



Classificação dos laboratórios de pesquisa biomédica baseada em publicações científicas: o caso do Instituto Oswaldo Cruz

Ricardo Barros Sampaio

Doutor; Universidade de Brasília, Brasília, DF, Brasil;
E-mail: rsampaio.br@gmail.com

Vitor Hugo da Silva Martins

Mestre; Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brasil;
E-mail: vitor.martins@ioc.fiocruz.br

Everard Christiaan Marie Noyons

Doutor; Universidade de Leiden, Leiden, NL, Holanda;
E-mail: noyons@cwts.leidenuniv.nl

Helena Célia de Souza Sacerdote

Doutora; Universidade de Brasília, Brasília, DF, Brasil;
E-mail: helenasacerdote@gmail.com

Resumo: Dada a importância do Instituto Oswaldo Cruz para as pesquisas biomédicas, em nível nacional e internacional, que visam ao desenvolvimento tecnológico, à inovação e ao fornecimento de serviços de referência para o diagnóstico de doenças infecciosas e genéticas e para o controle vetorial, um mapeamento da produtividade dos seus laboratórios pode contribuir para o planejamento dos recursos, considerando-se as necessidades e as prioridades das demandas de saúde regionais. O problema de pesquisa considerou o relato dos gestores a respeito da divisão dos laboratórios não se mostrar eficiente para atender às necessidades de investimento em pesquisas de saúde, devido à grande pulverização de recursos. O objetivo foi avaliar e apresentar as potencialidades de colaboração, de interação e de integração entre os laboratórios de pesquisa, considerando as publicações científicas dos seus pesquisadores, as áreas de pesquisa e as palavras-chave relacionadas aos laboratórios, também as áreas de conhecimento e os *medical subject headings* encontrados nas bases ISI *Web of Science* e PubMed, respectivamente. A metodologia utilizada compreendeu a análise de redes quanto à colaboração e à classificação científica das publicações; e a análise de textos instrumentalizada pelo software IRAMuTeQ. Como conclusão, foram apresentadas possibilidades de agrupamentos com diferentes abordagens. No entanto, esclarece-se que foram utilizadas apenas informações acerca de publicações científicas. Isso pode implicar um viés nos resultados encontrados, havendo, portanto, necessidade de um parecer de especialistas acerca desses resultados. A sugestão é que a pesquisa seja aprofundada, considerando os programas de pós-graduação do Instituto Oswaldo

Cruz e de seus professores, além de trabalhos de dissertação e de tese defendidos nos últimos anos (como o E-lattes) e que os seus resultados sejam apresentados em outras publicações.

Palavras-chave: Colaboração científica. Análise de redes. IRAMuTeQ. Fiocruz. Instituto Oswaldo Cruz.

1 Introdução

As pesquisas biomédicas focalizam o estudo de doenças locais típicas e de interesse universal e possuem

[...] particularidades que influem diretamente nas condições de saúde e na evolução das doenças de cada região, tais como condições climáticas, saneamento básico, desenvolvimento econômico e até mesmo aspectos culturais. (HOPPEN *et al.*, 2017, p. 53).

Por isso, elas são relevantes em contextos locais (HOPPEN *et al.*, 2017, p. 53). Os citados autores defendem que a colaboração entres pesquisadores brasileiros tem uma tendência de proximidade geográfica, que pode se materializar na concentração da produtividade científica em regiões específicas.

Exemplo disso é que no Brasil, as pesquisas bacteriológica e sanitária se iniciaram em 1899, devido à necessidade de produção de soros preventivos e curativos para a peste bubônica no porto de Santos. O laboratório criado para esse fim começou como parte do Instituto Bacteriológico do Estado de São Paulo; no ano de 1901, se tornou independente e foi denominado Instituto Serumterápico do Estado de São Paulo e, finalmente, assumiu o nome de Instituto Butantan (HOPPEN *et al.*, 2017).

Na mesma época, outros centros de pesquisas biomédicas foram criados, como o Instituto Pasteur do Rio de Janeiro, em 1888, com o objetivo de produzir vacinas contra a raiva; e o Instituto Adolfo Lutz, em 1893, em São Paulo (HOPPEN *et al.*, 2017).

Em 1900, foi criado o Instituto Oswaldo Cruz (IOC), que é a maior unidade de pesquisa dentro da Fiocruz, uma instituição pública de pesquisa médica no Brasil. O IOC diversificou suas ações e tornou-se um complexo que

gera conhecimento e produtos, e que presta serviços na área biomédica. Essa unidade trabalha nas áreas de pesquisa, desenvolvimento tecnológico e inovação; no fornecimento de serviços de referência para o diagnóstico de doenças infecciosas e genéticas; além de realizar controle vetorial. A base de ação do instituto são seus 72 laboratórios de pesquisa, desenvolvimento tecnológico e inovação, dedicados ao estudo e à geração de produtos e de insumos para o combate e a prevenção de várias doenças (FIOCRUZ, 2020a).

Dada a importância do IOC para as pesquisas biomédicas, em nível nacional e internacional, o problema de pesquisa deste trabalho considerou a afirmação dos diretores da instituição, que exerciam essa função quando esta pesquisa foi feita, de que a divisão dos laboratórios não se mostrava eficiente para atender às necessidades de investimento em pesquisas de saúde, devido à grande pulverização de recursos. Nesse cenário, um mapeamento da produtividade dos laboratórios poderia contribuir para o planejamento dos recursos, considerando as necessidades e as prioridades das demandas regionais.

Dessa forma, este trabalho se propõe a avaliar e a apresentar as potencialidades de colaboração, de interação e de integração entre os laboratórios do IOC, de modo a subsidiar a gestão na análise e no planejamento de recursos para pesquisas em saúde, de maneira otimizada. O objetivo é descobrir formas mais efetivas ou eficientes de se agrupar os laboratórios de pesquisa do IOC, com base nos dados publicados em periódicos científicos associados aos pesquisadores da entidade e aos seus colaboradores. Para isso, serão analisadas as publicações científicas realizadas pelos seus pesquisadores, as áreas de pesquisa e as palavras-chave relacionadas aos laboratórios do instituto, além de áreas de conhecimento e do *medical subject headings* (MeSH) encontrados nas bases ISI *Web of Science* (WoS) (WEB..., 2020) e PubMed (PUBMED, 2020), respectivamente.

