ACH2197 - Análise de Redes Sociais

Análise da Rede Social Acadêmica Brasileira:

Um Estudo de Caso

Luciano Antonio Digiampietri

Roteiro

- Contextualização
- Introdução
- Conceitos Básicos
- Obtenção e Organização dos Dados
- Resolução de Nomes
- Análise de Grupos
- Predição de Relacionamentos
- Análises Adicionais

Análise de Redes Sociais

- Área que tem **atraído a atenção da comunidade** da computação nos últimos anos, especialmente após o crescimento das **redes sociais online**;
  - Banco de Dados; Teoria dos Grafos; Análise de Redes Complexas; Inteligência Artificial; Processamento de Língua Natural
- Relacionada também a **Linked Data** e **Big Data**;
- Surgimento de **periódicos e eventos específicos**;
- Alguns **grupos de pequisa** se destacam espalhados pelo mundo.

Grupo de Análise de Redes Sociais e Cientometria

- Grupo interdisciplinar sediado na EACH-USP e com colaboração de professores do:
  - Centro de Matemática, Computação e Cognição - UFABC
  - Departamento de Ciência da Computação - UFMG
  - Departamento de Biblioteconomia e Documentação - ECA-USP

Introdução/Motivação

- Atualmente há uma grande quantidade de **informações disponíveis** sobre diversos tipos de atividades acadêmicas;
- Muitas dessas atividades são realizadas de maneira **colaborativa**;
- O Brasil possui uma base que integra boa parte dessa informação: a base da **Plataforma Lattes**.

Introdução/Motivação

- Diversos processos envolvem a **análise** (comparativa) **de pesquisadores ou grupos** de pesquisadores;

- Idealmente, **políticas científicas** devem ser construídas baseadas num profundo entendimento da situação nacional e seus resultados devem ser cuidadosamente avaliados;

- Em meio à enorme quantidade de informações, identificar **quais artigos ler** ou para quais **pesquisadores solicitar colaboração** são atividades complexas.

Objetivos

- **Caracterização/Entendimento** da rede social acadêmica nacional;
- Especificação (ou extensão) de **metodologias** para tratar os problemas relacionados;
- Desenvolvimento e testes de diferentes **algoritmos/soluções** para tratar problemas de análise de redes sociais e cientometria.

Conceitos Básicos

Rede social

- **Rede social** é uma estrutura composta por **indivíduos** (pessoas ou organizações) que são **conectados** por um ou mais tipos de relações, por exemplo, amizade, crença ou trabalho.

- Pode ser representada como um **grafo**.

- **Análise de redes sociais**: originada na antropologia social e sociologia considera que, muitas vezes, **o entendimento das relações entre os indivíduos é mais importante do que o entendimento das características individuais**.

Rede Social Acadêmica

- **Rede social acadêmica** é uma rede social que representa relações **acadêmicas** entre os indivíduos.
- Tipicamente os indivíduos são **professores**, **alunos**, **pesquisadores** ou **instituições** de pesquisa/**grupos** de pesquisa.
- **Relacionamentos** mais frequentemente utilizados:
  - Colaboração (tipicamente de coautoria);
  - Orientação;
  - Ensino (aluno x professor).

Métricas da análise de redes sociais

- Componente conexo / componente gigante
- Diâmetro
- Média dos caminhos mínimos
- Tamanho da clique máxima
- Densidade
- Assortatividade
- Centralidade /centralização
- Coeficiente de aglomeração (clusterização)

Plataforma Lattes e Currículo Lattes

*"... **padrão nacional** no registro da vida pregressa e atual dos estudantes e pesquisadores do país, e é hoje adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa do País"*

CNPq

- Mais de **7,3 milhõe**s de currículos, disponíveis nos formatos HTML e XML;

- Estruturados em **oito seções** principais: dados gerais, formação, atuação, projetos, produções, eventos, orientações e bancas.

Análise bibliométrica

- **Bibliometria**: estudo dos aspectos quantitativos da produção, disseminação e uso da informação.
- **Cientometria**: análise quantitativa da atividade de pesquisa científica, estudando tanto os recursos e os resultados quanto a organização e as técnicas de produção científica.
- Diversos tipos de medidas, relacionadas a:
  - Número de publicações;
  - Autores por publicação;
  - Qualificação dos veículos de publicação;
  - Número de citações.

Obtenção, organização e refinamento dos dados

- Atualmente, os currículos Lattes podem ser obtidos de duas maneiras principais:
  - Via ***website*** do CNPq, nos formatos XML ou HTML, sendo necessário passar por um *captcha*.
  - **Institucionalmente**: cada instituição pode solicitar ao CNPq os currículos de seus pesquisadores.

- Existem algumas ferramentas disponíveis para auxiliar no processo de obtenção e organização dos currículos:
  - scriptLattes
  - LattesMiner

Obtenção, organização e refinamento dos dados

- Limpeza
- Enriquecimento
- Construção de atributos

Atualização, corretude e completude dos dados

- São três aspectos muito importantes para qualquer tipo de análise de dados;
- Nenhum deles é assegurado ao se utilizar dados de currículos Lattes.

Resolução de entidades

- Resolução de entidades (*entity resolution*) é o processo de determinar se **duas referências** a objetos do mundo real se referem ou não ao **mesmo objeto**.

- Em redes sociais acadêmicas visa a identificar se:
  - Duas referências a publicações se referem à mesma **publicação**;
  - Duas referências a pessoas se referem à mesma pessoa.

Análise de grupos

- Análises bibliométricas x Análises de redes sociais

- Grupos podem ser definidos por:
  - região geográfica;
  - áreas de atuação;
  - local de trabalho (instituição ou departamento);
  - programa de pós-graduação;
  - etc.

Análise de tendências

- Modelagem e análise de um conjunto de dados de forma a **entender o comportamento** desses dados e **prever valores futuros**.

- Análise de tendências de **publicações científicas**:
  - Identificação de **tópicos/assuntos/subáreas**;
  - Modelar e prever o comportamento utilizando não apenas as **publicações** mas também as **fontes produtoras** de informação.

Dinâmica da Rede

- Costuma ser analisada de acordo com o **surgimento ou exclusão de arestas** entres os indivíduos da rede.

- Pode considerar **outros aspectos**:
  - surgimento ou a exclusão de indivíduos (nós);
  - a variação de atributos relacionados às arestas (por exemplo, variação no peso das arestas);
  - variação nos atributos estruturais da rede;
  - etc.

Predição de relacionamentos

- Visa a prever relacionamentos futuros em uma rede social.
  - Pode ser utilizada tanto para **encontrar amigos** que ainda não estavam ligados em numa rede social online, quanto para **potencializar a realização de trabalhos** em uma comunidade científica ou em empresas.

  - Tipicamente é utilizada para prever relacionamentos **novos/inéditos**, mas pode ser utilizada para prever a reincidência de relacionamentos ou o fim de um relacionamento.

Atividades realizadas e resultados

- Obtenção e Organização dos Dados
- Resolução de Nomes
- Análise de Grupos
- Predição de Relacionamentos
- Resultados Adicionais

Obtenção e Organização dos Dados

- Obtenção dos identificadores dos currículos
- Processamento inicial dos currículos
- Banco de dados relacional
- Enriquecimento do conjunto de dados
- Análise da atualização dos dados

Obtenção dos Identificadores dos CVs

- Uso dos motores de busca do Google e da Microsoft (com base nos nomes das pessoas procuradas);
- Uso da interface de busca por currículos do CNPq (por área);
- Uso da interface de busca por currículos do CNPq (por pessoa);

Obtenção dos identificadores dos CVs - Resultados

- Utilizando os **motores de busca** do Google e da Microsoft:
  - 82% dos 1.002 currículos de professores procurados foram encontrados;
- Uso da **interface de busca por currículos do CNPq**:
  - "Ciência da Computação": **2.600** currículos retornados
  - Estes currículos apontavam para, direta ou indiretamente, para mais de **1.200.000** outros currículos;
- Uso da interface de busca por currículos do CNPq:
  - ***String* vazia de busca**:3,2 milhões de currículos em 2013; 4,2 no primeiro semestre de 2015; e **7,3 milhões** em março de 2022.

Processamento inicial dos currículos

- Processamento inicial dos arquivos HTML:
  - Conversão para arquivos XML;
  - Criação de um banco de dados relacional.

- **1.236.548 currículos**;
  - 1.378.885 projetos de pesquisa
  - 3.250.846 registros sobre formação/titulação;
  - 3.256.019 de registros de áreas de atuação;
  - 4.329.993 registros de orientações;
  - 11.529.218 publicações (incluindo redundâncias)

Enriquecimento do conjunto de dados

- Enriquecimento de **parte** da base de dados.
- **Programas de pós-graduação**:
  - Ciência da Computação: professores de cada programa e notas dos programas nos triênio 2004-2006 e 2007-2009;
- Informações sobre **veículos de publicação**:
  - fator de impacto JCR, índice SJR e Qualis.
- **Citações** dos artigos completos:
  - Google Scholar e Microsoft Academic Search.
- **Medidas derivadas**:
  - Índices G e H.

Análise da atualização dos dados

- Verificação de há quanto tempo cada currículo foi atualizado
  - Por área de atuação;
  - Por formação máxima;

- Estimativa da informação faltante

Análise da atualização dos dados

Dias desde a última atualização por áreas

Grande Área	Total CVs	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Sem Grande Área	1.006.088	124	328	637	978	1173	1425	1718	2121	2631	5782
Ciências Agrarias	141.73	31	63	103	168	278	441	661	1043	2066	5779
Ciências Biológicas	175.384	28	59	97	151	247	398	613	1001	1944	5783
Ciências da Saúde	459.825	45	95	158	257	397	530	741	1093	1921	5780
Ciências Exatas e da Terra	266.693	40	78	130	218	344	501	747	1187	2251	5796
Ciências Humanas	417.334	37	78	130	201	316	478	670	975	1783	5779
Ciências Sociais Aplicadas	439.33	47	99	168	275	418	561	776	1146	2009	5795
Engenharias	192.495	44	90	151	258	403	566	851	1450	2633	5814
Linguística, Letras e Artes	157.597	40	83	137	218	333	487	684	1012	1834	5599
Outros/Multidisciplinar	337.017	49	86	125	179	242	330	420	469	708	5780
Mundo Lattes	3.187.710	55	112	196	333	486	715	1045	1523	2293	5814

(DIGIAMPIETRI et al., 2014b)

Análise da atualização dos dados

Dias desde a última atualização por formação

Maior Formação/Titulação	Total CVs	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Sem Formação Declarada	108.758	621	793	1202	2021	2290	2442	2584	2804	3202	5814
Ensino Fundamental Primeiro Grau	15.508	123	176	434	561	697	783	841	921	1156	4157
Ensino Médio Segundo Grau	121.554	112	196	368	485	621	697	781	849	1114	4284
Curso Técnico Profissionalizante	30.94	205	552	678	781	878	1044	1329	1762	2604	5786
Graduação	2.203.076	70	141	247	384	523	791	1138	1561	2239	5796
Mestrado	428.264	34	74	124	187	299	478	747	1195	2093	5787
Doutorado	279.61	19	34	55	79	114	167	281	533	1314	5780
Produtividade	13.787	10	18	22	32	42	58	78	111	134	2534

(DIGIAMPIETRI et al., 2014b)

Média dos artigos publicados (3 anos)

