A Cliometrics and Complexity Perspective on Ancient Greek Culture

Laurent Gauthier

A Cliometrics and Complexity Perspective on Ancient Greek Culture*

Laurent Gauthier†

August 11, 2022

Abstract

Ancient history appears to be the *persona non grata* of cliometrics: not a mention in the field's manual (Diebolt and Haupert 2019). Recent research combining cliometrics and complexity (Gonzales-Feliu and Parent 2016; Abry et al. 2022; Bastidon and Parent 2022), if it has not addressed Antiquity either, nevertheless opens the door to a new perspective on ancient Greece. By focusing on primary non-economic sources thanks to complexity sciences, a cliometric approach allows us to frame into economic terms the massive and complex cultural material we gathered from the Ancients. Concentrating on the generating mechanisms for various large-scale textual sources, in a complexity economics perspective, we consider how distinct constraints and objectives lead to measurable differences between judicial speeches, poetry, and epigraphic texts, for example. Based on thousands of inscriptions, we account for whom the Ancients worshiped with a shared-resources paradigm, and show how people's names and naming strategies exhibit patterns consistent with a high degree of conformity. Finally, we show that the religious formulae the Greeks used to address their gods, in spite of their short and condensed nature, share the features of a language, and appear to be constructed as an optimal communication medium.

Keywords: Ancient Greece, cliometrics, complexity, power laws, networks

* **Draft working paper** - I wish to thank Antoine Parent, from Paris 8 University, for his numerous comments.

† Email: laurent.o.gauthier@gmail.com. Laboratoire d'Économie Dionysien, Université Paris 8 Saint-Denis Vincennes, EA 3391 - Bâtiment D, 2, rue de la Liberté, 93526 Saint-Denis, France.

Diebolt and Haupert (2020) have recently argued that modern cliometrics had helped further both economics and history by providing carefully grafted datasets, thanks to the importance of data for the cliometricians' uses, and thanks to their advanced econometric techniques. This is certainly not apparent in the study of ancient history, seemingly the *persona non grata* of cliometrics. Cliometrics have indeed essentially ignored the period. As a matter of fact, the field's handbook (Diebolt and Haupert 2019) does not contain a single reference to Antiquity. Surely, the notion of an ancient economy is still hotly debated, even among historians (Hobson 2014), and there is little economics-related data pertaining to the period. It seems hence logical that the advances in cliometrics touted in Diebolt and Haupert (2020) have not offered much help to ancient history. On the other hand, the conjunction of cliometric perspectives with complexity sciences into a new intellectual endeavor, cliometrics & complexity, has significantly widened the reach of economic history in general. Using complexity economics notions, whether they relate to networks, to time series, or to the emergence of patterns, has renewed the cliometric perspective on economic or financial questions in the contemporary period (Bastidon et al. 2019, 2020; Bastidon and Parent 2022; Abry et al. 2022). While the use of networks is not something new in contemporary history (Lemercier 2012), the cliometrics & complexity approach, by considering complex generating mechanisms based on the agents' behavior, offers deeper insights. Combining complexity economics, such as networks and non-linear econometrics, with institutional economics also offers new insights in existing cliometric issues (Le Riche, Parent, and Zhang 2022). Stepping further away from considerations about the economy *stricto sensu*, Gonzales-Feliu and Parent (2016) examined military strategy from the perspective of an optimal allocation on a road network, in an effort to deconstruct the logic that prevailed in particular military choices in the past.

Staying in the path of cliometrics & complexity, we propose an economics-inspired analysis of various facets of ancient Greek culture, serving as an illustration of using new methodological angles in the study of the past through economics. We consider that advanced econometric methods do not have to be restricted to the study of the economy, but instead that large amounts of data can always benefit from quantitative scrutiny and be put into perspective through economic models. We apply econometric and economic analyzes, inspired by complexity models, to several important datasets pertaining to ancient Greece, directly derived from the Ancients' literary or epigraphic production. The epistemic distinctions between economics and historiography have been examined in Gauthier (2022), and one of the salient conclusions was that if economics are to contribute results that can be intelligible to historians, and further their understanding of the past, it is important to focus on primary rather than secondary sources. Focusing on primary sources, which are large datasets, recoups well with cliometrics and complexity approaches, designed to analyze the emergence of patterns. Historians usually consider primary sources at the atomic level. In ancient history, he philological analysis of a single inscription, or the occurrence of a particular name in a text, can often give rise to a detailed commentary, based on the historian's knowledge of the context in which the historical documents were produced. Economic analysis of a particular situation, typically using game theory, may too shed light on such historical analysis: this is the focus of analytic narratives, such as Mongin (2018), looking at strategy at the battle of Waterloo, or Crettez and Deloche (2018), considering Caesar's decision to step into the Senate where he was killed. However, cliometrics can offer a new perspective by focusing on primary sources through the lens of complexity economics, based on fundamentally different hermeneutics. Given the categorical nature of the data, one can indeed examine it in a distributional perspective, and ask what generating mechanism may have produced it, relying on complexity theory. The analysis of generating mechanisms, and the identification of the conditions for their optimality in terms of human behavior, nevertheless belong to economic analysis.

Given the importance of properly identifying the nature of the distributions observed in the data, we begin by discussing some common distributions applicable to categorical data, their generating mechanisms and practical aspects in their fitting. Then, we will turn to the analysis of the main textual sources, literary texts and inscriptions, as well as a particular subset of inscriptions concerning votive acts. We will examine them under the light of economics, either applied to linguistics or to religion sciences. Finally, we will examine some primary sources that can be structured into networks: the records of people's names, along with parent-child relationships, and full religious invocations, known as divine onomastic sequences, recorded from inscriptions, which create a network of gods and qualifiers. The list of primary sources pertaining to ancient Greece we discuss here is not intended to be fully comprehensive, but covers a fairly wide range of domains. For the most part, the primary sources of interest to us are not designed to be electronically available in bulk: this is a reflection of the manner in which classicists use the data. As a result, it is necessary to program specialized software to exploit the websites where the information is available as very large numbers of narrow subsets. Simply extracting and arranging these datasets requires programming expertise and powerful systems; this may explain why, so far, they have not often been exploited in bulk by historians.

1 Identifying Distributional Patterns and Mechanisms

Most of the data sources we will discuss here contain large amounts of data of a categorical nature. Before we begin to drill into them, and try to understand various aspects of ancient Greek culture under this light, it makes sense to establish what kind of patterns one may naturally expect from such data. Indeed, one could observe apparent regularities; what is the most direct manner in which to visualize them? It is worth noting that the analysis of distributions is not particularly common in cliometrics, nor when in relation with complexity economics, as the data is rarely categorical in nature. The empirical analysis of distributions of unstructured data representing complex social or physical phenomena is nevertheless part and parcel of complexity sciences (Boccara 2010). Most primary sources for ancient history being categorical or textual, the most direct handle to approach them is through a distributional perspective. In some cases, when dating is available, this may be combined with a diachronic analysis, but we will restrict ourselves to synchronic analyzes for the sake of simplicity. Since most traditional economic data is precisely situated through time, the question of a particular regime change, such as in Bastidon et al. (2020) or Abry et al. (2022), can lead to interesting insights, but we do not flow this angle here.

1.1 Some Useful Distribution Classes

Given the nature of the data at hand, we will typically seek to analyze the relationship between the size of some category (whether it be the number of times a word appears in a text corpus, or the number of votive acts a given god received, for example) and its rank. This can also equivalently be understood as considering the counter cumulative empirical probability: for an item $i \in [1..I]$ (a word, a god) with size s_i, its rank r_i is the number of items with a size greater than s_i, so that $r_i = |\{j \in [1..I] : s_j \geq s_i\}|$. Hence the empirical cumulative probability for the random variable representing the size S is $\mathbb{P}[S \geq s_i] = \frac{r_i}{I}$.

The kind of metrics we observe tend to be positive, and for the most part, their frequencies are strictly decreasing as a function of the observed values. While there is an infinite number of random distributions that may account for these patterns, such positive observations are often compared with power laws, exponential distributions or truncated power laws, because these distributions are found in a large number of empirical phenomena and have strictly decreasing densities. These distributions are defined as follows:

- If X follows a power law of parameter[1] α and minimal value $x_{\min}$, we have

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x>x_{\min}} \frac{\alpha}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-1-\alpha} dx.$$

  A particular case of power law is Zipf's law, where $\alpha = 1$ and in which case the cumulative probability $\mathbb{P}[X > x]$ has the form $\frac{1}{x}$;

- If X follows an exponential distribution of parameter λ, shifted to take values above $x_{\min}$, we have

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x>x_{\min}} \lambda e^{-\lambda(x-x_{\min})} dx;$$

- If X follows a truncated power law of parameters α and λ, with minimal value $x_{\min}$, we have

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x>x_{\min}} \frac{\lambda^{-\alpha}}{\Gamma\left(-\alpha, \lambda x_{\min}\right)} e^{-\lambda x} x^{-1-\alpha} dx.$$

When considering distributions for which the frequency may not be decreasing, the lognormal distribution, as the exponential of a Gaussian, is a good candidate:

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x>0} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx.$$

The lognormal distribution can be shown to be asymptotically equivalent to a power law, for intermediate values (Saichev, Malevergne, and Sornette 2010, 12).

Power laws are fat-tailed. When one talks of a fat-tailed distribution for some positive variable X, it means that $\mathbb{P}[X > x]$ does not go quickly to zero as x grows; there is some non-negligible probability of extreme values for X. A fat-tailed distribution for a continuous variable X translates into a particularly wide distribution. When we consider categorical data, such as names or words, then it is not the data *per se* that follows a certain distribution, but the number of observations in each category. If S represents the number of observations in each such bucket, then saying that S is fat-tailed means that there should be particularly few buckets with a large number of observations, and many with very few observations. Hence, it is important to note that, for categorical data, a fat-tailed distribution of occurrences rather translates in heavily concentrated observations.

Power laws are commonly used in complexity science, because they frequently appear in nature

[1] We follow the convention, more common in economics than in physics, of specifying the parameter as the exponent of the cumulative distribution, rather than that of the density. We hence have the same convention as Gabaix (1999), but that is different from that of Clauset, Shalizi, and Newman (2009), for example.

and in many human phenomena, and may be produced by numerous types of underlying processes (Boccara 2010; Mitzenmacher 2004). Power laws have indeed been applied in many aspects of economics (Gabaix 2016). The generating mechanisms for these distributions are generally associated with preferential attachment in network formation (Barabási and Albert 1999), or with random growth processes (Gabaix 1999). Power laws can also be shown to correspond to the limiting behavior of extreme fluctuations (Alfarano and Lux 2010). One particular case of power law is Zipf's law, with parameter α set to 1. Zipf initially observed that the use of words in human language, measured as the relationship between the frequency of occurrence of words in natural language and their frequency rank, followed this particular distribution (Zipf 1949, 19–55). This pattern has been extensively studied in computational linguistics, and many models accounting for the emergence of this pattern in language have been proposed, although some have argued they cannot truly explain why language exhibits it (Piantadosi 2014). Power laws, including Zipf's law, fundamentally recoup with economics in the sense that they can be shown to emerge as the result of optimization. This first came to light when Mandelbrot (1953) showed that Zipf's law was optimal for communication, by considering there is a cost of using a word depending on the number of characters in it, and that the informational content in the word depends on its probability of use. More complex modeling of the communication process, factoring in the cost of encoding and decoding information, confirms the optimality of Zipf's law for the use of words, as it reduces coding errors (Ferrer i Cancho and Solé 2003).

