

CLASSIFICAÇÃO DOS LABORATÓRIOS DE PESQUISA BIOMÉDICA BASEADA EM PUBLICAÇÕES CIENTÍFICAS: um estudo de caso do Instituto Oswaldo Cruz

Ricardo Barros Sampaio
Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil
rsampaio.br@gmail.com

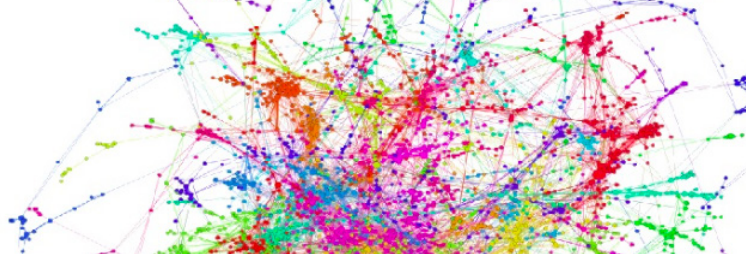
Vitor Hugo Martins
Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil
vitor.martins@ioc.fiocruz.br

Ed Noijons
Leiden University, Leiden, Países Baixos
noyons@cwts.leidenuniv.nl

1 INTRODUÇÃO

A pesquisa científica nos últimos anos evoluiu para uma abordagem mais transdisciplinar, onde a experiência de diferentes áreas é necessária para resolver problemas difíceis e em períodos de tempo mais curtos. No entanto, na prática, a capacidade de facilitar a colaboração entre pesquisadores de diferentes subáreas da pesquisa biomédica ou mesmo classificar seu trabalho é uma tarefa difícil.

O problema de pesquisa trabalhado neste projeto se pautou na seguinte pergunta: existem formas mais efetivas ou eficientes de se agrupar os laboratórios de pesquisa do Instituto Oswaldo Cruz (IOC)? O trabalho realizado teve como objetivo avaliar e apresentar as potencialidades de colaboração, interação e integração entre os laboratórios de pesquisa tomando como base as publicações científicas realizadas pelos seus pesquisadores, as áreas de pesquisa e palavras-chaves relacionadas aos laboratórios do instituto, além das áreas de conhecimento e *Medical*



Subject Headings (MeSH) encontrados nas bases ISI *Web of Science* (WoS) e *PubMed* respectivamente.

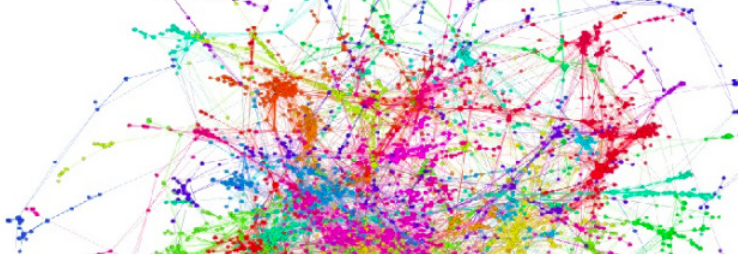
O IOC é a maior unidade de pesquisa dentro da Fiocruz, uma instituição pública de pesquisa médica no Brasil. De acordo com seu site, o instituto foi criado em 1900 e diversificou suas ações para um complexo que gera conhecimento, produtos e serviços na área biomédica. Trabalha nas áreas de pesquisa, desenvolvimento tecnológico e inovação e fornecimento de serviços de referência para o diagnóstico de doenças infecciosas e genéticas e controle vetorial. A base de ação do Instituto são seus 72 laboratórios de pesquisa, desenvolvimento tecnológico e inovação, dedicados ao estudo e geração de produtos e insumos para várias doenças.

A pesquisa desenvolvida pelos autores analisou o IOC, e seus 72 laboratórios e 456 pesquisadores afiliados, e com base em suas publicações nos últimos anos, propôs diferentes classificações usando técnicas comuns à pesquisa científica. Embora as publicações não sejam o único produto desenvolvido por esta instituição de pesquisa, que possui escolas técnicas, especializações e pós-graduações, nos concentramos nas publicações dos seus pesquisadores para construir um argumento sobre como explorar novas possibilidades para arranjos e classificações dos laboratórios.

O trabalho se justifica pela necessidade de uma melhor gestão de recursos de pesquisa, e sua alocação de forma justa e com maior clareza a todos os envolvidos.

2 METODOLOGIA

O relatório utilizou os dados das pesquisas desenvolvidas e publicadas em periódicos científicos que estavam associadas aos pesquisadores da entidade e seus colaboradores. Os dados foram obtidos da plataforma Coleta, base de dados utilizados pelos colaboradores do IOC para registrar seus trabalhos científicos, da base de dados e citações WoS e da base de dados PubMed mantida pela NCBI. Foram utilizados na pesquisa os softwares *Gephi* para visualização das redes, o *VosViewer* para tratamento de publicações científicas de bases indexadas e o *IRaMuTeQ* para análise léxica de textos e palavras-chaves.



