

LEIS DE POTÊNCIA EM SISTEMAS NATURAIS E ARTIFICIAIS

Marcus Gonçalves Silva



Mestrado em Engenharia Electrotécnica e de Computadores
Área de Especialização de Sistemas e Planeamento Industrial
Departamento de Engenharia Electrotécnica

Instituto Superior de Engenharia do Porto

2011

Este relatório satisfaz, parcialmente, os requisitos que constam da Ficha de Disciplina de
Tese/Dissertação, do 2º ano, do Mestrado em Engenharia Electrotécnica e de
Computadores

Candidato: Marcus Silva, Nº 1960581, 1960581@isep.ipp.pt
Orientação científica: J. A. Tenreiro Machado, jtm@isep.ipp.pt
Co-Orientador: António Mendes Lopes, aml@fe.up.pt



Mestrado em Engenharia Electrotécnica e de Computadores
Área de Especialização de Sistemas e Planeamento Industrial
Departamento de Engenharia Electrotécnica

Instituto Superior de Engenharia do Porto

5 de Novembro de 2011

Dedico este trabalho à minha Família.

Agradecimentos

Quero expressar o meu agradecimento ao professor António Mendes Lopes e ao professor Tenreiro Machado. O apoio demonstrado, principalmente na parte final, foi importante, pois sem ele não seria provável a conclusão deste trabalho. Agradeço também toda a disponibilidade demonstrada.

Resumo

A Lei de Potência é uma particularidade de um sistema não linear, revelando um sistema complexo próximo da auto-organização. Algumas características de sistemas naturais e artificiais, tais como dimensão populacional das cidades, valor dos rendimentos pessoais, frequência de ocorrência de palavras em textos e magnitude de sismos, seguem distribuições do tipo Lei de Potência. Estas distribuições indicam que pequenas ocorrências são muito comuns e grandes ocorrências são raras, podendo porém verificar-se com razoável probabilidade. A finalidade deste trabalho visa a identificação de fenómenos associados às Leis de Potência. Mostra-se o comportamento típico destes fenómenos, com os dados retirados dos vários casos de estudo e com a ajuda de uma meta-análise. As Leis de Potência em sistemas naturais e artificiais apresentam uma proximidade a um padrão, quando os valores são normalizados (frequências relativas) para dar origem a um meta-gráfico.

Palavras-Chave

Lei de Potência, Lei de Zipf, Lei de Pareto, Fenómenos naturais, Fenómenos artificiais, Meta-gráfico, Meta-análise.

Abstract

The Power Law is a feature of a nonlinear system, revealing a chaotic system near the self-organization. Some characteristics of natural and artificial systems, such as population size of cities, the value of personal income, the frequency of occurrence of words in texts and magnitude of earthquakes, follow a power law distribution. These distributions indicate that small occurrences are very common and large events are rare, but may occur with reasonable probability. The purpose of this work concerns the identification of phenomena associated with power laws. It is demonstrated a typical behavior of these phenomena, using data from several case studies and with the help of a meta-analysis. Power laws in natural and artificial systems have a proximity to a particular patterns, when values are normalized (relative frequencies) to give origin to a meta-graph.

Keywords

Power law, Zipf Law, Pareto Law, natural phenomena, artificial phenomena, Meta-graph, Meta-analysis.

Índice

AGRADECIMENTOS	I
RESUMO	III
ABSTRACT	V
ÍNDICE	VII
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABELAS	XI
ACRÓNIMOS.....	XIII
1. INTRODUÇÃO	1
1.1. CONTEXTUALIZAÇÃO	1
1.2. OBJECTIVOS.....	2
1.3. CALENDARIZAÇÃO	2
1.4. ORGANIZAÇÃO DO RELATÓRIO	3
2. LEIS DE POTÊNCIA EM SISTEMAS NATURAIS E ARTIFICIAIS.....	5
2.1. CRITICALIDADE AUTO-ORGANIZADA	5
2.2. LEI DE POTÊNCIA	6
2.3. O QUE É UMA LEI DE POTÊNCIA	7
2.4. FUNÇÃO LEI DE POTÊNCIA.....	8
2.5. FUNÇÃO DE PROBABILIDADE	8
2.6. LEI DE ZIPF	9
2.7. LEI DE PARETO	9
2.8. LEI DE BENFORD.....	10
2.9. APLICAÇÕES DE LEI DE POTÊNCIA	12
2.10. O FUTURO DA PESQUISA SOBRE LEIS DE POTÊNCIA.....	14
3. ESTUDOS DE LEI DE POTÊNCIA.....	17
3.1. CONSIDERAÇÕES GERAIS	17
3.2. POPULAÇÃO PORTUGUESA.....	18
3.3. OS MAIS RICOS DO MUNDO DA FORBES 2011	20
3.4. MAGNITUDE DOS SISMOS.....	22
3.5. ÁREA ARDIDA EM INCÊNDIOS	24
3.6. FREQUÊNCIA DE PALAVRAS EM TEXTOS.....	27
3.7. SITES MAIS VISITADOS.....	35
3.8. PREJUÍZOS CAUSADOS POR CHEIAS NOS ESTADOS UNIDOS	36
3.9. PREJUÍZOS E MORTES CAUSADOS POR TORNADOS	37

3.10.	MORTES CAUSADAS POR ACIDENTES COM AVALANCHES	39
3.11.	DIÂMETRO DE CRATERAS DA LUA	41
3.12.	DIÂMETRO DE CRATERAS E MONTANHAS EM MARTE	42
3.13.	ELEVAÇÃO DOS VULCÕES.....	43
4.	COMPARAÇÃO DOS ESTUDOS REALIZADOS	45
4.1.	ANÁLISE DE RESULTADOS.....	45
4.2.	GRÁFICO AVANÇADO DE RESULTADOS	48
5.	CONCLUSÕES	53
	REFERÊNCIAS DOCUMENTAIS.....	55
	HISTÓRICO	57

Índice de Figuras

Figura 1	Relação entre o número de vítimas e o número de conflitos [3]	7
Figura 2	Distribuição da lei de Benford para os dígitos de 1 a 9 [5]	11
Figura 3	Procedimento de tratamento de dados.....	18
Figura 4	População das cidades portuguesas.....	19
Figura 5	Gráfico Log-Log da População das Cidades Portuguesas.....	19
Figura 6	Gráfico Forbes 2011 dos mais ricos do mundo	21
Figura 7	Gráfico Log-Log Forbes 2011.....	21
Figura 8	Gráfico ilustrativo da ocorrência de sismos [15].....	22
Figura 9	Gráfico Log-Log Magnitude dos Sismos	24
Figura 10	Gráfico Log-Log área ardida 2009.....	25
Figura 11	Gráfico Log-Log área ardida 2010.....	26
Figura 12	Cabeçalho do <i>software Letter Frequencies</i>	27
Figura 13	Resultado de contagem de palavras.....	27
Figura 14	Gráfico Log-Log “A prenda do Cisne”	28
Figura 15	Gráfico Log-Log “A Senhora dos livros”	29
Figura 16	Gráfico Log-Log “As sapatilhas de Sofia”.....	29
Figura 17	Gráfico Log-Log “O menino Golfinho”.....	30
Figura 18	Gráfico Log-Log “O pastor amoroso”	31
Figura 19	Gráfico Log-Log “Os mil pássaros de Sadako I”	32
Figura 20	Gráfico Log-Log “Os mil pássaros de Sadako II”.....	32
Figura 21	Gráfico Log-Log “O pequeno comerciante que vendia sorrisos”	33
Figura 22	Gráfico Log-Log “Pão de Tâmaras e nozes”.....	34
Figura 23	Gráfico Log-Log <i>sites</i> visitados	36
Figura 24	Gráfico Log-Log custos cheias nos Estados Unidos	37
Figura 25	Gráfico Log-Log custos tornados nos Estados Unidos em 2010	38
Figura 26	Gráfico Log-Log Mortes em tornados nos Estados Unidos entre 2000 e 2010	39
Figura 27	Gráfico Log-Log Mortes em avalanches nos Estados Unidos e Canadá.....	40
Figura 28	Gráfico Log-Log diâmetro de crateras na Lua	41
Figura 29	Gráfico Log-Log diâmetro de crateras em Marte.....	42
Figura 30	Gráfico Log-Log diâmetro de montanhas em Marte.....	43
Figura 31	Gráfico Log-Log Elevação dos vulcões com mais de 1000m	44
Figura 32	Gráfico relação de (α) e (c)	47
Figura 33	Meta-gráfico resultante dos vários estudos	49
Figura 34	Gráfico por grupos de resultados.....	50

Índice de Tabelas

Tabela 1	Calendarização do trabalho	3
Tabela 2	Lista dos mais ricos do mundo [14]	20
Tabela 3	Frequência / Magnitude de sismos [15]	23
Tabela 4	Resumo do número de ocorrências e correspondente área ardida em 2009 [16]	25
Tabela 5	Resumo do número de ocorrências e correspondente área ardida em 2010 [17]	26
Tabela 6	Valores de α	35
Tabela 7	Valores resultantes dos estudos	48

Acrónimos

LP	–	Lei de Potência
Log-Log	–	Gráfico Logarítmico
TXT	–	Texto
Km	–	Quilómetro
m	–	Metro
ha	–	Hectare
EUA	–	Estados Unidos da América

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO

Certas características de sistemas naturais e artificiais, tais como dimensão populacional das cidades, valor dos rendimentos pessoais, transacções bolsistas, frequência de ocorrência de palavras em textos, severidade de guerras e actos terroristas e magnitude de tremores de terra, seguem distribuições do tipo Lei de Potência (LP). Estas distribuições indicam que pequenas ocorrências são muito comuns e grandes ocorrências são raras, podendo porém, verificar-se com razoável probabilidade. Ou seja, há, por exemplo, imensas cidades pequenas mas poucas grandes cidades ou, então, há muitos terremotos de fraca magnitude mas alguns, raros, que assumem magnitudes elevadas. As LP são uma das propriedades marcantes dos sistemas complexos. Um dos modelos que explica o seu aparecimento é a criticalidade auto-organizada. Por exemplo, numa floresta, a rede de árvores forma um estado crítico onde uma pequena faísca pode alastrar de modo a destruir toda a floresta, ou destruir apenas um pequeno número de árvores, ou não ter efeito nenhum. A distribuição dos incêndios segue, assim, uma LP [1].

1.2. OBJECTIVOS

O objectivo principal deste trabalho é estudar fenómenos em sistemas naturais e artificiais que sigam distribuições do tipo LP, bem como o estudo e aquisição de conhecimentos sobre LP. Dada a complexidade inerente a este objectivo, sentiu-se a necessidade de o subdividir em múltiplas tarefas de realização mais simples, tais como:

- Aquisição de conhecimentos de várias LP, tais como a Lei de Lotka, Lei de Yule, Lei de Gutenberg-Richter, entre outras;
- Estudo de vários fenómenos onde se aplicam LP;
- Escolha das LP a apresentar no estudo;
- Desenvolvimento do estudo baseado em dados;
- Relação existente entre fenómenos naturais e artificiais e as LP;
- Elaboração de meta-análise sobre os vários fenómenos.

1.3. CALENDARIZAÇÃO

Sendo o estudo e aquisição de conhecimentos a motivação deste trabalho, a sua prossecução conduziu à calendarização apresentada na Tabela 1. Esta inclui um conjunto de tarefas, como por exemplo: o estudo de bibliografia, em artigos científicos e informação adquiridos na internet; desenvolvimento do estudo; elaboração de análise; elaboração do relatório final. O estudo de bibliografia foi sem dúvida a parte mais importante e demorada, pois o conhecimento pessoal do candidato sobre o assunto era reduzido.

Tabela 1 Calendarização do trabalho

ID	Nome da Tarefa	Início	Conclusão	Duração																														
					Fev 2011				Mar 2011				Abr 2011				Mai 2011				Jun 2011				Jul 2011				Ago 2011					
					6-2	13-2	20-2	27-2	6-3	13-3	20-3	27-3	3-4	10-4	17-4	24-4	1-5	8-5	15-5	22-5	29-5	5-6	12-6	19-6	26-6	3-7	10-7	17-7	24-7	31-7	7-8	14-8		
1	Recolha de Bibliografia	01-02-2011	28-02-2011	4s																														
2	Estudo de Bibliografia	01-03-2011	22-04-2011	7,8s																														
3	Desenvolvimento dos estudos	28-03-2011	31-05-2011	9,4s																														
4	Análise de dados	01-06-2011	08-07-2011	5,8s																														
5	Elaboração do relatório	01-07-2011	15-09-2011	11s																														
6	Entrega relatório	15-09-2011	16-09-2011	,4s																														

1.4. ORGANIZAÇÃO DO RELATÓRIO

No Capítulo 1, faz-se uma introdução ao trabalho efectuado. No capítulo seguinte são apresentados os principais pontos correspondentes à introdução teórica sobre o tema. No terceiro capítulo, é desenvolvido o estudo sobre LP. No capítulo 4 é feita a comparação entre os vários estudos realizados. Por último, no quinto capítulo são reunidas as principais conclusões e perspectivados futuros desenvolvimentos.

2. LEIS DE POTÊNCIA EM SISTEMAS NATURAIS E ARTIFICIAIS

2.1. CRITICALIDADE AUTO-ORGANIZADA

É comum pensar que para se atingirem pontos críticos é necessária intervenção externa. Contudo, às vezes a criticalidade é atingida espontaneamente pela natureza, fenómeno denominado criticalidade auto-organizada. Esta parece surgir quando as partes de um sistema se afastam lentamente do estado de equilíbrio, onde as acções de cada parte individual são dominadas pelas interacções com as demais partes do sistema.

As florestas são um bom exemplo de criticalidade auto-organizada. A rede de árvores forma um estado crítico onde, por exemplo, uma pequena faísca pode alastrar de modo a destruir toda a floresta, ou destruir apenas um pequeno número de árvores, ou não ter efeito algum. O modelo proposto para a propagação dos incêndios pode também ser usado no estudo da propagação de doenças em populações humanas, bastando trocar “árvores”

por “pessoas” e “fogo” por “vírus”. De facto, Rhodes e Anderson explicam desta forma a distribuição de epidemias nas ilhas Faroe do Atlântico Norte [2][3].

2.2. LEI DE POTÊNCIA

Muitos acontecimentos que os cientistas medem têm uma dimensão típica ou de “escala”. Um exemplo simples é a altura dos seres humanos. A maioria dos adultos mede aproximadamente 1,80m. Existe alguma variação em torno deste valor, nomeadamente em função do sexo, mas não existem pessoas com 0,10m ou 5,00m de altura. Todavia nem todas as coisas mensuráveis estão distribuídas em torno de um valor típico. Algumas variam numa magnitude muito grande, como é o caso do comportamento tipo do tamanho das cidades ou da magnitude dos sismos [4].

A LP é uma das assinaturas comuns de um sistema dinâmico não linear, revelando um sistema complexo que está a um passo da auto-organização. Com a LP é possível expressar a auto-similaridade de grande e pequena dimensão, ou seja, unir diferentes tamanhos e comprimentos. A LP comum para todas as dimensões demonstra a auto-consistência interna do fractal e a sua unidade em todos os limites. A distribuição LP resulta da união de todas as leis e processos em todas as escalas [5].

A relação de escala aplica-se largamente em LP e foca características importantes dos sistemas. Em geral, não é evidente se uma perturbação tem um efeito maior ou menor num determinado sistema. O efeito de propagação de qualquer perturbação de um sistema pode variar de zero até infinito, pois existe uma imprevisibilidade inerente. Ao utilizar as LP, deve-se observar que os dados estatísticos para um fenómeno que obedece a uma LP (crescimento exponencial) é voltada para a parte inferior do intervalo, enquanto o fenómeno de saturação (crescimento logarítmico) tende a ser para a parte superior do intervalo [5].

2.3. O QUE É UMA LEI DE POTÊNCIA

Vilfredo Pareto (1987, 1906) proferiu que o princípio da economia tende a distribuir de acordo com uma lei. Alguns chamam-lhe a regra de Pareto, também conhecida por princípio 80/20, segundo o qual 20% da população concentra 80% da riqueza de um país. Em termos gerais, a observação de Pareto é que em todos os países, uma fracção pequena da sociedade, controla uma elevada quantidade de riqueza. Este é o primeiro exemplo de uma LP, que nos últimos anos atraiu a atenção dos investigadores nas diversas áreas. Distribuição de potência (*Power law distribution*), lei de Zipf e lei de Pareto, são usadas para descrever eventos nos quais uma variável aleatória encontra valores altos com pouca frequência, enquanto os valores médios ou baixos aparecem com muita frequência. Por exemplo, existem poucas cidades grandes e muitas povoações pequenas, em relação ao número de habitantes. Existem poucas palavras que se utilizam com muita frequência num texto, enquanto outras se usam poucas vezes. Na Lua, a frequência de crateras pequenas é elevada, e as crateras grandes são menos frequentes. Usando a escala de Richter, são poucos os grandes terremotos, e muito frequentes os de baixa intensidade. Existem poucos géneros biológicos com muitas espécies, e muitos géneros com poucas espécies. Sucessivamente, podem encontrar-se fenómenos naturais e artificiais, nos quais a variável de estudo segue uma distribuição de potência [6].

Quando um fenómeno segue uma LP, ver-se-á uma distribuição semelhante à da Figura 1.

