



XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB)

**GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação**

**CARACTERIZAÇÃO DAS REDES DE CONVITES DE BANCAS DE DEFESA  
DE TESES E DISSERTAÇÕES DO BRASIL**

***CHARACTERIZATION OF BRAZILIAN MASTERS AND DOCTORAL  
COMMITTEES NETWORKS***

**Diogo Fornaziero Segura Ramos<sup>1</sup>, Jesus Pascual Mena-Chalco<sup>2</sup>**

**Modalidade da apresentação: Comunicação Oral**

**Resumo:** O objetivo deste trabalho é apresentar uma caracterização topológica das redes de convites efetivas de bancas de defesa de teses de doutorado e mestrado. Ao todo foram analisados 607.389 registros de defesa correspondentes a trabalhos de mestrado e doutorado. Para tanto, foram utilizados procedimentos computacionais para: (i) a coleta, extração e normalização de informação dos registros de defesa, (ii) criação das redes de convites através de grafos direcionados sem pesos, (iii) caracterização das redes usando métricas topológicas clássicas, e (iv) visualização de redes. Nos resultados apresentamos as características principais das redes obtidas para as 9 grandes áreas, e também para as 80 áreas do conhecimento, no período de 1987 a 2011. Adicionalmente evidenciamos assimetrias nas relações de convites em todas as grandes áreas do conhecimento. Nós acreditamos que este trabalho pode ser usado como um primeiro passo para a caracterização, em nível macro e meso, das interações pouco exploradas no contexto acadêmico.

**Palavras-chave:** Redes de convites, bancas de defesa, redes complexas, mapeamento da ciência.

**Abstract:** *The objective of this work is to present a topological characterization of networks of effective invite to juries of thesis and dissertation defenses. There were 607,389 records of defenses analysed, corresponding to works of master's and doctorate degrees. For such, there were used computational procedures for (i) collect, extract and normalize the defense records' information, (ii) creation of the invite networks through weightless directed graphs, (iii) characterization of the networks using classical topological metrics, and (iv) network visualization. In the results we present the main characteristics of the networks drawn for the 9 major Brazilian knowledge areas, and also*

---

1 UFABC

2 UFABC

*for the 80 knowledge areas, from 1987 to 2011. We additionally show asymmetries in the invite relations of all knowledge big areas. We believe this work can be used as a first step towards a characterization, at the macro and middle levels, of the little explored interactions in the academic context.*

**Keywords:** *Network of invites, jury defense, complex networks, science mapping.*

## 1 INTRODUÇÃO

Uma rede de colaboração científica pode ser vista como uma rede social, pois é uma coleção de atores (pessoas, grupos ou organizações) em que subconjuntos de conhecidos estão ligados por relações explícitas ou implícitas (NEWMAN, 2001). Uma relação explícita geralmente refere-se à colaboração efetiva entre atores (e.g., coautoria, orientação acadêmica). Já uma relação implícita refere-se comumente à colaboração não documentada entre atores (e.g., participação em projeto de pesquisa, avaliação de trabalhos).

As redes sociais que descrevem a interação entre cientistas podem ser criadas usando diferentes fontes de dados ou recursos. Uma fonte comumente utilizada para a criação de uma rede de interação científica explícita é o conjunto de coautoria de publicações acadêmicas. Nesse tipo de rede de coautoria, os autores são representados por vértices e a participação em coautoria é representada por uma ligação ou aresta. Uma outra fonte de dados pouco explorada para examinar a interação entre cientistas é a relacionada com a participação em bancas de defesa de trabalhos de pós-graduação. Essa interação representa, em outras palavras, a interação efetiva entre cientistas, reunidos para avaliar um trabalho acadêmico, onde um dos autores importantes, além do aluno avaliado, é o orientador (ou presidente da banca de avaliação). Nesse contexto, algumas perguntas interessantes de pesquisa que subjazem são: Como é a estrutura das redes de convites de bancas de defesa no Brasil? Quais as principais características nesse tipo de estruturas? Existem padrões nas formas de convite realizadas por pesquisadores de diferentes áreas de conhecimento?

Pensando nas perguntas anteriormente apresentadas é que, neste trabalho, realizamos uma caracterização topológica da rede de convites de bancas de defesa de trabalhos de pós-graduação, usando como fonte para a criação desta rede de interação (colaboração) científica, todas as bancas de defesa de dissertação de mestrado e teses de doutorado no Brasil, disponíveis no banco de Teses e Dissertações da CAPES e registradas no período 1987-2011.

As informações obtidas através da caracterização topológica de uma rede, da perspectiva de convites efetivos em bancas, representam insumos importantes para avaliação e tomada de decisão acadêmica. As informações podem ser usadas para, por exemplo, comparar a interação de pesquisadores de diferentes áreas do conhecimento, identificar assimetrias na participação em bancas, identificar subgrupos de interação multidisciplinar (i.e., participação de pesquisadores de diferentes áreas em uma banca), e determinar a evolução temporal da interação dos pesquisadores participantes em bancas.

## **2 REFERENCIAL TEÓRICO**

A análise de redes de colaboração científica não é novidade. Newman (2001) caracteriza diversas redes de coautoría usando bases de dados de artigos científicos considerando dois autores ligados quando ambos estão presentes como coautores em um mesmo artigo. Newman (2001) demonstra que as redes de colaboração científicas analisadas seguem o padrão de “mundo pequeno”, no qual há uma distância pequena entre dois autores escolhidos aleatoriamente.

Por outro lado, Cunha e Neto (2016) analisaram 1002 bancas de defesa de mestrado *stricto sensu* em Ciências Contábeis no Brasil realizadas no período de 2002 a 2010. Estas 1002 bancas foram escolhidas pois possuíam na sua composição pelo menos um dos 241 docentes identificados no corpo docente dos 18 programas de pós-graduação em Ciências Contábeis, recomendados pela CAPES até o ano de 2010. Este trabalho é um dos mais importantes na literatura atual que aborda os conceitos de redes em bancas de defesa trabalhos de pós-graduação.