A análise de redes de coautoria se concentra em padrões de relações entre os autores, permitindo identificar a disponibilidade e o fluxo de troca de recursos entre eles (WASSERMAN; FAUST, 1994). Os recursos podem ser de vários tipos, incluindo aqueles tangíveis, como bens, serviços ou dinheiro, e

aqueles intangíveis, como informação e conhecimento (HAYTHORNTHWAITE, 1996).

O papel das universidades e das demais instituições de ciência e tecnologia (C&T) tem sido considerado elemento-chave no processo de inovação nos segmentos baseados em ciência (CORIAT; WEINSTEIN, 2002). Tal entendimento tem impulsionado iniciativas que buscam induzir o desenvolvimento local, com estímulos à transformação das atividades de pesquisa em inovação.

Apesar de diversas abordagens teóricas e explicações acerca do tema, parece haver o consenso de que, em indústrias/segmentos intensivos em ciência e com altos níveis de crescimento (CORIAT; WEINSTEIN, 2002), as redes de colaboração devem ser tratadas e analisadas como dispositivos organizacionais para a coordenação dos processos de aprendizagem por agentes com diferentes competências e habilidades, uma vez que são setores com elevado grau de complexidade, envolvendo um conjunto vasto de profissionais de diferentes áreas.

A análise das redes de colaboração do IOC pode contribuir para a compreensão do padrão de colaboração, do fluxo de conhecimento entre pesquisadores e laboratórios e do modo como os pesquisadores (autores) por áreas consideradas fundamentais para a geração de novas tecnologias se relacionam. Essa análise constitui, portanto, uma ferramenta importante para a gestão estratégica da unidade, que poderá servir de modelo para outras instituições de pesquisa em saúde.

Também o uso dos métodos de análise de conteúdo permite identificar significados de um texto com base nos princípios da análise semântica distributiva automática de dados textuais. Com a análise de conteúdo, também se podem reorganizar sequências de textos e gerar estatísticas de palavras por segmentação ou por parágrafos (SACERDOTE, 2018), o que permite conhecer os temas ou assuntos das pesquisas em saúde em cada laboratório.

Esta seção contextualizou a pesquisa e apresentou seus objetivos, o problema de pesquisa, a justificativa da pesquisa e o método proposto. A seção 2 descreve, mais detalhadamente, os métodos utilizados no estudo; a seção 3

apresenta a análise dos dados e a discussão dos resultados, seguida pela conclusão.

2 Metodologia

A pesquisa utilizou o estudo de caso, que, de acordo com Yin (2010, p. 39), esse tipo de método deveria ser utilizado quando se “[...] desejasse entender um fenômeno na vida real em profundidade, mas esse entendimento englobasse importantes condições contextuais — porque eram altamente pertinentes ao seu fenômeno de estudo.” O método envolveu um modelo exploratório de análise de redes, em que não se tem uma hipótese ou definição pré-concebida dos resultados a serem encontrados, mas, sim, a busca por padrões que possam explicar relações entre as unidades de pesquisa ou entre as áreas do conhecimento. Como complemento, foi utilizada a análise de conteúdo (ou léxica) para atingir os objetivos propostos.

2.1 Análise de redes

Existem diferentes métodos e ferramentas de modelagem de redes, que permitem construir sistemas reais multidimensionais e que são utilizados nas mais diversas áreas do conhecimento para demonstrar a topologia de redes. Esses métodos e ferramentas permitem mensurar as propriedades estruturais envolvidas na rede como a conectividade (como e com quais nós, as ligações se estabelecem) e a centralidade (qual nó possui a melhor conexão ou a maior influência). Cada propriedade é utilizada para a caracterização topológica que, por sua vez, permite a identificação das propriedades das redes (WASSERMAN; FAUST, 1994).

Embora, para analisar redes, sejam usados indicadores de conectividade/coesão da rede e de centralidade/importância dos atores na rede, nesta pesquisa, utilizam-se as métricas acerca de tamanho da rede, conforme se apresenta no Quadro 1.

Quadro 1 – Indicador – tamanho da rede

Métricas (indicadores)	Definição	Significado neste estudo
Tamanho da rede		
Nós	Atores presentes na rede	Quantidade de laboratórios na rede de laboratórios e quantidade de laboratórios e palavras-chave na rede de dois modos
Ligações	Relações entre os nós	Quantidade de relações de coautoria entre os laboratórios presentes na rede e número de vezes que determinada palavra-chave foi utilizada por um laboratório

Fonte: Elaborado pelos autores.

2.2 Análise de conteúdo (léxica)

Para fazer a análise do conteúdo, foi utilizado o software IRAMuTeQ, cujo nome significa “Interface R para Análise Multidimensional de Textos e Questionários”. O IRAMuTeQ é baseado em uma metodologia desenvolvida por Max Reinert na década de 1990, no laboratório de Psicologia Social da École des Hautes Études em Sciences Sociales (IRAMUTEQ, 2020). O software vem sendo amplamente utilizado em sociologia, psicologia, ciências políticas, saúde, entre outras áreas, para a análise quantitativa de dados textuais ou textual-estatísticos (GUARNACCIA *et al.*, 2016).

O software IRAMuTeQ tem a capacidade de analisar textos em diferentes línguas com base nos seus dicionários internos. Após a entrada de dados no software, é realizada uma contagem das palavras e a classificação morfológica de cada uma delas. Como o software não é capaz de interpretar palavras compostas, no caso das palavras-chave, fez-se necessário juntá-las, trocando os espaços vazios entre as palavras por *underscore*.

Com o texto analisado estatisticamente, o usuário pode realizar análises com base nas classificações do próprio texto ou utilizar o algoritmo de classificação desenvolvido por Reinert, também conhecido como classificação hierárquica descendente (CHD). Os resultados das análises são apresentados em formato visual, utilizando a análise fatorial de correspondência (AFC), para organizar sua distribuição espacial em um plano cartesiano. As palavras, então,

são agrupadas, utilizando a classificação desejada, apresentadas em cores diferentes para cada grupo e dispostas no plano cartesiano por meio da AFC.

A visualização dos resultados é um fator de relevância no processo de pesquisa; e, é indicada a sua validação por parte dos especialistas/pesquisadores da unidade, para a devida interpretação desses resultados. As análises apresentadas foram feitas com base nos resultados obtidos e em informações de áreas do conhecimento e de categorizações prévias acerca dos respectivos laboratórios.

