

ANÁLISE DE COOCORRÊNCIA DE PALAVRAS NA PESQUISA BRASILEIRA EM HIV/AIDS INDEXADA NA WEB OF SCIENCE NO PERÍODO 1993-2020

CO-WORD ANALYSIS IN BRAZILIAN HIV/AIDS RESEARCH INDEXED ON THE WEB OF SCIENCE IN THE 1993-2020 PERIOD

Rafael Antunes dos Santos^a
Eliseo Berni Reategui^b
Sonia Elisa Caregnato^c

RESUMO

Objetivo: A pesquisa apresenta um estudo exploratório-descritivo sobre a produção científica brasileira em HIV/AIDS indexada na Web of Science no período 1993-2020. O estudo analisou, com base em indicadores bibliométricos de coocorrência de palavras-chave, uma percepção alinhada à expressão autoral da pesquisa brasileira em HIV/AIDS a partir de informações obtidas no campo de metadados Descriptor (DE) dos 2895 artigos recuperados. **Metodologia:** A pesquisa procurou ampliar o conhecimento das dimensões relativas às temáticas de pesquisa em HIV/AIDS, mas também se preocupou em demonstrar o potencial dos procedimentos da análise de coocorrência de palavras. Os resultados ainda são amparados por indicadores de atividade científica acerca dessa produção ao longo dos anos. Obteve medidas de frequência; de intensidade, pela identificação de clusters; e de similaridade, a partir do Cosseno de Salton em sete agrupamentos temporais: 1993-1997, 1998-2001, 2002-2005, 2006-2009, 2010-2013, 2014-2017 e 2018-2020. **Resultados:** Apesar de manter crescimento exponencial até 2011, a pesquisa brasileira em HIV/AIDS demonstra sinais de crescimento negativo nos últimos agrupamentos. A análise de coocorrência de palavras identificou as palavras-chave do núcleo básico do campo do HIV/AIDS em cada período a partir da definição de uma ocorrência mínima nos agrupamentos. O mainstream de palavras-chave é definido pela intensidade e similaridade dos pares de palavras associadas ao longo do tempo. **Conclusões:** O estudo conclui que as palavras-chave usadas por autores podem demonstrar aspectos da evolução de uma área de pesquisa, assim como foi observado na pesquisa brasileira em HIV/AIDS do período 1993-2020.

Descritores: HIV/AIDS. Brasil. Produção científica. Bibliometria. Análise de

^a Doutorando em Informática na Educação e Mestre em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: rderafa@gmail.com

^b Doutor em Ciência da Computação pela University of London. Docente da Faculdade de Educação e do Programa de Pós-Graduação em Informática na Educação (UFRGS). E-mail: eliseoreategui@gmail.com

^c Doutora em Information Studies pela University of Scheffield. Docente da Faculdade de Biblioteconomia e Comunicação, do Programa de Pós-Graduação em Comunicação e do Programa de Pós-Graduação em Ciência da Informação (UFRGS). E-mail: sonia.caregnato@ufrgs.br

coocorrência de palavras.

1 INTRODUÇÃO

Os primeiros sinais de novas informações identificadas nas estatísticas de saúde sobre HIV/AIDS foram publicados em 1981 no *Morbidity and Mortality Weekly Report*, um boletim que fornece dados epidemiológicos sobre a progressão de doenças em nível mundial (BASTOS, 2002). A AIDS (*Acquired Immunodeficiency Syndrome*) foi reconhecida em 1982, nos Estados Unidos. O seu agente etiológico foi denominado HIV (*Human Immunodeficiency Virus*) em 1983, na França. A pesquisa sobre AIDS pode estar circunscrita num quadro de superação da ciência, “[...] envolto por fenômenos novos e insuspeitados de emergência por descobertas via pesquisa científica” (KUHN, 2017, p. 77), reconhecidamente identificados por “anomalias”. A descoberta da AIDS ocasionou um entrelaçamento de variáveis médicas, sociais, políticas, culturais e simbólicas sem precedentes em relação a outras doenças como a cólera ou a tuberculose (PARKER, 2000).

O Brasil continua a ser um dos países que mais investe em políticas públicas de controle epidêmico e valorização das ações em saúde pública no âmbito do HIV e da AIDS, apesar destas informações serem pouco veiculadas na mídia (JESUS, 2018). No entanto, a epidemia da AIDS no Brasil está longe de ser controlada (BRASIL, 2020). O país ainda possui taxas elevadas de infecção em diversos estratos sociais. Este é um desafio que a ciência tem se proposto a investigar, pois o enfrentamento a essa epidemia ainda é um tema urgente para a agenda nacional de pesquisa.

A produção científica sobre HIV/AIDS acompanha o percurso da doença em termos epidemiológicos. As taxas de literatura científica sobre AIDS também cresceram exponencialmente. No campo da avaliação da pesquisa científica, esse aumento motiva os pesquisadores a realizarem estudos bibliométricos (GANHOR *et al.*, 2020). A maioria dos estudos métricos de informação sobre AIDS analisam os principais periódicos, países e campos disciplinares, através da identificação da visibilidade da literatura. Essas informações servem de apoio

aos pesquisadores, agentes de saúde e responsáveis por programas de políticas públicas em HIV/AIDS. No entanto, há poucos estudos que empregam técnicas bibliométricas para estudar os temas das pesquisas publicadas.

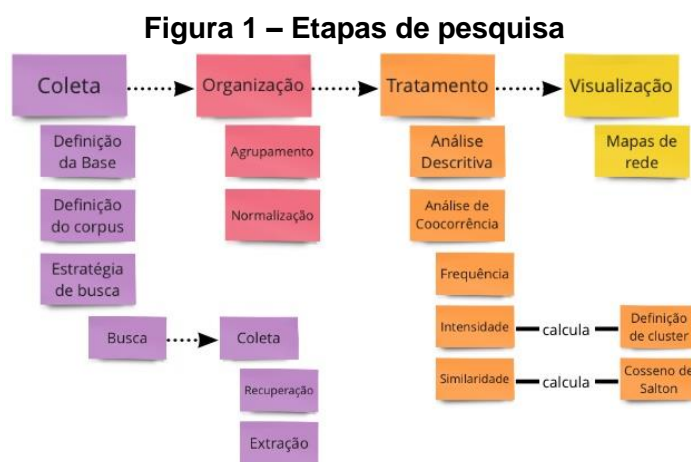
Assim, buscou-se um mapeamento das características de coocorrência de palavras em artigos publicados por autores de instituições de pesquisa brasileiras no período de 1993 a 2020. A análise de coocorrência de palavras é um sistema de descoberta que pode gerar conhecimento sobre áreas de pesquisa, dependendo dos níveis de investigação: palavras-chave definidas por autores, descritores do processo de indexação ou ainda títulos, resumos e textos. A análise de coocorrência de palavras, então, baseia-se nos pressupostos de que os autores científicos escolhem seus termos de um repertório abrangente e codificado, que eles usam termos diferentes sempre que postulam relações não triviais entre os conceitos, e que o reconhecimento das mesmas relações de diferentes autores diz algo sobre a estrutura temática de uma área de pesquisa (DE BELLIS, 2009).

