

Estado da publicação: Não informado pelo autor submissor

Coocorrência de Palavras em Títulos de Publicações da Comunidade Acadêmica de Artes

Rafael Werner, Thiago Pinto, Angelo Loula, João Queiroz

<https://doi.org/10.1590/SciELOPreprints.10957>

Submetido em: 2024-12-18

Postado em: 2025-01-05 (versão 1)

(AAAA-MM-DD)

A moderação deste preprint recebeu o endosso de:

João B. Rocha-Junior (ORCID: <https://orcid.org/0000-0002-6925-9729>)

COOCORRÊNCIA DE PALAVRAS EM TÍTULOS DE PUBLICAÇÕES DA COMUNIDADE ACADÊMICA DE ARTES

WORD CO-OCCURRENCE IN TITLES OF PUBLICATIONS FROM THE ARTS ACADEMIC COMMUNITY

Rafael Werner

<https://orcid.org/0000-0002-8618-8993>

Thiago Pinto

<https://orcid.org/0009-0007-9379-4864>

Angelo Loula

<https://orcid.org/0000-0001-7802-1731>

João Queiroz¹

<https://orcid.org/0000-0001-6978-4446>

¹ Correspondência: joao.queiroz@ufjf.br

RESUMO

O título é considerado o componente mais importante de uma publicação acadêmica. Ele comunica aos leitores, com frequência, o objeto da pesquisa, seus limites, métodos, e propósitos. **Objetivo:** Examinamos diversas propriedades relacionadas aos títulos de artigos, livros e capítulos, mono e multi-autorais, da comunidade acadêmica de artes em Minas Gerais (MG). Tais publicações incluem, ao menos, um(a) autor(a) filiado(a), entre 2013 e 2023, a um programa de pós-graduação (PPG) de Artes, no estado de MG. **Metodologia:** A metodologia usada (*network analysis*) baseia-se na identificação de propriedades estatísticas e topológicas, relacionadas à coocorrência de palavras. São examinadas propriedades de conectividade, centralidade e distributividade, além do tamanho dos títulos. As análises de redes de coocorrência de palavras se concentraram nos títulos de publicações em inglês. **Resultados:** Sobre a variedade de idiomas, do total de 7484 publicações, 85,5% foram publicados em português, 8,74% em inglês e 5,76% em outros idiomas. A caracterização da coocorrência de palavras dos títulos, com base nas arestas repetidas e paralelas normalizadas, indica 17,45% de palavras repetidas e paralelas em sequência. Arestas repetidas e paralelas ocorrem em maior proporção em publicações multi-autorais do que mono-autorais, e, em artigos, capítulos e livros, respectivamente. Baseados na razão entre as palavras-tipos (*types*), e palavras-ocorrências nos títulos (*tokens*), a taxa de diversidade de vocabulário é 47,6% em artigos, 58% em capítulos e 75% em livros. **Conclusões:** Identificamos maior autonomia para decisão autoral sobre títulos de livros, do que sobre artigos acadêmicos e capítulos, que certamente sofrem maior restrição relacionada a políticas editoriais. Os resultados parecem exibir um aumento da coerção exercida pelo grupo, ou sub-grupo da comunidade, sobre um fenômeno ou tópico investigado. Por outro lado, títulos de artigos multi-autorais tendem a refletir uma atividade autoral distribuída (*group mind*), onde se observa aumento no número de palavras e na variedade lexical dos termos da pesquisa.

Descritores: Coocorrência de palavras. Títulos. Análise de redes. Arte.

1 INTRODUÇÃO

O título de um trabalho acadêmico é considerado seu componente mais importante (Garcia et al., 2019; Jamali; Nikzad, 2011). Ele comunica diretamente a seus leitores o objeto (conteúdo, propósito, e limites) de uma pesquisa. O título também é usado para indexação e referência em bancos de dados (Izzo et al., 2022). A estrutura do título, tamanho, e forma, podem variar consideravelmente entre áreas e disciplinas científicas (Hudson, 2016). Há, inclusive, tendências que são identificadas quando comparamos ciências duras e brandas (*hard e soft sciences*) (Nagano, 2015). Nas ciências duras (*hard sciences*), há mais “palavras de conteúdo” (*content words*) (Haspelmath, 2003), sugerindo maior precisão e delimitação às suas audiências. Nas ciências brandas (*soft sciences*), segundo Nagano (2015), os títulos atuam de modo a

atrair uma maior variedade de leitores.

Há análises sobre muitos padrões e atributos dos títulos, como, por exemplo, sinais de pontuação, e os resultados sugerem diversas tendências, boa parte delas ainda sem explicação. O uso dos caracteres “dois pontos”, e “ponto de interrogação”, tende a diminuir com o aumento do número de autores (Hudson, 2016). Em Engenharia, Biologia e Economia, é mais usual a utilização de títulos únicos; em Sociologia, Educação, Medicina, História e Geografia, é mais comum a divisão do título em duas unidades, título e subtítulo (Nagano, 2015). Outros trabalhos analisam diferentes propriedades estruturais como frequência de frases compostas, completas, nominais e interrogativas (Soler, 2007). Muitas análises definem métricas estruturais relacionadas ao número de palavras, se substantivas, informativas ou não informativas, comprimento, composição e formato do título, número de caracteres, presença ou ausência de caracteres alfanuméricos, uso de acrônimos, entre outras. Os autores também têm analisado a extensão dos títulos. Hudson (2016, p. 871), entre eles, conclui que o tamanho dos títulos é diretamente proporcional ao aumento do número de autores em quase todas as disciplinas. Sobre uma conexão entre o tamanho do título e seu impacto, não existe consenso entre os pesquisadores.

Para Hartley (2008), diversos tipos, ou classes, de títulos devem ser distinguidos, títulos que: (1) anunciam um assunto geral, (2) particularizam um tema específico, (3) indicam uma pergunta orientadora, (4) declaram as descobertas, (5) indicam a resposta a uma pergunta que será revelada, (6) anunciam a tese e indicam a direção do argumento, (7) enfatizam a metodologia utilizada na pesquisa, (8) sugerem diretrizes e/ou comparações, (9) chamam a atenção para questões surpreendentes e eficazes, (10) atraem por aliteração, trocadilhos, e alusões literárias. Outros aspectos, segundo Nair e Gibbert (2016), incluem: se o título é declarativo, descritivo ou interrogativo; se especifica um método ou protocolo; se identifica resultados ou conclusões; se especifica um contexto, região geográfica, empresa, ou setor de empresa; se faz uso de recursos linguísticos heterodoxos e se são divertidos. Títulos divertidos ou jocosos tendem a ser mais curtos e apresentam uma pequena

correlação com o aumento no número de downloads dos artigos. Entretanto, não há correlação significativa com o número de citações (Subotic; Mukherjee, 2014).

Em projetos de bibliometria, e cientometria, análises de redes (*network analysis*) são usadas como ferramentas analíticas das dinâmicas de produção, disseminação e evolução do conhecimento (artístico, científico, filosófico) (Chen Song, 2019). Elas fornecem uma linguagem formal para estudar sistemas de diversas ontologias (biológicas, sociais, tecnológicas e informacionais), e podem variar em escopo e propósitos (análise, previsão, controle). As análises, e sua vertente teórico-matemática, que é a teoria de grafos (*graph theory*), têm apresentado, com sucesso, resultados originais e robustos sobre fenômenos tão diversos quanto propagação de doenças virais (Loyal; Chen, 2020), sistemas distribuídos de energia (Bauwens et al., 2022), biologia sistêmica (Zhu et al., 2021) e ecologia (Albery et al., 2021). Uma rede é um “catálogo de componentes de um sistema”, chamados nós, ou vértices, e interações entre eles, chamados links, ou arestas (Barabási, 2016). As redes capturam padrões de interação entre as partes de um sistema, e representam estatísticas e topologias de seu comportamento e estrutura (Newman, 2004).

A análise de publicações científicas por meio de redes de colaboração de palavras é um método valioso para descoberta de muitos padrões (Radhakrishnan et al., 2017). Esta abordagem modela relacionamentos linguísticos com base na frequência de ocorrência de palavras. As redes capturam as associações entre palavras sucessivas sem especificar a natureza dos relacionamentos linguísticos, gramaticais, sintáticos ou semânticos (Ciampelli et al., 2023). Ao modelar redes de coocorrência de palavras, torna-se possível revelar a estrutura e os padrões associados de termos dentro de um campo específico de pesquisa. As redes podem fornecer insights sobre a transformação de estruturas intelectuais, ao longo do tempo, e facilitar a identificação de disciplinas proeminentes e padrões de colaboração (Deng Xia, 2020). Embora trate-se de uma metodologia frequentemente praticada em muitas áreas e domínios (Lozano et al., 2019; Yuan et al., 2022; Araújo et al., 2011; Jackson et al., 2018), investigações sobre títulos acadêmicos,

especialmente quando baseadas em análise de redes, seguem quase totalmente inéditas em Humanidades, e particularmente em Artes.