Making the generating mechanisms more general, for example in network formation, results in a richer class of distributions, for which the straight power laws form approximations (Jackson and Rogers 2007). The more general truncated power law, with an exponential tail, can be associated with random group formation. Indeed, Baek, Bernhardsson, and Minnhagen (2011) show that forming groups off of any kind of categorical data in order to minimize the cost of identifying a given element results in a truncated power law distribution for the group size. Exponential distributions are generally associated with random times, or surprise, due to their "lack of memory" through conditioning. Ely, Frankel, and Kamenica (2015) modeled the notion of surprise maximization from an economic perspective, but do not characterize the optimal distribution of surprising events in general. Borwein, Borwein, and Maréchal (2000), who solve for the so-called "hangman's paradox" (how to best surprise a prisoner on the day of their execution), show that over long periods of time, the maximum surprised is reached with a partially exponential distribution. Connections have been found between exponential and power law distributions: the emergence of power laws can also be related to exponential processes by combining random event times with random observation times (Reed and Hughes 2003). Exponential distributions can also emerge in networks: in a non-equilibrium network, continuously growing, if the new nodes appearing link to the existing nodes following a uniform distribution, then the network's degree distribution is exponential, with a parameter driven by the average degrees in the network (Deng et al. 2011).

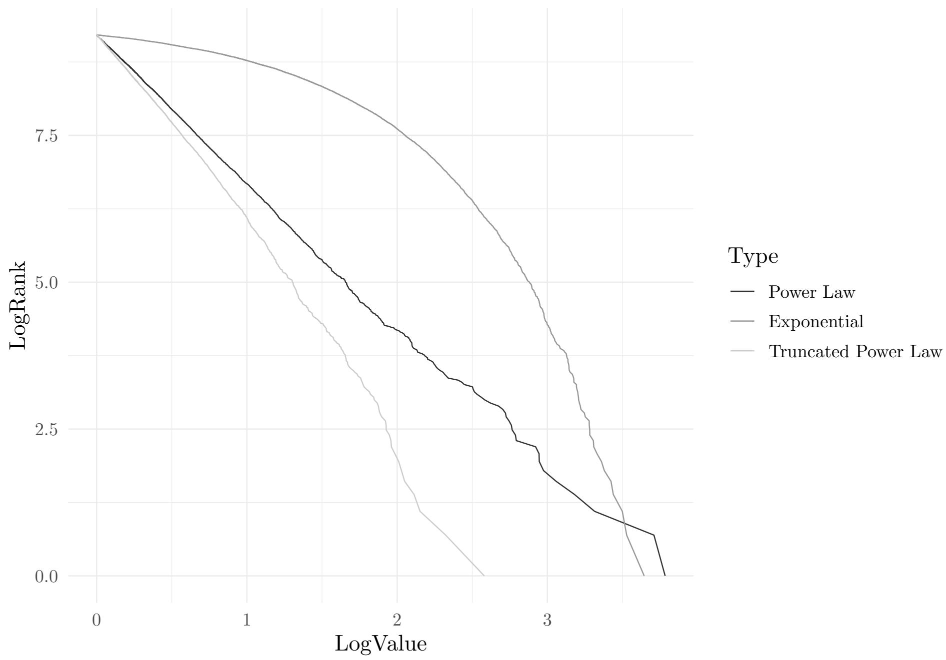
Considering the lognormal distribution, which we mentioned earlier, simply as the exponential of a normal distribution does not help much in understanding it, unless one has a specific reason for the exponentiation, such as exponential growth. Limpert, Stahel, and Abbt (2001) have noted the widespread occurrences of the lognormal distribution in biological phenomena and growth mechanisms, and point out that it is a good candidate for situations where the minimal size cannot be the most commonly occurring. Some economic data that were initially thought to follow Zipf's law have been shown to follow lognormal distributions, such as firm and city sizes (Saichev, Malevergne, and Sornette 2010, 2–5).

1.2 Empirical Distribution Comparisons

Figure 1 plots simulations of the first three types of distributions mentioned above, for a large number of random draws. The parameters are chosen so that the curves are relatively close to each other on the logarithmic scale. Their shapes are clearly distinguishable: the straight line of the power law and the curvature of the exponential stand out. The truncated power law's distribution plot shows a straighter part on the left side (for high frequency, low value cases), and a curved part on the right side, where x is large and the exponential behavior of the exponential dominates the density. The exponential's effect hence affects the tail, the largest observations, while the lower values retain a power law-like behavior. In this example, one may not need advanced statistical methods to establish the nature of these distributions: we can clearly see that the log/log plot, while heavily compressing the scale of the ranks and outcomes, does not betray the underlying density. It is not just any randomly generated numbers that, once ranked and scaled, will exhibit the patterns visible in Figure 1.

However, Clauset, Shalizi, and Newman (2009) offered a detailed treatment of the statistical methodology that can help empirically determine the type of probability law followed by some data, specifically in the case of power laws. They have shown that it is paramount to carry out proper distribution fit comparisons using a maximum likelihood approach, rather than simply carrying out a linear regression on the data on a logarithmic scale. According to their analysis, fitting a regression line on a log/log plot, such as was done in Gabaix (2016), for example,

Figure 1: *Comparison of Power Law, Exponential and Lognormal Distributions on a Large Sample*



Note: The parameters for the distributions are $\alpha = 2.5$ and $\lambda = 0.25$. The samples contain 10000 draws.

results in a gross over-estimate of the probability that the data follows a power law. Using the methodology from this implementation by Alstott, Bullmore, and Plenz (2014), we can compare the fits of the three distributions, as shown in Table 1.

We can see that the parameters estimates are very close to their true values, if we know what distribution to look up in each case. The goodness-of-fit ratios are clearly able to disambiguate the choice between power laws or truncated laws and the exponential. When we simulate a truncated power law, the goodness-of-fit also indicated a preference for that form over a simple power law. However, when we simulate a power law, the tests cannot cleanly distinguish it from a truncated power law. In this case, however, the estimated parameter in either case is close to that of the underlying simulation, and the decay rate is very small. Given that the truncated distribution with a small λ can come arbitrarily close to the straight power law, this is not surprising.

1.3 Dealing with Sample Size

If the categories we are interested in are words in a large text corpus, or individual persons in some historical record, or individual physical inscriptions across a large geographic space, then

Table 1: *Summary Statistics on Distribution Fits*

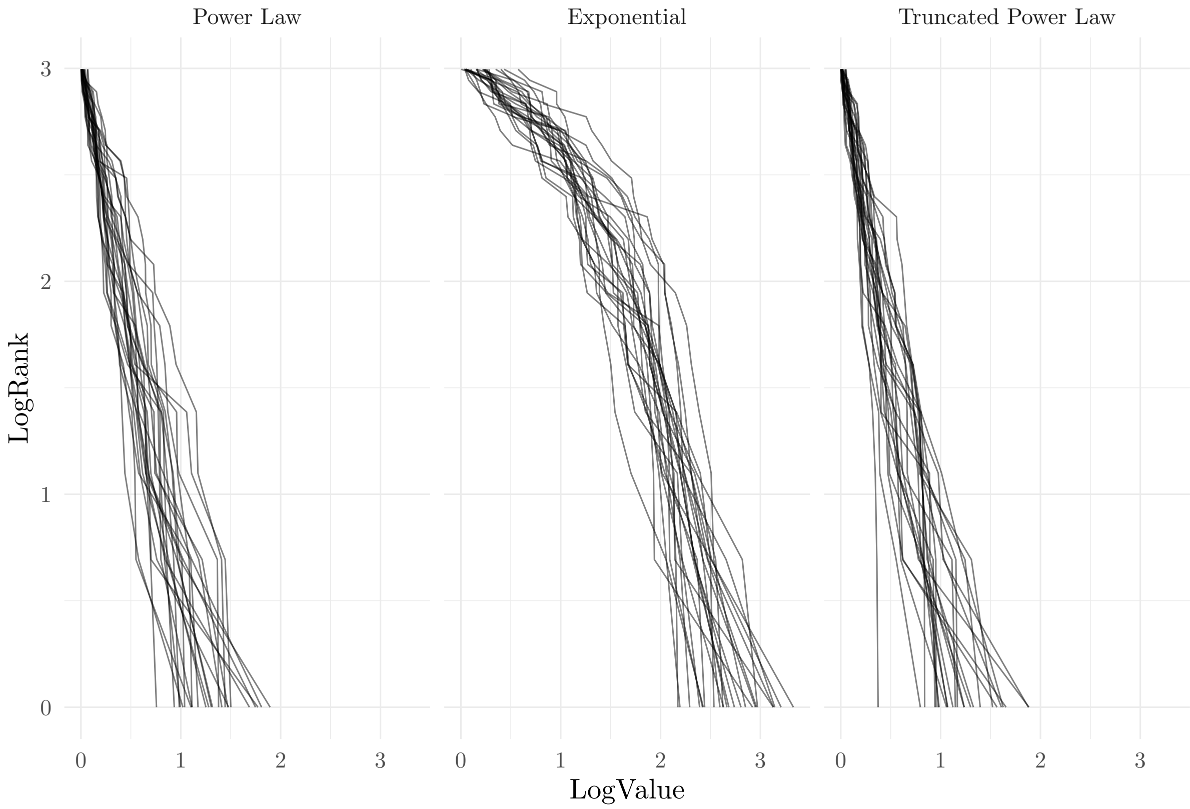
Statistic	Exponential	Power Law	Truncated Power Law
Lambda Exp	0.250	1.526	2.073
Alpha Pow	0.745	2.515	3.015
Alpha Trunc	0.000	2.493	2.666
Lambda Trunc	0.099	0.008	0.167
Trunc vs Pow R	80.357	0.332	2.464
Trunc vs Pow p	0.000	0.545	0.000
Trunc vs Exp R	-18.986	9.770	10.001
Trunc vs Exp p	0.000	0.000	0.000
Pow vs Exp R	-46.547	9.718	9.020
Pow vs Exp p	0.000	0.000	0.000

Note: The distribution names in the tests are abbreviated as follows: *Exp* = exponential, *Pow* = (pure) power, *Trunc* = power law with exponential decay. *R*: ratio of goodness-of-fit; a positive number means that the first law of the two is preferred. *p*: significance level; the probability that the preference would be due to randomness. The same abbreviations are used in other comparable tables.

the number of these categories is very large, and sample sizes as well. However, in some cases one may not find so many different categories, when considering individual gods worshiped in a given city, or distinct characters across ancient Greek theater, for example. Figure 2 shows the empirical cumulative distributions for a series of small samples from the same distributions as those represented in figure 1. Even with few observations in each case, and the variability across samples, we can see that the exponential distribution and power laws are markedly different. The behavior of the exponential for low ranks and low values gives it away. Since the effect of truncation on the power law is only apparent on the much scarcer high value and high rank outcomes, the distinction between the truncated and pure power law is naturally more difficult.

Distribution fits can be applied to each small sample, and the resulting statistics once aggregated are displayed in Table 2. The median significance levels for the goodness-of-fit are not very high across the board: there is a high probability of the difference in distributions to be fortuitous. For power law and exponential distribution draws, the median parameter estimate comes out close to the real value, but that is not the case for the truncated power law; here again a consequence of the additional degree of freedom in fitting to a small number of observations. These results illustrate that, with a reduced sample, it may still be possible to distinguish between exponential and power laws (just as one could see it on the chart, in Figure 2), but the more flexible nature of truncated power laws makes them more difficult to determine.

Figure 2: *Comparison of Power Law, Exponential and Lognormal Distributions on Small Samples*



Note: The parameters for the distributions are $\alpha = 2.5$ and $\lambda = 0.25$. The 25 samples for each law contain 20 draws.

