UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

REDES SOCIAIS DE COLABORAÇÃO ACADÊMICA: TENDÊNCIAS, DETERMINANTES E FRICÇÕES DO PROCESSO DE FORMAÇÃO DE COAUTORIAS ENTRE PESQUISADORES NO BRASIL

Maximiliano Barbosa da Silva

Orientador: Prof. Dr. Marcos de Almeida Rangel

SÃO PAULO
2013

Prof. Dr. João Grandino Rodas
Reitor da Universidade de São Paulo

Prof. Dr. Reinaldo Guerreiro
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Joaquim José Martins Guilhoto
Chefe do Departamento de Economia

Prof. Dr. Márcio Issao Nakane
Coordenador do Programa de Pós-Graduação em Economia

MAXIMILIANO BARBOSA DA SILVA

REDES SOCIAIS DE COLABORAÇÃO ACADÊMICA: TENDÊNCIAS, DETERMINANTES E FRICÇÕES DO PROCESSO DE FORMAÇÃO DE COAUTORIAS ENTRE PESQUISADORES NO BRASIL

Tese apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Marcos de Almeida Rangel

Versão Corrigida

(Versão original disponível na Faculdade de Economia, Administração e Contabilidade)

SÃO PAULO

2013

FICHA CATALOGRÁFICA
Elaborada pela Seção de Processamento Técnico do SBD/FEA/USP

Silva, Maximiliano Barbosa da
Redes sociais de colaboração acadêmica: tendências, determinantes e fricções do processo de formação de coautorias entre pesquisadores no Brasil / Maximiliano Barbosa da Silva. -- São Paulo, 2013.
216 p.

Tese (Doutorado) – Universidade de São Paulo, 2013.
Orientador: Marcos de Almeida Rangel.

1. Redes sociais 2. Co-autoria 3. Fricção 4. Co-autoria – Formação - Brasil I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 302.3

Aos meus pais,

José Mario e Maria Aparecida,

e à minha irmã,

Juliana.

AGRADECIMENTOS

Agradeço aos meus pais, José Mario e Maria Aparecida, pelo apoio incondicional e pelos exemplos de dedicação e de perseverança.

Agradeço à minha irmã, Juliana, pelo incentivo, que sempre significou uma importante carga extra de motivação.

Agradeço ao meu orientador, Dr. Marcos de Almeida Rangel, cujos conselhos foram muito além daqueles referentes à valiosa ajuda na elaboração desta tese.

Agradeço ao Dr. Gilberto Tadeu Lima, ao Dr. Marcos de Almeida Rangel e ao Dr. Ricardo de Abreu Madeira pelos comentários no exame de qualificação, que em muito contribuíram para o desenvolvimento deste trabalho.

Agradeço ao Dr. Bruno Cara Giovannetti, ao Dr. Enlinson Henrique Carvalho de Mattos, ao Dr. Jesús Pascual Mena-Chalco, ao Dr. Marcos de Almeida Rangel e ao Dr. Ricardo de Abreu Madeira, pelos comentários e sugestões que contribuíram significativamente para a versão revisada desta tese.

Agradeço aos meus colegas do IPE/USP por contribuírem para um ambiente acadêmico vibrante.

Agradeço ao pessoal da secretaria de pós-graduação, Cida, Francisco, Luciene e Lucimara, pelas orientações, compreensão e competência.

Agradeço à CAPES/Observatório da Educação - Projeto 3313, pelo suporte financeiro durante o programa de doutorado.

RESUMO

Nesta tese, estudam-se as tendências, os determinantes e as fricções do processo de formação de coautorias na produção de artigos acadêmicos entre pesquisadores no Brasil. Esta tarefa é dividida em três artigos.

O primeiro artigo analisa a evolução temporal da incidência e da extensão de coautoria entre pesquisadores de diferentes campos do conhecimento. Os resultados obtidos confirmam a tendência, reportada na literatura, de crescimento do percentual de artigos coautorados e do número médio de autores por artigo. Em seguida, investigam-se os determinantes deste tipo de colaboração formal entre acadêmicos. Encontra-se que mulheres são mais propensas à coautoria do que homens, que trabalhos não puramente teóricos são mais propícios à colaboração e que a probabilidade de colaboração por parte de um pesquisador aumenta com o seu nível de dedicação em alguns campos do conhecimento.

O segundo artigo estuda a evolução da rede social de coautoria em nove grandes áreas do conhecimento ao longo das últimas quatro décadas. Mostra-se que em quatro das nove grandes áreas do conhecimento analisadas, a rede social de coautoria pode ser considerada um mundo pequeno. Adicionalmente, ao decompor o processo de formação dos grafos de coautoria em seus componentes aleatórios e sociais, conclui-se que existem fricções significativas no pareamento de pesquisadores e que estas variam entre os campos do conhecimento, ao longo do tempo e entre populações de pesquisadores.

O terceiro artigo investiga como a distância e o número de menores caminhos entre dois pesquisadores inseridos em uma rede social de coautoria afetam a probabilidade de que eles estabeleçam um elo de colaboração entre si. Oito grandes áreas do conhecimento são estudadas. Os resultados mostram que em quatro destes campos existem efeitos de proximidade em rede sobre a probabilidade de emergência de um novo link de coautoria e que estes efeitos são fortes. Além disto, a estratégia econométrica adotada permite concluir que estes efeitos decorrem do fluxo de informação sobre atributos individuais e qualidade de pareamentos passados através das redes de colaboração acadêmica.

Além destas três contribuições principais, o capítulo introdutório desenvolve uma extensa revisão da literatura sobre redes econômicas e sociais com o objetivo de motivar o grande interesse da profissão pelo tema.

ABSTRACT

This dissertation studies the trends, determinants and frictions of the formation of coauthorship links among researchers in Brazil for the production of academic articles. This task is divided into three papers.

The first paper studies the evolution over time of the incidence and extension of coauthorship among researchers in various fields of knowledge. The results confirm the trend of growth, reported in the literature, in the percentage of coauthored papers and in the average number of authors per paper. Next, the determinants of formal collaboration among academics are investigated. It is found that women are more inclined to collaborate than men, that non-purely theoretic papers are more collaboration prone and that the probability of collaboration is increasing with the number of hours devoted by an individual to research in some fields of knowledge.

The second paper studies the evolution of the coauthorship network in nine broad fields of knowledge over the last four decades. It is shown that in four of the nine broad fields of knowledge analyzed, the coauthorship network can be considered a small world. In addition to that, by decomposing the network formation process into its random and social components, it can be concluded that there exist significant matching frictions among researchers and that they differ across fields, time and population.

The third paper investigates how the distance and the number of shortest paths between two researchers in a coauthorship network affect the probability that they initiate a collaboration. Eight broad fields of knowledge are studied. The results show that in four of them there are social network proximity effects on the probability of the emergence of a new coauthorship link and that these effects are strong. Moreover, the econometric strategy permits to conclude that the effects result from the flow of information about individual attributes and about past matching quality through the links of the network of academic collaboration.

In addition to these three main contributions, the introductory chapter presents an extensive review of the social and economic networks literature with the purpose of motivating the great interest on this subject.

SUMÁRIO

1. INTRODUÇÃO: A ECONOMIA DAS REDES SOCIAIS 7
1.1. Introdução 7
1.2. Conceitos básicos de teoria dos grafos e indicadores de estrutura 10
1.3. Estrutura empírica das redes sociais 17
1.3.1. Fatos estilizados 17
1.3.2. Redes sociais de pesquisadores 22
1.4. Formação de redes sociais: grafos aleatórios 23
1.4.1. Modelo de Erdös e Rényi 23
1.4.2. Variação dinâmica do modelo de Erdös e Rényi 24
1.4.3. Modelo de ligação preferencial 25
1.4.4. Modelo híbrido 26
1.5. Formação de redes sociais: modelos econômicos 28
1.5.1. Redes sociais de pesquisadores 29
1.5.2. Crime 30
1.5.3. Organização industrial 31
1.5.3.1. Modelo de Cournot 31
1.5.3.2. Modelo de Bertrand 32
1.6. Importância das interações intra-redes sociais 32
1.6.1. Educação: identidade e *peer effects* 32
1.6.2. Mercado de trabalho 35
1.6.3. Bens públicos 37
1.6.4. Compartilhamento de risco 38
1.6.5. Difusão 39
1.6.6. Crime 39
1.6.7. Programas sociais 40
1.6.8. Finanças 40
1.6.9. Governança corporativa 41
1.7. Identificação de influências intra-redes sociais 42
1.8. Considerações finais 44

2. TENDÊNCIAS E DETERMINANTES DA COAUTORIA DE ARTIGOS ACADÊMICOS 47
2.1. Introdução 47
2.2. Dados 50
2.3. Evolução da incidência e da extensão da coautoria 52
2.3.1. Grandes áreas do conhecimento 52
2.3.2. Áreas do conhecimento 53
2.4. Estratégia econométrica e definição das variáveis 56
2.5. Estatísticas descritivas 59
2.6. Resultados econométricos: determinantes da coautoria 61
2.6.1. Análise ao nível do artigo 61
2.6.2. Análise ao nível do autor 64
2.7. Conclusões 67

3. REDES SOCIAIS DE COAUTORIA, INTEGRAÇÃO DA COMUNIDADE ACADÊMICA BRASILEIRA E FRICÇÕES DE PAREAMENTO..... 69
3.1. Introdução..... 69
3.2. Dados.....70
3.3. Fatos estilizados..... 73
3.4. Comunidade acadêmica brasileira como um mundo pequeno.....74
3.4.1. Ciências agrárias.....77
3.4.2. Ciências biológicas.....80
3.4.3. Ciências da saúde.....84
3.4.4. Ciências exatas e da terra.....88
3.4.5. Ciências humanas..... 91
3.4.6. Ciências sociais aplicadas.....94
3.4.7. Engenharias..... 98
3.4.8. Linguística, letras e artes..... 101
3.4.9. Outros..... 104
3.5. Distribuição de grau e mecanismo de formação..... 108
3.6. Conclusões..... 119

4. EFEITOS DE REDES SOCIAIS NA ACADEMIA BRASILEIRA..... 121
4.1. Introdução..... 121
4.2. Dados e definição das variáveis..... 124
4.3. Estatísticas descritivas.....129
4.4. Estratégia econométrica..... 136
4.5. Resultados econométricos.....139
4.5.1. Efeitos de rede social.....139
4.5.2. Variáveis de controle.....147
4.6. Conclusões..... 149

5. CONSIDERAÇÕES FINAIS..... 151

REFERÊNCIAS.....153

APÊNDICE 1 - ESTATÍSTICAS DESCRITIVAS..... 161

APÊNDICE 2 - CLASSIFICAÇÃO DAS ÁREAS DO CONHECIMENTO..... 163

APÊNDICE 3 - INCIDÊNCIA DE COAUTORIA..... 165

APÊNDICE 4 - EXTENSÃO DE COAUTORIA..... 175

APÊNDICE 5 - ESTIMAÇÕES REFERENTES AO CAPÍTULO 2..... 189

APÊNDICE 6 - MESMA AMOSTRA POOLED E EFEITOS FIXOS..... 209

LISTA DE TABELAS

Tabela 1.1 - Estatísticas de posicionamento individuais.....17
Tabela 1.2 - Associação positiva nas rede sociais..... 20
Tabela 1.3 - Frequência de amizades em uma escola americana.....21
Tabela 1.4 - Frequência de amizades em uma escola holandesa..... 21
Tabela 1.5 - Redes sociais de coautoria por campo de pesquisa.....22
Tabela 2.1 - Incidência de coautoria (%): grandes áreas do conhecimento.....165
Tabela 2.2 - Incidência de coautoria (%): ciências agrárias.....166
Tabela 2.3 - Incidência de coautoria (%): ciências biológicas.....167
Tabela 2.4 - Incidência de coautoria (%): engenharias.....168
Tabela 2.5 - Incidência de coautoria (%): ciências exatas e da terra..... 169
Tabela 2.6 - Incidência de coautoria (%): ciências humanas..... 170
Tabela 2.7 - Incidência de coautoria (%): linguística, letras e artes..... 171
Tabela 2.8 - Incidência de coautoria (%): ciências da saúde..... 172
Tabela 2.9 - Incidência de coautoria (%): ciências sociais aplicadas..... 173
Tabela 2.10 - Incidência de coautoria (%): outros.....174
Tabela 2.11 - Extensão de coautoria: grandes áreas do conhecimento.....175
Tabela 2.12 - Extensão de coautoria: ciências agrárias.....176
Tabela 2.13 - Extensão de coautoria: ciências biológicas.....177
Tabela 2.14 - Extensão de coautoria: engenharias.....178
Tabela 2.15 - Extensão de coautoria: ciências exatas e da terra..... 179
Tabela 2.16 - Extensão de coautoria: ciências humanas.....180
Tabela 2.17 - Extensão de coautoria: linguística, letras e artes..... 181
Tabela 2.18 - Extensão de coautoria: ciências da saúde..... 182
Tabela 2.19 - Extensão de coautoria: ciências sociais aplicadas..... 183
Tabela 2.20 - Extensão de coautoria: outros.....184
Tabela 2.21 - Estatísticas descritivas: 1971-2009.....185
Tabela 2.22 - Estatísticas descritivas: 2000-2006.....186
Tabela 2.23 - Estatísticas descritivas: 2007-2008.....187
Tabela 2.24 - Probabilidade de coautoria: ciências agrárias.....189
Tabela 2.25 - Probabilidade de coautoria: ciências biológicas.....190
Tabela 2.26 - Probabilidade de coautoria: engenharias..... 191
Tabela 2.27 - Probabilidade de coautoria: ciências exatas e da terra.....192
Tabela 2.28 - Probabilidade de coautoria: ciências humanas..... 193
Tabela 2.29 - Probabilidade de coautoria: linguística, letras e artes.....194
Tabela 2.30 - Probabilidade de coautoria: ciências da saúde.....195
Tabela 2.31 - Probabilidade de coautoria: ciências sociais aplicadas.....196
Tabela 2.32 - Número de autores: ciências agrárias..... 197
Tabela 2.33 - Número de autores: ciências biológicas.....198
Tabela 2.34 - Número de autores: engenharias.....199
Tabela 2.35 - Número de autores: ciências exatas e da terra..... 200
Tabela 2.36 - Número de autores: ciências humanas.....201
Tabela 2.37 - Número de autores: linguística, letras e artes..... 202
Tabela 2.38 - Número de autores: ciências da saúde..... 203
Tabela 2.39 - Número de autores: ciências sociais aplicadas..... 204
Tabela 2.40.1 - Probabilidade de coautoria ao nível do autor: *pooled cross-section*..... 205
Tabela 2.40.2 - Probabilidade de coautoria ao nível do autor: *pooled cross-section*..... 206
Tabela 2.41 - Probabilidade de coautoria ao nível do autor: efeitos fixos.....207

Tabela 3.1 - Número de pesquisadores: ciências agrárias..... 77
Tabela 3.2 - Grau médio: ciências agrárias.....78
Tabela 3.3 - Porcentagem dos maiores componentes: ciências agrárias.....79
Tabela 3.4 - Distância média e diâmetro: ciências agrárias.....79
Tabela 3.5 - Aglomeração (%): ciências agrárias..... 80
Tabela 3.6 - Número de pesquisadores: ciências biológicas.....81
Tabela 3.7 - Grau médio: ciências biológicas.....82
Tabela 3.8 - Porcentagem dos maiores componentes: ciências biológicas.....82
Tabela 3.9 - Distância média e diâmetro: ciências biológicas.....83
Tabela 3.10 - Aglomeração (%): ciências biológicas..... 84
Tabela 3.11 - Número de pesquisadores: ciências da saúde..... 84
Tabela 3.12 - Grau médio: ciências da saúde.....85
Tabela 3.13 - Porcentagem dos maiores componentes: ciências da saúde..... 86
Tabela 3.14 - Distância média e diâmetro: ciências da saúde.....87
Tabela 3.15 - Aglomeração (%): ciências da saúde.....87
Tabela 3.16 - Número de pesquisadores: ciências exatas e da terra..... 88
Tabela 3.17 - Grau médio: ciências exatas e da terra.....89
Tabela 3.18 - Porcentagem dos maiores componentes: ciências exatas e da terra..... 89
Tabela 3.19 - Distância média e diâmetro: ciências exatas e da terra.....90
Tabela 3.20 - Aglomeração (%): ciências exatas e da terra.....91
Tabela 3.21 - Número de pesquisadores: ciências humanas.....91
Tabela 3.22 - Grau médio: ciências humanas..... 92
Tabela 3.23 - Porcentagem dos maiores componentes: ciências humanas.....93
Tabela 3.24 - Distância média e diâmetro: ciências humanas..... 93
Tabela 3.25 - Aglomeração (%): ciências humanas.....94
Tabela 3.26 - Número de pesquisadores: ciências sociais aplicadas..... 95
Tabela 3.27 - Grau médio: ciências sociais aplicadas.....95
Tabela 3.28 - Porcentagem dos maiores componentes: ciências sociais aplicadas.....96
Tabela 3.29 - Distância média e diâmetro: ciências sociais aplicadas.....97
Tabela 3.30 - Aglomeração (%): ciências sociais aplicadas..... 97
Tabela 3.31 - Número de pesquisadores: engenharias.....98
Tabela 3.32 - Grau médio: engenharias..... 99
Tabela 3.33 - Porcentagem dos maiores componentes: engenharias.....99
Tabela 3.34 - Distância média e diâmetro: engenharias..... 100
Tabela 3.35 - Aglomeração (%): engenharias.....101
Tabela 3.36 - Número de pesquisadores: linguística, letras e artes..... 101
Tabela 3.37 - Grau médio: linguística, letras e artes.....102
Tabela 3.38 - Porcentagem dos maiores componentes: linguística, letras e artes.....103
Tabela 3.39 - Distância média e diâmetro: linguística, letras e artes.....103
Tabela 3.40 - Aglomeração (%): linguística, letras e artes.....104
Tabela 3.41 - Número de pesquisadores: outros.....105
Tabela 3.42 - Grau médio: outros..... 105
Tabela 3.43 - Porcentagem dos maiores componentes: outros.....106
Tabela 3.44 - Distância média e diâmetro: outros..... 107
Tabela 3.45 - Aglomeração (%): outros.....107
Tabela 3.46 - Aleatoriedade das redes sociais de coautoria.....114
Tabela 4.1 - Estatísticas descritivas..... 130
Tabela 4.2.1 - Regressão logística (coeficientes): *pooled cross section*.....141
Tabela 4.2.2 - Regressão logística (coeficientes*): pooled cross section*.....142
Tabela 4.3.1 - Regressão logística (coeficientes): efeitos fixos.....144

Tabela 4.3.2 - Regressão logística (coeficientes): efeitos fixos.....145
Tabela 4.4 - Impacto marginal da distância (%).....147
Tabela 4.5.1 - Regressão logística (coeficientes): *pooled cross-section*.....209
Tabela 4.5.2 - Regressão logística (coeficientes): *pooled cross-section*.....210
Tabela A.1 - Estatísticas descritivas..... 162

LISTA DE FIGURAS

Figura 1.1 - Grafo e sua versão simplificada..... 12
Figura 1.2 - Menores caminhos e diâmetro.....14
Figura 1.3 - Componentes.....14
Figura 1.4 - Conjunto independente maximal e clique..... 15
Figura 3.1 - Grafos de coautoria..... 76
Figura 3.2 - Distribuição de grau: grafo irrestrito.....110
Figura 3.3 - Distribuição de grau: subgrafo Lattes/USP.....111
Figura 3.4 - Aleatoriedade do grafo irrestrito..... 117
Figura 3.5 - Aleatoriedade do subgrafo Lattes/USP..... 118
Figura 4.1 - Grafo de coautoria Lattes/USP.....123
Figura 4.2 - Proximidade média antes da primeira colaboração.....132
Figura 4.3 - Número médio de geodésicas antes da primeira colaboração.....133

1 INTRODUÇÃO: A ECONOMIA DAS REDES SOCIAIS

1.1 Introdução

Parte significativa das interações diretas entre os agentes não ocorre de forma irrestritamente multilateral, em que cada agente interage com todos os demais simultaneamente. Ao invés disto, com frequência, as interações envolvem diretamente apenas um subconjunto pequeno da população, que se auto-organiza em redes sociais cujas estruturas são relevantes para a compreensão de inúmeros fenômenos de interesse.

Embora as redes sociais sejam críticas em um amplo conjunto de ambientes, apenas recentemente, com várias décadas de atraso em relação aos sociólogos, este assunto foi incorporado ao conjunto de temas de investigação sistemática dos economistas. Em parte, este atraso remete à evolução do escopo da ciência econômica. Como discutido por Manski (2000), durante a maior parte do século XX, os economistas privilegiaram o rigor analítico em detrimento da abrangência de temas de investigação. Naquela época, que se estende até a década de 1970, os economistas estudavam predominantemente mercados centralizados, em que as trocas, mediadas por preços, ocorrem simultaneamente e multilateralmente. Neste tipo de abordagem, as interações diretas entre os agentes não constituem um tema de interesse intrínseco, posto que seu foco restringe-se às interações anônimas, via preços. A partir da década de 1970, a evolução da ciência econômica passou a caracterizar-se também pelo aumento da sua abrangência. Neste processo, ao mesmo tempo em que assegura o rigor analítico, a ciência econômica passa a incorporar novos temas, muitos dos quais originalmente identificados como pertencentes a outras disciplinas, como sociologia, ciências políticas, educação, antropologia e psicologia. É neste contexto de alargamento do escopo da ciência econômica que se enquadra o recente interesse da profissão pelo estudo das interações sociais e das redes sociais. De fato, o interesse por estes temas tem crescido em ritmo tão acelerado que não é nenhum exagero afirmar que atualmente constituem dois dos tópicos mais vibrantes em economia, tanto pelo lado teórico quanto pelo lado empírico.

O fato de muitas interações entre os agentes serem inseridas em ambientes sociais não é inócuo, ou seja, o recente interesse dos economistas pelo estudo das redes sociais não se deve apenas a curiosidade intelectual. Exemplos de situações em que as interações entre os

agentes e, consequentemente, os resultados observados são moldados pela estrutura das redes sociais são variados, incluindo influências entre estudantes quanto a desempenho escolar, envolvimento com o crime e consumo de drogas. A estrutura de uma rede social também determina a capacidade e a estabilidade de arranjos de compartilhamento de risco entre seus membros e o nível de provisão de bens públicos. Além disto, as suas conexões funcionam como canais de transporte de informação sobre oportunidades de emprego, novos produtos, novas tecnologias e preços. Com efeito, a maioria dos mercados não é centralizada, observação esta que motiva a literatura dos modelos de *search* e *matching* (ROGERSON; SHIMER; WRIGHT, 2005).

Além de contribuírem para o entendimento das interações dentro de redes sociais dadas, os economistas podem ajudar a entender os mecanismos de formação destas redes. Atualmente, existe ampla literatura empírica que mostra que muitas redes sociais apresentam características incompatíveis com um mecanismo de formação inteiramente aleatório. O que se verifica é que as redes sociais observadas resultam, ao menos parcialmente, de decisões individuais quanto ao estabelecimento de conexões. Na medida em que estas decisões são tomadas buscando o auto-interesse, os conceitos e ferramentas desenvolvidos pelos economistas são úteis para modelar a formação destas estruturas. É importante destacar que este ramo da literatura beneficia-se de contribuições advindas de diversos campos do conhecimento tais como: teoria dos grafos aleatórios[1] e mecânica estatística[2], que estudam processos dinâmicos *ad hoc* capazes de reproduzir a topologia de redes observadas nos dados. Conceitos advindos da sociologia e da psicologia social também têm sido incorporados aos modelos econômicos. Note que, embora estas duas abordagens (econômica e grafos aleatórios/mecânica estatística) sejam alternativas, elas não são antagônicas, razão pela qual são comuns os modelos híbridos que as combinam, fazendo com que o estudo da formação de redes seja interdisciplinar.

O presente trabalho contribui para a literatura em redes sociais ao investigar a formação de elos de coautoria entre pesquisadores para a produção de artigos acadêmicos no Brasil. A importância do tema é evidente na medida em que os artigos acadêmicos contêm a fronteira do conhecimento, que pode ser revertido em avanços tecnológicos (que são o elemento fundamental do crescimento econômico) e em melhorias na qualidade de vida. Especificamente, estudam-se as redes sociais de colaboração acadêmica em nove grandes áreas do conhecimento: ciências agrárias; ciências biológicas; ciências da saúde; ciências exatas e da terra; ciências humanas; ciências sociais aplicadas; engenharias; linguística, letras e artes; e outros. Com base em riquíssimas informações contidas nos currículos Lat-

[1] Ver Bollobás (2001).
[2] Ou física-estatística. Dois excelentes livros sobre o tema são: Ellis (2005) e Yeomans (1992).

tes[3] de 62516 indivíduos associados à Universidade de São Paulo, desenvolvem-se análises empíricas com o objetivo de entender a formação dos seus links de coautoria. Esta tarefa está distribuída em três artigos.

O primeiro artigo investiga a evolução temporal da incidência e da extensão de coautoria entre pesquisadores de diferentes campos do conhecimento. Tal como reportado na literatura, encontra-se que o percentual de artigos produzidos com a colaboração de um ou mais autores tem crescido. Além disto, verifica-se que o número médio de autores por artigo tem aumentado. Em seguida, procura-se entender os determinantes da coautoria ao nível do artigo e ao nível os pesquisador. Concomitantemente, estuda-se a relação entre coautoria e qualidade das publicações.

No segundo artigo, contribui-se para a sociologia da ciência ao descrever a evolução ao longo do tempo da estrutura das redes sociais de coautoria nas nove grandes áreas do conhecimento. Em concordância com pesquisas anteriores (GOYAL; LEIJ; MORAGA-GONZÁLEZ, 2006; MOODY, 2004; NEWMAN, 2001), mostra-se que a comunidade acadêmica tem convergido para aquilo que se convenciona chamar de um mundo pequeno. Em seguida, com o objetivo de aprender sobre os mecanismos subjacentes às estruturas de rede observadas nos grafos de colaboração acadêmica estudados, estima-se o modelo dinâmico de formação de redes sociais proposto por Jackson e Rogers (2007). Desta forma, encontram-se evidências de que o pareamento entre pesquisadores está sujeito a fricções e que estas manisfestam-se diferentemente entre os campos do conhecimento.

No terceiro artigo, avança-se para um entendimento mais profundo das interações sociais existentes dentro da comunidade acadêmica. Seguindo a estratégia econométrica proposta por Fafchamps, Goyal e van der Leij (2010), mostra-se que as conexões entre os indivíduos funcionam não apenas como meios de transmissão de conhecimento acadêmico, mas também como instrumentos de superação de assimetrias de informacionais inerentes ao processo de pareamento de pesquisadores em pelo menos quatro grandes áreas do conhecimento: ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra. Estes resultados reforçam o argumento de que, se as conexões sociais são importantes para o pareamento de agentes mesmo em um ambiente em que há grande disponibilidade de informação e poucas fricções de encontro, então elas devem desempenhar papel de destaque em muitas outras situações que envolvem a colaboração de dois ou mais agentes.

Antes de avançar para estas três contribuições principais do trabalho, é conveniente que se dedique o restante deste capítulo introdutório a uma revisão abrangente da literatura econômica em interações sociais e redes sociais. Com isto, pretende-se oferecer uma pers-

[3]Para uma descrição geral da base de dados utilizada, ver apêndice 1.

pectiva geral dos conceitos básicos, das técnicas utilizadas e dos desafios encontrados. Espera-se desta maneira justificar o entusiasmo da profissão pelos temas. Aqui, é muito importante que se faça uma observação: este ponto do trabalho tem como objetivos fornecer subsídios para uma melhor compreensão dos artigos apresentados adiante e motivar o interesse pelo tema das redes econômicas e sociais. Assim, a apresentação do restante deste capítulo introdutório privilegia abrangência e facilidade de leitura em detrimento de formalizações: as demonstrações de alguns resultados são omitidas para que não se desvie atenção dos objetivos principais.

Em decorrência de as redes sociais serem modeladas como grafos, na seção 1.2 discutem-se alguns conceitos básicos da teoria dos grafos e indicadores de estrutura, que serão úteis para o entendimento da discussão subsequente. Em seguida, na seção 1.3, com o intuito de ganhar uma noção sobre a estrutura das redes sociais observadas empiricamente, apresentam-se alguns fatos estilizados. A seção 1.4 trata de modelos de formação de redes pela abordagem de grafos aleatórios, em que se procura obter as propriedades de um grafo gerado por um processo arbitrário. Já na seção 1.5, avança-se para os trabalhos em que as redes sociais são endógenas. Neste ponto estudam-se alguns modelos econômicos de formação de redes, que são fortemente baseados em teoria dos jogos. Uma distinção elucidativa entre as duas abordagens, apontada por Jackson (2005), é que os modelos de grafos aleatórios ajudam a entender *como* uma rede social com certas características pode ser formada, enquanto que os modelos de teoria dos jogos são úteis para compreender *porque* esta rede com tais caracteríticas deve ser observada. A seção 1.6 discute alguns trabalhos que procuram mostrar como a estrutura das redes sociais nas quais os agentes interagem afeta os resultados observados em uma ampla variedade de ambientes. Na seção 1.7 discutem-se brevemente as dificuldades e as possibilidades em se estimar influências sociais. O debate de temas de interesse, que podem ser proficuamente abordados sob a perspectiva das redes sociais, ocorre ao longo de todo o texto. A seção 1.8 é dedicada às considerações finais.

1.2 Conceitos básicos de teoria dos grafos e indicadores de estrutura

Redes sociais são modeladas como grafos. Por este motivo, é conveniente apresentar alguns conceitos básicos da teoria dos grafos que serão necessários para o entendimento do que se segue. O propósito desta seção é apenas fornecer algumas informações que podem ajudar o leitor não familiarizado com alguns conceitos utilizados ao longo desta tese a compreender com mais facilidade as dicussões adiante. A apresentação aqui é limitada

ao mínimo necessário para a leitura do trabalho. O leitor familiarizado com a teoria dos grafos e com as medidas de estrutura de redes sociais comumente utilizadas pode pular esta seção sem qualquer prejuízo.

Formalmente, um grafo não-direcionado[4] é um par $G = (V, A)$ em que V e A são conjuntos disjuntos, para os propósitos deste trabalho, finitos, cujos elementos são chamados, respectivamente, de vértices (ou nós) e arestas. A ideia é que cada aresta estabelece uma conexão entre dois vértices, não necessariamente distintos. No caso de não haver arestas múltiplas entre dois vértices e não haver laços (arestas que conectam um vértice a si mesmo), diz-se que o grafo é simples. Neste caso, é conveniente dizer que $A \subseteq V^{(2)} := \{\{i, j\} : i, j \in V \text{ e } i \neq j\}$ e as arestas também são chamadas de links. Neste trabalho, quando um grafo não é simples, ele é simplificado, ou seja, arestas múltiplas redundantes e laços são desconsiderados. Assim, sejam $i, j \in V$, se existe uma aresta ij entre eles, diz-se que $ij \in A$ ou, por vezes convenientemente, $ij \in G$.

A figura 1.1 apresenta um grafo e a sua versão simplificada. No grafo original, não-simplificado, as arestas múltiplas redundantes (duas entre a e b, uma entre e e f e uma entre i e j) estão destacadas com tracejados longos na cor vermelha e o laço (incidente em d) com linhas verdes tracejadas. No grafo simplificado, o qual exclui as arestas múltiplas redundantes e o laço, os conjuntos de vértices e de arestas são dados por $V = \{a, b, c, d, e, f, g, h, i, j, k\}$ e $A = \{ab, ac, bd, be, de, df, dg, dh, ef, fg, fh, gh, ij, jk\}$, respectivamente. No restante desta seção, todos os exemplos referem-se a este grafo simplificado.

[4]Neste trabalho, todas as redes sociais estudadas são grafos não-direcionados já que as relações de coautoria são recíprocas. Por este motivo, a apresentação neste ponto é restrita apenas a este tipo de estrutura, ou seja, grafos direcionados não são contemplados.

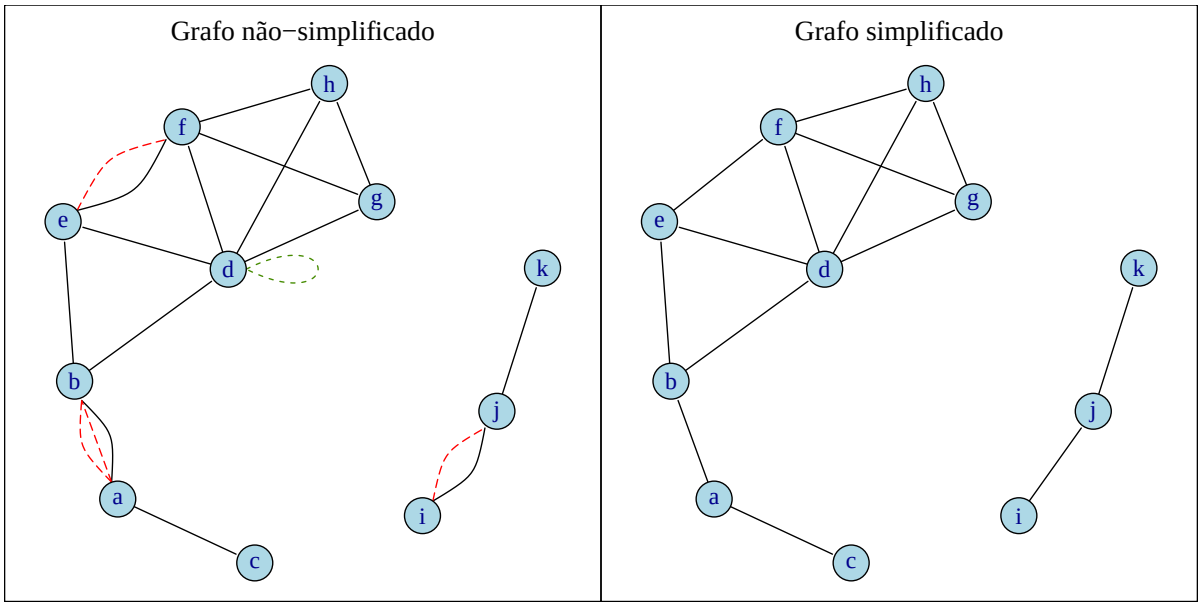

Figura 1.1 - Grafo e sua versão simplificada

Existem outras maneiras alternativas de definir um grafo simples e não-direcionado. Uma bastante útil é como uma dupla ordenada $G = (V, \mathbf{g})$ consistindo de um conjunto não-vazio de vértices e uma matriz $\mathbf{g}$, chamada de matriz de adjacência, onde $\mathbf{g}(i,j) \in \{0,1\}$, $\mathbf{g}(i,j) = \mathbf{g}(j,i)$ para todo $i,j \in V$ e $\mathbf{g}(i,i) = 0$ para todo $i \in V$. Da forma como está sendo definido, G também é não-ponderado já que apenas a existência de uma aresta é de interesse ($\mathbf{g}\,(i,j) = 1$, se a aresta $ij \in A$, e $\mathbf{g}\,(i,j) = 0$, caso contrário). Se a intensidade das arestas também é de interesse, $\mathbf{g}\,(i,j) \in \mathbb{R}$ e, neste caso, G é um grafo ponderado a cujas arestas são atribuídas capacidades. A matriz de adjacência do grafo simplificado da figura 1.1 é a seguinte:

$$
\mathbf{g} \;=\; \begin{array}{cc}
\left(\begin{array}{ccccccccccc}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}\right) & \begin{array}{c} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \\ k \end{array} \\
\begin{array}{ccccccccccc} a & b & c & d & e & f & g & h & i & j & k \end{array} &
\end{array} .
$$

Um grafo simples $G = (V, \mathbf{g})$ é dito completo se seus vértices são dois a dois adjacentes, isto é: $\mathbf{g}(i,j) = 1$ para todo $i, j \in V$ tal que $i \neq j$. Seja $i \in V$ um nó, a sua vizinhança é definida como o conjunto de nós $j \in V$, tais que existe uma aresta com extremos em i e em j, ou seja: $N_i(G) := \{j \in V : ij \in A\}$. O grau de i é igual à cardinalidade da sua vizinhança, $d_i(G) := |N_i(G)|$. Um grafo $G = (V, A)$ é dito regular se $d_i(G) = d$ para todo $i \in V$. Por exemplo, no grafo simplificado da figura 1.1, a vizinhança do vértice b é $N_b(G) = \{a, d, e\}$ e o seu grau é 3. Além disto, aquele grafo não é completo nem regular.

A densidade de um grafo G é igual ao número de arestas presentes em G dividido pelo número máximo de arestas possíveis em G, o que é o mesmo que o grau médio de seus vértices dividido por $n - 1$. Ou seja:

$$D(G) := \frac{\sum_{i \in V} d_i(G)}{n\,(n-1)}. \tag{1.1}$$

No caso do grafo simplificado da figura 1.1, a densidade é de ${}^{2,545}/_{(11-1)} = 25,45\%$.

Um caminho em G entre os vértices $i, j \in V$, é uma sequência de arestas $P(i,j) = (i_1 i_2, ..., i_{K-1} i_K)$, tal que $i_{k-1} i_k \in A$ para todo $k \in \{2, ..., K\}$, com $i_1 = i$ e $i_K = j$, sendo que cada nó em $P(i,j)$ é distinto. Se existe um caminho entre i e j, diz-se que j é alcançável por i e que i e j são caminho-conectados. Um grafo G é dito conexo se todos os seus vértices são alcançáveis, ou seja, se existe um caminho partindo de todo $i \in V$ a cada $j \in V$, com $i \neq j$. O comprimento de um caminho entre os nós $i, j \in V$ é igual ao número de arestas entre eles, ou seja: $l(i,j) = |P(i,j)|$. Já a distância entre estes dois nós é o comprimento do menor caminho entre eles: $dist(i,j) := \min |P(i,j)|$. Se i e j não são caminho-conectados, $dist(i,j) = \infty$. O diâmetro de um grafo é definido como a maior distância entre dois pares quaisquer de seus vértices. Quando um grafo não é conexo, é comum definir o diâmetro como a maior distância no seu maior componente (definido a seguir). Na figura 1.2, o painel esquerdo destaca com linhas vermelhas tracejadas os dois menores caminhos entre os vértices a e f: $P_1^*(a,f) = (ab, bd, df)$ e $P_2^*(a,f) = (ab, be, ef)$. Assim, a distância entre os vértices a e f é 3, já que este é o número de arestas em $P_1^*(a,f)$ e em $P_2^*(a,f)$. O painel direito da figura 1.2 ressalta com linhas vermelhas tracejadas um caminho que define o diâmetro do grafo: $P^*(c,f) = (ca, ab, bd, df)$, o qual tem comprimento 4.

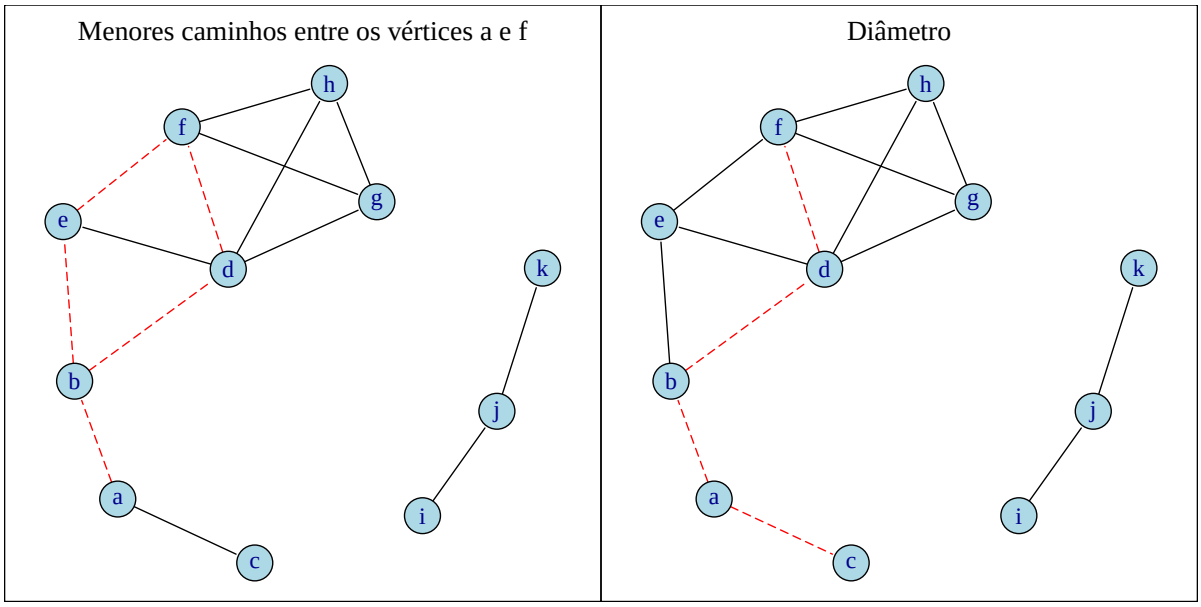

Figura 1.2 - Menores caminhos e diâmetro

Os componentes de um grafo $G = (V, A)$ são os seus subgrafos conexos maximais. Formalmente, um componente do grafo $G = (V, A)$ é um subgrafo $G' = (V', A')$ tal que $V' \subseteq V$, $V' \neq \emptyset$, $A' \subseteq A$, G' é conexo e, se $i \in V'$ e $ij \in A$, então $j \in V'$ e $ij \in A'$. Um grafo é dito conexo se ele só tem um componente. Na figura 1.3 estão destacados os dois componentes do grafo simplificado da figura 1.1. O primeiro componente, com arestas cheias e vértices circulares azuis, contém os vértices $V_1 = \{a, b, c, d, e, f, g, h\}$ e as arestas entre eles, enquanto que o segundo componente, com arestas tracejadas e vértices quadriculados vermelhos, é definido pelos vértices $V_2 = \{i, j, k\}$ e pelas arestas entre eles.

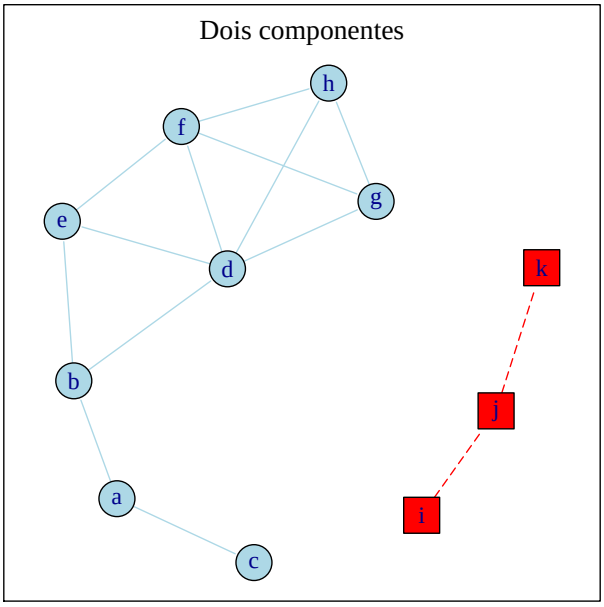

Figura 1.3 - Componentes

Um conjunto independente é um subconjunto $I \subseteq V$ tal que nenhum elemento de I está conectado a outro elemento de I. Um conjunto independente é maximal, $I_{\max}$, se ele não é subconjunto próprio de nenhum outro conjunto independente. Uma clique de G é um subgrafo completo maximal de G. Na figura 1.4, o painel esquerdo destaca em quadrados vermelhos um conjunto independente maximal do grafo: $I = \{a, e, g, i, k\}$. Já o painel direito ressalta também em quadrados vermelhos e com arestas tracejadas de mesma cor uma clique $C = (\{d, f, g, h\}, \{df, dg, dh, fg, fh, gh\})$.

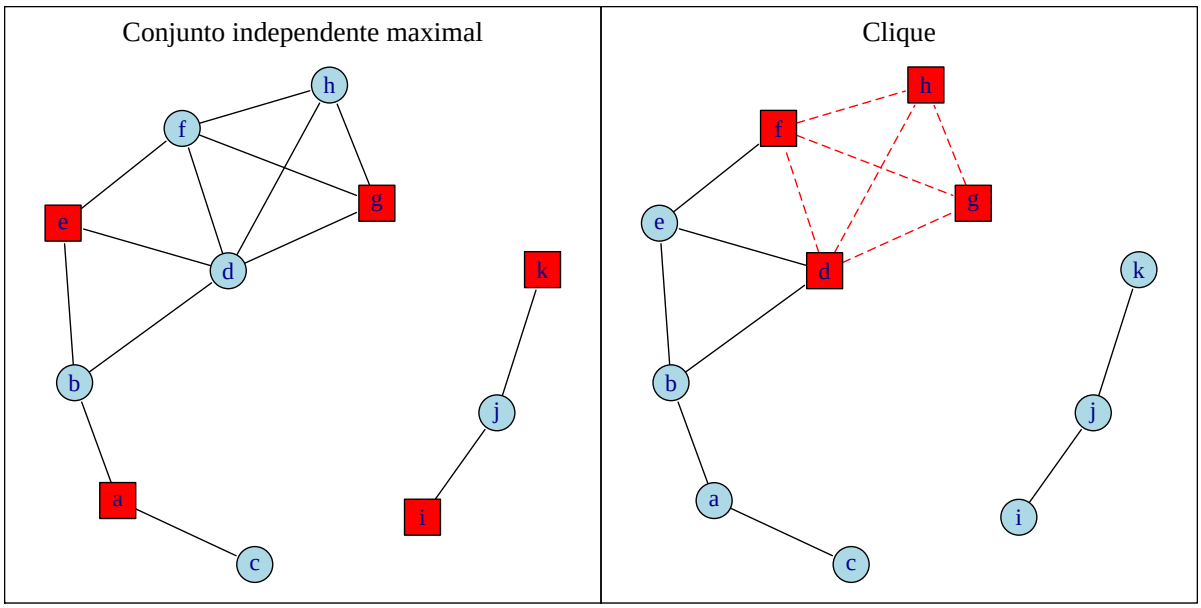

Figura 1.4 - Conjunto independente maximal e clique

Uma forma natural de medir aglomeração de nós em um grafo consiste em contar o número e o tamanho das cliques nele presentes. Embora esta seja uma medida óbvia, ela tem a fragilidade de ser muito sensível à retirada de conexões. Por exemplo, ao se retirar uma aresta de uma clique original composta de quatro nós, obtêm-se duas cliques com três nós cada. Em outras palavras, a retirada de uma única conexão em uma clique pode alterar drasticamente esta medida de aglomeração. Uma alternativa frequentemente utilizada é o coeficiente de aglomeração total, CT, o qual é dado pela proporção das vezes em que i sendo vizinho de j e j sendo vizinho de k, também i é vizinho de k:

$$CT\left(\mathbf{g}\right) := \frac{\sum\limits_{i;j\neq i;k\neq j;k\neq i} g(i,j)g(j,k)g(i,k)}{\sum\limits_{i;j\neq i;k\neq j;k\neq i} g(i,j)g(j,k)}. \tag{1.2}$$

Uma terceira medida de aglomeração é o coeficiente de aglomeração médio, CMe, o qual é a média do coeficiente de aglomeração individual, CI_i, dado pela razão na fórmula abaixo:

$$CMe\,(\mathbf{g}) := \frac{1}{n}\sum_{i}\frac{\sum\limits_{j\neq i;k\neq j;k\neq i} g(i,j)g(i,k)g(j,k)}{\sum\limits_{j\neq i;k\neq j;k\neq i} g(i,j)g(i,k)}. \tag{1.3}$$

No grafo estudado nesta seção, o coeficiente de aglomeração total é de 60% e o coeficiente de aglomeração médio é de 52,08%.

Por vezes, a centralidade de um indivíduo em uma rede social é de interesse. Por este motivo, agora se apresentam duas medidas amplamente utilizadas. É importante ressaltar que existem muitas medidas de centralidade, cada uma das quais captando diferentes aspectos da posição ocupada por um vértice de um grafo. Sendo assim, a escolha de uma medida de centralidade em uma aplicação específica depende essencialmente do contexto analisado. A primeira medida aqui apresentada é a medida de centralidade intermediária[5], a qual é calculada da seguinte maneira. Para cada par de vértices j e k, encontre todos os menores caminhos entre eles e compute a fração destes menores caminhos que passam pelo vértice $i \notin \{j,k\}$. Em seguida obtenha a média desta fração entre todos os vértices j e k. Esta é a medida de centralidade intermediária de i, ou seja:

$$C_i^b := \frac{1}{(n-1)\,(n-2)\,/2}\sum_{jk:k\neq j,i\notin\{j,k\}}\frac{|P_i^*(j,k)|}{|P^*(j,k)|}, \tag{1.4}$$

onde $P^*(j,k)$ é o conjunto de menores caminhos entre j e k e $P_i^*(j,k)$ é o conjunto de menores caminhos entre j e k que passam por i. O indicador de centralidade C_i^b mede o papel intermediador do nó i.

Uma segunda medida de centralidade é a centralidade autovetor[6], que carrega consigo uma noção de prestígio. A ideia desta medida, e de suas variantes, é que a centralidade de um vértice depende da centralidade de seus vizinhos. Considere o caso em que a centralidade de um indivíduo é proporcional à centralidade de seus vizinhos: $C_i^e = \alpha\sum_{j\in V}\mathbf{g}_{ij}C_j^e$, que em notação matricial pode ser escrito como $\mathbf{C^e} = \alpha\mathbf{g}\mathbf{C}^e$. O vetor de centralidade $\mathbf{C}^e$ é um autovetor de $\mathbf{g}$ e $1/\alpha$ é o autovalor correspondente. Como $\mathbf{C}^e$ não é único e em geral não faz sentido valores negativos em $\mathbf{C}^e$, costuma-se utilizar o autovetor associado ao maior autovalor[7] de $\mathbf{g}$ e o normalizar de tal forma que a soma de seus elementos seja igual a 1 ou que a sua norma euclidiana seja unitária.

[5] *Betweeness centrality.*

[6] Medidas deste tipo são utilizadas em ferramentas de busca como o Google.

[7] Pelo teorema de Perron-Frobenius, sabe-se que o maior autovalor de uma matriz não-negativa, como $\mathbf{g}$, é um número real e que o seu autovetor associado é não-negativo.

A tabela 1.1 apresenta algumas estatísticas individuais de posicionamento[8] dos vértices no grafo simplificado da figura 1.1.

Tabela 1.1 - Estatísticas de posicionamento individuais

Vértice	Grau	Aglomeração individual	Centralidade intermediária	Centralidade autovetor
a	2	0,00	0,13	0,08
b	3	0,33	0,22	0,27
c	1	NaN	0,00	0,02
d	5	0,50	0,19	0,53
e	3	0,67	0,03	0,35
f	4	0,67	0,02	0,47
g	3	1,00	0,00	0,39
h	3	1,00	0,00	0,39
i	1	NaN	0,00	0,00
j	2	0,00	0,02	0,00
k	1	NaN	0,00	0,00

A definições apresentadas nesta seção bastam para a compreensão do restante do trabalho. Conforme outros conceitos se fizerem necessários, eles serão introduzidos convenientemente. A seguir, discutem-se algumas regularidades encontradas nas redes sociais observadas empiricamente.

1.3 Estrutura empírica das redes sociais

1.3.1 Fatos estilizados

Um dos fatores que contribuem para o interesse nas rede sociais é a observação de que muitas delas possuem estruturas comuns incompatíveis com o modelo puro do tipo Erdös e Rényi (1959, 1960 e 1961). Isto indica, por um lado, que a sociedade não se organiza de forma aleatória, mas sim que esta organização resulta, ao menos parcialmente, de decisões individuais. O desafio então é desenvolver e testar modelos que combinem decisões individuais, teoria dos grafos aletórios e mecânica-estatística capazes de reproduzir as estruturas de redes verificadas. Por outro lado, a regularidade observada na topologia das redes empíricas permite que se coloquem restrições sobre os tipos de redes sociais relevantes. Como os resultados das interações intra-rede dependem da estrutura desta, a observação empírica permite que se reduza significativamente o número de redes possíveis.

[8]Norma do autovetor, associado ao maior autovalor, com centralidade unitária.

Em alguns contextos esta capacidade pode ser importante já que com n vértices, pode-se gerar $2^{\binom{n}{2}}$ grafos não-direcionados distintos e, portanto, com apenas $n = 24$ supera-se o número estimado de átomos presentes no universo observável[9]! Sendo assim, esta seção discute algumas regularidades sobre a estrutura empírica das redes sociais.

O primeiro fato estilizado é que a distância média entre nós tende a ser pequena, assim como o diâmetro, onde por pequena entende-se algo menor ou igual ao logaritmo natural do número de nós da rede. Um dos primeiros estudos que se propõem a medir a distância social entre indivíduos deve-se a Milgram (1967), que pôs em prática um experimento bastante criativo. O autor solicitou a um grupo de residentes dos estados do Kansas e do Nebraska que fizessem chegar uma carta a um determinado morador do estado de Massachussets, sobre o qual foram informados o nome, a profissão e algumas coordenadas aproximadas do endereço. Para atingir este objetivo, os participantes deveriam enviar a carta a alguém que eles conhecessem bem e que poderia conhecer o destinatário ou re-enviar a carta para outra pessoa que poderia fazê-la chegar até o destino final ou repassar para outro indivíduo e assim sucessivamente. Ao término do experimento, verificou-se que cerca de 25% das cartas atingiram o destino final, o que é uma taxa de resposta bastante alta para este tipo de experimento. Além disto, a mediana do número de links percorridos desde a origem até o destinário foi de apenas 5, sendo 12 o número máximo. Este resultado indica que a distância social entre as pessoas é supreendentemente baixa, sobretudo ao se considerar que os participantes não conheciam a estrutura da rede social e, por este motivo, dificilmente a carta percorreu o caminho mais curto. Por outro lado, é necessário cautela na interpretação deste resultado porque quanto maior o número de passos intermediários entre o remetente inicial e o destinatário, maior a probabilidade de que a carta não seja passada adiante ao longo da cadeia de remetentes. Sendo assim, por este argumento, as cartas que atingiram o destino final são aquelas cujos remetentes iniciais e o destinatário final estavam a uma distância relativamente pequena. De todo modo, outros trabalhos procuram medir a distância entre os vértices de diferentes redes sociais e tipicamente encontram distância média e diâmetro baixos. Por exemplo, ao estudarem redes sociais de coautoria de artigos acadêmicos, Grossman (2002) acha uma distância média 7,6 e um diâmetro de 27 entre matemáticos, Newman (2001) encontra distância média de 5,9 e diâmetro de 20 entre físicos e Goyal et al. (2004) obtêm distância média de 9,5 e diâmetro de 29 entre economistas. Na internet, Adamic (1999), em uma rede com 50 milhões de páginas, encontra distância média de 3,1 e Backstrom et al. (2012) descobrem uma distância média de 4,74 entre 721 milhões de usuários do Facebook.

O baixo número de passos necessários para que um vértice alcance outro é evidência de que a sociedade organiza-se em um *mundo pequeno*, que, informalmente, significa que a

[9]Estima-se que o número de átomos no universo observável esteja em torno de 10^{80}.

distância média e o diâmetro das redes sociais são pequenos. Outra característica comum a muitas redes sociais é alto nível de aglomeração, tipicamente muito maior do que aquele que seria observado caso elas fossem formadas por sorteios aleatórios sobre o conjunto de possíveis links entre um conjunto dado de vértices. Nas redes de coautoria citadas acima, a aglomeração entre matemáticos encontrada por Grossman (2002) é de 0,15 *vis-à-vis* 0,00002 no caso de formação aleatória, Newman (2001) encontra 0,09 *vis-à-vis* 0,00001 e Goyal et al. (2006) obtêm 0,19 *vis-à-vis* 0,00002. Adamic (1999) calcula uma aglomeração de 0,11 *vis-à-vis* 0,0002 e MacRae (1960), que estuda amizades entre prisioneiros, encontra 0,31 *vis-à-vis* 0,0134. Segundo Newman (2002), níveis tão elevados de aglomeração não estão presentes em outros tipos de redes, como as biológicas e as tecnológicas.

Outra regularidade relativa a aglomeração, segundo Jackson e Rogers (2007), é que a aglomeração entre vizinhos de um dado nó, em pelo menos algumas redes sociais, é inversamente relacionada com o grau deste nó. Ou seja, vizinhos de um nó com grau elevado são menos provavelmente ligados uns aos outros do que vizinhos de um nó com baixo grau.

A distribuição de grau de muitas redes sociais apresenta caldas gordas, isto é, a frequência de vértices com grau muito baixo e a frequência de vértices com grau muito alto costumam ser maiores do que suas correspondentes quando o grafo é formado segundo o modelo de Erdös-Rényi. Por vezes, tem sido afirmado que a distribuição de grau das redes sociais segue uma lei de potência, $\Pr(d) = \alpha d^{\beta}$ (PRICE, 1965 e ALBERT; JEONG; BARABÁSI, 1999). Esta afirmação, na maioria das vezes, é baseada em uma análise visual da distribuição de grau em um diagrama com escala logarítmica[10]. Sendo esta aparentemente linear, admite-se que a distribição de grau é bem aproximada por uma lei de potência. Claramente este tipo de análise é pouco cuidadoso já que muitas distribuições podem parecer lineares quando desenhadas em escala logarítmica. Adicionalmente, os trabalhos que se dispõem a estimar com mais cuidado a distribuição de grau encontram que a lei de potência não é aquela que melhor se ajusta aos dados (PENNOCK, 2002 e JACKSON; ROGERS, 2007). No segundo artigo desta tese, atenção especial é dedicada à estimação da distribuição de grau de pesquisadores inseridos em redes sociais de coautoria. Mostra-se que leis de potência não descrevem bem esta variável.

Muitas redes sociais tendem a apresentar o que se chama de associação positiva, ou seja, pessoas altamente conectadas tendem a ter como vizinhos pessoas também com grau elevado (NEWMAN, 2002). De acordo com Newman (2002), esta parece ser uma característica distintiva entre as redes sociais e outros tipos de rede, como as tecnológicas e as

[10]Note que ao tomar o logaritmo de ambos os lados da lei de potência encontra-se a seguinte função linear: $\ln(\Pr(d)) = \alpha + \beta \ln(d)$.

biológicas, que apresentam associação negativa, em que vértices de alto grau tendem a ter vizinhos pouco conectados, como pode ser visto na tabela 1.2.

Tabela 1.2 - Associação positiva nas redes sociais

Coautoria			Outras redes sociais	
Biologia	Física	Matemática	Atores de filmes	Diretores de empresas
0,127	0,363	0,120	0,208	0,276

FONTE: NEWMAN, 2002.

Newman e Park (2003) mostram que as duas características típicas das redes sociais, alta aglomeração e associação positiva, podem resultar de uma organização social em comunidades. Embora este tipo de organização social possa ter diversos motivos como, por exemplo, restrições geográficas à formação de links, uma outra característica comum às redes sociais pode ter um papel significativo. Trata-se da tendência de indivíduos com alguma característica comum estarem conectados em uma frequência superior à fração dos indivíduos com a referida característica na população. Esta tendência de associação entre pessoas parecidas em algum atributo é observada há séculos e a ela se dá o nome de homofilia (LAZARSFELD; MERTON, 1954). Nas redes sociais estudadas, verifica-se homofilia com relação a idade, nível de educação, renda, desempenho escolar, profissão, opiniões políticas, etnia e outras variáveis. Para medir homofilia, uma forma possível é a seguinte. Denote por s_i o número médio de amigos que indivíduos do tipo i possuem dentro do mesmo grupo (amigos que são do tipo i) e por d_i o número médio de amigos que indivíduos do tipo i possuem fora do grupo (amigos de tipo diferente de i). Assim, um índice simples de homofilia é definido da seguinte maneira:

$$H_i := \frac{s_i}{s_i + d_i}. \tag{1.5}$$

A tabela 1.3 reporta estatísticas referentes à composição étnica das amizades para as diferentes raças encontradas em uma escola americana, as quais foram calculadas a partir da base de dados *Add Health*. Repare que, embora os estudantes brancos representem 51% do total de estudantes, 85% dos amigos destes estudantes também são brancos. Por outro lado, os negros compõem 38% do corpo discente, mas 85% de seus amigos são negros. Estas duas informações sugerem que raça é uma variável importante no estabelecimento de elos de amizades entre estudantes porque, se ela não o fosse, seria de se esperar que a proporção de amigos de cada etnia que um certo grupo possui fosse (aproximadamente) igual à proporção de pessoas daquela etnia na população total.