Neste trabalho, analisamos os títulos das publicações e a rede de coocorrência de palavras em títulos de publicações (mono e multi-autorais) da comunidade de pesquisa de artes de MG, especificamente, de publicações quando ao menos um dos autores possui filiação, entre 2013 e 2023, a um Programa de Pós-Graduação de Artes em MG. Foram extraídos títulos de publicações em periódicos, livros e capítulos de livros a partir do currículo Lattes dos pesquisadores. Cada título foi processado para construção de uma rede de colaboração de palavras na forma de grafos de palavras conectadas com a palavra subsequente, analisados para extração de atributos que caracterizam a rede. Também foram obtidas informações descritivas das publicações como idioma, tipo de publicação, número de autores, estrutura do título, e uso de caracteres especiais.

Os resultados iniciais revelam diversas propriedades e características estruturais: (1) os títulos são majoritariamente de publicações em português, seguido de publicações em inglês, com variação conforme existência de coautoria; (2) 17,45% das arestas da rede analisada são arestas repetidas ou paralelas, e é maior o número de arestas paralelas e repetidas, em publicações multi-autorais; (3) baseados na razão entre as palavras-tipos (*types*), e palavras-ocorrências (*tokens*), a taxa de diversidade de vocabulário é 47,6% em artigos, 58% em capítulos e 75% em livros; (4) o tamanho médio dos títulos é de 9,5 palavras, sendo maior a média dos títulos de artigos multi-autorais.

2 MÉTODOS E MATERIAIS:

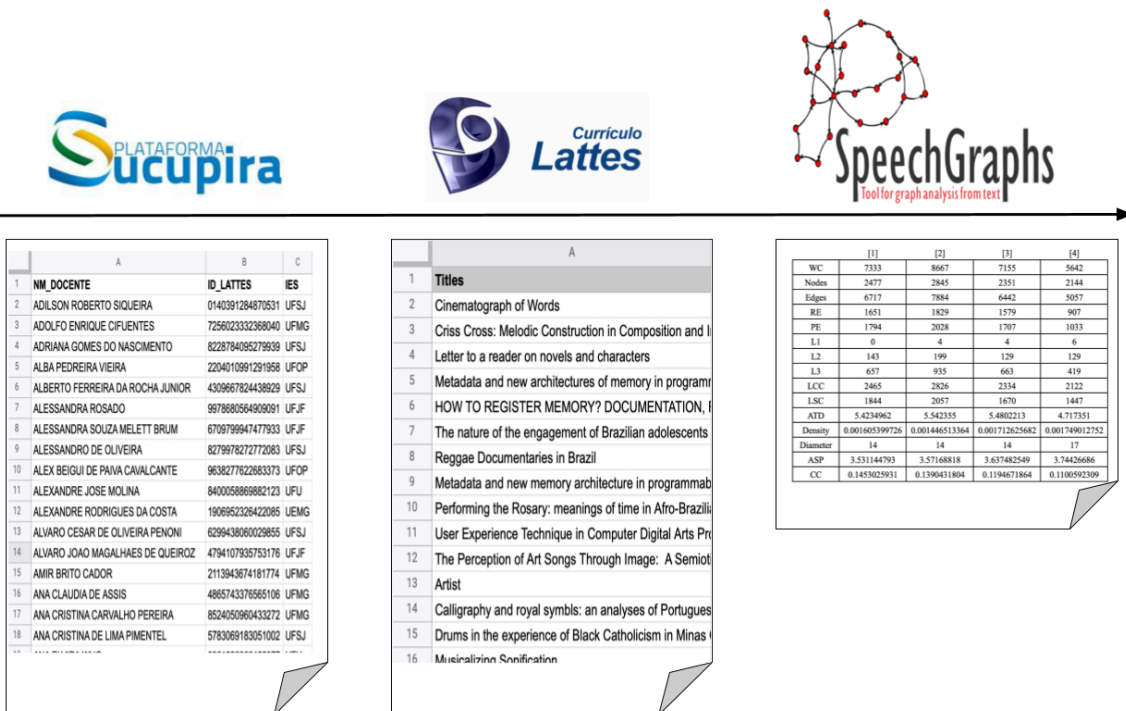
ANÁLISES DE TÍTULOS E REDE DE COCORRÊNCIA DE PALAVRAS

A escolha e definição do grupo amostral, que são docentes vinculados aos Programas de Pós-Graduação de Artes de Minas Gerais, está relacionada, principalmente, à identificação da baixa colaboração em pesquisas em Artes comparando-a a outras áreas. A relação dos docentes foi obtida na Plataforma Sucupira, que reúne dados do Sistema Nacional de Pós-Graduação. Foram identificados, para compor a amostra, docentes vinculados aos Programas de

Pós-Graduação de Artes de MG entre 2013 e 2023. Foram reunidos, em seguida, os currículos dos docentes na plataforma Lattes do CNPq. Com os currículos no formato XML, foram extraídos, por meio da ferramenta Lucy Lattes, dados de todas as publicações em periódicos, livros e capítulos de livros, incluindo o título e dados relacionados ao idioma e autores de cada publicação de trabalhos publicados. O número de autores foi utilizado para classificar as publicações em mono-autorais ou multi-autorais.

O processo de detecção do idioma foi automatizado. As listas de nós e de arestas, de títulos em inglês, foram obtidas através do Speech Graph, sem remoção de *stop words*. Estas listas foram utilizadas para geração do grafo no Gephi. A identificação de “palavras de função” e “palavras de conteúdo” ocorreu com base na listagem da frequência ordenada de ocorrência de palavras do Speech Graph. Os títulos também foram analisados, de um a um, para identificar a estruturação de subtítulos. Para identificação do uso de caracteres especiais utilizamos procedimentos algorítmicos.

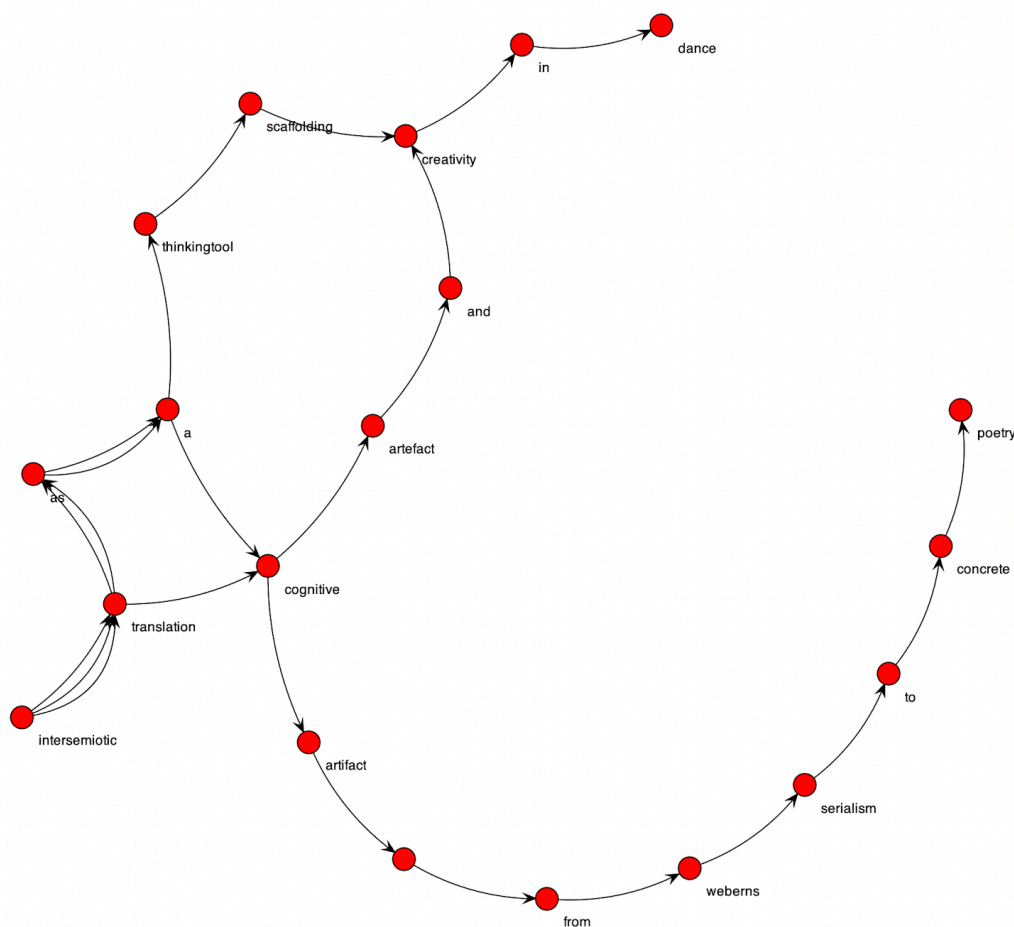
Figura 1 – As bases de dados e sistemas de modelagem usados para análise de títulos em inglês.



