

# ANÁLISES DE REDES DE COOCORÊNCIA TEMÁTICA PARA DEFINIÇÃO DE ÁREAS DE CONCENTRAÇÃO E LINHAS DE PESQUISA EM PROGRAMAS DE PÓS-GRADUAÇÃO: o caso do PPGCINF/UnB

João de Melo Maricato  
Universidade de Brasília, Brasília, DF, Brasil  
jmmaricato@gmail.com

Veruska da Silva Costa  
Universidade de Brasília, Brasília, DF, Brasil  
veruskadasilvacosta@gmail.com

Fernando César Lima Leite  
Universidade de Brasília, Brasília, DF, Brasil  
fernandodfc@gmail.com

## 1 INTRODUÇÃO

A presente pesquisa lança mão de técnicas bibliométricas de coocorrência temática e da Análise de Redes Sociais (ARS) com o objetivo de entender a proximidade da produção científica de um grupo de professores. Apoiar-se na Organização da Informação e na Teoria dos Grafos e propõe uma metodologia para identificação de categorias temáticas e a análise de suas coocorrências, a partir de artigos publicados por professores/pesquisadores. Tem com isso o intuito de contribuir para a definição de áreas de concentração e linhas de pesquisa em programas de pós-graduação.

Maricato e Noronha (2013) utilizaram coclassificação de artigos e patentes para identificar relações temáticas entre ciência e tecnologia. Lucas (2014) estudou coocorrência de palavras-chave e grupos temáticos em organização do conhecimento. Morillo, Bordons e Gómez (2001) analisaram cocategorias para estudar relações interdisciplinares



em química. Ding, Chowdhury e Foo (2001) utilizaram coocorrência de palavras-chave para entender a estrutura intelectual da área de recuperação da informação. Os estudos dessa natureza são considerados complexos e a construção de mapas lexicais da ciência ainda encontra muitos obstáculos (MARSHAKOVA-SHAIKEVICH, 2005).

Emirbayer e Goodwin (1994) defendem que a ARS não é uma teoria formal ou unitária, mas uma ampla estratégia de investigação de estruturas sociais. Newman (2003) define uma rede como sendo “um conjunto de vértices ou nós conectados por arestas”. Essas permitem estudar estruturas e sistemas cuja descrição, sem conceitos relacionais, se tornaria praticamente impossível. Podem-se utilizar técnicas de ARS a pessoas, organizações, países ou temáticas.

A proposta metodológica foi testada em um grupo de professores/pesquisadores vinculados à Faculdade de Ciência da Informação (FCI) da Universidade de Brasília (UnB), mas pode ser adaptada às necessidades de outros programas de pós-graduação.

## 2 PROCEDIMENTOS METODOLÓGICOS

Para testar a metodologia, utilizaram-se dados da produção científica dos professores/pesquisadores vinculados à FCI da UnB. Foram considerados somente vinculados aos cursos de graduação e pós-graduação, pesquisadores colaboradores, voluntários e aposentados com algum vínculo efetivo com a faculdade, que possuísem ao menos um artigo científico em periódicos brasileiros da área de Ciência da Informação ou publicação de trabalhos nos anais do Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (ENANCIB), entre 2010 e 2016. Não foi realizada delimitação de pesquisadores em virtude da sua titulação. Os dados que subsidiaram as análises foram extraídos da base de dados ABCDM. A busca foi realizada a partir da afiliação institucional dos professores/pesquisadores. Os dados foram complementados a partir da consulta aos currículos na Plataforma Lattes/CNPq <lattes.cnpq.br/>. Foram identificados 39 professores/pesquisadores, 232 artigos e 1.332 palavras-chave.



As palavras-chave dos artigos foram distribuídas em cada uma das 60 categorias, em razão de sua proximidade de significado. Cada artigo passou a ser descrito por no mínimo uma e no máximo cinco categorias temáticas. Palavras-chave com o mesmo significado e redundantes foram consideradas uma única vez. O processo de indexação dos 232 artigos nas 60 categorias temáticas resultou em 935 coocorrências temáticas. Para tratamento e análise dos dados foram utilizados os softwares *OpenRefine*, Microsoft Excel e, para métricas e visualização de redes, os softwares NETDRAW e UCINET.