Tabela 1.3 - Frequência de amizades em uma escola americana

	Etnia dos estudantes			
	Branca	Negra	Hispânica	Outras
Quantidade na escola	131	96	13	15
Porcentagem na escola	51	38	5	6
Porcentagem de amigos por etnia				
Branca	**86**	7	47	74
Negra	4	**85**	46	13
Hispânica	4	6	**2**	4
Outras	6	2	5	**9**

FONTE: CURRARINI; JACKSON; PIN, 2009.

A tabela 1.4 apresenta a porcentagem de estudantes das diferentes origens em uma escola holandesa assim como a porcentagem de amigos que cada um destes grupos étnicos possui. Como se pode notar, para todos os grupos o percentual de amigos de mesma etnia é maior do que a fração de cada um deles na população.

Tabela 1.4 - Frequência de amizades em uma escola holandesa

	Etnia dos estudantes				
	Holandesa	Marroquina	Turca	Surinamesa	Outra
Quantidade na escola	850	62	75	100	230
Porcentagem na escola	65	5	6	8	17
Holandesa	**79**	24	11	21	47
Marroquina	2	**27**	8	4	5
Turca	2	19	**59**	8	6
Surinamesa	3	8	8	**44**	12
Outra	13	22	14	23	**30**

FONTE: BAERVELDT et al., 2004

Homofilia é uma característica importante das redes sociais porque ela reflete o grau de segregação existente na sociedade, o que possui implicações sobre fluxo de informação entre grupos, mobilidade e comportamentos (FRYER, 2010). A tendência das pessoas em se relacionarem com indivíduos similares podem ter consequências relevantes em ambientes variados. Por exemplo, Gompers et al. (2012) mostram que investidores em *syndication* tendem a estabelecer parcerias homofílicas entre si tanto em termos de habilidade quanto em termos de afinidade, e que o segundo tipo implica desempenho reduzido dos projetos de investimento aos quais se associam. As consequências da homofilia é um tema que, de fato, tem sido objeto de intensa investigação recente em governança corporativa, onde a preocupação é centrada na maneira pela qual as diretorias e os conselhos de administração das empresas são formados e como isto se reflete sobre a rentabilidade das mesmas (LANDIER et al., 2012). Hong e Page (2004) desenvolvem um modelo em que grupos de trabalho mais heterogêneos podem apresentar desempenho superior ao grupo, de mesmo tamanho, composto pelos indivíduos mais hábeis, porém mais parecidos na maneira como

procuram resolver problemas. Com isto, eles formalizam ideias advindas da psicologia cognitiva e da teoria das organizações que estabelecem uma correlação entre diversidade de identidade e diversidade funcional. Outro trabalho que procura estabelecer um elo entre a estrutura das redes sociais (no caso específico de amizade) e performance de um grupo de trabalho é proposto por Bandiera, Barankay e Rasul (2008), em que os autores argumentam com relação à maneira pela qual contratos influenciam a formação destas redes e a forma de interação intra-rede (*risk-sharing*).

1.3.2 Redes sociais de pesquisadores

Neste ponto, é útil apresentar as propriedades encontradas em algumas redes sociais de coautoria, que constituem o objeto de análise principal desta tese. Na tabela 1.5 apresentam-se alguns indicadores da estrutura da rede de coautoria de pesquisadores de diferentes áreas do conhecimento. É interessante notar que o grau médio (número médio de coautores) varia consideravelmente entre os campos de pesquisa analisados. Por outro lado, quando se consideram a distância média entre dois autores e o diâmetro do maior componente, os valores encontrados são bastante parecidos. Ao se comparar o índice de aglomeração total entre as áreas, percebe-se que biologia, economia e matemática são parecidos nesta dimensão, enquanto que física possui um índice de aglomeração total notavelmente mais elevado. Por fim, verifica-se que o maior componente do grafo de colaboração de biólogos e de matemáticos contém uma fração bastante alta da população. O ponto que deve ser destacado com a análise da tabela 1.5 é que, embora haja elementos idiossincráticos à organização social da atividade de científica em cada uma das quatro disciplinas consideradas, elas têm características comuns, tais como: pequena distância média, pequeno diâmetro no maior componente e alta aglomeração[11].

Tabela 1.5 - Redes sociais de coautoria por campo de pesquisa

	Campo de pesquisa			
	Biologia	Economia	Matemática	Física
Número de pesquisadores	1.520.521	81.217	253.339	52.909
Grau médio	15,5	1,7	3,9	9,3
Distância média	4,9	9,5	7,6	6,2
Diâmetro no maior componente	24	29	27	20
Aglomeração total	0,09	0,16	0,15	0,45
Fração do maior componente	0,92	0,41	0,82	0,45

FONTE: JACKSON, 2008

[11] Como ficará claro adiante, estas características permitem dizer que estas redes sociais podem ser classificadas como mundos pequenos.

1.4 Formação de redes sociais: grafos aleatórios

1.4.1 Modelo de Erdös e Rényi

Erdös e Rényi (1959, 1960, 1961) desenvolvem um modelo estático de formação de grafo bastante simples, mas que serve como referencial muito útil para entender algumas propriedades observadas nas redes sociais tais como distribuição de grau, aglomeração e distâncias entre nós. Como este referencial será de grande utilidade para o artigo do capítulo 3, é conveniente apresentar o modelo e discutir brevemente algumas de suas propriedades. Seja $V = \{1, \ldots, n\}$ um conjunto fixo de vértices e seja A o conjunto de todas as arestas possíveis entre os elementos de V. Suponha que as arestas de A são sorteadas aleatoriamente e independentemente com probabilidade $p \in (0, 1)$. Sendo este um modelo binomial, a distribuição de grau entre vértices é, para n grande[12], aproximadamente:

$$\Pr(d) = \binom{n-1}{d} p^d (1-p)^{n-1-d}. \tag{1.6}$$

Para n grande e p pequeno, a distribuição de grau (1.6) aproxima-se de uma Poisson:

$$\Pr(d) = \frac{e^{-(n-1)p} \left[(n-1)\, p\right]^d}{d!}. \tag{1.7}$$

Note que a aglomeração em um grafo gerado desta maneira é igual p, já, que sendo o sorteio das arestas independentes, $\Pr(\mathbf{g}_{jk} = 1 \mid \mathbf{g}_{ij} = \mathbf{g}_{ik} = 1) = \Pr(\mathbf{g}_{jk} = 1) = p$. Agora, para entender a definição de mundo pequeno proposta em Goyal et al. (2006), a qual é utilizada no artigo do capítulo 3, é conveniente apresentar algumas propriedades assintóticas, para número de vértices muito grande, do modelo de Erdös e Rényi. Denotando por $\tilde{d}(n)$ o grau esperado em um grafo Erdös-Rényi com n vértices, para $\tilde{d}(n) > (1+\varepsilon)\ln(n)$, então conforme n cresce, a fração de vértices com grau próximo[13] do grau médio tende a 1 e quase certamente o grafo é conexo (existe um componente gigante, englobando uma fração não-desprezível dos vértices). Nestas condições, se $\tilde{d}(n)/n \to 0$, então a distância média converge em probabilidade para $\ln(n)/\ln(\tilde{d})$. As condições impostas por Goyal et al. (2006) para que um grafo seja um mundo pequeno são as seguintes: (i) o número de vértices é muito grande em relação ao grau médio; (ii) existe um caminho conectando uma parcela

[12] Note que para n pequeno, embora as arestas sejam sorteadas aleatoriamente, existe correlação de grau entre vértices. Contudo, conforme n cresce, esta correlação desaparece.

[13] Pelos limites de Chernoff, $\Pr(\tilde{d}/3 \leq d_i \leq 3\tilde{d},\ \text{para todo } i) \geq \left(1 - e^{-\tilde{d}}\right)^n$.

grande da população, o que tecnicamente se refere como um componente gigante; (iii) A distância média no componente gigante é pequena: da ordem do logarítimo natural do número de nós na rede; (iv) aglomeração é alta: muito maior do que o número médio de coautores dividido pelo número de nós (aproximadamente igual à densidade do grafo).

Resumindo, informalmente, a definição de mundo pequeno proposta por Goyal et al. (2006) exige que o grafo tenha as mesmas propriedades que um grafo Erdös-Rényi conexo grande com baixa densidade, porém com alta aglomeração. Ao observarem a evolução da rede de coautoria dos economistas nas décadas de 1970, 1980 e 1990, os autores encontram que o grau médio nas redes consideradas é bastante pequeno em relação ao número total de nós, a distância dentro com componente gigante é pequena e os indicadores de aglomeração são altos. Por fim, os autores encontram que, na década de 1970, o componente gigante continha apenas 15% dos economistas, porém na década de 1990, esta cobertura subiu para 40%. Nestas condições, eles concluem que a rede social de economistas é um mundo pequeno emergente.

1.4.2 Variação dinâmica do modelo de Erdös e Rényi

Nesta subseção apresenta-se uma simples variação dinâmica do modelo de Erdös e Rényi, em que novos vértices são adicionados ao grafo ao longo do tempo. Assim, suponha que a cada período t nasce um novo vértice, o qual forma, com a mesma probabilidade, m links com os nós pré-existentes no grafo. Assim, a probabilidade de que um vértice receba uma nova aresta no período $t+1$ é:

$$\Pr\left(d_i(t+1) = d_i(t) + 1\right) = \frac{m}{t}. \tag{1.8}$$

Assim, em t, o grau esperado de um vértice i, denotado por $\tilde{d}_i$, é dado por:

$$\tilde{d}_i(t) = m + \sum_{j=i+1}^{t} \frac{m}{j-1}, \tag{1.9}$$

que, para i grande, pode ser aproximado por:

$$\tilde{d}_i(t) \cong m\left[1+\ln\left(\frac{t-1}{i-1}\right)\right] \cong m\left[1+\ln\left(\frac{t}{i}\right)\right]. \tag{1.10}$$

Nestas condições, a distribuição de probabilidade do grau esperado é dada por:

$$F_t(\tilde{d}) = 1 - e^{-\frac{(\tilde{d}-m)}{m}}. \tag{1.11}$$

1.4.3 Modelo de ligação preferencial

Nesta subseção, discute-se brevemente o modelo de ligação preferencial de formação de grafos. Apresenta-se este modelo porque isto facilitará a compreensão do segundo artigo desta tese. O modelo de ligação preferencial tem recebido considerável atenção por gerar uma distribuição de grau com bom ajustamento àquelas de diferentes redes sociais observadas (PRICE, 1965 e BARABÁSI; ALBERT, 1999).

No modelo de ligação preferencial, um novo vértice nasce a cada período de tempo[14]. Ao nascer, cada vértice forma m links com os nós pré-existentes aleatoriamente, porém com probabilidades diferentes. Especificamente, a probabilidade de um novo vértice, t, conectar-se com um nó i é proporcional à razão entre o grau de i e a soma dos graus no grafo imediatamente anterior ao nascimento do novo vértice t:

$$\Pr\left(d_i(t+1) = d_i(t)+1\right) = \frac{d_i(t)}{2t}. \tag{1.12}$$

Para encontrar a distribuição de grau esperado resultante deste processo, utiliza-se a sua aproximação de campo médio, em que o tempo é contínuo e a variação de grau ocorre deterministicamente à taxa esperada[15]. Desta forma, a função que descreve a trajetória de grau esperado ao longo do tempo é obtida resolvendo a seguinte equação diferencial:

$$d_i^{'}(t) = \frac{d_i(t)}{2t},\ d_i(i) = m. \tag{1.13}$$

[14]É possível escrever o modelo com o nascimento de mais de um vértice por período. Porém, apenas para fixar a ideia fundamental do modelo, aqui se apresenta apenas a versão mais simples.

[15]Existem alguns teoremas que mostram que a aproximação de campo médio funciona bem para o modelo de ligação preferencial.

A solução da equação (1.13) é:

$$d_i(t) = m\left(\frac{t}{i}\right)^{\frac{1}{2}}. \tag{1.14}$$

Sendo assim, a função de distribuição de d, $\Pr\left(d_i(t) < d\right)$, é:

$$F_t(d) = 1 - m^2 d^{-3}, \tag{1.15}$$

cuja função densidade de probabilidade é dada por:

$$f_t(d) = 2m^2 d^{-3}, \tag{1.16}$$

que é uma lei de potência.

1.4.4 Modelo híbrido

Um processo intermediário entre o modelo dinâmico do tipo Erdös e Rényi e o modelo de ligação preferencial é proposto por Jackson e Rogers (2007). Este é um modelo interessante porque, embora bastante simples, é capaz de gerar diversas regularidades observadas nas redes sociais. Além disto, o modelo permite estimar o grau de aleatoriedade em uma rede social observada a partir da distribuição de grau de seus vértices.

Como nos modelos dinâmicos discutidos anteriormente, a cada período t um novo vértice é adicionado ao grafo. Este vértice, ao nascer, identifica aleatoriamente e sem reposição m_r outros vértices, seus parentes. O vértice entrante então pode estabelecer um link com estes nós inicialmente identificados, o que acontece com probabilidade p_r. O novo vértice também encontra m_n outros nós na união da vizinhança de seus parentes e estabelece uma aresta com eles com probabilidade p_n. Nestas condições, a probabilidade de um vértice i pré-existente com grau $d_i(t)$ ser incidido por uma nova aresta em $t+1$ é aproximadamente igual a:

$$\Pr\left(d_i(t+1) = d_i(t)+1\right) = p_r \frac{m_r}{t} + p_n \left[\frac{m_r d_i(t)}{t}\right]\left[\frac{m_n}{m_r\left(p_r m_r + p_n m_n\right)}\right]. \tag{1.17}$$

Definindo $m := p_r m_r + p_n m_n$, a equação (1.17) pode ser escrita da seguinte maneira:

$$\Pr\left(d_i(t+1) = d_i(t)+1\right) = \frac{p_r m_r}{t} + \frac{p_n m_n d_i(t)}{mt}. \tag{1.18}$$

Aplicando a aproximação de campo médio, tem-se que o grau esperado do vértice i evolui de acordo com a seguinte equação diferencial:

$$d_i^{'}(t) = \frac{p_r m_r}{t} + \frac{p_n m_n d_i(t)}{mt}, \tag{1.19}$$

cuja solução é dada por:

$$d_i(t) = (d_0 + rm)\left(\frac{t}{i}\right)^{\frac{1}{1+r}} - rm, \tag{1.20}$$

onde $r := \frac{p_r m_r}{p_n m_n}$ é a razão entre o número de links que são formados aleatoriamente e o número de links formados socialmente. Para os propósitos do artigo do capítulo 3, este é o parâmetro de maior interesse. Valores elevados de r indicam que uma rede social é formada de maneira essencialmente aleatória, enquanto que valores baixos indicam que conexões sociais pré-existentes são importantes para o estabecimento de novos links. Assim, em certo sentido, r é uma medida das fricções inerentes aos processo de pareamento entre os vértices. Com base na equação (1.20) encontra-se a distribuição de grau esperado:

$$F_t(d) = 1 - \left(\frac{d_0 + rm}{r + rm}\right)^{1+r}, \tag{1.21}$$

para $d > d_0$. Tomando o logaritmo e rearranjando a equação (1.21), encontra-se a equação que será estimada no artigo do capítulo 3:

$$\ln\left(1 - F_t(d)\right) = (1+r)\left[\ln\left(d_0 + rm\right) - \ln\left(d + rm\right)\right]. \tag{1.22}$$

Vale notar que $\lim_{r\to 0} F_t(d) = 1 - {}^{d_0}/_{d}$, que é uma distribuição livre de escala. Além disto, repare que $\lim_{r\to\infty} \ln(1 - F_t(d)) = {}^{(d_0-d)}/_{m}$, que é uma distribuição exponencial.

1.5 Formação de redes sociais: modelos econômicos

A abordagem de grafos aleatórios, discutida brevemente na seção anterior, é interessante porque, através de ajustes nos processos de formação de links, permite gerar modelos capazes de reproduzir características observadas nos grafo empíricos. Desta forma, é possível aprender *como* certas estruturas de grafos podem ser obtidas a partir de um processo arbitrário. Por outro lado, a abordagem de grafos aleatórios deixa a desejar quando se pergunta *porque* certas estruturas são observadas. A abordagem econômica, em geral fundamentada na teoria dos jogos, é apropriada para responder este tipo de questionamento. Além disto, os modelos econômicos, ao introduzirem explicitamente os *payoffs* dos agentes, permite que se obtenham avaliações de bem-estar. Sendo assim, esta seção é dedicada aos trabalhos em que a estrutura do grafo no qual os agentes interagem é um resultado de equilíbrio. Entretanto, antes de prosseguir, deve ser frisado que, frequentemente, existe conflito entre eficência e estabilidade das redes. Para compreender este ponto, convém discutir alguns conceitos de eficiência e de estabilidade empregados e, então, dar um exemplo em que redes eficientes não são estáveis e redes estáveis não são eficientes.

Seja $u_i(\mathbf{g})$ a utilidade líquida que um indivíduo i obtém no grafo (com matriz de adjacência) $\mathbf{g}$. Então: (i) Diz-se que $\mathbf{g}$ é eficiente no sentido de Pareto se não existe um outro grafo $\mathbf{g}'$ tal que $u_i(\mathbf{g}') \geq u_i(\mathbf{g})$ para todo i, sendo desigualdade estrita para pelo menos um i; (ii) Diz-se que $\mathbf{g}$ é eficiente se $\mathbf{g} \in \arg\max_{\mathbf{g}} \sum_{i=1}^{|V|} u_i(\mathbf{g})$; (iii) Diz-se que $\mathbf{g}$ é estável por pares se para todo $ij \in \mathbf{g}$ (isto é: $\mathbf{g}(i,j) = 1$), $u_i(\mathbf{g}) \geq u_i(\mathbf{g} - ij)$ e $u_j(\mathbf{g}) \geq u_j(\mathbf{g} - ij)$, e para todo $ij \notin \mathbf{g}$, se $u_i(\mathbf{g} + ij) > u_i(\mathbf{g})$ então $u_j(\mathbf{g} + ij) < u_j(\mathbf{g})$.

Note que todo grafo eficiente é eficiente de Pareto, mas nem todo grafo eficiente de Pareto é eficiente. Considere o seguinte exemplo. Suponha que existam 4 agentes e que a utilidade recebida de um link por um agente é $u_i(ij) = 4 - d_j(\mathbf{g})$, onde $d_j(\mathbf{g})$ denota o grau do indivíduo j em $\mathbf{g}$. Na figura abaixo, que apresenta uma sequência de matrizes de adjacência, inicia-se com o grafo vazio e se adicionam arestas sequencialmente sempre que algum par de vértices tiver o incentivo para fazer isto. Como se pode perceber, apenas o grafo completo é estável por pares sem ser eficiente. Além disto, o grafo eficiente não é

estável.

$$\underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{\text{Ineficiente, Ineficiente de Pareto e Instável}} \rightarrow \underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{\text{Ineficiente, Ineficiente de Pareto e Instável}} \rightarrow$$

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{Ineficiente, Eficiente de Pareto e Instável}} \rightarrow \underbrace{\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{Ineficiente, Eficiente de Pareto e Instável}} \rightarrow$$

$$\underbrace{\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}}_{\text{Eficiente, Eficiente de Pareto e Instável}} \rightarrow \underbrace{\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}}_{\text{Ineficiente, Eficiente de Pareto e Instável}} \rightarrow$$

$$\underbrace{\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}}_{\text{Ineficiente, Eficiente de Pareto e Estável}}$$

O motivo desta separação entre eficiência e estabilidade é que quando dois agentes decidem conectar-se eles ignoram o efeito negativo que causam sobre o bem-estar dos seus vizinhos, ou seja: eles geram externalidades negativas porque, e esta é uma lição geral, quando se trabalha em redes "nada" é independente, de tal forma que as decisões de um indivíduo "sempre" (exceto por vértices isolados) impactam sobre os demais.

1.5.1 Redes sociais de pesquisadores

Jackson e Wolinsky (1996) desenvolvem um modelo que pode ser inserido em um contexto de colaboração acadêmica. No modelo, suponha que existam n pesquisadores e que cada indivíduo possui uma unidade de tempo que pode ser alocada entre diferentes projetos de

pesquisa. Cada pesquisador dedica o mesmo número de horas a cada um de seus projetos de pesquisa e cada projeto envolve a colaboração de até dois pesquisadores. Assume-se que o valor que um indivíduo recebe de um dado projeto é igual à soma das horas dedicadas por ele, pelo seu colaborador e por um termo de sinergia, igual ao produto das horas dedicadas por eles. Assim, o payoff total recebido por um pesquisador i inserido em uma rede social de colaboração com matriz de adjacência $\mathbf{g}$ é dada por:

$$u_i(\mathbf{g}) = \sum_{j:ij\in\mathbf{g}} \left(\frac{1}{d_i(\mathbf{g})} + \frac{1}{d_j(\mathbf{g})} + \frac{1}{d_i(\mathbf{g})d_j(\mathbf{g})} \right), \tag{1.23}$$

para $d_i(\mathbf{g}) > 0$ e, se $d_i(\mathbf{g}) = 0$, $u_i(\mathbf{g}) = 1$. Não há custos associados à criação ou manutenção dos elos de coautoria. Entre outros resultados, é possível mostrar que: (i) se n é par, então o grafo eficiente consite de $n/2$ pares separados e (ii) se $n \geq 4$, então o grafo de equilíbrio é ineficiente e pode ser dividido em cliques, cada uma com um número diferente de indivíduos.

1.5.2 Crime

Em Baccara e Bar-Issac (2008) estuda-se a estruturação ótima de grupos criminosos. Neste modelo, a organização criminosa depara-se com o *trade-off* de por um lado promover a cooperação entre seus membros e por outro minimizar a sua vulnerabilidade à investigação policial. No modelo, cooperação é crescentemente facilitada à medida que a informação sobre os seus membros é compartilhada entre os elementos da organização. Quando se sabe dos antecedentes criminiais de um determinado indivíduo, este passa a cooperar porque pode ser punido pelos agentes que conhecem as suas informações privadas. Porém, de acordo com o modelo, se um membro da organização é capturado pela polícia, então ele revela as informações que possui acerca dos outros membros que ele conhece e a polícia, de posse deste conhecimento, captura os membros delatados. Desta forma, percebe-se que o compartilhamento da informação sobre os antecedentes criminais dos membros da organização a torna mais vulnerável. A resolução deste *trade-off* depende da estratégia de investigação da polícia e produz a estrutura ótima da rede de criminosos.

1.5.3 Organização industrial

As firmas relacionam-se de diferentes maneiras. Não apenas elas compram e vendem produtos entre si ao longo da cadeia produtiva que conecta demandantes e fornecedores, mas também elas muitas vezes se envolvem em relacionamentos mais sofisticados como de colaboração em pesquisa e desenvolvimento. Nestas condições, a pesquisa em organização industrial pode beneficiar-se bastante com a incorporação de grafos aos modelos. Com o objetivo de ilustrar as novas intuições que podem ser conquistadas por esta incoporação, abaixo se discutem dois modelos básicos de competição oligopolística, Cournot e Bertrand, em que conexões com outras empresas reduzem o custo marginal de cada firma. Estes modelos clássicos modificados foram propostos por Goyal e Joshi (2003)[16].

1.5.3.1 Modelo de Cournot

Suponha que o custo marginal da firma i seja dado por $c_i(\mathbf{g}) = \gamma_0 - \gamma n_i(\mathbf{g})$, onde $n_i(\mathbf{g}) = |N_i(\mathbf{g})|$. Ou seja, o custo marginal decresce linearmente no número de firmas com as quais a firma i mantém algum relacionamento de colaboração. Assuma também que $\gamma_0 > \gamma(n-1)$, onde n é o número de firmas, para garantir que o custo marginal é sempre positivo. Admita que as firmas deparam-se com uma função de demanda inversa dada por $p = \alpha - \sum_{i=1}^{n} q_i$, $\alpha > 0$, onde q_i denota a quantidade produzida pela firma i. Então, para α suficientemente grande, é fácil ver que no equilíbrio de Cournot, cada firma produz

$$q_i^*(\mathbf{g}) = \frac{1}{n+1}\left\{\alpha - \gamma_0 + \gamma\left[n_i(\mathbf{g})\,n - \sum_{j\neq i} n_j(\mathbf{g})\right]\right\}, \tag{1.24}$$

o que lhe rende um lucro de $[q_i^*(\mathbf{g})]^2$. Sendo assim, se o custo de estabelecimento de elos de colaboração é suficientemente baixo, deve-se esperar uma rede empresarial completa, em que cada firma colabora com todas as demais.

[16] Foi criado recentemente na internet um sítio de relacionamento entre empresas brasileiras. Este sítio chama-se Emprelink (www.emprelink.com.br) e já conta com milhares de empresas cadastradas. Em princípio, esta pode ser uma boa fonte de dados para trabalhos empíricos em organização industrial que incorporem redes de relacionamento entre firmas.

1.5.3.2 Modelo de Bertrand

No modelo de Bertrand, se uma firma tiver custo marginal menor que todas as demais, então ela fixará seu preço em um nível inferior ao custo marginal da segunda firma mais produtiva (de menor custo marginal) de tal forma a maximizar o seu lucro. Se mais de uma firma tiver o menor custo marginal, então estas escolherão seus preços no nível do custo marginal comum e assim dividirão o mercado, obtendo lucro zero. Agora, se o custo de estabelecer uma parceria é positivo, então as únicas firmas que podem estar dispostas a fazer isto são aquelas que possuem lucros positivos. Porém, no equilíbrio de Bertrand, no máximo uma firma pode ter lucro maior do que zero. Consequentemente, se as firmas competem de acordo com o modelo de Bertrand, a única rede de colaborações empresariais que se pode observar é descrita por um grafo vazio, sem nenhuma conexão.

1.6 Importância das interações intra-redes sociais

Nesta seção, discutem-se alguns trabalhos que procuram mostrar como a estrutura das redes sociais, tomadas parametricamente, afeta diversos resultados de interesse.

1.6.1 Educação: identidade e *peer effects*

Um dos maiores desafios para a promoção da educação é entender porque muitas vezes se observam grandes diferenças em resultados educacionais apesar de pequenas diferenças na quantidade de recursos empregados. Pesquisas recentes indicam que conceitos da sociologia conciliados a métodos usados em economia e técnicas de teoria dos grafos são promissores no sentido de oferecer respostas bem estruturadas para este questionamento e, assim, ajudar na formulação de melhores políticas educacionais.

Em educação, existe ampla literatura empírica que associa o desempenho de cada aluno individual à performance média de seus colegas de classe, aqueles com os quais ele se relaciona mais intensamente no âmbito escolar. Um dos objetivos desta linha de pesquisa é contribuir com o debate relativo ao efeito sobre o aprendizado decorrente de políticas de agrupamento de estudantes com desempenho acadêmico similar. Os defensores destas políticas argumentam que, ao separar os melhores alunos daqueles com maiores dificuldades,

há ganhos de performance acadêmica para todos porque os professores, ao se depararem com turmas mais homogêneas, são capazes de ajustar a metodologia e o material de ensino de tal forma a atender melhor as necessidades de cada grupo. Por outro lado, os críticos destas políticas de agrupamento dizem que, ao separar os estudantes em turmas distintas com base na performance acadêmica observada, perdem-se os chamados *peer effects*, segundo os quais os piores alunos beneficiam-se da convivência com aqueles que estão no topo da distribuição de notas. Adicionalmente, os críticos argumentam que a separação de estudantes pode ser prejudicial à auto-estima daqueles com mais dificuldades de aprendizado. Sendo assim, o agrupamento de alunos poderia beneficiar os estudantes mais avançados em detrimento daqueles mais atrasados.

Embora sujeita a alguns questionamentos, a evidência empírica indica que o agrupamento de estudantes possui resultados positivos. Duflo et al. (2008), a partir de um experimento aleatório realizado com estudantes primários do Quênia, encontram que o agrupamento de estudantes é benéfico para todos, indicando que classes mais homogêneas são mais fáceis de serem ensinadas. Porém, os autores também concluem que existem *peer effects* significantes, pois, nas escolas nas quais a política de agrupamento não foi implantada, os alunos que foram alocados aleatoriamente a turmas compostas por colegas com desempenho acadêmico inicial elevado apresentaram ganhos suficientemente maiores do que aqueles que ficaram em turmas mais fracas.

Calvó-Armengol et al. (2010) avançam no estudo de *peer effects* em relação aos demais trabalhos existentes ao incorporarem a topologia da rede social de estudantes explicitamente à análise. Diferentemente da maioria dos artigos encontrados, que simplesmente associam o desempenho escolar de cada aluno à performance média de seus colegas de classe, os autores permitem que estas influências sejam heterogêneas dentro da sala de aula a depender da posição de cada aluno na rede social de estudantes. De acordo com o modelo desenvolvido, os resultados obtidos por um indivíduo, em um ambiente genérico, são proporcionais à sua medida de centralidade de Katz-Bonacich[17]. Então, com base em uma amostra de estudantes, os autores estimam que o impacto decorrente de um aumento de um desvio padrão naquela medida de centralidade aumenta a performance do aluno em 7% de um desvio padrão das notas.

Vale notar que a família é a primeira rede social da qual um indivíduo participa e que muitos trabalhos reconhecem que ela desempenha um papel decisivo na performance escolar das crianças e, futuramente, no desempenho dos mesmos no mercado de trabalho.

[17] A medida de centralidade de Katz-Bonacich é uma variante da medida de centralidade autovalor, em que a importância de um vértice não depende apenas da importância de seus vizinhos, mas também da importância de outros vértices a ele caminho-conectados a maiores distâncias.

De fato, a evidência empírica demonstra que a base familiar tem um papel valioso na produção de habilidades cognitivas e não-cognitivas das crianças, conforme destacado por Todd e Wolpin (2007) em que se estudam os determinantes dos resultados de exames de alcance cognitivo em crianças. De acordo com os resultados apresentados pelos autores, diferenças no nível de habilidade da mãe são responsáveis por aproximadamente metade das diferenças de desempenho inter-raciais naqueles exames. Além disto, uma conclusão relevante do trabalho é que insumos familiares medidos em termos de indicadores de dedicação materna à educação dos filhos também são significativos, pois ao se igualar tais insumos familiares ao nível médio daqueles recebidos por crianças brancas reduzem-se as diferenças de performance em relação a negros e hispânicos entre 10% e 20%.

O que os estudos discutidos acima mostram é que a qualidade e a estrutura da rede social em que um indivíduo insere-se são importantes para explicar os resultados por ele obtidos, em particular o desempenho escolar. Sendo assim, seria interessante desenvolver uma pesquisa com dados brasileiros que permita identificar estes efeitos de rede entre estudantes já que, até onde se sabe, não existem trabalhos desta natureza. De posse de informações referentes às amizades dos estudantes, seria possível calcular diversos indicadores que descrevem a estrutura desta rede social. Destes indicadores, de particular interesse, são os indicadores de homofilia, os quais permitem verificar se há algum tipo de segregação entre estudantes e em que dimensões. Esta informação é importante porque maiores níveis de segregação devem estar associadas a *peer effects* menores ou restritos por grupos. Afinal, na medida em que a comunicação entre grupos distintos é restrita, a transferência de conhecimentos e de experiências também fica prejudicada. Nestas condições, políticas de agrupamento de estudantes podem ser benéficas, pois, se os estudantes separam-se em grupos espontaneamente, manter todos os grupos em uma mesma sala de aula com o objetivo de explorar *peer effects* tem resultados modestos diante da melhor adequação dos métodos de ensino para cada grupo. Isto porque, se a comunicação e a influência entre grupos é limitada em decorrência, por exemplo, de afirmação de determinadas identidades sociais, então os esperados *peer effects* provenientes dos bons alunos em direção aos alunos mais fracos podem não se realizar em magnitude satisfatória. Seria também possível investigar como estas amizades são formadas. Neste ponto, cabe mostrar como a estrutura da rede observada nos dados emerge dos objetivos sociais e acadêmicos dos estudantes e quais as suas consequências em termos de desempenho escolar. Ao se avançar no entendimento dos objetivos perseguidos pelos estudantes na formação de suas amizades e da maneira pela qual a busca destes objetivos traduz-se em termos de comportamento, em particular em termos de dedicação escolar, pode-se imaginar formas mais eficientes de incentivar os estudantes a dedicarem maior atenção às atividades acadêmicas. Entre estes objetivos está a forma como os jovens desejam inserir-se no círculo social, ou seja, que posição eles desejam ocupar na rede social da qual fazem parte. Pelo fato

de este posicionamento relacionar-se com o desempenho escolar e com outros comportamentos, fica evidente a importância de obter-se informações sobre a rede de amizades de estudantes conjuntamente com informações sobre performance dos mesmos em testes de aprendizado.

Em um artigo que contribui nesta direção, Akerlof e Kranton (2002) argumentam que o propósito principal dos estudantes é a construção de suas identidades. Estes possuem dotações de diferentes características valorizadas diferentemente por cada grupo social existente na escola, tais como: estudiosos, atletas, populares e desinteressados. Cada grupo possui um ideal de comportamento do qual os seus membros procuram aproximar-se ajustando, por exemplo, o nível de dedicação aos estudos. Fryer e Torelli (2010) mostram que nos EUA os estudantes negros e hispânicos, com o intuito de afirmarem as suas respectivas identidades raciais, rejeitam comportamentos associados aos brancos, dentre os quais se relaciona um bom desempenho escolar.

1.6.2 Mercado de trabalho

As redes sociais desenpenham papel de destaque no funcionamento do mercado de trabalho porque as conexões interpessoais funcionam como canais de transmissão de oportunidades de emprego e de reputações, facilitando assim o pareamento entre empregados e empregadores. Esta observação, evidente para qualquer pessoa que já tenha olhado para o mercado de trabalho, é corroborada por extensa evidência empírica. Ioannides e Loury (2004) revisam esta literatura e identificam, entre outros fatos estilizados, que: (i) utilização de amigos, parentes e outros contatos na procura por empregos é generalizada; (ii) a utilização de amigos na procura por empregos é produtiva e (iii) utilização de amigos na procura por empregos frequentemente, mas nem sempre, varia com a localização e com as características demográficas.

Myers e Shultz (1951) obtêm que 62% dos trabalhadores entrevistados por eles encontraram seus primeiros empregos por intermédio de contatos sociais, enquanto que 23% destes trabalhadores encontraram o primeiro emprego por conta própria e os restantes 25% o fizeram através de anúncios ou agências de emprego. Rees e Shultz (1970) mostram que as redes de contato pessoais constituem um canal importante para que trabalhadores de diferentes especialidades encontrem os seus empregos. Os autores encontram que 37,3% dos digitadores, 23,5% dos contadores, 73,8% dos carregadores, 65,5% dos faxineiros e 57,4% dos eletricistas encontraram seus empregos através de contatos sociais. Ou seja, as redes sociais são importantes para o funcionamento do mercado de trabalho para muitas

profissões.

Um atributo importante de um link é a sua intensidade. Em um dos primeiros trabalhos a levar este aspecto em consideração, Granoveter (1973) analisa informação acerca da maneira pela qual as pessoas em diferentes profissões encontraram os seus respectivos empregos. Encontra-se que a ampla maioria dos empregos não foi encontrada através de um contato social forte, com o qual há interação ao menos duas vezes por semana. Ao contrário, 55,7% dos empregos foram encontrados através de relacionamentos de intensidade média, em que há interação ao menos uma vez por ano e menos de duas vezes por semana, e 27,6% foram encontrados por intermédio de links fracos, quando a frequência de interação é de no máximo uma vez por ano. Para explicar este resultado, o autor argumenta que há grande aglomeração entre os indivíduos conectados por links fortes comparativamente aos relacionamentos mais esporádicos. Por este motivo, os links fracos unem grupos que, sem eles, seriam segregados. Desta forma, os links fracos transmitem informação, inclusive de oportunidade de emprego, até distâncias maiores do que os links fortes. Uma outra maneira de analisar a explicação proposta por Granoveter (1973) é através de navegação sobre a rede social. Como o subgrafo formado apenas pelas conexões fortes é altamente aglomerado, é mais fácil que uma informação fique aprisionada em um subconjunto pequeno da população.

A lição imediata que se extrai de Granovetter (1973) é que a posição dos indivíduos em uma dada rede social e a composição de seus contatos em termos de intensidade explicam diferenças na capacidade de obter um emprego. Pessoas mais conectadas, de preferência com elos fracos, tendem a ter mais informação sobre vagas disponíveis de emprego. Outra lição importante é que diferentes estruturas de rede social podem ter consequências diversas sobre, por exemplo, duração do desemprego. Imagine, por exemplo, duas redes sociais com o mesmo número de indivíduos, sendo que a primeira é um grafo conexo, enquanto que a segunda é formada por diversas cliques. Então, pelo fato uma informação sobre uma vaga de emprego poder viajar entre qualquer par de vértices do primeiro grafo, é de se esperar que a duração média do desemprego seja menor se a sociedade for organizada de acordo com o este grafo do que se for organizada de acordo com o segundo. Além disto, pelo fato de o desemprego ser menos duradouro no primeiro caso, os trabalhores possuem maior capacidade de extrair ganhos de comércio nas barganhas bilaterais sobre salário com os empregadores. De fato, Montgomery (1991) encontra que a utilização de contatos sociais no mercado de trabalho tem grande impacto sobre desigualdade salarial.

Em decorrência de vagas de trabalho serem encontradas por intermédio de contatos sociais, existe uma correlação positiva tanto entre emprego quanto entre salário de pessoas conectadas (TOPA, 2001 e WAHBA; ZENOU, 2005). A razão para a associação positiva

entre emprego de pessoas conectadas é que a quantidade de informação sobre oportunidades de emprego que um indivíduo desempregado recebe tende a ser maior quando os seus conhecidos encontram-se empregados do que quando estão desempregados. Disto decorre que a duração do desemprego deve ser correlacionada entre pessoas conectadas, o que fornece uma explicação alternativa para histerese do desemprego, isto é: a crescente probabilidade de um indivíduo continuar desempregado à medida que o tempo de desemprego se prolonga. Um trabalhador que se encontra desempregado tende a ter poucos conhecidos empregados e, consequentemente, menor a probabilidade de que ele seja informado sobre vagas de emprego no futuro.

1.6.3 Bens públicos

Bramoullé e Kranton (2007) desenvolvem um modelo de provisão de bens públicos, cujo principal interesse está nos incentivos individuais de agentes inseridos em redes sociais em inovar ou em gerar novas informações. De acordo com os autores, como a transmissão dos resultados obtidos nestas atividades é frequentemente local, deve-se investigar como a estrutura das redes sociais afeta o nível e a forma de provisão destes bens, assim como o bem-estar da sociedade. O primeiro resultado importante do artigo diz que redes sociais podem levar a especialização. Em qualquer rede social, existe um equilíbrio em que alguns indivíduos empregam esforço para prover o bem-público enquanto os demais não contribuem com nada. Neste modelo, a decisão de um indivíduo entre contribuir para a provisão do bem público e pegar carona depende da sua posição na rede social. Além disto, em toda rede social, o único tipo de equilíbrio estável é o especializado. Finalmente, em muitas redes sociais este tipo de equilíbrio é o único existente. O segundo resultado afirma que a especialização pode gerar ganho de bem-estar em relação a um equilíbrio, necessariamente instável, em que os custos incorridos na inovação são compartilhados entre todos os indivíduos da rede. O terceiro resultado do trabalho é que o acréscimo de uma nova conexão à rede possui consequências de bem-estar. Por um lado, esta nova conexão é benéfica por aumentar o acesso a uma dada quantidade de bem público. Por outro, ela altera os incentivos que cada agente tem em contribuir para a sua oferta. Desta forma, o bem-estar geral pode ser maior em redes socias desconexas, ou seja, redes sociais em que não existe ao menos um caminho ligando cada indivíduo a qualquer outro indivíduo da rede.

Uma das fragilidades dos modelos em que os agentes possuem informação completa sobre a rede em que irão interagir, como em Bramoullé e Kranton (2007), é a existência de múltiplos equilíbrios. Porém, uma maneira simples de contornar este problema é adicionar

informação incompleta de cada indivíduo com relação à rede na qual ele irá interagir no futuro, conforme demonstrado por Galeotii et al. (2010), em que se estudam dois tipos de relacionamento entre os agentes: substituição estratégica (como no caso de provisão de bens públicos) e complementaridade (como, por exemplo, a escolha sobre a utilização de um editor de texto por coautores).

1.6.4 Compartilhamento de risco

Não é incomum que as pessoas, quando abatidas por choques negativos de renda, recorram à ajuda de indivíduos nos seus círculos de contatos sociais. O uso deste artefato é amplamente documentado sobretudo em vilas rurais remotas, onde o acesso a mecanismos formais de seguro é drasticamente restrito ou inexistente. Nestas localidades, o que os estudos encontram é que, embora arranjos de seguro mútuo sejam frequententemente utilizados, eles são incompletos entre os residentes das vilas. Fafchamps e Lund (2003) encontram que, em localidades rurais filipinas, os indivíduos (unidades familiares) não fazem seguro mútuo ao nível da vila, mas sim com um subconjunto dos seus residentes com os quais cada um deles se relaciona por elos de amizade ou de parentesco.

Em Ambrus, Mobius e Szeidl (2010), a capacidade de compartilhamento de risco depende da expansividade da rede social. Os autores observam que compartilhamento de risco ocorre em nível local, em que as pessoas ajudam seus familiares, amigos e vizinhos próximos. Apesar disto, a evidência empírica mostra que estes arranjos frequentemente atingem um resultado muito próximo de compartilhamento de risco global. Com o objetivo de entender como é possível que obrigações locais de transferências, quando agregadas, gerem compartilhamento de risco praticamente global, desenvolvem um modelo de compartilhamento de risco em redes sociais. Encontra-se que seguro global é difícil de ser obtido porque exige um nível muito alto que conectividade, o qual não é encontrado em redes sociais observadas, mas que é possível obter seguro quase perfeito para redes sociais "menos conexas", tais como as observadas. Adicionalmente, mostra-se que o consumo de um agente é mais fortemente correlacionado com o consumo de agentes socialmente mais próximos.

1.6.5 Difusão

Os indivíduos transmitem informações, opiniões e comportamentos através de suas redes sociais. Assim, um indivíduo pode recomendar aos seus amigos a leitura de um determinado livro ou a adoção de certa tecnologia, pode alertar sobre o surgimento de novos produtos e sugerir uma profissão. Entre pesquisadores, teorias, métodos e temas de pesquisa podem ser disseminados através da comunidade científica. Assim, nesta subseção, tendo em vista a importância do tema, revisam-se alguns trabalhos sobre difusão em redes sociais[18].

Banerjee et al. (2012) estudam como a participação em um programa de microfinanciamento dissemina-se através das redes sociais em vilas no sul da Índia. Os autores encontram que o posicionamento dos indivíduos aos quais o programa é inicialmente introduzido, pontos de injeção, é importante. Mais especificamente, obtém-se que a participação no programa é crescente com a medida de centralidade autovalor destes indivíduos.

Bandiera e Rasul (2006) estudam a decisão de início de cultivo de girassol no norte do Moçambique. Tendo em mente que fazendeiros trocam informações sobre formas de cultivo entre si, os autores procuram estimar a probabilidade de implantação da cultura de girassol por um fazendeiro como função do número de seus familiares e amigos que tomaram esta decisão. Os autores encontram que esta relação tem o formato de "U" invertido e que a decisão de adotar o cultivo de girassol por um fazendeiro é menos sensível ao número de amigos e familiares que tomam a mesma decisão quando o fazendeiro tem melhores informações sobre esta cultura. Estes resultados indicam a existência de aprendizado social entre os fazendeiros. Em trabalho relacionado, Conley e Udry (2010) estudam como a adoção de novas tecnologias no cultivo de abacaxi em Gana. Especificamente, os autores investigam como a utilização de certos insumos (como fertilizantes) por parte de um agricultor é influenciada pelas decisões de seus vizinhos. Os autores encontram evidências de que um fazendeiro procura imitar as práticas adotadas por seus vizinhos que obtiveram resultados muito satisfatórios.

1.6.6 Crime

Segundo Reiss (1980, 1988), 2/3 dos crimes são cometidos com a participação de dois ou mais indivíduos. Glaeser, Sacerdote e Scheinkman (1996) procuram entender a variância dos indicadores de criminalidade entre cidades americanas, a qual segundo os autores não

[18] Uma excelente revisão sobre o tema pode ser encontrada em Jackson e Yariv (2010).

pode ser explicada apenas por características econômicas e sociais locais tais como renda per capita, taxa de desemprego e nível educacional da população. De acordo com os autores não mais do que 30% desta variância entre cidades pode ser atribuída a variações nestes atributos. Sendo assim, para que se consiga explicar as grandes diferenças nas taxas de criminalidade observadas mesmo entre cidades parecidas naquelas caracteríticas, é necessário introduzir correlação positiva no comportamento dos indivíduos. Pelo fato de a decisão de um indivíduo em entrar para o crime poder influenciar a mesma decisão das pessoas com quem ele se relaciona - conjunto este assumido unitário e composto por um vizinho no modelo proposto no trabalho, as taxas de criminalidade de uma cidade podem diferir substancialmente daquelas previstas pelas suas características sócio-econômicas, gerando assim maior variância entre localidades.

1.6.7 Programas sociais

Angelucci et al. (2010) encontram que a estrutura das redes locais em vilas do México influencia diferentemente as respostas das famílias beneficiadas pelo programa de incentivo à educação Progresa, implantado naquela país. De acordo com este artigo, as famílias que estão inseridas em uma rede familiar respondem de forma distinta ao programa daquelas unidades familiares isoladas de seus parentes, pois apenas as crianças das primeiras aumentam o atendimento do nível escolar secundário. O motivo para esta resposta positiva das famílias em redes familiares é a redistribuição de recursos entre as famílias de uma mesma rede.

1.6.8 Finanças

A última crise financeira mundial deixou mais uma vez claro o elevado grau de interdependência entre as instituições atuantes neste mercado. Os bancos, por exemplo, estão conectados diretamente uns aos outros pelo mercado interbancário e indiretamente pelos seus depositantes comuns. Assim, o sistema financeiro pode ser convenientemente representado por um grafo ponderado, onde a capacidade[19] de uma aresta conectando duas instituições financeiras mede o valor monetário da relação de crédito entre elas. Naturalmente, a estrutura desta rede financeira é relevante para avaliar o risco sistêmico, ou seja: a vulnerabilidade do sistema financeiro à propagação de choques localizados em algumas instituições individuais (ALLEN; BABUS, 2007). No contexto do mercado interbancário,

[19] Valor atribuído a uma aresta.

para o qual existem dados (embora não sejam públicos), seria interessante entender como a dispersão das taxas de juros cobradas e os valores negociados dependem da posição das instituições envolvidas na rede.

1.6.9 Governança corporativa

Um dos campos nos quais há grande interesse pelas redes sociais é o da governança corporativa. Aqui se deseja entender como grupos de trabalho e corpos diretivos são formados e quais as implicações sobre variáveis de interesse tais como: performance e fraudes contábeis.

Gompers, Mukharlyamov and Xuan (2012) encontram que a parceria entre investidores de risco é fortemente homofílica tanto em termos de habilidade (por exemplo, se os indivíduos formaram-se em uma universidade de grande prestígio) quanto em termos de afinidade (por exempo, se os individuos pertencem a um mesmo grupo étnico, estudaram na mesma universidade ou trabalharam juntos em uma firma comum). Os autores encontram que existem diferenças importantes entre as parcerias por habilidade e as parcerias por afinidade. Enquanto as primeiras aumentam a probabilidade de sucesso do investimento, as segundas reduzem-na. Além disto, esta relação entre os tipos de homofilia e a performance do investimento resulta das escolhas feitas pelos investidores e não por seleção em que as parcerias fundadas em habilidade concentram-se nos projetos mais promissores relativamente às parecerias fundadas em afinidade.

Cohen, Frazzini e Malloy (2008) estudam como as decisões de investimento em ações por gestores de fundos mútuos são influenciadas pelas suas conexões educacionais com executivos das empresas de capital aberto. Os resultados revelam que os gestores de fundos alocam mais recursos em firmas cujos executivos estão a eles conectados na rede educacional do que em firmas com as quais os gestores não estão conectados através desta rede e que este incremento na participação de ações conectadas no portfolio é crescente com a intensidade da conexão educacional entre um gestor e um executivo. Além disto, os retornos dos investimentos realizados nas firmas com as quais os gestores de fundos possuem elo na rede educacional são significantemente maiores do que os retornos dos investimentos realizados nas firmas com as quais eles não estão conectados. De acordo com as evidências, estes resultados devem-se a fluxo de informação privilegiada dentro da rede educacional.

Com a série de fraudes documentadas no início dos anos 2000, algumas medidas foram

tomadas com o propósito de evitar que elas se repitam. Estas medidas enfatizam independência econômica entre o diretores do conselho e o CEO. Chindambaran, Kedia e Prabhala (2012) estudam como diferentes tipos de conexões entre o CEO de uma empresa e os membros externos do conselho de administração impactam sobre a incidência de fraudes financeiras. Especificamente, investigam-se dois tipos de conexões: as de origem profissional e as de origem não-profissional. Os links profissionais são aqueles criados no ambiente de trabalho pelo serviço contemporâneo em alguma empresa. Já por elos não-profissionais entendem-se as interações em organizações sem fins lucrativos, em instituições de caridade, em clubes ou o compartilhamento de uma mesma alma mater. Os autores encontram que, enquanto as conexões não-profissionais entre o CEO e os diretores aumentam a probabilidade de fraudes, as conexões profissionais reduzem-na. Argumenta-se que os dois tipos de conexões possuem efeitos diferentes sobre a ocorrência de fraudes porque, na medida em que eles oferecem diferentes conhecimentos a respeito do CEO, eles dotam os diretores com capacidade diferenciada de monitoramento. Estes resultados são significantes estatisticamente e economicamente. De fato, os impactos destes links são maiores do que de outras variáveis apontadas na literatura de governança corporativa, inclusive são mais importantes do que a independência econômica entre os diretores e o CEO, que tem sido o foco principal das recentes regulações impostas após os diversos casos de fraudes verificadas no início dos anos 2000. Em trabalho relacionado, Black et al. (2013) estudam como a intensidade do relacionamento, medida pelo tempo de convivência, entre o CEO e o CFO de empresas relaciona-se com a probabilidade de fraudes contábeis. Resultados preliminares sugerem que esta relação é negativa, ou seja, quanto maior o tempo de relacionamento, menor a probabilidade de má conduta.

1.7 Identificação de influências intra-redes sociais

A agenda de pesquisa em interações sociais não se restringe à descrição da estrutura das diversas redes de relacionamento e a modelos teóricos que relacionam topologia de rede a comportamento. Além destas duas frentes de trabalho, há uma importante agenda de pesquisa econométrica que é muito bem resumida por Blume e Durlauf (2005) na seguinte passagem:

> Um terceiro componente no novo programa de pesquisa tem sido a investigação sistemática de questões econométricas. Este trabalho econométrico foca primariamente na determinação das condições sob as quais vários tipos de interações sociais podem ser econometricamente identificadas. Argumentos de identificação neste contexto equivalem a perguntar sob quais

> condições de dados e modelo pode o papel dos efeitos de interações sociais ser distinguido de outras influências sobre o comportamento.[20]

Para entender a dificuldade de identificação das interações sociais, abaixo se listam as três maneiras pelas quais se observam comportamento comuns entre indivíduos de um mesmo grupo:

1. Efeitos endógenos: o comportamento médio do grupo ao qual um indivíduo pertence afeta as suas escolhas;
2. Efeitos exógenos (ou efeitos contextuais): o comportamento de um indivíduo varia com características predeterminadas do grupo em relação a decisões individuais sobre este comportamento;
3. Efeitos correlacionados: indivíduos de um mesmo grupo comportam-se de maneira similar porque possuem características individuais comuns ou estão inseridos no mesmo ambiente institucional, independentemente de quaisquer influências entre eles. Um caso particular de efeitos correlacionados ocorre quando o próprio comportamento de interesse define o grupo.

Para perceber a diferença entre os três efeitos, analise o caso em que se estudam os determinantes do desempenho escolar de estudantes. Diz-se que há efeitos endógenos se a performance de um aluno influencia a performance dos demais. Note que, se efeitos endógenos estão presentes, conforme destacado por Manski (1993), surge o que o autor chama de "multiplicador social" porque o desempenho de um aluno afeta os desempenhos dos demais e estes influenciam de volta a performance do primeiro. Verificam-se efeitos exógenos quando o desempenho dos estudantes varia conforme a profissão de seus pais. O caso de efeitos correlacionados é observado quando alunos agrupam-se de acordo com habilidade. Nesta situação, a correlação positiva entre notas de amigos deve-se a homofilia na formação de elos de amizade e não a influências recíprocas, como quando há efeitos endógenos.

O problema do reflexo, conforme colocado por Manski (1993, 2000), é que os dados não permitem distinguir se o comportamento do grupo influencia o comportamento de cada um de seus membros ou se o comportamento do grupo é simplesmente a agregação do

[20] *A third component of the new social interactions research program has been the systematic investigation of econometric issues. This econometric work primarily focuses on the determination of conditions under which various types of social interactions may be econometrically identified. Identification arguments in this context amount to asking under what conditions on data and model can the role of social interactions effects be distinguished from other influences on behavior.*

comportamentos individuais. Em outras palavras, é difícil separar efeitos endógenos de efeitos exógenos e de efeitos correlacionados. Metaforicamente, o problema do reflexo assemelha-se à dificuldade em interpretar o movimento de uma pessoa e de sua imagem refletida em um espelho. Será que o espelho simplesmente reflete os movimentos da pessoa ou será que a imagem no espelho também influencia os seus movimentos?

Recentemente, importantes avanços nesta linha de pesquisa têm sido obtidos. De particular interesse para os propósitos do estudo das redes sociais é a conclusão de Bramoullé, Djebbari e Fortin (2009), segundo a qual, uma vez que se relaxa a hipótese de que as interações sociais ocorrem em grupos, entre todos os indivíduos, o problema do reflexo desaparece. Em outras palavras, o problema do reflexo existe quando as interações sociais ocorrem em cliques. Porém, na realidade, as redes sociais têm estruturas muito mais complexas. Assim, levando-se em consideração a natureza local das interações sociais, genericamente é possível identificar as influências intra-rede.

1.8 Considerações finais

O objetivo desta introdução é revisar a extensa literatura dedicada às redes sociais, a qual tem-se expandido rapidamente nos últimos anos em diferentes áreas do conhecimento, com especial destaque em economia. São inúmeras as situações em que a estrutura das conexões sociais desempenha papel relevante. Alguns exemplos aqui discutidos ressaltam a indicação a vagas de emprego, *peer effects*, educação, associação ao crime, adoção de tecnologia, provisão de bens públicos, compartilhamento de risco e vulnerabilidade do sistema financeiro. De fato, a quantidade de temas contemplados por esta ampla linha de pesquisa, sugere fortemente que o estudo das redes sociais e econômicas constitui um objeto de investigação para o desenvolvimento de uma extensa agenda de pesquisa por muitos anos adiante. As possibilidades para trabalhos futuros não se devem apenas à variedade de ambientes que podem ser estudados. Do ponto de vista teórico há muitas questões a serem exploradas com relação a modelos de formação de rede e modelos de interação intra-rede. Diversos resultados novos têm sido obtidos, oferecendo novas perspectivas a questões antigas e permitindo a compreensão de fenômenos para os quais outras abordagens mais tradicionais não são meios adequados para análise. Do ponto de vista empírico, ainda há importantes desafios a serem superados quanto a estratégias de identificação de parâmetros e quanto a organização de dados. Nos artigos que se seguem, o autor desta tese dá seus primeiros passos dentro deste campo de pesquisa.

O primeiro destes artigos mostra como a incidência e a extensão de coautoria têm evoluído em essencialmente todas as áreas do conhecimento nas últimas quatro décadas. Este artigo também estuda os determinantes da coautoria. No segundo artigo, obtém-se a evolução temporal da topologia das redes de coautoria entre pesquisadores. Calculam-se diversos indicadores de estrutura e atenção especial é dedicada à estimação da distribuição de grau dos autores. Sabendo que redes sociais de coautoria não são grafos Erdös-Rényi, o capítulo subsequente, terceiro artigo, procura entender melhor os mecanismos de interação social subjacentes à formação destas redes de colaboração acadêmica. Especificamente, mostra-se que informações sobre características individuais dos pesquisadores fluem ao longo das conexões de algumas destas redes facilitando o pareamento entre indivíduos de tipos identificados (*screened*). Por fim, são elaboradas algumas últimas considerações.

2 TENDÊNCIAS E DETERMINANTES DA COAUTORIA DE ARTIGOS ACADÊMICOS

2.1 Introdução

A produção de artigos acadêmicos não é uma atividade solitária. Ao contrário, no desenvolvimento de um projeto de pesquisa, um indivíduo beneficia-se da colaboração de seus pares pela troca informal de ideias, pela discussão em seminários, pelos pareceres recebidos e pela formação de coautorias. Esta característica colaborativa do trabalho científico vem intensificando-se ao longo das últimas décadas na forma de coautoria em diversas áreas do conhecimento, conforme inicialmente destacado por Price (1963) e confirmado por inúmeras contribuições mais recentes (BABCHUK; KEITH; PETERS, 1999; BARNETT; AULT; KASERMAN, 1988; HUDSON, 1996; LABAND; TOLLISON, 2000; MCDOWELL; MELVIN, 1983; MOODY, 2004; PIETTE; ROSS, 1992 e PRESSER, 1980).

Algumas explicações para este movimento em direção à coautoria têm sido aventadas. Uma delas sugere que o aumento do conhecimento, por gerar a necessidade de especialização, incentiva o trabalho em equipe (JONES, 2009). Associado ao aumento do conhecimento, o crescimento do número de pesquisadores também impulsiona a coautoria por elevar a oferta de potenciais colaboradores (MCDOWELL; MELVIN, 1983; BARNETT; AULT; KASERMAN, 1988; HUDSON, 1996). Uma outra hipótese, levantada por Barnett, Ault e Kaserman (1988), é que o custo de oportunidade do tempo de um pesquisador tem crescido pela maior ênfase em publicação para efeitos de remuneração e evolução na carreira. Esta política, segundo os autores, tem deslocado a forma de colaboração entre os pesquisadores. Enquanto outrora um pesquisador contava com revisões e comentários cuidadosos de seus colegas antes de submeter um determinado artigo para publicação, mais recentemente, este tipo de colaboração informal, que seria reconhecida por um agradecimento, tem sido substituído por coautoria. Os autores também argumentam que, com o aumento no número de artigos submetidos e com o aumento do custo de oportunidade do tempo dedicado pelos pareceristas dos periódicos, estes dedicam-se menos à revisão minuciosa de cada artigo a eles solicitada análise. Isto, por sua vez, eleva a incerteza dos autores destes artigos quanto às perspectivas de eventual publicação. Uma resposta possível para esta maior incerteza consiste em produzir mais artigos, o que pode ser alcançado via coautoria. Hudson (1996) considera a hipótese de que, em economia,

o crescimento da coautoria pode estar associado a um suposto favorecimento de projetos de pesquisa colaborativos por algumas fontes de financiamento.