Títulos de trabalhos acadêmicos podem ser modelados como redes de coocorrência de palavras. Nesse nível de descrição, significado, tamanho e função gramatical das palavras são desconsiderados, assim como subsistemas, estruturas morfológicas, e outros componentes linguísticos e paralinguísticos. Quando esse nível é transformado em grafo, as palavras são representadas por nós e a sequência de coocorrência de palavras por arestas direcionadas. O título torna-se uma sequência de nós que começa no início de uma sentença, com arestas apontando para a próxima palavra-tipo (*type*), ou retorna a uma palavra que se repete (*token*), até o seu final (ver Figura 2). Através da análise de redes, é possível modelar diversas relações linguísticas como a ocorrência de palavras, frequência e repetição de palavras em textos, relações de coincidência, associações semânticas, padrões e relações temáticos. Ao examinar a estrutura de redes de palavras em textos, títulos ou corpus, é possível identificar, visual e estatisticamente, clusters ou coleções de palavras, e palavras mais conectadas e centrais, permitindo a análise de grandes volumes de dados.

Os títulos de publicações coautorais em periódicos, livros e capítulos de livros foram tratados manualmente, sendo excluídas duplicidades e informações excedentes registradas em campo indevido. As duplicidades podem acontecer quando pesquisadores coautores incluem a mesma publicação em seus currículos. Informações excedentes foram encontradas no texto de alguns títulos que incluem também informações sobre o meio de publicação. O conjunto revisado de títulos, em inglês, de artigos dos pesquisadores também foi processado pelo Speech Graphs. Esta ferramenta constrói redes de coocorrência de palavras como grafos direcionados, e não direcionados, em que os nós são as palavras e os arcos ligam nós que representam palavras que ocorreram em uma certa sequência (Figura 2).

Figura 2 – Exemplo de uma rede de coocorrência gerada no Speech Graphs.



Legenda: Os nós representam as palavras de títulos de três trabalhos e as setas indicam a sequência de ocorrências. Os três títulos modelados nessa rede de co-ocorrência são:
 (1) “Intersemiotic translation as a cognitive artifact – from Webern’s serialism to concrete poetry”
 (2) “Intersemiotic translation as a thinking-tool – scaffolding creativity in dance”
 (3) “Intersemiotic Translation, Cognitive Artefact, and Creativity”.

Como uma ferramenta computacional de análise de redes, o Speech Graphs usa o texto como entrada e métricas de rede como saída. Redes de coocorrência de palavras apresentam padrões que, em análise de redes, são interpretados através de muitas métricas — recorrência (arestas repetidas e paralelas), conectividade e centralidade (maior componente conectado, grau, densidade), distributividade (coeficiente de agrupamento, número de comunidades, modularidade), tamanho (número de arestas e nós, diâmetro, caminho mais curto). Diversos atributos foram calculados. A tabela 1 apresenta a relação entre os atributos de grafo e o que eles representam em redes de coocorrência.

Tabela 1 – Relação entre métricas de redes e o que representam em redes de coocorrência de palavras.

Atributo	Definição em análise de rede	Redes de coocorrência de palavras	Propriedades
Nós (N)	Número de nós no grafo.	Cada nó representa uma palavra no título de artigos, livros e capítulos.	Diversidade lexical
Arestas (E, <i>edges</i>)	Número de arestas no grafo.	Representa relações lexicais e de coocorrência e da sequência em que as palavras ocorrem no título.	Coocorrência e Sequência de ocorrência
Arestas repetidas (RE)	Número de arestas com a mesma direção ligando o mesmo par de nós.	A soma das arestas repetidas ligando as mesmas duas palavras.	Coocorrência
Arestas paralelas (PE)	Número de todas as arestas paralelas (com a mesma direção e com direções opostas) que ligam o mesmo par de nós.	A soma das arestas paralelas que ligam as mesmas duas palavras.	Coocorrência
Maior componente fortemente conectado (LSC)	Número de nós no sub grafo máximo em que todos os pares de nós podem ser alcançados a partir de qualquer outro, pertencente ao sub grafo direcionado.	Número de palavras mutuamente conectadas por um caminho de arestas.	Quão mutuamente conectadas as palavras estão
Grau total médio (ATD)	Soma das arestas de entrada e saída de cada nó dividida pelo número de nós.	Soma do grau total de todas as palavras, dividido pelo número de palavras.	Recorrência
Coefficiente de Agrupamento (CC)	Medida do grau em que os nós em um grafo tendem a se agrupar.	Número de palavras que estão diretamente ligadas a uma determinada palavra, X e também como essas palavras estão diretamente ligadas entre si.	Tendência de agrupamento, distributividade
Densidade	Número de arestas dividido por arestas possíveis.	Número de links diretos de palavras dividido por todos os links de palavras possíveis considerando todas as palavras.	Conectividade
Diâmetro	Comprimento do mais longo dos caminhos mais curtos entre todos os pares de nós em uma rede.	Comprimento do caminho que liga o par de palavras mais distante do texto.	Tamanho da rede
Caminho mais curto médio (ASP)	Comprimento médio do caminho mais curto entre pares de nós de uma rede.	Média dos caminhos mais curtos entre cada par de palavras.	Tamanho da rede

Outra análise realizada foi a avaliação das palavras dos títulos distinguindo “palavras de função” (*function words*) e “palavras de conteúdo” (*content words*) (Haspelmath, 2003), que são palavras que informam. Elas se diferenciam de palavras de função (*function words*), necessárias à estruturação gramatical. As palavras de função, também chamadas de *stop words*, foram identificadas com o auxílio da visualização de grau gerada com o Gephi e das listagens ordenadas geradas com o Speech Graph. As palavras de conteúdo também são analisadas para identificar um subconjunto das palavras de maior grau (mais utilizadas). Para o cálculo do grau de ocorrência das palavras com maior grau, foi realizada busca pelos radicais destas palavras: brazil (brazilian, afrobrazilian, brazilological, brazil's); music (musics, musical, ethnomusicology, dance/music, musician's, musicology, musicalizing); semiotic (intersemiotic, semiosis, semiotics, neurosemiotic, biosemiotic, semiotranslation); art (arts, artistic, artist, artists', artist's, artworks, artwork, artifact, artifacts, artefacts, artefact, art/pedagogy/research, artventurous); epilepsy (epileptiform, epileptic, epilepticus, epistemic, epileptologists, epileptogenesis, antiepileptic); model (modeling, models, modes, modularity, multimodal, multimodality, nonmodularity); analysis (analyses, analytical, metaanalysis, metaanalytic); peirce's (peirce, peircean); dance (dances, danceunlocking, danceable, dancing, dance/music, videodance, streetdance); cognitive (cognition); system (systems, systematic); translation (translating, transcriptions, semiotranslation); motor (motion, oculomotor).

3 RESULTADOS

A comunidade possui um histórico de 7484 publicações, 85,5% escritas em português, 8,74% em inglês e 5,76% em outros idiomas (Tabela 2). O percentual de publicações multi-autorais é maior em inglês, com uma diferença de 24,12% em relação às publicações multi-autorais em português, e 36,66% em em relação às publicações multi-autorais em espanhol.

Tabela 2 – Principais idiomas utilizados em publicações da comunidade.

