



**UNIVERSIDADE FEDERAL DE PELOTAS - UFPEL
PROGRAMA DE PÓS-GRADUAÇÃO EM
ORGANIZAÇÕES E MERCADOS**

MESTRADO EM ECONOMIA APLICADA

MÁRCIO TACELI TAVEIRA

**Capital Humano, Trabalho e P&D no Brasil.
Análise de Uma Rede Complexa: 2006-2016**

**PELOTAS
2019**

MÁRCIO TACELI TAVEIRA

**Capital Humano, Trabalho e P&D no Brasil.
Análise de Uma Rede Complexa: 2006-2016**

Dissertação apresentada ao curso de Mestrado em Economia Aplicada da Universidade Federal de Pelotas – UFPEL, como requisito parcial para obtenção do título de Mestre em Economia.

Área de Concentração: Economia
Orientador: Marcelo de Oliveira Passos

**PELOTAS
2019**

MÁRCIO TACELI TAVEIRA

**Capital Humano, Trabalho e P&D no Brasil.
Análise de Uma Rede Complexa: 2006-2016**

Dissertação apresentada como requisito parcial
para obtenção do grau de Mestre em Economia
pelo Programa de Pós-Graduação em
Organizações e Mercados da Universidade
Federal de Pelotas (UFPEL)

Aprovada em ____/____/____

BANCA EXAMINADORA

Cristiano Aguiar de Oliveira (membro externo)
Universidade Federal do Rio Grande

Daniel de Abreu Pereira Uhr (membro interno)
Universidade Federal de Pelotas

Marcelo de Oliveira Passos (orientador)
Universidade Federal de Pelotas

Sumário

1. Introdução	9
2. Revisão de Literatura sobre redes complexas	13
2.1 Definição de Grafo e de Rede Complexa	13
2.2 Propriedades topográficas das redes complexas	13
2.3 Mecânica estatística de redes complexas	15
2.3.1 Medidas ao nível dos agentes	15
2.3.2 Medidas ao nível da rede	16
2.4 O <i>PageRank</i>	17
2.5 Clusters, agrupamentos hierárquicos e classes de modularidade	19
3. As hipóteses de Acemoglu e Autor (2014) e Alon (2018).....	22
4. Fontes, metodologia, softwares e distribuições de layout	24
4.1. Fontes dos dados.....	24
4.1.1 Cálculo dos pesos	24
4.2 Distribuições de Yifan-Hu e Yifan-Hu Proporcional	25
4.3 Por que utilizar redes complexas para analisar capital humano?	25
5. Resultados	27
5.1 Análise ao nível da rede – Classes de Modularidade (CMs)	27
5.2 Análise ao nível dos vértices/agentes: <i>hubs</i> , graus ponderados, <i>PageRanks</i> e centralidade dos autovetores	31
5.2.1 Existência de <i>Hubs</i>	32
5.2.2 Graus e graus ponderados	34
5.2.3 <i>PageRanks</i>	40
5.2.4 Centralidade do autovetor	46
6. Considerações Finais	48
7. Referências	51

Lista de Gráficos

Figura 1 - Todas as <i>web pages</i> se conectam apenas com a <i>web page</i> α	15
Figura 2 - Todas as <i>web pages</i> conectam-se com mais de uma outra <i>web page</i>	16
Figura 3 - Web Page β sem nenhum link.	17
Figura 4 - Evolução temporal das redes conforme suas classes de modularidade e pesos das arestas.	25
Figura 5 - Evolução temporal das redes conforme seus graus ponderados de entrada e pesos das arestas.	30
Figura 6 - Taxa de emprego formal (carteira assinada) em 2009 e em 2014 de mestres titulados no Brasil a partir de 1996, por grande área do conhecimento (em %).	31
Figura 7 - Evolução dos PageRanks conforme aumento da instrução formal: 2006/2015.....	33
Figura 8 - Evolução temporal das redes conforme seu <i>pagerank</i> e pesos das arestas.	37

Lista de Tabelas

Tabela 1 – Classificação Brasileira de Ocupações do MTE.	22
Tabela 2 – Alterações nos dois clusters do mercado de trabalho das indústrias brasileiras intensivas em PD&I.	22
Tabela 3 – Classes de modularidade e clusters das 4 redes complexas do mercado de trabalho de ocupações intensivas em PD&I no Brasil: 2006-2016.	23
Tabela 4 – Graus ponderados médios e <i>hubs</i> das 4 redes complexas do mercado de trabalho de ocupações intensivas em PD&I no Brasil: 2006-2016.	26
Tabela 5 – Crescimento médio e desvios-padrão dos graus de entrada ponderados nos triênios de 2006-08, 2009-11, 2012-14 e no biênio de 2015-16.	28
Tabela 6 – Graus ponderados de entrada (GPEnts) e variação entre o primeiro e o último períodos nos triênios de 2006-08, 2009-11, 2012-14 e no biênio de 2015-16.	28
Tabela 7 – Número de empregados entre os doutores e os mestres titulados no Brasil a partir de 1996, por seção da Classificação Nacional de Atividades Econômicas (CNAE) dos estabelecimentos empregadores, 2009 e 2014	28
Tabela 8 - PageRanks (PRs) e variação entre o primeiro e o último períodos nos triênios de 2006-08, 2009-11, 2012-14 e no biênio de 2015-16.	34
Tabela 9 – Variações dos PageRanks (PRs) entre o 1º triênio de 2006-08 e o último biênio de 2015-16, incluindo e excluindo os cargos de DP&D e Ger.	35

Lista de Siglas

ARS – Análise de Redes Sociais

Bio - Biólogos e afins

BiotGen - Biotecnologistas, geneticistas, pesquisadores em metrologia e especialistas em calibrações meteorológicas

CBO – Classificação Brasileira de Ocupações

CM – Classe de Modularidade

D – Doutorado

DP&D - Diretores de P&D

ECiv - Engenheiros Civis etc.

EM – Ensino Médio

FiQm - Físicos, químicos e afins

GDP/LOC – é o nível real de PIB per capita

Ger – Gerentes de P&D

GMM – *Generalized Methods Moment*

IBGE – Instituto Brasileiro de Geografia e Estatística

Info - Profissionais de informática

M – Mestrado

Mat - Matemáticos, estatísticos e afins

Mecat - Engenheiros mecatrônicos

MIT – *Massachusetts Institute of Technology*

MTE – Ministério do Trabalho e Emprego

P&D – Pesquisa e Desenvolvimento

PD&I – Pesquisa, Desenvolvimento e Inovação

Pesq - Pesquisadores

PIB – Produto Interno Bruto

PNAD – Pesquisa Nacional por Amostra de Domicílios

PR – Page Rank

PSID - *Panel Study of Income Dynamics*

PTF – Produtividade Total dos Fatores

R² – Coeficiente de determinação

RAIS – Relação Anual de Informações Sociais

Rede SF – Rede *Scale Free* (Rede Sem Escalas)

SC – Superior Completo

SENAI – Serviço Nacional de Aprendizagem Industrial

SI – Superior Incompleto

TPF – Total Factor Productivity

Capital Humano, Trabalho e P&D no Brasil. Análise de Uma Rede Complexa: 2006-2016

Resumo: O principal objetivo deste estudo é analisar a relação entre a acumulação de capital humano e as variações salariais das profissões consideradas "totalmente alocadas em pesquisa, desenvolvimento e inovação (PD & I)", segundo a classificação do Gusso (2006). Conforme Tchernis (2010) é necessária uma investigação mais aprofundada dos efeitos do capital humano no crescimento salarial, bem como a construção de novas medidas de capital humano que reflitam competências transferíveis entre alguns, mas não todos, os postos de trabalho. Nesse sentido, empregamos o método da mecânica estatística de redes complexas, a partir da qual analisamos essa relação no período 2006-2016. Como esse método é não-paramétrico e fortemente influenciado pela teoria dos grafos e pela topologia, não testamos a causalidade entre essas variáveis e optamos por encontrar evidências de relações topológicas significativas e calcular estatísticas de redes complexas. Assim, esse trabalho desenvolve uma análise com base em dois objetivos específicos: (i) testar a hipótese de que existe uma relação entre o nível educacional e a variação salarial nas ocupações de empresas industriais com maior capacidade inovativa e que investem em PD & I; e ii) verificar se há evidências que suportem as hipóteses Acemoglu e Autor (2014) e de Alon (2018). Acemoglu e Autor (2014) sugerem que indivíduos que frequentam a mesma quantidade de anos na mesma escola podem ter níveis assimétricos de capital humano devido a incentivos salariais que melhoram o processo de learning by doing. Alon (2018) considera que o aumento do grau de especialização dos trabalhadores norte-americanos respondeu por cerca de 29% do aumento do prêmio pago pelas habilidades e por 25-30% do aumento do nível educacional no período de 1965-2005. Para lidar com tais hipóteses, desenvolvemos medidas estatísticas derivadas de quatro redes complexas baseado no Yifan Hu proporcional (Hu, 2006) algoritmo proporcional para os 2006-08, 2009-2011 e 2012-2014 triênios e um para o biênio 2015-2016. Usando esses dados, comparamos as variações salariais resultantes do investimento em capital humano avançado e o aprendizado-por-fazer entre ocupações

Palavras – Chave: inovação na indústria, capital humano, redes complexas, mercado de trabalho, economia computacional.

Códigos JEL: J31, C14, C45, C81

Abstract: The main objective of this study is to analyze the relationship between the accumulation of human capital and the wage variations of the professions considered as "fully allocated in research, development and innovation (PD & I)", according to the classification of Gusso (2006). According to Tchernis (2010), a more in-depth investigation of the effects of human capital on wage growth is required, as well as the construction of new human capital measures that reflect transferable skills between some, but not all, jobs. In this sense, we use the method of statistical mechanics of complex networks, from which we analyze this relationship over the period 2006-2016. Since this method is non-parametric and strongly influenced by graph theory and topology, we did not test the causality between these variables and chose to find evidence of significant topological relationships and compute complex network statistics. Thus, this work develop an analysis based on two specific objectives: (i) test the hypothesis that there is a relationship between the educational level and the wage variation in the occupations of industrial companies with greater innovative capacity and that invest in PD & I; and ii) verify if there is evidence to support the hypotheses formulated by Acemoglu and Autor (2014) and Alon (2018). Acemoglu and Autor (2014) suggest that individuals who attend the same amount of years in the same school may have asymmetric levels of human capital due to wage incentives that improve the learning by doing process. Alon (2018) considers that the increase in the degree of specialization of American workers

accounted for about 29% of the increase in the premium paid for the skill and 25-30% of the increase in educational level in the period 1965-2005. To deal with such hypotheses, we developed statistical measures derived from four complex networks based on the Yifan Hu Proportional (Hu, 2006) proportional algorithm for the 2006-08, 2009-2011 and 2012-2014 triennials and one for the 2015-2016 biennium. Using this data, we compare the wage variations resulting from advanced human capital investment and learning-by-doing between occupations

Keywords: industry innovation, human capital, complex networks, labour market, computational economics.

JEL Codes: J31, C14, C45, C81.

1. Introdução

O objetivo geral deste trabalho é O principal objetivo deste trabalho é, com base na mecânica estatística de redes complexas (*statistical mechanics of complex networks*), analisar as relações existentes entre o acúmulo de capital de humano e as variações salariais entre as profissões consideradas como “integralmente alocadas em PD&I”, conforme a classificação de Gusso (2006).

A ideia inicial da dissertação era trabalhar com Renúncia Fiscal e Gasto em Pesquisa, Desenvolvimento e Inovação (PD&I) na Indústria Brasileira. Mais especificamente, foi proposto inicialmente tentar identificar se existe e qual seria a relação de Causalidade (no sentido de Granger)¹ entre as variáveis acima. Primeiramente, buscamos na literatura textos similares. Os achados sobre o tema retornaram apenas resultados em termos de estimação de parâmetros e composição dos gastos com P&D, mas nada muito próximo do objetivo inicial, que compreendia essencialmente análise de políticas públicas.

Para análise acima, os manuais de Econometria e/ou Séries Temporais, sugerem que primeiro se efetivem os testes de estacionariedade das séries temporais². Utilizamos os dados de Renúncia Fiscal e Gastos em PD&I separados por Macrorregião, por Estados e por Grande Setor com a periodicidade de 2006-2014, tal qual como estão disponíveis nos Relatórios da Lei do Bem elaborados pelo MCTIC (findados em 2015, após a suspensão da Medida Provisória que regia a Lei do Bem). Os testes foram feitos através do software estatístico RStudio, em que não foi possível encontrar evidências de estacionariedade dos dados, tornando inviável a análise através da Causalidade de Granger. Foi necessário então encontrar abordagens alternativas para trabalhar com o tema acima mencionado.

Dos trabalhos encontrados, o que mais se assemelhou ao exame proposto, foi o de Kannebley Júnior et al. (2016)³, em que os autores fazem uma análise econométrica sobre a relação entre o gasto com P&D e Renúncia Fiscal (com o gasto em P&D como variável dependente) em meio ao efeito da chamada “Lei do Bem”⁴. O modelo escolhido pelos autores foi o de dados em painel, mais especificamente um *diferenças em diferenças* com um *propense score matching* para dividir os setores da industrial nacional por intensidade tecnológica. Dentre as várias metodologias de construção e estimação das variáveis explicativas, pôde-se observar que quase não houve tratamento para a variável de remuneração dos trabalhadores “integralmente alocados em P&D”. O texto ainda argumenta que dos gastos totais com P&D, cerca de 60% são dispendidos através do pagamento de

¹ Ver Granger (1969) para detalhamento do conceito.

² Procedimento sugerido por Enders (2009) em *Applied econometric time series*, 3ª edição.

³ Para mais informações ver a referência número XX (KANNEBLEY JÚNIOR, S.; SHIMADA, E.; DE NEGRI, F. Efetividade da lei do bem no estímulo aos dispêndios em P&D: uma análise com dados em painel. **Pesquisa e Planejamento Econômico**, v. 46, n. 3, 2016.

⁴ Uma síntese pode ser encontrada no site da Associação Brasileira de Gestão em Inovação (ABGI): <http://brasil.abgi-group.com/lei-do-bem/>

salários e bônus (de produtividade, por exemplo). Entretanto, mesmo representando uma grande parte da variável dependente, o texto não apresenta uma discussão detalhada sobre as remunerações⁵.

