

SÉRGIO SALUSTIANO DA SILVA

Crossover Métrico: A integração de indicadores altmétricos para a análise da produção científica

Dissertação de Mestrado
Março de 2019



**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ
ESCOLA DE COMUNICAÇÃO – ECO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E
TECNOLOGIA – IBICT
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO –
PPGCI**

SÉRGIO SALUSTIANO DA SILVA

Crossover Métrico: a integração de indicadores
altmétricos para a análise da produção científica

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Federal do Rio de Janeiro e Instituto Brasileiro de Informação em Ciência e Tecnologia para a obtenção do título de Mestre em Ciência da Informação.

Orientador: Prof. Dr. Fábio Castro Gouveia

**RIO DE JANEIRO
2019**

Silva, Sergio Salustiano da.

S586c Crossover métrico: A integração de indicadores altmétricos para a análise da produção científica /Sérgio Salustiano da Silva. - Rio de Janeiro, 2019. 161 f.

Orientador: Fábio Castro Gouveia. Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Escola da Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, 2019.

1. Crossover. 2. Cibermetria. 3. Métricas Alternativas. 4. Altméria. 5. Redes Sociais. I. Gouveia, Fábio Castro, orient. II. Título.

**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ
ESCOLA DE COMUNICAÇÃO – ECO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E
TECNOLOGIA – IBICT
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO –
PPGCI**

SÉRGIO SALUSTIANO DA SILVA

Crossover Métrico: a integração de indicadores alométricos para a análise da produção científica

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Federal do Rio de Janeiro e Instituto Brasileiro de Informação em Ciência e Tecnologia para a obtenção do título de Mestre em Ciência da Informação.

Aprovado em 29 de março de 2019.

Prof^o. Dr. Fábio Castro Gouveia, – Orientador
PPGCI –IBICT/UFRJ -ECO / FIOCRUZ

Prof^a. Dr^a. – Eloísa da Conceição Príncipe de Oliveira
IBICT

Prof^a. Dr^a. - Kizi Mendonça de Araújo
FIOCRUZ

AGRADECIMENTOS

Ao Doutor e Professor Fábio Gouveia (orientador = amigo e incentivador) que acreditou no meu projeto, me deu liberdade para realizar as experimentações que eu acreditava ser necessárias e soube o momento de puxar o freio e me colocar de volta nos trilhos.

Aos professores pelas produtivas conversas de corredor, ao corpo técnico que sempre esteve disponível para atender minhas solicitações e aos amigos que fiz durante estes anos.

Aos departamentos de Informática da PUC-Rio e ao IESP-UERJ que me acolheram como aluno externo.

*And those who were seen dancing were thought to be
Insane by those who could not hear the music.
(Autor desconhecido)*

SILVA, Sergio Salustiano da. **Crossover métrico**: a integração de indicadores alométricos para a análise da produção científica. 2019. 148 f. Dissertação (Mestrado) - Curso de Pós-graduação em Ciência da Informação, Programa de Pós-graduação em Ciência da Informação, Universidade Federal do Rio de Janeiro - Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2019.

RESUMO: Os ambientes digitais oferecem uma infinidade de possibilidades para os atores envolvidos com o desenvolvimento e financiamentos de pesquisas. Entre as novas oportunidades está o acompanhamento quase em tempo real da repercussão de um conteúdo científico por meio de métricas e indicadores que sinalizam uma possível relevância e/ou interesse do assunto abordado para diversos tipos de avaliações. Dentro desse escopo surgiu a Almetria como um novo subcampo dos estudos métricos da informação cuja finalidade é medir a disseminação de documentos científicos na web e nas redes sociais. Os estudos de sua viabilidade e aplicação ainda estão em fase de desenvolvimento, e na maioria dos casos são comparados com vários outros indicadores que surgiram ao longo dos anos. Uma das principais características destas métricas alternativas é de oferecer informações complementares as métricas consideradas tradicionais. Porém, mesmo com a crescente popularização desse subcampo ainda existe a necessidade do aprofundamento teórico e metodológico. Por esse motivo, para a pesquisa foram utilizados os aportes conceituais e metodológicos da Cibermetria por ser um subcampo consolidado e que carrega em seu escopo todos os elementos que também compõem a Almetria. Assim, a pesquisa buscou utilizar esse conhecimento de informações que podem ser extraídas da web e das redes sociais, além de levar em consideração as particularidades e peculiaridades de cada ambiente para propor a utilização de técnicas de crossover para desenvolver um modelo que possa auxiliar nos estudos métricos da informação. Os resultados preliminares apontam para a viabilidade da utilização do crossover, ao mesmo tempo, possibilitaram o desenvolvimento três Índices (Exposição, Interação e Apropriação), que são utilizados como forma de contribuir com os estudos que buscam direcionar formas de olhar a quantidade de dados disponibilizados pelos e nos ambientes digitais.

Palavras-chave: Crossover. Métricas Alternativas. Almetria. Cibermetria. Redes Sociais.

SILVA, Sergio Salustiano da. **Crossover métrico**: a integração de indicadores altmétricos para a análise da produção científica. 2019. 148 f. Dissertação (Mestrado) - Curso de Pós-graduação em Ciência da Informação, Programa de Pós-graduação em Ciência da Informação, Universidade Federal do Rio de Janeiro - Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2019.

ABSTRACT

Digital environments offer a multitude of possibilities for the actors involved in research development and funding. Among the new opportunities is the almost real-time monitoring of the repercussion of a scientific content through metrics and indicators that sign a possible relevance and/or interest of the subject addressed to various types of assessments. Within this scope altmetrics emerged as a new subfield of information metrics studies whose purpose is to measure the dissemination of scientific documents on the web and social networks. Studies of its viability and application are still in development phase, and in most cases are compared with several other indicators that have emerged over the years. One of the main features of these alternative metrics is to offer complementary information to the considered traditional metrics. However, even with the increasing popularization of this subfield there is still the need for theoretical and methodological deepening. For this reason, the conceptual and methodological contributions of Cybermetrics were used for the research, being a consolidated subfield and which carries in its scope all the elements that also compose the altmetrics. Thus, the research sought to use the information knowledge that can be extracted from the web and social networks, in addition to taking into account the particularities and quirks of each environment to propose the use of crossover techniques for develop a model that can assist in the information metrics studies. The preliminary results point to the feasibility of using the crossover, at the same time, enabled the development of three index (exposure, interaction and appropriation), which are used as a way to contribute to the studies that seek direct ways of looking at the amount of data made available by and in digital environments.

Palavras-chave: Crossover. Alternative Metrics. Altmetric. Cibermetric. Social Media.

LISTA DE SIGLAS

ALM	Article Level Metrics
AMEC	Associação Internacional de Mensuração e Avaliação da Comunicação
API	Application Programming Interface
AVA	Ambientes Virtuais de Aprendizagem
B2B	Business to Business
B2C	Business to Consumer
CGU	Conteúdo Gerado pelo Usuários
DOI	Digital Object Identifier
FI	Fator de Impacto
FIW	Fator de Impacto Web
JIF	Journal Impact Factor
NSF	National Science Foundation
ORCID	Open Researcher and Contributor ID
PMID	PubMed Unique Identifier
PLoS	Public Library of Science
Q&A	Questions and Answers
RSS	Rich Site Summary
TICs	Tecnologias de Informação e Comunicação
UGC	User Generated Content
URL	Uniform Resource Locator
WIF	Web Impact Factor
WoS	Web of Science
WWW	World Wide Web

LISTA DE FIGURAS

Figura 1 - O tweet de Jason Priem, considerado o momento em que o termo "altmetrics" foi criado	17
Figura 2 - Modelo de categorias e tipos de atos referentes a objetos de pesquisa (documentos acadêmicos e agentes)	32
Figura 3 - Infográfico com as principais redes sociais entre os anos de 1995 a 2011.....	34
Figura 4 - Representação gráfica do Modelo AIDA proposto por Kotler e Keller	39
Figura 5 - Mapa Mundi das Redes Sociais em janeiro de 2018	44
Figura 6 - Tela de entrada da ferramenta Altmetric	47
Figura 7 - Reverberação de uma publicação nas redes sociais segundo o Altmetric.com	48
Figura 8 - Tela de entrada da ferramenta Plum Analytics.....	49
Figura 9 - Reverberação de uma publicação nas redes sociais segundo o Plum Analytics.....	50
Figura 10 - Tela de entrada da ferramenta Impactstory	50
Figura 11 - Tela que apresenta a reverberação da produção científica	51
Figura 12 - Tela de entrada para o sistema da PLOSALM	52
Figura 13 - Divisão de categorias utilizadas para agrupar as redes sociais.....	58
Figura 14 - Adaptação do Modelo AIDA para o trabalho com os indicadores	60
Figura 15 - Esquema para a análise de correlação entre os indicadores.....	75

LISTA DE GRÁFICOS

Gráfico 1 - Quantidade de envolvimento por tempo de publicação – dataset 2016	65
Gráfico 2 - Quantidade de envolvimento por tempo de publicação – dataset 2017	65
Gráfico 3 - Volume de dados por redes que compõem o Top 100 dos documentos com maior relevância no ano de 2016, segundo totalização realizada pela ferramenta Altmetric	69
Gráfico 4 - Volume de dados por redes que compõem o Top 100 dos documentos com maior relevância no ano de 2017, segundo totalização realizada pela ferramenta Altmetric	69
Gráfico 5 - Somatório dos envolvimento para o Indicador Exposição para cada um dos documentos do dataset de 2016 na ordem do ranking original	72
Gráfico 6 - Somatório dos envolvimento normalizados do Indicador Exposição para cada um dos documentos do dataset de 2016 na ordem do ranking original	72
Gráfico 7 - Somatório dos envolvimento para o Indicador Interação para cada um dos documentos do dataset de 2016 na ordem do ranking original	73
Gráfico 8 - Somatório dos envolvimento normalizados do Indicador Interação para cada um dos documentos do dataset de 2016 na ordem do ranking original	73
Gráfico 9 - Somatório dos envolvimento para o Indicador Apropriação para cada um dos documentos do dataset de 2016 na ordem do ranking original	74
Gráfico 10 - Somatório dos envolvimento normalizados do Indicador Apropriação para cada um dos documentos do dataset de 2016 na ordem do ranking original	74
Gráfico 11 - Distribuição da relação entre os Indicadores sem a realização da normalização para o dataset de 2016	75
Gráfico 12 - Distribuição da relação entre os Indicadores com base em dados normalizados por logaritmo natural para o dataset de 2016	76
Gráfico 13 - Histogramas, coeficientes de correlação e diagramas de dispersão de $r =$ Pearson, $\rho =$ Spearman, e $\tau =$ Kendall Tau, entre os três indicadores normalizados: Exposição, Interação e Apropriação para o dataset de 2016	77
Gráfico 14 - Somatório dos envolvimento para o Indicador Exposição para cada um dos documentos do dataset de 2017 na ordem do ranking original	79
Gráfico 15 - Somatório dos envolvimento normalizados do Indicador Exposição para cada um dos documentos do dataset de 2017 na ordem do ranking original	80
Gráfico 16 - Somatório dos envolvimento para o Indicador Interação para cada um dos documentos do dataset de 2017 na ordem do ranking original	80
Gráfico 17 - Somatório dos envolvimento normalizados do Indicador Interação para cada um dos documentos do dataset de 2017 na ordem do ranking original	81
Gráfico 18 - Somatório dos envolvimento para o Indicador Apropriação para cada um dos documentos do dataset de 2017 na ordem do ranking original	81
Gráfico 19 - Somatório dos envolvimento normalizados do Indicador Apropriação para cada um dos documentos do dataset de 2017 na ordem do ranking original	82
Gráfico 20 - Distribuição da relação entre os Indicadores com base em dados normalizados por logaritmo natural para o dataset de 2017	83
Gráfico 21 - Histogramas, coeficientes de correlação e diagramas de dispersão de $r =$ Pearson, $\rho =$ Spearman, e $\tau =$ Kendall Tau, entre os três indicadores normalizados: Exposição, Interação e Apropriação para o dataset de 2017	84

LISTA DE QUADROS

Quadro 1 - Modelo de hierarquia de respostas baseada nos estudos de Kotler e Keller.....	38
Quadro 2 - Fatores atribuídos as redes sociais pela ferramenta Altmetric.....	48
Quadro 3 – Parâmetros para a construção de um crossover.....	56
Quadro 4 - Indicadores e respectivas redes.....	62
Quadro 5 - Apresentação dos períodos que compreendem cada dataset.....	64
Quadro 6 - Distribuição dos conteúdos, por áreas do saber, dos Top 100 de 2016 e 2017	70
Quadro 7 - Comparativo por redes do engajamento do público nos anos de 2016 e 2017	78
Quadro 8 - Ranking com os 10 primeiros colocados no Top 100 do ano de 2016, tabualizado pelo indicador Apropriação	86
Quadro 9 - Ranking com os 10 primeiros colocados no Top 100 do ano de 2017, tabualizado pelo indicador Apropriação	88

LISTA DE EQUAÇÕES

Equação 1 - Fórmula utilizando Desvio Padrão	63
Equação 2 - Fórmula de depreciação.....	66

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Proposta e Hipótese	18
1.2	Objetivo Geral	20
1.3	Objetivo específicos	20
1.4	Justificativa	20
2	DAS MÉTRICAS TRADICIONAIS ÀS MÉTRICAS ALTERNATIVAS.....	23
2.1	Desenvolvimento dos estudos métricos na web.....	24
2.2	Estudos Métricos e a Cibermetria	27
3	A CIBERMETRIA E OS ESTUDOS MÉTRICOS.....	29
3.1	Bases conceituais dos estudos cibernétricos	29
3.2	Aplicação cibernétrica nos estudos altnétricos.....	31
4	MÍDIAS SOCIAIS, REDES SOCIAIS, SITES DE REDES SOCIAIS	34
4.1	Redes sociais e o marketing como suporte aos estudos métricos	36
4.2	A extração de dados nas redes sociais	40
4.3	Limitação e mensuração dos dados nas redes sociais.....	42
5	FUNDAMENTOS PARA UM CROSSOVER MÉTRICO	54
5.1	Crossover métrico	55
6	METODOLOGIA	58
6.1	Desenvolvimento de Indicadores	58
6.2	Procedimentos de coleta dos dados	61
6.3	Tratamento dos dados	62
6.4	Definição da janela de tempo	64
7	CROSSOVER E INTERPRETAÇÃO DOS DADOS.....	68
7.1	Modelagem dos Indicadores	71
7.2	Prova de conceito do Crossover métrico	78
7.3	Construção dos Rankings	85
8	CONSIDERAÇÕES FINAIS	90
	REFERÊNCIAS	93
	APÊNDICE A – CROSSOVER DO TOP 100 DO ANO DE 2016.....	102
	APÊNDICE B – CROSSOVER DO TOP 100 DO ANO DE 2017	114
	APÊNDICE C - Ranking do Top 100 do ano de 2016	126
	APÊNDICE D - Ranking do Top 100 do ano de 2017.....	135
	APÊNDICE E – FOLHA DE PROGRAMAÇÃO DO R PARA OBTENÇÃO DOS RESULTADOS	144

1 INTRODUÇÃO

Nas últimas duas décadas, os estudos baseados em dados originados dos ambientes virtuais ganharam espaço influenciados pelo aumento no volume de dados deixados por usuários de serviços e ferramentas digitais, caracterizados por O'Reilly (2009) como beta-perpétuo e permeados por uma lógica onde o público consumidor passou a ser um aliado no desenvolvimento dos conteúdos e das aplicações, que possibilitaram a “democratização das ferramentas de produção” (ANDERSON, 2006, p. 38) e favorecem a popularização nas formas de distribuição. Esse cenário é descrito por Kennedy (2016) como o momento em que a dataficação deixou de ser restrita e possibilitou que usuários comuns pudessem realizar a mineração de dados e transformar informações em inteligência.

Nesse novo mundo, ainda em desenvolvimento, indivíduos de diferentes orientações ideológicas conquistaram status e autoridade para influenciar política e economicamente o mundo real e o virtual, em uma verdadeira simbiose homem-máquina. Esse novo sujeito, cujo poder de influência foi identificado rapidamente por empresas que passaram a adotar os conteúdos gerados pelos usuários¹ (*User Generated Content* - UGC) como uma forma de ratificar os seus produtos e serviços, também virou objeto de atenção de pesquisadores que passaram a se debruçar sobre esse novo ambiente com o objetivo de extrair indicadores que trouxessem informações mais detalhadas sobre como acontecia a reverberação da produção científica. Em ambos os casos, o efeito foi de buscar nos dados, ferramentas para o desenvolvimento e aperfeiçoamento de produtos, serviços e métricas capazes de demonstrar os possíveis impactos sociais e no desenvolvimento científico, que foram definidos por Setzer (1999), como

uma sequência de símbolos quantificados ou quantificáveis. (...) um texto é um dado, também são dados fotos, figuras, sons gravados e animação, pois todos podem ser quantificados ao serem introduzidos em um computador, a ponto de se ter eventualmente dificuldade de distinguir a sua reprodução com o original. (SETZER, 1999, p. 1)

Essa nova área de atuação, da exploração e/ou mineração dos dados, composta por Academia, empresas de pesquisa, agências de publicidade, comunicação e assessoria de imprensa passou a acompanhar de forma contínua o que se fala delas e de temas relacionados, como forma de colher insumos para estratégias e novos negócios, e identificação de novas formas de se fazer pesquisas. Consultorias especializadas totalmente baseadas em dados sociais

¹ O termo alcançou popularidade em 2005, e é aplicado para descrever as várias formas de conteúdo criados por usuários finais (VICKERY; WUNSCH-VINCENT, 2007). Ver também Chiluka; Andrade e Pouwelse, (2011); Agichtein, et al. (2008) e Kaplan e Haenlein (2010)

digitais também foram criadas ao longo dos últimos anos, entregando estudos de percepção de imagem, análise de performance, monitoramento de crises digitais e uma miríade de serviços.

Em paralelo, uma variedade de metodologias e abordagens foram aperfeiçoadas e utilizadas como, aplicação de estatística descritiva, análise etnográfica, análise de redes e links, entre outras (ACKLAND, 2013; HALFPENNY; PROCTER, 2015). No entanto, Pang, Lee e Vaithyanathan (2002), ao estudar o início do monitoramento das redes sociais e o desenvolvimento de técnicas e metodologias para a mensuração dos conteúdos publicados nestes ambientes, observaram que a utilização destas técnicas e a evolução dos estudos métricos com dados de ambientes digitais foi favorecido pela mudança na forma como era realizado o acesso aos dados, que nos estágios iniciais de desenvolvimento das redes sociais tinham às conversas em ambientes virtuais rapidamente disponibilizadas pelas empresas tendo como única exigência para a liberação desses volumes de dados e metadados a criação de um perfil na rede.

Neste contexto, o Orkut² é apontado como marco para a mudança das políticas de disponibilização dos dados de usuários, por não permitir a sua captura por ferramentas de monitoramento. Além de ter sido o primeiro site de rede social a trazer os conceitos de celebridades digitais e monetização de perfis e comunidade. No entanto, a evolução da monetização das ferramentas sociais como veículo de mídia foi iniciada com o Myspace³ em 2003, primeira ferramenta a permitir a segmentação comportamental em anúncios publicitários nos perfis (ITO, 2017). As demais ferramentas sociais que surgiram posteriormente, trouxeram a visão comercial da necessidade de disponibilizar informações para demonstrar por meio de números o volume de acessos e interações como forma rentabilizar a rede e manter suas operações.

Esse modelo foi o adotado pelo Facebook, rede lançada no mesmo ano do Orkut (2004) mas somente para estudantes da Universidade Harvard e que posteriormente foi ampliando a sua abrangência para outras universidades americanas até o final de 2005 quando abriu o cadastro para qualquer interessado. A diferença entre ambas as redes foi a disponibilização dos dados. Enquanto o Orkut durante a sua existência não permitia a indexação do seu conteúdo nem mesmo por buscadores, o Facebook quando abriu o ingresso do grande público lançou uma API (*Application Programming Interface*) para qualquer usuário baixar dados públicos

² <https://www.orkut.com>

³ <https://www.myspace.com>

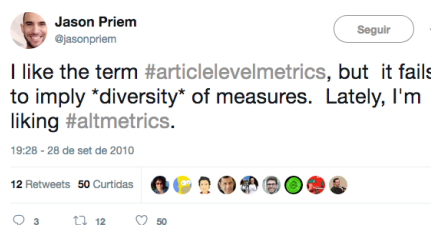
dos usuários. Essa estratégia estava alinhada com o mercado que passou a exigir a profissionalização na divulgação de dados de acesso das redes.

No entanto, ao mesmo tempo que isto representava uma abertura de acesso aos dados, se configurava como uma restrição ao exigir conhecimento de programação e das funções específicas que tornavam os dados acessíveis, assim permitindo-se utilizar características do software menos evidentes ao utilizador tradicional. A chamada por Interface de Programação de Aplicações ou API permite além da comunicação e automação de troca de dados entre sistemas, o controle e com isso a cobrança pelo acesso a funcionalidades adicionais (dados de acesso, tipo de conteúdo, métricas das redes e dos usuários) de empresas e demais interessados. Porém, também foram as APIs que permitiram o desenvolvimento de pesquisas, pois facilitaram em muitos casos, mesmo que de forma limitada, a extração e compartilhamento de dados provenientes das redes sociais.

Essas mudanças embora tenham tido como foco restringir o acesso de empresas publicitárias aos dados, gerou reflexos em pesquisas acadêmicas que também observaram o potencial das informações digitais para estudos de dispersão e engajamento de público com documentos científicos, conforme afirma Piwowar (2013, p. 9) ao identificar que as “citações não são a única maneira de representar o impacto de um artigo de pesquisa. Alguns indicadores alternativos têm sido objeto de pesquisa na webometria e na bibliometria há anos, incluindo contagens de downloads e menções em patentes.” (Tradução nossa)⁴.

Com o crescimento dos estudos métricos na web e com o objetivo de atender essa demanda surgiu a “*altmetrics*” (PRIEM; TARABORELLI; GROTH; NEYLON, 2010), em português, *altmetria* (GOUVEIA, 2013), ou métricas alternativas cuja atribuição do nome aconteceu por meio de um post publicado no Twitter em 28 de setembro de 2010 (Figura 1) e tem como principal escopo medir a disseminação de documentos científicos nas mídias sociais.

Figura 1 - O tweet de Jason Priem, considerado o momento em que o termo "altmetrics" foi criado



Fonte: <https://twitter.com/jasonpriem/status/25844968813>

⁴ “Citations are not the only way to represent the impact of a research article. A few alternative indicators have been the subjects of webometrics and bibliometrics research for years, including download counts and mentions in patents.” (PIWOWAR, 2013, p. 9)

Para Araújo e Furnival (2016) esse contexto ganha relevância devido

a visibilidade e atenção online tem sido considerada um imperativo dos últimos tempos e mesmo se, por algum motivo um pesquisador não estiver inserido no ambiente online do nexo digital que caracteriza a vida do século XXI, suas pesquisas estão, seja por meio das revistas eletrônicas, ou das bases de dados e repositórios, ou websites pessoais ou de compartilhamento de arquivos (como Figshare, Slideshare etc.). E os usuários estão “citando” essas pesquisas nos grupos coletivos em que participam. (ARAÚJO; FURNIVAL, 2016, p. 74)

Dessa forma, podemos concluir que o início dos estudos com dados oriundos dos meios digitais surgiu com o objetivo de tentar oferecer informações mais detalhadas, quando usamos como base comparativa nos atuais modelos de aferição, e ao mesmo tempo, subsidiar pesquisadores, agências de fomento e publicações com dados sobre a disseminação e/ou impacto das publicações de artigos científicos.

1.1 Proposta e Hipótese

A pesquisa parte da premissa de que os novos meios permitem novas formas para trabalhar os dados provenientes das redes sociais e permitem a utilização destas informações para aferir de forma mais precisa o impacto de uma publicação científica. Esse pressuposto se baseia na ideia de que as atuais tentativas de transpor índices, relatórios e rankings da produção científica são cercadas por vícios de origem e apresentam uma visão distorcida do engajamento do público nos ambientes digitais.

Um ponto que nos remete para essa hipótese é de que empresas e pesquisadores continuam a utilizar os dados sociais para os mais variados fins. Entretanto, muitos destes trabalhos ignoram ou não levam em consideração as arquiteturas das mídias sociais que são frequentemente insensíveis à granularidade da sociabilidade humana (HULL; LIPFORD; LATULIPE, 2011) e tratam as relações como se fossem únicas, ignorando as profundas diferenças entre os tipos de relação social (familiar, profissional, colegial, comercial, cívico, etc.).

Ao se levar o foco somente para as pesquisas científicas, é possível observar o crescimento de cientistas adotando as mídias sociais como um dos canais para divulgar, discutir e compartilhar os resultados de suas pesquisas, como observa Bollen, van de Sompel, Hagberg e Chute (2009)

os dados de uso [na web] permitem que a atividade científica seja observada imediatamente após a publicação, ao invés de esperar que as citações surjam em publicações e sejam incluídas em bancos de dados de citações. Um processo que, com

atrasos médios de publicação, pode levar vários anos. (BOLLEN; VAN DE SOMPEL; HAGBERG; CHUTE, 2009, p. 2 - tradução nossa)⁵

A opinião está em linha com o aventado por Bossy (1995), ao identificar que os meios digitais trouxeram uma nova forma de acompanhar e quantificar a dispersão do conteúdo científico produzido nestes ambientes, sendo, no entanto, necessário o entendimento do que será medido e com qual finalidade.

As mudanças trazidas pela rede são profundas e devem ser integradas em um novo modelo para o comportamento das comunidades científicas. Esta deve ser a nossa primeira tarefa - a base sobre a qual poderemos construir meios de observar o fluxo de informação e circulação acadêmica. (BOSSY, 1995, sem paginação, tradução nossa)⁶

A visão é complementada por Piwowar (2013) ao apontar que

Durante décadas, a métrica mais comum para avaliar o impacto da pesquisa foi o número de vezes que um artigo de pesquisa é citado por outros artigos. [...] No entanto, à medida que a comunicação acadêmica se torna cada vez mais on-line, mais indicadores se tornam disponíveis (PIWOWAR, 2013, p. 9 - tradução nossa)⁷

Essa prática demonstra ainda ser predominante e os estudos, na maioria dos casos, são para realizar comparações entre possíveis métricas de ambientes digitais com as tradicionais, como o realizado por Bollen, van de Sompel, Hagberg e Chute (2009), que analisou as possíveis novas medidas de impacto que influenciariam os rankings científicos; e a análise realizada por Hoffmann, Lutz e Meckel (2016), ao estudar a possível relação entre dados gerados pelos meios digitais e os indicadores considerados tradicionais. Um dos poucos estudos que buscaram construir um novo modelo de mensurar o real impacto do de ferramentas sociais nas produções científicas foi produzido por Haustein, Bowman e Costas (2015), que propôs a criação de indicadores como forma avaliar com maior precisão o envolvimento do público dentro de ambientes digitais.

Com base nestas questões sobre como medir o impacto da produção científica, sem abandonar o refinamento e rigor da mensuração das métricas tradicionais, mas conciliando com todo o conhecimento e informações que podem ser extraídas das mídias sociais e também levando em consideração suas particularidades e peculiaridades, é que desenvolvemos essa pesquisa.

⁵ “The resulting usage data allows scientific activity to be observed immediately upon publication, rather than to wait for citations to emerge in the published literature and to be included in citation databases such as the JCR; a process that with average publication delays can easily take several years.” (BOLLEN; VAN DE SOMPEL; HAGBERG; CHUTE, 2009, p. 2)

⁶ “The changes brought about by the network are profound and must be integrated in a new model for scientific communities behavior. This should be our first task - the foundation on which we will be able to build means of observing the flow of academic information and circulation.” (BOSSY, 1995, sem paginação)

⁷ “For decades, the most common metric for evaluating research impact has been the number of times a research article is cited by other articles. However, as scholarly communication moves increasingly online, more indicators have become available” (PIWOWAR, 2013, p. 9)

1.2 Objetivo Geral

Analisar a possibilidade de construir um modelo que possa auxiliar nos estudos métricos da informação a partir dos assim chamados dados alométricos com base no conceito de crossover, que será utilizado para modelar um ou mais indicadores capazes de traduzir de forma mais detalhada como é o consumo da produção científica nos ambientes digitais, neste caso baseado em parâmetros do marketing.

1.3 Objetivo específicos

- Definir categorias em que podem ser enquadrados os conteúdos/canais na mensuração dos ambientes digitais;
- Apontar as métricas e métodos para a avaliação alternativa;
- Propor uma metodologia para a modelagem dos dados;
- Descrever as melhores práticas para a análise e ranqueamento da produção científica por meio da captura de dados digitais.

1.4 Justificativa

Essa pesquisa se qualifica como importante instrumento ao se juntar a outros estudos que buscam dimensionar o impacto das métricas alternativas para os segmentos que buscam conhecer a repercussão, o alcance e a utilização de pesquisas científicas em ambientes digitais, além de apresentar como diferencial um olhar mais acurado sobre os dados extraídos das ferramentas sociais.

Por esse motivo a pesquisa é baseada em camadas ou partes, que juntas irão oferecer o resultado esperado. A técnica de crossover apresenta várias acepções, e está presente nas mais variadas áreas do conhecimento. Na biologia é possível observar como um fenômeno na recombinação e a troca aleatória de material genético durante a meiose, no campo das ciências sociais a apropriação do termo é utilizada apenas como adjetivo para identificar a combinação/fusão de teorias (DELLA PORTA; KEATING, 2008; GILLIAM, 1996), para citar modelos metodológicos de combinação de dados (ARCHIBALD; RADIL; ZHANG; HANSON, 2015), e para citar a combinação de processos no desenvolvimento de estudos científicos (ABBASI; CHEN; SALEM, 2008; COVER; THOMAS, 1991).

No marketing, as pesquisas nomeadas com este termo têm definições distintas e o utilizam quando marcas de campos diferentes, ou não, se unem com o objetivo de produzir um novo produto para atingir consumidores de diferentes áreas (JI; SHEN, 2013). Esse termo se

popularizou principalmente com a massificação das redes sociais, com os novos modelos de campanhas de marketing e com a necessidade ampliar os resultados, sendo que muitas vezes é utilizando na recombinação de produtos da mesma marca.

Partindo dessa premissa, e confrontando como se utilizam as métricas no mercado e academia, em especial na área da Ciência da Informação, foi possível trazer o conceito de crossover como uma forma de produzir um novo indicador métrico capaz, não de substituir, mas de complementar a visualização do alcance e uso das publicações científicas, que seriam impossíveis de se extrair somente com as métricas tradicionais baseadas na análise de citação. Embora seja claro que as citações são um indicador importante, mas que registram apenas um pedaço da trajetória da pesquisa, as metrias alternativas oferecem como benefícios uma compreensão mais sutil do impacto, ao mostrar quais produtos acadêmicos são lidos, discutidos, salvos e citados; e reduzem a janela de tempo entre publicação e repercussão do conteúdo mostrando o impacto em dias. Além disso permite ampliar o entendimento de onde (redes sociais, sites, blogs), como (vídeo, texto, matérias) e por quem (estudiosos, profissionais, publico geral) a pesquisa está sendo utilizada.

Metodologicamente, a pesquisa se apoiou nas novas racionalidades que nos permitem interpretar o quanto os fluxos de informação têm impacto social junto à comunidade científica e no restante da sociedade. Ziman (1979) esclarece que:

Na maioria das vezes o campo que está sendo investigado fica fora do escopo de teorias solidamente firmadas. Nossa definição de ciência implica que devem existir vastas áreas de ignorância onde o conhecimento público não tem jurisdição, isto é, onde não existem modelos de teoria e observação aprovados unanimemente. (ZIMAN, 1979, p. 63)

A observação feita por Ziman nos traz os elementos sobre as limitações que os instrumentos, métodos, e teorias têm diante de novos fenômenos que surgem em consequência dos avanços tecnológicos, das pressões sociais que incidem sobre a ciência e que demandam uma possível urgência em identificar formas e formatos de avaliação. Assim, a temática proposta nesta pesquisa se aproxima do pensamento, pelo fato de não ser possível atribuir necessariamente o atributo de científico aos meios analisados. Porém, entende-se que interpretações sobre esses dispositivos de ligação contribuem para delinear implicações acerca dos usos em processos de comunicação compostos por pesquisadores, tendo em vista as motivações pelas quais as redes sociais são ambientes que estão delineando um novo modelo no fazer pesquisa.

Como afirmou Bossy (1995, não paginado) “a diversidade de novos padrões de comunicação na rede eletrônica às vezes rompe as fronteiras entre circulação formal e informal,

entre atividades que ocorrem dentro e fora dos laboratórios”⁸. É exatamente alicerçado nessa premissa que esta pesquisa se desenvolve ao avaliar, respeitar e entender a importância das métricas tradicionais, mas também com a visão que as métricas sociais ou alternativas são um novo campo, que apesar de em muitos casos se apropriar das bases teóricas e conceituais dos estudos métricos, necessita de um novo aporte metodológico com vistas ao novo campo que se apresenta, com particularidades, características e dinâmica distintas.

Assim, a pesquisa está organizada em oito capítulos: no segundo é apresentado a evolução dos estudos métricos com o objetivo de situar em qual subcampo dos estudos métricos está inserida. No terceiro é discutido a atual fase dos estudos cibernétricos e suas aplicações em junção com a Almetria. No quarto capítulo é realizada a conceituação das redes sociais como um ambiente de estudo que precisa ser delimitado em sua estruturação como ferramenta, apropriação pelo público e pelos cientistas. O quinto capítulo apresenta os fundamentos para a aplicação da técnica do crossover. No sexto capítulo está descrito a metodologia, o tratamento, a coleta e a análise dos dados. No sétimo é apresentada a aplicação do crossover e a análise quantitativa dos dados. O oitavo capítulo apresenta as considerações finais da pesquisa.

⁸ “The diversity of new patterns of communication on the electronic network blurs sometimes the frontiers between formal and informal circulation, between activities taking place inside and outside laboratories.” (BOSSY, 1995)

2 DAS MÉTRICAS TRADICIONAIS ÀS MÉTRICAS ALTERNATIVAS

Meios para identificar a “popularidade” de periódicos e artigos científicos fazem parte dos círculos acadêmicos desde do início do século XX. Gross e Gross em um artigo publicado na Science, em 1927, demonstraram um sistema que buscava solucionar o problema de “quais periódicos eram necessários em uma biblioteca para preparar o aluno para o trabalho avançado, levando em consideração também os materiais necessários para a estimulação e desenvolvimento intelectual de suas capacidades?” (GROSS; GROSS, 1927, p. 386, tradução nossa)⁹. Um ponto importante desse trabalho é que posteriormente, e atualmente, tem sido utilizado como base para vários argumentos contrários ao Fator de Impacto (FI) proposto por Eugene Garfield, em 1955¹⁰, e repousa no fato que o indicador foi desenvolvido com foco no atendimento das necessidades de pesquisadores americanos. Esse modelo acabou por privilegiar as publicações em língua inglesa que passaram a ser cada vez mais citadas nos EUA e no exterior.

Outro ponto de questionamento do FI, é devido a métrica possuir cálculo específico para publicações indexadas pela Web of Science (WoS). Além disso, a fórmula é simples, o que para muitos pesquisadores é considerada fácil de manipular (LARIVIERE; SUGIMOTO, 2018; MONASTERSKY, 2005; RAMSDEN, 2009), e a lógica do FI influencia a prática da autocitação e as publicações de pesquisas de forma fracionada com o objetivo de ampliar o campo de citação. Esse desmembramento dos resultados de pesquisas passou a ser conhecido como “*Salami Science*” e segundo Thyer (2008) refere-se a prática de publicar vários artigos derivados de um único estudo. A prática tem seus defensores e críticos, porém o ponto de consenso é a crescente necessidade de apresentar resultados originados pelo número de publicações realizadas no período de um ano.

[...] cientistas e suas instituições são avaliados a partir de fórmulas matemáticas que levam em conta três ingredientes, combinados ao gosto do freguês: número de trabalhos publicados, quantas vezes esses trabalhos foram citados na literatura e qualidade das revistas (medida pela quantidade de citações a trabalhos publicados na revista). (REINACH, 2013, p. 402)

Esse modelo é o que prevaleceu nas últimas décadas e ainda é o padrão utilizado pelas agências de fomento. Mesmo com o desenvolvimento do campo dos estudos métricos, que tiveram como influenciador a evolução das bases computadorizadas, a contagem de citações como fator determinante para apontar o sucesso de uma pesquisa se mantém. No entanto, o

9 “What files of scientific periodicals are needed in a college library successfully to prepare the student for advanced work, taking into consideration also those materials necessary for the stimulation and intellectual development of the faculty?” (GROSS; GROSS, 1927, p. 386)

¹⁰ Ver também: Citation indexes for science: a new dimension in documentation through association of ideas. (GARFIELD, 1956). Disponível em: <http://www.garfield.library.upenn.edu/papers/31.html>

amadurecimento e surgimento de outros subcampos como a Cibermetria e Webometria permitiram uma ampla diversidade de estudos de citações, especialmente mapeando domínios científicos, além de ter criado a possibilidade de acompanhar o crescimento, a difusão, a especialização, a colaboração, o impacto e a obsolescência de literatura e conceitos, nas mais diversas áreas (BJÖRNEBORN; INGWERSEN, 2004).

A mudança de paradigma de ver as citações como padrão de excelência lentamente está acontecendo, como pode ser observado nos órgãos de financiamento do Reino Unido que definiram que indicadores para o cálculo do impacto da excelência de uma pesquisa “incluam todos os tipos de benefícios e impactos sociais, econômicos e culturais além do meio acadêmico, decorrentes de excelentes pesquisas” (DEPARTMENT FOR EMPLOYMENT AND LEARNING, 2011, p. 4 - tradução nossa)¹¹.

Além disso, outras métricas, principalmente, em nível de artigo, promoveram o desenvolvimento nos campos da Bibliometria, o surgimento da Informetria, Cientometria e Webometria. Segundo Vanti (2002, p. 153) “Todas têm funções semelhantes, mas, ao mesmo tempo, cada uma delas propõe medir a difusão do conhecimento científico e o fluxo da informação sob enfoques diversos.”. Mais recentemente, a Altmetria também passou a ser utilizada em estudos que buscam identificar a dispersão dos resultados de pesquisas em ambientes digitais.

2.1 Desenvolvimento dos estudos métricos na web

Com o crescimento da web e a perspectiva de estudos quantitativos na rede surgiram novas terminologias para esse campo de pesquisa emergente, como: Netometric (BOSSY, 1995), Webometry (ABRAHAM, 1997), Internetometrics (ALMIND; INGWERSEN, 1997), Webometrics (ALMIND; INGWERSEN, 1997), Web Bibliometry (CHAKRABARTI; JOSHI; PUNERA; PENNOCK, 2002), e Webmetrics, termo usado principalmente em Ciência da Computação (DHYANI; NG; BHOWMICK, 2002).

Na Ciência da Informação um dos primeiros estudos webométricos foi realizado por Larson (1996) cujo objetivo foi a realização de uma análise co-citação por links utilizando a ferramenta Altavista, principal buscador na época. Nos resultados o pesquisador chegou à conclusão que as descobertas realizadas apresentaram uma nova forma para mapear o terreno

¹¹ “will include all kinds of social, economic and cultural benefits and impacts beyond academia, arising from excellent research” (DEPARTMENT FOR EMPLOYMENT AND LEARNING, 2011, p. 4)

intelectual, mas ao mesmo tempo identificou a necessidade de aperfeiçoamento dos métodos para conseguir realizar um estudo com o grande volume de conteúdos disponíveis na rede.

Os métodos utilizados por Larson (1996) para a realização do seu experimento foram uma adaptação do modelo aplicado nos mapeamentos de autores e revistas. Por esse motivo, é destacado esse como marco para os estudos métricos na web dentro da Ciência da Informação. Essa adaptação de metodologia pode ter sido um dos caminhos que levaram Ingwersen (1998) a identificar uma possível similaridade entre links e citações e por consequência ter apresentado uma releitura do Fator de Impacto, que chamou de Fator de Impacto Web¹² (FIW, em inglês: Web Impact Factor - WIF) cujo propósito também é o de avaliar ou oferecer um indicador de performance para as páginas web.

Nos estudos métricos que passaram a ser realizados, com o objetivo de tentar observar a possibilidade de utilizar esses dados como métricas de produtividade, pesquisadores passaram a analisar os dados da web, principalmente, no estudo de links, com o objetivo de construir uma ponte, uma associação entre as métricas existentes com os dados extraídos dos ambientes digitais, como foram os casos de estudos realizados por Smith (2004), Bar-Ilan (2005), Noruzi (2005) e Thelwall (2006, 2011). Nesses estudos foram avaliadas a utilização da contagem de links de forma comparativa com o número de citações e relevância da publicação ou alcance da produtividade dos pesquisadores.

O que se pode concluir desses estudos de análise de links quando comparados às citações, é que os mesmos utilizam abordagens instrumentais similares, e não chegaram a considerar as funcionalidades e/ou possíveis potencialidades das redes sociais (no sentido amplo) e as ligações ou conexões não somente restritas ao ambiente científico. Isso acontece devido as citações serem parte de uma relação argumentativa e são influenciadas por uma intervenção analítica típica da produção científica, enquanto os links, diferentemente, são utilizados sem padrões metodológicos e cuidados com a sua utilização, como o estudo de Smith (2004) que analisou em que medida os links são análogos as citações tradicionais.

Para Lang; Gouveia e Leta (2008, p. 140), “o link, além de se configurar como unidade central de informação para estudos webométricos, é um importante indicador do impacto e da posição que determinado site ocupa no espaço da rede (...) funciona como legitimador desses sites para os mecanismos de busca”. No entanto, como lembram Frago; Recuero e Amaral (2011, p. 143).

¹² Embora a definição do Fator de Impacto Web seja comumente atribuída a Ingwersen (1998). Ela foi utilizada pela primeira vez por Rodríguez i Garín (1997) na *Revista española de documentación científica*.

A herança das Ciências da Informação, em especial a Biblioteconomia, confere centralidade à comparação entre hiperlinks e citações, e com isso, coloca ênfase em um subconjunto das possíveis motivações e usos desses conectivos. Apesar da aparente rigidez da analogia, ela preserva razoável grau de flexibilidade, uma vez que as citações, como os links, não são todas iguais: referências ao trabalho de outros podem endossar, reconhecer como fonte de conceitos ou dados, criticar, analisar, etc. (FRAGOSO; RECUERO; AMARAL, 2011, p. 143)

O pensamento das autoras expõe a necessidade de uma nova discussão sobre como devem ser vistos os conteúdos compartilhados nas redes sociais, pois é possível observar o princípio de uma desconstrução dos conceitos relativos aos suportes e fronteiras sobre os quais repousavam os fluxos da comunicação científica e que foram totalmente subvertidos pela lógica de participação em que atuam diferentes atores sociais.

Essas possíveis limitações dos estudos métricos em ambientes digitais ganham maior dimensão nas redes sociais, principalmente, por nesses ambientes cada sujeito fazer a apropriação das ferramentas sociais da forma que mais lhe convém, o que impede a definição de normas rígidas. Por consequência, possíveis comparações entre links e citações devem considerar o propósito e os meios nos quais esses conectivos são inseridos.

As citações não ocorrem apenas em canais formais, uma vez que são feitas pelos pesquisadores em diferentes situações nas quais a avaliação prévia não é uma característica do processo e nesse sentido é correto afirmar que o ato de ligar partes de textos é inerente às produções textuais na ciência, independentemente das formas de apresentação, níveis de linguagem, discursos produzidos, meios e canais utilizados pelos pesquisadores. Contudo, Thelwall (2004, p. 3, tradução nossa) lembra que a “[. . .] análise de links adota e adapta técnicas existentes da Ciência da Informação para a meta-análise dos documentos através da investigação de conexão entre documentos”¹³.

