

PESQUISA EM REDE: ANÁLISE PRELIMINAR DOS GRUPOS DE PESQUISA DO CNPq EM 2014

Bernardo Alves Furtado¹

1 INTRODUÇÃO

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) disponibiliza, em seu *site*, os resultados dos censos do Diretório dos Grupos de Pesquisa no Brasil.² Os censos contêm informações referentes aos 35.424 grupos em atividade, em 30 de dezembro de 2014 e anos anteriores, sendo que cada um deles possui informações básicas do próprio grupo, de todos os pesquisadores e estudantes vinculados e, quando existente, de empresas vinculadas ao grupo.

Uma busca preliminar na literatura que analisa grupos de pesquisa do CNPq revelou poucos trabalhos, todos com recortes de análise restritos a áreas específicas (Villela e Almeida, 2012; Ferraz *et al.*, 2014; Martins *et al.*, 2014; Rocha *et al.*, 2014; Santos *et al.*, 2014). Não é do conhecimento do autor trabalho que tenha feito análise mais sistemática do conjunto completo dos grupos de pesquisa, muito menos utilizando os dados de 2014.

Nesse sentido, este artigo apresenta uma primeira leitura dos dados, identificando preliminarmente aspectos de interesse, em especial referentes à associação de grupos de pesquisa com empresas.

Além desta introdução, este artigo possui outras quatro seções. A seção 2 apresenta um panorama inicial, com uma leitura geral dos dados disponíveis na página de grupos do CNPq. A seção 3 discute a metodologia operada para a análise dos dados, consideradas as redes propostas. A seção 4 comenta os resultados da análise de redes. Por fim, a seção 5 tece algumas considerações finais e possíveis desdobramentos da proposta a partir dos dados da base disponibilizada.

2 PANORAMA INICIAL

O próprio CNPq disponibiliza algumas informações gerais. Será apresentado um resumo desses grandes números no intuito de situar o leitor.

Os grupos de pesquisa concentram-se no Sudeste (44%), seguido do Sul e do Nordeste, com aproximadamente 20% em cada região. As ciências humanas concentram 21% dos grupos e são seguidas pelas ciências da saúde, sociais aplicadas e engenharias e computação, com 16%, 14% e 13%, respectivamente. São várias as áreas do conhecimento representadas, sendo que a educação destaca-se, com 9% dos grupos, seguida de medicina, com quase 5%. A maioria dos grupos é recente ou muito recente, com 61% deles criados a partir de 2007.

Há grupos de vários tamanhos, com o grupo típico contando com três a sete pesquisadores. Há maior presença masculina entre os líderes de grupos, enquanto a presença feminina é maior no conjunto de pesquisadores. A faixa etária mais representativa, numericamente, é a de 35 a 39 anos, seguida da faixa imediatamente anterior, 30 a 34 anos.

Os pesquisadores com grau de doutor representam 65% do total. Pesquisadores estrangeiros são 3% do total, sendo que quase 25% dos estrangeiros encontram-se em São Paulo e, predominantemente, nas ciências exatas e da terra. Proporcionalmente, os estrangeiros são mais relevantes em física, área na qual representam quase 10% do total, e na matemática, com 11%.

O número de estudantes de cada grupo varia bastante, com 9% de grupos sem estudante algum, 7% com cinco e mais de 20% dos grupos com mais de quinze estudantes. Os estudantes são, na grande maioria, graduandos (124 mil), seguidos de doutorandos e mestrandos, ambos com cerca de 50 mil.

1. Técnico de planejamento e pesquisa e coordenador na Diretoria de Estudos e Políticas Setoriais de Inovação, Regulação e Infraestrutura (Diset) do Ipea; Bolsista de produtividade do CNPq.

2. Disponível em: <<http://goo.gl/K2QE7>>.

A grande maioria dos grupos não tem suporte técnico algum (83%). Pouco mais de 11% contam com um técnico.

Os grupos mais produtivos por pesquisador doutor em periódicos nacionais são os das ciências agrárias, seguidos pelos grupos das ciências da saúde e das ciências sociais aplicadas. Em periódicos internacionais, a produção é maior nas ciências biológicas, seguidas das ciências da saúde e das ciências exatas e da terra.