O estudo aqui apresentado leva em consideração a produção disponível na coleção principal da base multidisciplinar *Web of Science* (WoS) mantida pela *Clarivate Analytics*. Para fazer parte da WoS, determinada publicação precisa cumprir uma série de requisitos como a avaliação por pares, a periodicidade confiável, a presença de um corpo editorial qualificado e internacional, e principalmente, indicar as referências dos trabalhos utilizados nas pesquisas (citações). A produção em questão possui a visibilidade internacional pelo fato de estar indexada na WOS, base de dados amplamente utilizada para realizar estudos métricos de informação (NÁJERA-SÁNCHEZ, 2020; ZUPIC; ČATER, 2015). Fundamentalmente, os estudos bibliométricos têm como objeto de estudo a informação bibliográfica (MACIAS-CHAPULA, 1998; OLIVEIRA; GRÁCIO, 2011; CURTY; DELBIANCO, 2020). Este artigo resgata e atualiza dados de pesquisa de mestrado (SANTOS, 2015), estudo até então não publicado em nenhum periódico científico. A pesquisa replica a metodologia do documento anteriormente citado com informações atualizadas e expandidas, para comunicar e contribuir com a compreensão sobre a utilização do termo HIV/AIDS na literatura científica do país ao longo do tempo.

O objetivo geral da pesquisa é analisar a produção científica brasileira em HIV/AIDS por meio do método da análise de coocorrência de palavras-chave de artigos de pesquisadores de instituições brasileiras de ensino e pesquisa. Mais especificamente, busca-se revelar as redes de coocorrência dos termos, analisar a frequência das palavras-chave e as matrizes de coocorrência destas palavras em virtude da intensidade, da formação de clusters e da similaridade entre as palavras. A pesquisa pretende contribuir para o conhecimento das dimensões relativas às temáticas de pesquisa em HIV/AIDS no cenário brasileiro.

2 METODOLOGIA

O estudo realizado é do tipo exploratório-descritivo com um caráter quantitativo e caracterizado no âmbito da bibliometria. A pesquisa utilizou como medidas os indicadores de atividade científica e os indicadores de relação (CALLON; COURTIAL; PENAN, 1995). A definição das etapas de pesquisa em uma metodologia desse nível responde a uma grande variedade de procedimentos. Os autores Khaldi e Prado-Gascó (2021) definem quatro etapas nesse processo: pesquisa, avaliação, síntese e análise. Nájera-Sánchez (2020) também estabelece quatro etapas: identificação do tema, visualização da rede temática, evolução do tema e análise do desempenho. Já Giannakos *et al.* (2020) indicam cinco etapas: busca, recuperação, padronização, análise e visualização. Na presente pesquisa, o esforço foi dividido em quatro etapas: coleta, organização, tratamento e visualização, conforme a Figura 1.



Fonte: Dados da pesquisa e Miro.

O *corpus* foi composto pelos artigos indexados na Coleção Principal da WoS. Definiu-se um recorte institucional e temporal compreendido pelas contribuições de pesquisadores de instituições de pesquisa brasileiras escritas entre 1993 e 2020 que contivessem nas palavras-chave, nos títulos, nos resumos ou nos descritores as expressões AIDS e HIV, considerando suas extensões. Definiu-se a formação do *corpus* como intencional por critérios pré-determinados. Para um artigo integrar o *corpus* deste estudo, foi preciso que estivesse descrito no campo de filiação (C1) dos autores as palavras *Brazil* ou *Brasil*. A estratégia de busca aconteceu em duas etapas. A primeira etapa utilizou expressão de busca $TS=(aids\ OR\ acquired\ immunodeficiency\ syndrome)\ AND\ TS=(hiv\ OR\ human\ immunodeficiency\ virus)$. Por meio dessa expressão, foram localizados artigos que considerassem as duas dimensões temáticas da pesquisa. A segunda etapa envolveu a coleta propriamente dita através da recuperação e da extração dos dados. A expressão de busca retornou 2895 (dois mil oitocentos e noventa e cinco) metadados de artigos com resultados de pesquisas brasileiras e outras, de caráter colaborativo com instituições de pesquisa internacionais. Após a coleta e a organização, os dados foram tratados segundo a metodologia do *Laboratorio de Estudios Métricos de la Información* (LEMI), da Universidade Carlos III de Madrid (2013), na Espanha, com o objetivo de caracterização do campo do HIV/AIDS brasileiro a partir de medidas relacionadas com o crescimento da produção científica. A outra etapa do tratamento foi a aplicação do método de coocorrência das palavras-chave com os *softwares Bibexcel e Excel*.

Devido ao amplo recorte temporal, buscou-se a manipulação dos dados em análises por agrupamentos temporais: 1993-1997; 1998-2001; 2002-2005; 2006-2009; 2010-2013; 2014-2017 e 2018-2020 (MORKUNAS; MOORE; DUNCAN, 2018). Nesta etapa, a referência para os agrupamentos foi o ano informado na publicação. Assim, a pesquisa definiu os indicadores que foram utilizados na análise quantitativa, adaptados de Callon, Courtial e Penan (1995) e Persson, Danell e Schneider (2009):

- a) contagem e *rank* de palavras: calcula a frequência de palavras com uso dos *softwares Bibexcel e Excel*;

b) intensidade e formação de *cluster*: calcula a força de associação entre as palavras segundo a conectividade do nó, com uso do *software Vosviewer*;

c) similaridade: calcula o coeficiente de um par de palavras-chave em potencial em diferentes agrupamentos temporais, com uso do *software Bibexcel*.

Na sequência, as matrizes simétricas serviram de insumo para as análises multivariadas. Assim, a pesquisa procedeu com a adaptação da matriz de dados brutos para uma matriz simétrica de similaridade bilateral com o uso de medidas do cosseno de Salton (RUIZ-BAÑOS *et al.*, 1999a; 1999b), escrita da seguinte forma:

$$S_{xy} = \frac{P_{xy}}{\sqrt{P_x \cdot P_y}}$$

Na fórmula acima, S é a medida de similaridade bilateral do cosseno de Salton; P_{xy} é a frequência de associação entre duas palavras x e y; P_x é a frequência total da palavra x e P_y a frequência total da palavra y. A similaridade foi usada atualmente em mensurações feitas por Khasseh *et al.* (2017) e por Vincent-Lamarre e Larivière (2021). Os tratamentos estatísticos possibilitaram a visualização dos mapas de palavras-chave que representam o campo multidisciplinar do HIV/AIDS com o auxílio do *software Vosviewer*.

3 RESULTADOS DA ANÁLISE DESCRITIVA

Os 2895 artigos indexados pela WoS representam os 28 anos de pesquisa em HIV/AIDS cobertos pela análise de coocorrência de palavras e apresentam uma média de publicação anual de 103,39 artigos. O cálculo da mediana estabelece o intervalo entre os anos 2006 e 2007 com 115,5 artigos. A média da taxa de crescimento da pesquisa brasileira em HIV/AIDS ainda pode ser considerada alta (14%) se comparada à média nacional de crescimento da literatura para todas as áreas científicas, em 10,7% (DELLAGOSTIN, 2021). No entanto, apesar de registrar aspectos de crescimento exponencial, a produção científica em HIV/AIDS do Brasil não cresceu nos últimos anos, mantendo

crescimento zero ou crescimento negativo a partir de 2012. A maior concentração de artigos se deu a partir do ano de 2007, apresentando taxas de contribuição relativa acima dos 5% em praticamente todos os casos. A Tabela 1 indica a distribuição dos artigos ao longo dos anos, juntamente com dados do seu crescimento absoluto e relativo, provendo um panorama do repertório de documentos recuperados.

