N° 2 2005

di Lorenzo Caselli

La funzione civile degli aziendalisti

saggi

Lorenzo **Caselli**

Dove vanno le nostre discipline. I contributi di Adalberto Alberici, Claudio Baccarani, Guido Corbetta, Gianfranco Rusconi, Dario Velo

Giovanni **Padroni**

Scenari della complessità e vincoli etici nell'organizzazione delle risorse umane

Pierpaolo **Andriani**

'De Rerum Pareti': On Power Laws and Organization Science

Nicoletta **Buratti**

Note a commento del saggio di P. Andriani "De Rerum Pareti: On Power Laws and Organizaion Science". Riflessioni sul contributo della teoria della complessità alla ricerca in campo manageriale

Gianni **Cozzi**

Marketing Territoriale o Marketing delle analogie? Alcune riflessioni da un esame sommario del caso genovese

Riccardo **Spinelli**

La e-internationalisation: una rilettura del modello dei vettori di internazionalizzazione nell'ottica dell'e-business

working paper

Clara **Benevolo**
Luca **Bianchi**

L'internazionalizzazione delle imprese in Cina: tra ostacoli e nuove opportunità. Il caso Esaote.

Sonia **Ruggiero**

I rapporti "Comune controllante - società controllata" nel settore dei servizi pubblici locali

Sara **Poggesi**

I servizi pubblici locali italiani alla luce del network management

Roberto **Garelli**

Bilancio d'esercizio e bilancio sociale nelle organizzazioni di volontariato

Sara **Campi**
Angelo **Gasparre**

Organizational Boundaries in Childcare Services System and the promotion of NPO Networks by Local Public Agencies in an Italian Metropolitan Area

Clara **Benevolo**
Riccardo **Spinelli**

Second Level Organisations (SLOs) in Voluntary Sector: an Italian Perspective

< indietro

Reg. Tribunale di Genova, n. 11/2004 del 31 maggio 2004 ISSN 1824-3576 Cod. CINECA E187020 p.iva 00754150100

IMPRESAPROGETTO
RIVISTA ON LINE DEL DITEA
n. 2 - 2005

‘De Rerum Pareti’: on Power Laws and Organisation Science

Pierpaolo Andriani

1 Introduction

Vilfredo Pareto can be credited with discovering one of the fundamental laws that regulates network-based complex systems. From the distribution of wealth to the size of cities, from the extinction of species to the distribution of matter in the universe, a seemingly mysterious law seems to dominate disparate networks. The first evidence of this distribution comes from the studies of wealth distribution in Western economies, carried out by Pareto at the end of the 19th century (Pareto, 1897). Pareto discovered that wealth distribution is inherently unfair. Large part of wealth is concentrated in a handful of people. At the opposite end of the spectrum, most people end up dividing a decreasing fraction of the total pie. Wealth seems to attract more wealth with a force proportional to its amount, in a way reminiscent of the biblical Matthew effect “…unto every one which hath shall be given … “(Luke). Commonly known as the 80/20 rule, this distribution has turned out to be as ubiquitous as mysterious. There is virtually no sector in natural and social sciences in which the Paretian distribution doesn’t play a central role.

I will in the following present some historical evidence about the emergence of the Paretian distribution in several fields. Then I will formally introduce the concept and characteristics of a Paretian (usually known as power law distribution). I will propose a unifying scheme for the disparate theories and models that are based on power laws. In the final part I will discuss (summarily) some implications, especially regarding the importance of extreme events and the type of approach we need to deal with extremes.

2. A look at the diffusion of the paretian distribution

The road opened by Pareto laid dormant until the second decade of the new century when Auerbach (Auerbach, 1913) observed that the distribution of sizes of cities obeyed a Paretian distribution: if one ranks cities in terms of number of inhabitants (size), one discovers that there is a fixed ratio between the number of cities with different size. This ratio holds true irrespective of time and place (although it may slightly change in different countries). Moreover, plotting (on a double logarithm graph) the rank of American cities against population, the graph shows a straight line with a slope of almost exactly -1. This means that for every city with 1 million inhabitants there are two with half a million and so on. A look at Figure 1 shows that this is a highly skewed distribution with most of the events packing around small values of the x-axis and very few points at the other extreme. In the middle the distribution is decreasing. The cities with an unusual large number of inhabitants generate a long tail in the distribution, which, as we shall see later, is the cause of a very interesting dynamics, whose meaning is still the object of a heated debate. With time it came to be realized that this behaviour (usually attributed to Zipf (Zipf, 1949)) is far from being isolated. This regularity is persistent in time and space (Krugman, 1996) and explaining it has been a challenge for economists and geographers for a long time. As Krugman puts it: *we are unused to seeing regularities this exact in economics – it is so exact that I find it spooky* (Krugman, 1996: 40).

In biology Yule (Yule, 1925), refining results obtained by Willis (Willis, 1922), observed that the number of species per genus follows a power law. Yule was expecting a distribution dominated by a typical scale. He found, instead, that species per genus distribution, likewise cities and wealth, follow a highly asymmetric distribution, that is, a Paretian distribution.

In 1916, Estoup (Estoup, 1916) and later Zipf (Zipf, 1949) found that a power law applies to language (word frequencies). Casti (Casti, 1994) shows that, whereas a monkey at a typewriter generates different words of equal length at equal probability, word usage in the English follows a perfect power law—if word usage frequencies and rank-order are plotted on double-log scales, the words, *the*, *of*, *and*, *to*, *I*, *or*, *say*, *really*, *quality* diminish at a perfect –1 slope.

