

# XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

ISSN 2177-3688

**GT-7 – Produção e comunicação em Ciência, Tecnologia & Inovação**

**A CONTRIBUIÇÃO DAS LINHAS E GRUPOS DE PESQUISA PARA A INSTITUCIONALIZAÇÃO DA PESQUISA NOS PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO NO BRASIL: UMA ANÁLISE COM IRAMUTEQ**

***THE CONTRIBUTION OF LINES AND RESEARCH GROUPS FOR THE INSTITUTIONALIZATION OF RESEARCH IN POSTGRADUATION PROGRAMS IN INFORMATION SCIENCE IN BRAZIL: AN ANALYSIS WITH IRAMUTEQ***

Leilah Santiago Bufrem - Universidade Federal de Pernambuco

Nancy Sánchez-Tarragó - Universidade Federal de Rio Grande do Norte

**Modalidade: Trabalho Completo**

**Resumo:** Objetiva compreender como se produz a articulação entre linhas de pesquisa dos Programas de Pós-graduação e as linhas dos Grupos de pesquisa em Ciência da Informação, esclarecendo as implicações destes elos para a institucionalização da Ciência da Informação brasileira. Realiza análise lexical com a ajuda do software IRaMuTeQ, complementada com uma análise do conteúdo das linhas de pesquisa de um corpus empírico a partir da documentação existente nas páginas institucionais de sete Programas de pós-graduação, reconhecidos e consolidados, utilizando como critério aqueles com teses defendidas até 2016. Descreve os Grupos de pesquisa em Ciência da Informação de instituições selecionadas e identifica as temáticas e sub-temáticas que se refletem nas linhas dos Grupos de pesquisa e nas linhas dos Programas para analisar a correspondência temática entre elas. Conclui que a discussão teórica sobre a institucionalização social e cognitiva da pesquisa nos Programas permite a compreensão dos dados empíricos sobre a constituição de linhas e grupos consolidados, assim como temas de pesquisa como atividades reconhecidas na constituição e delimitação do campo. Observa que a visualização possível graças aos mapas dos grupos e linhas de pesquisa, a partir das análises lexicais, permite a percepção das temáticas correspondentes aos grupos institucionalizados mais representativos e que a articulação entre linhas de pesquisa dos Programas de Pós-graduação e as linhas de pesquisa dos grupos de pesquisa em Ciência da Informação é perceptível em algumas instituições.

**Palavras-Chave:** institucionalização da pesquisa; ciência da informação; Brasil; grupos de pesquisa; linhas de pesquisa.

**Abstract:** It aims to understand how the articulation between research lines of the Postgraduate Programs and the research lines of the Information Science research groups is produced, clarifying the implications of these links for the institutionalization of Brazilian Information Science. It performs lexical analysis with the help of IRaMuTeQ software, complemented with a content analysis of the research lines of an empirical corpus from the existing documentation on the institutional pages of seven recognized and consolidated postgraduate programs, using as criteria those with theses defended until 2016. It describes the research groups in Information Science of selected institutions and identifies the themes and sub-themes that are reflected in the Research Groups and Program lines to analyze the thematic correspondence between them. It concludes that the theoretical discussion about the social and cognitive institutionalization of research in Postgraduate Programs

allows the understanding of empirical data on the constitution of consolidated lines and groups, as well as research themes as activities recognized in the constitution and delimitation of the field. It observes that the visualization possible thanks to the group maps and lines of research, based on lexical analyzes, allows the perception of the themes corresponding to the most representative institutionalized groups and that the articulation between lines of research in the Postgraduate Programs and the lines of research in Information Science research groups is noticeable in some institutions.

**Keywords:** institutionalization of research; information Science; Brazil; research groups; research lines.

## 1 INTRODUÇÃO

Os estudos da Sociologia da Ciência vêm atribuindo às instituições a garantia da organização e das estruturas que viabilizam a produção e as trocas científicas. Analisando esse processo institucional, Shinn e Ragouet (2008, p. 18) enfatizam a contribuição das academias “para o desenvolvimento de normas, para a ancoragem social dos modelos de excelência e para a instauração de uma hierarquia no interior das comunidades científicas”. Entretanto, se considerada a concepção de Ben-David (1977), analisada por Schwartzman (1987, p. 68), a sedimentação do campo científico realizou-se mediante a autonomia da comunidade científica, a garantia à competição entre instituições e entre pares, a descentralização institucional e as novas especialidades. Segundo Schwartzman (1987), os estudos históricos de Bem-David, com foco no contexto cultural, político e institucional do surgimento e das transformações das universidades e das atividades científicas, defendem a constituição de um "papel social" para os cientistas, pelo qual sua atividade fosse reconhecida, prestigiada e protegida (SCHWARTZMAN, 1987, p. 68). Esse movimento emancipatório e de institucionalização científica é caracterizada por suas dimensões cognitiva e social, o que ocorre segundo Whitley (1984), de acordo com critérios, entre os quais aqui se destacam a formação de comunidades científicas, de grupos e linhas de pesquisa, como apoio à formação da identidade do campo e ao sentido da prática científica.

Pode-se argumentar, entretanto, que os esforços em prol da institucionalização de um campo nem sempre resultam na ação coordenada dos pesquisadores relativamente às regras impostas. Esse argumento fortalece-se com a contribuição de Shinn e Ragouet (2008) para os quais a conduta oscilante caracteriza o agente científico e demonstra comportamento seletivo, ora em conformidade com as normas, ora em dissonância completa. Para os pesquisadores, essa conduta volúvel é incompatível com a existência de um *ethos* comum, uma vez que o apelo de ganho pessoal e a ambição podem desfavorecer o espírito crítico com relação a seus próprios resultados (SHINN; RAGOUET, 2008, p. 22).

A institucionalização da pesquisa ocorre alicerçada nas trocas e interdependências entre os sistemas conceituais (teorias, objetos de estudo, linhas de pesquisa) e as estruturas formais que lhe outorgam identidade e visibilidade ao campo (cursos de graduação e pós-graduação, associações profissionais, eventos, publicações, grupos de pesquisa). A

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

problemática presente nos contextos dos programas de pós-graduação (PG) no Brasil tem sido reflexo das mudanças conjunturais e dos acontecimentos mobilizadores das relações de força, influenciando as transformações e processo de ensino e pesquisa. Quando se considera que órgãos de fomento e de avaliação das pesquisas institucionais vêm fortalecendo a prática de criação de grupos e de linhas de investigação, percebe-se que o movimento espontâneo de produção e comunicação científica parece se restringir aos interesses, pessoais, institucionais ou locais.

No caso do Brasil, o desenvolvimento histórico da Ciência da Informação (CI) está vinculado ao estabelecimento dos cursos de Pós-graduação, sendo a definição da área de concentração e suas respectivas linhas de pesquisa elementos importantes para sua institucionalização cognitiva e social. A área de concentração constitui a vocação histórica do Programa, sustentada por um domínio temático conformado por linhas de pesquisa. Estas linhas expressam a especificidade de produção de conhecimento de uma área de concentração e sustentam-se nos docentes/pesquisadores do corpo permanente do Programa. Por outra parte, os grupos de pesquisa são equipes de pesquisadores e estudantes organizados formalmente em torno de uma ou mais linhas de pesquisa e de uma área do conhecimento, com o objetivo de desenvolver pesquisa científica. Levando em consideração que os grupos de pesquisa são constituídos, maiormente, por docentes/pesquisadores vinculados aos Programas de Pós-graduação, cuja produção científica coletiva visa atender requisitos de avaliação institucional, seria de esperar que as ementas e linhas específicas destes grupos estivessem articuladas às ementas das áreas de concentração e linhas de pesquisa dos Programas aos quais estão vinculados, contribuindo assim a uma maior institucionalização da pesquisa do campo.