Nesse sentido, dada a escassez de literatura específica sobre as remunerações da indústria nacional intensivas em PD&I, julgou-se que este problema de pesquisa poderia ser explorado por meio de técnicas estatísticas não-paramétricas de redes complexas. Partindo da teoria do capital humano, explicitada por Schultz (1961), utilizamos essa metodologia ainda pouco utilizada no país para investigar, pelo lado dos trabalhadores, qual a relação entre os aumentos de capacitação técnica (seja ela formal ou informal, como no caso do *learning by doing*) e a evolução dos diferenciais de salário na indústria brasileira intensiva em PD&I.

Entre os estudos econômicos aplicados mais recentes que utilizam a teoria do capital humano, é usual encontrar trabalhos baseados em agregados macro e/ou microeconômicos. Normalmente, estes trabalhos, alguns dos quais citados nessa dissertação, utilizam métodos econométricos mais tradicionais, como a econometria de dados em painel e em *cross-section*. Nesse sentido, este trabalho visa preencher uma lacuna na literatura, uma vez que não há qualquer outra obra que aborde a relação entre os incentivos e a acumulação de capital humano pela formação técnica e o comportamento das variações salariais. Entretanto, pode-se encontrar na literatura internacional, sobretudo nos trabalhos de Matthew Jackson, uma alternativa metodológica para estudo do problema, por meio da *network science* ou análise de redes complexas⁶.

Além disso, um dos temas mais analisados pelos especialistas em economia do trabalho é a formação dos salários. Quando esse assunto se relaciona com o diferencial de salários inter e intraindustriais, o problema passa a pertencer também à esfera dos estudos da economia industrial. Entre os anos 1980 e 1990, alguns economistas industriais e microeconométricos, dedicaram-se a esses estudos. Após controlarem por características de país, firma, tipo de indústria e indivíduos, encontraram resultados de uma regularidade considerável na dispersão dos salários. Entre esses trabalhos, podemos citar Krueger e Summers (1987,1988) referentes aos EUA e Gitleman e Wolf (1993) para 14 países da OCDE.

Nesse sentido, a primeira seção faz uma breve revisão da literatura empírica sobre redes complexas, enfatizando os conceitos referentes às métricas de mecânica estatística que serão utilizadas ao longo do texto.

A segunda seção trata das considerações de Acemoglu e Autor (2014), para os quais o debate sobre os incentivos para o investimento em capital humano conduz à descoberta de cinco fatores, dos quais destacaremos dois: (i) escolaridade - que é o ponto fundamental deste artigo e da maioria da literatura empírica sobre capital humano; e (ii) qualidade escolar e investimentos/incentivos não escolares - indivíduos que cursaram a mesma quantidade de anos em uma mesma escola podem também possuir diferentes níveis de capital humano por causa de investimentos ou incentivos em outros tipos de capital humano que não o ensino formal (um deles pode ter obtido maior nível de educação por causa do *learning-by-doing*, variável diretamente afetada pelos incentivos que uma remuneração mais alta podem proporcionar, ou por ter feito sido submetido à circunstâncias que lhe fizeram aprimorar uma habilidade como escrever ou falar melhor). Há na literatura um debate sobre a importância dessas habilidades “não observáveis” e seus impactos na estrutura salarial. A dificuldade reside na aferição desses componentes do capital humano, uma vez que não há dados confiáveis sobre eles. Assim, nesta segunda seção, delineamos a intenção de contribuir com uma análise para o caso brasileiro de como se configura o perfil de remunerações das ocupações na indústria intensiva em PD&I. Entendemos que tais remunerações são mecanismos de incentivo para a aquisição tanto de ensino formal quanto de *learning-by-doing*.

⁵ Podemos encontrar análises um pouco mais antigas em Arbache e De Negri (2002,2004).

⁶ Para um detalhamento maior, ver “Social and Economic Networks” de Matthew O. Jackson (2008).

A terceira seção explicita as fontes dos dados, os softwares utilizados e as distribuições de layout de Yifan Hu e Yifan Hu Proporcional. Estas distribuições são úteis para a visualização das redes e complementam a análise estatística da seção seguinte.

A quarta seção descreve os resultados obtidos por meio das medidas (ou métricas) da mecânica estatística de redes complexas detalhadas na revisão da primeira seção. Na opinião de Tchernis (2010) é necessária uma pesquisa mais detalhada sobre os efeitos do capital humano no crescimento salarial. Este autor advoga a necessidade de construção de novas medidas de capital humano que reflitam competências transferíveis entre alguns, mas não todos, os postos de trabalho. Nesse sentido, produzimos estas novas medidas a partir do emprego do método da mecânica estatística de redes complexas. Essas medidas quantificam as relações das topografias das redes e examinam as seguintes hipóteses: (i) se existem na indústria brasileira evidências de uma relação entre o grau de instrução e a variação salarial nos cargos das empresas industriais com maior capacidade inovativa e que dependem fundamentalmente dos investimentos na formação de capital humano intensivo em PD&I? (ii) se há evidências, considerando as estatísticas não-paramétricas do método proposto de correlações na topografia das redes obtidas que mostrem que há relações entre a acumulação de capital humano das referidas profissões e variações maiores nas suas remunerações, *coeteris paribus*? (iii) se é possível encontrar evidências que suportem a constatação de Acemoglu e Autor (2014) de que os indivíduos que cursaram a mesma quantidade de anos em uma mesma escola podem ter níveis assimétricos de capital humano por causa de incentivos salariais que tornam atraente o processo de *learning-by-doing*? Para lidar com tais hipóteses, desenvolvem-se estatísticas derivadas de quatro redes complexas baseadas no algoritmo de Yifan Hu proporcional (Hu, 2006) para os triênios de 2006-08, 2009-2011 e 2012-2014 e uma para o biênio de 2015-2016. Com base nestas métricas, comparamos as variações salariais resultantes do investimento em capital humano avançado e do *learning-by-doing* entre as ocupações.

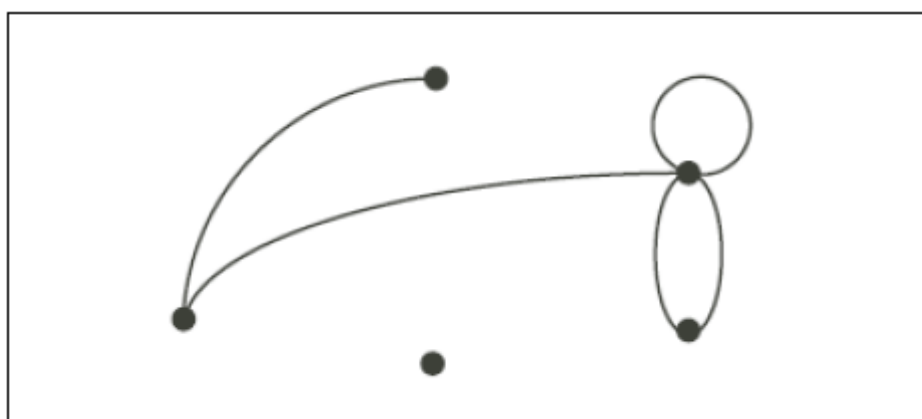
A quinta seção resume as última seção considerações finais.

2. Revisão da literatura sobre análise de redes complexas

2.1. Grafos e redes complexas

Por se tratar de um método ainda pouco utilizado pelos pesquisadores brasileiros, cabe uma breve introdução ao tema. Segundo Ruohonen (2013), uma rede complexa é um grafo no qual há um conjunto de tamanho considerável de vértices (ou nós) e outro de arestas (ou arcos) que fazem a ligação entre esses vértices. As arestas cumprem a função de expressar alguma relação entre dois vértices, conforme o fenômeno que é modelado.

Figura 1 – Grafo formado por um conjunto de vértices e arestas



Fonte: Ruohonen (2013)

A rede ou o grafo podem ser direcionados ou não. Um grafo direcionado é chamado de dígrafo (de digraph, ou directed graph). Nele, cada aresta possui um sentido, o qual conecta um vértice de origem a pelo menos um outro vértice de destino. Exemplos de dígrafos são os que descrevem curtidas em redes sociais, chamadas telefônicas ou mensagens em aplicativos de comunicação. Os dígrafos são chamados de cíclicos quando são existe(m) caminho(s) de um vértice para ele mesmo, sendo acíclicos, em caso contrário.

Formalmente, para ele, um grafo é um par de conjuntos (G), isto é, $G = (V, E)$, onde V é um conjunto de elementos chamados vértices e E é um conjunto de pares de vértices não ordenados denominados linhas ou arestas (o E , vem do inglês *edge*).

Segundo Metz et al. (2007), o tipo de rede que mais comumente representa as relações entre os seres humanos é o a rede complexa. Os autores argumentam que as redes complexas são estruturas que não seguem um padrão regular (por exemplo, um tetraedro poderia ser a representação visual de uma rede regular). Ainda não há um consenso na literatura sobre o que é exatamente padrão regular para redes. Nem tampouco um conceito universalmente aceito sobre qual a constituição exata ou desejável dessas redes.

Os autores ainda mencionam a falta de consenso sobre a definição desse tipo de redes. Mas já existe consenso sobre o fato de que elas possuem características próprias que não estão presentes nas redes regulares. Essas características revelariam, portanto, o processo de formação das redes, assim como a utilidade delas para a análise de um determinado problema. Para contornar essa diversidade de interpretações, utilizaremos nesse trabalho os conceitos e definições de Albert e Barabási (2002), que são aqueles nos quais a literatura mais relevante usualmente se apoia.

2.2. Análise de Redes Sociais

A análise de redes sociais (que representam as interações sociais entre indivíduos) já é uma área central e bem desenvolvida na sociologia, com sociedades, *journals*, conferências e décadas de pesquisa dedicadas especificamente a esse campo⁷. Ao mesmo tempo, a Teoria dos Grafos amadureceu na matemática no mesmo período⁸. Ainda que os estudos em redes tenham prosperado na sociologia por mais de seis décadas, a economia começou a explorar essa vertente com certa significância apenas perto dos anos 2000 (veremos exemplos mais à frente). Sua disseminação na ciência da computação e na física (principalmente através da mecânica estatística) tem ajudado a dar escala ao método durante as duas últimas décadas⁹.

Em relação à importância prática da chamada *network science*, as aplicações voltadas à análise de redes de relacionamentos (sinônimo para redes sociais) desempenham papel central em uma extensa pluralidade de interações sociais, econômicas e políticas. Por exemplo, muitos mercados, funcionam não como instituições centralizadas e anônimas, mas como sistemas que envolvem uma multiplicidade de trocas bilaterais ou negociações de contratos. Ioannides e Loury (2004) mostraram que grande parte dos empregos é preenchida por pessoas que foram informadas sobre o emprego por meio de um contato próximo. Este fato teria então, consequências para os padrões de emprego, desigualdade salarial entre os grupos sociais e até na mobilidade social desses indivíduos. Portanto, entender a estrutura da rede social e como ela influencia as interações humanas, não é importante apenas para a ciência (e para as ciências sociais em particular), é também importante para um bom entendimento sobre o funcionamento da maioria das relações humanas.

A abordagem de redes (*network approach*) ganhou evidência nos últimos, segundo Albert e Barabási (2002) por vários motivos. O primeiro deles diz respeito à evolução da

⁷ O artigo considerado seminal (pela literatura em sociologia) é o de SCOTT (1988).

⁸ Uma exposição detalhada da abordagem matemática pode ser encontrada em WEST (1996).

⁹ Para contribuições mais recentes, amplas e sistematizadas da matemática, ver Watts e Strogatz (1998) e para a mecânica estatística, Albert e Barabási (2002).

capacidade computacional (softwares, hardwares e internet), bem como a crescente disponibilidade de uma diversidade quase infinita de informações sobre praticamente qualquer aspecto do cotidiano. O argumento dos autores parte do princípio de que esses avanços possibilitaram um aumento considerável da velocidade de aquisição de dados, bem como a emergência de bases cada vez maiores que descrevem a topologia das redes reais. Isto viabilizou a investigação de questões anteriormente inalcançáveis. Outro ponto importante é que a popularização das redes de informação (por exemplo, a *world wide web*), vêm conscientizando a comunidade acadêmica (em geral) sobre a importância das redes em suas vidas. O segundo motivo diz respeito à diminuição das barreiras entre os diferentes campos de estudo. Essa tendência despertou o interesse dos pesquisadores em diversos tipos de bases de dados, permitindo-os a descobrir as propriedades genéricas das redes complexas referentes a eles. Além disso, enquanto os métodos de análise estatísticos mais tradicionais (como a Econometria e a Estatística Bayesiana) apresentam o rigor e a validação/aceitação metodológicas já consolidados, enquanto a análise de redes contribui principalmente com a uniformização e aplicação do mesmo método à diferentes campos de pesquisa (devido à sua natureza inter e multidisciplinares), facilitando a difusão do conhecimento em uma escala consideravelmente maior do que os métodos tradicionais. Somando isso à tendência da busca da compreensão de fenômenos cada vez mais gerais (que é justamente a proposta da análise de redes, o método vem ganhando cada vez mais adeptos.

2.3. Propriedades topográficas das redes complexas

Não são todos os grafos que recebem a classificação de rede complexa. Apenas aqueles que exibem algumas propriedades topográficas específicas, que não se encontram em grafos simples, são considerados assim. Uma rede para ser considerada complexa deverá representar um *small world* ou ser do tipo *scale free*¹⁰. Além disso, na maioria das vezes elas apresentam algumas propriedades específicas (a figura 2 exibe um exemplo de rede complexa).

As redes complexas são caracterizadas por algumas propriedades que são úteis para a investigação das várias facetas das redes e atendem à diversas finalidades. Nessa seção, destacamos aquelas que são mais utilizadas pela literatura empírica, conforme Metz, Calvo et al. (2007).