### 3 RESULTADOS E DISCUSSÕES

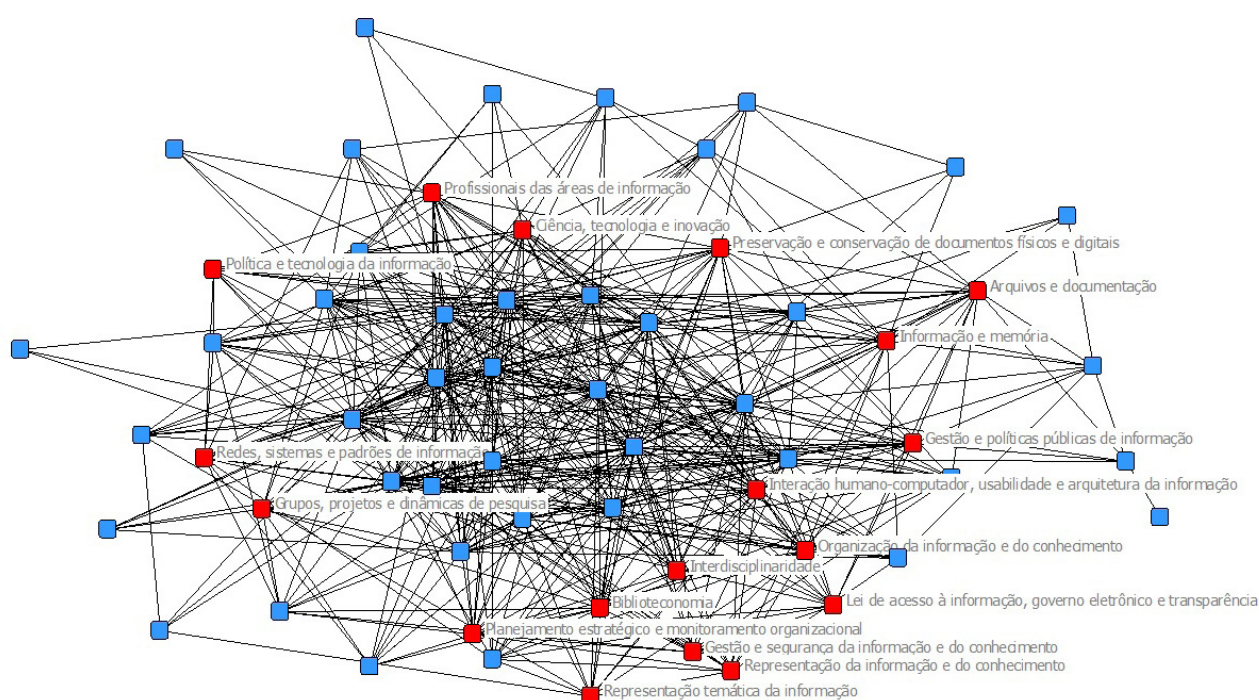
Após a realização de uma série de testes, constatou-se não ser adequada a definição de linhas de pesquisa de programas de pós-graduação focando nos nós temáticos centrais e tampouco nos nós temáticos periféricos. Os temas com maior Grau de Centralidade, que se aglutinam no centro de uma rede, são transversais e presentes nas pesquisas de uma grande parte dos professores/pesquisadores selecionados. Os periféricos são abordados por um conjunto muito pequeno de indivíduos. Desse modo, analisaram-se as categorias temáticas intermediárias da rede que conectam as categorias nucleares às categorias periféricas. Partiu-se da premissa de que os subgrupos temáticos não tão densos quanto os do núcleo e não tão dispersos quanto os periféricos poderiam trazer um mapa temático mais pertinente. Assim, a estratégia metodológica considerou os critérios: 1) identificar categorias temáticas com graus de centralidade entre 16 e 50 (de 0 a 100); 2) o número de ocorrências deveria ser no máximo 21 categorias temáticas (das 60); e 3) as categorias com os maiores Graus de Intermediação deveriam ser maiores ou iguais a 23 (onde 0 é ausência de Grau de Intermediação). A partir desses critérios, foram identificadas e estudadas detalhadamente 18 categorias temáticas.

