UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Análise de impacto de artigos em subcampos de periódicos científicos

Xiomara Sulvey Quispe Chacon

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Xiomara Sulvey Quispe Chacon

Análise de impacto de artigos em subcampos de periódicos científicos

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra em Ciências – Ciências de Computação e Matemática Computacional. *EXEMPLAR DE DEFESA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Diego Raphael Amâncio

USP – São Carlos
Fevereiro de 2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

C431a Chacon, Xiomara Sulvey Quispe
Análise de impacto de artigos em subcampos de
periódicos científicos / Xiomara Sulvey Quispe
Chacon; orientador Diego Raphael Amancio. -- São
Carlos, 2021.
68 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2021.

1. Sucesso de citação. 2. Fator de Impacto. 3.
Impacto de periódico. 4. Métricas de periódico. I.
Amancio, Diego Raphael, orient. II. Título.

Xiomara Sulvey Quispe Chacon

**Impact analysis of research articles in subfields of scientific
journals**

Dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Master in Science. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Diego Raphael Amâncio

USP – São Carlos
February 2021

RESUMO

XIOMARA, S. Q. CHACON. **Análise de impacto de artigos em subcampos de periódicos científicos**. 2021. 68 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

O fator de impacto tem sido amplamente utilizado nos últimos anos para avaliar a visibilidade e o prestígio dos periódicos científicos. Embora seja útil comparar periódicos, as especificidades da visibilidade dos subcampos em periódicos são negligenciados sempre que a visibilidade é medida apenas no nível do periódico. Neste trabalho, analisamos a visibilidade dos subcampos em um subconjunto de mais de 450.000 artigos de Física. Mostramos que a visibilidade dos subcampos não é regular no conjunto de dados considerado. Em anos específicos, a variabilidade no fator de impacto dos subcampos em um periódico atingiu 75% da média do fator de impacto dos subcampos. Também descobrimos que as diferenças de visibilidade de subcampos no mesmo periódico podem ser ainda maiores do que as diferenças de visibilidade entre diferentes periódicos. Nossos resultados mostram que o impacto dos subcampos é um fator importante para a visibilidade dos periódicos.

Palavras-chave: Sucesso de citação, Fator de Impacto, Impacto do periódico, Métricas do periódico.

ABSTRACT

XIOMARA, S. Q. CHACON. **Impact analysis of research articles in subfields of scientific journals**. 2021. 68 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

The impact factor has been extensively used in the last years to assess journals visibility and prestige. While it is useful to compare journals, the specificities of subfields visibility in journals are overlooked whenever visibility is measured only at the journal level. In this work, we analyze the subfields visibility in a subset of over 450,000 Physics papers. We show that the visibility of subfields is not regular in the considered dataset. In particular years, the variability in subfields impact factor in a journal reached 75% of the average subfields impact factor. We also found that differences of subfields visibility in the same journal can be even higher than differences of visibility between different journals. Our results show that subfields impact is an important factor accounting for journals visibility.

Keywords: Citation success, Impact Factor, Journal impact, Journal metrics.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de matriz de adjacência de um grafo não direcionado	18
Figura 2 – Exemplo de matriz de adjacência de um grafo direcionado	18
Figura 3 – Caminho mínimo de conexão (linha tracejada) entre os vértices i e j do grafo.	19
Figura 4 – Exemplo da medida de grau em redes não direcionadas e direcionadas.	20
Figura 5 – Exemplos de betweenness entre vértices de redes não direcionadas e direcionadas.	21
Figura 6 – Exemplo de um clique e triplo conectado em uma rede.	22
Figura 7 – O coeficiente de agrupamento fornece uma medida da interconectividade na vizinhança do vértice. Como exemplo, o vértice central da figura possui um coeficiente de agrupamento $C = 1$ se todos os seus vizinhos estiverem conectados e $C = 0$ se nenhuma interconexão estiver presente.	23
Figura 8 – Exemplo de uma rede exibindo estrutura de comunidade, com três grupos de nós com conexões internas densas e conexões mais esparsas entre grupos.	24
Figura 9 – Rede gerada pelo modelo binomial de Erdos e Renyi com $p = 0,01$	26
Figura 10 – Ilustração esquemática do modelo alternativo do <i>small-world</i>	26
Figura 11 – Evolução do modelo Barabási-Albert (BA) com base no conceito de “conexão preferencial” dos novos nós que se ingressam na rede.	27
Figura 12 – Exemplo de cálculo do Fator de Impacto.	31
Figura 13 – Definição do índice h	32
Figura 14 – Evolução do número de subcampos do periódico PRD, considerando o número de artigos e o número de citações.	44
Figura 15 – Evolução do número de subcampos, considerando o número de artigos e o número de citações.	45
Figura 16 – Evolução do fator de impacto para subcampos no periódico PRL.	46
Figura 17 – Evolução do índice de sucesso comparando o impacto de todos os pares de subcampos do periódico PRB.	47
Figura 18 – Evolução do índice de sucesso comparando o impacto de todos os pares de subcampos do periódico PRE.	48
Figura 19 – Evolução do índice de sucesso comparando o impacto de todos os pares de subcampos do periódico PRA.	49
Figura 20 – Histograma do índice de sucesso máximo de subcampos associados aleatoriamente a artigos do periódico PRA.	50

Figura 21 – Evolução do índice de sucesso comparando o impacto dos subcampos dos periódicos PRA e PRE.	51
Figura 22 – Evolução do índice de sucesso comparando o impacto dos subcampos dos periódicos PRC e PRB.	52
Figura 23 – Histograma do índice de sucesso máximo de subcampos associados aleatoria- mente a artigos dos periódicos PRA e PRE.	53

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Motivação e objetivos	14
1.2	Contribuições	14
1.3	Organização	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Matemática de redes	17
2.1.1	<i>Caminhos e conectividade</i>	19
2.2	Medidas de centralidade	20
2.2.1	<i>Grau</i>	20
2.2.2	<i>Betweenness</i>	20
2.3	Medidas estruturais	22
2.3.1	<i>Coeficiente de aglomeração</i>	22
2.3.2	<i>Modularidade</i>	24
2.4	Modelos de redes	25
2.4.1	<i>Erdős-Rényi</i>	25
2.4.2	<i>Watts-Strogatz</i>	25
2.4.3	<i>Barabási-Albert</i>	27
3	CONCEITOS PRÉVIOS	29
3.1	Cienciometria	29
3.2	Indicadores de impacto de citação	29
3.2.1	<i>Indicadores para periódicos</i>	31
3.2.2	<i>Indicadores para autores</i>	32
3.3	Códigos PACS	33
4	TRABALHOS RELACIONADOS	35
4.1	Prestígio de periódicos	35
4.2	Prestígio de autores	36
4.3	Dinâmica de citações	37
5	METODOLOGIA	39
5.1	Descrição dos dados	39
5.2	Identificação dos subcampos	39

5.2.1	<i>Diversidade</i>	40
5.3	Comparação do impacto dos subcampos	40
6	RESULTADOS E DISCUSSÃO	43
6.1	Estatísticas dos subcampos	43
6.2	Impacto dos subcampos	45
6.3	Comparação dos subcampos intra-periódicos	47
6.3.1	<i>Teste de robustez</i>	50
6.4	Comparação dos subcampos inter-periódicos	50
6.4.1	<i>Teste de robustez</i>	52
6.5	Conclusões do capítulo	52
7	CONCLUSÃO	55
7.1	Trabalhos futuros	56
7.2	Publicação	57
REFERÊNCIAS		59
APÊNDICE A	FIGURAS SUPLEMENTARES	63
APÊNDICE B	LISTA DE CÓDIGOS PACS	67

INTRODUÇÃO

A ciência é um conceito amplo que inclui uma variedade de atividades humanas, como biólogos observando bactérias usando microscópios, astrônomos olhando através de telescópios para observar galáxias e físicos resolvendo equações em um quadro negro. Essencialmente, é uma forma de conhecimento acumulado através do esforço do ser humano em compreender o universo em que vivemos. Diferentemente do passado, quando apenas um pequeno número de pessoas tinha a oportunidade de trabalhar como cientistas, a ciência moderna é avançada por muitos pesquisadores de diferentes disciplinas. E muito esforço está sendo feito para acelerar a publicação dos resultados da pesquisa. Centenas de periódicos foram criados na última década e milhares de artigos científicos são publicados todos os dias. Juntamente com o desenvolvimento da própria ciência, a ciência da ciência (Science of Science - SciSci) tornou-se um importante campo de pesquisa, que procura compreender, quantificar e prever a evolução futura da ciência (FORTUNATO *et al.*, 2018).

A visibilidade de artigos científicos, autores, periódicos e conferências tem sido usada nos últimos anos como um recurso importante para quantificar o impacto da pesquisa (WALTMAN, 2016). O número de citações tem sido um quantificador importante para medir o prestígio e, por esse motivo, muitas medições de impacto de pesquisa foram elaboradas com base na contagem de citações (REDNER, 1998). No nível do autor, por exemplo, o índice h tem sido amplamente usado como proxy para a visibilidade científica (BORNMANN; DANIEL, 2007). Um índice de citação importante para periódicos é o fator de impacto (FI), que fornece o número médio de citações recebidas por artigos em um periódico nos últimos 2 anos. Em muitos casos, os pesquisadores usam o fator de impacto do periódico (e outros atributos do periódico) para identificar o local mais relevante para divulgar suas pesquisas.

Apesar de que o fator de impacto tenha sido usado principalmente no nível do periódico (ALBERTS, 2013), as diferenças na visibilidade dos subcampos de um mesmo periódico têm sido, em grande parte, esquecidas. Considere, por exemplo, uma revista de interesse geral, publi-

cando em muitos subcampos diferentes em um periódico de alto impacto, alguns subcampos podem ter uma visibilidade menor, por outro lado, em um periódico de baixo impacto, subcampos específicos podem ser mais visíveis do que o periódico como um todo. As diferenças na visibilidade dos campos principais é um fenômeno bem conhecido (LI *et al.*, 2013). Essa é uma das razões pelas quais os conjuntos de dados acadêmicos (como a *Web of Science*) classificam os periódicos em diferentes assuntos. Por exemplo, a diferença nas citações acumuladas por artigos de Bioquímica e Matemática é normalmente muito grande (WALTMAN *et al.*, 2011). Tais diferenças têm motivado diversos estudos na área de normalização de campos de pesquisa.

Essa área de pesquisa desenvolveu alguns procedimentos de normalização para fornecer ferramentas e métodos para comparar artigos, periódicos ou cientistas de diferentes áreas. Uma medida de normalização bem conhecida é a *normalized citation score* (BORNMANN; MARX, 2015), definida como o número real de citações que um artigo recebeu normalizado pelo número esperado de citações que um artigo típico na área considerada recebe. Variações dessa medida também foram propostas para normalizar a contagem de citações de artigos e periódicos em diferentes campos importantes. Alguns estudos sugerem o uso de distribuições universais entre os campos para normalizar as contagens de citações (RADICCHI; FORTUNATO; CASTELLANO, 2008). Outros estudos sugerem uma abordagem mais simples que normaliza as contagens de citações por *z-scores* (LUNDBERG, 2007).