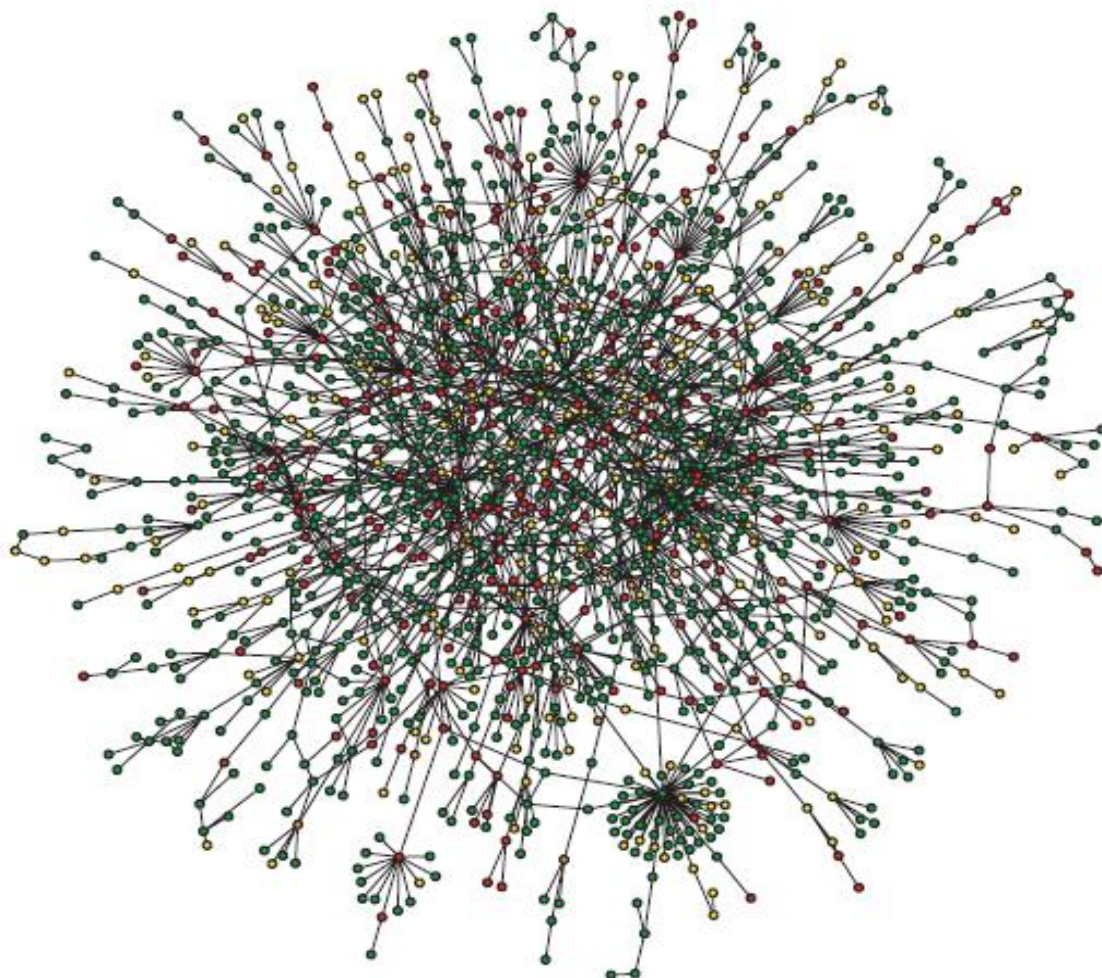
- **Coefficiente de aglomeração:** os clusters (comunidades ou agrupamentos) das redes podem ser quantificados por intermédio do coeficiente de aglomeração. Esta medida ou métrica da estatística de redes reflete o fenômeno da transitividade, que existe quando um vértice A se conecta com um vértice B, o qual está conectado a um vértice C, o que aumenta a probabilidade deste vértice C conectar-se com o A. Portanto, uma rede com muita transitividade indica a presença de uma alta quantidade de triângulos, que são os conjuntos de três vértices conectados entre si. O coeficiente de aglomeração CA de uma rede é definido por (1):

$$CA = \frac{3 \times \#\Delta}{\#v} \quad (1)$$

onde $\#\Delta$ refere-se ao número de triângulos na rede e, $\#v$ representa o número de vértices triplamente conectados, isto é, aqueles cujas arestas não estejam direcionadas para o outro par de nós. O número 3 no numerador indica que cada triângulo possui três triplas, o que assegura que o CA tenha um valor entre 0 (zero) e 1 (um).

¹⁰ Ver Barabási e Bonabeau (2003).

Figura 2 – Mapa das proteínas que interagem entre si em uma célula de fermento biológico



MAP OF INTERACTING PROTEINS in yeast highlights the discovery that highly linked, or hub, proteins tend to be crucial for a cell's survival. Red denotes essential proteins (their removal will cause the cell to die). Orange represents proteins of some importance (their removal will slow cell growth). Green and yellow represent proteins of lesser or unknown significance, respectively.

Fonte: Barabási (2003)

- **Distribuição de Graus:** o grau de um vértice qualquer em uma rede define o número de arestas que conectam ou estão ligadas àquele vértice. Portanto, a distribuição de graus é uma função de distribuição probabilística que indica a probabilidade de um determinado vértice apresentar um grau fixo. Uma forma de mensurar esta distribuição é utilizando uma função de distribuição cumulativa, tal como expressa em (2)

$$P_k = \sum_{k'=k}^{\infty} p_{k'} \quad (2)$$

Sendo P_k a função cumulativa de distribuição de probabilidades e $p_{k'}$ a fração de nós da rede com grau k .

No caso de um dígrafo, cada vértice possui um grau de entrada e outro de saída. Portanto, temos que a centralidade do grau (também chamada de valência) é definida pela seguinte expressão:

$$k_i = \sum_{j=1}^n a_{ij}, \quad 0 < k_i < n \quad (1) \quad \text{e} \quad k_v = |N_v| \quad 0 < k_v < n \quad (3)$$

Onde a_{ij} é a entrada da i -ésima linha e j -ésima coluna da matriz de adjacência A . E N_v é a vizinhança do agente (nó ou vértice) V .

Para redes direcionadas temos:

k_i^+ = grau de entrada (número de agentes de entrada, isto é, número das arestas ou relações que começam no agente v).

k_i^- = grau de saída (número de agentes de saída, isto é, número das arestas ou relações que terminam no agente v).

$$k_i^+ = \sum_{j=1}^n a_{ij} \quad (3), \quad k_i^- = \sum_{j=1}^n a_{ji} \quad (4)$$

A medida do grau em redes direcionadas é também conhecida como prestígio. É uma expressão muito usada em ARS (análise de redes sociais).

Existem dois tipos de prestígio: (i) o de suporte; e (ii) o de influência. O de suporte é o grau de entrada e o de influência é o grau de saída. Em redes pesadas (ou ponderadas) a força é equivalente ao grau. Ela é igual à soma dos pesos das arestas adjacentes a um dado agente (ou das relações vinculadas a este agente). Tal como em (5):

$$k_i^w = \sum_{j=1}^n a_{ij}^w \quad (5)$$

É importante ressaltar que a distribuição de graus nas redes aleatórias segue a distribuição de Poisson, enquanto em diversas reais redes reais a distribuição de graus segue uma Lei de Potência, em que

$$p_{k'} \sim k^{-\alpha} \quad (6)$$

onde α é uma constante qualquer.

- **Resiliência:** indica a capacidade de uma rede manter sua funcionalidade enquanto seus vértices vão sendo removidos. A resiliência está diretamente associada à distribuição dos graus dos vértices, dado que tal remoção pode gerar perda de conexões entre pares de vértices, aumentando a distância no caminho entre um vértice e outro.
- **Misturas de Padrões:** alguns tipos de redes caracterizam-se por padrões misturados, nos quais os vértices podem representar diferentes tipos de objetos. Um exemplo típico é o caso de redes de cadeias alimentares, onde há vértices que representam animais carnívoros coexistindo com outros vértices que denotam animais herbívoros e ainda vértices que representam plantas. Quando há padrões misturados, a probabilidade de conexão entre os vértices se dá em função do seu tipo. Assim, neste exemplo existiriam arestas ligando animais herbívoros às

plantas e animais carnívoros aos animais herbívoros. Todavia, nelas seria raro encontrar conexões entre herbívoros-herbívoros ou carnívoros-plantas. As redes sociais também podem apresentar misturas de padrões, pois elas são formadas por vértices de pessoas com diferentes hábitos, preferências, profissões, etnias, nível educacional etc. Portanto, elas apresentam uma tendência clara de haver mais conexões entre vértices do mesmo tipo, dado que as pessoas estão habituadas a relacionarem-se com outras que possuam semelhanças nas características citadas (Newman, 2003). Newman (2003) também percebeu que praticamente todas as redes sociais apresentam esses padrões misturados, ao passo que outros tipos de redes não apresentam tal mistura.

- **Correlação de Graus:** indica se as arestas de uma rede conectam-se com vértices similares. A correlação de graus é utilizada sobretudo em redes com padrões misturados, no sentido de testar a probabilidade de conexão entre vértices de tipos diferentes.

2.4. Mecânica estatística de redes complexas

Utilizar medidas estatísticas para analisar redes complexas permite compreender e avaliar a estrutura da rede sem a necessidade de conhecer a sua representação gráfica. O objetivo dessas medidas é quantificar a estrutura das redes, para possibilitar ao analista a compreensão do comportamento dos fenômenos econômicos que geraram essas redes.

As medidas estatísticas podem ser ao nível dos agentes (nós) ou ao nível da rede complexa como um todo.

2.4.1. Medidas ao nível dos agentes

As medidas ao nível dos agentes utilizadas são:

1. Centralidade do grau ou valência (já definida na seção anterior); e
2. Centralidade do autovetor.

A equação (7) fornece a centralidade do autovetor:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j \quad (7)$$

Onde x_i/x_j representa a centralidade do agente i/j ; a_{ij} denota a matriz de adjacência A ($a_{ij} = 1$ se os agentes i e j estão conectados por uma aresta e $a_{ij} = 0$ se não estão) e λ indica o maior autovetor da matriz A .

A centralidade do autovetor é uma medida proposta por Bonacich (1987) e fundamenta-se na noção de que a centralidade de um agente é definida pela centralidade dos agentes com os quais se relaciona (via trocas, transações etc.). Assim, o poder ou status econômico-financeiro de um agente é definido pelo poder ou status econômico-financeiro de seus *alters*. Os *alters* são os agentes diretamente relacionados ao agente central (também chamado de ponto focal ou ego). A centralidade do autovetor é a combinação linear das centralidades dos seus vizinhos de primeira ordem.

2.4.2. Medidas ao nível da rede

Quanto às medidas ao nível da rede, existem três conceitos fundamentais da teoria dos grafos devem ser conhecidos antes da apresentação delas: (i) o caminho; (ii) a distância geodésica (caminho mais curto); e (iii) a excentricidade.

Um caminho é uma sequência de agentes (vértices ou nós) nos quais pares consecutivos de agentes não repetidos estão ligados por uma relação (conexão ou aresta). O primeiro agente de um caminho é o agente inicial e o último agente é o agente final.

A distância geodésica (ou caminho mais curto) é denotada por $d(i,j)$ e corresponde ao caminho mínimo entre os agentes i e j .

A excentricidade é a maior distância geodésica entre um agente v e qualquer outro agente no grafo, conforme a expressão seguinte:

$$\varepsilon_v = \max_{u \in V(G) \setminus v} d(v, u) \quad (8)$$

Isto posto, as estatísticas ao nível da rede que servem de base para nossa análise são:

(i) Diâmetro/raio; (ii) distância geodésica média; (iii) grau médio; e (iv) densidade.

O diâmetro (D) é a excentricidade máxima do conjunto de agentes que definem a rede. Já o raio (R) é a excentricidade mínima deste conjunto de agentes.

As redes esparsas costumam apresentar diâmetro maior do que as redes completas. Isto ocorre porque as redes esparsas possuem caminhos menores entre pares de agentes.

Para certo tipo de redes reais, o diâmetro efetivo diminui ao longo do tempo (Leskovec et al, 2005). Isto contraria a sabedoria convencional de diâmetros progressivamente menores. Esta medida fornece uma ideia de proximidade de pares de agentes na rede complexa. Ela mede a distância entre dois agentes no pior dos casos.

$$D = \max \{ \varepsilon_v : v \in V \} \quad (12) \quad R = \min \{ \varepsilon_v : v \in V \} \quad (9)$$

A distância geodésica média é denotada por l . É medida para todas as combinações de pares de agentes numa rede e expressa a noção de quão longe estão dois agentes um do outro, em média. Ela ajuda a medir a eficiência do fluxo de transações, trocas ou relações econômico-financeiras no interior da rede. Ela é expressa por (14):

$$l = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j} d(i, j) \quad (10)$$

Onde $d(i, j)$ representa a distância geodésica, ou comprimento do caminho mais curto, entre os agentes i e j . E $\frac{1}{2}n(n-1)$ indica o número máximo de relações (ou arestas, ou vínculos) numa rede composta por n agentes. Esta equação não vale para o caso de uma rede que apresenta mais do que uma componente conectada. Isto ocorre porque a $d(i, j)$ é infinita quando não há caminho na rede que conecte dois agentes. Nesses casos, é melhor recorrer à média harmônica da distância geodésica (equação 11):

$$l^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} \frac{1}{d(i, j)} \quad (11)$$

A média harmônica da distância geodésica é à medida que possibilita a transformação de distâncias infinitas em distâncias nulas.

O grau médio, definido por (12), é apenas a média dos graus de todos os agentes da rede. Conforme Costa et al. (2011), o grau médio é uma medida de conectividade global da rede.

$$\bar{k} = \frac{1}{n} \sum_{i=1}^n k_i \quad (12)$$

De acordo com Wasserman e Faust (1994), uma díade assimétrica é um par de agentes conectados por uma relação (aresta) cuja direção assume o sentido de um dos agentes, mas não de ambos. De modo contrário, uma díade mútua é um par de agentes conectados por duas ligações, cada uma delas indo a uma direção diferente. Por exemplo $a \rightarrow b$ e sendo a e b dois agentes (ou nós, ou vértices) em uma rede.