A categorização de áreas dentro das publicações da base de dados WoS teve como referência o trabalho de Waltman, Van Eck e Noyons (2010), que utiliza técnicas de mapeamento e agrupamento derivadas do princípio de modularidade. A análise leva em consideração a informação dentro de cada artigo em contraste com a classificação categórica disponibilizada pela instituição, que usa a revista como base de classificação.

O software IRaMuTeQ (Interface R para Análise Multidimensional de Textos e Questionários) foi utilizado para a análise de conteúdo, com base em uma metodologia desenvolvida por Max Reinert na década de 1990 no laboratório de psicologia social da Ecole des hautes études en Ciências Sociais (GUARNACCIA et al., 2015; RODRIGUES et al., 2015).

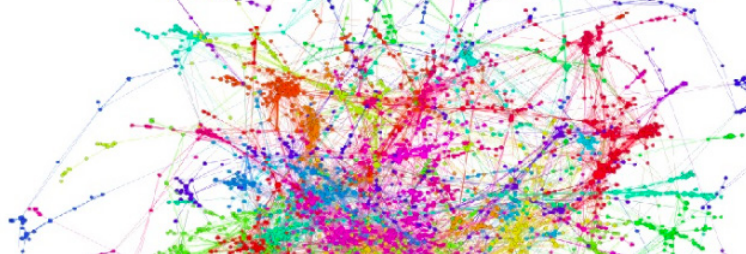
O número de laboratórios avaliados teve uma pequena variação dependendo da fonte de dados utilizada, mas que foi normalizada em 72 ativos e um não mais em funcionamento. O número de colaboradores funcionários do IOC foi de 616 sendo que destes 456 foram considerados como em função de pesquisa. A definição quanto a função de pesquisa não está relacionada ao seu enquadramento funcional na Fiocruz, mas sim as suas atividades no IOC.

Foi utilizada na pesquisa a análise das redes de coautores que aborda os padrões de relações entre autores, permitindo a identificação da disponibilidade e fluxo de troca de recursos entre eles (WASSERMAN; FAUST 1994; NEWMAN, 2001).

A pesquisa teve início com a análise das publicações por pesquisadores e laboratórios do IOC formando as redes de colaboração. Em seguida foi realizada uma análise das áreas de conhecimento das revistas seguida de uma análise semântica e de coocorrência de palavras. Estas etapas estão descritas na sequência apresentada a seguir.

3 RESULTADOS

Quanto aos dados de publicação, esses foram coletados através do repositório do IOC do período de 2010 a 2016 no total de 3.115 publicações. Essas publicações foram divididas nos respectivos laboratórios de acordo com a Tabela 1. Várias publicações foram colocadas em mais



de um laboratório no caso de colaboração entre elas e alguns dos laboratórios deixaram de existir durante o tempo analisado. Essas publicações e seus padrões de colaboração entre os laboratórios foram utilizados para a análise da rede.

TABELA 1 - NÚMERO DE PUBLICAÇÕES POR LABORATÓRIO

Laboratory	Publications	Laboratory	Publications	Laboratory	Publications
LABPMR	134	LADTV	57	LABE	34
LIMUNOFAR	134	LEMC	57	LICV	34
LEAS	130	LESM	53	LTBBF	34
LEAS	130	LPM	51	LEDOC	33
LAHAN	116	LABFLA	50	LAFICAVE	32
LBC	109	LAMICEL	49	LMMV	32
LABAIDS	105	LBqT	49	LBCS	31
LABIMDOE	103	LABTRIP	48	LSO	29
LPT	98	LABHR	47	LABENT	27
LNIRTT	95	LABIMI	46	LIMP	27
LVCA	93	LABINFLA	45	LABMAL	26
LIVEDIH	92	LAPIH	45	LAEFiB	25
LIPMED	88	LABPAT	44	LBI	25
LITEB	81	LABMAM	43	LGH	25
LAPSA	78	LABZOO	43	LABTOXO	21
LIP	78	LBE	43	LHPP	21
LGMM	76	LIC	42	LIV	21
LABDP	73	LAGFB	41	LABIFIV	19
LATOX	71	LUC	41	LEE	13
LABFISI	69	LABMOF	40	LEMEF	13
LVRS	67	LCC	40	LEV	13
LATHEMA	64	LaBECFar	39	LEIP	8
LABDIP	63	LHPV	39	LECEG	7
LAHEP	63	LVM	37	LFB	7
LPL	60	LABMPV	35		

Fonte: Elaborado pelos autores, 2018.



Utilizamos também como elemento de análise a base de dados da WoS e buscamos todas as publicações que tiveram IOC e suas diferentes nomenclaturas encontradas na WoS como organização afiliada pelos autores. Com a pesquisa realizada foi encontrado um total de 3.162 artigos no período de 1972 a 2016 e apenas 1.037 de 2010 a 2016. Esse corpus foi usado para a área de classificação de pesquisa. O número total de publicações foi menor que o extraído do próprio repositório do IOC. Essa diferença ocorre por uma série de motivos, mas para o propósito desta publicação, deixamos o argumento para uma próxima análise.