	Mono-autoral	Multi-autoral	Total	% multi-autoral
Português	3520	2879	6399	44,99%
Inglês	202	452	654	69,11%
Espanhol	102	49	151	32,45%
Francês	70	23	93	24,73%
Alemão	20	4	24	16,67%
Italiano	17	2	19	10,53%
Outros	48	96	144	66,67%
Total	3979	3505	7484	46,83%

Na tabela 3, exibimos uma comparação quantitativa por tipo de publicação: artigos, livros e capítulos, em inglês e português. Para essas categorias, distinguimos publicações multi-autorais e mono-autorais. Os artigos em periódicos representam 52% das publicações em português e 75% das publicações em inglês. A diferença entre as publicações em inglês (452, multi-autorais x 202, mono-autorais), e em português (2879, multi x 3520, mono), também parece-nos relevante – em inglês, as publicações mais frequentes aparecem em artigos multi-autorais (54%); em português são predominantemente artigos mono-autorais (31%). Existe, portanto, uma maior concentração de artigos publicados em inglês (75%) do que em português (52%).

Tabela 3 – Quantidade e percentual de tipo de publicações em inglês e português e em mono e multi autoria.

	INGLÊS				PORTUGUÊS			
	Mono	Multi	Total	% multi	Mono	Multi	Total	% multi
ARTIGOS	137	353	490	72,04%	1979	1352	3331	40,59%
LIVROS	7	29	36	80,56%	1214	659	986	66,84%
CAPÍTULOS	58	70	128	54,69%	327	868	2082	41,69%
Total	202	452	654	69,11%	3520	2879	6399	44,99%

3.1 ESTATÍSTICA DOS TÍTULOS E DA COCORRÊNCIA DE PALAVRAS

Medidas de coocorrência de palavras se relacionam com a conectividade e com a distributividade da rede (arestas), e com a recorrência e a centralidade dos nós (palavras). Deve-se observar que existe uma perspectiva local, e outra global: medidas locais descrevem “a vizinhança de um nó ou a ocorrência de subgrafos (componentes)”, enquanto medidas globais descrevem “propriedades estatísticas de toda a rede” (Mota et al., 2012). A perspectiva local de recorrência e centralidade é observada principalmente na medida relativa ao grau total médio. A perspectiva global da rede, que se refere à conectividade, e à distributividade, é observada nas medidas de nós e arestas, arestas paralelas e repetidas, e na densidade. Também são identificadas características diretamente relacionadas aos títulos, como tamanho e tamanho médio. (Tabela 4).

Tabela 4 – Parâmetros de rede dos 654 títulos em inglês.

		Artigos		Capítulos		Livros	
		Mono	Multi	Mono	Multi	Mono	Multi
Títulos	T	137	353	58	70	7	29
Número de palavras	WC	1381	4095	559	708	52	234
Tamanho dos títulos	WC/T	10	12	10	10	7	8
Palavras distintas	W	772	1612	348	381	44	153
Palavras distintas por títulos	W/T	5,6	4,6	6	5,4	6,3	5,3
TTR (<i>type token ratio</i>)	W/WC	55,9%	39,4%	62,3%	53,8%	84,6%	65,4%
Arestas	E	1244	3742	501	638	45	205
Arestas por título	E/T	9,1	10,6	8,6	9,1	6,4	7,1
Arestas repetidas	RE	109	820	32	96	0	19
Arestas paralelas	PE	126	861	35	103	3	21
Média de arestas repetidas e paralelas pelo total de arestas	$((RE+PE)/2)/E$	9,4%	22,5%	6,7%	15,6%	3,3%	9,8%

Maior componente conectado	<i>LCC</i>	750	1606	339	368	32	151
Maior componente fortemente conectado	<i>LSC</i>	491	1172	156	211	3	70
Grau total médio	<i>ATD</i>	3,2	4,6	2,9	3,3	2	2,7
Grau por título	<i>ATD/T</i>	0,024	0,013	0,050	0,048	0,292	0,092
Densidade	<i>D</i>	0,004	0,002	0,008	0,007	0,044	0,016
Diâmetro	<i>DI</i>	19	14	17	23	14	18
Caminho mais curto médio	<i>ASP</i>	4,9	3,9	5,2	4,7	5,4	5,7
Coefficiente de agrupamento	<i>CC</i>	0,091	0,098	0,077	0,064	0,032	0,031

Baseados nos parâmetros da rede de palavras dos títulos de produções bibliográficas em inglês (Tabelas 1, 3 e 4), identificamos:

- O principal indicador de coocorrência de arestas entre palavras são as arestas paralelas e as arestas repetidas. A média de arestas repetidas e paralelas $((RE+PE)/2)$, dividida pelo número de títulos (T), indica a coocorrência de 17,45% de pares repetidos de palavras, e, em média, de 16% em publicações multi-autorais e de 6% em publicações mono-autorais.
- A razão entre a quantidade de palavras-tipos distintas (W , *types*) e de palavras-ocorrências (WC , *tokens*), uma medida usual em linguística computacional e processamento de linguagem natural, representa a diversidade de vocabulário. Obtém-se uma taxa de diversidade de 47,6% em artigos, 58% em capítulos e 75% em livros. Comparando as publicações multi e mono-autorais, o percentual médio de diversidade do vocabulário em mono-autorias é 16,5% superior em artigos, 8,4% em capítulos e 19,2% em livros.
- A média de palavras distintas por título (W/T) é maior em livros, capítulos e artigos, respectivamente, indicando maior variedade de palavras na mesma ordem. Para artigos, capítulos e livros é menor a quantidade de palavras distintas por título (W/T) em multi-autorias do

que em mono-autorias, indicando que em mono-autorias há menos restrições à escolha das palavras.

- A quantidade de palavras dos maiores componentes conectados (*LCC*), é aproximadamente a quantidade de palavras distintas, significando que cada rede de títulos, enquanto grafo não direcionado, está quase toda conectada, com poucas palavras fora do maior componente conectado. Por outro lado, há uma quantidade mais baixa de palavras dos maiores componentes fortemente conectados (*LSC*), indicando que as redes formadas pelos grafos direcionados de palavras (considerando a ordem das palavras) são redes com menor conexão.
- A densidade, de todos os tipos de publicação, encontra-se mais próxima de 0 do que de 1. A rede é, portanto, mais esparsa do que densa. Possuem mais densidade de conexão as palavras nos títulos de livros, capítulos e artigos, respectivamente.
- O tamanho médio dos títulos em inglês (*WC/T*) é de 9,5 palavras, sendo maior a média dos títulos de artigos multi-autoriais, 10 palavras, em relação aos mono-autoriais, 9 palavras. Esse resultado confirma os resultados de Hudson (2016), para quem o tamanho dos títulos é diretamente proporcional ao aumento do número de autores em quase todas as disciplinas.
- O tamanho da rede pode ser representado pela média dos caminhos mais curtos (Mota et al., 2018). Considerando a média dos caminhos mais curtos por tipo de publicação, a rede é maior para livros do que para capítulos e artigos. Considerando a média das redes de publicações em mono e multi-autoria, são maiores as redes de títulos mono-autoriais.
- Pode-se interpretar o tamanho da rede através do diâmetro, que mede o caminho mais longo entre os caminhos mais curtos. Com base no diâmetro, é maior o tamanho da rede de títulos de capítulos do que de artigos e livros. Considerando os tipos de publicações em multi e mono-autoria, as menores são as redes de artigos multi-autoriais e de livros mono-autoriais, e a maior é de capítulos multi-autoriais.

- Um indicador importante, relacionado à ocorrência de palavras, à conectividade e à distributividade da rede, é o grau total médio. O grau de um nó representa a quantidade de conexões que ele possui com os demais nós da rede. O grau total médio é a média do grau de todos os nós da rede. O grau total médio é maior para artigos, capítulos e livros respectivamente. Ao examinar o grau total médio, vemos que as palavras de títulos em inglês que apresentam maior grau médio ocorrem em artigos multi-autorais, e, o menor, em livros mono-autorais. Essa diferença apresenta correlação positiva com o número de publicações. No entanto, isso não equivale a uma rede de coocorrência mais conectada, já que a densidade média da rede é maior em livros, capítulos e artigos respectivamente.