Tabela 1 – Crescimento da literatura brasileira em HIV/AIDS na WoS (1993-2020)

Anos de publicação	Frequência absoluta	Frequência relativa (%)	Frequência acumulada absoluta	Frequência acumulada relativa (%)	Taxa de Crescimento (%)
1993	13	0,45	13	0,45	
1994	29	1,00	42	1,45	123
1995	18	0,62	60	2,07	-38
1996	32	1,11	92	3,18	78
1997	40	1,38	132	4,56	25
1998	43	1,49	175	6,04	8
1999	42	1,45	217	7,50	-2
2000	51	1,76	268	9,26	21
2001	41	1,42	309	10,67	-20
2002	37	1,28	346	11,95	-10
2003	55	1,90	401	13,85	49
2004	60	2,07	461	15,92	9
2005	74	2,56	535	18,48	23
2006	95	3,28	630	21,76	28
2007	148	5,11	778	26,87	56
2008	156	5,39	934	32,26	5
2009	136	4,70	1070	36,96	-13
2010	178	6,15	1248	43,11	31
2011	191	6,60	1439	49,71	7
2012	191	6,60	1630	56,30	0
2013	168	5,80	1798	62,11	-12
2014	153	5,28	1951	67,39	-9
2015	160	5,53	2111	72,92	5
2016	159	5,49	2270	78,41	-1
2017	158	5,46	2428	83,87	-1
2018	173	5,98	2601	89,84	9
2019	144	4,97	2745	94,82	-17
2020	150	5,18	2895	100,00	4
Total de documentos: 2895					
Média de documentos por ano: 103,39 / Mediana: 115,5					
Média da taxa de crescimento: 14%					

Fonte: Dados da pesquisa.

A análise do crescimento da pesquisa brasileira em HIV/AIDS a partir de agrupamentos temporais evidenciou desaceleração com a projeção de taxas

negativas de crescimento e a diminuição das frequências. As estimativas de crescimento da literatura por agrupamentos podem ser observadas na Tabela 2:

Tabela 2 – Crescimento da literatura brasileira em HIV/AIDS por agrupamentos

Anos	Frequência absoluta	Média	Frequência relativa (%)	Frequência acumulada absoluta	Frequência acumulada relativa (%)	Taxa de Crescimento (%)
1993-1997	132	26,4	4,56	132	4,56	
1998-2001	177	44,2	6,11	309	10,67	25
2002-2005	226	56,5	7,81	535	18,48	28
2006-2009	535	133,7	18,48	1070	36,96	137
2010-2013	728	182	25,15	1798	62,11	36
2014-2017	630	157,5	21,76	2428	83,87	-13
2018-2020	467	155,6	16,13	2895	100,00	-26
Total de documentos: 2895						
Média de documentos por série temporal: 413,5 / Mediana: 467						
Média da taxa de crescimento: 14%						

Fonte: Dados da pesquisa.

Estes resultados mostram o padrão de crescimento/desaceleração da pesquisa. Já a compreensão desses resultados requereria outros estudos específicos que pudessem atender a complexidade da questão. Alguns indícios podem ser interpretados na literatura cientométrica sobre o tema, entretanto, na maioria dos casos, os dados de origem diferem quanto à sua presença nas bases de dados (GINGRAS, 2016).

4 RESULTADOS DA ANÁLISE DE COCORRÊNCIA DE PALAVRAS

As palavras utilizadas na análise de coocorrência foram as palavras-chave indicadas pelos próprios autores nos artigos recuperados. Esta adoção gera uma representação ajustada à percepção que os autores têm em relação ao campo semântico (LIBERATORE; HERRERO-SOLANA, 2013). A execução da análise de coocorrência de palavras demandou um esforço de adequação num processo de redução de palavras-chave. Tal operação resultou num conjunto de 2398 (dois mil trezentos e noventa e oito) artigos, sendo 80,4 % do resultado da busca. A Tabela 3 informa os dados do processo de redução.

Tabela 3 – Processo de redução de palavras-chave

	1993 1997	1998 2001	2002 2005	2006 2009	2010 2013	2014 2017	2018 2020
Número de documentos	132	177	226	535	728	630	467
Documentos sem palavras-chave	57	46	74	88	141	108	53
Proporção relativa	44%	26%	33%	16%	19%	17%	11%
Documentos com palavras-chave	75	131	152	447	589	521	411
Proporção relativa	56%	74%	67%	84%	81%	83%	89%
Unidades de palavras	346	655	807	2271	2871	2539	2126
Média de unidades por registro	4,61	5	5,3	5,08	4,87	4,86	5,13
Palavras-chave	233	429	509	1149	1360	1235	1150
Frequência mínima analisada	2	2	2	3	3	3	3
Palavras-chave analisadas	29	61	77	115	144	140	106
Pares associados	84	227	318	775	904	818	513

Fonte: Dados da pesquisa.

A análise de coocorrência de palavras, então, atribuiu um conjunto de palavras e de palavras associadas, considerando aquelas que obtiveram um escore mínimo de frequência nos respectivos agrupamentos temporais (ZHANG; HUA; YUAN, 2018). Tal esforço foi necessário pois alguns artigos não tinham registro de palavras-chave. Na tabela anterior, a proporção de artigos com palavras-chave de autor aumenta conforme observamos a evolução nos agrupamentos temporais.

A intensidade é o indicador de formação de *cluster* e calcula a força de associação entre as palavras. Uma palavra-chave que reúne muitas associações com outras palavras pode ser considerada um nó agregador formando o *cluster* devido à maior conectividade. Essa informação se dá a partir da frequência entre palavras associadas com os grafos formados no *Vosviewer*. O indicador também foi usado em estudos posteriores (KHALDI; PRADO-GASCÓ, 2021; LEUNG; SUN; BAI, 2017; ZUPIC; ČATER, 2015). A Tabela 4 indica as palavras-chave mais frequentes do conjunto da pesquisa e a Tabela 5 informa as principais frequências e forças de nó entre as palavras-chave dos períodos.

Tabela 4 – Indicativo de maiores frequências

Palavra-chave	Total	1993 1997	1998 2001	2002 2005	2006 2009	2010 2013	2014 2017	2018 2020
<i>HIV</i>	822	14	27	28	127	221	221	184
<i>AIDS</i>	644	33	52	50	125	167	112	105
<i>Acquired immunodeficiency syndrome</i>	408	13	25	20	109	114	78	49
<i>Brazil</i>	285	16	13	34	64	59	50	49
<i>HIV/AIDS</i>	182	x	5	10	30	59	51	27
<i>Antiretroviral therapy</i>	138	x	4	7	29	25	41	32
<i>HIV-1</i>	115	4	7	9	23	30	26	16
<i>Tuberculosis</i>	112	8	10	9	18	26	25	16
<i>HIV infection</i>	102	5	5	18	19	24	24	7
<i>Epidemiology</i>	100	5	7	5	16	30	16	21

Fonte: Dados da pesquisa.

Tabela 5 – Primeiros agrupamentos e principais frequências e forças de nós

Palavra-chave	Agrupamento							
	1993 1997		1998 2001		2002 2005		2006 2007	
	(F)	(C)	(F)	(C)	(F)	(C)	(F)	(C)
<i>Adherence</i>					4	5		
<i>Anti-HIV agents</i>							11	44
<i>Antiretroviral therapy</i>			4	7	7	7	29	19
<i>Autopsy</i>					4	7		
<i>Children</i>							9	6
<i>Co-infection</i>							9	10
<i>Diagnosis</i>			4	4			9	15
<i>Epidemiology</i>	5	11	7	13	5	10	16	41
<i>Haart</i>					5	11	20	19
<i>Health vulnerability</i>							8	27
<i>Hepatitis C</i>							9	20
<i>Highly active antiretroviral therapy</i>							22	15
<i>HIV infection</i>	5	10	5	3			8	7
<i>HIV infections</i>	3	5			10	27	11	41
<i>HIV-1</i>	4	6	7	6	9	12	23	20
<i>Opportunistic infections</i>			4	5				
<i>Polymerase chain reaction</i>					4	9		
<i>Prevalence</i>							8	14
<i>Prevention & control</i>							23	89
<i>Reverse transcriptase</i>					4	4		
<i>Risk factors</i>			4	12			9	27
<i>Sexually transmitted diseases</i>							9	17
<i>Socioeconomic factors</i>					5	18		
<i>Tuberculosis</i>	8	18	10	15	9	7	18	26
<i>Women</i>							10	21
Legenda: (F) = Frequência no agrupamento (C) = Força do nó no agrupamento (n) = Frequência para a palavra-chave	(3)		(4)		(4)		(8)	

Fonte: Dados da pesquisa.

drug uses”, palavras como “*Elisa*”, “*Risk factors*” ou “*Risk reductions*” dizem muito sobre o ambiente de pesquisa da época. No Brasil, nesses anos, já havia importantes programas de combate à epidemia, tanto estatais quanto organizados pela sociedade civil, que centravam os estudos nas questões de diagnóstico e assistência aos infectados enfermos (PARKER, 2000).