Atualmente, agendas relativas às necessidades em pesquisa desdobram-se em editais de financiamento específicos para temáticas previamente determinadas, conforme tipos de instituição, ou visando interesses da política de desenvolvimento de um país ou região. Assim, pode-se questionar se os projetos de criação de grupos e de definições de linhas de pesquisa vêm ocorrendo de modo a ampliar as possibilidades de intervenção ou, ao contrário, tendem a afunilar o espectro temático das opções de pesquisas. Pode-se também perguntar se interessa à sociedade a atuação de pesquisadores com linhas de investigação definidas ou preparadas para atender a qualquer tema ou problema de pesquisa relacionado

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

aos campos em que atuam a partir do domínio de diferentes referenciais teóricos metodológicos.

Como parte da necessária introspeção crítica sobre aspectos epistemológicos no campo da CI, procura-se com este trabalho discutir a institucionalização social e cognitiva da pesquisa nos Programas de Pós-graduação em Ciência da Informação no Brasil. Objetiva-se compreender como se produz a articulação entre linhas de pesquisa dos Programas de Pós-graduação e as linhas de pesquisa dos grupos de pesquisa em Ciência da Informação, esclarecendo as implicações destes elos para a institucionalização da Ciência da Informação brasileira. Para isso, se realiza uma análise lexical e de conteúdo dos textos das linhas de pesquisa dos grupos de pesquisa em Ciência da Informação vinculados a instituições com Programas de Pós-graduação em Ciência da Informação e das ementas das linhas dos respectivos Programas.

## **2 SOBRE ÁREAS DE CONCENTRAÇÃO, LINHAS E GRUPOS DE PESQUISA**

Com base nas definições emanadas da Capes, concebe-se a área de concentração como o conjunto que se sustenta por um domínio temático conformado por linhas de pesquisa. Seria, portanto, avocação inicial e histórica do Programa, devendo

[...] indicar, de maneira clara, a área do conhecimento à qual pertence o programa, os contornos gerais de sua especialidade na produção do conhecimento e na formação esperada. É desejável que apresente uma denominação abrangente, pois não se espera que os programas alterem sua área de concentração, a menos no caso de que venha a ser objeto de forte reestruturação. Um Programa pode ter uma ou mais áreas de concentração.(BRASIL, 2016, p. 2)

Um Programa de Pós-graduação (PPG) pode ter uma ou mais áreas de concentração, a depender de sua estrutura, vocação e perfil dos pesquisadores dele participantes. Essa definição se expressa na proposta dos cursos, respeitadas as especificidades das suas modalidades. Quanto às linhas de pesquisa (LP), elas expressam a especificidade de produção de conhecimento de uma área de concentração e sustentam-se nos docentes/pesquisadores do corpo permanente do programa. Portanto, já que representam um recorte específico e bem delimitado dentro da(s) área(s) de concentração, harmonizam-se em quantidade e qualidade com a vocação do programa, a área de concentração e a dimensão e área de competência acadêmica do corpo permanente de docentes.

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

Segundo a concepção da Capes, as LPdevem garantir uma distribuição equilibrada entre os docentes, os projetos de pesquisa do Programa e a articulação de suas ementas com o corpus temático constituído pela produção científica, articulando-se entre si e com as orientações, as disciplinas e as áreas do Programa. Representam “temas aglutinadores de estudos científicos que se fundamentam em tradição investigativa, de onde se originam projetos cujos resultados guardam afinidades entre si” (BRASIL, 2014). São as Linhas de pesquisa que, segundo a Capes (BRASIL, 2016):

- a) determinam os rumos do que vem sendo investigado num dado contexto ou realidade;
- b) limitam as fronteiras de um campo específico de conhecimento;
- c) oferecem orientação teórica para as pesquisas;
- d) estabelecem procedimentos adequados ao processo de construção do conhecimento;
- e) estruturam a ideia da produção circunscrita a um mesmo âmbito temático ou teórico;
- f) constituem em unidades de análise de desempenho especialmente dos cursos de graduação e pós-graduação, da produção dos pesquisadores nos órgãos de fomento à pesquisa, ou em outras instituições.

Seria nas linhas de pesquisa, por meio da convergência e divergência de ideias, das discussões entre os pares, que se materializariam as disciplinas e os currículos, os projetos de pesquisas e as consequentes produções dos professores nas suas diferentes modalidades. Portanto, toda produção intelectual do pesquisador/docente situa-se em uma ou alguma linha de pesquisa. Entretanto, percebe-se que ainda permanecem noções fracas, decorrentes de múltiplos entendimentos e formas de utilização e registro. Essas trocas entre pares têm sido institucionalizadas no Brasil por meio dos grupos de pesquisa.

Considera-se o grupo de pesquisa (GP) como um coletivo de pesquisadores e estudantes de diversas áreas ou de linhas de pesquisa definidas no interior de uma área de conhecimento, organizados com o objetivo de desenvolver investigações científicas. Para ele converge o envolvimento profissional e permanente com atividades de pesquisa pelas quais o trabalho se organiza. Os GP são, portanto, unidades básicas previamente habilitadas para o planejamento e o acompanhamento das atividades de pesquisa e podem compartilhar instalações e equipamentos em atividades de pesquisa individuais ou integradas, nele inseridas.

Os GP, atuantes em Programas de pós-graduação, têm se constituído em espaços de convergência e de diálogo das atividades de pós-graduação e de graduação, por meio de

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

reuniões de pesquisa, promoção de eventos e delineamento de publicações. Para a compreensão desses grupos em seus espaços de pesquisa, consideram-se dois aspectos. Por um lado, que são constituídos em sua maioria por docentes/pesquisadores vinculados aos Programas de Pós-graduação, cuja produção científica coletiva visa atender requisitos de avaliação institucional. Por outro lado, os grupos podem ser analisados em sua vinculação a linhas específicas que os constituem, assim como pelas relações que estabelecem com as áreas de concentração e linhas de pesquisa dos Programas aos quais seus professores estão vinculados, contribuindo assim para a institucionalização da pesquisa do campo. Como espaços de construção e socialização do conhecimento, esses grupos ampliam-se à medida que a CI se institucionaliza no país e que as agências de fomento passam a interferir positivamente para a consolidação de novos programas de pós-graduação em suas modalidades e níveis, como parte do processo de reestruturação produtiva, iniciado na década de 1990, período que o Diretório dos Grupos de Pesquisa do Brasil (DGPB) oficializou os grupos junto ao CNPq. Esta forma de organização contribuiu para o fortalecimento das áreas específicas e da pesquisa em seus diversos domínios.