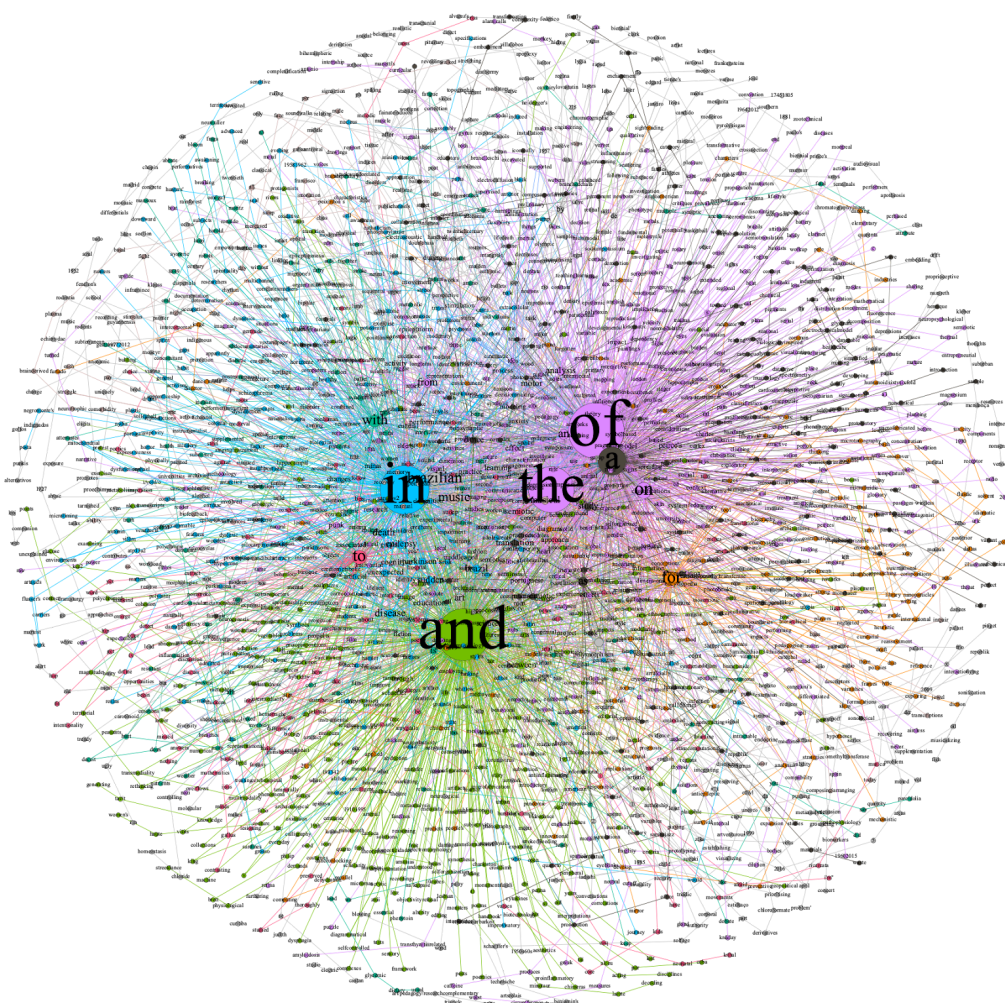
Sobre a ocorrência de arestas repetidas e paralelas, considerando que o total de publicações é diferente em cada subgrupo analisado, a diferença entre a ocorrência média destas arestas é 13,1% maior em artigos multi-autorais do que em artigos mono-autorais; 8,9% maior em capítulos e 6,5% maior em livros. Em outras palavras, arestas repetidas e paralelas ocorrem em maior proporção em publicações multi-autorais do que mono-autorais. Considerando os tipos de publicações, elas ocorrem em maior proporção em artigos, capítulos e livros, nesta ordem. Esse resultado pode estar relacionado à maior flexibilidade, ou autonomia, para decisão autoral sobre títulos de livros. Artigos acadêmicos e capítulos certamente sofrem maior restrição relacionada a políticas editoriais. Esses resultados parecem exibir um aumento da coerção exercido pelo grupo, ou subgrupo da comunidade, sobre um fenômeno ou tópico investigado. Quando solo, a variação tende a ser maior, ou a recorrência dos termos nos títulos tende a ser menor, um constrangimento daquilo que é comum aos interesses do grupo força um reaparecimento de termos nos títulos.²

² Trabalhos ulteriores devem explorar essa tendência (confirmando ou refutando Fleck) sobre o poder exercido sobre a variação de temas e tópicos, e revelados nos títulos e de outras propriedades relacionadas à coocorrência (e.g., arestas paralelas, arestas repetidas e grau total médio).

3.2 OCORRÊNCIA DE PALAVRAS DE FUNÇÃO E DE PALAVRAS DE CONTEÚDO

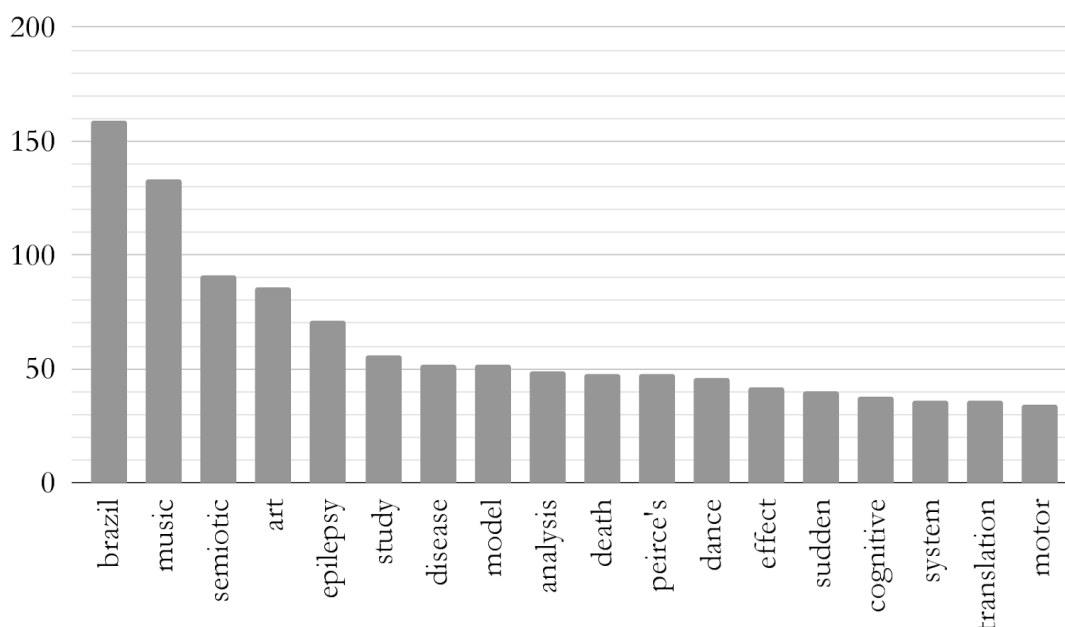
Na rede de coocorrência de palavras, abaixo (Figura 4), vemos que as palavras de função são as que possuem maior grau (*of, and, in, the, a, to, for, as, on, or, with*). Alinhados à conclusão de Hudson (2016, p. 871), de que o tamanho dos títulos é diretamente proporcional ao aumento do número de autores, em quase todas as disciplinas, identificamos que, cada novo autor, a depender da organização estabelecida para a pesquisa, corresponde a entrada de novos termos teóricos, novas abordagens, metodologias, ideias e preferências, que reflete-se no título.

Figura 4 – Rede de ocorrência de palavras dos títulos de artigos em coautoria em inglês.



O grau em que ocorrem “palavras de conteúdo” (*content words*) reflete o nível de atenção e coesão dos temas abordados nas publicações em determinados campos, como vemos no gráfico abaixo (Figura 5). Entre as palavras que mais ocorrem aparecem “music” e “art”. Embora não seja um termo que designe uma área de conhecimento sob estudo, o termo “semiotic” aparece em cinco ocorrências a mais do que arte. “Peirce”, “dance”, “cognitive”, “system” e “translation”, também são termos utilizados por pesquisadores com publicações relacionadas à semiótica. Nota-se que a comunidade possui publicações relacionadas à saúde, como se vê com as palavras “epilepsy”, “disease”, “death”, “effect”, “sudden” e “motor”. Ao analisar os currículos dos docentes, foi observado que as publicações foram realizadas em áreas próximas à musicoterapia, neurociência, biomedicina e educação física. “Brazil”, “study”, “model” e “analysis”, indicam abordagens metodológicas declaradas nos títulos.

Figura 5 – Palavras de conteúdo com maior ocorrência nos títulos em inglês.



Legenda: Possíveis variações de uma palavra foram consideradas no número de ocorrências. Foram consideradas palavras com grau superior a 30.

Estes são resultados iniciais, e têm o propósito, aqui, de indicar uma orientação metodológica que deve ser melhor desenvolvida em trabalhos futuros.