No agrupamento 1998-2001, a atenção da pesquisa volta-se para a identificação das principais enfermidades diagnosticadas e palavras-chave como “*Diagnosis*”, “*Risk factors*”, “*Opportunistic infections*”, “*Chaga’s disease*” e “*Tuberculosis*” aumentam no período. Interessante observar a palavra-chave “*Tuberculosis*” em particular. Trata da atenção dos cientistas à uma das principais doenças oportunistas no período. Em muitos pacientes, a tuberculose se aproveita da baixa imunidade do indivíduo, o debilita fisicamente e inicia um processo de abertura para novas enfermidades oportunistas. O número de casos relacionados à tuberculose aumentou devido à coinfeção com o vírus HIV em pacientes com AIDS (GALVÃO, 2000).

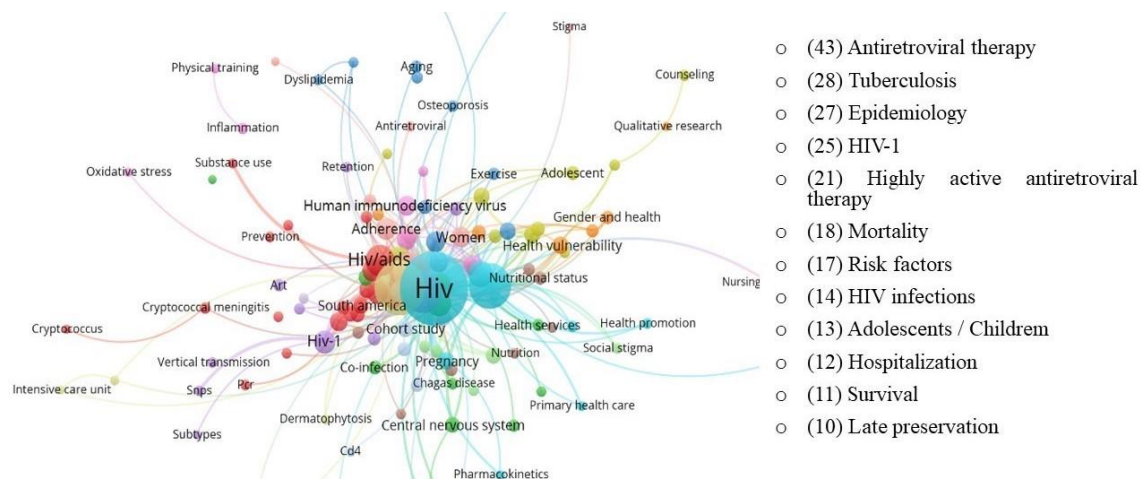
No terceiro agrupamento, 2002-2005, observa-se palavras-chave que mantêm coocorrências significativas como “*Antiretroviral therapy*”, “*Sexually transmitted diseases*”, “*HIV infections*”, “*Transmission*”, “*Socioeconomic factors*”, “*Candida spp*”, “*Human papillomavirus*”, “*Progressive multifocal leukoencephalopathy*”, “*Haart*” e “*HHV-8*”. O incremento de novas expressões pelos autores e o aumento da intensidade significa que existem mais rotas de uma palavra para outra na rede e novos *hubs* assumem, também, a proporcionalidade das conexões. Esse aspecto se observa, inclusive, nos agrupamentos posteriores. Nesse período se reafirma o percurso evolutivo da área de pesquisa em torno do HIV/AIDS e um direcionamento das intenções dos pesquisadores a temas como fatores socioeconômicos, aderência ao tratamento e profilaxia da doença, novas tipologias para o vírus HIV (genotipagem) e polimorfismos do vírus (CCR2 e CCR5).

No quarto período de agrupamentos, 2006-2009, a rede de palavras-chave se torna muito mais intensa. Uma hipótese pode ser entendida como a mudança de comportamento dos autores e dos editores de periódicos em relação ao uso de palavras-chave nos artigos, já que o período representa a

maior proporcionalidade de uso até então, na ordem de 83,6% do conjunto aglomerado do período. Esse dado determina um escore maior de frequência para inclusão das palavras-chave nas matrizes de coocorrência. As novas palavras surgidas são: “*Anti-HIV agents*”, “*Health vulnerability*”, “*Supply*”, “*Attitudes*”, “*Cross-sectional studies*”, “*Human rights*”, “*Mortality*”, “*Prenatal care*”, “*Questionnaires*”, “*Qualitative research*”, “*Health promotion*”, “*Oral manifestations*”, “*Pregnancy*”, “*Quality of life*” e “*Vertical transmission*”. Além disso, palavras-chave expressas em períodos anteriores reaparecem com maior força, devido ao agrupamento ser maior. Uma característica perceptível no aumento de palavras de agrupamento para agrupamento indica que as centralidades dos mapas têm diminuído conforme a passagem do tempo. O predomínio da frequência de palavras usadas na estratégia de busca não é mais um fator preponderante nas mensurações de coocorrência. Percebe-se, neste período, um direcionamento diversificado do conjunto de palavras-chave. A força dos nós aumenta, inclusive, em palavras fora da lista de alta frequência, como por exemplo, “*Incidence*”, “*Gender and health*”, “*HCV*”, “*Cryptococcus neofarms*”, “*Substance abuse intravenous*”, “*Sexual behavior*”, “*Sexual partnes*”, “*Health promotion*”, “*Health education*”, “*Sexuality*”, “*Prejudice*” e “*National health programs*”. As forças observadas identificam ainda mais a multidisciplinaridade característica da pesquisa sobre HIV/AIDS. Um exemplo desta dinâmica se observa na palavra-chave “*Prevention & control*” que possui alta intensidade na definição de *cluster* (89) apesar da frequência ser relativamente baixa (23), se comparada às palavras HIV (127) e AIDS (125) que ocupam o topo da lista de frequência neste agrupamento.