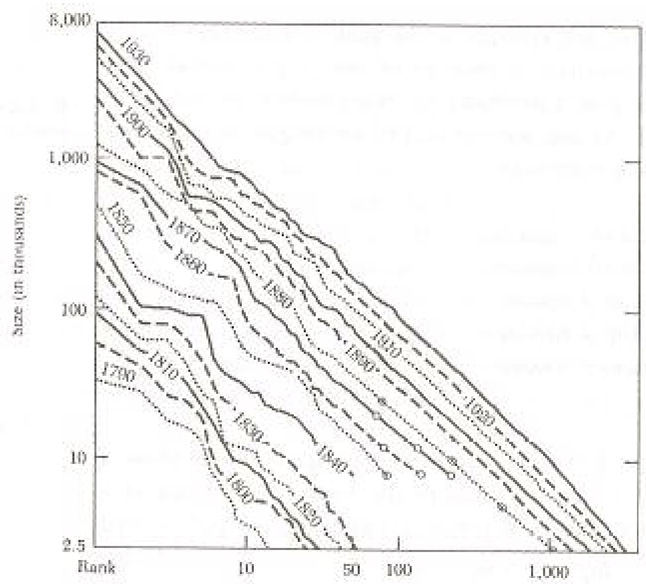

Figure 3.8: U.S.A. 1790-1930. Communities of 2500 or more inhabitants, ranked in decreasing order of population size. [From ref. (13)]

Figure 1: Auerbach Law (from (West *et al.*, 1995)

One of the most interesting development for power law distribution comes from an unlikely source: seismology. Richter and Gutenberg (Gutenberg *et al.*, 1944), generalising a previous result by Omori (Omori, 1895), published a famous paper in which they present what is still today one of most compelling evidence for the existence and relevance of power laws: the distribution linking frequency and magnitudes of earthquakes.

In finance and economics, power laws were first noted by Pareto (Pareto, 1897). Power laws were 'rediscovered' in the 20th century by Mandelbrot (Mandelbrot, 1963) and spurred a wave of interest in finance (Fama, 1965); (Montroll *et al.*, 1984). However, the rise of the "standard" model of efficient markets (signified by Portfolio Theory (Markowitz, 1959), the Capital Asset Pricing Model (Sharpe, 1964), and the Black-Scholes (Black *et al.*, 1973) Option Pricing Theory) sent power law models into obscurity from which they emerged in the 90s as a reaction to the occurrence of catastrophic events, such as the 87 financial market crash, that the standard model finds difficult to explain (Bouchard *et al.*, 1998). The case against the "standard" model is set by Mandelbrot (Mandelbrot *et al.*, 2004) with a simple observation:

…By the conventional wisdom, August 1998 simply should never have happened…. The standard theories…would estimate the odds of that final, August 31, collapse, at one in 20 million—an event that, if you traded daily for nearly 100,000 years, you would not expect to see even once. The odds of getting three such declines in the same month were even more minute: about one in 500 billion (p. 4)…. [An] index swing of more than 7 percent should come once every 300,000 years; in fact, the twentieth century saw forty-eight such days (p. 13).

The reason for the discrepancy between reality and theory lies in the crucial assumption by Finance Orthodoxy: variations in price are statistically independent, and normally distributed. These assumptions allow the use of calculus, modern probability and statistical theory, and give rise to a vast edifice of extreme mathematical sophistication. However, they conflict with reality. The price of virtually any stock or commodity exhibits a punctuated equilibrium behavior, in which chaotic and turbulent periods alternate with stable ones (Mandelbrot, 1963); (Fama, 1965); (Bouchard *et al.*, 1998); (Moss, 2002). Punctuated equilibria indicates that the system is self-critically organised, that is, the system exhibits both long-range and long-term correlations. Price variations are neither independent nor normally distributed. As we shall see later in section, self-critically organised systems are based on power laws.

Another interesting and recent example of power laws in the econosphere appears in the book, *Hollywood Economics*, by De Vany (De Vany, 2004). Throughout the book he shows that the occurrence of unprofitable and highly profitable movies are Pareto distributed. He demonstrates that ‘fat tails’ dominate the movie industry. Fat tails are generated by extreme events that should be negligeable in a Gaussian world. The consequence of the Paretian ‘obedience’ is the inherent chaotic behaviour of the industry. Movies don’t seem to show any significant correlation between any of the variables used to predict final profits. Budget is uncorrelated with earnings, the star system gives no indication about final success. Forecasting is an empty word. The movie industry is chaotic. The only recognisable pattern, writes De Vany, is the universal attractor of the Paretian distribution.

The movie industry is not the only information-based industry dominated by extreme events. Pharma companies seem to depend on a surprising small number of blockbuster drugs (Viagra for Pfizer, Zantac for Glaxo, etc.). Though we don’t have information, it wouldn’t come as a surprise that the distribution of profits in the pharma sector follows a power law and the industry were similar to Hollywood. A similar pattern seems common in the publishing industry too

Just a word of caution. Not all power laws are Paretian. For instance, the emergence of power laws in biology is related to scaling. Since the time of Galileo, it was known that the change of a property, say metabolism, is related to the change in another, say mass, according to a simple formula. D’Arcy Thompson (D'Arcy

Thompson, 1917) discusses scaling extensively in his 1917 masterpiece. In 1932 Kleiber (Kleiber, 1932) published his general power law result that connects the scaling of mass with metabolism via a ¾ exponent. This is remarkable as it was thought that metabolism had to scale with volume, therefore with a 1/3 exponent. The reason for this behaviour is related to a fractal branching pattern of distribution of resources. The explanatory and predictive power of the ¾ relationship is astonishing. It applies across an astonishing 27 orders of magnitudes, from submolecular entities to whales (West *et al.*, 2004). It is important to remark that allometric power laws represent a different and non Paretian type of power law. Allometric laws describe reciprocal change in conjugated variables. The fact that the variables are connected by a third invariable quantity, the slope of the power law equation, point to the presence of a conservation principle or fundamental constraint, of which the allometric power law represents the signature. Paretian power laws instead are concerned with distribution of events, that is, with statistics.