A opinião é complementada por ctar que

[...] uma analogia com as tradicionais citações bibliográficas que remetem a outros textos, os *links* têm sido considerados por alguns autores como citações, pois cumpriam função semelhante no meio eletrônico, ao estabelecerem relações entre páginas web

Uma visão semelhante é apresentada por Smith (2004) a partir de seus estudos, onde o pesquisador identificou que 20% dos links da amostra poderiam ser considerados equivalentes as citações. Essa descoberta, no âmbito da comunicação científica é importante por permitir interpretações sobre a dinâmica e as relações de poder similares as tradicionais (um número restrito de pesquisadores detém um grande volume de citações) nos ambientes digitais.

¹³ link analysis is to adopt and adapt existing information science techniques for the meta-analysis of documents through investigating inter-document connections. (THELWALL, 2004, p. 3)

Outra interpretação é proposta por Kelleher e Miller (2006) ao definir os blogs como uma forma equivalente aos periódicos científicos on-line, no qual seus autores apresentam conhecimentos sobre seus domínios profissionais, incluindo os avanços nas pesquisas, referências e observações. Esse argumento ao equiparar os canais ignora as competências necessárias aos blogueiros, o público para quem essas informações/conteúdos são direcionadas, e o fato de que estes ampliam a comunicação científica para não pesquisadores. Primo (2010) ao analisar a potencialidade dos blogs para a comunicação científica identifica que os blogueiros pesquisadores seguem a mesma tendência dos blogueiros comuns e os utilizam como mecanismos de obtenção de benefícios como a autopublicidade. O resultado para essa autopromoção é o destaque que será conquistado diante dos seus pares, como exemplifica Barabási (2009) ao registrar que em uma economia informacional em rede a unidade pseudo-monetária é baseada nos links e citações recebidas.

2.2 Estudos Métricos e a Cibermetria

A Cibermetria, segundo Björneborn (2004, p. 14), é “o estudo dos aspectos quantitativos da construção e uso de recursos de informação, estruturas e tecnologias em toda a Internet, com base em abordagens bibliométricas e informétricas”¹⁴. (THELWALL; VAUGHAN; BJÖRNEBORN, 2006) afirmam que

o campo da Cibermetria ultrapassa os limites da bibliometria porque algumas atividades no ciberespaço não são normalmente registradas, mas comunicadas de forma síncrona, como em salas de bate-papo. Os estudos cibernéticos de tais atividades ainda se encaixam no campo genérico da informação como o estudo dos aspectos quantitativos da informação¹⁵ (THELWALL; VAUGHAN; BJÖRNEBORN, 2006)

Enquanto a Almetria é definida por Priem (2015) como o "estudo e uso de medidas de impacto acadêmico com base na atividade em ferramentas e ambientes em linha" (tradução nossa)¹⁶, enquanto o Manifesto Altmetrics refere-se a eles como elementos de traços online de impacto acadêmico (PRIEM; TARABORELLI; GROTH; NEYLON, 2010). Haustein (2016) chama a atenção para a ausência de definição menos abstrata sobre a Almetria. A pesquisadora

¹⁴ “The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet, drawing on bibliometric and informetric approaches.” (BJÖRNEBORN, 2004, p. 14)

¹⁵ “the field of cybermetrics exceeds the boundaries of bibliometrics because some activities in cyberspace are not normally recorded, but communicated synchronously, as in chat rooms. Cybermetric studies of such activities still fit in the generic field of informetrics as the study of the quantitative aspects of information” (THELWALL; VAUGHAN; BJÖRNEBORN, 2006)

¹⁶ the study and use of scholarly impact measures based on activity in online tools and environments (PRIEM, 2015)

também alerta para a confusão que existe entre a Altméria e métricas de nível de artigo (ALM – do inglês *Article Level Metrics*).

Ambos os conceitos dos subcampos apontam para uma forte relação entre ambos ao trabalharem com dados os mesmos dados, como apontam Vanti (2005) e Araújo (2015a) ao destacarem que a Cibermetria utiliza em seus estudos: Comunicações formais e informais, entre quaisquer grupos sociais (científicos ou não) de qualquer forma, registrados em: bases de dados, páginas web, URLs, microblogs, blogs, salas de bate papo, *mailing lists*, comunidades virtuais, grupos de discussão, *muds*, ambientes virtuais de aprendizagem (AVA) e sites de redes sociais. Enquanto que a Altméria tem como foco a Comunicação científica (comunidades científicas, assuntos científicos, artigos, periódicos, pesquisadores, citações) em microblogs, blogs, comunidades virtuais, grupos de discussão e sites de redes sociais.

Dessa forma, é oportuno pontuar a Cibermetria como o subcampo mais adequado para se desenvolver um crossover métrico devido a sua amplitude e por no nosso entendimento abranger os estudos que são realizados pela Altméria. Ao mesmo tempo, pode-se afirmar que a Cibermetria por oferecer como escopo de trabalho um ambiente mais amplo e não somente dados provenientes de ferramentas sociais possibilita a análise multidimensional. Além dos dados que são tratados e/ou vistos como métricas alternativas serem oriundos tantos de páginas web quanto das redes sociais. Ao mesmo tempo, um dos pontos principais do crossover é a identificação do conteúdo que será operacionalizado e os links surgem como único elo comum em ambos os cenários e nos próprios estudos métricos na web foram eles as pontes que indicavam e indicam a importância de um conteúdo.

No capítulo foi traçado um panorama histórico da evolução nos modelos de utilizados para medição de resultados, com foco especial nos ambientes digitais. Ao mesmo tempo, foi introduzido os principais fundamentos da Cibermetria que será a base teórica para o desenvolvimento da pesquisa. Entre os principais pontos tratados destacam-se (I) a evolução dos modelos de medição que pontuamos como ponto inicial o estudo realizado por Gross e Gross (1927) e como ele serviu de base para os estudos métricos, (II) o começo das pesquisas sobre métricas digitais na década de 1990 e (III) a relação dos estudos métricos com a Cibermetria que será detalhado no próximo capítulo.

3 A CIBERMETRIA E OS ESTUDOS MÉTRICOS

No processo de evolução da Ciência da Informação como área do conhecimento, como campo permeável a constantes mudanças, e como campo de pesquisa e ensino é possível observar uma riqueza ímpar de possíveis metodologias e abordagens para o desenvolvimento de diferentes processos, métodos e técnicas de coleta, tratamento e recuperação da informação (SOUZA; STUMPF, 2009). A Cibermetria neste contexto se apresenta como uma importante ferramenta para medir as novas formas de envolvimento com as publicações nos ambientes digitais, por ser utilizada como forma de avaliação dos recursos disponíveis na internet com o intuito de melhorar o fluxo das informações, por meio de análises quantitativas e qualitativas.

Ao se analisar a Ciência da Informação pela Cibermetria, observa-se um campo recente e pouco utilizado. Shiri (1998) a define como o estudo de medição e análise de todos os tipos de informação e meios de comunicação que existem no ciberespaço utilizando-se de técnicas bibliométricas, cientométricas e informétricas. Além disso, ela engloba a webometria e a altmetria, que demonstra ter a capacidade de compreender aspectos comunicacionais emergentes. Gouveia (2016) aponta que o subcampo emergente da altmetria conseguiria capturar todos os vestígios do ciclo de vida da pesquisa expressado por todo tipo de audiência, indo dos cientistas, passando por jornalistas e chegando ao público em geral.

A cibermetria, portanto, recorre às tradicionais técnicas para a medição das informações produzidas no âmbito do ciberespaço. Isso acontece devido as técnicas utilizadas serem compostas de um modelo completo e multifacetado que integram métricas quantitativas e qualitativas cujo objetivo é encontrar potenciais correlações entre os dados e/ou ambientes analisados. Com base nesse pressuposto pode-se entender como método científico cibernético aquele cuja finalidade é quantificar os produtos em ambientes digitais e os atores envolvidos com o propósito de medir os aspectos comunicativos e suas articulações na web.

Assim, a cibermetria utiliza os mesmos métodos que os demais subcampos, porém em um âmbito mais específico que são os ambientes virtuais. O objeto de estudo da cibermetria é a informação produzida ou trocada no universo digital, por meio de aplicações como wikis, blogs, microblogs, RSS, redes sociais, videoblog, gerenciadores de bookmarks, mensagens instantâneas, marcadores sociais tags, entre outras ferramentas e mecanismos que tem o tráfego, armazenamento e compartilhamento de informação como o seu produto principal.

3.1 Bases conceituais dos estudos cibernéticos

Os estudos métricos na web carregam em sua base a utilização de técnicas e modelos

tradicionais, adaptados para o novo ambiente e dessa forma sem gerar rupturas ou grandes impactos na forma como a Academia e pesquisadores observam os resultados apresentados. Um exemplo, desse modelo é a Cibermetria, que apesar de ter em seu escopo o objetivo de estudar o universo digital, ainda se prende nos estudos das citações. Observa-se como importantes tais estudos, mas estes são apenas uma parte para descobrir como está sendo realizada a disseminação e a apropriação de uma publicação, mas também não são o único ponto que poderia e deveria ser estudado.

O mundo digital abriu diversas possibilidades como o estudo de metadados, a possibilidade de refinamento das pesquisas que podem permitir a identificação do tipo de conteúdo que é mais relevante para os pesquisadores, em qual época do ano determinado assunto pode ter mais aderência de público, dentre várias outras possibilidades, incluindo para o autor uma avaliação quase em tempo real de como a sua pesquisa está sendo observada por seus pares.

Neste contexto, Gouveia (2013) afirma que

podemos dizer que hoje os blogs e microblogs, como o Twitter, desempenham um papel de revisão por pares pós-publicação onde os debates sobre os resultados de pesquisa se aquecem e servem inclusive de orientação para a mídia no momento em que uma destas pesquisas se torna pauta. (GOUVEIA, 2013, p. 222)

Nwagwu e Onyancha (2015, p. 8) ampliam o debate e citam que a ciência está tornando-se mais centrada no público devido ao surgimento de novos meios, formais e informais, de disseminar e comunicar a ciência, na maior parte do tempo obscurecendo ou obliterando as fronteiras entre as unidades estabelecidas de comunicação científica. As transformações ainda estão em andamento e apontam para uma ruptura com os sistemas tradicionais de medição, como é possível ser observado no modelo de revisão por pares, que também é impactado pelos comentários dos leitores, por meio de uma comunicação informal no período pré ou pós-impressão.

O atual modelo de revistas, como produto de informação, foi desmembrado e hoje qualquer pessoa com acesso a internet pode acessar partes específicas, trechos ou artigos completos de um estudo ou publicação. A ciência foi pulverizada pelos meios sociais abrindo caminho e possibilidades de interação com fontes e conteúdos, contribuindo e influenciando o seu conhecimento e a sua prática.

No âmbito da Cibermetria, que tem suas bases no emprego de técnicas matemáticas e estatísticas para quantificar sites ou seus componentes e conceitos (JANA; CHATTERJEE, 2004, p. 191), se encontra o mesmo processo de revolução pelo qual passou a Cientometria, com a necessidade de amplificar o seu escopo de estudos de forma que possa, também,

acrescentar os estudos qualitativos de dados e metadados originados de ambientes sociais. Um dos caminhos é utilizar e/ou se apropriar dos métodos, ainda em estudo, da Altmétrie de forma a conseguir agregar a visão do público e consumo das pesquisas.

Essa premissa alinha-se com a visão de Nwagwu e Onyancha (2015, p. 8) que observa a ciência se tornando cada vez mais centrada no público, com novas formas de disseminar e comunicar a ciência. Assim, vêm-se utilizando ferramentas formais, informais e inovadoras que têm contribuído para diluir ou eliminar as fronteiras existentes entre revistas, artigos e ideias, e cada vez mais reconhecer novas e diferentes formas de validar os produtos que resultam dos esforços da investigação.

3.2 Aplicação cibernétrica nos estudos alométricos

A Cibernétrie e Altmétrie são áreas de estudo complementares. Enquanto a primeira utiliza principalmente metadados para a realização de estudos com foco mais técnico, a segunda se apropria de informações disponíveis ou disponibilizadas pelas ferramentas. Porém, estas apresentam baixa profundidade sobre o conteúdo, estando em estado bruto e sendo muitas vezes e erroneamente confundidas e apresentadas como “dados” e apresentadas como dados abertos ou *open data*.

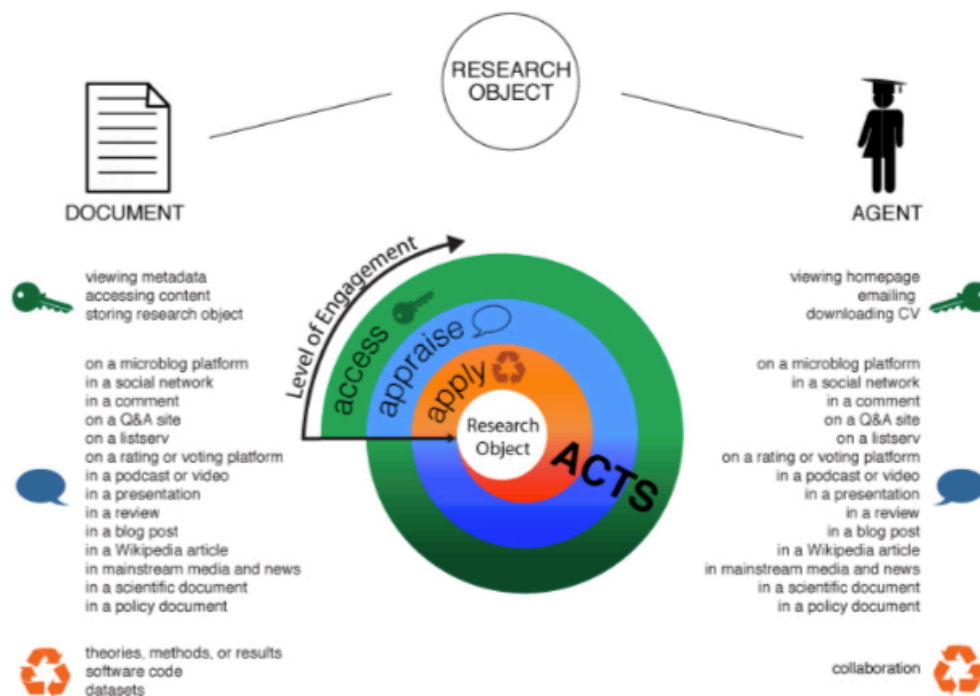
No entanto, é importante fazer a ressalva de que o conjunto de informações trabalhadas pela Altmétrie somente receberão a nomenclatura de “dados” após serem catalogados e para serem classificados como *open data*, necessitam estar disponíveis em estado bruto (sem nenhum tipo de manipulação) para que outros interessados possam utilizar. Atualmente, as mais variadas ferramentas sociais, mesmo as que oferecem APIs para a captura das informações, impõem limitações ao tipo de conteúdo e volume de conteúdo que o usuário ou pesquisador poderá utilizar. Dessa forma, a Altmétrie utiliza de técnicas cibernétricas, para a captura, modelagem e posterior tratamento das informações para que sejam transformadas em um conjunto de dados.

Dessa forma, os estudos alométricos estão baseados em citações provenientes das redes sociais e estão em constante reformulação devido à incorporação de novas fontes de dados. Para Mazov e Gureev (2015), as publicações de artigos em revista é, ainda, o foco principal a ser medido, enquanto livros, anais de congressos, apresentações, material em vídeo, matrizes de dados, códigos de programas, e outros tipos de informação científica estão a ser sucessivamente incluídos na recente área de pesquisa.

Os estudos das métricas alternativas têm demonstrado grande vitalidade, principalmente, com a adoção por várias universidades e periódicos. No entanto, para o pesquisador o trabalho ainda é árduo e cheio de obstáculos. Entre a definição, o que se deseja pesquisar e a captura dos dados, várias etapas podem inviabilizar um trabalho científico.

Ao partir dessa premissa Haustein, Bowman e Costas (2015) postularam um modelo (Figura 2), baseado na Teoria Social cujo objetivo é permitir categorizar e posteriormente quantificar a dispersão e engajamento do público com conteúdos provenientes dos ambientes digitais. No modelo proposto um documento científico pode receber duas classificações principais: Documento Acadêmico ou Agente Acadêmico, ambas se subdividem em três funções “Access”, “Appraise” e “Apply”, que concentram os tipos de atividades exercidas pelo público ao consumir um documento científico.

Figura 2 - Modelo de categorias e tipos de atos referentes a objetos de pesquisa (documentos acadêmicos e agentes)



Fonte: Haustein, Bowman e Costas (2015)

A construção de protótipos que possam incluir as redes sociais em suas medições se tornou a nova barreira dos estudos métricos, e o modelo apresentado por Haustein, Bowman e Costas (2015) demonstra ser apenas uma variação da proposta por Colledge (2014) que define os indicadores em quatro categorias: Atividade Acadêmica, Comentário Acadêmico, Comentários de Publicação e Mídia de Massa. Em ambos os modelos, devido a forma como os autores propõem o envolvimento do público com as publicações somente seria possível a

realização de uma análise se os metadados de acesso fossem públicos. Fato que tornaria a utilização dos modelos propostos restritos às publicações científicas.

Para evitar a barreira do acesso aos metadados, Loach Evans (2015) afirmaram que é possível, a partir de dados altmétricos, obter uma classificação (*ranking*) razoável de uma revista, pois, em sua maioria, estes dados parecem fornecer informação útil por possuírem uma correlação aceitável com o FI. Da mesma forma, aqueles dados são suficientemente diferentes, o que pode indicar diferentes tipos de impacto. Os seus resultados sugerem que diferentes métodos de avaliação, aplicados a um conjunto de dados, produzem relativamente pouca variação e podem, assim, proporcionar uma medida da incerteza de qualquer classificação da revista. Os padrões encontrados precisam de confirmação e isto necessita de grandes intervalos de tempo, a fim de se compreender melhor as razões sociais por trás daqueles padrões. Por esse prisma postulou-se que o aporte dos estudos cibernétricos para a análise de dados altmetricos podem promover e proporcionar uma maior clareza nos dados e inclusive preparar para a utilização dos metadados oferecido por algumas ferramentas, como uma nova camada de estudos altmetricos.

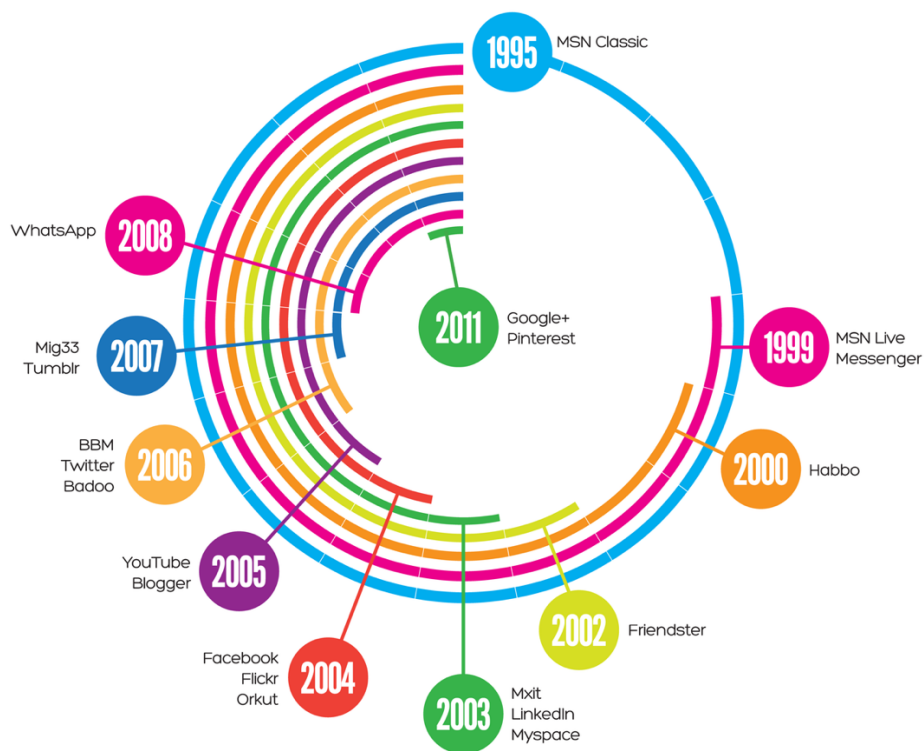
Neste capítulo foi apresentada as bases conceituais da Cibermetria e como pode ser uma importante ferramenta para medir as novas formas de envolvimento com as publicações nos ambientes digitais. Foi apontado como a Cibermetria e a Altmtria são áreas de estudo complementares devido a primeira utilizar metadados para a realização de estudos com foco mais técnico, a segunda se apropria de informações disponíveis ou disponibilizadas pelas ferramentas, mas que ainda necessita de amadurecimento e por esse motivo foi apresentada como um suporte aos estudos cibernétricos. Com a base e entendimento do atual estudo das métricas em ambientes digitais, o próximo capítulo irá apresentar e se aprofundar nas ferramentas sociais e como os conceitos do Marketing podem ser trabalhados em conjunto com os estudos métricos.

4 MÍDIAS SOCIAIS, REDES SOCIAIS, SITES DE REDES SOCIAIS

A consolidação e massificação de ferramentas sociais a partir de 1995 gerou como reflexo o interesse de pesquisadores dos mais variados segmentos para os rastros digitais deixados pelo público ao utilizar o ambiente. No entanto, um dos primeiros entraves para a utilização destes meios foi a diversidade de recursos, arquitetura, conceitos, terminologias, possibilidades, tipos de conexão, conteúdo e a apropriação que cada público/segmento realiza nestes ambientes.

Essas aplicações foram e ainda são importantes, pois oferecem insumos que são um pouco da história contemporânea e auxiliam no entendimento de como o público se relaciona entre si e com os conteúdos disponibilizados nestes ambientes. Ao direcionar um olhar para o mercado e Academia, também, é possível traçar um paralelo com o impacto que essas ferramentas proporcionaram na forma de mensurar o impacto, o alcance, a dispersão, o engajamento, a centralidade e a densidade das relações do público. Além de ser um registro da evolução tecnológica, como pode ser observado na Figura 3 que apresenta um infográfico com a evolução das ferramentas sociais segundo Colic (2012).

Figura 3 - Infográfico com as principais redes sociais entre os anos de 1995 a 2011



Fonte: Colic (2012)

No entanto, a nomenclatura destas ferramentas sociais demonstra carregar uma compreensão indistinta dos termos "Mídia Social", "Redes Sociais" e "Sites de Redes Sociais", que precisam ser distinguidas com maior profundidade. No entanto, não existe uma maneira sistemática de categorizar as diferentes aplicações. Além disso, novas ferramentas sociais aparecem todos os dias, por isso é importante que qualquer esquema de classificação leve em consideração as aplicações que podem estar disponíveis e as que podem surgir futuramente. Segundo Ramalho (2010), as mídias sociais estão distribuídas em diversas categorias de utilização, mas a classificação ou nomenclatura de muitas delas ainda é tema de discussão. Contudo, Kotler e Keller (2011, p. 546), defendem que existem três plataformas principais de mídias sociais: as comunidades e fóruns online, os blogs e as redes sociais. Sterne (2010) e Giglio (2010), dividem os ambientes digitais em seis categorias e alertam que algumas apresentam sobreposições em termos de finalidade e audiências. Elas podem ser divididas em: projetos colaborativos, blogs, comunidades de conteúdo, sites de redes sociais, mundos de jogos virtuais e mundos sociais virtuais.

Erlandson; Erb (2013) destacam que as "mídias sociais se referem aos canais de comunicação eletrônica utilizados para criar e distribuir conteúdo que permite o envolvimento entre criadores de conteúdo e usuários finais"¹⁷ (ERLANDSON; ERB, 2013, p. 85, tradução nossa).

As redes sociais na visão de Wasserman (1994, p. 21) são definidas como um conjunto de dois elementos: atores (pessoas, instituições ou grupos) e suas informações relacionais. Essa definição é seguida por Degenne e Forsé (1999) ao afirmar que um dos pontos que identificam as redes sociais são suas conexões (interações ou laços sociais).

Enquanto que os sites de redes sociais são definidos por Boyd e Ellison (2007) como sistemas que permitem: a construção de uma persona através de um perfil ou página pessoal; a interação através de comentários; e a exposição pública da rede social de cada ator.

Com base nestes apontamentos pode-se afirmar que Mídias Sociais é o ambiente social como um todo e dentro deste ambiente estão inseridas as Redes Sociais. Assim, com base nessas definições, e por identificar que as nomenclaturas (Mídias Sociais, Redes Sociais e Sites de Redes Sociais) acabam por aprofundar a dúvida sobre os ambientes trabalhados devido as sutis diferenças e em muitos casos até se sobrepõem em finalidades, optou-se para melhor compreensão da pesquisa utilizar o termo "Rede Social" para definir toda a interação, identificação do público e conteúdos gerados nessas plataformas. A opção foi influenciada,

¹⁷ "Social media refers to channels of electronic communication used to create and distribute content that allows for interaction between content creator and end users" (ERLANDSON; ERB, 2013, p. 85)

principalmente, por ser possível observar nos estudos científicos sobre esses ambientes, a quase normatização do termo redes sociais, como avalia Oliveira (2013)

[...] as redes sociais estão presentes em todos os níveis e segmentos da sociedade e, na ciência, não é diferente. Elas possibilitam maior interação entre os atores envolvidos no processo – autores, leitores e editores - de maneira rápida, imediata e interativa, apontando para novas práticas de comunicação e informação, ampliando a visibilidade e alcance das pesquisas realizadas e sua disseminação para a comunidade específica e sociedade em geral. (OLIVEIRA, 2013, p. 199)

Com base nesses esclarecimentos e devido o objeto de estudo desta pesquisa ser os dados originados pelo público dentro destes ambientes, que a partir deste ponto a nomenclatura “ferramentas sociais” será substituída por “redes sociais”. Esse aporte sobre a forma como será tratado os ambientes digitais, se faz necessária, devido a ampla discussão que será apresentada nos próximos capítulos e que necessitará do entendimento prévio da opção por utilizar o termo Rede Social, em detrimento de Mídia Social.

4.1 Redes sociais e o marketing como suporte aos estudos métricos

As métricas alternativas são uma realidade no fazer ciência. Periódicos, pesquisadores, instituições começaram a identificar as possibilidades oferecidas pelas redes sociais na visibilidade dos resultados, em grande parte influenciados pela Altmtria, que por meio do seu ranking ou “Índice de Atenção”¹⁸, tem se vendido como uma possibilidade para identificar artigos e autores com maior destaque entre os seus pares. Ferreira e Caregnato (2014) destacam que pesquisas facilmente acessíveis podem conquistar boa posição de visibilidade. Nesse quesito em específico, a visibilidade está associada a exposição, reconhecimento e utilização do material disponibilizado.

Um dos possíveis caminhos para se conseguir essa atenção do público online pode ser com a utilização de estratégias de marketing. Segundo Bizzocchi (2002, p. 4) “o princípio básico a ser adotado pelo marketing científico é o de que o conhecimento, e sobretudo o discurso que o contém, é um produto como outro qualquer” e que tem como objetivo popularizar a ciência por meio da apresentação ou divulgação de pesquisas para um público novo e com pouco conhecimento científico, com a utilização de ferramentas do marketing tradicional. No entanto, para dar esse passo deve ficar claro na estratégia que “o marketing científico não é a aplicação do conhecimento científico ao fazer mercadológico, mas sim o

¹⁸ O “Índice de Atenção” ou “*Attention Score*” é um indicador da ferramenta Altmtric.com que fornece um resumo rápido da atenção online recebida por um artigo. (nota do autor)

inverso disso: é a aplicação do conhecimento mercadológico ao fazer científico” (BIZZOCCHI, 2002, p. 5).

Bomfá, Freitas, Silva e Bornia (2009) complementa ao afirmar que instituições científicas também utilizam o marketing na consolidação da imagem da ciência, para estimular a confiança e com o objetivo de atrair investimentos nas pesquisas. Com base nessa premissa é possível fazer a apropriação dos estudos de Kotler e Keller (2011) ao destacar que o propósito do marketing é centrado em ampliar a abrangência do público atingido, além de atender e satisfazer às necessidades e desejos dos consumidores. Inclusive, não seria errado afirmar a possibilidade em trabalhar uma pesquisa como produto, termo utilizado por Gouveia (2013) ao exemplificar que a *National Science Foundation* (NSF) alterou o termo “publicações” para “produtos” no formulário descritivo dos pesquisadores que submetem os seus projetos para a captação de recursos na hora de buscar formas para disseminar estes conteúdos. O uso do termo também é apoiado por Björk (2007) ao propor um novo modelo de distribuição da comunicação científica, na qual utiliza um processo de produção para descrever como seria a melhor efetividade na distribuição de resultados de pesquisa, apontada como produto, para gerar maior rentabilidade de seus resultados.

Essa quebra de paradigma ao entender o fazer científico não apenas como ciência e sem nenhuma relação com o mercado é esclarecida por Kotler (1998) ao apresentar a importância do marketing em conquistar e fidelizar clientes com o objetivo de garantir a sobrevivência da organização no mercado onde a competitividade se encontra cada vez mais acirrada. Na atual configuração científica o que se busca é a atenção do leitor, que é uma mercadoria cada vez mais escassa e como agravante ainda existe relações e interações entre os indivíduos com interesses comuns, trocando informações sobre o que é interessante e em que vale a pena despender atenção.

Conseqüentemente, Kotler e Keller (2011) definiram micromodelos de comunicação de marketing voltados para as respostas específicas dos consumidores à comunicação, onde existe uma hierarquia de respostas que levam em consideração a curva de aprendizagem ou de adoção por qual passa o consumidor. Essa hierarquia é composta de três estágios: um estágio cognitivo, um afetivo e um estágio comportamental, conforme apresentado no micromodelo do Quadro 1.

Em relação aos objetivos da comunicação de marketing e os elementos fundamentais de uma comunicação efetiva, nos quais se concentram as respostas específicas do consumidor, estão presentes quatro modelos de hierarquia de respostas clássicos (Quadro 1). Nele é possível observar que todos são modelos lineares e sequenciais, baseados no pressuposto que todos os possíveis consumidores passam pelos estágios: cognitivo, relacionado à consciência e ao

conhecimento; afetivo, associado à simpatia, preferência e convicção; e estágio comportamental, relativo à ação efetiva.

Quadro 1 - Modelo de hierarquia de respostas baseada nos estudos de Kotler e Keller

Estágios	Modelo AIDA	Modelo hierarquia de efeitos	Modelo inovação-adoção	Modelo de comunicação
Estágio cognitivo	Atenção ↓	Conscientização ↓ Conhecimento ↓	Conscientização ↓	Exposição ↓ Recepção ↓ Resposta Cognitiva ↓
Estágio afetivo	Interesse ↓ Desejo ↓	Simpatia ↓ Preferência ↓ Convicção ↓	Interesse ↓ Avaliação ↓	Atitude ↓ Intenção ↓
Estágio comportamental	Ação	Compra	Experimentação ↓ Adoção	Comportamento

Fonte: Kotler e Keller (2011)

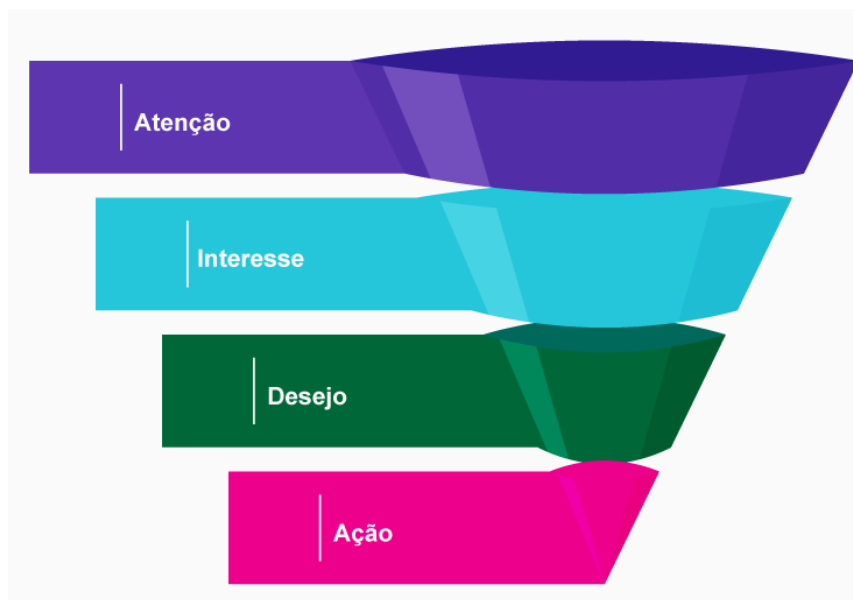
Segundo Wijaya (2012), dentre estes micromodelos, o modelo AIDA¹⁹, ou funil de vendas, foi o que demonstrou a melhor adaptação aos ambientes digitais e demonstrou o interesse de pesquisadores sobre o modelo estratégico de consumo alicerçado em quatro pontos: Atenção, Interesse, Desejo e Ação. Estes pontos demonstraram facilidade na adequação e transposição para o marketing digital e dessa forma passou a ser adotado como uma ferramenta auxiliar no marketing nas redes sociais ao apontar gargalos de comunicação.

Um dos pontos principais do modelo AIDA é a sua representação gráfica em formato de funil, conforme pode ser observado na figura 4. No marketing o funil é uma das principais ferramentas para acompanhar a retenção do público em cada etapa dentro de um funil de vendas, pois toda ação publicitária parte do pressuposto que apenas uma parcela, geralmente inferior a 10%, chegará ao final do funil. Dessa forma, cada camada vai representar o volume

¹⁹ AIDA é um modelo que tenta explicar como funciona o comportamento humano em relação à aquisição de um produto ou serviço. O modelo AIDA, determinou uma série de passos que descrevem o processo que um comprador de um determinado produto passa antes de fechar a venda. (nota do autor)

de pessoas que estão abandonando a intenção de comprar ou utilizar determinado serviço ou produto.

Figura 4 - Representação gráfica do Modelo AIDA proposto por Kotler e Keller



Fonte: Adaptação do modelo proposto por Kotler Keller (2011)

Por esse motivo que o modelo AIDA, atualmente, é um das preferidos para os estudos de conversão nas redes sociais, pois os seus pontos de medição permitem visualizar em qual etapa o consumidor desiste de chegar ao ato de compra. Adaptações do modelo AIDA expandiram o número de etapas (BARRY, 1987) e foram projetadas para acondicionar os desenvolvimentos teóricos, incluindo a satisfação do cliente, representado pelo o modelo AIDAS (BARRY; HOWARD, 1990) e o Modelo AISDALSLove (WIJAYA, 2012) que inseriu as fases de “*Search*” (“busca” após Interesse), a fase em que os consumidores pesquisam ativamente informações sobre marca/produto/serviço; “*Like/Dislike*” (“gostar/não gostar” após Ação) como um dos elementos na fase pós-compra; o “*Share*” quando se compartilha as experiências com outros usuários; e “*Love or Hate*” que serviria para expressar o nível de satisfação do consumidor. Essas novas fases que criaram inseriram novas camadas para os micromodelos²⁰ e auxiliam, principalmente, na mensuração qualitativa de resultados.

Outros modelos de aferição de resultados que migraram do marketing tradicional para o online foram: o modelo de Lindenmann ou Régua de Efetividade (LINDENMANN, 1993), onde é definido como uma régua, com um conjunto de diretrizes ou padrões que o profissional pode seguir se quiser medir a efetividade da comunicação. No modelo, dividido em três níveis

²⁰ Para um aprofundamento sobre os micromodelos consultar: Kotler e Keller (2011) e Vakratsas e Ambler (1999)

(Básico, Intermediário e Avançado), é exemplificado quais métricas podem ser utilizadas em cada fase do relacionamento entre cliente e produto. O modelo Macnamara (MACNAMARA, 1999) segue a mesma tendência de Lindenmann, mas aprofunda os estágios e pontos que podem ser medidos. O modelo de Bartolomeu e Bagnall (JEFFREY, 2013) possui três principais adoções: Pago, Institucional ou Autoral e Espontânea, onde para o pesquisador, uma comunicação eficiente deve ter como premissa qual foi a fonte da publicidade, para depois se definir o que será medido.

Esses modelos ajudam a compreender o tipo de informação que deve ser utilizada para cada estágio, ou seja, até mesmo o marketing científico pode focar sua divulgação em apenas um estágio do processo. Em muitos casos, o único objetivo é criar consciência da empresa; da marca; de um periódico; ou de um artigo. Isto é, se o objetivo for influenciar o estágio cognitivo, a mensagem deve despertar a atenção e falar sobre coisas de interesse do receptor. Se o objetivo for o estágio afetivo, onde são despertados o interesse e o desejo, a mensagem deve destacar vantagens e benefícios. Para influenciar no desejo, é necessário que a mensagem justifique e detalhe as razões para a “compra” e, se possível quantifique e explicita os seus benefícios. Para influenciar o estágio comportamental (geralmente objetivo de marcas conhecidas e consolidadas) a mensagem deve trazer vozes de comando, ênfase na oportunidade e, ainda, fatos que possam trazer credibilidade à marca. Com essas informações é possível traçar a melhor forma de se comunicar.

Mesmo com vários modelos que podem auxiliar a construção de uma matriz para aferir os resultados da dispersão de conteúdo na redes sociais, Shimp (2002), avalia que esses modelos de hierarquia são baseados em levar o público de uma meta para outra, de forma linear ou até alcançar o topo da escada ao levar o estado inicial de inconsciência de uma marca até o seu estágio final que é a aquisição do produto. Mas, os ambientes digitais, com todas as suas particularidades, têm suas características e principalmente possibilidades de etapas serem suprimidas no processo de aquisição de um produto. E no âmbito do científico, Araújo salienta que “A escolha das estratégias de marketing científico digital precisa refletir o tipo de imagem que se quer transmitir, bem como atender ao perfil dos clientes/usuários que possui ou que se pretende alcançar.” (ARAÚJO, 2015b, p. 72).

4.2 A extração de dados nas redes sociais

Os trabalhos sobre organização e funcionamento de redes sociais ainda são pouco investidos dos contributos teóricos e metodológicos do conceito de redes sociais e de seus

contextos epistemológicos de formulação e de estudo (MARTELETO, 2010). Por esse motivo os estudos relacionados a Ciência da Informação demonstram potencial para várias análises, sendo inclusive, com novos cenários a partir de abordagens cibernéticas, sendo possível partir de uma análise que vai desde as conexões até as informações que as mantêm.

Embora as existências de várias nuances possam parecer como limitações para a pesquisas nas redes sociais, o fator comum é que todas essas nomenclaturas têm como foco a apropriação que cada pessoa faz destas ferramentas sociais. Boyd e Ellison (2007) observam que a atenção não deve ser na busca de como se formam as conexões, mas no modo como permitem a visibilidade e a articulação. As autoras utilizam o termo apropriação como um sistema subjetivo, mas real, que é utilizado pelo público para manter e dar sentido as redes sociais e que é uma consequência direta do uso. As redes quando apropriadas podem acabar por subverter a sua função primordial, como é o caso dos blogs e do Twitter. Schmidt (2007) cita como exemplo os blogs que não seriam redes sociais, mas a forma como foram apropriados com links e sistemas de comentários, além da personalização dos perfis acabou por permitir essa classificação.

Pensamento que é complementado por Dourish (2003) ao destacar,

Apropriação é o modo pelo qual as tecnologias são adotadas, adaptadas e incorporadas à prática de trabalho. Isso pode envolver personalização, no sentido tradicional (isto é, a reconfiguração explícita da tecnologia de modo a corresponder a necessidades locais), mas também pode simplesmente envolver fazer uso da tecnologia para fins além daqueles para os quais foi originalmente projetada, ou para servir a novos fins. (DOURISH, 2003, p. 2, tradução nossa)²¹.

Outro elemento importante dentro da apropriação é a percepção que os diversos sites de redes sociais necessariamente não representam redes independentes entre si. Com frequência, um mesmo ator social pode utilizar diversas redes com diferentes objetivos. Da mesma forma, deve ser levado em consideração que diferentes ferramentas poderiam servir a diferentes propósitos para um mesmo usuário.

Dessa forma, com os vários recursos tecnológicos e práticas de uso, as redes sociais acabam por ser apoio de uma ampla gama de interesses pessoais e culturas que emergem em torno destes ambientes de forma variada. A maioria suporta a manutenção de redes sociais pré-existentes, mas outras ajudam os estranhos a se conectar com base em interesses compartilhados, opiniões políticas ou atividades. Algumas ferramentas atendem a um público diversificado, enquanto outros atraem pessoas com base em linguagem comum ou identidades

²¹ “Appropriation is the way in which technologies are adopted, adapted and incorporated into working practice. This might involve customisation in the traditional sense (that is, the explicit reconfiguration of the technology in order to suit local needs), but it might also simply involve making use of the technology for purposes beyond those for which it was originally designed, or to serve new ends.” (DOURISH, 2003, p. 2)

compartilhadas por raça, sexo, religião ou nacionalidade. Os sites também variam na medida em que incorporam novas ferramentas de informação e comunicação, como conectividade móvel, blogs e compartilhamento de fotos/vídeos. Um fator importante que também deve-se levar em consideração é que as redes sociais são modificadas em relação ao tempo e apropriação. Assim, uma das falhas de abordagem é não observar estes ambientes como um elemento em constante mutação, pois as redes sociais são dinâmicas e estão sempre em transformação e são fortemente influenciadas pelas interações.

A partir desse cenário observou-se que o modelo adequado para taxonomia própria de exploração científica das redes sociais, deve estar aliada com as métricas que podem ser utilizada por cada grupo de ferramentas. Mas, também, não é possível ignorar que em muitos casos, devido as suas funcionalidades e apropriações realizadas pelo público, algumas atividades podem se sobrepôr. Exatamente por esse motivo, foi levado em consideração a utilização de métricas consolidadas no mercado e referendadas por associações de classe, como a Declaração de Barcelona para os Princípios de Medição, publicada em 2010, pela Associação Internacional de Mensuração e Avaliação da Comunicação (AMEC, na sigla em inglês)²², que compilou um conjunto de sete diretrizes para medir a eficácia das campanhas; e o Conclave (2013)²³, uma coalizão de empresas *Business to Business* (B2B) e *Business to Consumer* (B2C), agências de relações públicas e de mídia social e associações industriais, que publicaram os padrões de medição de mídia, em junho de 2013.

4.3 Limitação e mensuração dos dados nas redes sociais

Os estudos das métricas alternativas têm demonstrado grande vitalidade, principalmente, com a adoção por várias universidades e periódicos. No entanto, para o pesquisador o trabalho ainda é árduo e cheio de obstáculos. Entre a definição, o que se deseja pesquisar e a captura dos dados, várias etapas podem inviabilizar um trabalho científico. Haustein (2016) observa que um dos problemas para as pesquisas com dados oriundos das redes sociais repousa na qualidade das informações disponibilizadas nas redes sociais.