Table 2: *Summary Statistics Across Small Sample Distribution Fits*

Statistic	Exponential (Med.)	Exponential (St. Dv.)	Power Law (Med.)	Power Law (St. Dv.)	Truncated Power Law (Med.)	Truncated Power Law (St. Dv.)
Lambda Exp	0.265	0.051	1.741	0.801	2.174	0.584
Alpha Pow	0.750	0.073	2.707	0.825	3.055	0.567
Alpha Trunc	0.000	0.000	2.001	1.194	1.477	1.428
Lambda Trunc	0.112	0.032	0.351	0.595	0.894	0.895
Trunc vs Pow R	4.280	1.787	0.347	0.523	0.840	0.988
Trunc vs Pow p	0.006	0.007	0.739	0.295	0.576	0.374
Trunc vs Exp R	-1.464	1.070	1.056	0.872	0.273	1.146
Trunc vs Exp p	0.119	0.158	0.222	0.269	0.371	0.276
Pow vs Exp R	-2.842	1.364	0.847	1.103	-0.094	1.467
Pow vs Exp p	0.045	0.093	0.357	0.296	0.336	0.321

Since the mechanics that can generate the observed distributions can be quite different, and correspond to distinct optimal behavior by agents, it is important to be able to differentiate between these distributions. The analysis of the controlled experiments above has shown us to what extent we may, in practice, rely on the statistical tests proposed by Clauset, Shalizi, and Newman (2009).

2 Ancient Greek Writings and Cliometrics

Texts, more than any other artifact from the past, are the prime raw material of historiography. The most natural primary source for the study of ancient Greece, logically, is the corpus of all recorded literature, covering the range from the Archaic period to late Antiquity. These texts have reached us, for the most part, through manuscripts which were copied through time; there is essentially no instance where we possess original literary writings. There are also numerous inscriptions, many with lacuna or missing characters, but they nevertheless constitute quite a large sample. Hence, the corpus of ancient Greek texts available today is the result of a combination of chance findings as well as specific choices that were made over more than 2,500 years about which works deserved conserving, and which ones did not. In spite of these layers of selection and filtering, the texts we have today belong to a fairly diverse set of genres. In this section, we will concentrate in turn on two large ancient Greek text corpora: the Diorisis, which gathers literary works, and the PHI, which gathers inscriptions. Then, we will discuss votive acts data extracted from inscriptions.

It is sometimes helpful to leverage historical work in order to categorize data from primary sources. One important resource for this purpose is the voluminous compilation by Hansen and

Nielsen (2004). These authors gathered information on a large variety of categorizations and metrics applicable to each *polis*, for instance such as its size, political regime, or affiliation. The POLIS database (Johnson and Ober 2014) is the computerized and augmented version of this inventory, covering over 1,000 *poleis* of the Greek world. The majority of *poleis* listed in the POLIS database have geographic coordinates and in many cases the primary sources can be mapped to the polity-level information from POLIS[2].

2.1 The Economy of Literary Genre

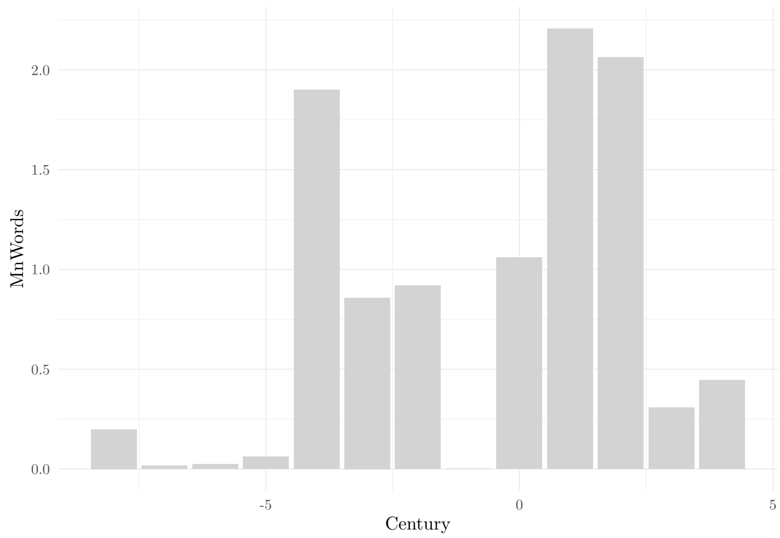
There are multiple electronically available resources compiling ancient Greek texts, such as the Thesaurus Linguae Graeca, or TLG, (Pantelia 2020), or the large Perseus website hosted by Tufts University (Crane 2012). These resources are nevertheless generally not absolutely comprehensive, do not allow users to download the data as one unique set, and only contain the raw text. For inflected languages such as ancient Greek, morphological inflection makes the identification of variations of a same word difficult. In order to associate any instance of a word to its lemma, the noninflected root, and hence to identify each word's role in a sentence, the words need to be categorized, with a so-called part-of-speech tagger (Ide 2004). The Diorisis is a centralized and comprehensive corpus of ancient Greek text that is already tagged (Vatri and McGillivray 2018b). In total, this corpus gathers 820 different works; it was originally composed in order to analyze semantic change in ancient Greek over time, with computational linguistics methods, and is aimed at historians and classicists. The Diorisis data contains one row per word or punctuation sign, for a total of over 14m rows. A lemma is mapped to each word, representing its root: the nominative form for a noun, or the singular first person of the present indicative for a verb in most cases. The particular declension or conjugation of the word is also specified. All the corpus's data is available online, in the form of one file per book (Vatri and McGillivray 2018a).

Since the Diorisis contains over 800 different works, it is useful to be able to categorize these works. A practical catalog has been compiled for that purpose by Tauber (2021) For each one of the works, it gives a year estimate, and a two-level literary genre category. Combining the catalog data with the Diorisis corpus, we can carry out a wide range of analyzes. Figure 3 shows the distribution of the number of words in Diorisis as a function of the year when the work is estimated to have been composed. We can see the large concentrations in the traditional classical

[2] The POLIS data itself, although many layers removed from primary sources, has been independently used in economics research: for example Fleck and Hanssen (2018) focused on political transition using this data exclusively.

period (5th and 4th centuries), in Hellenistic Greece afterwards, and during the Roman Empire in the early Christian period.

Figure 3: *Distribution of the Number of Words by Work Period*

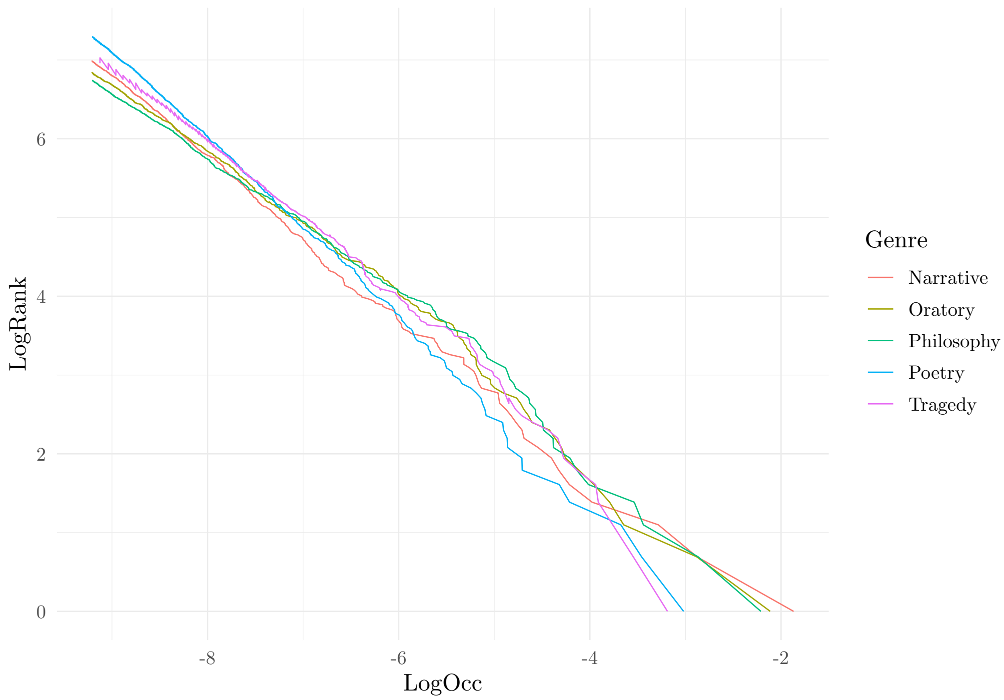


Note: The number of words is given in millions.

Typical linguistic patterns can also be observed on the data. Figure 4 illustrates this relationship in the Diorisis corpus, grouped according to literary genre according to Tauber's catalog. The left hand side of the curves, representing the tail of the distribution, are indeed quite straight and with a slope close to 1, and it appears reasonable to call these distributions Zipf laws. We know that Mandelbrot (1953) showed the optimality of the general pattern of Zipf's law in language, and Ferrer i Cancho and Solé (2003) have shown that power laws may emerge from a least-effort optimization in communications, and that its emergence may be associated with the use of symbolic concepts. We can note, however, that there are some variations in the power law coefficients in our data. Poetry, comedy and tragedy all seem to exhibit thinner tails than the other genres, with their most frequent words being less common than for other genres.

Table 3 shows the application of the distribution statistics discussed in Section 1.2 to the data displayed in Figure 4. We can see that in all five cases, the distribution is identified as a power law or truncated power law, without a strong distinction between the parameters for these two. The parameter α is close to 1: lower than 1 for prose (fatter tail), and higher for poetry (thinner tail). We can presumably relate this pattern to the fact that some bodies of literature typically resort to a broader range of vocabulary than others. Studies on contemporary languages have exhibited similar differences by genre, as well as across languages: see Grabska-Gradzińska et al.

Figure 4: *Log/Log Cumulative Distribution on Diorisis Corpus for Various Literary Genres*



Note: The data includes the lemmas for each genre that have more than 0.01% of occurrences. The horizontal axis is the logarithm of normalized frequency of each lemma, and the vertical axis is the logarithm of the lemma's rank.

(2012) for example, who focused on English and Polish. A cost-benefit analysis may allow us to form a more precise understanding of the differences we can observe in the distribution fits for these various literary genres. Poetry, in ancient languages, is mostly defined by metric constraints, as Devine and Stephens (1976) illustrated in detail in the case of Homer: the rhythm of long and short vowels must follow a particular pattern. This constraint creates the need for more varied vocabulary, given a message to communicate, beyond the aesthetic search for rare words. It is possible for poets to sometimes deviate from the norm, for a particular literary effect, or because they could not otherwise express their meaning, but such a deviation has an aesthetic cost. Tragedy is written in verses (Saïd and Trédé [1990] 1999), but in spite of being subject to the same type of constraints as poetry in general, it has a steeper distribution (smaller α). Unlike poetry in general, tragedy must be able to express vivid dialogs, and be clearly understandable. Hence, the difference in the fitted distribution parameter between the simple narrative and tragedy may capture the effect of metric constraints, while the difference between poetry and tragedy may capture the ability to use more flowery vocabulary as the coding/decoding constraints related to the message are less pregnant. Oratory texts, mostly Athenian judicial speeches, also need to strike the right balance between the need to be understood by the judges (an assembly of hundreds of citizens) and the necessity to use specialized vocabulary, in particular in financial affairs. Nevertheless, the potentially flowery language of the orators is significantly closer to simple narrative texts than to poetry. While it is difficult to construct an *a priori* model that could account for the magnitude of these differences, a formal model of optimal communication could allow us to scale the differences across genres, and potentially draw comparisons with genre-specific corpora in modern languages.

Table 3: *Summary Statistics on Distribution Fits Across Literary Genres*

Statistic	Narrative	Oratory	Philosophy	Poetry	Tragedy
Lambda Exp	0.001	0.001	0.001	0.003	0.003
Alpha Pow	0.792	0.852	0.820	1.134	1.091
Alpha Trunc	0.747	0.808	0.777	1.099	0.874
Lambda Trunc	0.000	0.000	0.000	0.000	0.000
Trunc vs Pow R	1.574	1.552	1.828	1.009	1.479
Trunc vs Pow p	0.000	0.013	0.011	0.198	0.074
Trunc vs Exp R	4.814	5.171	6.071	4.980	3.510
Trunc vs Exp p	0.000	0.000	0.000	0.000	0.000
Pow vs Exp R	4.739	5.095	5.984	4.928	3.086
Pow vs Exp p	0.000	0.000	0.000	0.000	0.002

Note: The data excludes words with less than 100 occurrences in text.