3 What a power law is

What Pareto discovered is a class of phenomena, where the ratio between the smallest and the largest event is gigantic. For instance, the ratio between the largest and the smallest urban agglomerate (from village to city) is more than 150.000 (Newman, 2005). The ratio between the smallest earthquake and the largest one is a staggering nine order of magnitude. We get a similar ratio between the largest and the smallest event in economics (largest vs smallest firm, order, profit, wealth, etc.). By contrast, the ratio of the heights of the shortest against the tallest human individual is 4.18 (Newman, 2005). Why are the differences between the extremes in the examples above so distinctly different? What type of dynamical forces underlies the first set of examples (cities, earthquakes and wealth - which we call the unbounded class)? Does it differ from the second set (heights of humans – which we call the constrained class)? If we plot the frequency against the size of events relative to the two classes, we observe two distinct and completely different patterns. In the case of the constrained class, the events are distributed in a bell shaped fashion, commonly known as the Gaussian or normal distribution. Most of the events will pack around the most common value, the mode, which, in a normal distribution, coincides with mean and median. All the others form the tails of the distribution and rapidly decay around the central value. This distribution is completely described by means of two parameters only: mean and variance. Though the normal distribution is only one of the statistical distributions (the other common ones are the binomial and the lognormal), it has rapidly become the dominant one. The great mathematician Poincare' so commented about it at the end of the 19th century: "All the world believes it firmly, because the mathematicians imagine that it is a fact of observation and the observers that it is a theorem of mathematics" (cited in (West *et al.*, 1995): 83).

The events of the 'unconstrained' class instead follow a completely different pattern: the plot of events frequency against events size (or rank) reveals a right skewed decreasing distribution with a very long tail. I will illustrate this point by using the example of the Italian industrial districts[1]. In Figure 2 I show frequency vs size of firms in twenty travel-to-work areas in Italy that, according to the Sforzi-Istat taxonomy (Cannari *et al.*, 2000), are classified as industrial districts. One can observe that most firms are small, but that there is an extremely long tail that makes the data representation very difficult on a linear scale. A double logarithmic transformation (Figure) shows an interesting feature of a power law distribution: it follows a straight line on a double log graph[2]. The type of power law we frequently find in nature and societies is called inverse power law with negative exponent between 1 and 3 (Newman, 2005). It indicates a decreasing distribution with a decreasing density of events. The low density of events in the right-hand part of the x-axis generates the noise seen in the tail of Figure 3. This is due to the fact that the bins in the extreme part of the tail contain only a few events, which causes significant fluctuations in adjacent bins. In order to improve resolution in the most important part of the graph, that is the tail, one can use a technique known as

[1]The agglomerations we consider are the so-called travel-to-work areas (TWAs), known in Italy as *sistemi locali del lavoro*. The dabaset is based on data from ISTAT and Cerved. TWAs are relatively self-contained economic and social units, calculated by dividing a national territory into units that maximize internal home-to-work commuting and minimize inter-TWA commuting ISTAT INdS. 1997. *I sistemi locali del lavoro*. ISTAT: Roma. The basic idea is that the higher the percentage of internal home-to-work commuting taking place within the boundaries of an area, the higher the chance of capturing within the area some territorially-specific social and industrial aspects. TWAs represent an algorithmic way to define the micro-units of analysis of economic geography and economic sociology. To test whether Italian industrial agglomerations follow a power law, I checked whether the logarithm of firm rank against the logarithm of firm size is a straight line and applied a linear regression approach to the different kinds of agglomerations—see Figure 2. The results are statistically significant (r = 0.997; p < 0.0001); they show that interconnected agglomerations of firms very strongly fit the rank/size power law distribution.

[2]This is easy to see: a power law function is given by:

$$f(x) = kx^{-\beta}$$

by taking the logarithm, we get:

$$\ln(f(x)) = \ln(kx^{-\beta}) = \ln(k) - \beta \ln(x)$$

which can be rewritten as:

$$z(y) = c - \beta y$$

This equation represents a linear equation where the exponent of the power law equation indicates the gradient of the straight line on the loglog graph.

logarithmic binning (Newman, 2005). Successive bins are multiplied by a constant factor[3] in order to increase the probability of including at least a few events. Figure 4 shows the result. A nearly perfect straight line indicates that the distribution of firms in Italian industrial districts is Pareto distributed[4]. A simple least square regression can help determine the degree of fitting to a power law.

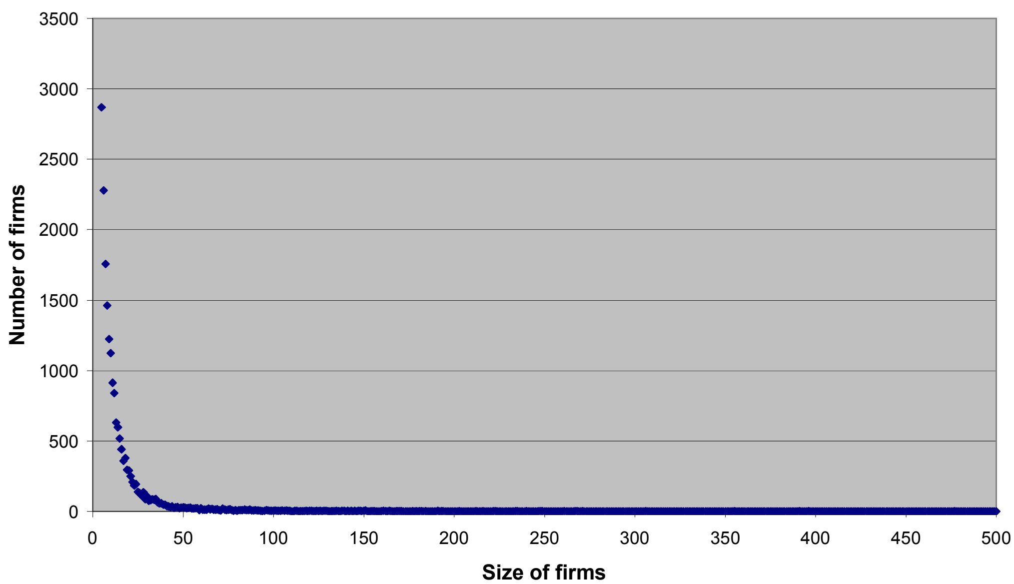

Figure 2: Distribution of firms in cluster areas (Sforzi-Istat classification) – Linear scale

[3]The values have to be normalised dividing them by the amplitude of the respective bin.
[4]This result confirms and extends results obtained by Axtell (Axtell RL. 2001. Zipf distribution of U.S. firm sizes. *Science* **293**: 1818-1820) and previously by Simon (Simon H. 1955. On a class of skewed distribution functions. *Biometrika* **42** (3/4): 425-440) about US firms.