Cerca de um quarto dos grupos relataram algum relacionamento com empresas. Sobressaem-se as ciências agrárias, as engenharias e as ciências biológicas, com 37% dos grupos conectados a, pelo menos, uma empresa. Proporcionalmente, os grupos com maior vínculo foram Biologia geral, engenharia aeroespacial e engenharia de materiais e metalúrgica, com cerca de metades dos grupos da área vinculados a empresas. Em números absolutos, destacam-se as áreas de química e agronomia, com mais de quatrocentos grupos.

Há pequena diferenciação entre a representação regional da proporção de grupos vinculados às empresas. Entre os estados, destacam-se Goiás, Santa Catarina e Pernambuco em termos proporcionais, e São Paulo e Rio de Janeiro em termos absolutos. Várias instituições sobressaem-se no vínculo de seus grupos a empresas: Universidade de São Paulo (USP), com 476 empresas; Universidade Estadual Paulista (Unesp), com 356; Fundação Oswaldo Cruz (Fiocruz), com duzentas; Universidade Federal de Pernambuco (UFPE), com 203; Universidade Federal do Rio de Janeiro (UFRJ), com 320; Universidade Federal do Rio Grande do Sul (UFRGS), com 248; e Universidade Federal de Minas Gerais (UFMG), com 217.

Vários grupos (dois terços) possuem parceria sem transferência de recursos. Entretanto, há várias parcerias com transferências de insumos (um terço), fornecimento de bolsas, recursos financeiros ou recursos humanos, entre outras.

3 METODOLOGIA

Os dados foram lidos da base original em XML para o formato Python 3.X., incorporados em classes e, em seguida, foram trabalhadas as estatísticas descritivas.³

As informações são disponibilizadas em dois arquivos XML. O primeiro (<número-de-identificação-do-grupo_2014>.XML) contém todas as informações referentes aos grupos – identificação, código de área, grande área, Unidade da Federação (UF), ano de início, sigla da instituição, latitude, longitude, município, pesquisadores líderes e, quando vinculado a alguma empresa, seu Cadastro Nacional de Pessoa Jurídica (CNPJ), classe da Classificação Nacional de Atividades Econômicas (CNAE) e situação jurídica. Além disso, contém ainda os pesquisadores e os estudantes com informações referentes ao código de identificação no CNPq, nome completo, sexo, identificação do orientador, quando for o caso, titulação máxima ou nível em estudo corrente. Embora esse primeiro arquivo contenha informações referentes ao pesquisador, foi utilizado o segundo arquivo (<número-de-identificação-do-grupo_2014_estendido>.XML), já que este continha ainda a produção dos pesquisadores para o período 2011-2014.

Vale ressaltar que os pesquisadores e alguns estudantes estão listados individualmente em todos os grupos a que pertencem. Assim, o CNPq adverte que a soma dos pesquisadores pode sofrer alterações devido à dupla contagem. Ilustrativamente, note que se somados os pesquisadores de todos os 35.424 grupos, temos um total de 179.023 pesquisadores únicos e 307.912 estudantes, ou seja, sem repetição de seus nomes nos diversos grupos de pesquisas em que participam.⁴

No intuito de verificar se há comportamento distinto entre os grupos do CNPq, procedeu-se à construção de rede entre os grupos, obedecendo aos seguintes critérios: seriam conectados os grupos caso houvesse pesquisador ou estudante que pertencesse a ambos os grupos simultaneamente, ou caso mais de um grupo estivesse conectado a empresas com o mesmo CNPJ.