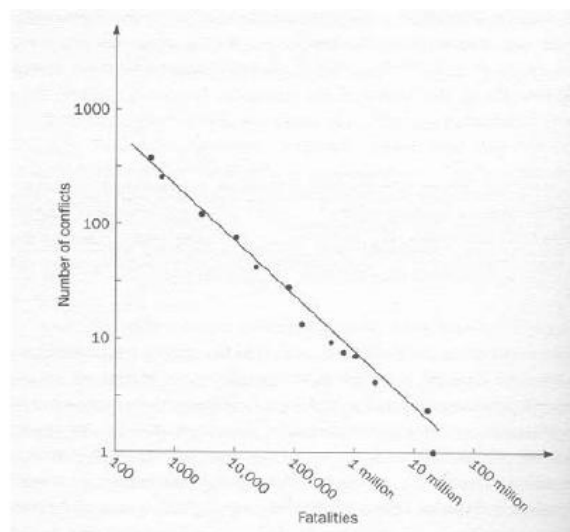


Figura 1 Relação entre o número de vítimas e o número de conflitos [3]

Este gráfico relaciona a frequência das guerras com o número de mortos em cada uma delas. Surpreende pelo perfeito ajuste a uma recta dos valores observados. Como se pode ver, ambos os eixos usam escalas logarítmicas [7].

2.4. FUNÇÃO LEI DE POTÊNCIA

Uma LP é uma função de “escala”, onde:

$$f(x) = cx^\alpha \quad c > 0 \text{ e } \alpha \in R \quad (1)$$

O parâmetro α é o índice ou “grandeza” da LP.

Aplicando logaritmos em ambos os membros da equação:

$$\log f(x) = \log c + \alpha \log x \quad c > 0 \text{ e } \alpha \in R \quad (2)$$

Numa LP inversa, o índice α é negativo, pelo que aparecerá uma linha recta com declive negativo num gráfico log-log.

2.5. FUNÇÃO DE PROBABILIDADE

Uma variável aleatória X segue uma LP se a sua função de distribuição de probabilidade acumulada (complementar) for da seguinte forma:

$$F(x) = P(X \geq x) = \frac{c}{\alpha-1} x^{-(\alpha-1)} \quad (3)$$

Com $\alpha > 0$ e $c > 0$.

A cauda declina assintoticamente com α , pelo que é comum chamar-lhe distribuição de cauda longa.

A distribuição de probabilidade simples é dada por:

$$P(X = x) = cx^{-\alpha} \quad (4)$$

2.6. LEI DE ZIPF

A designação Lei de Zipf deve-se ao trabalho de George Kingsley Zipf (1902-1950). É uma das leis de escala e estabelece que a frequência de ocorrência de determinado evento, P , em função da sua classificação, i , é um caso da LP. A lei de Zipf foi inicialmente aplicada ao estudo da frequência de ocorrência de palavras em texto [5]. Na lei de Zipf $\alpha = 2$.

Benoit Mandelbrot demonstrou que uma LP mais generalizada é obtida pela adição de constantes à variável e ao expoente (c e d). A Lei de Zipf é, assim, o caso especial em que as duas constantes são iguais a zero:

$$P(X = x) = (c + x)^{-(\alpha+d)} \quad (5)$$

Existem evidências de que a distribuição da população e outras actividades sócio-económicas, em diferentes escalas, são distribuídas de acordo com a lei de Zipf, e que tais distribuições de escala estão associados a sistemas em crescimento [5].

2.7. LEI DE PARETO

A Lei de Pareto (ou princípio 80/20) foi criada pelo economista italiano Vilfredo Pareto. Esta lei diz que 80% das consequências ocorrem de 20% das causas. Para se entender melhor este princípio, exemplifica-se nos seguintes pontos [8]:

- 80% das peras → produzidas por 20% das árvores;
- 80% das terras → cultivadas por 20% da população;

- 80% dos lucros → gerados por 20% dos funcionários;
- 80% dos problemas → motivados por 20% dos clientes;
- 80% das decisões → originadas por 20% de uma reunião;
- 80% dos crimes → causados por 20% dos criminosos.

No século XIX, Pareto comprovou que a maioria da receita e das riquezas ia para uma minoria de pessoas. Havia, portanto, uma forte relação matemática entre a proporção de pessoas e a renda recebida por este grupo. Se 20% recebiam 80% da riqueza, podia prever-se que 10% teriam 65% da riqueza e 5% ficariam com 50% [8] [9].

Mais tarde, em 1949, o professor de filologia (linguística) de Harvard George K. Zipf descobriu o popular princípio do menor esforço, segundo o qual as pessoas tendiam a minimizar seus trabalhos de modo que 20% ou 30% de quaisquer recursos corresponderiam a 70% ou 80% do resultado. Zipf usou estatísticas populacionais, livros de filologia (linguística) e comportamentos industriais para mostrar a recorrência constante desse padrão de desequilíbrio. Na verdade, Zipf elaborou o princípio descoberto por Pareto [9].

2.8. LEI DE BENFORD

A Lei de Benford, também conhecida como a lei do primeiro dígito ou fenómeno do primeiro dígito, afirma que a distribuição do primeiro dígito dos números não é uniforme. Esta lei aplica-se se os números sob pesquisa não forem aleatórios, mas forem de alguma forma socialmente ou naturalmente relacionados. Geralmente aplica-se a dados não dimensionais, o que significa que a distribuição de probabilidade dos dados não varia com uma mudança de escala ou dados [5].

Na lei de Benford, os números que aparecem em Tabelas de estatística, listagens e outros, têm como primeiro dígito mais provável o um e menos provável o nove.

Muitas situações fizeram com que existam cortes em séries de números. Por exemplo, os endereços de rua começam a partir do um e acabam em qualquer valor. Estas condicionantes levam à lei de Benford [5]:

$$P(X = x) = \log \left(1 + \frac{1}{x} \right) \quad (6)$$

A base do logaritmo também pode ser diferente de 10. De acordo com esta equação as probabilidades de dígitos 1, ..., 9 encontram-se entre 0.301, ..., 0.045, conforme mostrado na Figura 2. A probabilidade de ocorrência do dígito um é seis vezes superior à do dígito nove [5].

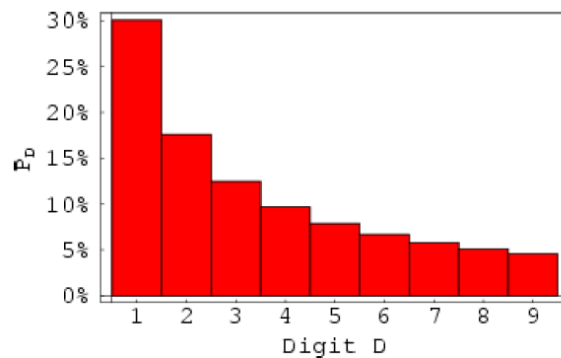


Figura 2 Distribuição da lei de Benford para os dígitos de 1 a 9 [5]

Os exemplos da lei de Benford são numerosos, incluindo os endereços; a área; a bacia dos rios; a população; a pressão; o peso molecular e atômico; a taxa de mortalidade; etc.

Esta lei é uma ferramenta poderosa e relativamente simples para aplicação em detecção de fraudes. As agências de impostos de vários países usam *software* de detecção baseado na lei de Benford [5].

2.9. APLICAÇÕES DE LEI DE POTÊNCIA

O mundo natural (sistemas naturais) é rico em distribuições de LP. Por exemplo, as LP definem a distribuição de eventos catastróficos em sistemas críticos auto-organizados.

A Lei de Gutenberg-Richter, usada em sismografia, é uma observação empírica de que os terremotos ocorrem com uma distribuição de LP. Todos os dias em várias partes do mundo ocorrem pequenos terremotos que são demasiado fracos para se detectarem sem a ajuda de instrumentos. Terremotos um pouco maiores, que chocalham pratos, têm uma frequência de ocorrência menor. Muito menos comum, são os que causam destruição em massa [5].

As LP são aplicadas para monitorizar emissões acústicas em materiais utilizados para pontes. Defeitos internos destes materiais fazem um *popping sound*, emissões acústicas sob *stress*. Estes materiais, contêm pequenas rachadelas que crescem com este *stress test*, até crescerem o suficiente para causar a falha do material. Isto pode causar a queda de edifícios, pontes ou outras estruturas. Este método de monitorização é uma forma simples e de custo reduzido de ensaios não destrutivos. Com a ajuda deste método também é possível desenvolver novas técnicas para melhoramentos dos materiais e estruturas menos vulneráveis a fissuras [5].

Alguns exemplos de LP em biologia e sociologia são as leis de Kleiber, que relacionam a taxa metabólica e a massa corporal em animais. A Lei Taylor da energia das flutuações da população e a lei de Yoda, que relaciona a densidade talos/hastes e o tamanho das plantas, são outros exemplos [5].

Os estudos bibliométricos eram inicialmente voltados para a medida de quantidade de edições e exemplares de livros. No entanto, foram-se expandindo para outros formatos de produção, como artigos de jornais, de autores e do estudo de citações. Três leis básicas regem os estudos bibliométricos, sendo elas: a Lei de Lotka, a Lei de Bradford e a Lei de Zipf. A Lei de Lotka foi desenvolvida a partir da descoberta de que uma larga porção da literatura científica é produzida por um reduzido número de autores. Já a lei de Bradford incide sobre os jornais. Bradford tinha o intuito de investigar a extensão na qual artigos de um assunto científico específico apareciam em jornais destinados a outros assuntos. Para isso analisava a distribuição dos estudos em termos de variáveis de proximidade ou de

afastamento. Desta análise Bradford formulou a lei, que chamou de lei da dispersão. Tal lei é enunciada da seguinte forma: dispondo jornais em ordem decrescente relativamente à publicação de artigos sobre um determinado tema, pode-se distinguir um “centro”, de jornais mais dedicados ao tema, e vários grupos ou zonas que incluem o mesmo número de artigos que o centro. A outra lei bibliométrica clássica é a Lei de Zipf, que descreve a relação entre palavras num determinado texto, suficientemente grande, e a ordem de série destas palavras. Segundo Zipf, listando as palavras que ocorrem num texto em ordem decrescente de frequência, a posição de uma palavra na lista multiplicada pela sua frequência é igual a uma constante. A partir daí Zipf formulou o princípio do menor esforço, segundo o qual existe uma economia na utilização de palavras. Ou seja, se a intenção é usar o mínimo, então uma mesma palavra vai ser adoptada várias vezes. Acrescenta-se ainda que a área mais importante da bibliometria é a análise de citações, que pode ser definida como “a parte da bibliometria que investiga as relações entre os documentos e os documentos citados, considerados como unidades de análise, no todo ou em diversas partes: autor, título, origem geográfica, ano e idioma de publicação, etc.” [10].

Em 1997, um *webmaster* de um site de recursos humanos estava curioso sobre se o número de visitas ao *site* por mês seguia o modelo da lei de Zipf. Uma análise rápida dizia que sim. Animado com a descoberta, queria saber se existia alguém com a mesma ideia, antes que comesse a escrever a sua publicação. A sua pesquisa acabou no departamento de ciências de computação da Universidade de Boston, onde o mesmo modelo de Zipf já tinha sido descoberto. Nos últimos anos, o estudo de comportamento de escala, no tráfego da internet já se tinha tornado num dos tópicos mais populares na ciência aplicada à computação [11].

Outra aplicação é o tamanho das empresas e os seus dados económicos. Uma empresa pode ser classificada pelo número de funcionários, receitas, lucros, valor de mercado, assim como muitas outras características. A classificação também pode ser feita dentro de determinada indústria ou certas localizações geográficas. A distribuição resultante é muitas vezes comparada com a lei de Pareto e indistinguível da lei de Zipf. A única questão é saber se o valor de α é igual a dois ou não. Um debate constante sobre os dados económicos é saber se estes são distribuídos como LP ou como logaritmo normal, ou talvez outra distribuição [11].

Estas são algumas aplicações das leis de potência, em sistemas naturais e artificiais. No Capítulo 3, estão mostrados alguns estudos efectuados.

2.10. O FUTURO DA PESQUISA SOBRE LEIS DE POTÊNCIA

“Lei de Potência... tem sido denominada de - A assinatura da atividade Humana” [12]

O estudo das leis de potência explodiu claramente na última década, com inúmeros artigos e livros muito populares. As LP estão de facto em toda a parte e, apesar deste notável sucesso, acredita-se que a investigação deste tema sofre de algumas lacunas. Lidar com esses problemas deverá levar a outra expansão de pesquisas. Estão a surgir inúmeros modelos de comportamentos de LP, ainda sem mecanismos ou abordagens que permitam a sua validação. Existem modelos e teorias, mais hesitantes e distantes das etapas seguintes [12].

Para desenvolver o raciocínio é importante uma classificação geral dos tipos de resultados que se pretendem alcançar [12]:

- Observação – Recolher dados sobre um comportamento de um sistema e demonstrar que seguem uma distribuição LP;
- Interpretação – Explicar o significado do comportamento LP do sistema;
- Modelo – Propor um modelo subjacente que explique o comportamento LP;
- Validação – Encontrar dados para validar e se necessário modificar o modelo;
- Controlo – Usar a compreensão do modelo para controlar, modificar e melhorar o comportamento do sistema.

O argumento é de que a maioria das pesquisas sobre LP tem-se centrado na observação, interpretação e obtenção de modelos. Quase nada foi feito sobre validação e controlo e, é nestes pontos, que se deve avançar nas pesquisas. Uma quantidade significativa de comportamentos e modelos tem levado a um histórico de várias pesquisas, como é o caso

de *Newman* [4]. A validação é importante, pois existem inúmeros modelos que produzem uma LP e, em muitos casos, mais que um pode ser usado para explicar um comportamento observado. Por exemplo, no caso de grandes mudanças, na distribuição de páginas *web* ainda haverá as mesmas características daqui a dois anos? Neste momento não se sabe o suficiente para responder à pergunta. Além disso, do ponto de vista prático, as pessoas gostariam de aplicar o modelo e usá-lo para controlar o sistema. Supondo que o *Google*, por exemplo, queria estimular as pessoas a criarem *links* de maior interesse e queria determinar se o poderia fazer com baixo custo associado. Isso exigiria não só validar o modelo subjacente, mas também a compreensão de como o modelo iria responder a mecanismos de controlo, como a publicidade, concursos ou pagamentos directos, que podem potenciar os *links* [12].

Numa nova direcção é possível dizer que as pessoas que fazem pesquisas sobre LP estão no caminho da validação. Um passo importante, talvez entre a validação e o modelo, é o trabalho não validado. A não validação pode tomar várias formas, inclusive mostrando que o modelo é insuficiente para explicar uma LP, fornecendo uma alternativa razoável ou mais convincente. Ou considerando propriedades adicionais além do comportamento da LP sugerido por um modelo, mostrando que essas propriedades não se encaixam nos dados. Existe ainda a obrigação de encontrar formas de controlar o comportamento do sistema. Em muitas áreas das ciências, simplesmente observando uma LP, obtém-se um resultado significativo. Para exemplificar, recorre-se novamente à amplitude dos terremotos, que segue uma distribuição de LP sem intervenção humana. Nos sistemas computacionais pode esperar-se, em muitos casos, a mudança do comportamento do sistema, pois é possível afectá-lo directamente, dando origem a um comportamento diferente. No mundo é possível controlar em certa medida as leis da natureza, só é necessário saber utilizar esse poder de forma eficaz [12].

A verdade é que neste momento a compreensão real dos processos que geram as leis de potência é tão limitada que não se sabe quais os mecanismos de controlo adequados. Os fundamentos nesta área ainda estão em aberto. Além disso, ao obter uma melhor compreensão dos mecanismos, é possível encontrar abordagens que actuam noutros domínios: redes sociais, economia, sistemas biológicos. Desta forma as descobertas podem ter um impacto poderoso [12].

Resumindo, estima-se que aconteceu a passagem da observação, interpretação e modelação para a validação e controlo das pesquisas em LP. A última década tem sido muito rica em pesquisas, com muitas descobertas notáveis na teoria e na prática. Acredita-se que uma mudança de mentalidade para a validação e controlo incentivará a continuidade nesta área de pesquisa [12].

3. ESTUDOS DE LEI DE POTÊNCIA

3.1. CONSIDERAÇÕES GERAIS

Ao longo deste capítulo são apresentados vários casos onde as LP se aplicam. Estes estudos foram realizados com base em dados disponíveis em *sites* e relatórios acessíveis por Internet.

Para a construção dos gráficos log-log, os dados são agrupados em intervalos, de acordo com o seu valor, sendo determinada a sua frequência. O procedimento é idêntico ao utilizado na construção de um histograma de frequências simples. Num segundo passo efectua-se uma normalização, em que os dados são divididos pelo maior valor observado e a frequência é dividida pelo maior valor de frequência registado.

Na Figura 3, é exibido um diagrama simples que representa o procedimento utilizado.