2.3 Delineamento da análise

O relatório utilizou os dados das pesquisas desenvolvidas e publicadas em periódicos científicos, associadas aos pesquisadores da entidade e aos seus colaboradores. Os dados foram obtidos da plataforma Coleta, base de dados utilizados pelos colaboradores do IOC para registrar seus trabalhos científicos (FIOCRUZ, 2020b); da base de dados e de citações *Web of Science*, da Thomson Reuters (atualmente Clarivate), e da base de dados PubMed, mantida pelo *National Center for Biotechnology Information* (NCBI), da *National Library of Medicine*. Além dessas bases de dados científicos, também serviram de fonte os currículos Lattes dos funcionários da unidade, segundo levantamento de 2016 do Observatório Fiocruz (OBSERVATÓRIO, 2020).

Os softwares utilizados na pesquisa foram Gephi, para visualização das redes; VosViewer, para tratamento de publicações científicas de bases indexadas (VAN ECK; WALTMAN, 2010); e o IRAMuTeQ, para análise de conteúdo e de palavras-chave.

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

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

As palavras-chave dos artigos da base Coleta e os MeSH dos artigos no PubMed foram analisados e apresentados em formato gráfico, que possibilita demonstrar algumas contribuições da análise de dados lexicais na extração de informações relevantes em documentos de texto.

O número de laboratórios avaliados apresentou uma pequena variação, dependendo da fonte de dados consultada, mas, foi normalizado em 72 laboratórios ativos; e um não mais em funcionamento. O número de colaboradores, funcionários do IOC, foi de 616, dos quais 456 tinham a sua função como de “pesquisa”. A definição da função de pesquisa não está relacionada ao enquadramento do funcionário na Fiocruz, mas, às suas atividades no IOC.

A investigação iniciou-se com a análise das publicações dos pesquisadores e das redes de colaboração formadas por laboratórios do IOC. Em seguida, foi realizada uma análise das áreas de conhecimento das revistas e, posteriormente, realizou-se uma análise semântica e de coocorrência de palavras.

2.4 Dados do currículo Lattes e gestão

É importante ressaltar que, a pesquisa foi realizada com base nas publicações científicas e no perfil dos pesquisadores, segundo os seus respectivos currículos Lattes. De acordo com a expectativa de resultado para reagrupamento dos laboratórios, o recorte adotado deve ser considerado, pois, retrata apenas um elemento do contexto científico no qual os laboratórios estão envolvidos.

Os dados coletados da plataforma Lattes indicam que o número de servidores, independentemente de estarem ocupando ou não função de pesquisa, estão agrupados em 72 laboratórios e em 23 departamentos, incluindo diretorias, vice-diretorias e demais centros. Do total de 616 funcionários, 456 foram

enquadrados em função de pesquisa e 35 indicaram somente o IOC como local de trabalho, em vez de um laboratório ou de um departamento. A distribuição da entrada de servidores na unidade (ou Fiocruz), em função de período, é a seguinte: 129 servidores de 2010 até 2019; 204, de 2001 a 2010; e 283 antes de 2000 (FIOCRUZ, 2020b).

Quanto à data de formação da sua mão de obra, o IOC tem 142 funcionários com titulação máxima informada, depois do ano de 2010; 268 funcionários com formação entre 2000 e 2010; e 204 funcionários com titulação máxima, anterior ao ano 2000. Esse pode ser um indicador de senioridade do corpo laboral da unidade.

2.5 Dados institucionais

O IOC, por intermédio da sua equipe de gestão, publicou a lista de laboratórios e as suas respectivas áreas do conhecimento, conforme definido pelos próprios laboratórios. As áreas do conhecimento totalizam 15 e são divididas em: viroses; doença sexualmente transmissível (DST); vigilância; Chagas; malária; leishmaniose; bacterianas; bioprodutos; crônico-degenerativas; dengue; educação; genômica; helmintoses; imuno; taxonomia. Cada laboratório se relaciona com uma ou mais dessas áreas.

Além dessas áreas do conhecimento, foi disponibilizado, também, o agrupamento sugerido pelos integrantes/representantes dos laboratórios, como forma de definir assuntos ou áreas do conhecimento convergentes. Essa classificação foi utilizada, assim como os demais dados desta pesquisa, como referência para a análise estatística de reagrupamento dessas áreas do conhecimento, conforme o Quadro 2.

Quadro 2 – Agrupamento dos laboratórios de pesquisa do IOC de acordo com os próprios laboratórios

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
LABDIP	LABECFAR	LABENT	LABINFLA	LABDP	LABFLA	LABAIDS	LABTOXO
LABE	LABFISI	LABMAM	LAEFIB	LABHR	LABIFIV	LADTV	
LEMEF	LABIMDOE	LABMPV	LATOX	LABMAL	LABIMI	LAHEP	
LIVEDIH	LEIP	LABZOO	LBC	LABPAT	LABMOF	LEV	
LNIRTT	LIC	LAGFB	LBE	LABPMR	LAFICAVE	LIPMED	
	LIP	LAHAN	LBI	LABTRIP	LATHEMA	LPT	
	LPL	LAMICEL	LBQT	LEDOC	LIV	LVCA	
	LPM	LAPIH	LEAS	LEE		LVM	
	LSO	LAPSA	LITEB	LESM		LVRS	
	LTBBF	LBCS		LFB			
	LUC	LCC		LGMM			
		LEMC		LHPP			
		LGH		LHPV			
		LICV					
		LIMP					
		LIMUNOFAR					
		LMMV					

Fonte: Elaborado pelos autores.

2.6 Dados da ferramenta Coleta

O IOC, como forma de organizar as suas publicações, possui um sistema de cadastramento de todos os artigos científicos publicados com os seus respectivos autores, denominado Coleta. O uso dos dados da ferramenta auxilia na coleta de informações científicas que, por vezes, podem não ser encontradas em buscas nas bases de dados científicos nacionais e internacionais (FIOCRUZ, 2020b).

A análise de redes realizada neste trabalho considera as publicações científicas com base nos dados retirados do Coleta e, com isso, infere-se que todas as publicações da unidade no período de 2010 a 2016 foram contempladas. Nesse período, foram encontradas 3.151 publicações, as quais estão divididas nos respectivos laboratórios conforme a Tabela 1. Percebe-se que o número total de publicações excede 3.151, pois, várias delas são contabilizadas mais de uma vez, no caso de a publicação ter sido feita em colaboração com outro laboratório.