Embora a tendência de aumento da incidência e da extensão da coautoria (LABAND; TOLLISON, 2000) seja generalizada, o envolvimento neste tipo de colaboração varia entre campos do conhecimento. A evidência mostra que há maior proporção de artigos coautorados e maior número de autores por artigo em campos de característica mais empírica ou quantitativa (HUDSON, 1996; MOODY, 2004). Segundo Hudson (1996), um dos motivos para o movimento dos economistas em direção à coautoria é o caráter crescentemente quantitativo da profissão. Para sustentar este argumento, o autor mostra que, dos periódicos por ele considerados, os dois mais quantitativos[1] apresentam maior proporção de artigos publicados em coautoria enquanto que a revista menos técnica e quantitativa[2] possui uma proporção bastante inferior de artigos coautorados.

Ao longo do tempo, os avanços da informática e das telecomunicações têm intensificado este diferencial entre trabalhos teóricos e trabalhos quantitativos. O aumento da capacidade dos computadores permite que trabalhos cada vez mais técnicos e intensivos em computação de dados sejam realizados e, como argumentado, estes trabalhos são mais frequentemente realizados em coautoria. Os desenvolvimentos das telecomunicações permitem que coautores comuniquem-se e troquem arquivos eletrônicos com facilidade. McDowell e Melvin (1983) propõem que quanto mais rápida a obsolescência de um campo de estudo, maior a incidência de coautoria. O mecanismo sugerido para esta relação é que, em um campo de estudo em rápida transformação, ser um generalista é muito custoso, pois as perdas de capital humano são muito elevadas. Nestas condições, é mais vantajoso especializar-se em pequenas subáreas e combinar os esforços de diferentes especialistas para a produção de novos conhecimentos.

Ao nível do autor, existem considerações de ciclo de vida que ajudam a entender o envolvimento em coautoria. Pelo fato de publicações em coautoria terem seus valores descontados relativamente a publicações individuais (SAUER, 1988), os pesquisadores mais jovens, para quem a necessidade de construir uma reputação é mais intensa para efeitos de contratação e promoção na carreira, teriam menos incentivos a trabalharem em equipes (MCDOWELL; MELVIN, 1983; PIETTE; ROSS, 1992). Outra observação é que, quando um pesquisador atua em um campo diferente daquele de sua especialidade, é mais provável que ele trabalhe em parceria (MCDOWELL; MELVIN, 1983; PIETTE; ROSS, 1992), o que se relaciona com a hipótese de Hudson (1996) de que a aumento da abrangência dos temas estudados em economia (MANSKI, 2000) ajuda a explicar o aumento da coautoria

[1] Review of Economics and Statistics e Econometrica.
[2] Economic History Review.

nesta disciplina.

Um questionamento imediato diz respeito à relação entre coautoria e a qualidade das publicações. A este respeito as evidências são contraditórias. Barnett, Ault e Kaserman (1988) e Presser (1980) encontram que trabalhos coautorados têm maior qualidade. Já Piette e Ross (1992) concluem que a qualidade de um artigo, medida pelo número de citações, é negativamente relacionada com coautoria.

Este artigo contribui para esta literatura sobre tendências e determinantes da coautoria. Utilizando informações dos currículos Lattes de pesquisadores relacionados à Universidade de São Paulo, este artigo caracteriza a evolução da incidência e da extensão da coautoria ao longo das últimas quatro décadas em essencialmente todos os campos de investigação acadêmica. Em seguida, exploram-se os determinantes da coautoria ao nível do artigo e ao nível do autor. Combinando os dados extraídos dos currículos Lattes com as notas atribuídas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) às revistas acadêmicas, procura-se oferecer novas evidências acerca da relação entre coautoria e qualidade das publicações.

Uma característica distintiva deste artigo é que as informações são obtidas diretamente dos currículos dos pesquisadores. Por este motivo, diferentemente dos trabalhos existentes na literatura, que limitam a investigação arbitrariamente a alguns periódicos selecionados, no presente artigo tem-se informação de todas as publicações realizadas pelos pesquisadores da amostra, o que torna a análise robusta a viés de seleção nos periódicos. Até onde se sabe, este é o primeiro artigo que estuda todas as principais áreas do conhecimento sem impor quais publicações devem ser consideradas.

O restante deste artigo está organizado da seguinte maneira. Na seção 2.2 descrevem-se os dados. Na seção 2.3 mostra-se como evoluem a incidência e a intensidade de coautoria nos diversos campos do conhecimento. A seção 2.4 descreve a estratégia econométrica e as variáveis utilizadas nas estimações. Na seção 2.5 apresentam-se as estatísticas descritivas das variáveis. A seção 2.6 discute os resultados econométricos. Por fim, a seção 2.7 é dedicada às conclusões.

2.2 Dados

A Plataforma Lattes é um sistema concebido e disponibilizado publicamente pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico[3] (CNPq) que coleta em uma única estrutura dados de currículos de pesquisadores, de grupos de pesquisa e de instituições envolvidas em ciência e tecnologia. O currículo Lattes de um pesquisador possui informações detalhadas sobre essencialmente todos os aspectos relevantes de sua atuação acadêmica. Atualmente, a Plataforma Lattes conta mais de um milhão e setecentos mil currículos[4], sendo que as informações nela contidas são utilizadas amplamente por diversos órgãos para fins de planejamento, gestão e fomento das atividades científica e de inovação no Brasil. Trata-se de um sistema muito bem sucedido e plenamente incorporado pela comunidade acadêmica brasileira, conforme destacado por Lane (2010) no trecho abaixo:

> A experiência brasileira com a base de dados Lattes (http://lattes.cnpq.br/english) é um exemplo poderoso de boa política. Ela provê dados de alta qualidade sobre cerca de 1,6 milhão de pesquisadores e sobre cerca de 4.000 instituições. A agência nacional do Brasil para financiamento reconheceu no final da década de 1990 que ela precisava de uma nova abordagem para avaliar as credenciais dos pesquisadores. Primeiro, ela desenvolveu uma 'comunidade virtual' de agências federais e pesquisadores para desenhar e desenvolver a infraestrutura Lattes. Em segundo lugar, ela criou os incentivos apropriados para que os pesquisadores e as instituições acadêmicas utilizem a base de dados: os dados são usados pela agência federal na tomada de decisões de financiamento e pelas universidades em suas decisões de estabilidade e promoção. Em terceiro lugar, ela estabeleceu um sistema de identificação única dos pesquisadores que garante que pessoas com nomes parecidos sejam consideradas corretamente. O resutado é uma das bases de dados de pesquisadores mais limpas existentes.[5]

Pelo fato de os currículos Lattes serem utilizados para avaliação de pesquisadores e de instituições, existem incentivos para que os pesquisadores atualizem seus currículos regularmente. Em particular, por ocasião da avaliação trienal dos programas de pós-graduação realizada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), os membros dos corpos docentes das universidades são estimulados a atualizarem os seus

[3] http://www.cnpq.br/

[4] Em 02/12/2010, a Plataforma Lattes continha 1.779.778 currículos.

[5] The Brazilian experience with the Lattes Database (http://lattes.cnpq.br/english) is a powerful example of good practice. This provides high-quality data on about 1.6 million researchers and about 4,000 institutions. Brazil's national funding agency recognized in the late 1990s that it needed a new approach to assessing the credentials of researchers. First, it developed a 'virtual community' of federal agencies and researchers to design and develop the Lattes infrastructure. Second, it created appropriate incentives for researchers and academic institutions to use the database: the data are referred to by the federal agency when making funding decisions, and by universities in deciding tenure and promotion. Third, it established a unique researcher identification system to ensure that people with similar names are credited correctly. The result is one of the cleanest researcher databases in existence.

currículos.

Dos currículos Lattes, utiliza-se o subconjunto Lattes/USP, composto pelos currículos Lattes dos indivíduos que citam a Universidade de São Paulo em pelo menos uma das seguintes três condições: (1) endereço profissional, (2) formação acadêmica/titulação ou (3) atuação profissional. Todos os indivíduos que indicam a Universidade de São Paulo em pelo menos uma destas três opções no currículo Lattes têm seus currículos carregados automaticamente para a base de dados utilizada, totalizando 62.516 currículos, dos quais são considerados apenas aqueles que possuem ao menos um artigo publicado. Faz-se este imposição para efeito de comparação dos resultados obtidos neste trabalho com os resultados encontrados em outras pesquisas sobre o tema, as quais, em sua ampla maioria, utilizam informações de cautoria provenientes de periódicos (BABCHUK; KEITH; PETERS, 1999; BARNETT; AULT; KASERMAN, 1988; HUDSON, 1996; LABAND; TOLLISON, 2000; MCDOWELL; MELVIN, 1983; MOODY, 2004; PIETTE; ROSS, 1992 e PRESSER, 1980). Sendo assim, na literatura, todos os indivíduos considerados possuem ao menos uma publicação. Por outro lado, este trabalho diferencia-se dos demais em um aspecto importante. Pelo fato de contar com uma amostra de currículos de pesquisadores, os resultados aqui encontrados não são afetados pela escolha específica do conjunto de revistas, afinal a composição da equipe de pesquisadores não é independente da qualidade daquelas. O período analisado compreende os anos de 1971 até 2009.

Seguindo a organização dos currículos Lattes, os campos do conhecimento são agregados em quatro níveis, definidos pela CAPES[6] literalmente da seguinte maneira:

> 1º nível - grande área: aglomeração de diversas áreas do conhecimento em virtude da afinidade de seus objetos, métodos cognitivos e recursos instrumentais refletindo contextos sociopolíticos específicos.
>
> 2º nível - área: conjunto de conhecimentos inter-relacionados, coletivamente construído, reunido segundo a natureza do objeto de investigação com finalidades de ensino, pesquisa e aplicações práticas.
>
> 3º nível - subárea: segmentação da área do conhecimento estabelecida em função do objeto de estudo e de procedimentos metodológicos reconhecidos e amplamente utilizados.
>
> 4º nível - especialidade: caracterização temática da atividade de pesquisa e ensino. Uma mesma especialidade pode ser enquadrada em diferentes grandes áreas, áreas e subáreas.

[6]Ver http://capes.gov.br/avaliacao/tabela-de-areas-de-conhecimento.

Nas análises a seguir, os artigos são classificados de acordo com as áreas do conhecimento à qual pertencem conforme reportadas pelo pesquisador. Alguns artigos contribuem para mais de uma área do conhecimento. Nestes casos eles são incluídos ao conjunto de artigos de cada uma destas áreas do conhecimento.

2.3 Evolução da incidência e da extensão da coautoria

2.3.1 Grandes áreas do conhecimento

Como discutido anteriormente, existe ampla evidência de que a produção acadêmica tem-se tornado cada vez mais colaborativa, com uma crescente proporção de artigos produzidos em coautoria e com número médio de autores por artigo também em crescimento. Nesta seção, mostra-se como a colaboração acadêmica entre pesquisadores contidos na base de dados Lattes/USP tem evoluído[7]. No apêndice 3 estão as tabelas que reportam a trajetória ao longo do tempo do percentual de artigos publicados com a colaboração formal de dois ou mais pesquisadores e no apêndice 4 as tabelas de número médio de autores por artigo nas nove grandes áreas do conhecimento e nas suas respectivas áreas[8].

Na tabela 2.1, apresenta-se a evolução da incidência de coautoria nas nove grandes áreas do conhecimento. Nela é possível notar que em algumas grandes áreas do conhecimento o percentual de artigos produzidos por mais de um autor é muito maior do que em outras. Historicamente, a incidência de coautoria nas ciências agrárias, ciências biológicas, ciências exatas e da terra, ciências da saúde e engenharias é tipicamente superior a 70%, enquanto que em linguística, letras e artes, este percentual nunca ultrapassou os 30%. Já nas ciências humanas e nas ciências sociais aplicadas, a fração de artigos coautorados não alcançou os 50% até meados da década de 2000. Outro fato que deve ser destacado é a tendência de crescimento da incidência de coautoria em todas as grandes áreas do conhecimento ao longo das quatro décadas, sendo mais evidente a partir do início dos anos 1990 quando o percentual de artigos coautorados passa a crescer quase monotonicamente. No último período nota-se que, com excessão de linguística, letras e artes, todas as grandes áreas

[7]Para reduzir o problema de dupla contagem de artigos, para cada artigo i eliminam-se os artigos $j \neq i$ tais que $0 \leq \frac{lev(i,j)}{max(comp(i),comp(j))} \leq 10\%$, onde $lev\,(i,j)$ é a distância de Levenshtein entre o título do artigo i e o título do artigo j e $comp\,(x)$ é o número de caracteres no título do artigo $x \in \{i,j\}$. Esta distância relativa de Levenshtein é calculada apenas entre artigos pertencentes a um mesmo campo do conhecimento. Para evitar que os artigos fossem eliminados segundo algum padrão, seus títulos foram ordenados de forma aleatória previamente.

[8]Por motivos de formatação, os resultados são apresentados em peridodicidade trienal, porém as tabelas com valores anuais estão disponíveis e podem ser obtidas com o autor sob solicitação.

do conhecimento são amplamente colaborativas no sentido de que a maioria dos artigos publicados possuem ao menos dois autores, sendo que em algumas destas grandes áreas a quase totalidade dos artigos produzidos é coautorada.

Na tabela 2.11 acompanha-se o comportamento intertemporal do número médio de autores por artigo em cada uma das grandes áreas do conhecimento. Analogamente ao que foi verificado na evolução da incidência de coautoria, existem diferenças importantes entre as grandes áreas do conhecimento no que diz respeito à sua extensão. Pela tabela percebe-se que as ciências agrárias, ciências biológicas, ciências exatas e da terra, ciências da saúde e engenharias, além de maior incidência, possuem maior extensão de coautoria. De fato, estes campos de estudo, em todo o período analisado, sempre tiveram ao menos dois autores por artigo em média. O mesmo não se verifica nas ciências humanas, nas ciências sociais aplicadas e em linguística, letras e artes. Nestas grandes áreas do conhecimento, em uma perspectiva histórica, cada artigo em média é obra de um único autor. Todavia, vale notar que, começando em meados da década de 2000, os artigos em ciências humanas e em ciências sociais aplicadas passaram a ter pouco mais de dois autores em média. Ao longo do tempo, as grandes áreas compartilham o fato de que a extensão da coautoria apresenta tendência de crescimento com taxas anuais médias entre 1,17% em linguística, letras e artes e 2,07% em ciências biológicas.

Analisando os padrões de coautoria das grandes áreas do conhecimento, parece haver evidência de que campos de estudos aplicados, com caráter quantitativo ou experimental, são mais propícios à coautoria. Esta hipótese será testada econometricamente mais adiante.

2.3.2 Áreas do conhecimento

Agora, aprofunda-se a análise a um nível de agregação dos campos do conhecimento mais baixo. Procura-se extrair conclusões sobre o comportamento intertemporal das áreas constituintes das grandes áreas do conhecimento. Pela tabela 2.2, observa-se que há homogeneidade nas ciências agrárias com relação à incidência de coautoria, que é bastante elevada em todo o período. A tabela 2.12 deixa claro que as áreas do conhecimento associadas às ciências agrárias também compartilham uma tendência ascendente no número médio de autores por artigo, que, nos últimos anos, é superior a quatro, com destaque para medicina veterinária e para zootecnia, que possuem extensão de coautoria superior a 6 e a 5, respectivamente.

Da mesma forma que nas ciências agrárias, a tabela 2.3 mostra que as áreas do conheci-

mento pertencentes às ciências biológicas caracterizam-se, sem exceção, por grande incidência de coautoria nas quatro décadas. Atualmente, praticamente todos os artigos nas áreas de ciências biológicas resultam de cooperação entre pesquisadores. Similarmente, pela tabela 2.13 percebe-se que a média de autores por artigo é alta, porém oscila em um intervalo entre áreas largo, de aproximandamente 3,1 em zoologia até 7,3 em imunologia.

Nas engenharias não se percebem diferenças significativas de incidência de coautoria entre as áreas (tabela 2.4), que é quase sempre bastante elevada, sendo maior do que 85% no triênio 2007-2009 em todas as áreas. Por outro lado, existe grande variação entre áreas relativamente à extensão da coautoria (tabela 2.14) com mínimo de 2,45 em engenharia de transporte e máximo de 5,42 em engenharia aeroespacial no último triênio.

Pela tabela 2.5 percebe-se que, com relação à participação em coautoria, as ciências exatas e da terra são parecidas com as engenharias no sentido de que o percentual de artigos coautorados é elevado em todas as áreas do conhecimento desde a década de 1980. Outra semelhança é a existência de grande variação no número médio de autores entre elas (tabela 2.15). De fato, enquanto atualmente um artigo em matemática possui em média 2,6 autores, um artigo em astronomia possui 5,31.

Nas ciências humanas não é tão comum a colaboração entre pesquisadores e a incidência de coautoria apresenta grande variação entre áreas, com mínimo de 26% em história e máximo de 74,4% em psicologia (tabela 2.6). A extensão de coautoria é concentrada na faixa de 2 até 2,8, porém as áreas de ciência política, filosofia, história e sociologia em média têm menos de 2 autores por artigo (tabela 2.16).

Com relação a linguística, letras e artes, nota-se nas tabelas 2.7 e 2.17 que tanto a incidência quanto a extensão de coautoria em cada uma das suas três áreas, apesar do crescimento ao longo do tempo, é baixa. Com efeito, um artigo em qualquer uma das três áreas possui em média menos de dois autores e o percentual de trabalhos coautorados é inferior a 40%.

Todas as áreas do conhecimento em ciências da saúde são altamente colaborativas no sentido de que elas possuem alto percentual de artigos coautorados, aproximadamente 100% (tabela 2.8) . Além disto, o número médio de autores por artigo em todas as áreas é elevado, variando entre 3,6 e 5,9 (tabela 2.18).

Nas ciências sociais aplicadas encontra-se maior heterogeneidade entre áreas quanto ao envolvimento em coautoria. Apesar do crescimento no percentual de artigos coautorados comum a todas as áreas, observa-se na tabela 2.9 que direito tem cerca de 28% dos artigos

produzidos em colaboração enquanto que administração tem mais de 90%. De todo modo, pode-se dizer que, no último período analisado, a incidência de coautoria nas áreas que compõem as ciências sociais aplicadas tipicamente se encontra entre 50% e 80%. Por outro lado, verifica-se na tabela 2.19 que a extensão de coautoria nestas áreas atualmente está entre 2 e 3 autores por artigo, mas que existem exceções, como direito, em que há 1,4 autor por artigo. Considerando apenas os padrões de envolvimento em coautoria, pode-se dizer que a área do direito assemelha-se mais às ciências humanas do que às ciências sociais aplicadas.

Os valores obtidos são parecidos com aqueles reportados em outros artigos. Newman (2001), por exemplo, estuda os padrões de coautoria em artigos publicados no período 1995-1999 nas ciências biomédicas, em física e na ciência da computação. Ele encontra que o número médio de autores nas ciências biomédicas é, no período por ele analisado, igual a 3,75. Nas ciências da computação, o número médio de autores por artigo é 2,22. Já em física teórica este valor é 2,53, sendo astrofísica 3,35 e em física de alta energia experimental 8,96.

É inegável que a produção acadêmica tem passado por importantes mudanças ao longo do tempo. Em particular, a crescente colaboração entre autores tem sido amplamente reportada em nível internacional. Nesta seção mostra-se que a academia brasileira não está alheia a esta tendência de crescimento da coautoria, pois tanto a sua incidência quanto a sua extensão vêm apresentando crescimento em todas as áreas do conhecimento. Este é um fenômeno interessante por pelo menos dois motivos. Em primeiro lugar pelo impacto que os desenvolvimentos na área acadêmica têm sobre as condições de vida da pessoas, o que torna a compreensão do funcionamento deste setor de atividade uma necessidade. Em segundo lugar, a formação de parcerias entre pesquisadores pressupõe a tomada de decisões coletivas na sua formação e no seu funcionamento. Estas questões de formação de grupos e de decisões coletivas têm implicações que vão além do funcionamento da academia, estendendo-se a inúmeras situações da interação humana. Sendo assim, do estudo da colaboração entre acadêmicos podem-se extrair lições aplicáveis a um espectro muito maior de ambientes.

No restante deste artigo, pretende-se contribuir em uma análise exploratória dos determinantes da coautoria.

2.4 Estratégia econométrica e definição das variáveis

Nos exercícios econométricos realizados a seguir, deseja-se entender os determinantes da coautoria ao nível do artigo e ao nível do pesquisador. Sendo assim, estimam-se modelos em que a variável dependente é uma *dummy* para ocorrência de colaboração (análise ao nível do artigo) e para participação em coautoria (análise ao nível do pesquisador). Estes modelos com variável dependente binária (*coautoria*) permitem avaliar o impacto de algumas variáveis explicativas sobre a probabilidade de coautoria; de que um artigo seja coautorado (análise ao nível do artigo) ou de que um pesquisador trabalhe em equipe (análise ao nível do pesquisador). Além dos modelos de resposta binária, estimam-se modelos em que a variável dependente é o número de autores de um artigo. Portanto, as variáveis dependentes são: *coutoria* e *número de autores*, esta denotada por *nautores*. Estas variáveis são obtidas com base nas informações referentes a artigos publicados contidas nos currículos dos indivíduos do conjunto de pesquisadores Lattes/USP. Para cada artigo no qual um autor pertencente à base de dados tem participação, calcula-se o número de autores envolvidos; se este valor é maior do que um, então a variável *coautoria* assume valor unitário, caso contrário ela é igual a zero.

Encontra-se na literatura uma grande variedade de hipóteses que pretendem explicar a coautoria, cada uma das quais dando destaque a um conjunto de variáveis explicativas. A base de dados Lattes/USP permite que se obtenham novas evidências quanto à importância destas variáveis. Ao nível do artigo, a primeira variável cujo impacto sobre a coautoria (incidência e intensidade) deseja-se estimar é a área do conhecimento à qual o artigo pertence. Nos currículos Lattes, cada artigo pode ser enquadrado em até três campos de cada um dos quatro níveis de especificidade, ou seja, um artigo pode pertencer a até três grandes áreas do conhecimento, três áreas do conhecimento, três subáreas e três especialidades. Para propósito das estimações, os artigos são separados primeiramente por grandes áreas do conhecimento. No caso de um artigo contribuir para mais de uma grande área do conhecimento, ele é incluído no conjunto de artigos de cada uma destas grandes áreas do conhecimento. Uma vez feita esta separação, ao nível de área do conhecimento, constroem-se variáveis *dummy* para cada uma delas. Ao nível de subárea do conhecimento, calculam-se quatro variáveis binárias relacionadas ao caráter teórico de uma publicação. Especificamente, quando em qualquer uma das três subáreas de um artigo aparece a palavra "aplicada" (no masculino ou no feminino e no singular ou no plural), a variável binária *aplicada* assume valor unitário. Procede-se analogamente para as variáveis *quantitativa*, *empírica* e *experimental*. Já a variável *não-teórico* agrega as quatro variáveis anteriores assumindo valor unitário quando ao menos uma delas é unitária e zero caso contrário. No exercício econométrico realizado adiante, estas variáveis são

utilizadas para testar a hipótese levantada por Hudson (1996) e por Moody (2004) de que trabalhos de cunho não puramente teórico são mais propensos à coautoria e que este viés pró-coautoria tem-se intensificado relativamente às pesquisas teóricas nas últimas décadas, possivelmente em decorrência dos avanços tecnológicos nas áreas de computação e de telecomunicações.

Uma das preocupações deste artigo é fornecer evidências sobre o relacionamento entre coautoria e qualidade das publicações. Para isto, os dados dos currículos Lattes/USP são fundidos com a avaliação Qualis de periódicos elaborada anualmente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Desde a sua implantação, em 1998, a avaliação Qualis de periódico passou por modificações, com variações no conjunto de revistas e com mudanças de metodologia. Enquanto no período de 1998 até 2006 as notas possíveis eram A, B e C com o qualificador de publicação internacional (I), a partir de 2007, os periódicos passaram a ser classificados em oito estratos: A1, A2, B1, B2, B3, B4, B5 e C. A fim de controlar para esta mudança de classificação, as estimações são feitas para o período 2000-2006 e para o período 2007-2008 separadamente. Além disto, em cada um dos dois períodos, com o intuito de controlar para mudanças de critério na avaliação de periódicos entre os anos, foram consideradas apenas as publicações cujas avaliações mantiveram-se estáveis no período 2000-2006 e 2007-2008.

Assim, ao nível do artigo, são estimados modelos de resposta binária, $\Pr(coautoria = 1 \mid \mathbf{x}) = f(\mathbf{x})$, e modelos lineares, $nautores = \mathbf{x}\beta + \varepsilon$, onde $\mathbf{x}$ é um vetor com as seguintes variáveis: *dummies* de área, *não-teórico*, *dummies* de nota Qualis e *dummies* de ano.

Para entender a participação individual em projetos de pesquisa colaborativos, também se utilizam informações demográficas e de campos de atuação dos pesquisadores. Com o intuito de verificar a existência de diferenças entre homens e mulheres no envolvimento em coautoria, constrói-se uma variável *dummy* de sexo, com valor unitário no caso de um pesquisador ser do gênero masculino e zero caso contrário.

Os currículos Lattes contêm informações sobre o campo de atuação dos indivíduos organizadas, como no caso dos artigos, nos quatro níveis de especificidade adotados pelo CNPq. Os indivíduos possuem grande liberdade no preenchimento destas informações de tal modo que, na base de dados, encontram-se autores com até trinta campos de atuação. A fim de testar se um autor recorre à coautoria quando uma de suas pesquisas relaciona-se a campos de pesquisas distintos daqueles de atuação, para cada artigo calcula-se o número de campos diferentes daqueles em que atua em cada um dos quatro níveis de especificidade. Desta maneira, definem-se quatro variáveis de desvio relativamente às áreas de atuação, sendo elas denotadas por: $\Delta ATgr$, $\Delta ATar$, $\Delta ATsub$ e $\Delta ATesp$. Ainda que

o termo não seja preciso, estes desvios em relação às áreas de atuação reportadas são chamados de desvios estáticos. Como a lista de campos de atuação pode não ser estável ao longo do tempo, já que novas áreas de atuação podem ser adicionadas à medida que os interesses de pesquisa de um pesquisador evolui, para cada nível de especificidade, controem-se variáveis de desvio em relação ao histórico de campos aos quais os artigos anteriores de um indivíduo contribuem. Mais precisamente, para os pesquisadores com ao menos dois artigos publicados, comparam-se os campos do artigo de ordem j com os campos de todos os artigos de ordem menor do que j e se obtém o número de campos do artigo de ordem j que não aparecem em nenhum dos artigos anteriores. As variáveis assim obtidas são denotadas por $\Delta ARgr$, $\Delta ARar$, $\Delta ARsub$ e $\Delta AResp$ e por vezes chamadas de desvios dinâmicos.

A partir do histórico de carreira de um pesquisador, define-se a variável binária *titular*, que é igual a um se dois anos antes da publicação de um artigo ele tinha o status de professor titular e é igual a zero caso contrário. Também a partir do histórico de carreira de um indivíduo, são obtidas as variáveis *carga* e *exclusiva*, onde a primeira corresponde ao número de horas semanais dedicadas pelo indivíduo à universidade e a segunda é uma variável binária de dedicação exclusiva. Como a variável *titular*, estas duas últimas são calculadas dois anos antes de cada publicação.

A fim de contemplar mudanças na participação em artigos coautorados ao longo do ciclo de vida de um pesquisador, calcula-se o número de anos transcorridos entre o início do seu primeiro curso de graduação e o ano de publicação dos seus artigos. Esta variável é denotada por *idade*. Outro aspecto considerado refere-se a ciclos de carreira, que se pretende captar pela variável *ordem*, que é a ordem de uma publicação no histórico de artigos de um pesquisador. Como discutido anteriormente, pelo fato de os créditos de uma publicação serem compartilhados entre os seus coautores (SAUER, 1988), os pesquisadores mais jovens, com o intuito de construírem uma reputação para efeitos de contratação e promoção na carreira, têm menos incentivos a trabalharem em equipes (MCDOWELL; MELVIN, 1983; PIETTE; ROSS, 1992). Por esta linha de argumentação, espera-se que o envolvimento em coautoria seja crescente com a experiência do pesquisador. Reforçando este efeito positivo, como as redes de relacionamento dos pesquisadores mais jovens são menores do que as redes de relacionamento dos mais experientes, a oferta de potenciais coautores é menor para os indivíduos em início de carreira. Logo, é menos provável que estes autores encontrem nas suas redes sociais bons colaboradores. Nas estimações econométricas, também se adicionam $idade^2$ e de $ordem^2$ a fim de captar não-linearidades.

Como *proxy* para habilidade cognitiva de um pesquisador, utiliza-se a variável binária *bolsa*, que assume valor um caso a um indivíduo tenha sido concedida uma bolsa de

estudos durante os seus estudos em nível de doutorado. Na medida em que, diante de limitações no número de bolsas de estudo, os departamentos procuram alocá-las segundo critérios meritocráticos, espera-se que os indivíduos mais habilidosos sejam beneficiados mais frequentemente com este tipo de financiamento do que os demais, o que induz uma correlação positiva entre habilidade cognitiva e recebimento de bolsa de estudo.

Tendo calculado todas estas variáveis explicativas, estimam-se modelos de resposta binária do tipo $\Pr(coautoria_{ij} = 1 \mid \mathbf{z}) = g(\mathbf{z})$, em que se obtém a probabilidade de que o artigo j do indivíduo i seja coautorado como função das variáveis do vetor $\mathbf{z}$. Este vetor inclui variáveis do indivíduo (*sexo*, *idade*, *bolsa*, *carga*, *titular*, *exclusiva*), desvios estáticos e dinâmicos e *dummies* de ano, além de j (*ordem*). Não se controla para a qualidade dos artigos. Os modelos de envolvimento individual em coautoria são estimados em duas especificações: *pooled* e efeitos fixos. O primeiro tipo de especificação é atraente porque permite obter coeficientes de variáveis fixas no tempo, como, por exemplo, *sexo* e *bolsa*. Porém, resta a possibilidade de que variáveis não-observadas invariantes no tempo, como, por exemplo, uma propensão inata à colaboração, sejam correlacionadas com as variáveis independentes. Neste caso, as estimativas em *pooled cross-section* são viesadas e inconsistentes. Para contornar este problema, utilizam-se especificações com efeitos fixos, que oferecem proteção a este tipo de endogeneidade.

2.5 Estatísticas descritivas

A tabela 2.21 apresenta as estatísticas descritivas para o período completo, 1971-2009. Além do percentual de artigos coautorados e da média de autores por artigo, que já foram discutidos detalhadamente nas seções anteriores, é possível analisar outras variáveis de interesse. Primeiramente, nota-se que, exceto por linguística, letras e artes, em todas as grandes áreas do conhecimento a maioria dos pesquisadores é do sexo masculino. Em média, no ano da publicação, transcorrem-se de 22 a 24 anos desde o início da primeira graduação do pesquisador. A ordem média dos artigos no histórico de publicação dos pesquisadores varia substancialmente entre as grandes áreas do conhecimento. Repare que, enquanto nas ciências agrárias em média um artigo é a trigésima contribuição de um pesquisador, em ciências sociais ele é a décima quinta. Entre os pesquisadores que possuem ou que buscam o título de doutorado, entre 30% e 61% deles receberam bolsa a depender da grande área do conhecimento. A proporção de professores titulares não ultrapassa os 9%. Com relação ao nível de dedicação à universidade, nota-se que esta varia entre cerca de 25 e 33 horas semanais em média e que, entre os professores, de

26% a 58% têm dedicação exclusiva. É interessante perceber que, nas ciências sociais aplicadas, em que os pesquisadores apresentam nível de dedicação relativamente baixo, tanto em termos de carga horária quanto em termos de exclusividade, é a grande área em que o número médio de artigos publicação é o mais baixo. Uma observação importante é que existem desvios estáticos, em relação aos campos de atuação, em todos os níveis de especificidade do conhecimento. Nas ciências da saúde, em nível de grande área, em média, existem 0,062 campos aos quais um artigo contribui e que não pertencem ao conjunto de campos de atuação citados pelo pesquisador. Nas ciências humanas, este número é de 0,116. À medida que se consideram níveis mais específicos de conhecimento, percebe-se uma tendência de aumento nas medidas de desvio estático, de tal forma que, ao nível de especialidade, existem, em média, nas ciências da saúde, 0,367 campos diferentes daqueles de atuação por artigo e, nas ciências humanas, 0,642. Já com relação aos desvios dinâmicos, eles são observados apenas ao nível mais específico, de especialidade, oscilando entre grandes áreas do conhecimento na faixa compreendida entre 0,079 e 0,257. Por fim, percebe-se que, uma pequena minoria dos artigos possuem ao menos uma das palavras aplicada, empírica, quantitiva ou experimental (flexionadas ou não em gênero ou número). De fato, não mais do que 12% dos artigos podem ser considerados não-puramente teóricos segundo a metodologia utilizada para medir esta característica.

Nas tabelas 2.22 e 2.23 são reportadas as estatísticas descritivas referentes aos períodos 2000-2006 e 2007-2008, considerando apenas os artigos cujas avaliações Qualis mantiveram-se constantes em cada um dos dois subperíodos. Um dos pontos que merece atenção é o fato de que, em comparação com o período completo de quatro décadas, não se percebe nos subperíodos mais recentes uma elevação do percentual de artigos não-teóricos na maioria das grande áreas do conhecimento. Este resultado, analisado isoladamente, contradiz a intuição de que, os avanços tecnológicos das últimas décadas, por facilitarem especialmente a computação e o compartilhamento de dados, deveriam induzir um aumento no percentual de trabalhos mais aplicados ou empíricos. No que se refere à qualidade dos artigos no subperíodo 2007-2008 é possível notar que há uma concentração na faixa intermediária B1-B4, mas que nas engenharias os artigos de qualidade B5 são os mais frequentes. Comparando as estatíticas de qualidade entre os dois subperíodos, parece que a nova avaliação consegue classificar com maior precisão as revistas, pois, em pelo menos quatro grandes áreas do conhecimento, a avaliação vigente até 2006 gerava uma concentração muito alta nos níveis mais elevados de qualidade (tabela 2.22)

2.6 Resultados econométricos: determinantes da coautoria

2.6.1 Análise ao nível do artigo

A seguir, discutem-se os resultados das estimações dos determinantes da coautoria ao nível do artigo. Primeiramente, investiga-se como a probabilidade de um artigo ser produzido em colaboração de dois ou mais autores responde a uma série de variáveis observadas. Para isto, estimam-se modelos de probabilidade linear em que a variável dependente assume valor um no caso de um artigo ser coautorado e zero caso contrário. Uma deficiência dos modelos de probabilidade linear é que, pelo fato dos efeitos marginais serem constantes, eles geram previsões fora do intervalo unitário para determinadas combinações de valores das variáveis independentes. Uma alternativa para evitar este problema é estimar modelos não-lineares, como logit ou probit, porém, esta prática não é adotada neste artigo porque, sobretudo nos anos mais recentes, a fração de artigos coautorados (variável dependente $coautoria = 1$) é muito alta em praticamente todas as áreas do conhecimento. Por este motivo, muitas combinações de variáveis geram previão perfeita de coautoria, violando a hipótese de que $0 < \Pr(coautoria = 1 \mid \mathbf{x}) < 1$.

As estimações são feitas para cada grande área do conhecimento separadamente e os efeitos parciais são reportados nas tabelas do apêndice 5. O conjunto de variáveis explicativas inclui *dummies* de ano, *dummies* de áreas do conhecimento, *dummy* de natureza não-puramente teórica e nota Qualis do periódico no qual o artigo foi publicado. Os modelos foram estimados para três períodos: (i) 1971-2009; (ii) 2000-2006, quando da avaliação Qualis antiga e (iii) 2007-2008, que corresponde à nova avaliação Qualis. Como o objetivo desta parte do trabalho é testar algumas hipóteses referentes aos determinantes da coautoria, a interpretação dos resultados é mais interessante se for feita variável a variável através das tabelas, pois permite que as lições gerais sejam destacadas e que as diferenças entre os campos do conhecimento sejam identificadas mais facilmente.

Para começar, considere o impacto da variável *não-teórico*. A literatura prevê que seu impacto seja positivo e que tenha crescido ao longo do tempo, possivelmente em decorrência dos avanços em computação e em telecomunicações. Os resultados obtidos mostram que o efeito sugerido não é geral, comum a todas as grandes áreas do conhecimento, posto que existem diferenças importantes entre os campos de pesquisa. Quando se analisam todas as quatro décadas, observa-se que, de fato, na maioria das grandes áreas do conhecimento, trabalhos não-puramente teóricos são mais provavelmente produzidos em coautoria, mas que existem exceções: ciências biológicas, ciências agrárias e linguística, letras e artes. Enquanto na primeira o impacto da variável *não-teórico* é (estatisticamente) nulo, nas

outras duas ele é negativo. Estudando os subperíodos mais recentes, nota-se que nos anos de 2000 até 2006, o impacto do caráter não-puramente teórico de um artigo sobre a ocorrência de coautoria essencialmente desaparece. De fato, apenas as ciências humanas apresentam um coeficiente positivo e significante e, ainda assim, apenas na especificação mais simples, sem controle de qualidade da publicação por notas Qualis. Por outro lado, no período 2007-2008, percebe-se que, enquanto nas demais grandes áreas o efeito de *não-teórico* é nulo, nas ciências da saúde e nas ciências exatas e da terra ele é positivo e significante a 1% nas duas especificações, com e sem notas Qualis. Em cada uma destas duas grandes áreas do conhecimento, ao comparar o coeficiente de *não-teórico* referente ao período completo com o seu correspondente no subperíodo 2007-2009, pode-se extrair evidências quanto à evolução do impacto desta variável ao longo do tempo (sem controlar para mudanças intertemporais nos demais parâmetros). Fazendo isto, percebe-se que nas ciências exatas e da terra o impacto de *não-teórico* sobre a probabilidade de coautoria mais do que triplicou. Já nas ciências da saúde, a mudança é mais modesta, se existente. Note que, em 2007-2008, na especificação sem notas Qualis, a estimação pontual do coeficiente associado a *não-teórico* tem uma pequena queda quando comparada à sua correspondente nas quatro décadas, mas que na especificação com notas Qualis ela é 34% maior.

Uma preocupação deste trabalho é investigar a relação entre coautoria e qualidade de um artigo. Por este motivo, estimam-se especificações em que notas Qualis de qualidade das publicações são utilizadas como variáveis explicativas. Naturalmente, muitas vezes, os autores envolvidos em uma pesquisa não são capazes de prever exatamente a qualidade da revista na qual o artigo dela resultante será publicado. Porém, eles devem ser capazes de avaliar o potencial dos trabalhos aos quais se dedicam. Assim, a nota Qualis de uma publicação pode ser interpretada como uma *proxy* para a avaliação que os autores possuem sobre uma pesquisa antes de sua conclusão. A direção do impacto de variações na qualidade esperada de um artigo sobre a coautoria não é claro a priori. Por um lado, quanto melhor a qualidade de um projeto de pesquisa, maior a sua atratividade para os pesquisadores. Assim, imagine um pesquisador com boas qualificações de posse de um bom projeto de pesquisa. Ele não deve ter dificuldade em encontrar um colaborador caso ele necessite. Em um ambiente de formação de parceria simples como este, em que um pesquisador procura por colaboradores em um projeto de pesquisa por ele poposto, pelo lado da oferta de coautores, espera-se uma relação positiva entre qualidade e coautoria. Por outro lado, mesmo neste ambiente, é possível que o pesquisador inspirado prefira trabalhar sozinho, ainda que haja grande oferta de colaboradores, para não dividir os créditos do trabalho. Esta decisão de trabalhar em equipe depende, entre outros fatores, da magnitude do desconto sofrido pelas publicações coautoradas, do potencial antecipado do projeto de pesquisa e da oferta de bons colaboradores. Imagine agora uma pesquisa que

tem sido desenvolvida em grupo. Se ao longo dos trabalhos percebe-se que a qualidade potencial da publicação resultante é suficientemente inferior à prevista inicialmente, é possível que um ou mais dos pesquisadores envolvidos abandone o projeto. Por este aspecto da estabilidade do grupo de pesquisa, espera-se que qualidade e coautoria estejam positivamente relacionados. Evidentemente, qualidade e coautoria não podem ser vistos como dois objetos separados, pois a própria qualidade de um projeto de pesquisa depende da combinação de habilidades e da sinergia entre os autores. Sendo assim, as estimativas obtidas devem ser interpretadas como resultantes de equilíbrio de um jogo de mercado com fricções de *matching* e de maneira nenhuma se propõe aqui o estabelecimento de uma relação de causalidade.

Os resultados não encontram uma relação clara entre qualidade e ocorrência de coautoria. Note que, a 5% de significância, nenhum coeficiente de qualidade é diferente de zero em ciências agrárias, ciências biológicas, engenharias, ciências sociais aplicadas e linguística, letras e artes. Nas ciências exatas e da terra, publicações de nível B nacional e C internacional têm probabilidade 25,5% e 27,4%, respectivamente, maior de serem coautoradas do que publicações de nível C nacional. Ou seja, nesta grande área do conhecimento, quando se considera a avaliação Qualis vigente no período 2000-2006, os resultados indicam que artigos de qualidade imediatamente superior à mínima, têm maior probabilidade de coautoria. Nas ciências da saúde, no período 2000-2006, nota-se que artigos de nível A internacional e A nacional e artigos de qualidade B nacional têm probabilidade maior (entre 7,3% e 8,8%) de coautoria *vis-à-vis* aqueles de qualidade mais baixa, C nacional. Por fim, nas ciências humanas, artigos do estrato intermediário B2 têm 25,5% menos probabilidade de coautoria do que artigos de qualidade C, segundo a atribuição de notas do período 2007-2008. Analisando todas as grandes áreas do conhecimento em conjunto, não é possível estabelecer uma relação bem definida entre incidência de coautoria e qualidade dos artigos, pois os efeitos encontrados variam entre os campos de pesquisa e entre níveis de qualidade.

Considere agora os resultados das regressões do número de autores nas mesmas variáveis do modelo de ocorrência de coautoria. De particular interesse são os coeficientes relativos à natureza não-puramente teórica dos artigos. As estimações sobre o período 1971-2009 mostram que, enquanto o número de autores em um artigo de ciências biológicas não responde à variável *não-teórico*, em todas as demais esta variável é significante e que, à exceção de linguística, letras e artes, o impacto desta variável, como sugerido na literatura, é positivo. Analisando os dois subperíodos mais recentes, observa-se que, exceto pelas ciências sociais aplicadas, artigos não-puramente teóricos têm o mesmo número esperado de autores que os demais. Nas ciências sociais aplicadas, no período 2000-2006, a especificação com notas Qualis indica que o impacto de *não-teórico* é positivo. Porém,

no período 2007-2008, o impacto é negativo.

Com referência ao impacto da qualidade dos artigos sobre o número de autores, no período 2000-2006, é difícil extrair alguma conclusão com algum nível de generalidade. Contudo, no período 2007-2008, os padrões tornam-se um pouco mais claros. Vale notar que em ciências agrárias, ciências biológicas e ciências da saúde, o coeficiente associado ao nível de qualidade mais alto, A1, é positivo e significante. Nas ciências da saúde, o mesmo é válido também para as publicações do estrato A2. Isto indica que há uma relação positiva entre número de autores e qualidade dos artigos nestas grandes áreas. Nas engenharias, todos os estratos de qualidade superiores a C apresentam maior número de autores. Este fato por si só já é suficiente para sugerir uma relação entre as duas variáveis, porém o resultado encontrado torna-se mais interessante quando se percebe que as estimativas pontuais dos coeficientes de qualidade atingem seus máximos nos estratos mais altos: A1 e A2. Nas ciências sociais aplicadas, ao contrário dos campos anteriores, artigos de qualidade A1 tendem a ter menos autores. Nas ciências exatas e da terra e nas ciências humanas, respectivamente, artigos de qualidade B4 e B2 (portanto, intermediários) têm menos autores do que os do estrato C.

2.6.2 Análise ao nível do autor

Agora, estudam-se os determinantes da coautoria ao nível do autor. Estimam-se os efeitos de variáveis observadas sobre a probabilidade do indivíduo trabalhar em parceria. Como antes, as estimações são feitas por grandes áreas do conhecimento. Nas tabelas 2.40.1 e 2.40.2 são reportados os resultados das estimações dos modelos de probabilidade linear em *pooled cross-section*. São estimadas duas especificações, uma sem *carga*, *exclusiva* e *titular* entre as variáveis explicativas e outra com estas variáveis. É interessante perceber que, com exceção das ciências sociais aplicadas, em todas as grandes áreas do conhecimento, em pelo menos uma especificação, homens são menos propensos à coautoria. Porém, esta diferença entre os gêneros desaparece quando as variáveis *carga*, *exclusiva* e *titular* são incluídas na equação de regressão nas ciências biológicas, ciências humanas, ciências da saúde e linguística, letras e artes, ainda que as estimativas pontuais continuem indicando uma menor probabilidade de coautoria entre os homens. De todo modo, parece seguro afirmar que as mulheres são mais propensas à colaboração em pesquisa do que homens.

A análise do impacto da função quadrática da variável *ordem* sobre a probabilidade de coautoria revela padrões interessantes. Como discutido anteriormente, alguns trabalhos (MCDOWELL; MELVIN, 1983; PIETTE; ROSS, 1992) sugerem que pesquisadores em

fases iniciais de carreira têm menor propensão à coautoria por conta da necessidade de construir uma reputação para efeitos de contratação e promoção na carreira. Sendo assim, pelo menos após as primeiras publicações, é de se esperar que a probabilidade de envolvimento em coautoria cresça à medida que um pesquisador acumule um histórico de contribuições à literatura. Outro fator relacionado atuando na mesma direção é que autores mais experientes dispõem de uma oferta maior de potenciais coautores. De fato, em todas grandes áreas do conhecimento, em ao menos uma das especificações, observa-se que a probabilidade de coautoria cresce com a ordem da publicação pelo menos a partir de um determinado ponto. Possíveis exceções a este padrão encontram-se em engenharias, ciências humanas e ciências sociais aplicadas nas estimações da especificação com as variáveis *carga*, *exclusiva* e *titular*. Nas engenharias os coeficientes de *ordem* e $ordem^2$ nesta especificação mais completa são significantes a 10%; nas ciências humanas, nesta mesma especificação, *ordem* não é siginificante e $ordem^2$ é a apenas 10%; nas ciências sociais aplicadas, o coeficiente de *ordem* é negativo e significante e o de $ordem^2$ é estatisticamente igual a zero. Assim, exceto pela segunda especificação em ciências sociais aplicadas, que prevê um padrão de decrescimento monotônico na probabilidade de coautoria como função da ordem de publicação, e sendo um pouco tolerante com os resultados da segunda especificação em engenharias e em ciências humanas, percebe-se uma diferença sugestiva entre os campos do conhecimento. Vale notar que nos campos mais aplicados, experimentais ou quantitativos (ciências agrárias, ciências biológicas, engenharias, ciências exatas e da terra e ciências das saúde), a probabilidade de coautoria responde à ordem das publicações em formato de "U" invertido. Já em ciências humanas, ciências sociais aplicadas e linguística, letras e artes, o formato desta função é de "U". Não é possível aqui afirmar os motivos desta diferença, mas é relevante o fato de que diferenças em temas e, consequentemente, na tecnologia de produção acadêmica sejam refletidos na participação em coautoria ao longo do ciclo de carreira de um pesquisador. Até onde se sabe, esta é uma contribuição nova à literatura.

Ao longo do ciclo de vida também parece haver efeitos importantes em algumas grandes áreas. A este respeito note que em ciências agrárias, ciências biológicas, engenharias e ciências da saúde existe alguma evidência de que a probabilidade de participação em artigos coautorados responde à idade de um pesquisador em formato de "U". Nos demais campos é difícil fazer qualquer inferência sobre a relação entre probabilidade de coautoria e idade de um pesquisador, pois, embora alguns coeficientes associados sejam significantes em pelo menos uma especificação, eles mudam de sinal entre uma e outra. Em todas as grandes áreas do conhecimento, os coeficientes associados às variáveis de idade são baixos, assim como os coeficientes de ordem.

Como *proxy* para habilidade cognitiva, as especificações dos modelos de envolvimento

individual em coautoria contêm a variável binária *bolsa*, que assume valor unitário se o indivíduo foi contemplado com bolsa de estudo em ao menos um curso de doutorado[9]. Nas ciências agrárias, nas ciências biológicas e nas ciências humanas não se obtém qualquer relação entre o recebimento de bolsa de doutorado e a probabilidade de coautoria. Já nas ciências exatas e da terra, a segunda especificação aponta para uma relação positiva enquanto que nas ciências sociais aplicada ela indica uma relação negativa. A magnitude dos impactos em ambas as áreas não são desprezíveis, com um acréscimo de quase 4% na primeira e uma redução de aproximadamente 10% na segunda.

Uma pergunta fundamental quando se investigam os determinantes da coautoria é se os pesquisadores procuram explorar complementaridades entre si. Os coeficientes associados aos desvios em temas de pesquisa podem ajudar a responder este questionamento. Se os pesquisadores buscam coautores com conhecimentos complementares aos seus, espera-se que os coeficientes das medidas de desvio sejam positivos, ou seja, quanto mais desvios, maior a probabilidade de coautoria. Analisando as tabelas 2.40.1 e 2.40.2 nota-se que, quando se controla para *carga*, *exclusiva* e *titular*, desvios dinâmicos de especialidade importam apenas nas ciências da saúde e nas ciências exatas e da terra, sendo, em ambas as grandes áreas do conhecimento, o impacto positivo da ordem de 3%.

Um outro aspecto explorado é o efeito do nível de dedicação do pesquisador sobre o seu envolvimento em coautoria. Para isto, as variáveis utilizadas são a carga horária dedicada pelo pesquisador e uma variável *dummy* de dedicação exclusiva, em ambos os casos medidas dois anos antes de cada publicação. Os resultados mostram que o coeficiente de *carga* é significante apenas em três grandes áreas do conhecimento: ciências agrárias, engenharias e ciências biológicas, sendo que nas duas primeiras o impacto é positivo, enquanto que na última ele é negativo. Já com relação à exclusividade na dedicação, encontra-se que, exceto pelas ciências humanas e pelas ciências sociais aplicadas, o coeficiente associado não é significante. Os impactos da variável *exclusiva* não são desprezíveis nas ciências humanas (9,8%) e nas ciências sociais aplicadas (-12,4%), porém apontam em direções opostas.

Por fim, analisa-se o impacto do status de professor titular sobre a probabilidade de envolvimento em coautoria. Percebe-se que apenas em ciências biológicas e em ciências humanas o coeficiente desta variável é significante e que em ambos os casos ele é negativo. Além disto, os impactos marginais não são pequenos, sendo de -29,1% na primeira grande área do conhecimento e de -12,3% na segunda.

Embora as estimações em *pooled cross section* permitam obter coeficientes de variáveis

[9]Na base de dados há indivíduos com até quatro títulos de doutorado.

estáveis no tempo, as estimativas podem ser inconsistentes em decorrência de correlações entre as variáveis explicativas e o resíduo. Por exemplo, é razoável supor que atributos não-observados da personalidade do pesquisador como curiosidade inata por novos temas sejam correlacionados com as medidas de desvios em temas de pesquisa. Por este motivo, estimam-se modelos de probabilidade linear em painel com efeitos fixos, o que fornece proteção contra endogeneidades decorrentes da correlação entre variáveis explicativas e características individuais do pesquisador invariantes no tempo. As variáveis explicativas inseridas no modelo são as de idade, as de desvio em temas de pesquisa e *dummies* de ordem. Outras variáveis não são contempladas por não apresentarem variação suficiente ao nível do pesquisador individual. Na tabela 2.41 reportam-se os resultados destas estimações. Nota-se que praticamente todos os coeficientes são insignificantes, com a única exceção de desvio dinâmico de especialidade em linguística, letras e artes, cuja estimativa pontual é de -5%.

O fato de efeitos idiossincráticos mudarem acentuadamente as conclusões obtidas nos modelos em *pooled cross-section* deve ser visto como uma oportunidade para o redirecionamento da investigação sobre os determinantes da coautoria. Muito esforço tem sido empregado na tentativa de identificação dos impactos de variáveis individuais observáveis, o que tem rendido algumas conclusões interessantes. Contudo, a formação de um grupo de pesquisa é uma decisão social, que envolve o consentimento de mais do que uma pessoa. Sendo assim, novos avanços na pesquisa sobre colaboração acadêmica devem enfatizar a compreensão dos mecanismos pelos quais se formam os grupos de pesquisa.

2.7 Conclusões

Este artigo investiga um dos mais importantes aspectos da produção acadêmica: a colaboração na forma de coautoria entre pesquisadores. Utilizando dados provenientes dos currículos Lattes/USP, mostra-se que a incidência e a extensão da coautoria têm crescido em todos os campos do conhecimento. Entretanto, apesar de todas as áreas do conhecimento compartilharem uma tendência de crescimento na incidência e na extensão da coautoria, sobretudo nos últimos quinze anos, existem diferenças importantes entre elas. Nas ciências agrárias, nas ciências biológicas, nas ciências da saúde, nas engenharias e nas ciências exatas e da terra, que são campos de investigação mais aplicados ou quantitativos, a quase totalidade dos artigos é produzida em parceria por dois ou mais pesquisadores e o número de autores por artigo é bastante alto. Nas áreas do conhecimento menos aplicadas e quantitativas, como ciências humanas, ciências sociais aplicadas e linguística,

letras e artes, a incidência e a extensão da coautoria são significativamente menores do que nas anteriores e, nestes campos de estudo, o crescimento destes indicadores iniciou-se posteriormente ao crescimento verificado nas demais.

Além de caracterizar o movimento em direção à coautoria, este artigo procura investigar os fatores que a favorecem. Neste ponto, os resultados obtidos variam muito entre os campos do conhecimento de tal forma que é difícil extrair conclusões gerais. Entretanto, algumas lições podem ser obtidas. Ao nível do artigo, trabalhos não-puramente teóricos têm mais autores pelo menos nas ciências da saúde e nas ciências exatas e da terra. Já ao nível do autor, os resultados indicam que homens são menos propensos à coautoria e que parecem existir efeitos de ciclos de carreira no envolvimento em coautoria. Uma novidade com relação a estes é que eles se expressam de maneira diferenciada entre os campos de pesquisa. Enquanto as grandes áreas do conhecimento mais aplicadas ou quantitativas a probabilidade de envolvimento em coautoria é uma função com formato de "U" invertido na ordem da publicação no histórico de carreira do pesquisador, nas demais ela é uma função em formato de "U".

Na medida em que a colaboração entre pesquisadores é um ingrediente fundamental na produção acadêmica, é necessário entender como se formam os grupos de pesquisa. Um dos objetivos deste artigo foi verificar o impacto de variáveis comumente apontadas como relevantes. Na maioria dos casos, não foi possível chegar a conclusões gerais de tal forma que ainda há muito o que avançar nesta linha de pesquisa. Acredita-se que um elemento importante para a compreensão da forma como os pesquisadores estabelecem parcerias reside na existência de fricções de pareamento. Se esta hipótese estiver correta, é possível que redes sociais desempenhem um papel relevante na redução destas fricções pelo fato de que suas conexões servem como canais ao longo dos quais informações sobre as características individuais dos pesquisadores e sobre os seus interesses de pesquisa podem navegar através da comunidade acadêmica facilitando o encontro de dois ou mais indivíduos.

3 REDES SOCIAIS DE COAUTORIA, INTEGRAÇÃO DA COMUNIDADE ACADÊMICA BRASILEIRA E FRICÇÕES DE PAREAMENTO

3.1 Introdução

As interações sociais entre pesquisadores constituem um insumo importante para a produção acadêmica e para a disseminação do conhecimento. Não obstante este fato e apesar do papel determinante dos avanços científicos e tecnológicos para a melhoria das condições de vida, em economia, apenas recentemente começam a surgir trabalhos que investigam tais interações. Este artigo contribui para esta literatura, primeiramente, ao caracterizar a evolução das redes sociais de coautoria de artigos acadêmicos ao longo de quatro décadas.

De maneira nenhuma se depreende que as interações sociais entre pesquisadores restringem-se àquelas inerentes à colaboração formal entre eles. É claro que uma rede social de coautoria em artigos acadêmicos está contida em uma rede social mais densa, que envolve um conjunto de tipos de relacionamentos mais amplo. Ainda assim, o estudo das redes sociais de coautoria é particularmente importante porque este tipo de relacionamento, colaborativo, pressupõe alto nível de interação entre pesquisadores através da qual se criam e se trocam conhecimentos. Sendo assim, as redes sociais de coautoria constituem um objeto de análise apropriado quando se deseja entender como a geração e a disseminação do conhecimento relacionam-se com a estrutura social da comunidade acadêmica. Um primeiro passo na busca deste entendimento naturalmente pede uma descrição da estrutura das redes de coautorias. Esta é a primeira tarefa à qual este artigo se propõe.

Outros trabalhos procuram descrever a estrutura de redes sociais de coautoria (GOYAL; LEIJ; MORAGA-GONZÁLEZ, 2006; MOODY, 2004; NEWMAN, 2001). Entretanto, a abrangência da base de dados aqui utilizada permite que se estudem praticamente todos os campos do conhecimento por um período longo sem que se escolham as revistas nas quais os artigos são publicados. Até onde se sabe, nenhum outro trabalho sobre redes sociais de coautoria de artigos acadêmicos reúne simultaneamente estas três características. Os resultados obtidos mostram que em pelo menos quatro das nove grandes áreas do conhecimento analisadas, a topologia da rede social de coautoria evoluiu para aquela de mundo pequeno, em que, informalmente, as distâncias entre os indivíduos são pequenas.

Além de descrever a evolução das redes sociais de coautoria, este artigo investiga a existência de fricções de pareamento entre os pesquisadores. Para isto, o mecanismo de formação destes grafos é decomposto em dois componentes: um aleatório e outro de intermediação social, em que um colaborador comum é determinante na formação de um elo de coautoria entre dois pesquisadores. Os resultados obtidos mostram que estas fricções estão presentes e que elas se manifestam diferentemente entre os campos do conhecimento estudados.

O restante do trabalho está organizado da seguinte maneira. Na seção 3.2 descrevem-se os dados. A seção 3.3 revisa alguns fatos estilizados referentes aos grafos de colaboração acadêmica. Na seção 3.4 mostra-se como a estrutura das redes sociais de coautoria em nove grandes áreas do conhecimento tem evoluído nas últimas quatro décadas. Na seção 3.5 estuda-se o mecanismo de formação das redes sociais de coautoria e se procura mensurar as fricções sociais de pareamento de pesquisadores. A seção 3.6 resume, conclui e propõe questões para exploração em trabalhos futuros.

