



Pesquisa Bibliométrica aplicada à pesquisa em Informática e Conhecimento

Prof. Dr. Domingos Márcio Rodrigues Napolitano – PPGI UNINOVE

Artigos para análise - Resenha

1. Cuccurullo, C., Aria, M., & Sarto, F. (2016). Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics*
2. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*
3. Salo, F., Injadat, M., Nassif, A. B., Shami, A., & Essex, A. (2018). Data mining techniques in intrusion detection systems: A systematic literature review. *IEEE Access*
4. Rajab, S., & Sharma, V. (2018). A review on the applications of neuro-fuzzy systems in business. *Artificial Intelligence Review*
5. De Oliveira, C. M., Terra, T. A. M., Azevedo Filho, E. T., Heymann, M. C., & MATIAS, I. D. O. (2018). Rfid nos cuidados com a saúde: Uma revisão de mapeamento sistemático. *Acta Biomedica Brasiliensia*
6. Ylijoki, O., & Porras, J. (2016). Conceptualizing big data: Analysis of case studies. *Intelligent systems in accounting, finance and management*
7. Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*

Resenhas e Seminários

Data	Dia	Tópico	Seminário	Apresnetador
16/09/2020	Qua	Aula 3 A- Redes Bibliométrica e Wos		
23/09/2020	Qua	Aula 3 B Vosviewr	Cuccurullo, C., Aria, M., & Sarto, F. (2016). Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. Scientometrics, 108(2), 595-611.	Domingos
30/09/2020	Qua	Aula 4 Análises Qunatitativas e Lei de Lotka Seminário 1	Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113.	Roger Amanda
07/10/2020	Qua	Aula 5 Lei de Zipf e Lei de Bradford Seminário 2	Salo, F., Injadat, M., Nassif, A. B., Shami, A., & Essex, A. (2018). Data mining techniques in intrusion detection systems: A systematic literature review. IEEE Access, 6, 56046-56058.	Fernando e Eli
14/10/2020	Qua	Aula 6 Fundamentando Teorias com Bibliometria Seminário 3	Rajab, S., & Sharma, V. (2018). A review on the applications of neuro-fuzzy systems in business. Artificial Intelligence Review, 49(4), 481-510.	Andersnn e Cintia
21/10/2020	Qua	Bibliomterix - Seminário 4	De Oliveira, C. M., Terra, T. A. M., Azevedo Filho, E. T., Heymann, M. C., & MATIAS, I. D. O. (2018). Rfid nos cuidados com a saúde: Uma revisão de mapeamento sistemático. Acta Biomedica Brasiliensia, 9(1), 48-61.	Casio e Sabrina
28/10/2020	Qua	Biblioshiny - Seminário 5	Ylijoki, O., & Porras, J. (2016). Conceptualizing big data: Analysis of case studies. Intelligent systems in accounting, finance and management, 23(4), 295-310.	Eduardo e Jorge
04/11/2020	Qua	Aplicando o Bibliometriz numa Pesquisa Seminário 6	Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. Information and Software Technology, 54(1), 41-59.	Vinnicius e Diego Candal
11/11/2020	Qua	Outras Ferramentas Bibliométricas Seminário 7		
18/11/2020	Qua	Acompanhameto do Trabalho aplicado		
25/11/2020	Qua	Acompanhameto do Trabalho aplicado		
02/12/2020	Qua	Apresnetação do Trabalho Aplicado		
09/12/2020	Qua	Tópicos avançados		

Seminários e Avaliações

Enviado memorial com requisitos de entrega dos trabalhos

Avaliação de Conceitos (Prova com consulta): 6/11

Aula para Retirada de Dúvidas e Reunião do Grupo: 13/11

Entrega dos trabalhos escritos: 21/11

Apresentação dos Trabalhos: 27/11

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

O propósito desta atividade é avaliação dos conhecimentos absorvidos pelo discente, para tanto serão realizados as seguintes avaliações:

Conhecimento dos conceitos dos aprendidos

Aplicação dos conhecimentos aprendidos

Desenvolvimento dos conceitos aprendidos.

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

Estas avaliações serão realizadas por meio de um trabalho prático em grupo, que consiste na realização de uma revisão sistemática da literatura com tema de escolha livre, porém recomenda-se que sejam temas ligados ao tema de pesquisa de um ou mais alunos do grupo.

O grupo terá no máximo 3 componentes, não pode haver mais de um aluno de doutorado no grupo.

O texto final deve ter entre 18 e 20 páginas contando com todos os elementos de uma produção científica, incluindo tabelas, figuras e referências.

Serão permitidas pesquisas de tipo secundárias, ou seja, revisões de artigos referentes às pesquisas primárias, ou ainda revisões de tipo terciária, revisões de RSL sobre um mesmo tema. Para maiores esclarecimentos recomenda-se consultar Kitchenham e Charters. (2007) disponível em nosso repositório.

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

RELATÓRIO DE PESQUISA

Introdução

Neste tópico deve haver uma contextualização do tema a ser pesquisado, o problema de pesquisa, na forma de uma questão a ser respondida, o objetivo geral da pesquisa, e breve descrição da metodologia aplicada para o atingimento deste objetivo.

Fundamentação Teórica

Descrição dos principais resultados encontrados na pesquisa de forma a estabelecer o referencial definido todos os elementos da questão e objetivo de pesquisa e demonstrando que ambos possuem sustentação teórica.