1.1 Motivação e objetivos

A motivação do presente trabalho de pesquisa é mostrar a existência de uma variabilidade significativa no impacto (visibilidade) dos subcampos em periódicos científicos ao longo do tempo e quais são os subcampos mais importantes para os periódicos, para que os pesquisadores juntamente com as informações do periódico possam tomar uma decisão mais informada sobre a visibilidade do tópico abordado em seu artigo. Apesar de que a maioria dos estudos atuais sobre a diferença de visibilidade de campos tenha se concentrado em subcampos principais, o objetivo principal é investigar o comportamento da visibilidade dos subcampos em periódicos científicos por meio de uma análise comparativa em dois níveis diferentes: dentro do mesmo periódico e em diferentes periódicos.

1.2 Contribuições

Alguns resultados interessantes foram encontrados para a análise realizada em 450.000 artigos de física publicados pela *American Physical Society* (APS).

- Primeiro, foi descoberto que existe uma variabilidade considerável no impacto dos subcampos em alguns casos. A variabilidade no fator de impacto dos subcampos pode chegar

até 75% do fator de impacto médio dos subcampos no periódico *Physical Review L* (PRL). Nos últimos anos, a variabilidade atingiu 50% do fator de impacto de PRL.

- Segundo, descobriu-se que as diferenças na visibilidade dos subcampos de um mesmo periódico podem ser maiores do que as diferenças na visibilidade dos periódicos. Esse resultado sugere que não só o local de publicação é um fator importante para prever a visibilidade dos artigos, mas também o subcampo dentro do periódico.

1.3 Organização

Este trabalho está organizado da seguinte forma: Os conceitos necessários para a compreensão do trabalho desenvolvido durante esta pesquisa são apresentados nos Capítulos 2 e 3. No Capítulo 4, apresentamos trabalhos relacionados com fatores que afetam as citações em periódicos. Também mencionamos alguns trabalhos que destacam o fato de que campos distintos têm comportamentos de citação distintos. A metodologia usada para comparar os subcampos é apresentado no Capítulo 5. O Capítulo 6 apresenta os resultados obtidos na comparação de subcampos de periódicos publicados pela APS. Finalmente, o Capítulo 7 discute as perspectivas para trabalhos futuros.

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresentamos os conceitos básicos da representação matemática de redes complexas na terminologia da teoria de grafos, um ramo da matemática que surgiu da prova de Euler (problema dos sete pontes de Königsberg); definimos também as medidas que caracterizam e quantificam a estrutura de uma rede complexa, incluindo os principais modelos de redes como Erdős-Rényi, Watts-Strogatz, Barabási-Albert.

2.1 Matemática de redes

Um grafo $G(V, E)$ é uma representação matemática das interações por pares entre $n \in \mathbb{Z}^+$ agentes individuais e é definido por um conjunto de *vértices* (ou nós) $V = \{1, 2, \dots, n\}$ e um conjunto de *arestas* (ou links) $E = \{(i, j); i, j \in V\}$. Um par (i, j) pertence a E se houver uma interação entre os agentes i e j e a cardinalidade do conjunto E é o número de arestas no grafo, e é denotada por $|E| = m$. Uma aresta (i, j) é chamada de *incidente* para ambos nós i e j que, neste caso, são denominados de *vizinhos*.

Um grafo pode ser representado univocamente por sua *matriz de adjacência* $A \in \{0, 1\}^{n \times n}$ de modo que

$$A(i, j) = \begin{cases} 1, & \text{se } (i, j) \in E, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.1)$$

A representação matricial de um grafo é particularmente útil porque permite analisar o grafo usando a teoria matricial e a álgebra linear.

Um grafo é chamado *não direcionado* se as arestas não tiverem orientação, ou seja, cada par $(i, j) \in E$ é considerado não-ordenado e $(i, j) = (j, i)$, como em redes de amizade (onde cada relacionamento é considerado recíproco). A matriz de adjacência associada a um grafo

não direcionado é simétrica, isto é $A = A^T$. Se a orientação das arestas é importante, o grafo é chamado de *direcionado* e uma aresta (i, j) tem uma origem i e um destino j . As redes de telecomunicação são um bom exemplo de rede direcionada, onde se faz uma distinção entre o chamador e o chamado. Em grafos direcionados, um vizinho j de um nó i é chamado de *filho* quando $(i, j) \in E$ e um *pai* quando $(j, i) \in E$. Observe que um vizinho pode ser filho e pai ao mesmo tempo. Uma aresta que conecta um nó a si mesmo, $(i, i) \in E$, é chamada de *auto-loop*.

As Figuras 1 e 2 apresentam a ilustração gráfica de grafos não direcionado e direcionado e suas matrizes de adjacência correspondentes.

Figura 1 – Exemplo de matriz de adjacência de um grafo não direcionado

Fonte: Elaborada pelo autor.

Figura 2 – Exemplo de matriz de adjacência de um grafo direcionado

Fonte: Elaborada pelo autor.

Estamos interessados nos subconjuntos de um grafo. Diz-se que um grafo $\tilde{G}(\tilde{V}, \tilde{E})$ é um *subgrafo* do grafo $G(V, E)$ se todos os vértices em $\tilde{V}$ pertencem a V e todas as arestas em $\tilde{E}$ pertencem a E , isto é, $\tilde{E} \subset E$ e $\tilde{V} \subset V$.

A caracterização de estruturas locais também está relacionada à identificação de *comunidades*. Em termos gerais, as comunidades são identificadas por subgrafos, onde os nós são altamente interconectados entre si e pouco conectados aos nós fora do subgrafo. A presença de subgrafos e

comunidades levanta a questão da modularidade nas redes. A definição de *modularidade* será apresentada na Subseção 2.3.2.

2.1.1 Caminhos e conectividade

Uma questão central na estrutura dos grafos é a *acessibilidade* dos vértices, ou seja, a possibilidade de ir de um vértice para outro seguindo as conexões dadas pelas arestas nas redes. Em uma rede conectada, todo vértice é acessível a partir de qualquer outro vértice. Os componentes conectados de um grafo definem muitas propriedades de sua estrutura física.

Para analisar as propriedades de conectividade, é definido um *caminho* P_{i_0, i_n} em um grafo $G(V, E)$ como uma coleção ordenada de $n + 1$ vértices $V_P = \{i_0, i_1, \dots, i_n\}$ e n arestas $E_P = \{(i_0, i_1), (i_1, i_2), \dots, (i_{n-1}, i_n)\}$, de modo que $i_\alpha \in V$ e $(i_\alpha, i_{\alpha+1}) \in E$, para todo α . Diz-se que o caminho P_{i_0, i_n} conecta os vértices i_0 e i_n . O *comprimento* do caminho P_{i_0, i_n} é n .

O conceito de “caminho” está na base da definição de distância entre vértices, de fato, enquanto os grafos geralmente não possuem métricas, a medida da distância natural entre dois vértices i e j é definida como o número de arestas atravessadas pelo caminho de conexão mais curto (veja Figura 3). Essa distância é chamada de *comprimento do caminho mais curto* e denotada como d_{ij} . Podemos ter múltiplos caminhos menores com o mesmo comprimento d_{ij} entre um par de vértices. O caminho mais curto nunca contém loops ou é interceptado.

Figura 3 – Caminho mínimo de conexão (linha tracejada) entre os vértices i e j do grafo.

Fonte: Elaborada pelo autor.

Em uma rede não direcionada $d_{ij} = d_{ji}$, isto é, a distância entre os vértices i e j é a mesma que a distância entre os vértices j e i . Em uma rede direcionada, muitas vezes $d_{ij} \neq d_{ji}$. Além disso, em uma rede direcionada a existência de um caminho do vértice i ao vértice j não garante a existência de um caminho de j para i .

$$d = \frac{1}{n(n-1)} \sum_{i,j} d_{ij}. \quad (2.2)$$

Outra definição efetiva do tamanho linear da rede é o *comprimento médio do caminho mais curto* (average shortest path length), definido como o valor médio de d_{ij} em todos os pares

possíveis de vértices na rede e pode ser calculado pela Equação 2.2.

2.2 Medidas de centralidade

As medidas de centralidade das redes complexas são funções que designam um valor numérico aos nós da rede conforme uma noção de prestígio pré-determinada (NEWMAN, 2018).

2.2.1 Grau

O grau k_i de um vértice i é definido como o número de arestas conectadas ao vértice i . Essa definição é clara para grafos não direcionados, já para grafos direcionados cada vértice tem dois graus, o grau de entrada $k_{in,i}$ do vértice i é definido como o número de arestas que chegam a i , enquanto o grau de saída $k_{out,i}$ é definido como o número de arestas que partem de i . A Figura 4 mostra exemplos do cálculo do grau para a rede não direcionada e para a rede direcionada

Figura 4 – Exemplo da medida de grau em redes não direcionadas e direcionadas.

Fonte: Elaborada pelo autor.

O grau de um vértice em um grafo direcionado é definido pela soma do grau de entrada e de saída, $k_i = k_{in,i} + k_{out,i}$. Em termos da matriz de adjacência o grau pode ser escrito como

$$k_{in,i} = \sum_j a_{ji}, \quad k_{out,i} = \sum_j a_{ij}. \quad (2.3)$$

Para um grafo não direcionado com uma matriz de adjacência simétrica, $k_{in,i} = k_{out,i}$. O grau de um vértice tem uma interpretação imediata em termos de centralidade, quantificando quão bem um elemento está conectado a outros elementos no grafo.

2.2.2 Betweenness

Em um grafo existem vértices que podem ser essenciais para conectar diferentes regiões da rede, agindo como pontes, assim como é possível que em uma rede não direcionada dois

caminhos mais curtos entre o mesmo par de vértices se sobreponham, compartilhando alguns vértices além dos vértices inicial e final (veja Figura 5a).

Para explicar quantitativamente o papel de tais vértices, o conceito de *betweenness* é definido como um número de caminhos mais curtos entre pares de vértices que passam por um determinado vértice (FREEMAN, 1977; NEWMAN, 2005). Mas precisamente, se σ_{hj} é o número total de caminhos mais curtos de h a j e $\sigma_{hj}(i)$ é o número desses caminhos mais curtos que passam pelo vértice i , o *betweenness* do vértice i é definido como

$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}} \quad (2.4)$$

O *betweenness* é uma medida que quantifica a influência de um nó na difusão de informações da rede, e também é aplicada a redes direcionadas. Em uma rede direcionada, o caminho mais curto entre dois vértices depende, geralmente, da direção em que o vértice se conecta com os outros vértices. O caminho mais curto de h a j é diferente do caminho mais curto de j a h , de fato pode haver um caminho em uma direção e nenhum caminho no outro (veja Figura 5b).

Portanto, é importante em uma rede direcionada incluir explicitamente as contagens de caminho em qualquer direção entre cada par de vértices. A Equação 2.4 pode ser utilizada sem modificação para o caso direcionado.

Figura 5 – Exemplos de *betweenness* entre vértices de redes não direcionadas e direcionadas.

Fonte: Elaborada pelo autor.