Tabela 1 – Número de publicações por laboratório de 2010 a 2016

<u>Laboratório</u>	<u>Número</u>	<u>Laboratório</u>	<u>Número</u>	<u>Laboratório</u>	<u>Número</u>
--------------------	---------------	--------------------	---------------	--------------------	---------------

	de pub		de pub		de pub
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: Elaborada pelos autores.

2.7 Dados das bases científicas *Web of Science* e *PubMed*

Como forma de aproveitar as classificações e as palavras-chave (MeSH) das bases internacionais, optou-se por uma análise dos resultados delas. Primeiro fez-se uma busca na WoS, que possui melhor qualificação dos dados, com a utilização dos nomes disponíveis para o IOC. Como resultado, foi encontrado um total de 3.126 artigos, no período de 1972 a 2016, dos quais apenas 1.037

eram de 2010 a 2016, um número bem menor quando comparado ao extraído da base Coleta.

Com o objetivo de compreender essa diferença, foi realizada uma busca manual no Coleta, para encontrar o número de publicações indexadas na base WoS. Com isso, descobriu-se haver, no período de 2013 a 2016, 1.825 artigos na base WoS. Essa diferença se deve ao fato de que a busca na base WoS depende da correta classificação dos autores e de sua filiação, ao passo que a busca pelos artigos, manualmente, independe da qualificação dos dados de filiação dos autores nesses artigos.

Para os resultados apresentados, quanto às áreas do conhecimento classificadas pela WoS, considerou-se a busca com os respectivos nomes do IOC, uma vez que não se tinha uma forma adequada (não manual) de busca dos artigos na base.

Uma vez com acesso aos artigos, foi realizada a busca por aqueles que tinham o PMID, que é o identificador da base PubMed. Do total de 3.126 artigos na WoS, 2.215 tinham PMID e eles foram utilizados como referência para uma análise dos MeSH. Com os artigos disponibilizados, foram avaliados os *subject headings* que somaram um total de 2.132 nomes, os quais serão analisados no próximo capítulo. Dentre os nomes com maior ocorrência, apresentamos os dez mais frequentes e, entre parênteses, o número de artigos que foram encontrados associados a cada nome: *trypanosoma_cruzi* (195); *molecular_sequence_data* (187); *phylogeny* (163); *insect_vectors* (153); *chagas_disease* (144); *polymerase_chain_reaction* (131); *genotype* (119); *prevalence* (116); *base_sequence* (100); *sequence_analysis_dna* (98).

Além dos MeSH, foram analisados também os *subheads*, ou classificadores, que somam um total de 81 e servem para melhor qualificar os MeSH quando se classifica um artigo. Os *subheads* mais comuns são apresentados, a seguir, considerando que o número entre parênteses representa a frequência com que esses classificadores foram encontrados nos artigos: *genetics* (643); *isolation_and_purification* (470); *immunology* (459); *parasitology* (451); *classification* (447); *epidemiology* (444); *physiology* (388); *metabolism* (350); *pharmacology* (286); *drug_effects* (257); *analysis* (244).

cada nó; se os nós são da mesma cor, isso é refletido na linha. A distribuição dos nós seguiu o algoritmo Force Atlas 2, da ferramenta Gephi, que tende a mostrar, com mais clareza, a formação dos grupos. No canto inferior esquerdo, está uma legenda com as cores dos *clusters*, seus respectivos números de identificação e a percentagem de nós da rede nesse *cluster* específico.

Os laboratórios do *cluster* 1, do Quadro 2, (LABDIP, LABE, LEMEF, LIVEDIH, LNIRTT), também estão no *cluster* 1, da rede (verde-escuro). O *cluster* 5, do mesmo quadro (LABDP, LABHR, LABMAL, LABPAT, LABPMR, LABTRIP, LEDOC, LEE, LESM, LFB, LGMM, LHPP, LHPV), também é encontrado no *cluster* 2, da rede (verde-claro). O mesmo acontece com o *cluster* 6, do Quadro 2, que está no *cluster* 7, da rede (vermelho). O grupo 7, do Quadro 2, que tem nove laboratórios no total, também tem uma proximidade com o representado na rede (cinza). Isso corrobora o sentimento inicial de agrupamento indicado pelo IOC, no início do trabalho. Algumas diferenças em laboratórios e *clusters* devem ser avaliadas.

A análise do padrão de colaboração nos laboratórios é algo que deve ser avaliado com um entendimento acerca das suas respectivas características. Em uma análise quantitativa, é possível notar, conforme apresentado na Tabela 2, que não há uma relação direta entre o número de publicações e o número de colaborações (grau). Assim, os cinco laboratórios com maior produtividade são diferentes dos cinco laboratórios com maior colaboração dentro da unidade.

Tabela 2 – Os cinco laboratórios com maior produtividade e os cinco laboratórios com maior colaboração

Laboratórios com maior produtividade				Laboratórios com maior colaboração			
Lab	Pub	Cluster	Grau	Lab	Pub	Cluster	Grau
LABPMR	134	2	20	LABIMDOE	103	0	37
LIMUNOFAR	134	4	19	LIVEDIH	93	1	28
LEAS	130	1	14	LESM	53	2	26
LAHAN	116	4	17	LIPMED	88	5	23
LBC	109	0	21	LIP	78	6	23

Fonte: Elaborada pelos autores.

Em uma análise quanto aos *clusters*, o maior deles é o *cluster* 4, na cor lilás, na parte superior da Figura 1, com 26% dos nós da rede. Dois dos laboratórios que mais publicam estão nesse *cluster* - LIMUNOFAR e LAHAN.

Quanto aos *clusters* dos laboratórios que mais colaboram, na parte direita da Tabela 2, nenhum dos laboratórios se encontra no mesmo *cluster*.

Quanto às colaborações entre os laboratórios, existem claras interações que merecem destaque. A Tabela 3 apresenta as colaborações entre laboratórios com peso dez (10) ou maior, ou seja, as colaborações cujos pesquisadores participaram do desenvolvimento e da publicação de dez ou mais trabalhos nos últimos seis anos, até 2019.