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

RELATÓRIO DE PESQUISA

Metodologia

Este capítulo deve conter os seguintes tópicos:

- Caracterização metodológica:
 - Estabelecer as características gerais da pesquisa, definir a tipologia e o formato da RSL, na qual pode ser usado Paré et al (2016)
 - Estabelecer o problema de pesquisa determinando uma questão de pesquisa
 - Determinar e justificar o objetivo de pesquisa
- Plano de trabalho
 - Definir a(s) base(s) a ser(em) pesquisada(s).
 - Definir questões de pesquisa bibliométricas
 - Definir critério de busca e seleção de artigos
 - Definir os métodos de análise
 - Definir os softwares a serem empregados, no mínimo deve-ser utilizado o VoSViewer. Porém pode-se utilizar também CiteNetScore e Bibliometrix ou Biblishiny.
 - Outros elementos que o grupo considere importantes

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

RELATÓRIO DE PESQUISA

Resultados e discussão

- Estrutura dos resultados pesquisados nas bases, quantidades de artigos por palavras chaves ou palavras de busca
- Evolução histórica das produções pesquisadas
- Evolução histórica das citações das produções pesquisadas
- Índices e indicadores
- Seleção de publicações para análise, discussão e quantidades analisadas
- Redes bibliométricas:
 - Redes de citação direta e discussão
 - Redes de co-citação e discussão
 - Redes de acoplamento bibliográfico e discussão
 - Redes de co-ocorrência de palavras chaves e discussão
 - Outras visualizações que o grupo considerar relevantes
- Outras análises quantitativas ou qualitativas
- Respostas às perguntas de Pesquisa Bibliométrica, cada uma deve ser respondida e discutida.
- Resposta à questão de pesquisa e atingimento dos objetivos determinados

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

Conclusões

- Conclusões finais do grupo.

Bibliografia

- Usar estilo de citação APA, ou ABNT, seguir o mesmo padrão para apresentação as figuras, quadros e tabelas.

Orientações para execução e entrega do trabalho final da Disciplina “Introdução à Bibliométrica para Pesquisa em Informática”

APRESENTAÇÃO DO RELATÓRIO DE PESQUISA

O grupo irá apresentar o conteúdo sumarizado da pesquisa em uma apresentação de no máximo 20 minutos, com 15 minutos de arguição.

O apresentador do trabalho será escolhido na hora da apresentação, **PODENDO SER QUALQUER ELEMENTO DO GRUPO.**

A sessão de perguntas consiste numa arguição com uma pergunta para cada componente do grupo. A resposta de cada um implicará num FATOR INDIVIDUAL, que será multiplicado pela nota final do obtida pelo relatório, conforme formulário de avaliação a seguir.

Aula 4 – Produtividade de Periódicos e Indicadores de Citação

Entendendo a produtividade de Periódicos a Lei de Bradford

A lei de Bradford foi formulado em 1934 por Samuel C. Bradford,

Ao analisar as revistas em que estes artigos foram publicados, ele encontrou uma regularidade, ou seja, uma relação inversa entre o número de artigos publicados em uma área de assunto e o número de revistas em que os artigos aparecem.

Isto significa que, numa determinada área, um pequeno número de publicações são responsáveis por uma parte considerável do total das publicações em uma área.

Entendendo a produtividade de Periódicos a Lei de Bradford

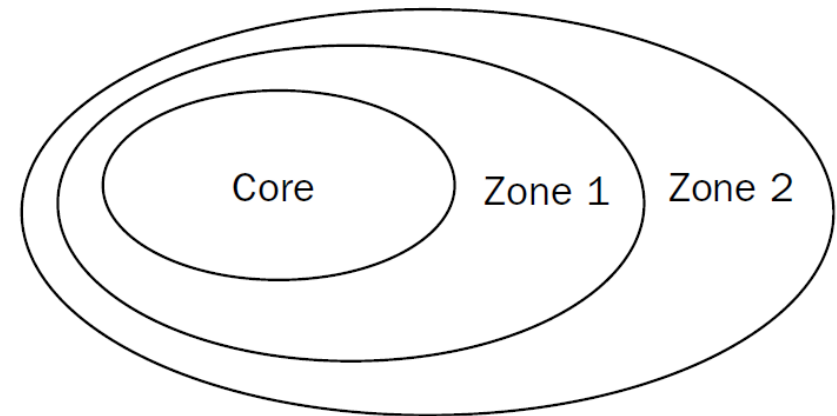
Esta lei sobre a produtividade revista pode ser facilmente representados em forma gráfica (Figura ao lado). R

evistas são classificados e divididos em grupos ou categorias, dependendo do número de artigos que eles apresentam.

O critério para o estabelecimento dessas zonas é que o número de artigos em cada zona tem de ser o mesmo.

No entanto, o número de revistas que publicam esses artigos não vai ser o mesmo em cada zona, como algumas revistas serão mais produtivas do que outras.

Bradford zones



Entendendo a produtividade de Periódicos a Lei de Bradford

Dado que a lei de Bradford classifica os periódicos de acordo com sua produtividade.

Um pequeno grupo de artigos estará localizado na primeira zona central, enquanto um número crescente de periódicos será encontrado em cada zona subsequente.