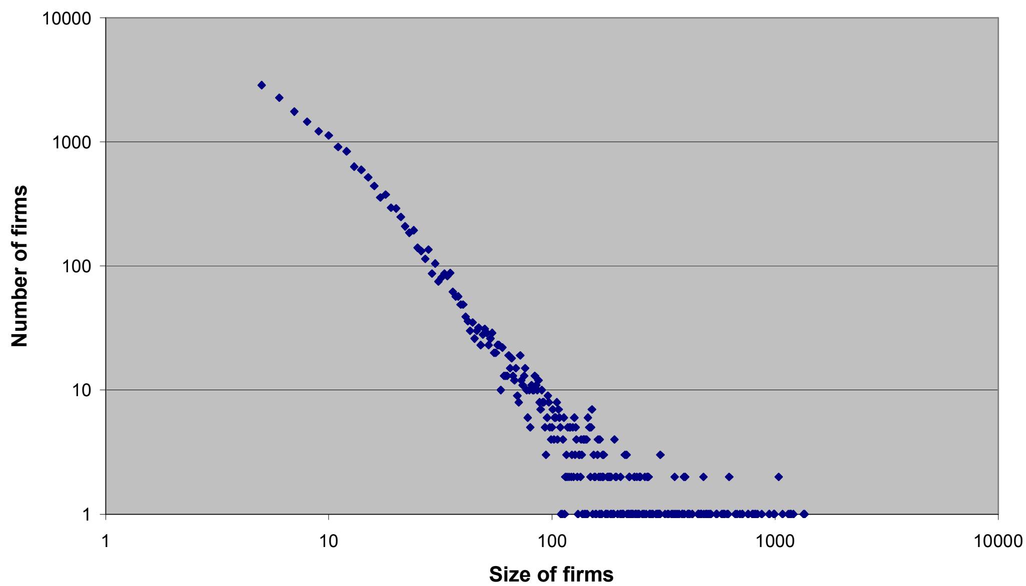

Figure 3: Distribution of firms in cluster areas (Sforzi-Istat classification) - Logarithmic scale

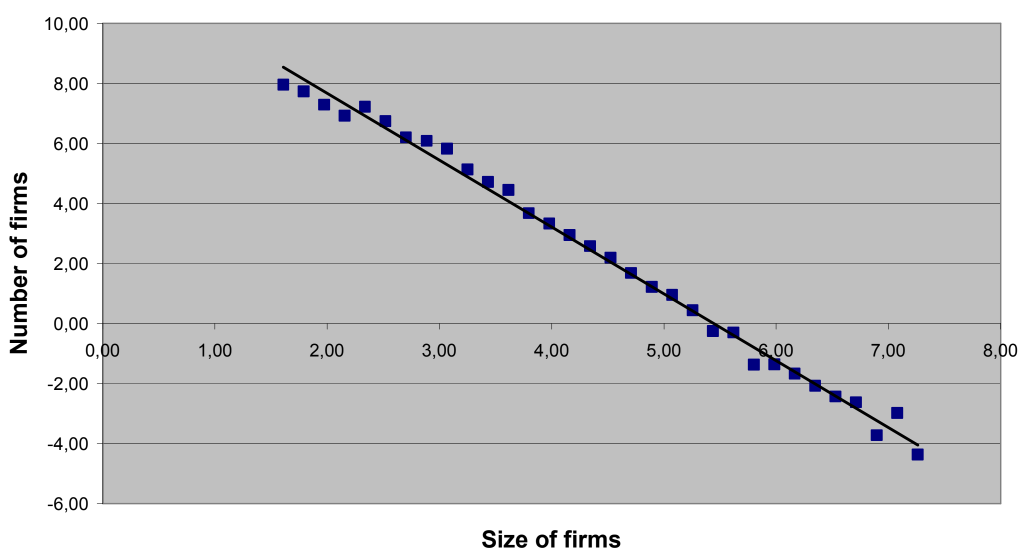

Figure 4: Distribution of firms in cluster areas (Sforzi-Istat classification) - Logarithmic binning

We immediately notice a crucial feature of power law distributions: they lack a typical scale, that is, there is no value that, as in the case of the normal distribution, could be taken to represent the average 'dimension' of the phenomenon. In other words, power law distributions have no meaningful mean. How is it possible? Although one can always calculate a mean, it is a volatile one. This is due to the fact that inverse power law distributions are the signature of systems that scale linearly (also called fractal systems, see section 4.1). These systems are part of a family of distributions named after the French mathematician Cauchy:

As a result of this linear scaling, the distribution of the average of N identically distributed Cauchy variables is the same as the original distribution. Thus, averaging Cauchy variables does not improve the estimate…. This is in stark contrast to all probability distributions with a finite variance, σ^2, for which averaging over N variables reduces the uncertainties by a factor $1/\sqrt{N}$. This nonstandard behavior of the Cauchy distribution is a consequence of its weakly decaying "tails" that produce too many " outliers" to lead to stable averages" (Schroeder, 1991): 159).

Contrary to a normal distribution, where the absence of a long tail makes the first two moments of the distribution (mean and variance) stable, in a power law distribution the extreme events in the tail cause a significant change in the mean and variance. Moreover, mean and variance do not converge even for very large distributions.

4 Three classes of power law theories

I think that the great variety of network-related power law phenomena can be classified in three classes. These are concerned with:

A. Geometric properties

Fractal theory has been the first to demonstrate the versatility and ubiquity of power laws and to build a rigorous mathematical framework about them. The impact of fractals has been profound although its progress has been slow. Fractals have been demonstrated to underpin chaos theory and chaotic attractors, provide a rigorous and more realistic theory about financial markets, and give new insights in virtually every field of knowledge, including the arts. Though the theory of fractals, at least formally, is not directly concerned with networks, it is nonetheless very useful in order to understand network's properties and dynamics, and has to be incorporated into this classification.

B. Spatial or structural properties of systems

We have two sets of theories: the former groups theories and models that study structural and/or spatial properties of networks assuming as unit of analysis either nodes or links. The rank-size rule focuses on nodes, which can be cities (density of population in a country), words (frequency in languages), profit of firms (production of wealth), etc. The latter group of models examines the network's connectivity pattern and derives generic features of networks from their connectivity topology. The two major models in this subcategory are the Scale-Free Networks theory by Barabasi and colleagues (Barabási, 2002; Barabási *et al.*, 1999) and the Small World theory (Watts, 2003; Watts *et al.*, 1998).