Para identificar em quais bancas de defesa de mestrado os 241 docentes participaram na área de Ciências Contábeis, Cunha e Neto (2016) extraíram essa informação do currículo de cada um dos 241 docentes contido na Plataforma *Lattes*. Os autores determinaram que uma ligação entre dois docentes acontece quando estes dois docentes participam de uma mesma banca de defesa de mestrado *stricto sensu* em Ciências Contábeis.

Ao dividirem o período de 2002 a 2010 em 3 triênios (2002-2004, 2005-2007 e 2008-2010), Cunha e Neto (2016) observaram que há um aumento no número de docentes e defesas de mestrado em função do triênio. Os autores atribuem este aumento ao aumento no número de programas de pós-graduação na primeira década dos anos 2000.

Apesar do aumento no número de docentes e ligações entre docentes na rede de bancas de defesa de mestrado, eles observaram que houve uma diminuição na densidade da rede e um aumento no número de componentes conexas, indicando que o fluxo de informação passa a ser mais lento e que surgem ilhas em que docentes não interagem com docentes de outras ilhas.

Ao calcular a centralidade de grau e centralidade de intermediação, Cunha e Neto (2016) encontraram docentes associados aos programas de pós-graduação que podem ser considerados com poder. Assim, naturalmente, a centralidade de grau pode ser um indicador de prestígio (MOODY, 2004) e a centralidade de intermediação pode indicar que vértices centralizados podem controlar o fluxo de informação entre vértices (ROSSONI, 2006). Foi evidenciado que, o grau de correlação calculado entre a centralidade de grau e a centralidade de intermediação é baixo, indicando que os vértices influentes privilegiam os seus grupos ao invés de colaborar com outros vértices influentes (NEWMAN, 2004).

Para caracterizar a colaboração entre os programas de pós-graduação das Ciências Contábeis, Cunha e Neto (2016) criaram uma rede de colaboração entre as instituições. Esta rede de colaboração entre instituições demonstrou que elas tendem a colaborar entre si, independente de sua posição ou distância geográfica.

Finalmente, usando os parâmetros desenvolvidos por Strogatz e Watts (1998) e o coeficiente de *Small World* de Spiro e Uzzi (2005), Cunha e Neto (2016) concluíram que a rede de colaboração em bancas de defesa de mestrado *stricto sensu* em Ciências Contábeis não apresenta um comportamento de “mundo pequeno”, diferentemente de Newman (2001) que encontra um comportamento de “mundo pequeno” em todas as comunidades científicas analisadas.

### **3 PROCEDIMENTOS METODOLÓGICOS**

Este trabalho trata-se de uma pesquisa exploratória e descritiva. O *corpus* da pesquisa envolve registros de teses e dissertações associadas a todos os programas acadêmicos e profissionalizantes de pós-graduação do Brasil e avaliados pela CAPES compreendendo 25 anos (1987-2011).

Em 2012 foram obtidos 607.389 registros do banco de Teses e Dissertações da CAPES<sup>3</sup> no período de 1987 (primeiro ano disponível para consulta no sistema web da CAPES) até 2011 (último ano disponível para consulta). Nesse período foram identificadas 136.619 teses de doutorado e 447.267 dissertações de mestrado (MENA-CHALCO; ROCHA, 2012).

Cada registro do banco de Teses e Dissertações da CAPES representa ou uma defesa de dissertação de mestrado ou uma defesa de tese de doutorado. Cada registro contém dados como o nome do aluno, o(s) nome(s) do(s) orientador(es), os nomes dos integrantes da banca, a(s) área(s) do conhecimento na qual o trabalho foi defendido, nome do programa de pós-graduação, palavras-chave, resumo, dentre outras informações.

A estruturação dos registros do banco de Teses e Dissertações da CAPES foi realizada através de uma estrutura de dados tabular que contém: (i) o nome do orientador principal, (ii) os nomes dos membros convidados para a banca de avaliação, (iii) a(s) área(s) do conhecimento do trabalho, e (iv) o ano de defesa no qual o trabalho foi defendido. Nos trabalhos que possuem mais de um orientador listado, assume-se que o primeiro nome da lista de orientadores é o orientador principal e o único que realiza o convite aos demais membros integrantes da banca de defesa.

É importante destacar que este conjunto de dados textuais, embora contenha dados estruturados (i.e., tabulados e organizados), são diferentes os padrões de escrita e acentuação que foram adotados para o preenchimento dos registros. Assim, é necessária a utilização de técnicas de de-duplicação de nomes a fim de identificar diferentes ou múltiplos registros que se referem a um mesmo objeto (ELMAGARMID et al., 2007)

Nesse contexto, durante a estruturação dos registros, os nomes dos membros integrantes da banca de defesa foram normalizados utilizando a seguinte sequência de ações:

- As letras foram transformadas em suas versões minúsculas para que nomes que tivessem somente diferença entre letras maiúsculas e minúsculas pudessem ser considerados como o mesmo nome;
- Textos entre os caracteres “<” e “>” e entre os caracteres “(“ e “)” foram removidos pois frequentemente representavam instituições ou outras informações que não pertencem ao nome do membro de banca convidado;
- Caracteres especiais (como “ŷ” e “ı”) foram removidos;
- Letras acentuadas foram substituídas por equivalentes sem acentos, como “ç” foi substituída por “c” e “é” foi substituída por “e”;

---

<sup>3</sup> O sistema web de consulta ao banco de teses e dissertações da CAPES está disponível em <http://www.capes.gov.br/servicos/banco-de-teses>. Acesso em 20 jun. 2016.