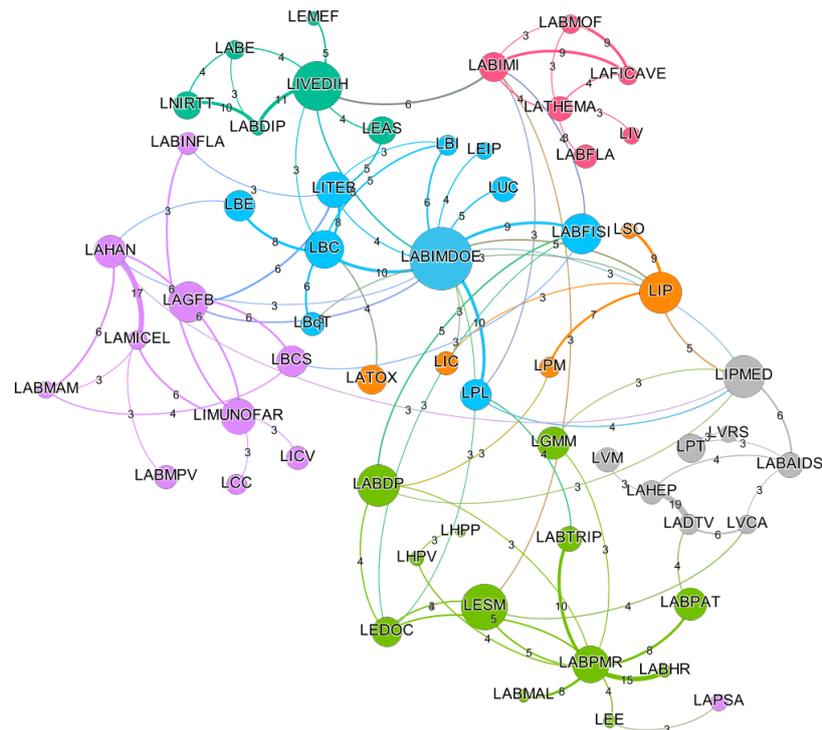
Tabela 3 – Colaborações entre laboratórios

Laboratórios		Número de colaborações
LAHEP	LADTV	19
LAHAN	LAMICEL	17
LABHR	LABPMR	15
LABDIP	LIVEDIH	11
LNIRTT	LABDIP	10
LABIMDOE	LPL	10
LABIMDOE	LBC	10
LABPMR	LABTRIP	10

Fonte: Elaborada pelos autores.

Na Figura 2, é apresentada a mesma rede de laboratórios da figura anterior, com um filtro para os laboratórios que têm, pelo menos, três colaborações com algum outro laboratório, no período de 2010 a 2016. Esse filtro de três ou mais colaborações pode ser aumentado, mas tem por finalidade avaliar o real nível de interação entre os laboratórios e não apenas algo pontual, como uma única publicação. Para fins de visualização e de interpretação das colaborações, foi adicionado, na figura, o peso das linhas representado pelos números entre os nós.

Figura 2 – Rede de colaboração dos laboratórios com grau 3 ou maior



Fonte: Elaborada pelos autores com dados de artigos armazenados na ferramenta Coleta.

Legenda: Os nós representam laboratórios; e as linhas, as colaborações entre eles. O tamanho do nó representa o grau ou número de laboratórios com os quais colaborou o laboratório que o nó representa. A cor dos nós foi gerada com base em um algoritmo de clusterização que considera as relações dentro e fora de determinado *cluster*. A espessura das linhas equivale ao número de colaborações entre os laboratórios, representado pelo número que se encontra sobre a linha, e a uma distância entre os dois nós conectados pela linha. A cor da linha reflete a cor dos nós ligados por ela; se os nós forem da mesma cor, essa cor se repetirá na linha. Se os nós forem de cores diferentes, a cor da linha será uma mistura das duas cores. A distribuição dos nós seguiu o algoritmo Force Atlas 2, da ferramenta Gephi, que prioriza os grupos com base no número de linhas que os une, ao distribuir os nós.

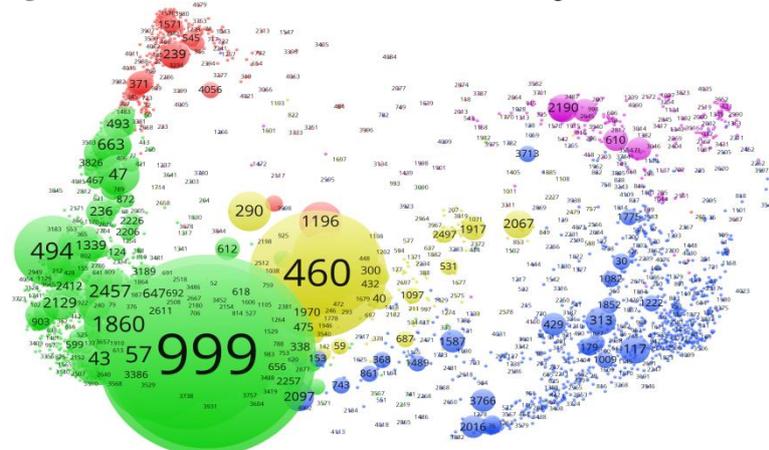
As colaborações apresentadas na Tabela são, novamente, representadas na Figura 2 e podem ser notadas quanto à sua espessura. Em uma avaliação pontual de áreas de interesse, no caso, os laboratórios LAHEP e LADTV, eles têm o maior número de colaborações e apenas o tema “vírus” em comum, segundo a classificação por áreas disponibilizada pela instituição. Esse é um número relativamente pequeno quando comparado com outros laboratórios, que têm muito mais áreas afins, embora o tema seja forte e recorrente o bastante para garantir essa colaboração.

3.2 Grandes áreas do conhecimento

Na análise por grandes áreas do conhecimento, foi utilizada a base da WoS e um modelo de classificação desenvolvido pelo Centro de Estudos de Ciência e Tecnologia (CWTS), da Universidade de Leiden, na Holanda. A base de citações da WoS adota um padrão para a classificação dos artigos indexados, que considera as revistas nas quais esses artigos foram publicados e que se chama *Web of Science Category*. A equipe de pesquisadores do CWTS, que trabalha fortemente com a base WoS, aprimorou essa classificação fazendo uma análise específica de cada artigo e os colocando em grandes grupos. A classificação reúne, em subgrupos, os artigos com base nas revistas em que foram publicados e nas palavras-chave mais comuns entre esses artigos (BUTER; NOYONS; VAN RAAN, 2010; VAN ECK; WALTMAN; NOYONS; BUTER, 2010).