Na altmetria, a qualidade dos dados é um grande desafio e transcende os erros e vieses conhecidos para dados de citação. No contexto das citações, os erros representam discrepâncias entre o ato e o evento registrado. No entanto, eles podem ser descobertos e medidos por meio de triangulação de diferentes agregadores de dados (isto é, índices de citação) ou por referência à fonte original (ou seja, a lista de referência em uma publicação). Embora as fontes bibliométricas sejam documentos estáticos, a maioria das fontes de dados no contexto da Altmetria é dinâmica e pode ser alterada ou excluída totalmente. Desta forma, precisão, consistência e replicabilidade podem ser

²² Disponível em: <https://amecorg.com/2012/06/barcelona-declaration-of-measurement-principles/>

²³ Disponível em: <http://paineublishing.com/wp-content/uploads/2013/10/Complete-standards-document.pdf>

identificadas como os principais problemas da qualidade dos dados altmétricos. (HAUSTEIN, 2016, p. 5, tradução nossa)²⁴

Entretanto, Maricato e Martins (2017) avaliam que os indicadores altmétricos possuem um grande potencial, mas ainda necessitam de uma maior reflexão sobre a sua utilização e sobre os seus fundamentos, pois nos estudos realizados sobre o subcampo “pouco conhecimento teórico é extraído dos seus resultados” (MARICATO; MARTINS, 2017, p. 50).

Ao olhar para o ponto prático da pesquisa com dados de redes sociais, alguns dos problemas apresentados para quem deseja realizar um trabalho com as redes sociais está na variável Tempo. A maioria das redes tem uma limitação na recuperação dos dados. Ferramentas como o Twitter²⁵ limitam a busca por conteúdos publicados em no máximo sete dias a partir da data de publicação. Para pesquisas que exigem um maior volume de dados existe a possibilidade de assinar planos para recuperação de dados postados a partir do ano de 2016. Gerenciadores de referências como o Mendeley permitem a captura de informações sobre o que está sendo arquivado, mas não oferece suporte para a data de adição do conteúdo na ferramenta. O mesmo problema ocorre na Wikipedia²⁶, para os Artigos Governamentais e alguns tipos de blogs.

Esse gargalo nas redes sociais acaba por influenciar diretamente qualquer tipo de trabalho científico, pois o pesquisador necessariamente vai precisar antever o que deseja estudar, quais redes oferecem suporte para a extração de dados, qual o limite permitido, como será feito o armazenamento dos dados sem influenciar na integridade e qual o período de captura. Esse último é importante, pois algumas ferramentas demoram até 48h para entregar todos os dados de uma busca. Dessa forma se o conjunto de dados (*dataset*) for de uma semana, será necessário configurar a captura para nove dias.

Em paralelo aos problemas da recuperação de dados nas redes sociais é possível observar estudos sobre como analisar, modelar e catalogar os dados extraídos dos ambientes digitais, mas nenhum *framework* de como se estruturar uma pesquisa com base nos parâmetros das redes. Os poucos existentes são baseados em adaptações rudimentares do mercado e são utilizados com limitado critério na definição de uma pesquisa, como a própria ferramenta

²⁴ “In altmetrics, data quality is a major challenge and transcends the known errors and biases for citation data. In the context of citations, errors mostly represent discrepancies between the act and the recorded event. These can be discovered and measured either through triangulation of different data aggregators (i.e., citation indexes) or by referring to the original source (i.e., the reference list in a publication). While bibliometrics sources are static documents, most data sources in the context of altmetrics are dynamic, which can be altered or deleted entirely. Accuracy, consistency and replicability can be identified as the main issues of altmetrics data quality.” (HAUSTEIN, 2016, p. 417)

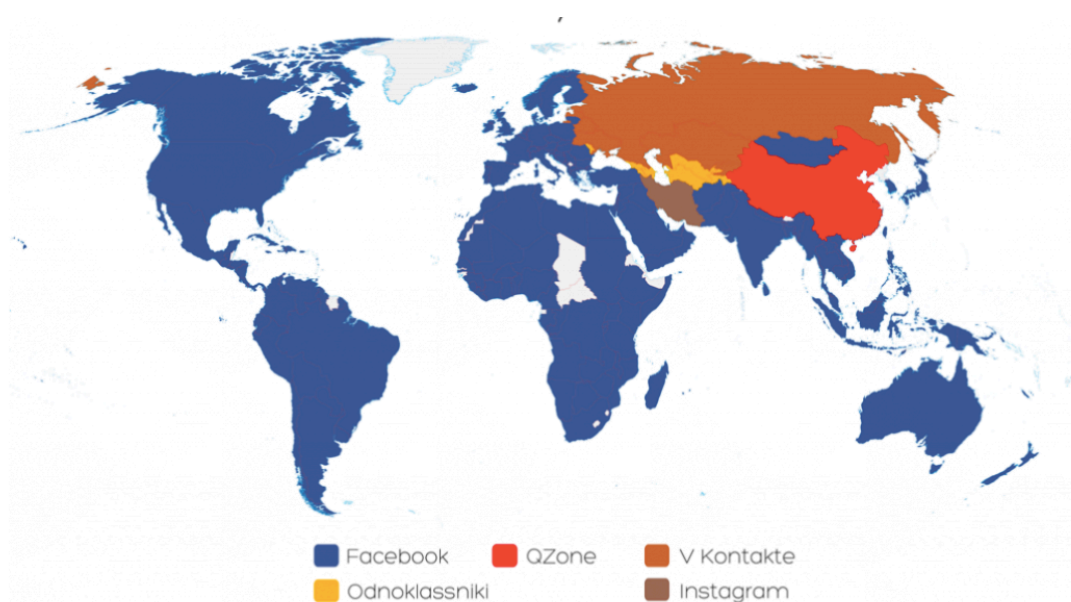
²⁵ Os parâmetros para utilização da API do Twitter podem ser visualizados no endereço: <https://developer.twitter.com/en/docs/tweets/search/overview>

²⁶ A API da Wikipedia aceita em seus parâmetros de buscas a identificação da data de criação e alteração dos verbetes. A documentação completa pode ser visualizada no endereço: <https://www.mediawiki.org/wiki/API:Query>

Altmetric.com que utiliza o padrão de atualização constante dos conteúdos capturados das redes sociais o que impossibilita pesquisas com recortes temporais, pois os dados são um reflexo do momento da consulta.

Além dos problemas com a estrutura dos dados, um outro que muitas vezes é relegado a segundo plano, principalmente, quando se trata de conteúdos que podem ser consumidos por pessoas de várias partes do mundo é a falta de unanimidade nas ferramentas sociais, ou seja, redes que podem ser sucesso no Brasil, podem não ter a mesma penetração na China ou Rússia. Além da realidade de cada país adota e utiliza as redes sócias de maneiras particulares, como pode ser observado na Figura 5 que apresenta a principal rede utilizada por país em janeiro de 2018.

Figura 5 - Mapa Mundi das Redes Sociais em janeiro de 2018



Fonte: Consenza (2018)

A diferença de predominância das redes no mundo levanta a questão sobre que qual a amplitude dos estudos métricos, baseados em métricas alternativas, principalmente pelos atuais sistemas de indexação de postagens, com forte apelo comercial, mas com um distanciamento das necessidades do universo acadêmico, pois incorporam em seus resultados uma visão excludente de ferramentas sociais que não são predominantes em países como Rússia e China, onde predominam o VKontakte e o Qzone, respectivamente. Ao mesmo, Haustein (2016) afirma ao analisar a qualidade dos dados disponibilizados pelas ferramentas sociais que “muitos fornecedores (por exemplo, Twitter, Facebook, Reddit) não são voltados para a academia, e a

qualidade dos dados alométricos não é sua prioridade”²⁷ (HAUSTEIN, 2016, p. 417, tradução nossa).

Essa realidade foi favorecida tanto pela profusão de dados que expandiu as possibilidades técnico-comunicacionais para armazenar e decodificar, como pelas explorações comerciais de *datasets*, de coletas por APIs, por *data scraping*²⁸ e por softwares de monitoramento que facilitaram a lógica da quantificação previamente notadas apenas por pesquisadores acadêmicos ou empresas com foco neste segmento. Nesse novo momento, as taxonomias para identificação e quantificação automática de alguns elementos nas mensagens coletadas seduz acadêmicos e mercado a mensurar tudo o que pode ser contado.

No entanto, mesmo com a variedade de indicadores alométricos é possível observar uma certa dificuldade na compilação de dados acerca de cada publicação e no seu posterior ranqueamento. Um dos motivos para essas dificuldades passam principalmente pela falta de um padrão.

Enquanto, essa normalização ou normatização não surge ferramentas que oferecem monitoramento com dados oriundos de redes sociais e outros ambientes como: *Altmetric*²⁹, *Plum Analytics*³⁰ e *Impact Story*³¹, se apresentam como solução para responder qual o impacto de um artigo ou pesquisa que ganha atenção no mundo digital. O funcionamento de todas é basicamente baseado no agrupamento dos dados compartilhados nas ferramentas sociais, desde que a publicação tenha incluído o número de identificação único do(s) artigo(s) como o DOI (*Digital Object Identifier*), o PMID ou o arXiv ID. Como retorno é oferecido aos usuários relatórios com as estatísticas de conjuntos de trabalhos. Porém, todos os resultados apresentados ainda são quantitativos, e essas ferramentas ainda não conseguiram incorporar em seus algoritmos a análise qualitativa das postagens, resolver problemas de *overlap* de perfis, dos *boots* de redes como o Twitter, do efeito "câmara de eco" ou das bolhas informacionais, e do *Dark Social*³². Dessa forma, os dados apresentados podem e devem ser considerados como uma visão superficial do que está sendo gerado a partir das pesquisas nas redes sociais.

²⁷ “That many providers (e.g., Twitter, Facebook, Reddit) are not targeted at academia and altmetrics data quality is thus not their priority.” (HAUSTEIN, 2016, p. 417)

²⁸ Data Scraping ou Web Scraping é o processo de extração de dados de um site e frequentemente realizada em serviços ou aplicativos web que não disponibilizam APIs para a captura de dados

²⁹ Disponível em www.altmetric.com

³⁰ Disponível em <http://plumanalytics.com/>

³¹ Disponível em <https://impactstory.org/>

³² O termo foi cunhado em 2012 por Alexis C. Madrigal, editor sênior da The Atlantic para se referir ao compartilhamento de conteúdo que ocorre fora do que pode ser medido por programas de análise da web. (Ver também MADRIGAL, 2012).

As demais ferramentas seguem o mesmo padrão e ao buscar mensurar as atividades ignoram as peculiaridades de cada ambiente e tentam sistematizar parâmetros, na maioria dos casos de forma questionável, sobre os dados oriundos dos ambientes sociais. Além disso, não apontam qual o grau de confiabilidade dos resultados apresentados. Em contrapartida o que se consegue observar é que esses vários modelos propostos pelas ferramentas, ainda são atrelados aos formatos tradicionais no qual se levava em consideração apenas a contagem de citações, e a contagem de links que são direcionados para o artigo. Apesar de ter uma lógica conceitual que permite entregar dentro de suas limitações um valor que quantifica a dispersão da produção, ainda assim, demonstra ser feita de forma amadora ou sem um respaldo de estudos que possam balizar a aplicação prática destes indicadores em estudos que envolvam dados oriundos de redes sociais.

Essa adaptação de indicadores que em muitas vezes demonstram limitar o conhecimento da produção científica entram em confronto direto com as métricas de mercado ou do marketing/comunicação que buscam ampliar o horizonte de saber apenas quantas pessoas foram impactadas por uma peça. Algumas bases indexadoras ou “publishers” como a SciELO³³ e a PLOS³⁴ incluíram em suas páginas dos artigos painéis de visualização que demonstram como está o envolvimento do público com a publicação, mas ainda os dados informados são restritos, ou limitados como o número de visualizações e downloads, taxa de permanência, distribuição de países com maior interação. Essa lógica similar as próprias redes sociais que oferecem dados, mas não os metadados, limitam o enriquecimento do debate, da pesquisa e do trabalho acadêmico com as métricas alternativas. Assim, apesar de existir uma variedade de ferramentas, ainda é necessário para o pesquisador ou biblioteca ter o seu próprio arsenal científico para conseguir identificar e avaliar os dados gerados nos ambientes sociais.

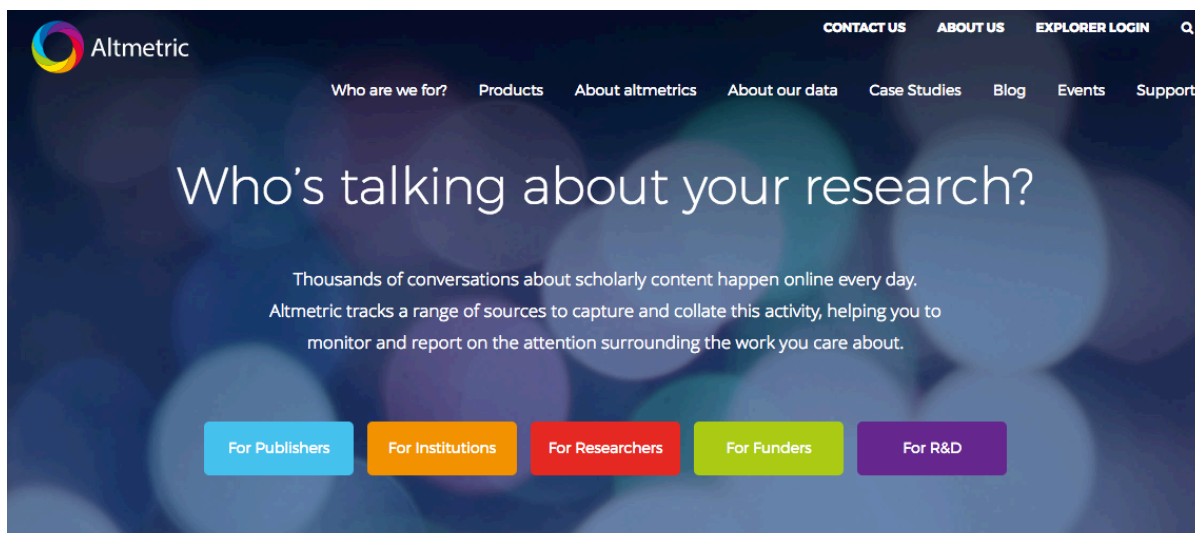
Uma das possíveis abordagens para a melhoria da qualidade dos dados seria torná-los abertos, através de uma API normalizada que permitisse auditorias para validar a qualidade dos dados e evitar eventuais manipulações ou até mesmo normatizar o marketing científico sobre que comportamentos podem ser considerados legítimos para a promoção de uma pesquisa científica. Entretanto, mesmo com esse cenário, nos últimos anos começaram a surgir agregadores de conteúdo ou ferramentas proprietárias com foco no segmento científico que oferecem a captura e armazenamento de dados. O principal diferencial das ferramentas existentes no mercado é o foco em manter um histórico de conteúdos gerados com foco em conteúdos científicos. Atualmente, as mais importantes são:

³³ Disponível em: <http://analytics.scielo.org>

³⁴ Disponível em: <http://journals.plos.org>

Altmetric – Fundada por Euan Adie em 2011, e com base em Londres, pretende tornar as altmetrias fáceis. Assume como missão, mapear e analisar a atividade, que ocorre em torno da literatura e produção acadêmica. Em 2012 lançou sua primeira versão autónoma (PRADHAN; DORA, 2015, p. 128).

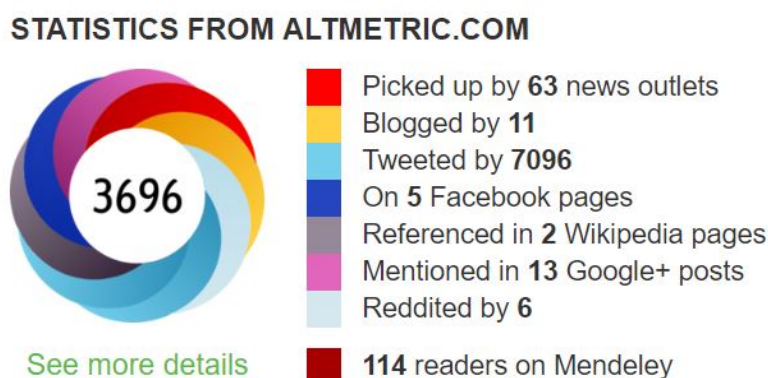
Figura 6 - Tela de entrada da ferramenta Altmetric



Fonte: <http://www.altmetric.com>

A Altmetric (Figura 6) possui cobertura de sites e redes sociais que atuam como fontes para o agrupamento dos conteúdos: Sites de notícias, blogs, Twitter, Facebook, Sina Weibo, YouTube, Reddit, Pinterest, LinkedIn, Open Syllabus, Google+, Documentos públicos, Wikipedia, Q&A, F1000, Mendeley e CiteULike. Os dados são classificados e pontuados com base no volume das menções que receberam. O impacto quantitativo é calculado por meio de um algoritmo, que irá avaliar o envolvimento do público com a publicação. Dessa forma, quanto maior for as republicações maior será a pontuação altmetrica, ou seja, o volume de menções tem uma forte influência na classificação e pontuação. Com o objetivo de promover uma melhor visualização dessa pontuação, foi desenvolvido um gráfico também chamado de “*donut*” ou “*badge*” (Figura 7) que mostra os ambientes com maior reverberação do conteúdo. As cores do gráfico representam a quantidade o volume de citações recebida em cada fonte (redes sociais) e busca representar o Índice de Atenção ponderada recebido pelo documento.

Figura 7 - Reverberação de uma publicação nas redes sociais segundo o Altmetric.com



Fonte: Altmetric

Com a utilização da ferramenta é possível ter uma percepção das menções recebidas em fontes consideradas não convencionais e que podem complementar aquelas baseadas em citações. No entanto, ainda não existem estudos que consigam comprovar que um impacto alto na ferramenta se traduzirá no volume de citações em outras publicações. A Altmetric é uma publicação comercial e disponibiliza pacotes de acesso, mas também oferece uma API aberta com a possibilidade de ser baixado os dados, com algumas limitações, sobre as publicações e a sua reverberação nos ambientes digitais (sites, redes sociais e agregadores de referências).

Porém os resultados apresentados pela Altmetric ponderam, classificam estes conteúdos e apresentam um ranking a partir de um “Índice de Atenção” recebido pelos artigos, utilizando como base uma pontuação que pode ser considerada arbitrária para as redes sociais (Quadro 2), pois não é explicitado a base que permitiu gerar os pesos para cada rede, como foi identificado a aderência, a capilaridade e possibilidade de disseminação em cada ambiente. No geral a pontuação Altmetric ou Índice de Atenção é uma medida que busca refletir a atenção que um artigo ou conjunto de dados recebeu nos ambientes digitais e com base nessa constatação tenta inferir: a visibilidade – quanto maior for o nível de envolvimento maior a pontuação; e a qualidade da atenção – segundo critérios da ferramenta algumas redes geram melhor pontuação para o autor e artigo.

Quadro 2 - Fatores atribuídos as redes sociais pela ferramenta Altmetric

Notícias	Blogs	Twitter	Facebook	Sina	Wikipedia	Q&A	F1000	YouTube
8	5	1	0.25	1	3	0.25	1	0.25

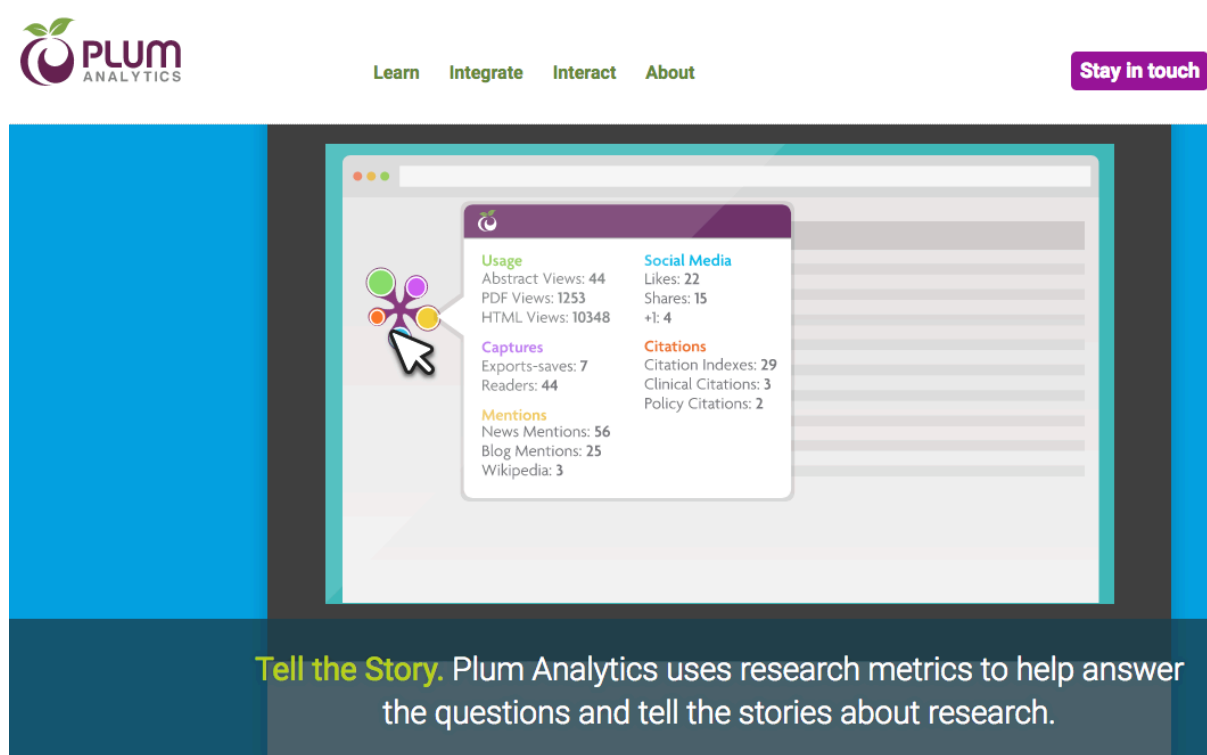
Fonte: Altmetric

Esse tipo de abordagem apresenta-se contrário a normatizações como ao *San Francisco Declaration on Research Assessment* ou simplesmente *Declaration on Research Assessment -*

DORA ((2010)), que entre as suas indicações para os diversos atores que trabalham e utilizam métricas de artigo e de periódicos aponta a necessidade de explicitação de todos os métodos utilizados para os cálculos de métricas e fornecimento de dados de forma aberta e transparente.

Plum Analytics – Foi fundada em 2011 por Andrea Michalek e Mike Buschman, e foi adquirida pela EBSCO Company em janeiro de 2014 (Figura 8). O seu produto PlumX é a ferramenta que apresenta os resultados da reverberação de pesquisas em ambientes digitais. A apresentação dos dados é por meio de cinco categorias de métricas: uso, capturas, menções, redes sociais e citações.

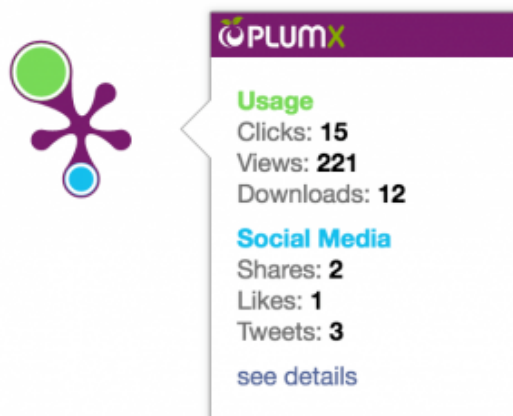
Figura 8 - Tela de entrada da ferramenta Plum Analytics



Fonte: <https://plumanalytics.com>

Os dados utilizados para apresentar a atenção que documentos científicos estão recebendo em ambientes digitais são originados de vários ambientes como: blogs, redes sociais, gerenciadores de referências, encurtadores de URL, gerenciadores de conteúdo, ferramentas de Q&A e plataformas de código fonte (Figura 9).

Figura 9 - Reverberação de uma publicação nas redes sociais segundo o Plum Analytics



Fonte: <https://plumanalytics.com>

A ferramenta, também, é de uso comercial e tem foco em atender universidades e instituições de pesquisa. Em 2013, anunciou a abertura da API, porém os dados são limitados de forma que o usuário tem apenas uma versão parcial da reverberação do conteúdo. Igualmente a ferramenta Altmetric, no Plum Analytics, é apresentado um gráfico que busca representar a atenção online recebida por documentos científicos.

Impactstory³⁵ – Ferramenta com foco em divulgar a reverberação de publicações para os pesquisadores (Figura 10). Na aplicação o usuário pode criar um perfil individual e monitorar a repercussão de sua produção acadêmica, como artigos, apresentações, pôsteres, vídeos, conjuntos de dados, programas de computador, entre outros.

Figura 10 - Tela de entrada da ferramenta Impactstory



Log in

Discover the online impact of your research.

Track buzz on Twitter, blogs, news outlets and more: we're like Google Scholar for your research's online reach. Making a profile takes just seconds:

 JOIN FOR FREE WITH TWITTER

 See an example profile

Fonte: <http://www.impactstory.com>

³⁵ Disponível em: <http://www.impactstory.org/>

A ferramenta foi lançada em código aberto, em 2011, por Jason Priem e Heather Piwowar, inicialmente com o nome de Total Impact. A partir de 2014, iniciou a cobrança de uma assinatura anual para os usuários, mas em 2016 voltou a ser oferecida gratuitamente, após ser remodelada e passar a ser um agregador de serviços de outras plataformas acadêmicas (Altmetric, BASE, Mendeley, CrossRef, ORCID e Twitter). Os dados de perfil e da produção acadêmica do usuário são sincronizados a partir do sistema de identificação de autores ORCID, de forma que cada vez que um produto de pesquisa é incluído no perfil de um autor nesse serviço, o ImpactStory automaticamente reconhece o novo item e passa a coletar as menções recebidas, mas o autor também pode auxiliar o sistema e incluir links de suas produções científicas (Figura 11).

Figura 11 - Tela que apresenta a reverberação da produção científica

The screenshot shows the ImpactStory profile for Ethan White, an Associate Professor at the University of Florida. The profile is divided into several sections:

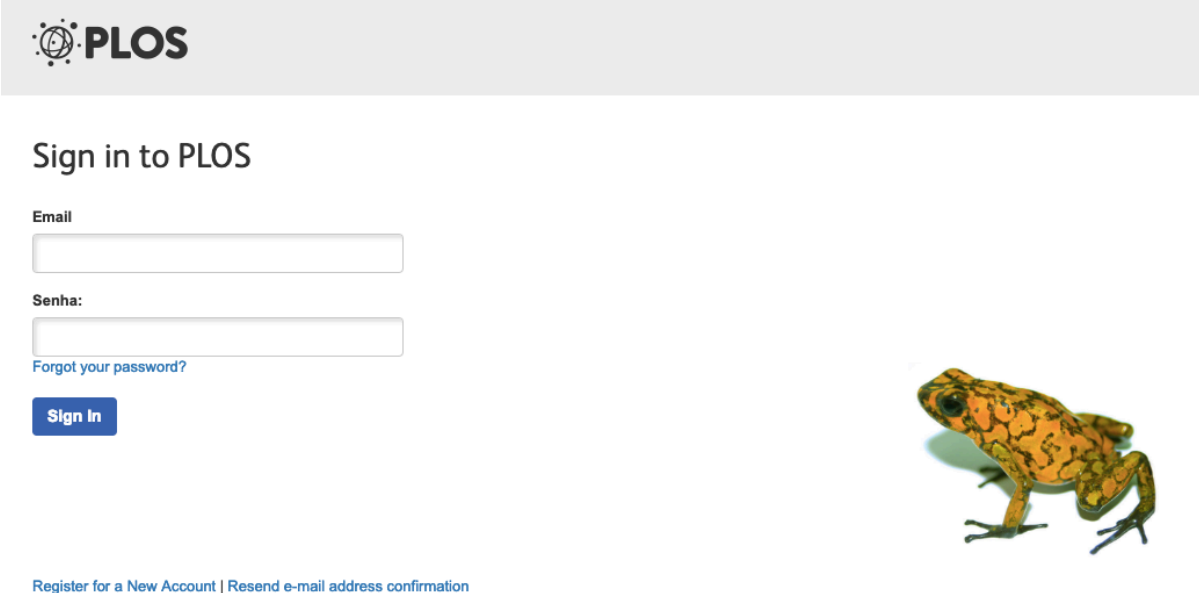
- OVERVIEW**: The main navigation tab.
- ACHIEVEMENTS**: A section with three items:
 - Wikitastic 90**: Your research is mentioned in 5 Wikipedia articles! Only 9% of researchers are this highly cited in Wikipedia. Your Wikipedia titles include *Holocene extinction*, *Quaternary extinction event*, *Ingelfinger rule* and 2 more.
 - Software Reuse 99**: Your research software keeps on giving. Your software impact is in the top 60 percent of all research software creators on Depsy.
 - Greatest Hit 99**: Your most discussed publication has been mentioned online 1.3 thousand times. Only 1% of researchers get this much attention on a publication. Your greatest hit online is *Best Practices for Scientific Computing*.
- MENTIONS**: Shows 2528 online mentions across 11 channels, with a breakdown: 2.3k on Twitter, 83 on Facebook, 61 on LinkedIn, 25 on ResearchGate, 5 on Mendeley, 5 on ORCID, 4 on Scopus, 3 on Crossref, and 3 on other channels.
- PUBLICATIONS**: Lists three publications:
 - Best Practices for Scientific Computing** (2014 PLoS Biology): 1328 mentions.
 - The Case for Open Preprints in Biology** (2013): 520 mentions.
 - The EcoData Retriever: Improving Access to Existing Ecological Data**: 84 mentions.

Fonte: ImpactStory

As menções encontradas são agregadas e apresentadas no perfil do pesquisador em conjunto com a lista de suas publicações e principais realizações acadêmicas.

Public Library of Science (PLOS)³⁶ foi uma das pioneiras no segmento das métricas alternativas e iniciou a divulgação de dados oriundos dos meios digitais em 2009. As “PLOS Article Level Metrics” ou métricas em nível de artigo PLoS (PLOS ALM) se apropriam de uma variedade de indicadores para oferecer volume de citações, visualizações, downloads e compartilhamentos em ambientes sociais.

Figura 12 - Tela de entrada para o sistema da PLOSALM



Fonte: <http://alm.plos.org/>

O diferencial da ferramenta e a disponibilização de métricas de envolvimento do público com os documentos científicos a partir de sua publicação, além de uma API para extração dos dados. No entanto, esta é limitada a extração de informações de apenas 10 documentos científicos por usuário registrado no sistema.

Com o interesse pela altmetria outros serviços estão surgindo e buscando o diferencial para concorrer neste mercado que demonstra alta possibilidade de consumo. Porém muitos ainda estão em busca do modelo ideal e estão limitadas a dados de ferramentas ocidentais e ignorando as particularidades de outros países.

No capítulo foram explorados os conceitos de Mídias Sociais, Redes Sociais e Sites de Redes Sociais com o objetivo de identificar e pontuar onde a pesquisa estaria enquadrada, além de apontar os motivos que levaram a adoção do termo "Redes Sociais" em todo o trabalho. Essa

³⁶ Disponível em: <http://alm.plos.org/>

identificação de nomenclatura serviu para um melhor entendimento da estrutura de cada ambiente. Em seguida, foi apresentado como o Marketing se alinha como uma ferramenta eficaz na mensuração do envolvimento do público com documentos científicos nos ambientes digitais.

A utilização do Marketing foi realizada por meio de conceitos de micromodelos de comunicação postulados por Kotler e Keller (2011) e dentro deste universo foi destacado o Modelo A.I.D.A. ou funil de vendas. O modelo destaca como é realizada o envolvimento do público e foi adaptada para o meio científico.

Na discussão também foi abordado a extração de dados das redes sociais, as suas limitações na mensuração de resultados e as principais ferramentas acadêmicas que servem de suporte para pesquisas científicas.

5 FUNDAMENTOS PARA UM *CROSSOVER* MÉTRICO

As pesquisas sobre o *crossover* têm início com os estudos culturais ao analisar a transformação de símbolos, principalmente, de grupos étnicos para produtos de consumo do *mainstream*. Nestes estudos foi constatado que diferente da tradução literal o termo *crossover* ultrapassa a sua tradução literal de “cruzamento” e é ampliado para fusão, penetração, introdução, transposição, assimilação e adaptação de símbolos restritos a um nicho ou gueto para novo mercado e/ou público. Mas, como observaram Grier, Brumbaugh e Thornton (2006), o *crossover* também pode ser utilizado para criar uma nova significação do mesmo produto e ser novamente reinserido no mesmo grupo de origem. A partir destas descobertas o conceito passou a ser adotado pelo marketing, no começo chamado de “*crossing over*”, para depois sofrer a contração e passar a ser designado apenas como “*Crossover*”³⁷.

No marketing, o conceito conquistou várias subdivisões como: *crossover* de produtos, *crossover* de serviço, *crossover* técnico, *crossover* de design, *crossover* regional, *crossover* emocional, *crossover* de marca, *crossover* de vendas, *crossover* de canal e *crossover* de comunicação baseado em campo de cooperação diversificado. Porém a definição mais utilizada é o *crossover* de produto que pode ser cruzado para gerar algo novo ou que possa atingir um mercado maior.

Embora exista um entendimento sobre os parâmetros básicos para a realização de um *crossover*, não existe uma taxonomia ou categorias sobre as diretrizes que devem nortear a construção desse novo produto. Um dos estudos que mais se aproximaram desse objetivo foi de Yang (2009) ao estudar os modelos de abordagem para a realização de um *crossover* e dividiu em três categorias:

- ✓ **Estratégia de produtos ou horizontal** – significa que a aliança ocorre entre diferentes categorias de produtos, mas que sejam complementares, com o objetivo de gerar algo novo. Dessa forma, é criada uma vantagem competitiva perante outros produtos ou marcas similares e cujo foco é atender uma necessidade pré-existente do mesmo público ou para um novo público;
- ✓ **Estratégia de canal de circulação de mercadorias ou longitudinal** – significa a cooperação baseada no compartilhamento de canal de circulação de produtos entre parceiros que demonstrem o mesmo grau de relevância com o objetivo de criar novas demandas ou abrir novos mercados. Esse tipo de técnica é utilizada principalmente no segmento corporativo;

³⁷ *Crossover* é um termo recorrente nessa pesquisa e daqui em diante o termo não será usado sem itálico

- ✓ **Estratégia de marketing ou interceptada** – significa a combinação de diferentes produtos ou marcas e que vai possibilitar novas experiências de consumo para o público. A estratégia tem como foco principal utilizar marcas ou produtos líderes em seus segmentos para oferecer uma nova experiência de consumo e o compartilhamento de valores e interesses comuns.

Por essa aparente variedade de áreas que podem ser trabalhadas pelo crossover, este é muitas vezes confundido com o co-branding³⁸. No entanto, a única semelhança é que seus participantes são múltiplas marcas, mas a estratégia de crossover atende mais a cooperação em diferentes campos que podem ser indústrias ou grupos de consumidores alvo e cujo objetivo final será a expansão para outras áreas. Garofalo (1994) complementa ao pontuar que para o sucesso do crossover primeiro deve-se observar que uma das marcas necessita ter uma boa aceitação em seu nicho, pois somente dessa forma será possível ampliar para um público mais heterogêneo. Por outro lado, Hsu e Hsu (2017) salientam que essa técnica do marketing auxilia no aumento da visibilidade do produto e ao mesmo tempo promove uma intenção de consumo ao oferecer mais benefícios, no comparativo com o produto original.

5.1 Crossover métrico

Com base no postulado como fundamentos para a realização do crossover de marketing é possível traçar os parâmetros para a construção de um crossover métrico para a área acadêmica.

³⁸ Co-branding é "a associação ou combinação, de curto ou longo prazo, de duas ou mais marcas individuais, produtos e/ou outros ativos distintos" (RAO; QU; RUEKERT, 1999, p. 259) cujo objetivo é o benefício mútuo ao oferecer novos produtos ou serviços que vão se beneficiar da reputação de cada parceiro

Quadro 3 – Parâmetros para a construção de um crossover

Estratégia	Objeto de estudo	Fontes de dados	Resultados
Produto (Artigos, Revistas e Jornais)	Métricas em nível de artigo e citação	Ferramentas proprietárias, APIs, Editoras	Indicador de visibilidade e ranqueamento das pesquisas de maior reverberação
Canal (Editoras)	Disseminação e apropriação do conteúdo	Metadados dos sites das editoras	Indicadores refinados de consumo das revistas, ou dos jornais, ou dos artigos
Marketing (Comunicação e Divulgação Científica)	Alcance das publicações	Metadados georefenciados das editoras	Indicador de impacto (quem, está consumindo o que, com qual frequência, de qual localidade e com qual reverberação) da produção científica

Fonte: Elaborado pelo autor

O Quadro 3 apresenta como o crossover utilizado no marketing poder ser adaptado para os estudos métricos. Segundo Grier, Brumbaugh e Thornton (2006), ao estudar o impacto de produtos éticos para uma economia global, foi observado que uma das principais formas para a inserção de um novo produto originado de um crossover é a manutenção de características (linguagem, símbolos ou formato de apresentação) do produto de maior aceitação para que não exista uma rejeição. Por esse motivo uma das características adotada neste crossover métrico é a manutenção de um ranqueamento da produção acadêmica, uma vez que um dos principais resultados pretendidos pelos utilizadores (pesquisadores, agências de fomento e universidades) desse tipo de métrica é o posicionamento dos resultados de sua pesquisa perante os demais.

Como forma de fácil aceitação do público-alvo desse novo indicador de performance da publicação acadêmica apoiou-se no modelo de Fishbein (1967) que posteriormente foi adaptado em um estudo dos efeitos do crossover realizado por Oliver e Bearden (1985), no qual é identificado que o público-alvo segue a tendência de manter o comportamento, mesmo quando expostos a produtos originados de um crossover por atender as preferências sociais e subjetivas e pelo fato de estímulos favoráveis, no caso o manutenção do ranqueamento, criar padrões estáveis de reação.

Um ponto de limitação para o crossover e para toda a pesquisa com métricas sociais e a dependência de dados oriundos de APIs ou extraídos de ferramentas proprietárias (Altmetric e Plum Analytics). Porém, vale destacar que mesmo os softwares que prometem entregar dados mais refinados, como o georeferenciamento de consumo dos artigos, são baseados na geolocalização informada pelo usuário, e não pelo IP de acesso.

Neste capítulo o objetivo foi de apresentar os fundamentos para a realização do Crossover Métrico, no qual foi utilizado o conceito proposto por Grier; Brumbaugh e Thornton (2006), ao pontuar que o crossover pode ser utilizado para criar uma nova significação do mesmo produto e ser novamente reinserido no mesmo grupo de origem. Em seguida, foram apresentadas as categorias na qual o crossover pode ser inserido, segundo Yang (2009) e definido qual o modelo adotado neste trabalho.

6 METODOLOGIA

A construção metodológica da pesquisa foi no método dedutivo, com abordagem exploratório e descritiva ao relacionar à análise do objeto e suas variáveis. Ao confrontar com a Ciência da Informação uma área em constante evolução e por seu caráter transdisciplinar, a pesquisa foi alicerçada na cibermetria por considerar para a análise dos dados, as inter-relações entre os conteúdos e os processos comunicativos, além de também se ater nas particularidades de consumo aqui observado e apropriado por meio dos conceitos oriundos do Marketing, nas ferramentas e nos serviços utilizados pelos usuários. Ao mesmo tempo, o subcampo inclui “a aplicação das tradicionais técnicas informétricas a qualquer tipo de informação disponível na internet” (SIQUEIRA, 2016, p. 136).

Com essa definição foram utilizados os fundamentos da cibermetria para compreender os aspectos comunicacionais dos canais com o propósito de mensurar a comunicação e o envolvimento da comunidade científica. Para se construir o que se definiu como “Crossover Métrico” vários aspectos foram analisados com o objetivo de obter um indicador que possibilitasse a comparação do desempenho científico de publicações (artigos, livros, entre outros) que pertencem a diferentes domínios científicos.

6.1 Desenvolvimento de Indicadores

Com base nos tipos de redes sociais que podem ser utilizadas para a extração de informações sobre a dispersão de documentos científicos nestes ambientes foi possível extrair três indicadores principais que servirão para agregar os serviços e ao mesmo tempo categorizar as redes sociais segundo as suas apropriações gerais conforme apresentada na Figura 13 e com sua modelagem explicadas nas páginas seguintes.

Figura 13 - Divisão de categorias utilizadas para agrupar as redes sociais



Fonte: Elaborado pelo autor

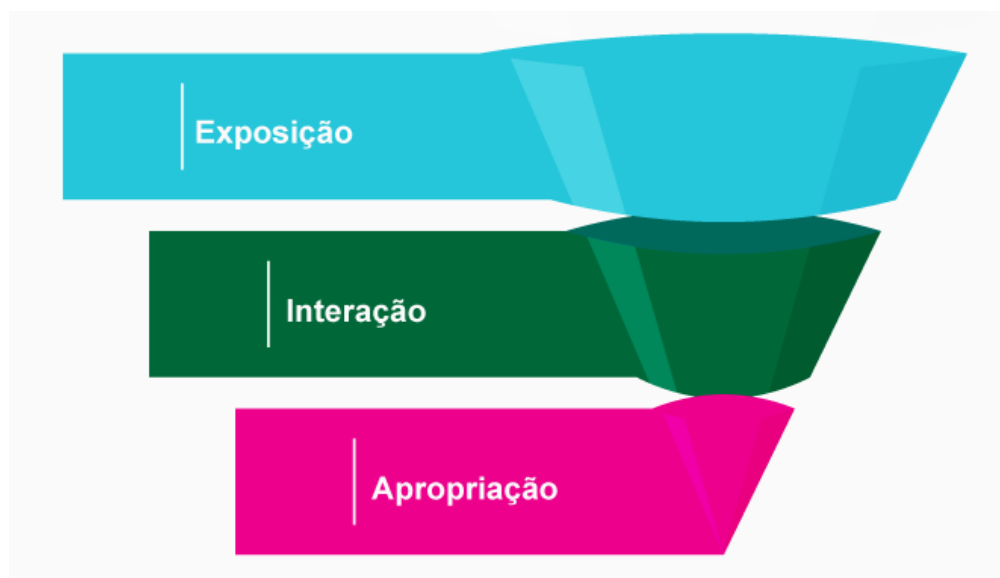
- ✓ **Exposição** – São considerados indicadores de exposição as publicações que sejam apenas para a divulgação científica. Também pode ser definido como o número de vezes que uma postagem recebe algum tipo de envolvimento do público. Esse conteúdo identificado como publicações de primeiro nível, geralmente tem o objetivo de apresentar, ou divulgar a publicação dos resultados de uma pesquisa, ou direcionar o público para um conteúdo específico, ou gerar conhecimento sobre um artigo ou pesquisa. Ocorre, geralmente, em redes sociais com pouca estrutura para se desenvolver uma argumentação ou debate linear sobre o conteúdo compartilhado e está presente em redes como Twitter, Facebook, Pinterest, Instagram e Google+;

- ✓ **Interação** - Definido como uma interação de segundo nível e envolve uma ação em resposta ao conteúdo disponibilizado. Esse tipo de envolvimento pode ser identificado por meio de comentários, incorporação de conteúdo, etc. Segundo pesquisa realizada por Mohammadi, Thelwall E Kousha (2016), salvar itens no Mendeley conta como uma indicação de leitura, pois 85% dos entrevistados declararam que salvaram artigos que tinham lido, pretendiam ler, ou tinham intenção de citar em publicações futuras. Esse tipo de envolvimento pode ser observado em gerenciadores de referências como o Mendeley, Citeulike, entre outros;

- ✓ **Apropriação** – Definido como a Apropriação de terceiro nível cuja ação pode ser pela apropriação da pesquisa original para o desenvolvimento ou aplicação na construção de verbetes em wikis; na abertura de tópicos para discussão da pesquisa em fóruns ou redes profissionais como o LinkedIn; para a realização de postagens em blogs; entre outros. A Apropriação se apresenta como o estágio da transformação que pode ser vista como uma citação ao trabalho original em artigos, palestras, desenvolvimento de produtos comerciais, matérias em revistas especializadas, entre outros.

