NCSC
TRENDS IN STATE COURTS

IMPROVING COURT STATISTICS BY EXPLORING THE SHAPE OF DATA

Adam S. Cohen
Planning and Program Evaluation Division, Hawai'i State Judiciary

Wesley Nakamoto
Planning and Program Evaluation Division, Hawai'i State Judiciary

Each year, millions of dollars in restitution are owed to victims of crime. Reports suggest that about 80 percent of that money is uncollected (e.g., Colorado Judicial Branch, 2020; Minnesota Restitution Working Group, 2015; Rex and Boyce, 2011). Based on this data, it would appear that the majority of victims are not restored, offenders are derelict in their responsibilities to victims, and court staff are unsuccessful at collecting restitution. But most of the unpaid restitution in a jurisdiction traces back to a small percentage of offenders who owe large sums of money. Most offenders, who typically owe far less restitution, pay victims in full. A deeper understanding of the data offers a more accurate and complete picture and leads to a substantially different conclusion: the majority of victims are restored, most offenders fulfill their responsibility, and court staff are largely effective at collecting restitution.

Instead of a bell curve, restitution and many other data sets follow a "power law." With power-law curves, large values are less common than small ones but frequent enough and large enough relative to the rest of the data that they account for a disproportionate share of the total. Examples

Courts rely on data to make well-informed decisions, but the shape of that data can vary, dramatically changing their interpretation. Distinguishing bell-shaped curves from "power law" curves can improve statistics and assist judges and court administrators grappling with important questions in need of evidence-based answers.

across a range of fields abound: 20 percent of customers generate 70 percent of sales (McCarthy and Winer, 2019); 10 percent of Twitter users produce over 95 percent of all political tweets (Pew Research Center, 2019); and 10 percent of Americans own 80 percent of the wealth in the United States (Saez and Zucman, 2016).

Failing to distinguish between bell curves and power laws can have implications for interpreting court data. If undetected, power laws can mask patterns that are essential to the questions that judges, court administrators, and researchers want to answer. Overlooking power laws can lead to invalid data analyses, unsound interpretations of those analyses, and ultimately suboptimal decision making among court leadership. By identifying data that follow power laws, courts can detect patterns that may better inform decision making.

Here we explain how to leverage power laws to learn from court data. First, we explain differences between bell-shaped and power-law distributions. Next, we show how data can be analyzed and interpreted when it is power-law distributed so that important patterns are not missed. Finally, we show how these analyses can help evaluate programs and inform decisions about allocating limited resources in state courts.

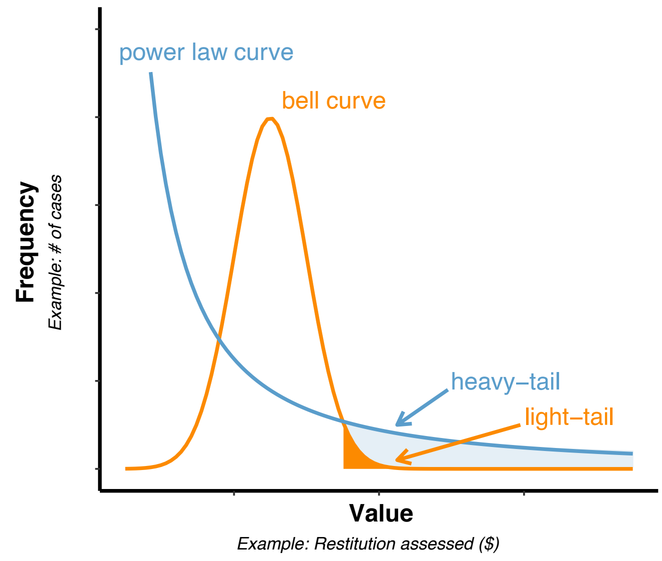
WHY IS IT IMPORTANT TO KNOW THE "SHAPE" OF COURT DATA?

The bell-curve shape of the normal distribution is the most familiar display of data on a graph (see Figure 1). Many variables measured both outside and inside the courtroom are bell shaped. Height, IQ, and the number of hearings in felony cases (Ostrom et al., 2020), for instance, are bell shaped. When data are bell shaped, the mean can be calculated to summarize the data, reducing it to a single number that is most representative of the entire data set. As one moves away from the mean at the peak of the bell curve and into the tails, the probability of extreme values rapidly approaches 0, creating thin or "light tails." Bell curves also tend to have a characteristic scale, spanning no more than a few orders of magnitude. For example, felony defendants have on average 6 hearings with most ranging between 1 to 12 hearings (Ostrom et al., 2020). This is just two orders of magnitude (10^0 to 10^1), and the probability of 100, 1,000, or 10,000 hearings is phenomenally low.