- Nomes que continham o caractere “,” ou “;” separando partes do nome tiveram a ordem das palavras invertidas e o caractere “,” ou “;” removido. Por exemplo, “Silva, Jose Carlos” foi transformado em “Jose Carlos Silva”;
- A abreviação “jr” foi transformada na palavra “junior”;
- Múltiplos espaços em branco em sequência foram reduzidos para apenas um espaço em branco e os espaços em branco no começo e fim de cada nome foram removidos.

As estruturas, agrupadas em seu formato tabular, foram selecionadas usando a área ou grande área de concentração desejada, baseada nas áreas e grandes áreas do conhecimento<sup>4</sup> (ao todo foram consideradas 80 áreas associadas a 9 grandes áreas do conhecimento). Adicionalmente, os registros que possuem o campo “orientador” vazio ou o campo “membro da banca” vazio não foram consideradas nas análises das redes de convites.

A rede de participação em banca (i.e., rede de convite efetivo na avaliação do candidato ao título de mestre ou doutor) de defesa de trabalhos é construída a partir do agrupamento de trabalhos em três níveis de especificidade: por área, grande área e período de tempo (e.g., 1987-2011).

Neste trabalho, uma rede de participação em banca é representada por um grafo direcionado e sem pesos. O orientador principal e os membros da banca são representados pelos vértices e um convite de participação pelo orientador principal para cada um dos membros da banca é representado por uma aresta direcionada partindo do orientador principal e chegando em cada um dos membros da banca. Por motivos de simplificação, somente uma aresta é usada para representar um ou mais convites, isto é, caso um membro de uma banca tenha sido convidado em outro trabalho pelo mesmo orientador, só uma aresta consta no grafo representando os convites. Veja na Figura 1 um exemplo de rede participação de bancas criadas considerando os registros na sua forma tabular.

No processo de construção iterativa da rede de convites em bancas, os nomes do orientador principal e dos membros da banca são comparados com todos os nomes que já estão presentes na rede utilizando uma medida de similaridade baseada na distância de Levenshtein (LEVENSHEIN, 1966)<sup>5</sup>. Dessa forma, nomes que tenham uma distância de

---

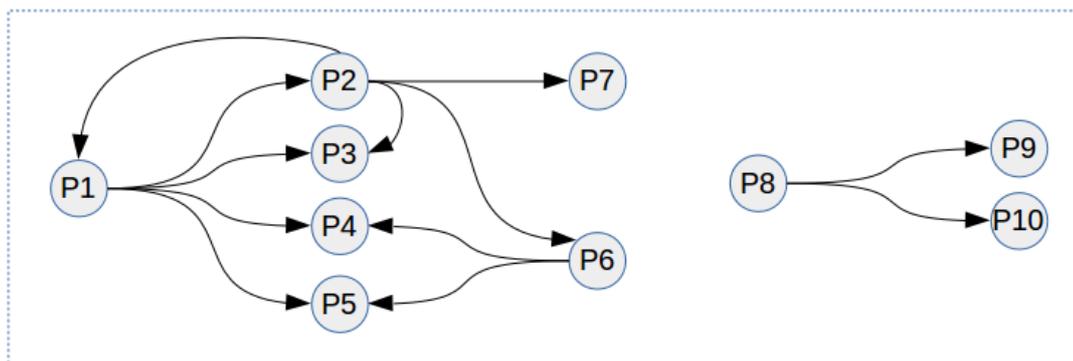
<sup>4</sup> A classificação das áreas e grandes áreas do conhecimento definidas pela CAPES está disponível em: <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>. Acesso em 15 abr. 2016.

<sup>5</sup> A distância de Levenshtein para duas cadeias (e.g., nomes completos de pesquisadores) pode ser entendida como o número mínimo de caracteres que devem ser substituídas ou removidos para transformar a primeira cadeia igual à segunda. Por exemplo “Mario Silva-Neto” e “Mario Silva Neto” estão a uma distância de

Levenshtein de 2 ou menos foram considerados como pertencentes à mesma pessoa, e representados pelo mesmo vértice no grafo. Note que, essa abordagem permite construir redes de bancas com nomes únicos, minimizando duplicações que podem ser gerados por registros com eventuais erros na escrita dos nomes.

T/D	Orientador	Membro da banca	Áreas T/D	Ano de defesa
Tese A	P1	P2, P3, P4, P5	A1, A2	Ano 1
Tese B	P2	P1, P3, P6, P7	A1	Ano 1
Dissertação C	P6	P4, P5	A2	Ano 2
Dissertação D	P8	P9, P10	A3, A4	Ano 3

Registros de trabalhos (T/D)



Rede de bancas de defesa

**Figura 1. Exemplo de rede de participação em bancas criado a partir de 4 registros de trabalhos (teses ou dissertações) acadêmicos. As setas, nas arestas, representam o convite do membro orientador (ou presidente) para os demais membros da banca (convidados efetivos).** Fonte: Elaborado pelos autores.

Finalmente, para a caracterização da rede de bancas de defesa foram consideradas 10 métricas ou indicadores globais<sup>6</sup> relacionadas apenas com informações de topologia de redes (FAUST; WASSERMAN, 1994):

- *Quantidade de vértices (pessoas)*, que é o número total de pesquisadores presentes na rede de participação em bancas;

Levenshtein igual a 1. Já “Luis Mendonca” e “Luiz Mendoca” estão a uma distância igual a 2.

<sup>6</sup> Uma métrica global pode ser entendida como uma característica geral da rede, isto é, uma medida que representa quantitativamente a rede como um todo.