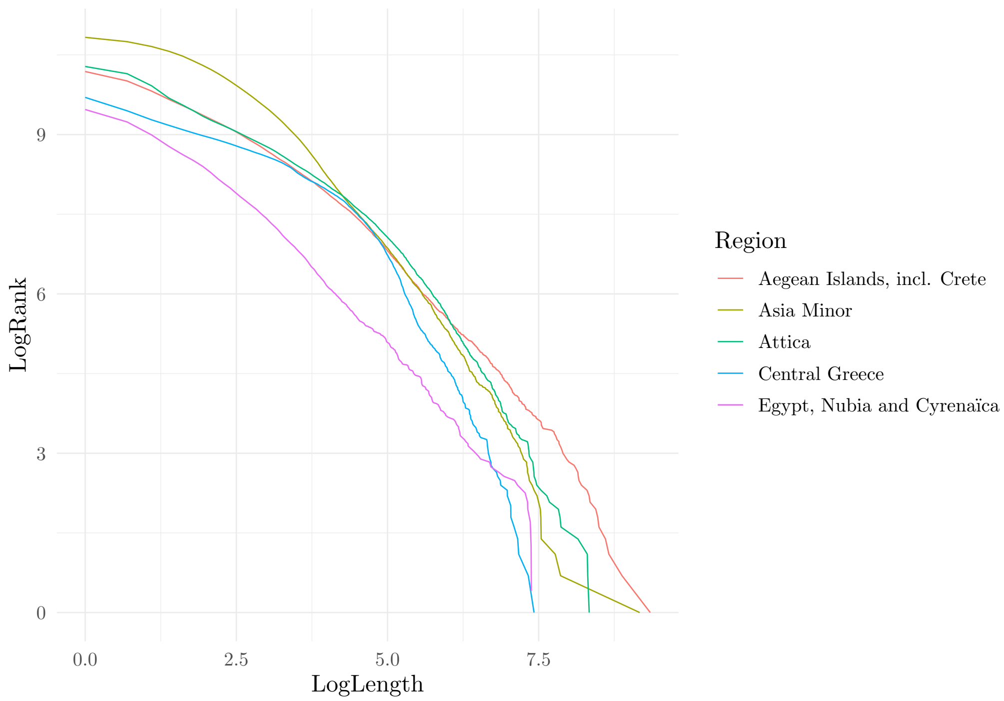
2.2 Epigraphy and Constraints

The ancient Greeks used to inscribe a wealth of information on stone and on various artifacts, and these inscriptions are found on monuments, on steles, on vases and in many other places. These writings have reached us in their original form, albeit with some lacunae. This makes for a different situation from literary texts, which, with the exception of papyri, were copied many times over as manuscripts, and very often curated. The literary tradition has kept works that were deemed of value, but many written documents, such as accounts or contracts, have been lost. Inscriptions therefore provide unique evidence from ancient Greece, since they constitute true original documents.

Epigraphists read and analyze these documents, and transcribing, editing and contextualizing the texts from archaeological material is a complex process. Many inscriptions are in a damaged or fragmentary state, and epigraphy is also concerned with forming the best hypotheses in order to fill in the missing information, based on some commonly observed recurrences (McLean 2002). Epigraphic sources have been collated into large volumes since the Renaissance, and there are few centralized editions for Greek or Roman Antiquity. Online, the largest and most comprehensive repository of Greek inscriptions is the Packard Humanities Institute (PHI)'s Searchable Greek Inscriptions website ("PHI Greek Inscriptions," n.d.). The PHI data available online is presented as one webpage for each inscription, that contains the inscription's text as well as some additional geographic information. Since there is no centralized database available containing the entirety of the PHI data, it must be processed page after page using an automated web browser. Since, for historians, inscriptions are usually analyzed one by one, or in small groups, epigraphic data has not been, to date, considered and analyzed as a bulk. Once all the inscriptions are electronically available in one place, it is possible to process them in order to map each recognizable word to a lemma, using the Classical Languages Toolkit (Johnson et al. 2019). Further, location information can be joined with the *polis*-level categorization data from the POLIS database.

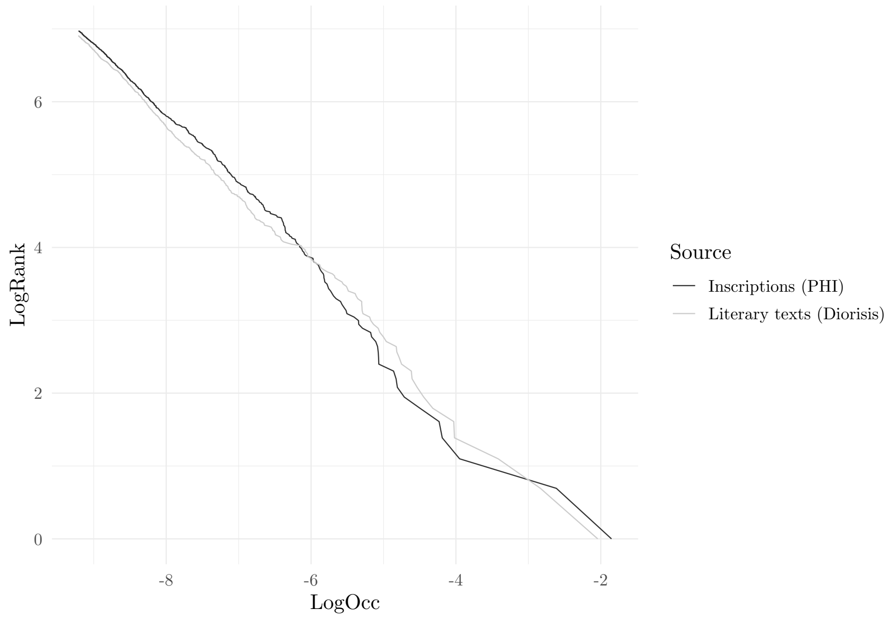
With the centralized PHI data, one can carry out cross-sectional analysis which would be otherwise impossible. Figure 5 shows the distribution of the number of words in the PHI inscriptions on a logarithmic scale, for a few regions where there are the largest number of inscriptions. We can see that the message length distribution is different across regions. The shapes of the curves seem consistent with an exponential distribution for some, and a power law for others. There are several factors that could affect the length of inscriptions. The material may certainly have been randomly broken, but in quite a few cases, fragments are paired back together, forming unique inscriptions. Mostly, it seems logical that the cost of putting messages to the stone and the nature of the messages are the drivers would affect the length of these messages. The features that differentiate these distributions the most from straight power laws appear to be the tangents for very short lengths, almost flat in most cases. This indicates that single word or two-word inscriptions are less frequent than a power law would imply, if it was to be consistent with the number of occurrences of longer messages. On the other hand, very long inscriptions are also shorter than a straight power law would imply. This cut-off, which may be captured by an exponential truncation, would seem to be necessary in order to reflect the constraint of material availability as well as the greater likelihood that longer inscriptions be broken and fragmentary. Short messages, in contemporary communication and advertising, are more effective, all else being equal (Baltas 2003). While ancient inscriptions are not the same thing as internet ads, they share the same aim of publicity. One may wonder, hence, what mechanisms lead to the distributions in Figure 5. While literary texts are essentially narrative, inscriptions sought to communicate specific information. They mostly recorded political decisions, treaties between cities, honors given, funeral rites, the freeing of slaves, or religious dedications (McLean 2002). Inscriptions therefore tend to express actions that some people take concerning other people; for example, a husband has a tombstone engraved in honor of his deceased wife, or the assembly honors a prominent citizen. As a result, the minimal information that can be communicated is of the form "A does B", which necessarily uses at least three words. On the other hand, engraving is costly, and more or less so depending on the region and on how common the "epigraphic habit" is there. There are many instances of single-word funeral inscriptions such as χαίρε ("be well"). The left-hand part of Figure 5 hence likely points to an optimization of the message in inscriptions, relative to production costs.

Figure 5: *Log/Log Cumulative Distributions of the Number of Words in Inscriptions for Selected Areas*



Interestingly, although inscriptions were generally short, the distribution of word lemma occurrences broadly matches that of a literary corpus. Figure 6 plots a comparison of the distribution of lemma occurrences in the Diorisis literary corpus and in the PHI corpus. They look very similar, which indicates that the language on inscriptions has comparable features to the standard literary Greek language.

Figure 6: *Log/Log Cumulative Distribution on Literary Texts and Inscriptions*



Note: The data includes the most common lemmas for each corpus, which account for more than 0.01% of occurrences. The horizontal axis is the logarithm of normalized frequency of each lemma, and the vertical axis is the logarithm of the lemma's rank.

Nevertheless, a more precise statistical fit, displayed in Table 4, shows that the use of words in inscriptions is less fat-tailed than in literary works: the inscriptions use rare words more often. This is presumably not related to a greater presence of poetry in these inscriptions, but rather to the more common appearance of random proper names, which would be much less likely in literature. In a nutshell, there is a greater amount of information content in inscriptions. In all the types of inscriptions we mentioned above, particular people are named, and these names are a substantial part of the information that is conveyed, making each inscription unique in that sense. The language of ancient Greek inscriptions, hence, may be seen as optimizing its informational content, given a certain cost of engraving. Econometric tests relating the vocabulary range in inscriptions with the size of the inscriptions and across regions could thus give a new perspective on the informational landscape of the ancient *polis*.

Table 4: *Summary Statistics on Distribution Fits*

Statistic	PHI	Diorisis
Lambda Exp	0.001	0.001
Alpha Pow	0.827	0.771
Alpha Trunc	0.789	0.743
Lambda Trunc	0.000	0.000
Trunc vs Pow R	1.351	1.968
Trunc vs Pow p	0.000	0.000
Trunc vs Exp R	4.522	7.291
Trunc vs Exp p	0.000	0.000
Pow vs Exp R	4.463	7.224
Pow vs Exp p	0.000	0.000

Note: The data excludes words with less than 100 occurrences in text.

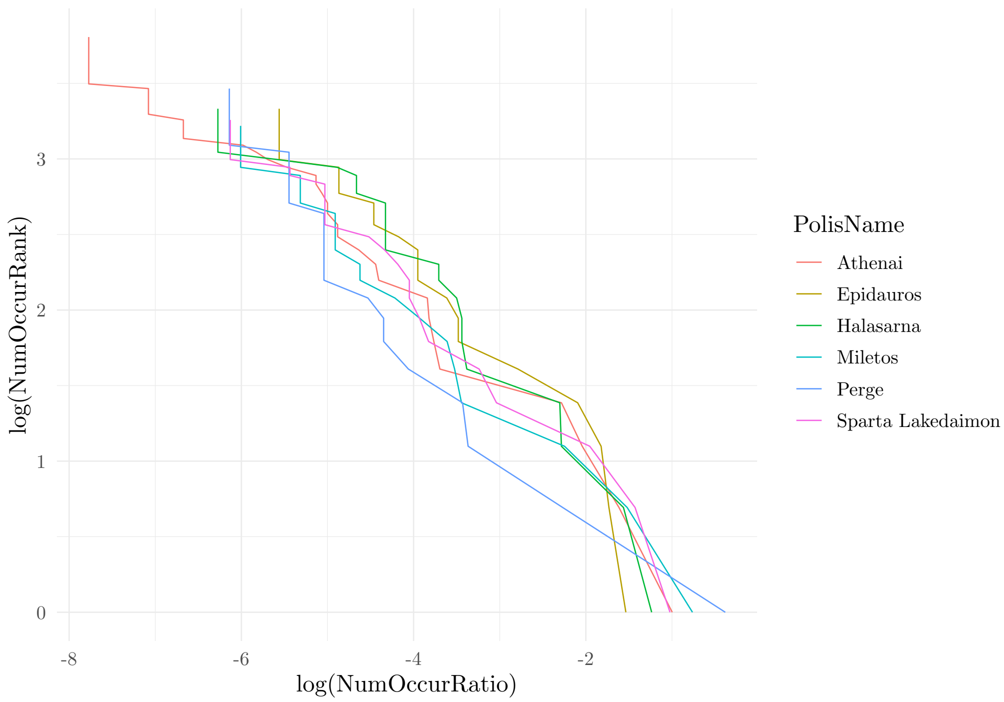
2.3 Gods and Market Shares

From the textual sources we have mentioned above, historians and classicists have created other, more complex corpora that are electronically available. We focus here on a particular example concerning religious worship. The study of ancient Greek religion and ritual is a significant part of historical research on the period. Certain inscriptions give us a glimpse of actual votive acts, and have been closely examined for this purpose. Epigraphic sources related to religious offerings or dedications typically contained the names of one or several gods along with some qualifications and reasons for the inscription. The centralization and analysis of such inscriptions have been carried out for over a century, starting with thousands of paper files in the 19th century. Bonnet and Lebreton (2019) discuss the historiography of the study of divine invocations, and explain how methods have evolved over time, converging towards the use of databases. One instance of such a database is the Base de Donnée des Epiclèses Grecques (BDEG), a project initiated in the early 2000s (Lebreton et al. 2014; Brulé and Lebreton 2007). It contains the information from thousands of epigraphic inscriptions or literary references to votive acts, including location, the god or gods names and the nature of the inscription. More specifically, researchers gathered and made available data about votive acts from all sources, periods and region they may come from. The god name to whom the dedication was addressed is recorded, typically in the first place in the invocation, with some additional qualification (commonly an adjective). The additional qualification, named cult-epithet, usually describes a particular feature of the god.