Esse tipo de categorização segue em consonância com uma adaptação do Modelo AIDA (Figura 14) no qual os três indicadores estão distribuídos conforme o tipo de ação que o público pode realizar nas redes sociais durante cada fase. Nesse modelo a premissa é que o público após o primeiro contato com um documento científico, siga para a segunda etapa (Interação) e posteriormente para a última etapa (Apropriação).

Figura 14 - Adaptação do Modelo AIDA para o trabalho com os indicadores



Fonte: Adaptação do modelo proposto por Kotler e Keller (2011)

Essa proposta de três níveis de envolvimento do público pode ser observada no trabalho de Haustein, Bowman e Costas (2015) que dividiram as redes sociais: Acesso, Avaliação e Aplicação o tipo de atividade realizada pelo público nas redes sociais. Entretanto, os pesquisadores também apontam que as apropriações realizadas pelo público podem ocorrer de forma espiralada, ou seja, um fluxo constante entre emissor e receptor, o que pode provocar ações difusas na forma de utilização do mesmo artigo e sobreposição de atividades. No entanto, diferente da visão apresentada no estudo de que a espiral demonstra uma tendência de aumento nos acessos com a passagem do público pelos níveis de “Aplicação” ou “Avaliação”, observa-se uma possível falta uma relação direta entre a Exposição e aumento da interação ou do engajamento.

Porém independente do postulado por Haustein, Bowman e Costas (2015), no entendimento alcançado durante o desenvolvimento da pesquisa foi observada a necessidade de uma apuração constante do que é medido, devido à natureza dos dados serem oriundos de um ambiente em permanente mudança, e os conjuntos de dados seguirem a mesma tendência. Assim o norteamento para a divisão dos três índices foi baseado em:

- ✓ O índice deve ser abrangente, de modo que, cada rede seja inserida somente em um grupo;
- ✓ O índice deve ser estruturado de forma a acomodar novas redes ou ambientes com mesmas características no futuro;

- ✓ O índice deve agrupar redes que partilhem as mesmas características (tipo de atividade, tipo de envolvimento do público, temporalidade de conteúdo, indexação).

Para a divisão das redes dentro dos índices foram considerados o objetivo do ambiente, a apropriação do público e o tipo de dado que poderia ser extraído e a definição de quais redes seriam incorporadas por cada indicador foi baseado no conteúdo apresentado no Capítulo 4.

6.2 Procedimentos de coleta dos dados

Os dados provenientes da ferramenta foram escolhidos por ser uma base referencial multidisciplinar que tem cobertura em todas as áreas do saber. No comparativo com a base da Plum Analytics, os dados apresentados pela Altmetric demonstraram melhor padrão de apresentação, extração de dados, capacidade de manipulação de informações e de identificação dos pontos de vulnerabilidade dos dados.

O *corpus* da análise foi constituído do Ranking de artigos com maior Altmetric Attention Score nos anos de 2016 e 2017³⁹, indicador que consolida os dados de fontes alométricas pela ferramenta Altmetric.com. Os critérios para a definição desse *dataset* foi para, em um primeiro momento ter uma base com cobertura de um maior número possível de redes sociais e ter possibilidade de avaliar a estabilidade e performance do crossover. Em um segundo momento, objetivou-se ter uma base com dados mais completos o que permitiria observar qual a melhor forma de tratar os dados, além de identificar possíveis falhas em replicações da pesquisa.

Assim, foram utilizados dois *datasetes*: o primeiro com o ranking do ano de 2016 cujo objetivo foi testar a possibilidade do desenvolvimento do crossover e o segundo ranking com dados do ano de 2017, que serviu para comprovar a aplicabilidade do crossover. Os dados coletados são registros fornecidos pela ferramenta Altmetric.com, com valores indexados até o dia 08 de dezembro de cada ano de publicação dos rankings. Dessa forma, pode existir uma distorção se os mesmos artigos e documentos ranqueados sejam visitados em data posterior.

Embora o volume de dados de cada dataset seja de apenas 100 registros, a adoção encontra respaldo em um estudo realizado por Eugene Garfield (1992), em um conjunto com outros pesquisadores, no qual foi levantada a hipótese da possibilidade de prever futuros indicados ou ganhadores do prêmio Nobel, somente com base nas citações das publicações. O

³⁹ A tabela de dados completa (dados brutos) pode ser conferida no endereço <http://bit.ly/crossoverCI>

recorte de de seis pesquisas permitiu concluir que o volume era suficiente para identificar os pesquisadores que tinham maior propensão a receber o prêmio.

Dessa forma, foram baixados os dados do ranking do Top 100 do ano de 2016, em 12/12/2016 e do do ano de 2017, em 15/12/2017. Em ambos os casos os dados de referentes aos gerenciadores de referências foram inseridos manualmente, devido a ferramenta Altmetric.com não utilizar as informações para o cálculo do Índice de Atenção e como consequência, também, ignorar nos rankings. Entretanto, observa-se que estas informações representam um importante tipo de envolvimento entre o público e os documentos científicos e por esse motivo foram incluídos na pesquisa.

6.3 Tratamento dos dados

Com base nas categorias: Exposição, Interação e Apropriação (conforme pode ser observado no Capítulo 6.1) a pesquisa foi direcionada para relacionar as teorias apresentadas para desenvolver entendimentos sobre o comportamento das variáveis com a aplicação de um crossover métrico, bem como a organização das informações em núcleos de conteúdos e posteriormente o relacionamento dessas análises para a elaboração das considerações finais.

Dessa forma, a primeira análise foi para verificar a estrutura dos dados, possíveis duplicatas e a ausência de alguma informação que pudesse inviabilizar a pesquisa. Em seguida foram realizados testes de qual o melhor tipo de modelagem aceitaria a premissa de construção dos Indicadores (Exposição, Interação e Apropriação), que compreenderam as seguintes redes apresentadas no Quadro 4:

Quadro 4 - Indicadores e respectivas redes

Indicadores	Redes
Exposição	Twitter, páginas do Facebook, Google+, Weibo e Vídeos
Interação	Mendeley, Connotea, Citeulike, Reddit
Apropriação	Noticias, Blogs, Páginas governamentais, Wikipedia, LinkedIn, Syllabi, RH, Q&A, Peer Review

Fonte: Elaborado pelo autor

A importância da realização dos testes foi para buscar a melhor união de redes diferentes. Esse tipo de procedimento demonstra ser comum na literatura de diversas áreas do saber ao observar que dados transformados podem auxiliar na resolução do problema da falta de normalidade e/ou de heterocedasticidade (HOAGLIN; MOSTELLER; TUKEY, 1983; ZAR, 2010).

As manipulações ou modelagem dos dados podem ser obtidas por meio de transformações lineares das variáveis originais, como soma ou subtração, onde os resultados apresentados pelas variáveis possuem peso semelhante. Outra possibilidade é a realização de transformações não lineares, como logaritmo ou funções trigonométricas, que trazem como resultados variâncias menores e, desta forma, alcançam mais facilmente a homocedasticidade (ZAR, 2010).

Para identificar qual o melhor tipo de transformação se adequaria os dados foi testada a modelagem por desvio padrão. Quando foi extraída a média (μ) e o desvio padrão (σ) para cada artigo e rede individualmente. Posteriormente os dados foram somados dentro de cada indicador, conforme pode ser observado na Equação 1.

Equação 1 - Fórmula utilizando Desvio Padrão

$$Ix = \frac{DN - \mu}{\sigma}$$

DN = A distribuição normal para cada conjunto de eventos altmetricos por documento científico;

μ = A média do total de eventos altmetricos por rede;

σ = O desvio padrão do total de eventos altmetricos por rede.

Os resultados, mesmo antes de aplicarmos qualquer modelagem de tempo, apresentaram desvios de normalidade assimétricos e quando realizado o teste de Kolmogorov-Smirnov chegou-se a conclusão que após a transformação dos dados os resultados não atingiram a normalidade. Dessa forma, o modelo foi descartado, pois conforme destacou Hair et al., (2009) o aceite desse tipo de resultado poderia gerar como um distanciamento e conclusões equivocados sobre os dados.

A segunda abordagem foi baseada em testes não paramétricos que utilizaram: logaritmo natural ou neperiano, logaritmo decimal e logaritmo binário. Entre as transformações realizadas, o logartimo natural demonstrou o melhor ajuste dos dados. Esse tipo de observação também foi extraída por Vanti, Costa e Silva (2013) ao postular o Fator de Impacto Web,

quando identificou que na modelagem dos dados que registram crescimento exponencial esse tipo de normalização era a mais adequada.

a transformação LOG reduz a faixa dinâmica dos valores, diminuindo a diferença entre eles. Acredita-se, ademais, que o logaritmo natural (ln) é o mais adequado neste caso pelo fato de ele ser comumente usado em cálculos matemáticos que apresentam números que crescem de forma exponencial, de modo a suprir a real necessidade aqui apresentada. (VANTI; COSTA; SILVA, 2013, p. 231 e 232)

Após os dados normalizados, os valores de cada rede foram somados para extrair os valores brutos dos Indicadores Exposição, Interação e Apropriação. Posteriormente, foi aplicado um fator de tempo para extrair o real valor dos Indicadores e validar o crossover de redes.

6.4 Definição da janela de tempo

A pesquisa utilizou dados compreendidos em janelas temporais delimitadas pelos dados dos Top 100 dos anos de 2016 e 2017. Em cada um, o recorte de tempo foi:

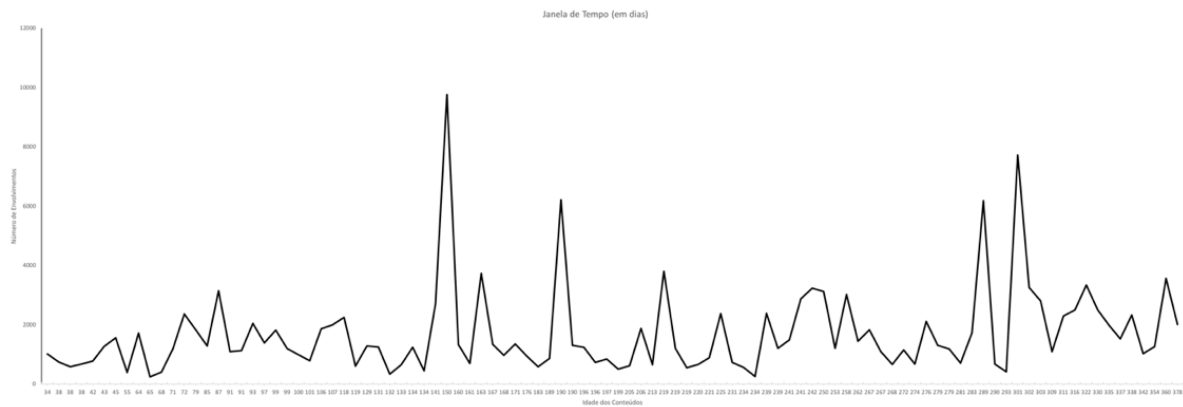
Quadro 5 - Apresentação dos períodos que compreendem cada dataset

Top 100	Janela Temporal
2016	Novembro de 2015 a novembro de 2016
2017	Dezembro de 2016 a novembro de 2017

Fonte: Elaborado pelo autor com base em dados da ferramenta Altmetric.com

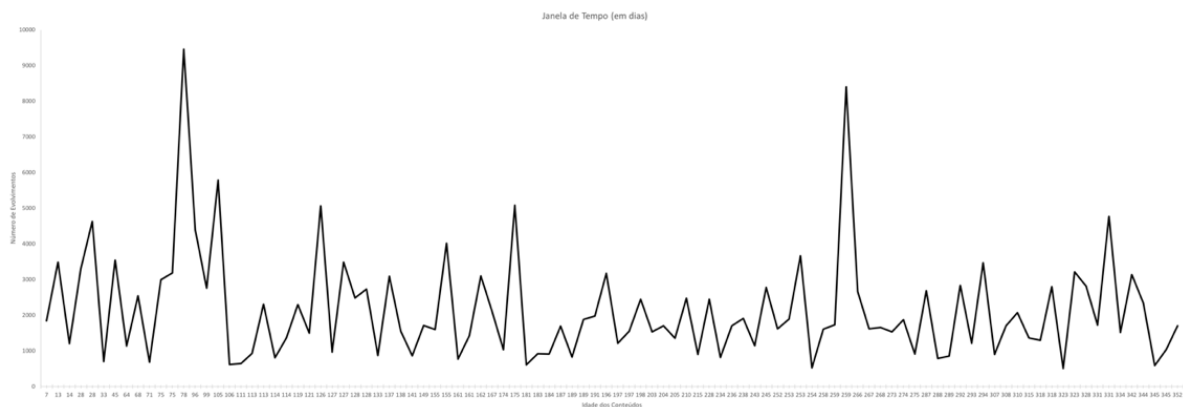
No entanto, a idade das publicações teve variações superior a um ano. Conforme observa-se nos Gráficos 1 e 2, que estão classificados por tempo de publicação (documento mais recente ao mais antigo) não existe uma relação entre o tempo de publicação e o envolvimento do público com os eventos altmétricos, ou seja, não é possível afirmar que o fator tempo tenha alguma influência sobre o tipo de envolvimento do público, com o consumo de documentos científicos.

Gráfico 1 - Quantidade de envolvimento por tempo de publicação – dataset 2016



Fonte: Dados da pesquisa

Gráfico 2 - Quantidade de envolvimento por tempo de publicação – dataset 2017



Fonte: Dados da pesquisa

Um ponto observado, ao trabalhar os indicadores com a base de tempo, também extraído da ferramenta Altmetric.com, foi a constatação de que o período de tempo de cada publicação para atingir o volume de citações pode ter sido influenciado por fatores que não estavam contemplados no escopo dessa pesquisa. Porém, pela observação dos dados é possível afirmar que os hábitos de publicação têm sofrido alterações ao longo dos anos, com um maior volume de publicações e reverberação nas redes sociais e como reflexo tem privilegiado conteúdos científicos mais recentes, conforme observou Lisée, Larivière; Archambault (2008) ao analisar a obsolescência de produções científicas por meio das citações.

Esse resultado foi comprovado em ambos *datasets* utilizados, ao observar a inexistência de conteúdos com pouco mais de um ano de publicação. Por isso, pode-se afirmar que o fator tempo demonstrou baixa influência no tipo de envolvimento do público com os Top 100, mas se apresenta com condições de ser utilizado como fator de depreciação na geração do ranqueamento dos conteúdos disponibilizados nos Top 100.

Para a realização dessa operação foi selecionado o algoritmo do site Hacker News⁴⁰, desenvolvido por Paul Graham, em 2009⁴¹. A escolha foi pela estabilidade da aplicação que foi adaptada para outros serviços e sites da web, que tinham interesse em ranquear os conteúdos pela relevância das interações, mas com a necessidade de aplicar um índice de depreciação.

A fórmula para a obtenção do ranking é apresentada em Equação 2:

Equação 2 - Fórmula de depreciação

$$(p - 1) / (t + 2) ^ f_i$$

p = Total de Interações

t = A diferença entre a data de publicação e data do fechamento dos rankings.

f_i = Fator de idade

p é subtraído por 1 para tentar minimizar a diferença da interação dos autores
f_i na fórmula original trabalha com a unidade de horas, mas devido ao escopo de nossa pesquisa será utilizado a unidade de dias, com o Fator de Idade (f_i) = 1.

Fonte: <https://www.hackernews.com>

A fórmula de depreciação serviu para atender dois objetivos: (I) deixar mais dinâmico o ranqueamento dos documentos científicos assegurando que os mais antigos possam ter um decaimento gradativo e (II) não gerar distorções nos comparativos com publicações recentes. Em ambos os casos também será possível observar os documentos que tiveram alto número de eventos altmetricos conseguiram, mesmo com o algoritmo de decaimento, manter um bom posicionamento no ranking.

Com a posse desses dados foi possível realizar um novo ranqueamento para demonstrar os documentos integrantes do Top 100, do ano de 2016 e 2017, que tiveram maior atenção do público. Os resultados do crossover estão demonstrados no Apêndice A e B, que foram classificados pelo Índice de Apropriação por ser o último nível dos indicadores e que representa como o trabalho original ganhou algum desdobramento em outras publicações. Esse tipo de classificação segue em consonância com a proposta da pesquisa e com a estratégia do crossover: apresentar um novo indicador, mas sem gerar rupturas com os tipos de resultados comumente buscados pelo público, neste caso as citações.

Na metodologia, buscou-se apresentar o roteiro do desenvolvido da pesquisa com a construção dos Indicadores (Interação, Exposição e Apropriação), baseados nos conceitos

⁴⁰ <https://www.hackernews.com>

⁴¹ O HackerNews é um site que tem o conteúdo produzido pelos usuários. Segundo Paul Graham, o algoritmo de ranqueamento foi desenvolvido com o objetivo de melhorar a usabilidade do site. Ver também Graham (2009)

abordados no Capítulo 4. Essa primeira parte serviu para demonstrar como os dados foram agrupados e posteriormente tratados.

Por último, discutiu-se a janela temporal e os procedimentos adotados para definir qual o melhor modelo de cálculo para extrair as informações sem ter o impacto de redes com alto envolvimento, como o Twitter no comparativo com redes de baixo envolvimento, como o Mendeley.

7 CROSSOVER E INTERPRETAÇÃO DOS DADOS

Durante a construção do Crossover foi possível identificar a necessidade da criação de Indicadores distintos para os conjuntos de atividades realizadas pelo público. O objetivo inicial era a construção de apenas um indicador que representasse todo o envolvimento do público, mas ao observar o tipo de envolvimento do público, com as publicações disponibilizadas nos ambientes digitais, constatou-se que a melhor forma de avaliar a reverberação dos conteúdos seria por meio de três indicadores conforme detalhado no Capítulo 3.

Ao postular a possibilidade de um crossover, optou-se por utilizar toda a listagem de documentos presentes no Top 100, de 2016, que incluiu uma diversidade de publicações. Ao avaliar os tipos de pesquisas ranqueadas na ferramenta Altmetric.com não foi observada a necessidade de proceder com a separação por áreas de saber ou tipo de publicação para identificar quais teriam maior atenção do público.

Os resultados apontaram uma mudança significativa na ordem dos artigos com melhor Índice de Apropriação⁴², conforme pode ser acompanhado no Apêndice A. Essa variação ocorreu por dois motivos: a normalização dos dados possibilitou que todas as redes tivessem seus pesos nivelados e ignorou a aplicação de pesos pré-determinados as redes. O segundo ponto, provavelmente o mais importante, foi a eliminação do ponto controverso e obscuro na forma como as métricas alternativas são mensuradas pela ferramenta Altmetric.com.

A análise individual por redes também demonstrou imprecisa, pois se fosse realizada seria superficial e não conclusiva, apresentando o Twitter (Gráfico 3 e

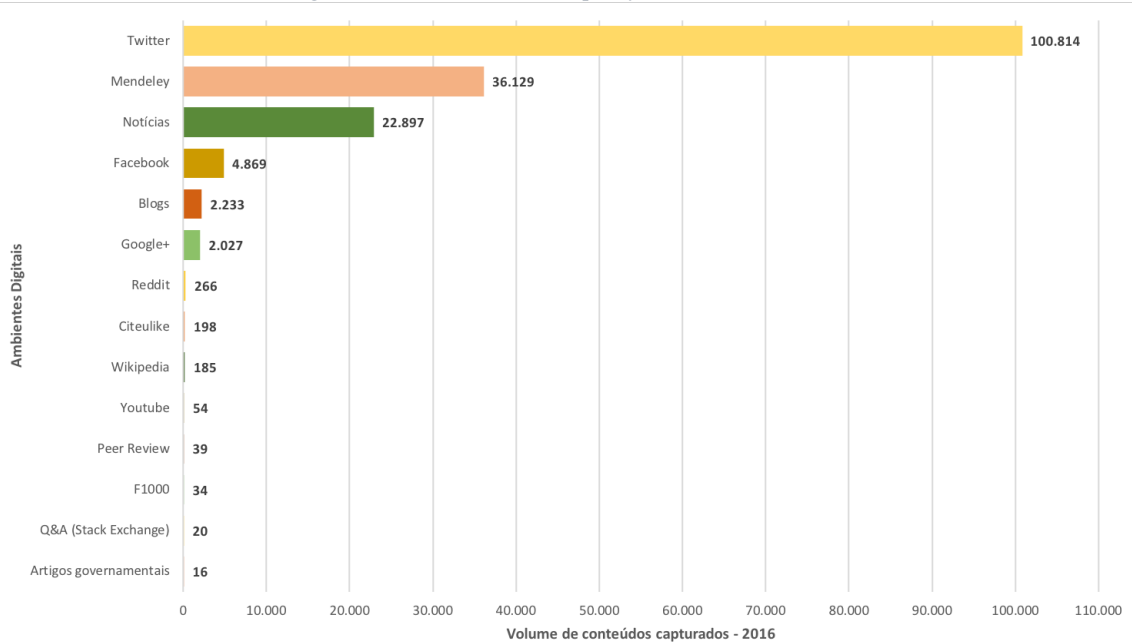
Gráfico 4) como rede de maior disseminação. Essa primeira afirmativa não é errada e foi comprovada por diversos estudos do impacto do microblog na divulgação científica (MALEKI, 2015), porém é limitador e excludente no comparativo com outras redes com o mesmo objetivo, mas com capilaridade de público menor, ou pelo menos, não acompanhada fielmente pelas ferramentas.

Um outro ponto também importante, é a diferença entre plataformas, como o Weibo (ferramenta semelhante ao Twitter) que tem grande influência na China, ou o Facebook, que devido a sua política de privacidade não permite o monitoramento e captura de dados de perfis, somente de páginas e grupos públicos. Com esse cenário é oportuno salientar que todo e

⁴² O Índice de Apropriação foi definido para ranquear os trabalhos, pois é o que mais se assemelha a uma citação tradicional. Em uma alusão a um periódico impresso se extrairia as mesmas fases: compra = exposição; leitura e comentários = Interação e publicação de artigo ou outro documento científico = a apropriação

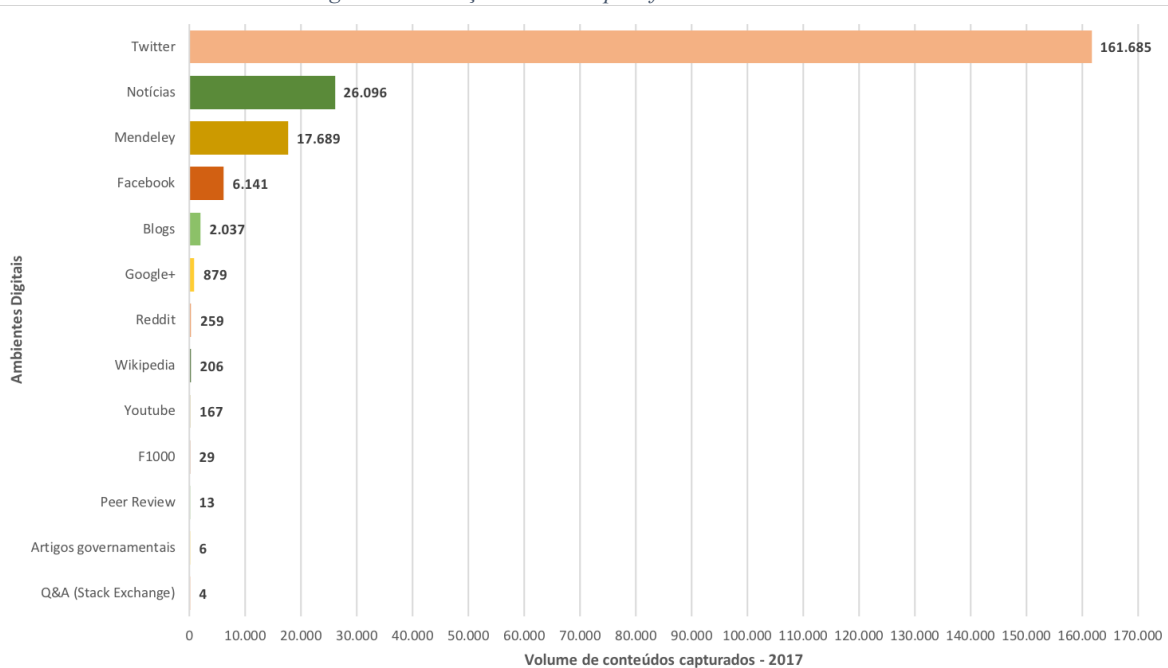
qualquer trabalho baseado em dados de redes sociais será apenas um retrato parcial da repercussão dos documentos científicos dentro destes ambientes.

Gráfico 3 - Volume de dados por redes que compõem o Top 100 dos documentos com maior relevância no ano de 2016, segundo totalização realizada pela ferramenta Altmetric



Fonte: Dados da pesquisa

Gráfico 4 - Volume de dados por redes que compõem o Top 100 dos documentos com maior relevância no ano de 2017, segundo totalização realizada pela ferramenta Altmetric



Fonte: Dados da pesquisa

Além disso, é necessário o cuidado ao comparar números obtidos em diferentes redes. Embora o propósito dessa pesquisa não seja o aprofundamento em como as diversas áreas estão

se apropriando dos ambientes digitais na divulgação acadêmica, deve-se levar em consideração como cada campo utiliza os ambientes digitais e com qual finalidade. Como destacam Patterson, Snyder e Ullman (1999) ao observarem a forma é modelo de publicação na Ciência da Computação que é fortemente baseada na publicação em anais de congressos e conferências, em detrimento de produções com foco em revistas especializadas. Um outro dado que ajuda a respaldar esse argumento é o perfil da área que se concentra em redes como Stack Exchange e Reddit, ao invés do Twitter. As áreas que demonstraram a maior reverberação e que compõem os dois datasets podem ser observados no Quadro 6.

Quadro 6 - Distribuição dos conteúdos, por áreas do saber, dos Top 100 de 2016 e 2017⁴³

Documentos publicados por áreas de interesse - 2016		Documentos publicados por áreas de interesse - 2017	
Áreas	Nº de Documentos	Áreas	Nº de Documentos
Ciências Médicas e da Saúde	49	Ciências Médicas e da Saúde	53
Ciências Biológicas	14	Ciências Biológicas	20
Estudos na Sociedade Humana	12	Ciências da Terra e do Meio Ambiente	9
Ciências da Terra e do Meio Ambiente	6	Estudos na Sociedade Humana	8
História e Arqueologia	6	História e Arqueologia	5
Ciências físicas	6	Ciências físicas	2
Pesquisa e Reprodutibilidade	4	Pesquisa e Reprodutibilidade	2
Ciências de Materiais	2	Ciências da Computação e Informação	1
Ciências da Computação e Informação	1		

Fonte: Dados da pesquisa

Os quadros registram o volume de publicações por áreas do saber que figuram no Top 100, do ranking da ferramenta Altmetric.com. O destaque fica nas publicações voltadas para a área de Ciências Médicas e Biológicas que concentram mais de 60% dos documentos científicos publicados em dois anos e que compuseram o Top 100, nos anos de 2016 e 2017.

Com a conjunção do cenário apresentado pelos Gráfico 3 e

Gráfico 4, onde o Twitter demonstra grande destaque seguido pelo Mendeley e Notícias invertendo posições no comparativo com as demais ferramentas, observa-se que mesmo com a diferença de envolvimento e utilização das redes não deve-se partir para o pensamento simplista de que existe uma enorme variedade de métricas que buscam suprir as lacunas do impacto dos meios digitais na produção científica e muito menos acreditar que seria mais prático buscar caminhos para a construção e a substituição de todas por um único indicador, como o Fator de Impacto. Como afirma Lin e Fenner (2013), essa facilidade em um primeiro momento poderia se mostrar vantajosa, mas ao longo do tempo seria impraticável se o objetivo for preservar a

⁴³ As Áreas de Interesse foram atribuídas pela ferramenta Altmetric.com (nota do autor)

amplitude das informações que essas métricas oferecem, bem como manter a conveniência de suas diferentes naturezas.

7.1 Modelagem dos Indicadores

Na análise da janela temporal foi identificada a falta de um padrão de tempo, e em uma avaliação detalhada dos datasets (Top 100 tanto de 2016 como o de 2017) também constatou-se que a métrica “Tempo” gerou baixo impacto no envolvimento do público, uma vez que é possível identificar que o engajamento do público com os documentos científicos não obedeceu um crescimento de acordo com uma ordem cronológica, ou seja, não é possível afirmar que as publicações mais antigas seriam beneficiadas com o aumento de engajamento do público. Assim, a métrica “Tempo”, conforme explicitado no Capítulo 6.4, ficou restrita para a confecção do Ranking do Crossover.

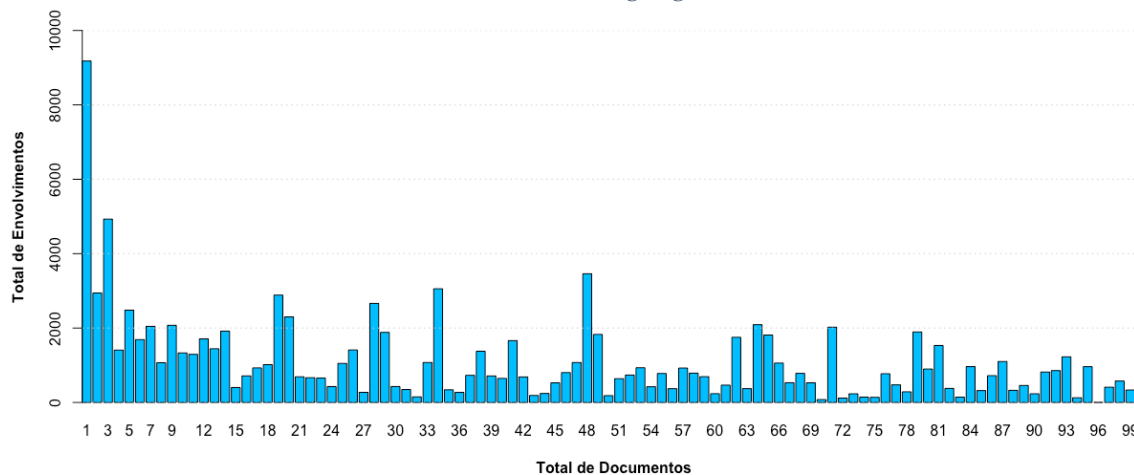
Com a certeza da falta de influência da métrica Tempo, a opção de modelagem dos indicadores ficou centrada somente no envolvimento do público com os documentos científicos. Essa opção aconteceu devido a falta de dados detalhados sobre o tipo de engajamento por rede, pois a ferramenta Altmetric.com ao divulgar o seu ranking anual disponibiliza apenas dados consolidados do período.

Embora seja possível acessar os eventos altmétricos de cada uma das publicações, essa opção apresentaria uma distorção na análise devido a ferramenta Altmetric.com não oferecer janelas temporais fixas⁴⁴. Assim, entre a data de captura dos dados, para a elaboração do Top 100, e a divulgação do ranking, os dados sofreram alterações, que inviabilizaria a análise detalhada do tipo de envolvimento por rede.

O primeiro ponto observado foi a avaliação realizada com o auxílio do Gráfico 5 (Dados não normalizados) e Gráfico 7 (Dados normalizados por logaritmo natural) sobre o impacto dos dados com e sem normalização. Essa primeira avaliação permite observar se existem distorções que pudessem invalidar a normalização. Os resultados reforçam que a transformação de dados, quando necessária, podem facilitar a interpretação dos resultados e reduzir as distorções na visualização da informação.

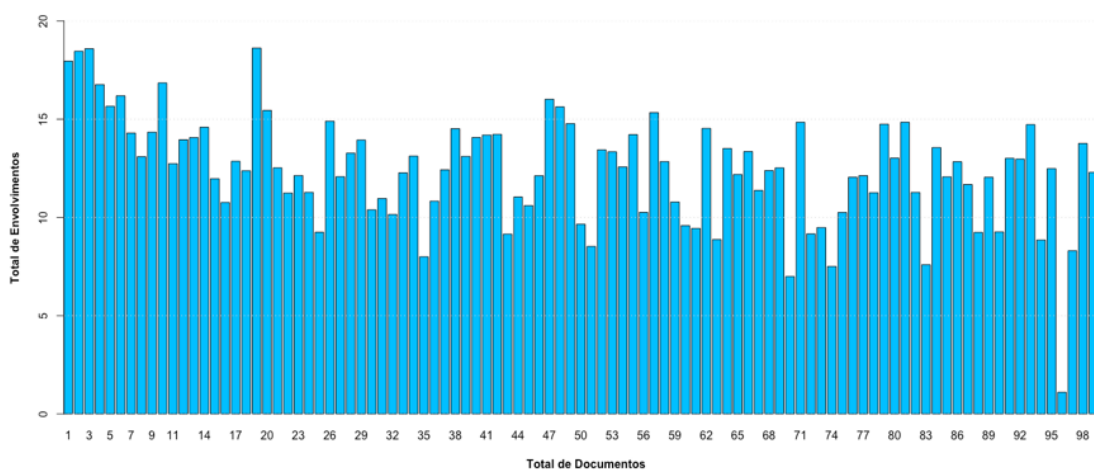
⁴⁴ As janelas temporais fixas permitiriam que os acessos aos dados de ferramentas como Altmetric.com tivessem o mesmo valor independente da época da consulta. No entanto, por optarem pela atualização dinâmica, os dados têm atualização constante. Dessa forma, os dados dos Top 100 tem pesos e valores diferentes dependendo da data de acesso. (Nota do autor)

Gráfico 5 - Somatório dos envolvimento para o Indicador Exposição para cada um dos documentos do dataset de 2016 na ordem do ranking original



Fonte: Dados da pesquisa

Gráfico 6 - Somatório dos envolvimento normalizados do Indicador Exposição para cada um dos documentos do dataset de 2016 na ordem do ranking original

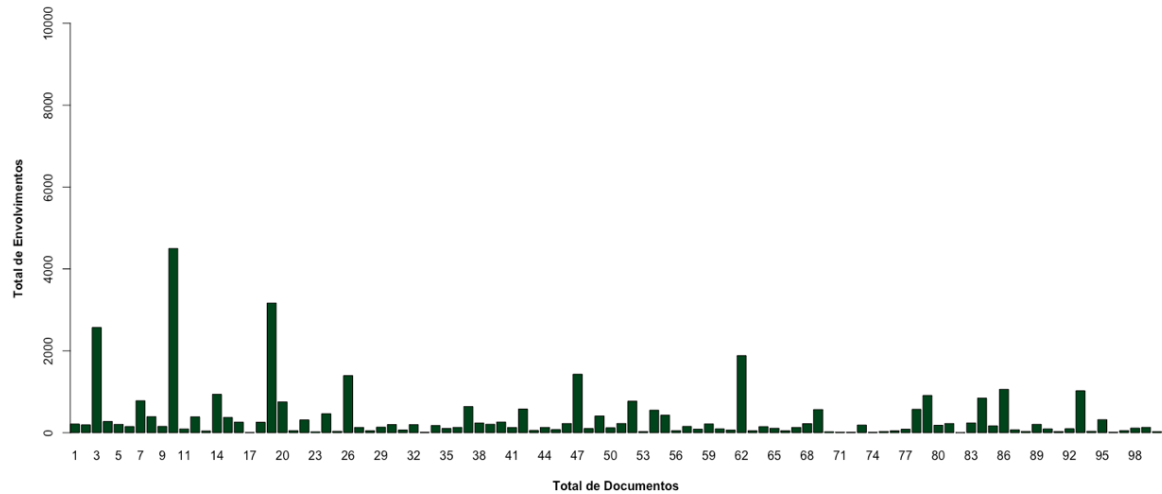


Fonte: Dados da pesquisa

Os gráficos demonstram de forma visual apenas o que poderá ser comprovado por meio de testes de correlação que serão aplicados adiante. Em ambas as visualizações se observou como foi o envolvimento do público com os documentos científicos. Enquanto no Gráfico 5 identifica-se apenas dois documentos que ultrapassam a escala dos cinco mil envolvimento, no Gráfico 6 essas oscilações são suavizadas e permitem observar com maior clareza como realmente foi realizado o envolvimento do público com os documentos científicos e destaca-se apenas dois documentos com baixo tipo de engajamento.

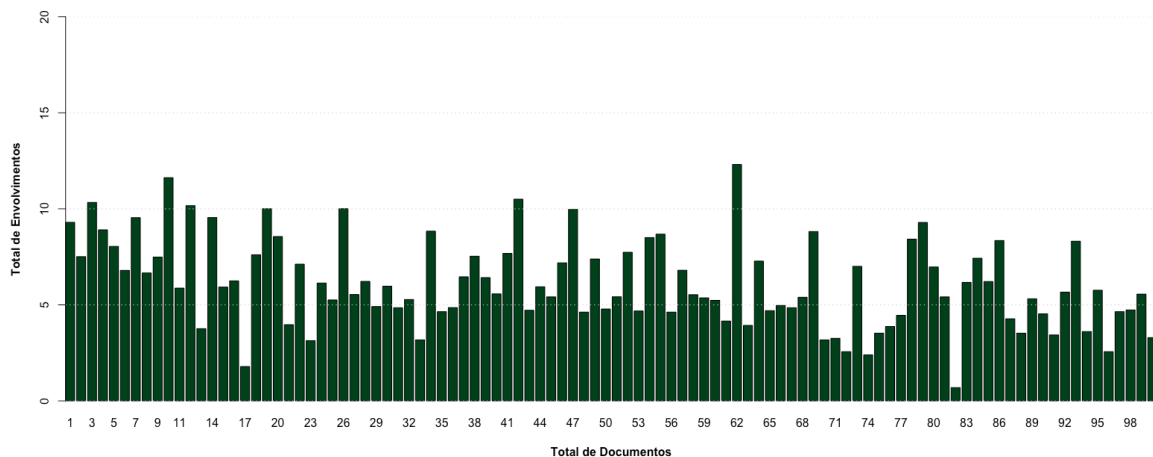
Ao avaliar isoladamente os gráficos de dados brutos (Gráfico 5 e Gráfico 7) é possível, também, constatar que a transferência do envolvimento não é direta entre as publicações. Muitos documentos apresentam alto engajamento no indicador Exposição, mas quando passam para o indicador Interação demonstram baixo desempenho.

Gráfico 7 - Somatório dos envoltimentos para o Indicador Interação para cada um dos documentos do dataset de 2016 na ordem do ranking original



Fonte: Dados da pesquisa

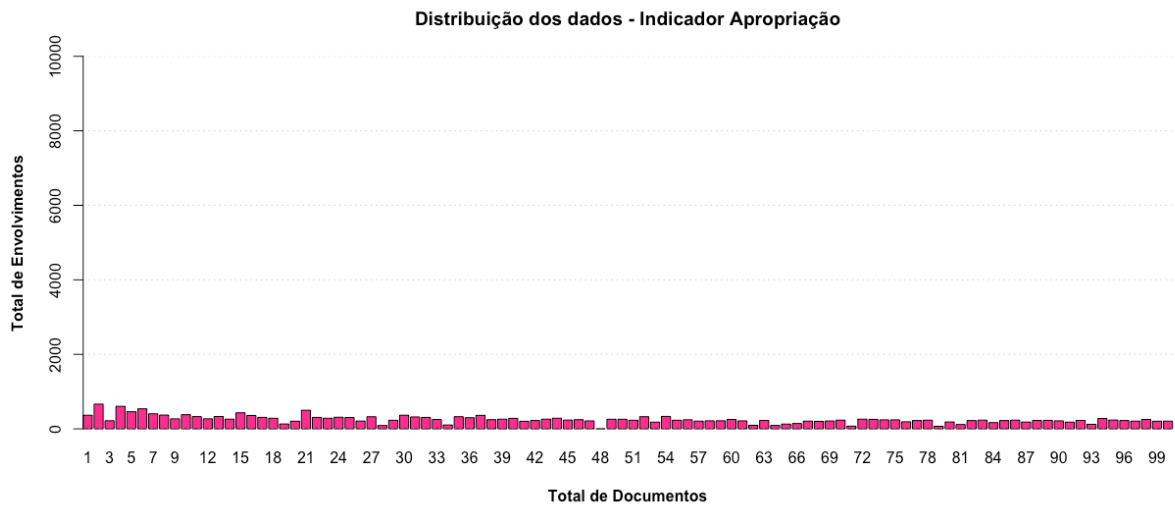
Gráfico 8 - Somatório dos envoltimentos normalizados do Indicador Interação para cada um dos documentos do dataset de 2016 na ordem do ranking original



Fonte: Dados da pesquisa

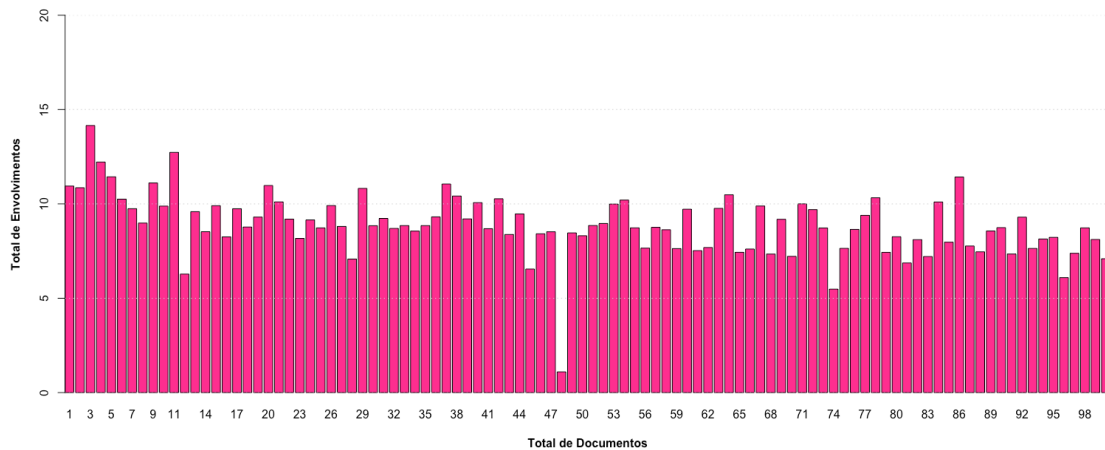
A normalização dos dados permitiu identificar com maior precisão a dispersão do envolvimento dentro dos indicadores. Como exemplo, pode-se utilizar o artigo “*United States Health Care Reform: Progress to Date and Next Steps*” (OBAMA, 2016), que ocupa a primeira posição no Top 100 de 2016, e registra alto volume no indicador Exposição, mas nos indicadores seguintes registra perda no envolvimento do público e demonstra, no comparativo com outras publicações, baixo engajamento.