2.3 Medidas estruturais

2.3.1 Coeficiente de aglomeração

O coeficiente de aglomeração (*clustering coefficient*) quantifica como o subgrafo formado pela vizinhança de um vértice se aproxima de um *clique* (i.e., um subgrafo totalmente conectado). Um exemplo de clique é mostrado na Figura 6a).

O tipo fundamental de relação entre vértices em uma rede é “conectado por uma borda”, se essa relação for transitiva, isso implica a propriedade de que, se o vértice i estiver conectado ao vértice j , e ao mesmo tempo j estiver conectado a k , com alta probabilidade i também será conectado a k (veja Figura 6b).

Figura 6 – Exemplo de um clique e triplo conectado em uma rede.

Fonte: Elaborada pelo autor.

Uma primeira definição de coeficiente de aglomeração, também conhecida como transitividade, baseia-se na seguinte definição para redes não ponderadas e não direcionadas.

$$T = \frac{3N_{\Delta}}{N_3}, \quad (2.5)$$

onde N_{Δ} é o número de triângulos na rede e N_3 é o número de triplos conectados. Um *triplo conectado* é um conjunto de três vértices i, j, k com arestas (i, j) e (j, k) , a aresta (i, k) pode estar presente ou não. O fator três no numerador surge porque cada triângulo é contado três vezes quando contamos os triplos conectados na rede e garante que $0 \leq T \leq 1$; Um *triângulo* é um conjunto de três vértices com arestas entre cada par de vértices. Assim N_{Δ} e N_3 são definidos como

$$N_{\Delta} = \sum_{k>j>i} a_{ij}a_{ik}a_{jk} \quad \text{e} \quad N_3 = \sum_{k>j>i} (a_{ij}a_{ik} + a_{ji}a_{jk} + a_{ki}a_{kj}) \quad (2.6)$$

onde os a_{ij} são os elementos da matriz de adjacência A e a soma é tomada sobre todos os triplos de vértices distintos i, j e k somente uma vez.

Uma segunda definição do coeficiente de aglomeração de um determinado vértice i é dada por:

$$C_i = \frac{N_{\Delta}(i)}{N_3(i)} \quad (2.7)$$

onde $N_{\Delta}(i)$ é o número de triângulos envolvendo vértice i e $N_3(i)$ é o número de triplos conectados tendo i como o vértice central:

$$N_{\Delta}(i) = \sum_{k>j} a_{ij}a_{ik}a_{jk} \quad e \quad N_3(i) = \sum_{k>j} a_{ij}a_{ik}. \quad (2.8)$$

Se k_i é o número de vizinhos do vértice i , então $N_3(i) = k_i(k_i - 1)/2$. $N_{\Delta}(i)$ conta o número de arestas entre os vizinhos de i . Representando o número de arestas entre vizinhos do vértice i como l_i , a Equação 2.7 pode ser reescrita como:

$$C_i = \frac{2l_i}{k_i(k_i - 1)}. \quad (2.9)$$

A definição acima, o coeficiente de aglomeração de um vértice i de uma rede é sempre menor ou igual a um, isto é $0 \leq C_i \leq 1$, uma vez que podem existir no máximo $k_i(k_i - 1)/2$ triângulos e no mínimo nenhum triângulo em torno de um vértice i da rede (COSTA *et al.*, 2007).

Na Figura 7, fornecemos uma ilustração de alguns exemplos simples de agrupamento de vértices com uma determinada vizinhança.

Figura 7 – O coeficiente de agrupamento fornece uma medida da interconectividade na vizinhança do vértice. Como exemplo, o vértice central da figura possui um coeficiente de agrupamento $C = 1$ se todos os seus vizinhos estiverem conectados e $C = 0$ se nenhuma interconexão estiver presente.

Fonte: Elaborada pelo autor.

2.3.2 Modularidade

A modularidade em uma rede é determinada pela existência de subgrafos específicos, chamados de *módulos* (ou comunidades), que são formados por conjuntos de vértices que possuem propriedades comuns, esses vértices são fortemente conectados dentro de suas comunidades e pouco conectados aos vértices de outras comunidades. Esta medida foi proposta por Girvan & Newman em (NEWMAN, 2006) com o intuito de caracterizar o quão boa era uma estrutura de comunidade em uma rede. Assim a *modularidade* é definida como

$$Q = \frac{1}{2m} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j), \quad (2.10)$$

onde m é a quantidade de arestas, a_{ij} é um elemento da matriz de adjacência A , k_i e k_j são os graus dos vértices i e j , respectivamente, c_i e c_j são as comunidades as quais os vértices i e j pertencem e $\delta(c_i, c_j)$ (delta de Kronecker) é 1 se i e j estão na mesma comunidade, e é 0 caso contrário.

Na definição acima, observamos que a fixação dos módulos da rede é feita *a priori*, assim a medida de modularidade em uma rede depende do seu conjunto de módulos e esta medida se modifica com a alteração destes módulos. Destacamos que a modularidade de uma partição é um valor escalar entre -1 e 1 que mede a densidade dos links dentro das comunidades em comparação aos links entre comunidades (CLAUSET; NEWMAN; MOORE, 2004). Um exemplo esquemático de comunidades em uma rede é apresentado na Figura 8.

Figura 8 – Exemplo de uma rede exibindo estrutura de comunidade, com três grupos de nós com conexões internas densas e conexões mais esparsas entre grupos.

Fonte: Elaborada pelo autor.

2.4 Modelos de redes

2.4.1 Erdős-Rényi

No século XX foram introduzidas na literatura o que hoje conhecemos como redes complexas as quais modelam sistemas em grafos. O primeiro modelo foi sugerido por Erdős e Rényi em 1959 e que ao longo dos anos ficou conhecido por grafos aleatórios ou modelos *ER*. O modelo possui dois parâmetros, o número total de vértices denotado por n e uma probabilidade p , no qual, cada par de vértices é conectado de acordo com essa probabilidade independente dos outros vértices.

Dado um vértice de um grafo aleatório. Ele está conectado com a mesma probabilidade p com cada um dos outros $(n - 1)$ vértices no grafo e, portanto, a probabilidade $p(k)$ de ter o grau exatamente k é dada pela distribuição binomial:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}. \quad (2.11)$$

Vemos que o grau médio de um vértice na rede é $z = (n - 1)p$, também podemos escrever isso como

$$P(k) = \binom{n-1}{k} \left[\frac{z}{n-1-z} \right]^k \left[1 - \frac{z}{n-1} \right]^{n-1} \simeq \frac{z^k}{k!} e^{-z}, \quad (2.12)$$

onde a última igualdade aproximada se torna exata no limite de n grande, isto é, redes aleatórias com grande número de vértices e um grau médio constante seguem uma distribuição de Poisson (NEWMAN; WATTS; STROGATZ, 2002). Mostramos um exemplo do modelo na Figura 9.

2.4.2 Watts-Strogatz

No ano 1998, Duncan Watts e Steven Strogatz, pesquisadores das Universidade de Columbia e Cornell respectivamente, mostraram que era possível ter redes complexas com coeficiente de clustering alto ao mesmo tempo em que mantinham o efeito *small-world*, isso significa que a presença de ciclos de ordem três (triângulos) em redes reais eram muito maiores do que no modelo aleatório proposto por Erdős e Rényi. Esse fato mostrou algo muito interessante: as redes reais, na verdade, não são completamente aleatórias e que por tanto existe algum tipo de lei de formação por trás da construção destas, dito modelo é conhecido como *small-world* de Watts-Strogatz ou simplesmente modelo *WS*.

O procedimento para obtê-lo é o seguinte: começar com uma rede regular de tamanho n com k vizinhos, logo cada aresta tem que ser redirecionada para qualquer outro vértice seguindo uma probabilidade p mas tendo em conta o seguinte: qualquer vértice não pode se ligar com ele

Figura 9 – Rede gerada pelo modelo binomial de Erdos e Renyi com $p = 0,01$

Fonte: Commons (2020).

mesmo, além disso, não pode haver mais de uma conexão entre um par qualquer de vértices da rede. Na Figura 10 mostramos este procedimento, e vemos que a rede com $p = 0$ é regular com arestas entre os vizinhos mais próximos e os próximos mais próximos, por tanto a conectividade média é $\langle k \rangle = 4$; a medida que o valor de p aumenta a rede fica cada vez mais desordenada, até que para $p = 1$ todas as arestas são religadas aleatoriamente (WATTS; STROGATZ, 1998).

Figura 10 – Ilustração esquemática do modelo alternativo do *small-world*.

Fonte: Elaborada pelo autor.

2.4.3 Barabási-Albert

Foi desenvolvido por Albert-László Barabási e Réka Albert em 1999, o modelo Barabási-Albert, resulta em redes *livre de escala* é também conhecido como modelo *BA*, é definido da seguinte forma:

Figura 11 – Evolução do modelo Barabási-Albert (*BA*) com base no conceito de “conexão preferencial” dos novos nós que se ingressam na rede.

Fonte: Elaborada pelo autor.

A rede começa com uma rede conectada inicial de m_0 nós, cujos links são escolhidos arbitrariamente, desde que cada nó tenha pelo menos um link. A rede desenvolve-se seguindo duas etapas:

- **Crescimento:** Novos nós são adicionados à rede, um de cada vez. Cada novo nó está conectado aos m ($\leq m_0$) nós existentes com uma probabilidade proporcional ao número de links que os nós existentes já possuem.
- **Ligação preferencial:** Os novos nós tendem a se vincular aos nós mais conectados. Formalmente, a probabilidade $\Pi(k)$ de que o novo nó se conecte ao nó i depende do grau k_i como

$$\Pi(k_i) = \frac{k_i}{\sum_i k_i} \quad (2.13)$$

Uma *ligação preferencial* é um mecanismo probabilístico: um novo nó é livre para conectar-se a qualquer nó da rede, seja um *hub* (nós fortemente conectados) ou um link único (BARABÁSI; PÓSFAL, 2016).

Os dois mecanismos, crescimento e a ligação preferencial desempenham um papel particularmente importante na formação da distribuição de grau de uma rede. Assim a distribuição de grau resultante do modelo BA é livre de escala, em particular, é uma lei de potência da forma

$$P(K) \sim k^{-3} \quad (2.14)$$

Para entender o comportamento do modelo BA apresentamos um exemplo na Figura 11, a sequência de imagens mostra nove etapas subsequentes do modelo BA. Os nós de cor amarelo marcam o nó recém-adicionado à rede, que decide onde conectar seus dois links ($m = 2$) usando ligação preferencial. Após t etapas do tempo, o modelo BA gera uma rede com $n = t + m_0$ nós e $m_0 + mt$ links.

CONCEITOS PRÉVIOS

Neste capítulo explicamos os principais indicadores de impacto baseados em citações e a classificação hierárquica de artigos científicos de física e astronomia (PACS).

3.1 Cienciometria

A cienciometria analisa os aspectos quantitativos da geração, propagação e utilização de informações científicas, a fim de contribuir para uma melhor compreensão do mecanismo das atividades de pesquisa científica. As principais questões de pesquisa incluem a medição do impacto de trabalhos de pesquisa e periódicos acadêmicos, por meio de textos (artigos) como unidades empíricas de análise e as citações como unidade de impacto da informação científica (LEYDESDORFF; MILOJEVIĆ, 2012).