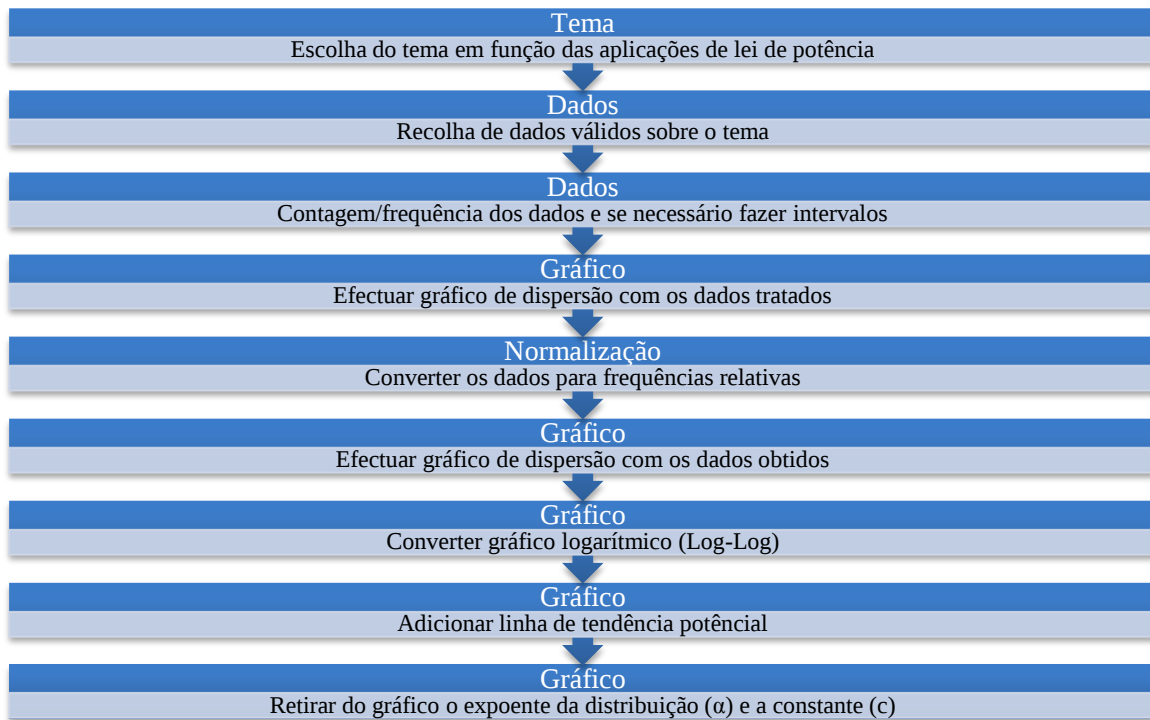


Figura 3 Procedimento de tratamento de dados

3.2. POPULAÇÃO PORTUGUESA

O objectivo deste estudo, sobre a população portuguesa, consiste em verificar se o número de pessoas por cidade segue uma distribuição do tipo LP. Para isso foram pesquisados no *site* do Instituto Nacional de Estatística (INE) dados sobre a população residente em Portugal continental no ano de 2009 (estimativas anuais de população residente) [13].

Com os dados extraídos de 280 cidades, verificou-se que a população variava entre 1.670 habitantes, para a cidade com menos população, e 479.884, para a cidade com maior população (sendo neste caso Lisboa).

Observou-se, assim, que existem muitas cidades com pouca população e poucas cidades com muita população (Figura 4). O gráfico escolhido foi do tipo “dispersão” e com valores absolutos. Pode verificar-se uma característica importante, que é a curvatura do gráfico.

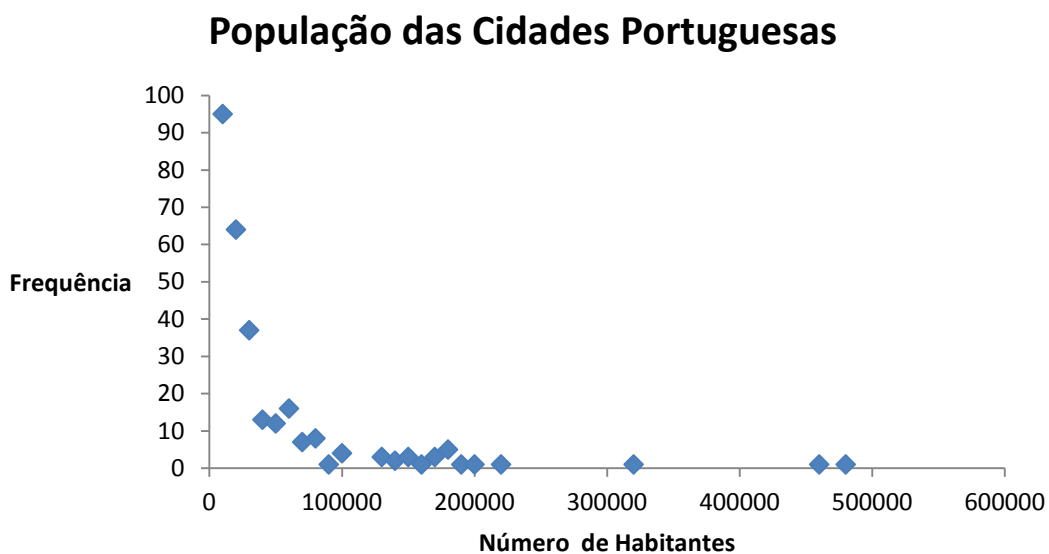


Figura 4 População das cidades portuguesas

Posteriormente os dados foram convertidos para valores relativos e feito um novo gráfico com escalas logarítmicas. Na Figura 5 é apresentado o gráfico Log-Log correspondente, onde foi ajustada uma linha de tendência, pelo método dos mínimos quadrados.

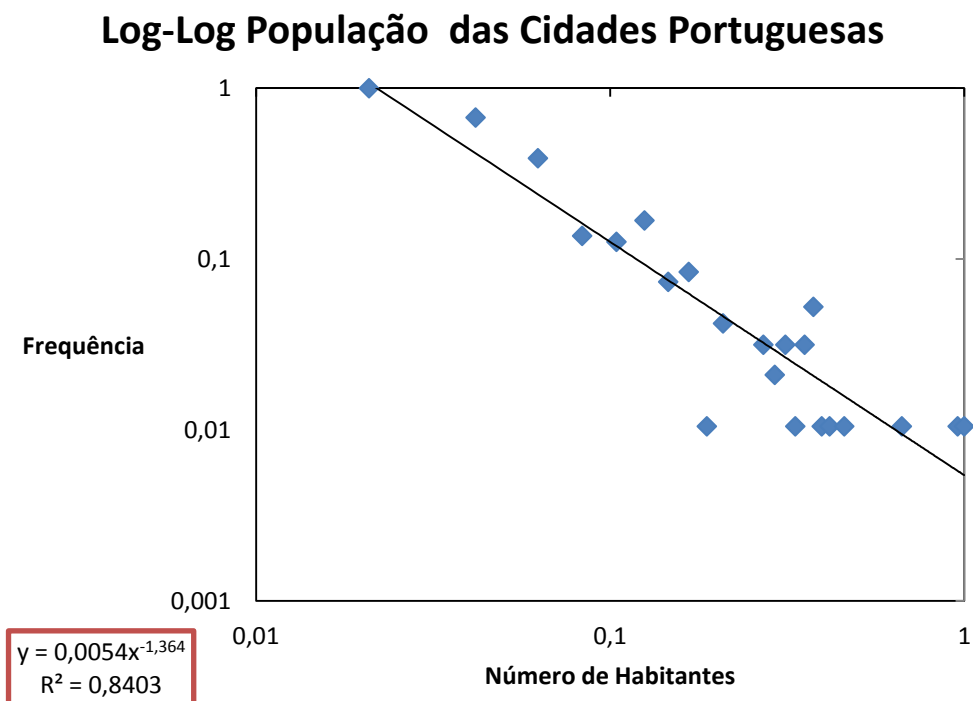


Figura 5 Gráfico Log-Log da População das Cidades Portuguesas

Neste estudo α tem o valor de 1,364, onde α é o módulo do expoente da distribuição LP.

3.3. OS MAIS RICOS DO MUNDO DA FORBES 2011

Anualmente a revista Forbes publica a lista dos mais ricos do mundo. Esta lista, também acessível no *site* da Forbes, inclui o ranking e valor do património [14]. A Tabela 2 mostra os três primeiros do ranking de 2011.

Tabela 2 Lista dos mais ricos do mundo [14]

Rank ▲		Name	Net Worth	Age	Source	Country of Citizenship
1		Carlos Slim Helu & family	\$74 B	71	telecom	Mexico
2		Bill Gates	\$56 B	55	Microsoft	United States
3		Warren Buffett	\$50 B	80	Berkshire Hathaway	United States

A listagem inclui diversos sectores de actividade e da economia, desde a informática/software de Bill Gates, passando pela cortiça, com Américo Amorim (no lugar 200, sendo o primeiro português da lista), entre outros sectores. Embora estas pessoas tenham feito as suas fortunas de diversas maneiras, a distribuição estatística da sua riqueza apresenta uma regularidade impressionante.

Os dados foram retirados dos 650 primeiros nomes da lista, tendo sido posteriormente tratados para elaboração dos gráficos. Como foi explicado anteriormente, o processo começa por se efectuar uma Tabela com valores absolutos, mostrando a frequência dos grupos com um determinado valor de riqueza. Na Figura 6 apresenta-se o resultado da primeira análise dos dados.

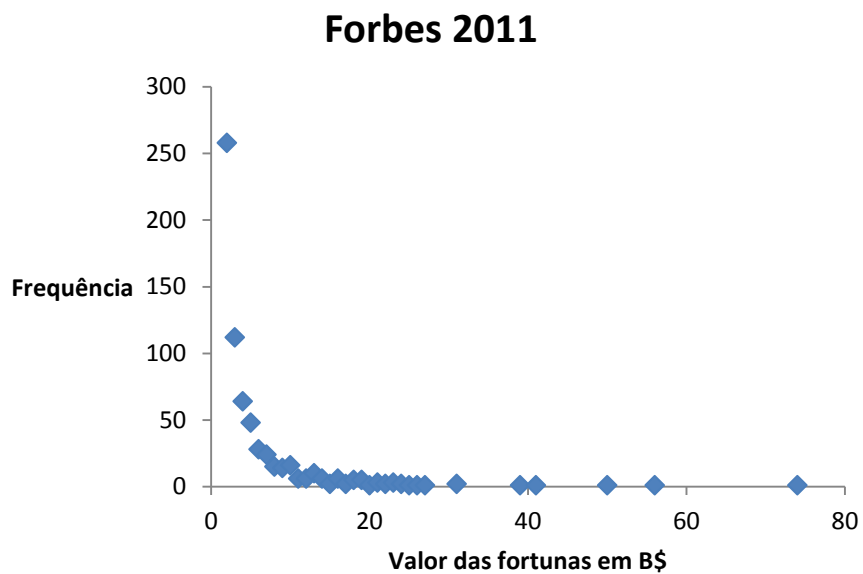


Figura 6 Gráfico Forbes 2011 dos mais ricos do mundo

Passa-se então para a normalização dos valores, convertendo-os em valores relativos. O gráfico que advém desta normalização é convertido para um gráfico Log-Log. Na Figura 7 é apresentado o gráfico correspondente.

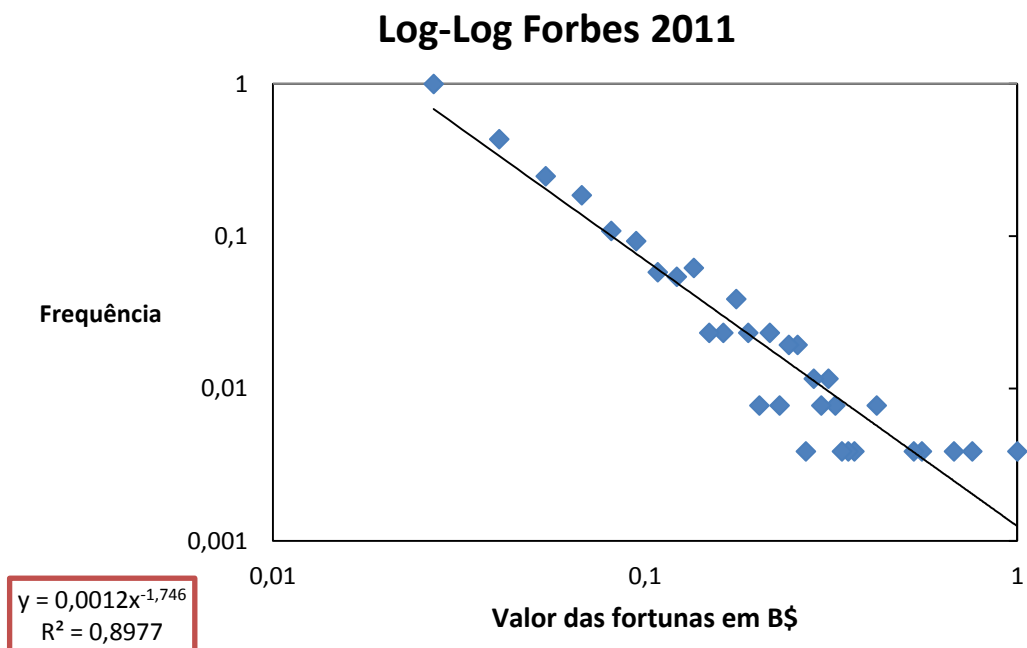


Figura 7 Gráfico Log-Log Forbes 2011

Neste gráfico normalizado observa-se também que a linha de tendência é uma recta com um declive negativo. Neste estudo α tem o valor de 1,746.

Mais uma vez é possível dizer que existem poucos Bilionários com muito dinheiro e muitos Bilionários com “pouco dinheiro”. Embora uma afirmação destas seja muito polémica, pois um Bilionário é sempre uma pessoa com muito dinheiro.

3.4. MAGNITUDE DOS SISMOS

A amplitude dos sismos é outro fenómeno que suporta uma LP, pois existem muitos sismos com uma amplitude baixa e poucos com uma amplitude elevada. Torna-se praticamente impossível prever quando ocorrerá um sismo e o nível de destruição que irá provocar. Os sismos distribuem-se em função da energia libertada, de acordo com a lei de Gutenberg-Richter. A escala de Richter atribui um número único para quantificar o nível de energia libertada por um sismo [3]. É uma escala logarítmica de base 10, obtida calculando o logaritmo da amplitude sísmica. Os grandes sismos são popularmente designados pelo termo terramoto.

Como curiosidade é apresentado, na Figura 8, um gráfico que representa a amplitude e o número de sismos, desde 1990 até 2010, no mundo.

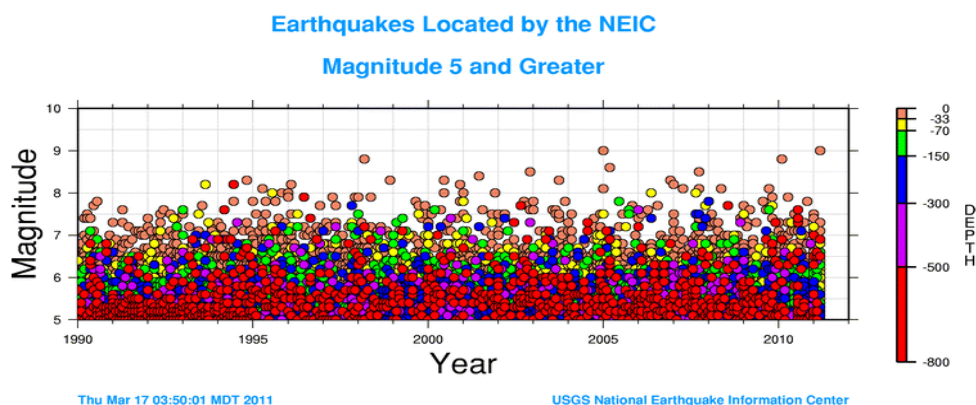


Figura 8 Gráfico ilustrativo da ocorrência de sismos [15]

Na Tabela 3, é apresentada a frequência de sismos por intervalo (na escala de Richter), baseada em observações desde o ano 1900 [15]. A Tabela começa a partir de 2, pois a actividade sísmica com valor inferior na escala de Richter por norma não é reportada.

Tabela 3 Frequência / Magnitude de sismos [15]

Intervalo de Amplitude	Média Anual
8 e superior	1
7 a 7,9	15
6 a 6,9	134
5 a 5,9	1319
4 a 4,9	13.000
3 a 3,9	130.000
2 a 2,9	1.300.000

O processo neste estudo é semelhante aos anteriores, no entanto um pormenor desta Tabela é o facto de a escala já estar em unidades logarítmicas. Efectuando a normalização dos valores, é possível construir o gráfico da Figura 9.

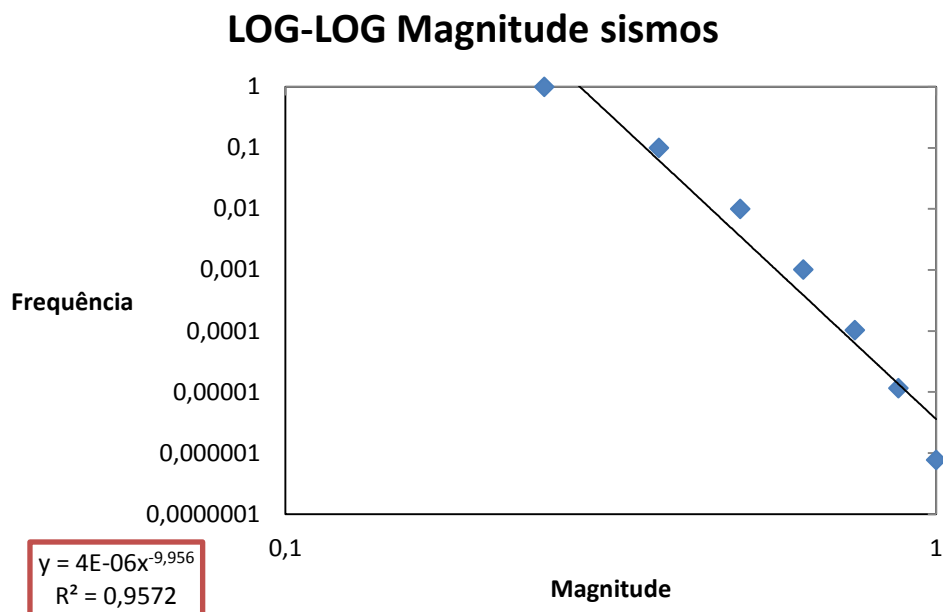


Figura 9 Gráfico Log-Log Magnitude dos Sismos

Neste gráfico observa-se mais uma vez o declive negativo da recta de ajuste resultante. Neste estudo, α tem o valor de 9,956.