A seguir, apresenta-se a visualização do indicador de intensidade nos três agrupamentos temporais seguintes, 2010-2013; 2014-2017; 2018-2020. Preferiu-se a visualização separada dos dados destes agrupamentos, pois a dinâmica a partir do primeiro muda em relação ao crescimento, desde a primeira fase deste estudo. Apesar do crescimento constante no número de publicações dos primeiros agrupamentos, observa-se queda nestes últimos períodos. Esta queda compromete a natureza exponencial do crescimento desta literatura. No período em que a literatura brasileira em HIV/AIDS diminui em sua quantidade,

Figura 4 – Intensidade no período 2014-2017 e força dos nós



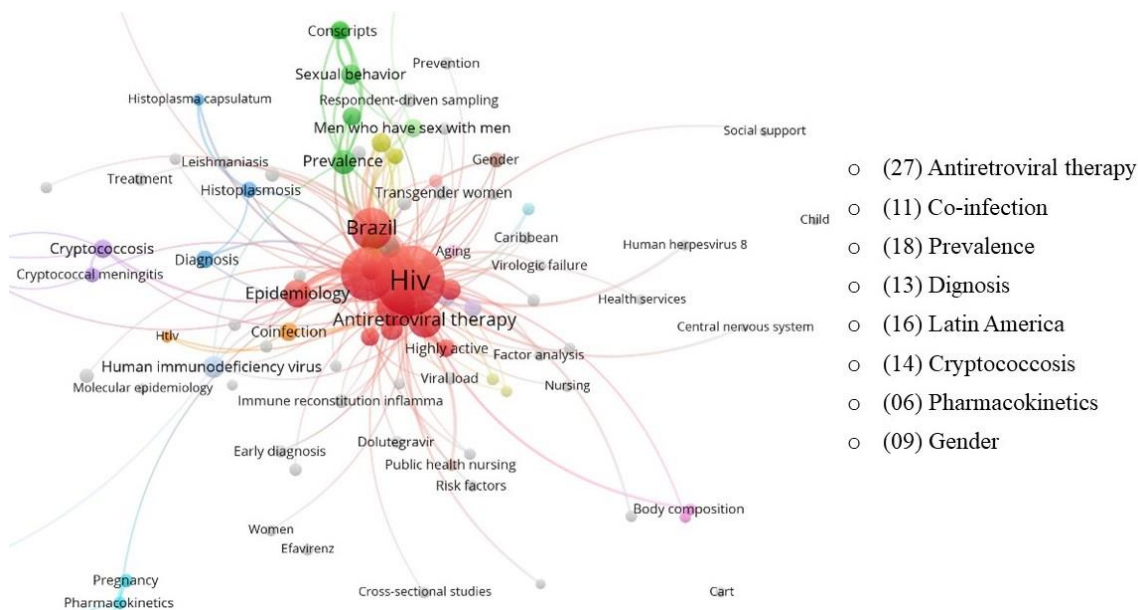
Fonte: Dados da pesquisa e Vosviewer.

Este sentido se confirma ao relatar o uso contínuo por parte dos pesquisadores de palavras-chave como “*Antiretroviral therapy*”, “*Highly active antiretroviral therapy*”, “*Highly active*” e “*Haart*”. A humanização da epidemia também evolui a partir do emprego de termos como “*Hospitalization*”, “*Survival*”, “*Late preservation*”, “*Treatment*”, “*Depression*”, “*Health vulnerability*”, “*Substance use*”, “*Screening*” e “*Obesity*”. Outra dimensão do agrupamento está na relação de palavras-chave identificadas com modelos de diagnóstico e métodos de tratamento: “*Co-infection*”, “*Lypodystrophy*”, “*SNPS*”, “*Special needs plans*”, “*Medication adherence*”, “*Spatial analysis*”, “*Co-hort studies*”, “*Antropometry*”.

A seguir, comenta-se o último agrupamento deste estudo, o período 2014-2020. Apesar de menor que os dois agrupamentos anteriores, este período também consolida alguns aspectos históricos da epidemia. No entanto, apesar de diversificado, torna-se preocupante a diminuição de trabalhos relacionados na pesquisa, já que a epidemia de HIV/AIDS não possui indicativos de diminuição, principalmente no Brasil. As palavras-chave mais significativas deste período são: “*Prevalence*”, “*Gender*”, “*Transgender women*”, “*MSM*”, “*Man who have sex with man*”, “*Sexual behavior*”, “*Aging*”. Novos desenvolvimentos científicos relacionados a medicamentos e tratamentos também possuem evidências importantes, como por exemplo, “*Prep*”, “*Pre-exposure prophylaxis*”, “*Adherence*”, “*Pharmacokinetics*”, além do contínuo estudo relacionado às comorbidades, como “*Co-infection*”, “*Histoplasmosis*”, “*Cryptococcosis*” e

“*Syphilis*”. Estas informações podem ser conferidas no grafo da Figura 5.

Figura 5 – Intensidade no período 2018-2020 e força dos nós



Fonte: Dados da pesquisa e *Vosviewer*.

As palavras-chave conectadas entre si em cada agrupamento temporal originam uma lista de frequência de pares de palavras associadas e servem de insumo para o cálculo do cosseno de Salton, usado para a definição de similaridade. Os pares de palavras associadas são dispostos em uma matriz de coocorrência MDS (*Maximum Distance Separable*) no *Bibexcel*. A organização dos dados permite a criação de uma matriz simétrica (.ma2) e uma matriz assimétrica (.ma5), que relacionam os pares de palavras em uma matriz; e a palavra-chave com o documento a que estão relacionadas na outra matriz. Assim, calcula-se a proporcionalidade a partir do índice de similaridade do par de palavras-chave. Com esta função, foi possível verificar a evolução da coocorrência de palavras no decorrer dos agrupamentos temporais, considerando a maior frequência entre os pares de palavras associadas. A coocorrência de palavras é uma das abordagens para construir uma medida de similaridade usando o conteúdo real dos documentos (ZUPIC; ČATER, 2015). Esta medida normalizada proporciona a constatação da similaridade mesmo entre agrupamentos com quantidades diferentes de artigos e escores de frequência de ocorrência e de coocorrência diferentes. O procedimento resultou em 144 pares de palavras significativos e um espectro de 96 palavras-chave. Na

sequência, se observa um estrato resumido dos pares de palavras associadas ao longo do tempo (28 coocorrências e um total de 24 palavras-chaves). Devido às limitações de espaço para a tabela, pelo menos uma das palavras do par não deve ser originária da expressão de busca. Também, a ocorrência do par deve estar descrita em mais de um agrupamento temporal, conforme a Tabela 6.

Tabela 6 – Similaridade dos pares de palavras-chave ao longo dos agrupamentos temporais

Palavras associadas		Agrupamentos temporais						
Palavra x	Palavra y	1993 1997	1998 2001	2002 2005	2006 2009	2010 2013	2014 2017	2018 2020
<i>Acquired immunodeficiency syndrome</i>	<i>Adolescents</i>				0,174	0,178		
<i>Acquired immunodeficiency syndrome</i>	<i>Epidemiology</i>	0,372	0,151	0,300	0,161	0,130	0,125	0,094
<i>Acquired immunodeficiency syndrome</i>	<i>Highly active antiretroviral therapy</i>				0,157	0,169	0,107	0,175
<i>Acquired immunodeficiency syndrome</i>	<i>HIV infections</i>		0,283	0,335	0,148	0,191	0,126	0,128
<i>Acquired immunodeficiency syndrome</i>	<i>Prevention & control</i>				0,415	0,156		
<i>Acquired immunodeficiency syndrome</i>	<i>Socioeconomic factors</i>			0,300	0,206			
<i>Acquired immunodeficiency syndrome</i>	<i>Tuberculosis</i>	0,294				0,122	0,086	0,107
<i>Acquired immunodeficiency syndrome</i>	<i>Women</i>				0,233	0,121	0,102	
<i>AIDS</i>	<i>Antiretroviral therapy</i>		0,208	0,160	0,166	0,092	0,207	0,104
<i>AIDS</i>	<i>Epidemiology</i>		0,105			0,147	0,071	0,149
<i>AIDS</i>	<i>Haart</i>			0,190	0,240	0,153	0,183	
<i>AIDS</i>	<i>HIV infection</i>	0,234			0,190		0,114	0,074
<i>AIDS</i>	<i>HIV-1</i>	0,174			0,093	0,098	0,093	0,098
<i>AIDS</i>	<i>Mortality</i>		0,240			0,106	0,079	0,113
<i>AIDS</i>	<i>Opportunistic infections</i>	0,246	0,139					
<i>AIDS</i>	<i>Prevention</i>		0,196	0,283			0,047	
<i>AIDS</i>	<i>Tuberculosis</i>	0,123	0,175	0,167	0,211		0,089	0,049
<i>Antiretroviral therapy</i>	<i>Highly active</i>				0,279	0,400	0,413	0,433
<i>Antiretroviral therapy</i>	<i>HIV</i>			0,286	0,044	0,135	0,250	0,207
<i>Brazil</i>	<i>HIV-1</i>	0,250			0,13	0,261	0,028	0,107
<i>Co-infection</i>	<i>HIV</i>		0,136		0,059	0,122	0,138	0,171
<i>Epidemiology</i>	<i>Tuberculosis</i>	0,316	0,239				0,047	0,055
<i>Haart</i>	<i>HIV</i>				0,198	0,135	0,093	
<i>Highly active antiretroviral therapy</i>	<i>HIV</i>				0,151	0,104	0,063	0,149
<i>HIV</i>	<i>Latin america</i>				0,044		0,148	0,162
<i>HIV</i>	<i>Quality of life</i>					0,082	0,179	0,208
<i>HIV</i>	<i>Tuberculosis</i>	0,204	0,061		0,063	0,112	0,152	0,146
<i>Human immunodeficiency virus</i>	<i>Tuberculosis</i>	0,289	0,218				0,052	