3. O código de leitura e análise dos dados está disponível e pode ser solicitado ao autor.

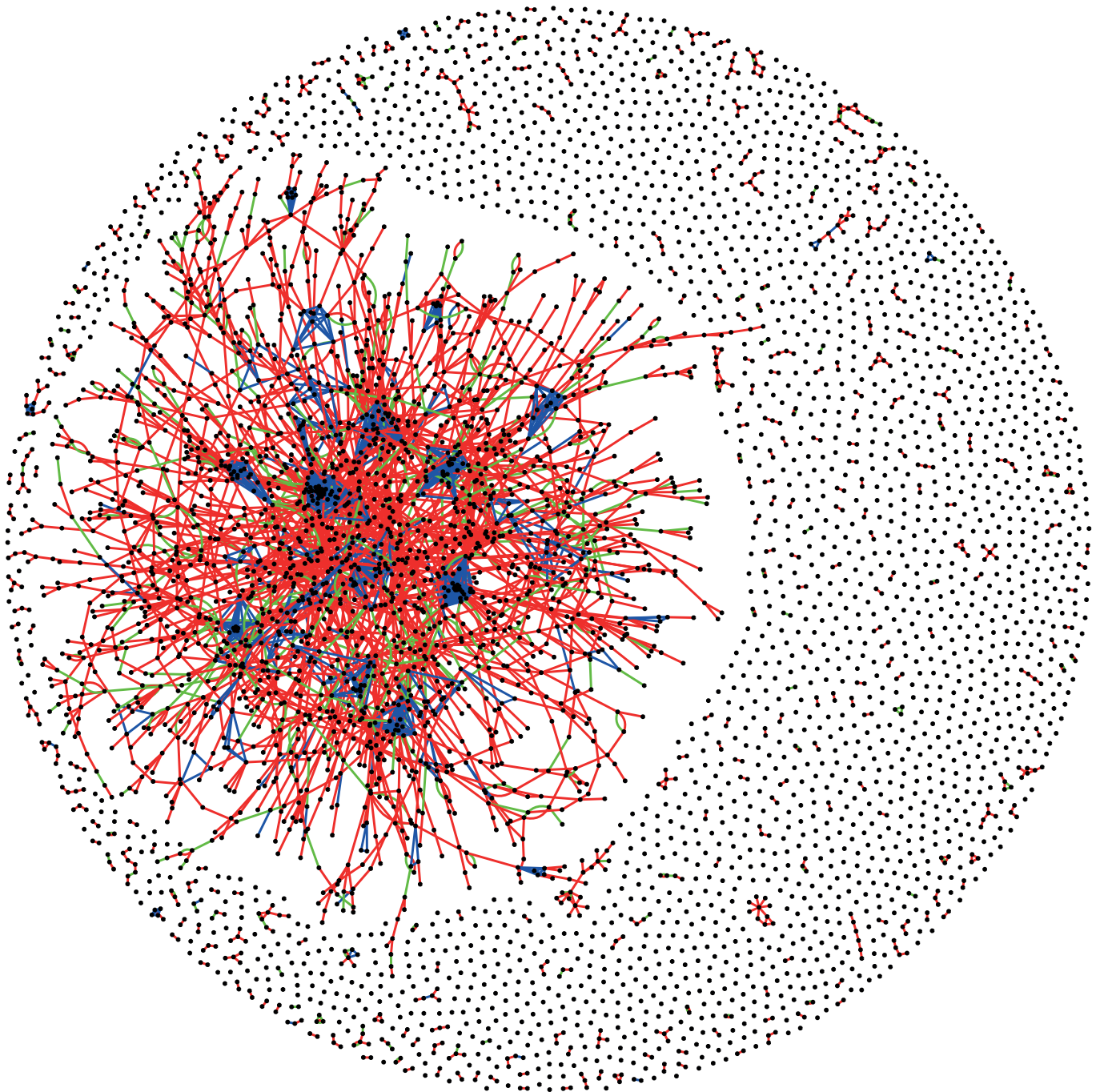
4. Na tabela que apresenta as informações por liderança, sexo e idade, o CNPq afirma que não há dupla contagem e indica o total de 180.262 pesquisadores. Ou seja, há 1.239 pesquisadores (0,007%) que não encontramos a partir da leitura dos dados originais. Ou, então, é possível que possa haver ainda pequeno nível de dupla contagem nos dados tabulados pelo CNPq.

4 RESULTADOS

Como resultado, foram obtidos dois conjuntos de grupos: *i)* grupos sem nenhuma conexão com os outros grupos; e *ii)* grupos com pelo menos uma conexão com outros grupos, seja a conexão por meio de um pesquisador, seja por meio de um estudante, ou seja, ainda, por empresa em comum. Adicionalmente, foram construídos os grupos com conexões apenas por meio de empresas, ou de pesquisadores ou de estudantes em comum. A título ilustrativo, apresenta-se, na figura 1, os grupos com alguma conexão (3.047 grupos), limitando-se aos 5.000 primeiros grupos do conjunto de dados, no intuito de facilitar a visualização. A figura A.1, no anexo A, apresenta a rede com 10.000 grupos.

FIGURA 1

3.047 grupos de pesquisa do CNPq com pesquisadores, estudantes ou empresas em comum



Fonte: Dados disponibilizados pelo CNPq. Disponível em: <<http://goo.gl/K2QEe7>>.

Elaboração do autor.