3.3 OUTROS ATRIBUTOS: SUBTÍTULOS E PONTUAÇÕES

Outro atributo relevante dos títulos é a sua estrutura como título único, e sua divisão em título e subtítulo. Na Figura 6, vemos a escala desenvolvida por Nagano (2015) com a proporção de títulos únicos por disciplina. É maior o percentual de títulos únicos (sem subtítulo), de publicações em Engenharia, Botânica e Economia; diferente do que vemos em Sociologia, Educação, Medicina, História, em que é mais usual a divisão do título em duas unidades, título e subtítulo. Em Geologia, o percentual de títulos, com ou sem subtítulos, é praticamente equivalente. Adicionamos aos resultados obtidos por Nagano (2015, Figura 6) o percentual de títulos únicos para as publicações em Artes, que somam 55,8% das publicações em inglês da comunidade de Artes.

Figura 6 – Proporção de títulos únicos por disciplina.



Legenda: Com 55,8% com título único, as publicações da comunidade de Artes estão próximas às médias de publicações da Economia, História, Geologia, Botânica, Medicina e Educação, e com distância aproximadamente similar da Engenharia e Sociologia. Os círculos vazios indicam *hard sciences* (Engenharias, Botânica, Geologia, Medicina) e os círculos preenchidos indicam *soft sciences* (Economia, História, Educação, Sociologia). Fonte: Adaptado de Nagano (2015).

Sobre a utilização de caracteres especiais, e sua importância, os autores têm sugerido que tratam-se de componentes cruciais na estruturação dos títulos. A função dos caracteres especiais, mais especificamente das pontuações, se relacionam com a modulação de ênfase, entonação, prosódia, e à restrições gramaticais e sintáticas, sendo uma função importante para “tornar gráfica a gramática” (Patt, 2013). Os caracteres especiais e de pontuação atuam como “artefatos cognitivos” e, em títulos da comunidade de

Artes, são usados de forma “não ortodoxa”.³ Mas, de modo geral, caracteres especiais são usados para satisfazer padrões estabelecidos. Na Tabela 5, são identificados os caracteres especiais mais frequentes nas publicações da comunidade.

Tabela 5 – Caracteres especiais nas publicações em inglês e ocorrência em mono e multi-autoria.

Caractere especial	Mono-autoral	Multi-autoral	Total	% multi
:	77	165	242	68%
-	30	142	172	83%
,	30	79	109	72%
()	11	26	37	70%
.	12	21	33	64%
?	7	20	27	74%
/	1	14	15	93%
“ ”	4	6	10	60%
&	2	1	3	33%
!	3	3	6	50%
Total	177	477	654	73%

Com base na Tabela 5, identificamos: (i) dois pontos (: *colon*) é a pontuação mais usual em títulos, sendo utilizado principalmente, em 84,7% das vezes em que ocorre, para separar título e subtítulo. Outros sinais de pontuação menos usuais na divisão de títulos e subtítulos são: hífen, ponto de interrogação, ponto de exclamação, ponto final. (ii) Títulos com subtítulos são predominantes nas publicações multi-autorais (68,2%). Este resultado difere daqueles obtidos por Lewison e Hartley (2005) de acordo com os quais títulos com dois pontos ocorrem mais frequentemente em trabalhos mono-autorais do que em multi-autorais. (iii) Apesar da alta ocorrência de hífen nos títulos, isso ocorre em apenas 14 títulos como divisor de título e subtítulo, ou seja, a grande

³ Os títulos são: [Des]envelhecer, uma reflexão sensório-cênica sobre a mulher e o envelhecimento; Po-éticas Pretas de [re]existência; Arte + Técnica = Artifício; Developing a Crossover Idiomatic Writing for the Double Bass: Composing/Arranging Playing, and ... Da-capo !

ocorrência de hífen deve-se a existência de palavras compostas (e.g. dance-unlocking, cross-modal, data-drive). (iv) A terceira pontuação mais utilizada é a vírgula. Já que os títulos sintetizam ideias, e que raramente surgem construções textuais discursivas, o uso da vírgula nos títulos ocorre majoritariamente como divisor de termos, elementos conceituais, e localidades (e.g. “Processes, artifacts, and semiosis”; “Rainforest Soundwalks: Ambiences of Bosavi, Papua New Guinea”). O uso da vírgula também aparece em estruturas estilísticas, como: “Today Your Style, Tomorrow The World”; “[...]Truth may hurt, but does deceit hurt more?”; “Brazilian Musics, Brazilian Identities”; “One struggle, one fight, all day, all night: Punk cartographies in the subway of São Paulo”. (v) O uso do parênteses ocorre, majoritariamente, para distinguir datas, locais e abreviações, mas ocorre, de forma não convencional, em casos como “Punk Fairytale: Popular Music, Media, and the (Re)Production of Gender”. (vi) O uso de ponto ocorre principalmente na finalização do título e em abreviações, mas também ocorre como divisor de título e subtítulo. (vii) O uso do ponto de interrogação (*question mark*) ocorre tanto em títulos únicos como em títulos com subtítulos. Nestes, ocorre tanto como divisor do subtítulo (e.g. “Never Mind the Pistols? The Legacy and Authenticity of the Sex Pistols in Portugal”), como no fim do subtítulo (e.g. “Artificial Emotions: Are We Ready for Them?”). (viii) A barra (/ *slash* ou *solidus*), ocorre como uma interseção ou proximidade entre termos (e.g. Composing/Arranging; agonist/antagonist; stroke/bleeding; singing/instrument; Art/pedagogy/research), no termo “Na+/K+” e no nome de uma publicação “I/VOID/O”.

CONCLUSÃO

Nós identificamos características e atributos observados nos títulos de produções bibliográficas da comunidade de Artes, através da análise de redes de coocorrência de palavras, e diversas descrições estatísticas. Como um complexo artefato semiótico, o título antecede, temporal e logicamente, uma fase fundamental da investigação, que é sua publicação e leitura, impressa ou virtual. Como um artefato, o título atua como um “atalho” cognitivo (*cognitive shortcut*) fundamental. Para Soler (2007, p. 91), “títulos em ciência espelham

um conjunto de requisitos que são cruciais para a construção, comunicação e progresso de novos conhecimentos”. Embora não mais subestimados, os títulos ainda recebem análises que tendem a ser superficiais, intuitivas, ou superficialmente descritivas, com ênfase em variáveis específicas (Berkenkotter; Huckin, 2016; Buxton; Meadows, 1977), ou normativas (Gastel; Day, 2012), mais interessadas em formas e estruturas eficazes associadas ao aumento de downloads e citações, em diversas áreas. É sabido que a escolha e a decisão sobre o título, em publicações acadêmicas, envolve “processos de negociação” entre autores e entre comunidades de pesquisa (Milojević, 2017). Mas ainda não sabemos claramente como tais processos têm lugar. As propriedades dos títulos variam consideravelmente entre disciplinas (Hudson, 2016) — “pertencer a alguma disciplina é o determinante mais forte para a extensão dos títulos e a ocorrência de diferentes formas de títulos” (Milojević, 2017) —, e entre áreas. Títulos funcionam como artefatos de identidade acadêmica, científica e disciplinar. Uma tipologia de diferentes classes e formas de títulos, em diferentes disciplinas, é capaz de sugerir a existência de diferentes normas estabelecidas em diversos campos. A existência de estruturas e propriedades linguísticas e paralinguísticas, em títulos de publicações, revelam padrões e tendências das investigações.

A definição de um idioma para divulgação dos resultados da pesquisa é fortemente constrangida por normas e convenções sociais, e pode ser considerada uma das principais variáveis entre diferentes comunidades de prática científica. Uma vez que o português é o idioma predominantemente usado, a atenção dada pela comunidade a publicações neste idioma sugere que ela não exhibe resultados de impacto internacional. Isso pode estar relacionado às dificuldades para publicar internacionalmente, e também a uma certa resistência da comunidade sobre o caráter hegemônico do inglês como idioma padrão em publicações acadêmicas.

Ao concentrar 85,5% das publicações no português, a comunidade reduz significativamente a disseminação global dos resultados da pesquisa. Do total de publicações, 8,74% são em inglês, indicando que a comunidade possui baixa adesão ao exibir seus resultados em revistas com impacto internacional.