Fonte: Dados da pesquisa e *Bibexcel*.

Na seção seguinte, comenta-se os principais resultados obtidos nas análises.

5 DISCUSSÃO DOS RESULTADOS

Do ponto de vista metodológico, a análise de coocorrência de palavras demonstrou satisfatoriamente que as palavras-chave atribuídas por autores contribuem para a verificação da estrutura temática de uma área de pesquisa. A agregação dos agrupamentos pela análise de *cluster* segundo critérios de classificação por proximidade permitem uma declaração acerca das prioridades temáticas nas pesquisas. E as relações resultantes das matrizes puderam ilustrar as palavras-chave em mapas de rede. As comparações entre os mapas de diferentes agrupamentos foram utilizadas para demonstrar o aspecto dinâmico do desenvolvimento científico no campo do HIV/AIDS.

Outro aspecto a ser observado nesta pesquisa é o relativo à evolução dos temas de pesquisa. No início do levantamento (1993-1997), os artigos versavam sobre temas como exames diagnósticos (“*Elisa*”), doenças oportunistas (“*Syphilis*”, “*Hepatitis B*”, “*Tuberculosis*”), formas diversas de infecção e grupos de risco (“*Injecting drug users*”), ou seja, a ciência brasileira se instrumentalizava sobre a nova epidemia. No segundo período (1998-2001), as palavras-chave já indicam de forma mais formalizada, conteúdos acerca da prevenção (“*Prevention*”), descobertas de subtipos do vírus HIV (“*Viral load*”). Também surgem termos que demonstram o aumento dos estudos sobre HIV em mulheres e crianças (“*Women*”, “*Gender*”, “*Breastfeeding*”) e a continuidade das pesquisas com doenças oportunistas e os primeiros estudos genéticos.

No período subsequente (2002-2005), a matriz semântica dos autores em HIV/AIDS indica a diminuição da mortalidade, uma atenção maior ao tratamento e novas profilaxias da doença (“*Antiretroviral therapy*”), ao conhecimento socioeconômico das populações (“*Age*”, “*Color*”), o avanço da pesquisa laboratorial rumo à cura (“*Genotyping*”, “*Drug resistance*”, “*Autopsy*”), e um reforço nos estudos sobre a genética do HIV. O quarto período de análise (2006-2009) é marcado por uma mudança de paradigma em que o olhar da doença passa do diagnóstico e prevenção aos cuidados com o indivíduo vivendo com

HIV/AIDS. Também, nesses anos, observa-se a mudança do viés fator de risco e grupo de risco para analisar os fatores do comportamento de risco, dado o aumento das incidências em novas populações (“*Women*”, “*Child*”, “*Adolescents*”, “*Vertical transmission*”, “*Sexual behavior*”). Em 2007, o governo federal avança na política de quebra de patentes relacionadas a medicamentos para as pessoas que vivem com HIV/AIDS, tornando o Brasil um dos poucos países que possui um programa de acesso universal a estes medicamentos essenciais ao combate da epidemia. Concomitante a este acontecimento, torna-se mais presente a palavra controle, atribuída em conjunto com prevenção (“*Prevention & control*”) e novas formas de tratamento (“*Haart*”; “*Highly active antiretroviral therapy*”).

No período 2010-2013, vê-se uma preocupação cada vez maior com cuidados e assistência em saúde ao indivíduo vivendo com HIV/AIDS (“*Quality of life*”, “*Nursing*”, “*Nursing care*”), além do aumento em todas as iniciativas de temas envolvendo o HIV e a AIDS. A lista de baixa frequência deste período, a mais numerosa e diversificada dentre todos os agrupamentos temporais indica que a ciência da AIDS continua evoluindo tematicamente. Nos dois últimos períodos, nota-se a continuidade da ascensão de novos termos e uma maior especificidade nas pesquisas, mesmo quando os agrupamentos indicam a diminuição no número de artigos de forma sistemática. No Brasil atual, a epidemia de HIV/AIDS possui características como a heterossexualização, a feminização, o envelhecimento, a interiorização e a pauperização. Ou seja, mulheres idosas heterossexuais de baixa renda e que vivem em cidades do interior têm se tornado as principais vítimas de infecção pelo HIV nos últimos tempos (JESUS, 2018).

Nos aspectos acerca do crescimento absoluto, se percebe que há indícios de crescimento exponencial intercalados por períodos de crescimento linear e crescimento negativo. Por exemplo: do segundo para o terceiro período de análise, de 1998-2001 a 2002-2005 há uma estabilização do crescimento. Na passagem para o quarto período, 2006-2009, ocorre um aumento exponencial que logo se estabiliza para crescer exponencialmente de novo no quinto período, 2010-2013. Nesse, a produção científica em questão sofre uma leve queda de

produtividade e mantém esta regularidade de crescimento negativo nos dois últimos períodos. No entanto, estes dados não afetam o crescimento da diversidade temática da pesquisa se observado o período integral do estudo.

6 CONCLUSÕES

A partir do diagnóstico feito com a análise de coocorrência de palavras, considera-se que o método pode ser útil nas mensurações de interesse científico que visam o mapeamento de áreas do conhecimento. O conceito, enquanto modelo de verificação da produção científica, pode contribuir com a identificação de temas de pesquisa, além de evidenciar a forma como evoluem. Assim, o estudo atinge seus objetivos, demonstrando as redes de coocorrência entre as palavras-chave. Os principais resultados elucidaram a estrutura conceitual do HIV/AIDS no cenário de pesquisa brasileira indexada na WoS, identificando as palavras-chave do núcleo básico do campo a partir de uma frequência mínima em cada agrupamento e a adoção de critérios de intensidade e de similaridade. A contribuição genuína desta pesquisa está na visualização da evolução temporal do léxico do HIV/AIDS na visão categorizada dos autores, fornecendo procedimentos acerca do método para o campo da Ciência da Informação e contribuindo com dados para análises no campo da Epidemiologia no Brasil.