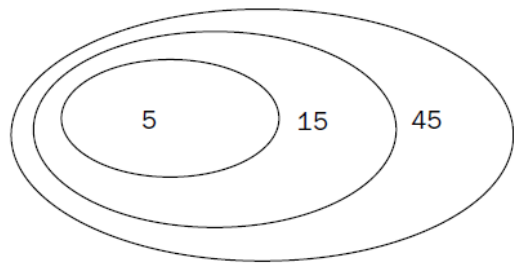
O primeiro grupo compreende os principais periódicos e conterá um determinado número de artigos.

Embora o número de artigos permaneça constante em todas as zonas, o número de periódicos aumentará entre as zonas.

A relação entre o número de periódicos em zonas subsequentes foi observada em aproximadamente $1: n: n^2 \dots$

Entendendo a produtividade de Periódicos a Lei de Bradford

Example of Bradford zones where $k = 3$



Isso significa que, para uma determinada área de assunto, devemos reunir todos os artigos publicados em um determinado período de tempo e listar os periódicos em que foram publicados.

Suponhamos que reunimos um total de 300 artigos publicados em 65 revistas. Se aceitarmos a hipótese de que nossa área de assunto se encaixa em uma lei de Bradford de três zonas, haverá aproximadamente 100 artigos em cada zona.

A tarefa é classificar os periódicos de acordo com sua produtividade e contar quantos periódicos pertencem à primeira zona (periódicos mais produtivos).

Suponha que a primeira zona compreenda cinco periódicos, que representam os periódicos mais produtivos (zona central) que publicam um total de 100 artigos na área de pesquisa.

O próximo passo envolve contar o número de periódicos na Zona 2: suponhamos que ele contenha 15 periódicos.

Por fim, imagine que o número de periódicos encontrados na terceira área seja 45.

Esse último grupo é formado pelos periódicos menos produtivos, pois o número total de artigos publicados por eles ainda será 100. Aplicando a formulação básica da lei de Bradford ao nosso exemplo produz a distribuição mostrada na Figura ao lado.

Simplificando a regularidade à formulação de Bradford fornece 1: 3: 9. Nesse caso, a constante n (geralmente denominada k) que multiplica o número de diários nas zonas será 3.

Entendendo a produtividade de Periódicos a Lei de Bradford

Dado que a lei de Bradford classifica os periódicos de acordo com sua produtividade.

Um pequeno grupo de artigos estará localizado na primeira zona central, enquanto um número crescente de periódicos será encontrado em cada zona subsequente.

O primeiro grupo compreende os principais periódicos e conterá um determinado número de artigos.

Embora o número de artigos permaneça constante em todas as zonas, o número de periódicos aumentará entre as zonas.

A relação entre o número de periódicos em zonas subsequentes foi observada em aproximadamente $1: n: n^2 \dots$

A Curva S

Considerando a lei de Bradford em maior profundidade, é possível analisar a natureza de sua distribuição. O gráfico da produtividade do periódico produzirá um gráfico semelhante ao mostrado na Figura ao lado.

Esta figura representa a forma de S como o gráfico típico de Bradford-Zipf (Brookes, 1969).

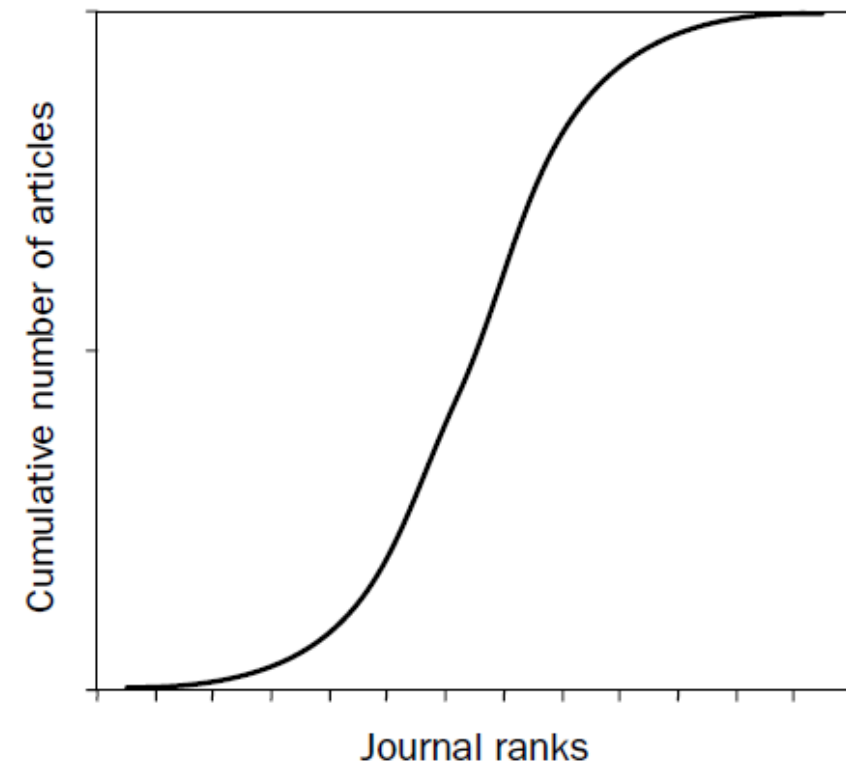
É o resultado da plotagem do número acumulado de artigos de cada periódico contra o logaritmo de sua classificação.

Os periódicos são classificados do mais ao menos produtivo.