3.2 Dados

A Plataforma Lattes é um sistema concebido e disponibilizado publicamente pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico[1] (CNPq), que coleta em uma única estrutura dados de currículos de pesquisadores, de grupos de pesquisa e de instituições envolvidas em ciência e tecnologia. O currículo Lattes de um pesquisador possui informações detalhadas sobre essencialmente todos os aspectos relevantes que descrevem a sua atuação acadêmica ao longo do tempo. Atualmente, a Plataforma Lattes conta mais de um milhão e setecentos mil currículos[2], cujas informações são utilizadas amplamente por diversos órgãos para fins de planejamento, gestão e fomento das atividades científica e de inovação no Brasil. Trata-se de um sistema muito bem sucedido e plenamente incorporado pela comunidade acadêmica brasileira, conforme destacado por Lane (2010) no trecho abaixo:

> A experiência brasileira com a base de dados Lattes (http://lattes.cnpq.br/english) é um exemplo poderoso de boa política. Ela provê dados de alta qualidade sobre cerca de 1,6 milhão de pesquisadores e sobre cerca de 4.000 instituições. A agência nacional do Brasil para financiamento reconheceu no final da década de 1990 que ela precisava de uma nova abordagem para avaliar as credenciais dos pesquisadores. Primeiro, ela desenvolveu uma 'comunidade virtual' de agências federais e pesquisadores para desenhar e desenvolver a infraestrutura Lattes. Em segundo lugar, ela criou os incentivos apropriados para que os

[1] http://www.cnpq.br/

[2] Em 02/12/2010, a Plataforma Lattes continha 1.779.778 currículos.

> pesquisadores e as instituições acadêmicas utilizem a base de dados: os dados são usados pela agência federal na tomada de decisões de financiamento e pelas universidades em suas decisões de estabilidade e promoção. Em terceiro lugar, ela estabeleceu um sistema de identificação única dos pesquisadores que garante que pessoas com nomes parecidos sejam consideradas corretamente. O resultado é uma das bases de dados de pesquisadores mais limpas existentes.[3]

Pelo fato de os currículos Lattes serem utilizados para avaliação de pesquisadores e de instituições, existem incentivos para que os pesquisadores atualizem seus currículos regularmente. Em particular, por ocasião da avaliação trienal dos programas de pós-graduação realizada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), os membros dos corpos docentes das universidades são estimulados a atualizarem os seus currículos.

Neste artigo, dos currículos Lattes, utiliza-se o subconjunto Lattes/USP, composto pelos currículos Lattes de indivíduos que citam a Universidade de São Paulo em pelo menos uma das seguintes três opções: (i) endereço profissional, (ii) formação acadêmica/titulação ou (iii) atuação profissional. Todos os indivíduos que indicam a Universidade de São Paulo em pelo menos uma destas três opções têm seus currículos carregados para a base de dados, que contabiliza 62516 indivíduos. Destes, são considerados ativos apenas aqueles que possuem ao menos um artigo publicado. Faz-se esta restrição para efeito de comparação dos resultados obtidos neste trabalho com os resultados encontrados em outras pesquisas sobre colaboração entre pesquisadores, as quais, em sua ampla maioria, utilizam informações de coautoria provenientes de periódicos (BABCHUK; KEITH; PETERS, 1999; BARNETT; AULT; KASERMAN, 1988; HUDSON, 1996; MCDOWELL; MELVIN, 1983; PIETTE; ROSS, 1992 e PRESSER, 1980). Sendo assim, na literatura, todos os indivíduos considerados possuem ao menos um artigo publicado. Por outro lado, este trabalho diferencia-se dos demais em um aspecto importante. Enquanto aqueles estudam a colaboração em artigos de revistas selecionadas, em geral de grande prestígio, aqui, pelo fato de se contar com uma amostra de currículos de pesquisadores, não há a necessidade de escolher os periódicos. Esta é uma vantagem relevante conferida pela base de dados, afinal, a composição da equipe de pesquisadores varia de acordo com a qualidade das pesquisas e, portanto, das revistas nas quais os artigos resultantes são publicados.

[3]The Brazilian experience with the Lattes Database (http://lattes.cnpq.br/english) is a powerful example of good practice. This provides high-quality data on about 1.6 million researchers and about 4,000 institutions. Brazil's national funding agency recognized in the late 1990s that it needed a new approach to assessing the credentials of researchers. First, it developed a 'virtual community' of federal agencies and researchers to design and develop the Lattes infrastructure. Second, it created appropriate incentives for researchers and academic institutions to use the database: the data are referred to by the federal agency when making funding decisions, and by universities in deciding tenure and promotion. Third, it established a unique researcher identification system to ensure that people with similar names are credited correctly. The result is one of the cleanest researcher databases in existence.

O período analisado compreende os anos de 1971 até 2009. Este período é dividido em triênios consecutivos, para cada um dos quais se constroem as redes sociais de coautoria, definidas pelo conjunto de pesquisadores que tiveram algum artigo publicado no triênio e pelas arestas que os conectam a seus coautores. Os pesquisadores são classificados em comunidades de acordo com as suas grandes áreas de pesquisa verificadas nos currículos. Nesta classificação, um pesquisador é associado a uma determinada grande área do conhecimento se ele possui ao menos um artigo publicado naquele campo, conforme reportado pelo próprio indivíduo. Os pesquisadores que possuem artigos em mais de uma grande área do conhecimento são alocados ao conjunto de indivíduos atuantes em cada um dos campos citados. Assim, obtém-se uma sequência com treze redes sociais de coautoria para cada uma das nove grandes áreas do conhecimento analisadas: ciências agrárias; ciências biológicas; ciências da saúde; ciências exatas e da terra; ciências humanas; ciências sociais aplicadas; linguística, letras e artes; e outros.

Dois pontos merecem destaque. O primeiro deles é que as redes sociais de coautoria são construídas utilizando as informações fornecidas pelos indivíduos acerca de seus respectivos artigos publicados, as quais incluem os nomes dos coautores. Entretanto, esta tarefa de construção dos grafos de colaboração não é trivial porque a caracterização do conjunto verdadeiro de pesquisadores neles inseridos é dificultada por quatro motivos: (i) existência de homônimos, (ii) erros de digitação, (iii) imprecisão na digitação e (iv) mesma pessoa reportada sob nomes diferentes. Para contornar estes problemas, dentro de cada área do conhecimento, foram comparados os nomes das pessoas citadas pela medida de distância relativa de Levenshtein. Ao fazer a comparação de nomes completos restrita por áreas do conhecimento, a possibilidade de que indivíduos de mesmo nome sejam confundidos é reduzida. O *matching* de nomes completos também permite amenizar o problema de erros de digitação. Contudo, a sua capacidade de superar as imprecisões de digitação e a mudança de nomes é limitada. Neste ponto, vale mencionar que os pesquisadores pertencentes ao conjunto Lattes/USP, de uma maneira geral, são mais cuidadosos ao preencherem os próprios nomes do que os nomes de seus coautores[4]. A segunda observação é que nem todos os artigos reportados nos currículos Lattes exigem revisão por pares. Entretanto, como argumentado por Newman (2001), embora estas revistas possam ser de qualidade científica inferior às revistas com revisão por pares, do ponto de vista da caracterização das conexões sociais entre pesquisadores, elas são igualmente boas.

[4]Para a melhoria da qualidade das informações contidas nos currículos Lattes, acredita-se que o CNPq poderia desenvolver mecanismos que induzam os pesquisadores a reportarem o nome de seus coautores com mais cuidado.

3.3 Fatos estilizados

Antes de prosseguir para a o estudo da evolução da estrutura das redes sociais de coautoria construídas a partir dos currículos Lattes/USP, é conveniente, para efeito de comparação, revisar os resultados obtidos em outros trabalhos que investigam os grafos de colaboração acadêmica. Em artigo recente, Newman (2001) analisa as redes sociais de coautoria em ciências biomédicas, física (inclusive subáreas) e ciências da computação no período de cinco anos compreendidos entre 1995 até 1999. Segundo o autor, um pesquisador em ciências biomédicas possui em média 18,1 colaboradores; em física teórica este indicador é de 9,7; em física experimental de alta energia o número médio de colaboradores é 173; e em ciência da computação o grau médio de um pesquisador é 2,22. Newman (2001) também conclui que estas comunidades científicas podem ser consideradas mundos pequenos, pois, nos grafos estudados, as distâncias médias entre dois cientistas são tipicamente baixas (entre 4 e 9,7) e variam logaritmicamente com o número de pesquisadores. Em todas as redes de coautoria, a aglomeração é alta, sugerindo que os pesquisadores facilitam de uma forma ou de outra o contato entre seus colaboradores. Além disto, sem exceção, as redes de coautoria possuem um componente gigante, que contém uma fração grande dos indivíduos, e o segundo maior componente é bem menor do que o primeiro, indicando que estas redes têm baixo risco de fragmentação. Outra característica comum é que as distribuições de grau dos pesquisadores são bem ajustadas por leis de potência com truncagem exponencial. Existem, entretanto, diferenças importantes entre as áreas. Em física experimental, o número médio de coautores que um pesquisador possui é muito maior do que aquele de um pesquisador das demais áreas. Já a comunidade de pesquisadores em ciências biomédicas apresenta aglomeração muito inferior do que as comunidades dos outros campos de pesquisa e as suas características médias são dominadas pela maioria das pessoas com baixo grau ao invés das poucas pessoas muito conectadas, tal como acontece nos outros campos. Uma possível explicação para este último fato é a prática comum nas ciências biomédicas de diretores de laboratórios serem listados como coautores dos artigos resultantes de pesquisas realizadas naqueles.

Goyal, Leij e Moraga-González (2006) estudam a evolução da rede de coautoria entre economistas no período compreendido entre os anos de 1970 e 2000. Neste estudo, os autores dividem o período em três subperíodos: 1970-1979, 1980-1989 e 1990-1999. Para cada um dos subperíodos, os autores obtêm a rede de coautoria, sendo que em cada umas destas três redes, dois economistas estão conectados se eles publicaram em parceria mútua ao menos um artigo acadêmico no período correspondente. Os autores então observam a evolução da rede de coautoria dos economistas nas décadas de 1970, 1980 e 1990. Eles encontram que o grau médio nas redes consideradas é bastante pequeno em relação ao

número total de nós, que a distância média dentro do componente gigante é pequena e que os indicadores de aglomeração são altos. Por fim, os autores encontram que, na década de 1970, o maior componente (com maior número de indivíduos) continha apenas 15% dos economistas, porém, na década de 1990, esta cobertura subiu para 40%. Nestas condições, eles concluem que a topologia da rede social de economistas converge para aquela de mundo pequeno.

No caso da rede de colaboração formal em sociologia, Moody (2004) conclui que, pelo fato de a distribuição do número de coautores não ser bem descrita por uma lei de potência, ela não é gerada por um processo de ligação preferencial. A conectividade da comunidade de pesquisadores desta disciplina não é dependente dos indivíduos com número de coautores muito alto, já que a fração de indivíduos no maior componente mantém-se elevada mesmo que se excluam todos os pesquisadores com grau bastante acima da média. Sendo assim, a difusão de informação nesta rede social não é dependente de algumas poucas estrelas.

3.4 Comunidade acadêmica brasileira como um mundo pequeno

Esta seção discute alguns indicadores de estrutura das redes sociais de coautoria de nove grandes áreas do conhecimento. Estes indicadores são calculados considerando o subgrafo restrito apenas aos pesquisadores pertencentes ao conjunto Lattes/USP e às conexões entre eles e para o grafo gerado pelos pesquisadores Lattes/USP, em que o conjunto de vértices é composto pelos pesquisadores pertencentes ao conjunto Lattes/USP e pelos coautores por eles citados, independentemente destes pertencerem ou não ao conjunto Lattes/USP. A estrutura da apresentação acompanha a de Goyal, Leij e Moraga-González (2006), em que os autores mostram a convergência da estrutura do grafo de colaboração entre economistas para aquela de mundo pequeno, a qual, pela definição dos autores, atende às seguinte quatro condições: (i) o número de vértices é muito grande em relação ao grau médio, $\eta(G)$; (ii) existe um caminho conectando uma parcela grande da população, o que tecnicamente se refere como um componente gigante, (iii) a distância média no componente gigante é pequena, da ordem do logaritmo natural do número de vértices do grafo, $|V|$ e (iv) a aglomeração, $C(G)$, é alta, muito maior do que o número médio de coautores dividido pelo número de vértices.

Antes de prosseguir, convém fazer alguns esclarecimentos. Em primeiro lugar, dos grafos excluem-se os laços e as arestas múltiplas, ou seja, trabalha-se com a versão simples do grafo irrestrito e do subgrafo Lattes/USP. Esta prática é adotada porque o objetivo

deste artigo é medir a integração da comunidade acadêmica e, caso os laços e as arestas múltiplas fossem mantidas, a produtividade de um pesquisador seria confundida com as suas parcerias. Considere, por exemplo, um indivíduo com grau dois no grafo original, não-simplificado. Não se sabe, somente a partir desta informação, se ele tem apenas uma publicação sem coautores (laço), se ele tem apenas uma publicação com dois coautores distintos, se ele tem duas publicações com um mesmo colaborador (aresta múltipla), ou se ele tem duas publicações, cada uma com um coautor distinto. Em segundo lugar, o grau é calculado apenas para os indivíduos pertencentes ao conjunto Lattes/USP, pois para estes tem-se informação sobre todos os seus coautores[5], enquanto que para os demais este indicador existe apenas na medida em que são citados como coautores pelos primeiros, sendo, por este motivo, subestimado se calculado a partir dos currículos Lattes/USP. Em terceiro lugar, pelo fato de nenhuma das redes sociais de coautoria analisadas ser conexa, as distâncias são calculadas para o maior componente de cada grafo. Note que a distância entre dois pesquisadores calculada no grafo irrestrito é maior ou igual à distância real entre eles na rede de coautoria, pois o grafo irrestrito, por ser calculado a partir dos currículos Lattes/USP omite possíveis conexões entre pesquisadores externos à USP.

Para ilustrar como os grafos de coautoria estudados são construídos, considere os grafos hipotéticos da figura 3.1. Em quadrados vermelhos são representados os pesquisadores externos ao conjunto Lattes/USP e em círculos azuis os indivíduos pertencentes a este conjunto. Entre os grafos dos painéis da esquerda e os da direita, removem-se as arestas múltiplas e o laço, representados por linhas tracejadas verdes, resultando assim em suas respectivas versões simplificadas, que são as utilizadas neste artigo. Como se pode perceber, os subgrafos Lattes/USP, nos painéis inferiores, ignoram por completo os pesquisadores que não fazem parte deste conjunto. No painel nordeste, é importante chamar atenção para a aresta fg, destacada em tracejado vermelho. Se os pesquisadores f e g são citados como coautores em um mesmo artigo por d ou por h, então fg é uma aresta do grafo. Porém, é possível que a aresta fg seja neglicenciada, embora na realidade ela exista, porque f e g são coautores em um artigo que não envolve nenhum pesquisador Lattes/USP.

[5]Exceto por omissões no preenchimento dos currículos Lattes.

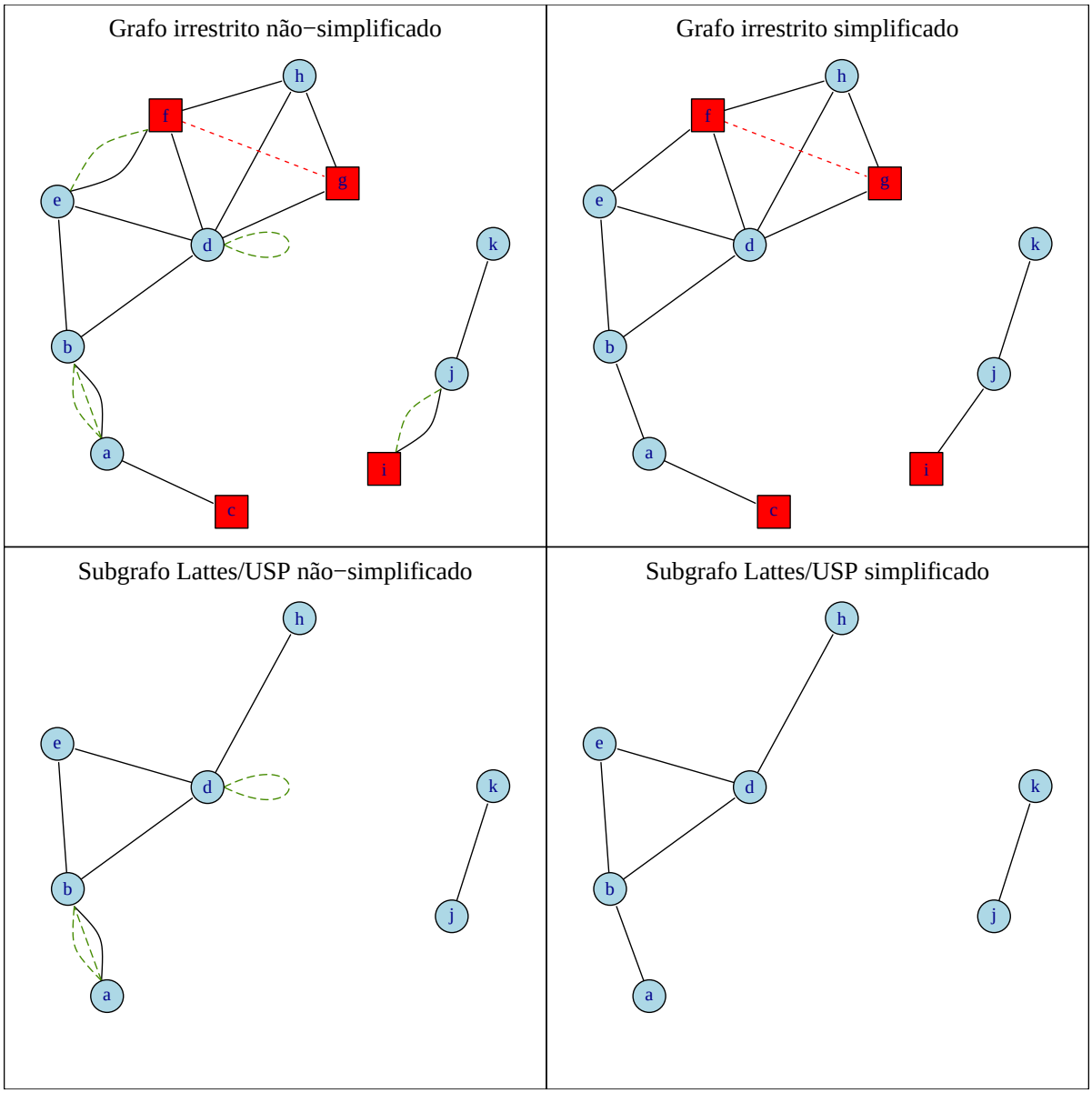

Figura 3.1 - Grafos de coautoria

Uma consideração importante na interpretação dos resultados a seguir é que alguns indivíduos considerados externos à população Lattes/USP podem na verdade pertencer a ela. Esta possibilidade existe por dois motivos. O primeiro é que o sistema de currículos Lattes foi implantado no final de década de 1990. Sendo assim, é possível que um indivíduo pertecente à população Lattes/USP cite entre seus coautores pesquisadores que tinham relação com a USP mas que, por exemplo, aposentaram-se antes de o sistema de currículos Lattes difundir-se. Como alguns destes pesquisadores, inativos a partir do final de década de 1990, podem não ter criado os seus currículos Lattes, eles são considerados externos ao conjunto de pesquisadores Lattes/USP. O segundo motivo é que, até recentemente, muitos estudantes (mesmo os de pós-graduação) não tinham currículo Lattes. Por este motivo, um ex-aluno que publicou um trabalho com um professor, mas que não

seguiu a carreira acadêmica após a formatura, pode não ter currículo Lattes e, portanto, ser considerado erroneamente externo à população Lattes/USP. Tendo em mente que o conjunto verdadeiro de pesquisadores relacionados à USP contém o conjunto de indivíduos pertencentes ao conjunto Lattes/USP, conclui-se que o cálculo de grau dos pesquisadores restrito ao subgrafo Lattes/USP está subestimado e que este erro deve ser maior para os períodos mais antigos.

3.4.1 Ciências agrárias

Em ciências agrárias, o número de pesquisadores produtivos pertencentes à Plataforma Lattes/USP cresce de forma estritamente monotônica em todo o período, assim como o número de seus coautores externos à população Lattes/USP. De fato, o crescimento observado é de 1532% entre os pesquisadores Lattes/USP e de 6104% entre os seus coautores externos.

Tabela 3.1 - Número de pesquisadores: ciências agrárias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	21761	9,99	1289	7,16
2004-2006	17926	9,79	1181	7,07
2001-2003	13492	9,51	983	6,89
1998-2000	9402	9,15	780	6,66
1995-1997	6088	8,71	630	6,45
1992-1994	4215	8,35	522	6,26
1989-1991	3086	8,03	430	6,06
1986-1988	2481	7,82	370	5,91
1983-1985	1919	7,56	289	5,67
1980-1982	1447	7,28	227	5,42
1977-1979	1094	7,00	174	5,16
1974-1976	751	6,62	134	4,90
1971-1973	409	6,01	79	4,37

Nesta grande área do conhecimento, o grau médio dos pesquisadores Lattes/USP aumenta de forma quase monotônica durante todo o período no grafo irrestrito e de maneira estritamente monotônica no grafo restrito aos pesquisadores Lattes/USP. A razão entre o grau médio dos pesquisadores pertencentes à Plataforma Lattes/USP e o total de indivíduos no grafo irrestrito, apresenta uma clara tendência de queda ao longo das quatro décadas, até que no triênio 2007-2009 atinge 0,14%. Já no grafo restrito aos indivíduos

Lattes/USP, esta razão tem crescido nos últimos cinco triênios, chegando em 2007-2009 a 0,27%. Sendo assim, o primeiro critério enumerado por Goyal, Leij e Moraga-González (2006) é atendido nas duas redes sociais de coautoria: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.2 - Grau médio: ciências agrárias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	30,06	0,14	3,47	0,27
2004-2006	25,69	0,14	3,09	0,26
2001-2003	21,84	0,16	2,44	0,25
1998-2000	16,76	0,18	1,72	0,22
1995-1997	12,35	0,20	1,16	0,18
1992-1994	9,86	0,23	0,95	0,18
1989-1991	8,53	0,28	0,82	0,19
1986-1988	7,74	0,31	0,74	0,20
1983-1985	7,83	0,41	0,73	0,25
1980-1982	7,05	0,49	0,50	0,22
1977-1979	6,86	0,63	0,47	0,27
1974-1976	5,88	0,78	0,45	0,33
1971-1973	4,53	1,11	0,13	0,16

Em ciências agrárias, a fração de indivíduos no maior componente cresce de forma estritamente monotônica desde 1989 no grafo irrestrito e desde 1986 no grafo Lattes/USP, com grande aceleração nos triênios 1995-1997 e 1998-2000, respectivamente. No último triênio, o maior componente dos grafos continha 93% e 64% dos respectivos pesquisadores. Logo, o segundo critério para um mundo pequeno também é respeitado nos dois grafos: *existe um componente que contém uma grande parcela dos vértices.*

Tabela 3.3 - Porcentagem dos maiores componentes: ciências agrárias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	92,79	0,28	64,3	0,54
2004-2006	89,65	0,28	60,63	0,59
2001-2003	85,78	0,50	47,81	2,03
1998-2000	76,23	1,40	38,08	2,31
1995-1997	66,49	1,38	10,16	5,40
1992-1994	24,27	20,50	6,90	2,49
1989-1991	18,60	17,47	4,88	3,26
1986-1988	37,00	6,21	4,32	2,70
1983-1985	36,69	5,94	4,84	2,08
1980-1982	34,35	4,63	2,64	2,64
1977-1979	35,10	6,49	4,02	3,45
1974-1976	18,91	7,99	6,72	3,73
1971-1973	6,11	5,13	3,80	2,53

Na rede social irrestrita, percebe-se que a distância média e o diâmetro crescem, não-monotonicamente, entre os triênios 1971-1973 e 1995-1997 e passam a apresentar um tendência mais nítida de decrescimento a partir do triênio 1998-2000. No grafo Lattes/USP, a distância média e o diâmetro crescem, não-monotonicamente, até o triênio 1998-2000. A partir deste ponto, com algumas oscilações, estes indicadores tendem a cair. Além disto, nos dois grafos, a distância média entre os pesquisadores é pequena, sendo tipicamente inferior ao $\ln(|V|)$. Nestas condições, o terceiro requisito para um mundo pequeno é atendido: *a distância média entre os vértices do componente gigante é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.4 - Distância média e diâmetro: ciências agrárias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	6,34	17	5,91	17
2004-2006	6,77	17	6,72	22
2001-2003	6,95	19	6,28	18
1998-2000	7,77	22	8,01	25
1995-1997	9,21	26	4,43	14
1992-1994	8,09	27	3,01	8
1989-1991	6,71	16	2,32	6
1986-1988	9,18	22	1,73	4
1983-1985	7,90	18	1,66	5
1980-1982	6,58	16	1,44	3
1977-1979	7,92	21	1,44	3
1974-1976	4,34	14	1,67	3
1971-1973	1,68	4	1,00	1

Com relação ao coeficiente de aglomeração individual médio, observa-se que, tanto no grafo irrestrito quanto no grafo Lattes/USP, este indicador apresenta uma tendência decrescente. Contudo, em ambas as redes sociais, este indicador é sempre bastante alto, muito superior à razão entre o grau médio e o número de vértices. Sendo assim, o quarto critério para um mundo pequeno também é atendido: *aglomeração é alta,* $C(G) \gg \eta(G)/|V|$.

Tabela 3.5 - Aglomeração (%): ciências agrárias

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	36,44	53,23
2004-2006	37,71	53,88
2001-2003	38,31	48,74
1998-2000	36,82	50,52
1995-1997	40,99	57,43
1992-1994	45,51	63,84
1989-1991	49,72	63,83
1986-1988	49,04	66,79
1983-1985	46,70	77,02
1980-1982	48,70	64,52
1977-1979	49,30	78,10
1974-1976	55,71	59,39
1971-1973	62,54	100,00

3.4.2 Ciências biológicas

O número de pesquisadores produtivos em ciências biológicas cresceu substancialmente entre o primeiro e o último triênio, porém merecem destaque alguns episódios de queda acentuada, sobretudo nos triênios 1998-2000 e 2001-2003.

Tabela 3.6 - Número de pesquisadores: ciências biológicas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	38525	10,56	1971	7,59
2004-2006	26038	10,17	1467	7,29
2001-2003	2714	7,91	167	5,12
1998-2000	4567	8,43	287	5,66
1995-1997	14080	9,55	1201	7,09
1992-1994	11724	9,37	1189	7,08
1989-1991	8627	9,06	1015	6,92
1986-1988	6256	8,74	809	6,70
1983-1985	4552	8,42	656	6,49
1980-1982	1856	7,53	265	5,58
1977-1979	1262	7,14	189	5,24
1974-1976	1698	7,44	304	5,72
1971-1973	1039	6,95	221	5,40

Ignorando os triênios 1998-2000 e 2001-2003, verifica-se que o grau médio dos pesquisadores Lattes/USP nesta mesma rede social de coautoria tem crescido ao longo do tempo. No que diz respeito à razão entre o grau médio e o número de pesquisadores em cada um dos grafos, verifica-se que no grafo restrito aos pesquisadores Lattes/USP ela é essencialmente estável, porém apresenta um certo crescimento nos últimos três triênios. Já na rede social de coautoria irrestrita, desconderando os triênios 1998-2000 e 2001-2003, nota-se uma clara tendência decrescente. Como nos dois grafos esta razão é sempre bastante pequena, conclui-se que o primeiro critério enumerado por Goyal, Leij e Moraga-González (2006) é atendido nas duas redes sociais de coautoria: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.7 - Grau médio: ciências biológicas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	31,66	0,08	2,16	0,11
2004-2006	24,42	0,09	1,21	0,08
2001-2003	16,41	0,60	0,12	0,07
1998-2000	16,28	0,36	0,18	0,06
1995-1997	15,20	0,11	0,94	0,08
1992-1994	12,73	0,11	0,92	0,08
1989-1991	10,36	0,12	0,74	0,07
1986-1988	8,82	0,14	0,59	0,07
1983-1985	7,61	0,17	0,47	0,07
1980-1982	6,77	0,36	0,14	0,05
1977-1979	6,46	0,51	0,15	0,08
1974-1976	5,39	0,32	0,22	0,07
1971-1973	4,18	0,40	0,17	0,08

Em ciências biológicas, a fração de pesquisadores pertencentes ao maior componente do grafo Lattes/USP cresce rapidamente entre 1995 e 2009, concluindo o período em 49%. Já no grafo irrestrito, a proporção de pesquisadores no maior componente inicia uma trajetória de crescimento sustentada por diversos triênios, que é interrompida entre 1998 e 2003, mas que é retomada nos triênios seguintes, até representar 94% dos indivíduos. Logo, o segundo critério enumerado por Goyal, Leij e Moraga-González (2006) é atendido nas duas redes sociais de coautoria: *existe um componente que contém uma grande parcela dos vértices.*

Tabela 3.8 - Porcentagem dos maiores componentes: ciências biológicas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	94,24	0,26	49,21	1,07
2004-2006	89,53	0,19	18,47	3,82
2001-2003	19,97	8,84	1,80	1,80
1998-2000	41,71	3,94	1,05	1,05
1995-1997	78,37	0,69	10,82	1,58
1992-1994	69,69	0,66	8,50	1,60
1989-1991	65,69	0,58	2,07	1,48
1986-1988	51,25	2,17	1,24	1,24
1983-1985	32,60	3,01	1,37	0,91
1980-1982	5,12	5,06	2,26	1,51
1977-1979	5,39	4,60	2,12	1,59
1974-1976	3,83	3,77	1,65	0,99
1971-1973	6,83	3,46	1,81	1,81

A rede social de coautoria Lattes/USP mostra crescimento da distância média e do diâmetro até o trênio 1992-1994. A partir do triênio seguinte, é difícil identificar uma tendência nestes indicadores por conta da grande volatilidade apresentada por eles. No grafo irrestrito, a distância média e o diâmetro aumentam inicialmente, porém, já no início da década de 1990, iniciam uma trajetória de queda. Como nos dois grafos a distância média é tipicamente inferior ao $\ln(|V|)$ e, quando superior, de uma ordem de magnitude comparável, afima-se que a terceira exigência para um mundo pequeno é respeitada: *a distância média entre os vértices do componente gigante é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.9 - Distância média e diâmetro: ciências biológicas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	6,21	18	7,43	18
2004-2006	6,50	20	8,31	24
2001-2003	5,80	17	1,09	2
1998-2000	8,54	23	1,10	2
1995-1997	7,93	26	6,24	16
1992-1994	8,47	23	6,61	19
1989-1991	10,17	30	2,27	7
1986-1988	11,39	28	1,68	5
1983-1985	11,17	29	1,52	4
1980-1982	3,15	11	1,52	3
1977-1979	2,49	7	1,28	3
1974-1976	2,60	8	1,20	2
1971-1973	2,71	9	1,17	2

O coeficiente de aglomeração individual médio do grafo restrito aos pesquisadores pertencentes ao conjunto Lattes/USP é tipicamente alto, superior a 50%, exceção feita ao triênio 1977-1979, com tendência de queda a partir do triênio 1989-1991 e com um pico no triênio 2001-2003, cujo comportamento já se verificou anormal. No grafo irrestrito, o coeficiente de aglomeração individual médio tem caído consistentemente ao longo do tempo, com valor mínimo de 38% no último triênio. Nestas condições, a quarta exigência para mundo pequeno também é respeitada: *aglomeração é alta,* $C(G) \gg \eta(G)/|V|$.

Tabela 3.10 - Aglomeração (%): ciências biológicas

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	37,85	51,49
2004-2006	44,12	51,56
2001-2003	53,35	75,00
1998-2000	44,29	50,00
1995-1997	41,64	51,81
1992-1994	45,92	60,14
1989-1991	49,26	63,79
1986-1988	50,27	64,44
1983-1985	50,95	55,48
1980-1982	54,46	38,89
1977-1979	55,12	0,00
1974-1976	57,01	57,41
1971-1973	64,40	77,78

3.4.3 Ciências da saúde

Em ciências da saúde, o número de pesquisadores produtivos pertencentes ao conjunto Lattes/USP cresce de maneira estritamente monotônica nos treze triênios e o mesmo é válido para o número de pesquisadores externos à USP citados pelos primeiros como seus coautores. O crescimento das duas populações de pesquisadores é de 1554% e de 5421%, respectivamente.

Tabela 3.11 - Número de pesquisadores: ciências da saúde

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	64159	11,07	3870	8,26
2004-2006	56969	10,95	3702	8,22
2001-2003	45176	10,72	3094	8,04
1998-2000	33155	10,41	2524	7,83
1995-1997	23081	10,05	2016	7,61
1992-1994	16327	9,70	1587	7,37
1989-1991	11293	9,33	1308	7,18
1986-1988	8174	9,01	1014	6,92
1983-1985	6150	8,72	806	6,69
1980-1982	4287	8,36	631	6,45
1977-1979	3240	8,08	476	6,17
1974-1976	2436	7,80	376	5,93
1971-1973	1326	7,19	234	5,46

O grau médio dos pesquisadores Lattes/USP cresce de forma estritamente monotônica em todo o período tanto no grafo restrito quanto no grafo irrestrito. Enquanto a razão entre o grau médio e número de pesquisadores Lattes/USP mantém-se relativamente estável em torno de 0,11% em todo o período, esta razão cai continuamente entre os triênios no grafo irrestrito, atingindo o patamar de 0,05% no último triênio. Sendo assim, o primeiro critério proposto por Goyal, Leij e Moraga-González (2006) é atendido nas duas redes sociais de coautoria: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.12 - Grau médio: ciências da saúde

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}$%	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}$%
2007-2009	33,97	0,05	4,37	0,11
2004-2006	29,31	0,05	3,82	0,10
2001-2003	26,64	0,06	3,34	0,11
1998-2000	22,70	0,07	2,75	0,11
1995-1997	18,77	0,08	2,15	0,11
1992-1994	16,39	0,10	1,84	0,12
1989-1991	12,89	0,11	1,40	0,11
1986-1988	10,86	0,13	1,04	0,10
1983-1985	9,65	0,16	0,83	0,10
1980-1982	8,29	0,19	0,72	0,11
1977-1979	8,16	0,25	0,63	0,13
1974-1976	7,41	0,30	0,51	0,14
1971-1973	5,69	0,43	0,30	0,13

O tamanho relativo do maior componente das ciências da saúde tem crescido em todos os períodos no caso do grafo irrestrito e, na rede social Lattes/USP, tem crescido desde meados da década de 1980. No último triênio, o maior componente do grafo Lattes/USP representava cerca de 76% dos seus indivíduos e o maior componente do grafo irrestrito continha mais de 96% dos pesquisadores. Com isto, conclui-se que a segunda exigência para caracterização de um mundo pequeno é respeitada: *existe um componente que contém uma grande parcela dos vértices.*

Tabela 3.13 - Porcentagem dos maiores componentes: ciências da saúde

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	96,73	0,15	76,25	0,13
2004-2006	94,95	0,29	69,31	0,35
2001-2003	93,91	0,14	63,61	0,48
1998-2000	90,71	0,20	56,93	0,63
1995-1997	87,16	0,58	45,59	0,74
1992-1994	81,59	0,60	39,07	1,58
1989-1991	77,38	0,85	15,29	4,36
1986-1988	70,28	1,08	5,13	3,25
1983-1985	57,56	1,30	5,96	2,73
1980-1982	48,87	1,26	7,77	1,43
1977-1979	39,94	3,09	6,93	1,47
1974-1976	17,53	14,20	4,26	1,60
1971-1973	7,99	6,86	3,85	2,14

Nesta grande área do conhecimento, observa-se uma tendência de crescimento da distância média e do diâmetro na rede social Lattes/USP até o início da década de 1990. Desde então, esta tendência inverte-se. Quando se analisam estes indicadores no grafo irrestrito, vê-se que a distância média cresce até o triênio 1980-1982 e dalí em diante passa a cair quase monotonicamente. Já o diâmetro tem um comportamento menos bem definido ao longo do tempo. Inicialmente, até o triênio 1980-1982, o diâmetro cresce. Em seguida, com mais volatilidade, ele cai entre 1983 e 1994. Entre 1995 e 2003, o diâmetro volta a crescer lentamente. Por fim, nos dois triênios finais, o diâmetro reduz-se. A distância média entre os pesquisadores é pequena, sendo na maioria dos triênios inferior ao $\ln(|V|)$. Nestas condições, a terceira condição para um mundo pequeno é atendida: *a distância média entre os vértices do componente gigante é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.14 - Distância média e diâmetro: ciências da saúde

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	5,73	17	6,09	17
2004-2006	5,94	19	6,22	19
2001-2003	6,23	20	6,55	19
1998-2000	6,44	18	7,45	19
1995-1997	6,67	19	8,80	30
1992-1994	6,77	18	9,57	30
1989-1991	7,92	29	7,50	20
1986-1988	7,75	24	3,39	10
1983-1985	8,78	30	3,48	9
1980-1982	11,15	31	3,94	11
1977-1979	10,31	27	2,75	6
1974-1976	5,68	14	2,23	7
1971-1973	3,65	11	1,72	4

O coeficiente de aglomeração individual médio apresenta uma tendência decrescente tanto no grafo irrestrito quanto no grafo Lattes/USP. Contudo, em ambas as redes sociais, este indicador é sempre bastante alto, muito superior à razão entre o grau médio e número de vértices, $C(G) \gg {}^{\eta(G)}/_{|V|}$. Sendo assim, o quarto critério para um mundo pequeno também é atendido.

Tabela 3.15 - Aglomeração (%): ciências da saúde

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	35,19	49,94
2004-2006	36,16	50,11
2001-2003	35,79	48,00
1998-2000	37,11	52,93
1995-1997	39,74	53,56
1992-1994	44,39	56,38
1989-1991	47,35	60,83
1986-1988	49,49	60,64
1983-1985	51,41	56,07
1980-1982	54,57	62,18
1977-1979	53,44	68,11
1974-1976	56,18	75,11
1971-1973	61,29	69,63

3.4.4 Ciências exatas e da terra

Nas ciências exatas e da terra, o número de pesquisadores produtivos cresce em todos os triênios, tanto no grafo Lattes/USP quanto no grafo irrestrito. Considerando todo o período, o crescimento observado nos dois conjuntos de cientistas é significativo, sendo de 1097% e de 3595%, respectivamente.

Tabela 3.16 - Número de pesquisadores: ciências exatas e da terra

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	27091	10,21	1903	7,55
2004-2006	25035	10,13	1878	7,54
2001-2003	19389	9,87	1585	7,37
1998-2000	14203	9,56	1301	7,17
1995-1997	10093	9,22	1112	7,01
1992-1994	6921	8,84	876	6,78
1989-1991	5204	8,56	740	6,61
1986-1988	3743	8,23	613	6,42
1983-1985	2707	7,90	467	6,15
1980-1982	2245	7,72	407	6,01
1977-1979	1711	7,44	302	5,71
1974-1976	1180	7,07	220	5,39
1971-1973	733	6,60	159	5,07

Com relação ao grau médio dos pesquisadores Lattes/USP, percebe-se uma clara tendência de crescimento em ambas as redes sociais de coautoria. A razão entre o grau médio e número de pesquisadores Lattes/USP mantém-se baixa, mas apresenta uma leve tendência de crescimento nos últimos seis triênios. Já no grafo irrestrito, esta razão, sempre inferior a 0,6%, cai entre os triênios de forma estritamente monotônica. Sendo assim, o primeiro critério de caracterização de um mundo pequeno é atendido nas duas redes sociais de coautoria: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.17 - Grau médio: ciências exatas e da terra

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	21,41	0,08	1,95	0,10
2004-2006	20,06	0,08	1,90	0,10
2001-2003	17,55	0,09	1,54	0,10
1998-2000	14,39	0,10	1,11	0,08
1995-1997	11,30	0,11	0,76	0,07
1992-1994	9,30	0,13	0,55	0,06
1989-1991	8,16	0,16	0,54	0,07
1986-1988	6,68	0,18	0,44	0,07
1983-1985	6,25	0,23	0,36	0,08
1980-1982	5,57	0,25	0,24	0,06
1977-1979	5,79	0,34	0,28	0,09
1974-1976	5,14	0,44	0,18	0,08
1971-1973	4,11	0,56	0,13	0,08

O maior componente tem crescido em tamanho relativo em todos os últimos triênios, sendo que este crescimento iniciou-se de maneira mais acelerada primeiramente no grafo irrestrito, meados da década de 1980, do que no grafo Lattes/USP, cujo crescimento mais acelerado começa no final da década de 1990, embora já ocorresse um crescimento gradual anteriormente. Em ambos os casos, no último triênio, o maior componente contém uma proporção não-desprezível dos pesquisadores, 88% no grafo irrestrito e 43% no grafo Lattes/USP. Com isto, conclui-se que a segunda condição listada por Goyal, Leij e Moraga-González (2006) é respeitada: *existe um componente que contém uma grande parcela dos vértices.*

Tabela 3.18 - Porcentagem dos maiores componentes: ciências exatas e da terra

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	88,20	0,24	43,25	0,84
2004-2006	85,49	0,33	42,33	1,17
2001-2003	80,34	0,47	29,91	1,58
1998-2000	75,47	0,54	11,68	2,84
1995-1997	59,47	1,75	2,25	1,80
1992-1994	52,72	2,93	1,94	1,37
1989-1991	42,41	2,08	1,62	1,22
1986-1988	17,77	5,85	1,47	1,31
1983-1985	8,05	6,98	1,29	1,28
1980-1982	9,18	7,35	1,47	0,98
1977-1979	10,11	6,25	1,66	1,32
1974-1976	11,53	5,85	1,36	1,36
1971-1973	8,32	6,68	1,89	1,26

As ciências exatas e da terra têm um comportamento intertemporal bastante distinto daqueles apresentados pelas grandes áreas do conhecimento anteriores com relação à distância média e ao diâmetro. Repare que, na rede social de coautoria Lattes/USP, há uma clara tendência de crescimento da distância média e do diâmetro. Já no grafo irrestrito, a distância média e o diâmetro partem de seus mínimos no primeiro triênio até atingirem os seus máximos entre o final da década de 1980 e o início da década de 1990, quando iniciam uma trajetória descendente. Nos dois grafos, a distância média entre os pesquisadores é pequena, porém merece atenção o fato de que nos últimos três triênios, a distância média no grafo Lattes/USP foi um pouco superior ao $\ln(|V|)$. Feitas estas considerações, conclui-se que a terceira condição para um mundo pequeno é atendida: *a distância média entre os vértices do componente gigante é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.19 - Distância média e diâmetro: ciências exatas e da terra

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	7,13	20	8,97	27
2004-2006	7,11	20	8,22	22
2001-2003	7,55	19	9,17	24
1998-2000	8,19	25	6,61	20
1995-1997	8,75	24	2,50	7
1992-1994	10,76	34	2,18	8
1989-1991	13,30	34	1,62	6
1986-1988	6,77	23	1,67	5
1983-1985	4,01	11	1,42	4
1980-1982	4,96	19	1,25	2
1977-1979	4,27	13	1,16	2
1974-1976	4,35	16	1,20	2
1971-1973	2,72	9	1,00	1

O coeficiente de aglomeração individual médio apresenta uma tendência decrescente tanto no grafo irrestrito quanto no grafo Lattes/USP. Entretanto, em ambas as redes sociais, este indicador é sempre bastante alto, muito superior à razão entre o grau médio e o número de pesquisadores. Sendo assim, o quarto critério para um mundo pequeno também é atendido: *aglomeração é alta,* $C(G) \gg {}^{\eta(G)}/_{|V|}$.

Tabela 3.20 - Aglomeração (%): ciências exatas e da terra

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	36,64	51,03
2004-2006	37,30	55,26
2001-2003	38,06	54,57
1998-2000	39,35	58,43
1995-1997	42,68	56,07
1992-1994	47,51	64,67
1989-1991	47,06	67,38
1986-1988	49,37	69,76
1983-1985	52,09	57,08
1980-1982	56,14	57,59
1977-1979	53,23	88,54
1974-1976	54,79	0,00
1971-1973	59,62	100,00

3.4.5 Ciências humanas

Como nas grandes áreas do conhecimento analisadas acima, o número de pesquisadores Lattes/USP produtivos tem crescido, assim como o número total de indivíduos não-pertencentes ao conjunto Lattes/USP e que são citados como coautores pelos primeiros. A população Lattes/USP cresceu 927% entre o primeiro e o último triênio e o conjunto de pesquisadores externos à USP teve um aumento de 3423%.

Tabela 3.21 - Número de pesquisadores: ciências humanas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	10319	9,24	1160	7,06
2004-2006	11593	9,36	1461	7,29
2001-2003	9936	9,20	1395	7,24
1998-2000	7302	8,90	1196	7,09
1995-1997	4905	8,50	952	6,86
1992-1994	3820	8,25	798	6,68
1989-1991	2745	7,92	609	6,41
1986-1988	2046	7,62	492	6,20
1983-1985	1502	7,31	382	5,95
1980-1982	1047	6,95	290	5,67
1977-1979	913	6,82	219	5,39
1974-1976	663	6,50	168	5,12
1971-1973	373	5,92	113	4,73

O grau médio dos pesquisadores Lattes/USP apresenta clara tendência de crescimento ao longo do tempo nos dois grafos: restrito à população Lattes/USP e irrestrito. Apesar da tendência de aumento no número de coautores entre triênios, a magnitude deste indicador é muito baixa se comparada ao número total de indivíduos, nas duas redes sociais consideradas. Há, porém uma diferença entre os dois grafos no que se refere à razão entre o número médio de coautores e o número total de pesquisadores. Enquanto no grafo Lattes/USP esta relação é relativamente estável em torno da média de 0,03% com uma fraca tendência de crescimento nos últimos triênios, no grafo irrestrito esta razão tende ao decaimento, terminando em 0,09%. De todo modo, a primeira condição que define um mundo pequeno é atendida: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.22 - Grau médio: ciências humanas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	9,32	0,09	0,41	0,04
2004-2006	8,26	0,07	0,41	0,03
2001-2003	7,34	0,07	0,37	0,03
1998-2000	6,05	0,08	0,27	0,02
1995-1997	4,76	0,10	0,20	0,02
1992-1994	4,43	0,12	0,20	0,02
1989-1991	4,03	0,15	0,17	0,03
1986-1988	3,58	0,17	0,17	0,04
1983-1985	3,14	0,21	0,06	0,02
1980-1982	2,92	0,28	0,10	0,04
1977-1979	3,62	0,40	0,13	0,06
1974-1976	3,14	0,47	0,08	0,05
1971-1973	2,33	0,62	0,00	0,00

Nas ciências humanas, o maior componente do grafo irrestrito tem aumentado a sua participação no total de pesquisadores desde o triênio iniciado em 1992, terminando no triênio 2007-2009 com mais de 50%. Em contrapartida, no grafo restrito aos pesquisadores Lattes/USP, a participação do maior componente não tem um comportamento tão bem definido. De todo modo, ele representa uma pequena parcela dos vértices. No último triênio o maior componente continha apenas pouco mais de 2% desta população. Sendo assim, o segundo critério de caracterização de um mundo pequeno é atendido apenas pelo grafo irrestrito.

Tabela 3.23 - Porcentagem dos maiores componentes: ciências humanas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	51,13	1,41	2,24	1,98
2004-2006	45,96	3,20	1,71	0,82
2001-2003	42,52	2,58	2,80	0,65
1998-2000	31,80	4,11	3,09	0,67
1995-1997	27,01	3,75	1,58	0,63
1992-1994	22,20	3,22	1,13	1,13
1989-1991	8,60	7,18	1,15	0,66
1986-1988	9,34	3,52	0,81	0,61
1983-1985	18,91	3,86	0,79	0,52
1980-1982	18,91	2,96	1,38	1,03
1977-1979	18,51	4,27	2,74	0,91
1974-1976	8,30	6,03	1,19	1,19
1971-1973	5,90	5,90	0,89	0,88

As ciências humanas demonstram um crescimento na distância média e no diâmetro da rede de coautoria Lattes/USP até o triênio 1998-2000. Desde então, a tendência destes indicadores é de redução. Por outro lado, no grafo irrestrito, estes indicadores têm crescido ao longo das quatro décadas consideradas, exceto por alguns triênios. Na rede social Lattes/USP e na rede social irrestrita a distância média é sempre inferior ao $\ln(|V|)$. Por este motivo, pode-se afirmar que a terceira condição proposta por Goyal, Leij e Moraga-González (2006) é atendida, ou seja: *a distância média entre os vértices do componente gigante (ou do maior componente) é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.24 - Distância média e diâmetro: ciências humanas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	9,19	27	2,91	9
2004-2006	8,28	27	2,70	10
2001-2003	8,65	23	3,73	11
1998-2000	8,46	21	4,48	13
1995-1997	6,42	18	1,97	5
1992-1994	6,22	14	1,59	4
1989-1991	3,46	9	1,36	4
1986-1988	3,16	10	1,13	3
1983-1985	5,18	13	1,00	1
1980-1982	3,91	9	1,17	2
1977-1979	3,09	6	1,30	2
1974-1976	1,99	6	1,00	1
1971-1973	1,58	4	NaN	0

Embora apresente uma tendência descendente, o coeficiente de aglomeração individual médio é sempre bastante alto nas duas redes sociais. Logo, o quarto critério para um mundo pequeno também é atendido: *aglomeração é alta,* $C(G) \gg {}^{\eta(G)}/_{|V|}$.

Tabela 3.25 - Aglomeração (%): ciências humanas

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	38,91	42,38
2004-2006	41,44	41,98
2001-2003	40,89	38,90
1998-2000	41,56	37,33
1995-1997	42,99	63,20
1992-1994	45,55	58,75
1989-1991	48,60	64,81
1986-1988	53,24	81,82
1983-1985	51,73	100,00
1980-1982	52,69	75,00
1977-1979	55,63	81,33
1974-1976	52,13	NaN
1971-1973	58,92	NaN

3.4.6 Ciências sociais aplicadas

Nas ciências sociais aplicadas, o número de autores Lattes/USP produtivos por triênio tem crescido de forma estritamente monotônica, assim como o número de coautores externos à USP citados por eles em cada triênio. O crescimento total observado nestas duas populações foi de 1980% e de 4923%, respectivamente.

Tabela 3.26 - Número de pesquisadores: ciências sociais aplicadas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	6117	8,72	1144	7,04
2004-2006	5078	8,53	1120	7,02
2001-2003	3659	8,20	907	6,81
1998-2000	2671	7,89	698	6,55
1995-1997	1786	7,49	584	6,37
1992-1994	1221	7,11	455	6,12
1989-1991	1000	6,91	346	5,85
1986-1988	801	6,69	302	5,71
1983-1985	642	6,46	237	5,47
1980-1982	455	6,12	170	5,14
1977-1979	383	5,95	132	4,88
1974-1976	192	5,26	83	4,42
1971-1973	154	5,04	55	4,01

O número médio de coautores dos pesquisadores Lattes/USP tem crescido consistentemente entre períodos nos dois grafos: restrito aos indivíduos associados à USP e irrestrito. Apesar deste crescimento, em ambos os casos, a razão entre o grau médio e o total de vértices do grafo é muito pequena, motivo pelo qual se pode afirmar que é atendido o primeiro requerimento para que esta comunidade seja considerada um mundo pequeno: *o número de vértices é muito grande relativamente ao grau médio.* Vale notar que, tal como observado nas grandes áreas anteriores, esta razão é decrescente no grafo irrestrito e mais estável no grafo Lattes/USP com leve tendência de crescimento nos três triênios mais recentes.

Tabela 3.27 - Grau médio: ciências sociais aplicadas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	5,98	0,10	0,80	0,07
2004-2006	4,64	0,09	0,62	0,06
2001-2003	3,79	0,10	0,38	0,04
1998-2000	3,20	0,12	0,20	0,03
1995-1997	2,38	0,13	0,20	0,03
1992-1994	1,96	0,16	0,17	0,04
1989-1991	2,02	0,20	0,09	0,03
1986-1988	1,82	0,23	0,09	0,03
1983-1985	1,84	0,29	0,08	0,03
1980-1982	1,78	0,39	0,07	0,04
1977-1979	2,07	0,54	0,09	0,07
1974-1976	1,40	0,73	0,02	0,03
1971-1973	1,84	1,19	0,00	0,00

O maior componente do grafo irrestrito e do grafo Lattes/USP comportam-se qualitativamente de modo semelhante, sobretudo nos quatro triênios mais recentes, em que ambos aumentam sua representatividade rapidamente até que em 2007-2009, o maior componente do grafo irrestrito continha 33% dos indivíduos e o do grafo Lattes/USP 13%. Como a fração dos vértices no maior componente do grafo irrestrito é de apenas um terço e do grafo restrito é baixa, não se pode afirmar que a segunda condição para mundo pequeno seja satisfeita para este conjunto de pesquisadores. Contudo, levando em consideração o rápido crescimento destas frações nos últimos triênios, espera-se que em pouco tempo este segundo requerimento seja atendido.

Tabela 3.28 - Porcentagem dos maiores componentes: ciências sociais aplicadas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	33,15	9,47	13,02	1,75
2004-2006	27,83	2,07	5,98	2,23
2001-2003	9,92	3,25	2,43	0,66
1998-2000	2,58	2,47	1,58	1,43
1995-1997	2,35	2,30	1,54	1,03
1992-1994	5,08	2,38	1,32	0,88
1989-1991	2,90	2,60	0,87	0,58
1986-1988	2,12	2,12	1,33	0,99
1983-1985	8,41	4,98	1,27	1,27
1980-1982	6,59	5,71	1,77	1,18
1977-1979	11,75	7,83	2,27	2,27
1974-1976	9,38	5,73	2,41	1,20
1971-1973	14,94	8,44	1,82	1,82

Nas ciências sociais aplicadas, o grafo Lattes/USP indica um crescimento da distância média e do diâmetro em praticamente todos os triênios. No grafo irrestrito, a distância média e o diâmetro comportam-se de maneira similar, com exceção do último triênio em que estes indicadores apresentam uma forte queda. Nos dois grafos, a distância média no último triênio é baixa, de magnitude comparável ao $\ln(|V|)$.

Tabela 3.29 - Distância média e diâmetro: ciências sociais aplicadas

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	8,69	22	5,95	14
2004-2006	13,85	33	5,08	14
2001-2003	4,67	13	2,40	8
1998-2000	2,17	9	2,22	7
1995-1997	1,89	7	1,74	5
1992-1994	2,03	5	1,36	3
1989-1991	1,70	4	1,06	2
1986-1988	1,55	4	1,19	2
1983-1985	1,88	4	1,18	2
1980-1982	1,62	4	1,00	1
1977-1979	1,86	4	1,25	2
1974-1976	1,50	4	1,00	1
1971-1973	1,94	5	NaN	0

O coeficiente de aglomeração individual médio no grafo Lattes/USP não apresenta uma tendência clara, mas deve ser ressaltado que nos últimos seis triênios, ele foi bastante alto, superior a 37%. Já no grafo irrestrito, este indicador tem uma tendência de queda, mas nos últimos três triênios ele se mantém relativamente estável. Como se pode perceber, o quarto critério para um mundo pequeno também é atendido: *aglomeração é alta,* $C(G) \gg {}^{\eta(G)}/_{|V|}$.

Tabela 3.30 - Aglomeração (%): ciências sociais aplicadas

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	40,24	44,62
2004-2006	39,73	40,27
2001-2003	39,94	44,42
1998-2000	41,06	37,78
1995-1997	48,61	55,33
1992-1994	47,73	39,29
1989-1991	42,71	0,00
1986-1988	58,30	58,33
1983-1985	57,03	0,00
1980-1982	54,82	100,00
1977-1979	57,88	0,00%
1974-1976	63,20	NaN
1971-1973	56,16	NaN

3.4.7 Engenharias

O conjunto de pesquisadores produtivos em cada triênio tem crescido ininterruptamente na rede social restrita aos indivíduos Lattes/USP e na rede social irrestrita. As taxas de crescimento destes conjuntos são, respectivamente de 1788% e 5790% entre os triênios 1971-1973 e 2007-2009.

Tabela 3.31 - Número de pesquisadores: engenharias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	12960	9,47	1133	7,03
2004-2006	11093	9,31	1115	7,02
2001-2003	8430	9,04	904	6,81
1998-2000	5795	8,66	728	6,59
1995-1997	3843	8,25	571	6,35
1992-1994	2495	7,82	416	6,03
1989-1991	1818	7,51	319	5,77
1986-1988	1366	7,22	258	5,55
1983-1985	922	6,83	197	5,28
1980-1982	703	6,56	139	4,93
1977-1979	523	6,26	101	4,62
1974-1976	396	5,98	80	4,38
1971-1973	220	5,39	60	4,09

O grau médio dos pesquisadores Lattes/USP vem aumentando consistentemente ao longo do tempo. Como, no último triênio, tanto no subgrafo Lattes/USP quanto no grafo irrestrito, a razão entre o grau médio e o número de pesquisadores é bastante baixa, 0,12% e 0,11%, respectivamente, conclui-se que a primeira condição para que a comunidade de pesquisadores em engenharias seja um mundo pequeno é atendida: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.32 - Grau médio: engenharias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	14,82	0,11	1,32	0,12
2004-2006	12,70	0,11	1,17	0,10
2001-2003	11,29	0,13	0,94	0,10
1998-2000	9,37	0,16	0,80	0,11
1995-1997	7,21	0,19	0,54	0,09
1992-1994	6,01	0,24	0,37	0,09
1989-1991	5,73	0,32	0,35	0,11
1986-1988	5,21	0,38	0,35	0,14
1983-1985	4,11	0,45	0,15	0,08
1980-1982	4,61	0,66	0,14	0,10
1977-1979	4,94	0,94	0,32	0,31
1974-1976	4,56	1,15	0,18	0,22
1971-1973	2,98	1,36	0,17	0,28

Nas engenharias também se verifica um rápido crescimento da participação do maior componente no total de pesquisadores do grafo irrestrito e do grafo Lattes/USP a partir de meados da década de 1990, de tal sorte que no último triênio, o maior componente do grafo irrestrito possuía 70% dos seus vértices e o grafo Lattes/USP continha quase 10%. Nestas condições, enquanto o grafo irrestrito satisfaz a segunda condição para mundo pequeno, o grafo Lattes/USP não satisfaz.

Tabela 3.33 - Porcentagem dos maiores componentes: engenharias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	70,01	0,89	9,89	6,35
2004-2006	60,96	1,08	4,93	3,23
2001-2003	50,76	1,86	3,98	2,77
1998-2000	34,25	6,19	2,89	1,65
1995-1997	21,18	4,53	1,58	1,23
1992-1994	6,53	3,61	1,68	1,68
1989-1991	8,20	5,45	1,57	1,57
1986-1988	10,83	3,00	2,33	1,94
1983-1985	14,64	4,77	3,05	1,52
1980-1982	13,37	5,12	2,16	2,16
1977-1979	12,62	8,60	3,96	3,96
1974-1976	7,07	6,31	6,25	2,50
1971-1973	9,09	7,73	5,00	3,33

Nas engenharias, como nas ciências sociais aplicadas, a distância média e o diâmetro na rede social Lattes/USP crescem consitentemente no período estudado, enquanto que na

rede social irrestrita, esta tendência de crescimento é interrompida no final do período. De qualquer modo, com exceção dos triênios 1998-2000 e 2001-2003 para o grafo irrestrito, a distância média é na maioria das vezes inferior ao $\ln(|V|)$ e, quando superior ao $\ln(|V|)$, apenas por pouca diferença. Em particular, no último triênio, a distância média nos dois grafos é um pouco inferior ao $\ln(|V|)$ e, portanto, neste triênio, a terceira condição para mundo pequeno é satisfeita: *a distância média entre os vértices do componente gigante (ou no maior componente) é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.34 - Distância média e diâmetro: engenharias

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	9,41	27	6,64	20
2004-2006	9,97	27	4,10	11
2001-2003	11,00	29	2,96	9
1998-2000	12,90	35	2,19	7
1995-1997	9,22	24	1,57	5
1992-1994	2,83	8	1,42	3
1989-1991	3,04	12	1,27	3
1986-1988	2,99	8	1,32	3
1983-1985	3,73	9	1,42	2
1980-1982	2,53	6	1,09	2
1977-1979	2,57	6	1,00	1
1974-1976	1,78	5	1,46	2
1971-1973	1,60	3	1,00	1

O coeficiente de aglomeração individual médio no grafo irrestrito apresenta tendência decrescente, mas é sempre superior a 39%. No grafo Lattes/USP, este coeficiente também é decrescente considerando todo o período, sendo que no último triênio ele se encontra em 54%. Como em ambas as redes, o coeficiente de aglomeração individual médio é bastante superior à razão entre o grau médio e o número de vértices, conclui-se que o quarto requerimento para mundo pequeno é atendido: *aglomeração é alta,* $C(G) \gg \eta(G)/|V|$.