But sometimes data follow a power law, not a bell curve. Data obeying a power law contain many small values and fewer large ones (Figure 1). But these large values out in the tail are more common than in a bell curve, creating "heavy tails." These heavy tails can dominate summary statistics. For instance, many power-law distributions have a mean that goes to infinity, and even when finite, the mean is nowhere near most of the data. Unlike bell curves, power-law data is "scale free," spanning many orders of magnitude, often from 10^0 to 10^7 if not higher.

Besides restitution, examples inside the courtroom include case filings, in which a minority of districts account for most filings across time and case types (Bak, 2006); violent crime, in which a small number of people commit a disproportionate amount of crime (Cook et al., 2004; Falk et al., 2014); and even case citations, in which a small percentage of cases receive the majority of citations in court decisions at state and federal levels, including the Supreme Court (Fowler et al., 2007; Post and Eisen, 2000; Smith, 2007).

FIGURE 1.
NORMAL BELL CURVE VS POWER LAW CURVE



NOTE: Normal "bell curve" distribution (orange) vs power-law distribution (blue) showing number of cases vs. restitution assessed. Restitution appears to follow a power law, not a bell-curve.

TABLE 1.
BELL CURVES VS POWER-LAW CURVES

	Normal Distributions ("Bell Curve")	Power-Law Distributions
Mean/ Median	Meaningful (finite and representative)	Not meaningful
Tails	Light tail (extreme values are rare)	Heavy tail (extreme values more frequent than in the bell-shaped normal distribution)
Scale	Few orders of magnitude Example: felony hearings range from 10^0 to 10^1	Many orders of magnitude Example: restitution ranges from 10^0 to 10^7, if not higher

When data are assumed to be bell shaped but in fact obey a power law, problems arise. The mean, which fails to represent the data, can mistakenly be used to summarize it, and even worse, entered into inferential statistics to test hypotheses. Although the median is an effective substitute for the mean when data are skewed, it is of little match for the heavy tail of power-law curves, where data live nowhere near the median, extending for many orders of magnitude beyond it. But awareness of power-law behavior can remind us to put means and medians aside and look to other statistics. Large values in the "heavy tails" are also problematic, skewing percentages, masking other informative patterns, and producing at best incomplete and at worst misleading answers to the questions we want addressed.

DETECTING COURT DATA THAT FOLLOW A POWER LAW

Since our goal here is to provide a nontechnical overview, we describe statistical approaches for detecting whether court data follow a power law in an Appendix. Judges and court administrators are encouraged to draw on the expertise of court researchers and statisticians at this step because detecting the shape of court data will ultimately help them to better understand the implications and limitations of their data.

ANALYZING AND INTERPRETING COURT DATA THAT FOLLOW A POWER LAW

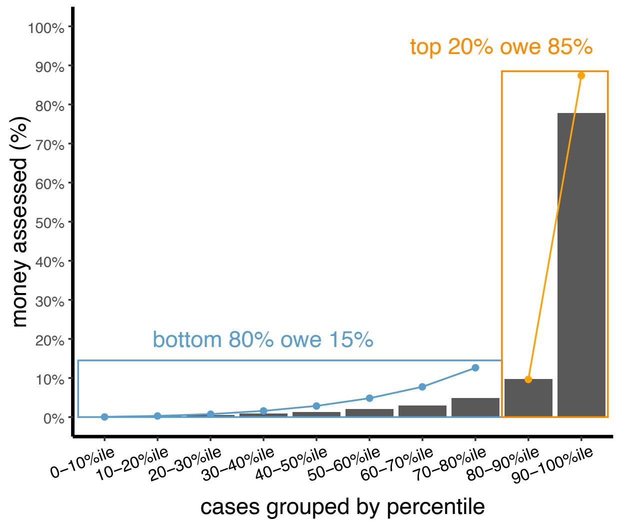
We have already seen the perils of analyzing restitution data while blind to its shape. But if data conform to a power law, how do we analyze it to best evaluate court performance? We can only offer general guidance because the specifics of an analysis will depend on the questions that need to be answered, but whatever form that analysis takes, it must respect the shape of the data. We propose two general approaches to analysis that may be useful for different data sets.