Artigos publicados em inglês tem potencial para atingir um maior número de citações do que aqueles publicados em outros idiomas (Di Bitetti; Ferreras, 2017), o que é benéfico à própria comunidade. Entre artigos, capítulos e livros, as publicações mais frequentes em inglês são de artigos multi-autorais (54%), e em português são de artigos mono-autorais (31%). Existe, portanto, uma maior concentração de artigos publicados em inglês (75%) do que em português (52%). Também foi identificado que existe uma tendência 24% maior de publicações multi-autorais serem realizadas em inglês.

A metodologia usada (análise de rede de coocorrência de palavras em publicações em inglês) revelou que: (1) a rede de títulos é praticamente toda conectada; (2) a razão entre a quantidade de palavras-tipos e de palavras-ocorrências (relação *type/token*), apresenta uma taxa de diversidade de vocabulário de 47,6% em artigos, 58% em capítulos e 75% em livros. (3) Do total de arestas, 17,45% são repetidas ou paralelas, e é maior o número de arestas paralelas e repetidas, em publicações multi-autorais. (4) As palavras dos títulos se repetem em 52,37% dos artigos, 41,9% dos capítulos e em 25% dos livros. Comparando as publicações multi e mono-autorais, a média de palavras repetidas é, respectivamente, 47,1% e 32,4%. (5) É maior a quantidade de nós nos títulos de publicações mono-autorais. (6) O tamanho médio dos títulos é de 9,5 palavras, sendo maiores os títulos de artigos multi-autorais, com a média de 12 palavras. (7) Apresentam maior grau médio os artigos multi-autorais (4,643), e o menor, livros mono-autorais (2,045). Essa diferença apresenta correlação positiva com o número de publicações. No entanto, isso não equivale a uma rede de coocorrência mais conectada, já que a densidade média da rede é mais esparsa do que coesa. (8) A média de arestas por título (E/T) é maior em artigos, seguido por capítulos e livros multi-autorais. (9) O número de arestas paralelas e repetidas é maior em publicações multi-autorais. Esse resultado pode estar relacionado à maior autonomia para decisão autoral de títulos em livros. (10) Dois pontos (: *colon*) é a pontuação mais usual em títulos, e, em percentual menor, hífen (5,6%), ponto de interrogação (3,2%), ponto de exclamação (0,8%) e ponto final (0,8%) também são usados como divisores de títulos e subtítulos. Pontos de

interrogação (*question marks*) são utilizados em 2% dos subtítulos em inglês. (11) Títulos com subtítulos são predominantes nas publicações multi-autorais (68,2%), e dois pontos é a pontuação mais utilizada (59,4%). Este resultado difere daqueles obtidos por Lewison e Hartley (2005) que afirmam que títulos com dois pontos ocorrem com maior frequência em trabalhos mono-autorais do que em coautorais.

Os resultados exibem um aumento da coerção exercida pelo grupo, ou sub-grupo, da comunidade, sobre um fenômeno ou tópico investigado. Quando solo, a variação tende a ser maior. Um constrangimento daquilo que é comum aos interesses do grupo aparentemente força o reaparecimento de termos nos títulos. A depender da organização estabelecida para a pesquisa, cada novo autor corresponde à entrada de novos termos teóricos, novas abordagens, metodologias, e preferências. Podemos também relacionar aos maiores títulos de artigos multi-autorais, em inglês e em português, a atividade autoral distribuída (*group mind*), que aumenta tanto o número de palavras, quanto a variedade lexical dos termos da pesquisa. Os resultados também sugerem que existe maior autonomia para decisão autoral sobre títulos de livros, artigos acadêmicos e capítulos certamente sofrem maior restrição relacionada a políticas editoriais.

Como já afirmamos, os resultados mostrados aqui são iniciais. Eles sugerem um amplo conjunto de questões sobre componentes tão cruciais quanto os títulos de publicações na comunidade acadêmica de Artes. É relevante que, baseado na metodologia usada, os resultados sejam rigorosamente comparados com diversas abordagens já desenvolvidas sobre a natureza e a estrutura de títulos formulados em áreas distantes.

REFERÊNCIAS

ALBERY, Gregory F.; KIRKPATRICK, Lucinda; FIRTH, Josh A.; BANSAL, Shweta. Unifying Spatial and Social Network Analysis in Disease Ecology. *Journal of Animal Ecology*, v. 90, n. 1, p. 45-61, 2021. Disponível em: <https://doi.org/10.1111/1365-2656.13356>.

ARAÚJO, Miguel B.; ROZENFELD, Alejandro; RAHBEK, Carsten; MARQUET, Pablo A. Using Species Co-occurrence Networks to Assess the Impacts of

Climate Change. *Ecography*, v. 34, n. 6, p. 897-908, 2011. Disponível em: <https://doi.org/10.1111/j.1600-0587.2011.06919.x>.

BARABÁS, Albert-László. *Network Science*. Cambridge: Cambridge University Press, 2016. Disponível em: <http://networksciencebook.com>.

BAUWENS, Thomas; SCHRAVEN, Daan; DREWING, Emily; RADTKE, Jörg; HOLSTENKAMP, Lars; GOTCHEV, Boris; YILDIZ, Özgür. Conceptualizing Community in Energy Systems: a Systematic Review of 183 Definitions. *Renewable and Sustainable Energy Reviews*, v. 156, p. 11-1999, 2022. Disponível em: <https://doi.org/10.1016/j.rser.2021.111999>.

BERKENKOTTER, Carol; HUCKIN, Thomas N. *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. London, New York: Routledge, 2016.

BUXTON, Andrew B.; MEADOWS, Arthur J. The Variation in the Information Content of Titles of Research Papers with Time and Discipline. *Journal of Documentation*, v. 33, n. 1, p. 46–52, 1977. Disponível em: <https://doi.org/10.1108/eb026633>.

CHEN, Chaomei; SONG, Min. Visualizing a Field of Research: a Methodology of Systematic Scientometric Reviews. *PLoS One*, v. 14, n. 10, e0223994, 2019. Disponível em: <https://doi.org/10.1371/journal.pone.0223994>.

CIAMPELLI, Silvia; DE BOER, Janna N.; VOPPEL, Alban E.; CORONA HERNANDEZ, Hugo; BREDEROO, Sanne G.; VAN DELLEN, Edwin; MOTA, Natalia B.; SOMMER, Iris E. C. Syntactic Network Analysis in Schizophrenia-spectrum Disorders. *Schizophrenia Bulletin*, v. 49, n. Supplement_2, p. S172-S182, 2023. Disponível em: <https://doi.org/10.1093/schbul/sbac194>.

DENG, Shengli; XIA, Sudi. Mapping the Interdisciplinarity in Information Behavior Research: a Quantitative Study using Diversity Measure and Co-occurrence Analysis. *Scientometrics*, v. 124, p. 489-513, 2020. Disponível em: <https://doi.org/10.1007/s11192-020-03465-x>.

DI BITETTI, Mario S.FERRERAS, Julián A. Publish (in English) or Perish: The Effect on Citation Rate of Using Languages other than English in Scientific Publications. *Ambio*, v. 46, n. 1, p. 121-127, 2017. Disponível em: <https://doi.org/10.1007/s13280-016-0820-7>.

GARCIA, D. C. F. GATTAZ, C. C.; GATTAZ, N. C. The relevance of title, abstract and keywords for scientific paper writing. *Revista de Administração Contemporânea*, v. 23, p. 1-9, 2019.

GASTEL, Barbara; DAY, Robert. A. *How to write and publish a scientific paper*. Cambridge: Cambridge University Press, 2012. Disponível em: https://assets.cambridge.org/97811076/70747/frontmatter/9781107670747_frontmatter.pdf.

HARTLEY, James. *Academic Writing and Publishing: a Practical Handbook*. London, New York: Routledge, 2008.

HASPELMATH, Martin. *The Geometry of Grammatical Meaning: Semantic Maps and Cross-linguistic Comparison*. In: TOMASELLO, Michael (org.). *The New Psychology of Language*. Mahwah: Lawrence Erlbaum, 2003.

HUDSON, John. *An Analysis of the Titles of Papers Submitted to the UK REF in 2014: Authors, Disciplines, and Stylistic Details*. *Scientometrics*, v. 109, n. 2, p. 871-889, 2016. Disponível em: <https://doi.org/10.1007/s11192-016-2081-4>.

IZZO, Flavio; OLIVEIRA, Elias; BADUE, Claudine. *Named Entities as a Metadata Resource for Indexing and Searching Information*. *Lecture Notes in Networks and Systems*. Springer Cham, v. 418, p. 838-848, 2022. Disponível em: https://doi.org/10.1007/978-3-030-96308-8_78.