Obs.: Pesquisadores em vermelho, estudantes em verde e empresas em azul.

O padrão estrutural da figura 1 já fornece algumas sugestões visuais acerca do comportamento típico de rede. Serão apresentados, primeiro, tais indícios para, em seguida, se apresentar os números da rede completa.

As ligações predominantes são entre pesquisadores. Ou seja, se há algum tipo de contaminação ou troca de experiências entre grupos, tal transferência ocorreria principalmente por meio dos pesquisadores. Além disso, há número relevante de estudantes que pertencem a mais de um grupo. Finalmente, as vinculações entre grupos e empresas, por sua vez, demonstram que, quando uma empresa está ligada a algum grupo, ela também está ligada aos grupos vizinhos, configurando um clique, ou seja, um subgráfico completo no qual todos os nós estão ligados entre si.

4.1 Rede completa

Neste caso em estudo, a rede apresenta 35.424 grupos (nós), com 285.793 conexões entre si (arestas). São identificados 307.912 estudantes, 179.912 pesquisadores únicos, além de 5.817 empresas distintas. Os grupos contêm, em média, 8,4 pesquisadores, com produção, também média, de 86,5 itens por grupo, em 2014. Estes itens referem-se à produção em periódicos nacionais, internacionais, livros, capítulos, resumos em revista ou em anais.⁵

A densidade da rede, dada como proporção entre as conexões possíveis e as existentes, é baixa (apenas 0,00046). Já o diâmetro da rede – caracterizado como a maior distância entre dois nós conectados – é de quatorze (grupos). Ou seja, a presença de pesquisadores, estudantes ou empresas em comum pode indicar a troca de informações entre até quatorze grupos de distância.

O índice ômega retrata o maior agrupamento no qual todos os nós estão interligados entre si, ou seja, o maior clique, ou subgráfico completo. O ômega para a rede completa é de 329, indicando que 329 grupos de pesquisa estão conectados à mesma empresa. Não surpreendentemente, esta empresa é a própria USP, representando quase 1% dos grupos, seguida da UFRJ (0,45%) e da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) (0,28%), e o próprio CNPq.

Entretanto, a análise dos códigos CNAE/Instituto Brasileiro de Geografia e Estatística (IBGE) dos grupos demonstra que 86% deles não indicaram nenhuma empresa vinculada.

Quanto às áreas, entre aquelas que indicaram alguma empresa, a maior presença é justamente a área de educação superior *P.85.3*, com 2.337 grupos, ou quase a metade dos grupos que indicaram vinculação. A segunda área mais presente entre as empresas foi a de *O.84.1*, indicativa de administração pública do estado e da política econômica e social, com 408 grupos. Ou seja, provavelmente vinculada a empresas, fundações e institutos estatais. A terceira área é justamente a de pesquisa e desenvolvimento científico *M.72.1*. Finalmente, as áreas seguintes são *S.94.3* – atividades de associações de defesa e direitos sociais, com noventa grupos; *J.62.0* – atividades dos serviços de tecnologia da informação, com 72; e *D.35.1* – geração e transmissão e distribuição de energia elétrica, com 44. A tabela 1 apresenta o resumo do número de grupos por subclasses da CNAE. O anexo B apresenta as subclasses e as divisões com mais de dez grupos.

TABELA 1
Número de grupos por subclasses da CNAE 2

Grupos	Divisões	Subclasses CNAE 2.2 – descrição
2.419	P	Educação
495	O	Administração pública, defesa e seguridade social
481	C	Indústrias de transformação
409	M	Atividades profissionais, científicas e técnicas
229	S	Outras atividades de serviços
137	Q	Saúde humana e serviços sociais
127	G	Comércio; reparação de veículos automotores e motocicletas
121	J	Informação e comunicação
102	A	Agricultura, pecuária, produção florestal, pesca e aquicultura
77	D	Eletricidade e gás
47	K	Atividades financeiras, de seguros e serviços relacionados
29	B	Indústrias extrativas
29	R	Artes, cultura, esporte e recreação

(Continua)

5. A base contém ainda dados referentes às produções de 2011, 2012 e 2013, não computados nesta análise. Adicionalmente, há censos para anos anteriores: 2010 e de dois em dois anos, até o ano 2000.