### **3 MÉTODOS**

A pesquisa apoia-se em corpus empírico construído a partir da documentação relativa a uma amostra dos Programas de pós-graduação em Ciência da Informação e emprega técnicas estatísticas multivariadas sobre textos para auxiliar a análise das linhas de pesquisa dos Grupos de pesquisa e dos Programas de Pós-graduação. Dentre os 27 programas de pós-graduação na área de Ciência da Informação no Brasil, a amostra ficou conformada por sete (7) Programas (Quadro 1), reconhecidos e consolidados, utilizando-se como critério aqueles com teses de doutorado defendidas até 2016. Esse critério enfatiza a trajetória histórica institucional de pesquisa reconhecida, principalmente, por agências reguladoras de pesquisa (CUSTÓDIO, 2018). De cada um procedeu-se a coleta das linhas de pesquisa e de suas respectivas ementas (busca realizada no dia 1 de julho de 2018 nos websites institucionais dos Programas).

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

**Quadro 1. Programas de Pós-Graduação em Ciência da informação reconhecidos e consolidados, com teses defendidas até 2016**

Instituição	Programa de Pós-graduação	Site
Universidade de Brasília (UnB)	Programa de Pós-graduação em Ciência da informação	<a href="http://www.ppgcinf.fci.unb.br/index.php">http://www.ppgcinf.fci.unb.br/index.php</a>
Universidade Federal do Rio de Janeiro (UFRJ) em convenio com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)	Programa de Pós-graduação em Ciência da informação	<a href="http://www.ppgci.ufrj.br/pt/">http://www.ppgci.ufrj.br/pt/</a>
Universidade Federal Fluminense (UFF)	Programa de Pós-graduação em Ciência da informação	<a href="http://www.ci.uff.br/ppgci/">http://www.ci.uff.br/ppgci/</a>
Universidade Estadual Paulista (UNESP/Marília),	Programa de Pós-graduação em Ciência da informação	<a href="http://www.marilia.unesp.br/#!/posci">http://www.marilia.unesp.br/#!/posci</a>
Universidade de São Paulo (USP)	Programa de Pós-graduação em Ciência da informação	<a href="http://www3.eca.usp.br/pos/ppgci">http://www3.eca.usp.br/pos/ppgci</a>
Universidade Federal de Minas Gerais (UFMG)	Programa de Pós-graduação em Ciência da informação	<a href="http://caronte.eci.ufmg.br/ppgci/">http://caronte.eci.ufmg.br/ppgci/</a>
	Programa de Pós-graduação em Gestão e Organização do Conhecimento	<a href="http://ppggoc.eci.ufmg.br/">http://ppggoc.eci.ufmg.br/</a>

**Fonte: Dados da pesquisa**

Em uma segunda etapa, se identificaram os Grupos de Pesquisa (GP) em Ciência da informação, cadastrados no Diretório de Grupos de Pesquisa mantido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e associados às instituições responsáveis pelos Programas selecionados (UnB, UFRJ, UFF, UNESP, USP e UFMG). A busca, realizada no dia 3 de julho de 2018, ampliou-se aos grupos cadastrados pelo Instituto Brasileiro de Informação Científica e Tecnológica (IBICT), sempre que fossem liderados por um pesquisador professor permanente do quadro do Programa de Pós-graduação em Ciência da Informação do convênio IBICT/UFRJ. Na opção “Consulta parametrizada” se realizou a busca pelo nome da instituição responsável pelo Programa. Foram recuperados todos os GP das instituições antes indicadas, cujas Grande área e Área de conhecimento foram predominantemente Ciências Sociais Aplicadas e Ciência da informação, respetivamente. Os GP deviam estar certificados, atualizados, e ser conformados por professores permanentes do Programa. A amostra ficou conformada pelos grupos indicados na Tabela 1.”

**Tabela 1. Quantidade de grupos de pesquisa em Ciência da Informação**

Instituição	Quantidade de Grupos
UnB	10
UFRJ	6
IBICT	8
UFF	4
UNESP	9
USP	4
UFMG	13
<b>Total</b>	<b>54</b>

Fonte: Dados da pesquisa

Para a caracterização de cada Grupo, se coletaram as seguintes variáveis: ano de criação do Grupo, titulação do (s) líder (es), quantidade de pesquisadores, afiliação institucional de cada pesquisador, quantidade de pesquisadores por instituição, quantidade de colaboradores estrangeiros cadastrados e afiliação institucional, quantidade de estudantes de graduação e de pós-graduação<sup>1</sup>, nome das linhas de pesquisa, objetivos da linha e palavras-chaves. As afiliações institucionais dos pesquisadores e colaboradores estrangeiros se identificaram entrando nas páginas do Curriculum Lattes de cada pesquisador.

Empregou-se a análise lexical, que utiliza cálculos efetuados sobre a co-ocorrência de palavras em segmentos de texto por meio de sistemas computacionais, o que permite facilitar a análise de grandes volumes de dados textuais. A detecção de regularidades nos vocabulários do texto sugere a existência de “campos contextuais” ou espaços semânticos específicos (NASCIMENTO; MENANDRO, 2006). Neste caso, a análise lexical teve como objetivo identificar as temáticas e sub-temáticas refletidas nas linhas dos Grupos de pesquisa em Ciência da Informação de instituições selecionadas e nas linhas dos Programas de pós-graduação em Ciência da Informação com doutorados consolidados, descobrir similitudes e diferenças entre as instituições e Programas, e descobrir se existe correspondência temática entre as linhas de pesquisa dos Programas e as linhas dos Grupos de pesquisa associados às instituições promotoras destes Programas. A análise lexical efetuada com a ajuda do software IRaMuTeQ, descrito mais a frente, foi complementada com uma análise do conteúdo das linhas de pesquisa.

Para a realização da análise lexical elaborou-se um corpus textual conformado pelo nome de cada linha, objetivo e palavras chaves, no caso dos Grupos de pesquisa, e nome da

<sup>1</sup>Incluídos no estudo só quando indicada a formação acadêmica

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

linha e ementa, no caso dos Programas de Pós-graduação. A primeira linha de cada bloco textual (por exemplo, um parágrafo conformado pelo nome da linha, seu objetivo e suas palavras chaves) foi introduzida por quatro asteriscos (\*\*\*\*), seguido de um espaço e outro asterisco que introduz cada uma das variáveis independentes. Um travessão baixo separa a variável de seus valores (chamados modalidades em IRaMuTeQ). A sequência é a seguinte: asterisco, código de identificação da variável, travessão baixo e código de valor da variável, como se mostra com o corpus seguinte, preparado com as linhas de pesquisa dos Grupos de pesquisa. A primeira codificação (\*INST\_) indica a variável Instituição responsável pelo Grupo, enquanto \*g\_ é a variável que identifica a numeração atribuída ao Grupo a que corresponde o trecho de texto específico.

\*\*\*\* \*INST\_UnB \*g\_G3UnB

Comunicação e Mediação da Informação. Os estudos desenvolvidos sob esta linha privilegiam a busca pelo entendimento dos fenômenos relacionados ao fluxo da informação, os atores que deles participam e os canais utilizados [...]

De maneira similar, procedeu-se com as linhas dos Programas. A primeira codificação (\*PPG\_) corresponde à variável que identifica o Programa de Pós-graduação, enquanto \*LPPG é a variável que identifica a linha específica.

\*\*\*\* \*PPG\_UnBCI \*LPPG\_1UnB

Pesquisa em Comunicação e Mediação da Informação. Reúne estudos que buscam conhecimentos nos níveis epistemológico, científico e prático sobre processos de comunicação em diversos contextos e setores da sociedade [...]