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	0,023	0,000	0,009	0,013	0,095	0,711	2,369	-
Ciências Agrárias	0,108	0,013	0,021	0,058	0,259	1,288	6,243	22,791
Ciências Biológicas	0,189	0,006	0,028	0,079	0,198	0,907	4,920	17,827
Ciências da Saúde	0,107	0,000	0,010	0,012	0,173	1,032	5,529	24,151
Ciências Exatas e da Terra	0,074	0,001	0,010	0,038	0,131	0,917	5,377	16,751
Ciências Humanas	0,036	0,003	0,021	0,037	0,157	1,082	4,318	12,214
Ciências Sociais Aplicadas	0,034	0,000	0,007	0,008	0,113	0,982	4,566	15,930
Engenharias	0,076	0,000	0,007	0,032	0,153	0,977	6,881	25,094
Linguística, Letras e Artes	0,044	0,000	0,020	0,039	0,120	0,949	3,331	7,570
Outros/Multidisciplinar	0,051	0,001	0,004	0,010	0,033	0,839	3,995	21,290

(DIGIAMPIETRI et al., 2014b)

Estimativa média de artigos faltantes

Grande Área/ Maior Formação/Titulação	Sem Formação Declarada	Ensino Fundamental	Ensino Médio	Curso Técnico Profissionalizante	Graduação	Mestrado	Doutorado	Produtividade
Sem Grande Área	0,051	0,000	0,009	0,011	0,065	0,287	1,185	-
Ciências Agrárias	0,178	0,035	0,005	0,041	0,189	0,694	1,337	1,281
Ciências Biológicas	0,307	0,015	0,016	0,095	0,145	0,426	1,087	0,982
Ciências da Saúde	0,152	0,000	0,006	0,015	0,128	0,455	1,087	1,294
Ciências Exatas e da Terra	0,126	0,003	0,004	0,053	0,098	0,467	1,343	1,152
Ciências Humanas	0,049	0,002	0,005	0,012	0,084	0,347	0,725	0,684
Ciências Sociais Aplicadas	0,045	0,000	0,002	0,012	0,071	0,355	0,786	0,970
Engenharias	0,173	0,000	0,006	0,054	0,137	0,637	1,569	1,606
Linguística, Letras e Artes	0,071	0,000	0,005	0,024	0,061	0,297	0,586	0,458
Outros/Multidisciplinar	0,035	0,000	0,001	0,003	0,010	0,219	0,581	1,021

(DIGIAMPIETRI et al., 2014b)

Resolução de Nomes

- Resolução de publicações
- Resolução de nomes de autores

Resolução de Publicações

- Pode considerar diferentes aspectos
  - Título:
    - Títulos iguais;
    - Filtros;
    - Casamento aproximado;
    - N-gramas;
  - Informações adicionais:
    - Ano;
    - Veículo de publicação;
    - Lista de autores;
    - Etc.

Resolução de Publicações

- Exemplo considerando apenas os títulos.
- Amostra: os **486 artigos** publicados pelos professores permanentes da programa de pós-graduação em Ciência da Computação do IME-USP no triênio 2007-2009.
- **96 artigos em coautoria**;
- **595 registros** nos currículos (**205** referentes aos 96 artigos em coautoria).

Estratégia	Pré-processamento	VP	FP	Precisão	Revocação
Casamento exato de títulos	nenhum	164	0	100,0%	80,0%
Casamento exato de títulos	filtro de acentos e pontuações	167	0	100,0%	81,5%
Substring dos títulos	filtro de acentos e pontuações	181	11	94,3%	88,3%
Distância de edição 1	nenhum	169	0	100,0%	82,4%
Distância de edição 2	nenhum	171	0	100,0%	83,4%
Distância de edição 3	nenhum	174	0	100,0%	84,9%
Distância de edição 4	nenhum	177	0	100,0%	86,3%
Distância de edição 5	nenhum	178	0	100,0%	86,8%
Distância de edição 6	nenhum	179	0	100,0%	87,3%
Distância de edição 7	nenhum	180	1	99,4%	87,8%
Distância de edição 8	nenhum	181	5	97,3%	88,3%
Distância de edição 9	nenhum	183	8	95,8%	89,3%
Distância de edição 10	nenhum	183	15	92,4%	89,3%
Distância de edição 1	filtro de acentos e pontuações	174	0	100,0%	84,9%
Distância de edição 2	filtro de acentos e pontuações	177	0	100,0%	86,3%
Distância de edição 3	filtro de acentos e pontuações	178	0	100,0%	86,8%
Distância de edição 4	filtro de acentos e pontuações	180	0	100,0%	87,8%
Distância de edição 5	filtro de acentos e pontuações	181	0	100,0%	88,3%
Distância de edição 6	filtro de acentos e pontuações	182	0	100,0%	88,8%
Distância de edição 7	filtro de acentos e pontuações	183	1	99,5%	89,3%
Distância de edição 8	filtro de acentos e pontuações	183	5	97,3%	89,3%
Distância de edição 9	filtro de acentos e pontuações	183	9	95,3%	89,3%
Distância de edição 10	filtro de acentos e pontuações	183	16	92,0%	89,3%

Estratégia	Pré-processamento	VP	FP	Precisão	Revocação
Distância de edição 2,5%	nenhum	173	0	100,0%	84,4%
Distância de edição 5%	nenhum	176	0	100,0%	85,9%
Distância de edição 7,5%	nenhum	178	0	100,0%	86,8%
Distância de edição 10%	nenhum	181	0	100,0%	88,3%
Distância de edição 12,5%	nenhum	181	0	100,0%	88,3%
Distância de edição 15%	nenhum	182	2	98,9%	88,8%
Distância de edição 17,5%	nenhum	183	10	94,8%	89,3%
Distância de edição 20%	nenhum	184	11	94,4%	89,8%
Distância de edição 22,5%	nenhum	186	11	94,4%	90,7%
Distância de edição 25%	nenhum	189	11	94,5%	92,2%
Distância de edição 2,5%	filtro de acentos e pontuações	177	0	100,0%	86,3%
Distância de edição 5%	filtro de acentos e pontuações	179	0	100,0%	87,3%
Distância de edição 7,5%	filtro de acentos e pontuações	180	0	100,0%	87,8%
Distância de edição 10%	filtro de acentos e pontuações	181	0	100,0%	88,3%
Distância de edição 12,5%	filtro de acentos e pontuações	181	0	100,0%	88,3%
Distância de edição 15%	filtro de acentos e pontuações	183	2	98,9%	89,3%
Distância de edição 17,5%	filtro de acentos e pontuações	184	11	94,4%	89,8%
Distância de edição 20%	filtro de acentos e pontuações	186	11	94,4%	90,7%
Distância de edição 22,5%	filtro de acentos e pontuações	190	11	94,5%	92,7%
Distância de edição 25%	filtro de acentos e pontuações	190	11	94,5%	92,7%

Resolução de Publicações

- Título e informações adicionais.

- Duas referências são consideradas compatíveis se:
  - os títulos forem compatíveis E
  - a lista de autores for compatível E
  - as demais informações forem compatíveis

Resolução de Publicações

- os títulos forem compatíveis:
  - são iguais ***OU*** (se a diferença entre o tamanho dos dois títulos for menor do que um terço da soma do tamanho dos títulos ***E*** ambos possuem mais de 10 caracteres e um estiver contido dentro do outro) ***OU*** (a Distância de Edição entre os dois títulos for menor do que 5)
- ***E*** a lista de autores for compatível:
  - Houver mais autores em comum do que diferentes, considerando-se apenas o casamento exato do último sobrenome de cada autor
- ***E*** as demais informações forem compatíveis:
  - Ao menos dois dos seguintes quatro campos forem iguais: ano de publicação, local, páginas e volume.

Resolução de Publicações

- Título e informações adicionais.

- Duas referências são consideradas compatíveis se:
  - os títulos forem compatíveis E
  - a lista de autores for compatível E
  - as demais informações forem compatíveis

- **Resultados (classe positiva):**
  - **Precisão: 98,9%**
  - **Revocação: 96,3%**

Resolução de Publicações

- O cálculo da **distância de edição** é custoso;

- Comparar todos os pares de títulos nem sempre é viável (ou mesmo desejado):
  - Comparar títulos de **anos próximos**;
  - Comparar títulos que inicial pelas **mesmas letras**;
  - Comparar títulos que possuam as **mesmas palavras**;
  - Comparar títulos de publicações do **mesmo tipo**;
  - Etc.

Resolução de Nomes de Autores

- Encontrar um autor específico em uma referência;
- Encontrar todas as referências de cada autor;

Encontrar autores específicos em referências

- Dados:
  - Banco de Dados Bibliográficos da Universidade de São Paulo (Dedalus)
  - Sistema de apoio à avaliação e a gestão institucional da USP (Tycho3)

- Objetivo:
  - Identificar **como cada autor está sendo referenciado**, quantificando cada tipo de variação e potenciais problemas de cadastro.

Encontrar autores específicos em referências - critérios utilizados

1. Busca pelo nome completo do docente;
2. Busca pelo nome do docente, permitindo-se que um ou mais nomes do meio estejam abreviados.;
3. Consideraram-se os mesmos critérios anteriores, porém permitindo-se a ausência ou excesso do último sobrenome (e neste caso, exige-se que o sobrenome anterior seja encontrado).
4. Combinação de diferentes variações …

Encontrar autores específicos em referências - dados

- Dados:
  - **12.628 registros bibliográficos**, correspondendo a produção bibliográfica de 2006 a 2010 de apenas quatro unidades da USP: Escola de Artes, Ciências e Humanidades (EACH), Escola de Comunicações e Artes (ECA), Faculdade de Educação (FE) e Instituto de Física de São Carlos (IFSC).

  - **1.137 autores docentes** diferentes.

Encontrar autores específicos em referências - resultados

- Resultados:
  - **74,2% dos registros** foram identificados pelo primeiro critério (casamento exato dos nomes).
  - **92,3% dos docentes** foram identificados corretamente pelo casamento exato dos nomes.
  - **99,5% dos registros** foram identificados e a validação manual realizada sobre 500 registros indicou a ausência de falso-positivos.
  - Dos 0,5% não identificados destaca-se o cadastro incorreto no **número de autores** (cerca de 20%) e casos em que o nome encontrado na lista de autores é um **apelido ou pseudônimo** (por exemplo, "Toninho" ao invés de "Antônio Augusto").

Encontrar autores específicos em referências - diferenças

Variações encontradas nas referências

Tipos de variações de nomes encontrados	Total de docentes	Total de ocorrências
Nomes a menos	185	1.466
Sobrenomes a menos	41	376
Abreviações	79	179
Nomes com diferenças	15	56
Sobrenomes parecidos	9	43
Sobrenomes a mais	7	37
Nomes parecidos	13	35
Nomes invertidos	0	0

(MUGNAINI et al. 2012b)

Encontrar todas as referências de cada autor

- Objetivo: dado um conjunto de publicações / referências bibliográficas:
  - **identificar todos os autores envolvidos** (e quais as publicações de cada autora);
  - dada uma lista de autores, **identificar todas as publicações desses autores**.

Encontrar todas as referências de cada autor

- Base de dados utilizada: **DBLP** (*Digital Bibliography & Library Project*)
  - mais de **3,1 milhões** de registros de **publicações**
  - mais de **1,6 milhões de autores**
- Amostra para testes:
  - os **48 professores** do programa de pós-graduação em Ciência da Computação da Unicamp
- Estratégia:
  - **combinação de diferentes métricas** utilizando algoritmos de inteligência artificial (classificação binária).

Encontrar todas as referências de cada autor

- Dada a lista dos 48 docentes, foram encontrados **82 registros** que possuíam o nome e o sobrenome dos docentes:
  - **29 identificações únicas** e corretas dos professores;
  - **17 blocos contendo entre 2 e 12 registros** de autores cada;
  - 2 professores não tiveram **nenhum registro** encontrado no DBLP.
- Para os **17 blocos** com mais de um registro foram calculadas 14 características para cada par de registros dentro do bloco (102 pares).