- *Quantidade de arestas (convites)*, que é a quantidade mínima de convites realizados. A aresta direcionada, e sem pesos, representa quem convidou quem, independentemente do período e do número de vezes de participação de cada membro na banca;
- *Diâmetro da rede*, que é o caminho mais longo dentre os mais curtos, isto é, a maior distância entre dois vértices quaisquer ao tentar encontrar a menor distância entre estes dois vértices. Aqui são excluídas as distâncias entre os vértices que não estão conectados pois convencionalmente sua distância é infinita. Adicionalmente, para o caso de redes com mais de uma componente conexa, o diâmetro da rede corresponde ao diâmetro da maior componente conexa. Esta métrica representa a maior das menores distâncias entre pessoas na rede;
- *Caminho médio*, que é a distância média mínima entre dois vértices quaisquer. Esta métrica representa, na média, quão distantes estão as pessoas na rede. As redes com características de “mundo pequeno” geralmente contém caminhos médios pequenos;
- *Grau médio de um vértice*, refere-se à quantidade média de arestas que um vértice tem, independente da direção da aresta. Esta métrica representa a quantidade média de pessoas diferentes que participaram de bancas de defesa de tese ou dissertação;
- *Grau máximo total de um vértice*, refere-se à quantidade máxima total de vértices de entrada e saída. Esta medida representa o maior número de convites que uma pessoa teve ou fez para pessoas diferentes.
- *Grau máximo de entrada*, refere-se à quantidade máxima de arestas de entrada que um vértice qualquer tem. Esta métrica representa o maior número de convites que um indivíduo obteve de pessoas diferentes.
- *Grau máximo de saída*, refere-se à quantidade máxima de arestas de saída que um vértice qualquer tem. Similarmente ao grau máximo de entrada, esta métrica representa o maior número de convites que um indivíduo fez à pessoas diferentes.
- *Tamanho da maior componente conexa*, refere-se à maior quantidade de vértices que estão conectados na rede. Esta métrica representa o maior grupo de pessoas que estão conectadas de forma direta ou indireta;
- *Porcentagem da maior componente conexa*, refere-se à razão da quantidade de vértices da maior componente conexa sobre o total de vértices. Esta métrica representa a porcentagem de pessoas que estão conectadas de forma direta ou indireta na rede.

## 4 RESULTADOS

Os resultados apresentados correspondem aos obtidos após o processamento das 607.389 teses e dissertações acadêmicas ou profissionalizantes registradas no Banco de Teses e dissertações da CAPES no período de 1987 e 2011.

As teses e dissertações foram classificadas por área e grande área do conhecimento. Ao todo, foram criados 89 redes de convites de bancas de defesa (80 áreas + 9 grandes áreas) seguindo os procedimentos metodológicos descritos na seção anterior. As redes foram armazenadas em arquivos que facilitam a leitura<sup>7</sup>, visualização e caracterização topológica.

#### **4.1 Redes de convites de bancas nas grandes áreas do conhecimento**

Na Tabela 1 são apresentadas as 10 métricas topológicas consideradas para as redes de convites no nível de especificidade de grande área (AGR=“Ciências agrárias”; BIO=“Ciências biológicas”; ENG=“Engenharias”; EXA=“Ciências exatas e da terra”; HUM=“Ciências humanas”; LIN=“Linguística, letras e artes”; MUL=“Multidisciplinar”; SAU=“Ciências da saúde”; SOC=“Ciências sociais aplicadas”). A quantidade de trabalhos, que é número total de registros válidos de defesas de dissertação ou teses utilizadas para gerar a rede de bancas, estão indicadas na segunda coluna. Um registro é válido se este contém um orientador e pelo menos um membro de banca.

A grande área com maior número de trabalhos corresponde à área de “Ciências humanas”. Já, a grande área com menor número de trabalhos é “Multidisciplinar”. Por outro lado, a grande área com maior número de pessoas envolvidas nas redes de convites é “Ciências da saúde”. A grande área com menor número de pessoas envolvidas nas redes é “Linguística, letras e artes”. Considerando o número de convites únicos, as grandes áreas de “Ciências humanas” e “Ciências da saúde” apresentam a maior quantidade.

Em relação aos diâmetros da redes (i.e., a maior das menores distâncias entre pessoas na rede) a grande área “Multidisciplinar” apresenta o maior valor (31). Já a grande área de “Engenharias” apresenta o menor valor (17).

**Tabela 1: Métricas topológicas obtidas para as redes de convite de teses e dissertações associadas às nove grandes áreas.** Fonte: Elaborado pelos autores.

---

<sup>7</sup> O formato adotado para este trabalho corresponde ao GraphML que é um formato similar ao XML que permite um fácil intercâmbio de dados estruturados como redes (grafos) (TAMASSIA, 2013).

Grande Área	Trabalhos (T/D)	Vértices (pessoas)	Arestas (convites)	Diâ-Metro	Caminho Médio	Grau				Maior C. Conexa	
						Média	Máx. Tot.	Máx. Ent.	Máx. Saí.	Tamanho	Porc.
AGR	54804	28478	101578	18	5.91	7.13	179	147	121	27868	97.86%
BIO	56820	35433	121484	17	5.74	6.86	181	79	116	34845	98.34%
ENG	67491	34850	120385	17	5.62	6.91	223	99	149	34154	98.00%
EXA	56751	32893	113010	19	6.37	6.87	165	84	119	32222	97.96%
HUM	95841	42711	167366	19	5.44	7.84	209	117	134	42199	98.80%
LIN	24811	13140	43450	21	5.66	6.61	146	62	100	12850	97.79%
MUL	13637	16172	27531	31	10.56	3.40	77	26	61	15137	93.60%
SAU	80498	43875	165713	18	5.72	7.55	217	62	197	43393	98.90%
SOC	69268	30805	108656	21	6.00	7.05	249	65	195	30196	98.02%

Independentemente do número de trabalhos, pessoas e convites, o caminho médio neste tipo de redes é próximo a 6, evidenciando assim o fenômeno dos seis graus de separação (MILGRAM; TRAVERS, 1969). Comportamento similar é obtido quando analisado a porcentagem da maior componente conexa (aproximadamente 98%). Note que a rede da grande área “Multidisciplinar” não segue esses padrões, provavelmente por ser uma área com poucos registros no banco de teses e dissertações.