Os textos foram revisados para corrigir erros de digitação e substituir sinais de pontuação que o software não admite, por exemplo, ponto e vírgula e o hífen que unia algumas palavras compostas (por exemplo, científico-técnico). Estes sinais foram substituídos por vírgula e travessão baixo, respectivamente. Seguindo estes parâmetros, todos os verbos com pronomes se passaram também a forma de próclise. Decidiu-se também unir por travessão baixo algumas palavras compostas para melhor identificá-las durante a análise, por exemplo, ciência\_da\_informação, ciência\_aberta, políticas\_públicas, entre outras.

Foi empregado o software IRaMuTeQ 7.0 alpha 2 (Interface de R para a Análise Multidimensional de Textos e Questionários), software livre desenvolvido por Pierre Ratinaud da Universidade de Toulouse, que funciona como interface do programa estatístico

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

R; deste utilizou-se a versão 3.1.2-win pois foi a que se mostrou mais estável para as análises multivariadas empregadas. O IRaMuTeQ permite realizar análise multidimensional de textos de diferente natureza como perguntas de resposta abertas de questionários, notícias, resumos, entre outras. O software permite gerar contextos (classes lexicais) caracterizados pelo seu vocabulário e pelos segmentos de texto que compartilham este vocabulário (REINERT, 1998, apud MAIA, 2017). Os segmentos de texto podem considerar-se o ambiente que rodeia cada palavra de mais ou menos três linhas, geralmente demarcado automaticamente pelo software em função do tamanho do texto. Com o software se buscam regularidades e correlações estatísticas a partir da frequência de ocorrência e coocorrência de palavras e seu contexto lexical.

O software traz um dicionário em português em formato de arquivo de texto com o vocabulário, suas possíveis variações de gênero e número e a categoria gramatical de cada palavra. O arquivo foi revisado pelas autoras para incorporar alguns termos do corpus textual que não apareciam ou apareciam com uma categoria diferente. Por exemplo, se adicionaram os termos webometria, webmetria e bibliometria, assim como diplomática. Neste último caso se especificou como categoria gramatical nome (substantivo) em lugar de adjetivo. As análises foram feitas utilizando esse dicionário personalizado. A partir deste dicionário o software também realiza a lematização, ou seja, os verbos são reduzidos ao infinitivo, os nomes ao singular e os adjetivos ao masculino singular. Para cada análise, o software permite escolher as categorias gramaticais nela incluídas. Essas são marcadas como formas ativas. Para a presente análise, selecionaram-se substantivos e adjetivos.

Dentre as diferentes análises que permite o software (desde cálculo de frequência de palavras, até análises multivariadas, como a Classificação Hierárquica Descendente (CHD), a Análise fatorial de Correspondência, e a Análise de similitude), nesta pesquisa nos auxiliamos das estatísticas léxicas básicas, que permitem identificar a frequência de ocorrência das palavras do corpus e a Classificação Hierárquica Descendente (CHD), que traz associadas também outras técnicas como a Análise fatorial de Correspondência. A CHD é um tipo de análise de cluster proposto por Reinert (1983). As técnicas de análise de cluster têm como objetivo criar grupos de elementos que compartilhem características semelhantes entre si, e que, por sua vez, se diferenciam de outros grupos formados. A CHD é uma técnica automática de criação de cluster, hierárquica (o objeto ou variável é classificado numa determinada etapa e não pode ser classificado novamente) e descendente, pois parte de um

só grupo que inclui todos os elementos e que se subdivide em grupos/classes em etapas subsequentes (RUIZ BUENO, 2017). Com a aplicação do CHD, os segmentos de texto (ST) do corpus são classificados em função dos seus respectivos vocabulários, e o conjunto deles é repartido segundo a frequência das formas lematizadas. A partir de matrizes cruzando formas lematizadas e segmentos de texto (em repetidos testes do tipo Qui-Quadrado), obtém-se uma classificação definitiva. A análise permite obter classes de ST que, ao mesmo tempo, apresentam vocabulário semelhante entre si, e vocabulário diferente dos segmentos das outras classes. A partir desses procedimentos, o software organiza a análise dos dados em um dendrograma ilustrativo das relações entre as classes e permite a descrição do léxico característico de cada classe e as relações com as variáveis estudadas (CAMARGO; JUSTO, 2016).

IRaMUTEQ também apresenta os resultados por meio de outra técnica multivariada, uma Análise Fatorial de Correspondência feita a partir da CHD, que representa num plano cartesiano as diferentes palavras e variáveis associadas a cada uma das classes da CHD (CAMARGO; JUSTO, 2013). A Análise de correspondência parte de uma matriz com os elementos dispostos em filas e colunas (neste caso, por exemplo, vocabulários segundo variáveis: Instituições, Grupos de pesquisa, Programas de PG e Linhas) e mediante o cálculo de Qui-quadrado proporciona um índice de distância entre as variáveis consideradas. Quanto menor a distância entre as variáveis, mais juntas estas serão representadas num espaço bidimensional (RUIZ BUENO, 2017).

## **4 RESULTADOS**

### **4.1 Descrição dos Grupos de pesquisa em Ciência da Informação**

Na Tabela 2, se apresenta uma caracterização resumida dos GP. Alguns elementos interessantes podem ser destacados. Por exemplo, a UFMG tem a maior quantidade de grupos de pesquisa, seguida pela UNB; contudo, a UNESP e a UFRJ têm maior quantidade de linhas de pesquisa. Embora a UFMG tenha dois PPG<sup>2</sup>, o PPGCI tem a maioria dos GP (11). Com relação às datas de fundação, a UNESP tem o GP ativo mais antigo (1990), dedicado às tecnologias da informação em sua relação com a organização, processamento e gestão da

---

<sup>2</sup>Programa de Pós-graduação em Ciência da Informação (PPGCI) e Programa de Pós-graduação em Gestão e Organização do Conhecimento (PPGGOC)

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

informação. Outro GP de longa data (1991) tem sede no IBICT e está dedicado aos estudos sobre Organização do Conhecimento. De 1994, também em atividade, é o GP dedicado à Comunicação Científica. Finalmente, o GP mais antigo da UFMG (1998) estuda a biblioteca escolar.

Todos os líderes dos GP são doutores e predominam amplamente professores vinculados aos PPG. A quantidade de participantes dos grupos é variável, com alguns grupos que reportam só 2 pessoas cadastradas enquanto outros chegam a reportar até 38. Contudo, não é possível constatar se efetivamente todos os participantes cadastrados estão ativos ou formam parte da história “acadêmica” do grupo. Em quase todos os grupos predominam os membros nacionais, da própria instituição, embora muitos grupos reportem parceria com outras instituições nacionais. Os colaboradores estrangeiros reportados são escassos, e predominam os parceiros da Espanha e Portugal. Os parceiros mais destacados de cada instituição são: no UNB, a UFF (participa em 3 grupos); na UFRJ, a FIOCRUZ (6 grupos); na Unesp, a Universidade Federal do Pará e a USP (6 grupos cada uma), e a Universidade Federal do Ceará e a Universidade Estadual de Londrina (4 grupos cada uma); na UFMG, a Universidade Federal de Rio Grande do Norte (3 grupos). A maioria dos grupos incluem estudantes de pós-graduação, e em menor número, estudantes de graduação.