JACKSON, Matthew A.; BONDER, Marc Jan; KUNCHEVA, Zhana; ZIERER, Jonas; FU, Jingyuan; KURILSHIKOV, Alexander; WIJMENGA, Cisca; et al. *Detection of Stable Community Structures within Gut Microbiota Co-occurrence Networks from Different Human Populations*. *PeerJ*, v. 6, e4303, 2018. Disponível em: <https://doi.org/10.7717/peerj.4303>.

JAMALI, Hamid. R.; NIKZAD, Mahsa. *Article Title Type and its Relation with the Number of Downloads and Citations*. *Scientometrics*, v. 88, n. 2, p. 653-661, 2011. Disponível em: <https://doi.org/10.1007/s11192-011-0412-z>.

LEWISON, Grant; HARTLEY, James. *What's in a Title? Numbers of Words and the Presence of Colons*. *Scientometrics*, v. 63, n. 2, p. 341-356, 2005. Disponível em: <https://doi.org/10.1007/s11192-005-0216-0>.

LOYAL, Joshua Daniel; CHEN, Yuguo. *Statistical Network Analysis: a Review with Applications to the Coronavirus Disease 2019 Pandemic*. *International Statistical Review*, v. 88, n. 2, p. 419-440, 2020. Disponível em: <https://doi.org/10.1111/insr.12398>.

LOZANO, Sebastián; CALZADA-INFANTE, Laura; ADENSO-DÍAZ, Belarmino; GARCÍA, Silvia. *Complex Network Analysis of Keywords Co-occurrence in the Recent Efficiency Analysis Literature*. *Scientometrics*, v. 120, n. 2, p. 609-629, 2019. Disponível em: <https://doi.org/10.1007/s11192-019-03132-w>.

MILOJEVIĆ, Staša. *The Length and Semantic Structure of Article Titles—Evolving Disciplinary Practices and Correlations with Impact*. *Frontiers in Research Metrics and Analytics*, v. 2, p. 2, 2017. Disponível em: <https://doi.org/10.3389/frma.2017.00002>.

MOTA, Natalia B.; VASCONCELOS, Nivaldo A.; LEMOS, Nathalia; PIERETTI, Ana C.; KINOUCI, Osame; CECCHI, Guillermo A.; COPELLI, Mauro. *Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis*. *PLoS One*, v. 7, n. 4, e 34928, 2012. Disponível em: <https://doi.org/10.1371/journal.pone.0034928>.

MOTA, Natália Bezerra, SIGMAN, Mariano; CECCHI, Guillermo; COPELLI, Mauro; RIBEIRO, Sidarta. The Maturation of Speech Structure in Psychosis is Resistant to Formal Education. *Schizophrenia*, v. 4, n. 1, pp. 1-10, 2018. Disponível em: <https://doi.org/10.1038/s41537-018-0067-3>

NAGANO, Robin. L. Research Article Titles and Disciplinary Conventions: a Corpus Study of Eight Disciplines. *Journal of Academic Writing*, v. 5, n. 1, pp. 133-144, 2015. Disponível em: <https://doi.org/10.18552/joaw.v5i1.168>

NAIR, Lakshmi Balachandran; Gibbert, Michael. What Makes a 'Good' Title and (How) Does it Matter for Citations? A Review and General Model of Article Title Attributes in Management Science. *Scientometrics*, v. 107, n. 3, pp. 1331-1359, 2016. Disponível em: <https://doi.org/10.1007/s11192-016-1937-y>

NEWMAN, Mark E. Coauthorship Networks and Patterns of Scientific Collaboration. *Proceedings of the National Academy of Sciences*, v. 101, n. suppl_1, pp. 5200-5205, 2004. Disponível em: <https://doi.org/10.1073/pnas.0307545100>

PATT, Sebastian. Punctuation as a means of medium-dependent presentation structure in English: Exploring the Guide Functions of Punctuation. Tübingen, Narr Verlag, 2013.

RADHAKRISHNAN, Srinivasan; ERBIS, Serkan; ISAACS, Jacqueline A.Kamarthi, Sagar. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PloS one*, v. 12, n. 3, e 0172778, 2017.

SOLER, Viviana. Writing Titles in Science: an Exploratory Study. *English for Specific Purposes*, v. 26, n. 1, pp. 90-102, 2007. Disponível em: <https://doi.org/10.1016/j.esp.2006.08.001>

SUBOTIC, Sinisa; MUKHERJEE, Bhaskar. Short and Amusing: the Relationship Between Title Characteristics, Downloads, and Citations in Psychology Articles. *Journal of Information Science*, v. 40, n. 1, pp. 115-24, 2014. Disponível em: <https://doi.org/10.1177/0165551513511393>

YUAN, Chenxi; LI, Guoyan; KAMARTHI, Sagar; JIN, Xiaoning; MOGHADDAM, Mohsen. Trends in Intelligent Manufacturing Research: a Keyword Co-occurrence Network based Review. *Journal of Intelligent Manufacturing*, v. 33, n. 2, pp. 425-439, 2022. Disponível em: <https://doi.org/10.1007/s10845-021-01885-x>

ZHU, Chunyan; CAI, Tingting; JIN, Ying; CHEN, Jiayun; LIU, Guoqiang; XU, Sheng; SHEN, Rong; et al. Artificial Intelligence and Network Pharmacology based Investigation of Pharmacological Mechanism and Substance Basis of Xiaoke Wan in Treating Diabetes. *Pharmacological Research*, v. 159, pp. 104935, 2020. Disponível em: <https://doi.org/10.1016/j.phrs.2020.104935>

ABSTRACT

The title is considered the most important component of an academic publication, as it often communicates the research object, its scope, methods, and objectives.

Objective: This study examines various properties related to the titles of articles, books, and book chapters—both single- and multi-authored—produced by the academic arts community in Minas Gerais (MG). The corpus includes publications with at least one author affiliated, between 2013 and 2023, with a postgraduate arts program (PPG) in the state of MG. **Methodology:** The methodology employs network analysis, focusing on statistical and topological properties related to word co-occurrence. Key features such as connectivity, centrality, and distributivity, as well as title length, were analyzed. The co-occurrence network analyses focused on titles of publications written in English. **Results:** Regarding the variety of languages, out of 7484 publications, 85.5% were in Portuguese, 8.74% in English, and 5.76% in other languages. The characterization of word co-occurrence in titles, based on normalized repeated and parallel edges, shows 17.45% of repeated and parallel words in sequence. These repeated and parallel edges occur more frequently in multi-authored publications compared to single-authored ones and in articles, chapters, and books, respectively. Based on the ratio between word types and word tokens in the titles, the vocabulary diversity rate is 47.6% for articles, 58% for chapters, and 75% for books. **Conclusions:** Greater autonomy in authorial decision-making was observed in book titles compared to academic articles and chapters, which are more constrained by editorial policies. The findings suggest an increase in coercion exerted by the group or sub-group within the academic community concerning specific phenomena or topics. Conversely, multi-authored article titles tend to reflect a distributed authorial activity (a "group mind"), characterized by an increase in both the number of words and lexical variety in the research terms.

Descriptors: Word Co-occurrence. Titles. Network Analysis. Art.

Agradecimentos

Queiroz agradece ao CNPq pelo apoio através da bolsa produtividade pesquisa (308355/2023-7), e do projeto regular de pesquisa (Grupos Emergentes; 404770/2023-1).

Declaração de contribuição dos autores (CRediT authorship contribution statement)

Rafael Werner – Curadoria de dados; Análise Formal; Investigação; Metodologia; Recursos; Validação; Visualização; Redação – rascunho original; Redação– revisão e edição.

Angelo Loula – Conceituação; Curadoria de dados; Análise Formal; Investigação; Metodologia; Recursos; Programas; Supervisão; Redação – revisão e edição.

Thiago Sena – Programas; Validação.

João Queiroz – Conceituação; Curadoria de dados; Análise Formal; Investigação; Metodologia; Administração do projeto; Recursos; Supervisão; Visualização; Redação – rascunho original; Redação – revisão e edição.

Declaração de conflito de interesse

Declaramos que não há qualquer conflito de interesse, em potencial, neste estudo.

Declaração de disponibilidade de dados da pesquisa

Todo o conjunto de dados de apoio aos resultados deste estudo foi disponibilizado através da Plataforma Sucupira e na Plataforma Lattes.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.