Encontrar todas as referências de cada autor

- **Quatro tipos de características**, baseadas em:
  - nomes dos autores;
  - rede social de coautorias;
  - mineração de texto (baseada nos títulos dos artigos)
  - datas de publicação dos artigos.

Encontrar todas as referências de cada

Tipo de característica	Característica	Descrição
características da rede social de coautorias	vizinhos em comum	número de vizinhos em comum na rede social acadêmica de todos os autores da DBLP
	são vizinhos	indica se o par de autores é ou não vizinho na rede social
características extraídas dos nomes	distância de edição	distância de edição entre os nomes dos dois autores
	distância relativa	proporção entre a distância de edição e o tamanho do menor nome dentre os autores
	primeiro nome diferente	indica se os autores têm seus primeiros nomes diferentes (um do outro)
	último nome diferente	indica se os autores têm seus últimos nomes diferentes (um do outro)
	proporção de diferentes nomes do meio	proporção (em relação ao número total de nomes) de nomes diferentes entre os autores, contabilizando nomes adicionais como diferentes
	proporção de diferentes abreviações	proporção (em relação ao número total de nomes) de nomes abreviados diferentes entre os autores, contabilizando abreviações adicionais como diferentes
	nomes invertidos	indica se há inversão da posição das partes dos nomes entre os autores
	nomes ou abreviações diferentes	indica a proporção de nomes efetivamente diferentes entre os autores (sem contar a presença de nomes adicionais)
características baseadas na mineração de texto	mineração de texto dos títulos dos artigos	métrica baseada em TFIDF que compara a frequência das palavras dos títulos entre os dois autores e entre o corpus formado pelos títulos de todos os autores avaliados
	log(MT)	logaritmo do valor resultante da mineração de texto
características baseadas nos anos de publicação dos artigos	intersecção do período de publicação	interseção entre os períodos de publicação dos dois autores
	distância em anos entre publicações	distância mínima em anos entre as publicações dos dois autores (apenas se a interseção for igual a zero)

Digiampietri, Barbosa e Linden (2015)

Encontrar todas as referências de cada autor

- Características mais correlacionadas com a classe:
  - As características relacionadas ao nome:
    - proporção de diferentes nomes do meio;
    - nomes ou abreviações diferentes;
    - último nome diferente
  - Vizinhos em comum

- Características mais selecionadas por seletores de atributos:
  - Relacionadas ao nome
  - Relacionadas aos títulos (mineração de texto)

Encontrar todas as referências de cada autor

Resultados para a classificação binária: duas referências se referem ou não a um mesmo autor

classe	VP	FP	Precisão	Revocação
F	1	0,333	0,957	1
T	0,667	0	1	0,667
Média Ponderada	**0,961**	**0,294**	**0,962**	**0,961**

classe	Medida-F	Área ROC
F	0,978	0,977
T	0,8	0,977
Média Ponderada	**0,957**	**0,977**

Digiampietri, Barbosa e Linden (2015)

Análise de Grupos

- Análise dos programas de pós-graduação em Ciência da Computação

- Análise de pesquisadores de acordo com sua localização geográfica

Análise dos programas de pós-graduação em CC

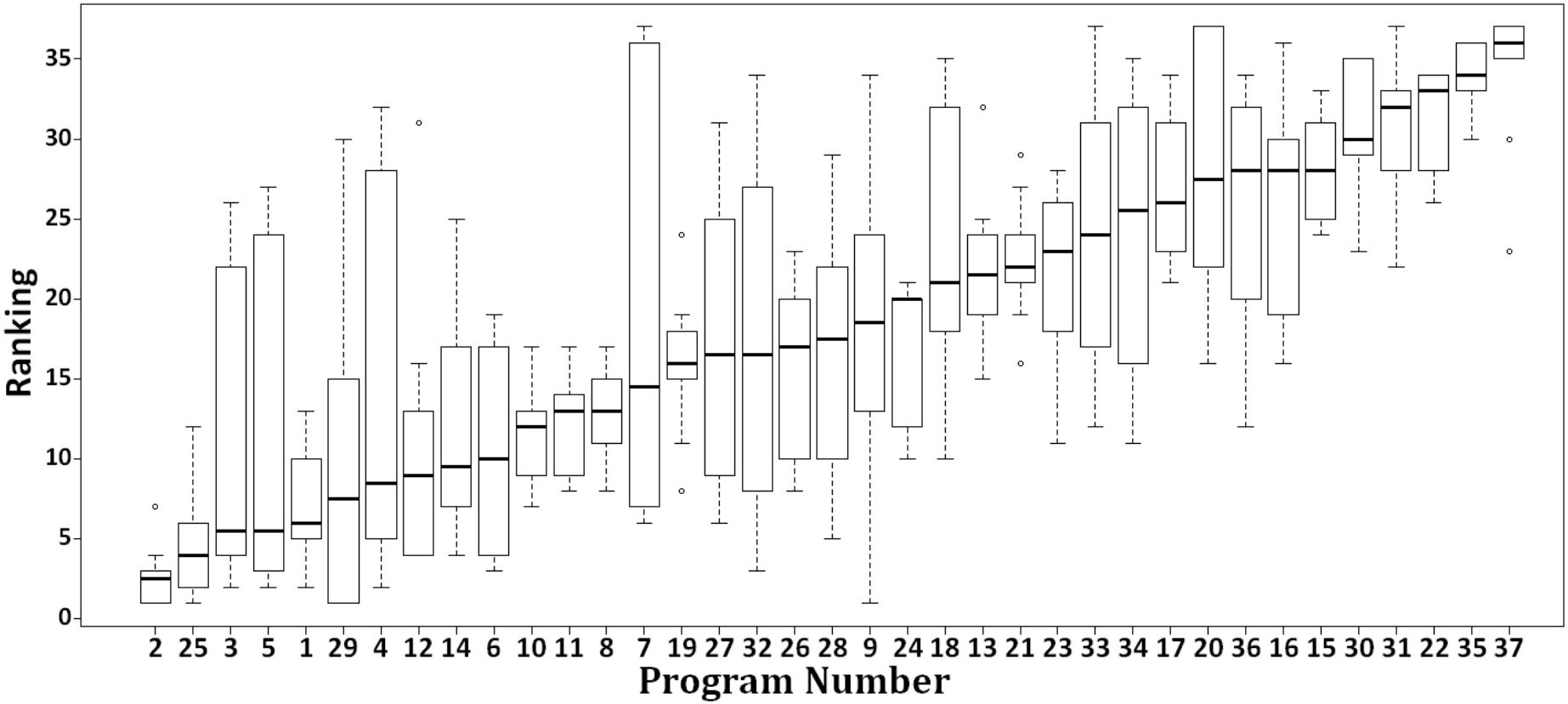
- Objetivo: estudar os programas de pós-graduação utilizando **diferentes perspectivas**.
  - **37 programas** brasileiros de pós-graduação em CC com mestrado acadêmico e/ou doutorado em ambos os triênios 2004-2006 e 2007-2009;
  - 732 professores permanentes;
  - 17.976 artigos completos (22,5% em periódicos);
  - 1.428 relações de coautoria (arestas no grafo)

Análise dos programas de pós-graduação em CC

- Medidas utilizadas
  - Citações (Microsoft Academic Search e Google Scholar);
  - Fatores de impacto/qualificadores: JCR, SJR e Qualis;
  - Índices H e G;
  - Número de artigos completos

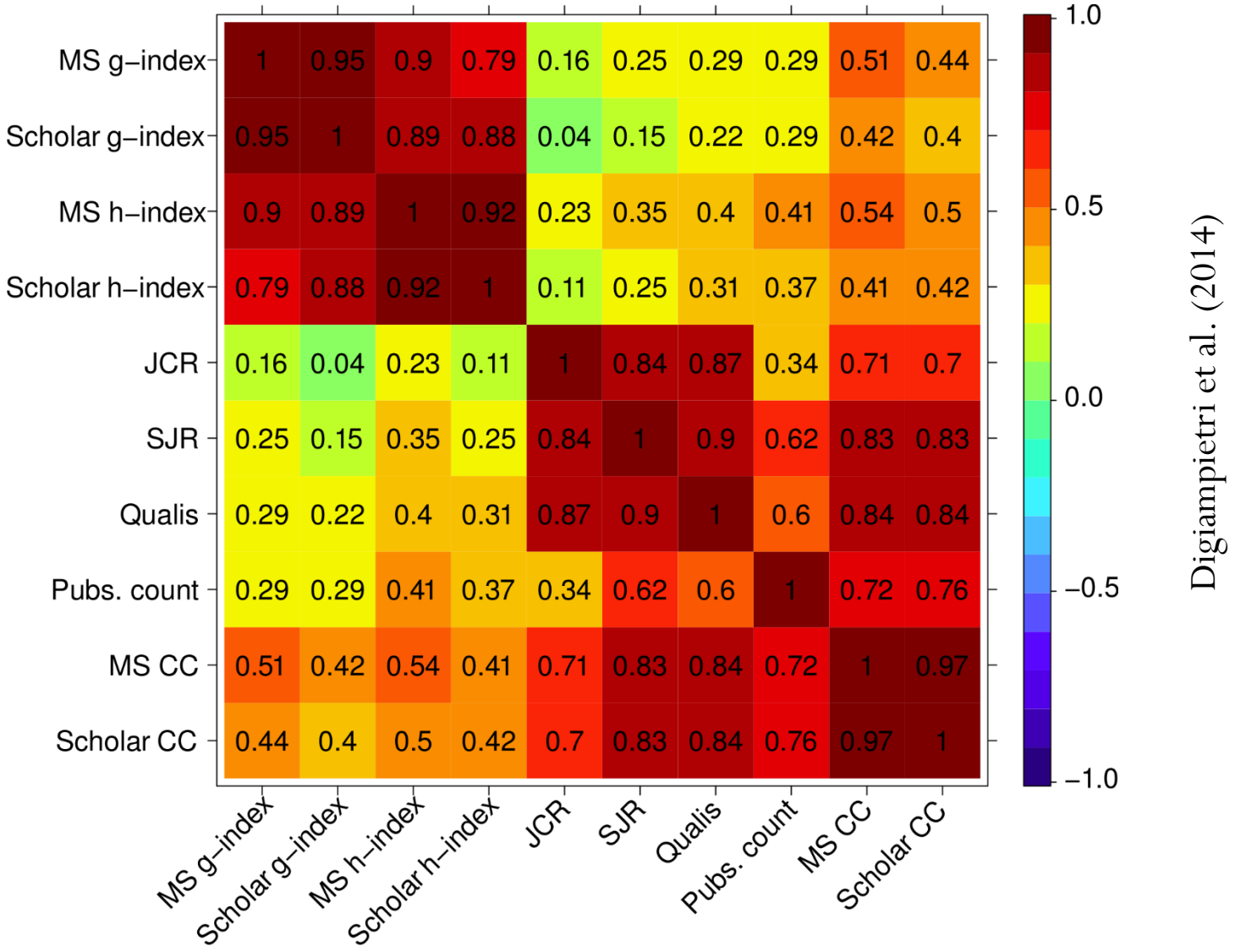
Análise dos programas de pós-graduação em CC