Dos artigos WoS também extraímos aqueles com o PMID, o identificador do banco de dados PubMed. Dos 3.126 artigos em WoS 2.215 continham o PMID e esses artigos foram utilizados como referência para a análise de coocorrência de MeSH. Entre os nomes mais comuns encontrados, apresentamos os 10 mais frequentes com a quantidade de artigos em que foram encontrados entre parênteses: trypanosoma_cruzi (195); Molecular_sequence_data (187); Filogenia (163); Vetores de insetos (153); Doença de Chagas (144); Reação em cadeia da polimerase (131); Genótipo (119); Prevalência (116); Sequência base (100); Análise de sequência de dados (98).

3.1 REDE DE COLABORAÇÃO

A análise e o agrupamento de colaboração em rede foram aplicados a todas as publicações do IOC de 2010 a 2016. A Figura 1 mostra o resultado e a classificação. No caso dos laboratórios LAHEP e LADTV, que possuem o maior número de colaborações, apenas os vírus são um tema comum de interesse. Um número relativamente pequeno de temas em comparação com outros laboratórios. Quanto aos clusters e áreas de conhecimento, parece haver uma forte relação entre esses dois tipos de agrupamento. Os laboratórios do cluster 1 da Tabela 1 (LABDIP, LABE, LEMEF, LIVEDIH, LNIRTT) também estão no cluster 1 da rede. O cluster 5 (LABDP, LABHR, LABM, LABPAT, LABPMR, LABTRIP, LEDOC, LEE, LESM, LFB, LGMM, LHPP, LHPV) também é encontrado no cluster 2 da rede. O mesmo acontece com o cluster 6 da Tabela



De acordo com a Figura 2, as duas áreas com maior concentração de publicações são as Ciências Biomédicas e da Saúde em verde e as Ciências da Vida e da Terra em amarelo. Mas mesmo dentro dessas áreas existe um acúmulo de publicações em certas revistas ou subáreas do conhecimento, como pode ser visto pelos nós de número 999 e 460. O entendimento sobre esse direcionamento das publicações pode apoiar avaliações tanto na reorganização dos laboratórios quanto na alocação de recursos.

3.3 COCORRÊNCIA DE MESH

Para a análise de coocorrência, utilizamos o software IRaMuTeQ e para plotá-lo usamos o *Gephi*. Os clusters mostrados na Figura 3 separaram-se em tópicos de visualização que eram comuns a alguns tipos de publicações e poderiam ser usados para compreender melhor o campo. Uma análise adicional com especialistas foi necessária para entender completamente as diferenças e validar sua subdivisão.

**FIGURA 3 - REDE DE COOCORRÊNCIA MESH**

Fonte: Elaborado pelos autores, 2018.

4 CONCLUSÃO

A pesquisa desenvolvida é apresentada em um formato preliminar com o objetivo de disponibilizar os dados utilizados nas análises e alguns resultados quantitativos que foram gerados nesse processo. A conclusão deste trabalho requer um envolvimento significativo dos



gestores e pesquisadores da unidade em relação à avaliação de resultados e discussão sobre o assunto. A análise estatística está em sua infância, e os elementos resultantes do envolvimento de especialistas direcionarão os ajustes necessários.

REFERÊNCIAS

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In: International AAAI Conference on Weblogs and Social Media, 3., 2009, San Jose, Califórnia, USA. Anais... AAAI: San Jose, Califórnia, USA, 2009.

CORIAT, B.; WEINSTEIN, O. Organizations, Firms and Institutions in the Generation of Innovation. **Research Policy**, v. 31, p. 273-290, 2002.

FERRARI, P. A.; GIANCARLO, M. Nonlinear Principal Component Analysis as a Tool for the Evaluation of Customer Satisfaction. **Quality Technology & Quantitative Management**, v. 7, n. 2, p. 117-132, 2010.

GUARNACCIA, C.; GIANNONE, F.; FALGARES, G., CALIGARIS, A. O.; SALES-WUILLEMIN, E. Differences in Social Representation of Blood Donation between Donors and Non-Donors: an empirical study. **Blood Transfus.**, v. 14, n. 6, p. 487-489, 2015. doi:10.2450/2015.0048-15.

NEWMAN, M. E. J. The Structure of Scientific Collaboration Networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 4, p. 404-409, 2001.

REINERT, M. Un logiciel d'analyse lexicale. **Cahiers de L'analyse des données**, v. 11, n. 4, p. 471-481, 1986.

vAN ECK, N. J.; WALTMAN, L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. **Scientometrics**, v. 84, n. 2, p. 523-538, 2010.

WALTMAN, L.; vAN ECK, N. J.; NOYONS, E. A Unified Approach to Mapping and Clustering of Bibliometric Networks. **Journal of Informetrics**, v. 4, n. 4, p. 629-635, 2010.

WASSERMAN, S.; FAUST, K. **Social Network Analysis: methods and applications**. Cambridge, ENG and New York: Cambridge University Press, 1994.