(Continuação)

Grupos	Divisões	Subclasses CNAE 2.2 – descrição
28	E	Água, esgoto, atividades de gestão de resíduos e descontaminação
28	N	Atividades administrativas e serviços complementares
22	F	Construção
21	H	Transporte, armazenagem e correio
3	L	Atividades imobiliárias
1	I	Alojamento e alimentação
1	U	Organismos internacionais e outras instituições extraterritoriais

Fonte: Censo 2014/CNPq. Disponível em: <<http://goo.gl/K2QEe7>>.
Elaboração do autor.

Quanto à natureza jurídica das empresas vinculadas a grupos de pesquisa, tem-se que a grande maioria dos grupos com informação de empresas associadas – 3.416 – está vinculada a autarquias, fundações, órgãos e empresas públicas (tabela 2). Sociedades empresariais, associações privadas e fundações privadas apresentam vínculo com 1.982 grupos de pesquisa.

TABELA 2

Número de grupos de pesquisa quanto à natureza jurídica da empresa associada

Grupos de pesquisa	Natureza jurídica da empresa associada
2.001	Autarquias federal, estadual ou municipal
1.199	Sociedades empresariais limitadas, anônimas, simples
838	Fundação federal, estadual ou municipal
541	Associação privada
359	Órgão público dos poderes federal, estadual ou municipal
242	Fundação privada
218	Empresa pública
129	Sociedade de economia mista
38	Cooperativa

Fonte: Censo 2014/CNPq. Disponível em: <<http://goo.gl/K2QEe7>>.
Elaboração do autor.

O maior grupo da base, com 295 pesquisadores, é denominado Pecus – Dinâmica de Gases de Efeito Estufa em Sistemas de Produção da Agropecuária Brasileira, sediado na Embrapa em São Carlos (São Paulo).

Os municípios com maior número de grupos estão listados no anexo C, sendo encabeçados pelas capitais, na seguinte ordem: São Paulo, Rio de Janeiro, Porto Alegre, Belo Horizonte, Recife, Curitiba e Salvador.

O coeficiente de agrupamento médio, por sua vez, é um indicador que denota a força de conexão entre os vizinhos de cada vértice e verifica se os grupos ligados a determinado grupo estão ligados entre si. Esse indicador para a rede completa é de 0,82.

A despeito dessa caracterização geral da rede, parece interessante verificar – ainda de forma exploratória – se grupos com maior conectividade distinguem-se em relação aos grupos não conectados.

Os grupos que apresentam pelo menos uma conexão têm, em média, 10,8 pesquisadores, ou dois a mais que a média geral. Os grupos sem conexão alguma apresentam, em média, 3,1 pesquisadores.⁶ A produção média total sobe 3 pontos nos grupos com conexões, para 89,6, e contrasta com uma média de apenas 4,8 para os grupos sem conexões. Do mesmo modo, a mediana da produção é 3 pontos maior para os grupos conectados (44) e de apenas 2 para os grupos sem conexão alguma. A produção por pesquisador geral é de 10,04; para os grupos com conexões, 10,35; e apenas 1,68 para os grupos não conectados.

4.2 Rede com conexões apenas entre empresas, pesquisadores ou estudantes

Os grupos formados por conexões restritas a empresas, pesquisadores ou estudantes, por construção, são menos densos. As conexões por empresas representam 17.103 arestas. Pesquisadores são responsáveis por conectar grupos por meio de 143.542 arestas, e os estudantes com 25.148. Nesse sentido, as densidades são, respectivamente, 0,00023, 0,00019 e 0,00004.