Esta pesquisa não servirá para a análise final do trabalho (capítulo seguinte), pois existe dúvida relativamente ao expoente da distribuição LP obtida. Como os dados não foram suficientes para o estudo coloca-se alguma incerteza na veracidade deste caso. Esta poderá ser resultante dos dados adquiridos, ou seja, por um lado o diminuto número de pontos e, por outro, os dados terem começado a partir de dois na escala de Richter.

3.5. ÁREA ARDIDA EM INCÊNDIOS

Durante o ano de 2009, em Portugal contabilizaram-se 26.136 ocorrências, das quais 22% correspondem a incêndios florestais e 78% a fogachos. Estas ocorrências corresponderam a 86.674 ha de área ardida. Mais de metade da área ardida em 2009 resultou de incêndios de grande dimensão (≥ 100 ha), (Tabela 4) [16].

Tabela 4 Resumo do número de ocorrências e correspondente área ardida em 2009 [16]

Ocorrências	< 1 ha (fogachos)	20.274 (78%)
	≥ 1 ha (incêndios)	5.862 (22%)
	≥ 100 ha (grandes incêndios)	133 (0,5%)
	TOTAL	26.136
Área Ardida	Povoamentos (ha)	24.107 (28%)
	Matos (ha)	62.567 (72%)
	TOTAL	86.674

O estudo que é apresentado diz respeito à frequência de fogos com área ardida superior a 100 ha (designados de grandes incêndios) e inferiores a 7.000 ha. Como pode observar-se pela Figura 10, para o ano de 2009, o valor de α é de 1,148.

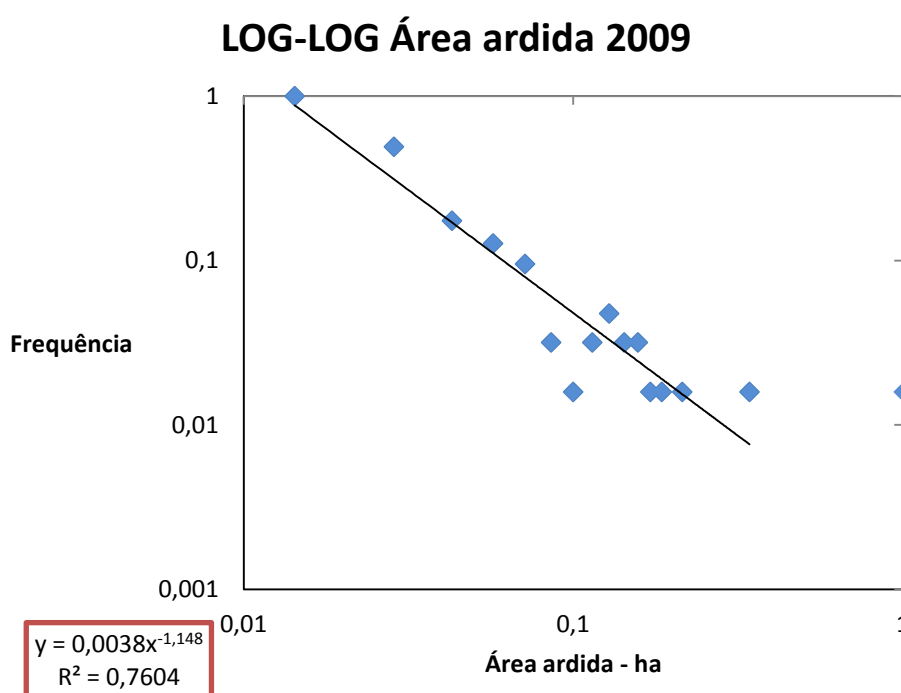


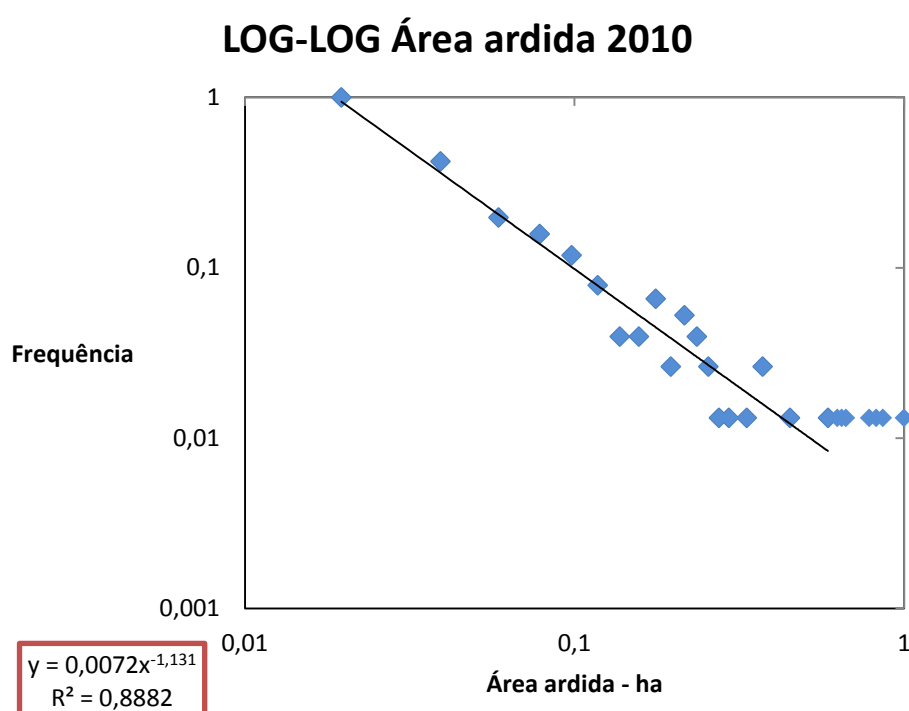
Figura 10 Gráfico Log-Log área ardida 2009

Durante o ano de 2010 contabilizaram-se, em Portugal Continental, 22.026 ocorrências, das quais 18% correspondem a incêndios florestais e 82% a fogachos. Mais de metade da área ardida em 2010 (101.809ha - 76% do total) resultou de incêndios de grande dimensão (≥ 100 ha), (Tabela 5) [17].

Tabela 5 Resumo do número de ocorrências e correspondente área ardida em 2010 [17]

PORTUGAL CONTINENTAL	Ocorrências	< 1ha (fogachos)	18.056 (82%)
		≥ 1ha (incêndios)	3.970 (18%)
		≥ 100ha (grandes incêndios)	186 (1%)
		TOTAL	22.026
	Área Ardida	Povoamentos (ha)	46.079 (35%)
		Matos (ha)	87.011 (65%)
		TOTAL	133.090

À semelhança do que foi feito para o ano de 2009, apresenta-se o gráfico correspondente aos grandes incêndios (superiores a 100 ha) e inferiores a 5.100 ha para o ano de 2010. A Figura 11 exibe o gráfico Log-Log da frequência da área ardida em 2010.



O valor α para o ano de 2010 é de 1,131.

Importante será referir que se obtêm dois valores de α semelhantes que posteriormente, no capítulo 4, servirão para a análise comparativa do estudo. É possível constatar que existem muitos incêndios de pequena área ardida e poucos incêndios de elevada área ardida.

3.6. FREQUÊNCIA DE PALAVRAS EM TEXTOS

Pretende-se com este estudo saber se a frequência de ocorrência de palavras, em determinados textos literários segue uma LP. Para isso, foram escolhidos vários textos enviados pelo Instituto Superior de Engenharia do Porto (ISEP) via e-mail. São histórias oferecidas à sexta-feira, do “1dois3 era outra vez”, enviadas todas as semanas. Faz parte deste projecto o envio de pequenas histórias para crianças e adolescentes, por vezes acompanhados de textos dirigidos a um público adulto, convidando à reflexão e também ao sonho.

Foi utilizado um software de contagem de palavras, com o nome *Letter Frequencies* 1.0.1 [18]. Neste *software* é necessária a conversão dos documentos para um formato TXT. A interface do software é apresentada nas Figuras 12 e 13.

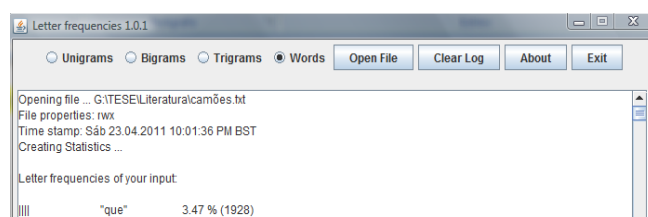


Figura 12 Cabeçalho do *software Letter Frequencies*

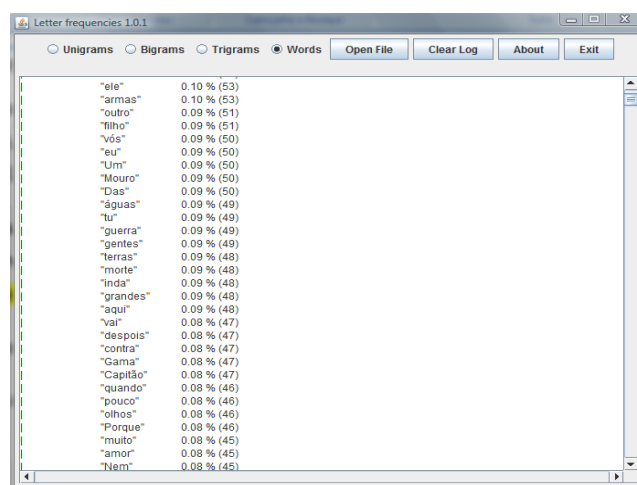


Figura 13 Resultado de contagem de palavras

O procedimento é o mesmo referido nos pontos anteriores.

Na Figura 14 apresenta-se o resultado da análise do texto “A prenda do cisne” [19], com um total de 913 palavras.

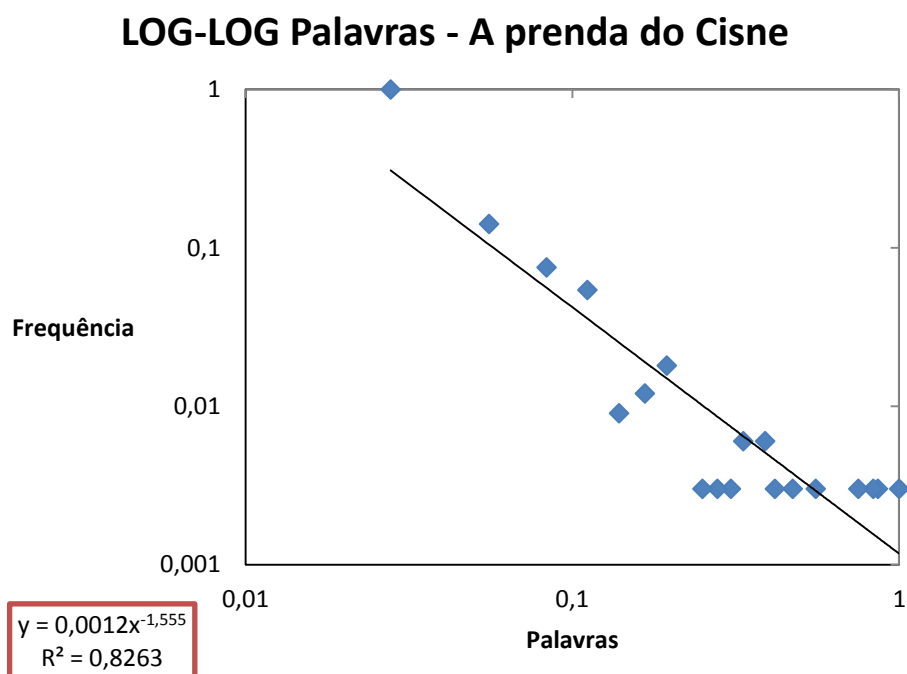


Figura 14 Gráfico Log-Log “A prenda do Cisne”

O título “A senhora dos livros” [20], com o gráfico Log-Log da Figura 15, contou com um total de 1.250 palavras.

LOG-LOG Palavras - A Senhora dos Livros

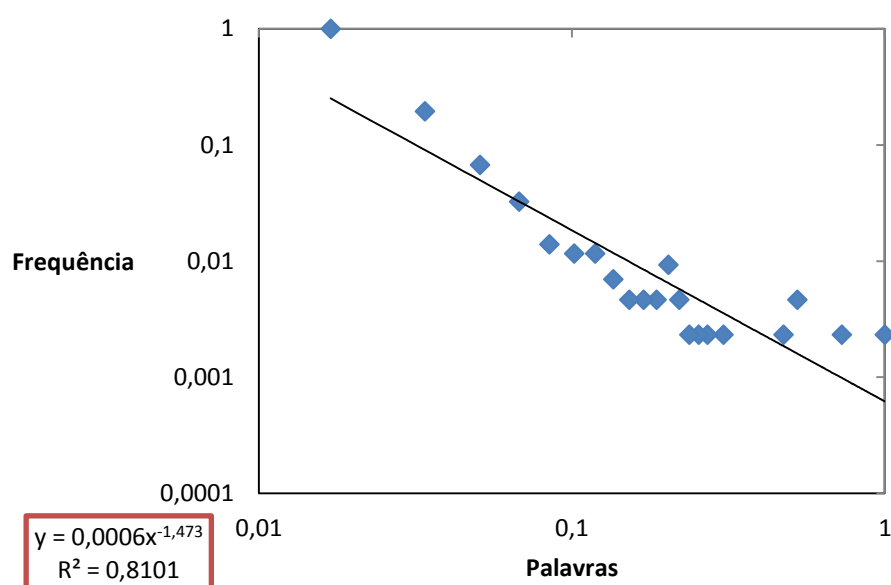


Figura 15 Gráfico Log-Log “A Senhora dos livros”

Na Figura 16, com um total de 1.188 palavras, foi estudado o texto “As sapatilhas de Sofia” [21].

LOG-LOG Palavras - As sapatilhas de Sofia

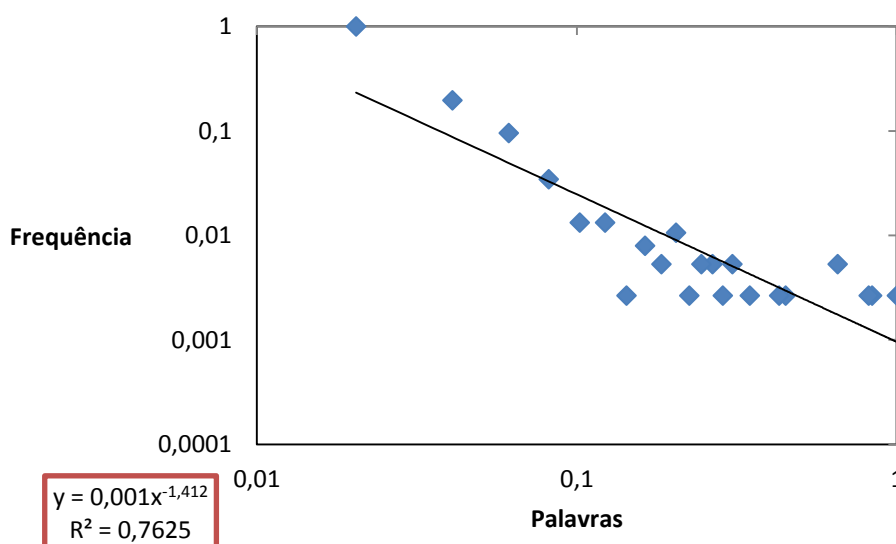


Figura 16 Gráfico Log-Log “As sapatilhas de Sofia”

Com um total de 1.532 palavras, é exposta a Figura 17, referente ao título “O menino Golfinho” [22].

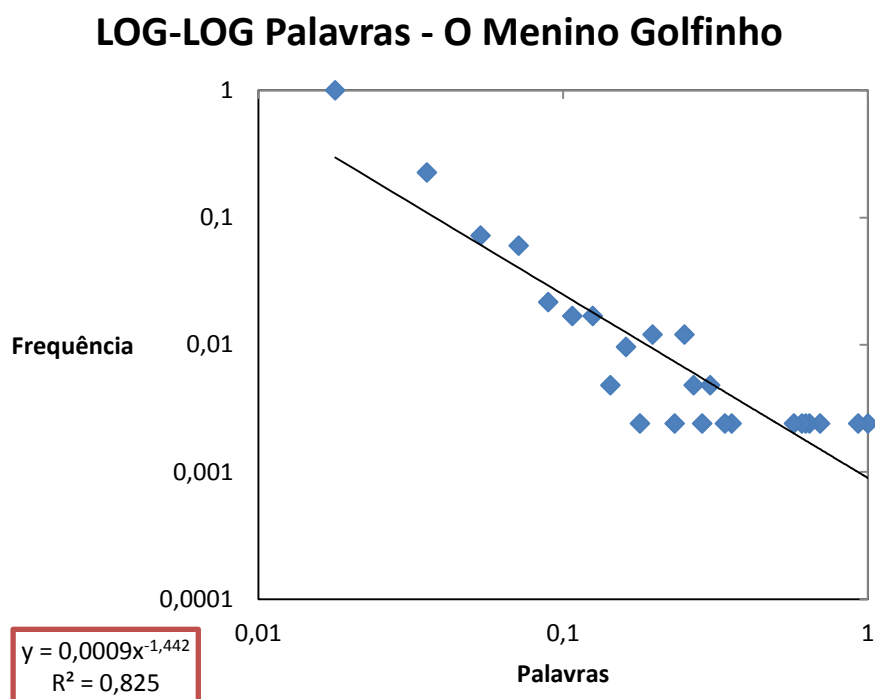


Figura 17 Gráfico Log-Log “O menino Golfinho”

O texto “O pastor amoroso” [23], apresentado no gráfico Log-Log da Figura 18, tem um total de 1.064 palavras.