Variação do ranqueamento de acordo com a métrica

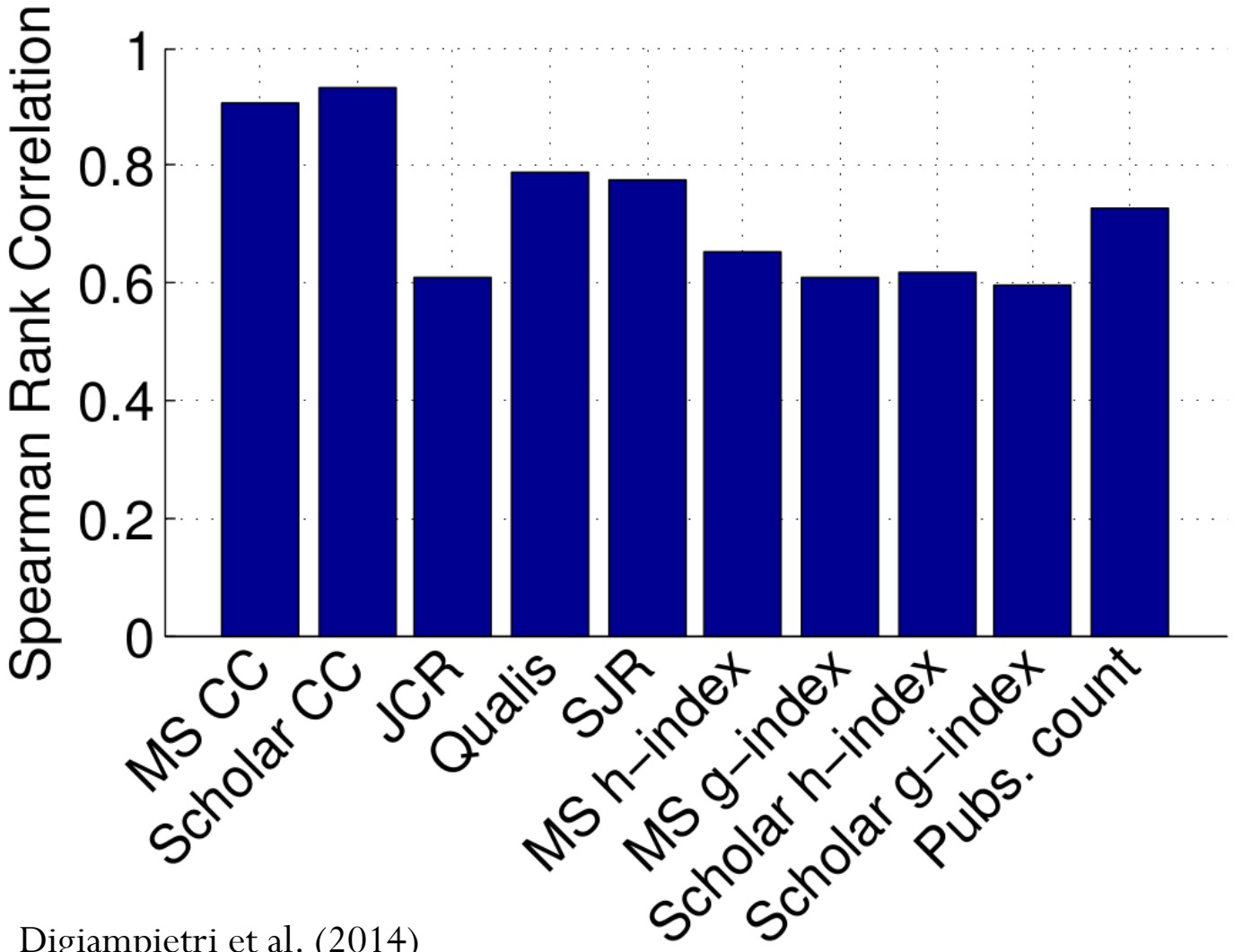


Digiampietri et al. (2014)

Análise dos programas de pós-graduação em CC

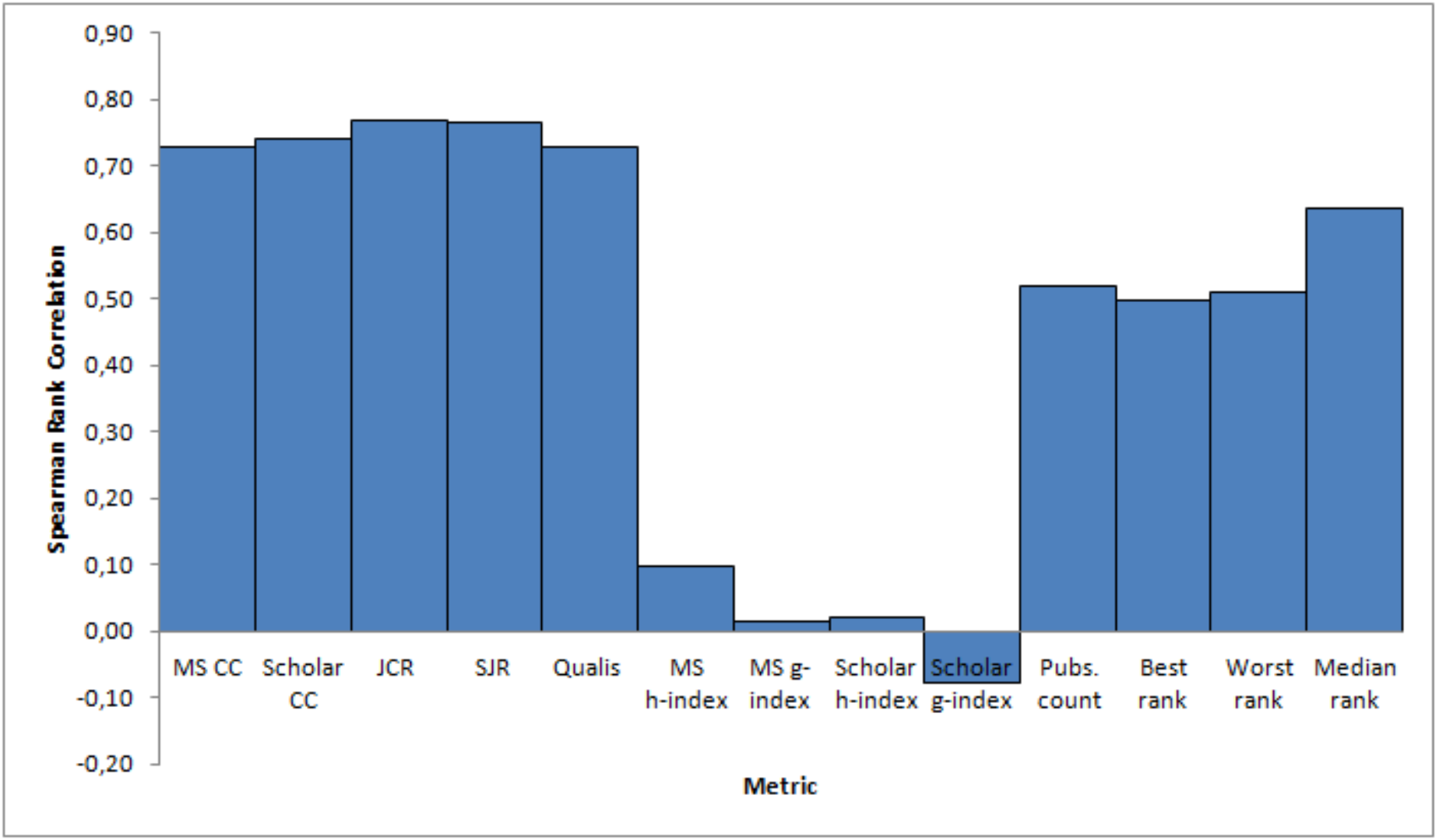


Análise dos programas de pós-graduação em CC – medidas mais correlacionadas entre si

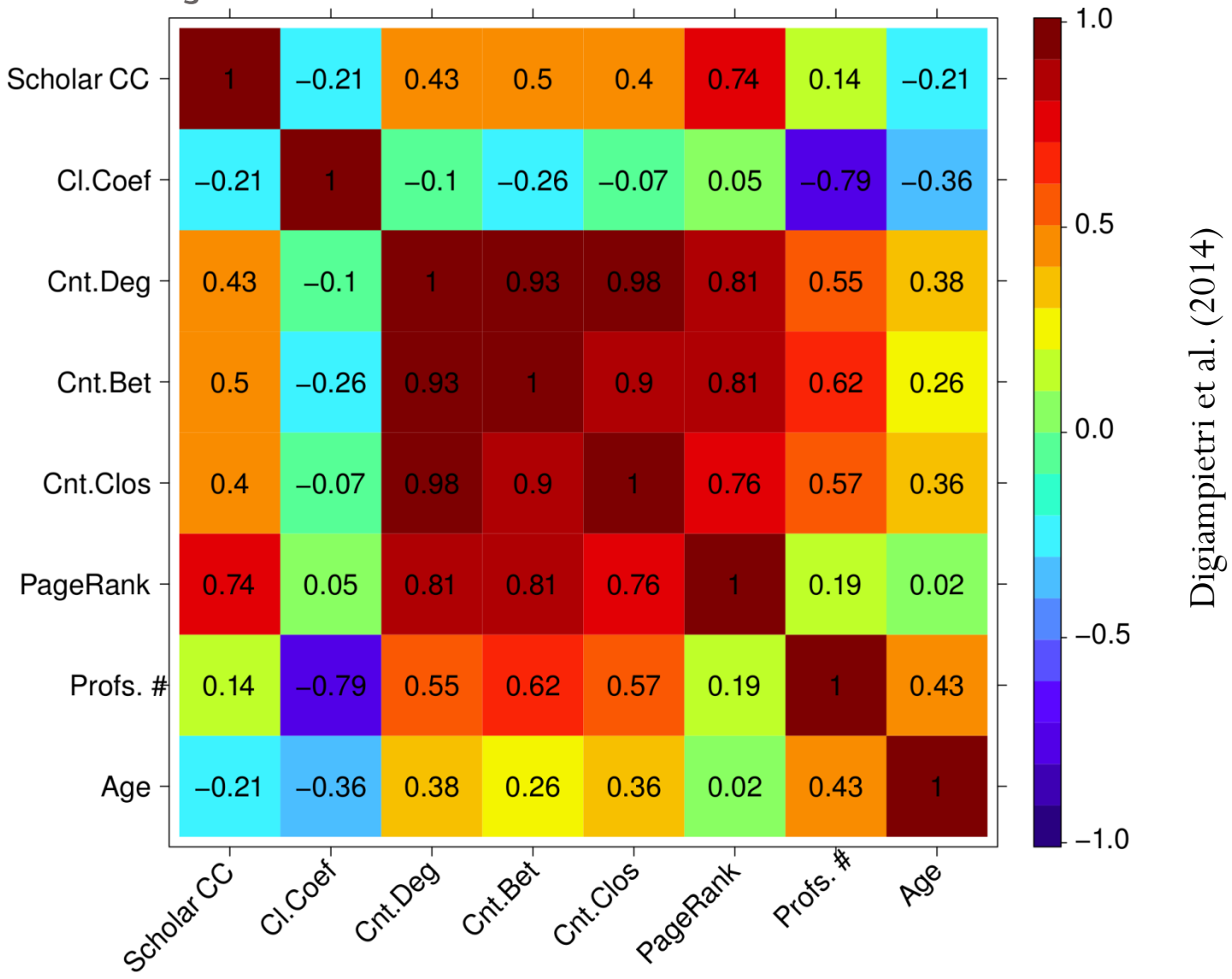


Digiampietri et al. (2014)