A densidade ρ quantifica o grau de conectividade que uma rede possui e é uma medida relevante ao nível da rede, pois ela expressa a proporção de arestas da rede em comparação com o número máximo possível de arestas. O valor mínimo da densidade é 0, quando a rede não possui arestas, e o valor máximo é 1, quando ela possui conectividade absoluta (uma rede completa ou um clique). Portanto, quando a densidade é alta, a rede é densa e, quando ocorre o contrário, a rede é esparsa.

$$\rho(G) = \frac{m}{m_{\max}}, 0 < \rho < 1 \quad (13)$$

Onde m representa o número de arestas presentes numa rede e $m_{\max}$ indica o número máximo de arestas nessa rede. No caso de as redes não serem direcionadas, o número máximo de arestas será dado por:

$$m_{\max}(nd) = \frac{n(n-1)}{2} \quad (14)$$

Onde $m_{\max}(nd)$ simboliza o número máximo de arestas em uma rede não direcionadas e n é o número de vértices da rede.

Porém, no caso de a rede ser direcionada, o número máximo de arestas é dado por: $m_{\max}(d) = n(n-1)$ (15), onde $m_{\max}(d)$ é o número máximo de arestas em uma rede direcionada.

2.5. O PageRank

Em determinados cenários o analista pode estar interessado em descobrir o nó dominante, com maior influência ou uma lista ordenada de nós com essas características. Para isso foram desenvolvidos algoritmos de análise de ligações, sendo o PageRank e o HITS os mais populares. Esses algoritmos exploram a relação existente entre as ligações e o conteúdo das páginas Web, com o intuito de melhorar a tarefa de recuperação de informações na Web, sendo extremamente importante no desenho de motores de busca eficientes.

Utilizaremos nesse trabalho apenas o PageRank, sendo *hub* e *authority* conceitos elementares para entendê-lo. No contexto da Web, um *hub* pode ser entendido como uma página Web que aponta para muitas outras páginas Web, ou seja, uma seleção de páginas Web que abordam um tema específico. A qualidade de um hub é geralmente determinada pela qualidade das *authorities* para as quais aponta. As *authorities* são páginas Web citadas por vários hubs diferentes, o que significa que sua importância é medida pelo número de ligações que recebem de outras páginas. Normalmente boas *authorities* são fontes confiáveis de informação sobre um determinado tema.

O *PageRank* é um algoritmo de análise de ligações que se baseia no conceito de centralidade do vetor próprio, utilizado pelo motor de busca da Google na medição de importância ou relevância das páginas da internet. A relevância de uma página é medida com

base no valor da informação transmitida por essa página, então as que são consideradas mais valiosas tendem a aparecer no topo dos resultados das pesquisas no Google.

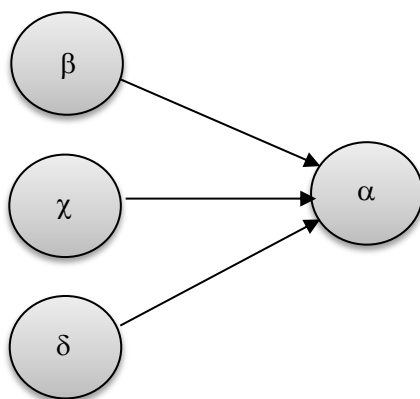
A ideia do algoritmo é que a informação da Web pode ser classificada de acordo com a popularidade da ligação, quanto maior o número de páginas ligadas a uma dada página Web, maior a sua popularidade. No entanto, a relevância dessas ligações também é importante. O *PageRank* mede a importância relativa de um conjunto de páginas Web, tendo por base não apenas a quantidade, mas sobretudo a qualidade das respectivas ligações.

Seja uma rede com somente 4 agentes/*web pages* α , β , χ e δ . Por simplicidade, ignoramos os vínculos entre um agente e ele mesmo e os vínculos múltiplos entre dois agentes.

Em um momento inicial do desenvolvimento do algoritmo, o somatório dos valores de PageRank para todas os nós (que no caso seriam *web pages* da internet) equivalia ao número de páginas da web. Porém, nas versões aperfeiçoadas do PageRank, seus valores passaram a ter uma distribuição probabilística no intervalo entre 0 e 1, expressando a probabilidade de um usuário chegar a uma determinada página, *acessando aleatoriamente os links* (ou vínculos, na linguagem de redes).

A primeira etapa do cálculo do algoritmo (que é um processo iterativo), assume-se que todas as *web pages* possuem o mesmo valor de PageRank. primeiro passo do processo de cálculo iterativo do PageRank, todas as páginas têm o mesmo valor de PageRank. Com somente 4 *web pages*, atribui-se o valor 0,25 para cada página (obviamente o somatório será igual a um).

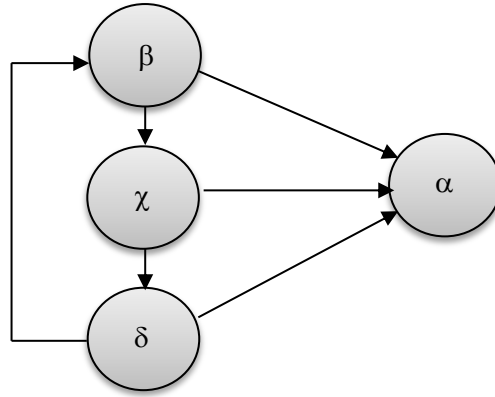
Figura 3 - Todas as *web pages* se conectam apenas com a *web page* α .



Na rede da figura 1, na segunda etapa do processo iterativo, cada conexão “envia” o valor 0,25 para o PageRank. Portanto, temos:

$$\text{Pr}(\alpha) = \text{Pr}(\beta) + \text{Pr}(\chi) + \text{Pr}(\delta) \quad (15)$$

Figura 4 - Todas as *web pages* conectam-se com mais de uma outra *web page*



Na rede da figura 2, na segunda iteração, metade do valor de β é transferido para α (0,125) e a outra metade vai para χ (0,125). Como a página δ conecta-se com 3 páginas, deve-se tomar a terça parte do valor que ela transfere para as demais. O PageRank fica assim:

$$\Pr(\alpha) = \frac{\Pr(\beta)}{2} + \frac{\Pr(\chi)}{1} + \frac{\Pr(\delta)}{3} \quad (16)$$

Isto é, a página referenciada com um link na internet contribui com o PageRank conforme o valor do PageRank da página com este link dividido pelo número de links que a página possui. Se denotarmos por $L(\bullet)$ o número total de links de uma web page, o nosso exemplo com 4 links terá a seguinte expressão.

$$\Pr(\alpha) = \frac{\Pr(\beta)}{L(\beta)} + \frac{\Pr(\chi)}{L(\chi)} + \frac{\Pr(\delta)}{L(\delta)} \quad (17)$$

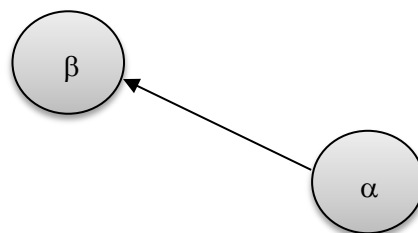
Portanto, podemos generalizar a expressão anterior e deduzir (18).

$$\Pr(\mu) = \sum_{\lambda \in B_\lambda} \frac{\Pr(\lambda)}{L(\lambda)} \quad (18)$$

Portanto, o valor de PageRank de uma web page μ , varia conforme os outros valores de PageRank de cada web page λ contida no conjunto B_λ dividido pelo número de links $L(\lambda)$ que existem em λ . Este conjunto B_λ reúne todas as páginas que possuem links para a página μ .

No caso de web pages sem links, o PageRank possui um processo iterativo bem menos trivial.

Figura 5- Web Page β sem nenhum link



Conforme Austin (2018), quando o algoritmo é aplicado à rede da figura 3, têm-se valores nulos para α e β . A cada iteração, a página β recebe uma parte do PageRank de α . Aqui neste exemplo, β receberia todo o PageRank de α . Mas esta é uma rede muito simples. Se ela fosse complexa e α possuísse links para outras páginas, β receberia apenas uma fração do PageRank de α . Mas neste exemplo β não possui links e não transmite nenhum valor para α ou qualquer outra página (só temos duas). Assim sendo, o PageRank é “enviado” para fora da rede.

2.6. Clusters, agrupamentos hierárquicos e classes de modularidade

Uma das características únicas das redes sociais é que possuem estrutura de comunidade. Normalmente essa propriedade emerge como consequência da heterogeneidade global e local da distribuição das arestas num grafo, então nesse tipo de rede é possível encontrar concentrações elevadas de arestas em determinadas regiões, e baixa concentração de arestas entre essas regiões.

Comunidades, ou *clusters*, são grupos de vértices densamente conectados com ligações esparsas entre eles. De acordo com Newman e Girvan (2004), existem duas linhas de investigação principais na descoberta de comunidades de redes. A primeira teve origem no Campo da Ciência da Computação e é conhecida como partição de grafos, enquanto a segunda foi essencialmente desenvolvida por sociólogos, sendo usualmente referida por *blockmodeling*, *agrupamento hierárquico* ou *detecção de estrutura de comunidades*. O processo básico subjacente aos algoritmos de detecção de comunidade baseia-se na divisão do grafo original num conjunto de subgrafos disjuntos, por via da otimização de uma dada função objetivo. O propósito de ambas as abordagens é descobrir grupos de vértices relacionados e, se possível, definir a respectiva organização hierárquica, tendo por base informação fornecida pela topologia da rede. Isto é geralmente realizado removendo iterativamente as arestas pontes que ligam grupos de vértices, conforme sugerido por Girvan e Newman (2002).

Na vida real é possível encontrar uma variedade de exemplos de grupos coesos, ou comunidades. A sociedade é um ambiente rico em encontrar comunidades, uma vez que as pessoas têm uma tendência natural para formar grupos. Esses grupos podem ser famílias, círculos de amigos, grupos religiosos ou de trabalho, cidades, nações, etc. Se também consideramos grupos formados por empresas, ou consumidores de um dado produto, é possível identificar comunidades com relevância para a área da Economia e da Gestão.

A importância de estudar estas comunidades é intuitiva em domínios como a ARS. Para sublinhar esta importância, Fortunato (2010) afirmou que a análise da suposição estrutural dos nós, em cada comunidade da rede, pode ajudar a identificar *atores centrais*, associados a funções de estabilidade e controle do grupo, bem como *atores intermediários*, que são aqueles que se localizam nas fronteiras das comunidades e desempenham um papel fundamental na disseminação e troca de informações e novas ideias, criando pontes entre comunidades.

O agrupamento hierárquico é uma classe de métodos para detectar clusters, ou grupos. Algoritmos hierárquicos geram estruturas de grupos inseridos dentro de grupos maiores que, por sua vez, se encontram inseridos em grupos ainda maiores, que são representados por dendrogramas que mostram a estrutura multinível da rede. Esses métodos são eficazes na solução de problemas de análise de grupos e problemas semelhantes, como a partição de grafos e a identificação de comunidades.

O agrupamento hierárquico é bastante intuitivo, sendo baseado na definição de semelhança. Primeiro é necessário escolher uma medida de semelhança (ou dissemelhança) para avaliar quão semelhantes são dois nós, de acordo com uma dada propriedade global ou

local¹¹. Em seguida, deve-se calcular a matriz de semelhança entre todos os pares de nós, independente desses nós estarem conectados entre si. Depois, é preciso selecionar um método para agrupar os nós: os *métodos aglomerativos*, que focam nas regiões mais densas da rede ao invés de focar nas ligações das fronteiras da rede¹², ou os *métodos divisivos*, que focam na identificação e remoção das ligações que conectam regiões densamente conectadas a rede, sobretudo as *pontes* e as *pontes locais* (Easley e Kleinberg, 2010)¹³. Conforme a escolha, uma medida de distância é selecionada para calcular a semelhança entre grupos¹⁴. O resultado final desse processo é um dendograma que ilustra a organização dos nós retornada pelo algoritmo hierárquico. Para selecionar os melhores métodos, uma estratégia é calcular o valor de modularidade (Newman e Girvan, 2004) para as comunidades e selecionar o número que maximiza essa função.

A otimização (classe) de modularidade é outro tipo de método utilizado para detectar comunidades em redes. A *modularidade Q* é uma função de qualidade que avalia e mede a importância de uma dada partição da rede em comunidades. Esta função é utilizada para comparar a qualidade das partições e também como uma função objetivo em problemas de otimização. Segundo Newman (2006), a modularidade é representada pela diferença normalizada entre o número de arestas observadas no interior de cada grupo de nós da rede e o número de arestas que seria provável observar no interior desse mesmo grupo numa rede equivalente onde as arestas são geradas aleatoriamente. A modularidade Q é calculada da seguinte forma:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (19)$$

Onde m indica o número de arestas; k_i e k_j representam respectivamente o grau dos vértices i e j ; A_{ij} é a entrada da matriz adjacência que indica que o número de ligações estabelecidas entre os vértices i e j ; $\frac{k_i k_j}{2m}$ representa o número esperado de arestas que deveria existir entre o par de vértices (i, j) ; c_i e c_j denotam os grupos a que os vértices i e j pertencem; e $\delta(c_i, c_j)$ representa o delta de Kronecker.

A modularidade Q pode assumir valores positivos e negativos. Se $Q > 0$, então existe a possibilidade de encontrar estrutura de comunidade em rede. Se Q for um número positivo e elevado, então a respectiva partição tem maior probabilidade de refletir a estrutura de comunidade verdadeira. De acordo com Clauset et al. (2004), uma modularidade que assuma valores superiores ou iguais a 0,3 é um bom indicador da existência de comunidade com significado na rede.

3. As hipóteses de Acemoglu e Autor (2014) e Alon (2018)

Daron Acemoglu é – além de economista conhecido por contribuições à teoria do crescimento e do capital humano, à moderna economia política, à economia do trabalho e à teoria institucionalista - pesquisador na área de redes complexas aplicadas à economia e finanças. Alguns de seus trabalhos utilizaram técnicas avançadas de redes complexas. Entre

¹¹ Entre os exemplos dessa medida estão a semelhança do cosseno, o índice de Jaccard, a distância Euclidiana, a distância de Manhattan e a distância de Hamming entre pares de linhas numa matriz de adjacência.

¹² O algoritmo *Walktrap* (elaborado por Pons e Latapy, 2005) é um exemplo deste tipo de método.

¹³ O conhecido algoritmo de Girvan e Newman (2002) é um exemplo bastante citado deste método.