**Tabela 2. Resumo descritivo dos Grupos de pesquisa em Ciência da Informação**

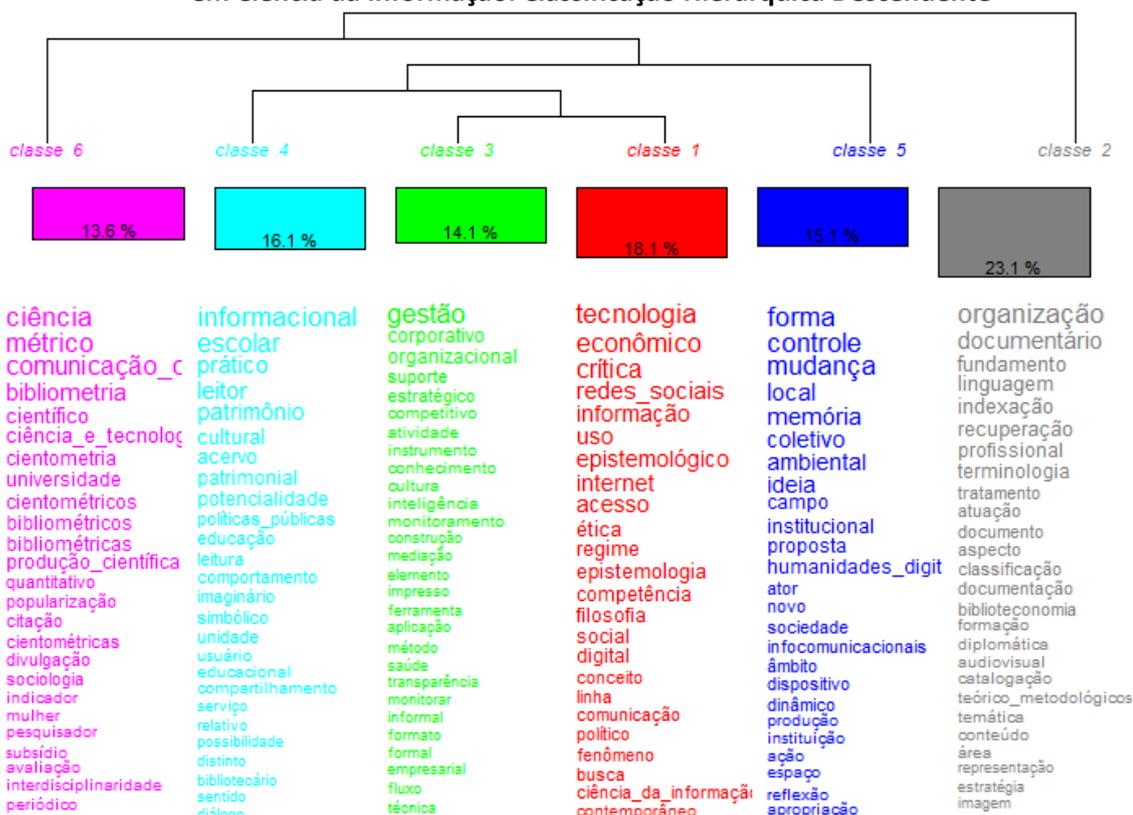
<b>Instituição</b>	<b>No. de GP</b>	<b>Período de criação</b>	<b>Filiação predominante dos líderes</b>	<b>No. de pesquisadores por grupo</b>	<b>No. de instituições envolvidas</b>	<b>No. de linhas de pesquisa</b>
<b>UNB</b>	10	1994-2014	PPGCI UNB	Entre 8 e 15 cada um	47 (45 Brasil e 2 estrangeiras)	23
<b>UFRJ</b>	6	2010-2018	PPGCI UFRJ/IBICT	Entre 2 e 26	48 (43 Brasil e 5 estrangeiras)	33
<b>IBICT</b>	8	1991-2013	PPGCI UFRJ/IBICT			
<b>UFF</b>	4	2007-2014	PPGCI UFF	Entre 4 e 15	22 (Brasil)	8
<b>UNESP</b>	9	1990-2018	PPGCI Unesp	Entre 11 a 38	27 (24 Brasil e 3 estrangeiras)	36
<b>USP</b>	4	2010-2016	PPGCI USP	Entre 2 e 11	11 (8 Brasil e 3 estrangeiras)	13
<b>UFMG</b>	13	1998-2017	PPGCI e PPGGOC	Entre 7 a 28	41 (40 Brasil e 1 estrangeira)	27

Fonte: Dados da pesquisa

#### 4.2 Análise das linhas dos Grupos de pesquisa

A análise lexical foi realizada sobre 54 textos conformados pelos títulos das linhas, seus objetivos e suas palavras chaves. A técnica de Classificação Hierárquica Descendente (CHD) classificou o corpus em seis classes conformadas por vocabulários característicos e homogêneos (Figura 1). O tamanho das palavras indica sua importância dentro desse “mundo léxico”. A classe 2 na cor cinza agrupa a maior porcentagem de vocabulário homogêneo (e distinto dos outros), associado às temáticas de organização da informação (indexação, linguagens documentárias, tratamento, terminologia, representação, classificação, catalogação), relacionadas tanto com a Biblioteconomia e a Documentação, quanto com a Arquivística e a Diplomática.

Figura 1: Dendograma das classes conformadas pelas linhas de pesquisa dos Grupos de pesquisa em Ciência da Informação. Classificação Hierárquica Descendente



Fonte: Dados da pesquisa

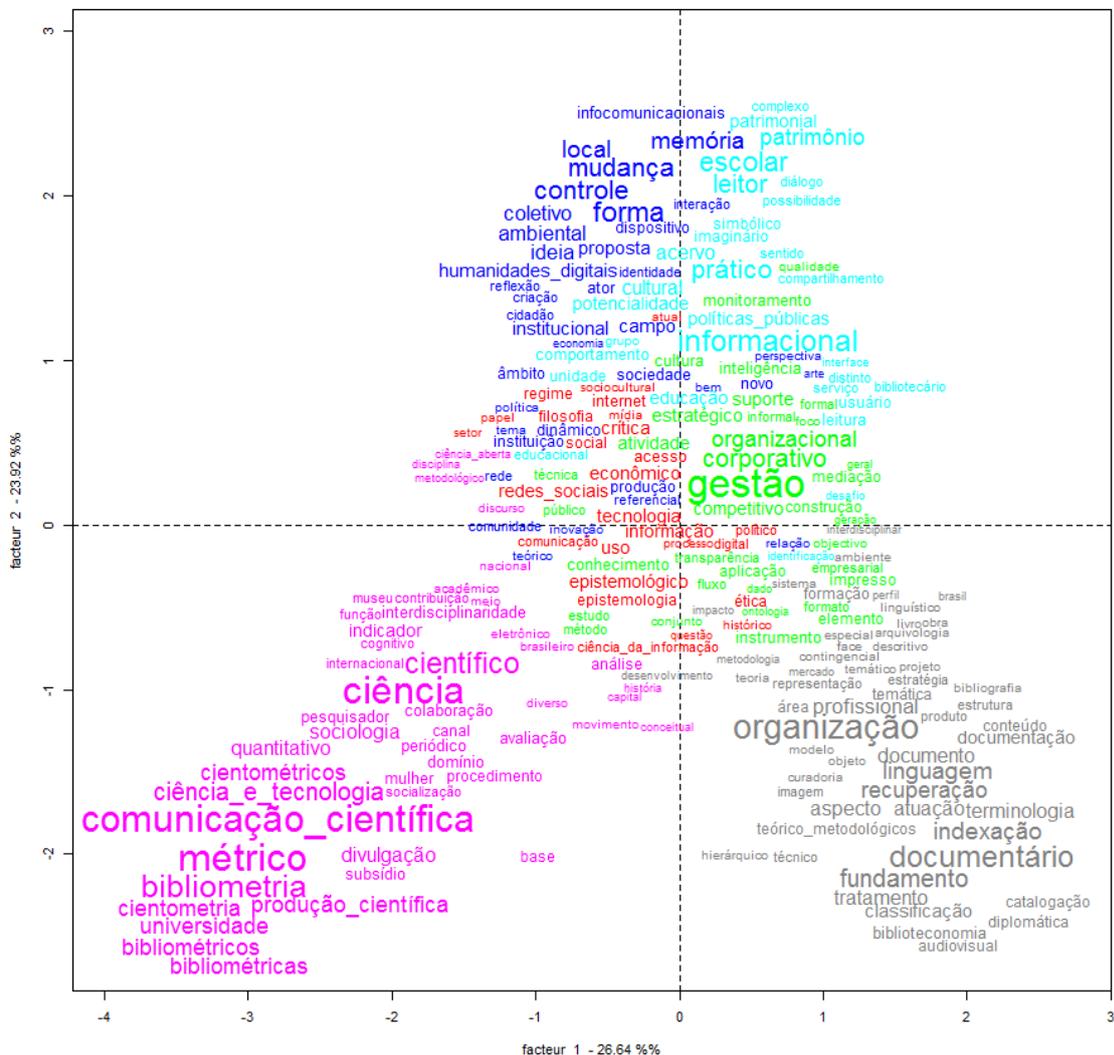
As outras classes são resultado de iterações e divisões subsequentes. A classe 6 (roxo) representa o vocabulário característico de estudos métricos e de comunicação científica (ciência, comunicação científica, estudos bibliométricos, cientométricos, produção científica, citação). A classe 5 (azul) agrupa vocabulários relacionados com humanidades