6. Embora não se possa afirmar se a diferença é estatisticamente significativa, a ordem de grandeza pareceu relevante.

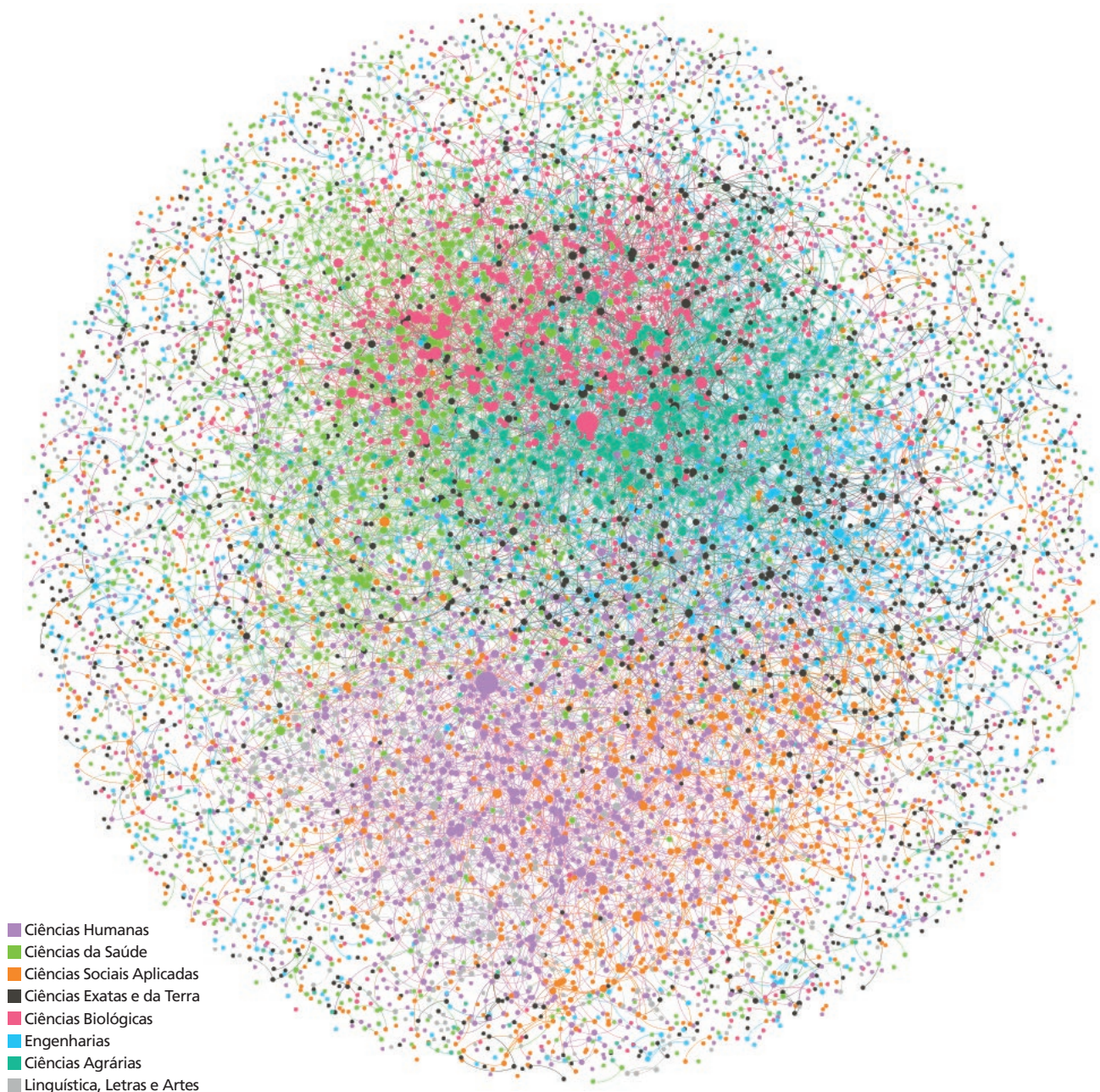
O ômega para empresas é o mesmo valor para a rede completa (329). Em relação à rede restrita aos pesquisadores, o ômega é de 24. Ou seja, há 24 grupos que estão completamente conectados entre si, por meio de pelo menos um de seus pesquisadores. A rede de estudantes tem um ômega de 7.

O diâmetro – ou a maior distância linear entre dois grupos – aumenta na medida em que a rede fica menos densa, passando a quinze grupos de distância para a rede somente de pesquisadores, e 34 para a rede de estudantes. Finalmente, o coeficiente de agrupamento – ou a probabilidade de conexão entre os grupos aos quais um grupo está ligado – diminui para 0,19 no caso dos pesquisadores, e 0,17 para os estudantes.

Finalmente, ainda a título ilustrativo, apresenta-se, na figura 2, a rede completa por pesquisadores. As cores dos grupos (nós) são indicativas das grandes áreas e o seu tamanho é referência ao número de grupos associados (o grau de cada nó). Nota-se certa proximidade entre os grandes grupos de ciências da saúde e biológicas, entremeadas pelas ciências agrárias, exatas e da terra e, menos concentradas, as engenharias. As ciências humanas aproximam-se das ciências sociais aplicadas.

FIGURA 2

Rede de pesquisadores



Fonte: Dados disponibilizados pelo CNPq. Disponível em: <<http://goo.gl/K2QEe7>>.