¹⁴ Alguns exemplos dessa medida: o *single linkage* (ou vizinho mais próximo), o *complete linkage* (ou vizinho mais afastado) e o *método de Ward*.

os quais, podemos citar Acemoglu, Daleh, Lobel e Ozdaglar (2008), Acemoglu, Carvalho, Ozdaglar e Tabaz-Salehi (2012) e Acemoglu, Ozdaglar e Tabaz-Salehi (2013).

Para Acemoglu e Autor (2014), o debate sobre os incentivos para o investimento em capital humano leva à identificação das fontes de diferenças de capital humano. Para ambos, tais fontes residem em cinco fatores, dos quais destacaremos dois:

- (1) Escolaridade: este é o ponto fundamental deste artigo e da maioria da literatura empírica sobre capital humano. Isto porque ela é a variável mais facilmente obtível dos investimentos realizados em capital humano. Todavia, há que se considerar que o R^2 das regressões de retornos que controlam a escolaridade é um tanto pequeno. E isso levanta a hipótese que de as diferenças de escolaridade importam para apenas uma pequena parcela das diferenças dos retornos. Portanto, Acemoglu e Autor (2014) consideram que há muito mais fatores que afetam o capital humano do que somente a escolaridade. Apesar disso, ambos admitem que as pesquisas sobre escolaridade são muito informativas quando se leva em conta que os mesmos fatores que influenciam nos investimentos escolares também tendem a afetar os investimentos não-escolares (que são de difícil mensuração). *Então, é possível inferir padrões úteis de escolaridade que podem existir também nos investimentos e incentivos não-escolares.*
- (2) Qualidade escolar e investimentos/incentivos não escolares: indivíduos de escolas diferentes podem ter níveis distintos de capital humano em razão da assimetria de qualidade de ensino existente entre as duas. Ou indivíduos que cursaram a mesma quantidade de anos em uma mesma escola podem também possuir diferentes níveis de capital humano por causa de *investimentos ou incentivos em outros tipos de capital humano que não o ensino formal* (um deles pode ter obtido maior nível de educação por causa do *learning-by-doing*, *variável diretamente afetada pelos incentivos que uma remuneração mais alta podem proporcionar*, ou por ter feito sido submetido à circunstâncias que lhe fizeram aprimorar uma habilidade como escrever ou falar melhor). Há na literatura um debate sobre a importância dessas habilidades “não observáveis” e seus impactos na estrutura salarial. A dificuldade reside na aferição desses componentes do capital humano, uma vez que não há dados confiáveis sobre eles.

Alon (2018) construiu um modelo de longo prazo de acumulação de capital humano motivado pelos incrementos no grau de especialização da força de trabalho. Em seu modelo os indivíduos aumentam a eficiência do tempo dedicado para a aquisição de capital humano, concentrando os investimentos em conjuntos mais restritos e específicos de habilidades (que podem ser obtidas de modo formal, pelos treinamentos oferecidos pelas empresas ou pelo *learning by doing*). Este autor analisou a evolução dos currículos secundários e pós-secundários nos Estados Unidos no período de 1870-2000, mostrando estas alterações nos níveis de especialização. Também obteve resultados que apontaram que a especialização foi responsável por cerca de 29% do aumento do prêmio pago pela habilidade e por 25-30% do aumento do nível educacional no período de 1965-2005. O efeito sobre o prêmio pago pela habilidade é em grande parte devido a um declínio na especialização no ensino médio, onde a formação profissional foi substituída por requisitos de graduação acadêmica. Assim, ele sugere a adoção de uma política educacional que amplie o espaço dedicado ao treinamento profissional especializado em escolas secundárias, considerando que esta medida pode ser útil para reduzir a desigualdade de renda.

Nesse sentido, essa dissertação, contribui com uma análise para o caso brasileiro de como se configura o perfil de remunerações das ocupações na indústria intensiva em PD&I, relacionando este perfil com o processo de acumulação de capital formal e chegando a

resultados que apontam a relevância do *learning by doing* (hipótese de Acemoglu e Autor, 2014). Como consideramos os níveis educacionais que vão do limite inferior do ensino médio (para verificar a hipótese de Alon, 2018) até o limite superior do doutorado (passando pelos níveis de ensino superior incompleto, completo e mestrado), os resultados deste trabalho permitem compreender, considerando cada uma das ocupações distribuídas por níveis educacionais, em que medida as remunerações são mecanismos de incentivo relevantes para a aquisição tanto de ensino formal quanto de *learning-by-doing*.

4. Fontes, metodologia, softwares e distribuição de layout

Na primeira subseção, estão as fontes dos dados. Na segunda, a metodologia utilizada, isto é, o algoritmo de Yifan Hu Proporcional utilizado para gerar as redes complexas e o cálculo dos pesos da rede complexa com arestas ponderadas. Na terceira, as métricas e as respectivas redes complexas são analisadas.

4.1. Fontes dos dados

Os dados referentes às médias salariais das ocupações, ao nível dos grupos e não dos indivíduos, foram obtidos do Cadastro Geral de Empregados e Desempregados da RAIS – MTE¹⁵

O software utilizado para a estimação das métricas de mecânica estatística da rede e também para a elaboração da rede complexa com o algoritmo de distribuição de Yifan Hu proporcional foi o Gephi 0.9.2¹⁶, que é um software livre muito utilizado para análises de redes complexas, teoria dos grafos, mecânica estatística e redes sociais.

4.1.1. Cálculo dos pesos

Dado que a principal maneira de fazer P&D nas empresas nacionais é investir na contratação de capital humano cada vez mais qualificado, analisamos as diferenças salariais entre as profissões consideradas como “integralmente alocadas em P&D” conforme a classificação proposta por Gusso (2006) e encontramos evidências de uma relação entre os níveis de capital humano (com a *proxy* do grau de instrução) e a variação salarial nestas profissões que fazem parte do complexo industrial brasileiro com capacidade de inovação.

A hipótese principal que verificamos é se o aumento da formação de capital humano dos trabalhadores implica em remuneração maior, *coeteris paribus*. Para tanto desenvolvemos quatro redes complexas não dinâmicas, comparando as diferenças salariais entre as profissões, graus de instrução e nível salarial que os indivíduos que nelas atuam, variando tanto os graus de instrução quanto as ocupações.

Para isso, adotamos este índice que calcula os “pesos” (ou “ponderações”) das arestas da rede complexa: $\frac{w_{h,j}}{w_{l,m}}$ (19) Onde, no denominador, temos: $w_{h,j}$ como o peso da aresta da remuneração de h , que representa a ocupação conquistada após o aumento do grau de instrução (j é este grau de instrução maior). E no denominador temos $w_{l,m}$ como sendo a ocupação do trabalhador quando este tinha o grau de instrução imediatamente mais baixo (representado por m).

Por óbvio, temos que $h > l$ e $j > m$ e, com o exposto, apresentam-se três situações possíveis:

¹⁵ http://bi.mte.gov.br/bgcaged/caged_rais_vinculo_id/login.php

¹⁶ The Open Graph Viz Platform - <https://gephi.org/>

- 1) $h = l, j = m$, neste caso não houve aumento de remuneração, nem nova ocupação nem melhora grau de instrução formal.
- 2) $h \neq l, j = m$, neste caso houve aumento da remuneração na nova ocupação, mas não houve melhora do grau de instrução. O trabalhador foi promovido sem melhorar sua instrução formal.
- 3) $h \neq l, j \neq m$, neste caso houve aumento da remuneração na nova ocupação e também melhora do grau de instrução formal.

4.2. Distribuições de Yifan Hu e Yifan Hu Proporcional

Conforme o tutorial de layouts do programa Gephi (2011)¹⁷, o algoritmo de layout de Yifan Hu multinível (Hu, 2006) é bastante rápido e gera grandes gráficos com boa qualidade visual e de processamento. Combina um modelo de força direcionada com uma técnica simples de grafos (em um algoritmo multinível) que reduz a complexidade dos mesmos. As forças repulsivas em um nó de um *cluster* de nós distantes são aproximadas pelo procedimento de Barnes-Hut, que os trata como um “super nó” (*super-node*). Ao contrário de outros algoritmos do Gephi, o de Yifan Hu é suspenso automaticamente após ser gerado. Esse algoritmo primeiro agrupa os nós e depois aplica uma lógica de força aos mesmos. Assim, como qualquer método de layout direcionados pela força (pesos das arestas) seus desenhos são baseados nas semelhanças e/ou diferenças nos dados. Os padrões podem ser ajustados para colocar mais ênfase na independência entre um e outro dos agentes (vértices) individuais. Ou então levando em conta a proximidade entre eles. O algoritmo de Yifan Hu multinível é bastante intuitivo, do ponto de vista visual, para tratar de dados que se referem a transições, hierarquias ou níveis. No caso desse trabalho, tratamos das transições entre os níveis de instrução (do menor para o maior), ocupações e remunerações (pesos ou forças das arestas) aproveitando a conveniência visual deste algoritmo. Nas palavras do próprio Yifan Hu (2006):

“Nós propomos um algoritmo de desenho de grafos que seja eficiente e de alta qualidade. Este algoritmo combina uma abordagem multinível, que efetivamente supera os mínimos locais, com a técnica de árvore óctupla de Barne e Hut, que aproxima de forma eficiente as forças de curto e de longo alcance. Nossos resultados numéricos mostram que o algoritmo é comparável em velocidade ao algoritmo de desenho de grafo multinível altamente eficiente de Walshaw e ainda gera melhores resultados para alguns dos problemas difíceis. Além disso, um esquema de amenização adaptativa para os algoritmos dirigidos pela força e um modelo de força repulsivo geral são propostos.” (Hu, 2006)¹⁸.

4.3. Por que utilizar redes complexas para analisar capital humano?

Já mencionamos que há necessidade na literatura de análises mais detalhada (e desagregadas) sobre os efeitos do capital humano no crescimento salarial. Citamos Tchernis (201), que considera ser necessária a elaboração de novas medidas de capital humano que reflitam competências transferíveis entre alguns, mas não todos, os postos de trabalho. Portanto, nossa contribuição para a literatura do capital humano reside na produção destas medidas a partir de uma metodologia inovadora e ainda pouco utilizada na literatura

¹⁷ <https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>

¹⁸ O algoritmo de Yifan Hu Proporcional é similar ao algoritmo Yifan Hu. A diferença é que o primeiro estabelece um deslocamento proporcional para distribuir os vértices na área do grafo. Em termos de velocidade de cálculo e precisão, não há muita diferença entre os dois.

econômica: a mecânica estatística de redes complexas, cujas técnicas são derivadas da teoria dos grafos e da ciência de redes (*network science*).

Usualmente, as técnicas econométricas utilizadas no estudo da teoria do capital humano são derivadas da estimação de parâmetros por meio da análise de séries temporais e da econometria de dados em painel e em *cross-section*. Nesse sentido, usualmente há também, além dos parâmetros, a estimação de relações de causalidade no sentido Granger-Newbold, o que faz com que estes trabalhos se baseiem em técnicas paramétricas.

A mecânica estatística de redes complexas é uma metodologia não-paramétrica. Com isto, as direções das relações entre variações salariais e capital humano, que conferem o sentido das arestas da rede, é dada pelo peso das referidas variações. Assim, utilizamos uma rede ponderada ou valorada, isto é, uma rede cujas ligações (arestas) são valoradas de acordo com as magnitudes das variações salariais. E os sentidos das “flechas” destas arestas fluem dos níveis de estudo menores para os maiores. Ou, pela ordem: do ensino médio para o nível superior incompleto, deste para o nível superior completo e, em seguida, para o mestrado e doutorado. Com isto, temos uma descrição não-paramétrica e uma análise puramente baseada nas medidas estatísticas da rede, as quais são calculadas com base no que a amostra realmente nos traz, sem nenhum tipo de estimação de parâmetros.

Ainda assim, consideramos que esta metodologia não pretende ser substituta das técnicas econométricas cujos usos já estão consolidados na literatura do capital humano. Ao contrário, defende-se o emprego da abordagem de redes como uma ferramenta bastante útil para a produção de evidências complementares que corroborem as conclusões dos bons artigos já publicados sobre o assunto. É dentro dessa perspectiva que a utilizamos neste trabalho.

5. Resultados

Para a análise dos resultados, utilizaremos a tabela 1, baseada em Gusso (2006) com os códigos referentes à CBO 02: (Classificação Brasileira de Ocupações do Ministério do Trabalho e Emprego).

Tabela 1 – Classificação Brasileira de Ocupações do MTE.

Grupo ocupacional		Códigos (CBO 02)
Pesquisadores	203	Pesquisadores
Engenheiros	202	Engenheiros mecatrônicos
	214	Engenheiros Civis etc.
Diretores e gerentes de PD&I	1237	Diretores de PD&I
	1426	Gerentes de PD&I
Profissionais científicos	201	Bioteecnologistas, geneticistas, pesquisadores em metrologia e especialistas em calibrações meteorológicas
	211	Matemáticos, estatísticos e afins
	212	Profissionais de informática
	213	Físicos, químicos e afins
	221	Biólogos e afins

Fonte: Classificação Brasileira de Ocupações do Ministério do Trabalho e Emprego.

A primeira subseção trata da análise ao nível da rede e busca detectar as comunidades existentes e as saídas e entradas das ocupações nestas comunidades. A métrica utilizada é o processo de otimização por classe de modularidade.

A segunda subseção aborda as métricas de mecânica estatística computadas ao nível dos agentes. Inicialmente, identificamos os *hubs* existentes (primeiro passo para verificar quais vértices de ocupação/nível de instrução beneficiaram-se de maiores variações salariais). Em seguida, calculamos os graus ponderados (médios e de entrada), e as centralidades dos autovetores (*eigenvector centralities*).

Em ambas as seções, testamos três hipóteses: (i) se existem na indústria brasileira evidências de uma clara relação entre grau de instrução e variação salarial nos cargos das empresas industriais com maior capacidade inovativa e que dependem fundamentalmente dos investimentos na formação de capital humano intensivo em P&D; (ii) se existem evidências que corroborem a análise de Acemoglu e Autor (2014) de que os indivíduos que cursaram a mesma quantidade de anos em uma mesma escola podem ter níveis assimétricos de capital humano por causa de incentivos salariais que tornam atraente o processo de *learning-by-doing*?; e (iii) se o aumento do grau de especialização dos trabalhadores brasileiros em indústrias intensivas em PD&I aumentou o prêmio pago pelas suas habilidades, conforme preconizou Alon (2018) para os Estados Unidos no período de 1965-2005.