Gráfico 9 - Somatório dos envoltimentos para o Indicador Apropriação para cada um dos documentos do dataset de 2016 na ordem do ranking original



Fonte: Dados da pesquisa

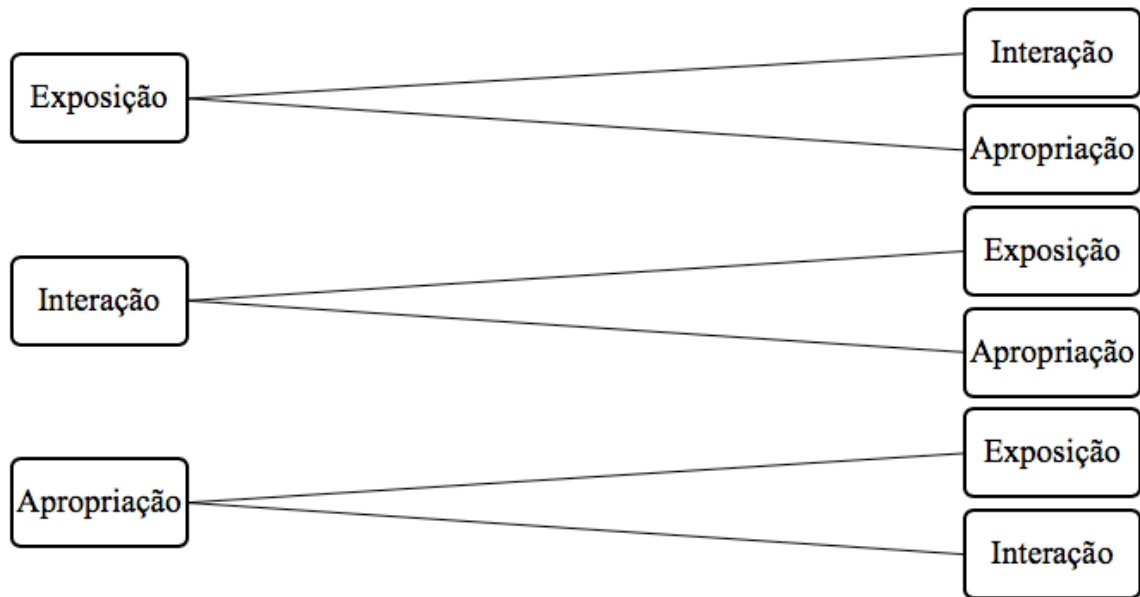
Gráfico 10 - Somatório dos envoltimentos normalizados do Indicador Apropriação para cada um dos documentos do dataset de 2016 na ordem do ranking original



Fonte: Dados da pesquisa

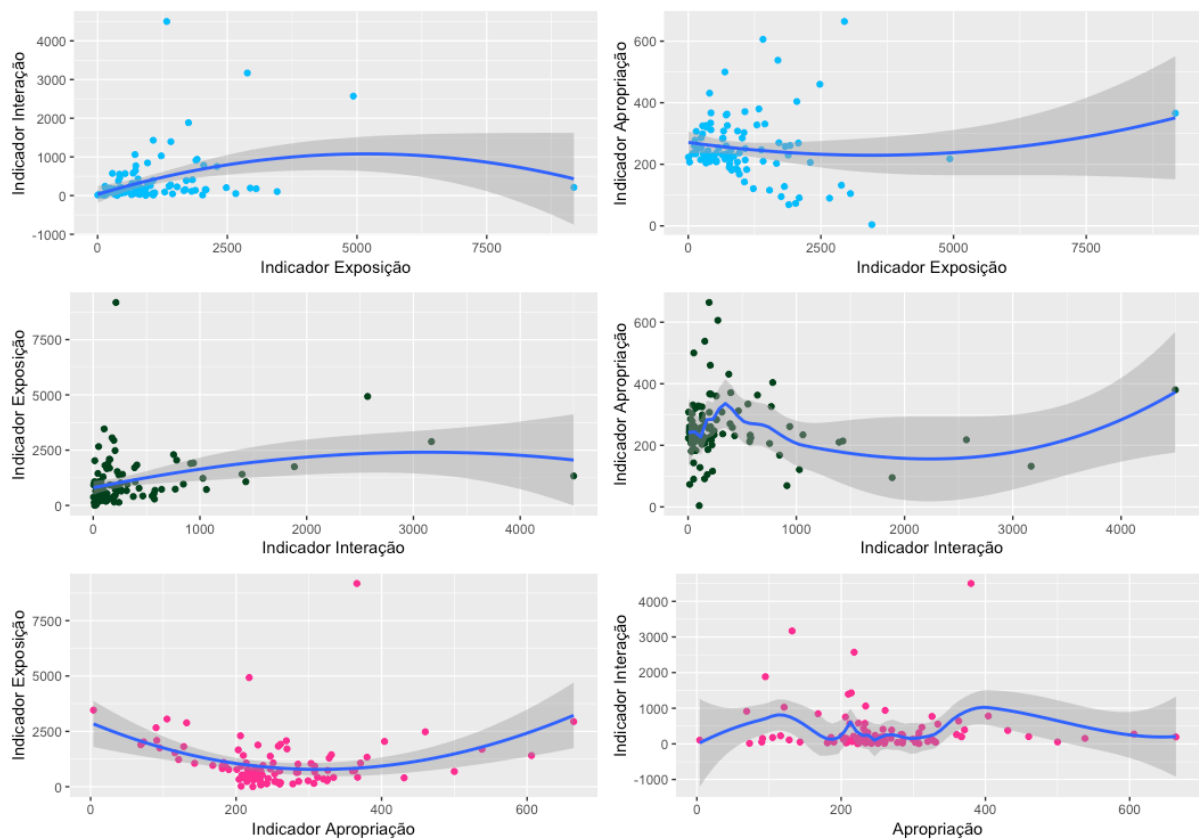
No entanto, ainda restou identificar a existência e tipo de correlação entre os indicadores, para observar a transmissão e qual o tipo de engajamento é realizado entre os indicadores. Para essa avaliação foi utilizado a correlação entre os Indicadores e o seu antecessor e predecessor, conforme demonstrado na Figura 15. Esse tipo de verificação possibilita observar se a transferência entre os indicadores é linear, ou seja, obedece a lógica de um funil que neste caso estaria compreendido em Exposição, Interação e Apropriação, ou é independente, sem seguir nenhuma sequência.

Figura 15 - Esquema para a análise de correlação entre os indicadores



Com base no esquema apresentado na Figura 15, foi gerado o Gráfico 11, que registra a correlação com os dados brutos, sem nenhum tipo de normalização, cujo objetivo foi verificar a distribuição dos dados em relação aos indicadores.

Gráfico 11 - Distribuição da relação entre os Indicadores sem a realização da normalização para o dataset de 2016

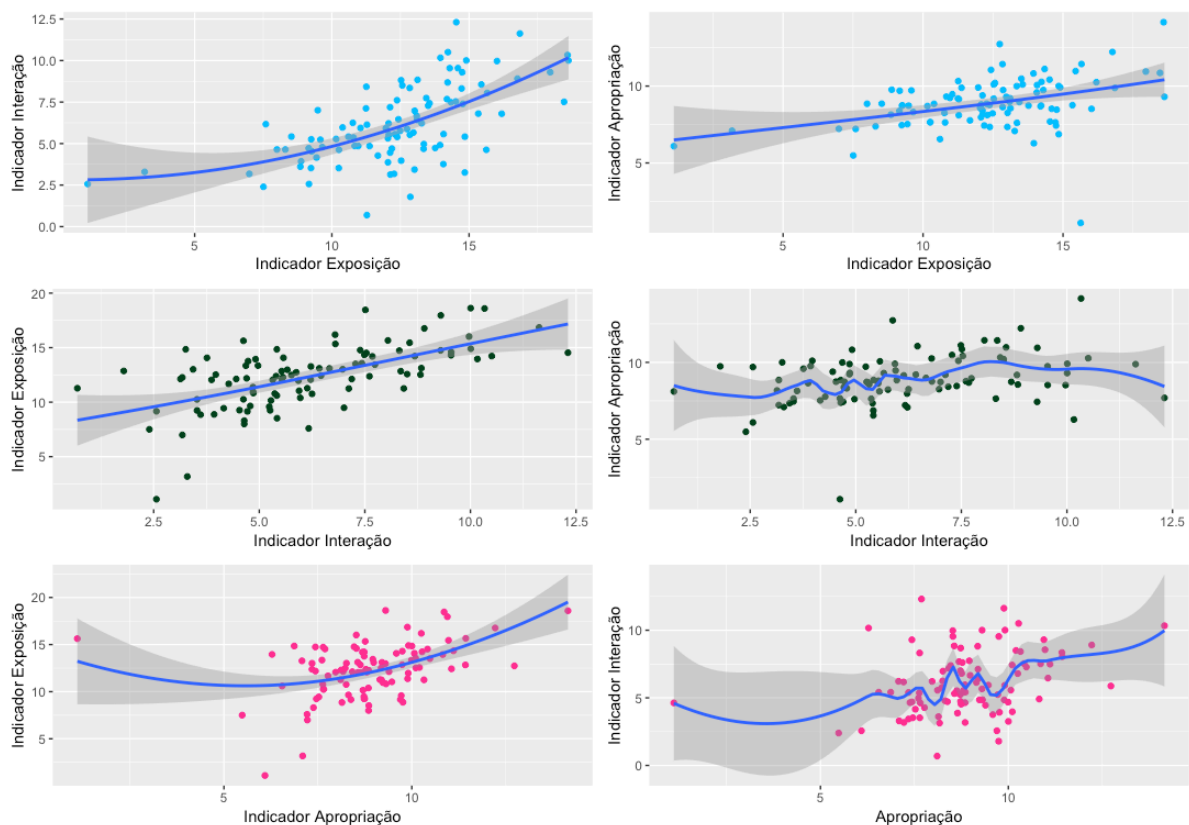


Fonte: Dados da pesquisa

Ao observar o comportamento dos dados no Gráfico 11 é possível identificar que existe alguma relação entre os Indicadores. Porém, a concentração de dados fica próxima ao cruzamento dos eixos x e y que não permite afirmar o nível de transferência do envolvimento entre os Indicadores. Com os dados é possível somente identificar a existência de outliers, que podem ser os responsáveis pela falta de linearidade.

Essa concentração em determinados pontos ao comparar com os gráficos 5, 7 e 9 dos dados não normalizados, desmonstram o impacto que algumas redes têm no envolvimento do público. Por esse motivo foi realizado o ajuste, por meio de normalização logarítmica natural, que também irá auxiliar na observação da diferença entre os valores extremos, cujo resultado pode ser observado no Gráfico 12.

Gráfico 12 - Distribuição da relação entre os Indicadores com base em dados normalizados por logaritmo natural para o dataset de 2016



Fonte: Dados da pesquisa

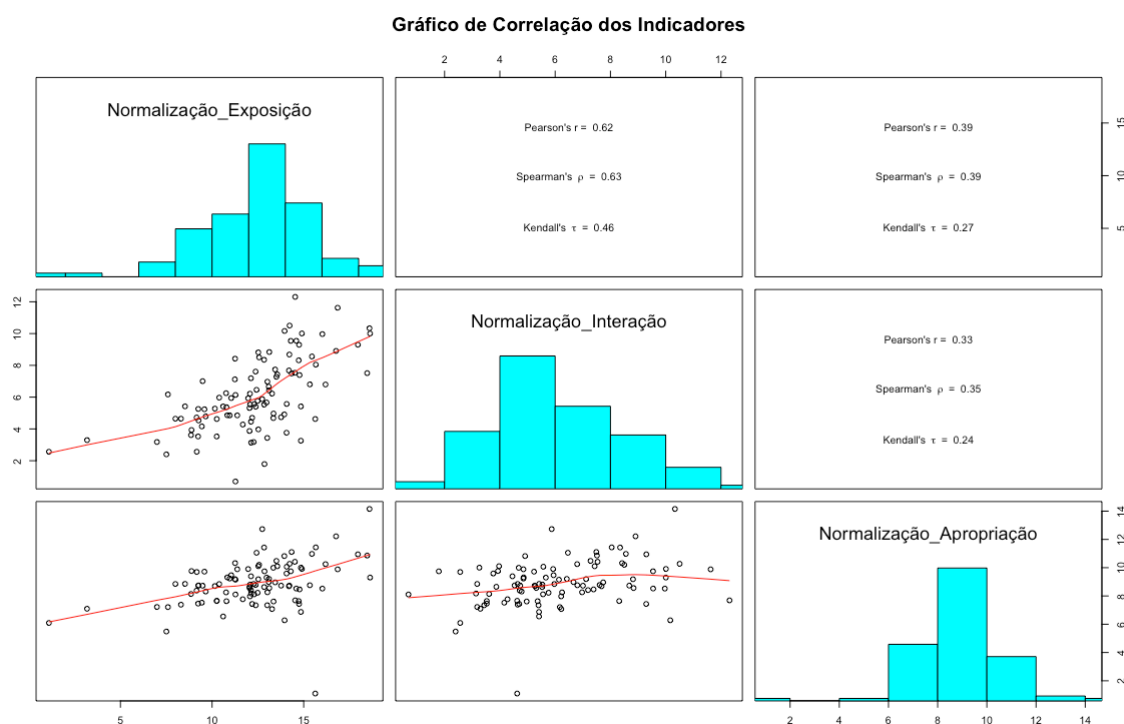
O Gráfico 12 ao ser analisado em conjunto com o Gráfico 13 permite observar que os indicadores normalizados Exposição e Interação demonstraram a maior correlação dos dados (0,62). Este percentual pode ser explicado devido ao alto volume de envolvimento que se observa nas redes de acesso e que compõe o indicador Exposição. Dessa forma, ao analisar pela perspectiva do modelo de funil, a retenção esperada nessa primeira correlação de indicadores

é alta, porém não se deve desprezar e/ou ignorar que esse o tipo de ação não requer um envolvimento propriamente aprofundado, pois o uso desse conteúdo é apenas de arquivamento para posterior leitura ou utilização.

Essa transferência parcial do envolvimento segue pelos indicadores Interação e Apropriação que registraram um coeficiente de correlação baixa (0,33), mas ainda reforça o conceito de funil na transferência de envolvimento do público.

O tipo de envolvimento que se observa em ambas as correlações representa a realidade do consumo de pesquisa que passa pela fase de conhecimento do material, salvar para posterior utilização ou referência posterior e utilização. Com essa descoberta é possível confirmar que utilização de três indicadores para demonstrar o envolvimento ou caminho seguido pelo público no consumo de documentos científicos demonstra-se viável.

Gráfico 13 - Histogramas, coeficientes de correlação e diagramas de dispersão de $r =$ Pearson, $\rho =$ Spearman, e $\tau =$ Kendall Tau, entre os três indicadores normalizados: Exposição, Interação e Apropriação para o dataset de 2016⁴⁵



Fonte: Dados da pesquisa

Outra observação que pode ser extraída do Gráfico 13 segue o postulado por Thelwall, Haustein, Larivière e Sugimoto (2013), sobre a tendência em associar as contagens de citações às métricas da web por meio de testes de correlação. Esse tipo de modelagem acontece devido as citações serem um indicador reconhecido do impacto acadêmico, assim supõe-se que qualquer outra medida que se correlacione positivamente, também pode ser associado a esse

⁴⁵ O gráfico 15 tem o objetivo de demonstrar a dispersão para cada conjunto de indicadores e apresentar por meio de testes de Pearson, Spearman e Kendal, o tipo de correlação obtida em cada teste. (Nota do autor)

impacto. Essa premissa serve para reforçar a viabilidade da construção do crossover, ao apresentar um indicador “Apropriação” e com elementos que o público demonstra ter familiaridade.

Dessa forma, com as análises do Top 100, do ano de 2016 foi possível modelar indicadores que podem auxiliar no entendimento de como funciona a dinâmica do consumo de conteúdos científicos nas redes sociais. Entre as principais descobertas está o fato de a janela temporal de doze meses não demonstrar influência na ação de Exposição, Interação ou Apropriação. A prova de conceito do modelo da metodologia de crossover métrico foi realizada com o Top 100 de 2017.

7.2 Prova de conceito do Crossover métrico

Para o teste de viabilidade do crossover utilizou a mesma metodologia aplicada no Top 100, do ano de 2016. Porém, antes de aplicar a metodologia foi realizada a verificação do comportamento do público com os documentos científicos nas redes sociais, conforme pode ser verificado no Quadro 7. Essa etapa foi para identificar a existência de crescimento, se todos as redes mantiveram o mesmo tipo de envolvimento nos dois anos e quais as redes foram excluídas e incluídas no acompanhamento da ferramenta.

Quadro 7 - Comparativo por redes do engajamento do público nos anos de 2016 e 2017

Redes	2016	2017
Artigos governamentais	16	6
Blogs	2.233	2.037
Citeulike	198	-
F1000	34	29
Facebook	4.869	6.141
Google+	2.027	879
Mendeley	36.129	17.689
Notícias	22.897	26.096
Peer Review	39	13
Q&A (Stack Exchange)	20	4
Reddit	266	259
Twitter	100.814	161.685
Wikipedia	185	206
Youtube	54	167
Total de envoltimentos	171.797	215.211

Fonte: Elaborado pelo autor

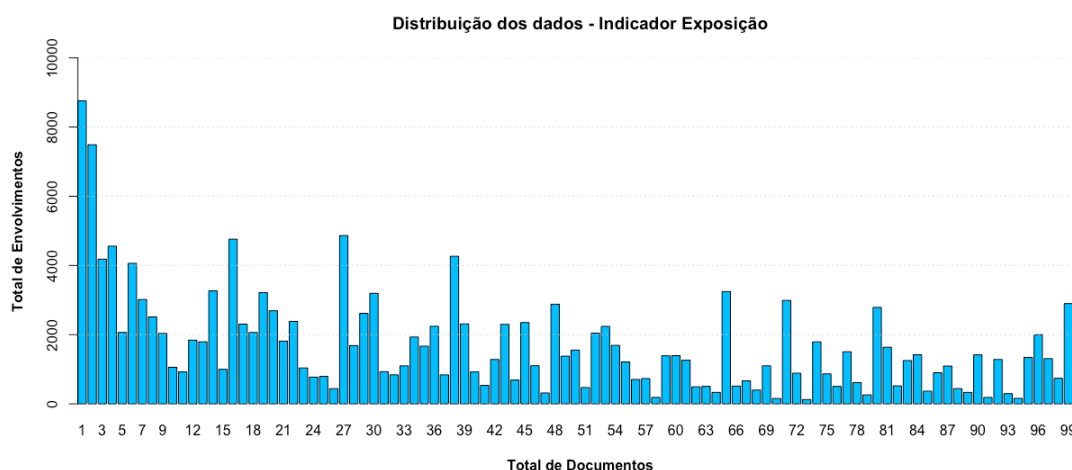
Entre os anos de 2016 e 2017, a ferramenta Altmetric.com deixou de acompanhar o Citeulike e Connotea, essa última não apresentada no quadro acima, pois em ambos os anos não registrou nenhum envolvimento. Mesmo com a exclusão de ambas as redes se observa um crescimento de 25,27% no envolvimento do público com documentos científicos, com movimentação principalmente no Twitter e Youtube que registraram os maiores aumentos no engajamento, enquanto o Mendeley registrou a maior retração.

As movimentações demonstram não gerar impactos nos indicadores, principalmente, pelo fato de terem sido modelados para se adaptarem as oscilações nos envolvimento do público com os conteúdos e por ambas as redes excluídas do monitoramento da ferramenta Altmetric.com não terem registrado expressividade no ano anterior.

Após essa está foi realizado o somatório das redes dentro de cada indicador: Exposição, Interação e Apropriação, conforme pode ser observado no **Erro! Fonte de referência não encontrada.**, Gráfico 14 e Gráfico 16. Em seguida, os dados foram normalizados por logaritmo natural, como forma de ter um melhor ajuste, além de poder identificar possíveis inconsistências.

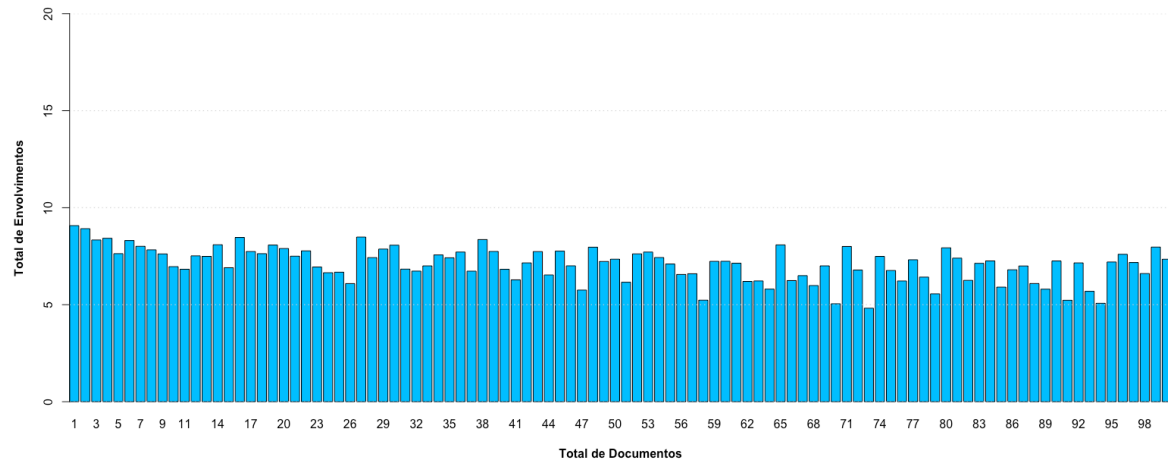
O primeiro ponto observado foi ao comparar o Gráfico 14 (Dados não normalizados) com o Gráfico 15 (Dados normalizados por logaritmo natural) foi que o somatório dos dados no Indicador Exposição, considerado o primeiro nível de envolvimento do público com as publicações científicas teve algumas oscilações no engajamento do público e mesmo com a normalização dos dados essa variação foi mantida. Porém, com menos distorções e com a possibilidade de analisar com maior profundidade os documentos que tiveram melhor desempenho no período.

Gráfico 14 - Somatório dos envolvimento para o Indicador Exposição para cada um dos documentos do dataset de 2017 na ordem do ranking original



Fonte: Dados da pesquisa

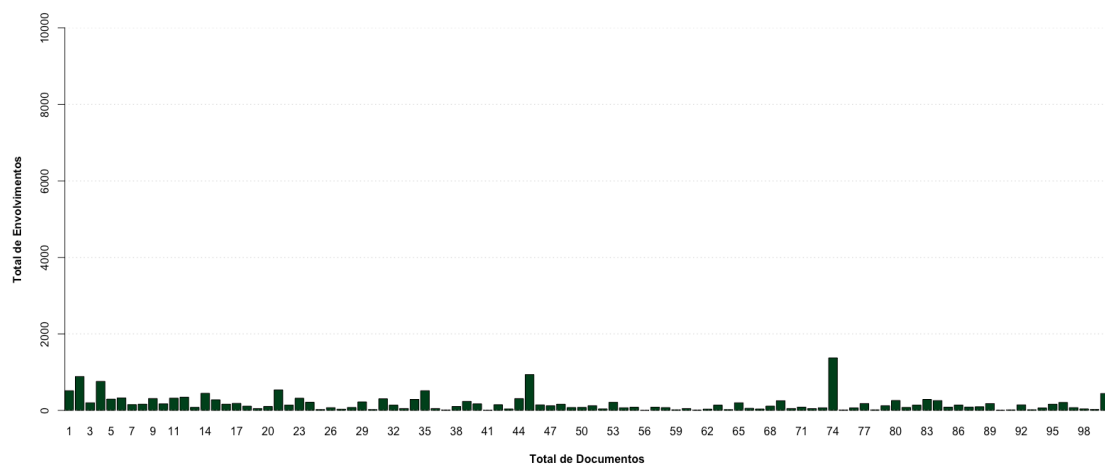
Gráfico 15 - Somatório dos envoltimentos normalizados do Indicador Exposição para cada um dos documentos do dataset de 2017 na ordem do ranking original



Fonte: Dados da pesquisa

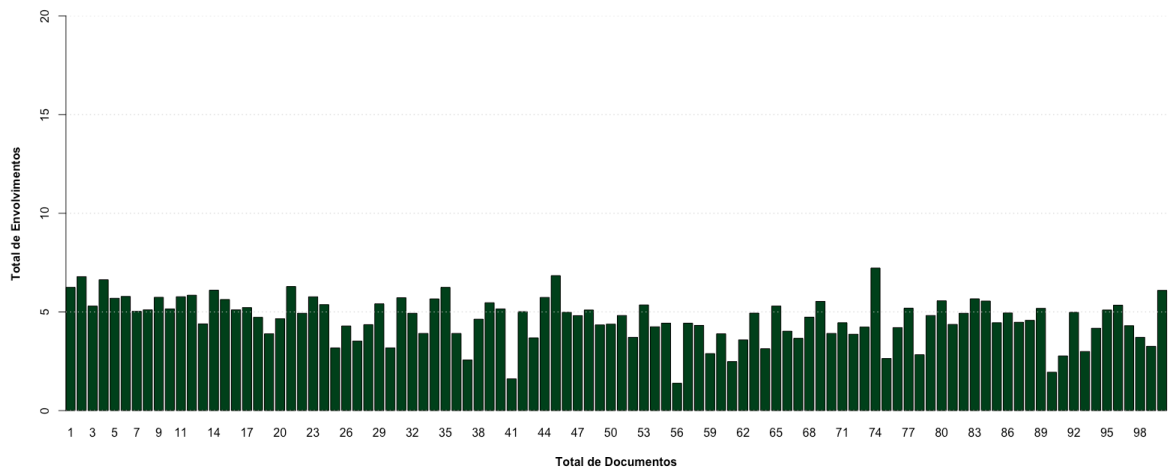
O envolvimento do público em um segundo nível de engajamento com os conteúdos continuou a demonstrar forte retração, como pode ser observado no Gráfico 16 (Dados não normalizados) e Gráfico 17 (Dados normalizados por logaritmo natural). Essa transferência é esperada, ao confrontar os resultados com o modelo AIDA, aponta para uma perda mais acentuada no comparativo com o ano anterior. A principal influência para essa alteração foi o baixo engajamento do público com a ferramenta Mendeley.

Gráfico 16 - Somatório dos envoltimentos para o Indicador Interação para cada um dos documentos do dataset de 2017 na ordem do ranking original



Fonte: Dados da pesquisa

Gráfico 17 - Somatório dos envoltimentos normalizados do Indicador Interação para cada um dos documentos do dataset de 2017 na ordem do ranking original

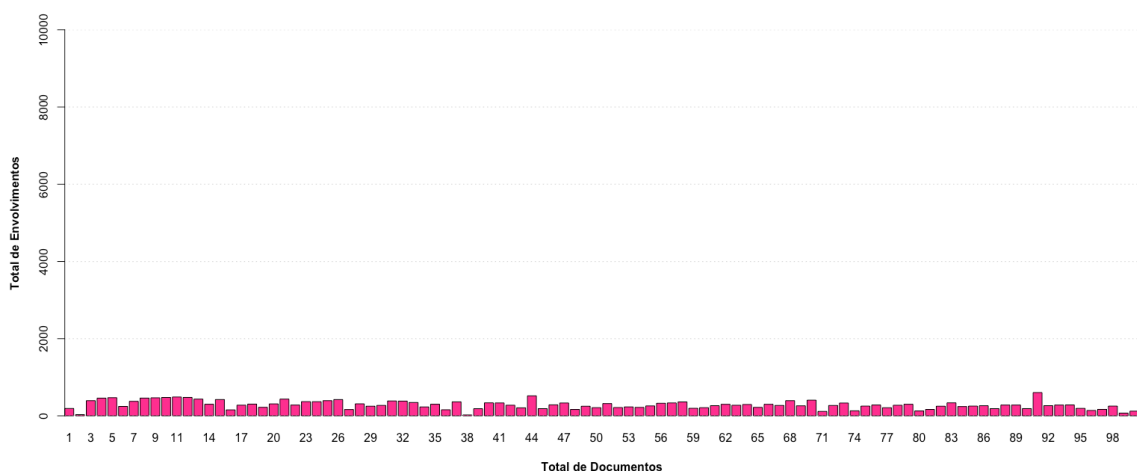


Fonte: Dados da pesquisa

A transferência entre os indicadores demonstrou estabilidade no comparativo com o dataset do ano de 2016 e no Gráfico 18 (Dados não normalizados) e Gráfico 19 (Dados normalizados por logaritmo natural) observa-se que a transferência de público ficou estável nos dois datasets e com transferência média entre o indicador Exposição e Interação de 25%, enquanto no indicador Interação para Apropriação foi de 41%.

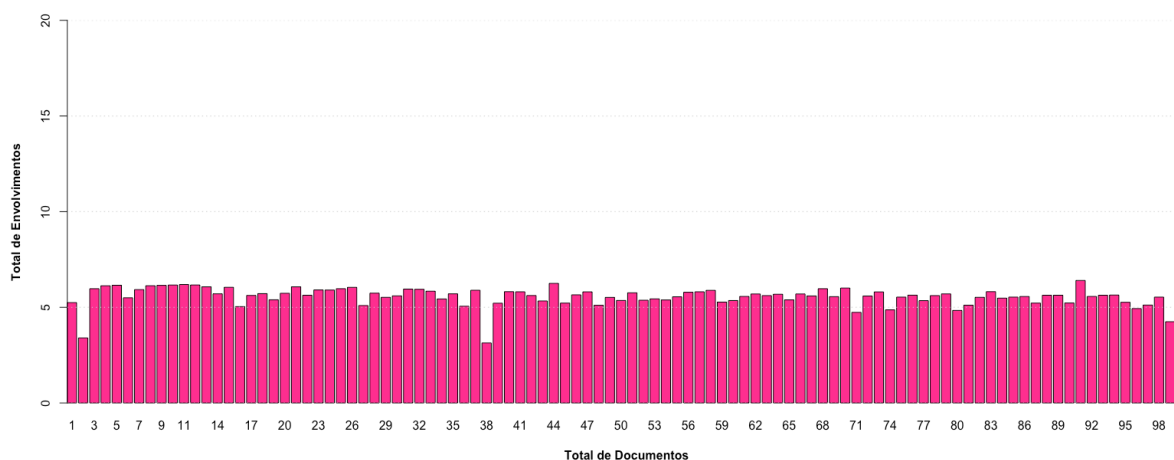
Os dados de transferência de envolvimento demonstram que o maior volume fica entre o público que chega ao segundo Indicador, ou seja, existe maior propensão de um documento científico ser citado se ele for utilizado em uma das redes que compõem o indicador Interação.

Gráfico 18 - Somatório dos envoltimentos para o Indicador Apropriação para cada um dos documentos do dataset de 2017 na ordem do ranking original



Fonte: Dados da pesquisa

Gráfico 19 - Somatório dos envoltimentos normalizados do Indicador Apropriação para cada um dos documentos do dataset de 2017 na ordem do ranking original

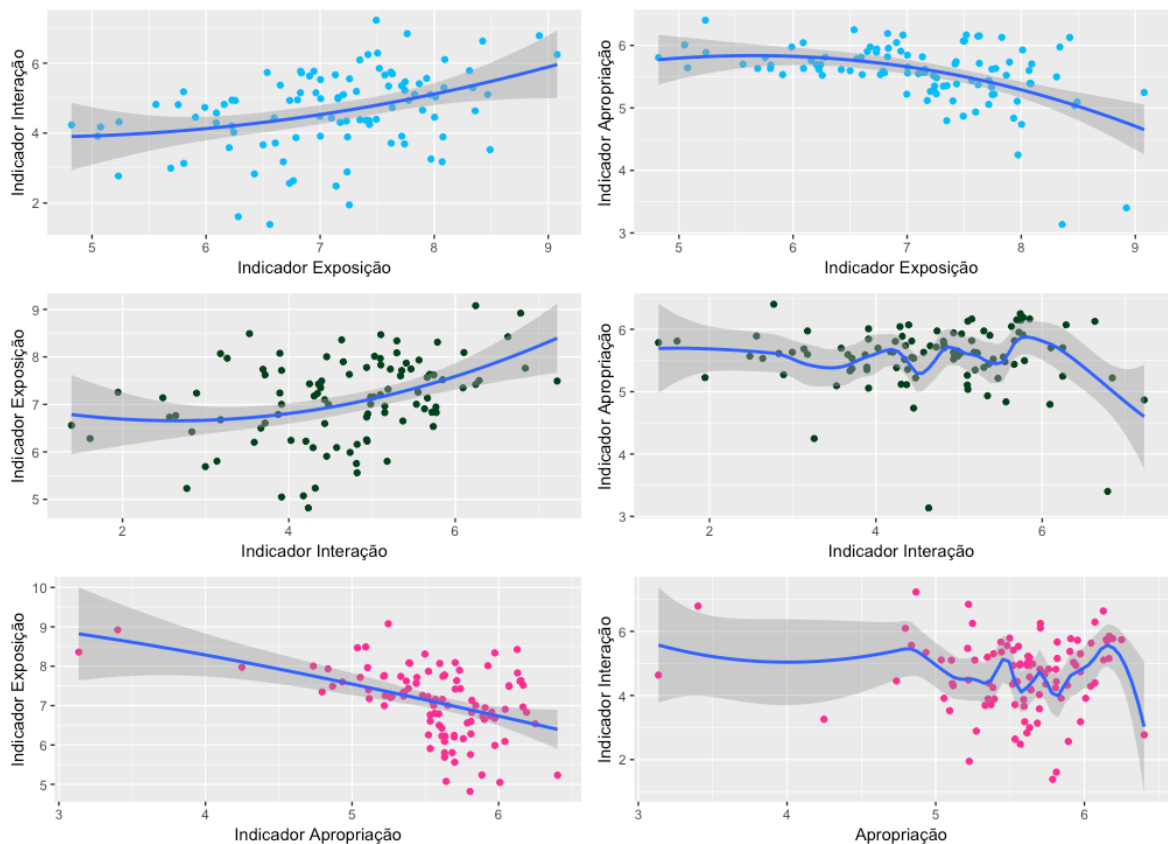


Fonte: Dados da pesquisa

Com base nos gráficos é possível observar que existe um consumo quase linear dos conteúdos entre os indicadores. A exceção é o Indicador Interação que registrou uma forte oscilação no envolvimento do público com as redes que compuseram o indicador. Com essa primeira modelagem observa-se que ambos os datasets (anos de 2016 e 2017) mantiveram como característica o modelo de funil na reverberação no conjunto de informações contidas nos agrupamentos que formaram os indicadores.

Ao observar a ausência de impactos ou distorções em ambos os datasets foi gerado o Gráfico 20, cujo objetivo é avaliar a força da associação entre os indicadores normalizados. Dessa forma, será possível avaliar o nível de transferência de envolvimento entre os indicadores e se realmente é possível inferir que no dataset do ano de 2017 o modelo AIDA também pode ser utilizado.

Gráfico 20 - Distribuição da relação entre os Indicadores com base em dados normalizados por logaritmo natural para o dataset de 2017

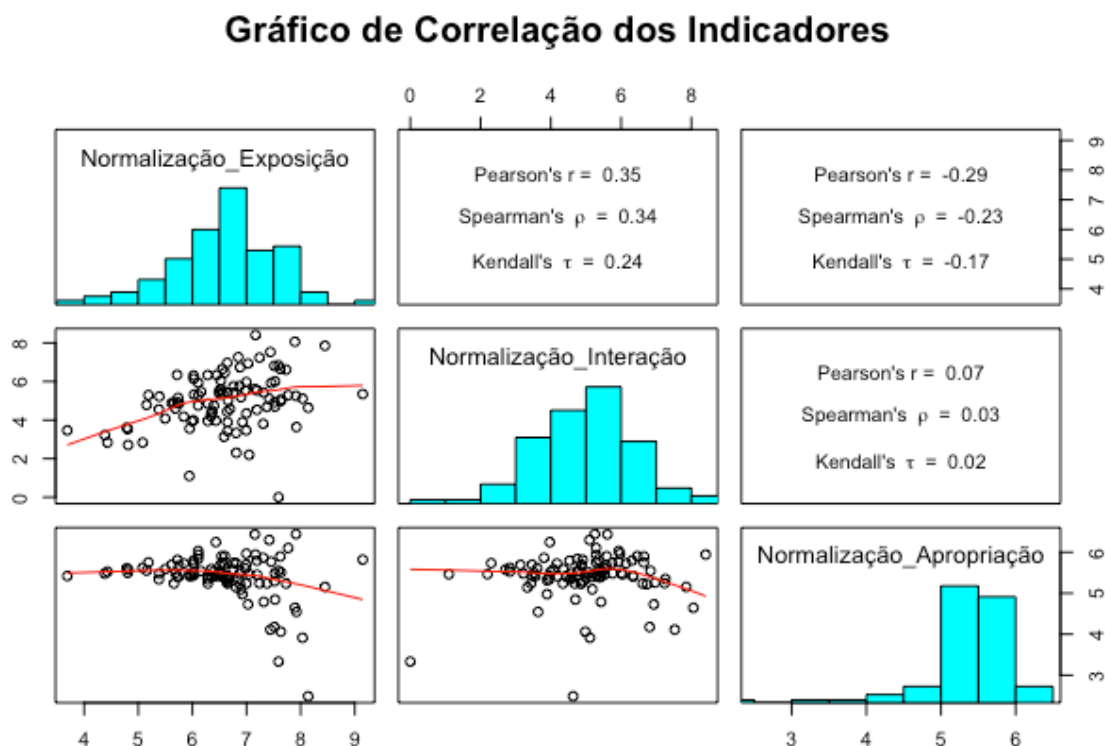


Fonte: Dados da pesquisa

Ao correlacionar os indicadores Exposição e Interação o coeficiente de correlação de Pearson para esses indicadores é de 0,44, o que demonstrou uma leve retração, no comparativo com o ano anterior, de pessoas de pessoas que após serem expostas a uma publicação utilizaram gerenciadores de referências para guardar o material para consulta ou utilização posterior.

A mesma retração foi observada entre os Indicadores normalizados Interação com Apropriação. Nesse caso foi registrado um coeficiente de correlação desprezível (-0,05), o que poderia inviabilizar o modelo de funil. No entanto, o envolvimento do público com o último indicador Interação, também registrou uma queda acentuada. Dessa forma, a transferência existiu e manteve o modelo de funil, constatação comprovada no Gráfico 21, que registra as correlações entre os indicadores.

Gráfico 21 - Histogramas, coeficientes de correlação e diagramas de dispersão de $r =$ Pearson, $\rho =$ Spearman, e $\tau =$ Kendall Tau, entre os três indicadores normalizados: Exposição, Interação e Apropriação para o dataset de 2017



Fonte: Dados da pesquisa

Os indicadores Interação e Apropriação apresentam envolvimento quase linear dos valores, que se distribuem sobre uma linha que descreve uma ligeira tendência de declínio de ambos os indicadores. Esta correlação, é importante ressaltar, mesmo sendo moderada registra uma certa linearidade de ambos os indicadores. Em contrapartida, os indicadores Exposição e Apropriação, utilizando-se como base os “ r ” de Pearson, apresentam comportamento antagônico, ou seja, enquanto cresce o envolvimento com o Indicador Apropriação, na mesma proporção existe a retração no Indicador Exposição.

Mesmo assim, os dados associados as correlações com o Indicador Apropriação registraram que a relação mesmo que moderada apresenta comportamento quase linear, como baixa graduação ascendente ou descendente, como pode ser observado no gráfico de dispersão. Isso permite concluir que ainda que baixa, há alguma correlação do Indicador Apropriação.

Pode-se concluir que os três indicadores demonstram capacidade de representar o tipo de envolvimento e/ou consumo realizado pelo público, nos ambientes digitais, e que mesmo em períodos de tempo distintos, o tipo de relação existente registra baixa variação. Esse tipo de verificação nos permite concluir como estáveis os indicadores e com possibilidades de ser aplicado em outros tipos de *datasets*, como de envolvimento diário, semanal e mensal.

7.3 Construção dos Rankings

A construção do Ranking tem por objetivo apresentar um outro tipo de visualização dos dados resultantes do Crossover. As possibilidades de apresentação podem ser baseadas em qualquer um dos três indicadores, pois cada um atenderia um tipo de público e/ou interesse específico. Assim, o Indicador Exposição poderia ser utilizado para observar quais os tipos de conteúdos estão registrando maior reverberação nas redes sociais, o Indicador Interação apresentaria quais tipos de conteúdos estão registrando maior interesse em ser armazenados para consulta posterior e o Indicador Apropriação apresentaria os conteúdos com maior volume de citações em mídias digitais.

Outra possibilidade de leitura dos rankings é por meio do funil do Modelo AIDA que demonstra apenas uma parte do público seguindo para a etapa seguinte. No marketing tradicional a utilidade desse tipo de ferramenta é para identificar as etapas de perda de interesse do público para um produto ou serviço. O objetivo de aplicar esse tipo de ferramenta no Marketing Científico segue a mesma premissa, mas com o diferencial de cada ponto poder ser trabalhado de forma individual ou servir para finalidades específicas, como a Exposição para Entidades Acadêmicas que tenham interesse em potencializar a divulgação ou comunicação das pesquisas realizadas pela Instituição; a Interação para pesquisadores que tenham um potencial engajamento com os seus trabalhos; e a Apropriação para agências de fomento que tenha o interesse de identificar os produtos de pesquisa que tiveram o maior volume de citações.

Para a realização desta pesquisa a opção pelo ranqueamento foi para o Indicador Apropriação, por encontrar similaridade com as métricas tradicionais e por ser um dos principais focos de interesse de pesquisadores. Os quadros abaixo são compostos por cinco colunas que apresentam os dados:

- **Posição no Ranking com Crossover:** Ordem dos artigos com melhor pontuação no Ranking;
- **Posição no Ranking da Altmetric.com:** O dado serve como comparativo da posição no artigo no Top 100 da ferramenta Altmetric.com;
- **Data da Publicação:** A data do artigo serve para demonstrar o baixo impacto desse indicador no ranqueamento dos documentos;
- **Título:** O nome dos artigos com melhor classificação;
- **Pontuação:** A pontuação que o documento recebeu após o crossover. Esse valor é baseado no indicador Apropriação.

Além de se enquadrar no conceito do crossover, que é oferecer um novo produto, mas com características similares aos produtos tradicionais como forma de evitar a rejeição do público alvo. O modelo de ranqueamento utilizado foi baseado no algoritmo do portal Hacker News, conforme apresentado no capítulo anterior. Os quadros abaixo, registram somente os 10 primeiros ranqueados no Top 100. A listagem completa é apresentada nos Apêndices C (Top 100 – 2016) e D (Top 100 – 2017).

Quadro 8 - Ranking com os 10 primeiros colocados no Top 100 do ano de 2016, tabualizado pelo indicador Apropriação

Posição no Rankink com Crossover	Posição no Ranking da Altmetric.com	Data de Publicação	Título	Pontuação
1	35	25/05/2016	Efficacy and Safety of an Injectable Combination Hormonal Contraceptive for Men	15,7
2	18	02/09/2016	The brain adapts to dishonesty	13,1
3	5	11/07/2016	Sugar Industry and Coronary Heart Disease Research: A Historical Analysis of Internal Industry Documents	12,7
4	88	04/10/2016	National Trends in Hospitalizations for Opioid Poisonings Among Children and Adolescents, 1997 to 2012	11,7
5	46	09/05/2016	1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America	11,7
6	11	01/07/2016	Evidence for a limit to human lifespan	11,5
7	61	13/01/2016	Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes: A Systematic Review and Meta-analysis	11,2
8	8	05/10/2016	Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial	11,0

9	97	01/01/2016	Online social integration is associated with reduced mortality risk	10,9
10	76	26/11/2015	Mutational signatures associated with tobacco smoking in human cancer	10,8

Fonte: Dados da pesquisa

Em ambos os rankings, mesmo com a aplicação do algoritmo de depreciação pelo tempo, que teve o objetivo de evitar distorções com o acúmulo de envolvimento visto que não há nos indicadores brutos algo que ajuste a depreciação da publicação após chegar ao público, foram observadas grandes distorções entre o Ranking proposto pela Altmetric e o Ranking originado pelo Crossover.