3.2 Indicadores de impacto de citação

Os indicadores de impacto de citações são indicadores de impacto científico baseados na análise das citações recebidas por publicações científicas. Os indicadores de impacto de citações podem fornecer informações sobre o impacto de publicações individuais, porém, mais frequentemente, proporcionam informações sobre o impacto de unidades de pesquisa, como pesquisadores, grupos de pesquisa, instituições de pesquisa, países ou periódicos (WALTMAN, 2016).

Por que estudar citações?

As medidas de citação podem ser tão simples como contar o número de citações recebidas por um autor, uma publicação ou um revista durante um determinado período de tempo e são uma ferramenta eficaz para comparar a produtividade da pesquisa e o impacto entre os autores,

instituições ou países. As citações são tratadas como reconhecimentos uniformemente positivos da contribuição feita pelo autor ou trabalho citado, de modo que quanto maior o número de citações que um autor, trabalho ou revista recebe, maior é o reconhecimento. O ato de citar implica um desejo por parte do autor de incluir uma dada referência em seu artigo e oferece informações significativas sobre a influência científica. O fato de citar um estudo também afetará a relação entre o artigo citado e o de citação, por exemplo, argumentando contra as descobertas do trabalho citado. Assim, a análise de citações é frequentemente usada para obter informações sobre o impacto e, mais frequentemente, a qualidade associada a um autor, uma publicação ou um país. No entanto, a qualidade de uma publicação, por exemplo, não deve se basear apenas no número de vezes que é citada, mas também em outras análises (WOLFRAM, 2003).

Vida de uma Citação

Uma das regularidades no comportamento da ciência é a obsolescência da literatura, que pode ser interpretado como o declínio no uso de documentos ao longo do tempo. Naturalmente, isso está relacionado à análise de citações, já que o uso de um documento é medido pelo número de citações que recebe. Quando um documento não é mais citado, ele se tornará obsoleto. Os padrões de citação podem variar enormemente dependendo dos documentos estudados (PRICE, 1963).

É possível que uma grande parte dos documentos nunca seja citada, enquanto outros receberão citações nos anos imediatamente após a publicação antes de se tornarem obsoletos. Também é possível que alguns documentos permaneçam incitados ou raramente citados nos anos seguintes à sua publicação, mas sejam posteriormente reconhecidos. Isso porque, com o tempo, mais pesquisadores se conscientizam do valor do documento e, assim, o número de citações aumentará. Relacionado à obsolescência da literatura, também é comum na análise de citações estudar a idade das citações. Isso é avaliado pelo conhecimento do ano da citação em comparação com o ano de publicação do documento.

Auto-citações do autor

A autocitação é uma prática comum em que os autores citam seu próprio trabalho publicado anteriormente. A autocitação ocorre sempre que o conjunto de co-autores do trabalho de citação e o do citado compartilham pelo menos um autor. No entanto, os dados referentes às autocitações devem ser tratados com cautela, pois a confiabilidade dessa análise pode ser afetada por homônimos e, portanto, por uma contagem errônea de autocitações. Ele também pode ser afetado por variações ortográficas ou erros de ortografia do nome de um autor, o que leva à autocitação não reconhecida (WOLFGANG; BART; BALÁZS, 2004).

Não há consenso sobre o papel das autocitações do autor na literatura. Alguns autores afirmam que as autocitações são um meio potencial de inflacionar artificialmente as taxas de

citação e, portanto, a própria posição do autor dentro da comunidade científica. No entanto, outros consideram que as autocitações são uma parte natural da comunicação científica.

3.2.1 Indicadores para periódicos

Fator de impacto (FI)

O fator de impacto (FI) de uma revista científica é uma medida que reflete o número médio anual de citações de artigos recentes publicados nessa revista. Foi proposto por Eugene Garfield em Garfield (1955), e utilizado como instrumento de avaliação da qualidade das publicações a partir dos anos sessenta.

O Fator de Impacto (FI) relaciona elementos de dois conjuntos com unidades diferentes: a citação e o artigo, por tanto o FI pode ser considerado o primeiro índice de citação, caracterizando as revistas científicas e relacionando o número de citações obtidas por um conjunto de artigos em uma revista ao número correspondente de artigos publicados nessa revista. Atualmente o fator de impacto é o indicador mais frequentemente usado em cienciometria, ajuda aos cientistas a estudar os processos de informação em ciências e a avaliar a estratégia de publicação e o impacto dos resultados científicos de indivíduos, equipes ou países.

Para o cálculo do Fator de Impacto leva-se em consideração o número de citações recebidas pelos artigos publicados em uma revista em determinado ano, dividido pelo número de artigos publicados na mesma revista nos dois anos anteriores (veja Figura 12).

Figura 12 – Exemplo de cálculo do Fator de Impacto.

Fonte: Elaborada pelo autor.

É calculado anualmente pelo *Institute for Scientific Information* (ISI) para as revistas indexadas em sua base de dados, e publicado por *Journal Citations Reports* (JCR).

3.2.2 Indicadores para autores

Índice Hirsch

O índice de Hirsch, mais comumente conhecido como índice h , foi proposto em [Hirsch \(2005\)](#) para quantificar a produção científica de um indivíduo. Especificamente, ele tentou quantificar o impacto cumulativo e a relevância da produção científica de um indivíduo. O índice é definido da seguinte forma: um cientista tem um índice h se: h de seus N artigos tiverem pelo menos h citações cada um; e os outros $(N - h)$ artigos têm no máximo h citações cada um (veja Figura 13).

Figura 13 – Definição do índice h .

Fonte: Adaptada de [Hirsch \(2005\)](#).

A boa recepção deste índice na comunidade científica deveu-se, em parte, ao trabalho de [Ball \(2005\)](#), [Holden \(2005\)](#) nos prestigiosos periódicos Nature e Science, respectivamente.

Os dados necessários para calcular o índice h são simples e fáceis de obter, eles são: os artigos publicados por um determinado autor durante sua carreira e o número de vezes que esses artigos foram citados. Ao classificar os artigos em ordem decrescente de acordo com o número de citações recebidas, podemos identificar o índice h como a classificação mais alta, de modo que as primeiras h publicações recebam pelo menos h citações ([ANDRÉS, 2009](#)).

3.3 Códigos PACS

O Physics and Astronomy Classification Scheme (PACS) é uma classificação hierárquica de artigos científicos de física e astronomia, com um propósito útil, permitindo:

- Aos pesquisadores e autores identificarem os periódicos que publicam ativamente nas áreas da física e astronomia.
- Aos periódicos identificarem as áreas em que professores e pesquisadores da comunidade da física são capazes de servir como referências e revisores em áreas específicas e especializadas da física e astronomia ([RABINOWITZ](#); [VOGEL](#), 2009).

Um código PACS compreende três elementos: um par de números de dois dígitos separados por um ponto seguidos por dois caracteres que podem ser letras (minúsculas ou maiúsculas) ou sinais (“-” ou “+”). No primeiro par de números do código, o primeiro dígito representa a categoria principal e o segundo dígito é o campo mais específico dentro dessa categoria, representando assim o primeiro e segundo nível de hierarquia respectivamente. O segundo par de números especifica uma categoria mais restrita (subcampos no terceiro nível de hierarquia) dentro do campo dado pelos primeiros dois dígitos. Os dois últimos caracteres do código fornecem uma caracterização ainda mais específica dos subcampos.

Por exemplo, a classificação do código PACS 42 . 30 . -d refere-se:

"40": “Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics” (categoria principal);

"42": “Optics” (campo);

"30": “Imaging and optical processing” (subcampo);

"-d": a presença de subcampos mais especializados é indicado pelo sinal "-"

- *Fourier optics*
- *Modulation and optical transfer functions*
- *Speckle and moiré patterns*
- *Phase retrieval*
- *Pattern recognition*
- *Image forming and processing*
- *Image reconstruction; tomography*

Uma lista dos subcampos mencionados neste trabalho é mostrada no Apêndice [B](#).

TRABALHOS RELACIONADOS

Este capítulo apresenta uma visão geral dos estudos focados na descoberta de padrões que contribuem e influenciam a visibilidade dos artigos científicos.

4.1 Prestígio de periódicos

Trabalho de Larivière e Gingras (2010)

Vários estudos empíricos tentaram medir a presença do efeito Matthew no nível de artigos, pesquisadores individuais, instituições ou países. No entanto, esses estudos raramente controlam a “qualidade” intrínseca dos artigos ou dos pesquisadores - artigos ou pesquisadores “melhores” (independentemente de sua definição) poderiam receber taxas de citação mais altas porque, na verdade, são de melhor qualidade.

Usando um método original para controlar o valor intrínseco dos artigos - artigos duplicados idênticos publicados em diferentes periódicos com diferentes fatores de impacto - Em Larivière e Gingras (2010) é mostrado que o periódico no qual os artigos são publicados exerce forte influência em suas taxas de citação, pois artigos duplicados publicados em periódicos de alto impacto obtêm, em média, duas vezes mais (o dobro) citações do que suas contrapartes idênticas publicadas em periódicos com menor fator de impacto. O valor intrínseco de um artigo, portanto, não é a única razão pela qual um determinado artigo é citado ou não; há um efeito Matthew específico associado aos periódicos e isso dá aos artigos publicados um valor agregado além de sua qualidade intrínseca.

Trabalho de Abramo, D'Angelo e Costa (2010)

Nos últimos anos, os bibliometristas têm prestado cada vez mais atenção aos problemas metodológicos da avaliação da pesquisa, entre eles a escolha dos indicadores mais adequados para avaliar a qualidade das publicações científicas e, portanto, para avaliar o trabalho de cientistas

individuais, grupos de pesquisa e organizações inteiras. Muita literatura tem sido dedicada à análise da robustez de vários indicadores, e muitos trabalhos alertam contra os riscos do uso de proxies facilmente disponíveis e relativamente simples, como o fator de impacto do periódico.

Em [Abramo, D'Angelo e Costa \(2010\)](#) é examinando se é válido que o uso do Fator de Impacto deva ser sempre evitado em favor de citações, ou se o uso do fator de impacto poderia ser aceitável, até mesmo preferível, em determinadas circunstâncias. A avaliação foi realizada observando todas as publicações científicas nas ciências exatas por universidades italianas, para o período de 2004-2007. As análises de sensibilidade de desempenho foram conduzidas com indicadores de mudança de qualidade e anos de observação.

Trabalho de [Bornmann e Leydesdorff \(2015\)](#)

Os modelos de regressão (binomial) aplicados em [Bornmann e Leydesdorff \(2015\)](#) utilizaram como variáveis independentes os indicadores bibliométricos: Fator de Impacto do Periódico (meio de publicação), número de autores, número de páginas e a citação estatística das referências utilizadas no artigo científico, já como indicadores de qualidade os julgamentos por pares. As citações em janelas de um a dez anos com a citação anual e as citações cumulativas como variáveis dependentes. Esses modelos mostraram que o Fator de Impacto do Periódico tem uma influência maior no impacto de citação de uma publicação do que a qualidade (medida por julgamentos de pares). No entanto, o número de páginas e a qualidade das referências são menos influentes. A influência do Fator de Impacto atinge o pico após três anos e, em seguida, diminui (na maioria das análises de regressão), mas permanece mais alta do que a influência dos julgamentos de qualidade mesmo após dez anos.