Tabela 3.35 - Aglomeração (%): engenharias

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	39,75	53,99
2004-2006	41,24	57,29
2001-2003	41,60	56,28
1998-2000	44,04	62,20
1995-1997	43,71	58,60
1992-1994	52,68	59,76
1989-1991	49,40	69,14
1986-1988	50,04	66,67
1983-1985	51,65	0,00
1980-1982	51,26	75,00
1977-1979	53,88	100,00
1974-1976	56,30	0,00
1971-1973	60,42	100,00

3.4.8 Linguística, letras e artes

O número de pesquisadores com ao menos uma publicação aumentou significativamente entre o triênio 1971-1973 e o triênio 2007-2009, sendo as taxas de crescimento 2787% no grafo Lattes/USP e 8712% na rede social irrestrita.

Tabela 3.36 - Número de pesquisadores: linguística, letras e artes

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	1410	7,25	433	6,07
2004-2006	1251	7,13	459	6,13
2001-2003	936	6,84	357	5,88
1998-2000	636	6,46	289	5,67
1995-1997	488	6,19	233	5,45
1992-1994	361	5,89	201	5,30
1989-1991	206	5,33	124	4,82
1986-1988	176	5,17	101	4,62
1983-1985	134	4,90	84	4,43
1980-1982	105	4,65	65	4,17
1977-1979	90	4,50	56	4,03
1974-1976	50	3,91	31	3,43
1971-1973	16	2,77	15	2,71

A razão entre o grau médio e o número de vértices comporta-se de maneira qualitativamente parecida com o que se observa nas grandes áreas do conhecimento estudadas

anteriormente. No grafo irrestrito, percebe-se uma clara tendência de queda e no grafo Lattes/USP, onde este indicador é mais estável, verifica-se uma tendência de crescimento nos últimos triênios. Em ambos os casos esta razão é bastante baixa no último triênio, motivo pelo qual se afirma que a primeira condição para mundo pequeno é respeitada: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.37 - Grau médio: linguística, letras e artes

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	2,83	0,20	0,25	0,06
2004-2006	2,01	0,16	0,21	0,05
2001-2003	1,73	0,18	0,08	0,02
1998-2000	1,31	0,21	0,08	0,03
1995-1997	1,17	0,24	0,05	0,02
1992-1994	0,82	0,23	0,01	0,00
1989-1991	0,67	0,32	0,00	0,00
1986-1988	0,84	0,48	0,04	0,04
1983-1985	0,63	0,47	0,02	0,03
1980-1982	0,62	0,59	0,00	0,00
1977-1979	0,61	0,67	0,00	0,00
1974-1976	0,61	1,23	0,00	0,00
1971-1973	0,07	0,42	0,00	0,00

Em linguística, letras e artes, a evolução do tamanho relativo do maior componente do grafo irrestrito e do grafo Lattes/USP é muito diferente das apresentadas pelas grandes áreas do conhecimento anteriores. Repare que não apenas o maior componente engloba uma pequena fração dos pesquisadores, como também esta fração tem uma tendência decrescente, embora entre os dois últimos triênios tenha crescido. Na rede social de coautoria irrestrita, no triênio 2007-2009, o maior componente respondia por 7% dos seus vértices e o maior componente da rede social Lattes/USP representava 1% dos seus pesquisadores. Nestas condições, a segunda exigência para mundo pequeno não é atendida.

Tabela 3.38 - Porcentagem dos maiores componentes: linguística, letras e artes

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	7,16	3,55	1,16	0,92
2004-2006	3,36	3,28	1,09	0,87
2001-2003	4,92	4,17	0,84	0,56
1998-2000	7,55	6,29	1,04	1,04
1995-1997	8,61	6,35	0,86	0,86
1992-1994	6,09	4,43	1,00	0,50
1989-1991	6,80	6,31	0,81	0,81
1986-1988	9,66	6,82	2,97	0,99
1983-1985	10,45	5,22	2,38	1,19
1980-1982	11,43	6,67	1,54	1,54
1977-1979	15,56	5,56	1,79	1,79
1974-1976	24,00	12,00	3,23	3,23
1971-1973	12,50	6,25	6,67	6,67

Em linguística, letras e artes, tanto no grafo restrito quanto no irrestrito, a distância média e o diâmetro têm uma tendência de crescimento. Entretanto, como a distância média em ambas as redes sociais de coautoria é sempre menor do que o $\ln(|V|)$, o terceiro requerimento para que esta comunidade de pesquisadores seja caracterizada como um mundo pequeno é respeitada: *a distância média entre os vértices do componente gigante (ou do maior componente) é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.39 - Distância média e diâmetro: linguística, letras e artes

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	2,27	5	1,21	3
2004-2006	1,77	6	1,22	3
2001-2003	1,75	4	1,07	2
1998-2000	1,72	4	1,15	2
1995-1997	1,66	3	1,00	1
1992-1994	1,60	4	1,00	1
1989-1991	1,56	3	NaN	0
1986-1988	1,53	2	1,33	2
1983-1985	1,39	2	1,00	1
1980-1982	1,35	2	NaN	0
1977-1979	1,47	2	NaN	0
1974-1976	1,45	2	NaN	0
1971-1973	1,00	1	NaN	0

Em todos os triênios em que pode ser calculado, o coeficiente de aglomeração individual médio é bastante alto nas duas redes sociais. Embora, no grafo irrestrito, este coeficiente

demonstre tendência de queda, mesmo no seu mínimo histórico, alcançado no último triênio, ele é de 42%. No grafo Lattes/USP, no último triênio, este coeficiente é de 79%. Diante destes valores, conclui-se que o quarto requerimento para mundo pequeno é atendido: *aglomeração é alta,* $C(G) \gg {}^{\eta(G)}/_{|V|}$.

Tabela 3.40 - Aglomeração (%): linguística, letras e artes

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	41,95	78,67
2004-2006	46,05	54,44
2001-2003	47,96	0,00
1998-2000	51,08	60,00
1995-1997	57,83	NaN
1992-1994	60,61	NaN
1989-1991	53,59	NaN
1986-1988	51,53	0,00
1983-1985	71,44	NaN
1980-1982	55,98	NaN
1977-1979	75,09	NaN
1974-1976	49,09	NaN
1971-1973	NaN	NaN

3.4.9 Outros

Esta grande área do conhecimento também tem experimentado um forte aumento no número de pesquisadores nas duas populações: Lattes/USP e de pesquisadores externos à USP. Com efeito, o grupo de pesquisadores Lattes/USP com publicações no triênio 2007-2009 é 683% maior do que o grupo de pesquisadores Lattes/USP no triênio 1971-1973. O número de coautores dos primeiros que não pertencem à Plataforma Lattes/USP tem um aumento de 4504%.

Tabela 3.41 - Número de pesquisadores: outros

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Nº Pesquisadores	ln (Nº Pesquisadores)	Nº Pesquisadores	ln (Nº Pesquisadores)
2007-2009	2212	7,70	94	4,54
2004-2006	1744	7,46	93	4,53
2001-2003	1488	7,31	80	4,38
1998-2000	1140	7,04	74	4,30
1995-1997	699	6,55	66	4,19
1992-1994	565	6,34	57	4,04
1989-1991	315	5,75	48	3,87
1986-1988	337	5,82	40	3,69
1983-1985	278	5,63	33	3,50
1980-1982	189	5,24	32	3,47
1977-1979	82	4,41	18	2,89
1974-1976	61	4,11	13	2,56
1971-1973	58	4,06	12	2,48

A razão entre o grau médio e o número de vértices é bastante pequena na rede social de coautoria Lattes/USP. Por outro lado, na rede social irrestrita, em que este indicador tem tendência decrescente, o razão entre o grau médio e o número de pesquisadores não é tão pequena. Ela parte de 6,61% e termina em 1,03%. Nestas condições e levando em consideração o decaimento deste indicador no grafo irrestrito, existe suporte para afirmar que a primeira condição para mundo pequeno é atendida: *o número de vértices é muito grande relativamente ao grau médio.*

Tabela 3.42 - Grau médio: outros

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$	Grau Médio	$\frac{\text{Grau Médio}}{\text{Nº Pesquisadores}}\%$
2007-2009	22,76	1,03	0,06	0,07
2004-2006	18,29	1,05	0,11	0,12
2001-2003	17,81	1,20	0,05	0,06
1998-2000	14,78	1,30	0,05	0,07
1995-1997	9,61	1,37	0,00	0,00
1992-1994	9,00	1,59	0,04	0,06
1989-1991	5,77	1,83	0,04	0,09
1986-1988	7,60	2,26	0,05	0,13
1983-1985	7,42	2,67	0,00	0,00
1980-1982	4,94	2,61	0,00	0,00
1977-1979	3,56	4,34	0,00	0,00
1974-1976	3,69	6,05	0,00	0,00
1971-1973	3,83	6,61	0,00	0,00

Em outros, nota-se que a participação do maior componente oscila em torno de uma média

de 16% no grafo irrestrito, sendo verificado um crescimento nos últimos quatro triênios. Já no grafo Lattes/USP, com bastante volatilidade, a participação do maior componente tende ao decaimento. No último triênio a participação do maior componente no grafo irrestrito era de 16,5% dos vértices, enquanto que no grafo Lattes/USP era de 2%. Sendo assim, a segunda condição para mundo pequeno não é atendida.

Tabela 3.43 - Porcentagem dos maiores componentes: outros

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Componente 1	Componente 2	Componente 1	Componente 2
2007-2009	16,55	9,90	2,13	2,13
2004-2006	16,34	9,58	2,15	2,15
2001-2003	15,39	10,62	2,50	2,50
1998-2000	14,30	12,98	4,05	1,35
1995-1997	9,30	8,87	1,52	1,52
1992-1994	13,10	8,85	3,51	1,75
1989-1991	19,37	15,87	4,17	2,08
1986-1988	21,36	7,12	5,00	2,50
1983-1985	17,99	10,79	3,03	3,03
1980-1982	9,00	7,41	3,13	3,13
1977-1979	15,85	13,41	5,56	5,56
1974-1976	21,31	13,11	7,69	7,69
1971-1973	20,69	18,97	8,33	8,33

Em outros, o grafo restrito não tem qualquer tendência na distância média e nem no diâmetro. Já no grafo irrestrito, nota-se que a distância média e o diâmetro cresceram durante as quatro décadas como um todo. Adicionalmente, nos dois últimos triênios a distância média tem caído. Como a distância média é sempre menor do que o $\ln(|V|)$, a terceira condição para mundo pequeno é atendida: *a distância média entre os vértices do componente gigante (ou do maior componente) é pequena, da ordem do* $\ln(|V|)$.

Tabela 3.44 - Distância média e diâmetro: outros

Triênio	Grafo Irrestrito		Subgrafo Lattes/USP	
	Distância Média	Diâmetro	Distância Média	Diâmetro
2007-2009	2,49	4	1,00	1
2004-2006	2,54	4	1,00	1
2001-2003	2,64	7	1,00	1
1998-2000	2,40	4	1,33	2
1995-1997	1,85	4	NaN	0
1992-1994	1,84	4	1,00	1
1989-1991	1,77	2	1,00	1
1986-1988	1,77	2	1,00	1
1983-1985	1,74	2	NaN	0
1980-1982	1,55	4	NaN	0
1977-1979	1,39	2	NaN	0
1974-1976	1,54	2	NaN	0
1971-1973	1,39	2	NaN	0

Em geral, nesta grande área do conhecimento, não é possível calcular o coeficiente de aglomeração individual médio para a rede social Lattes/USP. Por este motivo, a avaliação da quarta exigência deve ser baseada no valor deste coeficiente no grafo irrestrito. Nesta rede social, o coeficiente de aglomeração individual médio é bastante superior à razão entre o grau médio e o número de vértices e apresenta, como nas demais grandes áreas do conhecimento, tendência de queda. Conclui-se com isto que a quarta condição para mundo pequeno é atendida*: aglomeração é alta,* $C(G) \gg {}^{\eta(G)}/_{|V|}$.

Tabela 3.45 - Aglomeração (%): outros

Triênio	Grafo Irrestrito	Subgrafo Lattes/USP
2007-2009	32,33	NaN
2004-2006	32,95	NaN
2001-2003	35,23	NaN
1998-2000	28,64	0,00
1995-1997	36,80	NaN
1992-1994	51,29	NaN
1989-1991	44,22	NaN
1986-1988	40,54	NaN
1983-1985	50,13	NaN
1980-1982	60,55	NaN
1977-1979	47,24	NaN
1974-1976	47,49	NaN
1971-1973	59,99	NaN

3.5 Distribuição de grau e mecanismo de formação

Esta seção analisa a distribuição de número de colaboradores que um pesquisador possui. Um dos motivos pelos quais existe interesse na distribuição de grau entre os vértices de um grafo é que ela pode ser informativa acerca da maneira pela qual ele é formado. De fato, como de se esperar, diferentes modelos de formação de redes possuem implicações distintas sobre a distribuição de números de arestas incidentes nos vértices. Por exemplo, em um grafo formado de acordo com o modelo estático de Erdös e Rényi, para n grande e p pequeno, a distribuição de grau é uma Poisson. Já a sua variação dinâmica, em que um vértice surge a cada período de tempo e se une aleatoriamente a m vértices pré-existentes, gera uma distribuição de grau esperado dada por $F_t(d) = 1 - e^{-\frac{d-m}{m}}$. Por outro lado, o modelo de ligação preferencial, em que um vértice ganha novos vizinhos a uma taxa proporcional à cardinalidade de sua vizinhança corrente, a distribuição de grau é uma lei de potência, $\Pr(x) = \alpha x^{-\beta}$.

As figuras 3.2 e 3.3 apresentam a distribuição de frequência do número de colaboradores dos pesquisadores Lattes/USP no grafo de colaboração irrestrito e no subgrafo Lattes/USP, respectivamente, para cada uma das grandes áreas do conhecimento, exceto "outros", que não é investigada deste ponto em diante. Nas figuras, são considerados os históricos completos de colaboração apenas dos indivíduos ativos ao final do período analisado, definidos como aqueles que possuem ao menos uma publicação entre os anos de 2004 e 2009, extremos incluídos. Para facilitar a visualização, adota-se a escala logarítmica[6] e, apenas para efeito de comparação, também é desenhada a distribuição de frequência de grau em um grafo gerado aleatoriamente com mesmo número de vértices e mesma densidade. No caso dos grafos aleatórios irrestritos, além de mesmo número de vértices e mesma densidade que seus correspondentes observados, impõe-se a condição de que, comparativamente às redes sociais de coautoria empíricas, haja o mesmo número esperado de arestas entre indivíduos Lattes/USP e entre estes e os externos. Já os grafos aleatórios contidos na figura 3.3 são simplesmente uma realização de um processo de Erdös e Rényi com mesmo número de vértices e mesma densidade que os subgrafos Lattes/USP observados correspondentes[7].

Em muitas aplicações, verifica-se que a distribuição de grau nas redes sociais possuem caudas mais gordas do que aquelas observadas em grafos Erdös-Rényi com mesmo número de vértices e igual densidade. Analisando as figuras 3.2 e 3.3, percebe-se que, de fato, em

[6] Como $\ln(0)$ não é bem definido, as distribuições apresentadas são para a variável grau $+ 1$.

[7] Evidentemente, o modelo estático de Erdös e Rényi não é uma descrição adequada do processo de formação das redes sociais de coautoria estudadas, pois este processo é dinâmico, enquanto o modelo é estático.

todas as grandes áreas do conhecimento, a frequência de indivíduos com grau bastante elevado e a frequência de indivíduos com grau muito baixo são maiores do que aquelas obtidas no grafo Erdös-Rényi correspondente. Por vezes, é proposto que a distribuição de grau de algumas redes sociais pode ser bem aproximada por uma lei de potência. Se isto é verdade, então no diagrama em escala logarítmica ela deve ter um formato (aproximadamente) linear. Como se pode perceber nas figuras 3.2 e 3.3, em nenhuma das grandes áreas do conhecimento, isto acontece. De fato, mesmo visualmente, nota-se que o ajuste da reta de regressão simples de $\ln(\text{freqüência})$ em $\ln(\text{grau}+1)$ não é acurado.

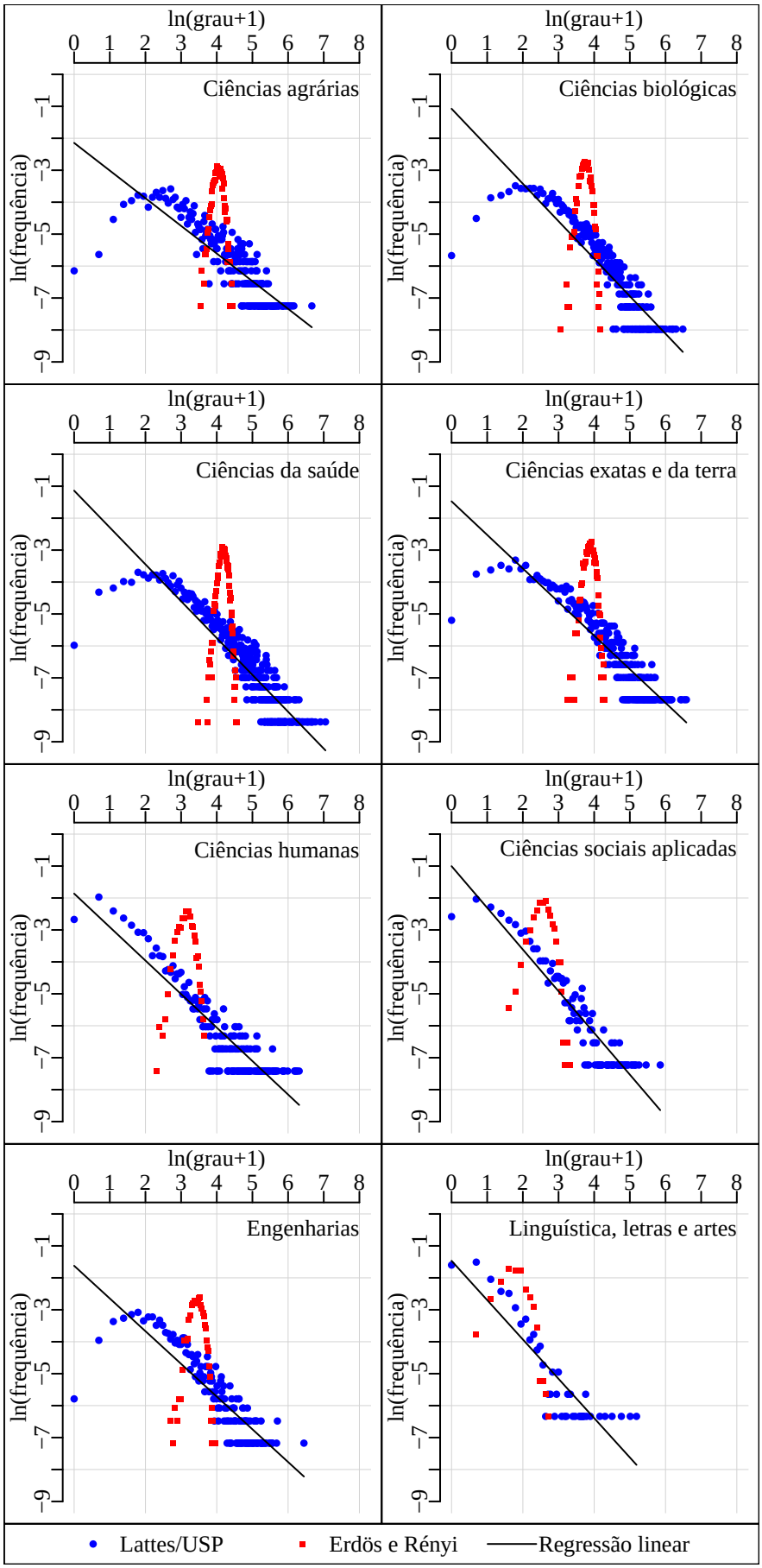

Figura 3.2 - Distribuição de grau: grafo irrestrito

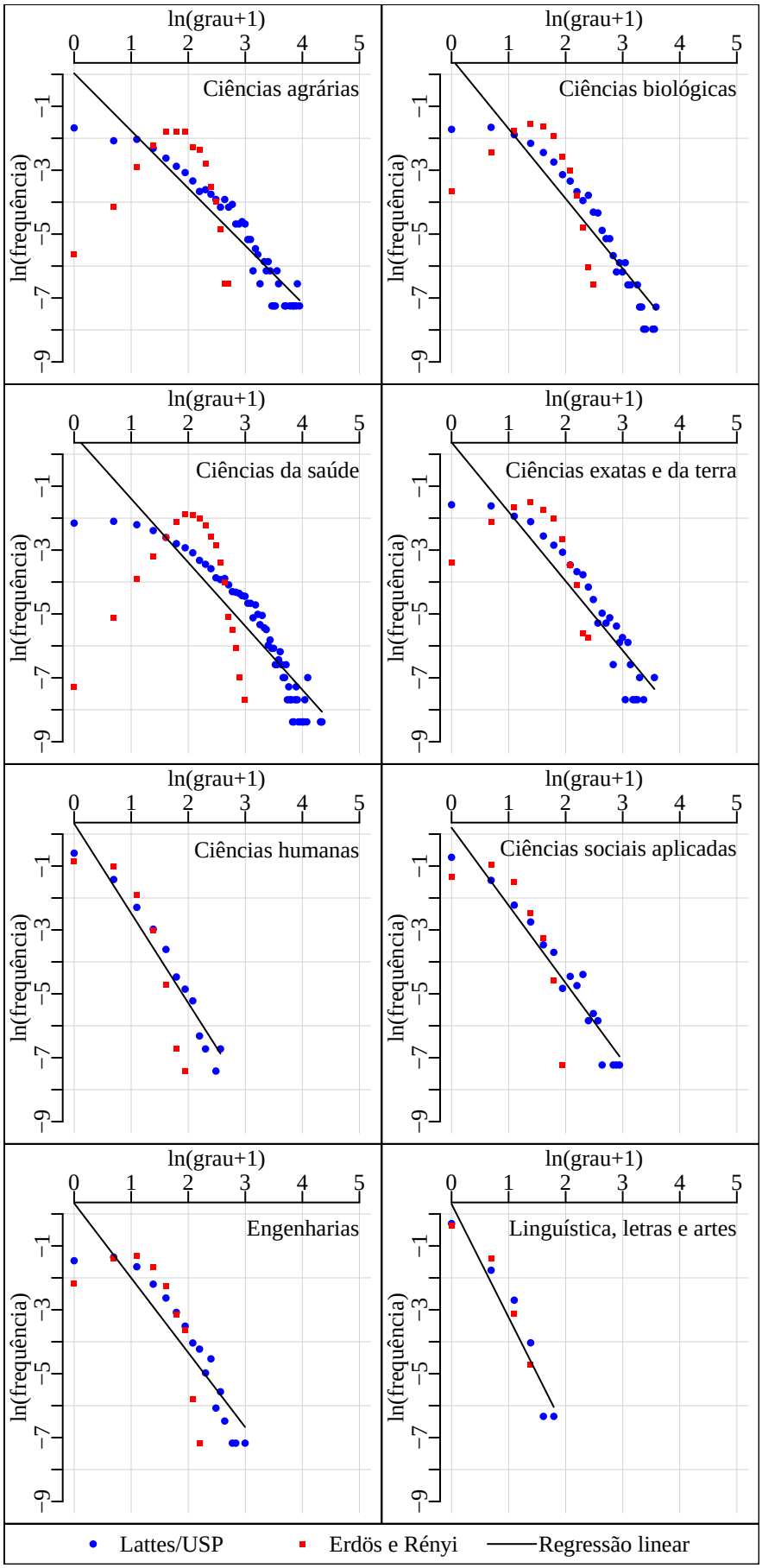

Figura 3.3 - Distribuição de grau: subgrafo Lattes/USP

Um dos objetivos deste artigo é verificar quanto à existência de fricções no processo de formação de elos de coautoria. Na medida em que existem incertezas com relação ao tipo (habilidades cognitivas e, em provavelmente maior grau, habilidades não-cognitivas) dos potenciais colaboradores, é possível que os indivíduos obtenham dentro de seus círculos de relacionamentos informações sobre atributos relevantes das demais pessoas. Uma forma indireta de obter alguma evidência da atuação deste mecanismo de referências é através do modelo proposto por Jackson e Rogers (2007). Resumidamente, neste modelo dinâmico de formação de redes sociais, um novo vértice é adicionado ao grafo ao longo do tempo. Ao nascer, ele se conecta aleatoriamente a alguns indivíuos (chamados de familiares) já presentes na rede social e a outros indivíduos na vizinhança de seus familiares. Conforme o tempo avança e novos vértices são adicionados ao grafo, um indivíduo i vai ganhando novos links seja porque ele é sorteado aleatoriamente pelos vértices entrantes, seja porque os novos vértices, na vizinhança de seus respectivos familiares com com quem i compartilha um link, encontram-no. Assim, a evolução do grau de um indivíduo i depende da quantidade de pessoas encontradas aleatoriamente e do número de pessoas encontradas socialmente. Neste modelo, a distribuição de grau, d, resultante é:

$$F_t(d) = 1 - \left(\frac{d_0 + rm}{d + rm}\right)^{1+r}, \tag{3.1}$$

para $d \geq d_0$ e para cada período t, onde d_0 é o grau inicial, m é metade do grau médio do grafo e r é a razão entre o número de links formados aleatoriamente e o número de links formados por encontros intra-rede. Assim, r fornece uma medida de aleatoriedade do grafo, a qual também pode ser interpretada como uma medida das fricções no pareamento entre os indivíduos. Quanto maior r, menor a necessidade de intermediação para que dois vértices i e j estabeleçam um link entre si.

Uma forma de estimar o parâmetro de aleatoriedade, r, é iterativamente. Para isto, primeiramente, como se está trabalhando com grafos não-direcionados, fixa-se $d_0 = m$, pois a cada período entra no conjunto de vértices um novo indivíduo que se une a m outros nós pré-existentes. Então, fixa-se um valor r_0 para o coeficiente de aleatoriedade. Em seguida, estima-se por mínimos quadrados ordinários a regressão de $\ln(1 - F_t(d))$ em $\ln(d_0 + r_0 m) - \ln(d + r_0 m)$. Desta forma, obtêm-se uma declividade β da qual se atualiza o valor inicial de r_0 para $r_1 = \beta - 1$. Então, substitui-se r_0 por r_1 na equação de regressão e se repete o procedimento até se encontrar um ponto fixo r^*. Outra possibilidade é estabelecer um *grid* de valores para r_0 e, para cada um destes valores, obter a atualização r_1 resultante da regressão de $\ln(1 - F_t(d))$ em $\ln(d_0 + r_0 m) - \ln(d + r_0 m)$. Então, escolhe-se o r_1^* que mais se aproxima do respectivo r_0^*, na vizinhança de r_0^*.

Na tabela 3.46 estão os resultados da estimação de r considerando a distribuição de grau dos pesquisadores Lattes/USP no grafo irrestrito e os resultados da estimação sobre a distribuição de grau dos mesmos pesquisadores no subgrafo Lattes/USP. Analisando estas tabelas, vale destacar o excelente ajustamento do modelo aos dados, com R^2 de no mínimo 90%. Outra observação importante é que e entre as grandes áreas do conhecimento existem diferenças significativas com relação à maneira pela qual as redes de coautoria são formadas. Por exemplo, no grafo irrestrito, enquanto nas ciências agrárias aproximadamente 67% dos links são formados aleatoriamente, nas ciências humanas este percentual é de 45%. Pela tabela também é possível perceber que as áreas do conhecimento mais quantitativas ou experimentais (ciências agrárias, ciências biológicas, ciências da saúde, engenharias e ciências exatas e da terra) possuem um grau de aleatoriedade maior na formação de suas redes de colaboração do que nas ciências humanas, nas ciências sociais aplicadas e em linguística, letras e artes. Este resultado é esperado considerando que a média de autores por artigo nestas grandes áreas mais quantitativas ou experimentais é maior do que nas outras três e se tendo em mente a forma com que frequentemente as pesquisas são desenvolvidas naqueles campos, muitas vezes com grande divisão do trabalho, realizado em laboratórios sob a coordenação de um grupo pequeno de pessoas. Nestas circunstâncias o nível de interação entre os pesquisadores envolvidos em uma pesquisa é frequentemente limitado e alguns dos coautores por vezes nem se conhecem. Se, na comunidade acadêmica, as conexões formadas socialmente refletem a aproximação de dois indivíduos intemediada por um coautor comum a eles, então é de se esperar que nestas ciências mais quantitativas ou experimentais, a participação deste método de formação de links seja relativamente baixa, afinal não raras as vezes um indivíduo possui informação insuficiente sobre os seus coautores ou influência suficiente sobre a formação de um grupo de pesquisa para que tenha condições de indicá-lo a potenciais colaboradores.

Tabela 3.46 - Aleatoriedade das redes sociais de coautoria

		Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
Grafo irrestrito	Número de autores USP	1405	2908	1306	2172	1667	565	4353	1377
	Número de autores total	54060	85726	33017	73805	34978	3278	171004	15313
	Grau médio: $2m$	56,68	42,32	31,09	48,21	23,10	5,35	65,39	12,56
	r estimado	2,01	1,88	1,61	1,64	0,84	0,96	1,58	1,11
	R^2	0,98	0,99	0,98	0,98	0,97	0,97	0,98	0,98
	Nº de observações	231	232	154	241	165	41	384	95
Subgrafo Lattes/USP	Número de autores USP	1405	2908	1306	2172	1667	565	4353	1377
	Grau médio: $2m$	5,64	3,71	2,25	3,40	0,99	0,39	7,38	1,36
	r estimado	2,17	2,92	3,92	2,77	1,94	1,65	2,82	1,74
	R^2	0,99	0,99	0,98	0,98	0,98	0,90	0,99	0,97
	Nº de observações	44	30	16	28	13	5	58	16

O argumento de que as conexões formadas socialmente possuem um conteúdo de interação social em que a formação de parcerias entre dois indivíduos é facilitada por coautores comuns que auxiliam na superação de fricções de pareamento ou na redução de assimetrias de informação entre as partes ganha força quando se comparam os resultados contidos na parte superior da tabela 3.46 com os seus correspondentes na parte inferior. Note que, sem exceção, a estimativa pontual do coeficiente de aleatoriedade é maior no subgrafo Lattes/USP do que no grafo irrestrito. Parece razoável supor que no subgrafo Lattes/USP os indivíduos possuem mais informação sobre os demais. Consequentemente, eles precisam recorrer menos à intermediação de terceiros antes de estabelecer uma parceria com um potencial coautor do que no caso do grafo irrestrito, que contém pessoas externas à USP com as quais os pesquisadores Lattes/USP em geral devem ter menos oportunidades de convivência regular e, portanto, mais dificuldade de aproximação e menos informação sobre características individuais e interesses de pesquisa. Pelo menos outros quatro motivos, não-excludentes, podem ser propostos para esta diferença no grau de aleatoriedade. O primeiro seria a existência de uma noção de comunidade entre os membros da USP, a qual prescreve certos padrões de comportamento considerados adequados. Associado a este primeiro motivo, uma segunda razão é que comportamentos considerados inadequados podem ser socialmente punidos mais severamente entre os membros do conjunto Lattes/USP. Uma terceira razão é que dentro da população Lattes/USP ocorram outros tipos de interações entre os indivíduos através das quais circulam informações a respeito de características individuais relevantes para o estabelecimento de parcerias além de intermediações por pessoas com quem futuros coautores nunca colaboraram na produção de artigos acadêmicos. Uma quarta possível explicação é que o status USP sirva de sinal sobre a qualidade dos pesquisadores, amenizando o problema de informação assimétrica entre as partes.

Para verificar a evolução da aleatoriedade da rede de coautoria em cada grande área do conhecimento, a equação (3.1) é estimada para cada período, sempre definindo um pesquisador como ativo se ele possui ao menos uma publicação no período que se estende do ano $t-5$ até o ano t. Com isto, perdem-se os cinco primeiros anos (1971-1975) como condições iniciais. As figuras 3.4 e 3.5 contêm os gráficos do coeficiente de aleatoriedade ao longo do tempo para o grafo irrestrito e para o subgrafo Lattes/USP de cada grande área do conhecimento.

Analisando os gráficos referentes à evolução da aleatoriedade dos grafos irrestritos (figura 3.4), percebe-se uma tendência de queda deste coeficiente nas redes sociais de coautoria em ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra. Em ciências humanas, a tendência geral ao longo das quatro décadas estudadas também é de decrescimento da aleatoriedade, porém nos últimos dez anos este comportamento parece

ter-se invertido. Um comportamento qualitativamente similar ao das ciências humanas é observado nas engenharias, em que, em um primeiro momento, a aleatoriedade tende a cair, mas, em seguida, esta tendência inverte-se. Além do nível e da velocidade de variação do coeficiente de aleatoriedade, o que diferencia o padrão de evolução deste nas ciências humanas e nas engenharias é que, nestas últimas, a inversão da tendência da aleatoriedade ocorreu antes, já (pelo menos de forma sustentada) na primeira metade da década de 1990. Nas ciências sociais aplicadas e em linguística, letras e artes, a aleatoriedade vem crescendo ao longo do tempo, apesar de esta última grande área do conhecimento ter experimentado uma queda no coeficiente de aleatoriedade na maior parte da década de 1990.

Os gráficos referentes ao subgrafo Lattes/USP (figura 3.5) mostram que o coeficiente de aleatoriedade tem crescido em todas as grandes áreas do conhecimento. É interessante perceber que, em ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra, o comportamento da aleatoriedade no subgrafo Lattes/USP é o oposto daquele verificado nos grafos irrestritos em que a aleatoriedade tem diminuído nestes campos. O fato de a trajetória temporal do coeficiente de aleatoriedade apresentar comportamentos diversos entre o grafo irrestrito e o subgrafo Lattes/USP sugere que as fricções do processo de formação de coautorias entre pesquisadores variam em sua natureza entre as duas populações. Em pesquisas futuras, seria interessante identificar quais são, e em que medida atuam, as dificuldades para o estabelecimento de parcerias em diferentes ambientes.

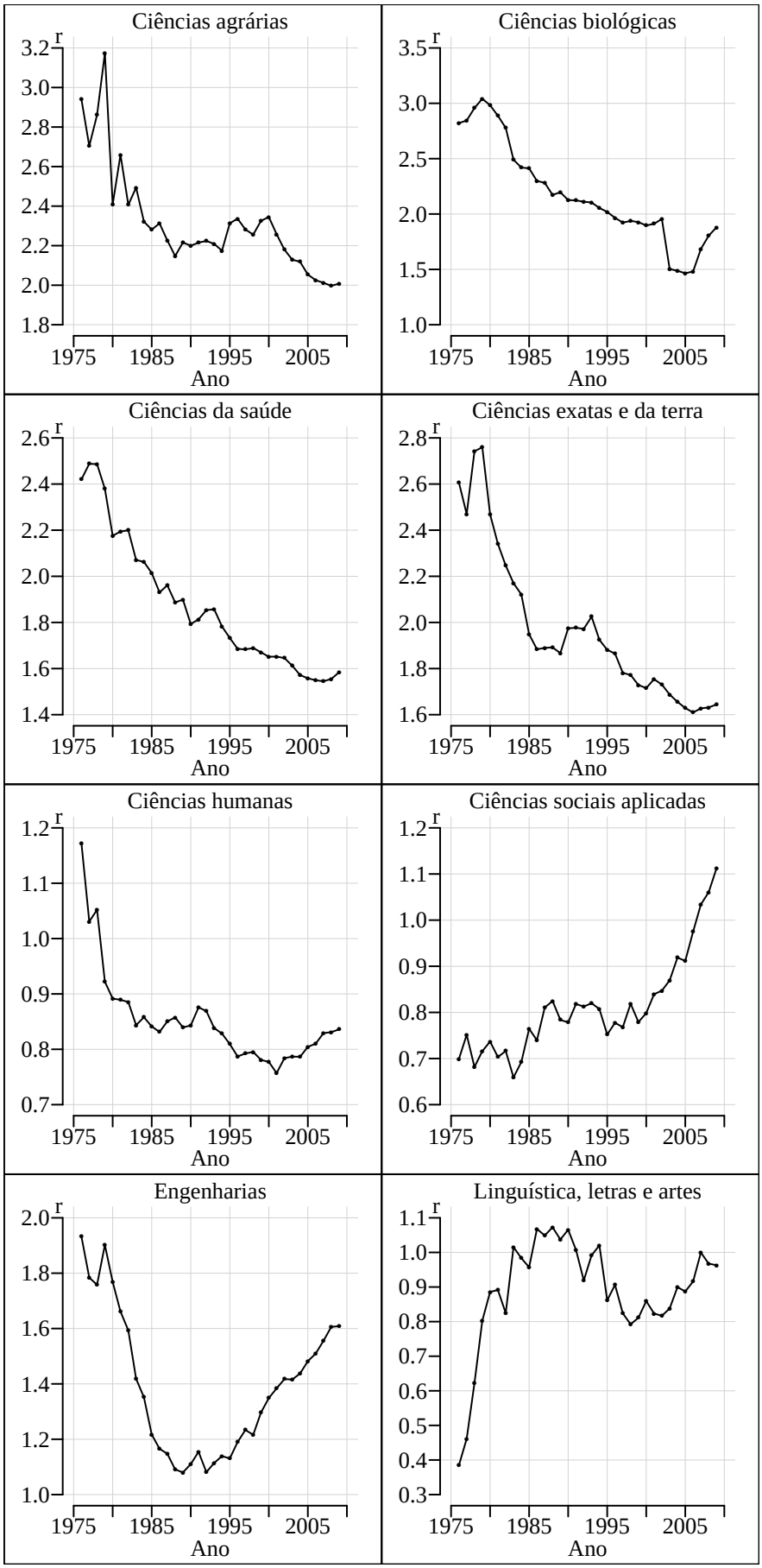

Figura 3.4 - Aleatoriedade do grafo irrestrito

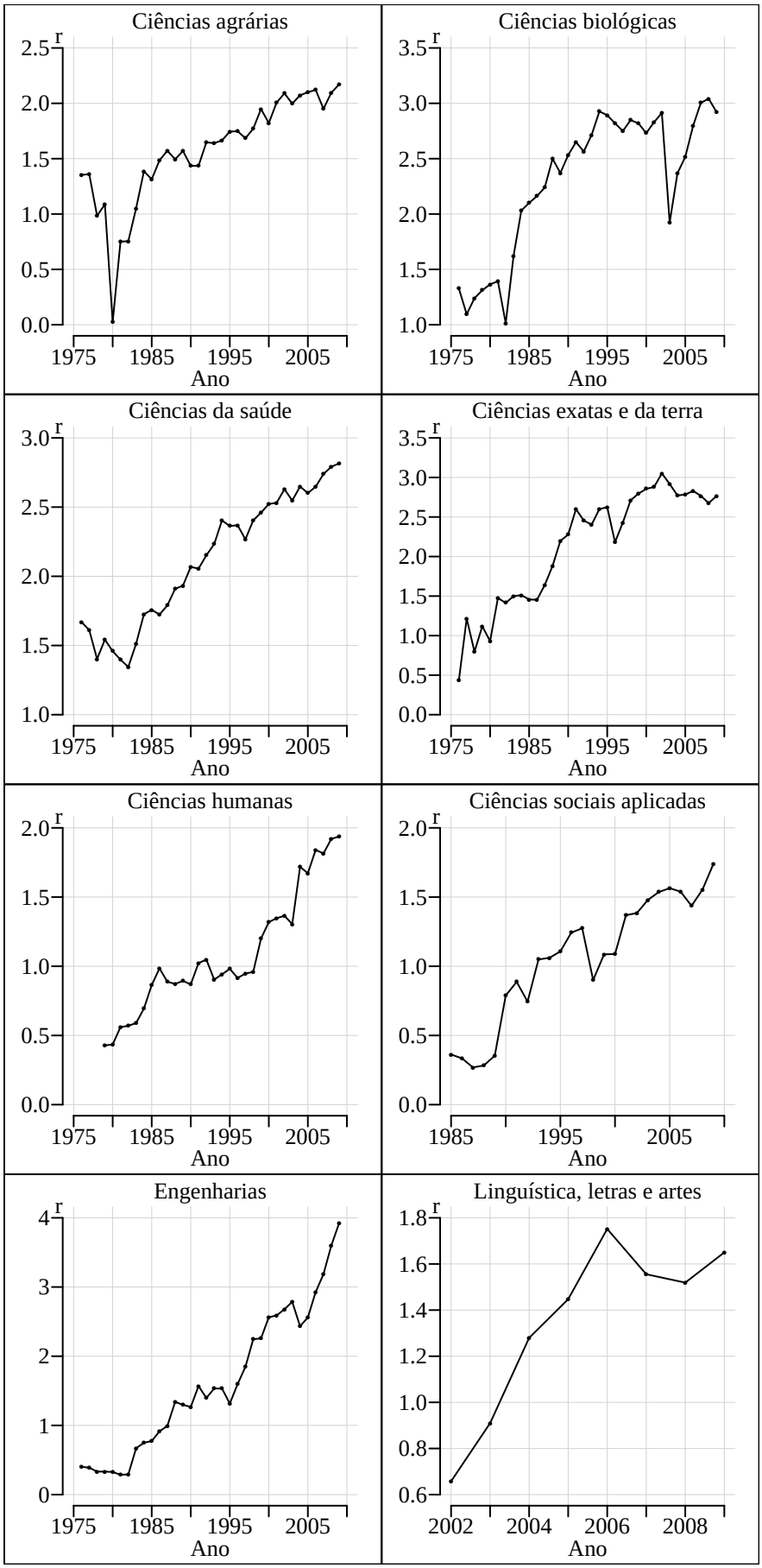

Figura 3.5 - Aleatoriedade do subgrafo Lattes/USP

3.6 Conclusões

Este artigo apresenta uma análise descritiva das redes sociais de colaboração na forma de coautoria entre pesquisadores para a produção de artigos acadêmicos nas nove grandes áreas do conhecimento consideradas pelo CNPq na Plataforma Lattes ao longo das últimas quatro décadas. Nota-se que existem diferenças importantes entre as grandes áreas com relação à estrutura de suas redes sociais de coautoria.

Ao longo do tempo, verifica-se que a estrutura das redes sociais de colaboração nas ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra convergem para aquela de mundo pequeno no triênio 2007-2009. No último triênio, as engenharias atendem a todas exigências para mundo pequeno no grafo irrestrito, mas não no subgrafo formado apenas por pesquisadores Lattes/USP. Ainda assim, considerando o crescimento da fração de pesquisadores no maior componente deste subgrafo, acredita-se que a topologia desta rede de coautoria deve preencher todas as exigências para mundo pequeno em um futuro próximo. Similarmente, nas ciências humanas, na rede social de coautoria irrestrita todas as condições para mundo pequeno são atendidas, porém o mesmo não se verifica no subgrafo Lattes/USP, pois neste a fração de pesquisadores no maior componente é bastante baixa. Como não há uma tendência clara de crescimento nesta fração no subgrafo Lattes/USP das ciências humanas, não se espera que em um futuro próximo este torne-se um mundo pequeno. A evolução da estrutura do grafo de coautoria nas ciências sociais aplicadas sugere que este deve tornar-se um mundo pequeno em breve, visto que o único requerimento não satisfeito é que a fração de pesquisadores no maior componente seja grande. Contudo, o aumento acelerado desta fração nos triênios recentes indicam que esta exigência deve ser atendida logo. Nas duas grandes áreas do conhecimento restantes, outros e linguística, letras e artes, embora as outras três condições que definem um mundo pequeno sejam respeitadas, a fração de pesquisadores no maior componente é pequena tanto no grafo irrestrito quanto no subgrafo Lattes/USP e não há indicação clara de crescimento deste indicador. Nestas condições, conclui-se que estas grandes áreas do conhecimento não são um mundo pequeno e não devem torna-se tão cedo.

Outra contribuição deste artigo é procurar extrair da distribuição de número de colaboradores informação à cerca da forma pela qual as redes de coautoria são formadas. Nesta parte do trabalho estima-se o modelo de Jackson e Rogers (2007). Os resultados indicam que a formação de elos de coautoria entre dois pesquisadores é frequentemente intermediada por coautores comuns, sugerindo a existência de fricções não-desprezíveis no pareamento de pesquisadores. Além disto, o papel da intermediação no estabelecimento de novas parcerias varia entre áreas, entre populações e ao ao longo de tempo. Até onde

se sabe, este é o primeiro trabalho a identificar estas diferenças.

A descrição da estrutura das redes sociais de coautoria não esgota o tema das interações sociais entre pesquisadores. Pelo contrário, ela deve ser encarada como um primeiro passo dentro de uma agenda de pesquisa, já que muitas questões ainda carecem de investigações mais cuidadosas. Em particular, deseja-se entender melhor como estas redes sociais são formadas. Considerando que a formação de grupos de pesquisa é sujeita a fricções de pareamento e a incertezas quanto às habilidades, sobretudo não-cognitivas, de potenciais colaboradores, as redes sociais podem ser utilizadas para promover a aproximação entre possíveis coautores, por exemplo, através da apresentação e da referência por um coautor comum. Os resultados obtidos neste artigo referentes ao grau de aleatoriedade dos grafos de coautoria sugerem que estes mecanismos de aproximação intra-rede devem ser importantes. Cabe agora aprofundar a investigação sobre o tema.

4 EFEITOS DE REDES SOCIAIS NA ACADEMIA BRASILEIRA

4.1 Introdução

A colaboração formal entre pesquisadores ocorre por diversas razões, tais como habilidades complementares, interesses de pesquisa comuns e afinidade. Embora estas variáveis ajudem a prever a compatibilidade entre potenciais membros de um grupo de pesquisa, ainda resta a questão de que parcerias mutuamente benéficas podem não se realizar em decorrência de fricções de pareamento e de assimetrias informacionais. De forma a amenizar estas dificuldades, os indivíduos podem beneficiar-se de suas redes de contatos sociais na medida em que seus conhecidos facilitam o encontro com outras pessoas e transmitem informações relevantes sobre atributos individuais de potenciais colaboradores.

A inserção em redes sociais do processo de pareamento é amplamente documentada em outros contextos, como no mercado de trabalho (GRANOVETTER, 1973; IOANNIDES; LOURY, 2004; MONTGOMERY, 1991; TOPA, 2001; WAHBA; ZENOU, 2005), porém ainda são poucos os trabalhos que procuram mensurar a sua importância para a colaboração acadêmica. Este artigo tem exatamente este objetivo: verificar em que medida a formação de novos elos de coautoria é facilitada pelas conexões pré-existentes entre os pesquisadores. Especificamente, procura-se mensurar a resposta da probabilidade de que dois pesquisadores venham a colaborar um com o outro na produção de um artigo acadêmico a mudanças no número e no comprimento das geodésicas[1] entre eles na rede social de coautoria. Com isto, contribui-se para o estudo empírico das redes sociais, para a economia das interações sociais e para a ciência da ciência.

É importante ressaltar que a principal referência deste artigo é Fafchamps, Goyal e van der Leij (2010), cuja estratégia econométrica é seguida de perto. Naquele trabalho, os autores encontram que, entre economistas, existem fortes efeitos de proximidade no sentido de que conforme a distância social (no grafo de colaboração) entre dois pesquisadores diminui, a probabilidade de que eles estabeleçam entre si um elo de colaboração inédito[2] aumenta substancialmente. Como argumentado pelos autores, a identificação de efeitos de proximidade nas redes sociais de coautoria em artigos acadêmicos é importante porque

[1] Uma geodésica entre dois vértices de um grafo é um menor caminho entre eles.
[2] Que não tenha sido realizado nos dez anos anteriores.

a formação de parcerias entre pesquisadores ocorre em um ambiente em que há pouca assimetria de informação sobre habilidades cognitivas e interesses de pesquisa dos indivíduos. Nestas condições, a presença destes efeitos indica que as redes de contatos sociais desempenham papel significativo no pareamento de agentes em situações em que informações individuais relevantes para o estabelecimento de parcerias não são tão detalhadas e facilmente acessíveis quanto no meio acadêmico.

Uma diferença entre este artigo e Fafchamps, Goyal e van der Leij (2010) é que aqui a análise não se restringe à rede social de colaboração entre economistas. Ao invés disto, estudam-se oito grandes áreas do conhecimento: ciências agrárias, ciências biológicas, ciências da saúde, ciências exatas e da terra, ciências humanas, ciências sociais aplicadas, engenharias e linguística, letras e artes. Para isto, utilizam-se dados extraídos das informações contidas nos currículos Lattes dos pesquisadores que citam a Universidade de São Paulo (USP) em qualquer uma das seguintes três opções: (1) endereço profissional ou (2) formação acadêmica/titulação ou (3) atuação profissional. Estes dados permitem construir as redes sociais de coautoria destes pesquisadores, as quais são então restritas apenas aos subgrafos Lattes/USP, que contêm apenas os indivíduos vinculados à USP em algumas das três opções enumeradas e os links entre eles.

A figura 4.1 exemplifica as redes sociais de coautoria estudadas. Suponha, apenas para efeito ilustrativo, que no painel esquerdo esteja desenhada uma rede social de relacionamentos entre pesquisadores. Em círculos azuis são representados os pesquisadores vinculados à USP e em quadrados vermelhos são representados os pesquisadores externos à USP. Existem relacionamentos de diferentes naturezas entre os pesquisadores das duas populações. Em arestas cheias pretas estão indicadas as relações de coautoria e em arestas tracejadas verdes são indicados outros tipos de relacionamentos. Como não existe informação disponível sobre estes outros relacionamentos, trabalha-se apenas com o grafo de coautoria, o que implica que se ignoram as arestas verdes tracejadas. Além disto, restringe-se este grafo de coautoria apenas às colaborações entre pesquisadores vinculados à USP. Assim, o grafo resultante, que, no caso exemplificado, constituiria o objeto de estudo, é o apresentado no painel direito da figura 4.1.

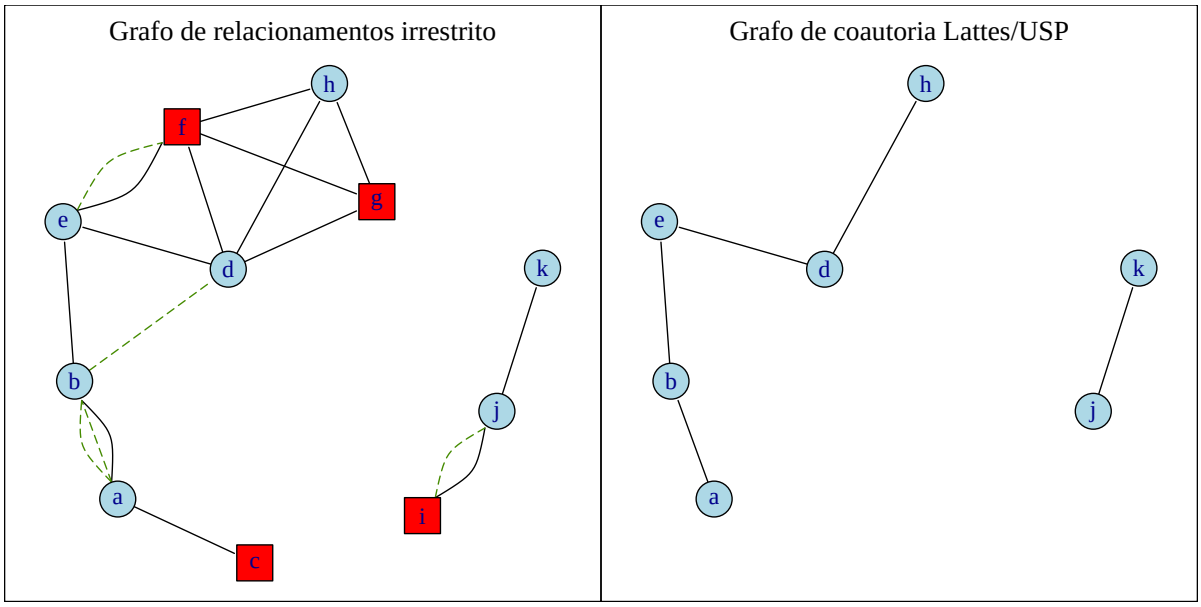

Figura 4.1 - Grafo de coautoria Lattes/USP

O fato de as redes sociais de coautoria serem restritas a pesquisadores com algum vínculo a uma mesma universidade implica que os links são formados em um ambiente extremamente favorável à colaboração, em que as fricções de pareamento e as assimetrias informacionais são mínimas, menos contudentes do que no ambiente em que os elos de coautoria estudados por Fafchamps, Goyal e van der Leij (2010) são formados. Um dos motivos para isto é que, considerando que todos os indivíduos inseridos nestas redes possuem filiação à mesma universidade em algum momento de suas carreiras, é razoável supor que, em geral, o conhecimento que eles têm sobre os seus potenciais colaboradores é ao menos tão bom quanto o que teriam em uma rede social irrestrita, que contém indivíduos de outras instituições. Este conhecimento mais preciso pode ser devido, por exemplo, à convivência frequente nas dependências da universidade. Outra razão pela qual a formação de parcerias entre os indivíduos considerados deve ser pouco afetada pelas fricções de pareamento e pelas assimetrias informacionais é que, pelo fato de os pesquisadores terem vínculo com a USP, em uma parcela significativa dos casos, no momento da primeira parceria entre dois indivíduos, a distância física entre eles era pequena. Isto tipicamente acontece quando os pesquisadores são contemporâneos na USP ao estabelecerem a primeira colaboração. Um terceiro motivo para a relativa facilidade na formação dos links de colaboração entre os pesquisadores estudados é a existência de um sentimento de identidade entre eles (COHEN; FRAZZINI; MALLOY, 2008). Como reportado em alguns estudos (GOMPERS; MUKHARLYAMOV; XUAN, 2012), muitos relacionamentos são altamente homofílicos em termos de local de educação, de participação em clubes e outras associações. Por fim, estes indivíduos por serem relacionados com a USP, que é a universidade brasileira de maior prestígio, carregam consigo uma reputação de qualidade que os torna mais atraentes para todos os demais, inclusive para outras pessoas associadas à

USP.

Os resultados encontrados mostram que existem diferenças importantes entre os campos do conhecimento estudados com relação aos efeitos de proximidade no estabelecimento de coautorias. Enquanto nas ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra, mudanças no número e no comprimento das geodésicas afetam a probabilidade de iniciação de uma parceria acadêmica, nas outras quatro grandes áreas do conhecimento estes efeitos não estão presentes. O impacto da distância é grande. Por exemplo, ao reduzir o número de links entre dois pesquisadores de 3 para 2, a probabilidade de que eles iniciem uma colaboração aumenta em 74,89% nas ciências agrárias, 43,69% nas ciências biológicas, 32,84% nas ciências da saúde e 62,58% nas ciências exatas e da terra. Já com relação ao impacto marginal do número de geodésicas, nas ciências biológicas ele é de 12,86% e nas ciências exatas e da terra de 11,18%. A estratégia econométrica adotada, proposta por Fafchamps, Goyal e van der Leij (2010), permite concluir que estes efeitos resultam dos fluxos de informações sobre atributos individuais e sobre qualidade de pareamentos passados através das redes de colaboração.

O restante do artigo está organizado da seguinte maneira. A seção 4.2 descreve os dados e as variáveis. A seção 4.3 discute as estatísticas descritivas. A seção 4.4 explica a estratégia econométrica. A seção 4.5 discute os resultados das estimações. A seção 4.6 conclui.

4.2 Dados e definição das variáveis

A fonte de dados utilizada é a Plataforma Lattes mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) que contém o currículo de essencialmente todos os pesquisadores ativos no Brasil. Deste total de mais de 1,7 milhão de currículos, utiliza-se o subconjunto Lattes/USP, que inclui o currículo de 62.516 indivíduos que citam a Universidade de São Paulo em ao menos uma das seguintes três condições: (1) endereço profissional ou (2) formação acadêmica/titulação ou (3) atuação profissional.

A partir do histórico de publicação dos indivíduos no período 1971-2009, para cada ano t, encontra-se o conjunto de pesquisadores ativos do conjunto Lattes/USP, V_t. Um pesquisador i é considerado ativo a partir do momento de sua primeira publicação, denotado por t_0^i, até o ano de sua última publicação, t_2^i. Note que o conjunto de pesquisadores ativos não é constante ao longo do tempo em decorrência da entrada de novos pesquisadores ativos e da saída de pesquisadores improdutivos. Para cada par ij de pesquisadores

ativos em t (ou seja: tais que $\max\{t_0^i, t_0^j\} \leq t \leq \min\{t_2^i, t_2^j\}$), a variável binária y_t^{ij} assume valor unitário se os pesquisadores colaboraram em t e zero caso contrário. Assim, denotando por t_1^{ij} o ano da primeira colaboração entre i e j, $y_t^{ij} = 0$ para todo t tal que $\max\{t_0^i, t_0^j\} \leq t < t_1^{ij}$ e $y_t^{ij} = 1$ para $t = t_1^{ij}$. Sendo t_2^{ij} o ano da última colaboração entre i e j, para $t \in (t_1^{ij}, t_2^{ij}]$, $y_t^{ij} = 1$ se existe uma coautoria entre i e j no ano t e $y_t^{ij} = 0$, caso contrário. No modelo econométrico estimado adiante, y será a variável dependente.

Agora, definem-se as variáveis explicativas. Primeiramente, para cada ano t obtém-se a rede social de coautoria G_t, composta pelo conjunto de pesquisadores do conjunto Lattes/USP que possuem ao menos uma publicação no período de $t - 9$ até t e pelos links entre eles. Laços e arestas repetidas são ignorados e, portanto, G_t é um grafo simples. Repare que, pelo fato de os nove primeiros anos de observações serem utilizados para o cálculo do primeiro grafo de colaboração, obtém-se uma sequência de redes de coautoria que se estende de 1980 até 2009. Para cada rede social de coautoria G_t calculam-se a distância e o número de menores caminhos entre cada par de seus vértices. Estas são as duas variáveis explicativas de maior interesse porque o estudo do impacto delas sobre a variável dependente permite fazer inferências sobre o fluxo de informações em G_t. É importante salientar que, por razões detalhadas na seção referente à estratégia econométrica, a distância e o número de geodésicas entre dois pesquisadores não foram calculados da maneira usual. Ao invés disto, para cada par de pesquisadores, estas variáveis foram calculadas excluindo-se do grafo G_t apenas a aresta que os conecta, caso ela esteja presente em G_t. Assim, a distância entre dois pesquisadores é sempre maior ou igual a dois. Quando não existe um caminho em G_t que conecte dois indivíduos, a distância entre eles é infinita. Para evitar este problema, trabalha-se com o inverso da distância, que se encontra no intervalo $[0, 1/2]$. Esta medida é chamada de proximidade. Portanto, as duas variáveis de rede utilizadas para explicar y_t^{ij} são proximidade e número de geodésicas, denotadas por p_t^{ij} e g_t^{ij}, respectivamente.