Ao se analisar graficamente os dados das 18 categorias selecionadas, chegou-se à Figura 1. Os nós em vermelho representam as 18 categorias temáticas selecionadas (intermediárias) e os em azul as demais temáticas (centrais e periféricas). Também foram calculados os coeficientes de



identidade das categorias temáticas selecionadas (Figura 1). A Tabela 1 apresenta, para fins de exemplificação, três das 18 categorias temáticas selecionadas; os respectivos graus de centralidade e de intermediação; o número de ocorrências das categorias temáticas; as categorias temáticas com maior coeficiente de identidade, com os respectivos coeficientes (cinco categorias com maiores coeficientes).

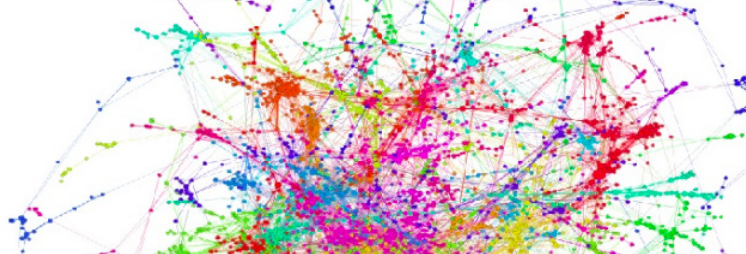
**FIGURA 1 - CATEGORIAS TEMÁTICAS COM MAIOR GRAU DE INTERMEDIÇÃO**



**Fonte:** Dados da pesquisa.

Realizou-se ainda a análise individualizada de cada uma das 18 categorias temáticas selecionadas em redes egocêntricas, ou seja, focando em cada um dos 18 nós e suas ligações diretas. A partir dos resultados, realizou-se o agrupamento dos professores/pesquisadores em três conjuntos de temas, buscando-se fazer o enquadramento destes em três linhas de pesquisa. Considerando o exposto, apresenta-se a Figura 2, que traz a rede temática do PPGCINF/UnB em três linhas de pesquisa possíveis.

A **Linha I**, representada na cor verde, possui relação forte com as seguintes categorias temáticas: Teoria, história e epistemologia em informação e comunicação; Universidades, instituições de ensino, cursos



e currículos; Preservação e conservação de documentos físicos e digitais; Museu, documento e memória; e Museologia.

**TABELA 1 - CATEGORIAS TEMÁTICAS SELECIONADAS (TRÊS DAS 18) E SUAS RESPECTIVAS CINCO CATEGORIAS COM MAIOR COEFICIENTE DE IDENTIDADE**