De acordo com a análise realizada, as publicações do IOC podem ser classificadas em cinco grandes grupos, conforme demonstra a Figura 2. As áreas do conhecimento e as suas respectivas cores na figura são: Ciências Biomédicas e da Saúde – *Biomedical and health sciences* (verde); Ciências da Vida e da Terra – *Life and earth sciences* (amarelo); Ciências Físicas e Engenharia – *Physical sciences and engineering* (azul); Ciências Sociais e Humanidades - *Social sciences and humanities* (vermelho); Matemática e Ciência da Computação – *Mathematics and computer science* (rosa).

Figura 2 – Grandes áreas do conhecimento dos artigos indexados na WoS



Fonte: Elaborada pelos autores com uso da ferramenta VosViewer.

Cada círculo representa um subgrupo dentro da grande área e a sua distribuição está associada à proximidade de assuntos. Quanto maior o tamanho do círculo, maior o número de artigos que correspondem àquele subgrupo. O círculo de número 999, o maior da figura, possui 152 artigos distribuídos nas revistas *Memórias do Instituto Oswaldo Cruz*; *Plos Neglected Tropical Diseases*; *Revista da Sociedade Brasileira de Medicina Tropical*; *Acta Tropica* e *American Journal of Tropical Medicine and Hygiene* e utiliza as palavras-chave com maior frequência: *Chagas Disease*; *Reduviidae*; *Hemiptera*; *Triatominae*; *Trypanosoma Cruzi Infection*.

A Tabela 4 apresenta o número de subgrupos e o número de artigos em cada uma das cinco grandes áreas. Desse modo, é possível observar que a grande área na qual se concentram as publicações são as Ciências Biomédicas e da Saúde e as Ciências da Vida e da Terra.

Tabela 4 – Subgrupos e artigos por grande área

Grande área	Grupos	Artigos
Biomédicas e da Saúde – <i>Biomedical and health sciences</i>	236	1196
Ciências da Vida e da Terra – <i>Life and earth sciences</i>	93	406
Ciências Físicas e Engenharia – <i>Physical sciences and engineering</i>	27	37
Ciências Sociais e Humanidades – <i>Social sciences and humanities</i>	12	22
Matemática e Ciência da Computação – <i>Mathematics and computer science</i>	7	14

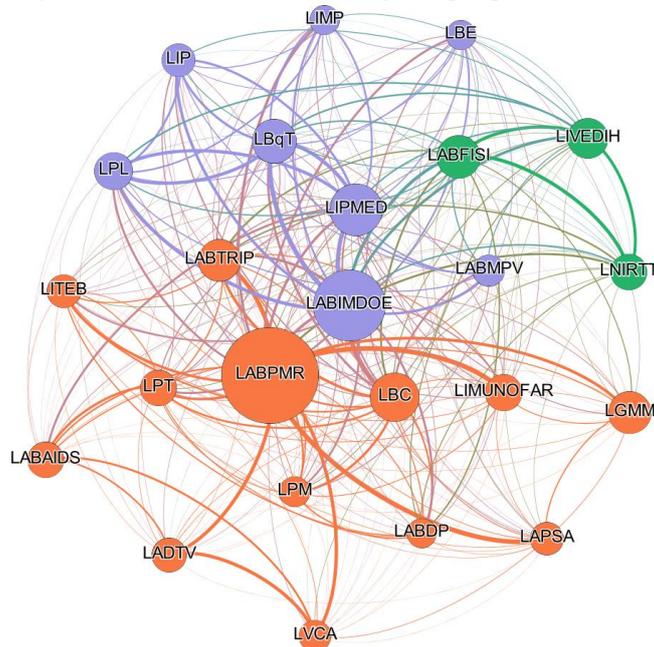
Fonte: Elaborada pelos autores.

3.3 Análise dos títulos

Quanto aos temas abordados pelos laboratórios em suas publicações, foi criada uma rede de dois modos, transformada para um modo, em que as palavras dos títulos foram utilizadas como relação entre os laboratórios. Além disso, uma análise de correlação, com o uso da ferramenta IRAMuTeQ, apresenta o distanciamento dos laboratórios com base no uso e na frequência dessas palavras.

Uma rede de dois modos é definida quando existe mais de um tipo de elemento; neste caso, palavras e laboratórios. Cada elemento de um grupo pode se conectar apenas a elementos do outro grupo. Ou seja, os laboratórios se relacionam com as palavras e elas, por sua vez, somente com os laboratórios. Na transformação para a rede de um modo, mantiveram-se os laboratórios, mas, as linhas agora são as palavras que unem dois laboratórios. A rede é apresentada na Figura 4.

Figura 3 – Rede de laboratórios ligados por palavras-chave



Fonte: Elaborada pelos autores.

Legenda: Cada nó representa um laboratório, e as linhas representam palavras que cada laboratório divide com os outros. O tamanho do nó representa o grau ponderado ou o número de relações com outros laboratórios. A cor dos nós foi gerada com base em um algoritmo de clusterização, que considera as relações dentro e fora de determinado *cluster*. A espessura das linhas representa o número de palavras comuns utilizadas pelos laboratórios. A cor da linha reflete a cor dos nós ligados por ela; se os nós são da mesma cor, essa cor se reflete na linha. A distribuição dos nós seguiu o algoritmo Force Atlas 2, da ferramenta Gephi, que, quando realizada, prioriza os grupos.