One option is to split data. Even though 80 percent or more of restitution might go uncollected in a jurisdiction, we now know this will be driven by a small number of individuals owing huge sums of restitution. How should the data be partitioned? There are more and less principled ways to do it. Sometimes, the data might have structure that suggests ways to divide it. In the case of restitution, we could consider offense severity, which is related to size of restitution. Felonies typically involve larger restitution amounts than misdemeanors and infractions, and this can be leveraged to disaggregate data. Simulated data disaggregated by severity suggest that even when only 20 percent of restitution money is collected overall, as much as 60 percent is collected from misdemeanants.

Another option is to look at multiple variables. For restitution, we can look at not just the percentage of money collected, which is distorted due to the heavy tails of the power law, but also the percentage of cases that are paid in full, or case-completion rate. The case-completion rate, which *CourTools* Measure 7b also recommends calculating (Schauffler and Ostrom, 2017), can range from 50 percent to 70 percent even when only 20 percent of restitution money has been collected (Colorado Judicial Branch, 2020; Minnesota Restitution Working Group, 2015). To put it another way, if *everyone* in the bottom 80 percent paid their restitution in full, and the top 20 percent paid nothing, monetary collection rates typically will only reach 15 percent, and often lower than that, even though the case completion rate is 80 percent. How much money is collected offers an incomplete and potentially misleading picture of victim restoration, offender rehabilitation, and the collection efforts of court staff. Other variables like the case-completion rate serve as a corrective lens, revealing a more complete picture of restitution.

A power-law analysis highlights the two sides of restitution data. For restitution, high case completion coexists with low monetary collection. This seems like a contradiction, but it is not. It is to be expected when you have "heavy-tailed" data that conform to a power law. The sooner we appreciate power laws and their presence in court data, the sooner an intuition will develop for this type of data.

FIGURE 2.
PERCENT OF RESTITUTION MONEY ASSESSED AMONG BOTTOM 80% AND TOP 20% OF CASES



NOTE: Graphs showing a theoretical sample of restitution assessed broken down by the bottom 80 percent (blue) and top 20 percent (orange). Bars represent amount owed within each decile. Lines represent cumulative amount owed within the bottom 80 percent and top 20 percent. Even if everyone in the bottom 80 percent—by definition, 80 percent of people owing restitution—paid in full, only 15 percent of restitution or less would be collected.

These options are not an excuse to present only part of the data. As misleading as it is to focus on only the low monetary-collection rates, it is equally misleading to focus on only the higher monetary-collection rates of a subgroup or only the higher case-completion rates. Power laws justify partitioning the data but also require presenting each as part of the whole. They justify looking at multiple variables but not selectively ignoring any. Also these analysis options do not indiscriminately lead to "better looking" court statistics; they lead to more accurate ones, whichever way the data fall. The overarching point is that knowing what kind of data you have—bell shaped vs. power law—improves data analysis, leading to more valid performance measures and the discovery of patterns that would otherwise be missed.

More broadly, how do analyses for power-law data fit into the larger toolbox of court statistics, especially *CourTools*, a set of court performance measures developed by the National Center for State Courts? They can be thought of as a complement to *CourTools*, extending the toolbox to an important class of data that traditional statistics are not designed to handle.

EMPLOYING POWER-LAW DATA IN PROGRAM EVALUATION AND DECISION MAKING

Accurately interpreting court data that behave according to a power law can help evaluate programs and inform decisions about allocating limited resources in state courts.

Responsive courts seek meaningful feedback from data on how well they are doing and then adapt based on this feedback (Ostrom and Hanson, 2010). All decisions that flow from this feedback depend on court statistics that are valid and comprehensive. To this end, distinguishing power-law data from bell-shaped and other distributions is critical, and several tools for helping with this determination have been provided (see Appendix). Once detected, state courts can leverage this knowledge to analyze and interpret data more carefully and in ways that allow researchers, judges, and court administrators to evaluate programs with a more accurate and complete set of performance measures.