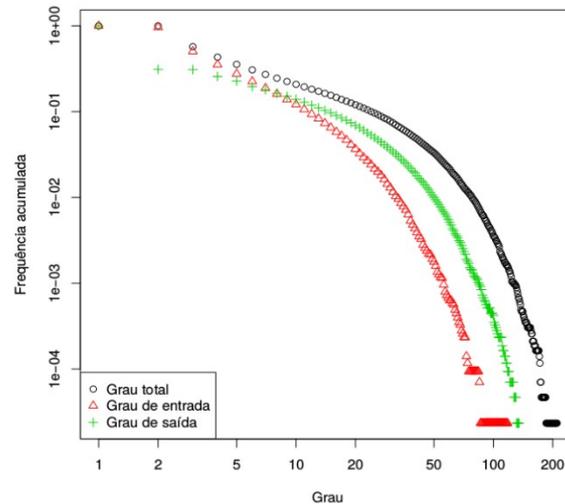
Finalmente, as informações sobre o grau médio e grau máximo total apresentam características próprias da interação de cada grande área. Considerando o grau máximo de entrada e grau máximo de saída podemos evidenciar assimetrias na forma dos convites, i.e., os graus máximos de saída são maiores aos graus máximos de entrada.

Na Figura 2 apresentam-se as distribuições cumulativas de grau para a rede de “Ciências humanas” (a grande área com o maior número de trabalhos registrados no conjunto de dados considerado). Essas distribuições cumulativas explicitam a assimetria evidenciada nos graus de entrada e de saída. Aqui é importante frisar que o mesmo comportamento foi identificado quando analisadas as distribuições cumulativas para todas as redes de convites para as outras grandes áreas do conhecimento.

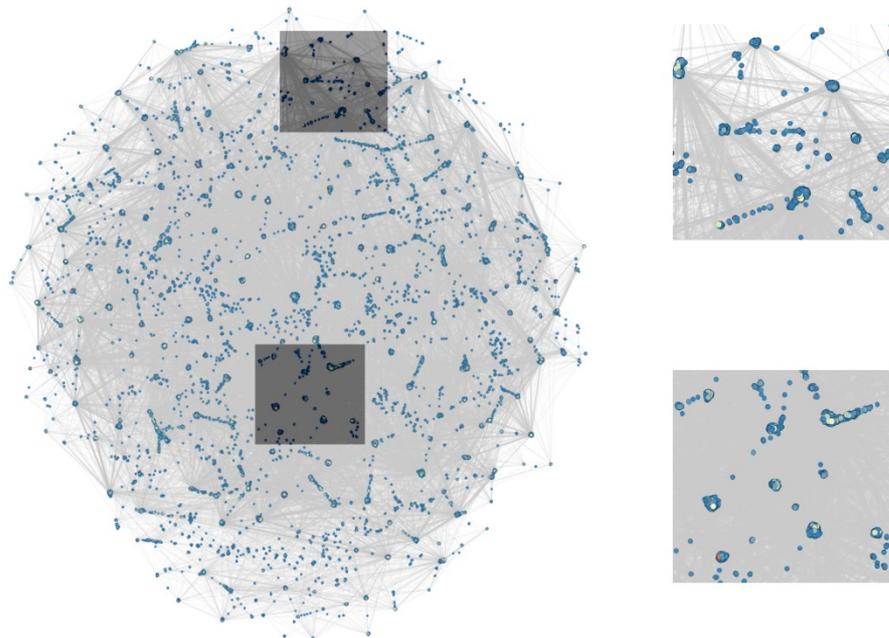
Além de cada grande área ser caracterizada por métricas topológicas em redes, também foram utilizadas ferramentas computacionais para visualizar cada rede. Para fins de exemplificação, na Figura 3 é apresentada a maior componente conexa da rede de “Ciências Humanas”, criada a partir da análise de 95841 trabalhos<sup>8</sup>. A maior componente conexa está composta de 42199 vértices (pessoas) e 166963 arestas (convites). O tamanho e cor do vértice é proporcional ao seu grau total (a colocação utilizada corresponde ao mapa de cor, de frio a quente). As duas regiões quadradas colocadas em destaque nessa figura evidenciam diferentes grupos de pesquisadores que participam de bancas de forma mais coesa. Em cada

<sup>8</sup> Para a visualização foi utilizado o algoritmo de desenho de grafos denominado OpenOrd com 2000 iterações, e fator de corte de 0,9 disponível no Programa Gephi 0.9.1 (<https://gephi.org/>).

agrupamento foram identificados pesquisadores com alto grau de entrada e alto grau de saída. Aparentemente esses agrupamentos estão compostos por pesquisadores da mesma ou similar área de atuação profissional (em uma futura etapa do projeto exploraremos detalhadamente esta relação).



**Figura 2. Distribuições cumulativas de graus de entrada, de saída e total para a rede de convites da grande área de “Ciências humanas”.** Fonte: Elaborado pelos autores.



**Figura 3. Rede de convites de defesa de teses e dissertações obtido para a grande área de “Ciências humanas”.** Fonte: Elaborado pelos autores.