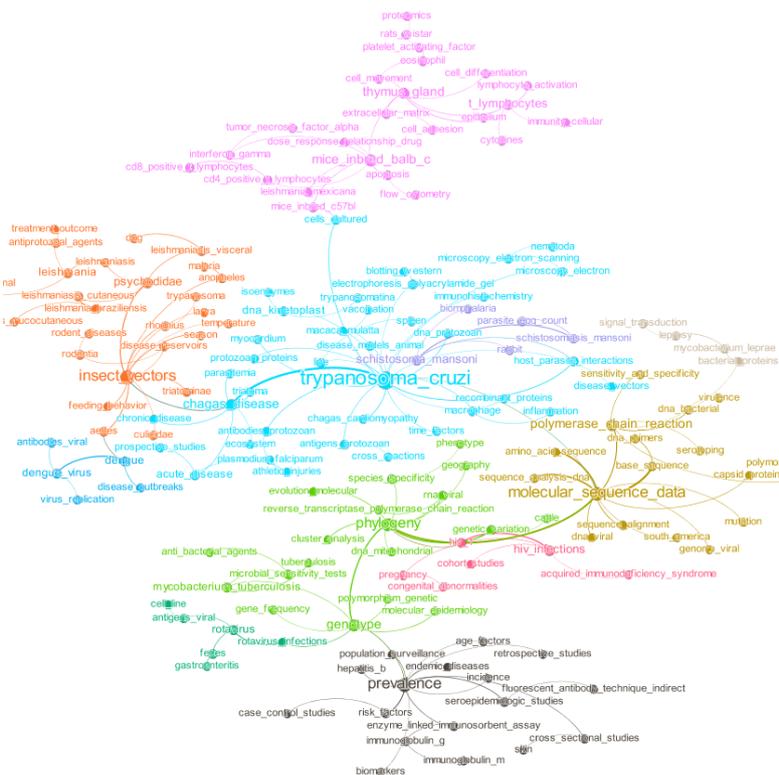
Nesta análise, o número de palavras que eram divididas pelos laboratórios era muito alto, devido ao número de publicações. Para que se pudesse obter uma visualização adequada da rede, optou-se por mostrar apenas os nós e as linhas que continham um número elevado de similaridades no uso

3.4 MeSH e *subheads*

Uma análise similar àquela das palavras dos títulos foi feita por meio dos MeSH e dos seus classificadores, ou *subheads*. Foi aplicada, neste caso, uma análise de coocorrência de palavras e as palavras foram agrupadas para facilitar o reconhecimento do tipo ou da área do conhecimento mais comum.

A coocorrência de palavras demonstrou que áreas específicas se agrupavam; e que essas áreas estavam em consonância com as áreas de interesse dos laboratórios. A palavra *Trypanosoma cruzi*, nome de uma espécie de protozoário, está ligada ao tema Leishmaniose, um dos mais estudados pelos pesquisadores da unidade. Outras áreas de interesse também foram identificadas como centrais, considerando o número de relações com outras palavras. Infelizmente, a validação desses termos junto aos especialistas, para relacionar a estruturação deles com os temas dos laboratórios, não pôde ser realizada. No entanto, entende-se que essa avaliação pode influenciar os temas centrais no agrupamento dos laboratórios.

Figura 5 – Rede MeSH com análise de similaridade e coocorrência



Fonte: Elaborada pelos autores.

com o LAMICEL e os seus pesquisadores participaram do desenvolvimento e da publicação de dez ou mais trabalhos nos últimos seis (6) anos, até 2019. Em primeiro lugar, quanto ao número de colaborações, ficou o LAHEP em colaboração com o LADTV.

Quanto às áreas de publicação, as grandes áreas de conhecimento, na qual se concentram as publicações, são as Ciências Biomédicas e da Saúde e as Ciências da Vida e da Terra.

Quanto à análise dos títulos e das palavras-chave relacionadas às redes dos laboratórios, emergiram três grupos, e o resultado sugere que entre os laboratórios LABPMR, LADTV e LVCA, a palavra destacada foi *vírus*; entre os laboratórios LABIMDOE, LBqT e LIPMED, a palavra *Leishmaniasis*; e entre os laboratórios LABFIFI, LIVEDIH e LNIRTT, a palavra *Hemiptera*.

Cabe ressaltar que, por terem sido utilizadas apenas informações acerca de publicações científicas, pode haver um viés nos resultados encontrados. Assim, sugere-se o aprofundamento da pesquisa, considerando os programas de pós-graduação da unidade e os seus professores, além de trabalhos de dissertação e tese defendidos nos últimos anos.

Para exemplificar alguns resultados decorrentes de tal aprofundamento, foi criado o E-lattes¹, uma plataforma resultante das pesquisas dos autores, que avalia todos os programas de pós-graduação, inclusive da Fiocruz, e não somente do IOC. No caso da Fiocruz, busca-se avaliar a sinergia entre todas as suas unidades e não somente entre laboratórios (SAMPAIO *et al.*, 2019).

Alguns achados dos Programas de Pós-Graduação do IOC são os seguintes perfis por área, que estão disponíveis no E-lattes: Biologia Celular e Molecular – 88 pesquisadores, 894 publicações e 856 orientações; Biologia Parasitária – 55 pesquisadores, 732 publicações e 599 orientações; Medicina Tropical – 50 pesquisadores, 611 publicações e 506 orientações; Ensino em Biociências e Saúde – 35 pesquisadores, 403 publicações e 556 orientações; Biologia Computacional e Sistemas – 27 pesquisadores, 277 publicações e 239 orientações; Biodiversidade e Saúde – 23 pesquisadores, 249 publicações e 210 orientações; Vigilância e Controle de Vetores – 27 pesquisadores, 309 publicações e 303 orientações (SAMPAIO *et al.*, 2019).

4 Conclusão

Considerando a importância do IOC para as pesquisas biomédicas, em nível nacional e internacional, este trabalho visou apresentar um mapeamento da produtividade dos laboratórios para verificar se isso pode contribuir para o planejamento dos recursos, de acordo com as necessidades e as prioridades das demandas regionais.

Assim, esta pesquisa teve como objetivo descobrir formas mais efetivas ou eficientes de agrupar os laboratórios de pesquisa do IOC, com base nos dados publicados em periódicos científicos associados aos pesquisadores da entidade e aos seus colaboradores.

A pesquisa desenvolvida foi apresentada em um formato preliminar, disponibilizando os dados utilizados nas análises e alguns resultados quantitativos que foram gerados nesse processo. Conclui-se que são necessários um envolvimento significativo dos gestores e dos pesquisadores da unidade na avaliação de resultados e uma discussão mais aprofundada acerca do assunto. A partir dos resultados, é possível explorar novas possibilidades de arranjos e de classificações dos laboratórios, uma melhor gestão de recursos de pesquisa e uma alocação desses recursos de forma justa e com maior clareza a todos os envolvidos.

Os resultados encontrados sugerem que pode haver uma diversidade de modos de avaliar ou de reestruturar os grupos de pesquisa, considerando as colaborações, os termos utilizados nos textos e a evolução do conhecimento. Esses são apenas alguns elementos que podem ajudar a caracterizar os laboratórios, não apenas os do IOC, mas, outros tipos de organizações de pesquisa. No entanto, mesmo se houver laboratórios com temas similares, garantir uma colaboração requer um esforço continuado.