Elaboração do autor.

Obs.: 1. Grandes áreas representadas por cores; tamanho dos nós indicativo do número de conexões (grau).

2. Figura reproduzida em baixa resolução e cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

5 CONSIDERAÇÕES FINAIS E POSSÍVEIS DESDOBRAMENTOS

Este artigo apresenta brevemente alguns números gerais referentes aos grupos de pesquisa disponibilizados pelo CNPq para o ano de 2014. Para além das estatísticas básicas já disponíveis e compiladas na seção 2, este trabalho indica que a presença de vínculo dos grupos com empresas é fortemente influenciada pela informação dos grupos referentes às próprias instituições de ensino. A despeito disso, há informações sobre 1.824 outros vínculos com empresas que parecem ser, de fato, empresas não ligadas exclusivamente ao ensino superior ou a instituições de pesquisa estatais. A caracterização dessas empresas e dos grupos vinculados a elas pode contribuir para o entendimento mais completo da vinculação “ciência e produção” para o caso brasileiro.

Dado que também há informação detalhada da produção dos pesquisadores nos grupos, é possível realizar análise da produção por grupos, grandes áreas e áreas de pesquisa, além de análises regionais. Por exemplo, a força relativa das ciências agrárias no Sudeste e no Centro-Oeste do Brasil.

Finalmente, este artigo cumpre a tarefa de analisar, de forma exploratória, a base de dados dos grupos de pesquisa do CNPq para o ano de 2014, contribuição, ao ver do autor, ainda não disponível na literatura.⁷ Adicionalmente, conta ainda com a verificação de possíveis investigações, a partir da análise preliminar da base.