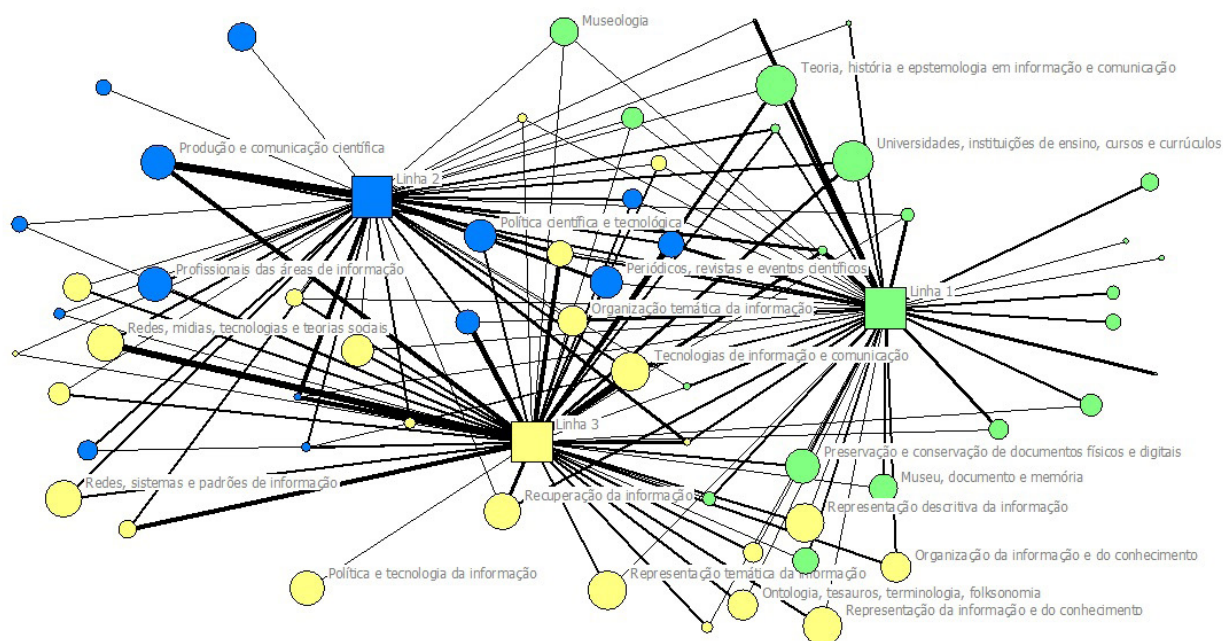
| Categorias temáticas selecionadas             | Grau Centralidade | Ocorrência de categoria | Grau de Intermediação | Categorias temáticas com maior Coeficiente de Identidade                                | Coeficiente de Identidade |
|-----------------------------------------------|-------------------|-------------------------|-----------------------|-----------------------------------------------------------------------------------------|---------------------------|
| Informação e memória                          | 34                | 19                      | 88                    | Teoria, história e epistemologia em informação e comunicação                            | 64%                       |
|                                               |                   |                         |                       | Documentos, diplomática, obras e documentação                                           | 60%                       |
|                                               |                   |                         |                       | Organização da informação e do conhecimento                                             | 60%                       |
|                                               |                   |                         |                       | Mediação, circulação, divulgação e disseminação da informação                           | 58%                       |
|                                               |                   |                         |                       | Ciências e conhecimento científico                                                      | 55%                       |
|                                               |                   |                         |                       | Grupos, projetos e dinâmicas de pesquisa                                                | 49%                       |
|                                               |                   |                         |                       | Ensino-aprendizagem                                                                     | 39%                       |
|                                               |                   |                         |                       | Interação humano-computador, usabilidade e arquitetura da informação                    | 37%                       |
|                                               |                   |                         |                       | Ontologia, tesouros, terminologia, folksonomia                                          | 24%                       |
| Política científica e tecnológica             | 33                | 18                      | 42                    | Documentos, diplomática, obras e documentação                                           | 59%                       |
|                                               |                   |                         |                       | Ciência, tecnologia e inovação                                                          | 58%                       |
|                                               |                   |                         |                       | Universidades, instituições de ensino, cursos e currículos de graduação e pós-graduação | 58%                       |
|                                               |                   |                         |                       | Ciências e conhecimento científico                                                      | 53%                       |
|                                               |                   |                         |                       | Grupos, projetos e dinâmicas de pesquisa                                                | 49%                       |
| Representação da informação e do conhecimento | 24                | 14                      | 24                    | Organização da informação e do conhecimento                                             | 60%                       |
|                                               |                   |                         |                       | Ontologia, tesouros, terminologia, folksonomia                                          | 47%                       |
|                                               |                   |                         |                       | Representação temática da informação                                                    | 41%                       |
|                                               |                   |                         |                       | Mediação, circulação, divulgação e disseminação da informação                           | 36%                       |
|                                               |                   |                         |                       | Organização temática da informação                                                      | 36%                       |

**Fonte:** Dados da pesquisa.

A **Linha 2**, em azul, possui relação forte com as seguintes categorias/temas: Produção e comunicação científica; Profissionais da área de informação; Política científica e tecnológica; Periódicos, revistas e eventos científicos. Por fim, a **Linha 3**, representada em amarelo, possui relação forte com as seguintes categorias/temas: Redes, mídias, teorias e tecnologias sociais; Redes, sistemas e padrões de informação; Recuperação da informação; Representação temática da informação. Como todos esses temas perpassam aspectos relacionados à comunicação, acredita-se que a **Área de concentração** do programa em análise poderia ser denominada: “Comunicação e Gestão da Informação”.