O Fator de Impacto do Periódico e o número de autores mostraram uma (forte) correlação com a contagem de citações de um artigo: quanto maior a reputação de um periódico e quanto mais autores um artigo tiver, mais forte será o aumento esperado no impacto do artigo, especialmente dois a três anos após sua publicação. Neste trabalho é mencionado que o Fator de Impacto pode ser considerado uma média móvel de dois anos, de modo que melhor indica o impacto de curto prazo, já o número de autores está associado a um efeito de público que também pode desaparecer após alguns anos. Esses resultados questionam a discrepância entre os indicadores baseados em algoritmos e os julgamentos qualitativos dos especialistas. O último parece menos preditivo para citações futuras do que uma combinação de construções algorítmicas.

4.2 Prestígio de autores

Trabalho de [Amjad et al. \(2017\)](#)

Jovens acadêmicos freqüentemente procuram trabalhar em colaboração com pesquisadores de ponta em seu campo na busca por uma carreira de sucesso. Embora o sucesso na academia

possa ser definido de forma diferente, todos concordam que o treinamento com um pesquisador conhecido pode ajudar a levar a uma carreira eficaz. O objetivo de estudo de [Amjad et al. \(2017\)](#) foi investigar se a colaboração com cientistas estabelecidos melhora, de fato as chances de sucesso dos acadêmicos iniciantes. Assim como também analisar o efeito da colaboração com uma estrela (com um pesquisador de sucesso) conhecida no sucesso de um jovem acadêmico. Os resultados sugerem que trabalhar com os principais especialistas pode levar a uma carreira de sucesso, mas não é o único caminho. Os pesquisadores que não tiveram a sorte de começar sua carreira com um pesquisador de elite ainda podiam ter sucesso com trabalho árduo e paixão. Essas descobertas surgiram de análises de dois grupos distintos de estudiosos bem conhecidos sobre a carreira de recém-chegados, sugerindo sua força e validade.

4.3 Dinâmica de citações

Trabalho de [Newman \(2001\)](#)

Em [Newman \(2001\)](#) é estudado empiricamente a evolução temporal das redes de colaboração científica em física e biologia. Nessas redes, dois cientistas são considerados conectados se forem co-autores de um ou mais artigos. Foi mostrado que a probabilidade de um par de cientistas que colaboram aumenta com o número de outros colaboradores que têm em comum, e que a probabilidade de um determinado cientista adquirir novos colaboradores aumenta com o número de seus antigos colaboradores. Esses resultados fornecem evidência experimental em favor de mecanismos previamente conjecturados para agrupamento e distribuições de grau de lei de potência em redes.

Trabalho de [Jeong, Néda e Barabási \(2003\)](#)

A hipótese de que nós altamente conectados aumentam sua conectividade mais rápido do que seus pares menos conectados é um fenômeno chamado de apego preferencial. Em [Jeong, Néda e Barabási \(2003\)](#), as medições em quatro redes complexas, a saber, rede de citações científicas, Internet, colaboração de atores e rede de coautoria científica, indicam que a taxa em que os nós adquirem conexões depende do grau do nó, oferecendo suporte quantitativo direto para a presença do apego preferencial. Descobrimos assim, que para os dois primeiros sistemas a taxa de apego preferencial depende linearmente do grau do nó, enquanto para os dois últimos a dependência segue uma lei de potência sublinear.

Trabalho de [Petersen et al. \(2011\)](#)

O efeito Matthew refere-se ao ditado escrito há cerca de dois mil anos no Evangelho de São Mateus: “Pois a todos aqueles que têm, mais será dado”. Mesmo dois milênios depois, esse idioma é usado pelos sociólogos para descrever qualitativamente a dinâmica do progresso indivi-

dual e a interação entre status e recompensa. Os estudos quantitativos de carreiras profissionais são tradicionalmente limitados pela dificuldade em medir o progresso e pela falta de dados sobre carreiras individuais. No entanto, em algumas profissões, existem métricas bem definidas que quantificam a longevidade, o sucesso e a capacidade da carreira, que juntas contribuem para a classificação geral de sucesso de um funcionário individual.

Em [Petersen *et al.* \(2011\)](#) são demonstradas evidências testáveis do antigo efeito Matthew de “O rico fica mais rico” em que a longevidade e o sucesso passado de um indivíduo levam a uma vantagem cumulativa no desenvolvimento posterior de sua carreira. O modelo de progresso de carreira estocástico exatamente solucionável incorpora quantitativamente o efeito Matthew e valida as previsões de modelo para várias profissões competitivas. O modelo é testado nas carreiras de 400.000 cientistas usando dados de seis periódicos de alto impacto, as descobertas foram confirmadas testando o modelo nas carreiras de mais de 20.000 atletas em quatro ligas esportivas. O modelo destaca a importância do desenvolvimento no início da carreira, mostrando que muitas carreiras são atrofiadas pela desvantagem relativa associada à inexperience.

METODOLOGIA

5.1 Descrição dos dados

No presente trabalho é utilizado o conjunto de dados fornecido pela American Physical Society (APS). As revistas da APS são os veículos de publicação mais comuns em física e cobrem todos os subcampos desta disciplina; eles portanto representam uma referência ideal para o estudo de padrões de citação de publicações dentro da física. Este conjunto de dados fornece citações para as revistas da APS como Physical Review Letters (PRL), Review of Modern Physics (RMP) e Physical Review A - E (PRA - E). Para nossa análise usamos apenas os dados de citação, nome da revista, código PACS e data de publicação de 450.000 artigos correspondentes ao período de 1984 a 2015, porém o conjunto de dados da APS também fornece informações adicionais, como DOI, título, nomes e afiliações dos autores.

5.2 Identificação dos subcampos

Embora não haja uma forma única de definir um campo ou disciplina, aqui usamos os códigos PACS (definidos na Seção 3.3) proporcionados pelos autores como uma representação de subcampos, a estrutura hierárquica dos códigos PACS permitem identificar progressivamente áreas e subcampos de pesquisa mais específicos, pelo que o PACS é o melhor indicador. Em particular focamos esta análise no terceiro nível hierárquico composto pelos quatro primeiros dígitos do código, este é um nível intermediário que permite analisar subcampos que não são nem muito gerais nem muito específicos.

Uma vez que códigos diferentes podem ocorrer simultaneamente em um determinado artigo identificamos até 5 códigos PACS por artigo, mas apenas aqueles que correspondem a subcampos diferentes no terceiro nível de hierarquia são considerados. Por exemplo, um dado artigo possui os códigos PACS [05.45.-a, 05.45.Ac, 05.45.Df, 05.40.Jc, 05.40.Fb] mas atribuiremos apenas os códigos 05.45 e 05.40 ao artigo.

5.2.1 Diversidade

Para capturar o número efetivo de subcampos relevantes considerando a distribuição dos artigos publicados nos subcampos e das citações recebidas pelos subcampos, calculamos a diversidade de subcampos por meio da entropia definida na Equação 5.1 (JR *et al.*, 2017):

$$D = \exp \left(- \sum_{i=1}^R \pi_i \ln \pi_i \right), \quad (5.1)$$

onde R é o número total de tipos de subcampos relevantes no conjunto de dados e π_i é a abundancia proporcional do i -ésimo tipo. Consideramos um subcampo relevante se pelo menos 50 artigos foram publicados nessa área nos últimos dois anos.

5.3 Comparação do impacto dos subcampos

A diferença na visibilidade dos subcampos foi calculada usando o chamado índice de sucesso de citação definido em Milojević, Radicchi e Bar-Ilan (2017). Esta medida fornece uma interpretação clara das diferenças no fator de impacto (e distribuição de citação) entre quaisquer dois grupos de artigos. O índice de sucesso S_{tr} foi projetado para comparar quantitativamente o sucesso de dois periódicos, mas o mesmo conceito pode ser estendido para comparar qualquer conjunto de artigos.

Dado dois conjuntos de artigos, os conjuntos de referência (r) e objetivo (t), o índice de sucesso S_{tr} é definido como a probabilidade de que um artigo retirado aleatoriamente do grupo-objetivo receba mais citações do que um artigo retirado do grupo de referência S_{tr} pode ser calculado diretamente a partir da distribuição de citações de t e r .

$$S_{tr} = \sum_{c=0}^{\infty} \left(P_t(c) + \frac{P_r(c)}{2} \right) P_r(c), \quad (5.2)$$

onde $P_t(c)$ é a fração de artigos em t com mais do que c citações e $P_r(c)$ a fração de artigos em r que receberam c citações.

O índice de sucesso também pode ser obtido considerando uma aproximação. Isso permite seu uso tomando como entrada o fator de impacto de periódicos (MILOJEVIĆ; RADICCHI; BAR-ILAN, 2017). Embora neste trabalho não tenhamos usado a distribuição de citação completa para calcular o índice de sucesso, a aproximação ainda é interessante para ser usada quando apenas os fatores de impacto estão disponíveis, ou seja, quando não há acesso à distribuição de citação completa dos periódicos (ou qualquer outro subconjunto de artigos) sendo comparados.

Uma relação entre o índice de sucesso de citação e o fator de impacto pode ser derivada da definição do fator de impacto e da definição do índice de sucesso na Equação 5.2. A seguinte

equação também pode ser usada para calcular o índice de sucesso de citação:

$$S_{tr} = \frac{f_0}{2} + \frac{1 - f_0/2}{1 + q\rho^{-k}}, \quad (5.3)$$

onde $\rho = FI_t/FI_r$ é a razão entre os fatores de impacto dos conjuntos t e r , respectivamente; $k = 1.23$, q é um fator de normalização e f_0 é a taxa de artigos não citados em r . Assim, q pode ser definido como $q = 1/(1 - f_0)$. O cálculo de S_{tr} da equação 5.3 pode ser simplificado sempre que f_0 é baixo (o que normalmente ocorre quando $FI_r > 10$) ou $FI_t > FI_r$. Nesse caso,

$$S_{tr} = \frac{1}{1 + \rho^{-k}}. \quad (5.4)$$

RESULTADOS E DISCUSSÃO

Neste capítulo, apresentamos os resultados obtidos da abordagem especificada no Capítulo 5 aplicada ao conjunto de dados da APS via classificação PACS apresentados na Seção 5.1. Nas Seção 6.1 é mostrado um análise dos subcampos relevantes dos periódicos da APS e na seção 6.2 o Fator de Impacto dos subcampos é calculado. Discutimos a visibilidade dos subcampos em um mesmo periódico na Seção 6.3 e em diferentes periódicos na Seção 6.4 para um período de tempo.

6.1 Estatísticas dos subcampos

O estudo começa com a análise da evolução do número de subcampos relevantes de periódicos. No geral, na maioria dos periódicos considerados, o número de subcampos aumenta ao longo dos últimos anos (curva azul da Figura 14). Também medimos o número de subcampos considerando a heterogeneidade no número de artigos publicados em cada subcampo. Para isso, calculamos a diversidade de subcampos (definida na Subseção 5.2.1), uma medida que tem sido usada em ciência de redes e outros campos (TUOMISTO, 2010; JR *et al.*, 2017; JOST, 2006; AMANCIO *et al.*, 2012).