Power laws can also reveal different courses of action available to courts. If courts are focused on improving monetary collection for restitution, knowing the data are power-law distributed suggests putting resources toward collecting from the infrequent big-dollar cases, focusing on the 85 percent or more of restitution owed by the top 20 percent. If courts are interested in restoring as many victims as possible, then courts could put more resources toward collecting from the more common small-dollar cases, focusing on the bottom 80 percent who owe 15 percent or less of the total restitution. Courts can then use this information alongside their goals and priorities to decide how best to allocate limited resources among these multiple courses of actions.

CONCLUSION

Courts increasingly rely on data to make decisions. We hope that by drawing attention to power laws, state courts can move closer to developing statistics that not only are valid but also provide deeper, more complete answers to questions of importance for decision making. State courts clearly stand to benefit, but so does the public. As the Conference of State Court Administrators observes, "State courts must be proactive in the measurement of their performance with empirical, credible tools. . . . State court leaders are in a position to provide empirical data on which the public can make judgments about the effectiveness of state court systems, rather than judgments based on inaccurate or anecdotal information" (COSCA, 2008).

REFERENCES

Bak, T. (2006). "Power-Law Distributions and the Federal Judiciary." 46 *Jurimetrics* 139.

Colorado Judicial Branch (2020). "Annual Statistical Report." Perma link https://perma.cc/RF4L-VP6W.

Cook, W., P. Ormerod, and E. Cooper (2004). "Scaling Behaviour in the Number of Criminal Acts Committed by Individuals." *Journal of Statistical Mechanics: Theory and Experiment*, no.7, P07003.

Conference of State Court Administrators. (2008). "White Paper on Promoting a Culture of Accountability and Transparency: Court System Performance Measures." Perma link: https://perma.cc/9YAM-5N6J.

Falk, Ö., M. Wallinius, S. Lundström, T. Frisell, H. Anckarsäter, and N. Kerekes (2014). "The 1% of the Population Accountable for 63% of All Violent Crime Convictions." 49 *Social Psychiatry and Psychiatric Epidemiology* 559.

Fowler, J. H., T. R. Johnson, J. F. Spriggs, S. Jeon, and P. J. Wahlbeck (2007). "Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court." 15 *Political Analysis* 324.

McCarthy, D. M., and R. S. Winer (2019). "The Pareto Rule in Marketing Revisited: Is It 80/20 or 70/20?" 30 *Marketing Letters* 139.

Minnesota Restitution Working Group. (2015). "Report to the Legislature." Office of Justice Programs, Minnesota Department of Public Safety, January. Perma link: https://perma.cc/ZYE4-K5AT.

Ostrom, B. J., and R. A. Hanson (2010). *Achieving High Performance: A Framework for Courts*. Williamsburg, VA: National Center for State Courts.

Ostrom, B. J., L. E. Hamblin, R. Y. Schauffler, and N. Raaen (2020). *Timely Justice in Criminal Cases: What the Data Tells Us*. Williamsburg, VA: National Center for State Courts.

Pew Research Center. (2019). "National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets." October. Perma link: https://perma.cc/DHN9-6HRS.

Post, D. G., and M. B. Eisen (2000). "How Long Is the Coastline of the Law? Thoughts on the Fractal Nature of Legal Systems." 29 *Journal of Legal Studies* 545.

Rex, J., and E. Boyce (2011). "The Vermont Model: A Victim-centered Approach to Restitution." Vermont Center for Crime Victim Statistics. In *Making Restitution Real: Five Case Studies on improving Restitution Collection*. Washington, DC: National Center for Victims of Crime. Perma link: https://perma.cc/L3JH-VVVQ.

Saez, E., and G. Zucman (2016). "Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data." 131 *Quarterly Journal of Economics* 519.

Schauffler, R., and B. Ostrom (2017). "ReTooling *CourTools*: Legal Financial Obligations and the New Measure 7." In D. W. Smith, C. F. Campbell, and B. P. Kavanagh (eds.), *Trends in State Courts 2017*. Williamsburg, VA: National Center for State Courts.

Smith, T. A. (2007). "The Web of Law." 44 *San Diego Law Review* 309.

APPENDIX: DETECTING POWER-LAW DATA

DETECTING COURT DATA THAT FOLLOW A POWER LAW

This Appendix is a guide to detecting power-law curves in court data and is intended for court researchers, statisticians, and anyone else working directly with the data.