Obviamente, a forma S de Bradford encontrada nas distribuições de dados reais não será exatamente igual à mostrada na Figura ao lado, pois os dados reais não se comportarão de maneira tão perfeita.

No entanto, existem evidências de que os dados se encaixam na forma de Bradford S (Tsay et al., 2000).

Bradford's S-shape



Exemplo da Lei de Bradford

Para ilustrar como calcular a lei de Bradford com dados reais, vamos usar a distribuição de produtividade das revistas encontrada na indexação automática por Pulgarín e Gil-Leiva (2004).

Esses autores estudaram a produtividade científica nessa área no período de 1956 a 2000, encontrando um total de 527 artigos.

Como a lei de Bradford se concentra na produtividade da revista, é necessário contar o número de artigos publicados por cada revista nessa área específica.

Uma tabela de frequências mostrando o número de periódicos que publicam um determinado número de artigos pode ser obtida.

A Tabela ao lado mostra a dispersão da literatura científica encontrada na área de indexação automática.

Table 4.1 Dispersion of scientific literature in *automatic indexing* area

No. of journals	No. of articles	Cumulative journals	Cumulative articles	Ln (cumulative journals)
1	66	1	66	0
1	40	2	106	0.6931
1	36	3	142	1.0986
1	28	4	170	1.3862
1	21	5	191	1.6094
1	18	6	209	1.7917
1	10	7	219	1.9459
1	8	8	227	2.0794
2	7	10	241	2.3025
4	6	14	265	2.6390
5	5	19	290	2.9444
6	4	25	314	3.2188
14	3	39	356	3.6635
27	2	66	410	4.1896
117	1	183	527	5.2094

Reprinted with permission from Pulgarín and Gil-Leiva (2004).

Entendendo as três leis da bibliometria

A lei de Lotka (1926): www.lis.uiuc.edu/~jdownie/biblio/lotka.html

- Produtividade de autores (pesquisadores); quantos pesquisadores escreveram 1, 2, 3 ... artigos?
- distribuição de frequência (probabilidade)

A lei de Bradford (1934): www.lis.uiuc.edu/~jdownie/biblio/bradford.html

- Scatter dos artigos de um determinado assunto a mais revistas
- grupos iguais-produção

A lei de Zipf (1949): www.lis.uiuc.edu/~jdownie/biblio/zipf.html

- A frequência de palavras no texto
- Posto de distribuição / Frequência

A Lei de Lotka

distribuição de frequência da produção científica

lei de autoria de Lotka descreve as frequências de publicação para autores de um dado domínio:

"... O número (de autores) fazendo n contribuições é de cerca de $1/n^2$ das pessoas que faz um; ... e a proporção de todos os colaboradores, que fazem uma única contribuição, é de cerca de 60 por cento"
(Lotka, 1926)

$f(n) = C / n^2$ - Onde $f(n)$ denota o número de autores dentro de um domínio que, em um determinado período, ter publicado n artigos

C é uma constante e denota o número de pesquisadores que têm produzido um artigo

O número de pesquisadores que publicaram n artigos é proporcional com $1 / n^2$ - isso é ...

Lei de Lotka

Se 100 pesquisadores publicaram um artigo cada um dentro de um determinado domínio, em seguida, a lei de Lotka prevê que:

$100/2^2 = 25$ pesquisadores, Vai ter publicado 2 artigos,

$100/3^2 = 11$ pesquisadores, Vai ter publicado artigos 3, e

$100/7^2 = 2$ pesquisadores publicaram 7 artigos ...

Porém a em grandes conjuntos de artigos e durante um período de tempo adequado (dependente de domínio) a Lei de Lotka é bastante precisa, mas não (necessariamente) de forma estatística.

Lei de Bradford

Lei de Bradford (dispersão) é muitas vezes usado como um guia para determinar o número de revistas do núcleo (core) dentro de um determinado assunto

A maioria dos artigos sobre um determinado assunto têm uma tendência a ser agrupados em um pequeno número de revistas

Os artigos restantes tendem a serem espalhados ao longo de um grande número de revistas

Enunciado da Lei de Bradford

“Se revistas científicas são dispostos em ordem decrescente de produtividade de artigos sobre um determinado assunto, eles podem ser divididos em um núcleo de periódicos mais particularmente dedicadas aos grupos de sujeitos e vários ou zonas que contêm o mesmo número de artigos como o núcleo, quando o número de periódicos no núcleo e zonas sucessivas será tão 1: N: N²”

(SC Bradford, 1934)

A proporção matemática entre o número de revistas no núcleo e na zona 2 é uma constante = n (multiplicador de Bradford); e a zona 2n²

Bradford expressa esta relação tão 1:n:n², Mas ele nunca definiu a lei matematicamente - foi ajustado a aprox. 5

formulação simples: uma regra de 80/20, ou seja, 20% dos periódicos na bibliografia abrange 80% dos artigos na bibliografia sobre um determinado assunto durante um determinado período de tempo

Exemplo Lei de Bradford

Exemplo: dados geofísicos de Bradford:

O número de artigos deve ser idealmente o mesmo no três zonas

A lei de Bradford não é estatisticamente exato, mas é muito difundida como uma regra de ouro-