Aqui é importante que se faça uma observação. Da forma como a variáveis foram definidas e as redes sociais de coautoria construídas, ignoram-se as defasagens entre o início de um projeto de pesquisa e a publicação do artigo dele derivado. Esta é uma fragilidade da análise desenvolvida porque o intervalo de tempo envolvido em todo o ciclo desde o começo de uma pesquisa até a sua publicação varia entre pesquisas, entre campos do conhecimento e ao longo do tempo.

Além das variáveis de rede, outros determinantes de qualidade de pareamento são importantes para explicar a colaboração entre dois autores. Para entender como estes determinantes são definidos, é necessário ter em mente que se deseja estimar a probabilidade de que um par ij de indivíduos esteja conectado em algum período t via uma relação de

coautoria, $y_t^{ij} = 1$. Assim, sendo o objeto de estudo o link (a sua presença ou ausência em G_t), o qual incide em dois pesquisadores distintos (já que os grafos estudados são simples), as variáveis explicativas são diáticas, referentes a pares de pesquisadores. Feita esta observação, o primeiro determinante considerado é a produtividade dos pesquisadores. Para medí-la, utiliza-se a avaliação Qualis de periódicos elaborada anualmente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Em decorrência de mudanças de metodologia e de variações no conjunto de periódicos avaliados ao longo de tempo, atribui-se a cada publicação a nota Qualis referente ao ano de 2008. Note que, ao fixar o valor das publicações em uma dada revista à nota da mesma em um determinado ano (no caso 2008), ignoram-se mudanças intertemporais em sua qualidade. Em 2008, as notas atribuídas aos periódicos pela avaliação Qualis eram distribuídas em oito estratos: A1, A2, B1, B2, B3, B4, B5 e C. A partir deste espectro, define-se arbitrariamente uma escala de qualidade que começa em 1 no caso dos periódicos de nota igual a C e que dobra conforme se avança a cada estrato sucessivo de qualidade de tal forma que uma publicação em um periódico A1 vale 128 pontos. Ao adotar o critério, que se admite arbitrário, de dobrar o valor das revistas entre estratos da avaliação Qualis/2008 segue-se Fafchamps, Goyal e van der Leij (2010), que usam uma outra fonte primária de avaliação, a qual inclui apenas três níveis de qualidade: AA, A e B. Nestas condições, definem-se os pontos recebidos por um pesquisador em cada uma de suas publicações da seguinte maneira:

$$q := \frac{\text{Índice de qualidade}}{\text{Número de autores} + 1}. \tag{4.1}$$

É comum na literatura (FAFCHAMPS; GOYAL; VAN DER LEIJ, 2010; LABAND; TOLLISON, 2000; MCDOWELL; MELVIN, 1983; SAUER, 1988) ponderar a qualidade dos artigos pelo seu número de páginas. Esta prática não é adotada neste trabalho porque implicaria a perda de muitas observações já que o conjunto de autores que informam as páginas inicial e final de seus artigos constitui uma minoria dentro da população de pesquisadores Lattes/USP. Tendo encontrado a sequência de pontos obtidos pelos pesquisadores em cada uma de suas publicações, calcula-se para cada ano t o número de pontos acumulados por eles nos nove anos anteriores, q_t^i, de onde se obtém a produtividade média de cada par de autores:

$$\bar{q}_t^{ij} := \frac{q_t^i + q_t^j}{2}. \tag{4.2}$$

Esta será uma das variáveis explicativas do modelo econométrico, discutido na seção 4.4. Intuitivamente, quanto maior a produtividade de um pesquisador, mais atraente ele

deve ser para potenciais coautores. Sendo assim, à medida que a produtividade de um par aumenta, maior deve ser a probabilidade de que este par forme uma parceria. Por outro lado, pesquisadores com alta produção devem ser mais seletivos na escolha de seus coautores. Sendo assim, é de se esperar que a probabilidade de um par estabelecer um elo de coautoria seja decrescente no módulo da diferença de produtividade entre os autores:

$$\Delta q_t^{ij} := \left| q_t^i - q_t^j \right| . \tag{4.3}$$

Uma terceira variável importante para a colaboração entre dois indivíduos é a intersecção de interesses pesquisa. Nos currículos Lattes, os indivíduos têm a possibilidade de listar até três campos de estudos aos quais cada uma de suas publicações contribui partindo desde o nível mais alto de agregação, de grande área do conhecimento, até o nível mais específico, de especialidade. Com base nestas informações, calcula-se o índice cosseno de similaridade[3], o qual, sendo $x_{t,f}^i$ a fração de artigos escritos pelo indivíduo i no campo f entre os anos $t-9$ e t, é definido da seguinte maneira:

$$\omega_{t,f}^{ij} := \frac{\sum_f x_{t,f}^i x_{t,f}^j}{\sqrt{\left(\sum_f (x_{t,f}^i)^2\right)\left(\sum_f (x_{t,f}^j)^2\right)}}, \tag{4.4}$$

$\sum_f x_{tf}^i = 1$. Esta medida varia entre 0, no caso de os pesquisadores nunca terem publicado em um campo de estudo comum, e 1 quando os pesquisadores publicaram nos mesmos campos de estudo e nas mesmas proporções. Em certo sentido, a medida de similaridade cosseno capta o grau de substitutabilidade entre os pesquisadores i e j.

O sinal do impacto de variações na similaridade em interesses de pesquisa sobre a probabilidade de que os pesquisadores i e j trabalhem em coautoria não é óbvio. Por um lado, é possível que os pesquisadores procurem explorar complementaridades em seus campos de expertise. Na medida em que estes se refletem nos campos aos quais eles contribuem, variações em $\omega_{t,f}^{ij}$ devem resultar em variações com sinal oposto na probabilidade de coautoria. Por outro lado, $\omega_{t,f}^{ij}$ pode ter um impacto positivo sobre colaboração por pelo menos outros três motivos: (i) colaboração entre professor e ex-aluno, (ii) aversão ao risco e (iii) facilidade de comunicação. No primeiro caso, é razoável supor que haja alta coincidência de interesses de pesquisa entre os dois. No segundo, levando-se em consideração que um projeto de pesquisa demora algum tempo para ser concluído, é possível

[3] A medida de similaridade cosseno é bastante utilizada em ciências da computação com o propósito de mineração de textos.

que algum grau de substitutabilidade entre colaboradores seja útil no sentido de permitir uma redistribuição do esforço empregado por eles ao longo de sua elaboração como resposta a choques, por exemplo, de disponibilidade de tempo. Desta forma, ao fazerem um seguro mútuo, os colaboradores conseguem evitar que o andamento de seus projetos de pesquisa seja comprometido. Já no terceiro caso, auto-explicativo, maior similaridade em campos de expertise pode facilitar a comunicação entre os colaboradores (ROGERS; BHOWMIK, 1971). Neste artigo, utiliza-se a medida de similaridade cosseno em três níveis de especificidade: grande área do conhecimento, área do conhecimento e subárea do conhecimento[4].

Uma outra variável explicativa para a coautoria é a propensão a colaborar. Na medida em que esta propensão é invariante no tempo, ela será absorvida pelo efeito idiossincrático de par adicionado ao modelo econométrico. Porém, a propensão a colaborar não é constante, já que existem evidências de ciclos de carreira na participação em coautoria (LEVIN; STEPHAN, 1991; MCDOWELL; MELVIN, 1983; PIETTE; ROSS, 1992). Por este motivo, como medida de propensão a colaborar do pesquisador i no ano t, calcula-se o número de coautores distintos que ele teve nove anos anteriores, n_t^i. Analogamente ao que foi feito para as variáveis de produtividade, para cada par de pesquisadores, calculam-se a média e a diferença absoluta da propensão a colaborar, denotadas respectivamente por: $\bar{n}_t^{ij}$ e Δn_t^{ij}.

É importante destacar que são calculadas duas medidas de propensão individual a colaborar, uma restrita a coautores pertencentes ao conjunto Lattes/USP e outra irrestrita, incluindo também pesquisadores externos a esta população. Faz-se esta distinção entre número total de colaboradores e número de colaboradores internos à população Lattes/USP por dois motivos. O primeiro é que, pelo fato de o sistema de currículos Lattes ter sido implantado no final de década de 1990, é possível que um indivíduo pertecente à população Lattes/USP cite entre seus coautores pesquisadores que tinham relação com a USP mas que, por exemplo, aposentaram-se antes de a utilização do sistema de currículos Lattes difundir-se. Como alguns destes pesquisadores, inativos a partir do final de década de 1990, podem não ter criado os seus currículos Lattes, eles são considerados externos ao conjunto de pesquisadores Lattes/USP. O segundo motivo é que, até recentemente, muitos estudantes, mesmo os de pós-graduação, não tinham currículo Lattes. Por esta razão, um ex-aluno que publicou um trabalho com um professor, mas que não seguiu a carreira

[4]Os pesquisadores possuem grande liberdade para adicionarem campos aos quais seus artigos contribuem nos níveis de especificidade inferiores ao de grande área do conhecimento. Por este motivo, é possível que um mesmo campo seja citado com nomes diferentes, por exemplo, por erros de digitação ou pequenas variações na grafia. Para contornar este problema, foi feito um *matching* aproximado de Levenshtein, em que dois campos de pesquisa de mesmo nível de especificidade são considerados idênticos se $0 \leq \frac{lev(i,j)}{max(comp(i),comp(j))} \leq 10\%$, onde $lev\,(i,j)$ é a distância de Levenshtein entre eles e $comp\,(x)$ é o número de caracteres no nome do campo $x \in \{i,j\}$.

acadêmica após a formatura, pode não ter currículo Lattes e, portanto, ser considerado erroneamente externo à população Lattes/USP. Tendo em mente que o conjunto verdadeiro de pesquisadores relacionados à USP contém o conjunto de indivíduos pertencentes ao conjunto Lattes/USP observado, a medida de propensão individual a colaborar é subsestimada. Mais ainda, por estes mesmos dois motivos, os próprios grafos de colaboração Lattes/USP utilizados são subgrafos das verdadeiras redes sociais de coautoria geradas pelas pessoas que possuem relacionamento com a USP e pelos elos de coautoria entre elas.

4.3 Estatísticas descritivas

Na tabela 4.1, apresentam-se as estatísticas descritivas das variáveis usadas neste artigo. Primeiramente, nota-se que o número de pares de pesquisadores que coautoram mais de um artigo é, em média, 25% do número de pares que colaboram ao menos uma vez. Em seguida, reportam-se as médias referentes às duas primeiras colaborações entre pares de pesquisadores de cada uma das grandes áreas do conhecimento. Nota-se que, a partir do momento em que dois pesquisadores são ativos, o tempo médio transcorrido até que o primeiro artigo coautorado entre eles é publicado varia entre 4,86 e 7,61 anos. Já o número de anos entre a primeira publicação colaborativa e a segunda encontra-se na faixa de 2,5 anos. Esta redução no tempo entre as duas primeiras coautorias pode ser devida a pelo menos três fatores: (i) o segundo artigo coautorado por um par pode ser derivado do primeiro, (ii) aproximação das áreas de interesse dos autores após eles terem colaborado e (iii) identificação da qualidade do pareamento, de tal forma que duplas com boa sinergia tendem a ser repetidas.

Com relação à proximidade entre os pesquisadores, fica claro que, entre os indivíduos caminho-conectados, ela é alta tanto na primeira quanto na segunda colaboração, refletindo a existência de, em média, dois ou três links intermediários. Além disto, não se percebe uma mudança significativa entre as duas primeiras publicações, embora a proximidade no ano anterior ao da publicação do segundo artigo coautorado seja sempre maior ou igual à proximidade no ano anterior da publicação do primeiro. Quanto ao número de geodésicas, fica evidente que não há muitos menores caminhos entre dois coautores já que esta estatística não ultrapassa três em média. Além disto, entre a primeira e a segunda colaboração apenas as ciências exatas e da terra, as ciências da saúde e as ciências sociais aplicadas apresentam suave redução no número de geodésicas entre pares caminho-conectados e apenas as duas primeiras entre quaisquer pares, caminho-conectados ou não.

Tabela 4.1 - Estatísticas descritivas

		Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
	Nº pares colab.	4029	5551	1523	3764	937	117	16438	946
	Nº pares colab. reincidentes	1049	1125	405	974	256	23	4603	209
1ª colaboração	Demora (em anos)	6,77	5,89	4,86	7,14	6,81	5,26	7,61	5,37
	Proximidade: todos os pesquisadores	0,29	0,20	0,13	0,20	0,07	0,05	0,32	0,11
	Proximidade: caminho-conectados	0,40	0,36	0,43	0,36	0,42	0,50	0,40	0,34
	Nº de geodésicas: todos os pesquisadores	1,91	1,08	0,41	1,10	0,19	0,09	2,25	0,43
	Nº de geodésicas: caminho-conectados	2,64	1,94	1,36	1,96	1,23	1,00	2,79	1,37
	Produtividade média	7,81	9,59	9,25	9,08	18,56	35,00	5,94	6,21
	Dif. produtividade	11,56	15,70	15,90	13,86	29,49	56,20	9,17	9,90
	Média de coautores: Total	109,95	108,99	41,74	82,56	28,44	14,35	142,23	18,22
	Dif. de coautores: Total	106,78	114,19	39,92	85,85	32,14	13,96	149,87	19,67
	Média de coautores: USP	0,03	0,03	0,03	0,03	0,01	0,00	0,04	0,08
	Dif. de coautores: USP	0,06	0,06	0,06	0,06	0,02	0,00	0,08	0,16
	Cosseno grande área	0,92	0,86	0,87	0.92	0.94	0,98	0,93	0,95
	Cosseno área	0,83	0,61	0,69	0,84	0,80	0,75	0,83	0,89
	Cosseno subárea	0,53	0,36	0,45	0,60	0,34	0,49	0,38	0,51
2ª colaboração	Demora (em anos)	2,30	2,42	2,58	2,45	2,53	2,80	2,51	2,44
	Proximidade: todos os pesquisadores	0,38	0,27	0,23	0,27	0,12	0,08	0,39	0,19
	Proximidade: caminho-conectados	0,47	0,45	0,48	0,45	0,45	0,50	0,46	0,44
	Nº de geodésicas: todos os pesquisadores	2,13	1,21	0,66	1,02	0,35	0,16	2,18	0,58
	Nº de geodésicas: caminho-conectados	2,66	2,02	1,41	1,72	1,32	1,00	2,61	1,35
	Produtividade média	8,26	9,56	8,86	9,67	15,60	40,09	6,13	8,42
	Dif. produtividade	11,89	14,92	14,41	14,36	22,74	56,74	9,11	12,57
	Média de coautores: Total	112,93	99,86	43,39	79,68	34,17	19,18	149,66	25,84
	Dif. de coautores: Total	114,94	105,30	45,38	86,54	38,55	14,94	157,91	29,87
	Média de coautores: USP	0,06	0,10	0,13	0,12	0,05	0,00	0,07	0,11
	Dif. de coautores: USP	0,11	0,20	0,27	0,25	0,11	0,00	0,13	0,18
	Cosseno grande área	0,94	0,87	0,84	0,92	0,83	0,85	0,94	0,97
	Cosseno área	0,86	0,70	0,78	0,88	0,80	0,84	0,86	0,90
	Cosseno subárea	0,63	0,44	0,55	0,68	0,42	0,51	0,47	0,59

Para ganhar uma percepção um pouco mais refinada sobre a relação entre as variáveis de rede (proximidade e número de geodésicas) e o início de uma relação de parceria acadêmica, é conveniente analisar os gráficos das figuras 4.2 e 4.3. A primeira mostra, para cada uma das oito grandes áreas do conhecimento, a aproximação média entre dois pesquisadores anteriormente à primeira publicação colaborativa entre eles. Em outras palavras, o que os gráficos da figura 4.2 apresentam é a proximidade média para cada ano anterior à primeira colaboração. Estes cálculos são feitos para os pares de pesquisadores que são caminho-conectados em G_{t-k} e para todo o conjunto de pares de pesquisadores em G_{t-k}, independentemente de existir um caminho entre eles no grafo. Como se pode perceber, quando não se restringe o cálculo da proximidade média aos pares de indivíduos caminho-conectados, em todas a grandes áreas do conhecimento, a defasagens não muito grandes, existe uma clara tendência de aumento da proximidade entre futuros colaboradores conforme o momento da primeira parceria entre eles fica mais perto. Porém, a velocidade desta aproximação varia entre os campos. Por outro lado, quando a proximidade média é calculada apenas entre pares de pesquisadores pertencentes a um mesmo componente de G_{t-k}, nota-se que a tendência de crescimento da proximidade entre futuros coautores desacelera, chegando a desaparecer ou a se tornar menos clara em engenharias, ciências humanas, ciências sociais aplicadas e linguística, letras e artes.

A segunda figura contém os gráficos entre o número médio de menores caminhos entre dois pesquisadores antes da primeira coautoria entre eles tanto para pares de pesquisadores caminho-conectados como para todos os pares, pertencentes ou não a um mesmo componente de G_{t-k}. A aparência geral dos gráficos com os cálculos irrestritos é similar à aparência dos gráficos correspondentes da figura 4.2, indicando que, pelo menos a defasagens não muito grandes, maior proximidade temporal em relação à primeira parceria está associada a maior número de geodésicas. Já quando se analisam apenas os pares caminho-conectados, a tendência de crescimento desaparece ou se enfraquece significativamente a defasagens menores do que seis em todos os campos, com exceção das ciências biológicas e das ciência sociais aplicadas. Nas ciências humanas, a pequenas defasagens, a relação entre o número de geodésicas e a aproximação do momento da primeira colaboração é negativa quando se consideram apenas pares de pesquisadores caminho-conectados.

Os gráficos das figuras 4.2 e 4.3 fornecem evidência preliminar da existência de fricções de pareamento entre pesquisadores da população Lattes/USP, pois pesquisadores que estão mais próximos ou que têm maior número de geodésicas entre eles tendem a colaborar mais rapidamente, pelo menos em alguns campos como ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra. Caso tais fricções de pareamento não existissem, o tempo transcorrido até a primeira colaboração entre dois pesquisadores deveria ser independentemente da distância e do número de geodésicas de um até o outro.

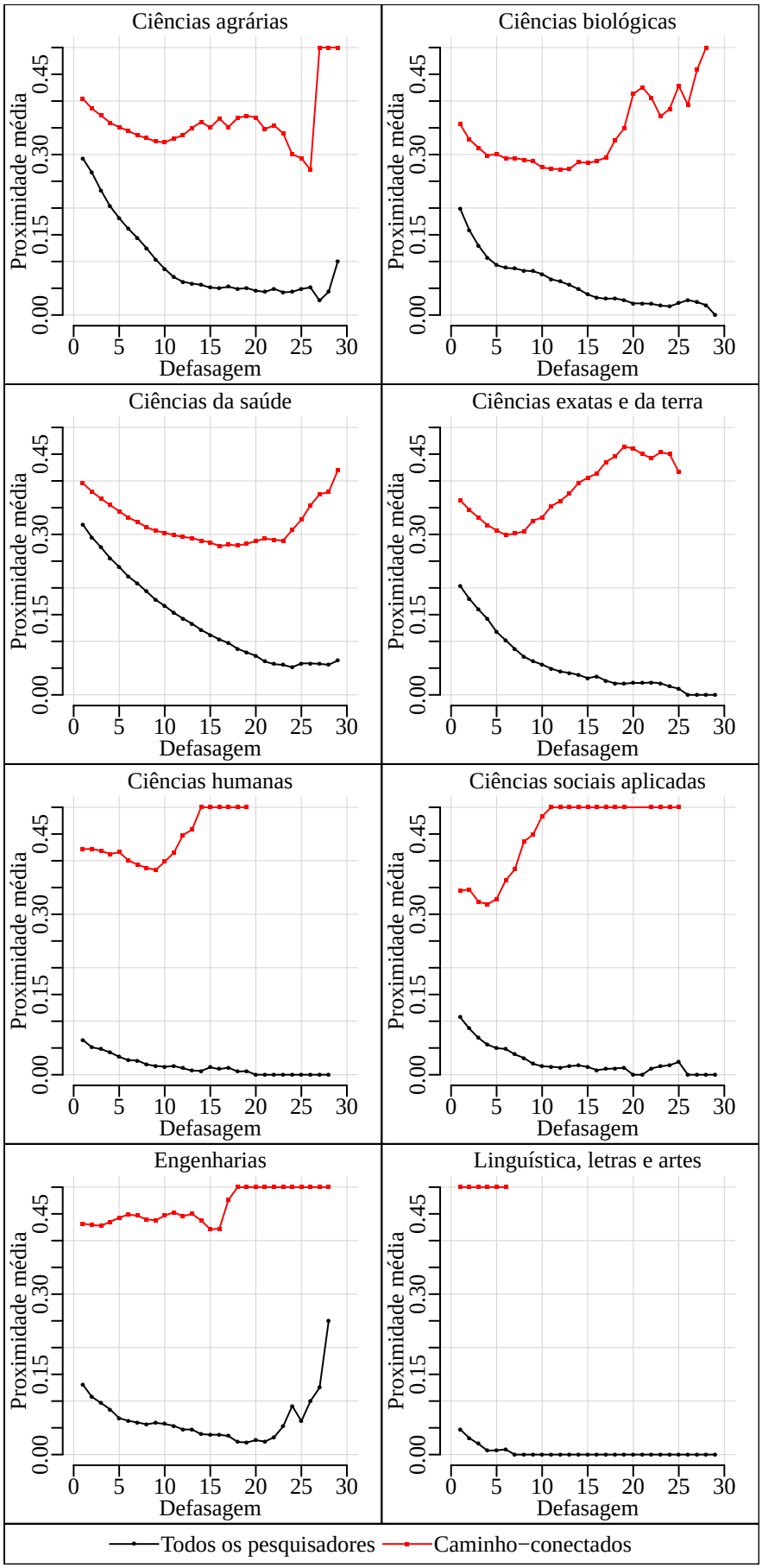

Figura 4.2 - Proximidade média antes da primeira colaboração

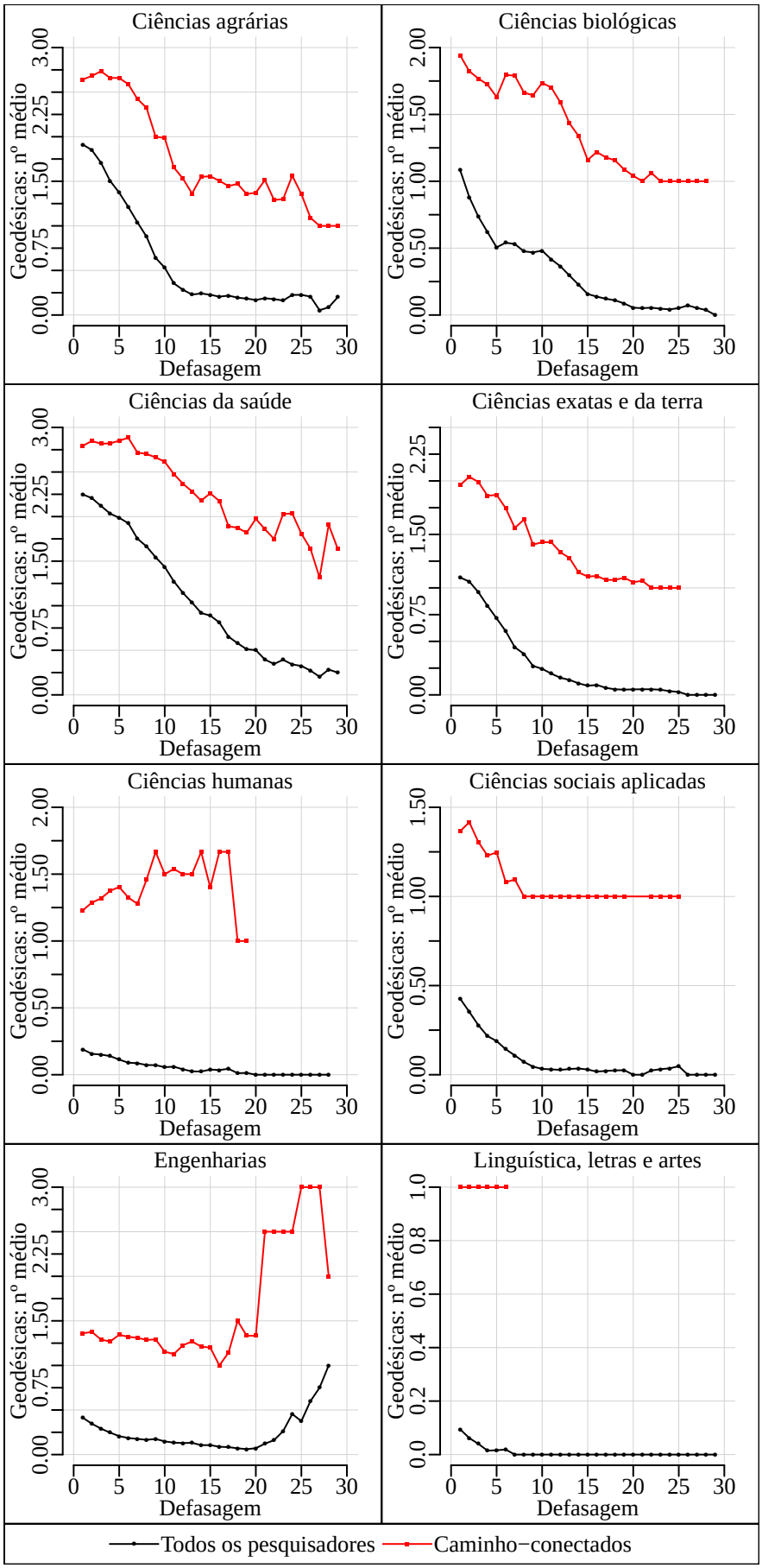

Figura 4.3 - Número médio de geodésicas antes da primeira colaboração

Voltando à tabela 4.1, ao analisar as linhas referentes à produtividade média dos pares nota-se que ela é, em média, baixa na maioria das grandes áreas do conhecimento. Convertendo a produtividade média acumulada em nove anos para os estratos da avaliação Qualis de 2008, encontra-se que nas ciências agrárias, nas ciências biológicas, nas engenharias, nas ciências exatas e da terra, nas ciências da saúde e nas ciências sociais aplicadas o valor total aproximado é equivalente a um artigo de nota B3. Nas ciências humanas, a mesma conversão indica uma produtividade acumulada média de classe B2 e em linguística, letras e artes este indicador é pouco maior do que o correspondente ao estrato B1. Por outro lado, a média da diferença de produtividade entre pares de coautores, não é desprezível, sendo em todas as grandes áreas do conhecimento maior do que a média da produtividade média. Analisados conjuntamente, os indicadores relativos à produtividade abrem questões interessantes para investigações futuras. Apesar de ainda em nível especulativo, uma possível maneira de racionalizar estes resultados é analisando mais profundamente a composição dos grupos de pesquisa. A baixa produtividade média entre pares pode resultar da participação de pelo menos um pesquisador em início de carreira, o qual não têm um histórico longo de publicações passadas. Já a diferença não-desprezível de produtividade entre coautores é compatível com alta frequência de colaboração entre pesquisadores iniciantes e pesquisadores mais experientes. Nos campos do conhecimento mais experimentais, como ciências agrárias, ciências biológicas, ciências da saúde e algumas áreas de ciências exatas e da terra, é possível que a baixa produtividade média e alta diferença de produtividade sejam devidas, por exemplo, ao envolvimento de estudantes em pesquisas coordenadas por um pesquisador sênior ou pela prática de atribuição de coautoria a chefes de laboratórios. Em outros campos do conhecimento, como nas ciências sociais aplicadas, as estatísticas de produtividade podem sugerir que a publicação coautorada entre orientadores e orientandos não seja incomum.

O aprofundamento da descrição dos participantes dos grupos de pesquisa é importante porque ajuda a entender como o conhecimento acadêmico é produzido e disseminado na comunidade de pesquisadores. Além disto, na medida em que esta auto-organização social da comunidade acadêmica responde às políticas adotadas nos departamentos, pode ser possível extrair conclusões sobre os impactos destas políticas sobre temas de interesse como o fluxo de informação através desta comunidade. De todo modo, o fato de a produtividade média dos pares em seis das oito grandes áreas do conhecimento estar na metade inferior do espectro de qualidade merece atenção. No melhor dos casos, em que este resultado é devido à coautoria de pesquisadores jovens, ainda assim se percebe que a produção destas pessoas tem difícil acesso às melhores revistas acadêmicas. Se realmente participação de pesquisadores jovens é um motivo da baixa produtividade, cabe uma reflexão das razões pelas quais a produção destes encontra dificuldades para aceitação nos periódiocos de maior prestígio, o que necessariamente passa por uma discussão dos programas de pós-

graduação e dos incentivos oferecidos pelas agências de fomento, avaliação e financiamento.

Com relação à similaridade nos interesses de pesquisa, nota-se que a medida cosseno, como esperado, reduz-se conforme se aumenta o grau de especificidade. Assim, a medida de similaridade cosseno ao nível de grande área do conhecimento é maior do que ao nível de área, que é maior do que ao nível de subárea. Entre a primeira e a segunda publicação coautorada, não se verifica um padrão claro de crescimento ou de diminuição da similaridade cosseno no nível de grande área do conhecimento já que, em cinco grandes áreas há uma suave elevação neste indicador e em outras três redução. Porém, aos níveis mais baixos de agregação, de área e subárea, a similaridade cosseno é maior na segunda colaboração do que na primeira, sendo a variação neste indicador maior no nível mais específico, de subárea do conhecimento. Em outras palavras, parece haver uma aproximação dos interesses de pesquisa de coautores após a primeira colaboração entre eles, embora o efeito reportado esteja contaminado pelo fato de que parte desta aparente aproximação é devida ao computo dos campos de contribuição do primeiro trabalho colaborativo entre o par de pesquisadores.

Para concluir, a análise da propensão a colaborar dos pares reforça a hipótese de que é frequente a colaboração entre pesquisadores experientes e pesquisadores principiantes. Analisando o histórico de coautoria total, não restrito à população Lattes/USP, a média entre pares da propensão média a coloborar é da mesma ordem de magnitude que a média entre pares da sua diferença absoluta. Este fato, conjugado com os altos valores encontrados, sugere que frequentemente as colaborações ocorrem entre pessoas com grande número de colaboradores pregressos e pessoas com poucos coautores passados. Olhando para os mesmos indicadores, porém restritos à população Lattes/USP, também se observa que há pouca diferença entre a média da propensão média a colaborar e a média de sua diferença absoluta. Além disto, enquanto os indicadores de propensão a colaborar irrestritos possuem algumas dezenas, estas mesmas variáveis restritas ao conjunto Lattes/USP são muito menores. Na medida em que esta discrepância é devida ao segundo motivo para a subestimação da propensão individual a colaborar discutido na seção 4.2, ela proporciona mais sustentação à hipótese de alta frequência de colaboração entre pesquisadores seniores e iniciantes.

4.4 Estratégia econométrica

O objetivo principal deste artigo é verificar se o início de uma relação de coautoria entre dois pesquisadores é facilitada pela rede de contatos interpessoais à qual eles pertencem. Para isto constroem-se as redes sociais de colaboração acadêmica e se estimam os impactos da proximidade e do número de menores caminhos entre dois pesquisadores sobre a probabilidade de que eles formem uma parceria no período seguinte. Antes de discutir as questões relativas a estas estimações, é importante fazer alguns comentários sobre o uso das redes de coautoria para o propósito deste trabalho. É claro que uma rede social de coautoria é um subgrafo de uma rede social mais ampla que engloba outros tipos de relacionamentos e que cujas arestas também podem servir como canais de transmissão de informação. Entretanto, como não se possuem dados que permitam construir uma rede social mais abrangente, trabalha-se com as redes de coautoria. Cabe então indagar quanto à validade desta restrição. O argumento a seu favor é que a proximidade e o número de geodésicas em uma rede de coautoria são *proxies* para as suas correspondentes em um grafo de relacionamentos mais geral (FAFCHAMPS; GOYAL; VAN DER LEIJ, 2010).

A estratégia econométrica adotada é idêntica à de Fafchamps, Goyal e van der Leij (2010), a qual se baseia em um experimento do tipo placebo. Especificamente, sendo p^{ij} a proximidade entre os pesquisadores i e j, g^{ij} o número de geodésicas entre eles e $\mathbf{m}^{ij}$ um vetor de variáveis de qualidade de pareamento, estima-se o seguinte modelo logit para a primeira colaboração:

$$\Pr(y_t^{ij} = 1 \mid y_s^{ij} = 0 \text{ para todo } s \text{ tal que } t > s \geqslant t_0) = f(p_{t-1}^{ij}, g_{t-1}^{ij}, \mathbf{m}_t^{ij}). \tag{4.5}$$

Se existem efeitos de rede social na formação de coautorias, os coeficientes estimados devem ser tais que $f_p(\cdot) > 0$ e $f_g(\cdot) > 0$. Em seguida, estima-se para as colaborações repetidas entre os indivíduos i e j o seguinte modelo logit:

$$\Pr(y_t^{ij} = 1 \mid y_s^{ij} = 1 \text{ para algum } s \text{ tal que } t > s \geqslant t_0) = h(p_{t-1}^{ij}, g_{t-1}^{ij}, \mathbf{m}_t^{ij}). \tag{4.6}$$

Se para o estabelecimento de novas parcerias os pesquisadores utilizam-se de informações disponíveis nos seus círculos sociais, então $h_p(\cdot) = 0$ e $h_g(\cdot) = 0$. Isto porque, após a primeira colaboração entre dois pesquisadores, eles já se conhecem bem e sabem avaliar a qualidade do pareamento entre eles. Sendo assim, espera-se que pouca informação nova relevante para a colaboração entre eles chegue até qualquer um deles através outras pessoas. Em outras palavras, se os coeficientes de proximidade e de número de geodésicas estão captando outro efeito que não o fluxo de informações relevantes para a iniciação de uma parceria em pesquisa, então eles devem ser significantes tanto para a primeira

colaboração quanto para as subsequentes. Aqui fica claro porque a distância entre dois pesquiadores é calculada excluindo-se o link, caso exista, entre eles. A razão é que, se este link fosse mantido no cálculo da distância, então não haveria variação nesta variável depois da primeira colaboração. O mesmo argumento vale para o cálculo do número de geodéscias entre dois pesquisadores. Após a primeira coautoria entre eles, há apenas um menor caminho, se calculado da maneira padrão, conectando-os.

A estratégia para identificação dos fluxos intra-rede de informações relevantes para o estabelecimento de parcerias consiste na variação dos impactos da proximidade e do número de geodésicas entre a primeira colaboração e as repetidas. Para identificar estes impactos é preciso primeiramente eliminar as fontes de endogeneidade. O primeiro controle que se insere nos modelos é a produtividade média dos pares de pesquisadores, $\bar{q}_t^{ij}$. Como pesquisadores altamente produtivos tendem a ter alto número de coautores, a proximidade e o número de menores caminhos entre eles e os demais também devem ser grandes. Isto gera uma correlação positiva entre as variáveis de rede e a produtividade do par. Por outro lado, pesquisadores com grande diferença de produção acadêmica, Δq_t^{ij}, tendem a estar mais distantes e a ter baixo número de menores caminhos entre eles. Por este motivo, aos modelos é adicionada a variável Δq_t^{ij}.

A similaridade entre interesses de pesquisa de dois pesquisadores é positivamente correlacionada com proximidade e com número de geodésicas entre eles, tanto antes quanto depois da primeira coautoria. Mais ainda, entre colaborações, os interesses de pesquisa de dois coautores tendem a se aproximar. Como a distância e o número de geodésicas entre dois pesquisadores i e j são calculados excluindo-se o link entre eles, caso exista, estas variáveis podem flutuar ao longo do tempo mesmo após a primeira colaboração. Sendo assim, pelo fato de os interesses de pesquisa de um par de pesquisadores aproximarem-se após a primeira parceria, surge a possibilidade de correlações espúrias entre as variáveis de rede e a probabilidade de coautorias repetidas. Isto acontece porque, com a convergência dos interesses de pesquisa entre os indivíduos i e j após a primeria coautoria entre eles, aumenta a probabilidade de que i colabore com algum coautor de j, que deve ter interesses de pesquisa parecidos com os de j. Assim, mesmo após a primeira colaboração entre i e j, eles devem-se aproximar na rede social de coautoria em decorrência da maior intersecção em interesses de pesquisa. Sendo assim, ao vetor de controles acrescentam-se as medidas de similaridade cosseno aos níveis de grande área do conhecimento, área do conhecimento e subárea do conhecimento.

Quanto maior a propensão a colaborar dos indivíduos, maior deve ser a proximidade e o número de geodésicas entre eles. Como esta propensão é, pelo menos em parte, variante no tempo, para se estimar os efeitos de rede social consistentemente, aos modelos são

adicionados dois controles de propensão média a colaborar, $\bar{n}_t^{ij}$: (i) o número médio de coautores totais e (ii) o número médio de coautores pertencentes ao conjunto Lattes/USP. Por fim, também são adicionadas aos modelos a diferença absoluta entre as propensões a colaborar dos indivíduos tanto restritas quanto irrestritas à população Lattes/USP.

As variáveis de controle de produtividade, propensão a colaborar e similaridade de interesses de pesquisa fornecem proteção para fontes de correlações espúrias variantes no tempo. Porém, ainda resta a possibilidade de que a criação de parcerias seja influenciada por outros fatores, tais como local de educação, orientador comum e habilidade inata. Pelo fato de que estes determinantes podem ser correlacionados com as demais variáveis explicativas dos modelos, ignorá-las resulta em viés e inconsistência dos estimadores. Para contornar este problema, como pares ij de indivíduos tais que $\max\{t_0^i, t_0^j\} < \min\{t_2^i, t_2^j\}$ são observados mais de uma vez, aos modelos é adicionado um efeito idiossincrático de par, os quais são estimados por logit condicional (efeitos fixos). Com relação a este estimador, uma observação importante deve ser feita. Conforme discutido detalhadamente por Fafchamps, Goyal e van der Leij (2010), a equação referente à primeira coautoria é equivalente a um modelo de duração com manifestação única. Em modelos deste tipo, os coeficientes dos regressores variantes no tempo podem ser estimados consistentemente se eles não apresentarem tendência temporal (ALLISON; CHRISTAKIS, 2006). Por este motivo, a exemplo do que fizeram aqueles autores, no modelo referente à primeira colaboração, remove-se a tendência temporal de todas as variáveis explicativas individualmente. Para isto, obtêm-se os resíduos da regressão de cada uma destas variáveis em um efeito idiossincrático de par e em uma função temporal, os quais então substituem as respectivas variáveis explicativas do modelo de primeira coautoria. Esta prática de remoção de tendências, tem como consequência perda de informação e, portanto, gera viés de atenuação. Uma segunda fragilidade do método adotado é que ele assume independência contemporânea entre pares de pesquisadores. Esta é uma hipótese forte porque um grupo de n coautores de um dado artigo pode ser decomposto em $n(n-1)/2$ pares distintos. Além disto, um mesmo autor, ao trabalhar simultaneamente com dois outros pesquisadores em projetos distintos deve dividir o seu tempo entre eles. Embora estas interdependências entre pares não afetem a estimação dos coeficientes, elas podem viesar os erros padrões, comprometendo as conclusões dos testes de hipóteses.

4.5 Resultados econométricos

4.5.1 Efeitos de rede social

A comparação dos coeficientes associados à proximidade e ao número de geodésicas entre a primeira colaboração e as subsequentes é a estratégia adotada para extração de evidências quanto ao fluxo de informações individuais e de qualidade de pareamento nas redes sociais de coautoria entre os pesquisadores do conjunto Lattes/USP.

Considere primeiramente os resultados das estimações dos modelos sem efeitos idiossincráticos de par reportados nas tabelas 4.2.1 e 4.2.2. Como se pode perceber, nas ciências agrárias, nas ciências biológicas, nas ciências da saúde e nas ciências exatas e da terra o coeficiente de proximidade é significante tanto para a primeira colaboração quanto para as subsequentes em ambas as especificações, com e sem controles variantes no tempo. Já o coeficiente de número de geodésicas tem um comportamento diferente entre estes campos do conhecimento. Enquanto nas ciências agrárias ele se mantém significante nos modelos de primeira colaboração e nos modelos de colaborações repetidas, nas ciências biológicas ele não é significante em nenhum dos modelos, nas ciências da saúde ele é significante apenas para o segundo e nas ciências exatas e da terra apenas no primeiro. Apesar destas diferenças, o teste de exclusão dos efeitos de rede (proximidade e número de geodésicas) mostra que estas variáveis são conjuntamente significantes tanto para o início de uma parceria quanto para a sua repetição nas especificações com e sem controles variantes no tempo. Nestas grandes áreas do conhecimento, quando significantes, os coeficientes associados à proximidade e ao número de geodéscias são sempre positivos, como esperado.

Nestas quatro grandes áreas do conhecimento (ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra) verifica-se que, embora o coeficiente de proximidade seja significante tanto na primeira colaboração quanto nas subsequentes, ele cai acentuadamente entre os dois momentos. Entretanto, o mesmo não pode ser afirmado em relação ao coeficiente de número de geodéscias, cujo comportamento entre colaborações não é claro. Outra observação importante é que a inclusão das variáveis de controle não tem grande efeito no coeficiente de proximidade nem para a primeira colaboração nem para as demais. Além disto, o impacto da inclusão das variáveis de controle sobre a estimativa pontual do coeficiente de proximidade nem sempre é de redução.

Ainda analisando os resultados dos modelos *pooled* logit, nota-se que o coeficiente de proximidade é insignificante em engenharias, ciências humanas, ciências sociais aplicadas e linguística, letras e artes tanto para a primeira colaboração quanto para as que se seguem,

independentemente da inclusão dos controles variantes no tempo. Nas engenharias e nas ciências sociais aplicadas, o coeficiente relativo ao número de geodésicas também é insignificante em todas as estimações por *pooled* logit. Já nas ciências humanas, o coeficiente de número de geodésicas é significante apenas nas especificações sem os controles. Curiosamente, para a primeira colaboração sua estimativa é negativa e para a segunda ela é positiva. Embora, analisados individualmente, os coeficientes apontem para a inexistência de efeitos de rede nestes campos, quando estudados em conjunto, esta conclusão só se mantém em linguística, letras e artes. Nas engenharias, nas ciências humanas e nas ciências sociais aplicadas, proximidade e número de geodésicas são conjuntamente significantes para primeira colaboração, mas não para as colaborações seguintes no modelo com todos os controles[5]. No apêndice 6, são apresentadas as tabelas com os resultados das estimações por *pooled* logit sobre a mesma amostra utilizada nas estimações do modelo com efeitos fixos, discutido a seguir. Como se pode perceber, embora haja variações nas estimativas pontuais dos coeficientes, as conclusões relativas à existência de efeitos de rede não se alteram quando comparadas às obtidas pelo mesmo modelo, porém estimado sobre a amostra irrestrita.

[5]Nas ciências sociais aplicadas, proximidade e número de geodésicas são conjuntamente insignificantes para colaborações subsequentes tanto na especificação com controles variantes no tempo quanto na especificação sem estes controles.

Tabela 4.2.1 - Regressão logística (coeficientes): *pooled cross section*

		Agrárias		Biológicas		Exatas		Saúde	
Primeira colaboração	p_{t-1}^{ij}	3,483***	2,990***	2,413***	2,469***	2,736***	2,517***	1,840***	1,694***
		(0,250)	(0,264)	(0,561)	(0,560)	(0,333)	(0,341)	(0,163)	(0,166)
	g_{t-1}^{ij}	0,044***	0,027**	0,070	0,073	0,085**	0,081**	-0,001	-0,002
		(0,014)	(0,013)	(0,058)	(0,058)	(0,037)	(0,037)	(0,006)	(0,006)
	$\bar{q}_t^{ij}$		0,022***		0,019**		0,037***		0,025***
			(0,008)		(0,009)		(0,009)		(0,006)
	Δq_t^{ij}		0,007		-0,008*		-0,003		-0,003
			(0,005)		(0,005)		(0,006)		(0,004)
	$\bar{n}_t^{ij}$: total		0,009***		-0,003		0,001		0,002***
			(0,002)		(0,002)		(0,002)		(0,000)
	$\bar{n}_t^{ij}$: USP		-0,026		-0,993		-2,155***		-1,416***
			(0,368)		(1,037)		(0,551)		(0,175)
	Δn_t^{ij}: total		-0,002**		0,001		-0,002		-0,001***
			(0,001)		(0,001)		(0,001)		(0,000)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		1,098***		0,234		0,883		0,567***
			(0,344)		(0,328)		(0,601)		(0,156)
	$\omega_{t,f}^{ij}$: área		0,315		0,387		0,803		0,865***
			(0,435)		(0,484)		(0,570)		(0,242)
	$\omega_{t,f}^{ij}$: subárea		-0,099		1,150**		0,329		0,049
			(0,322)		(0,575)		(0,413)		(0,178)
	Constante	-1,932***	-1,955***	-2,305***	-2,314***	-1,970***	-1,991***	-1,969***	-1,976***
		(0,030)	(0,031)	(0,048)	(0,049)	(0,036)	(0,037)	(0,015)	(0,015)
	Excl. ef. rede	220,93	138,21	25,27	26,20	86,90	68,99	128,42	104,92
		[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]
	Observações	9.375	9.375	5.640	5.640	6.974	6.974	39.763	39.763
		Agrárias		Biológicas		Exatas		Saúde	
Colaborações subsequentes	p_{t-1}^{ij}	0,614***	0,611**	0,817***	0,882***	0,793***	0,546***	0,908***	1,067***
		(0,138)	(0,240)	(0,114)	(0,275)	(0,113)	(0,212)	(0,079)	(0,156)
	g_{t-1}^{ij}	0,064***	0,092***	0,020	0,036	0,016	0,027	0,053***	0,072***
		(0,009)	(0,015)	(0,013)	(0,040)	(0,013)	(0,028)	(0,004)	(0,007)
	$\bar{q}_t^{ij}$		0,002		-0,007		-0,006		0,003
			(0,006)		(0,005)		(0,004)		(0,002)
	Δq_t^{ij}		-0,005		0,003		0,002		-0,003**
			(0,004)		(0,004)		(0,003)		(0,002)
	$\bar{n}_t^{ij}$: total		-0,003***		-0,001		-0,002		-0,001***
			(0,001)		(0,001)		(0,001)		(0,000)
	$\bar{n}_t^{ij}$: USP		1,372***		1,872***		1,613***		7,369***
			(0,421)		(0,391)		(0,376)		(0,357)
	Δn_t^{ij}: total		0,003***		0,001		0,002***		0,001***
			(0,000)		(0,001)		(0,001)		(0,000)
	Δn_t^{ij}: USP								-3,151***
									(0,198)
	$\omega_{t,f}^{ij}$: gr. área		-0,881***		-0,474**		-0,514**		-0,671***
			(0,199)		(0,208)		(0,256)		(0,105)
	$\omega_{t,f}^{ij}$: área		-0,382*		0,703***		-0,273		0,043
			(0,197)		(0,236)		(0,311)		(0,090)
	$\omega_{t,f}^{ij}$: subárea		1,030***		0,892***		0,735***		0,680***
			(0,124)		(0,204)		(0,167)		(0,053)
	Constante	-0,337	-0,679**	-0,340	-1,131***	-0,878**	1,682	-0,523***	0,339
		(0,385)	(0,282)	(0,407)	(0,407)	(0,437)	(1,144)	(0,177)	(0,572)
	Excl. ef. rede	77,87	47,28	60,17	14,39	53,25	8,56	276,96	141,77
		[0,000]	[0,000]	[0,000]	[0,001]	[0,000]	[0,014]	[0,000]	[0,000]
	Observações	31.633	8.748	29.486	5.077	28.347	5.813	140.972	41.708

*** p<0,01; ** p<0,05; * p<0,1.

Tabela 4.2.2 - Regressão logística (coeficientes): *pooled cross section*

		Engenharias		Humanas		Linguística		Sociais	
Primeira colaboração	p_{t-1}^{ij}	0,571	0,324	1,656	1,200	23,998	31,680*	1,513	1,132
		(0,879)	(0,922)	(2,429)	(2,694)	(14,785)	(16,659)	(1,578)	(1,572)
	g_{t-1}^{ij}	0,412*	0,406	-1,897**	-1,866*			0,565	0,558
		(0,235)	(0,249)	(0,925)	(0,990)			(0,405)	(0,389)
	$\bar{q}_t^{ij}$		0,028		0,011		-0,004		0,049*
			(0,018)		(0,012)		(0,015)		(0,029)
	Δq_t^{ij}		-0,002		-0,001		-0,008		-0,027
			(0,010)		(0,008)		(0,009)		(0,022)
	$\bar{n}_t^{ij}$: total		0,008		-0,003		-0,011		0,042**
			(0,007)		(0,014)		(0,163)		(0,017)
	$\bar{n}_t^{ij}$: USP		-0,095		0,883				0,602
			(0,737)		(1,225)				(1,790)
	Δn_t^{ij}: total		0,003		0,002		-0,067		-0,024*
			(0,005)		(0,009)		(0,055)		(0,013)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		0,042		0,169		-6,110		-0,549
			(0,430)		(0,711)		(6,583)		(1,645)
	$\omega_{t,f}^{ij}$: área		1,192		3,347***		0,304		0,751
			(0,916)		(1,181)		(7,705)		(0,967)
	$\omega_{t,f}^{ij}$: subárea		-0,408		0,057		1,565		-1,359
			(0,858)		(0,808)		(5,409)		(0,921)
	Constante	-1,338***	-1,347***	-1,833***	-1,870***	-1,106***	-1,149***	-1,390***	-1,408***
		(0,066)	(0,068)	(0,094)	(0,099)	(0,297)	(0,326)	(0,093)	(0,095)
	Excl. ef. rede	9,36	7,13	6,63	7,29	2,63	3,62	10,80	8,11
		[0,009]	[0,028]	[0,036]	[0,026]	[0,105]	[0,057]	[0,005]	[0,017]
	Observações	1.227	1.227	877	877	69	69	797	797
Colaborações subsequentes		Engenharias		Humanas		Linguística		Sociais	
	p_{t-1}^{ij}	0,418*	-0,225	0,485	0,100	0,546	5,212*	0,464*	-0,574
		(0,218)	(0,478)	(0,391)	(1,147)	(0,905)	(2,914)	(0,261)	(0,797)
	g_{t-1}^{ij}	0,055	0,071	0,269**	0,041		0,005	0,032	0,259
		(0,055)	(0,126)	(0,107)	(0,425)		(0,012)	(0,056)	(0,227)
	$\bar{q}_t^{ij}$		-0,007		-0,013		-0,009		-0,035**
			(0,009)		(0,009)		(0,008)		(0,017)
	Δq_t^{ij}		0,002		0,003		-0,024		0,009
			(0,005)		(0,005)		(0,059)		(0,010)
	$\bar{n}_t^{ij}$: total		0,001		0,008				0,009
			(0,003)		(0,007)				(0,008)
	$\bar{n}_t^{ij}$: USP		1,328***		1,167				0,755
			(0,493)		(0,840)				(0,679)
	Δn_t^{ij}: total		0,001		-0,005		0,095*		0,000
			(0,002)		(0,005)		(0,051)		(0,005)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		-0,564*		-1,250**		-9,337*		1,046
			(0,311)		(0,570)		(5,347)		(0,953)
	$\omega_{t,f}^{ij}$: área		0,207		-0,018		8,831*		-0,529
			(0,360)		(0,531)		(4,606)		(0,585)
	$\omega_{t,f}^{ij}$: subárea		0,752***		0,799**		2,763		0,205
			(0,267)		(0,402)		(2,077)		(0,277)
	Constante	-1,928*	-0,820*	-1,887	-0,039	-0,419	-1,164	-1,339***	-0,919
		(1,100)	(0,472)	(1,160)	(0,810)	(0,327)	(4,395)	(0,127)	(1,183)
	Excl. ef. rede	14,84	0,32	26,83	0,11	0,36	3,20	5,69	1,32
		[0,001]	[0,851]	[0,000]	[0,945]	[0,546]	[0,074]	[0,058]	[0,517]
	Observações	11.478	1.811	6.599	681	466	70	6.146	667

*** p<0,01; ** p<0,05; * p<0,1.

As estimações por *pooled* logit não controlam para determinantes invariantes no tempo da coautoria. Desta forma, variáveis como local de educação, afinidade e interesses de pesquisa estáveis são delegadas para o resíduo. Pelo fato de estas variáveis serem correlacionadas com outras variáveis explicativas do modelo, ignorá-las resulta em viés e incosistência do estimador por *pooled* logit. Sendo assim, considere os resultados das estimações por logit condicional (que são computacionalmente idênticas às estimações por efeitos fixos) reportados nas tabelas 4.3.1 e 4.3.2.

Tabela 4.3.1 - Regressão logística (coeficientes): efeitos fixos

		Agrárias		Biológicas		Exatas		Saúde	
Primeira colaboração	p_{t-1}^{ij}	4,029***	3,354***	2,077***	2,175***	3,051***	2,916***	1,831***	1,704***
		(0,355)	(0,382)	(0,753)	(0,785)	(0,494)	(0,514)	(0,198)	(0,203)
	g_{t-1}^{ij}	0,051**	0,030*	0,101**	0,121**	0,117**	0,106**	0,002	0,002
		(0,020)	(0,018)	(0,050)	(0,056)	(0,049)	(0,051)	(0,008)	(0,008)
	$\bar{q}_t^{ij}$		0,027**		0,050**		0,054***		0,023*
			(0,013)		(0,024)		(0,015)		(0,014)
	Δq_t^{ij}		0,007		-0,052***		-0,003		-0,011
			(0,008)		(0,017)		(0,009)		(0,009)
	$\bar{n}_t^{ij}$: total		0,016***		-0,008*		-0,005		0,004***
			(0,003)		(0,005)		(0,005)		(0,001)
	$\bar{n}_t^{ij}$: USP		-0,072		0,080		-0,793**		-0,323***
			(0,372)		(1,216)		(0,398)		(0,115)
	Δn_t^{ij}: total		-0,004**		-0,002		-0,001		-0,002***
			(0,002)		(0,004)		(0,004)		(0,000)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		1,125**		0,239		0,773		0,845***
			(0,448)		(0,702)		(0,521)		(0,213)
	$\omega_{t,f}^{ij}$: área		0,791		0,311		1,034		1,118***
			(0,578)		(0,898)		(0,823)		(0,325)
	$\omega_{t,f}^{ij}$: subárea		-0,322		1,689**		0,619		0,079
			(0,439)		(0,717)		(0,661)		(0,224)
	Excl. ef. rede	145,79	82.93	14,52	16,37	49,97	39,20	86,20	70,43
		[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]
	Observações	5.126	5.126	1.780	1.780	3,273	3,273	21.395	21.395
	Nº pares	958	958	322	322	628	628	3.706	3.706
Colaborações subsequentes		Agrárias		Biológicas		Exatas		Saúde	
	p_{t-1}^{ij}	0,079	0,332	-0,056	-0,290	0,282	0,030	0,201*	-0,426
		(0,231)	(0,512)	(0,210)	(0,598)	(0,195)	(0,463)	(0,114)	(0,265)
	g_{t-1}^{ij}	-0,010	0,021	-0,035	0,036	-0,058**	0,040	0,010**	0,005
		(0,015)	(0,031)	(0,029)	(0,083)	(0,025)	(0,052)	(0,005)	(0,013)
	$\bar{q}_t^{ij}$		-0,001		-0,018*		-0,016**		-0,004
			(0,009)		(0,010)		(0,008)		(0,004)
	Δq_t^{ij}		-0,011**		0,009		0,013**		-0,002
			(0,006)		(0,008)		(0,006)		(0,003)
	$\bar{n}_t^{ij}$: total		-0,002		-0,002		-0,003		-0,000
			(0,002)		(0,002)		(0,002)		(0,000)
	$\bar{n}_t^{ij}$: USP		0,337		-0,115		0,728		10,315***
			(0,580)		(0,658)		(0,626)		(0,079)
	Δn_t^{ij}: total		0,002		0,002		0,005**		0,000
			(0,001)		(0,002)		(0,002)		(0,000)
	Δn_t^{ij}: USP								-4,881***
									(0,164)
	$\omega_{t,f}^{ij}$: gr. área		0,099		-1,030		0,237		-0,203
			(0,580)		(0,645)		(0,823)		(0,303)
	$\omega_{t,f}^{ij}$: área		0,047		-0,427		-1,454		-0,466
			(0,788)		(0,939)		(1,562)		(0,388)
	$\omega_{t,f}^{ij}$: subárea		-0,793		-0,087		-0,509		0,196
			(0,487)		(0,567)		(0,824)		(0,211)
	Excl. ef. rede	0,79	1,53	1,82	0,30	6,38	0,59	7,19	2,85
		[0,674]	[0,465]	[0,403]	[0,859]	[0,0411]	[0,744]	[0,278]	[0,240]
	Observações	19.535	4.506	16.319	2.217	17.104	2.663	87.077	21.739
	Nº pares	2.307	708	1.668	311	1.976	475	9.027	3.204

*** p<0,01; ** p<0,05; * p<0,1.

Tabela 4.3.2 - Regressão logística (coeficientes): efeitos fixos

		Engenharias		Humanas		Linguística	Sociais	
Primeira colaboração	p_{t-1}^{ij}	1,164	0,037	-72,808	-70,671		8,472	7,513
		(1,866)	(2,118)	(52,499)	(51,811)		(5,347)	(8,503)
	g_{t-1}^{ij}	0,707	0,818	-6,895	-6,687		-2,309	-1,808
		(0,510)	(0,533)	(4,452)	(4,149)		(2,266)	(3,272)
	$\bar{q}_t^{ij}$		0,032		-0,002			0,088*
			(0,024)		(0,022)			(0,050)
	Δq_t^{ij}		-0,006		0,007			-0,070
			(0,012)		(0,013)			(0,071)
	$\bar{n}_t^{ij}$: total		0,021**		0,049			0,117**
			(0,011)		(0,049)			(0,058)
	$\bar{n}_t^{ij}$: USP		0,313		3,262			0,826
			(0,918)		(2,439)			(1,519)
	Δn_t^{ij}: total		0,001		-0,014			-0,057*
			(0,006)		(0,025)			(0,029)
	Δn_t^{ij}: USP							
	$\omega_{t,f}^{ij}$: gr. área		0,379		0,442			0,500
			(0,835)		(3,220)			(3,389)
	$\omega_{t,f}^{ij}$: área		1,771		3,617***			-0,130
			(1,197)		(1,253)			(3,409)
	$\omega_{t,f}^{ij}$: subárea		-0,330		-0,466			-1,761
			(1,029)		(0,910)			(1,291)
	Excl. ef. rede	7,37	5,52	4,90	6,36		4,39	1,77
		[0,025]	[0,063]	[0,086]	[0,042]		[0,111]	[0,413]
	Observações	572	572	386	386		269	269
	Nº pares	152	152	88	88		80	80
Colaborações subsequentes		Engenharias		Humanas		Linguística	Sociais	
	p_{t-1}^{ij}	0,030	0,063	-0,291	-0,291		-0,443	1,253
		(0,329)	(1,073)	(0,595)	(0,595)		(0,000)	(0,000)
	g_{t-1}^{ij}	-0,088	-0,007	-0,010	-0,010		-0,072	-0,697
		(0,110)	(0,228)	(0,148)	(0,148)		(0,000)	(0,000)
	$\bar{q}_t^{ij}$		-0,012					-0,026
			(0,016)					(0,000)
	Δq_t^{ij}		0,002					0,040
			(0,010)					(0,000)
	$\bar{n}_t^{ij}$: total		-0,019**					0,046
			(0,007)					(0,000)
	$\bar{n}_t^{ij}$: USP							0,966
								(0,000)
	Δn_t^{ij}: total		-0,005					-0,010
			(0,006)					(0,000)
	Δn_t^{ij}: USP							
	$\omega_{t,f}^{ij}$: gr. área							-0,608
								(0,000)
	$\omega_{t,f}^{ij}$: área							8,620
								(0,000)
	$\omega_{t,f}^{ij}$: subárea							1,950
								(0,000)
	Excl. ef. rede	0,77	0,00	0,47	0,47		2,59	1,15
		[0,680]	[0,998]	[0,791]	[0,791]		[0,275]	[0,564]
	Observações	6.990	838	3.465	3.465		2.818	237
	Nº pares	847	172	416	416		401	58

*** p<0,01; ** p<0,05; * p<0,1.

Estudando os resultados das estimações dos modelos com efeitos idiossincráticos de par, observa-se que em nenhuma grande área do conhecimento é significante o coeficiente de proximidade nas colaborações subsequentes. Porém, em quatro grandes áreas do conhe-

cimento este coeficiente é estatisticamente diferente de zero para a primeira colaboração. Estas grandes áreas do conhecimento são ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra. Já em ciências humanas, ciências sociais aplicadas e engenharias, o coeficiente de proximidade não é significante na primeira colaboração. Em linguística, letras e artes, a especificação com efeitos idiossincráticos não foi possível de ser estimada.