The data has been entered in over 11,000 forms, with in principle one for each observation of a divinity and epithets. The BDEG presents its data as separate webpages: it is designed to allow researchers to visualize a form at a time, and in order to analyze the data in bulk, it is necessary to programmatically collate, parse and compile all the data from all these webpages.

Location information is provided in some cases with coordinates, and in other cases only with the gallicized Greek name. Once the names are manually mapped to the corresponding anglicized Greek names, location information can be joined with the data from the POLIS database. The geographic spread of the data covers the entire Greek world, and there appears to be a definite concentration on some particular gods. The divinities in the Greek pantheon were indeed far from equally treated by humans. The distribution of votive acts across individual gods for a few specific locations is shown in Figure 7.

Figure 7: *Log/Log Cumulative Distribution of the Number of Votive Acts Among Gods, for Selected Poleis*



Note: NumOccurRatio is the ratio of the number of votive acts for a given god relative to the total for the *polis*.

The somewhat straight lines are indicative of a power law, as confirmed by the statistical tests from Table 5, which is carried out on the entire dataset, not grouped by location. The estimate gives us approximately $\alpha = \frac{1}{2}$, so that the distribution has a fatter tail than Zipf's law: in each city there are a few gods that attract large numbers of dedications, but not the same gods are the most popular in every city. How can one make sense of these patterns? Research in religion sciences based on economics has tended to focus on contemporary religion economics (Iannaccone 1998) on one hand, and, on the other hand, on the conditions leading to the existence of religion, or of certain features of ritual (Bulbulia 2010; Bulbulia and Frean 2010; Henrich 2009). In the economic modeling of the existence of religion, a cult tends to be considered as a club good, the benefit of which accrues as there are more participants. This type of model is not adapted to a polytheist world where worship is distributed. If one considers the benefits from the gods as a finite quantity, shared by worshipers, then a game equilibrium can consist of mixed strategies, hence potentially accounting for the fact that all gods get at least some worship. Given there are several “products”, the fact that votive acts are distributed according to a power law could be envisioned as a “natural” pattern in goods consumption, as Kohli and Sah (2006) empirically showed that was the case in market shares. Other, more fundamental, explanations can be envisioned by considering a form of preferential attachment: the more popular gods attract more people, so that after a large number of cycles, one may observe a power law.

Table 5: *Summary Statistics on Distribution Fits for Votive Acts*

Statistic	BDEG
Lambda Exp	0.013
Alpha Pow	0.482
Alpha Trunc	0.510
Lambda Trunc	0.000
Trunc vs Pow R	1.710
Trunc vs Pow p	0.060
Trunc vs Exp R	7.175
Trunc vs Exp p	0.000
Pow vs Exp R	7.106
Pow vs Exp p	0.000

In spite of the fact that the primary sources we have examined above were essentially unstructured categorical data, we have seen that a detailed analysis of their distributions and their drivers, using simple economic-driven paradigms, can shed light on many aspects of ancient Greek society. The perspective we acquire thanks to the application of complexity methods to primary sources treated in bulk, in a cliometric context, gives us new insights into the data, even though these texts and inscriptions have been closely studied by humanists and classicists since the Renaissance.

3 Networks, Complexity, and Naming Gods and Men

The use of network theory by historians has expanded over the years (Lemercier 2012; Karila-Cohen et al. 2018), and has given rise to renewed epistemological debates around which kinds of historical data could be conceived of as networks. Network analysis in historiography is usually carried out in a descriptive way, to account for a set of relationships extracted from a given document corpus. The application of network analyzes to literary texts, and to classical texts in particular, has also been researched, most often by physicists or computer scientists. For example, Kydros, Notopoulos, and Exarchos (2015) looked at the networks between characters in mythology, and in their book, Kenna, MacCarron, and MacCarron (2017) discuss the analysis of networks stemming from various mythological and historical texts. With a less quantitative but more illustrative logic, Rydberg-Cox (2011) was interested in networks in the specific context of the Greek tragedy as well as their visualization. Finally, the works of Waumans, Nicodème, and Bersini (2015), Rochat (2014) and Elson, Dames, and McKeown (2010) provide a general perspective on the use of networks in literature. In all these approaches, however, the precise nature of the links that put individuals (characters in a play, or historical characters in a narrative) in a relationship is not well determined: what exactly does it mean, in terms of a relationship, that two names appear in the same sentence? Networks are a frequent tool in complexity economics, and in particular in cliometrics: for example in order to account for international market integration through time (Bastidon et al. 2019; Bastidon and Parent 2022).

In this section, we concern ourselves with some ancient Greek data which naturally enough represents relations, and have therefore been tackled by historians as networks, in a few instances at least. These particular data are mostly compiled from inscriptions: they are the Lexikon of Greek Personal Names, and the Mapping Ancient Polytheisms database. Both rely on arduous work carried out by specialists, having sifted through hundreds of thousands of inscriptions in order to produce the resulting electronic data.

3.1 The Economy of Names: Onomastic Networks

In ancient Greece, something as fundamental as naming a baby worked in a very different fashion from what we are familiar with today: one would only get a single anthroponym, a unique name. To this unique name could be associated a patronym, the name of one's father. In some cases, a person could acquire a nickname through their actions or their physical aspect, but the single name remained a core principle. The choice of a name by the parents carried meaning, and could reflect through etymological links the child's belonging to the broad family. Giving to a child the name of a grand-parent, papponymy, was also quite common. Names were often formed through derivational morphology: using one or two nouns or adjectives to make a name. In the study of ancient history, the analysis of names, onomastics, often goes hand in hand with prosopography, that is the gathering of all available historical information on particular individuals. For persons who have been considered as important in historiography, one can usually get sizable volumes of information. For less well-known people, it is a complex historical inquiry (Karila-Cohen 2016).

The available epigraphic sources provide anthroponyms, sometimes with a patronym. In order to go from lists of names to the notion of individuals, it is necessary to relate these inscriptions to each other. Prosopographers rely on instances of identical, or closely related names in a given geographic area, and are dated around the same time, so that they can transform these names into references to particular individuals. A systematic perspective on how to carry out this research was proposed in Bresson (1981), who suggested the use of certain family relationships in a network context, combined with the naming conventions that appeared to be most common, in order to build a family tree. This logic initially applied to Rhodes has been generalized (Karila-Cohen 2018). In his his seminal study, he made it a central hypothesis that the practices observed in Rhodes in the modern period and up to contemporary times obeyed similar rules to those practiced by the Ancients. He showed among other things the way in which papponymy was applied, with the names from the father and mother's sides alternating. Using this historical information, one can therefore create onomastic networks: someone with such name had a child with such name. Then, with prosopographic work, one can create prosopographic networks: such person had such person as a child. There is more onomastic data available than prosopographic, and the prosopographic reconstruction relies on many assumptions.

Large volumes of onomastic and prosopographic data have been made electronically available through the Lexicon of Greek Personal Names (LGPN) housed by Oxford University (Parker, Yon, and Depauw 1996). The project started in the early 1970s, and has led to the publication of references to hundreds of thousands of names. The electronic interface to the LGPN is not designed for its data to be processed and analyzed in bulk. It is designed as a tool to query a name or a name root, and observe its occurrences. It is impossible to simply download the entire data in a structured form in one batch; it has to be reconstructed from the data pertaining to each possible name. We created a centralized dataset containing all these elements, in a structured fashion in the sense that all the data is stacked together in unique data tables, and so that it may be joined with additional data sources, such as the POLIS database through geographic information. We found close to 40,000 unique names spread across about 350,000 individual entries. The relationship table includes approximately 250,000 links.

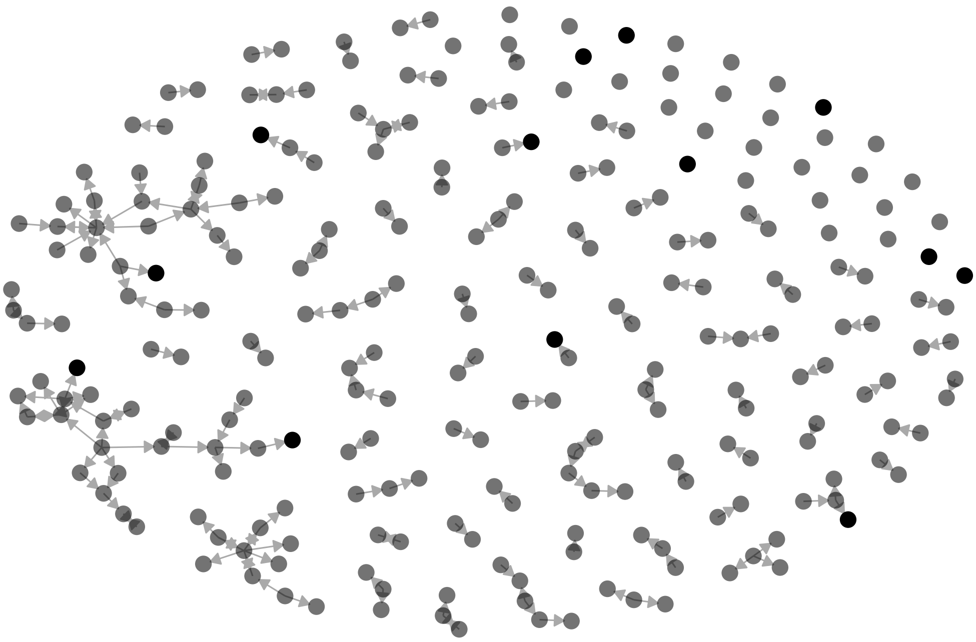
Before looking into the network aspect of naming relationships, we examine the distribution of men and women names. Women were essentially excluded from political life, and inscriptions very often reported official acts, so that the the number of observations is biased towards men. Indeed, the number of feminine names in the dataset is a tenth of that for men. Table 6 shows distribution fits for the names of men and women, across the entire dataset. The parameters for women indicate a flatter distribution: women names are less concentrated. The distribution fits in Table 6 also tell us that the distribution is more likely to be a truncated power law than an exponential, with a very small exponential parameter λ. Baek, Kiet, and Kim (2007) have shown that family names distributions across the world tended to follow power laws, but not always, and that the parameters were not constant from one country to the other. Focusing on the contrary on given names, specifically in the US and over time, Li (2012) showed that a simple power law could not fully account for the distribution names. In general, Baek, Kiet, and Kim (2007) estimated that the power law coefficients in contemporary data on last names were generally around 2. We can see that the ancient Greek names exhibit a different behavior, with coefficients closer to 1. The lower α is indicative of a fatter distribution tail: names were more concentrated in ancient Greece than they are in the contemporary world. In modern times, names are family names, and are hence quite idiosyncratic, while given names are not used as a unique identification mechanism. The names in ancient Greece were given, or chosen, so that they were presumably selected from an existing stock, and affected by fashion and taste, which should translate into a higher degree of concentration.