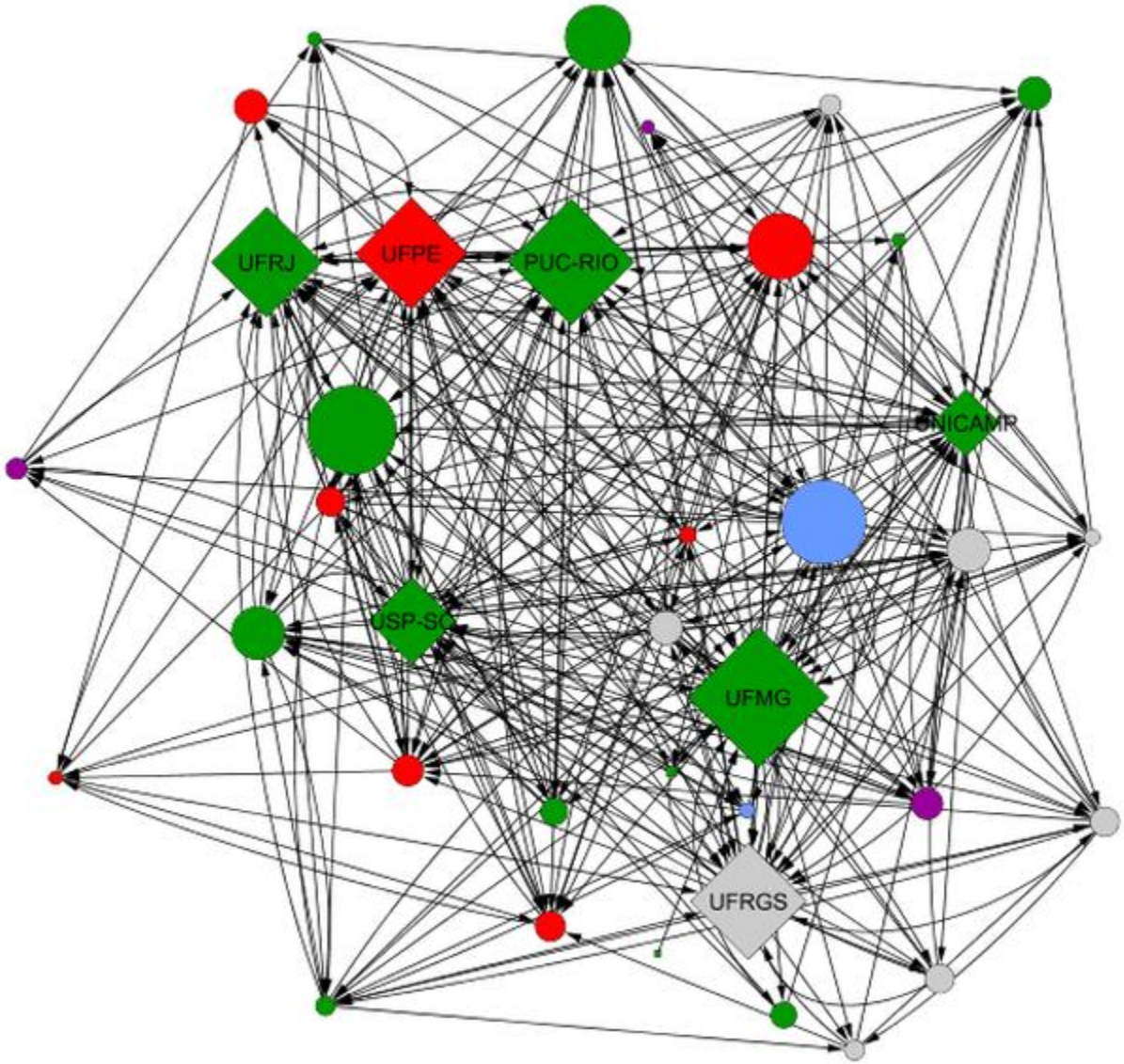
Análise dos programas de pós-graduação em CC – correlação com a nota do programa



Análise dos programas de pós-graduação em CC

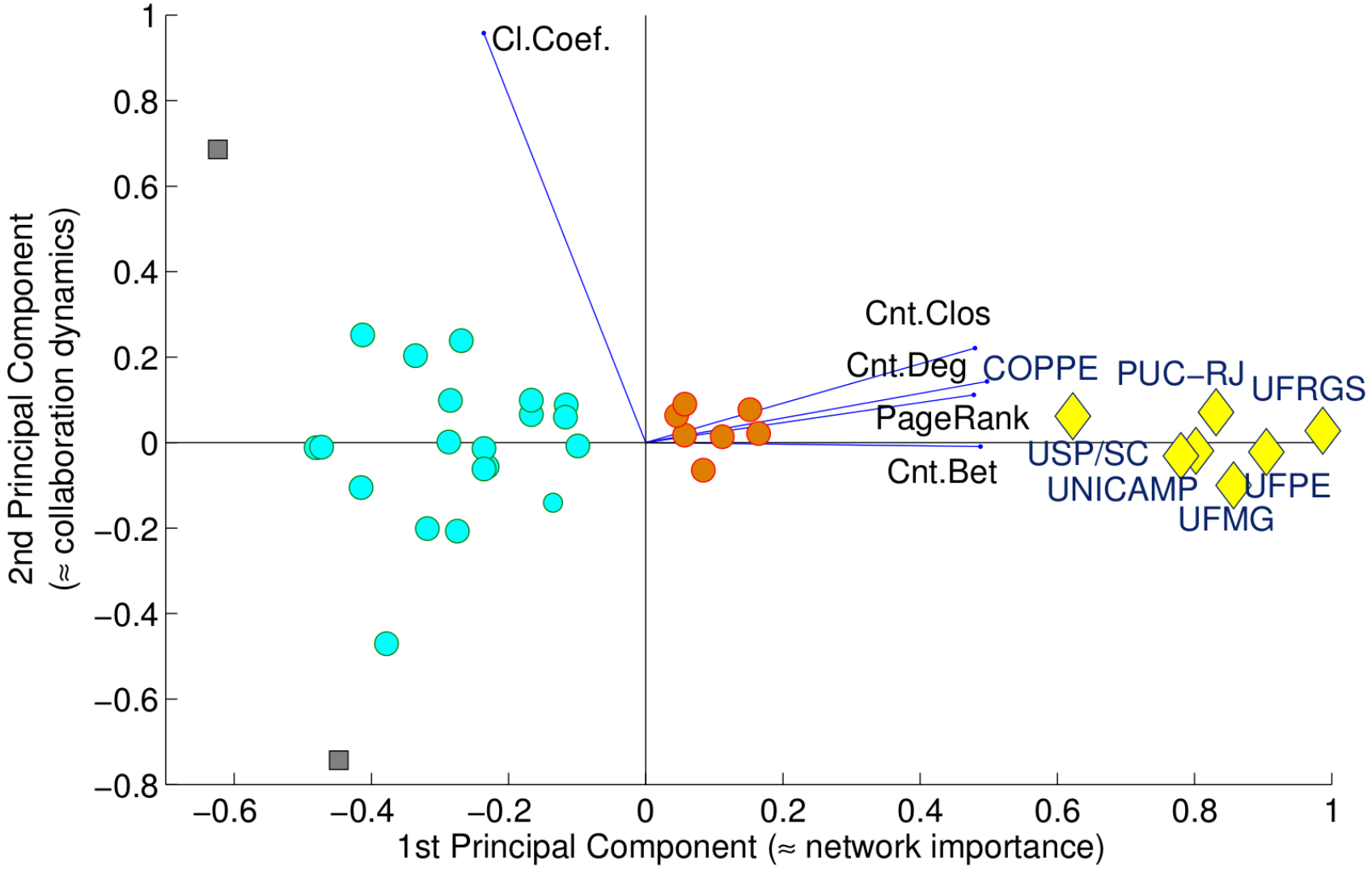


Análise dos programas de pós-graduação em CC



Digiampietri et al. (2014)

Análise dos programas de pós-graduação em CC



Digiampietri et al. (2014)

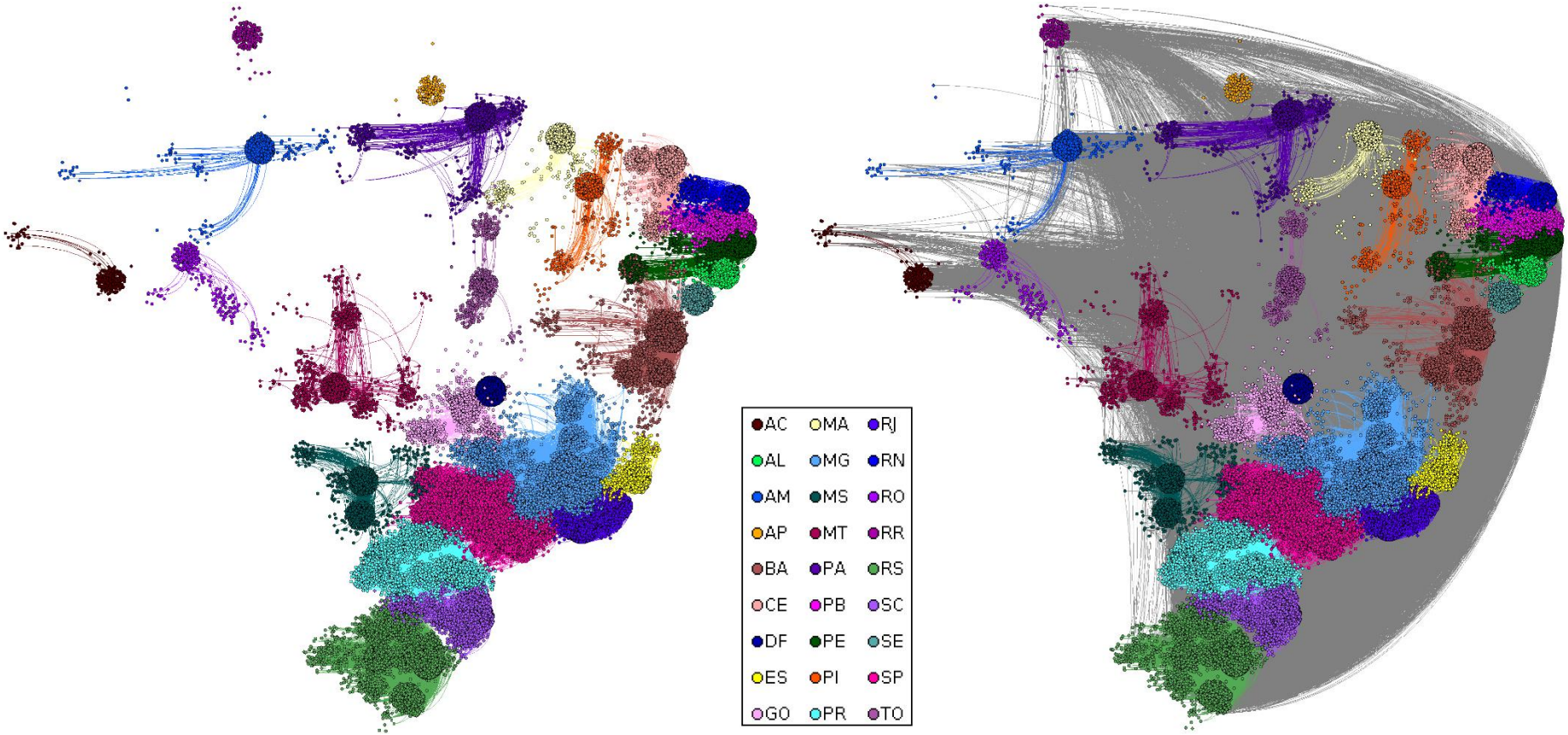
Análise de pesquisadores de acordo com sua localização geográfica

- Estudo dos pesquisadores brasileiros de acordo com seu endereço profissional;
  - Análise das redes estaduais e nacional;
  - Análise das coautorias intra e inter-estados;
  - Análise das áreas de atuação dos pesquisadores;

Análise de pesquisadores de acordo com sua localização geográfica

- A partir dos 3,2 milhões de CVs Lattes cadastrados em 2013 foram identificados **156.278** currículos de **doutores** com endereço profissional no **Brasil**.