C. Temporal and evolutionary properties

This group covers theories that describe network's dynamic. In this case, a power law emerges in the rate of change of some behavioural properties of a natural/social system, subjected to a perturbation of some kind. Whereas the class B focuses on the type of distribution of the network-forming elements (nodes or links), this class analyses network's emergent collective behaviour. Classical examples are Bak and Chen's *self-organised criticality* (Bak, 1996; Bak *et al.*, 1988) and phase transition models in physics. In both cases the emergence of a power law is due to a form of critical behaviour. However, in SOC the system evolves spontaneously towards the critical threshold, whereas in phase transition models some parameters must be externally fine-tuned to achieve criticality.

4.1 Fractals

Fractal geometry was developed by Mandelbrot (Mandelbrot, 1975) to make sense of the rough, recursive and irregular shapes of most natural objects, from cauliflowers to coastlines, trees, and galaxies. As Mandelbrot (Mandelbrot, 1975): 1) writes: "*Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line*." The coasts of England or Norway exemplify fractality: the length of the coast profile depends with inverse linear proportionality on the length of the ruler—i.e., the smaller the ruler, the longer the coast. The length of coast is indefinite although the area within the curve is finite. A fractal (Mandelbrot *et al.*, 2004): 118) is: "a pattern or shape whose parts echo the whole." Fractals are self-similar objects. As for power laws, some of the original ideas regarding fractals go back to the 19th century. Fractals were controversial from the very beginning. Charles Hermite commented that mathematicians should be "turning away in fear and horror from this lamentable plague of fucntions with no derivatives" (Schroeder, 1991)

Fractals are not idle mathematical curiosities. Fractals and power laws are found from atoms (~10^{10} meters) to galactic megaparsecs (~10^{22} m)—across a range of 32 orders of magnitudes (Baryshev *et al.*, 2002). In biology, West and Brown (West *et al.*, 2004) demonstrate an allometric power law relationship between the mass and metabolism of virtually any organism and its components—based on fractal geometry of distribution of resources—across an astounding 27 orders of magnitude. Self-similarity is key to a fundamental property of fractals and power laws: linear scalability. Power law systems do not exhibit a characteristic scale and consequently enjoy some peculiar statistical properties.

4.2 Spatio-structural properties of systems: the new science of networks

This category groups spatio-structural properties of networks, assuming as unit of analysis either nodes or links. I mention two: (1) rank-size rules focusing on nodes, which can be cities (size of population), words (frequency in languages), profits of firms (production of wealth), etc.; and (2) connectivity patterns that derive generic features of networks from their connectivity topology.

The legendary Hungarian mathematician Paul Erdos, in introducing random network theory, assumed links are randomly distributed across nodes and form a bell-shaped distribution, wherein most nodes have a typical number of links with the frequency of remaining nodes rapidly decreasing on either side of the maximum. Watts and Strogatz (Watts *et al.*, 1998) show, instead, that many real networks follow the *small world* phenomenon, whereby society is visualized as consisting of weakly connected clusters, each having highly interconnected members within. This structure allows cohesiveness (high clustering coefficient) and speed/spread of information (low path length) across the whole network.

In their initial *small world* model, Watts and Strogatz also assume that links are Gaussian distributed. Barabási and colleagues (Barabási, 2002), however, studying the world wide web, find that the structure of the Web shows a power law distribution, where most nodes have only a few links and a tiny minority—the hubs—are disproportionately very highly connected. The system is scale-free, that is, no node can be taken to represent the scale of the system. Defined as *Scale-free Networks* (SFNs), the relative distribution shows infinite variance and the absence of a stable mean. It turns out that *small world* networks are scale-free and follow power laws (Song, Shlomo, and Makse, 2005). SFNs appear in fields as disparate as epidemiology, metabolism of cells, Internet, and networks of sexual contacts (Barabási, 2002).

4.3 Self-organized criticality

Self-organized criticality describes the dynamic behaviour of a network, not the distribution of its parts as in the previous case. In this category, a power law characterizes the nature of the behavioral properties of a system subjected to a perturbation of some kind. Whereas the former category focuses on the type of distribution of the network-forming elements (nodes and links), this one analyses a network's emergent collective behavior. Classical examples are Bak's (1996) *Self-Organized Criticality* (SOC) and phase transition models in physics. In both cases the emergence of a power law is due to a state of critical connectivity. However, in SOC the system evolves spontaneously towards the critical threshold, whereas in phase transition models the order parameters must be fine-tuned by an external agent (i.e., energy source) to achieve criticality.

This group of models is symbolized by Bak's (Bak, 1996) sandpile experiments. A sandpile subjected to an infinitesimal external perturbation (such as adding a single sand grain) evolves toward a critical state, characterized by a critical slope, whereby any additional perturbation induces a systemic reaction that can span any order of magnitude, with a frequency distribution expressed by a power law. This behavior is counter-intuitive. We generally assume a linear relationship between perturbation size and system's reaction, i.e., small causes yield small effects. This is true before SOC is attained. Before criticality, each falling grain has a constant probability of displacing an adjacent grain. The probability of an avalanche therefore scales exponentially with the number of sand grains. This makes large avalanches highly unlikely. Instead, at criticality, a power law distribution is a consequence of the global connectivity of the sandpile, which is a result of the accumulating interdependencies among the sand grains. In other words, SOC dynamics arise when an emergent system of links connects local pockets into a coevolving whole such that small and local fluctuations can be amplified to achieve systemic effects. As Bak (Bak, 1996): 60) writes: "In the critical state, the sandpile is the functional unit, not the grain of sand." Mathematically this means that the behavior of the avalanches obeys a power law of the type: $F \sim S^{-\alpha}$, where F represents avalanche frequency with size S.