Cabe ressaltar que os achados desta pesquisa foram transmitidos aos gestores do IOC com sugestões de implementação/organização dos seus laboratórios. No entanto, ainda não foi possível obter informações acerca de uma possível ação nesse sentido, por parte da gestão.

Devido à utilização exclusiva de informações de publicações científicas nesta pesquisa, sugere-se a continuidade dos estudos baseados no E-lattes, para a análise dos programas de pós-graduação da unidade e de seus professores e de trabalhos de dissertação e de tese defendidos nos últimos anos, que seriam comparados com os resultados encontrados nesta pesquisa. Também se sugere, por fim, que tais achados sejam apresentados em outras publicações.

Referências

- BUTER, R. K.; NOYONS, E. C. M.; VAN RAAN, A. F. J. Identification of converging research areas using publication and citation data. **Research Evaluation**, [s. l.], v. 19, n. 1, p. 19-27, 1 mar. 2010. Oxford University Press (OUP). Disponível em: <http://dx.doi.org/10.3152/095820210x492503>. Acesso em: 04 set. 2019.
- CORIAT, B.; WEINSTEIN, O. Organizations, firms and institutions in the Generation of Innovation. **Research Policy**, [s. l.], v. 31, n. 2, p. 273-290, 2002.
- FIOCRUZ. **O instituto**. Rio de Janeiro: Instituto Oswaldo Cruz, 2020a.
- FIOCRUZ. **Coleta**: sistema integrado de coleta de dados do IOC. Rio de Janeiro: Instituto Oswaldo Cruz, 2020b.
- GUARNACCIA, C. *et al.* Differences in social representation of blood donation between donors and non-donors: an empirical study. **Blood Transfusion**, [s. l.], v. 14, n. 6, p. 487-489, 2016.
- HAYTHORNTHWAITE, C. Social network analysis: an approach and technique for the study of information exchange. **Library & Information Science Research**, [s. l.], v. 18, n. 4, p. 323-342, 1996.
- HOPPEN, N. H. F. *et al.* Distribuição geográfica da produção e colaboração científica brasileira nas Ciências Biomédicas. **Em Questão**, Porto Alegre, v. 23, n. especial, p. 50-73, 2017.
- IRAMUTEQ. **Manual do aplicativo IRAMuTeQ**. [S. l.: s. n.], 2020.
- NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences of the United States of America**, Santa Fe, v. 98, n. 2, p. 404-409, 2001.
- OBSERVATÓRIO Fiocruz. [S. l.: s. n.], 2020. Disponível em: <http://observatorio.fiocruz.br/indicadores>. Acesso em: 20 fev. 2020.

PUBMED. [S. l.: s. n.], 2020. Disponível em:
<https://www.ncbi.nlm.nih.gov/pubmed>. Acesso em: 20 fev. 2020.

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

SACERDOTE, H. C. S. **A mediação segundo Feuerstein e o uso da informação em educação on-line**. 2018. Tese (Doutorado em Ciência da Informação) - Curso de Pós-graduação em Ciência da Informação, Universidade de Brasília, Brasília, 2018.

SAMPAIO, R. B. *et al.* E-lattes: a 11 new framework in R language for analysis of the lattes curriculum. *In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS, 17., 2019, Rome, Italy. Anais [...]. Rome: ISSI Society, 2019. p. 2660-2661.*

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

VAN ECK, N. J. WALTMAN, L.; NOYONS, E. C. M.; BUTER, R. K. Automatic term identification for bibliometric mapping. **Scientometrics**, [s. l.], v. 82, n. 3, p. 581-596, 11 fev. 2010. Springer Science and Business Media LLC. Disponível em: <https://doi.org/10.1007/s11192-010-0173-0>. Acesso em: 22 out. 2019.

WALTMAN, L.; VAN ECK, N. J.; NOYONS, E. A unified approach to mapping and clustering of bibliometric networks. **Journal of Informetrics**, [s. l.], v. 4, n. 4, p. 629-635, 2010.

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

WEB of science. [s. l.]: Clarivate, 2020. Disponível em:
<https://webofknowledge.com/>. Acesso em: 20 fev. 2020.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010.

A classification of biomedical research laboratories based on scientific papers: the case of the Oswaldo Cruz Institute

Abstract: Given the importance of the Oswaldo Cruz Institute, within Brazil and worldwide, for biomedical research, for technological development and for innovation, and given its major role in providing standard services for the diagnosis of infectious and genetic diseases and for vector control, it is suggested that a mapping of the productivity of its laboratories might contribute to its resource management, according to the needs and priorities of regional health demands. This study took into consideration the account of the managers of that institution that the division of laboratories was not efficient in meeting the investment needs in health research, due to the great dispersion of resources. The goal of this study was to evaluate and to present the potential for collaboration, interaction and integration between Oswaldo Cruz Institute's research facilities by examining the scientific papers published by their researchers and the research areas and keywords related to the laboratories, as well as the areas of knowledge in Web of Science and the MeSH (Medical Subject Headings) indexing the articles found in the PubMed database. We analyzed collaboration networks and a scientific research classification, and we used the software IRAMuTeQ to make quantitative analysis of texts. As a conclusion, possibilities of groupings with different approaches were presented. However, in addition to the need for expert opinion on the results, the study also clarifies that only information about scientific publications was used, which may imply a bias in the results found. We suggest the continuation of the deepening of the research, considering the graduate programs of the Institute and its professors, in addition to dissertation and thesis works defended in recent years (such as E-lattes) and featured in other publications.

Keywords: Scientific Collaboration. Network analysis. IRAMuTeQ. Fiocruz. Instituto Oswaldo Cruz.

Recebido: 14/03/2020

Aceito: 13/06/2020

Como citar:

SAMPAIO, Ricardo Barros; MARTINS, Vitor Hugo da Silva; NOYONS, Everard Christiaan Marie; SACERDOTE, Helena Célia de Souza. Classificação dos Laboratórios de Pesquisa Biomédica baseada em publicações científicas: o caso do Instituto Oswaldo Cruz. **Em Questão**, Porto Alegre, v.27, n. 1, p. 308-335, 2021. Doi: <https://doi.org/10.19132/1808-5245271.308-335>

ⁱ O E-lattes está disponível em: <http://elattes.fiocruz.br>.