Table 6: *Summary Statistics For Name Distributions Fits by Gender*

Statistic	Man	Woman
Lambda Exp	0.104	0.284
Alpha Pow	0.654	0.783
Alpha Trunc	0.747	0.946
Lambda Trunc	0.001	0.003
Trunc vs Pow R	18.134	12.879
Trunc vs Pow p	0.000	0.000
Trunc vs Exp R	33.355	22.523
Trunc vs Exp p	0.000	0.000
Pow vs Exp R	32.623	21.094
Pow vs Exp p	0.000	0.000

Using the parental relationships between names in a given location, we can build a network of these links. We implemented a comparable method as that laid out in Karila-Cohen (2018), where she focused on particular demes (suburbs) in Athens, but applied it to the entire LGPN data. We constitute a network where the edges are unique names, and the vertices are parental relationship entries in the LGPN: the resulting network is hence onomastic from the perspective of the edges, and prosopographic from the perspective of the vertices. Restricting the LGPN entries to the same demes as in Karila-Cohen (2018), we find approximately the same nodes and edges. As an illustration, Figure 8 displays the network thus obtained, reduced to the demes Oion Kerameikon and Oion Dekeleikon in Athens. We can see that a majority of nodes are linked, and even for this small network there is a wide range of components size. This network effectively displays the ancient Greek's child naming strategy in Oion, as a function of the father's name. The darker nodes, representing the most common names, are either isolated or are the leaves in this directed network. There are many observations of names for which there is no record of a child or a parent, so that, in the case of Oion at least, it may have been the case that the people with the most common names were, all else being equal, less likely to see their children appear in inscriptions.

Figure 8: *Naming Network for the Oion Demes*



Note: The arrows represent the naming relationship (, and the darker the nodes, the more frequent the name is.

Building all local networks at the *polis* level in the same manner, we can create a large network covering the entire dataset, for both men and women, and for the whole Greek world across time periods. Focusing on this large network, Table 7 displays the degree distribution fit, that is, the distribution of the number of different names that have been chosen by a person with a given name, in a given location. The number of distinct names chosen by each name appears to be a fat-tailed power law, with a coefficient around $\frac{1}{3}$. Let us assume there is a large network composed of individuals, nodes, each one with a name attribute, such that a new person appearing has no material effect on population counts. A naming relationship is a link created between a parent and the child, the new person. Such a prosopographic network can be collapsed into an onomastic network, where all the individuals are gathered in a single name node and the links grouped by the nodes from which and to which they go, as with the networks analyzed in Figure 8 and Table 7. If the individuals are distributed among attributes following a given distribution, and if the naming links are distributed uniformly, then the both the in- and out-degree distribution of the nodes in the name network should follow that given distribution. We saw that names were distributed according to a power law, potentially truncated, according to Table 6, and the in-degree distribution for the name network is also a power law, according to Table 7. However, their parameters are quite different, the degrees being more concentrated than a random name selection would suggest. There is therefore a certain strategy at play in the selection of names, according to our data. This pattern could be accounted for by preferential attachment: if new links are more likely to attach to the nodes as a function of their existing number of links, then we would expect a greater concentration in the resulting degree distribution than in the underlying name distribution. The ancient Greeks' naming strategy, concerning people whose father name was mentioned along theirs in inscriptions, hence, would reflect a certain conformism.

Table 7: *Summary Statistics on Distribution Fits for the Number of In-Degrees in the Naming Network*

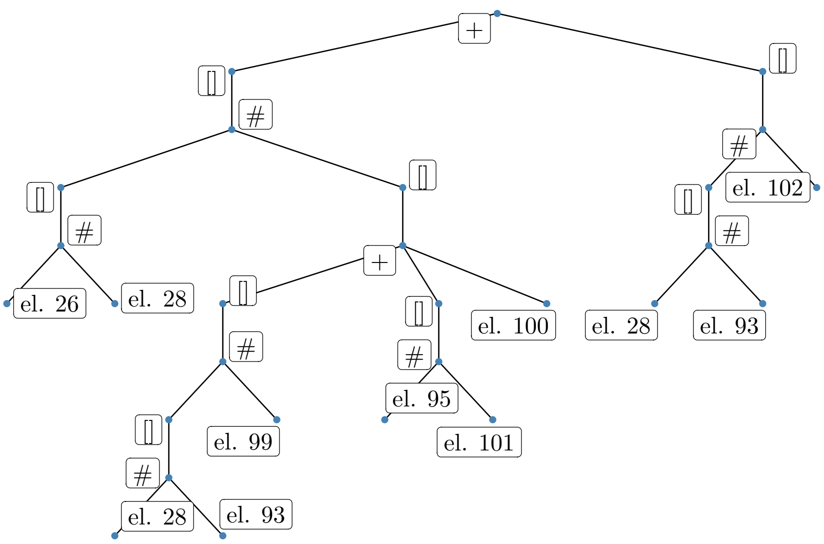
Statistic	LGPN
Lambda Exp	0.002
Alpha Pow	0.383
Alpha Trunc	0.388
Lambda Trunc	0.000
Trunc vs Pow R	2.427
Trunc vs Pow p	0.197
Trunc vs Exp R	7.769
Trunc vs Exp p	0.000
Pow vs Exp R	7.706
Pow vs Exp p	0.000

3.2 How to Speak to the Gods? The Economy of Divine Formulae

The BDEG data we discussed earlier represented information in an unstructured fashion on the way ancient invocations addressed the gods, in addition to the god's name: the forms in that database contained one or two epithets at most in each case. On a large scale, this data is useful in order to track the number of votive acts, but it is difficult to exploit if one wants to precisely understand the way in which the gods were invoked. The more recent large-scale MAP project (Bonnet 2017) offers a more precise and entirely revamped representation of the underlying data, mostly derived from inscriptions, covering the Greek and Semitic ancient worlds, and allows for the close study of the formulas the Ancients used to address divine beings. The MAP is used by researchers for the philological study of dedications, and allows them to easily find series of invocations that mention a particular god or qualifier. To understand the relational logic which structures these divine powers, MAP takes into account a wide range of divine onomastic sequences, combinations of divine names or elements (names, epithets, titles, propositions), some shared by several gods, others specific to a particular god. Indeed, the simple act of performing a rite in ancient Greece involved addressing one or more gods with a complex series of qualifiers, the epicleses, ordered and expressed in a deliberate manner. These qualifiers, which were sometimes common between certain gods, effectively created a network between all the gods; and the gods also created a system of relations between the qualifiers. These onomastic sequences taken as a whole form a representation of the way in which the Greeks conceived of polytheism. The detailed study of onomastic sequences has been going on for many years, as we alluded to earlier, and recent research in this area includes Brulé (1998), Brulé (2005), Bonnet and Belayche (2017), who establish various assumptions implicit in how onomastic sequences are constructed. The idea of putting these sequences within the framework of network analysis is in fact at the basis of the MAP project, and another cornerstone is the extraction of onomastic formulas from the sources which, thanks to a particular syntax, can account for the great complexity of these inscriptions (Bonnet and Lebreton 2019).

The data that constitutes MAP, derived from the inscriptions by a team of specialists, is highly structured and accessible in bulk. Each material item, such as a stele or part of a monument for example, is an entry in a source data table, with detailed information such as its publication or location. A source contains one or several testimonies, addresses to the gods, each entered in a testimony table, so that each testimony, usually in the form of a sentence referring to divinities and qualifiers. These sentences are converted to formulaic expressions akin to mathematical formulas expressing the links between the various elements in the testimony. These elements, mostly adjectives, nouns or god names, are also centralized in a specific element table. In order to understand how to construct a network that captures the semantic relationships between gods, it is necessary to examine the logic of onomastic formulas. One example of a testimony would be number 100, from Egypt, in the MAP database, from an inscription stating: “Beside the Lords Gods Priô the Greatest God and Horegebthis and Isis Rhesakemis and the Greatest Gods Who are with them” (MAP translation). This dedication is encoded as a formula referencing elements like the gods (Prion or Isis) or qualifications (Great), as well as the connections between them (such as “and”, for example). In this case, this is recorded as the following formula: `[[{26}#{28}]#[[{99}#[{28}#{93}]]+{100}+[{95}#{101}]]]+[{102}#[{28}#{93}]]`, where the numbers are the identifications of particular elements, and the operators are the MAP researchers' interpretation of the logic of the text. All these formulae can be systematically converted into tree network representations, as shown in Figure 9 for the case at hand. Any operator-based expression could in fact be converted to a tree representation.

Figure 9: *Syntactic Network Representation of Testimony #100*

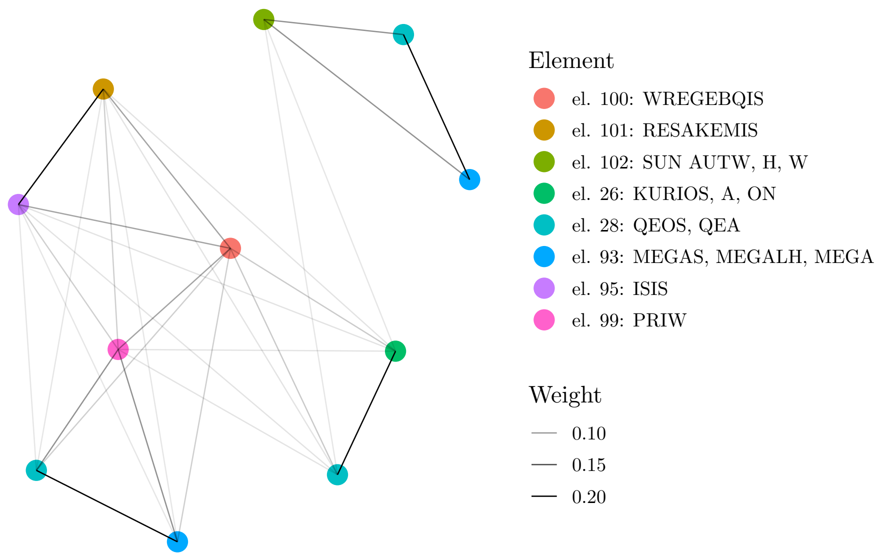


Note: Every node represents either an element or an operator and the links represent the application of an operator to one or several elements.

In order to better seize the nature of the links created between the elements and the gods, one approach is to allocate a certain distance to all the links in Figure 9 depending on the operators, so that the edge connecting two entities related by "and", for example, would have a lesser distance than a simple apposition. Then, the total distance from any entity to any other entity in the formula can be computed, along the syntactic network. Finally, the strength of the connection between two entities, which can be used as a weight in the graph connecting all the entities appearing in a formula, can be chosen as the inverse of the distance. Figure 10 shows a representation of the resulting graph for Testimony 100, where only the links above a certain weight are shown. This way of approaching onomastic formulas through the syntactic networks they can generate is inspired from quantitative linguistics (Mehler et al. 2016).

One may build the large disconnected network generated by all the known onomastic formulas.

Figure 10: *Distance Network Representation of Testimony #100*



Note: All the links between the nodes with a weight above 0.05 are represented, and they are weighed as a function of the formulaic distance: closer relationships are marked by denser lines. The nodes representing entities are in grey, and qualifications in blue. The elements are encoded in latinized Greek capitals.

Looking at all the gods in all testimonies, we can focus on the weighted degree distribution, as shown in Figure 11, which represents the strength with which each occurrence of each god in invocations is related to other gods. This degree distribution does not appear as a straight line and it is therefore presumably not a simple power law, and not generated by a basic preferential attachment mechanism.