We describe two approaches to detecting power-law distributions: the visualization and the fit-test-compare approach. Which one to use depends on weighing a tradeoff. The visualization approach is simple but more error prone and susceptible to diagnosing a power law where none exists. The fit-test-compare approach is more complex but more rigorous, requiring stronger evidence to establish a power law.

The Visualization Approach

The visualization approach involves little more than replotting the data on a log-log plot. If the data fail to form a straight line, it does not follow a power law. Assuming we have already measured our variable of interest (e.g., restitution assessed) and graphed the frequency distribution, we take the logarithm of the values on the x and y axes to create a log-log plot. Why do this? A power law is described by,

$$p(x) \propto \frac{C}{x^{\alpha}}.$$

Letting $y = p(x)$, taking the logarithm of each side, and using the power rule of logarithms, we get

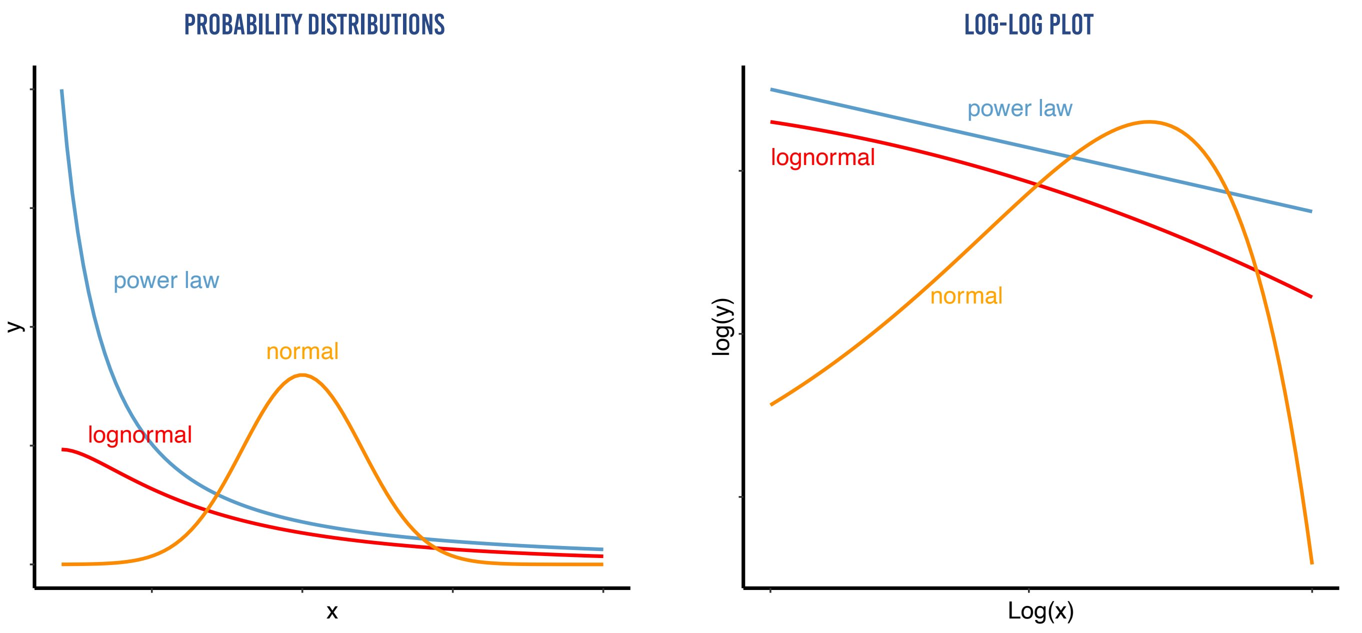
$$\log(y) = -\alpha \log(x) + \log C.$$

But notice that this has the same form as the equation of a line, $y = mx + b$, where m is the slope and b is the y-intercept:

$$\underbrace{\log(y)}_{y} = \underbrace{-\alpha}_{m} \underbrace{\log(x)}_{x} + \underbrace{\log C}_{b}$$

This tells us that if the data follows a power law, plotting log(y) on the y-axis against log(x) on the x-axis—a log-log plot—should yield a straight line with slope -α. If the data do not obey a power law, then the data will not follow a straight line.

SUPPLEMENTAL FIGURE 1.



Probability distributions for the normal, lognormal, and power law (left panel). Log-log plots of the same distributions (right panel). The orange curve on the log-log plot is inconsistent with a power law. The blue and red lines are consistent with a power law, but only the blue line comes from an actual power-law distribution.