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

digitais, mudanças sociotécnicas, políticas e normas de controle de informação). A classe 4 (azul ciano) reúne vocabulários relacionados com práticas informacionais, leitores, bibliotecas escolares, patrimônio, memória e políticas públicas. Já as classes 1 e 3 apresentam as maiores semelhanças, o que se visualizará com maior clareza na Figura 2. Contudo, é possível reconhecer vocabulário típico das temáticas relacionadas com gestão da informação e do conhecimento e inteligência competitiva na classe 3 (verde), enquanto a classe 1 (vermelho) reúne léxico associado às tecnologias da informação, as redes sociais, internet e regimes de informação.

O mapa da Figura 2 permite apreciar a existência de duas temáticas bem diferenciadas: à esquerda, os estudos métricos e de comunicação científica, que apresentam como referência central o fluxo de informação científica, incluindo aspectos da comunicação e colaboração entre cientistas e da divulgação do conhecimento científico; à direita, organização da informação. Na parte central, outras temáticas, cujas fronteiras estão menos demarcadas (gestão da informação, tecnologias, redes sociotécnicas, patrimônio memória, usuários e práticas informacionais).

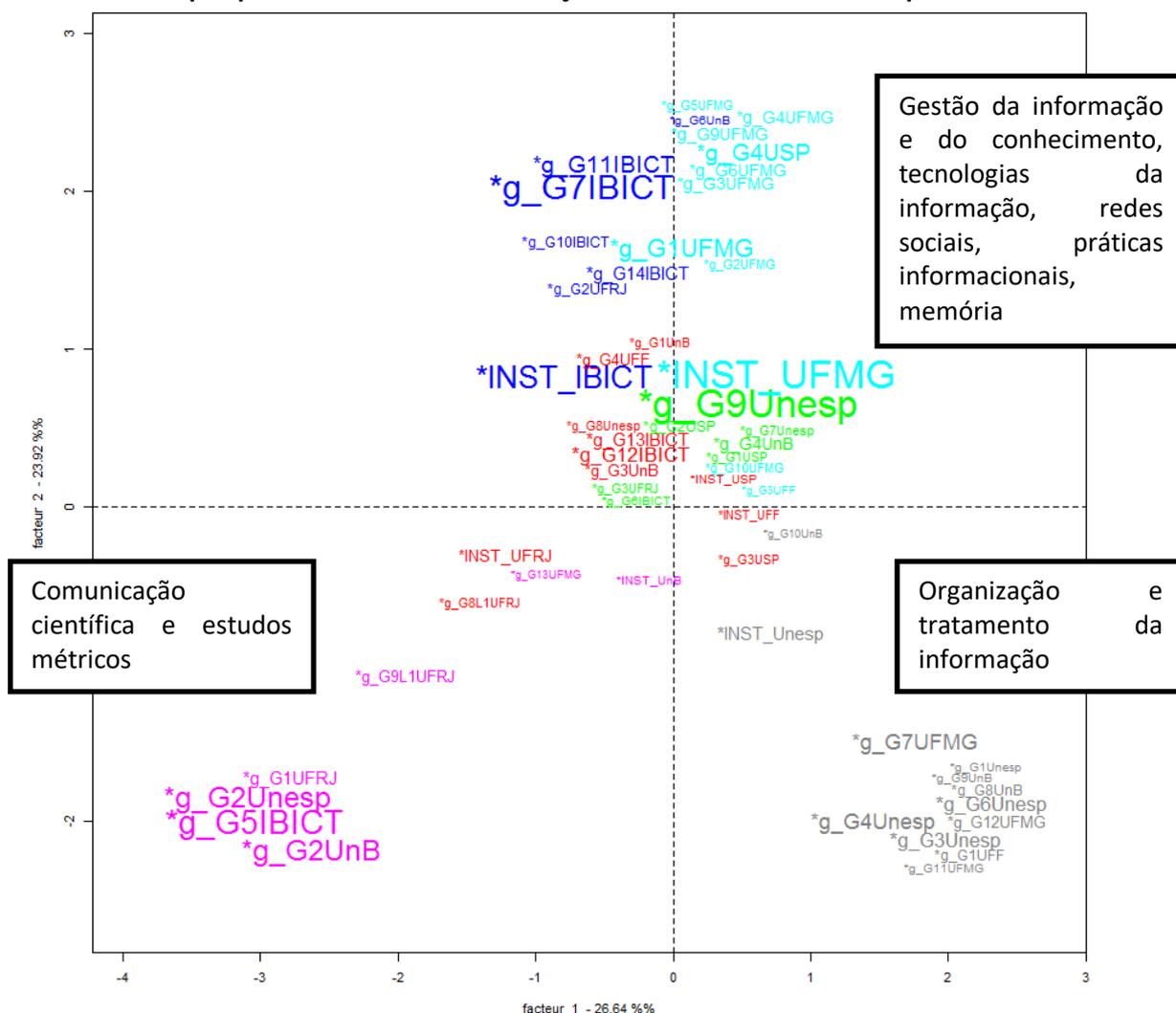
Figura 2: Mapa das seis classes a partir do vocabulário das linhas dos Grupos de pesquisa em Ciência da informação. Análise fatorial de correspondência



Fonte: Dados da pesquisa

A interseção entre as figuras 2 e 3 permite identificar quais as linhas de pesquisa mais associadas com as temáticas identificadas. É possível também identificar as temáticas predominantes nas instituições, pelo tamanho e posição no plano dos códigos das variáveis. Por exemplo, na Figura 3, a variável *Inst.\_Unesp*, que representa os grupos da UNESP como um todo, está na cor cinza, sugerindo forte inclinação pela temática Organização e tratamento da informação. De fato, das 36 linhas de pesquisa em Ciência da informação desta instituição, 12 abordam aspectos relacionados com a organização da informação e do conhecimento.