Com relação ao coeficiente de número de geodésicas, considerando as especificações mais completas com efeitos idiossincráticos de par e controles variantes no tempo, também se percebe que em nenhuma grande área do conhecimento ele é significante quando se trata das colaborações repetidas. Além disto, apenas em ciências biológicas e em ciências exatas e da terra o coeficiente associado ao número de geodésicas é significante para a primeira colaboração quando se estuda o modelo mais completo.

No modelo de colaborações repetidas, o teste de significância conjunta dos coeficientes de proximidade e de número de geodéscias não rejeita a hipótese nula na especificação mais completa em todas as grandes áreas do conhecimento. Além disto, apenas em ciências exatas e da terra a hipótese de que não há efeitos de rede na repetição de parcerias acadêmicas é rejeitada na especificação com efeitos idiossincráticos, mas sem os demais controles, e ainda assim por pequena margem. Já com relação ao modelo de primeira coautoria, o teste de exclusão da proximidade e do número de geodéscias mostra que a hipótese nula é rejeitada na especificação com todos os controle variantes no tempo em ciência agrárias, ciências biológicas, ciências humanas, ciências da saúde e ciências exatas e da terra. Com base nestes resultados, conclui-se que a proximidade e o número de geodésicas entre dois pesquisadores são irrelevantes para a repetição de uma coautoria entre eles. Adicionalmente, como em ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra, estas variáveis não podem ser excluídas do modelo mais completo, conclui-se que há fluxo de informação sobre qualidade de coautores dentro das suas redes sociais de colaboração acadêmica. Em ciências humanas, a evidência a este respeito é frágil e os coeficientes, individualmente insignificantes, apontam no sentido oposto do esperado. Já nos demais campos do conhecimento não existe evidência de que estes fluxos de informação estejam presentes.

O impacto da proximidade é forte. Como reportado na tabela 4.4, ao reduzir a distância de 3 para 2, por exemplo, a chance (*odds-ratio*) de iniciar uma colaboração aumenta em 74,89% nas ciências agrárias, 43,69% nas ciências biológicas, 32,84% nas ciências da saúde e 62,58% nas ciências exatas e da terra. Já com relação ao impacto marginal do número de geodésicas, nas ciências biológicas ele é de 12,86% e nas ciências exatas e da terra de 11,18%.

Tabela 4.4 - Impacto marginal da distância (%)

	Agrárias	Biológicas	Exatas	Saúde
$3 \rightarrow 2$	74,89	43,69	62,58	32,84
$4 \rightarrow 3$	32,25	19,87	27,51	15,26
$5 \rightarrow 4$	18,26	11,49	15,70	8,89
$6 \rightarrow 5$	11,83	7,52	10,21	5,84
$7 \rightarrow 6$	8,31	5,32	7,19	4,14
$8 \rightarrow 7$	6,17	3,96	5,35	3,09
$9 \rightarrow 8$	4,77	3,07	4,13	2,39
$10 \rightarrow 9$	3,80	2,45	3,29	1,91

Até onde se sabe, este é o primeiro trabalho a encontrar diferenças entre campos do conhecimento no papel desempenhado pelas redes sociais de coautoria na emergência de novos elos de colaboração. Por um lado, este resultado é esperado, considerando que o recurso à coautoria varia entre os campos de pesquisa. De fato, os campos com maior incidência e extensão de coautoria coincidem com aqueles em que as pessoas utilizam informações disponíveis nas suas redes de colaboração para selecionarem seus futuros coautores, o que é intuitivo já que nestes campos a colaboração é um elemento muito importante na tecnologia de produção acadêmica. Por outro lado, este resultado é válido apenas na medida em que a proximidade e o número de geodéscias nas redes de coautoria são boas *proxies* para as suas correspondentes nas redes sociais de relacionamentos mais amplas. Tendo em mente que, por serem menos colaborativas as produções acadêmicas em engenharias, ciências humanas, ciências sociais aplicadas e linguística, letras e artes, nestas grandes áreas do conhecimento, as diferenças de densidade entre as redes sociais de coautoria e as redes sociais de relacionamentos mais amplas devem ser maiores do que nas outras quatro grandes áreas do conhecimento. Sendo assim, é possível que a falta de evidências de fluxos de informação facilitadoras de pareamento nestes quatro campos menos colaborativos deva-se ao fato de que as *proxies* utilizadas não sejam suficientemente boas.

4.5.2 Variáveis de controle

Nesta subseção interpretam-se os resultados referentes às variáveis de controle das estimações por logit condicional. Considere primeiramente os coeficientes das variáveis de similaridade em interesses de pesquisa. Percebe-se que, para a repetição de colaborações, estes coeficientes são insignificantes em todos os três níveis de agregação considerados.

Para a primeira colaboração, estes coeficientes também são insignificantes nos casos das engenharias, ciências sociais aplicadas, ciências exatas e da terra e linguística, letras e artes. Por outro lado, nas ciências agrárias, ciências biológicas, ciências da saúde e ciências humanas ao menos um dos três coeficientes de similaridade no interesse de pesquisa é estatisticamente diferente zero. Nestas grandes áreas, quando significantes para a primeira colaboração, os coeficientes de similaridade são positivos, indicando que pesquisadores mais parecidos em termos de campos de atuação têm maior probabilidade de iniciarem uma parceria.

O fato de a probabilidade de estabelecimento de uma primeira colaboração aumentar com a similaridade dos pesquisadores em quatro grandes áreas do conhecimento é interessante e merece atenção até porque levanta questões para pesquisas futuras. Para entender porque os coeficientes de similaridade são positivos, pode ser útil caracterizar detalhadamente quais são as colaborações estabelecidas. É possível, por exemplo, que este resultado decorra de alta frequência de colaboração entre orientadores e orientandos, cujos interesses de pesquisa devem ter grande intersecção. Outra possibilidade é que os pesquisadores procurem colaboradores parecidos para que, havendo alto grau de substitutabilidade entre eles, seja possível suavizar esforço ao longo do tempo ou redistribuí-lo entre os coautores a depender de choques, por exemplo, de disponibilidade de tempo. Ou seja, neste caso, os pesquisadores fariam um seguro mútuo entre si. Evidentemente, o efeito colateral deste tipo de arranjo é que, elevado nível de substitutabilidade entre os pesquisadores pode reforçar problemas de risco moral, com baixo nível de esforço empregado por eles.

Com referência à produtividade média, o seu coeficiente é significante para a primeira colaboração em três grandes áreas do conhecimento: ciências agrárias, ciências biológicas e ciências exatas e da terra. Nos três casos, à medida que a produtividade média aumenta, também cresce a probabilidade de coautoria, como esperado. Porém, quando se trata das colaborações subsequentes, o coeficiente de produtividade média só é significante nas ciências exatas e da terra e é negativo.

Para a primeira colaboração, o coeficiente de diferença de produtividade é significante apenas nas ciências biológicas, sendo negativo. Por outro lado, nas colaborações subsequentes, este coeficiente é estatisticamente diferente de zero apenas nas ciências agrárias e nas ciências exatas e da terra. No primeiro campo de estudo, o impacto da diferença de produtividade é negativo e no segundo positivo.

4.6 Conclusões

Este artigo investiga em que medida informação sobre pesquisadores individuais e sobre qualidade de pareamentos passados circulam dentro das redes sociais de coautoria de oito grandes áreas do conhecimento. Neste estudo, a estratégia econométrica adotada é idêntica à apresentada em Fafchamps, Goyal e van der Leij (2010). Apesar da similaridade, este trabalho distingue-se daquele em alguns pontos importantes. Em primeiro lugar, não se restringe a análise à rede de coautoria entre economistas. Ao contrário, estudam-se oito grandes áreas do conhecimento: ciências agrárias, ciências biológicas, ciências da saúde, ciências exatas e da terra, ciências humanas, ciências sociais aplicadas, engenharias e linguística, letras e artes. Em segundo lugar, neste artigo, o ambiente em que os links são formados é extremamente favorável à colaboração, porque, pelo fato de todos os pesquisadores terem algum vínculo com uma mesma universidade (USP), as fricções de pareamento e as assimetrias informacionais são mínimas, menos contudentes do que no ambiente em que se formam os elos de coautoria estudados por Fafchamps, Goyal e van der Leij (2010).

Os resultados encontrados indicam que, apesar das grandes facilidades para formação independente de links de coautoria, em quatro grandes áreas do conhecimento (ciências agrárias, ciências biológicas, ciências da saúde e ciências exatas e da terra) existem efeitos de rede sobre a formação de novas parcerias. Em ciências humanas a evidência destes efeitos é fraca. Já nas engenharias, nas ciências sociais aplicadas e em linguística letras e artes não há evidências de que estes efeitos estão presentes. A existência destes efeitos, mesmo em um ambiente com fricções mínimas, reforça ainda mais o argumento de Fafchamps, Goyal e van der Leij (2010) de que fricções de pareamento desempenham papel importante em outras situações que envolvam a formação de times, ou seja, elas permeiam o funcionamento da economia de uma maneira geral.

5 CONSIDERAÇÕES FINAIS

Neste trabalho, o autor dá seus primeiros passos dentro da agenda de pesquisa em redes econômicas e sociais. Como discutido, este é um campo de estudo interdisciplinar, que se beneficia de conceitos e técnicas de diversas áreas do conhecimento e que tem recebido enorme atenção dos pesquisadores, em particular dos economistas, nos últimos anos. Em parte, este interesse deve-se à variedade de ambientes que podem ser inseridos em redes sociais, permitindo com isto obter novas perspectivas sobre questões abordadas de outras maneiras. Além disto, ao se levar em consideração as interconexões entre os agentes, é possível superar algumas dificuldades empíricas, como o problema do reflexo.

Nesta tese, o interesse concentra-se no mecanismo de formação das redes sociais de coautoria entre pesquisadores com alguma relação com a Universidade de São Paulo. A importância em estudar as colaborações entre pesquisadores é evidente levando-se em consideração que delas emergem inovações tecnológicas que são fundamentais para o crescimento econômico e para a melhoria da qualidade de vida. Em especial, esta tese procura mensurar as fricções inerentes ao processo de formação de elos de coautoria. Percebe-se que, mesmo em um ambiente extremamente propício à colaboração, estas fricções fazem-se presentes, indicando que elas devem ter papel importante em muitas outras situações que envolvam a formação de equipes. Em um dos artigos desenvolvidos, verifica-se que parte destas fricções são devidas a assimetrias informacionais acerca da qualidade de potenciais colaboradores e que para mitigá-las os pesquisadores beneficiam-se dos fluxos de informações sobre atributos individuais e sobre qualidade de pareamentos passados ao longo dos elos de coautoria.

Uma recomendação de política que se extrai é que, diante das fricções à formação de novas parcerias, parece produtivo incentivar mecanismos que facilitem a interação entre pesquisadores. Exemplos de instrumentos desta natureza inclui congressos, seminários e grupos de trabalho direcionados. Desta forma, os pesquisadores ganham oportunidades para familiarizarem-se com os estudos desenvolvidos pelos seus pares, permitindo identificar interesses comuns e possibilidades de colaboração. Adicionalmente, em tais ambientes, as pessoas, ao interagirem diretamente, podem apurar suas percepções acerca de atributos individuais relevantes de potenciais colaboradores.

Muitas questões ainda podem ser investigadas. Em particular, parece interessante estudar

como as interconexões entre os pesquisadores ajudam a entender produtividade acadêmica e evolução na carreira. Estes são alguns dos temas que se pretende explorar no futuro.

REFERÊNCIAS

AKERLOF, George A.; KRANTON, Rachel E.. Identity and Schooling: Some Lessons for the Economics of Education. **Journal of Economic Literature,** v. 40, n. 4, p.1167-1201, Dec. 2002.

ALLEN, Frankin; BABUS, Ana. **Networks in Finance**. 2007.

ALLISON, Paul D.; CHRISTAKIS, Nicholas. Fixed Effects Methods for the Analysis of Non-Repeated Events. **Sociological Methodology,** v. 36, n. 1, p.155-172, Aug. 2006.

AUSTEN-SMITH, David; FRYER JUNIOR, Roland Gerhard. An Economic Analysis of 'Acting White'. **The Quarterly Journal of Economics,** v. 120, n. 2, p.551-583, May 2005.

BABCHUK, Nicholas; KEITH, Bruce; PETER, George. Collaboration in sociology and other scientific disciplines: A comparative trend analysis of scholarship in the social, physical, and mathematical sciences. **The American Sociologist,** v.30, n. 3, p.5-21, Fall 1999.

BACCARA, Mariagiovanna; BAR-ISSAC, Heski. How to organize crime. **The Review of Economic Studies,** v. 75, n. 4, p.1039-1067, Oct. 2008.

BAERVELDT, Chris et al. Ethnic boundaries and personal choice: Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks. **Social Networks**, v. 26, n. 1, p.55-74, Jan. 2004.

BANDIERA, Oriana; RASUL, Imran. Social Networks and Technology Adoption in Northern Mozambique. **The Economic Journal,** v. 116, n. 514, p.862-902, 2006.

BANDIERA, Oriana; BARANKAY, Iwan; RASUL, Imran. Social capital in the workplace: Evidence on its formation and consequences. **Labour Economics**, v. 15, n. 4, p.725-749, Aug. 2008

BANERJEE, Abhijit; CHANDRASEKHAR, Arun G.; DUFLO, Esther; JACKSON, Matthew O. **The Diffusion of Microfinance.** NBER Working Paper 17743.

BARABÁSI, Albert-lászló; ALBERT, Réka. Emergence of Scaling in Random Networks. **Science,** v. 286, n. 5439, p.509-512, 15 Oct. 1999.

BARNETT, Andy H.; AULT, Richard W.; KASERMAN, David L.. The Rising Incidence of Co-authorship in Economics: Further Evidence. **The Review of Economics and Statistics,** v. 70, n. 3, p.539-543, Aug. 1988.

BERTRAND, Marianne; LUTTMER, Erzof. P.; MULLAINATHAN, Sendhil. Network Effects and Welfare Cultures. **The Quarterly Journal of Economics,** v. 115, n. 3, p.1019-1955, Aug. 2000.

BLOCH, Francis; DUTTA, Bhaskar. **Communication Networks with Endogenous Link Strength.** May 15th, 2007. Disponível em: <http://2007.icfsn.net/docs/francis_bloch.pdf>. Acesso em: 21 nov. 2009.

BLOCH, Francis; GENICOT, Garance; RAY, Debraj. Informal insurance in social networks. **Journal of Economic Theory,** v. 143, n. 1, p.36-58, Nov. 2008.

BLUME, Lawrence E.; DURLAUF, Steven N. **Identifying Social Interactions**: A Review, July 22nd, 2005. Disponível em: <http://www.ssc.wisc.edu/econ/archive/wp2005-12.pdf>. Acesso em: 04 out. 2009.

BOLLOBÁS, Béla. **Random Graphs.** 2. ed. Cambridge: Cambridge University Press, 2001. 500 p.

BRAMOULLÉ, Yann; DJEBBARI, Habiba; FORTIN, Bernard. Identification of Peer Effects through Social Networks. **Journal of Econometrics,** v. 150, n. 1, p.41-55, May 2009.

BRAMOULLÉ, Yann; KRANTON, Rachel. Risk-Sharing Networks. **Journal of Economic Behavior & Organization,** v. 64, n. 3-4, p.275-294, Nov.-Dec. 2007.

BROCK, William N.; DURLAUF, Steven N.. Interactions-Based Models. In: LEAMER, Edward E.; HECKMAN, James J. (Eds.). **Handbook of Econometrics.** North Holland, 2001. Cap. 54, p. 3297-3380.

CHIDAMBARAN, N. K.; KEDIA, Simi; PRABHALA, Nagpurnanand R.. **CEO-Director Connections and Corporate Fraud**: Not Just Whether You are Connected but How. Disponível em: <http://www.sec.gov/divisions/riskfin/seminar/prabhala051712.pdf>. Acesso em: 28 dez. 2012.

CABRALES, Antonio; CALVÓ-ARMENGOL, Antoni; ZENOU, Yves. **Effort and Synergies in Network Formation.** Feb. 1st, 2007. <http://2007.icfsn.net/docs/antoni_calvo-armengol.pdf>. Acesso em: 21 nov. 2009.

CALVÓ-ARMENGOL, Antoni; PATACCHINI, Eleonora; ZENOU, Yves. Peer Effects and Social Networks in Education. **The Review of Economic Studies,** p.1-47, Forthcoming 2010. <http://www.restud.com/uploads/papers/MS%2011863-3%20manuscript.pdf>. Acesso em: 6 nov. 2009.

CASELLA, Alessandra; HANAKI, Nobuyuki. Why Personal Ties Cannot Be Bought. **The American Economic Review,** v. 96, n. 2, p.261-264, May 2006.

COHEN, Lauren; FRAZZINI, Andrea; MALLOY, Christopher. The Small World of Investing: Board Connections and Mutual Fund Returns. **Journal of Political Economy**, v. 116, n. 5, p.951-979, Oct. 2008.

COLEMAN, James S.. Social Capital in the Creation of Human Capital. **The American Journal of Sociology,** v. 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure, p.S95-S120, 1988.

CONLEY, Timothy G.; UDRY, Christopher R. Learning about a New Technology: Pineapple in Ghana. **The American Economic Review,** v. 100, n. 1, p.35-69, March 2010.

CURRARINI, Sergio; JACKSON, Matthew; PIN, Paolo. An Economic Model of Friendship: Homophily, Minorities and Segregation. **Econometrica,** v. 77, n. 4, p.1003-1045, July 2009.

DUFLO, Esther; DUPAS, Pascaline; KREMER, Michael. **Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.** NBER Working Paper Nº 14475, p.1-51, Nov. 2008. <http://papers.nber.org/papers/w14475.pdf>. Acesso em 28 abr. 2009.

DURLAUF, Steven N.. **How can Statistical Mechanics Contribute to Social Sciences.** <http://www.ssc.wisc.edu/econ/archive/wp9913.pdf>. Acesso em: 4 out. 2009.

DURLAUF, Steven N.. **A Framework for the Study of Individual Behavior and Social Interactions.** <http://www.ssc.wisc.edu/econ/archive/wp2016r.pdf>. Acesso em: 4 out. 2009.

ELLIS, Richard S.. **Entropy, Large Deviations, and Statistical Mechanics.** Springer-verlag, 2005. 364 p.

ENDERSBY, James W.. Collaborative Research in the Social Sciences: Multiple Authorship and Publication Credit. **Social Science Quarterly,** v. 77, n. 2, p.375-392, June 1996.

ERDÖS, Paul; RÉNYI, Alfréd. On Random Graphs. **Publicationes Mathematicae Debrecen,** v. 6, p.290-297, 1959.

ERDÖS, Paul; RÉNYI, Alfréd. On the Evolution of Random Graphs. **Publication of The Mathematical Institute of The Hungarian Academy of Sciences,** v. 5, p.17-61, 1960.

ERDÖS, Paul; RÉNYI, Alfréd. On the Strength of Connectedness of a Random Graph. **Acta Mathamatica Academy of Sciences of Hungarica,** v. 12, p.261-267, 1961.

FAFCHAMPS, Marcel; GOYAL, Sanjeev; LEIJ, Marco J. Van Der. Matching and

Network Effects,. **Journal of The European Economic Association**, v. 8, n. 1, p.203-231, Jan. 2010.

FAFCHAMPS, Marcel; LUND, Susan.. Risk Sharing Networks in Rural Philippines. **Journal of Development Economics**, v. 71, n. 2, p.261-287, Aug. 2003.

FRYER JUNIOR, Roland Gerhard. Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century. **Journal of Economic Perspectives,** v. 21, n. 2, p.71-90, Spring 2007.

FRYER JUNIOR, Roland Gerhard; TORELLI, Paul. An Empirical Analysis of 'Acting White'. **Journal of Public Economics,** v. 94, n. 5-6, p.380-396, June 2010.

GALEOTTI, Andrea et al. Network Games. **The Review of Economic Studies,** v. 77, n. 1, p.218-244, Jan. 2010.

GLAESER, Edward L.; SACERDOTE, Bruce; SCHEINKMAN, José Alexandre. Crime and Social Interactions. **The Quarterly Journal of Economics,** v. 111, n. 2, p.507-548, May 1996.

GOMPERS, Paul; MUKHARLYAMOV, Vladimir; XUAN, Yuhai. **The Cost of Friendship.** Working Paper 18141. Disponível em: <http://www.nber.org/papers/w18141>. Accesso em : 20 de Junho de 2012.

GOYAL, Sanjeev; JOSHI, Sumit. Networks of Collaboration in Oligopoly. **Games and Economic Behavior,** v. 43, n. 1, p.57-85, April 2003.

GOYAL, Sanjeev; LEIJ, Marco J. van der; MORAGA-GONZÁLEZ, José Luis. Economics: An Emerging Small World. **Journal of Political Economy,** v. 114, n. 2, p.403-412, Apr. 2006.

GRANOVETTER, Mark S.. The Strength of Weak Ties. **The American Journal of Sociology,** v. 78, n. 6, p.1360-1380, May 1973.

GRANOVETTER, Mark S.. Network Sampling: Some First Steps. **The American Journal of Sociology,** v. 81, n. 6, p.1287-1303, May 1976.

HAYES, Rachel M.; OYER, Paul; SCHAEFER, Scott. Coworker Complementarity and the Stability of Top-Management Teams. **The Journal of Law, Economics, & Organization**, v. 22, n. 1, p.184-212, Apr. 2006.

HUDSON, John. Trends in Multi-Authored Papers in Economics. **The Journal of Economic Perspectives,** v. 10, n. 3, p.153-158, Summer 1996.

IOANNIDES, Yannis M.; LOURY, Linda Datcher. Job Information Networks, Neigh-

borhood Effects and Inequality. **Journal of Economic Literature**, v. 42, n. 4, p.1056-1093, Dec. 2004.

JACKSON, Matthew O.. **The Stability and Efficiency of Economic and Social Networks.** June 18th, 2001. <http://www.stanford.edu/jacksonm/netstab.pdf>. Acesso em: 21 mar. 2009.

JACKSON, Matthew O.. **A Survey of Models of Network Formation:** Stability and Efficiency, November 29th, 2003. <http://www.stanford.edu/jacksonm/netsurv.pdf>. Acesso em: 11 abr. 2009.

JACKSON, Matthew O.. **The Economics of Social Networls.** December 13th, 2005. <http://www.stanford.edu/~jacksonm/netect.pdf>. Acesso em: 10 abr. 2010.

JACKSON, Matthew O.. **Social and Economic Networks.** Princeton, NJ: Princeton University Press, 2008.

JACKSON, Matthew O.; ROGERS, Brian W.. The Economics of Small Worlds. **Journal of the European Economic Association,** v. 3, n. 2-3, p.617-627, April/May 2005.

JACKSON, Matthew O.; ROGERS, Brian W.. Meeting Strangers and Friends of Friends: How Random are Social Networks. **The American Economic Review,** v. 97, n. 3, p.890-915, June 2007.

JACKSON, Matthew O.; YARIV, Leeat. **Diffusion, Strategic Interaction, and Social Structure.** June, 2010. <http://www.stanford.edu/~jacksonm/diffusionchapter.pdf>. Acesso em: 17 abr. 2012.

JONES, Benjamin F.. The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder. **The Review of Economic Studies,** v. 76, n. 1, p.283-317, Jan. 2009.

KUHNEN, Camelia M.. Business Networks, Corporate Governance, and Contracting in the Mutual Fund Industry. **The Journal of Finance**, v. LXIV, n. 5, p.2185-2220, Oct. 2009.

LABAND, David N.; TOLLISON, Robert D.. Intellectual Collaboration. **Journal of Political Economy,** v. 108, n. 3, p.632-662, June 2000.

LANDIER, Augustin et al. Bottom-Up Corporate Governance. **Review of Finance**, v. 17, n. 1, p.161-201, Jan. 2012.

LANE, Julia. Let's make science metrics more scientic. **Nature,** v. 464, n. 7288, p.488-489, 24 Mar. 2010.

LEVIN, Sharon G.; STEPHAN, Paula E.. Research Productivity Over the Life Cycle: Evidence for Academic Scientists. **American Economic Review,** v. 81, n. 1, p.114-132, Mar. 1991.

MANSKI, Charles F.. Identification of Endogenous Social Effects: The Reflection Problem. **The Review of Economic Studies,** v. 60, n. 3, p.531-542, July 1993.

MANSKI, Charles F.. Economic Analysis of Social Interactions. **The Journal of Economic Perspectives,** v. 14, n. 3, p.115-136, Summer 2000.

MCDOWELL, John M.; MELVIN, Michael. The Determinants of Co-Authorship: An Analysis of the Economics Literature. **The Review of Economics and Statistics,** v. 65, n. 1, p.155-160, Feb. 1983.

MONTGOMERY, James D.. Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. **The American Economic Review**, v. 81, n. 5, p.1408-1418, Dec. 1991.

MOODY, James. The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. **American Sociological Review**, v. 69, n. 2, p.213-238, Apr. 2004.

NEWMAN, Mark E. J.. The structure of scientic collaboration networks. **Proceedings of The National Academy of Sciences of The United States of America,** v. 98, n. 2, p.404-409, Jan. 2001.

NEWMAN, Mark E. J.. Assortative mixing in networks. **Physical Review Letters,** v. 89, n. 20, 208701, 28 Oct. 2002.

NEWMAN, Mark E. J.. Mixing patterns in networks. **Physical Review Letters**, v. 67, n. 2, p.026126 27 Feb. 2003.

NEWMAN, Mark E. J.. The Structure and Function of Complex Networks. **SIAM Review**, v. 45, n. 2, p.167-256, 2003.

NEWMAN, Mark E. J.; PARK, Juyong. Why social networks are different from other types of networks. **Physical Review Letters,** v. 68, n. 3, p.036122, 22 Sep. 2003.

PIETTE, Michael J.; ROSS, Kevin L.. An Analysis of the Determinants of Co-authorship in Economics. **Journal of Economic Education,** v. 23, n. 3, p.277-283, summer 1992.

PRESSER, Stanley; ROSS, Kevin L.. Collaboration and the Quality of Research. **Social Studies of Science,** v. 10, n. 1, p.95-101, Feb. 1980.

PRICE, Derek John de Solla. **Little Science, Big Science**. New York: Columbia

University Press, 1963. 119 p.

ROGERS, Everett M.; BHOWMIK, Dilip K.. Homophily-Heterophily: Relational Concepts for Communication Research. **Public Opinion Quarterly,** v. 34, n. 4, p.523-538, Winter 1970.

ROGERSON, Richard; SHIMER, Robert; WRIGHT, Randall. Seach-Theoretic Models of the Labor Market: A Survey. **Journal of Economic Literature.** v. 43, n. 4, p.959-988, Dec. 2005.

SAUER, Raymond D.. Estimates of the Returns to Quality and Coauthorship in Economic Academia. **Journal of Political Economy,** v. 96, n. 4, p.855-866, Aug. 1988.

SCHWESER, Carl. The Economics of Academic Publishing. **Journal of Economic Education,** v. 14, n. 1, p.60-64, winter 1983.

TODD, Petra E.; WOLPIN, Kenneth I.. The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. **Journal of Human Capital,** v. 1, n. 1, p.91-136, Winter 2007. <http://www.journals.uchicago.edu/doi/pdf/10.1086/526401>. Acesso em: 09 fev. 2009.

TOPA, Giorgio. Social Interactions, Local Spillovers and Unemployment. **Review of Economic Studies,** v. 68, n. 2, p.261-295, Apr. 2001.

TOWNSEND, Robert. Risk and Insurance in Village India. **Econometrica,** v. 62, n. 3, p.539-591, May 1994.

UDRY, Christopher, Risk and Saving in Northern Nigeria. **The American Economic Review.** v. 85, n. 5, Dec. 1995.

WAHBA, Jackline; ZENOU, Yves. Density, Social Networks and Job Search Methods: Theory and Application to Egypt. **Journal of Development Economics,** v. 78, n. 2, p. 443-473, Dec. 2005.

WATSON, Tara; MCLANAHAN, Sara. **Marriage Meets the Joneses: Relative Income, Identity, and Marital Status.** Working Paper Nº 14773, p.1-59, Mar. 2009. Disponível em: <http://www.nber.org/papers/w14773>. Acesso em: 11 nov. 2009.

YEOMANS, Julia Mary. **Statistical Mechanics of Phase Transitions.** Oxford University Press, 1992. 168 p.

APÊNDICE 1 - ESTATÍSTICAS DESCRITIVAS

Na tabela A.1, abaixo, apresentam-se algumas estatísticas descritivas da população de indivíduos pertencentes ao conjunto Lattes/USP, definidos como aqueles que citam a Universidade de São Paulo em pelo menos uma das seguintes três opções: (i) endereço profissional, (ii) formação acadêmica/titulação ou (iii) atuação profissional. Todos os indivíduos que indicam a Universidade de São Paulo em pelo menos uma destas três opções têm seus currículos carregados para a base de dados, que contabiliza 62516 indivíduos. As estatísticas da tabela A.1 estão separadas por grandes áreas do conhecimento às quais os pesquisadores são associados se eles possuem ao menos um artigo publicado naqueles campos, conforme reportado pelos próprios indivíduos em seus currículos. Os pesquisadores que possuem artigos em mais de uma grande área do conhecimento são alocados ao conjunto de indivíduos de cada um dos campos citados.

Tabela A.1 - Estatísticas descritivas

	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
Número de indivíduos	4508	9762	6449	10018	10899	4682	12367	8434
Brasileiros (%)	98	97	96	96	98	98	99	99
Homens (%)	51	43	74	67	45	40	38	56
Titulação na USP	3796	7904	5608	8667	9841	4248	10043	7594
Graduação na USP	1876	3251	2532	4381	5487	2494	4575	4298
Graduação só na USP	1815	3107	2359	4079	4710	2105	4355	3605
Pós-graduação na USP	2757	6311	4359	6461	7211	2996	7149	5337
Pós-graduação só na USP	2063	5032	3320	4848	5416	2384	5928	3994
Mestrado na USP	2055	4354	3417	4795	5377	2355	5043	3930
Mestrado só na USP	2048	4331	3170	4745	5276	2322	5009	3808
Doutorado na USP	1675	4187	2634	4005	4219	1652	4611	3116
Doutorado só na USP	1570	3903	2475	3691	3981	1560	4370	2939
Graduação concluída (%)	86	91	88	88	89	85	90	90
Mestrado concluído (%)	59	55	63	58	59	55	52	56
Doutorado concluído (%)	37	41	37	39	32	29	37	30
Ano de conclusao da graduação 1 (Média)	1996	1996	1994	1996	1994	1994	1994	1994
Ano de conclusao do mestrado 1 (Média)	1999	1999	1998	1998	1999	2000	1999	1999
Ano de conclusao do doutorado 1 (Média)	1997	1998	1998	1997	1998	1999	1998	1998
Atuação na USP	2495	5603	3032	5056	4942	1997	6404	3449

APÊNDICE 2 - CLASSIFICAÇÃO DAS ÁREAS DO CONHECIMENTO

Abaixo constam as áreas do conhecimento pertencentes às nove grandes áreas do conhecimento.

Ciências agrárias: agronomia, ciência e tecnologia de alimentos, engenharia agrícola, medicina veterinária, recursos florestais e engenharia florestal, recursos pesqueiros e engenharia de pesca, zootecnia;

Ciências biológicas: biofísica, biologia geral, bioquímica, botânica, ecologia, farmacologia, fisiologia, genética, imunologia, microbiologia, morfologia, parasitologia, zoologia;

Ciências da saúde: educação física, enfermagem, farmácia, fisioterapia e terapia ocupacional, fonoaudiologia, medicina, nutrição, odontologia, saúde coletiva;

Ciências exatas e da terra: astronomia, ciência da computação, física, geociências, matemática, oceanografia, probabilidade e estatística, química;

Ciências humanas: antropologia, arqueologia, ciência política, educação, filosofia, geografia, história, psicologia, sociologia, teologia;

Ciências sociais aplicadas: administração, arquitetura e urbanismo, ciência da informação, comunicação, demografia, desenho industrial, direito, economia, economia doméstica, museologia, planejamento urbano e regional, serviço social, turismo;

Engenharias: engenharia aeroespacial, engenharia biomédica, engenharia civil, engenharia de materiais e metalúrgica, engenharia de minas, engenharia de produção, engenharia de transportes, engenharia elétrica, engenharia mecânica, engenharia naval e oceânica, engenharia nuclear, engenharia química, engenharia sanitária;

Linguística, letras e artes: linguística, letras, artes;

Outros: bioética, ciências ambientais, defesa, divulgação científica.

Tabela 2.1 - Incidência de coautoria (%): grandes áreas do conhecimento

Triênio	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais	Outros
1971 - 1973	89,39	82,39	69,23	75,60	32,63	0,00	77,43	24,39	100,00
1974 - 1976	82,76	87,52	68,37	76,82	33,33	8,82	82,07	19,82	0,00
1977 - 1979	88,86	88,17	75,19	85,66	30,42	7,61	87,13	29,88	100,00
1980 - 1982	91,52	88,62	84,77	85,11	28,86	13,41	86,64	21,35	75,00
1983 - 1985	89,47	90,12	68,64	87,41	28,27	12,50	86,94	16,37	80,00
1986 - 1988	89,71	90,79	67,63	87,38	32,44	9,43	88,68	22,90	60,00
1989 - 1991	88,70	91,42	78,69	88,12	29,15	12,00	87,63	24,57	44,44
1992 - 1994	92,53	91,91	78,81	89,28	29,69	12,53	87,69	24,03	82,14
1995 - 1997	92,08	93,05	87,70	91,44	31,12	19,47	88,87	29,82	70,97
1998 - 2000	95,03	94,97	90,60	93,54	36,20	14,24	90,53	36,10	64,10
2001 - 2003	96,81	96,26	94,37	94,89	41,88	16,32	92,88	40,35	80,95
2004 - 2006	98,06	97,29	94,58	95,98	51,73	24,07	94,96	54,89	72,22
2007 - 2009	99,19	98,18	96,47	96,25	57,38	25,99	97,07	67,71	71,43

Tabela 2.2 - Incidência de coautoria (%): ciências agrárias

Triênio	Agronomia	C&T Alimento	Eng. Agrícola	Veterinária	Florestal	Pesca	Zootecnia
1971 - 1973	87,50	93,75	0,00	93,94	100,00	50,00	92,31
1974 - 1976	85,53	81,58	37,50	89,80	40,00	100,00	75,61
1977 - 1979	90,31	100,00	92,31	89,47	68,00	100,00	78,85
1980 - 1982	92,55	94,12	76,19	88,52	83,33	100,00	95,51
1983 - 1985	88,39	93,44	80,95	94,17	79,25	100,00	89,84
1986 - 1988	89,04	86,11	70,83	94,41	85,71	85,71	90,98
1989 - 1991	92,18	94,94	65,00	91,12	59,09	88,24	90,43
1992 - 1994	94,24	91,85	77,97	95,13	82,54	62,50	93,53
1995 - 1997	93,68	94,09	90,29	90,66	78,21	97,67	92,83
1998 - 2000	93,74	95,35	92,59	96,62	87,36	97,01	97,83
2001 - 2003	96,88	96,30	90,57	97,31	91,67	94,44	98,89
2004 - 2006	97,68	99,07	97,44	98,49	94,81	96,51	98,62
2007 - 2009	99,39	98,54	100,00	99,37	96,32	98,00	100,00

Tabela 2.3 - Incidência de coautoria (%): ciências biológicas

Triênio	Biof.	Bio. Geral	Bioq.	Bot.	Eco.	Farmaco.	Fisio.	Gen.	Imuno.	Microbio.	Morfo.	Parasito.	Zoo.
1971 - 1973	100,00	50,00	94,55	68,42	80,00	78,79	94,12	89,47	100,00	92,31	83,02	95,00	40,91
1974 - 1976	100,00	50,00	98,13	57,14	83,33	83,02	89,09	92,86	97,14	100,00	92,59	90,00	56,45
1977 - 1979	97,06	72,73	98,58	74,19	84,21	86,21	94,44	85,37	97,30	92,86	93,42	94,44	63,01
1980 - 1982	93,02	100,00	96,46	73,68	79,66	75,82	92,74	91,40	95,74	96,36	95,48	93,94	55,10
1983 - 1985	97,73	57,14	96,24	75,51	75,41	86,13	96,39	94,20	96,55	96,20	95,83	93,90	61,74
1986 - 1988	90,62	80,00	92,90	81,25	80,65	90,91	92,54	90,38	96,98	99,42	97,00	88,62	72,12
1989 - 1991	93,59	91,67	94,24	83,12	89,47	93,38	93,45	92,91	96,46	91,43	96,77	98,60	71,62
1992 - 1994	93,10	98,48	94,34	79,82	77,24	94,06	95,60	92,94	95,52	93,98	97,12	93,46	73,64
1995 - 1997	95,48	93,10	95,41	87,26	85,43	96,64	94,09	93,51	96,35	93,56	96,88	94,12	77,84
1998 - 2000	92,77	93,02	96,37	89,84	88,45	97,52	95,21	97,06	96,96	95,71	97,63	97,14	85,03
2001 - 2003	98,61	93,24	98,22	90,36	91,38	97,68	96,01	98,16	97,46	97,32	98,62	99,15	90,32
2004 - 2006	99,25	96,79	98,97	91,85	94,85	98,99	97,29	98,76	98,32	98,68	97,91	99,14	91,43
2007 - 2009	99,22	97,03	99,37	97,09	95,15	100,00	98,11	98,16	99,58	99,76	99,26	99,10	94,20

Tabela 2.4 - Incidência de coautoria (%): engenharias

Triênio	Aero.	Biom.	Civil	Mat. e Metal.	Minas	Prod.	Transp.	Elétrica	Mec.	Nav. e Ocean	Nuclear	Química	Sanit.
1971 - 1973	0,00	NA	58,33	86,21	50,00	50,00	100,00	60,00	0,00	100,00	100,00	50,00	100,00
1974 - 1976	NA	100,00	14,29	86,67	50,00	NA	66,67	66,67	40,00	100,00	100,00	69,23	44,44
1977 - 1979	NA	100,00	28,57	79,55	0,00	25,00	66,67	85,71	0,00	50,00	100,00	92,86	87,50
1980 - 1982	NA	100,00	64,29	83,72	66,67	20,00	100,00	81,25	83,33	83,33	100,00	88,24	93,33
1983 - 1985	50,00	86,67	45,16	77,27	77,78	35,71	66,67	74,19	60,00	77,78	94,12	79,17	54,05
1986 - 1988	100,00	84,21	36,17	75,90	31,25	47,62	88,89	78,72	66,67	40,00	93,75	73,44	64,71
1989 - 1991	NA	82,22	67,31	91,40	50,00	50,00	80,00	83,87	81,82	80,00	94,12	75,00	72,73
1992 - 1994	NA	82,93	60,00	93,55	66,67	62,50	88,24	90,38	75,00	90,00	95,00	88,41	67,74
1995 - 1997	100,00	92,59	87,92	93,39	73,17	77,36	76,67	90,48	92,52	93,75	96,88	89,55	72,37
1998 - 2000	83,33	91,72	84,36	95,24	89,19	77,44	82,61	91,81	95,65	91,30	93,10	95,31	79,59
2001 - 2003	100,00	98,90	91,59	96,71	100,00	86,19	86,96	94,21	96,09	96,77	97,62	98,52	84,93
2004 - 2006	95,65	93,10	94,59	96,05	95,74	87,30	92,00	97,05	92,54	91,67	95,83	97,75	93,14
2007 - 2009	100,00	100,00	90,91	99,49	96,77	90,48	95,00	97,98	97,12	85,71	96,43	98,29	98,35

Tabela 2.5 - Incidência de coautoria (%): ciências exatas e da terra

Triênio	Astronomia	Computação	Física	Geociências	Matemática	Oceanografia	Prob. e Estat.	Química
1971 - 1973	55,00	20,00	80,70	73,33	13,33	87,50	80,00	92,65
1974 - 1976	39,13	100,00	79,80	77,94	28,57	94,44	61,54	85,09
1977 - 1979	83,87	66,67	88,83	85,57	48,57	81,48	76,19	93,33
1980 - 1982	67,50	41,67	89,97	88,89	55,81	75,61	78,79	91,70
1983 - 1985	80,00	71,43	91,00	77,70	62,86	81,82	82,69	93,84
1986 - 1988	86,05	86,96	91,19	81,91	54,29	76,60	81,71	93,21
1989 - 1991	90,62	84,21	90,73	84,03	58,95	77,78	84,76	94,34
1992 - 1994	91,18	73,08	93,43	79,93	63,48	68,57	83,33	94,74
1995 - 1997	92,64	86,25	94,18	86,65	73,08	89,19	75,52	96,19
1998 - 2000	94,08	88,36	94,99	88,68	77,64	96,88	92,02	96,55
2001 - 2003	94,30	93,95	95,29	93,31	84,91	97,01	93,23	97,05
2004 - 2006	97,96	93,98	97,40	94,90	80,79	95,36	92,80	98,11
2007 - 2009	94,37	95,05	97,95	95,36	87,06	94,95	96,11	97,38

Tabela 2.6 - Incidência de coautoria (%): ciências humanas

Triênio	Antropologia	Arqueologia	Política	Educação	Filosofia	Geografia	História	Psicologia	Sociologia	Teologia
1971 - 1973	33,33	100,00	0,00	42,86	6,67	100,00	6,67	43,48	60,00	NA
1974 - 1976	25,00	20,00	22,22	56,67	16,67	16,67	10,53	39,13	40,00	NA
1977 - 1979	20,00	9,09	27,27	55,81	6,25	57,14	15,79	42,86	6,90	NA
1980 - 1982	18,18	50,00	20,00	27,78	7,14	22,22	24,49	39,45	21,57	NA
1983 - 1985	20,51	23,08	20,45	37,62	3,85	18,18	11,11	45,95	25,97	0,00
1986 - 1988	25,00	11,11	22,03	34,55	30,61	40,00	17,65	48,15	23,85	NA
1989 - 1991	15,28	31,03	20,73	22,40	20,93	24,14	12,82	53,65	26,67	NA
1992 - 1994	20,20	25,93	32,77	31,35	18,46	26,09	10,38	46,07	17,41	0,00
1995 - 1997	20,00	25,51	26,62	34,81	15,52	25,00	13,75	46,09	18,22	0,00
1998 - 2000	25,59	30,97	29,65	40,42	12,35	27,43	12,50	53,33	21,03	10,00
2001 - 2003	22,34	47,62	24,21	46,46	20,81	39,31	17,60	61,38	21,40	0,00
2004 - 2006	37,59	60,53	29,69	54,76	24,16	50,72	21,50	70,03	28,00	0,00
2007 - 2009	40,91	57,58	32,65	64,43	32,08	56,00	25,95	74,40	42,16	66,67

Tabela 2.7 - Incidência de coautoria (%): linguística, letras e artes

Triênio	Artes	Letras	Linguística
1971 - 1973	0,00	0,00	NA
1974 - 1976	0,00	9,52	9,09
1977 - 1979	8,70	6,90	7,14
1980 - 1982	5,56	13,73	25,00
1983 - 1985	20,69	11,94	7,41
1986 - 1988	1,52	7,97	17,95
1989 - 1991	7,04	9,94	27,27
1992 - 1994	5,71	13,77	23,00
1995 - 1997	9,59	25,62	18,95
1998 - 2000	19,23	11,02	19,80
2001 - 2003	13,07	13,00	23,81
2004 - 2006	23,16	17,46	33,56
2007 - 2009	28,46	20,07	38,64

Tabela 2.8 - Incidência de coautoria (%): ciências da saúde

Triênio	Edu. Física	Enfermagem	Farmácia	Fisio. e Ocup.	Fonoaudio.	Medicina	Nutrição	Odontologia	S. Coletiva
1971 - 1973	100,00	71,43	77,78	NA	100,00	80,78	72,73	76,61	56,10
1974 - 1976	100,00	62,86	97,83	NA	100,00	88,61	54,76	74,37	68,93
1977 - 1979	41,67	80,00	91,67	66,67	100,00	89,09	71,43	88,69	80,67
1980 - 1982	36,36	75,68	92,00	50,00	75,00	88,55	79,38	87,85	77,78
1983 - 1985	84,62	68,63	96,70	88,89	100,00	89,85	82,95	85,27	70,63
1986 - 1988	40,91	70,35	96,81	88,89	84,00	91,27	88,66	91,02	76,44
1989 - 1991	62,96	72,34	95,04	57,14	70,83	91,21	81,51	86,95	74,50
1992 - 1994	65,00	70,35	85,64	77,42	73,33	91,57	89,81	89,29	74,75
1995 - 1997	70,31	74,07	86,08	75,73	75,34	91,10	91,15	92,43	78,06
1998 - 2000	81,36	80,68	90,55	85,23	89,36	90,88	91,00	96,17	81,21
2001 - 2003	89,51	87,20	96,56	90,46	92,78	91,81	93,50	97,11	88,29
2004 - 2006	92,10	92,49	98,87	96,03	92,37	94,58	94,43	97,42	90,54
2007 - 2009	93,02	97,63	99,65	99,07	98,52	96,68	98,03	98,15	93,48

Tabela 2.9 - Incidência de coautoria (%): ciências sociais aplicadas

Triênio	Adm.	Arquit.	C. Info.	Comun.	Demog.	Indust.	Dir.	Econ. Dom.	Econ.	Museo.	Urb. e Reg.	Serv. Social	Tur.
1971 - 1973	0,00	0,00	20,00	NA	33,33	NA	0,00	NA	53,85	NA	50,00	NA	0,00
1974 - 1976	8,33	20,00	4,55	33,33	33,33	0,00	8,82	NA	38,10	NA	66,67	NA	NA
1977 - 1979	29,41	31,58	25,00	16,67	71,43	NA	11,43	NA	42,86	0,00	40,00	NA	0,00
1980 - 1982	14,63	12,12	33,33	0,00	66,67	0,00	3,33	NA	36,54	0,00	46,67	0,00	0,00
1983 - 1985	12,90	16,07	14,29	29,41	20,00	0,00	6,52	NA	23,81	NA	44,44	NA	0,00
1986 - 1988	29,23	16,44	28,12	14,63	36,36	0,00	4,44	NA	36,36	33,33	46,15	NA	10,00
1989 - 1991	29,41	25,53	59,62	6,85	60,00	NA	8,53	50,00	26,50	40,00	27,78	0,00	20,00
1992 - 1994	34,85	16,91	44,44	12,50	20,00	0,00	9,41	NA	31,45	40,00	50,00	0,00	10,53
1995 - 1997	46,30	18,18	34,88	17,32	13,33	0,00	11,31	0,00	46,64	20,00	27,03	25,00	11,11
1998 - 2000	53,73	25,15	46,48	13,17	25,00	0,00	18,24	100,00	59,75	20,00	25,00	20,00	18,92
2001 - 2003	63,41	36,65	59,49	14,64	84,62	37,50	20,36	NA	54,62	54,55	41,51	22,22	27,08
2004 - 2006	82,11	42,80	58,24	23,13	66,67	7,14	17,82	100,00	74,45	38,10	53,03	30,23	54,17
2007 - 2009	91,94	52,63	71,88	48,08	50,00	71,43	28,05	NA	76,60	80,00	76,47	100,00	55,56

Tabela 2.10 - Incidência de coautoria (%): outros

Triênio	Bioética	Ciências Ambientais	Divulgação Científica
1971 - 1973	NA	NA	NA
1974 - 1976	100,00	NA	NA
1977 - 1979	100,00	NA	NA
1980 - 1982	100,00	100,00	100,00
1983 - 1985	NA	66,67	NA
1986 - 1988	0,00	NA	100,00
1989 - 1991	0,00	40,00	NA
1992 - 1994	57,14	50,00	0,00
1995 - 1997	57,14	20,00	0,00
1998 - 2000	84,62	57,89	60,00
2001 - 2003	50,00	66,67	71,43

Tabela 2.11 - Extensão de coautoria: grandes áreas do conhecimento

Triênio	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais	Outros
1971 - 1973	2,939	2,707	2,169	2,478	1,547	1,000	3,133	1,415	2,000
1974 - 1976	2,952	3,028	2,439	2,557	1,698	1,088	3,569	1,306	1,000
1977 - 1979	3,223	3,076	2,383	2,990	1,462	1,152	3,592	1,415	11,000
1980 - 1982	3,092	3,217	2,649	2,874	1,534	1,244	3,733	1,312	3,500
1983 - 1985	3,254	3,257	2,288	3,036	1,426	1,174	3,813	1,288	3,800
1986 - 1988	3,202	3,460	2,370	3,079	1,570	1,160	4,015	1,404	2,600
1989 - 1991	3,262	3,657	2,755	3,346	1,529	1,153	4,080	1,398	1,778
1992 - 1994	3,443	3,899	2,575	3,574	1,580	1,199	4,183	1,404	4,786
1995 - 1997	3,601	4,072	2,919	3,803	1,626	1,300	4,161	1,478	3,677
1998 - 2000	3,797	4,349	3,131	3,918	1,758	1,222	4,190	1,561	2,846
2001 - 2003	4,317	4,682	3,390	4,148	1,873	1,317	4,412	1,657	4,032
2004 - 2006	4,696	5,169	3,488	4,354	2,125	1,334	4,701	1,952	3,111
2007 - 2009	5,147	5,461	3,720	4,489	2,227	1,438	5,138	2,247	2,800

Tabela 2.12 - Extensão de coautoria: ciências agrárias

Triênio	Agronomia	C&T Alimento	Eng. Agrícola	Veterinária	Florestal	Pesca	Zootecnia
1971 - 1973	2,734	2,938	1,000	3,364	3,500	2,500	2,846
1974 - 1976	2,925	2,842	1,625	3,612	2,000	3,000	2,659
1977 - 1979	3,159	3,231	3,000	3,930	2,520	3,500	2,846
1980 - 1982	3,103	2,882	2,286	3,525	2,458	4,000	3,191
1983 - 1985	3,229	3,115	2,571	3,718	2,491	2,500	3,352
1986 - 1988	3,070	2,639	2,167	3,820	2,667	3,500	3,484
1989 - 1991	3,276	3,241	2,250	3,696	2,318	3,941	3,278
1992 - 1994	3,344	3,081	2,712	3,880	2,730	2,500	3,950
1995 - 1997	3,420	3,167	3,311	4,006	2,885	3,651	4,021
1998 - 2000	3,479	3,359	3,370	4,477	2,943	3,493	4,203
2001 - 2003	3,873	3,564	3,387	5,114	3,133	3,750	5,239
2004 - 2006	4,310	3,987	3,966	5,538	3,818	3,860	5,348
2007 - 2009	4,614	4,423	4,308	6,273	4,272	4,580	5,547

Tabela 2.13 - Extensão de coautoria: ciências biológicas

Triênio	Biof.	Bio. Geral	Bioq.	Bot.	Eco.	Farmaco.	Fisio.	Gen.	Imuno.	Microbio.	Morfo.	Parasito.	Zoo.
1971 - 1973	3,067	1,500	2,800	2,158	2,300	2,727	3,382	2,868	3,667	3,692	2,868	2,500	1,591
1974 - 1976	3,000	1,750	3,290	2,000	2,667	2,528	3,527	2,871	4,086	3,800	3,185	3,700	1,855
1977 - 1979	3,882	1,909	3,319	2,710	2,553	3,069	3,333	2,732	3,622	3,524	3,000	3,241	1,904
1980 - 1982	2,814	2,667	3,551	2,105	2,475	2,846	3,250	2,935	4,170	3,655	3,477	3,909	1,857
1983 - 1985	3,636	1,714	3,380	2,367	2,426	2,985	3,380	3,275	4,181	3,646	3,304	3,829	2,013
1986 - 1988	3,609	2,667	3,445	2,422	3,430	3,364	3,373	3,300	4,025	4,064	3,489	3,699	2,436
1989 - 1991	3,846	2,889	3,709	2,662	3,556	3,511	3,429	3,409	4,441	4,192	3,806	4,587	2,183
1992 - 1994	4,069	3,227	4,172	2,569	2,848	3,953	3,821	3,754	4,779	4,177	4,159	4,601	2,310
1995 - 1997	4,110	3,690	4,615	2,599	3,211	4,276	3,957	4,171	4,791	4,358	4,067	4,226	2,580
1998 - 2000	4,313	3,829	5,470	2,979	3,479	4,298	4,149	4,446	5,077	4,633	4,112	5,763	2,705
2001 - 2003	5,639	4,527	5,613	3,193	3,420	4,753	4,594	5,206	5,484	4,966	4,752	5,340	2,697
2004 - 2006	5,724	5,199	6,267	3,507	3,899	5,523	4,895	6,142	6,291	5,440	4,939	6,288	2,925
2007 - 2009	5,504	4,802	6,495	3,913	3,755	6,397	5,226	6,081	7,317	5,988	5,602	6,235	3,095

Tabela 2.14 - Extensão de coautoria: engenharias

Triênio	Aero.	Biom.	Civil	Mat. e Metal.	Minas	Prod.	Transp.	Elétrica	Mec.	Nav. e Ocean	Nuclear	Química	Sanit.
1971 - 1973	1,000	NA	1,750	2,310	1,750	2,000	2,000	3,000	1,000	3,000	4,000	1,750	2,000
1974 - 1976	NA	4,000	1,286	2,333	1,667	NA	1,667	4,167	1,800	2,333	4,333	2,654	1,667
1977 - 1979	NA	3,500	1,929	2,364	1,000	1,500	2,000	2,357	1,000	1,500	2,692	2,857	2,375
1980 - 1982	NA	2,833	2,357	2,395	1,833	1,400	2,000	2,750	2,333	2,167	3,667	2,647	2,667
1983 - 1985	1,500	3,200	1,871	2,273	2,222	1,357	2,667	2,355	2,100	3,000	3,294	2,583	1,838
1986 - 1988	2,000	3,316	1,511	2,422	1,500	1,952	2,444	3,000	1,905	1,600	3,125	2,406	2,294
1989 - 1991	NA	3,022	2,019	3,290	1,889	1,667	2,200	2,968	2,473	2,300	4,882	2,719	2,333
1992 - 1994	NA	3,000	1,848	3,315	1,933	1,859	2,235	2,731	2,310	2,450	3,400	3,043	2,548
1995 - 1997	2,000	3,296	2,449	3,247	2,098	2,217	2,267	3,071	2,884	2,375	3,938	3,328	2,408
1998 - 2000	2,667	3,624	2,436	3,532	2,514	2,286	2,174	3,185	3,256	2,870	4,276	3,465	2,602
2001 - 2003	3,800	4,149	2,654	3,787	2,684	2,398	2,739	3,582	3,222	2,903	5,071	3,696	3,027
2004 - 2006	3,130	4,517	2,746	4,017	2,957	2,552	2,640	3,727	3,061	3,167	4,500	3,925	3,486
2007 - 2009	5,423	5,416	2,855	4,158	3,774	2,651	2,450	3,770	3,532	3,000	4,679	4,143	3,967

Tabela 2.15 - Extensão de coautoria: ciências exatas e da terra

Triênio	Astronomia	Computação	Física	Geociências	Matemática	Oceanografia	Prob. e Estat.	Química
1971 - 1973	1,700	1,200	2,711	2,600	1,400	2,750	3,600	2,765
1974 - 1976	1,565	2,000	2,657	2,426	1,286	2,389	2,692	2,956
1977 - 1979	2,548	2,000	3,112	2,979	1,743	2,481	2,429	3,418
1980 - 1982	2,325	1,500	3,000	3,020	1,860	2,293	2,667	3,187
1983 - 1985	2,960	2,286	3,090	3,043	1,886	2,818	2,885	3,217
1986 - 1988	2,686	2,478	3,223	3,298	1,814	2,489	2,671	3,318
1989 - 1991	2,750	2,474	3,663	3,479	1,884	2,815	2,800	3,445
1992 - 1994	3,529	2,538	4,006	3,118	1,939	2,671	2,569	3,651
1995 - 1997	3,798	2,812	4,550	3,405	2,111	2,865	2,483	3,712
1998 - 2000	3,743	3,178	4,551	3,411	2,378	3,508	3,046	3,898
2001 - 2003	4,082	3,577	4,732	3,936	2,582	3,366	3,620	4,141
2004 - 2006	4,374	3,732	4,911	3,944	2,513	4,033	3,712	4,519
2007 - 2009	5,310	3,396	4,861	4,085	2,587	5,222	3,767	4,868

Tabela 2.16 - Extensão de coautoria: ciências humanas

Triênio	Antropologia	Arqueologia	Política	Educação	Filosofia	Geografia	História	Psicologia	Sociologia	Teologia
1971 - 1973	1,333	2,000	1,000	1,857	1,067	2,000	1,067	1,870	1,800	NA
1974 - 1976	1,500	1,200	1,222	2,100	1,167	1,500	1,158	1,804	2,350	NA
1977 - 1979	1,200	1,091	1,591	2,023	1,062	1,571	1,184	1,619	1,069	NA
1980 - 1982	1,591	1,800	1,333	1,811	1,214	1,222	1,286	1,615	1,314	NA
1983 - 1985	1,308	1,269	1,205	1,673	1,038	1,318	1,143	1,658	1,338	1,000
1986 - 1988	1,250	1,111	1,271	1,628	1,347	2,080	1,235	2,006	1,330	NA
1989 - 1991	1,181	1,448	1,341	1,424	1,395	1,483	1,197	2,107	1,327	NA
1992 - 1994	1,364	1,407	1,429	1,627	1,231	1,551	1,179	1,987	1,244	1,000
1995 - 1997	1,450	1,520	1,344	1,729	1,276	1,464	1,175	2,065	1,251	1,000
1998 - 2000	1,559	1,752	1,457	1,829	1,185	1,602	1,211	2,182	1,329	1,100
2001 - 2003	1,415	2,105	1,379	2,031	1,389	1,800	1,227	2,326	1,368	1,000
2004 - 2006	2,085	2,645	1,526	2,209	1,349	1,812	1,371	2,602	1,427	1,000
2007 - 2009	2,106	2,667	1,490	2,235	1,368	2,270	1,411	2,815	1,588	2,000

Tabela 2.17 - Extensão de coautoria: linguística, letras e artes

Triênio	Artes	Letras	Linguística
1971 - 1973	1,000	1,000	NA
1974 - 1976	1,000	1,095	1,091
1977 - 1979	1,130	1,138	1,214
1980 - 1982	1,056	1,255	1,500
1983 - 1985	1,276	1,179	1,093
1986 - 1988	1,015	1,142	1,321
1989 - 1991	1,070	1,111	1,377
1992 - 1994	1,109	1,182	1,420
1995 - 1997	1,151	1,367	1,346
1998 - 2000	1,300	1,157	1,327
2001 - 2003	1,248	1,265	1,446
2004 - 2006	1,299	1,277	1,420
2007 - 2009	1,358	1,416	1,576