Table 8 shows that the degree distributions appears as an exponential, with a parameter λ equal to 2.5. We know that the dedications recorded in votive acts, organized as a function of the main divinity to which they were addressed, follow power laws, potentialy truncated. It is hence somewhat surprising that the connections made between gods and other gods or qualifiers clearly follow exponential distributions. Indeed, if connections were random between power-law distributed acts towards gods, then the number of degrees would also have the same distribution, following an argument comparable to what we suggested about onomastic networks. Exponential degree distributions can emerge when nodes appear and get attached randomly among the existing nodes. In that case, Table 8 may indicate that invocations are expressed by merging in some new elements in the existing mix, without necessarily following a pre-existing pattern. In other words, invocations are in essence innovative, more than a simple reshuffling of known qualifiers and gods associations. In a broader context, this could be taken as a symptomatic feature of ancient polytheist religion, lacking the kind of dogma known in monotheist religions, and allowing worshipers to make official through an inscription any message they wanted.

Figure 11: *Degree Distribution of Divinities in Distance-Weighted Formulae Networks*

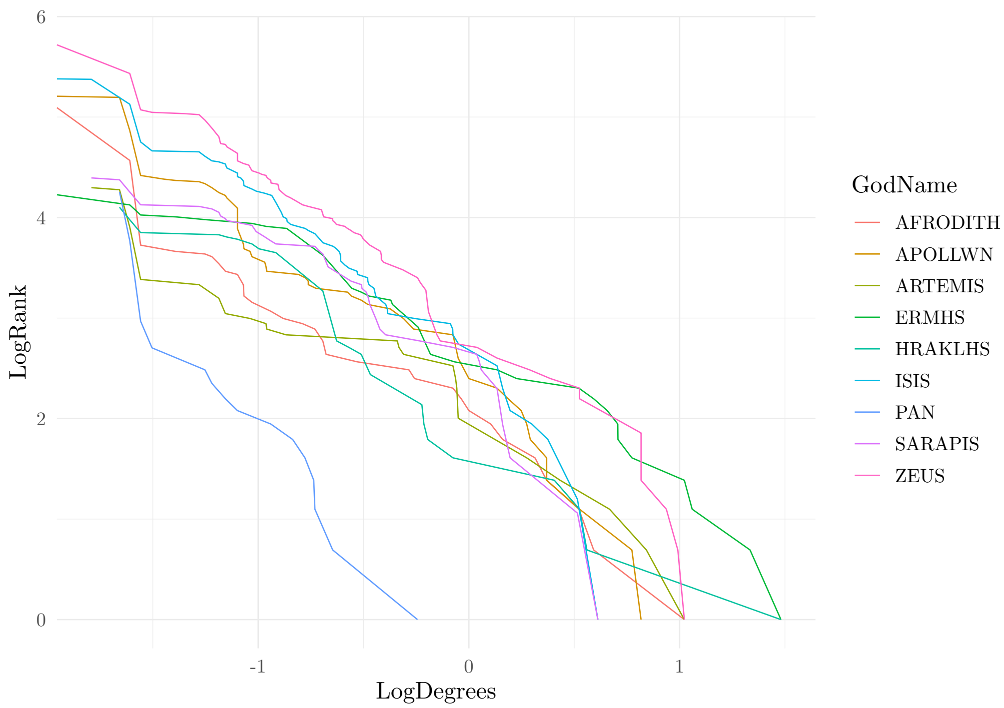
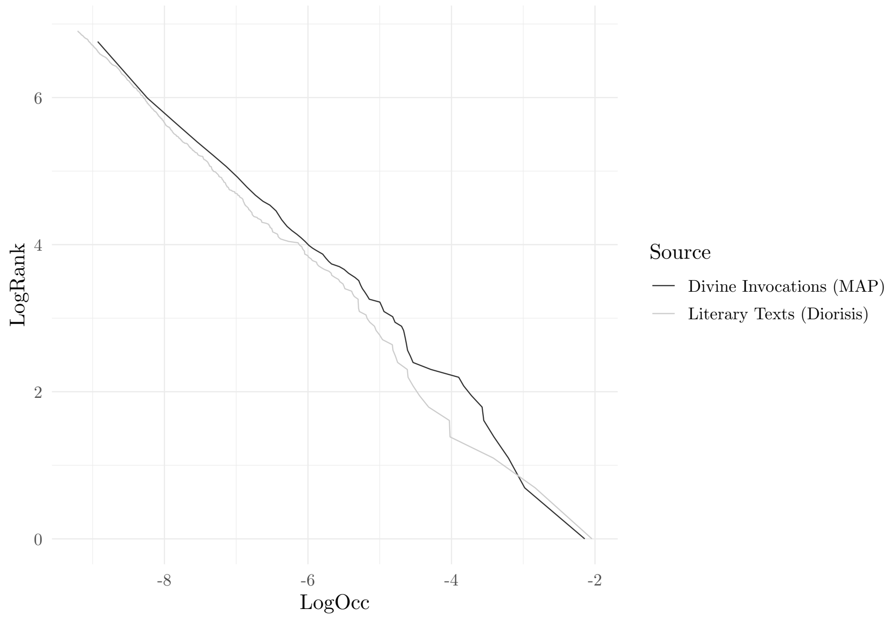


Table 8: *Summary Statistics on Degree Distribution Fit*

Statistic	MAP
Lambda Exp	2.413
Alpha Pow	0.173
Alpha Trunc	0.000
Lambda Trunc	0.323
Trunc vs Pow R	337.738
Trunc vs Pow p	0.000
Trunc vs Exp R	-70.928
Trunc vs Exp p	0.000
Pow vs Exp R	-104.786
Pow vs Exp p	0.000

One important question, potentially related to the network structure between gods and qualifiers, is whether the "language" used in these invocations possesses characteristics of a natural language. In the applications of economics and compleity to linguistics, many detailed analyzes can indeed be carried out on grammar, in particular when reflecting the network structure of syntactic relationships (Čech, Mačutek, and Liu 2016). We will restrict ourselves here to a simple comparison, looking at the distribution of terms, for the sake of simplicity. Elements, which correspond to common names or proper names, are the natural equivalent of lemmas in POS tagging. Figure 12 plots the occurrence distribution for literary texts from the Diorisis corpus, and for the onomastic formulae in MAP, and they appear quite close to each other. In fact, distribution fits tell us they bot follow power laws with the same parameter, to the second decimal. In spite of a very formulaic nature, the term distribution of divine invocations resembles that of natural language. While this may not hold if one were to look more closely at the syntactic structure of both languages, it is consistent with a notion of random terms being merged into divine formulae, as we discussed above, since these terms would be pulled from the natural language. Further, as Ferrer i Cancho, Riordan, and Bollobás (2005) have shown, when a communication language exhibits Zipf's law, one can theoretically expect the emergence of syntactic patterns. Hence, the inclusion of qualifiers in divine formulae, as random as it may be, would not be inconsistent with the appearance of the logical syntax-like structure which we have observed.

Figure 12: *Log/Log Cumulative Distribution of Diorisis and MAP Corpora*



Note: The data includes the most common lemmas for each corpus, which account for more than 0.01% of occurrences. The horizontal axis is the logarithm of normalized frequency of each lemma, and the vertical axis is the logarithm of the lemma's rank.

The networks we have analyzed here represent very different things: how the ancient Greeks named their children on one hand, and how the they addressed the gods in the dedications they inscribed on stone. In spite of these fundamental differences, the methodology for analyzing them remains the same, and involves a series of particular metrics, although we only considered degree distributions. These networks turned out to have fairly different characteristics, for which we offered some explanations, but fully developed theoretical models would be necessary to truly account for the formation of these networks.

4 Conclusion

Cliometrics have not been much concerned with ancient history, simply because there is little appropriate economic data form the period that one may analyze; in fact, whether there was an economy is debatable. Ancient history is indeed particular in the sense that there is little clean data of the form usually exploited in cliometrics. There are no clean time series of anything for the ancient times, maybe barring geological or astronomical information. Instead, there are fairly large volumes of unstructured data. This data is unstructured not only because it gathers texts and very diverse categorical information, but also because historians, who created all these information sources, tend to exploit it in this form.

On the other hand, one can apply economic principles to the analysis of potentially every aspect of human life in the ancient times, and that is what we have done here for ancient Greece. Cliometrics and complexity have extended the range of traditional cliometrics through a different approach to data, and we have followed in their path. By relying on complex systems analysis, we can develeop different hermeneutics of data, and do not need them to be structured. This has allowed us to operate on the same primary sources that historians use, albeit with a different perspective.

The behavioral economics explanations, based on game theory or optimization arguments, that we have provided, effectively open new ways of inquiry into ancient Greece. Considering how genre affected texts, how costs and information were reflected in inscriptions, how worship could self-organize, how names could be picked in a fundamentally different manner than what we know, and how the language spoken to the gods was built according to specific principles, we have not looked at economics through history, but we have instead looked at history through economics. This change of perspective, with significant help from complexity sciences, should largely open the distant past to cliometrics.

5 References

Abry, Patrice, Cécile Bastidon, Pierre Borgnat, Pablo Jensen, and Antoine Parent. 2022. “The Contours of a Cliophysics: How Can Econophysics Enrich Cliometrics? Case Studies in Debt Issues and Global Capital Markets.” *Frontiers in Physics* 10: 894400.

Alfarano, Simone, and Thomas Lux. 2010. “Extreme Value Theory as a Theoretical Background for Power Law Behavior.” 1648. *Kiel Working Papers*. Kiel Working Papers. Kiel Institute for the World Economy. https://ideas.repec.org/p/kie/kieliw/1648.html.

Alstott, Jeff, Ed Bullmore, and Dietmar Plenz. 2014. “Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions.” *PLOS ONE* 9 (1): e85777.

Baek, Seung Ki, Sebastian Bernhardsson, and Petter Minnhagen. 2011. “Zipf’s Law Unzipped.” *New Journal of Physics* 13 (4): 043004.

Baek, Seung Ki, Hoang Anh Tuan Kiet, and Beom Jun Kim. 2007. “Family Name Distributions: Master Equation Approach.” *Physical Review E* 76 (4): 046113.

Baltas, George. 2003. “Determinants of Internet Advertising Effectiveness: An Empirical Study.” *International Journal of Market Research* 45 (4): 1–9.

Barabási, Albert-László, and Réka Albert. 1999. “Emergence of Scaling in Random Networks.” *Science* 286 (5439): 509–512.

Bastidon, Cécile, Michael Bordo, Antoine Parent, and Marc Weidenmier. 2019. “Towards an Unstable Hook: The Evolution of Stock Market Integration Since 1913.” Working Paper w26166. Cambridge: National Bureau of Economic Research. https://doi.org/10.3386/w26166.

Bastidon, Cécile, and Antoine Parent. 2022. “Cliometrics of World Stock Markets Evolving Networks.” *Annals of Operations Research* To appear.

Bastidon, Cécile, Antoine Parent, Pablo Jensen, Patrice Abry, and Pierre Borgnat. 2020. “Graph-Based Era Segmentation of International Financial Integration.” *Physica A: Statistical Mechanics and Its Applications* 539: 122877.

Boccara, Nino. 2010. *Modeling Complex Systems*. Graduate Texts in Physics. New York: Springer.

Bonnet, Corinne. 2017. “Mapping Ancient Polytheisms.” 2017. https://map-polytheisms.huma-num.fr/.

Bonnet, Corinne, and Nicole Belayche, eds. 2017. *Puissances divines à l’épreuve du comparatisme : constructions, variations et réseaux relationnels*. Bibliothèque de l’École des Hautes Études, sciences religieuses, volume 175. Turnhout: Brepols.

Bonnet, Corinne, and Sylvain Lebreton. 2019. “Mettre les polythéismes en formules ? À propos de la base de données Mapping Ancient Polytheisms.” *Kernos* 32: 267–296.