Análise de pesquisadores de acordo com sua localização geográfica



Digiampietri et al. (2014a)

	Nós	Arestas	Nós no Componente Gigante	Porcentagem de Nós no Componente	Densidade	Grau Médio	Coeficiente de Clusterização	Assortatividade de Grau	Centralização de Grau	Centralização de Proximidade	Diâmetro	Tamanho da Clique Máxima
AP	**161**	**55**	40	24.84%	**0.00427**	0.683	**0.057**	**-0.308**	**0.424**	0.0393	7	**3**
RR	279	62	**15**	**5.38%**	0.00160	**0.444**	0.287	-0.077	0.128	0.0041	**5**	4
AC	317	110	24	7.57%	0.00220	0.694	**0.342**	0.164	0.071	0.0025	8	4
RO	411	203	92	22.38%	0.00241	0.988	0.320	-0.015	0.090	0.0081	10	5
TO	587	267	116	19.76%	0.00155	0.910	0.323	0.150	0.101	0.0070	11	7
PI	1026	774	345	33.63%	0.00147	1.509	0.212	-0.084	0.115	0.0039	12	6
AL	1109	850	379	34.17%	0.00138	1.533	0.214	-0.053	0.096	0.0033	13	5
MA	1154	807	334	28.94%	0.00121	1.399	0.323	**0.208**	0.043	0.0032	17	6
SE	1193	1165	485	40.65%	0.00164	1.953	0.255	0.117	0.058	0.0032	15	8
AM	1583	1595	706	44.60%	0.00127	2.015	0.195	-0.052	0.043	0.0022	13	7
MT	1718	1182	585	34.05%	0.00080	1.376	0.217	0.062	0.044	0.0021	17	7
MS	1828	2058	808	44.20%	0.00123	2.252	0.188	0.002	0.044	0.0018	16	6
ES	1969	1832	843	42.81%	0.00095	1.861	0.251	0.033	0.040	0.0019	19	7
PA	2576	3399	1405	54.54%	0.00102	2.639	0.194	-0.104	0.054	0.0013	15	8
RN	2627	3413	1372	52.23%	0.00099	2.598	0.213	-0.072	0.036	0.0015	16	8
GO	2987	4071	1458	48.81%	0.00091	2.726	0.232	0.039	0.029	0.0012	20	11
PB	3488	5163	1741	49.91%	0.00085	2.960	0.208	-0.031	0.042	0.0010	17	9
CE	3561	8288	2195	61.64%	0.00131	4.655	0.209	0.076	0.039	0.0011	19	10
PE	4842	10060	3106	64.15%	0.00086	4.155	0.178	0.036	0.028	0.0008	18	8
BA	5357	7291	2570	47.97%	0.00051	2.722	0.182	0.021	0.027	0.0007	19	8
DF	5421	6791	2567	47.35%	0.00046	2.505	0.191	0.026	0.020	0.0007	19	8
SC	5578	9751	3376	60.52%	0.00063	3.496	0.181	0.030	0.022	0.0006	**22**	8
PR	10307	21472	6577	63.81%	0.00040	4.166	0.185	0.013	0.014	0.0003	16	11
RS	13012	39254	9621	73.94%	0.00046	6.034	0.167	0.080	0.019	0.0002	16	13
MG	15234	45208	10137	66.54%	0.00039	5.935	0.158	0.067	0.021	0.0002	21	11
RJ	20639	50368	13638	66.08%	0.00024	4.881	0.146	0.081	0.012	0.0002	18	**18**
SP	**47314**	**179092**	**35836**	**75.74%**	**0.00016**	**7.570**	0.130	0.057	**0.007**	**0.0001**	18	14
Brasil	**156278**	**641825**	**118678**	**75.94%**	**0.00005**	**8.214**	**0.116**	**0.054**	**0.004**	**0.000**	**18**	**24**

Digiampietri et al. (2014a)

Análise de pesquisadores de acordo com sua localização geográfica

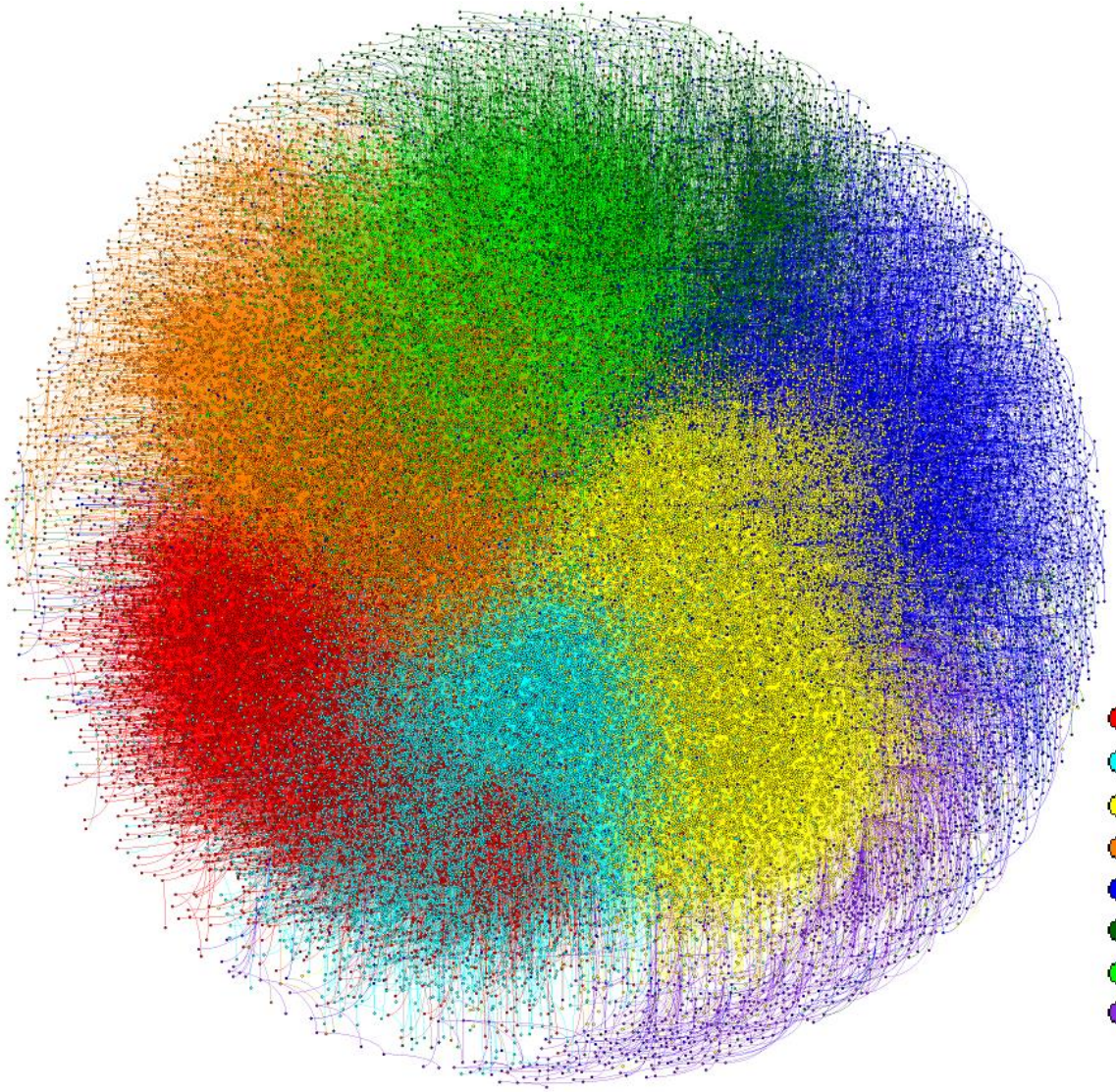
	AC	AL	AM	AP	BA	CE	DF	ES	GO	MA	MG	MS	MT	PA	PB	PE	PI	PR	RJ	RN	RO	RR	RS	SC	SE	SP	TO
AC	9%	1%	2%	0%	3%	3%	4%	1%	1%	0%	18%	1%	2%	2%	2%	1%	1%	5%	6%	2%	1%	0%	4%	2%	1%	28%	0%
AL	0%	17%	1%	0%	3%	4%	2%	1%	1%	1%	7%	0%	1%	1%	7%	11%	0%	3%	9%	2%	0%	0%	4%	2%	3%	22%	0%
AM	0%	0%	22%	0%	2%	2%	3%	1%	1%	1%	8%	1%	2%	4%	2%	2%	0%	5%	8%	1%	1%	0%	3%	2%	1%	27%	0%
AP	0%	1%	1%	10%	1%	1%	2%	1%	2%	1%	9%	1%	0%	14%	3%	3%	0%	7%	7%	2%	1%	1%	3%	1%	1%	28%	0%
BA	0%	1%	0%	0%	29%	2%	3%	1%	1%	0%	11%	1%	1%	1%	2%	4%	1%	3%	8%	2%	0%	0%	4%	2%	1%	23%	0%
CE	0%	1%	1%	0%	2%	38%	2%	0%	1%	1%	5%	0%	1%	2%	4%	4%	2%	2%	6%	4%	0%	0%	3%	1%	1%	17%	0%
DF	0%	0%	1%	0%	3%	2%	27%	1%	5%	0%	11%	1%	1%	1%	1%	2%	0%	4%	7%	1%	0%	0%	5%	2%	1%	23%	0%
ES	0%	0%	1%	0%	2%	1%	2%	20%	1%	0%	23%	1%	1%	1%	1%	2%	0%	3%	15%	1%	0%	0%	3%	1%	0%	21%	0%
GO	0%	0%	0%	0%	1%	1%	8%	1%	25%	0%	15%	1%	1%	1%	1%	1%	0%	4%	5%	1%	0%	0%	3%	1%	0%	27%	1%
MA	0%	1%	1%	0%	3%	6%	2%	1%	1%	17%	8%	1%	0%	3%	5%	3%	2%	4%	6%	3%	0%	0%	2%	1%	1%	30%	0%
MG	0%	0%	1%	0%	3%	1%	3%	2%	2%	0%	44%	1%	1%	1%	1%	1%	0%	3%	7%	1%	0%	0%	3%	2%	1%	20%	1%
MS	0%	0%	1%	0%	2%	1%	3%	1%	2%	0%	10%	20%	2%	1%	0%	1%	0%	8%	5%	1%	0%	0%	6%	2%	0%	33%	0%
MT	0%	0%	1%	0%	2%	2%	3%	1%	2%	0%	16%	2%	14%	1%	1%	1%	0%	6%	6%	1%	0%	0%	7%	2%	1%	30%	0%
PA	0%	1%	2%	1%	2%	3%	3%	1%	1%	1%	8%	1%	1%	29%	2%	2%	1%	3%	8%	1%	0%	0%	4%	2%	1%	23%	0%
PB	0%	2%	1%	0%	3%	5%	2%	1%	1%	1%	6%	0%	1%	1%	29%	12%	2%	2%	4%	7%	0%	0%	3%	1%	2%	14%	0%
PE	0%	2%	1%	0%	4%	3%	2%	1%	1%	0%	5%	0%	0%	1%	8%	37%	1%	2%	6%	3%	0%	0%	3%	1%	2%	16%	0%
PI	0%	0%	1%	0%	3%	11%	2%	0%	1%	2%	10%	1%	1%	1%	6%	7%	16%	2%	4%	3%	0%	0%	2%	1%	1%	24%	1%
PR	0%	0%	1%	0%	1%	1%	2%	0%	1%	0%	5%	1%	1%	1%	1%	1%	0%	37%	4%	1%	0%	0%	6%	6%	1%	28%	0%
RJ	0%	0%	1%	0%	2%	1%	2%	1%	1%	0%	8%	1%	1%	1%	1%	2%	0%	3%	53%	1%	0%	0%	4%	2%	0%	15%	0%
RN	0%	1%	1%	0%	3%	6%	2%	0%	1%	1%	6%	0%	0%	1%	8%	6%	1%	3%	6%	25%	0%	0%	4%	2%	1%	21%	0%
RO	1%	0%	2%	0%	2%	3%	5%	1%	2%	0%	13%	1%	2%	1%	2%	1%	0%	6%	8%	1%	10%	0%	7%	2%	0%	28%	0%
RR	0%	0%	3%	0%	1%	2%	3%	1%	1%	1%	23%	1%	1%	2%	5%	3%	1%	4%	9%	3%	1%	6%	6%	2%	1%	20%	0%
RS	0%	0%	0%	0%	1%	1%	2%	0%	1%	0%	4%	1%	1%	1%	1%	1%	0%	5%	5%	1%	0%	0%	53%	8%	1%	14%	0%
SC	0%	0%	1%	0%	1%	1%	2%	0%	1%	0%	5%	1%	1%	1%	1%	1%	0%	11%	6%	1%	0%	0%	18%	31%	1%	17%	0%
SE	0%	2%	1%	0%	5%	3%	2%	1%	1%	1%	10%	1%	1%	1%	4%	6%	1%	4%	6%	2%	0%	0%	5%	2%	16%	26%	0%
SP	0%	0%	1%	0%	2%	1%	2%	1%	1%	0%	7%	1%	1%	1%	1%	1%	0%	5%	5%	1%	0%	0%	3%	2%	1%	61%	0%
TO	0%	0%	1%	0%	2%	3%	3%	1%	4%	0%	20%	1%	1%	1%	3%	2%	1%	6%	5%	1%	0%	0%	6%	3%	1%	24%	10%