#### 4.2 Redes de convites de bancas nas áreas do conhecimento

Na Tabela 2 são apresentadas as 10 métricas topológicas para as redes de convites no nível de especificidade de área do conhecimento. Ao todo foram avaliadas 80 áreas, sendo que para a área de Ciências Ambientais não foram identificados trabalhos, dentro do período

considerado, para a criação das redes. Em uma análise similar à de grandes áreas do conhecimento, podem ser observadas características topológicas similares para grupos de áreas diferentes. Nesse sentido, foi realizado uma análise de fatores (DARLINGTON et al., 1973) nos dados da Tabela 2. A análise fatorial permite sintetizar as informações contidas em variáveis originais em um conjunto menor de variáveis (i.e., fatores).

**Tabela 2: Métricas topológicas para as rede associadas a todas as áreas.**

Grande Área	Área	Trabalhos (T/D)	Vértices (pessoas)	Arestas (convites)	Diâ-Metro	Caminho Médio	Grau			Maior C. Conexa		
							Média	Máx. Tot.	Máx. Ent.	Máx. Sai.	Tamanho	Porc.
AGR	Agronomia	20455	12259	39511	17	5.63	6.45	153	63	110	11942	97.41%
	Ciência e Tecnologia de Alimentos	7555	5885	14517	14	5.13	4.93	130	53	86	5654	96.07%
	Engenharia agrícola	3818	3495	7385	14	5.33	4.23	107	35	91	3147	90.04%
	Medicina Veterinária	9467	6735	17808	15	5.10	5.29	100	49	75	6348	94.25%
	Recursos Florestais e Eng. Florestal	3773	4119	8097	15	5.37	3.93	96	46	69	3687	89.51%
	Recursos Pesqueiros e Eng. Pesca	1648	1566	3183	16	5.36	4.07	70	28	52	1452	92.72%
	Zootecnia	7730	4818	13409	15	5.20	5.57	95	43	73	4575	94.96%
BIO	Biofísica	1695	2070	4191	17	5.48	4.05	72	39	61	1909	92.22%
	Biologia Geral	1685	3920	4529	31	8.43	2.31	52	10	46	2875	73.34%
	Bioquímica	7587	8283	18323	20	5.88	4.42	109	37	87	7861	94.91%
	Botânica	4883	4760	11003	17	5.72	4.62	81	39	64	4542	95.42%
	Ecologia	8017	7233	17598	16	5.25	4.87	170	62	108	6910	95.53%
	Farmacologia	4322	5139	9483	16	5.67	3.69	118	37	101	4789	93.19%
	Fisiologia	4059	4505	8876	15	5.42	3.94	111	25	90	4188	92.96%
	Genética	6467	6471	14309	15	5.29	4.42	115	42	99	6277	97.00%
	Imunologia	2554	3397	6260	15	5.53	3.69	102	34	68	3203	94.29%
	Microbiologia	4700	5447	10318	18	6.16	3.79	75	33	60	5093	93.50%
	Morfologia	3473	4620	8154	17	5.81	3.53	147	31	116	4259	92.19%
	Oceanografia	1134	1178	2389	16	5.56	4.06	44	21	38	1127	95.67%
Parasitologia	1966	2764	4827	18	6.03	3.49	68	28	44	2617	94.68%	
	Zoologia	5125	4702	10727	14	5.45	4.56	83	37	65	4512	95.96%
SAU	Educação Física	4171	3275	7147	13	4.95	4.36	127	40	87	2983	91.08%
	Enfermagem	7945	5228	14873	14	4.57	5.69	145	50	105	5026	96.14%
	Farmácia	4051	4544	8392	21	5.96	3.69	77	31	60	4286	94.32%
	Fisioterapia e Terapia ocupacional	2061	3101	4492	18	5.87	2.90	68	30	58	2520	81.26%
	Fonoaudiologia	1783	1491	2937	11	4.35	3.94	104	32	73	1241	83.23%
	Medicina	28874	22408	69780	17	5.56	6.23	170	46	141	22004	98.20%
	Nutrição	3096	4077	7214	16	5.86	3.54	106	32	94	3723	91.32%
	Odontologia	15045	7622	27253	14	4.84	7.15	125	61	92	7469	97.99%
	Saúde Coletiva	11866	10584	24946	18	5.47	4.71	148	44	111	10129	95.70%
EXA	Astronomia	598	715	1530	13	4.58	4.28	42	21	27	617	86.29%
	Ciência da Computação	13743	7868	23807	17	4.98	6.05	133	61	97	7376	93.75%
	Física	8781	7428	20125	16	5.43	5.42	155	80	94	7056	94.99%
	Geociências	10051	7520	21134	15	5.38	5.62	92	44	61	7318	97.31%
	Matemática	5849	4735	11092	17	5.90	4.69	75	36	54	4293	90.67%
	Probabilidade e Estatística	1692	1906	3363	12	4.66	3.53	81	31	53	1436	75.34%
	Química	15327	9647	32013	14	4.93	6.64	163	61	119	9380	97.23%
HUM	Antropologia	3829	3633	7632	13	4.75	4.20	104	34	82	3321	91.41%
	Arqueologia	564	503	960	13	4.09	3.82	62	17	46	460	91.45%
	Ciência Política	5023	5171	9357	23	6.64	3.62	115	36	82	4434	85.75%
	Educação	37390	20002	65391	16	5.08	6.54	169	111	130	19546	97.72%
	Filosofia	5536	3894	9479	13	4.57	4.87	195	70	125	3485	89.50%
	Geografia	6363	4734	11435	13	4.79	4.83	141	46	95	4525	95.59%
	História	11731	7610	21259	12	4.72	5.59	141	56	95	7324	96.24%
	Psicologia	17167	12429	29619	18	5.21	4.77	163	48	117	11596	93.30%
	Sociologia	8272	7681	16995	15	5.38	4.43	118	38	89	7355	95.76%
		Teologia	1952	1246	2764	11	4.31	4.44	81	28	61	1151