REFERÊNCIAS

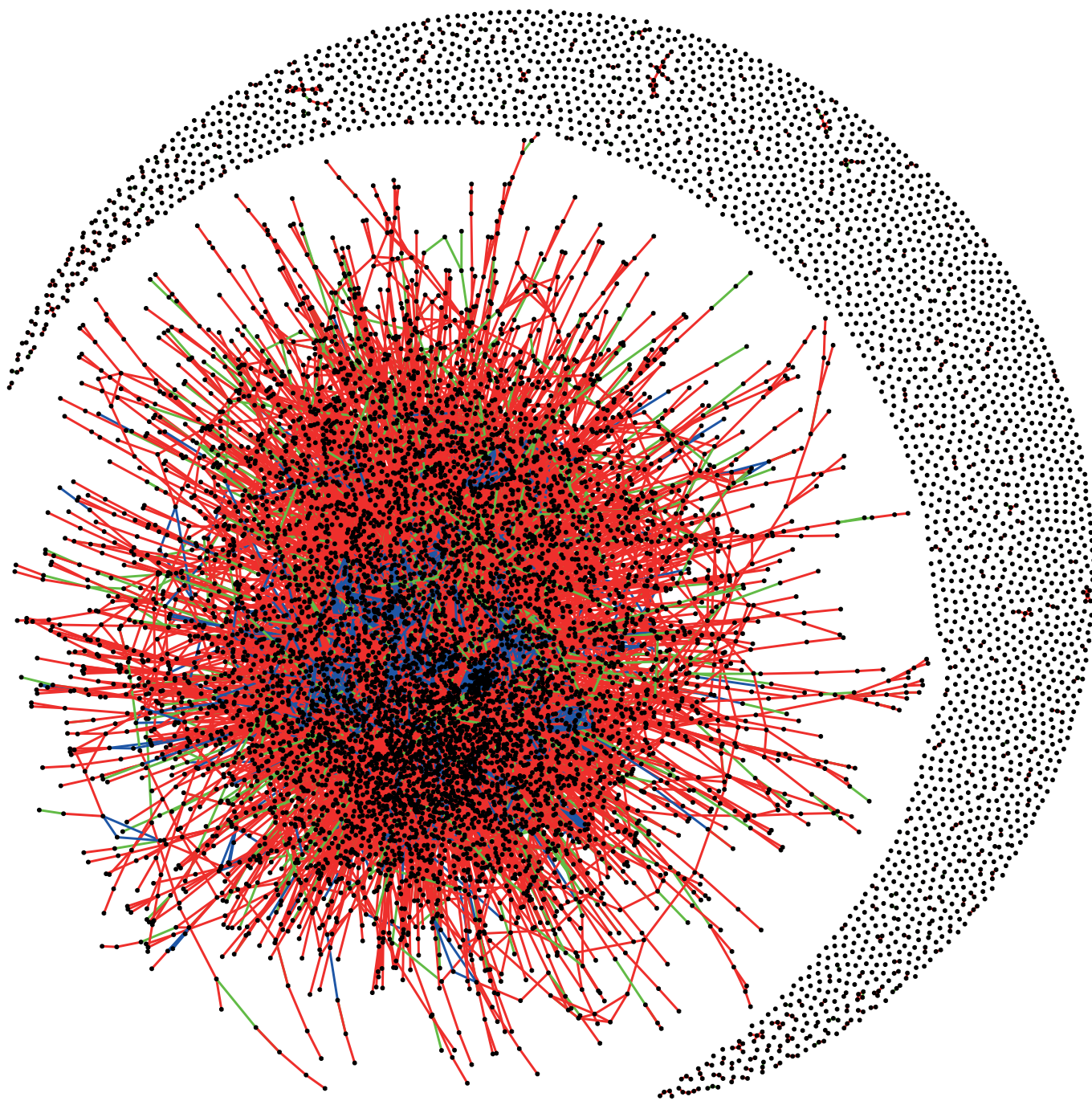
- FERRAZ, R. R. N. *et al.* Análise e gestão de análise de redes de colaboração entre pesquisadores de programas de pós-graduação stricto sensu com a utilização da ferramenta computacional scriptLattes. **Perspectivas em Gestão & Conhecimento**, v. 4, p. 133-147, 2014.
- MARTINS, A. M.; CARVALHO, C. A. S.; ANTUNES-ROCHA, M. I. Pesquisa em representações sociais no Brasil: cartografia dos grupos registrados no CNPq. **Psicologia: teoria e prática**, v. 16, n. 1, p. 104-114, 2014.
- ROCHA, A. M. *et al.* Pesquisas acadêmicas sobre biodiesel no Brasil: uma análise dos grupos do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). **Revista Tecnologia e Sociedade**, v. 10, n. 20, p. 71-95, 2014.
- SANTOS, S. S. C. *et al.* Enfermagem e idoso: grupos de pesquisa cadastrados no diretório do CNPq. **Enfermagem em Foco**, v. 4, n. 2, p. 80-83, 2014. Disponível em: <<http://goo.gl/yjAvvS>>.
- VILLELA, T. C.; ALMEIDA, C. C. R. Relações universidade-empresa no estado de Mato Grosso no período 2002-2008: um estudo comparativo a partir do diretório de grupos de pesquisa do CNPq. **RAI: Revista de Administração e Inovação**, v. 9, n. 2, p. 223-252, 2012.

7. Especialmente, no âmbito dos pesquisadores da Diset do Ipea.

ANEXO A

FIGURA A.1

Figura de rede com 10.000 grupos



Elaboração do autor.

ANEXO B

TABELA B.1

Lista dos grupos por subclasses CNAE

Grupos	Código subclasse CNAE 2.2
2.412	P.85
476	O.84
286	M.72
227	S.94
123	Q.86
96	J.62
85	G.46
77	D.35
72	A.01
65	C.10
62	C.21
48	M.74
47	M.71
46	C.20
43	K.64
40	C.19
39	G.47
35	C.26
30	C.33
24	N.82
23	C.32
21	C.23
20	C.24
20	A.02
19	C.17
19	C.28
19	O
19	C.25
17	R.91
17	E.36
14	M.70
13	C.27
13	C.22
13	C.29
12	B.08
11	F.41
11	Q.88
10	C.16
10	A.03
10	H.49

Elaboração do autor.

ANEXO C

TABELA C.1

Grupos por municípios

Município	Número de grupos
São Paulo	2.745
Rio de Janeiro	2.654
Porto Alegre	1.367
Belo Horizonte	1.227
Recife	937
Curitiba	904
Salvador	839
Campinas	827
Brasília	801
Fortaleza	613
Belém	602
Niterói	582
São Carlos	541
Londrina	539
Florianópolis	517
Manaus	452
Santa Maria	445
Maringá	426
João Pessoa	380
Natal	352
Vitória	337
Goiânia	326
Ribeirão Preto	325
Campo Grande	308
Juiz de Fora	302
Uberlândia	301
Pelotas	300
São Luís	292
Teresina	289

Elaboração do autor.