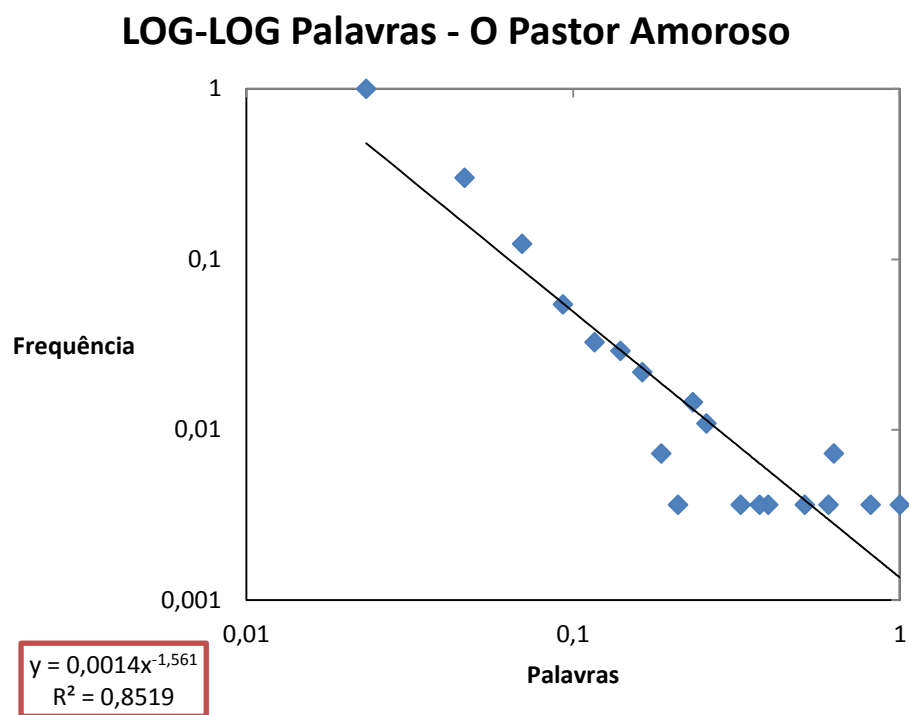


Figura 18 Gráfico Log-Log “O pastor amoroso”

Nas Figuras 19 e 20, são apresentados os gráficos correspondentes aos títulos “Os mil pássaros de Sadako I” [24] e “Os mil pássaros de Sadako II” [25], respectivamente. O primeiro com um total de 3.094 palavras e o segundo com 2.733 palavras.

Log- Log Palavras - Os mil pássaros de Sadako 1

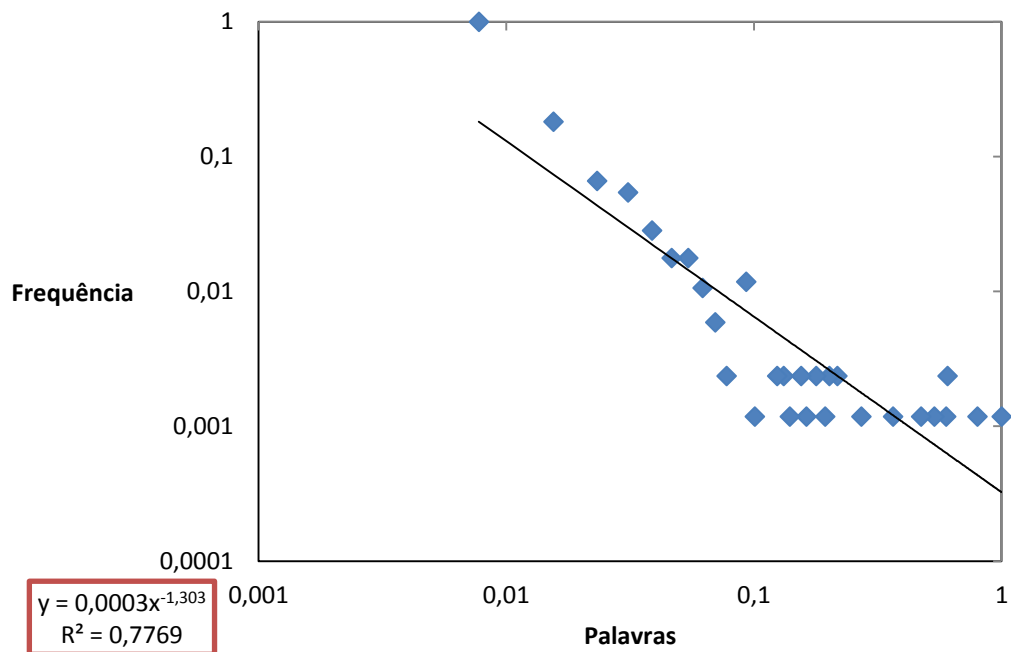


Figura 19 Gráfico Log-Log “Os mil pássaros de Sadako I”

LOG-LOG Palavras - Os mil pássaros de Sadako 2

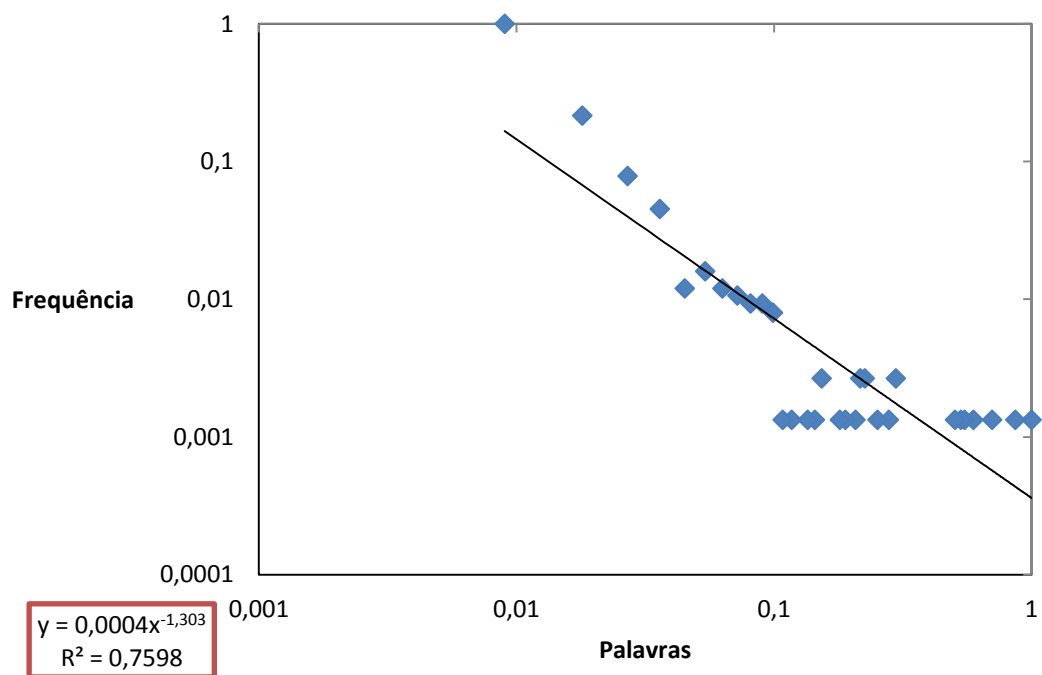


Figura 20 Gráfico Log-Log “Os mil pássaros de Sadako II”

Totalizando 1.848 palavras, o texto “O pequeno comerciante que vendia sorrisos” [26] é apresentado no gráfico da Figura 21.

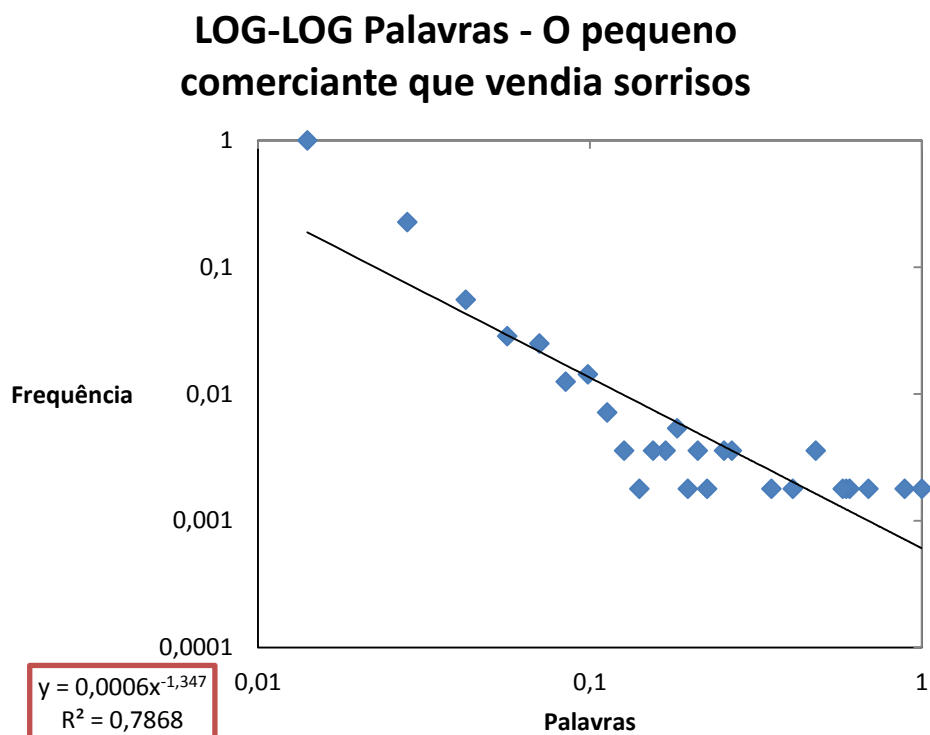


Figura 21 Gráfico Log-Log “O pequeno comerciante que vendia sorrisos”

Por último, na Figura 22, apresenta-se o resultado da análise do texto o “Pão de Tâmaras e nozes”, com 1.255 palavras [27].

LOG-LOG Palavras - Pão de Tâmaras e nozes

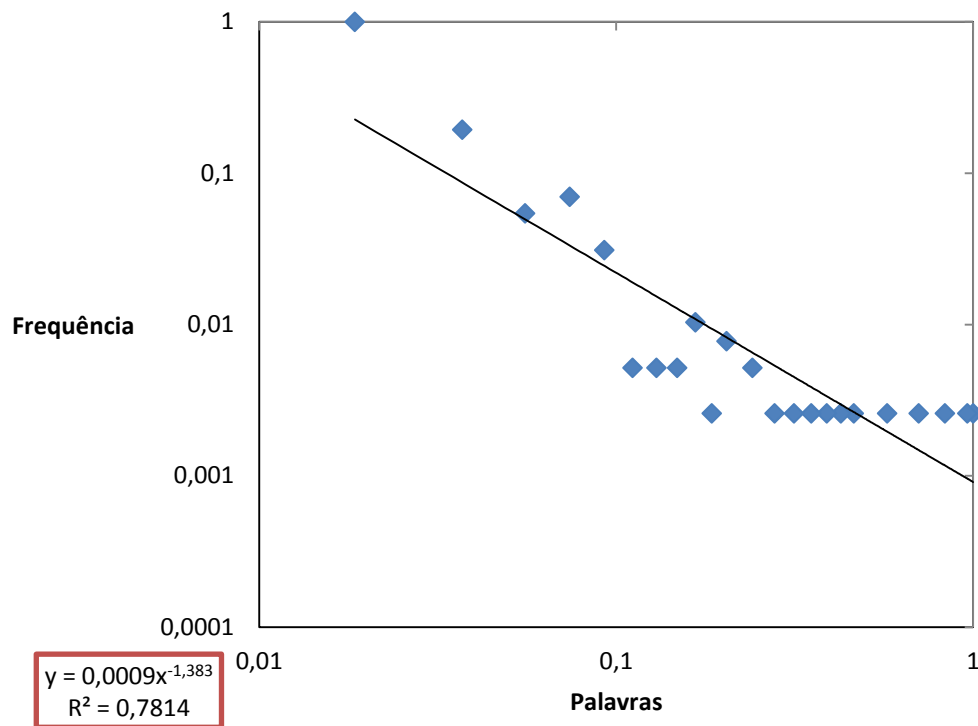


Figura 22 Gráfico Log-Log “Pão de Tâmaras e nozes”

Para finalizar este estudo, são compilados na Tabela 6 os vários valores de α .

Tabela 6 Valores de α

Palavras	c	α	Literatura
913	0,0012	1,555	A prenda do Cisne
1.250	0,0006	1,473	A senhora dos Livros
1.188	0,0010	1,412	As Sapatilhas de Sofia
1.532	0,0009	1,442	O menino Golfinho
1.064	0,0014	1,561	O Pastor amoroso
1.848	0,0006	1,347	O pequeno comerciante que vendia sorrisos
3.094	0,0003	1,303	Os mil pássaros de Sadako - Parte 1
2.733	0,0004	1,303	Os mil pássaros de Sadako - Parte 2
1.255	0,0009	1,383	Pão de Tâmaras e nozes

3.7. SITES MAIS VISITADOS

Neste ponto pretende-se identificar uma LP na internet. Para isso foi necessário identificar os mil sites mais visitados, fornecidos pela *ad planner top1000* da Google [28]. Como seria de prever o Facebook.com foi o mais procurado, com 590.000.000 de visitas mensais.

Seguindo o mesmo procedimento dos casos anteriores a LP é representada na Figura 23, com o gráfico Log-Log correspondente.

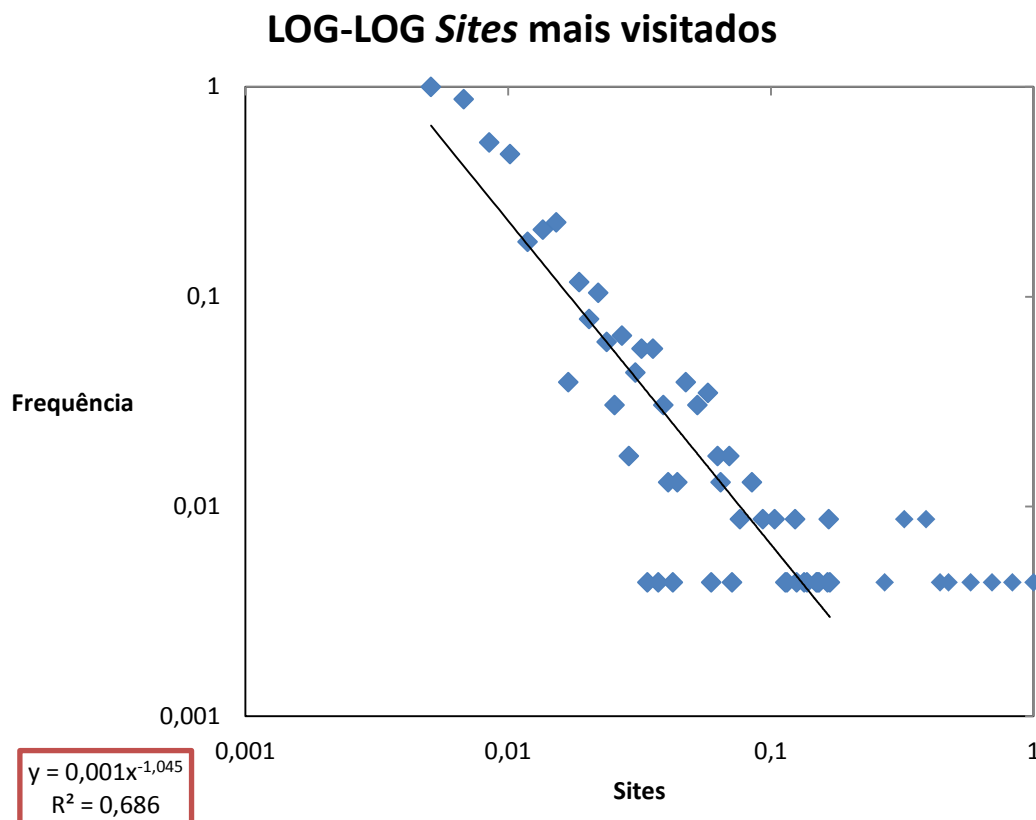


Figura 23 Gráfico Log-Log sites visitados

O valor de α é de 1,045 para a LP de visitas a sites da internet.

3.8. PREJUÍZOS CAUSADOS POR CHEIAS NOS ESTADOS UNIDOS

Os prejuízos causados por cheias são muito elevados todos os anos. O caso dos Estados Unidos não é excepção, muito pelo contrário, pois é um lugar muito afectado por catástrofes deste tipo. Com base em dados fornecidos pela *National Weather Service*, desde 1904 a 2010 os prejuízos anuais ascendem a muitos biliões de dólares [29]. O ano de 2005 foi o que registou mais prejuízos, que ascenderam a 49,66 B\$.

Estes prejuízos com fenómenos naturais seguem uma LP, apresentada no gráfico Log-Log da Figura 24.