De acordo com o índice de diversidade, se todos os campos tiverem o mesmo tamanho, a diversidade dos subcampos corresponde ao número total de subcampos. Por outro lado, se apenas alguns subcampos tiverem a maioria dos artigos publicados, a diversidade dos subcampos será muito menor do que o número total de subcampos. No índice de diversidade, essa heterogeneidade é medida por meio da entropia (JR *et al.*, 2017).

A diversidade de subcampos medida em termos do número de artigos publicados para cada subcampo é mostrada na Figura 14 (curva amarela). Note que existe uma heterogeneidade no número de artigos em cada subcampo, uma vez que a diversidade dos subcampos é muito menor do que o número total de campos. Em 2015, quase 60 subcampos foram encontrados;

Figura 14 – Evolução do número de subcampos do periódico PRD, considerando o número de artigos e o número de citações.

Fonte: Elaborada pelo autor.

entretanto, a diversidade aponta para *efetivamente* apenas 40 subcampos. Também calculamos a diversidade de subcampos considerando o número de citações recebidas por subcampos (curva laranja). Para o periódico PRD, a Figura 14 mostra que ambas as medições de diversidade são semelhantes. Isso sugere que as citações recebidas pelos subcampos seguem o respectivo número de artigos publicados.

Um cenário diferente é observado para os periódicos PRE e PRL nos últimos anos. Enquanto as curvas de diversidade para número de artigos e citações seguem o mesmo comportamento, a diversidade de subcampos considerando o número de artigos é maior do que a diversidade medida por meio de citações.

Por exemplo como é mostrado na Figura 15a, no ano 2015, foram identificados 115 subcampos no periódico PRL. Nesse mesmo ano, a diversidade em termos de número de artigos era de cerca de 90 subcampos. Por outro lado, a diversidade de subcampos medidos em termos de citações foi apenas cerca de 78 subcampos. Este resultado sugere que alguns subcampos são mais citados do que outros e essa diferença não pode ser explicada apenas pelo tamanho dos subcampos na PRL.

Os subcampos identificados no periódico PRE (Figura 15b) em 2015 foram 65, porém a diversidade em termos de artigos foi de 48 subcampos e a diversidade de subcampos medidos em termos de citações foi cerca de 41 subcampos. Para investigar melhor as diferenças na visibilidade dos subcampos, na Seção 6.3 comparamos os subcampos no mesmo periódico usando o índice de sucesso de citação.

Figura 15 – Evolução do número de subcampos, considerando o número de artigos e o número de citações.

(a) Evolução do número de subcampos do periódico PRL

(b) Evolução do número de subcampos do periódico PRE

Fonte: Elaborada pelo autor.

6.2 Impacto dos subcampos

Para analisar como a visibilidade dos subcampos varia ao longo do tempo, calculamos o Fator de Impacto (FI) anual de cada subcampo definido pelo código PACS. Na Figura 16a, mostramos a evolução do fator de impacto médio dos subcampos (curva azul) e a evolução do fator de impacto do periódico PRL (curva celeste). Como usamos o conjunto de dados da APS, estamos limitados às citações recebidas por seus periódicos. Como consequência, os fatores de impacto podem não ser os mesmos relatados pela *Clarivate Analytics*¹. À semelhança de outros resultados da literatura, esta amostragem não afeta a comparação de dados de citação de periódicos e subcampos (MILOJEVIĆ; RADICCHI; BAR-ILAN, 2017).

O FI médio dos subcampos é consistente com o FI do periódico. Isso acontece para todos os periódicos considerados. No entanto observamos uma variabilidade do impacto dos subcampos ao longo do tempo, o que significa que diferentes subcampos são mais (ou menos) visíveis do que o periódico como um todo. Tal variabilidade fica evidente ao se analisar o coeficiente de variação do FI dos subcampos. A Figura 16b revela que, nos anos mais recentes, o desvio típico é de aproximadamente 50% do FI médio.

¹ <https://clarivate.com/>

Figura 16 – Evolução do fator de impacto para subcampos no periódico PRL.

Fonte: Elaborada pelo autor.

Uma heterogeneidade ainda maior de impacto dos subcampos ocorreu em 1990. Naquele ano, um desvio típico de 75% do FI médio foi observado. Esse resultado sugere que diferentes subcampos publicados pela PRL podem ter visibilidade diferente. Outros periódicos da APS exibem um comportamento semelhante, porém os coeficientes de variação do impacto dos subcampos são normalmente abaixo de 0.50, resultados mostrados no Apêndice A.

6.3 Comparação dos subcampos intra-periódicos

Para investigar mais a fundo o quão diferente é o impacto dos subcampos dentro de um periódico, usamos o índice de sucesso de citação (definido na Seção 5.3). Esta medição fornece uma interpretação mais clara em relação à diferença de visibilidade (isto é, fator de impacto) entre dois subcampos. Em outras palavras, dados os subcampos A e B , o índice de sucesso S_{AB} comparando-os dá a probabilidade de que um artigo escolhido aleatoriamente de A seja mais citado do que um artigo escolhido de B .

Figura 17 – Evolução do índice de sucesso comparando o impacto de todos os pares de subcampos do periódico PRB.

Fonte: Elaborada pelo autor.

Para entender as diferenças de visibilidade, para cada periódico, medimos o índice de sucesso entre todos os pares de subcampos no mesmo periódico. Como $S_{AB} + S_{BA} = 100\%$ em nossa análise, para cada par A e B , consideramos apenas o índice máximo entre S_{AB} e S_{BA} .

Observamos na Figura 17, que para o periódico PRB o índice de sucesso médio (comparando todos os pares de subcampos) varia entre 55 e 57%. Isso significa que normalmente o fator de impacto dos subcampos neste periódico são semelhantes. No entanto, em casos particulares, alguns subcampos são muito mais visíveis do que outros. Os valores mais altos do índice de sucesso são destacados na curva vermelha. Para efeito de comparação, mostramos também o índice de sucesso obtido na comparação dos periódicos PRL e PRB (curva verde).

Normalmente, a diferença de visibilidade entre PRL e PRB é mais significativa do que a

diferença de visibilidade de subcampos do periódico PRB. No entanto, para pares *específicos* de subcampos, a diferença de visibilidade entre subcampos é mais significativa do que a diferença de impacto entre periódicos (PRL e PRB). Em 2015, o índice de sucesso comparando PACS 85.25 (Superconducting devices) e PACS 85.75 (Magnetoelectronics and spintronics) atingiu quase 85%, enquanto a diferença de visibilidade entre PRB e PRL foi de cerca de 65%. Uma lista de códigos PACS é mostrada no Apêndice B.

Figura 18 – Evolução do índice de sucesso comparando o impacto de todos os pares de subcampos do periódico PRE.

Fonte: Elaborada pelo autor.

Na Figura 18 mostramos a distribuição dos índices de sucesso ao comparar todos os pares de subcampos do periódico PRE (boxplots). Em todos os anos considerados, a maioria dos índices de sucesso está abaixo de 60%, revelando que não há diferença significativa de impacto nos subcampos do periódico PRE para a maioria dos pares considerados. No entanto, conforme observado para o PRB (ver Figura 17), alguns pares de subcampos exibem fatores de impacto muito distintos.

Em 2015, a comparação entre o PACS 89.20 e 85.75 rendeu um índice de sucesso superior a 85%. Isso é muito maior do que o índice de sucesso obtido ao comparar os periódicos PRA e PRE. Nos últimos anos, podemos ver que a diferença de impacto interna típica é menor do que a diferença de impacto entre PRA e PRE (ver curva verde). Diferentemente da Figura 17, em termos de visibilidade, parece que a escolha do subcampo dentro do periódico PRE é mais importante do que a escolha entre o PRA e o PRE em determinados anos. Isso fica claro, por

exemplo, em 1995 e 1996, quando mais de 75% de todos os pares de subcampos tinham uma diferença de visibilidade maior do que a observada entre PRA e PRE.

Figura 19 – Evolução do índice de sucesso comparando o impacto de todos os pares de subcampos do periódico PRA.

Fonte: Elaborada pelo autor.

A Figura 19 mostra a distribuição de índices de sucesso comparando todos os pares de subcampos do periódico PRA. Também mostramos a evolução do índice de sucesso comparando os periódicos PRA e PRB. Essa comparação é mostrada em duas cores diferentes: verde; o que significa que o impacto do PRB é maior do que o impacto do PRA; e celeste, para representar o oposto.

Ao contrário da análise anterior, observamos um comportamento interessante após 1997. A diferença de impacto entre PRA e PRB torna-se muito pequena, conforme revelado por índices de sucesso tipicamente abaixo de 55%. Ao mesmo tempo, a diferença de impacto dos subcampos de PRA é tipicamente maior que 55%, mais uma vez significando que as diferenças de visibilidade interna (subcampo) podem ser mais importantes do que as diferenças de visibilidade observadas entre periódicos. Em casos extremos, o índice de sucesso comparando dois subcampos do periódico PRA atinge a quase 90%.

6.3.1 Teste de robustez

Para analisar se as diferenças observadas na visibilidade dos artigos podem ser descritas por acaso, foi realizado o seguinte teste de robustez. Para um determinado ano, associamos aleatoriamente códigos PACS aos artigos, mas mantendo a distribuição original do tamanho do conjunto de artigos e a cardinalidade existente entre os subcampos. Em seguida, analisamos para cada atribuição aleatória o índice de sucesso máximo.

Figura 20 – Histograma do índice de sucesso máximo de subcampos associados aleatoriamente a artigos do periódico PRA.

Fonte: Elaborada pelo autor.

A distribuição dos índices de sucesso é mostrada na Figura 20. Esta figura mostra os resultados obtidos para o periódico PRA em 2013 e 2015. A linha tracejada vermelha representa o valor correspondente do índice de sucesso observado nos dados reais, 80 e 79 correspondentemente. Esse resultado sugere que as diferenças na visibilidade dos subcampos não podem ser explicadas por uma divisão aleatória dos artigos em subcampos. Devido que ambos os anos, o perfil observado para atribuição de códigos aleatórios é claramente incompatível com os dados reais.

6.4 Comparação dos subcampos inter-periódicos

Enquanto nas Figuras 17, 18 e 19 focamos nossa análise na comparação de subcampos em um *mesmo* periódico, agora comparamos o fator de impacto de subcampos em diferentes periódicos. A comparação entre todos os pares de subcampos dos periódicos PRA e PRE é mostrada na Figura 21. Um padrão interessante observado aqui é que a curva do índice de sucesso comparando o impacto dos dois periódicos (curva celeste e verde) segue o mesmo comportamento da mediana comparando pares de subcampos.

Portanto, em geral, a comparação de visibilidade do PRA e do PRE é compatível com a comparação da visibilidade dos respectivos subcampos. No entanto, subcampos particulares têm visibilidade muito distinta, conforme revelado pela dinâmica da curva vermelha na Figura 21. Particularmente, em 2015, foi encontrada uma grande diferença de visibilidade ao comparar PACS 05.70 (do PRA) e 47.65 (do PRE). Observe que essa diferença de visibilidade é muito maior do que a diferença de visibilidade típica entre PRA e PRE.