Para testá-las, exploramos as estatísticas derivadas de quatro redes complexas baseadas no algoritmo de Yifan Hu proporcional (Hu, 2006) para os triênios de 2006-08, 2009-2011 e 2012-2014 e uma para o biênio de 2015-2016. Com elas, comparamos a evolução das variações salariais resultantes do investimento em capital humano avançado e do *learning-by-doing* entre as ocupações.

5.1. Análise ao nível da rede - Classes de modularidade (CMs)

Esta subseção calcula as classes de modularidade para detectar os sub-digrafos (doravante chamados de *clusters*) das redes e a entrada/saída das ocupações destes *clusters*.

Em relação às classes de modularidade, em todos os períodos identificaram-se três *clusters* que, todavia, mostraram algumas alterações a cada transição de período. Analisaremos a natureza de cada *cluster* e as suas transições, pois elas revelam informações úteis sobre aspectos mais desagregados da rede, ao nível de seus vértices (tabela 2).

A classe de modularidade aponta os vértices capazes de formar *clusters* dentro da rede e mede também a intensidade das ligações dentro do *cluster* em questão. As modificações no cluster 1, mostram uma variação significativa nas remunerações das ocupações de EM (ensino médio).

Tabela 2 – Alterações nos dois clusters do mercado de trabalho das indústrias brasileiras intensivas em PD&I

Períodos	Modificações no cluster 1 (CM = 0)	Modificações no cluster 2 (CM = 1)	Modificações no cluster 3 (CM = 2)
2006-08	BiotGen (SI), Pesq (SI), Info (SI), ECiv (SI), Bio (SI), Pesq (SC), Mat (SC), FiQm (SC), ECiv (SC), DP&D (SC) e Ger (SC).	<u>Todas as ocupações de EM + Mat (SI), FisQuim (SI), Mecat (SI), Ger (SI) e DP&D (SI).</u>	Todas as ocupações com M e D + Bio (SC), Info (SC) e Mecat (SC).
2009-11	<u>Todas as ocupações de EM + Mecat (SI), Mat (SI), ECiv (SI), DP&D (SI) e Ger (SI).</u>	<u>BiotGen (SI), Pesq (SI), Info (SI), FiQm (SI), Bio (SI), BiotGen (SC), FiQm (SC), ECiv (SC), DP&D (SC) e Ger (SC).</u>	Todas as ocupações com M e D + Bio (SC), Info (SC), Mat (SC) , Mecat (SC) e Pesq (SC)
2012-14	<u>BiotGen (SI), Info (SI), FiQm (SI), Bio (SI), BiotGen (SC), Pesq (SC).</u>	<u>Todas as ocupações do EM + Mecat (SI), Mat (SI), ECiv (SI), DP&D (SI) e Ger (SI).</u>	Todas as ocupações com M e D + Bio (SC), Info (SC), Mat (SC) e

	<u>FiQm (SC), ECiv (SC), DP&D (SC) e Ger (SC).</u>	<u>Mecat (SC).</u>
2015-16	Todas as ocupações do EM, Mat (SI), ECiv (SI), DP&D (SI) e Ger (SI).	Todas as ocupações com M e D + Bio (SC), Info (SC) e Mat (SC).
	BiotGen (SI), Mecat (SI), Pesq (SI), Info (SI), FiQm (SI), Bio (SI), BiotGen (SC), Mecat (SC), Pesq (SC), FiQm (SC), ECiv (SC), DP&D (SC) e Ger (SC)	

Fonte: estatísticas dos autores a partir das redes complexas elaboradas no Gephi 0.9.2.

Obs: Os itens sublinhados são os que deixaram o cluster no período seguinte. Os itens em negrito são os que ingressaram no cluster no período atual.

Percebe-se que nos triênios de 2006-08 e 2012-14, tais ocupações fizeram parte do *cluster*, o que não ocorreu no triênio de 2009-11 e no biênio 2014-15. No quarto trimestre de 2008 e no primeiro semestre de 2009, os efeitos da Grande Recessão foram mais fortes na economia brasileira. De modo análogo, o PIB brasileiro caiu 9% em 2014-2017, o que pode ter afetado a remuneração¹⁹ das ocupações de EM desse *cluster* 1.

Tabela 3 – Classes de modularidade e clusters das 4 redes complexas do mercado de trabalho de ocupações intensivas em PD&I no Brasil: 2006-2016

Código	Ocupação	2006-08 CM1	2009-11 CM2	2012-14 CM3	2015-16 CM4	2006-16 CM média
201EM	BiotGen	1	0	1	0	0,50
202EM	Mecat	1	0	1	0	0,50
203EM	Pesq	1	0	1	0	0,50
211EM	Mat	1	0	1	0	0,50
212EM	Info	1	0	1	0	0,50
213EM	FiQm	1	0	1	0	0,50
214EM	ECiv	1	0	1	0	0,50
221EM	Bio	1	0	1	0	0,50
1237EM	DP&D	1	0	1	0	0,50
1426EM	Ger	1	0	1	0	0,50
201SI	BiotGen	0	1	0	1	0,50
202SI	Mecat	1	0	1	1	<u>0,75</u>
203SI	Pesq	0	1	0	1	0,50
211SI	Mat	1	0	1	0	0,50
212SI	Info	0	1	0	1	0,50
213SI	FiQm	1	1	0	1	<u>0,75</u>
214SI	ECiv	0	0	1	0	0,25
221SI	Bio	0	1	0	1	0,50
1237SI	DP&D	1	0	1	0	0,50
1426SI	Ger	1	0	1	0	0,50
201SC	BiotGen	2	1	0	1	<u>1,00</u>
202SC	Mecat	2	2	2	1	<u>1,75</u>
203SC	Pesq	0	2	0	1	<u>0,75</u>
211SC	Mat	0	2	2	2	<u>1,50</u>
212SC	Info	2	2	2	2	2,00
213SC	FiQm	0	1	0	1	0,50

¹⁹ E até mesmo o nível de emprego, embora não tenhamos os dados sobre ele.

214SC	ECiv	0	1	0	1	0,50
221SC	Bio	2	2	2	2	2,00
1237SC	DP&D	0	1	0	1	0,50
1426SC	Ger	0	1	0	1	0,50
201M	BiotGen	2	2	2	2	2,00
202M	Mecat	2	2	2	2	2,00
203M	Pesq	2	2	2	2	2,00
211M	Mat	2	2	2	2	2,00
212M	Info	2	2	2	2	2,00
213M	FiQm	2	2	2	2	2,00
214M	ECiv	2	2	2	2	2,00
221M	Bio	2	2	2	2	2,00
1237M	DP&D	2	2	2	2	2,00
1426M	Ger	2	2	2	2	2,00
201D	BiotGen	2	2	2	2	2,00
202D	Mecat	2	2	2	2	2,00
203D	Pesq	2	2	2	2	2,00
211D	Mat	2	2	2	2	2,00
212D	Info	2	2	2	2	2,00
213D	FiQm	2	2	2	2	2,00
214D	ECiv	2	2	2	2	2,00
221D	Bio	2	2	2	2	2,00
1237D	DP&D	2	2	2	2	2,00
1426D	Ger	2	2	2	2	2,00

Fonte: estatísticas das redes complexas estimadas pelos autores no Gephi 0.9.2.

Obs: os *hubs* estão realçados em cinza claro e seus graus ponderados médios estão em negrito.

Outra hipótese, é que pode ter ocorrido uma natural substituição decorrente da evolução da carreira desses trabalhadores associada à mobilidade desse mercado trabalho. Assim, os trabalhadores com EM (ensino médio) no primeiro triênio podem ter obtido o grau de SC (superior completo) no segundo e/ou trocado de emprego, o que pode ter impulsionado a demanda por novas contratações no terceiro triênio.

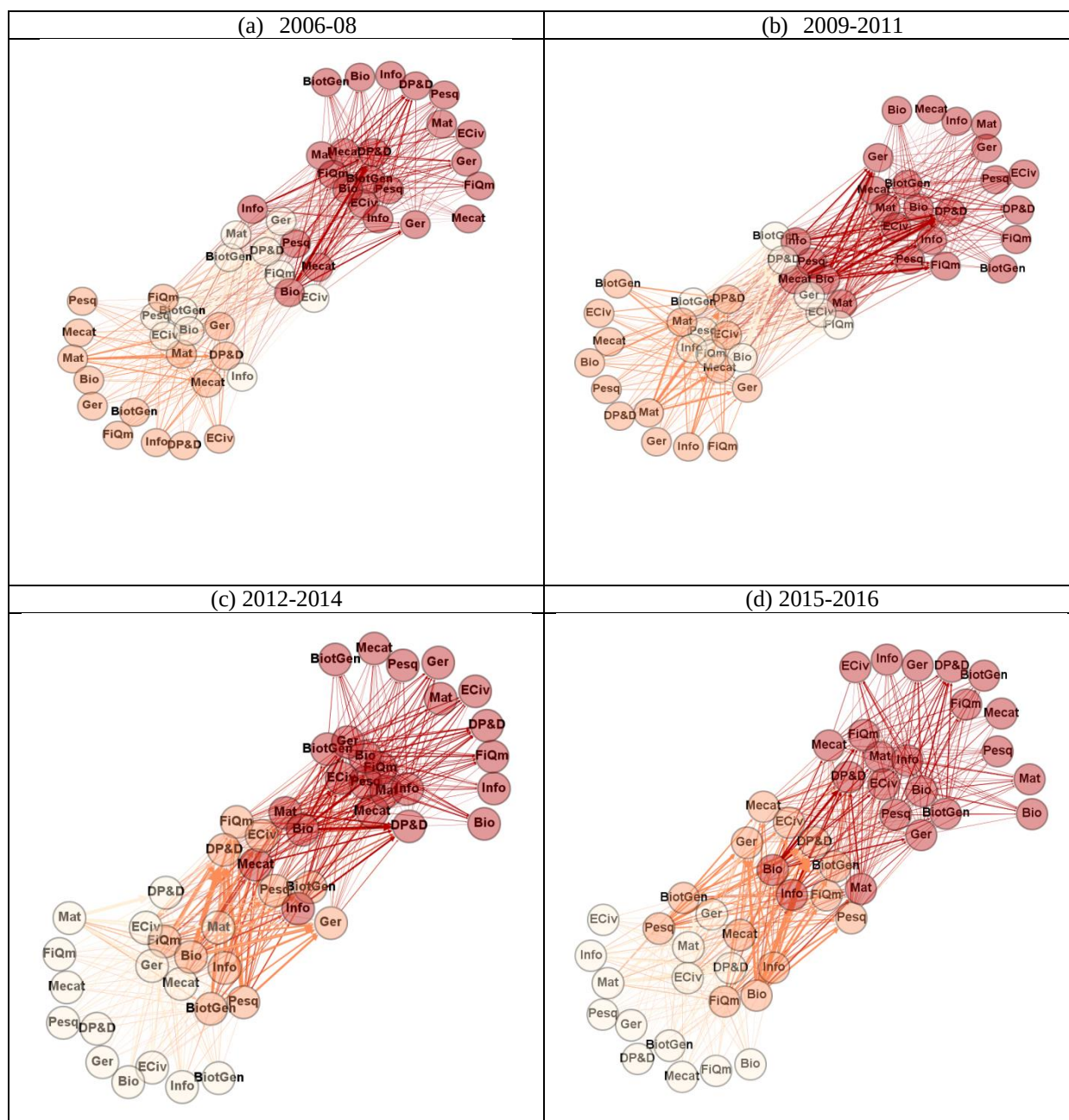
Conforme dados do Censo da Educação Superior do MEC e Inep, em 2016, 34.366 cursos de graduação foram ofertados em 2.407 instituições de educação superior (IES) no Brasil para um total de 8.052.254 estudantes matriculados. Segundo os dados do Censo, as 197 universidades existentes no país equivalem a 8,2% do total de IES, mas concentram 53,7% das matrículas em cursos de graduação.

Em 2016, o número de matrículas na educação superior (graduação e sequencial) continuou crescendo, mas essa tendência desacelerou quando se comparam os dados com os de anos anteriores, pois entre 2006 e 2016, houve aumento de 62,8%, com uma média anual de 5% de crescimento. Porém, em relação a 2015, a variação positiva foi de apenas 0,2%.

Essa hipótese é corroborada pela maior presença de ocupações com SI (superior incompleto) no segundo triênio e no biênio 2015-16, no qual também aumentaram bastante as ocupações com SC (superior completo).

A figura 1, delineada a partir da distribuição de layout de Yifan-Hu proporcional, ajuda a visualizar a dinâmica de mobilidade salarial e ocupacional expressa na tabela 3. As redes da figura 1 sugerem que os graus de instrução M e D (mestrado e doutorado) podem formar um único cluster, que é justamente o mais forte da rede. Nele também estariam os profissionais com ensino superior completo identificados como “Biólogos e afins - (212SC)”, “Profissionais de Informática – 221(SC)” e “Matemáticos, Estatísticos e afins – (211SC)”.

Figura 6 – Evolução temporal das redes conforme suas classes de modularidade e pesos das arestas



Fonte: Redes complexas elaboradas pelos autores no Gephi 0.9.2.

Legenda 6 - cores das redes

	Classes de modularidade 2 (cluster 3)
	Classes de modularidade 2 (cluster 2)
	Classes de modularidade 2 (cluster 1)

Outro resultado é que nos períodos ímpares (1 e 3), os empregados com EM (ensino médio) mostram evidências muito fracas de formação de um *cluster*. Já nos períodos pares (2 e 4), eles se agrupam mais fortemente. O mesmo vale para os Diretores e Gerentes de P&D com superior incompleto (1237SI e 1426SI).

Por fim, para as ocupações com SI e SC (superior incompleto e completo) não há agrupamento definido. O que pode se dizer é que há uma alternância de entrada para SI entre os graus de modularidade 2 e 1, e os de SC entre 0 e 1.