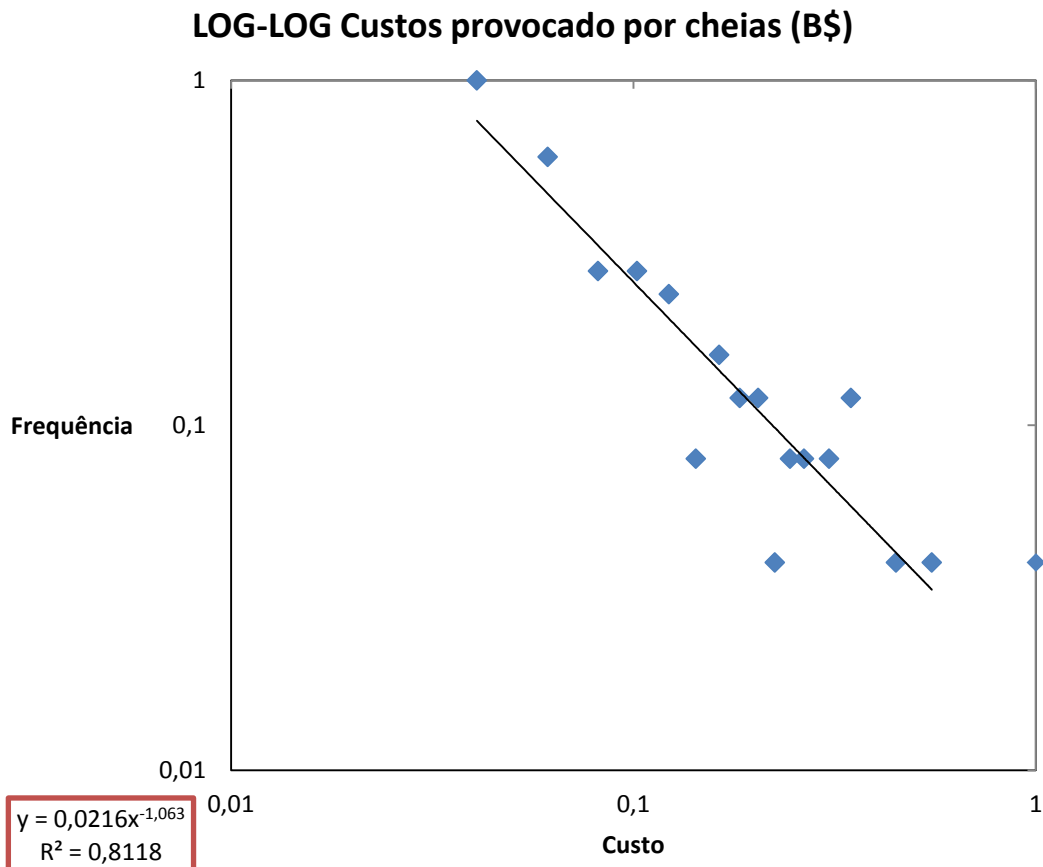


Figura 24 Gráfico Log-Log custos cheias nos Estados Unidos

É retirado deste gráfico mais um valor de $\alpha = 1,063$.

3.9. PREJUÍZOS E MORTES CAUSADOS POR TORNADOS

Durante o ano de 2010, nos Estados Unidos, ocorreram 1.282 tornados e 45 mortos. Os prejuízos causados por estes tornados foram objecto de estudo, para identificar uma LP. O tornado que mais danos causou teve um impacto de 750 milhões de dólares de prejuízos [30].

O gráfico da Figura 25 representa a frequência relativa do número de tornados com determinado prejuízo. É representado numa escala Log-Log o gráfico resultante e a LP associada.

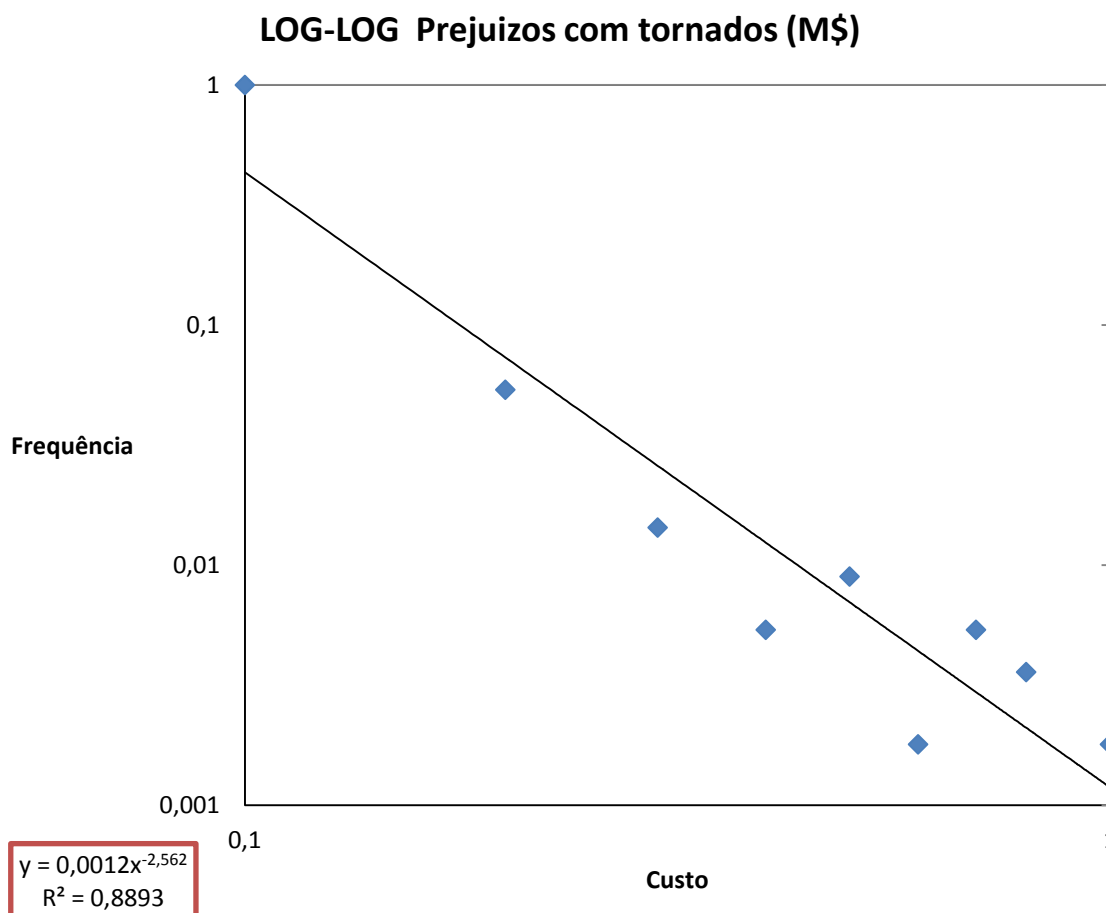


Figura 25 Gráfico Log-Log custos tornados nos Estados Unidos em 2010

O expoente da distribuição, α , tem o valor de 2,562. Mais uma vez neste caso, pode-se afirmar que existem muitos tornados a provocar poucos prejuízos e poucos tornados a provocar muitos prejuízos.

As mortes causadas por tornados nos Estados Unidos desde o ano de 1950 é outro ponto abordado no *TornadoHistoryProject.com*. Foi verificado que entre os anos 2000 e 2010 ocorreram aproximadamente 250 tornados, que provocaram mortes. Existe o registo de um tornado com 24 mortes e o registo de 130 tornados com uma morte [30].

O gráfico Log-Log da Figura 26 representa a frequência relativa do número de mortes por tornado, nos Estados Unidos entre 2000 e 2010. A LP expressa o expoente da distribuição de valor 1,611.

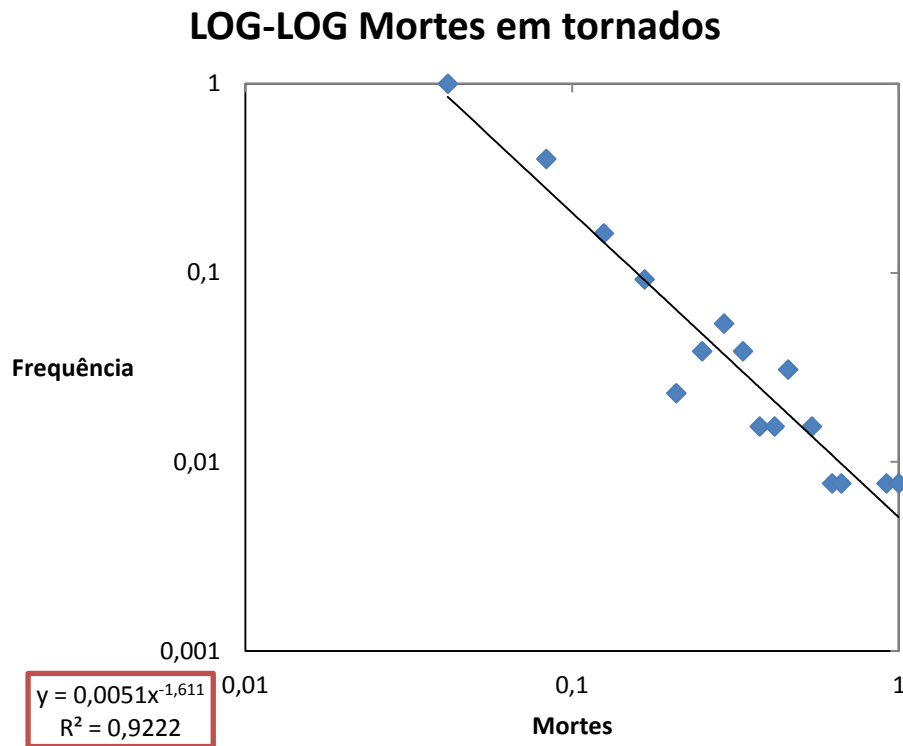


Figura 26 Gráfico Log-Log Mortes em tornados nos Estados Unidos entre 2000 e 2010

3.10. MORTES CAUSADAS POR ACIDENTES COM AVALANCHES

Uma avalanche é uma massa de neve que se desprende de um cume ou montanha. Esta grande massa de neve tem a força de arrastar à sua frente rochas, gelo, árvores, entre outros. A massa de neve é constituída por camadas que se acumularam ao longo do tempo. As camadas são paralelas à superfície do solo, em que cada uma indica uma condição meteorológica durante a qual a neve se acumulou. Para ocorrer uma avalanche é necessário que uma camada mais fraca (mais instável) se separe das outras [31].

Todos os anos nas zonas onde se pratica esqui, e não só, morrem pessoas em resultado de estas avalanches. Só desde o início do ano de 2011 já morreram 20 pessoas nos Estados Unidos e Canadá. O objectivo deste estudo é verificar se este tipo de fenómeno natural segue uma LP. Para isso foi consultado o registo da *U.S. Avalanche Accidents Reports* entre os anos de 1998 e 2010, no Canadá e Estados Unidos [32].

A frequência relativa de acidentes mortais com avalanches é mostrada no gráfico Log-Log da Figura 27, entre 1998 e 2010. Do gráfico surge o valor de α correspondente a 2,729.

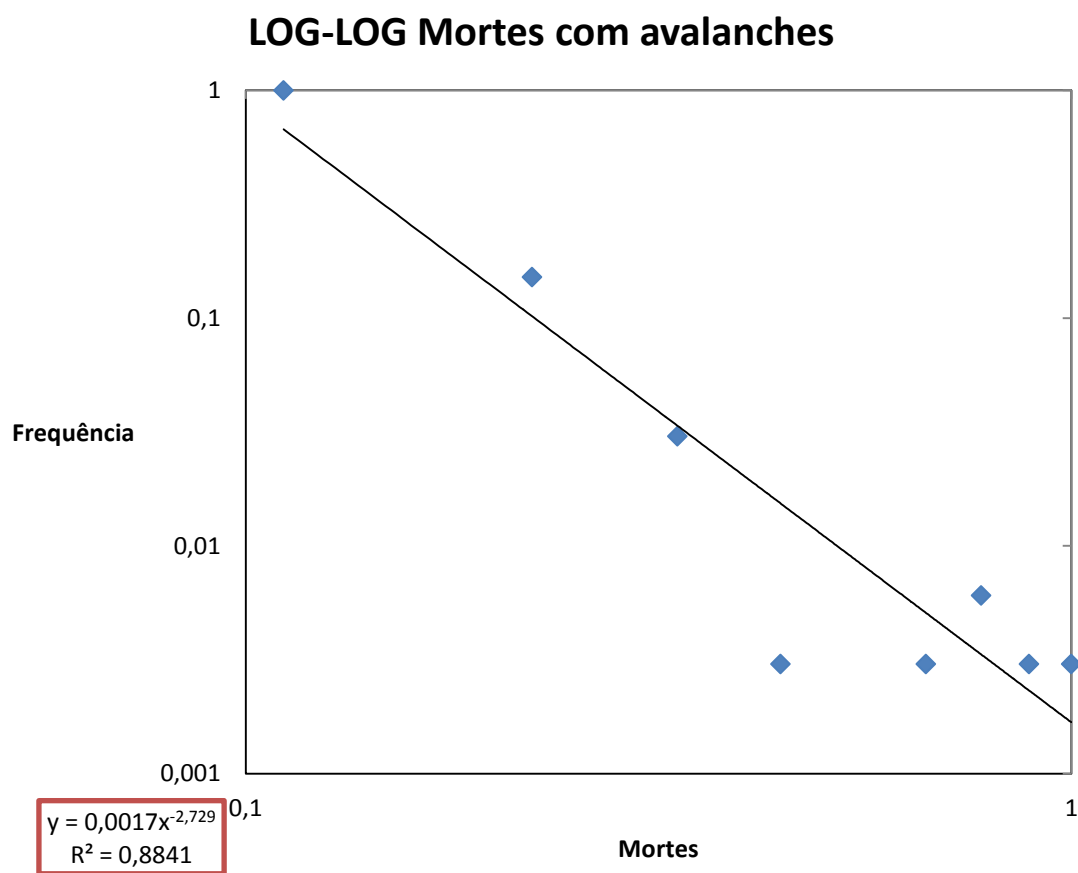


Figura 27 Gráfico Log-Log Mortes em avalanches nos Estados Unidos e Canadá

3.11. DIÂMETRO DE CRATERAS DA LUA

A Lua é o maior satélite natural do sistema solar. A origem da Lua é incerta, mas as semelhanças no teor dos elementos encontrados tanto na Lua quanto na Terra indicam que ambos os corpos podem ter tido uma origem comum [33][34].

A maioria das crateras da superfície lunar são devidas a impactos de meteoritos ou asteróides. A cratera com o diâmetro maior registada na listagem da Wikipédia é de 591 Km, com o nome de Hertzsprung [34] [35]. Embora na mesma fonte se afirme que a maior encontrada até hoje tem um diâmetro de 2.240 Km [35].

O estudo tem como base uma Tabela de 1.600 crateras [34]. O gráfico da Figura 28 correspondente ao Log-Log do diâmetro de crateras da Lua. O parâmetro α da LP tem o valor de 3,011.

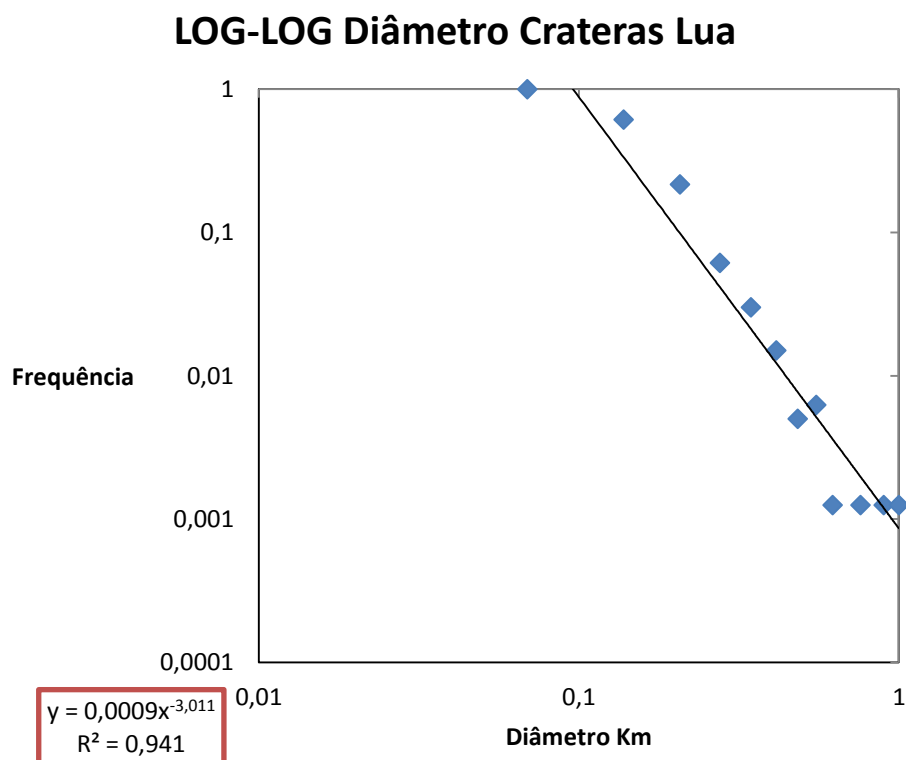


Figura 28 Gráfico Log-Log diâmetro de crateras na Lua

3.12. DIÂMETRO DE CRATERAS E MONTANHAS EM MARTE

Marte é o quarto planeta a contar do Sol, situando-se entre a Terra e a cintura de asteróides, a 1,5 UA do Sol (ou seja, a 1,5 vezes a distância da Terra ao Sol). De noite, aparece como uma estrela vermelha, sendo que os romanos lhe deram o nome do seu deus da guerra [36].

Existem milhares de crateras em Marte com diâmetro maior que 1 Km, mas apenas mil têm um nome. Os nomes foram dados pela *International Astronomical Union*, depois de uma petição de cientistas. As crateras só podem ter nomes de pessoas falecidas [36] [37].

É com base na listagem de 1.002 crateras fornecidas pela Wikipedia que este estudo é desenvolvido. A maior cratera que aparece nesta listagem tem 471 Km, encontrada em 1973 e chamada de *Schiaparelli* (*Giovanni Schiaparelli*) [37].

O diâmetro das crateras em Marte segue uma LP, tendo o expoente da distribuição o valor 2,723. Pode ver-se no gráfico Log-Log da Figura 29 o declive negativo da recta encontrada.

