technique on graphs [7,17] and the comparison with other estimators of power laws in empirical data [23].

To conclude, our pipeline is an attempt to perform statistical testing while considering its limits both theoretical and due to noisiness of data. We argue that this is fundamental to reliably test assumptions on real-world examples.

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. J. Am. Stat. Assoc. **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 Stat. Mech. Appl. **281**(1–4), 69–77 (2000)
6. Barabási, A.L., et al.: Network Science. Cambridge University Press, Cambridge (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. Boldi, P., Rosa, M., Santini, M., Vigna, S.: Layered label propagation: a multiresolution coordinate-free ordering for compressing social networks. In: Srinivasan, S., Ramamritham, K., Kumar, A., Ravindra, M.P., Bertino, E., Kumar, R. (eds.) Proceedings of the 20th International Conference on World Wide Web, pp. 587–596. ACM Press (2011)

9. Boldi, P., Vigna, S.: The WebGraph framework I: compression techniques. In: Proceedings of the Thirteenth International World Wide Web Conference (WWW 2004), pp. 595–601. ACM Press, Manhattan (2004)

10. Broder, A., et al.: Graph structure in the web. Comput. Netw. **33**(1–6), 309–320 (2000)

11. Clauset, A., Shalizi, C.R., Newman, M.E.: Power-law distributions in empirical data. SIAM Rev. **51**(4), 661–703 (2009)

12. Clauset, A., Young, M., Gleditsch, K.S.: On the frequency of severe terrorist events. J. Conflict Resolut. **51**(1), 58–87 (2007)

13. 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)

14. Erdös, P., et al.: On random graphs

15. Hardy, M.: Pareto's law. Math. Intell. **32**(3), 38–43 (2010)

16. Leskovec, J., Sosič, R.: SNAP: a general-purpose network analysis and graph-mining library. ACM Trans. Intell. Syst. Technol. (TIST) **8**(1), 1 (2016)

17. Lovász, L., et al.: Random walks on graphs: a survey. Combinatorics **2**(1), 1–46 (1993). Paul erdos is eighty

18. Massey Jr., F.J.: The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc. **46**(253), 68–78 (1951)

19. Maydeu-Olivares, A., Garcia-Forero, C.: Goodness-of-fit testing. Int. Encycl. Educ. **7**(1), 190–196 (2010)

20. 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 (2002). https://link.aps.org/doi/10.1103/PhysRevLett.88.138701

21. Newman, M.E.: Power laws, Pareto distributions and Zipf's law. Contemp. Phys. **46**(5), 323–351 (2005)

22. 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. Proc. Natl. Acad. Sci. **99**(8), 5207–5211 (2002)

23. Resnick, S.I.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. ORFE. Springer, Newyork (2007). https://doi.org/10.1007/978-0-387-45024-7

24. Stumpf, M.P., Porter, M.A.: Critical truths about power laws. Science **335**(6069), 665–666 (2012)

25. Stute, W., Manteiga, W.G., Quindimil, M.P.: Bootstrap based goodness-of-fit-tests. Metrika **40**(1), 243–256 (1993)