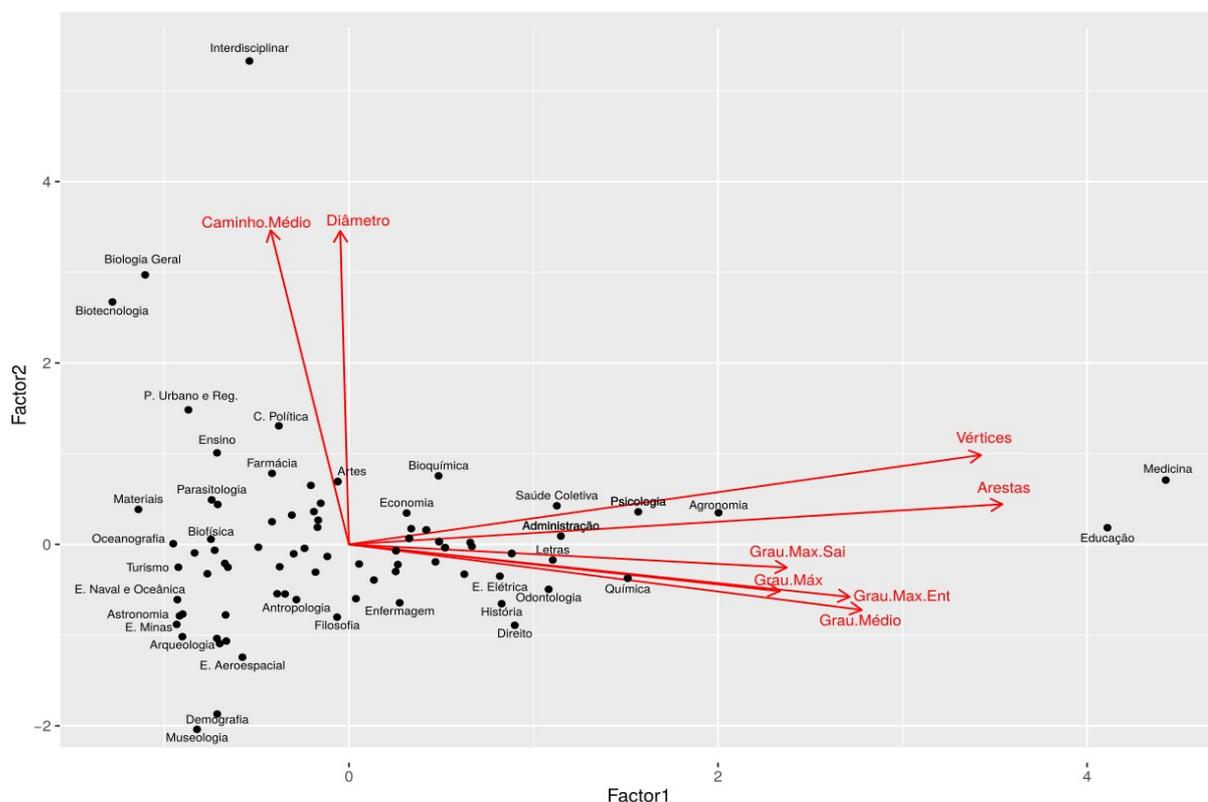
**Tabela 2: Métricas topológicas para as rede associadas a todas as áreas. (Cont.)**

SOC	Administração	18711	9009	27482	18	5.12	6.10	172	53	119	8544	94.84%
	Arquitetura e Urbanismo	5040	3958	9790	16	4.89	4.95	127	37	101	3559	89.92%
	Ciência da Informação	2210	2333	4245	15	5.25	3.64	88	36	63	2009	86.11%
	Comunicação	7765	5866	13825	17	4.87	4.71	245	50	195	5500	93.76%
	Demografia	481	804	1100	8	3.21	2.74	72	25	50	381	47.39%
	Desenho Industrial	616	657	1193	12	4.72	3.63	64	24	46	612	93.15%
	Direito	18623	6261	23852	13	4.46	7.62	215	65	156	5863	93.64%
	Economia	9046	6790	16638	17	5.75	4.90	105	37	81	6277	92.44%
	Museologia	133	261	300	7	2.98	2.30	27	7	20	148	56.70%
	Planejamento Urbano e Regional	2395	3231	4914	23	7.05	3.04	80	19	67	2727	84.40%
	Serviço Social	3755	3048	6345	13	5.03	4.16	158	51	107	2813	92.29%
	Turismo	1000	1231	1826	15	5.19	2.97	73	27	47	1004	81.56%
ENG	Engenharia Aeroespacial	1471	1506	3278	11	4.10	4.35	139	54	85	1352	89.77%
	Engenharia biomédica	2134	2607	4175	14	5.26	3.20	80	23	65	2309	88.57%
	Engenharia Civil	11029	6872	19652	14	5.12	5.72	154	61	105	6628	96.45%
	Engenharia de Materiais e Met.	6874	5384	13034	13	5.12	4.84	129	33	114	5284	98.14%
	Engenharia de Minas	520	620	989	12	4.47	3.19	51	14	37	557	89.84%
	Engenharia de Produção	9553	6839	15129	18	5.33	4.42	202	53	149	6370	93.14%
	Engenharia de Transportes	1495	1345	2581	12	4.20	3.84	79	29	54	1155	85.87%
	Engenharia Elétrica	11858	7666	22358	14	4.96	5.83	117	48	92	7337	95.71%
	Engenharia mecânica	8762	7155	17315	17	5.35	4.84	98	43	66	6718	93.89%
	Engenharia Naval e Oceânica	741	838	1402	14	4.67	3.35	55	17	45	666	79.47%
	Engenharia Nuclear	1538	1821	3071	13	5.42	3.37	114	24	90	1721	94.51%
	Engenharia Química	7405	6764	15385	16	5.40	4.55	142	45	110	6163	91.11%
Engenharia Sanitária	5371	5311	10574	19	5.66	3.98	95	43	67	4972	93.62%	
LIN	Artes	7249	5632	12993	20	5.94	4.61	73	36	53	5297	94.05%
	Letras	15696	8183	27549	16	5.07	6.73	146	61	100	8001	97.78%
	Linguística	2320	2198	4486	14	5.70	4.08	59	20	45	2099	95.50%
MUL	Biotecnologia	1296	2551	3257	24	9.23	2.55	34	15	30	2287	89.65%
	Ensino	3351	3163	6403	20	6.72	4.05	77	23	58	2824	89.28%
	Interdisciplinar	6901	8806	13634	41	11.03	3.10	70	19	61	7922	89.96%
	Materiais	471	983	1192	17	6.01	2.43	22	11	18	821	83.52%