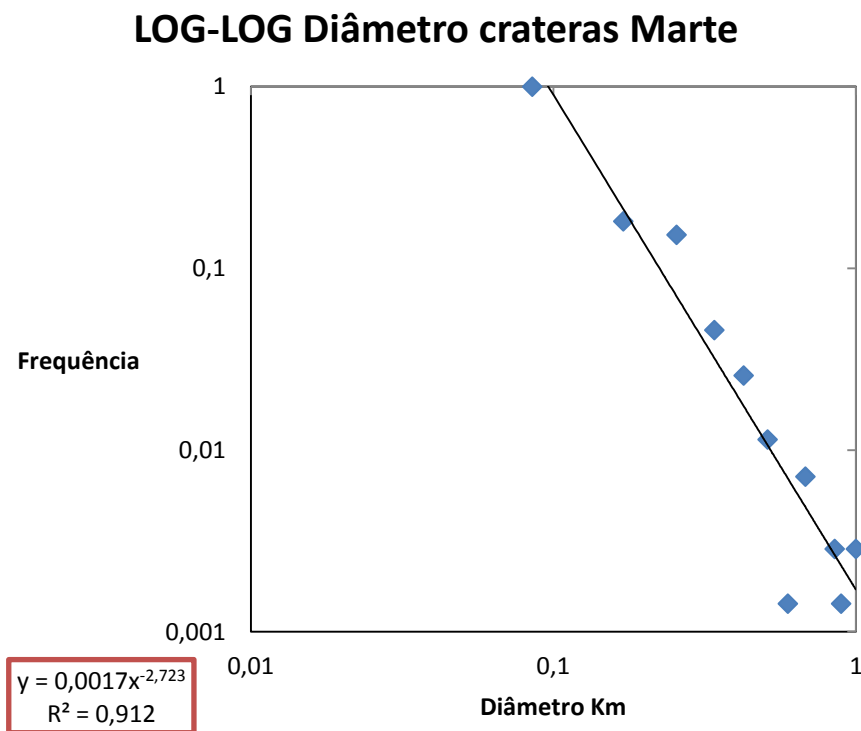


Figura 29 Gráfico Log-Log diâmetro de crateras em Marte

Em Marte existem também montanhas, originadas por vulcões, embora em muito menor número que as crateras. Neste caso foi analisado o diâmetro de 51 montanhas, listadas na Wikipédia. A montanha de maior diâmetro, com 1.840 Km, tem o nome de *Tharsis Montes* [38].

É possível verificar pelo gráfico Log-Log na Figura 30, o declive negativo da recta, com o valor α de 1,488. É mais uma vez identificada uma LP para o estudo do diâmetro das montanhas em Marte.

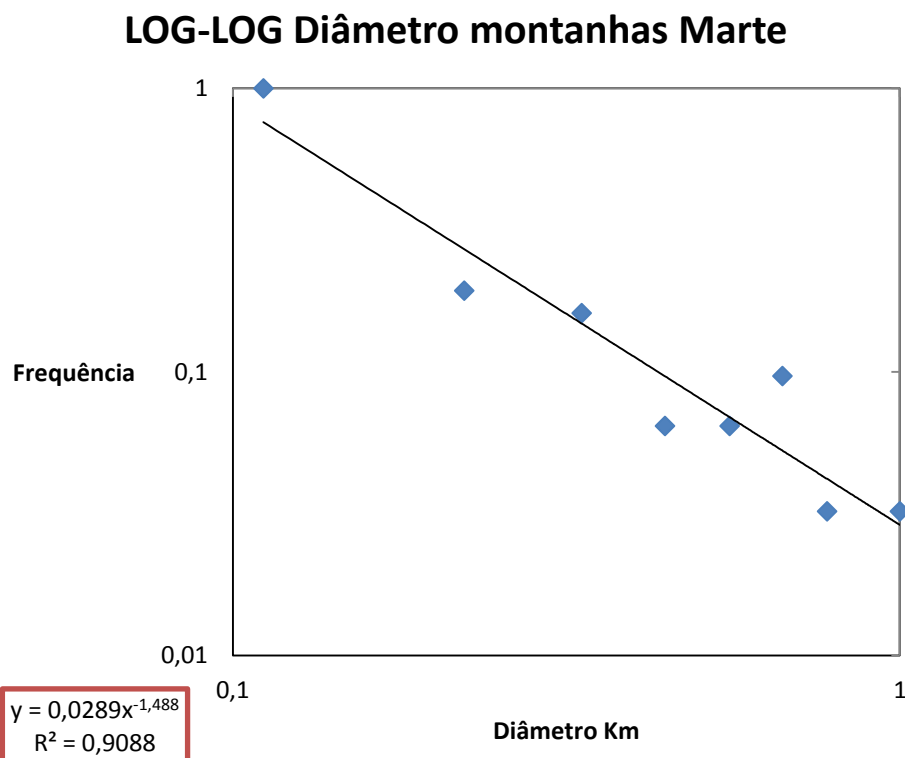


Figura 30 Gráfico Log-Log diâmetro de montanhas em Marte

3.13. ELEVAÇÃO DOS VULCÕES

O conceito de vulcão tem sido difícil de normalizar. A definição de vulcão varia com a sua abertura e medida. Entende-se por vulcão uma abertura na superfície da terra, através da qual se dá a erupção do magma, dos gases e das cinzas associadas [39].

Na base de dados referente aos vulcões existem dados sobre a altura. Foi um item interessante para testar uma LP, pois existem muitos vulcões com pequena elevação e poucos com grande elevação. O vulcão com maior elevação está catalogado como tendo 6.887 m, denominado “*Ojos del Salado, Nevados*”, situado entre o Chile e Argentina [39].

Para este último estudo são considerados apenas os vulcões com mais de 1000 m, num total de 744 em todo o mundo. O resultado é apresentado no gráfico Log-Log da Figura 31, com o expoente da distribuição α de valor 2,227.

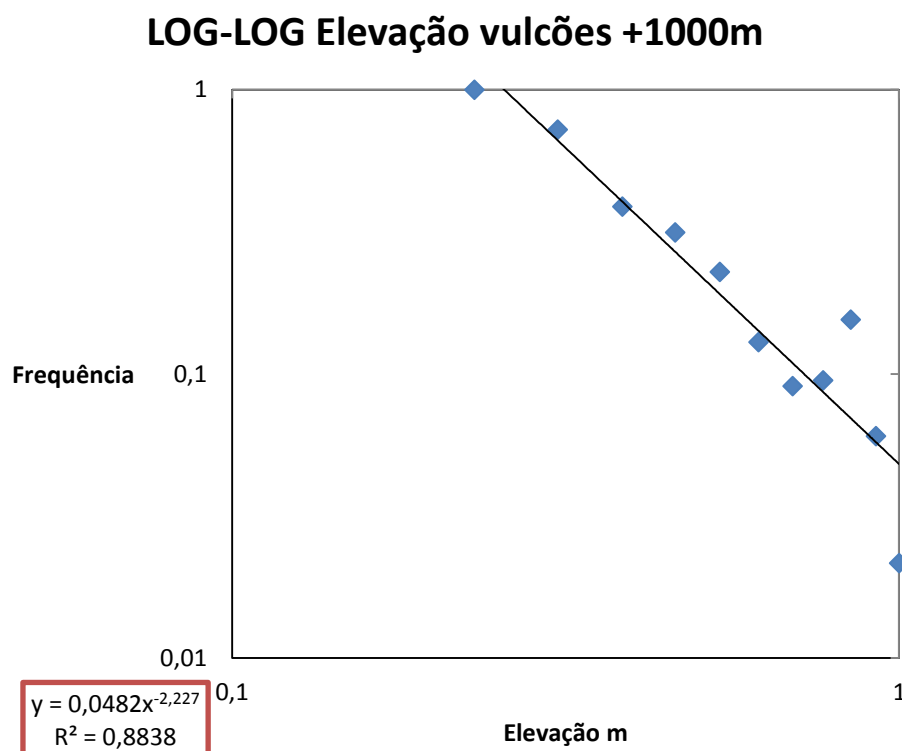


Figura 31 Gráfico Log-Log Elevação dos vulcões com mais de 1000m

4. COMPARAÇÃO DOS ESTUDOS REALIZADOS

4.1. ANÁLISE DE RESULTADOS

Como foi possível observar ao longo do capítulo 3, todos os exemplos estudados seguem uma LP. Exemplos esses que definem as características básicas da LP inversa, que são:

- Declive negativo da recta de ajuste ao gráfico Log-Log;
- Valor de $\alpha > 0$;

Importante será referir que nem todos os estudos realizados para este trabalho seguiam uma distribuição da LP, como por exemplo a frequência de letras num texto literário (em português). Ou seja, nem todos os temas em que se suponha existir uma LP, na verdade ela foi verificada.

No caso da população das cidades portuguesas, duas observações devem ser feitas:

- A lei de Zipf, que nada mais é do que um caso particular lei de Pareto, e onde o expoente α é igual a 2, não se aplica às cidades portuguesas. Para isso seria necessário que a segunda cidade tivesse metade da população da primeira, a terceira um terço da primeira, e assim sucessivamente.
- A LP (inversa) é claramente identificada no estudo, pois os valores seguem a distribuição desejada. Assim, com um valor $\alpha = 1,364$ e uma constante $c = 0,0054$ pode afirmar-se que este estudo segue uma distribuição de LP.

O estudo sobre os mais ricos do mundo de 2011, destacados na revista Forbes, apresenta também um resultado interessante. Tem-se um $\alpha = 1,746$ e uma constante $c = 0,0012$. Ou seja, tem-se um valor de α a divergir nitidamente da unidade, notando uma acentuada discrepância entre as várias posições do ranking dos mais ricos.

No tema “Magnitude dos sismos” é verificado um valor do expoente α bastante elevado. Este poderá ter proveniência da falta de um número considerável de dados. Embora a frequência média anual das várias magnitudes estejam “supostamente” correctas, existe a falta de registos para magnitudes inferiores a dois.

No caso da frequência de palavras em textos, foi efectuado o estudo em vários textos e observou-se que em todos eles existe uma distribuição de LP. Mais uma vez, pode afirmar-se que os textos em português escolhidos não convergem para a lei de Zipf. No entanto com textos literários maiores poderia haver uma aproximação a esta lei.

Pretende-se com estes vários casos, descobrir se existe algum relacionamento com o tamanho dos textos. Na Figura 32, recorrendo à Tabela 6, foi elaborado um gráfico para verificação do resultado. É de notar que os valores já estão todos normalizados e, por isso, é possível colocá-los todos no mesmo gráfico.

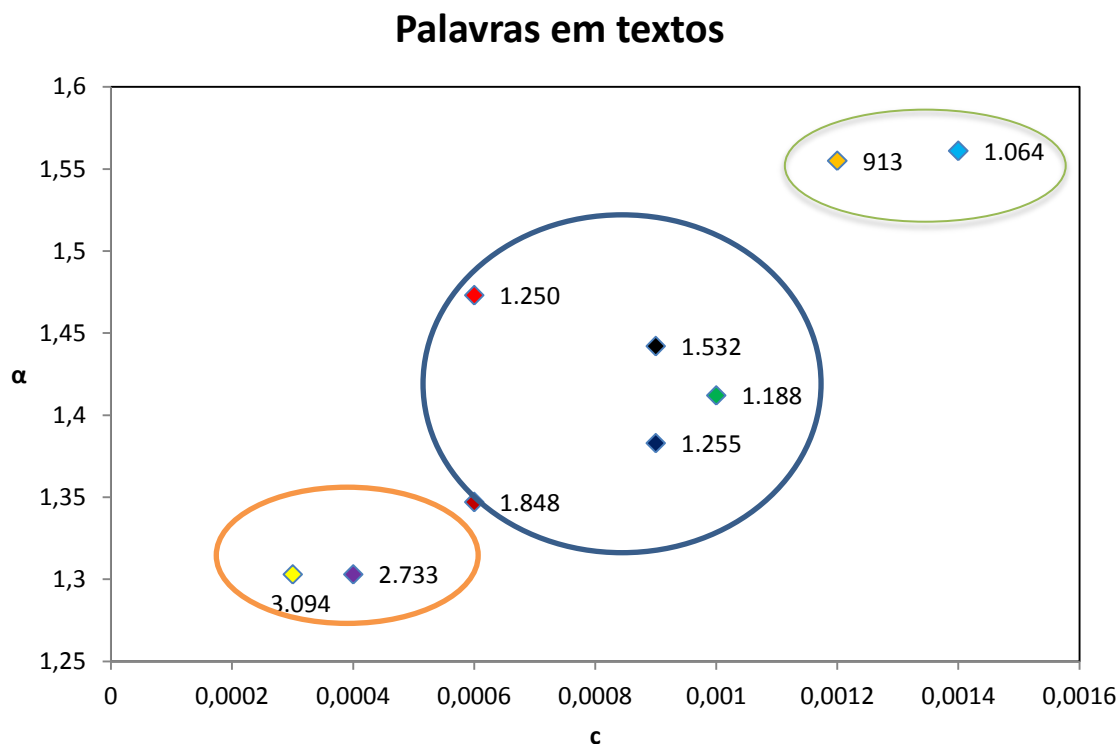


Figura 32 Gráfico relação de (α) e (c)

Pela análise do gráfico, pode constatar-se que os textos com maior número de palavras se situam na parte inferior e os de menor número de palavras se situam na parte superior. Na parte central é representado o grupo de pontos correspondentes aos textos em que o número de palavras está compreendido entre 1.100 e 1.900. Deste grupo não se consegue afirmar que exista uma relação entre o valor de α e o número de palavras.

Nos restantes estudos efectuados foi possível encontrar a LP associada. Cada qual com o seu expoente da distribuição e com a sua constante c . Valores esses que irão servir de base para a análise no ponto 4.2.

É tentador encontrar grandes causas por trás de terríveis terremotos ou extinções em massa, é também tentador encontrar grandes pessoas por trás dos grandes eventos históricos. Entretanto, é possível que a única causa geral para tais eventos seja a organização interna de uma situação crítica, que faz com que eventos raros sejam não apenas possíveis mas inevitáveis. Os fundamentos de uma situação crítica reflectem-se em leis estatísticas

simples. Ou seja, leis de potência, que não possuem uma escala característica, revelam a ausência de uma “grandeza” esperada para o próximo evento.

4.2. GRÁFICO AVANÇADO DE RESULTADOS

Este é um ponto relacionado com a meta-análise dos resultados. A meta-análise utiliza técnicas estatísticas para combinar, num determinado resumo, os resultados de estudos independentes voltados a uma única questão que são as LP. Todas as normalizações efectuadas ao longo do estudo foram feitas a pensar nesta análise, dado que os estudos tinham grandezas diferentes e só assim seria possível compilar tudo num único gráfico.

Na Tabela 7, foram sintetizados os resultados obtidos nos vários estudos realizados. Os valores das constantes c aparecem na primeira coluna e os expoentes, α , surgem na segunda coluna.

Tabela 7 Valores resultantes dos estudos

c	α	Estudos
0,0012	1,555	Palavras "A prenda do Cisne"
0,0006	1,473	Palavras "A senhora dos Livros"
0,0010	1,412	Palavras "As Sapatilhas de Sofia"
0,0009	1,442	Palavras "O menino Golfinho"
0,0014	1,561	Palavras "O Pastor amoroso"
0,0006	1,347	Palavras "O pequeno comerciante que vendia sorrisos"
0,0003	1,303	Palavras "Os mil passaros de Sadako - Parte 1"
0,0004	1,303	Palavras "Os mil passaros de Sadako - Parte 2"
0,0009	1,383	Palavras "Pão de Tamaras e nozes"
0,0010	1,045	Sites mais visitados
0,0216	1,063	Prejuizos causados por cheias EUA (1903-2009)
0,0051	1,611	Mortes em tornados EUA (2000-2010)
0,0009	3,011	Diâmetro Crateras Lua
0,0289	1,488	Diâmetro Crateras Marte
0,0017	2,723	Diâmetro montanhas Marte
0,0017	2,729	Mortes em acidentes com Avalanches (1998-2010 no Canada e EUA)
0,0482	2,227	Elevação dos Vulcões com mais de 1000m
0,0038	1,148	Área ardida incêndios em Portugal 2009
0,0072	1,131	Área ardida incêndios em Portugal 2010
0,0012	1,746	Lista Forbes dos mais ricos do mundo 2011
0,0054	1,364	População nas cidades Portuguesas 2010
4E-06	9,956	Magnitude Terramotos
0,0012	2,562	Prejuizo estragos com tornados até 10 M\$

Com base nestes dados foi efectuado o gráfico de dispersão, da Figura 33. Como se pode observar, o gráfico resultante apresenta no eixo vertical os valores de α e no eixo do horizontal os valores de c .

A análise correspondente ao tema “Magnitude dos Sismos” foi excluída, porque resultou num valor inesperado do expoente α . Este distanciamento poderá ser resultante dos dados adquiridos, ou seja, por um lado o diminuto número de pontos e, por outro, os dados terem começado a partir de dois na escala de *Richter*. Como os dados não foram suficientes para o estudo coloca-se alguma incerteza na veracidade deste caso.