SOC is very common in nature (Buchanan, 2000). From the dynamics of earthquakes, the succession of booms and busts in economic cycles (Krugman, 1996), to the dynamics of supply chains (Scheinkman *et al.*, 1994), there seems to be a common pattern across disparate fields. A few implications follow. First, the fact that a self-critical system spontaneously tunes itself towards a self-critical state (Bak *et al.*, 1991; Kauffman, 1995), where "…the system organizes itself towards the critical point where single events have the widest possible range of effects" (Cilliers, 1998): 97), makes reductionism inappropriate for the study of SOC. Second, extreme events ceases to be the outlyers of the Gaussian distribution, which we are safe to ignore, but become instead the necessary outcomes of the system's connectionism. Interdependence generates outcomes that depend on the amplification provided by the occurrences of joint probabilities For instance, the conventional explanation regarding mass extinctions (e.g., dinosaurs at the end of Cretaceous Period) is imputed to exogenous events (asteroid or eruptions or both). Instead, according to SOC, internal causes may have been progressively amplified until a catastrophic chain reaction took place (Gould, 2000); (Raup, 1999): 217–218). Fundamentally, self-organized criticality (SOC) is a theory about endogenously initiated nonlinear change in systems.

5. Indipendence versus interdipendence

So far our separation of power law phenomena into geometric, spatio-structural, dynamic phenomena begs the question whether the different phenomena described by power laws share a common property. Mandelbrot (Mandelbrot, 1963); quoted in 2004: 170) writes:

…The cotton story shows the strange liaison among different branches of the economy, and between economics and nature. That cotton prices should vary the way income does; that income variations should look like Swedish fire-insurance claims; that these, in turn, are in the same mathematical family as formulae describing the way we speak, or how earthquakes happen—this is, truly, the great mystery of all.

Simon (Simon, 1955): 425) pointed to a common probability mechanism:

[The power law's] *appearance is so frequent, and the phenomena in which it appears so diverse, that one is led to the conjecture that if these phenomena have any property in common, it can only be a similarity in the structure of the underlying probability mechanism*.

Others also argue that the appearance of power laws points to common underlying dynamic and coevolutionary mechanisms (Bak, 1996; Lee *et al.*, 1998; Shin *et al.*; West *et al.*, 2004). Stanley, a founder of econophysics, writes:

If the same empirical laws hold for the growth dynamics of both countries and firms, then a common mechanism might describe both processes. (Stanley *et al.*, 1996): 3277)

Whatever your stance on this point, organizations can make use of the commonality. In fact:

"*relatively simple patterns, known as power laws and observed in disparate settings from astrophysics to evolutionary biology, as well as in human society,* *suggest strategies by which well-managed organizations can deal with uncertainty and navigate the discontinuities of contemporary business"* (Buchanan, 2004)

In a separate paper (Andriani *et al.*, 2005) Bill McKelvey and I advance the hypothesis that the explanation for the predominance of power laws is *interdependence among agents* (data points) that—with some probability—leads to fat tails and consequent power law effects. The long tail effects are more evident in seismology, social networks, economics and finance, but plenty of evidence indicates they appear everywhere including organizations.

5.1 *Averages instead of tails*

Linear thinking is engrained in our mentality. Scientific and mathematical models are based on the concepts of equilibrium and linearity[5], which allows efficient causality to operate, equations to be solved and forecasting modeling to be elaborated. Economics, for instance, is almost theistic in the (scarcely verified) hypothesis that economic phenomena trend toward (general) equilibrium. The operational arm of science and economics, calculus, is limited to continuity. One of the implications of treating reality as mainly continuous and linear is the ostracisation of innovation, treated as a *black box*. Innovation with its content of *creative destruction* (Schumpeter, 1942) is seen as a disturbance to the economic system, inevitably followed by the restoration of equilibrium. Given the dominance of linear thinking and economics, it is no surprise therefore that as economists and math modelers have more and more come to dominate business and management studies and practices, equilibrium models and linear thinking have also come to dominate over nonlinear phenomena in organisational studies.

What is the relevance of this philosophical digression for our power law argument? By focusing on systems in equilibrium, we implicitly accept that the number of possible states that our system can attain is limited and that the search time following the onset of instability is short compared to the 'equilibrium' time. For this to be true, the system must be internally weakly connected[6] so that many internal events can effectively be considered as independent. Weak connectivity plays well with linearity as it allows interdependencies to be treated as second order effects (by means of perturbation analysis, see for instance (West *et al.*, 1995)) and

[5]Linearity means two things: 1)proportionality between cause and effect and 2)superposition, that is that the dynamic of a system can be reconstructed by summing up the effects of the single causes acting on the single components (Nicolis Prigogine).

[6]Otherwise, following Kauffman's NK landscape, the system would collapse into a complexity catastrophy and would incessantly circle into its state phase with a very low probability of finding attractors (Kauffman, 1995).

with Gaussian statistics, as it excludes, *ipso facto*, the amplification mechanisms that give rise to the 'fat tails' of the Paretian distribution. If we take 100 companies belonging to the same sector and approximately of the same size and we assume independence, and plot a variable, say profit, we expect (according to linear thinking) to find that most events will pack around the most common one, the average, in a rapidly decaying distribution that follows a bell (or Gaussian) curve. A bell shaped distribution is the most studied statistical distribution and it underlies much of our understanding of the natural and social world. However the crux of the point above is independence of events. In real life, those companies will benchmark against each other, imitate those perceived as successful, exchange information, organize cartels, organize mergers and undergo acquisitions, compete for limited resources, etc. In a word, they are interdependent and not independent! The statistical distribution governing interconnected situation doesn't give rise to a bell distribution but instead to a power law distribution.

The two distributions are radically different. The main feature of the normal distribution can be entirely characterized by its mean and variance. A power law instead doesn't even show a proper mean let alone a variance, which is infinite (Newman, 2005). A power law therefore has no value that can be assumed to represent the typical features of the distribution (it is therefore defined as *scale-free*) and no standard deviations.

There are two major implications for this:

1. The dream of social science of building robust frameworks that allow social scientists to predict the evolution of social phenomena get shattered by the absence of statistical regularities in phenomena dominated by persistent interconnectivity. In fact, absent mean and standard deviation, the very basis for probabilistic assessment of outcomes is denied. This point reflects the more pervasive and structural issue of nonlinearity and emergence in complex systems. Linearity assumes the divisibility of systems into modules whose dynamic can be inquired irrespective of the context. This point gives rise to the independence assumption, which, we have seen playing a key role in generating bell shaped distribution.
2. Power law tails decay more slowly than those of normal distributions. These "fat" tails affect systems' behaviors in significant ways. Extreme events, that in a normal distribution world could be safely ignored, are not only more common than expected but also of vastly larger magnitude and consequential. For instance financial market drops of 10% in one day should occur once every 500 years according to a normal distribution. They instead tend to occur once every five years (Mandelbrot *et al.*, 2004).