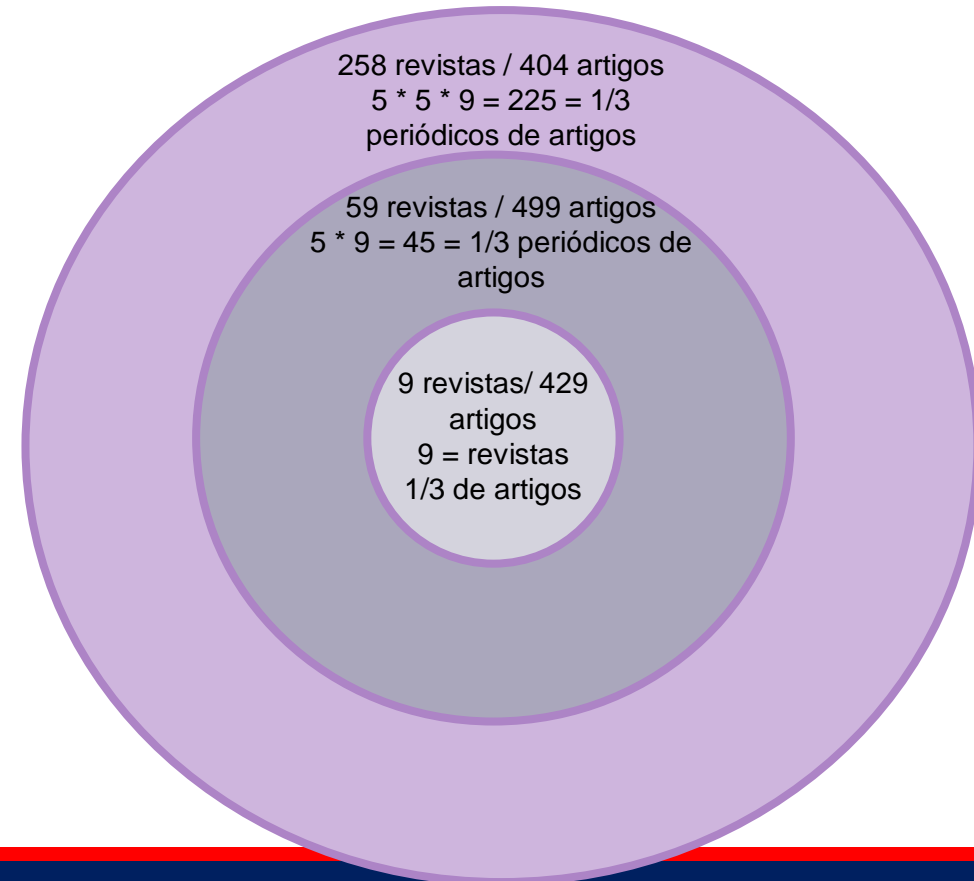
	Zone 1	Zone 2	Zone 3
Art	429	499	405
Tss	$9=a$	59	225
	1	n	n2
	1a	na	n2a
	$(1*9)$	$(5*10)$	$(25*10)$