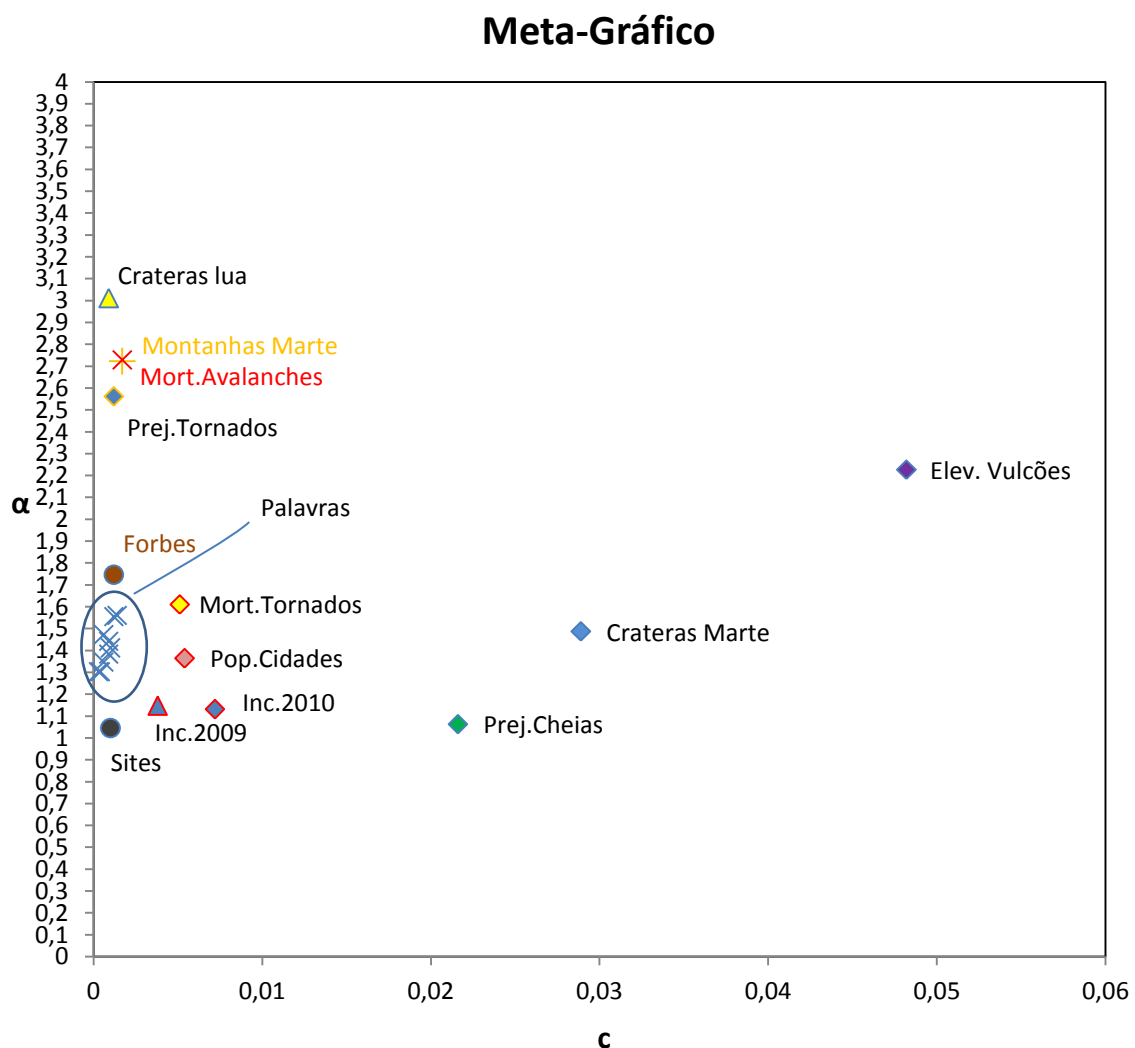


Figura 33 Meta-gráfico resultante dos vários estudos

Os pontos alusivos à frequência de palavras em textos (o pequeno grupo com um círculo) estão todos muito seguidos e não se desviam muito daquela área, indiciando um padrão para este tema. Os pontos que residem na área do gráfico mais à direita, que é o caso do “Prejuízos Cheias”, “Crateras Marte” e “Elevação dos Vulcões” sugerem também um modelo próprio para estes casos.

Na Figura 34 são examinados mais a fundo as diversas áreas do gráfico, tentando-se estabelecer alguma relação com os grupos de pontos.

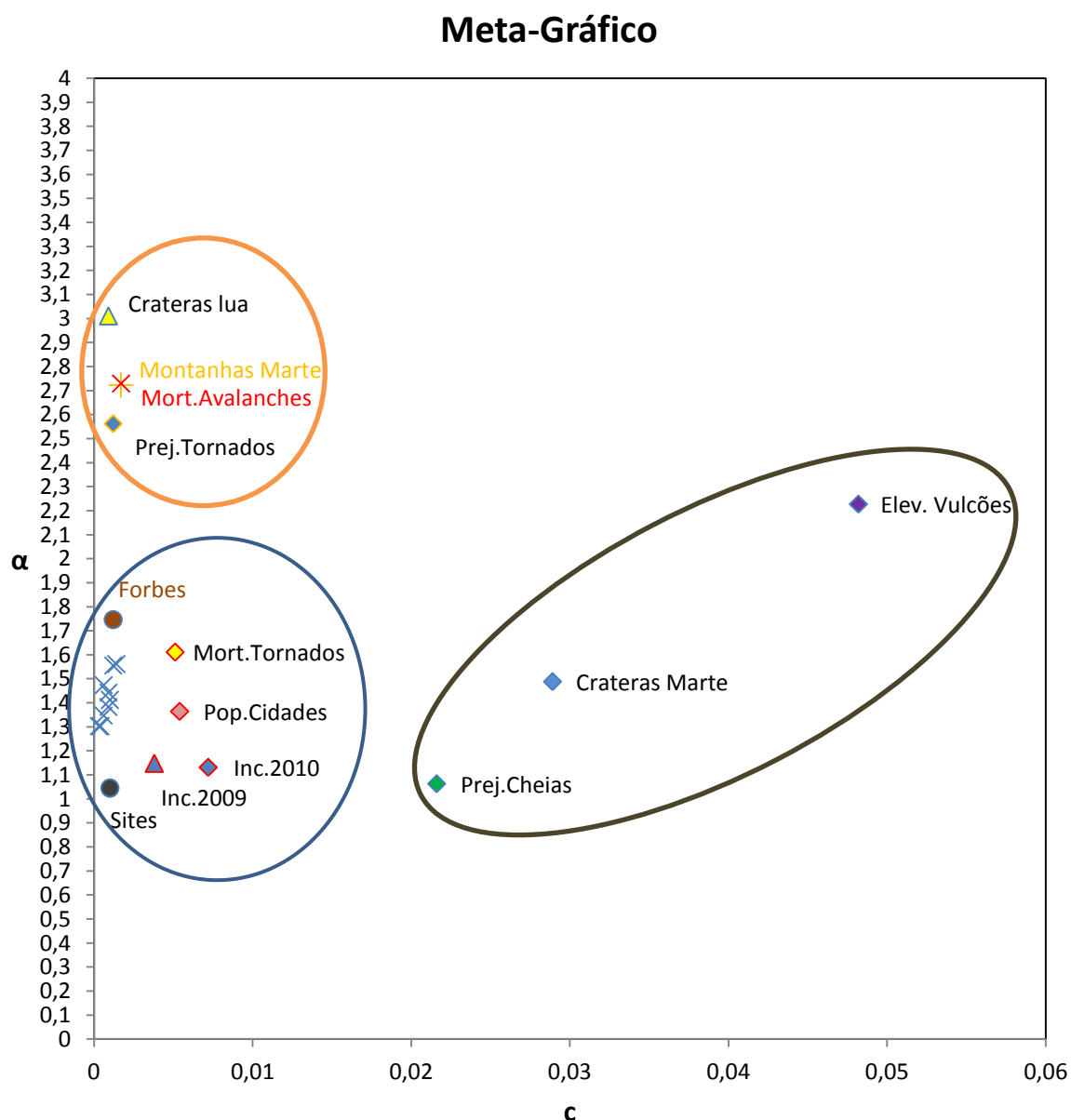


Figura 34 Gráfico por grupos de resultados

O objectivo deste agrupamento é identificar se existem particularidades entre fenómenos naturais e fenómenos artificiais.

O grupo de três pontos mais à direita do gráfico foi identificado como sendo um grupo de LP em fenómenos naturais. Grosso modo, seria presumível afirmar que uma distribuição LP correspondente a um fenómeno natural teria o valor de c superior a 0,015. No entanto não é possível certificar essa conclusão sem que haja mais dados.

Reuniu-se outro grupo que seguia uma LP em fenómenos naturais, um grupo de quatro pontos composto por “Prejuízos com Tornados”, “Mortes com Avalanches”, “Montanhas de Marte” e “Crateras na Lua”. Seria aceitável, também pela ocupação do grupo na área do gráfico, dizer que para evidenciar este tipo de fenómeno o valor α teria de ser superior a 2. Poder-se-ia ainda complementar com a observação do parágrafo anterior e atestar: para uma LP com a constante c superior a 0,015 e/ou α superior a 2, está-se perante um fenómeno natural. Mais uma vez, não se pode certificar esta conclusão sem que haja mais estudos.

Por último, o grupo mais controverso é constituído pelos pontos das “palavras”, “Forbes”, “incêndios”, “cidades”, “sites” e “mortes com tornados”. Este é o grupo destinado aos fenómenos artificiais conforme se pode observar pelo gráfico. Os nove pontos relativos ao estudo das palavras encontram-se bastante sólidos, e tudo leva a supor que exista um padrão para este fenómeno artificial. Nesta circunstância é um pouco excessivo afirmar que a LP de um fenómeno artificial terá um α compreendido entre 1 e 2 e uma constante c dentro do intervalo $[0; 0,015]$. No entanto tudo leva a admitir que esta conclusão se aplica e se possa considerar um padrão, mas existe um ponto no meio que não é um fenómeno artificial. Este ponto é referente à “morte com tornados”, e põe em causa este padrão.

O estudo sobre mortes com tornados nos EUA entre 2000 e 2010 deveria estar nos grupos destinados aos fenómenos naturais e não no grupo dos fenómenos artificiais. Este facto pode ter várias explicações, entre elas dados insuficientes e um intervalo temporal reduzido.

5. CONCLUSÕES

O principal objectivo deste trabalho foi a identificação de LP em fenómenos naturais e artificiais. Este reconhecimento foi conseguido com a realização de vários estudos e onde se concluiu que existem muitas e diversificadas áreas onde as LP se aplicam. Fenómenos críticos aparecem associados ao conceito de criticalidade auto-organizada, na qual um factor de grandeza poderá divergir de um sistema natural ou artificial.

Foram descritos ao longo do trabalho mecanismos de identificação para explicar a ocorrência de LP. Estimando o parâmetro α é possível identificar à partida uma LP. De todos os estudos expostos, a totalidade apresentou um α superior a um.

As distribuições LP têm características muito próprias, em que a principal é o declive negativo da recta e a aproximação à lei de Pareto. Embora se tentasse identificar a lei de Zipf no estudo correspondente ao tamanho das cidades em Portugal, este não seguia essa particularidade. A desigualdade existente na sociedade fica muito evidente no estudo relativo aos mais ricos do mundo. Na realidade existem muitas pessoas pobres e poucas pessoas ricas. Isso é comprovado pela distribuição resultante.

Tentou-se estabelecer um padrão para fenómenos naturais e artificiais, com a realização do “meta-gráfico”. No entanto a relação não foi conseguida da totalidade, mas ficou iniciado

um bom caminho, que poderá ir mais longe com a inclusão de mais dados e estudos. Este tipo de meta-análise poderá abrir portas a um novo conceito no interior das LP.

Quanto às limitações encontradas para a realização deste trabalho, destaca-se a falta de dados. Estes dados são de difícil acesso, e seria interessante existir um *site* mundial onde encontrar mais dados.

Para pesquisas futuras, seria um desafio interessante realizar mais estudos de forma a conseguir expandir o “meta-gráfico” e tirar conclusões mais exactas. A análise da rede social profissional LinkedIn, assim como outras redes semelhantes, seriam outras pesquisas a realizar.

O estudo da distribuição da LP é uma área em que existe um interesse considerável nas pesquisas da actualidade. Os mecanismos e as explicações apresentadas aqui certamente darão algumas reflexões. Há muito trabalho a ser feito tanto no campo experimental como teórico, antes de se poder entender os processos físicos de condução destes sistemas. Sem dúvida, existirão muitas descobertas interessantes para serem realizadas.

Referências Documentais

- [1] Machado, T. (2010). Descrição Trabalho. Porto: ISEP.
- [2] Rhodes, Chris J. e Anderson, Roy M. (1996). Power laws governing epidemics in isolated populations. *Nature* 381: pp. 600-2.
- [3] Iram Gleria, R. M. (2004). Sistemas Complexos, Criticalidade e leis de potência. *Revista Brasileira de Ensino de Física* , pp. 99-108.
- [4] Newman, M. E. (2006). Power laws, Pareto distribution and Zipf's law. Department of Physics and Center for the Study of Complex Systems, University of Michigan.
- [5] Komulainen, T. (2003). Session 10. In *Self-Similarity and Power Laws* (pp. 109-122). Helsinki University of Technology.
- [6] Aljure, Y., & Gallego, J. (2010). Desigualdad y leyes de potencia. *Cuadernos de Economía*, 29(53).
- [7] Regueiro, M. D. (2007). Incendios, Complejidad, Inteligencia y Linealidad. *Centro de Formación e Recursos de Lugo*.
- [8] André, S. (2010). BC-0506: Comunicação e Redes. *Leis de Potência*. Universidade Federal do ABC.
- [9] Clauset, A., & Shalizi, C. R. (June de 2010). Scaling: welcome the inverse power law.
- [10] Rosa, A. F., Teixeira, G. M., Mendes, A. C., & Abreu, S. M. (2010). Gerenciamento de resultados: Análise Sociométrica e Bibliométrica dos autores de referência.
- [11] Li, W. (2003). Zipf's Law Everywhere. *Glottometrics* , pp. 14-21.
- [12] Mitzenmacher, M. (2005). Editorial: The Future of Power Law Research. *Internet Mathematics* , pp. 525-534.
- [13] Instituto Nacional de Estatística. (Janeiro de 2010). <http://www.ine.pt>. Obtido em 2011.
- [14] Forbes. (2011). <http://www.forbes.com/wealth/Billionaires/list>. Obtido em 2011.
- [15] USGS. (s.d.). <http://earthquake.usgs.gov/earthquakes/eqarchives/year/eqstats.php>. Obtido em 2011.
- [16] Direção de Unidade de Defesa da Floresta. (2010). *Relatório anual de áreas ardidas e ocorrências 2009*. Autoridade Florestal Nacional.
- [17] Direção de Unidade de Defesa da Floresta. (2011). *Relatório anual de áreas ardidas e ocorrências 2010*. Autoridade Florestal Nacional.
- [18] Borkmann, D. (s.d.). http://www.imn.htwk-leipzig.de/~dborkman/offtopic/letter_frequency/letter.html. Obtido em 2011.
- [19] Seabrooke, B. (1995). A Prenda do Cisne. London: Walker Books.
- [20] Henson, H. (2010). A Senhora dos Livros. Barcelona: Editorial Juventud.

- [21] Lipp, F. (2007). *As sapatilhas de Sofia*. Barcelona: Intermón Oxfam.
- [22] Morpurgo, M., & Foreman, M. (2004). *O menino Golfinho*. London: Andersen Press.
- [23] Silva, E. L. (2006). *O pastor amoroso*. In F. Pessoa. Rio de Janeiro.
- [24] Coerr, E. (2003). *Os mil pássaros de Sadako I*. Toulouse: Éditions Milan.
- [25] Coerr, E. (2003). *Os mil pássaros de Sadako II*. Toulouse: Éditions Milan.
- [26] Carquain, S. (2008). *O pequeno comerciante que vendia sorrisos*. Paris: Ed. Albin Michel.
- [27] Mayer-Skumanz, L. (2003). *Pão de tâmaras e nozes*. Wien: Herder Verlag.
- [28] Google. (s.d.). <http://www.google.com/adplanner/static/top1000/>. Obtido em 2011.
- [29] National Weather Service. (s.d.). http://www.nws.noaa.gov/oh/hic/flood_stats/Flood_loss_time_series.shtml. Obtido em 2011.
- [30] Lietz, J. (s.d.). <http://www.tornadohistoryproject.com/browse>. Obtido em 2011, de TornadoHistoryProject.
- [31] Wikipedia. (2011). <http://en.wikipedia.org/wiki/Avalanche>. Obtido em 2011.
- [32] American avalanche association. (s.d.). <http://avalanche.org/accidents.php>. Obtido em 2011, de U.S. Avalanche Accidents Reports.
- [33] Neukum, G., Konig, B., & Arkani-Hamed, J. (1974). A study of lunar impact crater size-distributions. *Earth, Moon, and Planets* , pp. 201-229.
- [34] Wikipedia. (2011). http://en.wikipedia.org/wiki/List_of_craters_on_the_Moon,_A-B. Obtido em 2011.
- [35] Wikipedia. (2011). <http://pt.wikipedia.org/wiki/Lua>. Obtido em 2011.
- [36] Wikipedia. (2011). [http://pt.wikipedia.org/wiki/Marte_\(planeta\)](http://pt.wikipedia.org/wiki/Marte_(planeta)). Obtido em 2011.
- [37] Wikipedia. (2011). http://en.wikipedia.org/wiki/List_of_craters_on_Mars. Obtido em 2011.
- [38] Wikipedia. (2011). http://en.wikipedia.org/wiki/List_of_mountains_on_Mars. Obtido em 2011.
- [39] National Museum of natural History. (2011). <http://www.volcano.si.edu/index.cfm>. Obtido em 2011, de Global Volcanism Program.

Histórico

- 4 de Setembro de 2011, Versão 1.0, <mailto:1960581@isep.ipp.pt>
- 11 de Setembro de 2011, Versão 1.1, <mailto:1960581@isep.ipp.pt>
- 31 de Outubro de 2011, Versão 1.2, <mailto:1960581@isep.ipp.pt>