**FIGURA 2 - REDE TEMÁTICA E PROPOSTA DE ORGANIZAÇÃO EM 3 LINHAS DE PESQUISA NO PPGCINF/UNB**



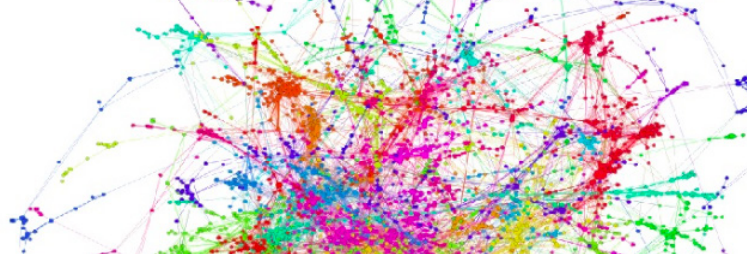
**Fonte:** Dados da pesquisa.

A partir da análise dos três aglomerados de categorias temáticas e professores, conclui-se que um programa de pós-graduação que tenha as características de produção científica apresentada pelo grupo de professores da FCI poderia ser dividido em três linhas de pesquisa, a saber: 1) “Gestão, organização e tecnologias da informação”; 2) “Produção, comunicação e mediação científica”; e, 3) “Política, história e cultura informacional”.

#### 4 CONSIDERAÇÕES FINAIS

Com a metodologia proposta e sua aplicação no caso do PPGCINF/UnB, os resultados fundamentam a proposição de uma área de concentração do programa denominada “Comunicação e Gestão da Informação” e três linhas de pesquisa: 1) “Gestão, organização e tecnologias da informação”; 2) “Produção, comunicação e mediação científica”; e, 3) “Política, história e cultura informacional”.

Com relação à proposta metodológica, alguns aspectos merecem ser discutidos. A seleção dos professores/pesquisadores deve ser cuidadosa-



mente realizada para se propor linhas de pesquisa coerentes. Quanto menor clareza sobre potenciais participantes do programa e suas produções, maiores os riscos de se tomar decisões equivocadas. Destaca-se que as decisões não podem ser pautadas apenas em critérios quantitativos. Outras opções de seleção de professores/pesquisadores, definição de categorias e de um número mínimo de artigos por professor poderiam ser consideradas. Portanto, outros estudos e abordagens devem ser realizados, com vistas a minimizar as limitações, tanto do ponto de vista da proposta metodológica, quanto da sua aplicação nos casos concretos.

## REFERÊNCIAS

- DING, Y.; CHOWDHURY, G. G.; FOO, S. Bibliometric cartography of information retrieval research by using co-word analysis. **Information processing & management**, v. 37, n. 6, p. 817-842, 2001.
- EMIRBAYER, M.; GOODWIN, J. Network analysis, culture, and the problem of agency. **American Journal of Sociology**, v. 99, n. 6, p. 1411-1454, May 1994.
- MARICATO, J. M.; NORONHA, D. P. Coclassificação em artigos e patentes em biodiesel. **Liinc em Revista**, v. 9, n. 1, p. 85-102, 2013.
- MARSHAKOVA-SHAIKEVICH, I. Bibliometric maps of field of science. **Information Processing & Management**, v. 41, n. 6, p. 1534-1547, 2005.
- MORILLO, F.; BORDONS, M.; GÓMEZ, I. An approach to interdisciplinarity through bibliometric indicators. **Scientometrics**, v. 51, n. 1, p. 203-222, 2001.
- NEWMAN, M. E. J. The structure and function of complex networks. **SIAM review**, v. 45, n. 2, p. 167-256, 2003.
- LUCAS, E. R. O. **Capital Social e Capital Científico na produção científica sobre Linguagens Documentárias e Sistemas de Organização do Conhecimento no campo da Knowledge Organization (KO) nos idiomas espanhol, francês e português**. São Paulo, 2014. 165 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2014.