Essa análise permitiu identificar apenas 2 dimensões latentes ou fatores que justificam as correlações observadas entre as 8 variáveis associadas às métricas topológicas das 79 redes de convites. As informações relacionadas com o tamanho e a porcentagem da maior componente conexa não foram consideradas para essa análise por estarem bem correlacionadas com o número total de vértices.

Na Figura 4 é apresentada um gráfico de dispersão para as redes de todas as áreas. Observe que as métricas topológicas são agrupadas com base em suas correlações. Foram identificados 2 grandes grupos. O primeiro grupo (no eixo “y”) relacionado com a métrica de caminho médio e diâmetro, e o segundo grupo (no eixo “x”) relacionado com as métricas de

número de vértices, número de arestas, e os graus dos vértices. É interessante notar a posição das áreas que estão nos extremos dos eixos (periferia).



**Figura 4. Gráfico de dispersão dos redes de convites de defesa de teses e dissertações para as 79 áreas do conhecimento obtidos a partir da análise de fatores.** Fonte: Elaborado pelos autores.

## 5 CONSIDERAÇÕES FINAIS

Neste trabalho, realizamos uma caracterização baseada na teoria dos grafos para as redes de participação em bancas de defesa de mestrado e doutorado. A participação considerada foi a de convite efetivo (i.e., participação efetiva na banca de defesa).

Ao todo analisamos mais de 600 mil registros de defesa correspondentes a trabalhos de mestrado e doutorado. Adotamos e desenvolvemos procedimentos computacionais para: (i) a coleta, extração e normalização de informação dos registros de defesa, (ii) criação das redes de convites através de grafos direcionados sem pesos, (iii) caracterização das redes usando métricas topológicas clássicas, e (iv) visualização de redes.

Por se tratarem de mais de 600 mil registros, a criação das redes de convites demandou muitos recursos computacionais. Em particular, a utilização da distância de

Levenshtein fez com que a complexidade computacional (tempo de processamento) aumentasse drasticamente. Foram necessários mais de 24 horas para que determinadas redes de convites fossem calculadas. Algumas otimizações foram feitas para diminuir a demanda por memória e técnicas de otimização, como *memoization*, não puderam ser usadas justamente pela limitação de memória computacional para o processamento.

Acreditamos que este trabalho servirá como ponto de partida para uma caracterização mais detalhada, no nível macro e meso das interações de convites de bancas de defesa de trabalhos acadêmicos, raramente exploradas na área de Ciência da Informação. Finalmente, este trabalho, como passos futuros, pretende caracterizar de forma temporal as redes de convites para responder perguntas relacionadas com a evolução das assimetrias. Assim, técnicas computacionais de reconhecimento de padrões e evolução em grafos serão também adotadas.

## REFERÊNCIAS

- ELMAGARMID, A. K.; IPEIROTIS, P. G.; VERYKIOS, V. S. Duplicate record detection: A survey. **IEEE Transactions on knowledge and data engineering**, v. 19, n. 1, p. 1-16, 2007.
- DARLINGTON, R. B.; WEINBERG, S. L.; WALBERG, H. J. Canonical variate analysis and related techniques. **Review of Educational Research**, v. 43, n. 4, p. 433-454, 1973.
- LEVENSHTAIN, V. I. Binary codes capable of correcting deletions, insertions, and reversals. **Soviet Physics Doklady**, v. 10, p. 707-710, Fevereiro, 1966.
- MENA-CHALCO, J. P.; ROCHA, V. Caracterização do banco de teses e dissertações da CAPES. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 4., 2014, Pernambuco. **Anais...** Recife: UFPE, 2014.
- MOODY, J. The structure of a social science collaboration network: disciplinary cohesion from 1963 to 1999. **American Sociological Review**, v. 69, n. 2, p. 213-238, Abril, 2004.
- NETO, J. E. B.; CUNHA, J. V. A. Colaboração Acadêmica em Bancas de Mestrado na Pós-Graduação Stricto Sensu em Contabilidade. **Contabilidade, Gestão e Governança**, v. 19, n. 1, p. 126-145, Janeiro/Abril, 2016.
- NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 2, p. 404-409, Janeiro, 2001.
- NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 1, p. 5200-5205, Abril, 2004.

ROSSONI, L. **A Dinâmica de relações no campo da pesquisa em organizações e estratégia no Brasil: uma análise institucional.** 2006. Dissertação – Universidade Federal do Paraná, Curitiba, 2006.

TAMASSIA, R.. **Handbook of graph drawing and visualization.** CRC press, 2013.

TRAVERS, J.; MILGRAM, S. An experimental study of the small world problem. **Sociometry**, p. 425-443, 1969.

UZZI, B.; SPIRO, J. Collaboration and creativity: the small world problem. **American Journal of Sociology**, v. 111, n. 2, p. 447-504, 2005.

WASSERMAN, S.; FAUST, K. **Social network analysis: Methods and applications.** Cambridge university press, 1994.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of “small-world” networks. **Nature**, v. 393, n. 4, p. 440-442, 1998.