Ao avaliar o impacto do tempo de publicação, constatou-se que o fator tempo gerou pouca influência, mas salientou discrepâncias como no Top 100, do ano de 2016. O artigo do Obama (2016) saiu da primeira posição e ficou ranqueado em 24º pelo Ranking do Crossover. É importante salientar que o artigo apresentava um Índice de Apropriação dissonante com sua posição no ranking da Altmetric. Porém, outros documentos também registraram mudanças significativas, como foi o caso do “*National Trends in Hospitalizations for Opioid Poisonings Among Children and Adolescents, 1997 to 2012*” que saltou da posição 88º, do ranking da Altmetric.com, para o quarto lugar no Crossover.

As variações observadas nos rankings foram influenciadas pelo tipo de reverberação que está sendo utilizada para o ranqueamento. Enquanto na ferramenta Altmetric.com essa tarefa é realizada com base em um somatório do total de envolvimento, conforme detalhado no Capítulo 4.3, no Crossover foi utilizado somente o final do funil ou os envolvimento que foram considerados como citações.

Essa forma de gerar o ranqueamento, que pode ser realizada com qualquer um dos indicadores isoladamente, permite separar o que pode ser considerado apenas um engajamento de impacto, como no caso do indicador Exposição e observar com maior exatidão o engajamento que se transformou em uma citação, como no caso do indicador Apropriação utilizado para a construção dos rankings.

Quadro 9 - Ranking com os 10 primeiros colocados no Top 100 do ano de 2017, tabualizado pelo indicador Apropriação

Posição no Rankink com Crossover	Posição no Ranking da Altmetric.com	Data de Publicação	Título	Pontuação
1	30	02/11/2017	Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons	14,7
2	72	01/11/2017	Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species	14,2
3	50	08/11/2017	Regeneration of the entire human epidermis using transgenic stem cells	13,2
4	7	01/10/2017	Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults	12,5
5	91	24/07/2017	Remote detection of widespread indigenous water in lunar pyroclastic deposits	11,6
6	6	18/10/2017	More than 75 percent decline over 27 years in total flying insect biomass in protected areas	10,1
7	85	13/10/2017	Psilocybin for treatment-resistant depression: fMRI-measured brain mechanisms	9,8
8	4	02/08/2017	Correction of a pathogenic gene mutation in human embryos	9,3
9	13	25/07/2017	Clinicopathological Evaluation of Chronic Traumatic Encephalopathy in Players of American Football	8,4
10	26	25/07/2017	Temporal trends in sperm count: a systematic review and meta-regression analysis	8,2

Fonte: Dados da pesquisa

Por outro lado, o mesmo não se refletiu no Ranking de 2017, que teve publicações mais recentes com a maior Apropriação. Os três primeiros lugares foram ocupados por conteúdos lançados semanas antes do fechamento do Ranking da ferramenta Altmetric, isso pode sugerir um aumento no consumo e reprocessamento de pesquisas em outros conteúdos, ou ainda, pode significar apenas um reflexo natural do crescimento das redes e ampliação do número de acadêmicos utilizando estes ambientes para consumir e divulgar as suas produções.

Independente dos motivos, o destaque dos rankings é a identificação da possibilidade de ter Indicadores que apresentem com maior precisão o impacto de publicações científicas e com ranqueamento que pode ser realizado por público distintos e com objetivos específicos.

8 CONSIDERAÇÕES FINAIS

A pesquisa partiu do pressuposto que existe a possibilidade de utilizar a técnica do crossover, por meio de dados extraídos de ambientes digitais, para o desenvolvimento de Indicadores que serviriam de suporte na aferição e ranqueamento da produção científica. A hipótese tem por base o crescimento na utilização de dados das redes sociais por pesquisadores e publicações científicas. Além do surgimento de ferramentas como a Altmetric.com e Plum Analytics, que utilizam peso arbitrário e acabam por promover uma forte distorção na percepção do que é relevante para o público e o nível de disseminação dos documentos científicos.

Nesse sentido foram gerados novos Indicadores (Exposição, Interação e Apropriação cujo foco foi identificar o comportamento, a dispersão e a possibilidade de apropriação de dados de diferentes redes para traduzir uma ação única, em relação a cada artigo ou documento científico.

O primeiro ponto para o prosseguimento do estudo foi identificar pesquisas que pudessem auxiliar no melhor direcionamento do crossover e modelagem das redes, de forma a não gerar distorções de interpretação dos dados. Nessa parte, foi encontrado autores como Haustein, Bowman e Costas (2015) que observam o debate sobre as métricas alternativas sendo baseado em dois argumentos centrais: a crença que esse conjunto de métricas pode servir para antecipar o impacto científico e, ao mesmo tempo, pode refletir apenas barulho ou popularidade. A conclusão dos pesquisadores é que ambos os aspectos podem estar certos, pois os argumentos expostos demonstram que as plataformas consideradas para o estudo podem influenciar em maior ou menor extensão da compreensão do que está sendo estudado e qual é a validade dos resultados para o desenvolvimento do campo.

Em paralelo também foi possível identificar a preocupação da falta de um conjunto concreto de quadros, modelos e teorias, que possam servir de suporte e enquadramento para as interpretações e utilizações destas novas métricas sociais. Neste sentido, foi observado que a maioria das pesquisas utilizam, principalmente, o Twitter e o Mendeley como objetos para estudos de dispersão e blogs para estudos de citação. Em ambos os casos são reforçadas as similaridades entre indicadores tradicionais e os obtidos dos meios digitais, ao invés de se pensar em métricas digitais e tradicionais como ferramentas complementares aos estudos métricos.

Ao obter esse panorama e com a consciência de que os ambientes digitais também permitem a disseminação de forma mais rápida do que é produzido nos ambientes acadêmicos,

e que artigos acabam sofrendo com a rápida obsolescência própria das redes, foi possível aperfeiçoar o escopo do projeto para refinar os Indicadores que seriam extraídos do crossover dos ambientes. Entre os resultados foi identificado fatos de interesse como correlações moderadas entre os Índices. Enquanto a expectativa inicial era que os resultados apresentassem correlação fraca ou desprezível devido a dinâmica de apropriação do conteúdo científico.

A aplicação da técnica de crossover aliada ao Modelo AIDA para a geração dos índices demonstrou ser capaz de entregar três formas distintas e complementares para a leitura do consumo da produção científica. Além de ter possibilitado a geração de um ranking dos documentos, conforme pode ser observado nos Apêndices C e D.

Nos três indicadores observou-se que a Exposição registra como principal característica a rápida disseminação dos conteúdos e tempo médio de circulação e envolvimento do público com o conteúdo foi de sete dias, com poucos pontos discrepantes, mas esses demonstraram não gerar impacto nos indicadores seguintes.

O Indicador Interação foi o que sofreu a maior oscilação entre os dois datasets. No entanto, também foi observado que esse cenário não impediu a análise e trabalho com os dados e demonstraram ser próprio das redes e do tipo de envolvimento do público, o que foi explicado no Capítulo 4 e explica o fenômeno da falta de linearidade entre os três indicadores e principalmente para publicações mais recentes.

o Indicador Apropriação reforçou que a transferência de envolvimento entre os indicadores é presente e que é fortemente influenciada por pesquisas pontuais. Esta serviria inclusive como um reflexo de um possível newsmaking científico que demonstrou privilegiar assuntos que estão em evidência, como as publicações relacionadas a área da Saúde. Ao se utilizar a Teoria do Newsmaking⁴⁶ pode-se observar que este é um assunto que tem alto índice de disseminação e interesse de público e foi responsável por quase 50% dos conteúdos publicados nos dois anos (2016 e 2017).

Além dos indicadores, o ranking serviu para demonstrar que existe a possibilidade de apresentar como são consumidas as publicações científicas (Exposição), qual a proporção do público que está utilizando gerenciadores de referencias para armazenar essas publicações (Interações) e quais conteúdos estão servindo como base para novas publicações (Apropriação).

Com as descobertas realizadas é possível destacar que a proposta e objetivo da pesquisa em gerar um modelo de indicadores utilizando o conceito de Crossover demonstrou-se viável e com capacidade de ser aplicada para acompanhamento de dispersão e engajamento da

⁴⁶ Ver também Mauro Wolf (1999)

produção científica. Os objetivos específicos de definir categorias para enquadrar os conteúdos/canais na mensuração dos ambientes digitais; indicar métricas e métodos para a avaliação alternativa; propor uma metodologia para a modelagem dos dados, e descrever as melhores práticas para a análise e ranqueamento da produção científica por meio da captura de dados digitais foram realizadas com a apresentação dos Indicadores (Exposição, Interação e Apropriação) e com o ranking.

Ao mesmo tempo pode-se apontar como um possível caminho que poderá subsidiar pesquisadores, agências de fomento e publicações com dados sobre a disseminação e/ou impacto das publicações de artigos científicos. Desta forma, se enquadrando como um ponto complementar nos debates sobre modelos e formatos de medição científica nos ambientes digitais.

No entanto, ainda não foi possível habilitar um ambiente estável para que os cálculos fossem realizados automaticamente por qualquer interessado em saber o impacto de sua pesquisa somente com base no DOI. Esse fator pode ser influenciado pelas alterações realizadas nos dados capturados entre os anos.

Essa limitação ainda impede uma ampla utilização dos indicadores para um levantamento diário e acaba por limitar seu uso a pessoas com conhecimento de ferramentas de modelagem de dados como R. Entretanto, com a Folha de Programação apresentada no Apêndice D, espera-se que esse impacto possa ser minimizado e a pesquisa replicada e aperfeiçoada.

Um próximo passo que pode ser desenvolvido como pesquisa complementar é a identificação do que é realmente audiência e repercussão, uma vez que não existe necessariamente uma relação possível de ser medida, sem a utilização dos metadados, entre o que realmente é consumido com o que realmente é utilizado para e na produção científica. Simultaneamente, pode-se aperfeiçoar o algoritmo de Ranking com o objetivo de entregar um sistema de ranqueamento, mais apurado e com baixa interreferência e/ou manipulação dos ambientes de digitais.

REFERÊNCIAS

- ABBASI, A.; CHEN, H.; SALEM, A. Sentiment analysis in multiple languages. **ACM Transactions on Information Systems**, v. 26, n. 3, p. 1–34, 2008.
- ABRAHAM, R. H. Webometry: Measuring the complexity of the world wide web. **World Futures**, v. 50, n. 1–4, p. 785–791, 1997.
- ACKLAND, R. **Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age**. 3. ed. London, United Kingdom: SAGE Publications Ltd, 2013.
- AGICHTEIN, E. et al. Finding high-quality content in social media. *In*: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON WEB SEARCH AND WEB DATA MINING - WSDM '08 2008, New York, USA. **Anais [...]**. New York, USA: ACM Press, 2008.
- ALMIND, T. C.; INGWERSEN, P. Informetric analyses on the world wide web: methodological approaches to 'webometrics'. **Journal of Documentation**, v. 53, n. 4, p. 404–426, 1997.
- ANDERSON, C. **A cauda longa: Do mercado de massa para o mercado de nicho**. 1. ed. : Elsevier, 2006.
- ARAÚJO, R. F. Estudos métricos da informação na web e o papel dos profissionais da informação. **Bibliotecas Universitárias: pesquisas, experiências e perspectivas**, Belo Horizonte, MG, v. 2, p. 42–64, 2015. a.
- ARAÚJO, R. F. Marketing científico digital e métricas alternativas para periódicos: da visibilidade ao engajamento. **Perspectivas em Ciência da Informação**, v. 20, n. 3, p. 67–84, 2015. b.
- ARAÚJO, R. F.; FURNIVAL, A. C. M. Comunicação científica e atenção online: em busca de colégios virtuais que sustentam métricas alternativas. **Informação & Informação**, v. 21, n. 2, 2016.
- ARCHIBALD, M. M.; RADIL, A. I.; ZHANG, X.; HANSON, W. E. Current Mixed Methods Practices in Qualitative Research: A Content Analysis of Leading Journals. **International Journal of Qualitative Methods**, v. 14, n. 2, p. 5–33, 2015.
- ASSOCIAÇÃO INTERNACIONAL PARA MEDIÇÃO E AVALIAÇÃO DA COMUNICAÇÃO (AMEC). Barcelona Declaration of Measurement Principles: Validated Metrics Social Media Measurement. *In*: EUROPEAN SUMMIT ON MEASUREMENT, 2ND 2010, Barcelona, ES. **Anais [...]**. Barcelona, ES: IPR - Institute For Public Relations, 2010. Disponível em: <https://amecorg.com/2012/06/barcelona-declaration-of-measurement-principles/>
- BAR-ILAN, J. Information hub blogs. **Journal of Information Science**, v. 31, n. 4, p. 297–307, 2005.
- BARABASI, A. **Linked - A Nova Ciencia Dos Networks**. São Paulo: Leopardo Editora, 2009.
- BARRY, T. E. The Development of the Hierarchy of Effects: An Historical Perspective. **Current Issues and Research in Advertising**, v. 10, n. 1–2, p. 251–295, 1987.

BARRY, T. E.; HOWARD, D. J. A Review and Critique of the Hierarchy of Effects in Advertising. **International Journal of Advertising**, v. 9, n. 2, p. 121–135, 1990.

BIZZOCCHI, A. Marketing científico: o papel do marketing na difusão da ciência. *In*: XXV CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO 2002, Salvador - BA. **Anais [...]**. Salvador - BA: Intercom, 2002. Disponível em: <http://www.portcom.intercom.org.br/pdfs/20069376494178684556805136387239562824.pdf>

BJÖRK, B.-C. A model of scientific communication as a global distributed information system. **Information Research: an international electronic journal**, p. 48, 2007.

BJÖRNEBORN, L. **Small-world link structures across an academic web space: a library and information science approach**. : Royal School of Library and Information Science, 2004.

BJÖRNEBORN, L.; INGWERSEN, P. Toward a basic framework for webometrics. **Journal of the American Society for Information Science and Technology**, v. 55, n. 14, p. 1216–1227, 2004.

BOLLEN, J.; VAN DE SOMPEL, H.; HAGBERG, A.; CHUTE, R. A Principal Component Analysis of 39 Scientific Impact Measures. **PLoS ONE**, v. 4, n. 6, p. e6022, 2009. Disponível em: <http://dx.plos.org/10.1371/journal.pone.0006022>. Acesso em: 12 jun. 2017.

BOMFÁ, C. R. Z.; FREITAS, M. do C. D.; SILVA, L. de J. O. L. Da; BORNIA, A. C. Marketing Científico Eletrônico: um novo conceito voltado para periódicos eletrônicos. **Estudos em Comunicação**, v. 5, 2009.

BOSSY, M. J. The Last of the Litter : Netometrics. 1995. Disponível em: <http://gabriel.gallezot.free.fr/Solaris/d02/2bossy.html>. Acesso em: 22 abr. 2017.

BOYD, D. M.; ELLISON, N. B. Social Network Sites: Definition, History, and Scholarship. **Journal of Computer-Mediated Communication**, v. 13, n. 1, p. 210–230, 2007.

CHAKRABARTI, S.; JOSHI, M. M.; PUNERA, K.; PENNOCK, D. M. The structure of broad topics on the web. *In*: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB - WWW '02 2002, New York, USA. **Anais [...]**. New York, USA: ACM Press, 2002.

CHILUKA, N.; ANDRADE, N.; POUWELSE, J. A Link Prediction Approach to Recommendations in Large-Scale User-Generated Content Systems. *In*: : Springer, Berlin, Heidelberg, 2011. p. 189–200.

COLIC, I. **Afrographique (An infographic depicting a short history of social networks)**. 2012. Disponível em: <https://afrographique.tumblr.com/post/28003426619/an-infographic-depicting-a-short-history-of-social>. Acesso em: 15 jan. 2018.

COLLEDGE, L. **Snowball Metrics Recipe Book**, Amsterdam: Snowball Metrics Program Partners, 2014. Disponível em: <https://www.snowballmetrics.com/wp-content/uploads/snowball-metrics-recipe-book-upd.pdf>. Acesso em: 7 jun. 2018.

COSENZA, V. **World map of social networks**. 2018. Disponível em: https://vincos.it/wp-content/uploads/2018/02/WMSN0118_1029.png. Acesso em: 2 dez. 2018.

COVER, T. M.; THOMAS, J. A. **Elements of information theory**. 2. ed. : John Wiley &

- Sons, Inc., 1991. Disponível em: http://www.cs-114.org/wp-content/uploads/2015/01/Elements_of_Information_Theory_Elements.pdf. Acesso em: 4 abr. 2017.
- DEGENNE, A.; FORSÉ, M. **Introducing social networks**. : SAGE Publications Ltd, 1999.
- DELLA PORTA, D.; KEATING, M. **Approaches and methodologies in the social sciences : a pluralist perspective**. : Cambridge University Press, 2008.
- DEPARTMENT FOR EMPLOYMENT AND LEARNING. **Decisions on assessing research impact**. : Higher Education Funding Council for England, 2011. Disponível em: https://dera.ioe.ac.uk/10002/1/01_11.pdf. Acesso em: 11 mar. 2017.
- DHYANI, D.; NG, W. K.; BHOWMICK, S. S. A survey of Web metrics. **ACM Computing Surveys**, v. 34, n. 4, p. 469–503, 2002.
- DOURISH, P. The Appropriation of Interactive Technologies: Some Lessons from Placeless Documents. **Computer Supported Cooperative Work (CSCW)**, v. 12, n. 4, p. 465–490, 2003. Disponível em: <http://link.springer.com/10.1023/A:1026149119426>. Acesso em: 13 jan. 2017.
- ERLANDSON, R. J.; ERB, R. A. **Technology for small and one-person libraries : a LITA guide**. : ALA TechSource, 2013.
- FERREIRA, A. G. C.; CAREGNATO, S. E. Visibilidade de revistas científicas: um estudo no Portal de Periódicos Científicos da Universidade Federal do Rio Grande do Sul. **Transinformação**, v. 26, n. 2, p. 177–190, 2014.
- FISHBEIN, M. **Readings in attitude theory and measurement**. : John Wiley & Sons Inc, 1967.
- FRAGOSO, S.; RECUERO, R.; AMARAL, A. **Métodos de pesquisa para internet**. : Editora Sulina, 2011.
- GARFIELD, E. **Citation Indexes - New Paths to Scientific Knowledge**. 1956. Disponível em: <http://www.garfield.library.upenn.edu/papers/31.html>. Acesso em: 24 abr. 2018.
- GARFIELD, E.; WELLJAMS-DOROF, A. Of Nobel class: A citation perspective on high impact research authors. **Theoretical Medicine**, p. 117–135, 1992. Disponível em: <http://link.springer.com/10.1007/BF02163625>
- GAROFALO, R. Culture versus commerce: the marketing of black popular music. **Public Culture**, v. 7, n. 1, p. 275–287, 1994.
- GIGLIO, F. D. Implicações dos meios digitais na comunicação de marketing. *In*: 2010, São Paulo. **Anais [...]**. São Paulo: XIII Semead - Seminários em Administração, 2010. Disponível em: <http://sistema.semead.com.br/13semead/resultado/trabalhosPDF/167.pdf>
- GILLIAM, A. A ideologia do crossover e a sua relação com o gênero. **Cadernos Pagu**, São Paulo, p. 227–240, 1996. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cadpagu/article/view/1869>. Acesso em: 11 jul. 2016.
- GOUVEIA, F. C. Altmetria: métricas de produção científica para além das citações. **Liinc em**

- Revista**, v. 9, n. 1, p. 214–227, 2013. Disponível em: <http://revista.ibict.br/liinc/article/view/3434>. Acesso em: 30 nov. 2016.
- GOUVEIA, F. C. A altmetria e a interface entre a ciência e a sociedade. **Trabalho, Educação e Saúde**, v. 14, p. 643–645, 2016.
- GRAHAM, P. **What I've Learned from Hacker News**. 2009. Disponível em: <http://www.paulgraham.com/hackernews.html>. Acesso em: 23 out. 2017.
- GRIER, S. A.; BRUMBAUGH, A. M.; THORNTON, C. G. Crossover Dreams: Consumer Responses to Ethnic-Oriented Products. **Journal of Marketing**, v. 70, n. 2, p. 35–51, 2006.
- GROSS, P. L.; GROSS, E. M. College libraries and chemical education. **Science (New York, N.Y.)**, v. 66, n. 1713, p. 385–9, 1927.
- HAIR, J. F. et al. **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009.
- HALFPENNY, P.; PROCTER, R. **Innovations in digital research methods**. London, United Kingdom: SAGE Publications Ltd, 2015.
- HAUSTEIN, S. Grand challenges in altmetrics: heterogeneity, data quality and dependencies. **Scientometrics**, v. 108, n. 1, p. 413–423, 2016. Disponível em: <http://link.springer.com/10.1007/s11192-016-1910-9>. Acesso em: 14 jan. 2018.
- HAUSTEIN, S.; BOWMAN, T. D.; COSTAS, R. Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In: **Theories of Informetrics: A Festschrift in Honor of Blaise Cronin**. Berlin: Cassidy R. Sugimoto (Ed.), 2015. p. 372–405.
- HOAGLIN, D. C. (David C.); MOSTELLER, F.; TUKEY, J. W. (John W. **Understanding robust and exploratory data analysis**. New York: John Wiley & Sons, 1983.
- HOFFMANN, C. P.; LUTZ, C.; MECKEL, M. A relational altmetric? Network centrality on ResearchGate as an indicator of scientific impact. **Journal of the Association for Information Science and Technology**, v. 67, n. 4, p. 765–775, 2016.
- HSU, Y.; HSU, W. J. The Impact of Crossover Clothing Brand on Consumer Purchase Intention. **International Review of Management and Business Research**, v. 6, n. 1, p. 105–121, 2017.
- HULL, G.; LIPFORD, H. R.; LATULIPE, C. Contextual gaps: privacy issues on Facebook. **Ethics and Information Technology**, v. 13, n. 4, p. 289–302, 2011. Disponível em: <http://link.springer.com/10.1007/s10676-010-9224-8>. Acesso em: 11 out. 2017.
- INGWERSEN, P. The calculation of web impact factors. **Journal of Documentation**, v. 54, n. 2, p. 236–243, 1998.
- ITO, L. de L. **Músicos Independentes na Internet: Novas Lógicas de Consagração Artística**. Curitiba: APPRIS, 2017.
- JANA, S.; CHATTERJEE, S. Quantifying Web-site visits using Web statistics: an extended cybermetrics study. **Online Information Review**, v. 28, n. 3, p. 191–199, 2004. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/14684520410543634>. Acesso em: 5 dez. 2016.

JEFFREY, A. Social media measurement: a step by Step Approach Using the amec valid Metrics Framework For the. *In: 2013, Anais [...]. : 16TH International Public Relations Research Conference, 2013. Disponível em: <https://instituteforpr.org/wp-content/uploads/Social-Media-Measurement-Paper-Jeffrey-6-4-13.pdf>. Acesso em: 12 abr. 2016.*

JI, X. F.; SHEN, H. N. The research of “crossover” marketing strategy. *In: 19TH INTERNATIONAL CONFERENCE ON INDUSTRIAL ENGINEERING AND ENGINEERING MANAGEMENT: ENGINEERING ECONOMICS MANAGEMENT 2013, Berlin, Heidelberg. Anais [...]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. Disponível em: http://link.springer.com/10.1007/978-3-642-38442-4_143. Acesso em: 16 maio. 2016.*

KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! The challenges and opportunities of Social Media. **Business Horizons**, v. 53, n. 1, p. 59–68, 2010.

KELLEHER, T.; MILLER, B. M. Organizational Blogs and the Human Voice: Relational Strategies and Relational Outcomes. **Journal of Computer-Mediated Communication**, v. 11, n. 2, p. 395–414, 2006. Disponível em: <https://academic.oup.com/jcmc/article/11/2/395-414/4617725>. Acesso em: 2 ago. 2016.

KENNEDY, H. **Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary**. 1. ed. : Palgrave Macmillan, 2016.

KOTLER, P. **Administração de marketing : análise, planejamento, implementação e controle**. 5. ed. São Paulo: Atlas, 1998.

KOTLER, P.; KELLER, K. L. **Marketing management**. 14. ed. Boston: Prentice Hall, 2011.

LANG, P. B.; GOUVEIA, F. C.; LETA, J. Relações Intra-Institucionais na Internet: um estudo exploratório com base em metodologias webométricas. **Perspectivas em Ciência da Informação**, v. 13, n. 3, p. 137–150, 2008.

LARIVIERE, V.; SUGIMOTO, C. R. The Journal Impact Factor: A brief history, critique, and discussion of adverse effects. **arXiv** , 2018. Disponível em: <http://arxiv.org/abs/1801.08992>

LARSON, R. R. Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. **Proceedings of the ASIS Annual Meeting**, v. 33, p. 71–78, 1996.

LIN, J.; FENNER, M. The many faces of article-level metrics. **Bulletin of the American Society for Information Science and Technology**, v. 39, n. 4, p. 27–30, 2013.

LINDENMANN, W. K. An effectiveness yardstick’ to measure public relations success. **Public Relations Quarterly**, v. 38, p. 7–9, 1993. Disponível em: <http://eportfolio.lib.ksu.edu.tw/user/T/0/T093000279/repository/公關效果評估5.pdf>

LISÉE, C.; LARIVIÈRE, V.; ARCHAMBAULT, É. Conference proceedings as a source of scientific information: A bibliometric analysis. **Journal of the American Society for Information Science and Technology**, v. 59, n. 11, p. 1776–1784, 2008.

LOACH, T. V.; EVANS, T. S. Ranking Journals Using Altmetrics. 2015. Disponível em: <http://arxiv.org/abs/1507.00451>. Acesso em: 4 maio. 2017.

MACNAMARA, J. Research in Public Relations: A review of the use of evaluation and formative research. **Asia Pacific Public Relations Journal**, v. 1, p. 107–133, 1999.

MADRIGAL, A. C. **Dark Social: We Have the Whole History of the Web Wrong - The Atlantic**. 2012.

MALEKI, A. PubMed and ArXiv vs. Gold Open Access: Citation, Mendeley, and Twitter Uptake of Academic Articles of Iran. **ISSI**, 2015.

MARICATO, J. D. M.; MARTINS, D. L. Almetria: complexidades, desafios e novas formas de mensuração e compreensão da comunicação científica na web social. **Biblios: Journal of Librarianship and Information Science**, n. 68, p. 48–68, 2017. Disponível em: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000300004&lng=es&nrm=iso&tlng=pt

MARTELETO, R. M. **Redes Sociais, Mediação e Apropriação de Informações: situando campos, objetos e conceitos na pesquisa em Ciência da Informação**. João Pessoa: Tendências da Pesquisa Brasileira em Ciência da Informação, 2010. v. 5

MAZOV, N. A.; GUREEV, V. N. Alternative approaches to assessing scientific results. **Herald of the Russian Academy of Sciences**, v. 85, n. 1, p. 26–32, 2015.

MOHAMMADI, E.; THELWALL, M.; KOUSHA, K. Can Mendeley bookmarks reflect readership? A survey of user motivations. **Journal of the Association for Information Science and Technology**, v. 67, n. 5, p. 1198–1209, 2016.

MONASTERSKY, R. The Number That's Devouring Science. **The Chronicle of Higher Education**, v. 52, n. 8, 2005. Disponível em: <https://www.chronicle.com/article/The-Number-Thats-Devouring/26481>

NORUZI, A. Fundamental Differences between Hyperlinks and Citations. **Webology**, v. 2, n. 2, 2005. Disponível em: <http://www.webology.org/2005/v2n2/editorial4.html>

NWAGWU, W. E.; ONYANCHA, B. Back to the Beginning — The Journal is Dead, Long Live Science. **The Journal of Academic Librarianship**, v. 41, n. 5, p. 669–679, 2015. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0099133315001056>. Acesso em: 1 maio. 2019.

O'REILLY, T. **What is web 2.0**. : O'Reilly Media, Inc., 2009. Disponível em: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>

OBAMA, B. United States Health Care Reform: Progress to Date and Next Steps. **JAMA**, v. 316, n. 5, p. 525–532, 2016.

OLIVEIRA, E. da C. P. Comunicação científica e redes sociais. In: ALBAGLI, S. (Org. . (Ed.). **Fronteiras da Ciência da Informação**. Brasília. Brasília: IBICT, 2013. p. 196–216.

OLIVER, R. L.; BEARDEN, W. O. Crossover Effects in the Theory of Reasoned Action: A Moderating Influence Attempt. **Journal of Consumer Research**, v. 12, n. 3, p. 324–340, 1985.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. *In: PROCEEDINGS OF THE ACL-02 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING - EMNLP '02 2002*, Morristown, NJ, USA.

Anais [...]. Morristown, NJ, USA: Association for Computational Linguistics, 2002.
Disponível em: <http://portal.acm.org/citation.cfm?doid=1118693.1118704>. Acesso em: 18 nov. 2016.

PATTERSON, D.; SNYDER, L.; ULLMAN, J. Best practices memo: evaluating computer scientists and engineers for promotion and tenure. **Computing Research News**, p. A-B, 1999.

PIWOWAR, H. Introduction altmetrics: What, why and where? **Bulletin of the American Society for Information Science and Technology**, v. 39, n. 4, p. 8–9, 2013. Disponível em: <http://doi.wiley.com/10.1002/bult.2013.1720390404>. Acesso em: 24 jun. 2016.

PRADHAN, P.; DORA, M. Altmetrics: An Alternative View-Point to Assess Scholarly Research Impact. **International Journal of Information Dissemination and Technology (IJIDT)**, v. 5, n. 2, p. 123–130, 2015.

PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C. **Altmetrics: a manifesto**. 2010. Disponível em: <http://altmetrics.org/manifesto/>. Acesso em: 18 maio. 2016.

PRIEM, J. Altmetrics (Chapter from Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact). 2015. Disponível em: <http://arxiv.org/abs/1507.01328>. Acesso em: 21 set. 2018.

PRIMO, A. Blogs e seus gêneros: avaliação estatística de 50 blogs em língua portuguesa. **MATRIZES**, v. 4, n. 1, p. 129–147, 2010.

RAMALHO, J. A. **Mídias sociais na prática**. 1. ed. : Elsevier, 2010.

RAMSDEN, J. J. Impact factors-a critique. **The Journal of Biological Physics and Chemistry**, v. 9, p. 139–140, 2009.

RAO, A. R.; QU, L.; RUEKERT, R. W. Signaling Unobservable Product Quality through a Brand Ally. **Journal of Marketing Research**, v. 36, n. 2, p. 258, 1999.

REINACH, F. Darwin e a prática da “Salami Science”. **Revista de Ciências Médicas e Biológicas**, v. 12, n. 4, p. 402–403, 2013.

RODRÍGUEZ I GAIRÍN, J. M. Valoración del impacto de la información en Internet: Altavista, el “Citation Index” de la red. **Revista española de Documentación Científica**, v. 20, n. 2, p. 175–181, 1997.

SCHMIDT, J. Blogging Practices: An Analytical Framework. **Journal of Computer-Mediated Communication**, v. 12, n. 4, p. 1409–1427, 2007.

SETZER, V. W. Dado, Informação, Conhecimento e Competência. **DataGramZero - Revista de Ciência da Informação**, 1999. Disponível em: <https://www.ime.usp.br/~vwsetzer/datagrama.html>. Acesso em: 11 out. 2016.

SHIMP, T. A. **Propaganda e promoção : aspectos complementares da comunicação integrada de marketing**. 5. ed. : Bookman, 2002.

SHIRI, A. A. Cybermetrics: a new horizon in information research. **Cybermetrics**, v. 50, p. 111, 1998.

SIQUEIRA, J. C. **Glossário de Neologismos da Ciência da Informação**, por Jéssica

Câmara Siqueira - Clube de Autores. 1. ed. : Clube de Autores, 2016.

SMITH, A. G. Web links as analogues of citations. **Information Research**, v. 9, n. 4, p. 4–9, 2004.

SOUZA, R. F. De; STUMPF, I. R. C. Ciência da Informação como área do conhecimento: abordagem no contexto da pesquisa e da Pós-Graduação no Brasil. **Perspectivas em Ciência da Informação**, v. 14, n. 3, p. 41–58, 2009.

STERNE, J. **Social media metrics : how to measure and optimize your marketing investment.** 1. ed. : John Wiley, 2010.

THE CONCLAVE. The standards for Influence, Opinion & Advocacy. *In: THE CONCLAVE ON SOCIAL MEDIA MEASUREMENT STANDARDS 2013*, **Anais [...]**.

THELWALL, M. Introduction. *In: MIKE THELWALL (Ed.). Link Analysis: An Information Science Approach.* 1. ed. San Diego, USA: Academic Press, 2004. p. 1–8.

THELWALL, M. Interpreting social science link analysis research: A theoretical framework. **Journal of the American Society for Information Science and Technology**, v. 57, n. 1, p. 60–68, 2006.

THELWALL, M. A comparison of link and URL citation counting. **Aslib Proceedings**, v. 63, n. 4, p. 419–425, 2011.

THELWALL, M.; HAUSTEIN, S.; LARIVIÈRE, V.; SUGIMOTO, C. R. Do Altmetrics Work? Twitter and Ten Other Social Web Services. **PLoS ONE**, v. 8, n. 5, p. e64841, 2013. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0064841>. Acesso em: 16 maio. 2018.

THELWALL, M.; VAUGHAN, L.; BJÖRNEBORN, L. Webometrics. **Annual Review of Information Science and Technology**, v. 39, n. 1, p. 81–135, 2006.

THYER, B. A. **Preparing Research Articles.** : Oxford University Press, 2008.

VAKRATSAS, D.; AMBLER, T. How Advertising Works: What Do We Really Know? **Journal of Marketing**, v. 63, n. 1, p. 26, 1999.

VANTI, N. Os links e os estudos webométricos. *In: Ciência da Informação.* : Instituto Brasileiro de Informação em Ciência e Tecnologia, 2005. v. 34.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v. 31, n. 2, p. 369–379, 2002.

VANTI, N.; COSTA, J. A. F.; SILVA, I. C. O. Da Nova fórmula revisada para o cálculo do fator de impacto web (FIW) | New revised formula for calculating Web Impact Factor. **Liinc em Revista**, v. 9, n. 1, 2013.

VICKERY, G.; WUNSCH-VINCENT, S. **Participative Web and user-created content : Web 2.0, wikis and social networking.** França: Organisation for Economic Co-operation and Development, 2007.

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications.** : Cambridge University Press, 1994.

WIJAYA, B. S. The Development of Hierarchy of Effects Model in Advertising. **International Research Journal of Business Studies**, v. 5, n. 1, p. 73–85, 2012.

WOLF, M.; FIGUEIREDO, M. J. V. de. **Teorias da comunicação**. 5 ed ed. Lisboa: Presenca, 1999.

YANG, P. Who flashes new profit growth now? **Shanghai Economy**, 2009.

ZAR, J. H. **Biostatistical analysis**. 5. ed. : Prentice-Hall/Pearson, 2010.

ZIMAN, J. M. **Conhecimento Público**. : Itatiaia, 1979.

APÊNDICE A – CROSSOVER DO TOP 100 DO ANO DE 2016

Dados Crossover								
Dados das Publicações						Indicadores		
Ranking	Data de Publicação	Título	Periódico	Tipo de conteúdo	Área	Exposição	Interação	Apropriação
1	27/10/2016	Efficacy and Safety of an Injectable Combination Hormonal Contraceptive for Men	The Journal of Clinical Endocrinology & Metabolism	Early Release Article	Medical & Health Sciences	16,5	5,0	15,7
2	24/10/2016	The brain adapts to dishonesty	Nature Neuroscience	Article	Medical & Health Sciences	46,7	11,8	13,1
3	12/09/2016	Sugar Industry and Coronary Heart Disease Research: A Historical Analysis of Internal Industry Documents	JAMA Internal Medicine	Special Communication	Research & Reproducibility	68,4	5,6	12,7
4	31/10/2016	National Trends in Hospitalizations for Opioid Poisonings Among Children and Adolescents, 1997 to 2012	JAMA Pediatrics	Original investigation	Medical & Health Sciences	17,0	1,8	11,7
5	26/10/2016	1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America	Nature	Letter	Medical & Health Sciences	38,2	10,7	11,7
6	05/10/2016	Evidence for a limit to human lifespan	Nature	Letter	Medical & Health Sciences	45,3	3,2	11,5
7	31/10/2016	Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes: A Systematic Review and Meta-analysis	JAMA Pediatrics	Original Investigation	Medical & Health Sciences	24,3	3,3	11,2

8	20/09/2016	Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial	JAMA	Original Investigation	Medical & Health Sciences	31,7	11,7	11,0
9	31/10/2016	Online social integration is associated with reduced mortality risk	PNAS	Article	Studies in Human Society	21,4	2,8	10,9
10	04/11/2016	Mutational signatures associated with tobacco smoking in human cancer	Science	Report	Medical & Health Sciences	43,9	2,7	10,8
11	14/10/2016	The Evolution of Galaxy Number Density at $z < 8$ and its Implications	The Astrophysical Journal	Article	Physical Sciences	4,6	0,5	10,2
12	30/08/2016	Neural mechanisms for lexical processing in dogs	Science	Report	Biological Science	10,6	4,9	9,0
13	03/05/2016	Medical error—the third leading cause of death in the US	British Medical Journal	Analysis	Studies in Human Society	39,2	2,6	8,8
14	14/09/2016	10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer	New England Journal of Medicine	Article	Medical & Health Sciences	18,7	8,8	8,6
15	27/09/2016	The terrorist inside my husband's brain	Neurology	Editorial	Medical & Health Sciences	54,6	12,4	8,6
16	28/09/2016	Association of Hormonal Contraception With Depression	JAMA Psychiatry	Original Investigation	Medical & Health Sciences	23,0	6,7	8,4

17	31/08/2016	The antibody aducanumab reduces A β plaques in Alzheimer's disease	Nature	Article	Medical & Health Sciences	36,0	1,1	8,2
18	01/10/2016	Validation of a Functional Pyelocalyceal Renal Model for the Evaluation of Renal Calculi Passage While Riding a Roller Coaster	Journal of the American Osteopathic Association	Article	Medical & Health Sciences	4,8	0,3	8,1
19	04/10/2016	Acne and telomere length. A new spectrum between senescence and apoptosis pathways	Journal of Investigative Dermatology	Article	Medical & Health Sciences	0,1	0,4	7,7
20	02/09/2016	Longer Contact Times Increase Cross-Contamination of Enterobacter aerogenes from Surfaces to Food	Applied & Environmental Microbiology	Article	Biological Science	26,6	0,9	7,6
21	26/05/2016	Report of Partial findings from the National Toxicology Program Carcinogenesis Studies of Cell Phone Radiofrequency Radiation in Hsd: Sprague Dawley® SD rats (Whole Body Exposure)	bioRxiv	Pre-print	Medical & Health Sciences	10,0	0,8	7,3
22	31/08/2016	Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures	Nature	Letter	Earth & Environmental Sciences	16,1	6,5	7,0
23	13/04/2016	Zika Virus and Birth Defects — Reviewing the Evidence for Causality	New England Journal of Medicine	Special Report	Medical & Health Sciences	21,0	1,9	6,7
24	11/07/2016	United States Health Care Reform: Progress to Date and Next Steps	JAMA	Special Communication	Studies in Human Society	164,9	3,8	6,6

25	01/07/2016	Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women	The Lancet	Article	Medical & Health Sciences	12,2	4,4	6,1
26	08/09/2016	Multi-locus Analyses Reveal Four Giraffe Species Instead of One	Current Biology	Report	Biological Science	17,0	6,0	6,1
27	01/02/2016	Evidence for a Distant Giant Planet in the Solar System	The Astronomical Journal	Article	Physical Sciences	14,2	2,8	6,1
28	27/07/2016	Human commensals producing a novel antibiotic impair pathogen colonization	Nature	Article	Biological Science	18,2	0,1	6,0
29	24/08/2016	A terrestrial planet candidate in a temperate orbit around Proxima Centauri	Nature	Letter	Physical Sciences	32,5	5,6	5,8
30	08/09/2016	Catastrophic Declines in Wilderness Areas Undermine Global Environment Targets	Current Biology	Report	Earth & Environmental Sciences	18,4	5,7	5,7
31	01/06/2016	The new world atlas of artificial night sky brightness	Science Advances	Article	Physical Sciences	19,8	67,1	5,6
32	29/08/2016	Perimortem fractures in Lucy suggest mortality from fall out of tall tree	Nature	Article	History & Archaeology	11,6	2,1	5,4

33	27/07/2016	Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa	South African Journal of Science	Article	History & Archaeology	2,5	0,7	5,4
34	30/03/2016	Contribution of Antarctica to past and future sea-level rise	Nature	Article	Earth & Environmental Sciences	4,8	4,5	5,1
35	30/06/2016	Emergence of healing in the Antarctic ozone layer	Science	Article	Earth & Environmental Sciences	4,6	2,2	5,1
36	12/08/2016	Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (<i>Somniosus microcephalus</i>)	Science	Report	Biological Science	40,9	3,0	5,0
37	10/04/2016	The Association Between Income and Life Expectancy in the United States, 2001-2014	JAMA	Special Communication	Studies in Human Society	25,2	9,6	5,0
38	11/08/2016	Long-Term Training with a Brain-Machine Interface-Based Gait Protocol Induces Partial Neurological Recovery in Paraplegic Patients	Scientific Reports	Article	Medical & Health Sciences	8,1	0,0	4,8
39	23/08/2016	The High Cost of Prescription Drugs in the United States: Origins and Prospects for Reform	JAMA	Special Communication	Studies in Human Society	39,0	3,0	4,7

40	23/06/2016	Transient Smartphone “Blindness”	New England Journal of Medicine	Correspondence	Medical & Health Sciences	10,7	0,4	4,7
41	26/05/2016	Escherichia coli Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First report of mcr-1 in the USA	Antimicrobial Agents and Chemotherapy	Letter	Biological Science	5,0	1,0	4,6
42	29/07/2016	A genetic assessment of the English bulldog	Canine Genetics and Epidemiology	Article	Biological Science	1,5	0,5	4,6
43	28/07/2016	Apollo Lunar Astronauts Show Higher Cardiovascular Disease Mortality: Possible Deep Space Radiation Effects on the Vascular Endothelium	Scientific Reports	Article	Medical & Health Sciences	7,3	1,0	4,4
44	01/08/2016	Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality	JAMA Internal Medicine	Original Investigation	Medical & Health Sciences	18,7	3,1	4,2
45	09/05/2016	An elastic second skin	Nature Materials	Article	Materials Science	2,0	2,7	4,2
46	21/04/2016	Night Watch in One Brain Hemisphere during Sleep Associated with the First-Night Effect in Humans	Current Biology	Report	Biological Science	3,4	1,6	4,1
47	24/06/2016	The social dilemma of autonomous vehicles	Science	Report	Studies in Human Society	17,7	0,2	4,1
48	20/07/2016	A multi-modal parcellation of human cerebral cortex	Nature	Article	Medical & Health Sciences	20,2	27,0	4,0
49	15/06/2016	Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement	JAMA	Recommendation statement	Medical & Health Sciences	9,1	1,8	4,0

50	29/02/2016	Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study	The Lancet	Article	Medical & Health Sciences	7,9	7,0	3,9
51	08/06/2016	Homo floresiensis-like fossils from the early Middle Pleistocene of Flores	Nature	Letter	History & Archaeology	3,6	1,5	3,9
52	13/04/2016	Restoring cortical control of functional movement in a human with quadriplegia	Nature	Letter	Medical & Health Sciences	5,3	5,8	3,9
53	02/05/2016	Temperate Earth-sized planets transiting a nearby ultracool dwarf star	Nature	Letter	Physical Sciences	3,2	1,7	3,8
54	23/05/2016	Revealing a 5,000-y-old beer recipe in China	PNAS	Article	History & Archaeology	2,7	0,8	3,7
55	16/03/2016	CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016	MMWR Recommendations & Reports	Report	Medical & Health Sciences	8,4	8,7	3,7
56	04/03/2016	Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth	Cell Stem Cell	Brief Report	Medical & Health Sciences	4,6	6,1	3,7
57	25/05/2016	Economic downturns, universal health coverage, and cancer mortality in high-income and middle-income countries, 1990—2010: a longitudinal analysis	The Lancet	Article	Studies in Human Society	7,6	1,1	3,4
58	06/09/2016	Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties	Annals of Internal Medicine	Article	Studies in Human Society	47,4	2,9	3,3