5.2. Análise ao nível dos vértices/agentes: *hubs*, graus ponderados, PageRanks e centralidade dos autovetores

Esta subseção estima as medidas de mecânica estatística ao nível dos agentes (ou vértices). Primeiramente, mostramos os *hubs* existentes. Logo após, calculamos os graus ponderados (médios e de entrada), os PageRanks e as centralidades dos autovetores (*eigenvector centralities*).

5.2.1. Existência de hubs

Um *hub* é um vértice em uma rede com uma série de conexões (ou arestas) cujo número ou o peso é muito acima da média dos outros vértices da rede. São, portanto, agentes com graus médios (ponderados ou não) elevados. A emergência de *hubs* decorre de uma propriedade de redes livres de escala (*scale free*). Ainda que os *hubs* não possam ser vistos em redes aleatórias, eles devem emergir de redes livres de escala. Tal emergência é fruto da distribuição de leis de potência. Eles possuem um impacto significativo na topologia da rede e são encontrados em redes reais como a internet, redes biológicas e as redes de neurônios.

É possível observar, pelos valores de graus ponderados médios acima de 28, que nas 4 redes há 26 *hubs* definidos, nos quais se encontram todas as ocupações com SI (superior incompleto) e SC (superior completo), metade das ocupações com M (mestrado) e uma única ocupação de EM (trabalhadores com ensino médio que estão em ocupações que exigem conhecimento matemático ou técnico). A existência desses hubs evidencia, de forma direta e utilizando mecânica estatística (que possui caráter não-paramétrico), a influência do acúmulo de capital humano na obtenção de maiores variações salariais nas ocupações marcadas com cor cinza e em negrito na tabela 4. Os graus ponderados médios são o resultado da soma dos graus de entrada e de saída de cada agente ou ocupação. Outro ponto é que o valor de corte do grau médio ponderado adotado como referência para classificar um *hub*, corresponde a um nível equivalente ao valor médio destes graus (que é de 27,976). Os agentes que atingem o nível médio ou têm valores acima da média, são incluídos em *hub* (tabela 4). Assim, podemos afirmar que quando uma ocupação/nível de instrução faz parte de um *hub*, é porque seus trabalhadores receberam uma variação salarial acima da média, o que reflete a influência mencionada. Finalmente, não há nenhuma ocupação com D que seja um *hub*, o que reforça a conclusão de que a obtenção do mestrado possui maior relevância, em termos de acréscimo salarial, do que o doutoramento nas contratações dos setores da indústria brasileira intensiva em PD&I.

Tabela 4 – Graus ponderados médios e *hubs* das 4 redes complexas do mercado de trabalho de ocupações intensivas em PD&I no Brasil: 2006-2016

Código	Ocupação	2006-08 GPs 1	2009-11 GPs 2	2012-14 GPs 3	2015-16 GPs 4	Hubs Média dos GPs
201EM	BiotGen	17,985	17,754	13,910	19,207	17,214
202EM	Mecat	11,942	14,076	14,007	15,096	13,780
203EM	Pesq	9,312	10,977	10,922	11,772	10,746
211EM	Mat	33,594	39,598	39,402	42,466	38,765
212EM	Info	18,379	20,443	17,984	26,241	20,762
213EM	FiQm	7,805	13,024	10,730	11,564	10,781
214EM	ECiv	14,760	11,437	13,733	14,801	13,683
221EM	Bio	12,914	15,222	15,147	16,324	14,902
1237EM	DP&D	6,953	8,011	8,470	10,378	8,453
1426EM	Ger	8,715	9,394	7,320	11,051	9,120

201SI	BiotGen	37,816	35,872	38,780	42,863	38,833
202SI	Mecat	31,012	35,008	37,391	44,543	36,988
203SI	Pesq	33,909	39,569	45,787	52,656	42,980
211SI	Mat	32,171	33,661	36,049	40,665	35,637
212SI	Info	32,952	37,663	41,725	46,233	39,643
213SI	FiQm	31,541	35,914	40,745	46,649	38,712
214SI	ECiv	32,505	33,495	36,593	41,529	36,030
221SI	Bio	43,177	52,159	61,971	71,685	57,248
1237SI	DP&D	35,864	40,986	41,915	52,825	42,898
1426SI	Ger	31,362	34,015	37,824	44,367	36,892
201SC	BiotGen	31,845	39,211	39,737	43,710	38,626
202SC	Mecat	32,480	40,712	38,097	41,676	38,242
203SC	Pesq	31,877	38,370	38,775	42,682	37,926
211SC	Mat	31,916	38,425	38,070	41,144	37,389
212SC	Info	31,855	39,929	38,261	41,062	37,777
213SC	FiQm	32,152	38,510	39,756	44,039	38,614
214SC	ECiv	32,145	38,439	39,820	43,493	38,474
221SC	Bio	39,654	51,303	44,443	47,125	45,631
1237SC	DP&D	50,510	56,234	66,409	75,703	62,214
1426SC	Ger	39,358	43,025	47,364	51,126	45,218
201M	BiotGen	26,692	25,960	27,185	28,842	27,170
202M	Mecat	29,639	26,234	26,946	28,060	27,720
203M	Pesq	25,512	26,749	26,684	28,556	26,875
211M	Mat	25,511	27,326	26,475	28,152	26,866
212M	Info	25,511	26,428	26,694	28,468	26,775
213M	FiQm	25,511	31,000	29,634	30,637	29,196
214M	ECiv	25,631	29,551	27,840	29,045	28,017
221M	Bio	27,413	25,884	28,854	31,742	28,473
1237M	DP&D	46,207	43,845	38,541	41,462	42,514
1426M	Ger	28,908	33,385	28,189	29,273	29,939
201D	BiotGen	10,400	7,707	9,518	10,458	9,521
202D	Mecat	4,836	1,819	6,688	8,139	5,370
203D	Pesq	12,143	8,337	10,699	10,573	10,438
211D	Mat	14,917	9,213	11,898	13,708	12,434
212D	Info	12,332	6,951	8,988	9,600	9,468
213D	FiQm	11,428	10,043	15,778	16,628	13,469
214D	ECiv	12,165	9,092	14,003	14,843	12,526
221D	Bio	12,111	9,344	11,829	12,025	11,327
1237D	DP&D	24,282	16,130	19,329	25,596	21,334
1426D	Ger	18,078	11,972	14,737	16,022	15,202

Fonte: estatísticas das redes complexas estimadas pelos autores no Gephi 0.9.2.

Obs: os *hubs* estão realçados em cinza claro e seus graus ponderados médios estão em negrito.

5.2.2. Graus e graus ponderados

O resultado dos graus de entrada, saída e grau completo (tabela 8, no anexo) são um tanto óbvios: os níveis superior (completo e incompleto) e o mestrado possuem graus 20, enquanto os outros dois (ensino médio e doutorado) possuem graus 10. Isso ocorre porque não consideramos o ensino fundamental (um nível abaixo do ensino médio) e porque não há grau

superior ao nível de doutorado. Se considerássemos o ensino fundamental, os nós com ocupação de ensino médio teriam grau 20. E se houvesse um nível de ensino superior ao de doutorado, teríamos também grau 20 para essas ocupações.

Quanto aos graus ponderados (tabela 5), a obviedade dos resultados dos graus simples da subseção anterior desaparece. Como as arestas estão ponderadas (no sentido de “valoradas” ou “pesadas”) pelo índice proposto na subseção 3.10, os graus mais pesados refletem a capacidade remuneratória das ocupações e níveis de instrução.

Assim, verifica-se o crescimento médio dos salários de acordo com o aumento do grau de instrução, evidenciando que o investimento em capital humano proporciona aumentos de salários que vão de 7,88% (quando ocorre o ingresso nos cursos superiores relacionados à PD&I, isto é, no nível SI – superior incompleto) até o pico de 51,58% (quando um trabalhador obtém um doutorado nesses cursos).

Tabela 5 – Crescimento médio e desvios-padrão dos graus de entrada ponderados nos triênios de 2006-08, 2009-11, 2012-14 e no biênio de 2015-16.

Períodos	SI		SC		M		D	
	Cresc. Médio	Desvpad	Cresc. Médio	Desvpad	Cresc. Médio	Desvpad	Cresc. Médio	Desvpad
06-08 para 09-11	13,42%	25,30%	9,81%	12,39%	47,21%	33,35%	-32,91%	13,57%
09-11 para 12-14	-4,09%	15,80%	23,17%	11,50%	-21,08%	11,03%	55,87%	75,53%
12-14 para 15-16	14,32%	11,16%	14,19%	3,88%	1,41%	7,46%	10,66%	10,03%
Período total 06-16	7,88%	-	15,72%	-	9,18%	-	11,21%	-
Variação acumulada	7,88%		24,84%		36,30%		51,58%	

Fonte: estatísticas dos autores a partir das redes complexas elaboradas no Gephi 0.9.2.

Também é possível observar que, quanto maiores os crescimentos médios das remunerações, maior tende a ser o aumento dos seus desvios-padrão (os valores estão em negrito, na tabela 5).

Analisando os graus ponderados de entrada (tabela 6), no tocante ao nível superior incompleto, como a nossa análise restringe-se à 4 períodos, o aumento do grau de instrução mostrou um efeito de 13,42 pontos percentuais maior na composição da variação salarial no segundo período em comparação ao primeiro.

Algumas ocupações registraram aumentos e outras quedas nas proporções salariais relativas aos aumentos do grau de instrução de EM (ensino médio) para SI (superior incompleto). No 3º período o aumento do grau de instrução passou a impactar 4,09 p.p. a menos do que o período anterior, em média, exceto para Gerentes de P&D (códigos 1426). Por fim, a última comparação indicou novamente um aumento na proporção salarial de 14,32 p.p., e aqui o destaque fica para os Gerentes e Diretores de P&D, que tiveram a maior participação na variação salarial.

Tabela 6 – Graus ponderados de entrada (GPEnts) e variação entre o primeiro e o último períodos nos triênios de 2006-08, 2009-11, 2012-14 e no biênio de 2015-16.

Código	Ocupação	2006 a 2008 GPEnt1	2009 a 2011 GPEnt2	2011 a 2014 GPEnt3	2015 a 2016 GPEnt4	Médias dos GPEnts	Var. (%)
201SI	BiotGen	7,804	11,441	12,226	14,656	11,532	87,8%
202SI	Mecat	17,999	22,687	13,710	13,182	16,894	-26,8%
203SI	Pesq	9,659	9,206	8,771	9,602	9,309	-0,6%
211SI	Mat	21,038	18,767	18,492	20,245	19,636	-3,8%
212SI	Info	10,374	10,165	10,347	12,118	10,751	16,8%
213SI	FiQm	11,965	11,405	10,866	11,895	11,533	-0,6%

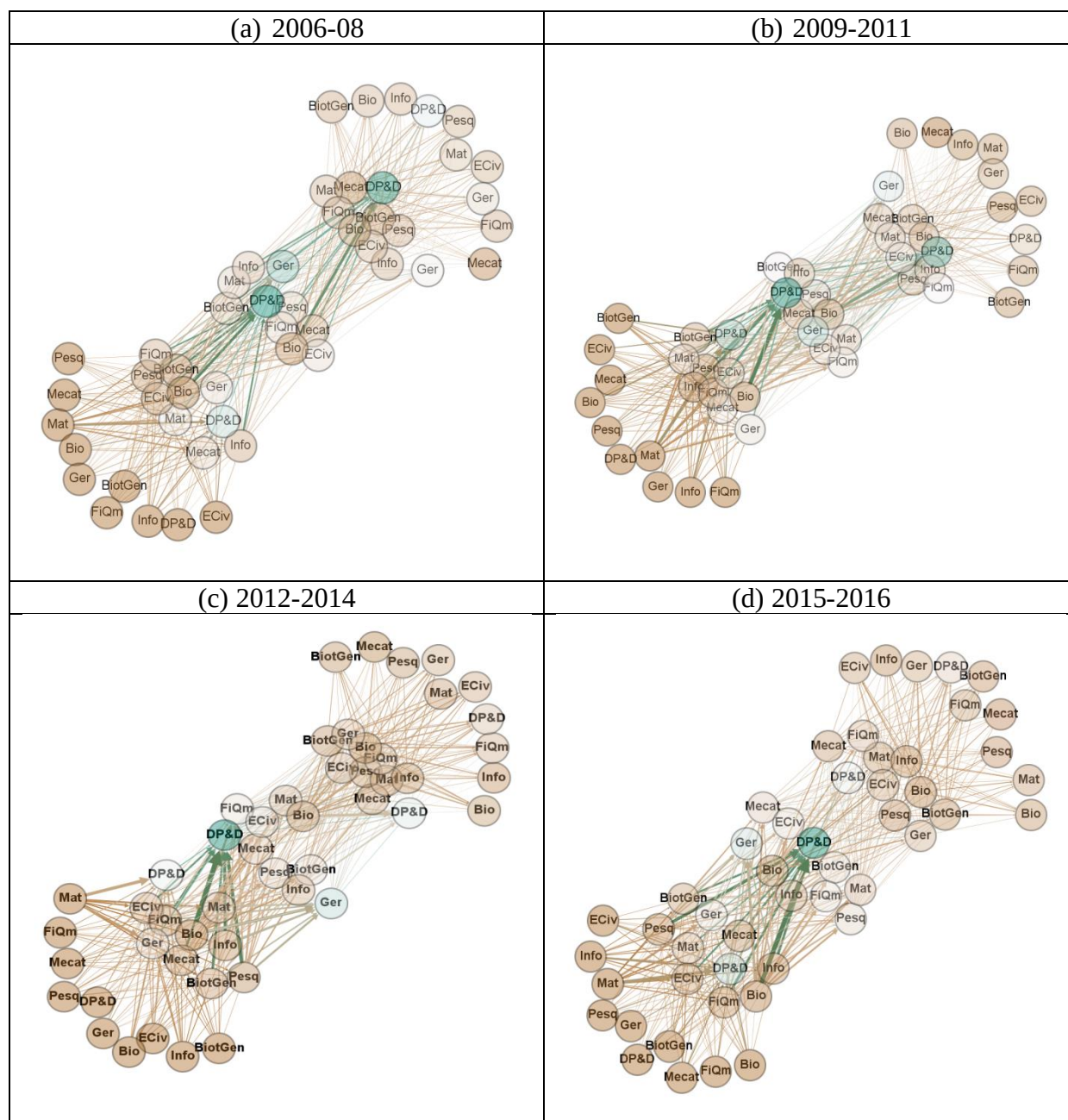
214SI	ECiv	10,783	17,731	15,118	16,550	15,046	53,5%
221SI	Bio	6,362	6,064	5,777	6,325	6,132	-0,6%
1237SI	DP&D	27,277	32,344	31,661	43,271	33,638	58,6%
1426SI	Ger	19,098	20,128	24,657	31,055	23,735	62,6%
201SC	BiotGen	16,712	23,672	25,698	29,557	23,910	76,9%
202SC	Mecat	12,947	13,535	20,489	24,827	17,949	91,8%
203SC	Pesq	17,003	18,718	23,272	27,434	21,607	61,3%
211SC	Mat	17,285	18,087	20,277	22,783	19,608	31,8%
212SC	Info	15,042	14,421	16,849	18,739	16,263	24,6%
213SC	FiQm	18,428	20,958	25,738	30,176	23,825	63,7%
214SC	ECiv	18,402	20,463	25,878	29,135	23,469	58,3%
221SC	Bio	7,986	8,618	10,689	11,863	9,789	48,6%
1237SC	DP&D	44,874	48,677	60,440	69,701	55,923	55,3%
1426SC	Ger	31,270	31,257	37,826	40,898	35,313	30,8%
201M	BiotGen	9,420	14,308	10,506	11,083	11,329	17,6%
202M	Mecat	7,275	15,427	15,981	13,892	13,144	90,9%
203M	Pesq	12,629	16,863	11,675	11,625	13,198	-7,9%
211M	Mat	12,681	18,130	13,174	15,220	14,801	20,0%
212M	Info	12,729	16,023	11,641	11,829	13,055	-7,1%
213M	FiQm	12,745	24,076	21,475	21,469	19,941	68,5%
214M	ECiv	14,053	21,958	18,225	18,276	18,128	30,0%
221M	Bio	8,690	12,058	8,690	8,450	9,472	-2,8%
1237M	DP&D	42,366	39,640	33,275	35,994	37,819	-15,0%
1426M	Ger	21,253	27,272	18,934	18,809	21,567	-11,5%
201D	BiotGen	10,400	7,707	9,518	10,458	9,521	0,6%
202D	Mecat	4,836	1,819	6,688	8,139	5,370	68,3%
203D	Pesq	12,143	8,337	10,699	10,573	10,438	-12,9%
211D	Mat	14,917	9,213	11,898	13,708	12,434	-8,1%
212D	Info	12,332	6,951	8,988	9,600	9,468	-22,2%
213D	FiQm	11,428	10,043	15,778	16,628	13,469	45,5%
214D	ECiv	12,165	9,092	14,003	14,843	12,526	22,0%
221D	Bio	12,111	9,344	11,829	12,025	11,327	-0,7%
1237D	DP&D	24,282	16,130	19,329	25,596	21,334	5,4%
1426D	Ger	18,078	11,972	14,737	16,022	15,202	-11,4%