Figura 3: Mapa dos Grupos e linhas de pesquisa a partir do vocabulário das linhas dos Grupos de pesquisa em Ciência da informação. Análise fatorial de correspondência



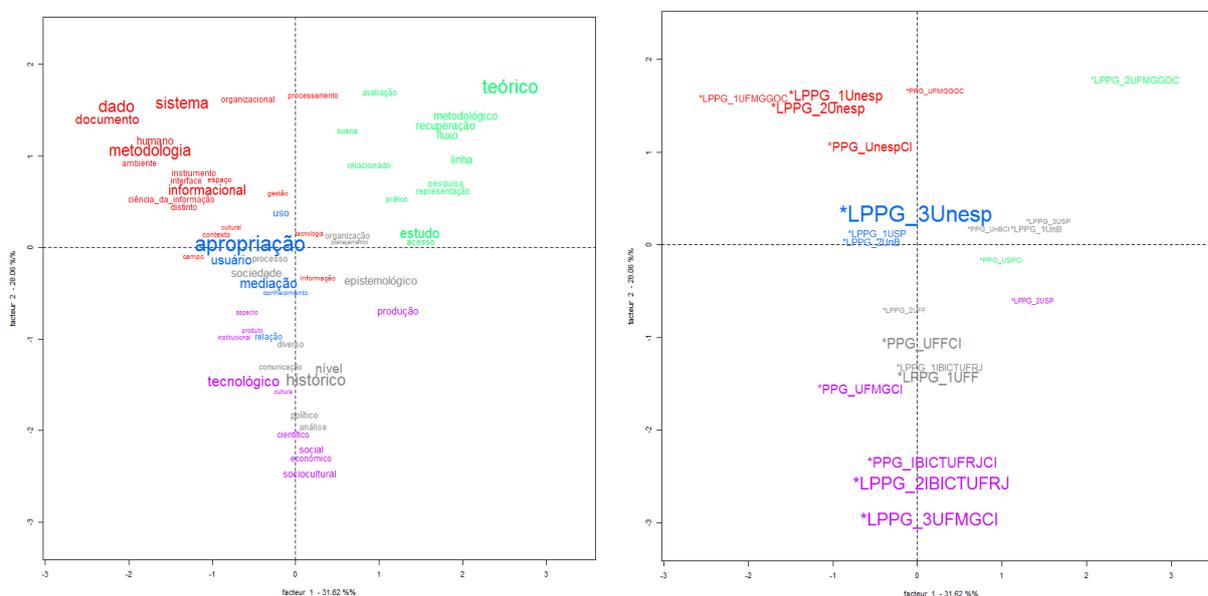
Fonte: Dados da pesquisa

De forma similar, o código que representa a UFMG está em azul ciano, no centro do plano, pois predominam nesta instituição os grupos com linhas relacionadas aos estudos de usuários, práticas informacionais, gestão do conhecimento, memória e patrimônio. Isso não significa que não sejam visualizados grupos específicos vinculados a outras temáticas, por exemplo, o código *g\_G7UFMG* em cinza representa o grupo Fundamentos teóricos, metodológicos e históricos da Organização da Informação, coordenado pela professora Cristina Dotta Ortega da UFMG. Na esquerda inferior do quadrante, encontram-se grupos dedicados à comunicação científica, com destaque para os grupos do IBICT, da UNESP e da UnB, com maior quantidade de linhas de pesquisa.

### 4.3 Análise das linhas de pesquisa dos Programas de Pós-graduação e suas ementas

A análise das linhas de pesquisa dos PPG foi realizada a partir de 17 textos (títulos das linhas e suas ementas). Foi aplicada também a técnica de Classificação Hierárquica Descendente (CHD), que classificou o corpus em cinco classes, embora tenha sido constatado como limitação o fato de o corpus textual ser muito mais reduzido que aquele construído com as linhas dos GP, o que levou possivelmente a conformar clusters mais heterogêneos. A Figura 4 é uma visualização derivada do CHD, por meio da análise fatorial de correspondência.

Figura 4: Mapas das classes a partir do vocabulário das linhas dos Programas de Pós-graduação em Ciência da informação. Análise fatorial de correspondência.



Fonte: dados da pesquisa

O mapa da esquerda representa as cinco classes nas quais foi classificado o corpus textual, diferenciando-as segundo cores; também a frequência está indicada pelo tamanho das palavras. No mapa bidimensional da direita, estão representados os PPG, ocupando os mesmos espaços que no mapa da esquerda. Por isso, ambas as representações se complementam. No centro do mapa destacam-se os termos *apropriação* e *mediação*[da informação] e *usuário*, utilizados nas ementas de algumas das linhas dos PPGCI da UNESP, da USP e da UnB. No entanto, o destaque em tamanho do primeiro termo deve-se ao uso redundante do termo na ementa da linha Gestão, Mediação e Uso da Informação, da UNESP.

#### **4.4 Análise da relação entre as linhas dos Grupos de pesquisa e as dos Programas de pós-graduação**

Tanto as visualizações obtidas como a análise de conteúdo das linhas permitiram uma aproximação à compreensão dos elos entre as linhas dos GP e as dos PPG. Essa relação é bastante clara em alguns casos. Por exemplo, na UNESP, os 9 nove grupos de pesquisa distribuem-se em 36 linhas de pesquisa que são absorvidas pelas quatro linhas de pesquisa do Programa: Produção e Organização da Informação (desdobrada em dois eixos, o de Produção e o de Organização); Informação e Tecnologia; Gestão, Mediação e Uso da Informação. Também no caso do PPGCI da UnB, suas duas linhas Pesquisa em Comunicação e Mediação da Informação e Pesquisa em Organização da Informação abrangem todas as linhas de pesquisa de seus GP. Estas relações são enfatizadas por alguns Programas desde seus próprios sites institucionais. Já no caso da UFRJ, percebe-se a predominância da Linha do Programa, Comunicação, Organização e Gestão da Informação e do Conhecimento, abrangendo cinco grupos de pesquisa criados a partir de 2012 e suas respectivas linhas. Esta linha do Programa, entretanto, relaciona-se desde 1991 com o Grupo de Pesquisa Organização do Conhecimento, ainda sob a égide do convênio IBICT/UFRJ, podendo ser considerado um grupo histórico, assim como suas relações com linhas dos dois Programas. Da mesma forma, a origem da Linha Configurações Socioculturais, Políticas e Econômicas da Informação, comum aos dois períodos, permanece em íntima relação com os grupos de pesquisa e suas respectivas linhas, com nítido predomínio nos cinco grupos do programa conveniado IBICT/UFRJ.