Digiampietri et al. (2014a)

	Porcentagem do Total	Ciências Agrárias	Ciências Biológicas	Ciências da Saúde	Ciências Exatas e da Terra	Ciências Humanas	Ciências Sociais Aplicadas	Engenharias	Linguística, Letras e Artes
AC	0,22%	**0,56%**	0,24%	0,12%	0,14%	0,32%	0,10%	0,04%	0,34%
AL	0,72%	0,83%	0,53%	0,60%	0,95%	0,70%	0,78%	0,47%	**1,03%**
AM	1,03%	1,17%	**2,13%**	0,80%	1,21%	0,84%	0,57%	0,73%	0,46%
AP	0,08%	0,11%	0,09%	0,06%	0,11%	**0,12%**	0,09%	0,01%	0,05%
BA	3,49%	3,63%	3,28%	3,33%	3,61%	4,09%	2,82%	2,12%	**5,75%**
CE	2,29%	**2,76%**	1,43%	2,44%	2,45%	2,61%	1,88%	2,38%	2,01%
DF	3,23%	2,87%	3,09%	2,21%	2,83%	4,03%	**5,84%**	2,61%	3,13%
ES	1,25%	1,50%	1,15%	0,95%	1,27%	1,24%	1,34%	**1,69%**	1,19%
GO	2,00%	**3,50%**	1,82%	1,55%	1,85%	2,70%	1,09%	1,35%	2,20%
MA	0,75%	**1,04%**	0,57%	0,74%	0,80%	1,00%	0,56%	0,52%	0,46%
MG	9,88%	**13,74%**	9,67%	8,30%	9,44%	8,67%	9,61%	11,56%	11,01%

Digiampietri et al. (2014a)

Rede de coautorias – doutores Brasil



- Ciências Agrárias
- Ciências Biológicas
- Ciências da Saúde
- Ciências Exatas e da Terra
- Ciências Humanas
- Ciências Sociais Aplicadas
- Engenharias
- Linguística, Letras e Artes

Digiampietri et al. (2014a)

Predição de Relacionamentos

- **Objetivo**: tratar problemas de predição de relacionamentos em redes sociais acadêmicas.
- **Estratégia**: tratar o problema de predição de relacionamentos como um **problema de classificação binária**.
  - Extração/cálculo de atributos específicos do domínio e estruturais da rede de coautorias.
- Tratar os problemas de predição de **novas coautorias** e geral de predição (**novas e reincidentes**).

Predição de Relacionamentos

- Metodologia
  - revisão da literatura correlata;
  - seleção da amostra;
  - cálculo dos atributos;
  - **filtragem horizontal dos dados;**
  - **balanceamento do conjunto de treinamento;**
  - especificação e desenvolvimento do sistema;
  - execução dos experimentos;
  - **análise dos resultados**.

Predição de Relacionamentos

- Atributos específicos, relacionados a:
  - publicação de artigos completos;
  - relações de orientação;
  - *vizinhos em comum*;
  - atuação nos mesmos programas de pós;
  - áreas de atuação/interesse;
  - distância geográfica (do endereço profissional).

- Atributos estruturais:
  - 12 atributos oriundos da literatura correlata.

Predição de Relacionamentos

- Resultados:
  - **Treinamento** utilizando dados de **1971 a 2010**
  - **Teste**: predição dos relacionamentos ocorridos de **2011 a 2015**

Predição de Relacionamentos – novos relacionamentos

	Taxa de verdadeiro-positivos	Taxa de falso-positivos	Precisão
Subáreas em comum	0,461	0,358	0,966
Vizinhos em comum	0,486	0,307	0,969
PA	0,742	0,674	0,964
Distância	0,790	0,831	0,961
CN	0,807	0,515	0,969
Katz 0,05	0,823	0,819	0,962
Específicos	0,973	0,963	0,963
Solução Proposta	**0,976**	**0,959**	**0,963**

	Revocação	F-Measure	Área ROC
Subáreas em comum	0,461	0,614	0,552
Vizinhos em comum	0,486	0,636	0,643
PA	0,742	0,835	0,514
Distância	0,790	0,866	0,540
CN	0,807	0,877	0,646
Katz 0,05	0,823	0,886	0,430
Específicos	0,973	0,968	0,616
Solução Proposta	**0,976**	**0,969**	**0,646**

Digiampietri et al. (2015)

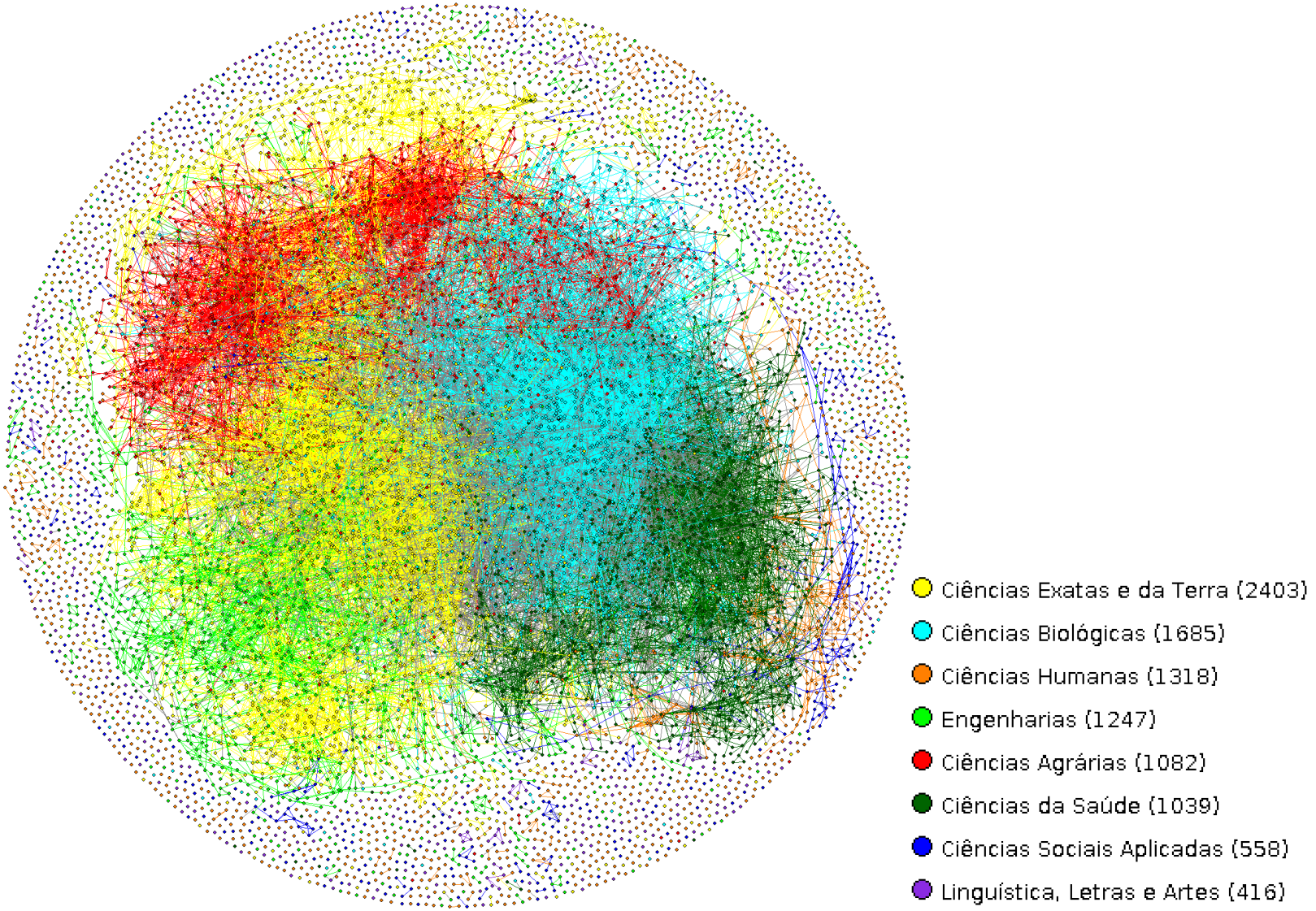
Resultados Adicionais

- Identificação de áreas de atuação
- Análise de tendências
- Relação orientador-orientado

Identificação de áreas de atuação

- **Objetivo**: dado um pesquisador em uma rede social acadêmica identificar sua **área de atuação**
- **Estratégia**: combinar **análise da rede social** com **mineração de textos** aplicada sobre os títulos das publicações
- **Amostra**: **bolsistas produtividade em** 2012 que declararam atuar em apenas uma grande área, área, ou subárea

Identificação de áreas de atuação



Miyata, Kano e Digiampietri (2013a)

Identificação de áreas de atuação

Grandes Áreas

	6 anos	7 anos	8 anos	9 anos	10 anos
MT	84,00%	84,82%	85,33%	86,67%	86,05%
V1	69,23%	70,67%	71,69%	72,31%	73,03%
V2	69,64%	70,67%	71,18%	71,59%	71,90%
MT+V2	89,03%	89,33%	89,95%	**90,56%**	90,26%

Áreas

	6 anos	7 anos	8 anos	9 anos	10 anos
MT	64,11%	65,62%	66,99%	67,26%	68,22%
V1	58,77%	61,10%	61,51%	61,51%	61,64%
V2	61,10%	62,88%	63,42%	63,84%	63,70%
MT+V2	80,27%	81,92%	81,78%	83,29%	**84,11%**

Subáreas

	6 anos	7 anos	8 anos	9 anos	10 anos
MT	23,03%	23,62%	26,53%	26,53%	26,24%
V1	46,94%	48,69%	48,10%	48,10%	47,81%
V2	52,19%	53,64%	54,81%	55,98%	56,27%
MT+V2	55,39%	58,02%	59,48%	**59,77%**	**59,77%**

Miyata, Kano e Digiampietri (2013a)