a é definido
arbitrariamente

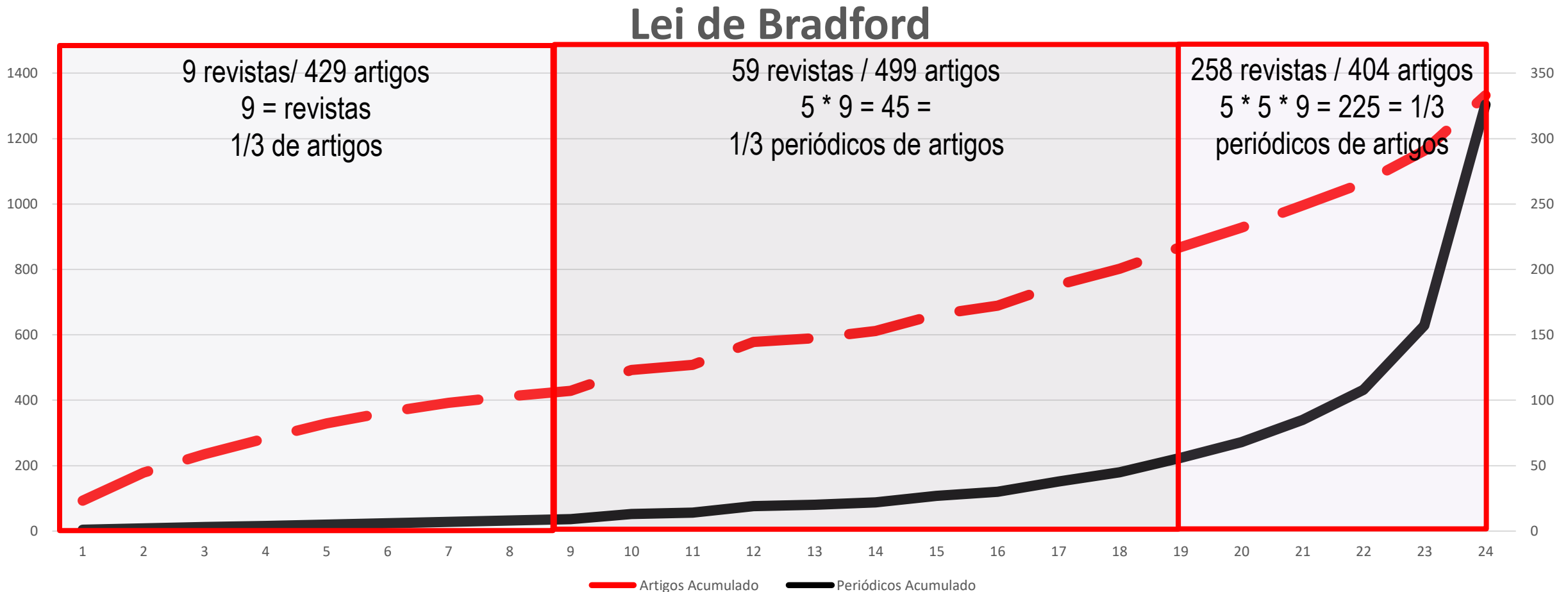
Aplicação da Lei de Bradford

Zona	Periódicos	Artigos	Periódicos Acumulado	Artigos Acumulado
Zona 0	1	93	1	93
	1	86	2	179
	1	56	3	235
	1	48	4	283
	1	46	5	329
	1	35	6	364
	1	28	7	392
	1	20	8	412
	1	17	9	429
Zona 1	4	16	13	493
	1	15	14	508
	5	14	19	578
	1	12	20	590
	2	11	22	612
	5	10	27	662
	3	9	30	689
	8	8	38	753
	7	7	45	802
	11	6	56	868
Zona 2	12	5	68	928
	17	4	85	996
	23	3	108	1065
	49	2	157	1163
	169	1	326	1332

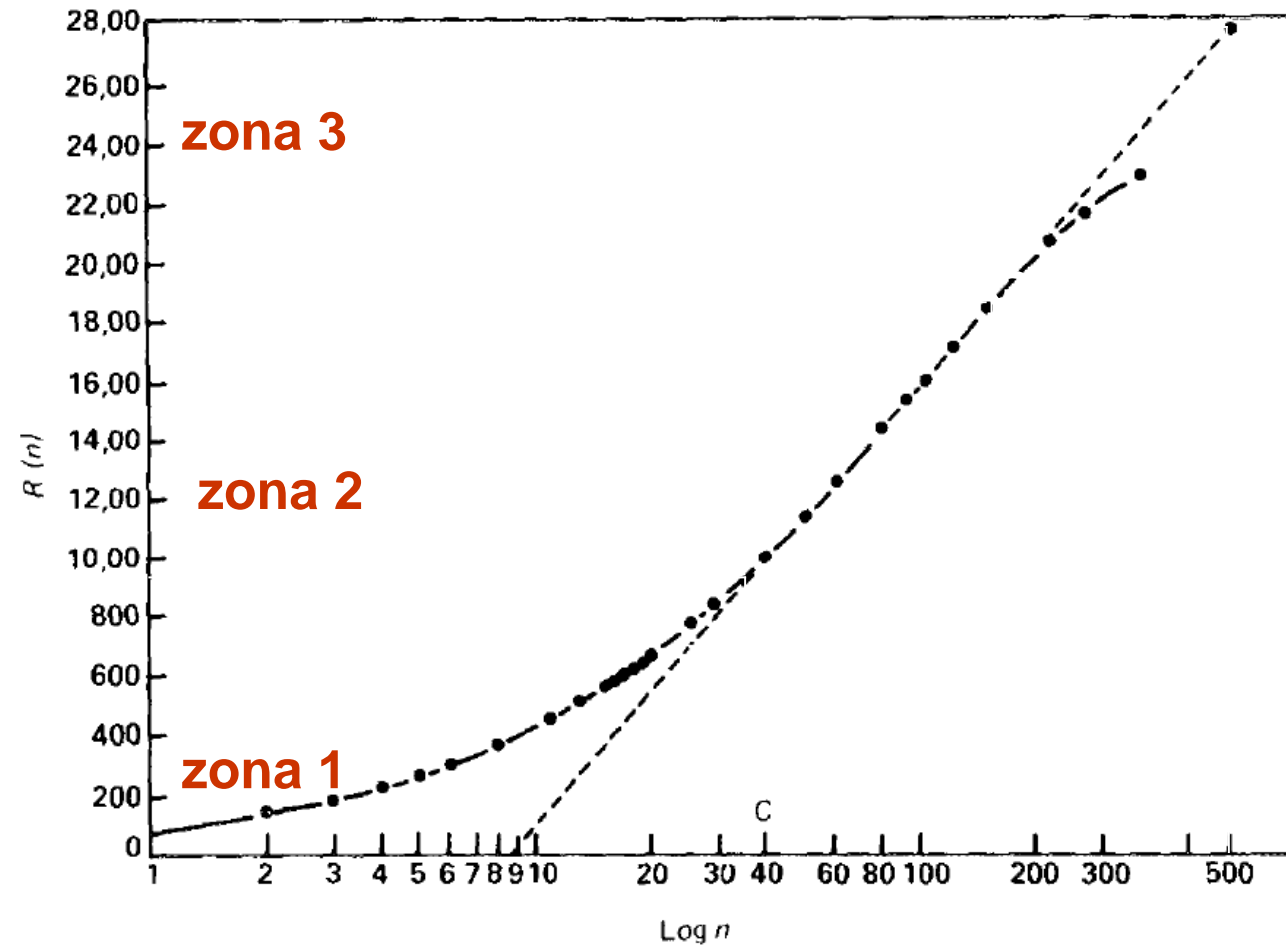
$$1, n, n^2 = 1, 5, 25$$



Distribuição Periódicos – Lei de Bradford



Visualização Gráfica



... além de Garfield 'Lei da Concentração' de 1971

Em vez de 'dispersão bibliométrica' Garfield sugeriu 'concentração bibliométrica' - em vez um axioma:

Cada disciplina tem um pequeno núcleo de revistas cobrindo uma grande porção de artigos (Bradford)

o 'cauda longa' de cada disciplina consiste nos núcleos de outras disciplinas!