Quanto aos programas da UFMG, observa-se a permanência e abrangência da Linha de Pesquisa Usuários, gestão do conhecimento e práticas informacionais, à qual estão vinculados cinco grupos de pesquisa e suas respectivas linhas, entre eles o Grupo de Estudos em Biblioteca Escolar, o mais antigo, de 1998, coordenado pela professora Bernadete Santos Campello. À linha do Programa Memória Social, Patrimônio e Produção do conhecimento estão vinculados quatro grupos de pesquisa, entre os quais o de Fundamentos teóricos, metodológicos e históricos da Organização da Informação, cujos desdobramentos temáticos se conformam a três linhas específicas de pesquisa. Composto o Programa de Pós-graduação em Gestão e Organização do Conhecimento, duas linhas de Pesquisa do Programa, Arquitetura & Organização do Conhecimento e Gestão & Tecnologia harmonizam-se com dois grupos e cinco linhas de pesquisa. A USP, contando com três linhas do

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

Programa, Apropriação Social da Informação, à qual se vincula apenas um grupo, apresentando afinidade temática em suas duas linhas de pesquisa; as linhas de Gestão de Dispositivos de Informação e de Organização da Informação e do Conhecimento, compondo uma constelação temática expressiva pela diversidade dos grupos a elas vinculados. A UFF, por sua vez, apresenta duas linhas de pesquisa do Programa, Informação, Cultura e Sociedade e Fluxos e Mediações Sócio Técnicas da Informação, cujas relações com os quatro grupos de pesquisa se verificam em linhas coerentes com o conjunto temático apresentado.

## **5 CONSIDERAÇÕES FINAIS**

A discussão teórica sobre a institucionalização social e cognitiva da pesquisa nos Programas de Pós-graduação em Ciência da Informação no Brasil permitiu a compreensão dos dados empíricos sobre a constituição de linhas e grupos consolidados, assim como temas de pesquisa como atividades reconhecidas enquanto formas de prestígio acadêmico, de acordo com as características que compõem o que se poderia denominar como *habitus* (Bourdieu) na constituição e delimitação do campo, fortalecendo laços e, desse modo, a ciência que se produz.

A visualização possível, graças aos mapas dos grupos e linhas de pesquisa, a partir das análises lexicais realizadas com o auxílio do software IRaMuTeQ, apresenta duas temáticas diferenciadas dos grupos mais representativos, um deles abrangendo os estudos métricos e de comunicação científica e o outro, os estudos sobre tratamento e organização da informação. Há um conjunto temático com delineamento menos nítido, que envolve temas como gestão da informação, tecnologias, redes sociotécnicas, patrimônio, memória, usuários e práticas informacionais. A articulação entre linhas de pesquisa dos Programas de Pós-graduação e as linhas de pesquisa dos grupos de pesquisa em Ciência da Informação é perceptível em algumas instituições, esclarecidas as implicações destes elos para a institucionalização da Ciência da Informação brasileira. Assim, as instituições abrigam grupos de pesquisa cujos membros distribuem-se em linhas que são absorvidas pelas linhas de pesquisa institucionalizadas nos Programas.

Contudo, as discussões efetivadas nos programas de pós-graduação refletem preocupações como o esclarecimento sobre a conceituação e a abrangência da área de concentração e das linhas de pesquisa, considerando-se a necessidade de se promover

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

maior organicidade aos programas. Aponta-se a incipiente evidência da relação entre produção científica, currículo, disciplinas e corpo docente na área de Ciência da Informação. Essa questão tem a ver com a necessidade de repensar os projetos de pesquisa para esclarecimento de seus elementos constitutivos, como contextualização teórica e empírica, problemática, definição do problema e dos objetivos e metodologias a ele relacionados.

Portanto, pesquisas futuras poderiam procurar os elos entre as linhas de pesquisa dos grupos, os Programas de Pós-graduação e a produção científica produzida pelos membros destes Programas, para verificar a articulação e organicidade entre linhas, produtos da pesquisa e suas manifestações nos currículos.

## **REFERÊNCIAS**

BEN-DAVID. J. **Centers of Learning**: Britain, Germany, the United States. New York, McGraw Hill, 1977.

BRASIL. Diretório dos Grupos de Pesquisa no Brasil. Lattes, 2014. Disponível em: <http://lattes.cnpq.br/web/dgp/glossario> Acesso em: 27 de julho de 2019

BRASIL. Ministério da Educação. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Requisitos para a Apresentação de Propostas de Cursos Novos (APCN). Diretoria de Avaliação, 2016. Disponível em: [https://www.capes.gov.br/images/documentos/Criterios\\_apcn\\_2semestre/Crit%C3%A9rios\\_de\\_APCN\\_2017\\_-\\_Artes\\_Musica.pdf](https://www.capes.gov.br/images/documentos/Criterios_apcn_2semestre/Crit%C3%A9rios_de_APCN_2017_-_Artes_Musica.pdf). Acesso em: 27 de julho de 2019.

CAMARGO, B. V., JUSTO, A. M. IRAMUTEQ: Um Software Gratuito para Análise de Dados Textuais. **Temas em Psicologia**. v. 21, n. 2, p. 513-518, 2013. DOI: 10.9788/TP2013.2-16

CAMARGO, B.V; JUSTO, A.M. **Tutorial para uso do software Iramuteq**. 2016. Disponível em: [http://www.iramuteq.org/documentation/fichiers/Tutorial%20IRaMuTeQ%20em%20portugues\\_17.03.2016.pdf](http://www.iramuteq.org/documentation/fichiers/Tutorial%20IRaMuTeQ%20em%20portugues_17.03.2016.pdf). Acesso em: 27 de julho de 2019.

CUSTÓDIO, P. Á. G. R. **As bases teóricas expressas nas teses dos programas de pós-graduação em ciência da informação no brasil: uma análise cientométrica**. 2018. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Marília, 2018.

MAIA, V.I. Produção científica versus linhas de pesquisa dos Programas de Pós-Graduação em Sociologia e Ciências Sociais no Brasil. **Latitude**, v. 12, n. 2, p. 637-689, 2017.

NASCIMENTO, Adriano Roberto Afonso do; MENANDRO, Paulo Rogério Meira. Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. **Estudos e pesquisas em psicologia**, Rio de Janeiro, v. 6, n. 2, 2006

**XX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2019**  
**21 a 25 de outubro de 2019 – Florianópolis – SC**

REINERT, A. Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. **Les cahiers de l'analyse des données**, v. 8, n. 2, p. 187–198, 1983.

REINERT, M. **Alceste**: Analyse de données textuelles. Manuel d'utilisateur. Toulouse: Image, 1998.

RUIZ BUENO, A. **Trabalhar com IRaMuTeQ**: pautas. 2017. Disponível em: [http://diposit.ub.edu/dspace/bitstream/2445/113063/1/Trabajar\\_con\\_IRAMUTEQ\\_PAUTAS.pdf](http://diposit.ub.edu/dspace/bitstream/2445/113063/1/Trabajar_con_IRAMUTEQ_PAUTAS.pdf). Acesso em: 27 de julho de 2019.

SHINN, T.; RAGOUET, P. **Controvérsias sobre a ciência**: por uma sociologia transversalista da atividade científica. São Paulo: Associação Filosófica Scientiae Studia; Editora 34, 2008.

SCHWARTZMAN, S. Relatório de uma visita ao Brasil. **Ciência Hoje**, São Paulo, v. 7, n. 37, nov., 1987. Disponível em: <http://www.schwartzman.org.br/simon/bendavid.htm>. Acesso em: 20 out. 2019.

WHITLEY, R. **The intellectual and social organization of the sciences**. Oxford: Oxford University Press, 1984.