Borwein, David, Jonathan M. Borwein, and Pierre Maréchal. 2000. "Surprise Maximization." *The American Mathematical Monthly* 107 (6): 517–527.

Bresson, Alain. 1981. "Règles de nomination dans la Rhodes antique." *Dialogues d'histoire ancienne* 7 (1): 345–362.

Brulé, Pierre. 1998. "Le langage des épiclèses dans le polythéisme hellénique (l'exemple de quelques divinités féminines). Quelques pistes de recherches." *Kernos* 11 (11): 13–34.

———. 2005. "Le polythéisme en transformation : les listes de dieux dans les serments internationaux en Grèce antique (V^{e}-II^{e} siècles)." In *Nommer les dieux : Théonymes, épithètes, épiclèses dans l'Antiquité*, edited by Nicole Belayche, Pierre Brulé, Gérard Freyburger, Yves Lehman, Laurent Pernot, and Antoine Prost, 143–173. Recherches sur les rhétoriques religieuses. Turnhout: Brepols.

Brulé, Pierre, and Sylvain Lebreton. 2007. "La Banque de données sur les épiclèses divines (BDDE) du Crescam : sa philosophie." *Kernos* 20: 217–228.

Bulbulia, Joseph. 2010. "Charismatic Signalling." *Journal for the Study of Religion, Nature and Culture* 3 (4): 518–551.

Bulbulia, Joseph, and Marcus Frean. 2010. "The Evolution of Charismatic Cultures." *Method & Theory in the Study of Religion* 22 (4): 254–271.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.

Crane, Gregory R. 2012. "Perseus Digital Library." April 2012. http://www.perseus.tufts.edu.

Crettez, Bertrand, and Régis Deloche. 2018. "An Analytic Narrative of Caesar's Death: Suicide or Not? That Is the Question." *Rationality and Society* 30 (3): 332–349.

Čech, Radek, Ján Mačutek, and Haitao Liu. 2016. "Syntactic Complex Networks and Their Applications." In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, edited by Alexander Mehler, Andy Lücking, Sven Banisch, Philippe Blanchard, and Barbara Job, 167–186. Understanding Complex Systems. Berlin: Springer.

Deng, Weibing, Wei Li, Xu Cai, and Qiuping A. Wang. 2011. "The Exponential Degree Distribution in Complex Networks: Non-Equilibrium Network Theory, Numerical Simulation and Empirical Data." *Physica A: Statistical Mechanics and Its Applications* 390 (8): 1481–1485.

Devine, Andrew M., and Laurence Stephens. 1976. "The Homeric Hexameter and a Basic Principle of Metrical Theory." *Classical Philology* 71 (2): 141–163.

Diebolt, Claude, and Michael Haupert, eds. 2019. *Handbook of Cliometrics.* 2nd ed. Springer Reference. Cham: Springer International Publishing.

———. 2020. “How Cliometrics Has Infiltrated Economics – and Helped to Improve the Discipline.” *Annals of the Fondazione Luigi Einaudi* 54 (1): 219–230.

Elson, David, Nicholas Dames, and Kathleen McKeown. 2010. “Extracting Social Networks from Literary Fiction.” In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 138–147. Uppsala: Association for Computational Linguistics.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica. 2015. “Suspense and Surprise.” *Journal of Political Economy* 123 (1): 215–260.

Ferrer i Cancho, Ramon, Oliver Riordan, and Béla Bollobás. 2005. “The Consequences of Zipf’s Law for Syntax and Symbolic Reference.” *Proceedings of the Royal Society B: Biological Sciences* 272 (1562): 561–565.

Ferrer i Cancho, Ramon, and Ricard V. Solé. 2003. “Least Effort and the Origins of Scaling in Human Language.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (3): 788–791.

Fleck, Robert K., and F. Andrew Hanssen. 2018. “What Can Data Drawn from the Hansen-Nielsen Inventory Tell Us About Political Transitions in Ancient Greece?” In *Ancient Greek History and Contemporary Social Science*, edited by Mirko Canevaro, Andrew Erskine, Benjamin D. Gray, and Josiah Ober, 213–238. Edinburgh: Edinburgh University Press.

Gabaix, Xavier. 1999. “Zipf’s Law for Cities: An Explanation.” *The Quarterly Journal of Economics* 114 (3): 739–767.

———. 2016. “Power Laws in Economics: An Introduction.” *Journal of Economic Perspectives* 30 (1): 185–206.

Gauthier, Laurent. 2022. “Putting Clio Back in Cliometrics.” *History and Theory* 61 (2): 289–311.

Gonzales-Feliu, Jesus, and Antoine Parent. 2016. “Clio-Combinatorics: A Novel Framework to Analyze Military Logistics Choices Using Operations Research Techniques.” In *Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations:* edited by Alberto Ochoa-Zezzatti, Jöns Sánchez, Miguel Gastón Cedillo-Campos, and Margain de Lourdes, 79–101. Advances in Logistics, Operations, and Management Science. Hershey: IGI Global.

Grabska-Gradzińska, Iwona, Andrzej Kulig, Jarosław Kwapień, and Stanisław Drożdż. 2012. “Complex Network Analysis of Literary and Scientific Texts.” *International Journal of Modern Physics C* 23 (07): 1250051/1–15.

Hansen, Mogens Herman, and Thomas Heine Nielsen. 2004. *An Inventory of Archaic and Classical Poleis*. Oxford: Oxford University Press.

Henrich, Joseph. 2009. “The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution.” *Evolution and Human Behavior* 30 (4): 244–260.

Hobson, Matthew S. 2014. “A Historiography of the Study of the Roman Economy: Economic Growth, Development, and Neoliberalism.” *Theoretical Roman Archaeology Journal* 0 (2013): 11–26.

Iannaccone, Laurence R. 1998. “Introduction to the Economics of Religion.” *Journal of Economic Literature* 36 (3): 1465–1495.

Ide, Nancy. 2004. “Preparation and Analysis of Linguistic Corpora.” In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 289–305. Malden: Blackwell Publishing.

Jackson, Matthew O., and Brian W. Rogers. 2007. “Meeting Strangers and Friends of Friends: How Random Are Social Networks?” *American Economic Review* 97 (3): 890–915.

Johnson, Kyle P., Patrick J. Burns, John Stewart, and Todd Cook. 2019. “The Classical Language Toolkit.” January 2019. http://cltk.org.

Johnson, Tim, and Josiah Ober. 2014. “POLIS.” 2014. http://polis.stanford.edu.

Karila-Cohen, Karine. 2016. “Prosopographia Attica 2.0 : base de données et raisonnement prosopographique.” *Revue historique* 680 (4): 869–904.

———. 2018. “Le graphe, la trace et les fragments : l’apport des méthodes quantitatives et des outils numériques à l’étude des élites civiques athéniennes.” *Annales HSS* 73 (4): 785–815.

Karila-Cohen, Karine, Claire Lemercier, Isabelle Rosé, and Claire Zalc. 2018. “Nouvelles cuisines de l’histoire quantitative.” *Annales HSS* 73 (4): 773–783.

Kenna, Ralph, Máirín MacCarron, and Pádraig MacCarron, eds. 2017. *Maths Meets Myths: Quantitative Approaches to Ancient Narratives.* Understanding Complex Systems. Cham: Springer International Publishing.

Kohli, Rajeev, and Raaj Sah. 2006. “Some Empirical Regularities in Market Shares.” *Management Science* 52 (11): 1792–1798.

Kydros, Dimitrios, Panagiotis Notopoulos, and Georgios Exarchos. 2015. “Homer’s Iliad – A Social Network Analytic Approach.” *International Journal of Humanities and Arts Computing* 9 (1): 115–132.

Lebreton, Sylvain, Jean-Baptiste Barreau, Karine Karila-Cohen, and Pierre Brulé. 2014. “Banque de Données des Épiclèses Grecques (BDEG).” 2014. https://epiclesesgrecques.univ-rennes1.fr/?lang=en.

Lemercier, Claire. 2012. "Formale Methoden der Netzwerkanalyse in den Geschichtswissenschaften: Warum und Wie?" *Österreichische Zeitschrift für Geschichtswissenschaften* 23 (1): 16–41.

Le Riche, Antoine, Antoine Parent, and Lei Zhang. 2022. "Introduction." In *Institutional Change and China Capitalism: Frontier of Cliometrics and Its Application to China*, edited by Antoine Le Riche, Antoine Parent, and Lei Zhang, xxvii–xlv. Transformations in Banking, Finance and Regulation 2. Hackensack: World Scientific.

Li, Wentian. 2012. "Analyses of Baby Name Popularity Distribution in U.S. For the Last 131 Years." *Complexity* 18 (1): 44–50.

Limpert, Eckhard, Werner A. Stahel, and Markus Abbt. 2001. "Log-Normal Distributions Across the Sciences: Keys and Clues." *BioScience* 51 (5): 341.

Mandelbrot, Benoît. 1953. "An Informational Theory of the Statistical Structure of Languages." In *Communication Theory*, edited by Willis Jackson, 486–502. Woburn: Butterworth.

McLean, Bradley H. 2002. *An Introduction to Greek Epigraphy of the Hellenistic and Roman Periods from Alexander the Great down to the Reign of Constantine (323 B.C.-A.D. 337)*. Ann Arbor: University of Michigan Press.

Mehler, Alexander, Andy Lücking, Sven Banisch, Philippe Blanchard, and Barbara Job, eds. 2016. *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Understanding Complex Systems. Berlin: Springer.

Mitzenmacher, Michael. 2004. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2): 226–251.

Mongin, Philippe. 2018. "A Game-Theoretic Analysis of the Waterloo Campaign and Some Comments on the Analytic Narrative Project." *Cliometrica* 12 (3): 451–480.

Pantelia, Maria. 2020. "Thesaurus Linguae Graecae." 2020. http://stephanus.tlg.uci.edu.

Parker, Robert, Jean-Baptiste Yon, and Mark Depauw. 1996. "Lexikon of Greek Personal Names." 1996. http://www.lgpn.ox.ac.uk.

"PHI Greek Inscriptions." n.d. https://inscriptions.packhum.org/.

Piantadosi, Steven T. 2014. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychonomic Bulletin & Review* 21 (5): 1112–1130.

Reed, William J., and Barry D. Hughes. 2003. "Power-Law Distributions from Exponential Processes: An Explanation for the Occurrence of Long-Tailed Distributions in Biology and Elsewhere." *Scientiae Mathematicae Japonicae* 8: 329–339.

Rochat, Yannick. 2014. "Character Networks and Centrality." PhD thesis, Lausanne: Université de Lausanne. http://infoscience.epfl.ch/record/203889.

Rydberg-Cox, Jeff. 2011. “Social Networks and the Language of Greek Tragedy.” *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1 (3): 1–11.

Saichev, Alex, Yannick Malevergne, and Didier Sornette. 2010. *Theory of Zipf’s Law and Beyond.* Lecture Notes in Economics and Mathematical Systems 632. Berlin: Springer.

Saïd, Suzanne, and Monique Trédé. (1990) 1999. *A Short History of Greek Literature.* Translated by Trista Selous. London: Routledge.

Tauber, James. 2021. “Jtauber/Diorisis.” Exploring the Diorisis Ancient Greek Corpus. 2021. https://github.com/jtauber/diorisis.

Vatri, Alessandro, and Barbara McGillivray. 2018a. “The Diorisis Ancient Greek Corpus.” May 2, 2018. https://figshare.com/articles/dataset/The_Diorisis_Ancient_Greek_Corpus/6187256.

———. 2018b. “The Diorisis Ancient Greek Corpus: Linguistics and Literature.” *Research Data Journal for the Humanities and Social Sciences* 3 (1): 55–65.

Waumans, Michaël C., Thibaut Nicodème, and Hugues Bersini. 2015. “Topology Analysis of Social Networks Extracted from Literature.” *PLOS ONE* 10 (6): e0126470.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Cambridge: Addison-Wesley.