Análise de tendências

- **Objetivo**: analisar a tendência nas **publicações científicas**

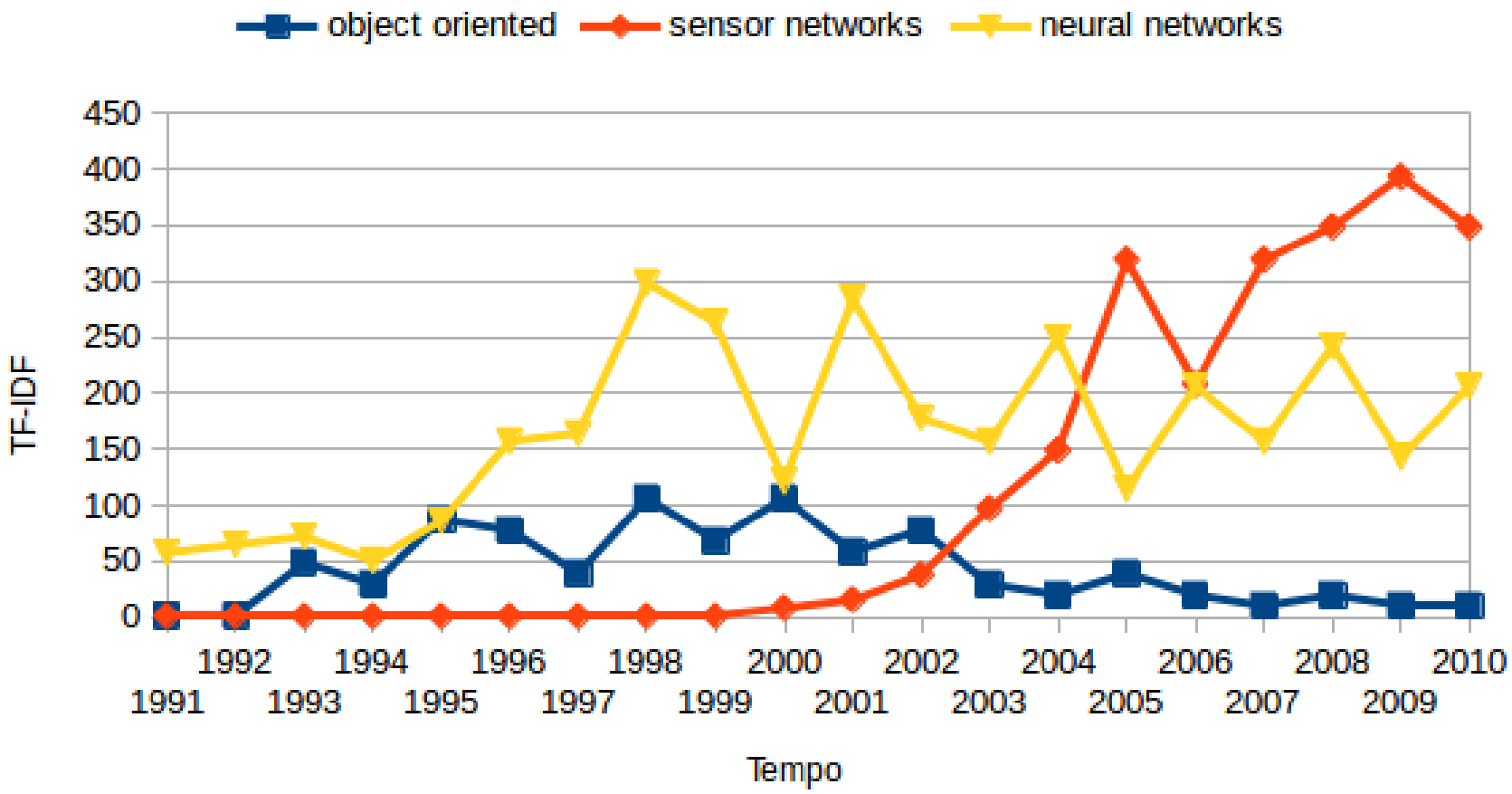
- **Estratégia**: combinar análises históricas baseadas em **documentos** (títulos) com **análise de redes sociais**

- **Amostra**:
  - Pesquisadores dos programas brasileiros de **pós-graduação** em Ciência da Computação
  - **Doutores** que atuam em CC no Brasil

Análise de tendências

- **Solução proposta**: prever o valor de TF-IDF para os termos selecionados, combinando-se:
  - a análise da série temporal do índice;
  - medidas das fontes geradores de informação (rede social);

Análise de tendências



Trucolo e Digiampietri (2014a)

Análise de tendências – programas de pós-graduação em CC

	Programa	Primeira tendência	Segunda tendência
1	PUC-RIO - Informática	product line	microscopy images
2	UFMG - Ciências da Computação	genetic programming	name disambiguation
3	UFRJ - Engenharia de Sistemas e Computação	hyperbolic smoothing	clustering method
4	UFPE - Ciências da Computação	software development	time series
5	UFRGS - Computação	sensor networks	eye fundus images
6	UNICAMP - Ciência da Computação	optimum-path forest	foresting transform
7	USP / SC - Ciências da Computação e Matemática Computacional	neural networks	time series
8	UFF - Computação	wave propagation	cellular automata
9	USP - Ciências da Computação	oriented relational	field-research oriented relational database
10	PUC / PR - Informática	arq scheme	music genre
11	PUC / RS - Ciência da Computação	promising protein receptor snapshots	infocomp ufla

Trucolo e Digiampietri (2014a)

Análise de tendências – doutores em CC atuando no Brasil - 2012

Termo	Obtido	Séries-paramétrico	Erro	Séries-não paramétrico	Erro	Modelo proposto	Erro
service discovery	135,17	441,52	306,35	58,43	76,74	123,39	11,77
based approach	155,19	424,16	268,97	249,40	94,21	161,10	5,91
information systems	147,32	334,29	186,97	182,08	34,76	148,37	1,05
supply chain	174,31	298,37	124,06	145,71	28,6	143,96	30,35
web services	225,28	297,74	72,46	190,14	35,14	201,05	24,23
product line	174,99	291,57	116,57	481,68	306,69	154,73	20,26
motion estimation	107,78	274,36	166,58	174,73	66,95	99,00	8,78
social network	249,05	269,42	20,38	327,70	78,65	198,94	50,11
business process	131,75	240,09	108,34	264,25	132,5	119,61	12,14
time series	150,79	217,76	66,97	196,08	45,29	147,03	3,76
neural network	213,36	178,86	34,51	565,81	352,45	198,85	14,51
sign language	108,21	176,83	68,62	76,97	31,24	101,69	6,52
são paulo	191,93	172,84	19,09	71,51	120,42	145,79	46,15
genetic programming	128,25	156,64	28,39	104,18	24,07	107,98	20,26
routing problem	101,11	147,16	46,05	195,61	94,5	83,75	17,36

Trucolo (2015)

Análise de tendências – doutores em CC atuando no Brasil

Posição	2012	2015		2020	
1	web service	neural network	2	neural network	0
2	social network	social network	0	social network	0
3	neural network	web service	-2	web service	0
4	based approach	information system	2	based approach	1
5	product line	based approach	-1	product line	2
6	information system	business process	4	business process	0
7	time series	product line	-2	supply chain	1
8	são paulo	supply chain	1	information system	-3
9	supply chain	service discovery	1	service discovery	0
10	service discovery	genetic programming	2	são paulo	1
11	business process	são paulo	-2	sign language	3
12	genetic programming	routing problem	3	routing problem	0
13	sign language	motion estimation	1	genetic programming	-3
14	motion estimation	sign language	-1	time series	1
15	routing problem	time series	-8	motion estimation	-2

Trucolo (2015)

Relação orientador-orientado

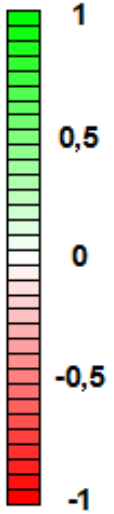
- **Objetivo**: analisar a **participação dos orientandos** nas publicações dos orientadores e a correlação entre algumas medidas.

- **Estratégia**: análise das informações sobre **publicações** e **orientações** dos currículos dos orientadores e da rede social acadêmica.

- **Amostra**: pesquisadores dos programas brasileiros de **pós-graduação em Ciência da Computação**.

Relação orientador-orientado

	total de publicações do orientador	artigos completos publicados em periódicos	trabalhos completos publicados em anais de congressos	resumos expandidos publicados em anais de congressos	resumos publicados em anais de congressos	livros publicados organizados ou edições	capítulos de livros publicados
total de supervisões	0,54	0,3	0,51	0,34	0,3	0,29	0,38
supervisão de pós-doutorado	0,33	0,41	0,28	0,08	0,09	0,34	0,18
tese de doutorado	0,67	0,69	0,61	0,13	0,19	0,48	0,45
dissertação de mestrado	0,64	0,44	0,68	0,13	0,07	0,44	0,42
iniciação científica	0,31	0,15	0,23	0,32	0,34	0,15	0,18
trabalho de conclusão de curso	0,14	-0,04	0,13	0,23	0,14	0,01	0,13
orientação de outra natureza	0,06	-0,02	0,02	0,14	0,18	-0,03	0,06
monografia de conclusão de curso de aperfeiçoamento ou especialização	-0,02	-0,07	-0,01	0,01	0,02	-0,04	0,01



Digiampietri, Mugnaini e Alves (2013)

Relação orientador-orientado

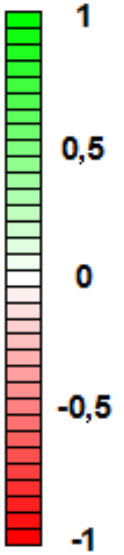
	artigos completos publicados em periódicos	trabalhos completos publicados em anais de congressos	resumos expandidos publicados em anais de congressos	resumos publicados em anais de congressos	livros publicados organizados ou edições	capítulos de livros publicados
supervisão de pós-doutorado	25,58%	57,76%	5,55%	5,47%	0,97%	4,67%
tese de doutorado	20,44%	63,63%	4,45%	5,86%	0,63%	4,99%
dissertação de mestrado	13,05%	71,29%	5,24%	6,80%	0,39%	3,22%
iniciação científica	8,71%	53,55%	8,80%	26,81%	0,12%	2,02%
trabalho de conclusão de curso	9,66%	64,37%	8,36%	14,09%	0,19%	3,33%
orientação de outra natureza	9,65%	53,42%	8,39%	25,09%	0,63%	2,82%
monografia de conclusão de curso de aperfeiçoamento ou especialização	18,38%	58,77%	7,80%	10,58%	0,28%	4,18%
média do orientador	**16,26%**	**58,99%**	**4,62%**	**9,74%**	**3,93%**	**6,47%**

Digiampietri, Mugnaini e Alves (2013)

Relação orientador-orientado

	total de artigos orientador	artigos completos publicados em periódicos	trabalhos completos publicados em anais de congressos	resumos expandidos publicados em anais de congressos	resumos publicados em anais de congressos	livros publicados organizados ou edições	capítulos de livros publicados	primeira autoria
Author Rank	0,82	0,6	0,8	0,31	0,24	0,52	0,58	-0,44
grau	0,7	0,7	0,66	0,35	0,36	0,22	0,19	-0,39

	total de supervisões	supervisão de pós-doutorado	tese de doutorado	dissertação de mestrado	iniciação científica	trabalho de conclusão de curso	orientação de outra natureza	monografia de conclusão de curso de aperfeiçoamento ou especialização
Author Rank	0,5	0,22	0,63	0,68	0,22	0,11	0,04	-0,02
grau	0,43	0,2	0,59	0,63	0,19	0,08	0,01	-0,04



Digiampietri, Mugnaini e Alves (2013)

ACH2197 - Análise de Redes Sociais

Análise da Rede Social Acadêmica Brasileira:

Um Estudo de Caso

Luciano Antonio Digiampietri