Tabela 2.18 - Extensão de coautoria: ciências da saúde

Triênio	Edu. Física	Enfermagem	Farmácia	Fisio. e Ocup.	Fonoaudio.	Medicina	Nutrição	Odontologia	S. Coletiva
1971 - 1973	6,600	2,571	3,222	NA	2,500	3,368	3,409	2,589	2,585
1974 - 1976	4,000	2,829	3,522	NA	3,000	4,130	2,167	2,704	3,000
1977 - 1979	2,417	3,289	3,542	2,000	4,000	3,786	3,048	3,167	3,353
1980 - 1982	3,000	3,203	3,120	2,667	2,000	4,005	3,371	3,224	3,212
1983 - 1985	4,385	2,569	3,593	2,556	2,333	4,079	3,045	3,380	3,091
1986 - 1988	2,045	2,907	3,894	3,222	3,400	4,407	3,464	3,284	3,288
1989 - 1991	2,407	2,587	4,184	2,393	2,333	4,599	3,118	3,237	3,298
1992 - 1994	2,962	2,715	3,574	2,758	2,711	4,733	3,688	3,391	3,426
1995 - 1997	3,078	2,615	3,671	3,194	3,342	4,718	3,721	3,425	3,364
1998 - 2000	3,538	2,789	3,721	3,460	3,149	4,755	3,605	3,715	3,723
2001 - 2003	4,222	2,856	4,417	3,637	3,365	4,998	4,000	4,040	4,183
2004 - 2006	4,123	3,114	5,087	4,364	3,659	5,437	3,921	4,499	4,127
2007 - 2009	4,090	3,588	5,634	4,623	3,911	5,930	4,547	4,979	4,668

Tabela 2.19 - Extensão de coautoria: ciências sociais aplicadas

Triênio	Adm.	Arquit.	C. Info.	Comun.	Demog.	Indust.	Dir.	Econ. Dom.	Econ.	Museo.	Urb. e Reg.	Serv. Social	Tur.
1971 - 1973	1,000	1,000	1,200	NA	1,333	NA	1,000	NA	2,077	NA	1,500	NA	1,000
1974 - 1976	1,167	1,333	1,045	1,333	1,500	1,000	1,088	NA	1,762	NA	1,667	NA	NA
1977 - 1979	1,412	1,474	1,417	1,167	1,714	NA	1,114	NA	1,673	1,000	1,800	NA	1,000
1980 - 1982	1,195	1,121	1,833	1,000	1,667	1,000	1,033	NA	1,500	1,000	1,733	1,000	1,000
1983 - 1985	1,242	1,286	1,571	1,294	1,300	1,000	1,109	NA	1,460	NA	1,611	NA	1,000
1986 - 1988	1,554	1,370	1,594	1,171	1,455	1,000	1,044	NA	1,645	1,333	1,615	NA	1,100
1989 - 1991	1,422	1,489	2,269	1,096	1,800	NA	1,109	2,000	1,316	2,000	1,389	1,000	1,400
1992 - 1994	1,576	1,338	2,000	1,198	1,300	1,000	1,106	NA	1,459	1,400	2,118	1,000	1,105
1995 - 1997	1,704	1,421	1,872	1,283	1,133	1,000	1,186	1,000	1,648	1,200	1,351	1,625	1,111
1998 - 2000	1,869	1,450	1,690	1,198	1,250	1,000	1,331	4,000	1,792	1,200	1,423	1,333	1,351
2001 - 2003	1,968	1,524	2,177	1,285	2,308	1,375	1,458	NA	1,812	1,545	1,623	1,456	1,396
2004 - 2006	2,393	1,703	2,330	1,340	2,000	1,071	1,327	2,000	2,236	2,286	1,864	1,837	1,688
2007 - 2009	2,786	1,992	2,297	1,635	2,000	2,571	1,434	NA	2,404	2,800	2,706	3,000	2,037

Tabela 2.20 - Extensão de coautoria: outros

Triênio	Bioética	Ciências Ambientais	Divulgação Científica
1971 - 1973	NA	NA	NA
1974 - 1976	4,000	NA	NA
1977 - 1979	4,000	NA	NA
1980 - 1982	2,000	5,000	4,000
1983 - 1985	NA	2,000	NA
1986 - 1988	1,000	NA	3,000
1989 - 1991	1,000	1,400	NA
1992 - 1994	2,857	2,500	1,000
1995 - 1997	2,143	1,200	1,000
1998 - 2000	2,769	2,316	5,600
2001 - 2003	1,500	2,571	3,714

Tabela 2.21 - Estatísticas descritivas: 1971-2009

	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
Coautoria	0,949	0,943	0,898	0,923	0,386	0,178	0,914	0,406
Nº autores	4,103	4,364	3,206	3,876	1,785	1,280	4,362	1,686
Sexo	0,760	0,580	0,859	0,764	0,526	0,390	0,615	0,687
Idade	22,724	22,487	24,101	23,477	23,349	23,964	24,946	23,548
Ordem	30,810	29,336	15,889	28,416	18,373	15,745	43,657	15,106
Bolsa	0,519	0,580	0,515	0,613	0,502	0,416	0,303	0,343
Titular	0,010	0,044	0,022	0,043	0,056	0,073	0,065	0,089
Carga	32,331	31,978	31,072	33,285	27,599	23,954	26,090	25,189
Exclusiva	0,399	0,504	0,468	0,579	0,368	0,352	0,294	0,258
$\Delta ATgr$	0,085	0,109	0,111	0,070	0,116	0,097	0,062	0,101
$\Delta ATar$	0,162	0,301	0,227	0,126	0,247	0,198	0,146	0,190
$\Delta ATsub$	0,145	0,239	0,162	0,118	0,239	0,245	0,177	0,224
$\Delta ATesp$	0,478	0,513	0,582	0,475	0,642	0,665	0,367	0,600
$\Delta ARgr$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta ARar$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta ARsub$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta AResp$	0,159	0,150	0,211	0,160	0,235	0,224	0,079	0,257
N-teórico	0,030	0,119	0,022	0,044	0,038	0,064	0,016	0,026
Aplicada	0,025	0,111	0,021	0,043	0,004	0,064	0,014	0,007
Empírica	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Quant.	0,003	0,006	0,000	0,001	0,000	0,000	0,000	0,020
Exper.	0,002	0,002	0,000	0,000	0,033	0,000	0,002	0,000

Tabela 2.22 - Estatísticas descritivas: 2000-2006

	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
Coautoria	0,992	0,977	0,940	0,963	0,545	0,057	0,919	0,885
Nº autores	4,712	5,064	3,233	4,203	1,851	1,057	4,008	2,327
Sexo	0,808	0,509	0,810	0,714	0,426	0,453	0,505	0,635
Idade	23,837	23,047	25,939	24,627	26,680	25,679	21,577	22,692
Ordem	38,318	26,591	16,000	32,701	20,812	17,245	32,403	10,115
Bolsa	0,700	0,756	0,534	0,792	0,558	0,471	0,413	0,423
Titular	0,000	0,000	0,000	0,000	0,040	0,000	0,037	0,077
Carga	37,382	35,235	38,182	40,000	28,600	21,600	23,710	26,667
Exclusiva	0,667	0,724	0,312	0,821	0,400	0,333	0,243	0,231
$\Delta ATgr$	0,063	0,109	0,060	0,024	0,050	0,096	0,060	0,019
$\Delta ATar$	0,203	0,323	0,172	0,081	0,099	0,192	0,159	0,038
$\Delta ATsub$	0,137	0,304	0,106	0,101	0,243	0,333	0,228	0,094
$\Delta ATesp$	0,474	0,653	0,624	0,438	0,490	0,250	0,500	0,435
$\Delta ARgr$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta ARar$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta ARsub$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta AResp$	0,151	0,155	0,188	0,185	0,163	0,000	0,091	0,357
N-teórico	0,022	0,068	0,017	0,033	0,010	0,113	0,000	0,038
Aplicada	0,022	0,059	0,017	0,033	0,000	0,113	0,000	0,019
Empírica	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Quant.	0,000	0,009	0,000	0,000	0,000	0,000	0,000	0,019
Exper.	0,000	0,000	0,000	0,000	0,010	0,000	0,000	0,000
Intern.	0,438	0,905	0,517	0,811	0,257	0,057	0,314	0,000
Nível A	0,945	0,923	1,000	0,910	0,871	0,302	0,270	0,827
Nível B	0,025	0,077	0,000	0,023	0,000	0,377	0,284	0,019
Nível C	0,030	0,000	0,000	0,066	0,129	0,321	0,446	0,154
Nível AI	0,438	0,905	0,517	0,767	0,257	0,000	0,235	0,000
Nível BI	0,000	0,000	0,000	0,020	0,000	0,000	0,051	0,000
Nível CI	0,000	0,000	0,000	0,023	0,000	0,057	0,027	0,000

Tabela 2.23 - Estatísticas descritivas: 2007-2008

	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
Coautoria	0,994	0,979	0,964	0,926	0,537	0,260	0,969	0,772
Nº autores	5,283	5,398	3,350	4,207	2,088	1,375	4,521	2,445
Sexo	0,802	0,576	0,848	0,760	0,470	0,531	0,468	0,647
Idade	23,340	25,070	27,873	26,657	24,281	23,925	25,496	25,434
Ordem	43,159	40,444	24,185	35,174	23,276	25,240	45,924	17,566
Bolsa	0,754	0,732	0,527	0,760	0,627	0,419	0,405	0,429
Titular	0,000	0,000	0,000	0,000	0,000	0,000	0,058	0,042
Carga	30,000	37,810	21,069	40,000	33,500	NaN	25,536	17,111
Exclusiva	0,368	0,459	0,312	0,600	0,258	0,400	0,163	0,125
$\Delta ATgr$	0,059	0,066	0,061	0,041	0,077	0,043	0,044	0,036
$\Delta ATar$	0,131	0,173	0,179	0,086	0,132	0,118	0,140	0,087
$\Delta ATsub$	0,179	0,182	0,144	0,135	0,199	0,153	0,201	0,125
$\Delta ATesp$	0,344	0,384	0,421	0,315	0,373	0,809	0,428	0,380
$\Delta ARgr$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta ARar$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta ARsub$	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
$\Delta AResp$	0,117	0,112	0,124	0,128	0,098	0,070	0,071	0,148
N-teórico	0,011	0,123	0,033	0,041	0,019	0,021	0,010	0,006
Aplicada	0,011	0,122	0,033	0,041	0,000	0,021	0,009	0,000
Empírica	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Quant.	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,006
Exper.	0,000	0,002	0,000	0,000	0,019	0,000	0,001	0,000
Est. A1	0,022	0,072	0,112	0,068	0,019	0,000	0,032	0,003
Est. A2	0,072	0,108	0,046	0,060	0,107	0,083	0,081	0,032
Est. B1	0,317	0,199	0,076	0,278	0,152	0,052	0,202	0,147
Est. B2	0,124	0,213	0,149	0,215	0,238	0,344	0,208	0,104
Est. B3	0,100	0,099	0,131	0,084	0,197	0,052	0,162	0,185
Est. B4	0,157	0,141	0,094	0,101	0,135	0,198	0,142	0,136
Est. B5	0,181	0,118	0,359	0,106	0,116	0,271	0,156	0,199
Est. C	0,026	0,050	0,033	0,087	0,036	0,000	0,017	0,194

APÊNDICE 5 - ESTIMAÇÕES REFERENTES AO CAPÍTULO 2

Tabela 2.24 - Probabilidade de coautoria: ciências agrárias

Variáveis	1971-2009	2000-2006		2007-2008	
C&T Alimento	0,003	0,012	0,014	-0,020	-0,016
	(0,005)	(0,012)	(0,012)	(0,033)	(0,022)
Eng. Agrícola	-0,062***	0,012	0,012	0,012	0,012
	(0,011)	(0,012)	(0,014)	(0,008)	(0,009)
Veterinária	0,005	0,009	0,009	0,013	0,009
	(0,004)	(0,011)	(0,012)	(0,009)	(0,006)
Eng. Florestal	-0,087***	-0,092	-0,095	0,014	0,019
	(0,012)	(0,091)	(0,094)	(0,010)	(0,014)
Eng. Pesca	-0,010	0,007	0,007	0,009	0,001
	(0,012)	(0,010)	(0,015)	(0,007)	(0,002)
Zootecnia	0,008*	0,015	0,015	0,013	0,013*
	(0,004)	(0,013)	(0,014)	(0,009)	(0,008)
Não-teórico	-0,024**	-0,107	-0,105	-0,001	0,008
	(0,011)	(0,099)	(0,098)	(0,003)	(0,014)
AI			-0,020		
			(0,021)		
AN			-0,018		
			(0,021)		
BN			-0,003		
			(0,027)		
A1					0,068
					(0,065)
A2					0,070
					(0,065)
B1					0,065
					(0,062)
B2					0,071
					(0,068)
B3					0,066
					(0,066)
B4					0,066
					(0,064)
B5					0,046
					(0,067)
Constante	0,999***	0,983***	1,000***	0,981***	0,920***
	(0,003)	(0,010)	(0,018)	(0,012)	(0,067)
Observações	20,718	365	365	540	540
R^2	0,043	0,081	0,083	0,021	0,050

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.

Variáveis dummy de ano omitidas da tabela.

Tabela 2.25 - Probabilidade de coautoria: ciências biológicas

Variáveis	1971-2009	2000-2006		2007-2008	
Biologia geral	-0,040***	-0,002	-0,001	0,000	0,005
	(0,010)	(0,162)	(0,164)	(0,003)	(0,010)
Bioquímica	0,007			-0,001	0,003
	(0,006)			(0,003)	(0,007)
Botânica	-0,087***	0,190	0,191	-0,062	-0,049
	(0,010)	(0,117)	(0,118)	(0,042)	(0,045)
Ecologia	-0,078***	0,186	0,179	-0,065**	-0,059**
	(0,008)	(0,118)	(0,117)	(0,029)	(0,029)
Farmacologia	-0,004			-0,000	0,002
	(0,006)			(0,003)	(0,005)
Fisiologia	-0,008			-0,020	-0,016
	(0,006)			(0,020)	(0,020)
Genética	-0,007	0,182	0,183	-0,021	-0,017
	(0,006)	(0,117)	(0,118)	(0,020)	(0,020)
Imunologia	0,010			-0,002	0,004
	(0,006)			(0,005)	(0,008)
Microbiologia	-0,006			-0,005	0,003
	(0,006)			(0,009)	(0,010)
Morfologia	0,013**			0,000	0,007
	(0,006)			(0,003)	(0,008)
Parasitologia	0,003			-0,002	0,004
	(0,007)			(0,003)	(0,008)
Zoologia	-0,147***			-0,043*	-0,035
	(0,008)			(0,024)	(0,025)
Não-teórico	0,007	0,028	0,027	0,011	0,012
	(0,004)	(0,020)	(0,020)	(0,018)	(0,020)
AI			-0,021		
			(0,024)		
BN			-0,004		
			(0,024)		
A1					0,020*
					(0,011)
A2					0,018*
					(0,009)
B1					0,003
					(0,016)
B2					-0,004
					(0,014)
B3					-0,002
					(0,024)
B4					0,004
					(0,015)
B5					-0,010
					(0,022)
Constante	0,818***	0,790***	0,809***	0,994***	0,985***
	(0,034)	(0,124)	(0,122)	(0,006)	(0,012)
Observações	40,746	220	220	583	583
R^2	0,058	0,156	0,157	0,033	0,036

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.26 - Probabilidade de coautoria: engenharias

Variáveis	1971-2009	2000-2006		2007-2008	
Eng. Biomédica	0,026			-0,005	-0,015
	(0,027)			(0,011)	(0,022)
Eng. Civil	-0,073***	-0,159	-0,172	-0,076**	-0,070*
	(0,027)	(0,106)	(0,143)	(0,036)	(0,040)
Eng. Minas	-0,082**			-0,014	0,005
	(0,033)			(0,020)	(0,048)
Eng. Produção	-0,115***			-0,094**	-0,090*
	(0,028)			(0,042)	(0,049)
Eng. Transportes	-0,022	-0,177	-0,198	-0,001	0,008
	(0,037)	(0,164)	(0,216)	(0,007)	(0,028)
Eng. Elétrica	0,011	0,072	0,066	-0,029	-0,035
	(0,026)	(0,046)	(0,064)	(0,028)	(0,032)
Eng. Mat & Metal	0,036	0,028	0,029	-0,005	-0,006
	(0,026)	(0,026)	(0,027)	(0,009)	(0,016)
Eng. Mecânica	-0,000	0,034	0,013	-0,049	-0,074
	(0,027)	(0,031)	(0,140)	(0,044)	(0,066)
Eng. Naval e Oceânica	0,008	0,073	0,054		
	(0,036)	(0,061)	(0,156)		
Eng. Nuclear	0,094***			-0,006	-0,017
	(0,029)			(0,012)	(0,024)
Eng. Química	0,025	0,008	0,009	-0,045	-0,045
	(0,026)	(0,027)	(0,029)	(0,042)	(0,049)
Eng. Sanitária	-0,065**	-0,022	-0,043	-0,007	0,009
	(0,028)	(0,045)	(0,142)	(0,010)	(0,017)
Não-teórico	0,057***	0,022	0,022	0,019	0,025
	(0,015)	(0,063)	(0,065)	(0,015)	(0,019)
AI			-0,022		
			(0,142)		
A1					-0,053
					(0,057)
A2					-0,046
					(0,059)
B1					-0,048
					(0,054)
B2					-0,117*
					(0,069)
B3					-0,069
					(0,061)
B4					-0,092
					(0,064)
B5					-0,079
					(0,063)
Constante	0,760***	0,717***	0,737***	0,997***	1,070***
	(0,081)	(0,134)	(0,202)	(0,004)	(0,057)
Observações	11,884	116	116	329	329
R^2	0,106	0,190	0,190	0,036	0,051

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.27 - Probabilidade de coautoria: ciências exatas e da terra

Variáveis	1971-2009	2000-2006		2007-2008	
Computação	-0,026**			0,331	0,323
	(0,012)			(0,228)	(0,209)
Física	0,035***	-0,026	-0,031	0,372*	0,394*
	(0,009)	(0,034)	(0,036)	(0,225)	(0,209)
Geociências	-0,021**	-0,043	0,015	0,264	0,302
	(0,010)	(0,036)	(0,024)	(0,228)	(0,213)
Matemática	-0,175***			0,180	0,209
	(0,014)			(0,236)	(0,222)
Oceanografia	-0,017	-0,018	-0,013	0,315	0,373*
	(0,013)	(0,046)	(0,050)	(0,238)	(0,216)
Prob.e Estat.	-0,046***			0,257	0,302
	(0,013)			(0,236)	(0,229)
Química	0,049***	-0,007	0,002	0,364	0,374*
	(0,009)	(0,027)	(0,025)	(0,224)	(0,206)
Não-teórico	0,045***	0,050*	0,012	0,156***	0,173***
	(0,010)	(0,030)	(0,024)	(0,050)	(0,057)
AI			0,230*		
			(0,118)		
BI			0,073		
			(0,198)		
CI			0,274**		
			(0,122)		
AN			0,205*		
			(0,122)		
BN			0,255**		
			(0,125)		
A1					0,035
					(0,058)
A2					-0,071
					(0,076)
B1					-0,007
					(0,040)
B2					0,005
					(0,047)
B3					-0,037
					(0,061)
B4					-0,076
					(0,062)
B5					-0,148*
					(0,082)
Constante	0,749***	0,961***	0,724***	0,598***	0,607***
	(0,043)	(0,041)	(0,125)	(0,223)	(0,208)
Observações	31,532	301	301	367	367
R^2	0,062	0,013	0,072	0,075	0,108

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.28 - Probabilidade de coautoria: ciências humanas

Variáveis	1971-2009	2000-2006		2007-2008	
Arqueologia	0,124***			0,425**	0,459**
	(0,024)			(0,170)	(0,183)
Ciência Política	0,015			-0,017	-0,077
	(0,019)			(0,016)	(0,072)
Educação	0,166***	0,779***	0,602***	0,570***	0,537***
	(0,015)	(0,136)	(0,166)	(0,043)	(0,072)
Filosofia	-0,073***	0,110	0,146	0,329***	0,307***
	(0,019)	(0,098)	(0,105)	(0,091)	(0,103)
Geografia	0,097***			0,462***	0,397***
	(0,022)			(0,085)	(0,106)
História	-0,094***			0,132**	0,114
	(0,016)			(0,062)	(0,084)
Psicologia	0,283***	0,857***	0,833***	0,687***	0,673***
	(0,015)	(0,110)	(0,113)	(0,052)	(0,070)
Sociologia	-0,012			0,400***	0,369**
	(0,017)			(0,131)	(0,148)
Teologia	-0,211***				
	(0,042)				
Não-teórico	0,178***	0,232**	0,045	0,159	0,107
	(0,019)	(0,114)	(0,090)	(0,123)	(0,121)
AI			0,020		
			(0,184)		
AN			-0,254		
			(0,199)		
A1					-0,241
					(0,186)
A2					-0,063
					(0,127)
B1					-0,208*
					(0,124)
B2					-0,255**
					(0,118)
B3					-0,200*
					(0,120)
B4					-0,141
					(0,122)
B5					-0,155
					(0,131)
Constante	0,311***	-0,254	-0,038	-0,000	0,208*
	(0,077)	(0,170)	(0,253)	(0,000)	(0,124)
Observações	17,766	101	101	421	421
R^2	0,133	0,281	0,316	0,146	0,163

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.29 - Probabilidade de coautoria: linguística, letras e artes

Variáveis	1971-2009	2000-2006		2007-2008	
Letras	0,011	-0,061	-0,052	-0,139	0,114
	(0,013)	(0,063)	(0,071)	(0,110)	(0,151)
Linguística	0,121***			0,375**	0,548***
	(0,017)			(0,148)	(0,189)
Não-teórico	-0,077***	-0,023	0,089	-0,170	-0,112
	(0,024)	(0,029)	(0,082)	(0,382)	(0,473)
CI			0,104		
			(0,328)		
AN			-0,080		
			(0,133)		
BN			-0,161		
			(0,114)		
A2					0,368
					(0,292)
B1					0,074
					(0,307)
B2					-0,176
					(0,238)
B4					-0,188
					(0,240)
B5					0,054
					(0,254)
B3					
Constante	-0,007	0,052	0,105	0,295**	0,157
	(0,009)	(0,055)	(0,080)	(0,118)	(0,232)
Observações	4,919	53	53	96	96
R^2	0,041	0,151	0,241	0,209	0,303

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.30 - Probabilidade de coautoria: ciências da saúde

Variáveis	1971-2009	2000-2006		2007-2008	
Enfermagem	0,010	-0,067	-0,022	0,053**	0,044*
	(0,010)	(0,078)	(0,087)	(0,025)	(0,023)
Farmácia	0,109***	0,131***	0,083*	0,068***	0,053**
	(0,010)	(0,048)	(0,050)	(0,023)	(0,022)
Fisio e TO	0,056***	0,023	-0,026	0,060**	0,043*
	(0,011)	(0,059)	(0,051)	(0,025)	(0,023)
Fonoaudiologia	0,056***	0,117**	0,101	0,069***	0,045**
	(0,012)	(0,049)	(0,065)	(0,024)	(0,021)
Medicina	0,093***			0,051*	0,050*
	(0,009)			(0,029)	(0,029)
Nutrição	0,073***	0,146***	0,092	0,052*	0,063**
	(0,011)	(0,056)	(0,059)	(0,027)	(0,028)
Odontologia	0,111***	0,077	0,070	0,044*	0,044*
	(0,009)	(0,051)	(0,052)	(0,025)	(0,024)
Saúde Coletiva	0,006	-0,014	-0,068	-0,018	-0,021
	(0,010)	(0,081)	(0,081)	(0,032)	(0,031)
Não-teórico	0,038***			0,036***	0,051***
	(0,007)			(0,014)	(0,020)
AI			0,073***		
			(0,025)		
BI			-0,065		
			(0,092)		
CI			0,000		
			(0,124)		
AN			0,088***		
			(0,033)		
BN			0,074**		
			(0,029)		
A1					0,014
					(0,044)
A2					0,017
					(0,044)
B1					0,015
					(0,044)
B2					-0,002
					(0,045)
B3					-0,025
					(0,046)
B4					-0,066
					(0,049)
B5					-0,031
					(0,047)
Constante	0,718***	0,828***	0,814***	0,929***	0,948***
	(0,034)	(0,071)	(0,074)	(0,023)	(0,047)
Observações	80,938	370	370	1,300	1,300
R^2	0,034	0,048	0,061	0,024	0,048

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.31 - Probabilidade de coautoria: ciências sociais aplicadas

Variáveis	1971-2009	2000-2006		2007-2008	
Arquit. e Urban.	-0,281***			-0,463***	-0,457***
	(0,015)			(0,089)	(0,088)
C. Informação	-0,093***	0,139	0,142	-0,167	-0,184
	(0,022)	(0,094)	(0,097)	(0,144)	(0,133)
Comunicação	-0,415***			-0,153	-0,212**
	(0,014)			(0,101)	(0,095)
Demografia	-0,055				
	(0,050)				
Des. Industrial	-0,408***			0,092***	-0,013
	(0,059)			(0,025)	(0,038)
Direito	-0,433***			-0,617***	-0,606***
	(0,012)			(0,064)	(0,089)
Econ. Doméstica	0,057				
	(0,198)				
Economia	-0,062***			-0,068	-0,091
	(0,014)			(0,071)	(0,075)
Museologia	-0,208***				
	(0,058)				
Plan. Urban. e Reg.	-0,139***			-0,325	-0,359
	(0,027)			(0,234)	(0,221)
Serviço Social	-0,419***			0,092***	0,041
	(0,029)			(0,025)	(0,043)
Turismo	-0,347***			-0,408	-0,386
	(0,031)			(0,361)	(0,372)
Não-teórico	0,171***	0,229*	0,210*	-0,341	-0,269
	(0,028)	(0,133)	(0,115)	(0,367)	(0,360)
AN			-0,066		
			(0,125)		
BN			-0,533		
			(0,429)		
A1					0,123*
					(0,073)
A2					-0,227
					(0,147)
B1					0,018
					(0,090)
B2					-0,036
					(0,089)
B3					0,000
					(0,087)
B4					0,136
					(0,083)
B5					0,082
					(0,087)
Constante	0,333***	1,000***	1,066***	0,908***	0,877***
	(0,096)	(0,000)	(0,125)	(0,025)	(0,073)
Observações	11,033	52	52	346	346
R^2	0,235	0,293	0,329	0,354	0,381

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.32 - Número de autores: ciências agrárias

Variáveis	1971-2009	2000-2006		2007-2008	
C&T Alimento	-0,286***	0,024	0,120	-0,467	-0,408
	(0,038)	(0,852)	(0,857)	(0,307)	(0,329)
Eng. Agrícola	-0,379***	-0,339	-0,371	-0,285	-0,219
	(0,055)	(0,230)	(0,247)	(0,324)	(0,324)
Veterinária	1,042***	0,892***	1,027***	1,726***	1,759***
	(0,035)	(0,231)	(0,234)	(0,258)	(0,286)
Eng. Florestal	-0,564***	-0,204	0,120	1,159	1,288
	(0,062)	(0,646)	(0,721)	(0,768)	(0,791)
Eng. Pesca	-0,106	0,943***	0,879***	0,549	0,601
	(0,096)	(0,244)	(0,271)	(0,567)	(0,606)
Zootecnia	0,765***	0,723***	0,761***	0,872***	0,814***
	(0,054)	(0,224)	(0,235)	(0,197)	(0,242)
Não-teórico	0,182**	-0,039	-0,085	-1,379*	-1,414*
	(0,080)	(0,684)	(0,711)	(0,753)	(0,729)
AI			1,148*		
			(0,665)		
AN			1,306*		
			(0,670)		
BN			1,604**		
			(0,727)		
A1					2,941***
					(0,798)
A2					0,601
					(0,672)
B1					0,700
					(0,533)
B2					0,656
					(0,581)
B3					0,211
					(0,577)
B4					0,044
					(0,584)
B5					0,128
					(0,594)
Constante	2,395***	4,116***	2,896***	4,618***	4,129***
	(0,164)	(0,209)	(0,673)	(0,188)	(0,568)
Observações	20,718	365	365	540	540
R^2	0,171	0,121	0,141	0,118	0,165

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.33 - Número de autores: ciências biológicas

Variáveis	1971-2009	2000-2006		2007-2008	
Biologia Geral	-0,722***	0,300	0,354	-0,691	-0,502
	(0,139)	(1,266)	(1,256)	(1,127)	(1,142)
Bioquímica	0,386***			2,467**	2,504**
	(0,113)			(1,046)	(1,009)
Botânica	-1,709***	1,051	1,125	-1,545**	-1,173
	(0,109)	(1,035)	(1,026)	(0,730)	(0,799)
Ecologia	-1,376***	-1,198	-1,694**	-1,505**	-1,337*
	(0,108)	(0,803)	(0,755)	(0,715)	(0,736)
Farmacologia	-0,127			1,042	0,915
	(0,104)			(0,727)	(0,761)
Fisiologia	-0,432***			-0,118	-0,056
	(0,104)			(0,712)	(0,740)
Genética	-0,049	0,952	0,954	0,821	0,876
	(0,108)	(0,733)	(0,709)	(0,785)	(0,820)
Imunologia	0,629***			2,496***	2,651***
	(0,108)			(0,780)	(0,824)
Microbiologia	0,067			0,121	0,297
	(0,109)			(0,714)	(0,770)
Morfologia	-0,306***			0,284	0,603
	(0,103)			(0,787)	(0,820)
Parasitologia	0,441***			1,355*	1,583*
	(0,132)			(0,803)	(0,862)
Zoologia	-2,102***			-2,094***	-1,772**
	(0,101)			(0,664)	(0,729)
Não-teórico	-0,062	0,907	0,844	-0,003	0,131
	(0,039)	(0,705)	(0,733)	(0,325)	(0,327)
AI			-1,964*		
			(1,094)		
BN			-1,119		
			(1,166)		
A1					2,071**
					(0,961)
A2					0,639
					(0,603)
B1					0,092
					(0,538)
B2					0,186
					(0,534)
B3					0,633
					(0,590)
B4					-0,159
					(0,546)
B5					-0,398
					(0,546)
Constante	3,156***	3,394***	5,213***	5,334***	4,883***
	(0,148)	(0,746)	(1,272)	(0,645)	(0,853)
Observações	40,746	220	220	583	583
R^2	0,136	0,122	0,137	0,212	0,244

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.34 - Número de autores: engenharias

Variáveis	1971-2009	2000-2006		2007-2008	
Eng. Biomédica	0,364			-0,102	-0,593
	(0,530)			(0,990)	(0,944)
Eng. Civil	-1,114**	0,893	0,175	-1,837***	-1,974***
	(0,527)	(0,557)	(0,762)	(0,651)	(0,685)
Eng. Minas	-1,000*			-2,466***	-2,887***
	(0,530)			(0,653)	(0,721)
Eng. Produção	-1,323**			-1,921***	-1,782***
	(0,526)			(0,650)	(0,679)
Eng. Transportes	-1,039*	0,693	-0,450	-1,733**	-1,770**
	(0,533)	(0,565)	(0,895)	(0,675)	(0,764)
Eng. Elétrica	-0,227	2,275**	1,970***	-1,049	-1,229*
	(0,526)	(0,873)	(0,721)	(0,682)	(0,695)
Eng. Mat & Metal	0,009	2,415***	2,485***	-0,732	-0,989
	(0,526)	(0,605)	(0,608)	(0,669)	(0,673)
Eng. Mecânica	-0,572	0,251	-0,887	-1,207	-0,687
	(0,527)	(0,363)	(0,773)	(0,736)	(0,825)
Eng. Naval e Oceânica	-0,707	0,318	-0,740		
	(0,539)	(0,647)	(0,882)		
Eng. Nuclear	0,777			0,667	0,372
	(0,540)			(0,856)	(0,832)
Eng. Química	-0,064	1,208**	1,248**	-0,158	-0,625
	(0,526)	(0,480)	(0,482)	(0,711)	(0,729)
Eng. Sanitária	-0,570	1,450***	0,321	-0,817	-0,771
	(0,528)	(0,538)	(0,847)	(0,664)	(0,687)
Não-teórico	0,268***	-0,408	-0,402	0,069	-0,023
	(0,102)	(0,334)	(0,368)	(0,516)	(0,447)
AI			-1,212		
			(0,733)		
A1					2,240***
					(0,503)
A2					2,320***
					(0,675)
B1					1,409**
					(0,592)
B2					1,142**
					(0,461)
B3					1,371***
					(0,471)
B4					1,783***
					(0,428)
B5					1,208***
					(0,408)
Constante	2,521***	0,816	1,951*	4,489***	3,190***
	(0,554)	(0,739)	(0,986)	(0,636)	(0,799)
Observações	11,884	116	116	329	329
R^2	0,158	0,262	0,287	0,201	0,277

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.35 - Número de autores: ciências exatas e da terra

Variáveis	1971-2009	2000-2006		2007-2008	
Computação	-0,694***			-3,037	-3,758**
	(0,103)			(1,963)	(1,761)
Física	0,602***	-0,360	-0,465	-1,519	-1,870
	(0,089)	(0,435)	(0,430)	(1,969)	(1,769)
Geociências	-0,159*	-0,464	0,841	-2,239	-2,372
	(0,092)	(0,456)	(0,680)	(1,968)	(1,766)
Matemática	-1,640***			-3,532*	-3,788**
	(0,090)			(1,969)	(1,771)
Oceanografia	-0,307***	-0,633	-0,541	-1,753	-1,753
	(0,113)	(0,521)	(0,535)	(2,024)	(1,802)
Prob.e Estat.	-0,848***			-3,188	-3,264*
	(0,099)			(1,967)	(1,772)
Química	0,224***	0,085	0,074	-1,005	-1,470
	(0,087)	(0,443)	(0,442)	(1,964)	(1,742)
Não-teórico	0,645***	0,097	0,087	0,302	0,406
	(0,068)	(0,638)	(0,520)	(0,290)	(0,329)
AI			2,537***		
			(0,659)		
BI			2,771*		
			(1,548)		
CI			3,514***		
			(1,288)		
AN			0,972**		
			(0,436)		
BN			1,204		
			(0,854)		
A1					-0,723
					(0,764)
A2					-0,515
					(0,801)
B1					-0,946
					(0,688)
B2					-0,757
					(0,719)
B3					-0,602
					(0,764)
B4					-2,006***
					(0,712)
B5					-1,377*
					(0,767)
Constante	2,701***	4,490***	1,889***	5,995***	7,216***
	(0,227)	(0,497)	(0,724)	(1,947)	(1,853)
Observações	31,532	301	301	367	367
R^2	0,081	0,050	0,101	0,140	0,175

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.36 - Número de autores: ciências humanas

Variáveis	1971-2009	2000-2006		2007-2008	
Arqueologia	0,312***			1,112**	1,241**
	(0,071)			(0,492)	(0,557)
Ciência Política	-0,139***			0,001	-0,457
	(0,046)			(0,046)	(0,358)
Educação	0,326***	1,159***	0,715*	1,030***	0,867***
	(0,044)	(0,301)	(0,370)	(0,113)	(0,238)
Filosofia	-0,286***	0,087	0,169	0,388***	0,256
	(0,047)	(0,195)	(0,194)	(0,120)	(0,206)
Geografia	0,137**			1,325**	0,991**
	(0,067)			(0,520)	(0,444)
História	-0,308***			0,180*	0,104
	(0,041)			(0,094)	(0,250)
Psicologia	0,615***	1,238***	1,198***	1,668***	1,620***
	(0,045)	(0,218)	(0,217)	(0,209)	(0,327)
Sociologia	-0,171***			0,787***	0,687*
	(0,042)			(0,287)	(0,382)
Teologia	-0,526***				
	(0,070)				
Não-teórico	0,598***	-0,232	-0,610	1,333	1,187
	(0,078)	(0,374)	(0,512)	(0,998)	(1,004)
AI			-0,191		
			(0,487)		
AN			-0,752		
			(0,490)		
A1					-1,957*
					(1,172)
A2					-1,660
					(1,059)
B1					-1,842*
					(1,042)
B2					-2,036**
					(1,033)
B3					-1,656
					(1,024)
B4					-1,610
					(1,061)
B5					-1,752
					(1,065)
Constante	1,425***	0,819**	1,496**	1,000***	2,842***
	(0,136)	(0,386)	(0,729)	(0,000)	(1,042)
Observações	17,766	101	101	421	421
R^2	0,105	0,144	0,185	0,111	0,156

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.37 - Número de autores: linguística, letras e artes

Variáveis	1971-2009	2000-2006		2007-2008	
Letras	0,043	-0,061	-0,052	-0,173	0,249
	(0,027)	(0,063)	(0,071)	(0,187)	(0,232)
Linguística	0,212***			0,541**	0,946***
	(0,033)			(0,269)	(0,325)
Não-teórico	-0,153***	-0,023	0,089	-0,476	-0,402
	(0,044)	(0,029)	(0,082)	(0,440)	(0,491)
CI			0,104		
			(0,328)		
AN			-0,080		
			(0,133)		
BN			-0,161		
			(0,114)		
A2					0,805*
					(0,438)
B1					0,470
					(0,482)
B2					0,035
					(0,279)
B4					-0,092
					(0,249)
B5					0,061
					(0,272)
Constante	0,971***	1,052***	1,105***	1,435***	1,003***
	(0,022)	(0,055)	(0,080)	(0,204)	(0,256)
Observações	4,919	53	53	96	96
R^2	0,022	0,151	0,241	0,141	0,196

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.38 - Número de autores: ciências da saúde

Variáveis	1971-2009	2000-2006		2007-2008	
Enfermagem	-0,780***	-1,558***	-1,284***	-0,263	-0,363*
	(0,064)	(0,405)	(0,457)	(0,214)	(0,209)
Farmácia	0,712***	1,120**	0,413	2,546***	2,238***
	(0,070)	(0,479)	(0,503)	(0,472)	(0,470)
Fisio e TO	0,012	-0,245	-0,457	0,509**	0,376*
	(0,079)	(0,458)	(0,514)	(0,228)	(0,219)
Fonoaudiologia	-0,413***	-0,022	-0,269	-0,045	-0,136
	(0,080)	(0,726)	(0,706)	(0,269)	(0,287)
Medicina	1,329***			1,868***	1,862***
	(0,062)			(0,457)	(0,403)
Nutrição	0,143**	1,877**	1,430*	0,253	0,530*
	(0,073)	(0,751)	(0,837)	(0,269)	(0,275)
Odontologia	0,291***	0,328	0,247	0,921***	0,860***
	(0,061)	(0,361)	(0,355)	(0,188)	(0,181)
Saúde Coletiva	0,193***	0,356	-0,140	0,720**	0,649**
	(0,068)	(0,598)	(0,646)	(0,286)	(0,276)
Não-teórico	0,343***			0,021	0,222
	(0,065)			(0,451)	(0,374)
AI			0,613**		
			(0,306)		
BI			-0,120		
			(0,420)		
CI			-0,264		
			(0,606)		
AN			1,833***		
			(0,609)		
BN			0,316		
			(0,338)		
A1					1,696***
					(0,591)
A2					1,987***
					(0,423)
B1					0,578
					(0,389)
B2					0,632
					(0,393)
B3					0,052
					(0,406)
B4					-0,035
					(0,392)
B5					-0,153
					(0,394)
Constante	2,315***	3,721***	3,540***	3,930***	3,555***
	(0,205)	(0,424)	(0,429)	(0,179)	(0,399)
Observações	80,938	370	370	1,300	1,300
R^2	0,098	0,145	0,167	0,091	0,174

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.39 - Número de autores: ciências sociais aplicadas

Variáveis	1971-2009	2000-2006		2007-2008	
Arquit. e Urban.	-0,455***			-1,167***	-1,086***
	(0,037)			(0,176)	(0,179)
C. Informação	0,019	-0,284	-0,297	-0,880***	-0,968***
	(0,061)	(0,213)	(0,219)	(0,296)	(0,322)
Comunicação	-0,750***			-0,855***	-1,036***
	(0,029)			(0,203)	(0,178)
Demografia	-0,225***				
	(0,082)				
Des. Industrial	-0,742***			0,188**	-0,215
	(0,106)			(0,095)	(0,158)
Direito	-0,716***			-1,378***	-1,320***
	(0,029)			(0,158)	(0,332)
Econ. Doméstica	0,302				
	(0,508)				
Economia	-0,181***			-0,126	-0,128
	(0,029)			(0,255)	(0,262)
Museologia	-0,247				
	(0,190)				
Plan. Urban. e Reg.	-0,225***			-0,242	-0,382
	(0,065)			(0,702)	(0,651)
Serviço Social	-0,580***			-0,812***	-0,855***
	(0,073)			(0,095)	(0,164)
Turismo	-0,655***			-1,312***	-1,285***
	(0,061)			(0,372)	(0,339)
Não-teórico	0,349***	0,463*	0,547**	-1,185***	-1,033***
	(0,063)	(0,272)	(0,264)	(0,434)	(0,336)
AN			0,296		
			(0,268)		
BN			-0,352		
			(0,311)		
A1					-0,716**
					(0,311)
A2					-0,557
					(0,392)
B1					0,002
					(0,346)
B2					-0,247
					(0,342)
B3					0,138
					(0,352)
B4					0,499
					(0,349)
B5					0,139
					(0,342)
Constante	1,602***	2,000***	1,704***	2,812***	2,716***
	(0,229)	(0,000)	(0,268)	(0,095)	(0,311)
Observações	11,033	52	52	346	346
R^2	0,154	0,278	0,299	0,242	0,276

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano omitidas da tabela.

Tabela 2.40.1 - Probabilidade de coautoria ao nível do autor: *pooled cross-section*

Variáveis	Agrárias		Biológicas		Engenharias		Exatas	
Sexo	-0,019***	-0,051***	-0,014***	-0,010	-0,026***	-0,055**	-0,027***	-0,034**
	(0,005)	(0,014)	(0,004)	(0,013)	(0,008)	(0,025)	(0,004)	(0,014)
Ordem	0,001***	0,000	0,000***	0,001**	0,001***	0,003*	0,001***	0,002***
	(0,000)	(0,001)	(0,000)	(0,001)	(0,000)	(0,002)	(0,000)	(0,001)
Ordem^2	-0,000***	-0,000	-0,000***	-0,000**	-0,000***	-0,000*	-0,000***	-0,000**
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Idade	-0,003***	-0,008**	-0,001	-0,006*	-0,006***	-0,007	0,001	-0,005*
	(0,001)	(0,003)	(0,001)	(0,003)	(0,001)	(0,005)	(0,001)	(0,003)
Idade^2	0,000**	0,000**	0,000	0,000***	0,000***	0,000	-0,000**	0,000
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Bolsa	-0,004	-0,001	-0,004	-0,015	0,010	0,036*	0,002	0,039**
	(0,004)	(0,014)	(0,005)	(0,011)	(0,008)	(0,022)	(0,004)	(0,017)
$\Delta AResp$	-0,007	-0,035*	0,006	-0,008	0,007	-0,000	-0,002	0,030**
	(0,006)	(0,018)	(0,005)	(0,016)	(0,008)	(0,022)	(0,005)	(0,014)
$\Delta ATgr$	0,012	0,023	-0,008	-0,027	-0,024	-0,104***	0,003	-0,014
	(0,011)	(0,040)	(0,008)	(0,023)	(0,016)	(0,040)	(0,011)	(0,024)
$\Delta ATar$	0,003	-0,004	0,004	-0,009	0,015	0,033	0,006	0,028*
	(0,008)	(0,025)	(0,005)	(0,015)	(0,012)	(0,024)	(0,008)	(0,015)
$\Delta ATsub$	-0,013*	-0,042	-0,004	-0,003	0,010	0,059**	-0,002	-0,030
	(0,008)	(0,027)	(0,006)	(0,018)	(0,011)	(0,030)	(0,007)	(0,022)
$\Delta ATesp$	0,002	0,011	0,005	0,007	0,002	-0,024	0,009***	-0,011
	(0,003)	(0,010)	(0,004)	(0,012)	(0,006)	(0,016)	(0,003)	(0,011)
Carga		0,003***		-0,001**		0,001**		0,000
		(0,001)		(0,000)		(0,001)		(0,001)
Exclusiva		-0,004		-0,010		0,029*		0,024
		(0,018)		(0,013)		(0,018)		(0,015)
Titular		0,028		-0,291***		-0,225*		-0,002
		(0,030)		(0,057)		(0,119)		(0,037)
Constante	0,996***	0,925***	0,853***	1,118***	0,649***	1,271***	0,917***	0,013
	(0,062)	(0,118)	(0,083)	(0,057)	(0,151)	(0,118)	(0,047)	(0,028)
Observações	10.819	1.375	10.590	1.703	6.611	1.121	17.099	1.928
R^2	0,038	0,163	0,078	0,215	0,097	0,181	0,064	0,121

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano e de áreas omitidas da tabela.

Tabela 2.40.2 - Probabilidade de coautoria ao nível do autor: *pooled cross-section*

Variáveis	Humanas		Linguística		Saúde		Sociais	
Sexo	-0,056***	-0,061*	-0,079***	-0,001	-0,019***	-0,009	0,008	0,005
	(0,014)	(0,033)	(0,028)	(0,094)	(0,004)	(0,009)	(0,018)	(0,046)
Ordem	-0,003***	-0,002	-0,006**	-0,015**	0,000***	0,000***	-0,005***	-0,009**
	(0,001)	(0,002)	(0,003)	(0,007)	(0,000)	(0,000)	(0,001)	(0,004)
Ordem^2	0,000***	0,000*	0,000**	0,000***	-0,000***	-0,000**	0,000***	0,000
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Idade	0,006**	-0,014*	-0,020***	0,008	-0,005***	-0,009***	-0,000	-0,007
	(0,003)	(0,008)	(0,007)	(0,022)	(0,001)	(0,002)	(0,003)	(0,010)
Idade^2	-0,000**	0,000**	0,000***	-0,000	0,000***	0,000***	0,000	0,000
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
Bolsa	-0,008	0,016	-0,105***	0,020	0,019***	0,011	0,033*	-0,106**
	(0,014)	(0,031)	(0,026)	(0,078)	(0,004)	(0,009)	(0,018)	(0,044)
$\Delta AResp$	0,011	0,051	-0,059**	-0,056	0,015***	0,033**	-0,031**	-0,068*
	(0,014)	(0,032)	(0,024)	(0,047)	(0,006)	(0,014)	(0,015)	(0,035)
$\Delta ATgr$	0,284***	0,100	0,075	0,068	0,010	0,039	-0,002	-0,173
	(0,032)	(0,068)	(0,073)	(0,161)	(0,010)	(0,027)	(0,046)	(0,109)
$\Delta ATar$	-0,065***	0,021	-0,079*	-0,052	0,010	-0,023	-0,020	0,062
	(0,021)	(0,047)	(0,042)	(0,161)	(0,007)	(0,018)	(0,032)	(0,083)
$\Delta ATsub$	0,039**	-0,017	0,085***	0,031	0,015***	-0,044***	0,014	0,141**
	(0,019)	(0,039)	(0,032)	(0,140)	(0,006)	(0,017)	(0,023)	(0,063)
$\Delta ATesp$	0,009	0,053**	-0,005	-0,057	-0,022***	-0,020**	0,044***	0,029
	(0,011)	(0,023)	(0,018)	(0,046)	(0,004)	(0,008)	(0,012)	(0,026)
Carga		-0,000		0,002		-0,000		0,001
		(0,001)		(0,002)		(0,000)		(0,002)
Exclusiva		0,098***		0,150*		0,021*		-0,124**
		(0,036)		(0,083)		(0,012)		(0,056)
Titular		-0,123**		0,240		0,025*		-0,053
		(0,053)		(0,161)		(0,014)		(0,055)
Constante	0,597***	0,120	0,380***	0,163	0,776***	0,040	0,645***	0,647***
	(0,062)	(0,136)	(0,100)	(0,404)	(0,061)	(0,051)	(0,237)	(0,137)
Observações	4.853	1.237	1.158	227	27.906	5.289	3.581	729
R^2	0,139	0,218	0,108	0,353	0,032	0,082	0,293	0,280

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.
Variáveis dummy de ano e de áreas omitidas da tabela.

Tabela 2.41 - Probabilidade de coautoria ao nível do autor: efeitos fixos

Variáveis	Agrárias	Biológicas	Engenharias	Exatas	Humanas	Linguística	Saúde	Sociais
Idade	-0,004	0,002	-0,002	0,003	0,015*	-0,002	-0,000	0,004
	(0,003)	(0,004)	(0,007)	(0,004)	(0,008)	(0,014)	(0,003)	(0,008)
Idade^2	0,000	0,000	0,000	-0,000	-0,000*	0,000	0,000	0,000
	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)	(0,000)
$\Delta AResp$	-0,007	0,000	0,010	-0,001	-0,015	-0,050**	0,008	0,002
	(0,005)	(0,006)	(0,008)	(0,005)	(0,013)	(0,025)	(0,006)	(0,014)
$\Delta ATsub$	-0,010	0,009	0,014	-0,001	0,022	0,041	0,006	-0,004
	(0,009)	(0,010)	(0,014)	(0,008)	(0,031)	(0,036)	(0,008)	(0,034)
$\Delta ATesp$	0,001	0,006	-0,001	0,007	0,004	-0,041	0,001	0,010
	(0,005)	(0,007)	(0,008)	(0,005)	(0,017)	(0,036)	(0,008)	(0,017)
Constante	1,042***	0,902***	0,144	0,884***	0,966***	0,932***	0,872***	0,939***
	(0,056)	(0,115)	(0,170)	(0,103)	(0,201)	(0,235)	(0,074)	(0,161)
Observações	11.247	10.941	6.757	17.281	4.998	1.190	29.483	3.669
Nº autores	699	923	671	1.169	620	144	1.326	465
R^2	0,035	0,027	0,080	0,024	0,083	0,152	0,014	0,081

Erros padrões robustos em parênteses; *** p<0,01; ** p<0,05; * p<0,1.

Variáveis dummy de ano e de ordem.

APÊNDICE 6 - MESMA AMOSTRA POOLED E EFEITOS FIXOS

Tabela 4.5.1 - Regressão logística (coeficientes): *pooled cross-section*

		Agrárias		Biológicas		Exatas		Saúde	
Primeira colaboração	p_{t-1}^{ij}	3,982***	3,408***	2,500***	2,565***	3,180***	2,937***	1,972***	1,811***
		(0,292)	(0,313)	(0,739)	(0,735)	(0,394)	(0,404)	(0,190)	(0,193)
	g_{t-1}^{ij}	0,055***	0,036**	0,112**	0,120**	0,082*	0,076*	0,000	-0,001
		(0,015)	(0,015)	(0,054)	(0,054)	(0,044)	(0,045)	(0,007)	(0,007)
	$\bar{q}_t^{ij}$		0,024***		0,024**		0,042***		0,026***
			(0,009)		(0,010)		(0,010)		(0,007)
	Δq_t^{ij}		0,008		-0,013**		-0,004		-0,003
			(0,006)		(0,006)		(0,007)		(0,005)
	$\bar{n}_t^{ij}$: total		0,010***		-0,004		0,001		0,003***
			(0,002)		(0,003)		(0,002)		(0,000)
	$\bar{n}_t^{ij}$: USP		0,105		-1,409		-1,512**		-0,950***
			(0,451)		-1,662		(0,761)		(0,209)
	Δn_t^{ij}: total		-0,003**		0,001		-0,002		-0,001***
			(0,001)		(0,002)		(0,002)		(0,000)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		1,347***		0,317		0,828		0,717***
			(0,412)		(0,461)		(0,736)		(0,189)
	$\omega_{t,f}^{ij}$: área		0,397		0,503		0,984		0,967***
			(0,519)		(0,691)		(0,720)		(0,289)
	$\omega_{t,f}^{ij}$: subárea		-0,134		1,525*		0,510		0,075
			(0,383)		(0,819)		(0,513)		(0,216)
	Constante	-2,185***	-2,224***	-2,727***	-2,765***	-2,255***	-2,279***	-2,226***	-2,230***
		(0,032)	(0,034)	(0,055)	(0,058)	(0,039)	(0,039)	(0,015)	(0,015)
	Excl. ef. rede	209,72	129,11	19,41	20,74	79,04	62,92	108,31	88,11
		[0.000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]	[0,000]
	Obs.	8.663	8.663	5.155	5.155	6.470	6.470	37.303	37.303
		Agrárias		Biológicas		Exatas		Saúde	
Colaborações subsequentes	p_{t-1}^{ij}	0,627***	0,616**	0,900***	0,898***	0,806***	0,527**	0,946***	1,075***
		(0,140)	(0,242)	(0,120)	(0,285)	(0,115)	(0,212)	(0,081)	(0,157)
	g_{t-1}^{ij}	0,070***	0,094***	0,030**	0,042	0,021	0,027	0,055***	0,074***
		(0,009)	(0,015)	(0,014)	(0,043)	(0,014)	(0,028)	(0,004)	(0,007)
	$\bar{q}_t^{ij}$		0,003		-0,009		-0,006		0,003
			(0,006)		(0,006)		(0,004)		(0,002)
	Δq_t^{ij}		-0,006		0,004		0,002		-0,003**
			(0,004)		(0,004)		(0,003)		(0,002)
	$\bar{n}_t^{ij}$: total		-0,003***		-0,000		-0,002		-0,001***
			(0,001)		(0,001)		(0,001)		(0,000)
	$\bar{n}_t^{ij}$: USP		1,472***		1,864***		1,507***		7,371***
			(0,498)		(0,394)		(0,399)		(0,357)
	Δn_t^{ij}: total		0,003***		0,001		0,002***		0,001***
			(0,000)		(0,001)		(0,001)		(0,000)
	Δn_t^{ij}: USP								-3,174***
									(0,198)
	$\omega_{t,f}^{ij}$: gr. área		-0,895***		-0,459**		-0,515**		-0,672***
			(0,200)		(0,213)		(0,258)		(0,105)
	$\omega_{t,f}^{ij}$: área		-0,390*		0,702***		-0,266		0,031
			(0,199)		(0,244)		(0,314)		(0,090)
	$\omega_{t,f}^{ij}$: subárea		1,043***		0,948***		0,741***		0,681***
			(0,125)		(0,211)		(0,168)		(0,053)
	Constante	-0,341	-0,691**	-0,352	-1,365***	-0,880**	1,669	-0,537***	0,350
		(0,385)	(0,287)	(0,408)	(0,441)	(0,437)	-1,144	(0,178)	(0,572)
	Excl. ef. rede	86,51	48,12	69,24	14,67	55,03	7,97	286,17	144,03
		[0,000]	[0,000]	[0,000]	[0,001]	[0,000]	[0,019]	[0,000]	[0,000]
	Obs.	30.987	8.678	28.239	4.994	27.869	5.788	138.686	41.488

*** p<0,01; ** p<0,05; * p<0,1.

Tabela 4.5.2 - Regressão logística (coeficientes): *pooled cross-section*

		Engenharias		Humanas		Linguística		Sociais	
Primeira colaboração	p_{t-1}^{ij}	0,010	-0,359	1,092	0,474	20,298	35,424*	2,977	2,020
		-1,158	-1,241	-2,885	-3,359	(15,065)	(21,014)	(2,374)	(2,357)
	g_{t-1}^{ij}	0,554*	0,530	-1,863*	-1,832*			0,157	0,300
		(0,305)	(0,328)	-1,005	-1,112			(0,627)	(0,625)
	$\bar{q}_t^{ij}$		0,033		0,013		-0,003		0,057
			(0,021)		(0,014)		(0,022)		(0,038)
	Δq_t^{ij}		-0,004		-0,000		-0,015		-0,037
			(0,012)		(0,009)		(0,015)		(0,031)
	$\bar{n}_t^{ij}$: total		0,013		0,003		-0,081		0,064***
			(0,008)		(0,017)		(0,234)		(0,023)
	$\bar{n}_t^{ij}$: USP		-0,234		0,895				1,058
			-1,002		-1,479				(2,569)
	Δn_t^{ij}: total		0,001		0,002		-0,092		-0,036**
			(0,006)		(0,011)		(0,083)		(0,018)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		0,226		0,049		-1,433		-0,353
			(0,552)		(0,904)		(8,482)		-2,425
	$\omega_{t,f}^{ij}$: área		1,386		3,975***		1,648		0,621
			-1,207		-1,364		(13,639)		-1,500
	$\omega_{t,f}^{ij}$: subárea		-0,502		0,061		1,492		-1,870
			-1,115		(0,994)		(9,133)		-1,198
	Constante	-1,737***	-1,771***	-2,092***	-2,158***	-1,698***	-1,892***	-1,941***	-1,986***
		(0,071)	(0,076)	(0,101)	(0,113)	(0,374)	(0,500)	(0,112)	(0,116)
	Excl. ef. rede	5,83	3,91	6,07	6,89	1,82	2,84	5,13	3,20
		[0,054]	[0,142]	[0,048]	[0,032]	[0,178]	[0,092]	[0,077]	[0,202]
	Obs.	1.009	1.009	768	768	50	50	633	633
Colaborações subsequentes		Engenharias		Humanas		Linguística		Sociais	
	p_{t-1}^{ij}	0,436**	-0,248	0,548	0,158	0,795	4,725**	0,468*	-0,580
		(0,222)	(0,483)	(0,396)	-1,146	(0,949)	(2,202)	(0,277)	(0,803)
	g_{t-1}^{ij}	0,057	0,065	0,271**	0,028			0,053	0,292
		(0,056)	(0,128)	(0,107)	(0,422)			(0,061)	(0,223)
	$\bar{q}_t^{ij}$		-0,007		-0,013		0,008		-0,033**
			(0,009)		(0,009)		(0,012)		(0,017)
	Δq_t^{ij}		0,002		0,003		-0,005		0,007
			(0,005)		(0,005)		(0,008)		(0,010)
	$\bar{n}_t^{ij}$: total		0,001		0,008		-0,004		0,010
			(0,003)		(0,007)		(0,041)		(0,008)
	$\bar{n}_t^{ij}$: USP		1,311***		1,149				0,575
			(0,493)		(0,841)				(0,720)
	Δn_t^{ij}: total		0,001		-0,005		0,117**		-0,001
			(0,002)		(0,005)		(0,045)		(0,005)
	Δn_t^{ij}: USP								
	$\omega_{t,f}^{ij}$: gr. área		-0,556*		-1,154**		-8,094*		1,071
			(0,312)		(0,573)		(4,326)		(0,961)
	$\omega_{t,f}^{ij}$: área		0,186		-0,064		4,841*		-0,557
			(0,362)		(0,533)		(2,789)		(0,578)
	$\omega_{t,f}^{ij}$: subárea		0,783***		0,828**		3,626*		0,155
			(0,268)		(0,407)		(2,129)		(0,277)
	Constante	-1,935*	-0,684	-1,907	-0,347	-0,286	0,007	-1,446***	-0,652
		-1,101	(0,478)	-1,165	(0,838)	(0,351)	-4,085	(0,136)	-1,190
	Excl. ef. rede	15,53	0,30	28,72	0,13	0,70	4,60	6,84	1,78
		[0,000]	[0,860]	[0,000]	[0,935]	[0,402]	[0,032]	[0,033]	[0,411]
	Obs.	11.249	1.792	6.488	673	449	68	6.001	658

*** p<0,01; ** p<0,05; * p<0,1.