Assim, a não limitado. de revistas em SCI

Lei de Zipf

A lei de Zipf é usada para prever frequências palavra em texto completo

A lei diz que se você (em um relativamente grande corpus de texto):

- “...listar as palavras que ocorrem dentro desse texto em ordem decrescente de frequência, a patente de uma palavra nessa lista multiplicado por sua frequência será igual a uma constante “. (Zipf, 1949)
- $r \times f = c$, Onde r é o número de classificação e f a frequência de cada termo
- (Acontece para descrever a parte linear de uma curva de Bradford)
- Nem a lei de Zipf é estatisticamente exato, mas é extremamente valioso para efeitos de ponderação do prazo nos motores de busca com indexação automática

Lei de Zipf

A lei de Zipf é usada para prever frequências palavra em texto completo

A lei diz que se você (em um relativamente grande corpus de texto):

- “...listar as palavras que ocorrem dentro desse texto em ordem decrescente de frequência, a patente de uma palavra nessa lista multiplicado por sua frequência será igual a uma constante “. (Zipf, 1949)
- $r \times f = c$, Onde r é o número de classificação e f a frequência de cada termo
- (Acontece para descrever a parte linear de uma curva de Bradford)
- Nem a lei de Zipf é estatisticamente exato, mas é extremamente valioso para efeitos de ponderação do prazo nos motores de busca com indexação automática

Lei de Zipf: frequência de ocorrência de palavras

A Lei de Zipf é também chamada de lei quantitativa fundamental da atividade humana.

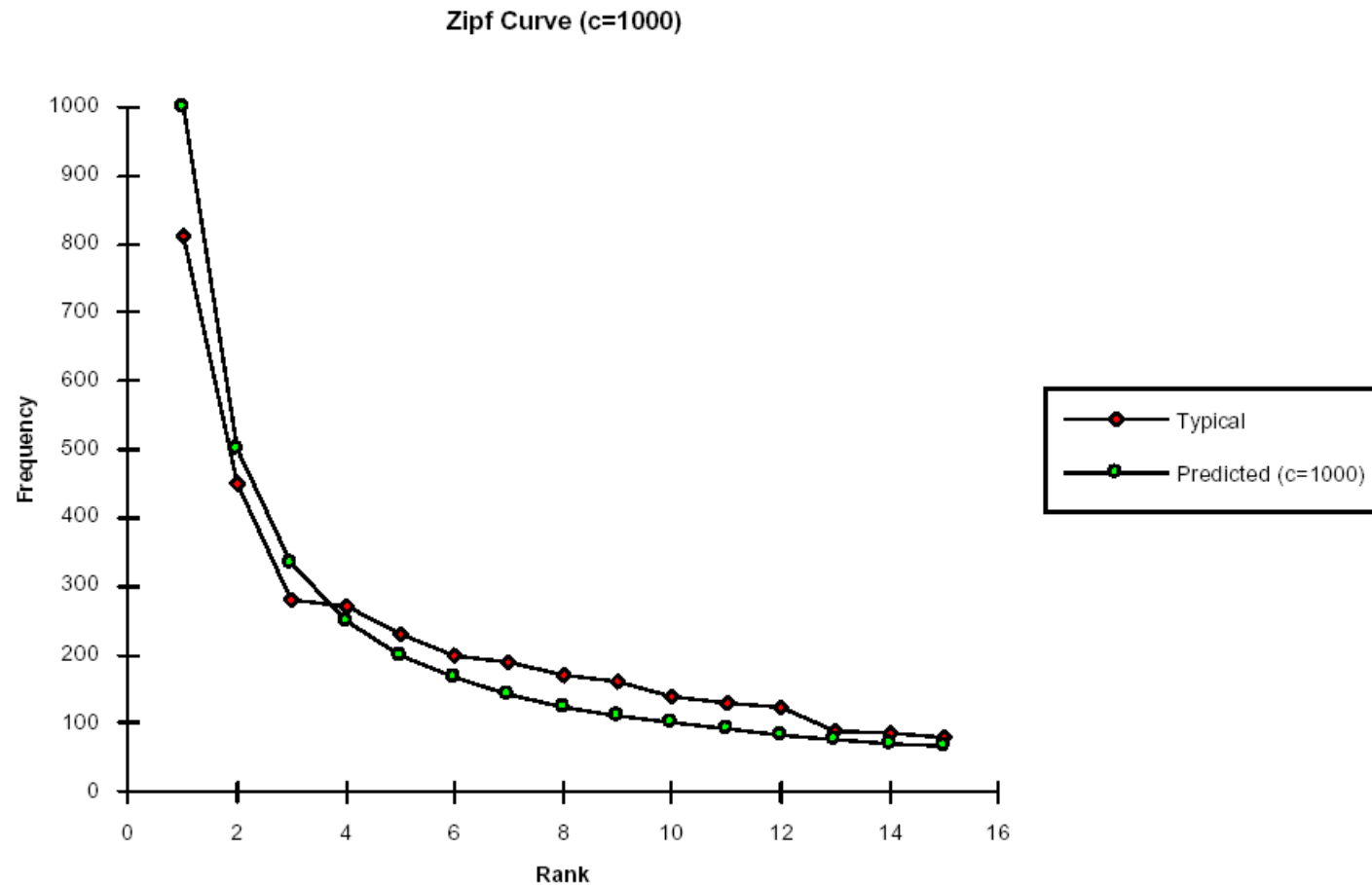
Segundo a Lei de Zipf subdivide-se em:

- primeira Lei de Zipf: diz que o produto da ordem de série de uma palavra multiplicado pela frequência de ocorrência é aproximadamente constante. É regida pela expressão matemática: $K = R \times F$, onde K = constante; R = ordem das palavras; F = frequência das palavras;
- segunda Lei de Zipf que enuncia que, em um determinado texto, várias palavras de baixa frequência de ocorrência (alta ordem de série) têm a mesma frequência.

Assim, a Lei de Zipf calcula uma constante em relação às frequências das palavras em um texto.

Por exemplo: se a palavra com maior frequência for citada mil vezes, a décima palavra da lista decrescente de frequências terá cem citações e a centésima palavra da lista terá dez citações.

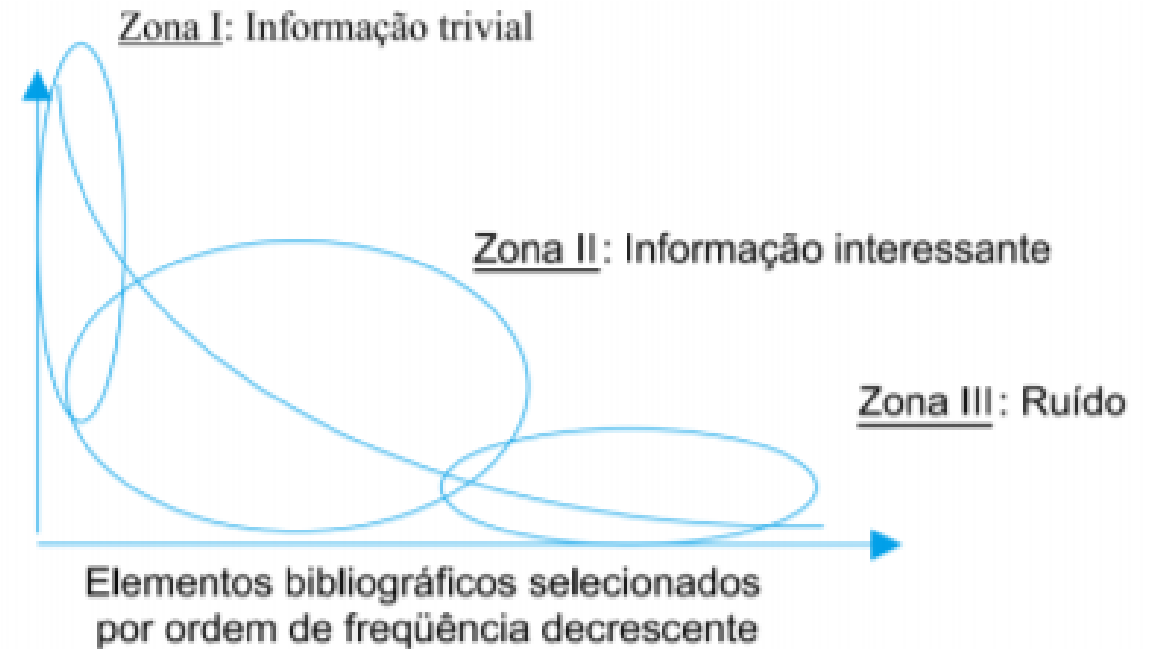
A Curva de Zipf



As Três Zonas da Lei de Zipf

Conforme classificação apresentada por Quoniam (1992), podem ser delimitadas na curva Zipf três zonas distintas, a saber:

- Zona I – informação trivial ou básica: define os temas centrais da análise bibliométrica;
- Zona II – informação interessante: localiza-se entre as Zonas I e III e mostra ora os temas periféricos, ora a informação potencialmente inovadora;
- Zona III – sinais fracos ou informação de ruído: essa zona tem como característica conter conceitos que podem ser emergentes (sinais fracos) ou apenas ruído estatístico.



Leitura e Discussão de Artigo:

25 anos de publicação em auditoria: análise bibliométrica sob o ponto de vista da Lei de Lotka, Lei de Zipf e Ponto de Transição (T) de Goffman

<https://forms.gle/jPq1Pe99tJ7TRwNZ9>

Vamos nos Separar em 3 Grupos

MELLO, Ivan Ribeiro et al. 25 anos de publicação em auditoria: análise bibliométrica sob o ponto de vista da Lei de Lotka, Lei de Zipf e Ponto de Transição (T) de Goffman. **Revista de Estudos Contábeis**, v. 8, n. 15, p. 45-65, 2018.

<http://www.uel.br/seer/index.php/rec/article/view/32197>

- 1- Grupo A Apresenta análise da Introdução e Conclusão
- 2- Grupo B Apresenta análise da Fundamentação Teórica e Referências
- 3- Grupo C Apresenta análise de Resultados