Figura 21 – Evolução do índice de sucesso comparando o impacto dos subcampos dos periódicos PRA e PRE.

Fonte: Elaborada pelo autor.

A relevância de comparar subcampos (ao invés de periódicos) pode ser observada ao comparar subcampos de periódicos PRC e PRB, conforme mostrado na Figura 22. De 1986 a 1994, o índice de sucesso comparando PRB e PRC é compatível com a mediana do índice de sucesso comparando os respectivos subcampos. No entanto, de 2002 a 2015, é claro que PRC e PRB têm valores de fator de impacto muito semelhantes, conforme revelado por valores de índice de sucesso muito próximos de 50% (ver curvas celeste e verde). Nesse mesmo período, porém, a diferença típica entre a visibilidade dos subcampos era próxima a 60%.

Mais uma vez, para determinados subcampos, o índice de sucesso de citação atingiu valores próximos a 85%. Enquanto em 2015 os periódicos têm o mesmo fator de impacto, os subcampos representados pelo PACS 12.38 e 61.46 apresentam uma diferença de impacto gerando um índice de sucesso próximo a 80%.

Figura 22 – Evolução do índice de sucesso comparando o impacto dos subcampos dos periódicos PRC e PRB.

Fonte: Elaborada pelo autor.

6.4.1 Teste de robustez

Um teste de robustez análogo ao realizado na Figura 20 também revelou que as diferenças de visibilidade não podem ser explicadas apenas pelo acaso. Os resultados correspondentes a este teste são mostrados na Figura 23. Enquanto mostramos os resultados para a comparação dos subcampos dos periódicos PRA e PRE em 2013 e 2015, resultados semelhantes foram encontrados para as outras comparações de visibilidade dos subcampos em diferentes periódicos.

6.5 Conclusões do capítulo

Ao todo, os resultados apresentados nas Seções 6.3 e 6.4 revelaram que os subcampos intra e inter-periódicos podem ter visibilidade muito distinta. Este é um resultado importante, pois pode ser usado como uma informação adicional em políticas de pesquisa. Do ponto de vista dos estudiosos, a quantificação da visibilidade dos subcampos poderia auxiliar os pesquisadores em suas decisões de carreira. Juntamente com outros atributos do campo, pode ser interessante saber de antemão a visibilidade potencial dos subcampos antes que esforços sejam feitos para aprender e produzir conhecimento no campo.

Outra conclusão interessante decorrente dos resultados obtidos diz respeito à comparação

Figura 23 – Histograma do índice de sucesso máximo de subcampos associados aleatoriamente a artigos dos periódicos PRA e PRE.

Fonte: Elaborada pelo autor.

dos periódicos. Nossos resultados mostram que, em alguns casos, a comparação direta de periódicos pode não fornecer todas as informações relevantes. Dois periódicos com fatores de impacto semelhantes podem ter valores de visibilidade muito distintos quando os subcampos são comparados. Essas informações podem ser usadas, por exemplo, para melhorar as previsões e avaliações que usam o fator de impacto (em combinação com outros índices de impacto de pesquisa estabelecidos).

CONCLUSÃO

A análise do impacto dos subcampos é relevante para fornecer informações mais detalhadas dos fatores que afetam a visibilidade dos periódicos. Neste trabalho, estudamos a variabilidade do impacto dos subcampos em um subconjunto de periódicos de física publicados pela APS. Apesar de suas limitações e deficiências, os fatores de impacto de periódicos são uma das ferramentas mais conhecidas e disponíveis para quantificar o impacto médio de citações de curto prazo ([ANTONOIANNAKIS, 2020](#); [BORDONS](#); [FERNÁNDEZ](#); [GÓMEZ, 2002](#); [WALLIN, 2005](#)).

As diferenças de impacto foram calculadas usando o índice de sucesso, uma medida que compara a distribuição de citações observada em dois conjuntos de artigos. De acordo com [Milojević, Radicchi e Bar-Ilan \(2017\)](#), essa medição também pode ser calculada, de forma simplificada, usando o fator de impacto dos conjuntos de artigos considerados. O índice de sucesso comparando dois subconjuntos de artigos A e B fornece a probabilidade de que um artigo escolhido aleatoriamente do subconjunto A receberá mais citações do que um artigo escolhido aleatoriamente do subconjunto B .

A identificação dos subcampos foi realizada utilizando o esquema de classificação fornecido pela APS, para logo comparar as diferenças de visibilidade dos subcampos de duas maneiras. Primeiro, os subcampos do mesmo periódico foram comparados, assim descobrimos que alguns subcampos são mais visíveis do que outros no mesmo periódico. Por exemplo, para artigos publicados no periódico PRB em 2015, o índice de sucesso comparando o código PACS 85.25 (superconducting devices) e 85.75 (magnetoelectronics and spintronics) foi encontrado em cerca de 85%; enquanto, no mesmo ano, o índice de sucesso comparando os periódicos PRB e PRL foi de apenas 65%. Também descobrimos que, em anos específicos, a diferença de visibilidade típica para subcampos de um periódico (por exemplo, subcampos do periódico PRE) pode ser maior do que a diferença de visibilidade dos periódicos (por exemplo, PRA vs. PRE em 1995 e 1996). Um segundo tipo de análise focou na comparação de subcampos em periódicos distintos.

Nesse caso, também descobrimos que quando os subcampos são considerados, as diferenças de visibilidade dos subcampos podem ser muito diferentes das diferenças de visibilidade dos respectivos periódicos.

De modo geral, nossos resultados sugerem que a visibilidade dos subcampos em periódicos especializados pode não ser uniforme e essa informação pode ser usada para melhor compreender os componentes que afetam o impacto dos periódicos. Essas informações podem ser utilizadas, por exemplo, por autores que pesquisam o local mais adequado e visível para o seu trabalho. Embora essas métricas de visibilidade de periódicos possam ajudar, uma visibilidade de subcampos mais detalhada pode permitir que os autores tomem decisões mais informadas.

Os editores também podem confiar nas informações de visibilidade dos subcampos para entender melhor como os periódicos evoluem ao longo do tempo. Isso poderia ser usado, por exemplo, para destacar as principais áreas dos periódicos. A identificação de subcampos de impacto relativamente baixo em um periódico pode ser importante para identificar fontes potenciais de práticas científicas de baixa qualidade (por exemplo, relatórios de revisores de baixa qualidade) adotados pelo periódico em um subcampo específico.

Deve-se ter em mente que, apesar das vantagens de se entender melhor a visibilidade dos subcampos, esse conhecimento também pode levar a comportamentos enviesados. Por exemplo, se o subcampo *A* é mais visível do que o subcampo *B* em um periódico, esta informação pode ser usada para aumentar o impacto do periódico, aceitando mais artigos no subcampo *A*. Condutas semelhantes foram relatadas em estudos recentes de [Martin \(2016\)](#) e devem ser evitadas caso a visibilidade dos subcampos se torne amplamente disponível.

7.1 Trabalhos futuros

Enquanto focalizamos a análise de visibilidade dos subcampos via classificação PACS, este trabalho poderia ser estendido considerando outras noções de subcampos, incluindo informações extraídas de textos ([ARRUDA; COSTA; AMANCIO, 2016](#)). Por exemplo, os subcampos podem ser identificados sem qualquer tipo de esquema de classificação. Em particular, a identificação de subcampos poderia ser realizada por meio da detecção de comunidades em redes de citação (ou co-citação) ([SILVA *et al.*, 2016](#)).

Os subcampos também poderiam ser identificados usando redes de palavras de coocorrência obtidas do título e/ou resumo ([CASTRO; STELLA, 2019](#); [AMANCIO; SILVA; COSTA, 2015](#); [STELLA *et al.*, 2019](#)). Redes de colaboração também podem ser usadas para detectar subcampos, mas, neste caso, uma estratégia de desambiguação é essencial para construir redes baseadas no autor ([AMANCIO *et al.*, 2012](#)). Finalmente, também pretendemos estender este estudo para analisar a variabilidade da visibilidade de subcampos em outros campos importantes.

7.2 Publicação

As principais contribuições desta pesquisa foram relatadas na seguinte publicação:

1. **Chacon, X.S.Q.**, Silva, T.C. & Amancio, D.R. Comparing the impact of subfields in scientific journals. *Scientometrics* 125, 625–639 (2020).

REFERÊNCIAS

- ABRAMO, G.; D'ANGELO, C.; COSTA, F. D. Citations versus journal impact factor as proxy of quality: could the latter ever be preferable? **Scientometrics**, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV ..., v. 84, n. 3, p. 821–833, 2010. Citado nas páginas 35 e 36.
- ALBERTS, B. **Impact factor distortions**. [S.l.]: American Association for the Advancement of Science, 2013. Citado na página 13.
- AMANCIO, D. R.; ALUISIO, S. M.; JR, O. N. O.; COSTA, L. d. F. Complex networks analysis of language complexity. **EPL (Europhysics Letters)**, IOP Publishing, v. 100, n. 5, p. 58002, 2012. Citado nas páginas 43 e 56.
- AMANCIO, D. R.; SILVA, F. N.; COSTA, L. d. F. Concentric network symmetry grasps authors' styles in word adjacency networks. **EPL (Europhysics Letters)**, IOP Publishing, v. 110, n. 6, p. 68001, 2015. Citado na página 56.
- AMJAD, T.; DING, Y.; XU, J.; ZHANG, C.; DAUD, A.; TANG, J.; SONG, M. Standing on the shoulders of giants. **Journal of Informetrics**, Elsevier, v. 11, n. 1, p. 307–323, 2017. Citado nas páginas 36 e 37.
- ANDRÉS, A. **Measuring academic research: How to undertake a bibliometric study**. [S.l.]: Elsevier, 2009. Citado na página 32.
- ANTONOIYIANNAKIS, M. Impact factor volatility due to a single paper: A comprehensive analysis. **Quantitative Science Studies**, MIT Press, v. 1, n. 2, p. 639–663, 2020. Citado na página 55.
- ARRUDA, H. F. de; COSTA, L. d. F.; AMANCIO, D. R. Using complex networks for text classification: Discriminating informative and imaginative documents. **EPL (Europhysics Letters)**, IOP Publishing, v. 113, n. 2, p. 28007, 2016. Citado na página 56.
- BALL, P. **Index aims for fair ranking of scientists**. [S.l.]: Nature Publishing Group, 2005. Citado na página 32.
- BARABÁSI, A.-L.; PÓSFAL, M. **Network science**. Cambridge: Cambridge University Press, 2016. ISBN 9781107076266 1107076269. Disponível em: <<http://barabasi.com/networksciencebook/>>. Citado na página 28.
- BORDONS, M.; FERNÁNDEZ, M.; GÓMEZ, I. Advantages and limitations in the use of impact factor measures for the assessment of research performance. **Scientometrics**, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV ..., v. 53, n. 2, p. 195–206, 2002. Citado na página 55.
- BORNMAN, L.; DANIEL, H.-D. What do we know about the h index? **Journal of the American Society for Information Science and technology**, Wiley Online Library, v. 58, n. 9, p. 1381–1385, 2007. Citado na página 13.