5.2 Statistics: obscuring rather than clarifying?

A power law world is dominated by extreme events that we are safe to ignore in a Gaussian-world. In fact, the 'fat' tails of power law distributions make large extreme events orders of magnitude more likely. In a 'normal' world, where distributions show a typical scale, extreme events are so different from the type and so rare that they don't significantly influence either the mean or the variance. Hence ignoring them is a safe strategy. However, insurance companies that used normal distributions to assess likelihood of extreme events got their fingers burned. The devastations followed flooding in Central Europe in 2003, the multiple yearly recurring cyclones hitting the American coasts, the multiple earthquakes of scale 7 and higher, etc. indicate that we are not in a 'normal' world. On the contrary, the action seems to be in the tails. In the movie industry, almost all the profit come from the blockbusters, that is the extreme events, with the majority of the movies contributing next to nothing to profitability and revenues (De Vany, 2004). If this is true, normal distribution statistics is obscuring the matter rather than clarifying it. The practice of looking for the mean, that gives an indication about the scale of the phenomenon, the reliance on variance to build confidence interval and therefore assess likelihood of single event; and even more the practice of excluding outlying events become irrelevant or openly wrong in a power law-world. We need a statistics that basically concentrate our attention on the 'fat' tails of the distribution. This statistics is the object of the following section.

5.3 Different kind of statistics?

A non-Gaussian world demands a statistics that takes into account path-dependency, nonlinearities, emergent properties of systems and the dynamics of multiple punctuated equilibria. The assumption of independence of events, which underlies the Gaussian world and the reductionist approach, leads to the wrong tools and conclusion when dealing with connectionist dynamics. The demise of reductionism causes the demise of predictability of single events. Given mean and variance, Gaussian statistics allows the prediction of the occurrence probability of the next event. Instead where extreme events dominate and variability is infinite, the maximum statistics can do is to indicate the shape of the distribution, that is, the general attractor toward and around which the events will tend to self-organize. This attractor is a universal one, which means that the connectionist dynamics underlying power laws causes the aggregate of events to follow a double log linear graph. The universality of the graph is confirmed by the fact that the relationships among events hold across time and space. Approximately the same graph with a very similar slope holds for American cities across 200 years (see Figure 1)and (Auerbach, 1913; Krugman, 1996)). Power laws for cities are found in other nations. Again a power law (or Paretian) distribution is found in the movie industry in every country in which research has been carried out (De Vany, 2004). This doesn't mean that the same features apply in different cultural contexts but that for instance the share of profit of blockbusters to unsuccessful movies is the same.

What are the main features of a generalized power law based statistics:

1. The mode (most frequent event) is smaller than the median (central point) which is smaller than the mean. Contrarily to the Gaussian, a power law distribution mean is strongly influenced by large extreme events.
2. Universality: the dynamics of connectionist phenomena lead to a universal power law distribution that is valid for the same variable across time and space. A power law distribution can therefore be considered as a universal attractor from which the dynamics of phenomena are attracted.
3. Scale-free: in general a power law distribution shows no typical scale (no mean). As De Vany (De Vany, 2004: 258) writes: "*in this world nothing is "typical" and every movie is unique".*
4. Infinite variability: the variance of the distribution is very large (approaching infinity). This means that the use of variance for forecasting leads to a probability distribution as wide as the original distribution
5. "*Nobody knows*" principle: the prediction of single events is meaningless. Events are to be intended as probability distribution. Prediction is possible only at the level of the aggregate of events.
6. Cascade dynamics: the power law results from a generalized self-organized criticality dynamics. As an event derives from the propagation of a signal under conditions of positive feedback, the logic of preferential attachment holds. For instance, in the case of information based cascades, success breed success.
7. A business of extremes: the important part of the statistics is in the tails. Extreme events are more frequent and disproportionate is size than in a Gaussian dominated world.
8. Self-similarity: the shape of the distribution looks the same at any scale. What this suggests is that common dynamical patters are in action at different levels. Whether we take the whole series of events or sample a part of it, we find the same pattern of large discontinuous events irregularly casted against a background of finer perturbations. In other words, we need a fractal statistics.
9. Linear amplification: fat tails result from the amplification of simple causes that could evolve to generate events of any size. The major difference between a Gaussian and a power law distribution is that the former 'tends' to quench events (in fact the assumption of independence kills at its root the

positive feedback that gives rise to large events) whereas the latter to ‘amplify’ them.

6. Conclusions

I’ve tried in this essay to do two things: first to credit Pareto for the discovery of what has become one of the greatest successes of complexity theory, that is, the power law (or Paretian) distribution. Pareto has opened a direction for research that has expanded in virtually every field of knowledge and contributed to change our view of the world. The second part of this essay looks at (albeit in a very concise way) the impact that the Paretian distribution has had on the way we conceptualise and carry out research on dynamic phenomena, especially with regards to the type of statistics that is appropriate to use.

In conclusion, Vilfredo Pareto’s merit is to have started, probably unwillingly, that trend of research that has demonstrated that behind the apparent randomness or stocasticity of real world phenomena there is a kind of hidden order, reflected by the ubiquity of the Paretian distribution.

Aknowledgement

This paper draws its inspiration and part of its material from a much longer paper that Bill McKelvey (from the Anderson School of Management, at UCLA) and I have written on a similar subject. In this paper I have focussed, in a necessary kaleidoscopic manner, on some specific aspects of power laws and tried to show how Pareto’s seminal intuition has developed into a fascinating field.