O estudo também propicia novas perguntas, pois se avalia a opção em estudar as razões da nova configuração de crescimento. Também se avalia o interesse em descobrir temas emergentes a partir das listas de baixa frequência e, por último, a configuração deste estudo com os títulos e resumos das publicações. Podem ser realizadas comparações entre as estruturas semânticas de cada tipo de metadado para ampliar ainda mais os conhecimentos acerca da ciência do HIV/AIDS no Brasil. Conforme o entendimento de diversos autores (CORRALES-GARAY; ORTIZ-DE-URBINA-CRIADO; MORA-VALENTÍN, 2019; GIANNAKOS *et al.*, 2020; SOLER-COSTA *et al.*, 2021) após os metadados serem estruturados para as análises métricas de informação, uma série de averiguações poderão ser executadas para ampliar os conhecimentos da atividade científica de um campo determinado, como o da pesquisa brasileira em HIV/AIDS.

REFERÊNCIAS

BASTOS, C. **Ciência, poder, ação**: as respostas à Sida. Lisboa: Imprensa de Ciências Sociais, 2002.

BIBEXCEL. [S. l.: s. n.], 2021. Disponível em:
<https://homepage.univie.ac.at/juan.gorraiz/bibexcel>. Acesso em: 23 ago. 2021.

BRASIL. Ministério da Saúde. **Boletim Epidemiológico**: HIV/AIDS. Brasília, DF, n. especial, dez. 2020. Disponível em:
<http://www.aids.gov.br/ptbr/pub/2020/boletim-epidemiologico-hivaid-2020>. Acesso em: 23 ago. 2021.

CALLON, M.; COURTIAL, J.P.; PENAN, H. **Cienciometria**: el estudio cuantitativo de la actividad científica: de la bibliometria a la vigilancia tecnológica. Espanha: Trea, 1995.

CORRALES-GARAY, D.; ORTIZ-DE-URBINA-CRIADO, M.; MORA-VALENTÍN, E. M. Knowledge areas, themes and future research on open data: a co-word analysis. **Government Information Quarterly**, [S. l.], v. 36, n. 1, p. 77-87, 2019. Disponível em:
<https://www.sciencedirect.com/science/article/abs/pii/S0740624X18303216?via%3Dihub>. Acesso em: 22 maio 2021.

CURTY, R. G.; DELBIANCO, N. R. As diferentes metrias dos estudos métricos de informação: evolução epistemológica, inter-relações e representações, **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 25, jan./dez. 2020. Disponível em:
<https://periodicos.ufsc.br/index.php/eb/article/view/74593>. Acesso em: 12 nov. 2022.

DE BELLIS, N. **Bibliometrics and citation analysis**: from the Science Citation Index to cybermetrics. USA: Scarecrow Press, 2009.

DELLAGOSTIN, O. A análise do fomento à pesquisa no país e a contribuição das agências federais e estaduais. **Revista Inovação e Desenvolvimento**, Recife, v. 2, n. 6, p. 6-12, ago. 2021. Disponível em:
<https://revistainovacao.facepe.br/index.php/revistaFacepe/article/view/62>. Acesso em: 12 set. 2021.

EXCEL. [S. l.: s. n.], 2021. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/excel?rtc=1>. Acesso em: 23 ago. 2021.

GALVÃO, J. **AIDS no Brasil**: a agenda de construção de uma epidemia. Rio de Janeiro: Editora 34, 2000.

GANHOR, J. P.; SILVA, D. R.; MEGLHIORATTI, F. A.; MALACARNE, V. Mapping of mobilizations of bibliometrics and scientometrics in science education research: a look at national journals. **Research, Society and**

Development, [S. l.], v. 9, n. 7, p. e805974849, 2020. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/4849>. Acesso em: 23 jun. 2021.

GIANNAKOS, M.; PAPAMITSIOU., Z.; MARKOPOULOS, P.; LEIA, J.; HOURCADE, J. P. Mapping child-computer interaction through co-word analysis. **International Journal of Child-Computer Interaction**, [S. l.], v. 23-24, p. 100-165, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2212868920300015>. Acesso em: 10 jun. 2021.

GINGRAS, Y. **Bibliometrics and research evaluation uses and abuses**. Cambridge: MIT Press Direct, 2016. *E-book*. Disponível em: <https://direct.mit.edu/books/book/4081/Bibliometrics-and-Research-EvaluationUses-and>. Acesso em: 12 jul. 2021.

JESUS, A. **Investigação de informação e comunicação: o boletim epidemiológico como dispositivo interdisciplinar de saúde**. 2018. 111 f. Dissertação (Mestrado em Informação e Comunicação em Saúde) – Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2018.

KHALDI, H.; PRADO-GASCÓ, V. Bibliometric maps and co-word analysis of the literature on international cooperation and migration. **Quality and Quantity**, [S. l.], v. 55, p. 1845-1869, 2021. Disponível em: <https://link.springer.com/article/10.1007/s11135-020-01085-4>. Acesso em: 22 jun. 2021.

KHASSEH, A.A.; SOHEILI, F.; MOGHADDAM, H. S.; CHELAK, A. M. Intellectual structure of knowledge in iMetrics. **Information Processing and Management**, [S. l.], v. 53, n. 3, p. 705-720, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0306457316303338?via%3Dihub>. Acesso em: 12 ago. 2021.

KUHN, T. **A estrutura das revoluções científicas**. 13. ed. São Paulo: Perspectiva, 2017.

LABORATÓRIO DE ESTUDIOS MÉTRICOS DE LA INFORMACIÓN (LEMI). **Análisis avanzados para la evaluación científica a partir de herramientas informétricas**. Madrid: UFRGS, 2013. Curso ministrado na Universidade Federal do Rio Grande do Sul, p. 19-21, nov. 2013.

LEUNG, X.Y.; SUN, J.; BAI, B. Bibliometrics of social media research: a cocitacion and co-word analysis. **International Journal of Hospitality Management**, [S. l.], v. 66, p. 35-45, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0278431917302281>. Acesso em: 25 jun. 2021.

LIBERATORE, G.; HERRERO-SOLANA, V. Caracterización temática de la investigación en ciencia de la información en Brasil en el período 2000-2009.

Transinformação, Campinas, v. 25, n. 3, p. 225-235, set./dez. 2013.

Disponível em:

<https://www.scielo.br/j/tinf/a/v5ywcpkjdzTQ7sNfXpVB6jd/abstract/?lang=es>.

Acesso em: 12 ago. 2021.

MACIAS-CHAPULA, C.A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998. Disponível em:

<https://revista.ibict.br/ciinf/article/view/794>. Acesso em: 12 nov. 2022.

MIRO. [S. l.: s. n.], 2021. Disponível em: <https://miro.com>. Acesso em: 12 out. 2021.

MORKUNAS, V. J.; MOORE, S.; DUNCAN, S. Y. A bibliometric analysis of the Journal of Public Affairs. **Journal of Public Affairs**, [S. l.], v. 19, n. 2, p. e1833, June 2018. Disponível em:

https://www.researchgate.net/publication/325762552_A_bibliometric_analysis_of_the_Journal_of_Public_Affairs. Acesso em: 25 jun. 2021.

NÁJERA-SÁNCHEZ, J. J. A systematic review of sustainable banking through a co-word analysis. **Sustainability**, [S. l.], v. 12, n. 1, 2020. Disponível em: <https://www.mdpi.com/2071-1050/12/1/278>. Acesso em: 12 jul. 2021.

OLIVEIRA, L. F. T.; GRÁCIO, M. C. C. Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tema estudos métricos na base Scopus. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 4, p. 16-28, out./dez. 2011. Disponível em:

<https://periodicos.ufmg.br/index.php/pci/article/view/22742>. Acesso em: 12 nov. 2021.

PARKER, R. G. **Na contramão da AIDS: sexualidade, intervenção, política**. São Paulo: Editora 34, 2000.

PERSSON, O.; DANELL, R.; SCHNEIDER, J.W. How to use Bibexcel for various types of bibliometric analysis. **ISSI Newsletter**, [S. l.], v. 5, n. esp., p. 9-24, jun. 2009.