- BORNMAN, L.; LEYDESDORFF, L. Does quality and content matter for citedness? a comparison with para-textual factors and over time. **Journal of Informetrics**, Elsevier, v. 9, n. 3, p. 419–429, 2015. Citado na página 36.
- BORNMAN, L.; MARX, W. Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? **Journal of Informetrics**, Elsevier, v. 9, n. 2, p. 408–418, 2015. Citado na página 14.
- CASTRO, N.; STELLA, M. The multiplex structure of the mental lexicon influences picture naming in people with aphasia. **Journal of Complex Networks**, Oxford University Press, v. 7, n. 6, p. 913–931, 2019. Citado na página 56.
- CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. **Physical review E**, APS, v. 70, n. 6, p. 066111, 2004. Citado na página 24.
- COMMONS, W. **File:Erds generated network-p0.01.jpg — Wikimedia Commons, the free media repository**. 2020. [Online; accessed 7-December-2020]. Disponível em: <https://commons.wikimedia.org/w/index.php?title=File:Erds_generated_network-p0.01.jpg&oldid=468326599>. Citado na página 26.
- COSTA, L. d. F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: A survey of measurements. **Advances in physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007. Citado na página 23.
- FORTUNATO, S.; BERGSTROM, C. T.; BÖRNER, K.; EVANS, J. A.; HELBING, D.; MILOJEVIĆ, S.; PETERSEN, A. M.; RADICCHI, F.; SINATRA, R.; UZZI, B. *et al.* Science of science. **Science**, American Association for the Advancement of Science, v. 359, n. 6379, 2018. Citado na página 13.
- FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, JSTOR, p. 35–41, 1977. Citado na página 21.
- GARFIELD, E. Citation indexes for science. **Science**, JSTOR, v. 122, n. 3159, p. 108–111, 1955. Citado na página 31.
- HIRSCH, J. E. An index to quantify an individual's scientific research output. **Proceedings of the National academy of Sciences**, National Acad Sciences, v. 102, n. 46, p. 16569–16572, 2005. Citado na página 32.
- HOLDEN, C. Random samples: data point-impact factor. **Science**, v. 309, n. 8, p. 1181c, 2005. Citado na página 32.
- JEONG, H.; NÉDA, Z.; BARABÁSI, A.-L. Measuring preferential attachment in evolving networks. **EPL (Europhysics Letters)**, IOP Publishing, v. 61, n. 4, p. 567, 2003. Citado na página 37.
- JOST, L. Entropy and diversity. **Oikos**, Wiley Online Library, v. 113, n. 2, p. 363–375, 2006. Citado na página 43.
- JR, E. A. C.; SILVA, F. N.; COSTA, L. d. F.; AMANCIO, D. R. Patterns of authors contribution in scientific manuscripts. **Journal of Informetrics**, Elsevier, v. 11, n. 2, p. 498–510, 2017. Citado nas páginas 40 e 43.

LARIVIÈRE, V.; GINGRAS, Y. The impact factor's matthew effect: A natural experiment in bibliometrics. **Journal of the American Society for Information Science and Technology**, Wiley Online Library, v. 61, n. 2, p. 424–427, 2010. Citado na página 35.

LEYDESDORFF, L.; MILOJEVIĆ, S. Scientometrics. **arXiv preprint arXiv:1208.4566**, 2012. Citado na página 29.

LI, Y.; RADICCHI, F.; CASTELLANO, C.; RUIZ-CASTILLO, J. Quantitative evaluation of alternative field normalization procedures. **Journal of Informetrics**, Elsevier, v. 7, n. 3, p. 746–755, 2013. Citado na página 14.

LUNDBERG, J. Lifting the crown—citation z-score. **Journal of informetrics**, Elsevier, v. 1, n. 2, p. 145–154, 2007. Citado na página 14.

MARTIN, B. R. **Editors' JIF-boosting stratagems—Which are appropriate and which not?** [S.l.]: Elsevier, 2016. Citado na página 56.

MILOJEVIĆ, S.; RADICCHI, F.; BAR-ILAN, J. Citation success index- an intuitive pair-wise journal comparison metric. **Journal of Informetrics**, Elsevier, v. 11, n. 1, p. 223–231, 2017. Citado nas páginas 40, 45 e 55.

NEWMAN, M. **Networks**. [S.l.]: Oxford university press, 2018. Citado na página 20.

NEWMAN, M. E. Clustering and preferential attachment in growing networks. **Physical review E**, APS, v. 64, n. 2, p. 025102, 2001. Citado na página 37.

_____. A measure of betweenness centrality based on random walks. **Social networks**, Elsevier, v. 27, n. 1, p. 39–54, 2005. Citado na página 21.

_____. Modularity and community structure in networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 103, n. 23, p. 8577–8582, 2006. Citado na página 24.

NEWMAN, M. E.; WATTS, D. J.; STROGATZ, S. H. Random graph models of social networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 99, n. suppl 1, p. 2566–2572, 2002. Citado na página 25.

PETERSEN, A. M.; JUNG, W.-S.; YANG, J.-S.; STANLEY, H. E. Quantitative and empirical demonstration of the matthew effect in a study of career longevity. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 108, n. 1, p. 18–23, 2011. Citado nas páginas 37 e 38.

PRICE, D. J. D. S. Little science, big science. 1963. Citado na página 30.

RABINOWITZ, H.; VOGEL, S. **The Manual of Scientific Style: a guide for authors, editors, and researchers**. [S.l.]: Elsevier, 2009. Citado na página 33.

RADICCHI, F.; FORTUNATO, S.; CASTELLANO, C. Universality of citation distributions: Toward an objective measure of scientific impact. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 105, n. 45, p. 17268–17272, 2008. Citado na página 14.

REDNER, S. How popular is your paper? an empirical study of the citation distribution. **The European Physical Journal B-Condensed Matter and Complex Systems**, Springer, v. 4, n. 2, p. 131–134, 1998. Citado na página 13.

- SILVA, F. N.; AMANCIO, D. R.; BARDOSOVA, M.; COSTA, L. d. F.; JR, O. N. O. Using network science and text analytics to produce surveys in a scientific topic. **Journal of Informetrics**, Elsevier, v. 10, n. 2, p. 487–502, 2016. Citado na página 56.
- STELLA, M.; NIGRIS, S. D.; ALORIC, A.; SIEW, C. S. Forma mentis networks quantify crucial differences in stem perception between students and experts. **PloS one**, Public Library of Science San Francisco, CA USA, v. 14, n. 10, p. e0222870, 2019. Citado na página 56.
- TUOMISTO, H. A consistent terminology for quantifying species diversity? yes, it does exist. **Oecologia**, Springer, v. 164, n. 4, p. 853–860, 2010. Citado na página 43.
- WALLIN, J. A. Bibliometric methods: pitfalls and possibilities. **Basic & clinical pharmacology & toxicology**, Wiley Online Library, v. 97, n. 5, p. 261–275, 2005. Citado na página 55.
- WALTMAN, L. A review of the literature on citation impact indicators. **Journal of informetrics**, Elsevier, v. 10, n. 2, p. 365–391, 2016. Citado nas páginas 13 e 29.
- WALTMAN, L.; ECK, N. J. van; LEEUWEN, T. N. van; VISSER, M. S.; RAAN, A. F. van. Towards a new crown indicator: An empirical analysis. **Scientometrics**, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV ..., v. 87, n. 3, p. 467–481, 2011. Citado na página 14.
- WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. **nature**, Nature Publishing Group, v. 393, n. 6684, p. 440, 1998. Citado na página 26.
- WOLFGANG, G.; BART, T.; BALÁZS, S. A bibliometric approach to the role of author self-citations in scientific communication. **Scientometrics**, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV ..., v. 59, n. 1, p. 63–77, 2004. Citado na página 30.
- WOLFRAM, D. **Applied informetrics for information retrieval research**. [S.l.]: Greenwood Publishing Group, 2003. Citado na página 30.

FIGURAS SUPLEMENTARES

Evolução do fator de impacto para subcampos

Evolução do fator de impacto para subcampos no periódico PRA.

Fonte: Elaborada pelo autor.

Evolução do fator de impacto para subcampos no periódico PRB.

Fonte: Elaborada pelo autor.

Evolução do fator de impacto para subcampos no periódico PRC.

Fonte: Elaborada pelo autor.

Evolução do fator de impacto para subcampos no periódico PRD.

Fonte: Elaborada pelo autor.

Evolução do fator de impacto para subcampos no periódico PRE.

Fonte: Elaborada pelo autor.

LISTA DE CÓDIGOS PACS

Estas informações suplementares relacionam o código dos subcampos mencionados neste manuscrito. Os primeiros dois dígitos correspondem aos primeiros níveis hierárquicos. Neste estudo, analisamos os subcampos no terceiro nível hierárquico. Este nível hierárquico compreende 4 dígitos.

1. **Código 05:** statistical physics, thermodynamics, and nonlinear dynamical systems, subcampo: thermodynamics (05.70).
2. **Código 11:** General theory of fields and particles, com subcampos: symmetry and conservation laws (11.30).
3. **Código 12:** specific theories and interaction models; particle systematics, com subcampos: quantum electrodynamics (12.20), quantum chromodynamics (12.38).
4. **Código 21:** nuclear structure, subcampo: nuclear forces (21.30).
5. **Código 25:** nuclear reactions: specific reactions, com subcampos: relativistic heavy-ion collisions (25.75).
6. **Código 25:** nuclear astrophysics, com subcampos: nuclear matter aspects of neutron stars (26.60).
7. **Código 37:** mechanical control of atoms, molecules, and ions, com subcampos: atoms, molecules, and ions in cavities (37.30).
8. **Código 47:** fluid dynamics, com subcampos: magneto-hydrodynamics and electro-hydrodynamics (47.65).
9. **Código 67:** quantum fluids and solids, com subcampos: ultracold gases, trapped gases (67.85).

10. **Código 62:** Mechanical and acoustical properties of condensed matter, com subcampo: mechanical properties of solids (62.20).
11. **Código 64:** equations of state, phase equilibria, and phase transitions, com subcampos: general studies of phase transitions (64.60), phase equilibria (64.75).
12. **Código 67:** quantum fluids and solids, subcampo: ultracold gases, trapped gases (67.85).
13. **Código 74:** superconductivity, com subcampos: cuprate superconductors (74.72), superconducting films and low-dimensional structures (74.78).
14. **Código 75:** magnetic properties and materials, com subcampos: general theory and models of magnetic ordering (75.10), spin transport effects (75.76).
15. **Código 85:** electronic and magnetic devices; microelectronics, com subcampos: superconducting devices (85.25), magnetoelectronics; spintronics: devices exploiting spin polarized transport or integrated magnetic fields (85.75).
16. **Código 87:** Biological and medical physics, com subcampos: biomolecules: types (87.14), subcellular structure and processes (87.16).
17. **Código 89:** Other areas of applied and interdisciplinary physics, subcampo: interdisciplinary applications of physics (89.20).