References

Andriani P, McKelvey B. 2005. Power Law Phenomena in Organizations: Causes and Consequences for Managers. Working paper. Durham Business School, University of Durham: Durham, UK

Auerbach F. 1913. Das Gesetz der Bevolkerungskoncentration. *Petermanns Geographische Mitteilungen* 59: 74-76

Axtell RL. 2001. Zipf distribution of U.S. firm sizes. *Science* 293: 1818-1820

Bak P. 1996. *How Nature Works: The Science of Self-Organized Criticality*. Copernicus: New York

Bak P, Chen K. 1991. Self-organized criticality. *Scientific American* 264(January): 46-53

Bak P, Tang C, Wiesenfeld K. 1988. Self-organized Criticality. *Phys. Rev. A* 38: 364-374

Barabási AL. 2002. *Linked: the new science of networks*. Perseus Publishing: Cambridge, MA

Barabási AL, Albert L. 1999. Emergence of scaling in random networks. *Science* 286: 509-512

Baryshev Y, Teerikorpi P. 2002. *Discovery of Cosmic Fractals*. World Scientific: River Edge, NJ

Black F, Scholes M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637-654

Bouchard JP, Sornette D, Walter C, Aguilar JP. 1998. Taming large events: Optimal portfolio theory for strongly fluctuating assets. *International Journal of Theoretical and Applied Finance* 1(1): 25-41

Buchanan M. 2000. *Ubiquity the science of history - or why the world is simpler than we think*. Weidenfeld & Nicolson: London

Cannari L, Signorini LF. 2000. Nuovi strumenti per la classificazione dei sistemi locali. In LF Signorini (Ed.), *Modelli di sviluppo locale: produzione, mercati e finanza: I risultati di una ricerca della Banca d'Italia sui distretti industriali italiani*. Meridiana Libri: Roma

Casti JL. 1994. *Complexification: explaining a paradoxical world through the science of surprise*. Abacus: London

Cilliers P. 1998. *Complexity and Postmodernism*. Routledge: London

D'Arcy Thompson. 1942. *On growth and form.* Cambridge University Press: Cambridge

De Vany A. 2004. *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge: New York

Estoup JB. 1916. *Gammes Stenographiques*. Institut Stenographique de France: Paris

Fama EF. 1965. The behavior of stock-market prices. *Journal of Business* 38 (1): 34-105
Gould JS. 2000. *Wonderful Life: the Burgess Shale and the nature of history*. Vintage: London
Gutenberg B, Richter RF. 1944. Frequency of earthquakes in california. *Bulletin of the Seismological Society of America* 34: 185-188
ISTAT INdS. 1997. *I sistemi locali del lavoro*. ISTAT: Roma
Kauffman SA. 1995. *At home in the universe*. Oxford University Press: Oxford
Kleiber M. 1932. Body size and metabolism. *Hilgardia* 6: 315-353
Krugman P. 1996. *The self-organizing economy*. Blackwell: Maiden, MA
Lee Y, Amaral LAN, Canning D, Meyer M, Stanley HE. 1998. Universal Features in the Growth Dynamics of Complex Organizations. *Physical Review Letters* 81(15): 3275-3278
Luke. *Holy Bible (19:26)* (King James Version ed.)
Mandelbrot BB. 1963. The Variation of Certain Speculative Prices. *Journal of Business* 36: 394-419
Mandelbrot BB. 1975. *Les object fractales*. Flammarion: Paris
Mandelbrot BB, Hudson RL. 2004. *The (mis)behaviour of markets: a fractal view of risk, ruin and reward*. Profile Books: London
Markowitz H. 1959. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press: New Haven: CT
Montroll E, Shlesinger M. 1984. On the wonderful world of random walks. In JL Lebowitz, EW Montroll (Eds.), *Nonequilibrium Phenomena II, From Stochastic to Hydrodynamics*. North Holland: Amsterdam
Moss S. 2002. Policy analysis from first principles. *Proceedings of the National Academy of Sciences* 99(Suppl. 3): 7267-7274
Newman MEJ. 2005. Power laws, Pareto distributions and Zipf's law. *arXiv:cond-mat/0412004* 2(9 January)
Omori F. 1895. On the aftershocks of earthquakes. *J. Coll. Sci.* 7: 111
Pareto V. 1897. *Cours d'economie politique*. Rouge: Lausanne et Paris
Raup DM. 1999. *The nemesis affair: a story of the death of dinosaurs and the ways of science*. Norton: New York
Scheinkman J, Woodford M. 1994. Self-organized criticality and economic fluctuations. *The American Economic Review* 84(2): 417-421
Schroeder M. 1991. *Fractals, Chaos, Power Laws*. Freeman & Co.: New York
Schumpeter JA. 1942. *Capitalism, Socialism and Democracy* (2nd ed.). George
Allen & Unwin: London
Sharpe WF. 1964. Capital asset prices: a theory of market equilibrium under conditions of risks. *Journal of Finance* 19 (3): 425-442
Shin CW, Kim S. 2004. Self-organized Criticality and Scale-free Properties in Emergent Functional Neural Networks, http://arxiv.org/PS_cache/cond-mat/pdf/0408/0408700.pdf ed.:

Simon H. 1955. On a class of skewed distribution functions. *Biometrika* 42 (3/4): 425-440

Stanley MHR, Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Salinger MA, Stanley HE. 1996. Scaling behavior in the growth of companies. *Nature* 379: 804-806

Watts DJ. 2003. *Six degrees: the science of a connected age*. Heinemann: London

Watts DJ, Strogatz SH. 1998. Collective dynamics of 'small-world' networks. *Nature* 393: 440-442

West BJ, Deering B. 1995. *The Lure of Modern Science: Fractal Thinking*. World Scientific: Singapore

West GB, Brown JH. 2004. Life's Universal Scaling Laws. *Physics Today* 57 (9): 36-42

Willis JC. 1922. *Age and Area: A Study in Geographical Distribution and Origin of Species*. Cambridge University Press

Yule GU. 1925. A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis. *Philos. Trans. R. Soc. London B* 213: 21-87

Zipf GK. 1949. *Human Behavior and the Principle of Least Effort*. Hafner Pub. Co: New York

Pierpaolo Andriani
Durham Business School
University of Durham
Mill Hill Lane, Durham, DH1 3LB
United Kingdom
Tel. +44 191 334 5385
Fax +44 191 334 5201
e-mail: pierpaolo.andriani@durham.ac.uk