59	17/05/2016	Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study	The Lancet Psychiatry	Article	Medical & Health Sciences	2,0	3,4	3,3
60	18/04/2016	Dinosaurs in decline tens of millions of years before their final extinction	PNAS	Article	Earth & Environmental Sciences	2,3	1,5	3,2
61	01/06/2016	Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy	Nature	Letter	Biological Science	7,9	8,5	3,1
62	02/04/2016	Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants	The Lancet	Article	Medical & Health Sciences	23,0	11,2	3,1
63	02/06/2016	The Genome Project-Write	Science	Policy Forum	Biological Science	7,9	1,9	3,1
64	01/05/2016	Persistent metabolic adaptation 6 years after “The Biggest Loser” competition	Obesity	Article	Medical & Health Sciences	6,0	2,7	3,0
65	20/06/2016	Pharmaceutical Industry–Sponsored Meals and Physician Prescribing Patterns for Medicare Beneficiaries	JAMA Internal Medicine	Original investigation	Studies in Human Society	17,8	1,2	2,9
66	03/05/2016	A Deletion in the Canine POMC Gene Is Associated with Weight and Appetite in Obesity-Prone Labrador Retriever Dogs	Cell Metabolism (Science Direct)	Short Article	Biological Science	3,1	1,3	2,9

67	30/07/2016	A chapter a day: Association of book reading with longevity	Social Science & Medicine	Short Communication	Medical & Health Sciences	21,2	1,0	2,8
68	15/03/2016	Gradual Versus Abrupt Smoking Cessation	Annals of Internal Medicine	Article	Medical & Health Sciences	4,2	0,6	2,8
69	11/04/2016	Neural correlates of the LSD experience revealed by multimodal neuroimaging	PNAS	Article	Medical & Health Sciences	8,5	7,1	2,8
70	22/02/2016	Temperature-driven global sea-level variability in the Common Era	PNAS	Article	Earth & Environmental Sciences	2,5	2,0	2,7
71	27/01/2016	Mastering the game of Go with deep neural networks and tree search	Nature	Article	Information & Computer Sciences	20,6	1,5	2,7
72	03/05/2016	Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011	JAMA	Original Investigation	Medical & Health Sciences	10,4	2,9	2,7
73	21/03/2016	Analysis and valuation of the health and climate change cobenefits of dietary change	PNAS	Article	Medical & Health Sciences	9,0	5,0	2,7
74	18/04/2016	Loneliness and social isolation as risk factors for CVD: implications for evidence-based patient care and scientific inquiry	Heart	Editorial	Medical & Health Sciences	0,3	0,3	2,6
75	19/02/2016	Happy heart syndrome: role of positive emotional stress in takotsubo syndrome	European Heart Journal	Article	Medical & Health Sciences	1,4	0,3	2,5

76	04/03/2016	Zika Virus Infection in Pregnant Women in Rio de Janeiro — Preliminary Report	New England Journal of Medicine	Article	Medical & Health Sciences	9,4	1,1	2,5
77	13/01/2016	Zika Virus in the Americas - Yet Another Arbovirus Threat.	New England Journal of Medicine	Perspective	Medical & Health Sciences	17,6	3,9	2,5
78	16/03/2016	The ‘Tully monster’ is a vertebrate	Nature	Letter	History & Archaeology	9,0	1,0	2,4
79	25/03/2016	Design and synthesis of a minimal bacterial genome	Science	Article	Biological Science	16,5	16,3	2,4
80	02/03/2016	A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features	Nature Communications	Article	Medical & Health Sciences	3,5	1,8	2,4
81	03/02/2016	Naturally occurring p16Ink4a-positive cells shorten healthy lifespan	Nature	Article	Medical & Health Sciences	2,9	5,8	2,4
82	09/03/2016	Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans	Nature	Letter	Medical & Health Sciences	3,7	1,5	2,3
83	11/02/2016	Observation of Gravitational Waves from a Binary Black Hole Merger	Physical Review Letters	Article	Physical Sciences	51,0	26,6	2,2
84	06/01/2016	Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study	British Medical Journal	Article	Studies in Human Society	9,1	3,0	2,2

85	10/02/2016	Zika Virus Associated with Microcephaly	New England Journal of Medicine	Brief report	Medical & Health Sciences	23,7	7,8	2,1
86	11/03/2016	A bacterium that degrades and assimilates poly(ethylene terephthalate)	Science	Report	Materials Science	10,5	0,3	2,0
87	26/11/2015	Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study	Lancet Infectious Diseases	Article	Biological Science	6,2	9,2	2,0
88	01/01/2016	Alcohol consumption as a cause of cancer	Addiction	For Debate	Medical & Health Sciences	7,6	0,3	1,7
89	20/12/2015	Does happiness itself directly affect mortality? The prospective UK Million Women Study.	The Lancet	Letter	Studies in Human Society	8,2	1,7	1,7
90	16/05/2016	Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults	JAMA Internal Medicine	Original investigation	Medical & Health Sciences	21,4	3,1	1,6
91	08/01/2016	The Anthropocene is functionally and stratigraphically distinct from the Holocene	Science	Review	History & Archaeology	9,1	8,0	1,6
92	28/06/2016	Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates	PNAS	Article	Medical & Health Sciences	29,5	31,7	1,6

93	27/04/2016	Natural speech reveals the semantic maps that tile human cerebral cortex	Nature	Article	Medical & Health Sciences	16,0	13,4	1,6
94	23/02/2016	The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)	JAMA	Special Communication	Medical & Health Sciences	30,8	33,9	1,4
95	21/01/2016	Data Sharing.	New England Journal of Medicine	Editorial	Research & Reproducibility	30,0	1,7	1,0
96	09/02/2016	Gender bias in open source: Pull request acceptance of women versus men	PeerJ	Pre-print	Studies in Human Society	27,4	0,5	0,9
97	05/01/2016	Biomechanical Characteristics of Hand Coordination in Grasping Activities of Daily Living	PLoS ONE	Article	Medical & Health Sciences	19,7	1,4	0,8
98	11/04/2016	A new view of the tree of life	Nature Microbiology	Letter	Biological Science	23,4	11,3	0,8
99	07/03/2016	The ASA's statement on p-values: context, process, and purpose	The American Statistician	Editorial	Research & Reproducibility	22,4	0,2	0,8
100	14/12/2015	Rejection of rejection: a novel approach to overcoming barriers to publication	British Medical Journal	Article	Research & Reproducibility	31,0	0,9	0,0

APÊNDICE B – CROSSOVER DO TOP 100 DO ANO DE 2017

Dados Crossover								
Dados das Publicações						Indicadores		
Ranking	Data de Publicação	Título	Períodico	Tipo de conteúdo	Área	Exposição	Interação	Apropriação
1	02/11/2017	Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons	Nature	Letter	History & Archaeology	174,0	1,3	14,7
2	01/11/2017	Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species	Current Biology	Report	Biological Sciences	47,3	2,5	14,2
3	08/11/2017	Regeneration of the entire human epidermis using transgenic stem cells	Nature	Article	Biological Sciences	97,1	4,9	13,2
4	01/10/2017	Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults	The Lancet	Article	Medical & Health Sciences	100,7	5,1	12,5
5	24/07/2017	Remote detection of widespread indigenous water in lunar pyroclastic deposits	Nature Geoscience	Article	Earth & Environmental Science	3,6	0,3	11,6
6	18/10/2017	More than 75 percent decline over 27 years in total flying insect biomass in protected areas	PLoS ONE	Article	Biological Sciences	169,5	13,6	10,1

7	13/10/2017	Psilocybin for treatment-resistant depression: fMRI-measured brain mechanisms	Scientific Reports	Article	Biological Sciences	14,3	3,3	9,8
8	02/08/2017	Correction of a pathogenic gene mutation in human embryos	Nature	Article	Biological Sciences	92,8	15,5	9,3
9	25/07/2017	Clinicopathological Evaluation of Chronic Traumatic Encephalopathy in Players of American Football	JAMA: Journal of the American Medical Association	Editorial	Medical & Health Sciences	34,8	1,6	8,4
10	25/07/2017	Temporal trends in sperm count: a systematic review and meta-regression analysis	Human Reproduction Update	Review	Medical & Health Sciences	8,5	1,4	8,2
11	05/09/2017	Zika virus has oncolytic activity against glioblastoma stem cells	The Journal of Experimental Medicine	Brief report	Medical & Health Sciences	6,7	3,2	7,7
12	01/08/2017	Colorectal Cancer Incidence Patterns in the United States, 1974–2013	JNCI: Journal of the National Cancer Institute	Article	Medical & Health Sciences	3,8	1,5	7,3
13	12/09/2017	Patterns of Sedentary Behavior and Mortality in U.S. Middle-Aged and Older Adults: A National Cohort Study	Annals of Internal Medicine	Article	Medical & Health Sciences	23,9	0,4	6,9
14	17/07/2017	Nonnutritive sweeteners and cardiometabolic health: a systematic review and meta-analysis of randomized controlled trials and prospective cohort studies	Canadian Medical Association Journal	Article	Medical & Health Sciences	20,4	0,9	6,4

15	25/04/2017	An extra-uterine system to physiologically support the extreme premature lamb	Nature Communications	Article	Biological Sciences	13,7	2,2	6,1
16	08/08/2017	Instagram photos reveal predictive markers of depression	EPJ Data Science	Article	Studies in Human Society	48,7	3,9	5,8
17	05/07/2017	Retained contact lenses	British Medical Journal	Article	Medical & Health Sciences	9,3	0,1	5,8
18	27/07/2017	Alcohol drinking patterns and risk of diabetes: a cohort study of 70,551 men and women from the general Danish population	Diabetologia	Article	Medical & Health Sciences	6,5	0,4	5,8
19	18/05/2017	Widespread Biological Response to Rapid Warming on the Antarctic Peninsula	Current Biology	Report	Earth & Environmental Science	2,2	0,7	5,7
20	08/09/2017	A female Viking warrior confirmed by genomics	American Journal of Physical Anthropology	Brief report	Biological Sciences	61,1	1,0	5,5
21	18/10/2017	Mastering the game of Go without human knowledge	Nature	Article	Information & Computer Sciences	74,8	57,4	5,4
22	10/07/2017	Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines	Proceedings of the National Academy of Sciences of the United States of America	Article	Biological Sciences	29,8	9,2	5,3
23	01/03/2017	Evidence for early life in Earth's oldest hydrothermal vent precipitates	Nature	Article	Earth & Environmental Science	10,1	3,5	5,3

24	02/05/2017	Chocolate intake and risk of clinically apparent atrial fibrillation: the Danish Diet, Cancer, and Health Study	Heart	Article	Medical & Health Sciences	10,5	0,3	5,2
25	07/06/2017	New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens	Nature	Letter	History & Archaeology	14,1	2,6	5,1
26	22/02/2017	Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1	Nature	Letter	Physical Sciences	19,7	3,7	5,1
27	07/06/2017	The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age	Nature	Letter	History & Archaeology	4,8	1,9	5,1
28	27/06/2017	The safety, immunogenicity, and acceptability of inactivated influenza vaccine delivered by microneedle patch (TIV-MNP 2015): a randomised, partly blinded, placebo-controlled, phase 1 trial	The Lancet	Article	Medical & Health Sciences	8,6	0,9	5,0
29	24/03/2017	Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention	Science	Report	Medical & Health Sciences	11,6	3,2	4,9
30	15/03/2017	Global warming and recurrent mass bleaching of corals	Nature	Article	Earth & Environmental Science	20,6	6,1	4,9
31	14/04/2017	Cassini finds molecular hydrogen in the Enceladus plume: Evidence for hydrothermal processes	Science	Article	Physical Sciences	5,0	1,4	4,9

32	11/07/2017	Association of Coffee Consumption With Total and Cause-Specific Mortality Among Nonwhite Populations Association of Coffee Consumption With Mortality Among Nonwhite Populations	Annals of Internal Medicine	Article	Medical & Health Sciences	11,9	0,7	4,8
33	24/04/2017	Polyethylene bio-degradation by caterpillars of the wax moth <i>Galleria mellonella</i>	Current Biology	Correspondence	Biological Sciences	10,0	2,8	4,7
34	25/05/2017	Trial of Cannabidiol for Drug-Resistant Seizures in the Dravet Syndrome	New England Journal of Medicine	Article	Medical & Health Sciences	10,2	0,0	4,7
35	13/06/2017	Health Effects of Overweight and Obesity in 195 Countries over 25 Years	New England Journal of Medicine	Article	Medical & Health Sciences	51,3	7,0	4,7
36	06/06/2017	Moderate alcohol consumption as risk factor for adverse brain outcomes and cognitive decline: longitudinal cohort study	British Medical Journal	Article	Medical & Health Sciences	40,9	1,6	4,7
37	27/01/2017	Gender stereotypes about intellectual ability emerge early and influence children's interests	Science	Report	Studies in Human Society	20,6	3,0	4,7
38	16/12/2016	Increases in Drug and Opioid-Involved Overdose Deaths — United States, 2010–2015	MMWR: Morbidity & Mortality Weekly Report	Report	Medical & Health Sciences	6,2	2,8	4,7
39	29/08/2017	Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study	The Lancet	Article	Medical & Health Sciences	214,8	12,6	4,6

40	19/06/2017	Global risk of deadly heat	Nature Climate Change	Letter	Earth & Environmental Science	20,7	2,4	4,4
41	15/05/2017	Exceptional and rapid accumulation of anthropogenic debris on one of the world's most remote and pristine islands	Proceedings of the National Academy of Sciences of the United States of America	Article	Earth & Environmental Science	6,6	1,7	4,4
42	30/06/2017	Estimating economic damage from climate change in the United States	Science	Article	Earth & Environmental Science	20,6	1,4	4,4
43	22/12/2016	Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!)	The Lancet	Comment	Medical & Health Sciences	18,6	2,8	4,3
44	01/07/2017	Production, use, and fate of all plastics ever made	Science Advances	Article	Earth & Environmental Science	44,7	3,8	4,3
45	21/02/2017	Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble	The Lancet	Article	Medical & Health Sciences	9,9	3,2	4,1
46	08/12/2016	A Feathered Dinosaur Tail with Primitive Plumage Trapped in Mid-Cretaceous Amber	Current Biology	Report	Biological Sciences	22,3	1,5	4,1

47	10/05/2017	Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies	The Lancet HIV	Article	Medical & Health Sciences	6,7	0,5	4,1
48	19/07/2017	Microbiome analysis and confocal microscopy of used kitchen sponges reveal massive colonization by Acinetobacter, Moraxella and Chryseobacterium species	Scientific Reports	Article	Biological Sciences	38,3	0,8	4,0
49	11/07/2017	Coffee Drinking and Mortality in 10 European Countries: A Multinational Cohort Study	Annals of Internal Medicine	Article	Medical & Health Sciences	57,8	0,9	3,9
50	19/04/2017	Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study	British Medical Journal	Article	Medical & Health Sciences	26,2	1,4	3,8
51	16/05/2017	A bioprosthetic ovary created using 3D printed microporous scaffolds restores ovarian function in sterilized mice	Nature Communications	Article	Medical & Health Sciences	7,1	1,9	3,8
52	17/03/2017	Coronary atherosclerosis in indigenous South American Tsimane: a cross-sectional cohort study	The Lancet	Article	Medical & Health Sciences	8,4	0,9	3,8
53	26/04/2017	A 130,000-year-old archaeological site in southern California, USA	Nature	Letter	History & Archaeology	14,3	1,9	3,7
54	06/03/2017	Social Media Use and Perceived Social Isolation Among Young Adults in the U.S.	American Journal of Preventive Medicine	Article	Studies in Human Society	1,4	0,8	3,7

55	12/05/2017	Poor human olfaction is a 19th-century myth	Science	Review	Studies in Human Society	17,7	2,0	3,6
56	26/01/2017	Interspecies Chimerism with Mammalian Pluripotent Stem Cells	Cell	Article	Biological Sciences	8,3	0,1	3,6
57	04/01/2017	Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study	The Lancet	Article	Medical & Health Sciences	7,9	1,3	3,6
58	19/12/2016	Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians	JAMA Internal Medicine	Original Investigation	Medical & Health Sciences	38,0	1,8	3,6
59	24/07/2017	Buying time promotes happiness	Proceedings of the National Academy of Sciences of the United States of America	Article	Studies in Human Society	21,1	1,7	3,6
60	14/02/2017	Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain	Annals of Internal Medicine	Clinical Guidelines	Medical & Health Sciences	13,1	3,0	3,5
61	19/12/2016	Pregnancy leads to long-lasting changes in human brain structure	Nature Neuroscience	Article	Biological Sciences	9,4	2,9	3,3
62	10/07/2017	Large-scale physical activity data reveal worldwide activity inequality	Nature	Letter	Medical & Health Sciences	41,3	4,2	3,3
63	26/03/2017	Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies	International Journal of Epidemiology	Miscellaneous	Medical & Health Sciences	5,2	1,1	3,3

64	13/06/2017	Time to Delivery of an Automated External Defibrillator Using a Drone for Simulated Out-of-Hospital Cardiac Arrests vs Emergency Medical Services	JAMA: Journal of the American Medical Association	Letter	Studies in Human Society	21,8	0,3	3,0
65	01/05/2017	Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015	The Lancet	Article	Medical & Health Sciences	25,5	3,8	3,0
66	09/01/2017	Association of “Weekend Warrior” and Other Leisure Time Physical Activity Patterns With Risks for All-Cause, Cardiovascular Disease, and Cancer Mortality	JAMA Internal Medicine	Original Investigation	Medical & Health Sciences	16,1	0,7	3,0
67	08/05/2017	Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers	JAMA Internal Medicine	Original Investigation	Medical & Health Sciences	22,9	0,9	2,9
68	12/07/2017	CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria	Nature	Letter	Biological Sciences	87,9	0,6	2,9
69	08/03/2017	Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus	Nature	Letter	Medical & Health Sciences	12,2	2,8	2,9
70	01/09/2017	The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study	The Lancet	Article	Medical & Health Sciences	75,1	2,1	2,8
71	30/01/2017	Meiofaunal deuterostomes from the basal Cambrian of Shaanxi (China)	Nature	Letter	History & Archaeology	5,1	0,7	2,8
72	31/01/2017	Brain–Computer Interface–Based Communication in the Completely Locked-In State	PLoS Biology	Article	Medical & Health Sciences	3,3	1,8	2,8

73	22/03/2017	A new hypothesis of dinosaur relationships and early dinosaur evolution	Nature	Article	Biological Sciences	16,4	3,0	2,8
74	15/02/2017	Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data	British Medical Journal	Article	Medical & Health Sciences	13,2	0,1	2,7
75	12/01/2017	Notes from the Field: Pan-Resistant New Delhi Metallo-Beta-Lactamase-Producing <i>Klebsiella pneumoniae</i>	MMWR: Morbidity & Mortality Weekly Report	Report	Medical & Health Sciences	5,9	0,2	2,6
76	01/01/2017	Sugar- and Artificially Sweetened Beverages and the Risks of Incident Stroke and Dementia	Stroke	Original Contributions	Medical & Health Sciences	22,3	1,3	2,6
77	27/12/2016	The global decline of cheetah <i>Acinonyx jubatus</i> and what it means for conservation	Proceedings of the National Academy of Sciences of the United States of America	Article	Biological Sciences	1,5	0,6	2,6
78	13/02/2017	Bioaccumulation of persistent organic pollutants in the deepest ocean fauna	Nature Ecology & Evolution	Article	Earth & Environmental Science	5,4	1,4	2,6
79	05/12/2016	Correlation between pubic hair grooming and STIs: results from a nationally representative probability sample	Sexually Transmitted Infections	Article	Medical & Health Sciences	2,6	0,2	2,5
80	01/01/2017	Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association	Circulation	Clinical Guidelines	Medical & Health Sciences	8,4	1,3	2,4
81	07/03/2017	Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans	Cell Metabolism	Report	Medical & Health Sciences	36,0	2,2	2,4

82	01/02/2017	Crossing kingdoms: Using decellularized plants as perfusable tissue engineering scaffolds	Biomaterials	Article	Biological Sciences	22,6	2,1	2,3
83	07/03/2017	Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States	JAMA: Journal of the American Medical Association	Original Investigation	Medical & Health Sciences	16,7	2,0	2,3
84	10/05/2017	Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline	British Medical Journal	Practice	Medical & Health Sciences	22,4	1,1	2,3
85	20/02/2017	Difference-in-Differences Analysis of the Association Between State Same-Sex Marriage Policies and Adolescent Suicide Attempts	JAMA Pediatrics	Original Investigation	Studies in Human Society	14,8	0,5	2,2
86	05/12/2016	Nut consumption and risk of cardiovascular disease, total cancer, all-cause and cause-specific mortality: a systematic review and dose-response meta-analysis of prospective studies	BMC Medicine	Article	Medical & Health Sciences	6,5	0,4	2,2
87	24/05/2017	When Will AI Exceed Human Performance? Evidence from AI Experts	arXiv	Article	Studies in Human Society	68,7	2,4	2,2
88	02/05/2017	Long term gluten consumption in adults without celiac disease and risk of coronary heart disease: prospective cohort study	British Medical Journal	Article	Medical & Health Sciences	17,3	1,0	2,2
89	28/11/2016	Associations of specific types of sports and exercise with all-cause and cardiovascular-disease mortality: a cohort study of 80?306 British adults	British Journal of Sports Medicine	Article	Medical & Health Sciences	12,0	0,7	2,2
90	02/03/2017	Gene Therapy in a Patient with Sickle Cell Disease	New England Journal of Medicine	Brief report	Medical & Health Sciences	15,5	0,1	2,0

91	01/04/2017	Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity	Journal of the Association for Consumer Research	Article	Biological Sciences	26,8	0,6	1,9
92	11/01/2017	Relation between resting amygdalar activity and cardiovascular events: a longitudinal and cohort study	The Lancet	Article	Medical & Health Sciences	12,9	1,6	1,8
93	25/01/2017	Dermatologist-level classification of skin cancer with deep neural networks	Nature	Letter	Biological Sciences	23,3	9,3	1,8
94	01/06/2017	The Code for Facial Identity in the Primate Brain	Cell	Article	Biological Sciences	23,1	6,6	1,8
95	01/09/2017	How will Brexit affect health and health services in the UK? Evaluating three possible scenarios	The Lancet	Health Policy	Medical & Health Sciences	72,8	0,6	1,7
96	03/05/2017	Centers for Disease Control and Prevention Guideline for the Prevention of Surgical Site Infection, 2017	JAMA Surgery	Special Communication	Medical & Health Sciences	37,1	3,5	1,7
97	27/12/2016	US Spending on Personal Health Care and Public Health, 1996-2013	JAMA: Journal of the American Medical Association	Original Investigation	Medical & Health Sciences	26,7	1,5	1,5
98	06/12/2016	Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis	JAMA: Journal of the American Medical Association	Original Investigation	Medical & Health Sciences	17,7	1,8	1,2
99	11/08/2017	A Solution of the P versus NP Problem	arXiv	Article	Research & Reproducibility	92,0	2,2	0,5
100	01/03/2017	Work organization and mental health problems in PhD students	Research Policy	Article	Research & Reproducibility	81,6	9,6	0,3

APÊNDICE C - Ranking do Top 100 do ano de 2016

Posição no Rankink com Crossover	Posição no Ranking da Altmetric.com	Data de Publicação	Título	Pontuação
1	35	25/05/2016	Efficacy and Safety of an Injectable Combination Hormonal Contraceptive for Men	15,7
2	18	02/09/2016	The brain adapts to dishonesty	13,1
3	5	11/07/2016	Sugar Industry and Coronary Heart Disease Research: A Historical Analysis of Internal Industry Documents	12,7
4	88	04/10/2016	National Trends in Hospitalizations for Opioid Poisonings Among Children and Adolescents, 1997 to 2012	11,7
5	46	09/05/2016	1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America	11,7
6	11	01/07/2016	Evidence for a limit to human lifespan	11,5
7	61	13/01/2016	Association Between Portable Screen-Based Media Device Access or Use and Sleep Outcomes: A Systematic Review and Meta-analysis	11,2
8	8	05/10/2016	Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial	11,0
9	97	01/01/2016	Online social integration is associated with reduced mortality risk	10,9
10	76	26/11/2015	Mutational signatures associated with tobacco smoking in human cancer	10,8

11	72	14/10/2016	The Evolution of Galaxy Number Density at $z < 8$ and its Implications	10,2
12	30	28/07/2016	Neural mechanisms for lexical processing in dogs	9,0
13	2	12/09/2016	Medical error—the third leading cause of death in the US	8,8
14	22	26/05/2016	10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer	8,6
15	12	02/04/2016	The terrorist inside my husband's brain	8,6
16	39	31/10/2016	Association of Hormonal Contraception With Depression	8,4
17	13	27/09/2016	The antibody aducanumab reduces A β plaques in Alzheimer's disease	8,2
18	74	03/02/2016	Validation of a Functional Pyelocalyceal Renal Model for the Evaluation of Renal Calculi Passage While Riding a Roller Coaster	8,1
19	96	15/06/2016	Acne and telomere length. A new spectrum between senescence and apoptosis pathways	7,7
20	25	25/03/2016	Longer Contact Times Increase Cross-Contamination of Enterobacter aerogenes from Surfaces to Food	7,6
21	21	24/10/2016	Report of Partial findings from the National Toxicology Program Carcinogenesis Studies of Cell Phone Radiofrequency Radiation in Hsd: Sprague Dawley® SD rats (Whole Body Exposure)	7,3

22	40	26/10/2016	Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures	7,0
23	6	01/02/2016	Zika Virus and Birth Defects — Reviewing the Evidence for Causality	6,7
24	1	13/04/2016	United States Health Care Reform: Progress to Date and Next Steps	6,6
25	16	20/09/2016	Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women	6,1
26	51	16/03/2016	Multi-locus Analyses Reveal Four Giraffe Species Instead of One	6,1
27	4	10/04/2016	Evidence for a Distant Giant Planet in the Solar System	6,1
28	17	23/06/2016	Human commensals producing a novel antibiotic impair pathogen colonization	6,0
29	38	29/02/2016	A terrestrial planet candidate in a temperate orbit around Proxima Centauri	5,8
30	59	30/07/2016	Catastrophic Declines in Wilderness Areas Undermine Global Environment Targets	5,7
31	10	27/07/2016	The new world atlas of artificial night sky brightness	5,6
32	77	11/08/2016	Perimortem fractures in Lucy suggest mortality from fall out of tall tree	5,4

33	94	27/04/2016	Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa	5,4
34	15	30/03/2016	Contribution of Antarctica to past and future sea-level rise	5,1
35	36	13/04/2016	Emergence of healing in the Antarctic ozone layer	5,1
36	29	02/05/2016	Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (<i>Somniosus microcephalus</i>)	5,0
37	7	03/05/2016	The Association Between Income and Life Expectancy in the United States, 2001-2014	5,0
38	82	21/03/2016	Long-Term Training with a Brain-Machine Interface-Based Gait Protocol Induces Partial Neurological Recovery in Paraplegic Patients	4,8
39	41	31/08/2016	The High Cost of Prescription Drugs in the United States: Origins and Prospects for Reform	4,7
40	23	31/08/2016	Transient Smartphone “Blindness”	4,7
41	31	24/08/2016	<i>Escherichia coli</i> Harboring <i>mcr-1</i> and <i>blaCTX-M</i> on a Novel IncF Plasmid: First report of <i>mcr-1</i> in the USA	4,6
42	70	28/09/2016	A genetic assessment of the English bulldog	4,6
43	63	07/03/2016	Apollo Lunar Astronauts Show Higher Cardiovascular Disease Mortality: Possible Deep Space Radiation Effects on the Vascular Endothelium	4,4

44	57	02/06/2016	Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality	4,2
45	32	23/08/2016	An elastic second skin	4,2
46	27	21/01/2016	Night Watch in One Brain Hemisphere during Sleep Associated with the First-Night Effect in Humans	4,1
47	33	21/04/2016	The social dilemma of autonomous vehicles	4,1
48	47	08/06/2016	A multi-modal parcellation of human cerebral cortex	4,0
49	98	04/03/2016	Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement	4,0
50	37	12/08/2016	Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study	3,9
51	60	03/05/2016	Homo floresiensis-like fossils from the early Middle Pleistocene of Flores	3,9
52	24	30/08/2016	Restoring cortical control of functional movement in a human with quadriplegia	3,9
53	44	18/04/2016	Temperate Earth-sized planets transiting a nearby ultracool dwarf star	3,8
54	43	14/12/2015	Revealing a 5,000-y-old beer recipe in China	3,7

55	52	08/09/2016	CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016	3,7
56	54	20/07/2016	Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth	3,7
57	45	28/06/2016	Economic downturns, universal health coverage, and cancer mortality in high-income and middle-income countries, 1990—2010: a longitudinal analysis	3,4
58	65	11/03/2016	Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties	3,3
59	83	27/07/2016	Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study	3,3
60	50	16/03/2016	Dinosaurs in decline tens of millions of years before their final extinction	3,2
61	69	06/09/2016	Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy	3,1
62	14	01/06/2016	Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants	3,1
63	67	31/10/2016	The Genome Project-Write	3,1
64	89	16/05/2016	Persistent metabolic adaptation 6 years after “The Biggest Loser” competition	3,0
65	87	02/03/2016	Pharmaceutical Industry–Sponsored Meals and Physician Prescribing Patterns for Medicare Beneficiaries	2,9

66	90	24/06/2016	A Deletion in the Canine POMC Gene Is Associated with Weight and Appetite in Obesity-Prone Labrador Retriever Dogs	2,9
67	66	03/05/2016	A chapter a day: Association of book reading with longevity	2,8
68	56	01/06/2016	Gradual Versus Abrupt Smoking Cessation	2,8
69	42	05/01/2016	Neural correlates of the LSD experience revealed by multimodal neuroimaging	2,8
70	73	15/03/2016	Temperature-driven global sea-level variability in the Common Era	2,7
71	9	27/01/2016	Mastering the game of Go with deep neural networks and tree search	2,7
72	68	11/04/2016	Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011	2,7
73	55	10/02/2016	Analysis and valuation of the health and climate change cobenefits of dietary change	2,7
74	100	20/06/2016	Loneliness and social isolation as risk factors for CVD: implications for evidence-based patient care and scientific inquiry	2,6
75	75	29/07/2016	Happy heart syndrome: role of positive emotional stress in takotsubo syndrome	2,5
76	92	06/01/2016	Zika Virus Infection in Pregnant Women in Rio de Janeiro — Preliminary Report	2,5

77	49	23/05/2016	Zika Virus in the Americas - Yet Another Arbovirus Threat.	2,5
78	58	04/03/2016	The 'Tully monster' is a vertebrate	2,4
79	26	27/10/2016	Design and synthesis of a minimal bacterial genome	2,4
80	85	19/02/2016	A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features	2,4
81	78	09/03/2016	Naturally occurring p16Ink4a-positive cells shorten healthy lifespan	2,4
82	99	18/04/2016	Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans	2,3
83	3	11/02/2016	Observation of Gravitational Waves from a Binary Black Hole Merger	2,2
84	95	08/01/2016	Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study	2,2
85	20	23/02/2016	Zika Virus Associated with Microcephaly	2,1
86	53	01/05/2016	A bacterium that degrades and assimilates poly(ethylene terephthalate)	2,0
87	86	17/05/2016	Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study	2,0

88	91	01/10/2016	Alcohol consumption as a cause of cancer	1,7
89	80	29/08/2016	Does happiness itself directly affect mortality? The prospective UK Million Women Study.	1,7
90	81	11/04/2016	Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults	1,6
91	84	22/02/2016	The Anthropocene is functionally and stratigraphically distinct from the Holocene	1,6
92	62	31/10/2016	Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates	1,6
93	93	20/12/2015	Natural speech reveals the semantic maps that tile human cerebral cortex	1,6
94	19	08/09/2016	The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)	1,4
95	34	26/05/2016	Data Sharing.	1,0
96	28	30/06/2016	Gender bias in open source: Pull request acceptance of women versus men	0,9
97	64	09/02/2016	Biomechanical Characteristics of Hand Coordination in Grasping Activities of Daily Living	0,8
98	79	01/08/2016	A new view of the tree of life	0,8
99	71	04/11/2016	The ASA's statement on p-values: context, process, and purpose	0,8
100	48	14/09/2016	Rejection of rejection: a novel approach to overcoming barriers to publication	0,0

APÊNDICE D - Ranking do Top 100 do ano de 2017

Posição no Rankink com Crossover	Posição no Ranking da Altmetric.com	Data de Publicação	Título	Pontuação
1	30	02/11/2017	Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons	14,7
2	72	01/11/2017	Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species	14,2
3	50	08/11/2017	Regeneration of the entire human epidermis using transgenic stem cells	13,2
4	7	01/10/2017	Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults	12,5
5	91	24/07/2017	Remote detection of widespread indigenous water in lunar pyroclastic deposits	11,6
6	6	18/10/2017	More than 75 percent decline over 27 years in total flying insect biomass in protected areas	10,1
7	85	13/10/2017	Psilocybin for treatment-resistant depression: fMRI-measured brain mechanisms	9,8
8	4	02/08/2017	Correction of a pathogenic gene mutation in human embryos	9,3
9	13	25/07/2017	Clinicopathological Evaluation of Chronic Traumatic Encephalopathy in Players of American Football	8,4

10	26	25/07/2017	Temporal trends in sperm count: a systematic review and meta-regression analysis	8,2
11	79	05/09/2017	Zika virus has oncolytic activity against glioblastoma stem cells	7,7
12	58	01/08/2017	Colorectal Cancer Incidence Patterns in the United States, 1974–2013	7,3
13	75	12/09/2017	Patterns of Sedentary Behavior and Mortality in U.S. Middle-Aged and Older Adults: A National Cohort Study	6,9
14	33	17/07/2017	Nonnutritive sweeteners and cardiometabolic health: a systematic review and meta-analysis of randomized controlled trials and prospective cohort studies	6,4
15	10	25/04/2017	An extra-uterine system to physiologically support the extreme premature lamb	6,1
16	17	08/08/2017	Instagram photos reveal predictive markers of depression	5,8
17	41	05/07/2017	Retained contact lenses	5,8
18	64	27/07/2017	Alcohol drinking patterns and risk of diabetes: a cohort study of 70,551 men and women from the general Danish population	5,8
19	70	18/05/2017	Widespread Biological Response to Rapid Warming on the Antarctic Peninsula	5,7
20	43	08/09/2017	A female Viking warrior confirmed by genomics	5,5

21	74	18/10/2017	Mastering the game of Go without human knowledge	5,4
22	35	10/07/2017	Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines	5,3
23	11	01/03/2017	Evidence for early life in Earth's oldest hydrothermal vent precipitates	5,3
24	25	02/05/2017	Chocolate intake and risk of clinically apparent atrial fibrillation: the Danish Diet, Cancer, and Health Study	5,2
25	40	07/06/2017	New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens	5,1
26	12	22/02/2017	Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1	5,1
27	47	07/06/2017	The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age	5,1
28	66	27/06/2017	The safety, immunogenicity, and acceptability of inactivated influenza vaccine delivered by microneedle patch (TIV-MNP 2015): a randomised, partly blinded, placebo-controlled, phase 1 trial	5,0
29	15	24/03/2017	Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention	4,9
30	21	15/03/2017	Global warming and recurrent mass bleaching of corals	4,9
31	68	14/04/2017	Cassini finds molecular hydrogen in the Enceladus plume: Evidence for hydrothermal processes	4,9

32	67	11/07/2017	Association of Coffee Consumption With Total and Cause-Specific Mortality Among Nonwhite Populations Association of Coffee Consumption With Mortality Among Nonwhite Populations	4,8
33	24	24/04/2017	Polyethylene bio-degradation by caterpillars of the wax moth <i>Galleria mellonella</i>	4,7
34	56	25/05/2017	Trial of Cannabidiol for Drug-Resistant Seizures in the Dravet Syndrome	4,7
35	14	13/06/2017	Health Effects of Overweight and Obesity in 195 Countries over 25 Years	4,7
36	20	06/06/2017	Moderate alcohol consumption as risk factor for adverse brain outcomes and cognitive decline: longitudinal cohort study	4,7
37	5	27/01/2017	Gender stereotypes about intellectual ability emerge early and influence children's interests	4,7
38	44	16/12/2016	Increases in Drug and Opioid-Involved Overdose Deaths — United States, 2010–2015	4,7
39	1	29/08/2017	Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study	4,6
40	42	19/06/2017	Global risk of deadly heat	4,4
41	51	15/05/2017	Exceptional and rapid accumulation of anthropogenic debris on one of the world's most remote and pristine islands	4,4
42	55	30/06/2017	Estimating economic damage from climate change in the United States	4,4

43	9	22/12/2016	Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!)	4,3
44	29	01/07/2017	Production, use, and fate of all plastics ever made	4,3
45	31	21/02/2017	Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble	4,1
46	8	08/12/2016	A Feathered Dinosaur Tail with Primitive Plumage Trapped in Mid-Cretaceous Amber	4,1
47	62	10/05/2017	Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies	4,1
48	52	19/07/2017	Microbiome analysis and confocal microscopy of used kitchen sponges reveal massive colonization by Acinetobacter, Moraxella and Chryseobacterium species	4,0
49	19	11/07/2017	Coffee Drinking and Mortality in 10 European Countries: A Multinational Cohort Study	3,9
50	18	19/04/2017	Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study	3,8
51	63	16/05/2017	A bioprosthetic ovary created using 3D printed microporous scaffolds restores ovarian function in sterilized mice	3,8
52	57	17/03/2017	Coronary atherosclerosis in indigenous South American Tsimane: a cross-sectional cohort study	3,8
53	46	26/04/2017	A 130,000-year-old archaeological site in southern California, USA	3,7

54	73	06/03/2017	Social Media Use and Perceived Social Isolation Among Young Adults in the U.S.	3,7
55	92	12/05/2017	Poor human olfaction is a 19th-century myth	3,6
56	37	26/01/2017	Interspecies Chimerism with Mammalian Pluripotent Stem Cells	3,6
57	32	04/01/2017	Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study	3,6
58	3	19/12/2016	Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians	3,6
59	87	24/07/2017	Buying time promotes happiness	3,6
60	83	14/02/2017	Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline From the American College of Physicians Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain	3,5
61	23	19/12/2016	Pregnancy leads to long-lasting changes in human brain structure	3,3
62	39	10/07/2017	Large-scale physical activity data reveal worldwide activity inequality	3,3
63	88	26/03/2017	Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies	3,3
64	59	13/06/2017	Time to Delivery of an Automated External Defibrillator Using a Drone for Simulated Out-of-Hospital Cardiac Arrests vs Emergency Medical Services	3,0

65	34	01/05/2017	Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015	3,0
66	28	09/01/2017	Association of “Weekend Warrior” and Other Leisure Time Physical Activity Patterns With Risks for All-Cause, Cardiovascular Disease, and Cancer Mortality	3,0
67	54	08/05/2017	Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers	2,9
68	27	12/07/2017	CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria	2,9
69	69	08/03/2017	Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus	2,9
70	71	01/09/2017	The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study	2,8
71	76	30/01/2017	Meiofaunal deuterostomes from the basal Cambrian of Shaanxi (China)	2,8
72	89	31/01/2017	Brain–Computer Interface–Based Communication in the Completely Locked-In State	2,8
73	84	22/03/2017	A new hypothesis of dinosaur relationships and early dinosaur evolution	2,8
74	61	15/02/2017	Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data	2,7
75	78	12/01/2017	Notes from the Field: Pan-Resistant New Delhi Metallo-Beta-Lactamase-Producing <i>Klebsiella pneumoniae</i>	2,6

76	22	01/01/2017	Sugar- and Artificially Sweetened Beverages and the Risks of Incident Stroke and Dementia	2,6
77	94	05/12/2016	Correlation between pubic hair grooming and STIs: results from a nationally representative probability sample	2,6
78	82	13/02/2017	Bioaccumulation of persistent organic pollutants in the deepest ocean fauna	2,6
79	93	27/12/2016	The global decline of cheetah <i>Acinonyx jubatus</i> and what it means for conservation	2,5
80	86	01/01/2017	Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association	2,4
81	65	07/03/2017	Enhanced Protein Translation Underlies Improved Metabolic and Physical Adaptations to Different Exercise Training Modes in Young and Old Humans	2,4
82	53	01/02/2017	Crossing kingdoms: Using decellularized plants as perfusable tissue engineering scaffolds	2,3
83	77	07/03/2017	Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States	2,3
84	81	10/05/2017	Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline	2,3
85	60	20/02/2017	Difference-in-Differences Analysis of the Association Between State Same-Sex Marriage Policies and Adolescent Suicide Attempts	2,2
86	98	05/12/2016	Nut consumption and risk of cardiovascular disease, total cancer, all-cause and cause-specific mortality: a systematic review and dose-response meta-analysis of prospective studies	2,2

87	16	24/05/2017	When Will AI Exceed Human Performance? Evidence from AI Experts	2,2
88	97	02/05/2017	Long term gluten consumption in adults without celiac disease and risk of coronary heart disease: prospective cohort study	2,2
89	49	28/11/2016	Associations of specific types of sports and exercise with all-cause and cardiovascular-disease mortality: a cohort study of 80?306 British adults	2,2
90	90	02/03/2017	Gene Therapy in a Patient with Sickle Cell Disease	2,0
91	36	01/04/2017	Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity	1,9
92	95	11/01/2017	Relation between resting amygdalar activity and cardiovascular events: a longitudinal and cohort study	1,8
93	45	25/01/2017	Dermatologist-level classification of skin cancer with deep neural networks	1,8
94	100	01/06/2017	The Code for Facial Identity in the Primate Brain	1,8
95	99	01/09/2017	How will Brexit affect health and health services in the UK? Evaluating three possible scenarios	1,7
96	80	03/05/2017	Centers for Disease Control and Prevention Guideline for the Prevention of Surgical Site Infection, 2017	1,7
97	48	27/12/2016	US Spending on Personal Health Care and Public Health, 1996-2013	1,5
98	96	06/12/2016	Prevalence of Depression, Depressive Symptoms, and Suicidal Ideation Among Medical Students: A Systematic Review and Meta-Analysis	1,2

99	38	11/08/2017	A Solution of the P versus NP Problem	0,5
100	2	01/03/2017	Work organization and mental health problems in PhD students	0,3

APÊNDICE E – FOLHA DE PROGRAMAÇÃO DO R PARA OBTENÇÃO DOS RESULTADOS

```

#Limpar tudo da lista de objetos
rm(list=ls())
Sys.setenv(TZ='GMT')
library(rmarkdown)

#Bibliotecas que serão utilizadas na modelagem dos dados
library(readxl) #Biblioteca que permite leitura de arquivos com extensão xls
library(ggplot2) #Biblioteca para geração de gráficos
library(gridExtra)
library(corrplot) #Biblioteca para geração de gráficos de correlação

## corrplot 0.84 loaded

library(PerformanceAnalytics) #Biblioteca para geração de gráficos de correlação

## Loading required package: xts
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##   legend

#Comando utilizado para importar as tabelas com os dados dos Top 100 - 2016
e Top 100 - 2017
url <- "https://www.dropbox.com/s/5jkdwmw2igxw93yt/Crossover2016.xlsx?dl=1"
destfile <- "Crossover2016.xlsx"
download.file(url, destfile)
Crossover2016 <- read_excel(destfile, sheet = "Top100_2016") #Comando que irá
importar a tabela do Top 100 de 2016