RUIZ-BAÑOS, R. BAILON-MORENO, R.; JIMENEZ-CONTRERAS, E.; COURTIAL, J. Structure and dynamics of scientific networks. Part II: the new Zipf's law, the clusters of co-citations and the model of the descriptor presence. **Scientometrics**, Budapest, v. 44, n. 2, p. 235-265, fev. 1999b. Disponível em: http://eprints.rclis.org/13849/1/Ruiz_Bagnos%2C_R_Structure_and_Dynamics_of_Scientific_networks_Part_II_The_new_Zipf%27s_law%2C_the_clusters_of_co-citations_and_the_model_of_the_descriptor_presence%2C_Scientometrics_1999%2C_44%282%29_235-265.pdf Acesso em: 12 jul. 2021.

RUIZ-BAÑOS, R.; BAILON-MORENO, R.; JIMENEZ-CONTRERAS, E.; COURTIAL, J. Structure and dynamics of scientific networks. Part I:

fundamentals of the quantitative model of translation. **Scientometrics**, Budapest, v. 44, n. 2, p. 217-234, fev. 1999a. Disponível em: [http://eprints.rclis.org/12997/1/Ruiz_Bagnos,_R_Structure_and_Dynamics_of_Scientific_networks_Part_I_Fundamentals_of_the_quantitative_model_of_translation_Scientometrics_1999_44\(2\)_217-234.pdf](http://eprints.rclis.org/12997/1/Ruiz_Bagnos,_R_Structure_and_Dynamics_of_Scientific_networks_Part_I_Fundamentals_of_the_quantitative_model_of_translation_Scientometrics_1999_44(2)_217-234.pdf). Acesso em: 12 jul. 2021.

SANTOS, R. A. **Análise de coocorrência de palavras na pesquisa brasileira em HIV/Aids indexada na Web of Science no período 1993-2013**. 2015. 167 f. Dissertação (Mestrado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/122180/000971529.pdf?sequence=1&isAllowed=y>. Acesso em: 15 jun. 2021.

SOLER-COSTA, R.; MORENO-GUERRERO, A.-J.; LÓPEZ-BELMONTE, J.; MARÍN-MARÍN, J.-A. Co-word analysis and academic performance of the term TPACK in Web of Science. **Sustainability**, [S. l.], v. 13, n. 3, p. 1481, 2021. Disponível em: <https://www.mdpi.com/2071-1050/13/3/1481>. Acesso em: 15 jun. 2021.

VINCENT-LAMARRE, P.; LARIVIÈRE, V. Textual analysis of artificial intelligence manuscripts reveals features associated with peer review outcome. **Quantitative Science Studies**, [S. l.], v. 2, n. 2, p. 662-677, 2021. Disponível em: <https://direct.mit.edu/qss/article/2/2/662/97556/Textual-analysis-of-artificial-intelligence>. Acesso em: 24 jun. 2021.

VOSVIEWER. [S. l.: s. n.], 2021. Disponível em: <https://www.vosviewer.com>. Acesso em: 23 ago. 2021.

WEB of Science. [S. l.: s. n.], 2021. Disponível em: <https://webofknowledge.com>. Acesso em: 17 jun. 2021.

ZHANG, Y.; HUA, W.; YUAN, S. Mapping the scientific research on open data: a bibliometric review. **Learned Publishing**, [S. l.], v. 31, n. 2, p. 95-106, 2018. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1110>. Acesso em: 12 maio 2021.

ZUPIC, I.; ČATER, T. Bibliometric methods in management and organization. **Organization Research Methods**, [S. l.], v.18, n.3, p. 429-472, 2015. Disponível em: https://www.researchgate.net/publication/254970530_Bibliometric_Methods_in_Management_and_Organization. Acesso em: 25 jun. 2021.

CO-WORD ANALYSIS IN BRAZILIAN HIV/AIDS RESEARCH INDEXED ON THE WEB OF SCIENCE IN THE 1993-2020 PERIOD

ABSTRACT

Objective: The research presents an exploratory-descriptive study on the Brazilian scientific production on HIV/AIDS indexed in the Web of Science in the period 1993-2020. The study analyzes, based on bibliometric indicators of co-word analysis, a perception aligned with the authorial expression of Brazilian research on HIV/AIDS, based on information obtained in the Descriptor (DE) metadata field of the 2895 retrieved articles. **Methodology:** The research sought to expand the knowledge of the dimensions related to the themes of research on HIV/AIDS, but it was also concerned with demonstrating the potential of the procedures of the co-word analysis. The results are still supported by indicators of scientific activity about this production over the years. Frequency measurements; of intensity, through the identification of clusters; and similarity, based on Salton's Cosine in seven temporal groups: 1993-1997, 1998- 2001, 2002-2005, 2006-2009, 2010-2013, 2014-2017 and 2018-2020 were obtained. **Results:** Despite maintaining exponential growth until 2011, Brazilian research on HIV/AIDS shows signs of negative growth in the last groups. The co-word analysis identified the keywords of the basic core of the field of HIV/AIDS in each period from the definition of a minimum occurrence in the groups. Keyword mainstream is defined by the intensity and similarity of associated word pairs over time. **Conclusions:** The study concludes that the keywords used by authors can demonstrate aspects of the evolution of a research area, as observed in Brazilian research on HIV/AIDS from the 1993-2020 period.

Descriptors: HIV/AIDS. Brazil. Scientific production. Bibliometric. Co-word analysis.

ANÁLISIS DE CO-PALABRA EN LA INVESTIGACIÓN BRASILEÑA SOBRE VIH/SIDA INDEXADA EN LA WEB OF SCIENCE EN EL PERÍODO 1993-2020

RESUMEN

Objetivo: La investigación presenta un estudio exploratorio-descriptivo sobre la producción científica brasileña sobre VIH/SIDA indexada en la Web of Science en el período 1993-2020. El estudio analizó, con base en indicadores bibliométricos de co-ocurrencia de palabras clave, una percepción alineada con la expresión autoral de la investigación brasileña sobre VIH/SIDA, a partir de la información obtenida en el campo de metadatos Descriptor (DE) de los 2895 artículos recuperados. **Metodología:** La investigación buscó ampliar el conocimiento de las dimensiones relacionadas con los temas de investigación en VIH/SIDA, pero también se preocupó por demostrar el potencial de los procedimientos para analizar la co-ocurrencia de palabras. Los resultados aún están respaldados por indicadores de actividad científica sobre esta producción a lo largo de los años. Medidas de frecuencia obtenidas; de intensidad, a través de la identificación de clusters; y semejanza, con base en el coseno de Salton en siete grupos temporales: 1993-1997, 1998-2001, 2002-2005, 2006-2009, 2010-2013, 2014-2017 y 2018-2020. **Resultados:** A pesar de mantener un crecimiento exponencial hasta 2011, la investigación brasileña sobre VIH/SIDA muestra signos de crecimiento negativo en los últimos grupos. El análisis de co-ocurrencia de palabras identificó las palabras clave del núcleo básico del campo de VIH/SIDA en cada período a partir de la definición de ocurrencia mínima en los grupos. La corriente principal de palabras clave se define por la intensidad y similitud de los pares de palabras asociados a lo largo del tiempo. **Conclusiones:** El estudio concluye que las palabras clave utilizadas por los autores pueden demostrar aspectos de la evolución de un área de investigación, como se observa en la investigación brasileña sobre VIH/SIDA en el período 1993-2020.

Descritores: VIH/SIDA. Brasil. Producción científica. Bibliometría. Análisis de co-palabras.

Recebido em: 05.01.2022

Aceito em: 21.11.2022