This approach is simple but not without drawbacks. First, it involves a subjective judgment about how well the data form a straight line. Some distributions suspected to follow a power law based on visualization do not in fact obey a power law when examined with quantitative, statistical tests (Broido and Clauset, 2019; Clauset, Shalizi, and Newman, 2009; Stumpf and Porter, 2012). Additionally, few empirical data sets follow a power law for all values of x, and a visualization provides no clear way of deciding where power-law behavior begins and ends in a data set. Not only that, one can easily be misled in their judgment about a straight line by confirmation bias, the tendency to see what one wants or expects to see.

Second, a straight line on a log-log plot does not uniquely identify power laws. Other models besides power laws generate straight lines in a log-log plot (see Supplemental Figure 1, right panel). A more objective approach is no help here: statistically analyzing the log-log plot (e.g., with least squares linear regression) would also fail to rule out other models that follow a straight line. A straight line is necessary but not sufficient for a power law. In other words, if the data do not fall on a straight line, we can assert with confidence that the data are not power-law abiding. However, if the data do fall on a straight line, we can only say the data are consistent with a power law, as well as with other distributions. The visualization approach allows us to reject but not confirm a power law.

Another visualization approach, the Pareto Q-Q plot, is not covered here but is a variation on the normal Q-Q plot. Software packages are available that help generate various Q-Q plots (e.g., Reynkens, 2020). The Q-Q plot can be used to visually compare power laws to different models, but, like the log-log plot, does not involve any formal quantitative tests.

The Fit-Test-Compare Approach

For those who dream about fitting power-law distributions to court data, testing the "goodness" of the fit, and comparing the fit of power laws to other probability distribution models, there is the fit-test-compare approach (Clauset, Shalizi, and Newman, 2009). The first step involves estimating parameters. For example, if we wanted to fit a straight line, y = mx + b, to a set of data, we would need to estimate the parameter values, m and b. In the same way, the power-law probability function is a general equation with its own parameters to be estimated. The first step asks the question, if the data followed a power law, what parameter values would produce the best fit to the data?

Although the power-law model is fit to the data in step one, it does not tell us whether the model provides a good fit. Step two asks, how plausible is the power-law model for the data? Statistical tests can answer how likely the data came from a power-law distribution compared to chance.

Even if the power law provides a plausible match for the data at step two, the power law might not be the only good match for the data. Step three asks, does the power-law model fit the data better than other models? The logic is that if the best-fit power-law model at step one is a plausible fit for the data at step two and if it also fits the data better than other models at step three, we would have strong quantitative evidence for a power law.

Although the fit-test-compare approach is more rigorous than the visualization approach, it is more complex. Fortunately, computer code (including packages for Matlab, Python, and R) for the fit-test-compare approach is available. To get started, see the following:

Matlab: https://aaronclauset.github.io/powerlaws/
(Perma link: https://perma.cc/Q3B9-PJ58)

Python: https://github.com/jeffalstott/powerlaw
(Perma link: https://perma.cc/35VL-YWEB)

R: https://github.com/csgillespie/poweRlaw
(Perma link: https://perma.cc/Z7Y5-YXZF)

Does It Matter Whether Court Data Follow a Power-Law Distribution vs. Other Heavy-Tailed Distributions?

For the practical purposes of the courts, often it will not matter whether data follow a power-law or other heavy-tailed distribution. After all, power laws are interesting to courts *because of their heavy-tails* and the alternative analyses they invite. Under these looser conditions, the log-log plot from the visualization approach may be sufficient for diagnosing "power law-like" behavior, even if it cannot guarantee the data obey a power law. In contrast, researchers working with formal theories and others needing more accurate diagnostics will likely benefit from the fit-test-compare approach, which can better distinguish between power laws and related distributions.

REFERENCES

Broido, A. D., and A. Clauset (2019). "Scale-Free Networks Are Rare." 10 *Nature Communications* 1.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). "Power-Law Distributions in Empirical Data." 51 *SIAM Review* 661.

Reynkens, T. (2020). "Using the ReIns Package." May 16. Perma link: https://perma.cc/GPA4-LCQS.

Stumpf, M. P., and M. A. Porter (2012). "Critical Truths About Power Laws." 335 *Science* 665.