```

#Com a tabela do Top 100 - 2016 importado houve a necessidade de renomear as colunas para facilitar a manipulação dos dados

```
names(Crossover2016) <- c("Posição", "Altmetric_ID", "Titulo", "Publicação",
, "Pontuação", "Data_Publicação", "DOI", "Link", "Tipo", "Area", "Tinha_Correções", "Dados_afiliação", "Resumo", "Verificado_menções", "Notícias", "Blogs", "Governo", "Twitter", "Peer_Review", "Weibo", "Facebook", "Wikipedia", "Google+", "Linkedin", "Reddit", "Pinterest", "F1000", "QNA", "Videos", "Mendeley", "Connotea", "Citeulike", "Sharedit")
```

#A tabela com os dados não continha a data de extração final dos dados que compuseram o Top 100 - 2016, como posteriormente será necessário estes dados utilizamos o comando para incluir a coluna com a data para todos os dados do Top 100 - 2016

```
Crossover2016$Adicionado <- "2016-12-08 12:50:45"
```

```
Crossover2016["Adicionado"] <- format(as.Date(Crossover2016$Adicionado,format="%Y-%m-%d %H:%M:%S"),format="%Y-%m-%d")
```

```
Crossover2016["Data_Publicação"] <- format(as.Date(Crossover2016$Data_Publicação,format="%Y-%m-%d %H:%M:%S"),format="%Y-%m-%d")
```

#Calcular a diferente entre a data de adição e a data de atualização

```
Crossover2016$Adicionado <- as.Date(Crossover2016$Adicionado)
Crossover2016["Dias"] <- abs(as.numeric(difftime((Crossover2016$Adicionado), as.Date(Crossover2016$Data_Publicação), units = "auto")))
```

#Apesar de não influenciar nos resultados foram excluídas colunas que não teriam utilidade para o estudo. No entanto, a exclusão foi somente por precisismo dos autores.

```
excluir <- c("Tinha_Correções", "Dados_afiliação", "Verificado_menções", "Sharedit")
Crossover2016 <- Crossover2016[,!(names(Crossover2016)%in% excluir)]
```

#Reordenar as colunas - O reordenamento teve sentido apenas estético para facilitar identificar os dados, mas não gerou qualquer influência nos resultados

```
Crossover2016 <- Crossover2016 [c("Posição","Altmetric_ID","Titulo","Publicação","Data_Publicação","Adicionado","Dias","DOI","Link","Pontuação","Tipo","Area","Resumo","Twitter","Facebook","Google+","Weibo","Pinterest","Videos","Mendeley","Connotea","Citeulike","Reddit","Notícias","Blogs","Governo","Wikipedia","F1000","QNA","Peer_Review","Linkedin")]
```

#A função "names" tem como objetivo visualizar a disposição dos dados dentro da tabela ou data frame depois do reordenamento

```
names(Crossover2016)
```

```
## [1] "Posição"      "Altmetric_ID"  "Titulo"
## [4] "Publicação"   "Data_Publicação" "Adicionado"
## [7] "Dias"         "DOI"           "Link"
## [10] "Pontuação"    "Tipo"          "Area"
## [13] "Resumo"       "Twitter"       "Facebook"
## [16] "Google+"      "Weibo"         "Pinterest"
## [19] "Videos"       "Mendeley"      "Connotea"
```

```
## [22] "Citeulike"      "Reddit"        "Notícias"
## [25] "Blogs"          "Governo"       "Wikipedia"
## [28] "F1000"          "QNA"           "Peer_Review"
## [31] "Linkedin"
```

#Exportação da Tabela que será utilizada como Apêndice A
write.csv2(Crossover2016, file = "Crossover2016rev.csv")

names(Crossover2016)

```
## [1] "Posição"      "Altmetric_ID"  "Titulo"
## [4] "Publicação"   "Data_Publicação" "Adicionado"
## [7] "Dias"          "DOI"           "Link"
## [10] "Pontuação"    "Tipo"          "Area"
## [13] "Resumo"        "Twitter"       "Facebook"
## [16] "Google+"      "Weibo"         "Pinterest"
## [19] "Videos"       "Mendeley"      "Connotea"
## [22] "Citeulike"    "Reddit"        "Notícias"
## [25] "Blogs"        "Governo"       "Wikipedia"
## [28] "F1000"        "QNA"           "Peer_Review"
## [31] "Linkedin"
```

#Com a distribuição correta dos dados foi realizado o somatório de cada coluna do indicador Exposição

```
Crossover2016["Exposição"] <- Crossover2016["Twitter"] +
  Crossover2016["Facebook"] +
  Crossover2016["Google+"] +
  Crossover2016["Weibo"] +
  Crossover2016["Pinterest"] +
  Crossover2016["Videos"]
```

#O mesmo cálculo foi realizado para o indicador Interação

```
Crossover2016["Interação"] <- Crossover2016["Mendeley"] +
  Crossover2016["Connotea"] +
  Crossover2016["Citeulike"] +
  Crossover2016["Reddit"]
```

#O último indicador foi Apropriação que também teve os dados das redes somados

```
Crossover2016["Apropriação"] <- Crossover2016["Notícias"] +
  Crossover2016["Blogs"] +
  Crossover2016["Governo"] +
  Crossover2016["Wikipedia"] +
  Crossover2016["F1000"] +
  Crossover2016["QNA"] +
  Crossover2016["Peer_Review"] +
  Crossover2016["Linkedin"]
```

#Somar as colunas dos três indicadores - O mesmo comando poderia ser utilizado na soma individual de todas as colunas das redes sociais

```
Crossover2016["Total"] <- Crossover2016["Exposição"] + Crossover2016["Inter
```

```

ação"] + Crossover2016["Apropriação"]

#A primeira visualização dos dados após a transformação tem como objetivo i
dentificar como estão distribuídos em um gráfico de barras
barplot(Crossover2016$Total, main="Distribuição dos dados", xlab="Total de
Documentos", ylab="Total de Postagens", col="Gold", cex.names = 0.3, xaxs =
"i")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Exposição, main="Distribuição dos dados", xlab="Total
de Documentos", ylab="Total de Postagens", col="DeepSkyBlue", cex.names = 0
.3, xaxs = "i", add=TRUE)

barplot(Crossover2016$Interação, main="Distribuição dos dados", xlab="Total
de Documentos", ylab="Total de Postagens", col="#00441B", cex.names = 0.8,
xaxs = "i", add=TRUE)

barplot(Crossover2016$Apropriação, main="Distribuição dos dados", xlab="Tot
al de Documentos", ylab="Total de Postagens", col="DeepPink1", cex.names =
0.8, xaxs = "i", add=TRUE)

legend("topright", legend=c("Total", "Exposição", "Interação", "Apropriação"),
col=c("Gold", "DeepSkyBlue", "#00441B", "DeepPink1"),
lty=c(1,1,1,1), title("example of line plot"), ncol=1)

#Em seguida foram gerados os gráficos individuais que proporcionam melhor v
isualização dos dados e que foram inseridos na Dissertação
barplot(Crossover2016$Total, names.arg = Crossover2016$Posição, font.axis =
1, main="Distribuição dos dados - Volume Total de Publicações - 2016", xlab
="Total de Documentos", ylab="Total de Envolvimentos", font.lab = 2, col.lab
= "black", ylim=c(0, 10000), col="Gold", cex.names = 0.8, xaxs = "i")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Exposição, names.arg = Crossover2016$Posição, font.ax
is = 1, main="Distribuição dos dados - Indicador Exposição", xlab="Total de
Documentos", ylab="Total de Envolvimentos", font.lab = 2, col.lab = "black"
, ylim=c(0, 10000), col="DeepSkyBlue", xaxs = "i")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Interação, names.arg = Crossover2016$Posição, font.ax
is = 1, main="", xlab="Total de Documentos", ylab="Total de Envolvimentos",
font.lab = 2, col.lab = "black", ylim=c(0, 10000), col="#00441B", xaxs = "i
")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Apropriação, names.arg = Crossover2016$Posição, font.
axis = 1, main="Distribuição dos dados - Indicador Apropriação", xlab="Tota
l de Documentos", ylab="Total de Envolvimentos", font.lab = 2, col.lab = "b
lack", ylim=c(0, 10000), col="DeepPink1", xaxs = "i")
grid(nx=NA, ny=NULL)

#Normalização das redes que compõem o indicador Exposição
Crossover2016["NTwitter"] <- log(Crossover2016["Twitter"])
Crossover2016["NFacebook"] <- log(Crossover2016["Facebook"])

```

```

Crossover2016["NGoogle+"] <- log(Crossover2016["Google+"])
Crossover2016["NWeibo"] <- log(Crossover2016["Weibo"])
Crossover2016["NPinterest"] <- log(Crossover2016["Pinterest"])
Crossover2016["NVideos"] <- log(Crossover2016["Videos"])

#A visualização dos dados possibilitou identificar que alguns resultados ti
veram resultado infinito. Então, estes dados serão convertidos para 0 (zero
)
Crossover2016$NTwitter[!is.finite(Crossover2016$NTwitter)] <- 0
Crossover2016$NFacebook[!is.finite(Crossover2016$NFacebook)] <- 0
Crossover2016$NGoogle+[!is.finite(Crossover2016$NGoogle+)] <- 0
Crossover2016$NWeibo[!is.finite(Crossover2016$NWeibo)] <- 0
Crossover2016$NPinterest[!is.finite(Crossover2016$NPinterest)] <- 0
Crossover2016$NVideos[!is.finite(Crossover2016$NVideos)] <- 0

#Normalização das redes que compõem o indicador Interação
Crossover2016["NMendeley"] <- log(Crossover2016["Mendeley"])
Crossover2016["NConnotea"] <- log(Crossover2016["Connotea"])
Crossover2016["NCiteulike"] <- log(Crossover2016["Citeulike"])
Crossover2016["NReddit"] <- log(Crossover2016["Reddit"])

#A visualização dos dados possibilitou identificar que alguns resultados ti
veram resultado infinito. Então, estes dados serão convertidos para 0 (zero
)
Crossover2016$NMendeley[!is.finite(Crossover2016$NMendeley)] <- 0
Crossover2016$NConnotea[!is.finite(Crossover2016$NConnotea)] <- 0
Crossover2016$NCiteulike[!is.finite(Crossover2016$NCiteulike)] <- 0
Crossover2016$NReddit[!is.finite(Crossover2016$NReddit)] <- 0

#Normalização das redes que compõem o indicador Apropriação
Crossover2016["NNotícias"] <- log(Crossover2016["Notícias"])
Crossover2016["NBlogs"] <- log(Crossover2016["Blogs"])
Crossover2016["NGoverno"] <- log(Crossover2016["Governo"])
Crossover2016["NWikipedia"] <- log(Crossover2016["Wikipedia"])
Crossover2016["NF1000"] <- log(Crossover2016["F1000"])
Crossover2016["NQNA"] <- log(Crossover2016["QNA"])
Crossover2016["NPeer_Review"] <- log(Crossover2016["Peer_Review"])
Crossover2016["NLinkedin"] <- log(Crossover2016["Linkedin"])

#A visualização dos dados possibilitou identificar que alguns resultados ti
veram resultado infinito. Então, estes dados serão convertidos para 0 (zero
)
Crossover2016$NNotícias[!is.finite(Crossover2016$NNotícias)] <- 0
Crossover2016$NBlogs[!is.finite(Crossover2016$NBlogs)] <- 0
Crossover2016$NGoverno[!is.finite(Crossover2016$NGoverno)] <- 0
Crossover2016$NWikipedia[!is.finite(Crossover2016$NWikipedia)] <- 0
Crossover2016$NF1000[!is.finite(Crossover2016$NF1000)] <- 0
Crossover2016$NQNA[!is.finite(Crossover2016$NQNA)] <- 0
Crossover2016$NPeer_Review[!is.finite(Crossover2016$NPeer_Review)] <- 0
Crossover2016$NLinkedin[!is.finite(Crossover2016$NLinkedin)] <- 0

#Com a distribuição correta dos dados foi realizado o somatório de cada co

```

Luna do indicador Exposição

```
Crossover2016["Normalização_Exposição"] <- Crossover2016["NTwitter"] +
  Crossover2016["NFacebook"] +
  Crossover2016["NGoogle+"] +
  Crossover2016["NWeibo"] +
  Crossover2016["NPinterest"] +
  Crossover2016["NVideos"]
```

#O mesmo cálculo foi realizado para o indicador Interação

```
Crossover2016["Normalização_Interação"] <- Crossover2016["NMendeley"] +
  Crossover2016["NConnotea"] +
  Crossover2016["NCiteulike"] +
  Crossover2016["NReddit"]
```

#O último indicador foi Apropriação que também teve os dados das redes somados

```
Crossover2016["Normalização_Apropriação"] <- Crossover2016["NNotícias"] +
  Crossover2016["NBlogs"] +
  Crossover2016["NGoverno"] +
  Crossover2016["NWikipedia"] +
  Crossover2016["NF1000"] +
  Crossover2016["NQNA"] +
  Crossover2016["NPeer_Review"] +
  Crossover2016["NLinkedIn"]
```

#Somar as colunas dos três indicadores - O mesmo comando poderia ser utilizado na soma individual de todas as colunas das redes sociais

```
Crossover2016["Normalização_Total"] <- Crossover2016["Normalização_Exposição"] + Crossover2016["Normalização_Interação"] + Crossover2016["Normalização_Apropriação"]
```

#A definição que o Logaritmo natural seria o mais adequado para a realização da pesquisa partiu de testes realizados com outros modelos de transformações

```
Crossover2016["TwitterLog"] <- log(Crossover2016["Twitter"])
Crossover2016["TwitterLog2"] <- log2(Crossover2016["Twitter"])
Crossover2016["TwitterLog10"] <- log10(Crossover2016["Twitter"])
Crossover2016["TwitterLog1p"] <- log1p(Crossover2016["Twitter"])
Crossover2016["TwitterLogb"] <- logb(Crossover2016["Twitter"])
```

*View(Crossover2016) #Comando utilizado para conferir os resultados da transformação dos dados**#Testes realizados para verificar como ficou a distribuição dos dados antes e depois da normalização**#Indicador Exposição*

```
qqnorm(Crossover2016$Exposição) #Dados do Indicador Exposição não normalizados
```

```
qqline(Crossover2016$Exposição, col="red")
```

```

qqnorm(Crossover2016$Normalização_Exposição) #Dados do Indicador Exposição
normalizados
qqline(Crossover2016$Normalização_Exposição, col="red")

#Indicador Interação
qqnorm(Crossover2016$Interação) #Dadoss do Indicador Interação não normaliz
ados
qqline(Crossover2016$Interação, col="red")

qqnorm(Crossover2016$Normalização_Interação) #Dados do Indicador Interação
normalizados
qqline(Crossover2016$Normalização_Interação, col="red")

#Indicador Apropriação
qqnorm(Crossover2016$Apropriação) #Dadoss do Indicador Apropriação não norm
alizados
qqline(Crossover2016$Apropriação, col="red")

qqnorm(Crossover2016$Normalização_Apropriação) #Dados do Indicador Apropria
ção normalizados
qqline(Crossover2016$Normalização_Apropriação, col="red")

#Após a realização dos testes nos indicadores foi gerado gráficos com os d
ados normalizados de cada indicador
barplot(Crossover2016$Normalização_Total, names.arg = Crossover2016$Posição
, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Envolve
mentos", col="Gold", col.lab = "black", ylim=c(0, 50), xaxs = "i")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Normalização_Exposição, names.arg = Crossover2016$Pos
ição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Env
olvimentos", col="DeepSkyBlue", col.lab = "black", ylim=c(0, 20), xaxs = "i
")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Normalização_Interação, names.arg = Crossover2016$Pos
ição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Env
olvimentos", col="#00441B", col.lab = "black", ylim=c(0, 20), xaxs = "i")
grid(nx=NA, ny=NULL)

barplot(Crossover2016$Normalização_Apropriação, names.arg = Crossover2016$P
osição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de E
nvolvimentos", col="DeepPink1", col.lab = "black", ylim=c(0, 20), xaxs = "i
")
grid(nx=NA, ny=NULL)

#Gráfico de correlação com índices não normalizados - 2016
p1 <- ggplot(Crossover2016, mapping = aes(x = Exposição, y = Interação)) +
geom_point(color = "firebrick4", fill = "Gold") + geom_smooth(span = 4) + x
lab("Indicador Exposição") + ylab("Indicador Interação");
p2 <- ggplot(Crossover2016, mapping = aes(x = Exposição, y = Apropriação))
+ geom_point(color = "firebrick4", fill = "DeepSkyBlue") + geom_smooth(span
= 4) + xlab("Indicador Exposição") + ylab("Indicador Apropriação");
p3 <- ggplot(Crossover2016, mapping = aes(x = Interação, y = Apropriação))
+ geom_point(color = "firebrick4", fill = "#00441B") + geom_smooth(span = 4
) + xlab("Indicador Interação") + ylab("Indicador Apropriação");

```

```

#p4 <- ggplot(Crossover2016, mapping = aes(x = Normalização_Exposição, y =
Normalização_Interação)) + geom_point (color = "DeepPink1", fill = "DeepPin
k1") + geom_smooth (span = 0.3) + xlab("Apropriação");
grid.arrange(p1,p2, p3, top = "", nrow=3, ncol=1)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

#Gráfico de correlação com índices normalizados - 2016
p1 <- ggplot(Crossover2016, mapping = aes(x = Normalização_Exposição, y = N
ormalização_Interação)) + geom_point(color = "goldenrod4", fill = "Gold") +
geom_smooth(span = 4) + xlab("Indicador Exposição") + ylab("Indicador Inter
ação");
p2 <- ggplot(Crossover2016, mapping = aes(x = Normalização_Exposição, y = N
ormalização_Apropriação)) + geom_point(color = "goldenrod4", fill = "DeepSk
yBlue") + geom_smooth(span = 4) + xlab("Indicador Exposição") + ylab("Indic
ador Apropriação");
p3 <- ggplot(Crossover2016, mapping = aes(x = Normalização_Interação, y = N
ormalização_Apropriação)) + geom_point(color = "goldenrod4", fill = "#00441
B") + geom_smooth(span = 4) + xlab("Indicador Interação") + ylab("Indicador
Apropriação");
#p4 <- ggplot(Crossover2016, mapping = aes(x = Normalização_Exposição, y =
Normalização_Interação)) + geom_point (color = "goldenrod4", fill = "DeepPi
nk1") + geom_smooth (span = 0.3) + xlab("Apropriação");
grid.arrange(p1,p2, p3, top = "", nrow=3, ncol=1)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

#Criação de um data.frame somente com os dados a ser trabalhados nos gráfic
os de Correlação
Cross16 <-Crossover2016[c("Exposição", "Interação", "Apropriação")]
Cross16n <-Crossover2016[c("Normalização_Exposição", "Normalização_Interaçã
o", "Normalização_Apropriação")]

#Gráficos de correlação de Pearson, Spearman e Kendall

panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, col.line="red
", lower.panel = panel.smooth, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))

  # Função para calcular os modelos de correlação
  rp = round(cor(x, y, method = "pearson", use = "complete.obs"),
            digits = 2)
  text(0.5, 0.75, paste("Pearson's r = ", rp))

  rs = round(cor(x, y, method = "spearman", use = "complete.obs"),
            digits = 2)

```

```

text(0.5, 0.5, bquote("Spearman's  $\rho$ " = " $\rho$ .(rs)"))

rk = round(cor(x, y, method = "kendall", use = "complete.obs"),
           digits = 2)
text(0.5, 0.25, bquote("Kendall's  $\tau$ " = " $\tau$ .(rk)"))
}

panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}

pairs(Cross16n, diag.panel = panel.hist, upper.panel = panel.cor)

pairs(Cross16n, diag.panel = panel.hist, upper.panel = panel.cor,
      lower.panel = panel.smooth, main = "Gráfico de Correlação dos Indicadores")

names(Crossover2016)

## [1] "Posição" "Altmetric_ID"
## [3] "Titulo" "Publicação"
## [5] "Data_Publicação" "Adicionado"
## [7] "Dias" "DOI"
## [9] "Link" "Pontuação"
## [11] "Tipo" "Area"
## [13] "Resumo" "Twitter"
## [15] "Facebook" "Google+"
## [17] "Weibo" "Pinterest"
## [19] "Videos" "Mendeley"
## [21] "Connotea" "Citeulike"
## [23] "Reddit" "Notícias"
## [25] "Blogs" "Governo"
## [27] "Wikipedia" "F1000"
## [29] "QNA" "Peer_Review"
## [31] "Linkedin" "Exposição"
## [33] "Interação" "Apropriação"
## [35] "Total" "NTwitter"
## [37] "NFacebook" "NGoogle+"
## [39] "NWeibo" "NPinterest"
## [41] "NVideos" "NMendeley"
## [43] "NConnotea" "NCiteulike"
## [45] "NReddit" "NNotícias"
## [47] "NBlogs" "NGoverno"
## [49] "NWikipedia" "NF1000"
## [51] "NQNA" "NPeer_Review"
## [53] "NLinkedin" "Normalização_Exposição"
## [55] "Normalização_Interação" "Normalização_Apropriação"
## [57] "Normalização_Total" "TwitterLog"

```

```

## [59] "TwitterLog2"          "TwitterLog10"
## [61] "TwitterLog1p"        "TwitterLogb"

View(Crossover2016)

#####

#Criação de um data.frame somente com os dados a ser trabalhados nos gráficos
os de Correlação
CrossMatrix2016 <-Crossover2016[c("Data_Publicação", "Adicionado", "Dias",
"Titulo", "Exposição", "Interação", "Apropriação")]

names(CrossMatrix2016)

## [1] "Data_Publicação" "Adicionado"      "Dias"            "Titulo"
## [5] "Exposição"       "Interação"       "Apropriação"

#Ranking
Crossover2016["Ranking_Exp"] <- (Crossover2016["Exposição"] - 1) / ((Crossover2016["Dias"] + 2) ^ 0.8)

Crossover2016["Ranking_Int"] <- (Crossover2016["Interação"] - 1) / ((Crossover2016["Dias"] + 2) ^ 0.8)

Crossover2016["Ranking_Apr"] <- (Crossover2016["Apropriação"] - 1) / ((Crossover2016["Dias"] + 2) ^ 0.8)

#Taxa de decaimento
lm(formula = log(CrossMatrix2016$Apropriação) ~ CrossMatrix2016$Dias)

##
## Call:
## lm(formula = log(CrossMatrix2016$Apropriação) ~ CrossMatrix2016$Dias)
##
## Coefficients:
##          (Intercept) CrossMatrix2016$Dias
##          5.735965      -0.001566

#Segunda parte - Prova de Conceito com dados do Top 100 de 2017

Crossover2017 <- read_excel(destfile, sheet = "Top100_2017") #Comando que irá importar a tabela do Top 100 de 2017
names(Crossover2017)

## [1] "Score"          "Altmetric ID"
## [3] "Title"         "Journal"
## [5] "Date"          "Subject"
## [7] "Description"   "OA"
## [9] "DOI"           "PubMed ID"
## [11] "arXiv ID"      "Number of news stories"
## [13] "Number of blog posts" "Number of policy documents"
## [15] "Number of tweets" "Number of peer reviews"
## [17] "Number of weibo posts" "Number of Facebook posts"
## [19] "Number of Wikipedia pages" "Number of Google+ posts"
## [21] "Number of LinkedIn posts" "Number of Reddit posts"

```

```
## [23] "Number of pins"           "Number of F1000 posts"
## [25] "Number of Q&A posts"      "Number of videos"
## [27] "Number of syllabi"        "Number of Mendeley readers"
```

#Diferente do Top 100 do ano de 2016, os dados do ano de 2017 não tiveram uma numeração que ordenasse a colocação de cada documento. Dessa forma, o primeiro passo foi criar essa coluna para que fosse possível consultar, em caso de necessidades, as alterações realizadas no ranking original disponibilizado pela ferramenta Almetric.com

```
Crossover2017 <- cbind(c(1:nrow(Crossover2017)),Crossover2017)
```

#Alterar os nomes das colunas 2017

```
names(Crossover2017) <- c("Posição","Pontuação","Almetric_ID","Titulo","Publicação","Data_Publicação","Area","Resumo","Acesso","DOI","PubMed ID","arXiv ID","Notícias","Blogs","Governo","Twitter","Peer_Review","Weibo","Facebook","Wikipedia","Google+","Linkedin","Reddit","Pinterest","F1000","QNA","Videos","Syllabi","Mendeley")
```

#Inclusão da coluna data da captura dos dados

```
Crossover2017$Adicionado <- "2017-12-08 12:50:45"
```

#Comando utilizado para formatar as colunas de data

```
Crossover2017["Adicionado"] <- format(as.Date(Crossover2017$Adicionado,format="%Y-%m-%d %H:%M:%S"),format="%Y-%m-%d")
```

```
Crossover2017["Data_Publicação"] <- format(as.Date(Crossover2017$Data_Publicação,format="%Y-%m-%d %H:%M:%S"),format="%Y-%m-%d")
```

#Calcular a diferente entre a data de adição e a data de atualização

```
Crossover2017$Adicionado <- as.Date(Crossover2017$Adicionado)
Crossover2017["Dias"] <- abs(as.numeric(difftime((Crossover2017$Adicionado), as.Date(Crossover2017$Data_Publicação), units = "auto")))
```

#Apesar de não influenciar nos resultados foram excluídas colunas que não teriam utilidade para o estudo. No entanto, a exclusão foi somente por precisismo dos autores.

```
excluir <- c("Acesso", "PubMed ID", "arXiv ID")
```

```
Crossover2017 <- Crossover2017[,!(names(Crossover2017)%in% excluir)]
```

#Com a distribuição correta dos dados foi realizado o somatório de cada coluna do indicador Exposição

```
Crossover2017["Exposição"] <- Crossover2017["Twitter"] +
  Crossover2017["Facebook"] +
  Crossover2017["Google+"] +
  Crossover2017["Weibo"] +
  Crossover2017["Pinterest"] +
  Crossover2017["Videos"]
```

#O mesmo cálculo foi realizado para o indicador Interação

```
Crossover2017["Interação"] <- Crossover2017["Mendeley"] +
  Crossover2017["Syllabi"] +
  Crossover2017["Reddit"]
```

#O último indicador foi Apropriação que também teve os dados das redes somados

```
Crossover2017["Apropriação"] <- Crossover2017["Notícias"] +
  Crossover2017["Blogs"] +
  Crossover2017["Governo"] +
  Crossover2017["Wikipedia"] +
  Crossover2017["F1000"] +
  Crossover2017["QNA"] +
  Crossover2017["Peer_Review"] +
  Crossover2017["Linkedin"]
```

#Somar as colunas dos três indicadores - O mesmo comando poderia ser utilizado na soma individual de todas as colunas das redes sociais

```
Crossover2017["Total"] <- Crossover2017["Exposição"] + Crossover2017["Interação"] +
  Crossover2017["Apropriação"]
```

#A primeira visualização dos dados após a transformação tem como objetivo identificar como estão distribuídos em um gráfico de barras

```
barplot(Crossover2017$Total, main="Distribuição dos dados", xlab="Total de Documentos",
  ylab="Total de Postagens", col="Gold", cex.names = 0.3, xaxs = "i")
grid(nx=NA, ny=NULL)
```

```
barplot(Crossover2017$Exposição, main="Distribuição dos dados", xlab="Total de Documentos",
  ylab="Total de Postagens", col="DeepSkyBlue", cex.names = 0.3, xaxs = "i", add=TRUE)
```

```
barplot(Crossover2017$Interação, main="Distribuição dos dados", xlab="Total de Documentos",
  ylab="Total de Postagens", col="#00441B", cex.names = 0.8, xaxs = "i", add=TRUE)
```

```
barplot(Crossover2017$Apropriação, main="Distribuição dos dados", xlab="Total de Documentos",
  ylab="Total de Postagens", col="DeepPink1", cex.names = 0.8, xaxs = "i", add=TRUE)
```

```
legend("topright", legend=c("Total", "Exposição", "Interação", "Apropriação"),
  col=c("Gold", "DeepSkyBlue", "#00441B", "DeepPink1"),
  lty=c(1,1,1,1), title("example of line plot"), ncol=1)
```

#Em seguida foram gerados os gráficos individuais que proporcionam melhor visualização dos dados e que foram inseridos na Dissertação

```
barplot(Crossover2017$Total, names.arg = Crossover2017$Posição, font.axis = 1,
  main="", xlab="Total de Documentos", ylab="Total de Envolvimentos", font.lab = 2,
  col.lab = "black", ylim=c(0, 10000), col="Gold", cex.names = 0.8, xaxs = "i")
grid(nx=NA, ny=NULL)
```

```
barplot(Crossover2017$Exposição, names.arg = Crossover2017$Posição, font.axis = 1,
  main="", xlab="Total de Documentos", ylab="Total de Envolvimentos", font.lab = 2,
  col.lab = "black", ylim=c(0, 10000), col="DeepSkyBlue", xaxs = "i")
```

```

= "i")
grid(nx=NA, ny=NULL)

barplot(Crossover2017$Interação, names.arg = Crossover2017$Posição, font.ax
is = 1, main="", xlab="Total de Documentos", ylab="Total de Envolvimentos",
font.lab = 2, col.lab = "black", ylim=c(0, 10000), col="#00441B", xaxs = "i
")
grid(nx=NA, ny=NULL)

barplot(Crossover2017$Apropriação, names.arg = Crossover2017$Posição, font.
axis = 1, main="", xlab="Total de Documentos", ylab="Total de Envolvimentos
", font.lab = 2, col.lab = "black", ylim=c(0, 10000), col="DeepPink1", xaxs
= "i")
grid(nx=NA, ny=NULL)

#Normalização das redes que compõem o indicador Exposição
Crossover2017["NTwitter"] <- log(Crossover2017["Twitter"])
Crossover2017["NFacebook"] <- log(Crossover2017["Facebook"])
Crossover2017["NGoogle+"] <- log(Crossover2017["Google+"])
Crossover2017["NWeibo"] <- log(Crossover2017["Weibo"])
Crossover2017["NPinterest"] <- log(Crossover2017["Pinterest"])
Crossover2017["NVideos"] <- log(Crossover2017["Videos"])

#A visualização dos dados possibilitou identificar que alguns resultados ti
veram resultado infinito. Então, estes dados serão convertidos para 0 (zero
)
Crossover2017$NTwitter[!is.finite(Crossover2017$NTwitter)] <- 0
Crossover2017$NFacebook[!is.finite(Crossover2017$NFacebook)] <- 0
Crossover2017$"NGoogle+"[!is.finite(Crossover2017$"NGoogle+")] <- 0
Crossover2017$NWeibo[!is.finite(Crossover2017$NWeibo)] <- 0
Crossover2017$NPinterest[!is.finite(Crossover2017$NPinterest)] <- 0
Crossover2017$NVideos[!is.finite(Crossover2017$NVideos)] <- 0

#Normalização das redes que compõem o indicador Interação
Crossover2017["NMendeley"] <- log(Crossover2017["Mendeley"])
Crossover2017["NSyllabi"] <- log(Crossover2017["Syllabi"])
Crossover2017["NReddit"] <- log(Crossover2017["Reddit"])

#A visualização dos dados possibilitou identificar que alguns resultados ti
veram resultado infinito. Então, estes dados serão convertidos para 0 (zero
)
Crossover2017$NMendeley[!is.finite(Crossover2017$NMendeley)] <- 0
Crossover2017$NSyllabi[!is.finite(Crossover2017$NSyllabi)] <- 0
Crossover2017$NReddit[!is.finite(Crossover2017$NReddit)] <- 0

#Normalização das redes que compõem o indicador Apropriação
Crossover2017["NNoticias"] <- log(Crossover2017["Notícias"])
Crossover2017["NBlogs"] <- log(Crossover2017["Blogs"])
Crossover2017["NGoverno"] <- log(Crossover2017["Governo"])
Crossover2017["NWikipedia"] <- log(Crossover2017["Wikipedia"])
Crossover2017["NF1000"] <- log(Crossover2017["F1000"])
Crossover2017["NQNA"] <- log(Crossover2017["QNA"])
Crossover2017["NPeer_Review"] <- log(Crossover2017["Peer_Review"])

```

```

Crossover2017["NLinkedin"] <- log(Crossover2017["Linkedin"])

#A visualização dos dados possibilitou identificar que alguns resultados ti
veram resultado infinito. Então, estes dados serão convertidos para 0 (zero
)
Crossover2017$NNoticias[!is.finite(Crossover2017$NNoticias)] <- 0
Crossover2017$NBlogs[!is.finite(Crossover2017$NBlogs)] <- 0
Crossover2017$NGoverno[!is.finite(Crossover2017$NGoverno)] <- 0
Crossover2017$NWikipedia[!is.finite(Crossover2017$NWikipedia)] <- 0
Crossover2017$NF1000[!is.finite(Crossover2017$NF1000)] <- 0
Crossover2017$NQNA[!is.finite(Crossover2017$NQNA)] <- 0
Crossover2017$NPeer_Review[!is.finite(Crossover2017$NPeer_Review)] <- 0
Crossover2017$NLinkedin[!is.finite(Crossover2017$NLinkedin)] <- 0

#Com a distribuição correta dos dados foi realizado o somatório de cada co
Luna do indicador Exposição
Crossover2017["Normalização_Exposição"] <- Crossover2017["NTwitter"] +
  Crossover2017["NFacebook"] +
  Crossover2017["NGoogle+"] +
  Crossover2017["NWeibo"] +
  Crossover2017["NPinterest"] +
  Crossover2017["NVideos"]

#O mesmo cálculo foi realizado para o indicador Interação
Crossover2017["Normalização_Interação"] <- Crossover2017["NMendeley"] +
  Crossover2017["NSyllabi"] +
  Crossover2017["NReddit"]

#O último indicador foi Apropriação que também teve os dados das redes soma
dos
Crossover2017["Normalização_Apropriação"] <- Crossover2017["NNoticias"] +
  Crossover2017["NBlogs"] +
  Crossover2017["NGoverno"] +
  Crossover2017["NWikipedia"] +
  Crossover2017["NF1000"] +
  Crossover2017["NQNA"] +
  Crossover2017["NPeer_Review"] +
  Crossover2017["NLinkedin"]

#A definição que o logaritmo natural seria o mais adequado para a realizaçã
o da pesquisa partiu de testes realizados com outros modelos de transformaç
ões
Crossover2017["Normalização_Exposição1"] <- log(Crossover2017["Exposição"])
Crossover2017["Normalização_Exposição2"] <- log2(Crossover2017["Exposição"
])
Crossover2017["Normalização_Exposição3"] <- log10(Crossover2017["Exposição"
])
Crossover2017["Normalização_Exposição4"] <- log1p(Crossover2017["Exposição"
])
Crossover2017["Normalização_Exposição5"] <- logb(Crossover2017["Exposição"]
)

```

```

)

#A principal transformação aconteceu por meio de Logaritmo natural dos indicadores. O objetivo foi gerar uma visualização harmônica e maior facilidade na construção dos indicadores
Crossover2017["Normalização_Exposição"] <- log(Crossover2017["Exposição"])
Crossover2017["Normalização_Interação"] <- log(Crossover2017["Interação"])
Crossover2017["Normalização_Apropriação"] <-log(Crossover2017["Apropriação"])
)
Crossover2017["Normalização_Total"] <- log(Crossover2017["Total"])
View(Crossover2017) #Comando utilizado para conferir os resultados da transformação dos dados

#A visualização dos dados possibilitou identificar que alguns resultados tiveram resultado infinito. Então, estes dados serão convertidos para 0 (zero)
)
Crossover2017$Normalização_Exposição[!is.finite(Crossover2017$Normalização_Exposição)] <- 0
Crossover2017$Normalização_Interação[!is.finite(Crossover2017$Normalização_Interação)] <- 0
Crossover2017$Normalização_Apropriação[!is.finite(Crossover2017$Normalização_Apropriação)] <- 0
Crossover2017$Normalização_Total[!is.finite(Crossover2017$Normalização_Total)] <- 0

#Testes realizados para verificar como ficou a distribuição dos dados antes e depois da normalização
#Indicador Exposição
qqnorm(Crossover2017$Exposição) #Dados do Indicador Exposição não normalizados
qqline(Crossover2017$Exposição, col="red")

qqnorm(Crossover2017$Normalização_Exposição) #Dados do Indicador Exposição normalizados
qqline(Crossover2017$Normalização_Exposição, col="red")

#Indicador Interação
qqnorm(Crossover2017$Interação) #Dados do Indicador Interação não normalizados
qqline(Crossover2017$Interação, col="red")

qqnorm(Crossover2017$Normalização_Interação) #Dados do Indicador Interação normalizados
qqline(Crossover2017$Normalização_Interação, col="red")

#Indicador Apropriação
qqnorm(Crossover2017$Apropriação) #Dados do Indicador Apropriação não normalizados
qqline(Crossover2017$Apropriação, col="red")

qqnorm(Crossover2017$Normalização_Apropriação) #Dados do Indicador Apropriação normalizados
qqline(Crossover2017$Normalização_Apropriação, col="red")

```

#Após a realização dos testes nos indicadores foi gerado gráficos com os dados normalizados de cada indicador

```
barplot(Crossover2017$Normalização_Total, names.arg = Crossover2017$Posição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Envolvimentos", col="Gold", col.lab = "black", ylim=c(0, 10), xaxs = "i")
grid(nx=NA, ny=NULL)
```

```
barplot(Crossover2017$Normalização_Exposição, names.arg = Crossover2017$Posição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Envolvimentos", col="DeepSkyBlue", col.lab = "black", ylim=c(0, 20), xaxs = "i")
grid(nx=NA, ny=NULL)
```

```
barplot(Crossover2017$Normalização_Interação, names.arg = Crossover2017$Posição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Envolvimentos", col="#00441B", col.lab = "black", ylim=c(0, 20), xaxs = "i")
grid(nx=NA, ny=NULL)
```

```
barplot(Crossover2017$Normalização_Apropriação, names.arg = Crossover2017$Posição, main="", font.lab = 2, xlab="Total de Documentos", ylab="Total de Envolvimentos", col="DeepPink1", col.lab = "black", ylim=c(0, 20), xaxs = "i")
grid(nx=NA, ny=NULL)
```

#Gráfico de correlação com índices não normalizados 2017

```
p1 <- ggplot(Crossover2017, mapping = aes(x = Exposição, y = Interação)) +
  geom_point(color = "firebrick4", fill = "Gold") + geom_smooth(span = 4) +
  xlab("Indicador Exposição") + ylab("Indicador Interação");
p2 <- ggplot(Crossover2017, mapping = aes(x = Exposição, y = Apropriação)) +
  geom_point(color = "firebrick4", fill = "DeepSkyBlue") + geom_smooth(span = 4) +
  xlab("Indicador Exposição") + ylab("Indicador Apropriação");
p3 <- ggplot(Crossover2017, mapping = aes(x = Interação, y = Apropriação)) +
  geom_point(color = "firebrick4", fill = "#00441B") + geom_smooth(span = 4) +
  xlab("Indicador Interação") + ylab("Indicador Apropriação");
grid.arrange(p1,p2, p3, top = "", nrow=3, ncol=1)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

#Gráfico de correlação com índices normalizados 2017

```
p1 <- ggplot(Crossover2017, mapping = aes(x = Normalização_Exposição, y = Normalização_Interação)) +
  geom_point(color = "goldenrod4", fill = "Gold") + geom_smooth(span = 4) +
  xlab("Indicador Exposição") + ylab("Indicador Interação");
p2 <- ggplot(Crossover2017, mapping = aes(x = Normalização_Exposição, y = Normalização_Apropriação)) +
  geom_point(color = "goldenrod4", fill = "DeepSkyBlue") + geom_smooth(span = 4) +
  xlab("Indicador Exposição") + ylab("Indicador Apropriação");
p3 <- ggplot(Crossover2017, mapping = aes(x = Normalização_Interação, y = Normalização_Apropriação)) +
  geom_point(color = "goldenrod4", fill = "#00441B") + geom_smooth(span = 4) +
  xlab("Indicador Interação") + ylab("Indicador Apropriação");
grid.arrange(p1,p2, p3, top = "", nrow=3, ncol=1)
```

```

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

#Criação de um data.frame somente com os dados a ser trabalhados nos gráficos
os de Correlação
Cross17 <-Crossover2017[c("Exposição", "Interação", "Apropriação")]
Cross17n <-Crossover2017[c("Normalização_Exposição", "Normalização_Interação",
"Normalização_Apropriação")]

#Gráficos de correlação de Pearson, Spearman e Kendall

panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, col.line="red",
lower.panel = panel.smooth, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))

  # Função para calcular os modelos de correlação
  rp = round(cor(x, y, method = "pearson", use = "complete.obs"),
            digits = 2)
  text(0.5, 0.75, paste("Pearson's r = ", rp))

  rs = round(cor(x, y, method = "spearman", use = "complete.obs"),
            digits = 2)
  text(0.5, 0.5, bquote("Spearman's ~rho~" = "~.(rs)))

  rk = round(cor(x, y, method = "kendall", use = "complete.obs"),
            digits = 2)
  text(0.5, 0.25, bquote("Kendall's ~tau~" = "~.(rk)))
}

panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}

pairs(Cross17n, diag.panel = panel.hist, upper.panel = panel.cor)
pairs(Cross17n, diag.panel = panel.hist, upper.panel = panel.cor,
      lower.panel = panel.smooth, main = "Gráfico de Correlação dos Indicadores")

names(Crossover2017)

## [1] "Posição" "Pontuação"
## [3] "Altmetric_ID" "Titulo"

```

```
## [5] "Publicação" "Data_Publicação"
## [7] "Area" "Resumo"
## [9] "DOI" "Notícias"
## [11] "Blogs" "Governo"
## [13] "Twitter" "Peer_Review"
## [15] "Weibo" "Facebook"
## [17] "Wikipedia" "Google+"
## [19] "Linkedin" "Reddit"
## [21] "Pinterest" "F1000"
## [23] "QNA" "Videos"
## [25] "Syllabi" "Mendeley"
## [27] "Adicionado" "Dias"
## [29] "Exposição" "Interação"
## [31] "Apropriação" "Total"
## [33] "NTwitter" "NFacebook"
## [35] "NGoogle+" "NWeibo"
## [37] "NPinterest" "NVideos"
## [39] "NMendeley" "NSyllabi"
## [41] "NReddit" "NNoticias"
## [43] "NBlogs" "NGoverno"
## [45] "NWikipedia" "NF1000"
## [47] "NQNA" "NPeer_Review"
## [49] "NLinkedin" "Normalização_Exposição"
## [51] "Normalização_Interação" "Normalização_Apropriação"
## [53] "Normalização_Exposição1" "Normalização_Exposição2"
## [55] "Normalização_Exposição3" "Normalização_Exposição4"
## [57] "Normalização_Exposição5" "Normalização_Total"
```

```
View(Crossover2017)
```

```
#####
```

#Ranking

```
Crossover2017["Ranking_Exp"] <- (Crossover2017["Exposição"] - 1) / ((Crossover2017["Dias"] + 2) ^ 0.8)
```

```
Crossover2017["Ranking_Int"] <- (Crossover2017["Interação"] - 1) / ((Crossover2017["Dias"] + 2) ^ 0.8)
```

```
Crossover2017["Ranking_Apr"] <- (Crossover2017["Apropriação"] - 1) / ((Crossover2017["Dias"] + 2) ^ 0.8)
```

#A construção da tabela dos Apêndices foi modelada no software Microsoft Excel, após a exportação dos dados realizado com o comando abaixo

```
write.csv2(Crossover2017, file = "Crossover2017Final.csv")
```