Fonte: estatísticas dos autores a partir das redes complexas elaboradas no Gephi 0.9.2.

Obs: as ocupações com maiores variações estão realçadas em cinza claro e seus valores estão em negrito.

Observando a figura 5 da rede complexa obtida com o algoritmo de Yifan Hu proporcional (2005), observamos que entre os profissionais com superior completo houve aumento médio da participação dos profissionais com SI (superior incompleto) na composição da variação salarial do grau de instrução SC (superior completo) de 9,81, 23,17 e 14,19 p.p em média. Isso indica que as pessoas que completam o nível superior conseguem obter salários maiores.

Figura 7 – Evolução temporal das redes conforme seus graus ponderados de entrada e pesos das arestas



Fonte: Redes complexas elaboradas pelos autores no Gephi 0.9.2.

Legenda 7 - cores das redes

	$GP_{Ent} \geq 32,00$
	$27,00 \leq GP_{Ent} < 32,00$
	$14,00 \leq GP_{Ent} < 27,00$
	$0 < GP_{Ent} < 14,00$

Quando estes profissionais saem da graduação para o mestrado, os valores encontrados para a composição da variação salarial foram de 47,21, -21,08 e 1,41 p.p. Para a comparação do segundo período em relação ao primeiro, os valores encontrados foram quase todos negativos (exceto CBO 202), o que indica uma queda da influência dos mestres na composição salarial geral.

Por que isso ocorre? Uma hipótese é que há uma parte considerável das profissões com SC (superior completo) com salários maiores do que outras com M (mestrado). Outra hipótese é que a diferença salarial pode ter caído consideravelmente entre os anos do triênio considerado.

Em relação aos doutores, houve uma forte retração na composição salarial do aumento do grau de instrução de M para D em 32,91 p.p. em média. Para os demais períodos, encontramos aumentos de 55,87 e 10,55 p.p (figura 5).

Em relação ao grau de saída ponderado, como o grau de entrada representa o quanto um grau de instrução afeta a variação salarial por ocupação, o de saída expressa quanto essa ocupação de menor grau instrucional foi afetada. Portanto, não faz sentido analisar o grau de saída, uma vez que queremos avaliar os efeitos dos aumentos e não das reduções do grau de instrução sobre a variação salarial (figura 5).

O *learning by doing* afeta a remuneração destas profissões. Entre as ocupações com maior capacidade remuneratória da indústria brasileira baseada em inovação e P&D, no triênio de 2006-2008, encontram-se os diretores de pesquisa e desenvolvimento com superior completo, que se beneficiaram de maiores variações salariais do que aqueles com mestrado (50,509 contra 46,207). Do mesmo modo, os biólogos com superior incompleto (43,177) aparecem em terceiro lugar nesse *ranking*, acima dos que possuem superior completo (39,654) e bem acima daqueles que possuem mestrado (27,413). Esse padrão repete-se para biotecnologistas, geneticistas, engenheiros civis, profissionais de informática, matemáticos, físicos, químicos e afins.

De certo modo, esta assimetria na remuneração dos profissionais mais experientes e com maior grau de *learning by doing* em relação aos com melhor educação formal, porém menos experientes, ajuda a compreender o ponto levantado por Acemoglu e Autor (2014), mencionado no item 2 deste trabalho.

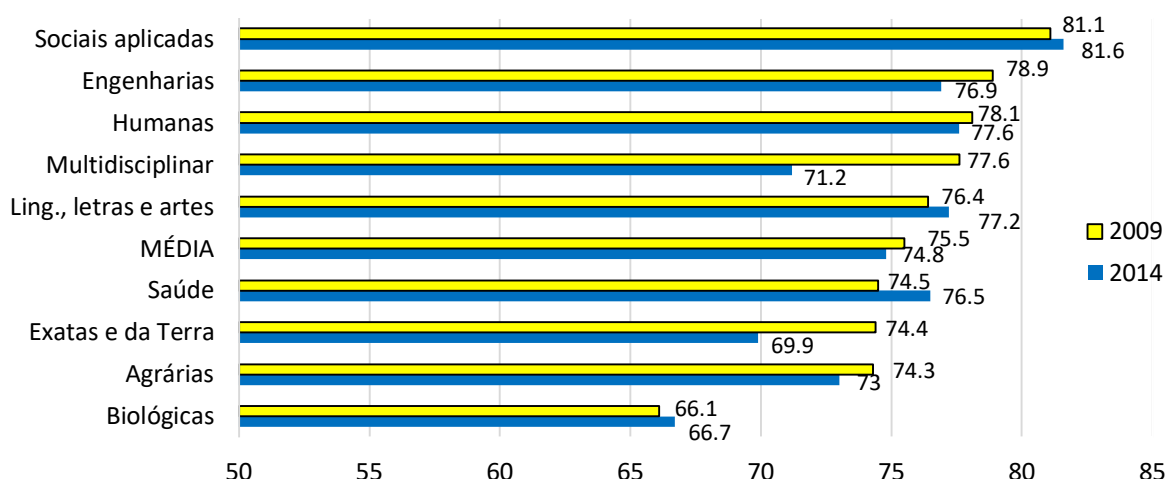
E também espelha o fato de que a formação mais intensiva de mestres e doutores no Brasil é um fenômeno recente. É possível levantar a hipótese de que estes profissionais com qualificações mais altas ainda não tiveram tempo suficiente de trabalho para conquistarem maiores remunerações do que aqueles com maior experiência e nível de *learning by doing*. Isso somado à escassez de mestres e doutores nesse mercado de trabalho e à importância da experiência nessas ocupações e temos um ambiente institucional que possibilita o fenômeno mencionado.

Ainda que, conforme o Ministério da Ciência, Tecnologia e Inovação, o número de mestres e doutores formados pelas universidades brasileiras tenha mais do que quadruplicado em 15 anos, o país passou de um número muito baixo de 13.219 em 1996 para 55.047 em 2011 (figura 6).

Assim, o aumento de 312%, embora tenha sido expressivo, não se mostrou suficiente para afetar decisivamente o mercado de trabalho dos profissionais de PD&I. Além disso, estes mestres e doutores também são absorvidos pelas administrações públicas federais e estaduais, com menor participação no setor privado.

Em relação à produtividade dos pesquisadores de engenharia, Fioroni (2015) aponta que ainda há uma produção científica muito concentrada em periódicos nacionais. Há ainda pouca inserção da pesquisa brasileira nos diversos campos da engenharia na literatura de ponta internacional.

Figura 8 – Taxa de emprego formal (carteira assinada) em 2009 e em 2014 de mestres titulados no Brasil a partir de 1996, por grande área do conhecimento (em %).



Fonte: Centro de Gestão e Estudos Estratégicos (CGEE) do Ministério da Ciência, Tecnologia e Inovação (MCTI), com dados da Coleta Capes 1996-2012.

Também há uma participação ainda insuficiente de mestres e doutores das áreas de exatas e da Terra e Engenharias (figura 6 e tabela 7).

Tabela 7 - Número de empregados entre os doutores e os mestres titulados no Brasil a partir de 1996, por seção da Classificação Nacional de Atividades Econômicas (CNAE) dos estabelecimentos empregadores, 2009 e 2014

Atividade Econômica (Seção da CNAE)	2009	2014	Particip. 2014 (%)	Taxa anual de crescimento (%)
Doutores Empregados				
Total	23.767	126.902	100	9,46
P Educação	53.989	94.535	74,49	9,79
O Adm. Pública, defesa e seguridade social	8.764	15.345	12,09	9,79
Q Saúde humana e serviços sociais	2.735	5.531	4,36	12,45
M Atividades profis., científicas e técnicas	3.581	4.472	3,52	3,77
C Indústrias de transformação	1.023	1.715	1,35	8,99
Outras Seções	3.675	5.304	4,19	6,31
Mestres Empregados				
Total	184.960	293.381	100	7,99
P Educação	79.391	123.699	42,16	7,67
O Adm. Pública, defesa e seguridade social	54.021	91.517	31,19	9,18
C Indústrias de transformação	8.505	14.433	4,92	9,21
Q Saúde humana e serviços sociais	7.294	13.770	4,69	11,17
M Atividades profis., científicas e técnicas	7.307	10.017	3,41	5,40
Outras seções	28.442	39.945	13,63	5,82

Fonte: Adaptado do Relatório: Mestres e Doutores 2015 - Estudos da demografia da base técnico-científica brasileira (CGEE).

No tocante à importância do capital humano avançado brasileiro, Paula, Uhr et al. (2018) partiram de um modelo de crescimento de Solow aumentado com capital humano, semelhante aos propostos por Lucas (1988) e Mankiw, Romer e Weil (1992) e das

contribuições de Nelson e Phelps (1966), os autores examinaram por meio de quais canais o capital humano afetou o crescimento econômico no Brasil, no período de 1996 a 2015. Foram testadas as seguintes hipóteses: o capital humano – agregado e desagregado em níveis básico e avançado - afeta o crescimento através (i) do canal de acumulação dos fatores; (ii) do canal de produtividade total dos fatores; ou (iii) de ambos os canais simultaneamente. Este trabalho propôs novas proxies para o estoque de capital humano, relacionadas a retornos educação formal, experiência e treinamento no local de trabalho, e que, portanto, não são afetadas pelas principais deficiências das medidas documentadas pela literatura, tais como as relacionadas às taxas de evasão escolar, à repetência e aos aspectos de produtividade da mão de obra (Barro e Lee, 1993; Mulligan e Sala- I-Martin, 1995). A análise foi feita com os dados para as unidades federativas do Brasil e a análise empírica teve como o método *two-step System-GMM* (Arellano e Bover, 1995; Blundell e Bond, 1998). Os resultados mostraram que tanto a medida de capital humano agregado quanto as medidas de capital humano desagregado em níveis básico e avançado impactaram no crescimento econômico por meio de ambos os canais, mas que os impactos por meio do canal de acumulação dos fatores e os relacionados ao capital humano em nível básico foram mais expressivos.

Aprofundando um pouco os itens C, Q e M da tabela 7, observamos um indicador da influência do capital avançado para a produtividade das empresas brasileiras, em uma perspectiva setorial. A tabela 8 retrata os setores com maior riqueza gerada por empregado (em dólares), uma *proxy* para a produtividade setorial.

Tabela 8 – Riqueza gerada por empregado dos setores que compõem uma amostra das 1000 maiores empresas brasileiras (em US\$) – 2013/2017

SETORES	2013	2014	2015	2016	2017	Médias	Ranking
Atacado	102.567	96.938	169.434	102.945	77.642	109.905	7
Autoindústria	44.141	42.586	33.060	46.199	52.534	43.704	18
Bens de capital	54.722	53.068	38.304	35.775	43.182	45.010	16
Bens de consumo	69.480	69.018	58.317	69.347	72.539	67.740	12
Eletroeletrônico	55.867	59.988	54.430	46.620	85.787	60.538	14
Energia	340.673	324.392	520.749	542.178	546.550	454.908	1
Farmacêutico	107.357	120.949	108.151	119.148	119.146	114.950	5
Indústria da Construção	57.876	63.817	51.314	66.927	66.177	61.222	13
Indústria digital	61.604	53.218	49.349	61.978	61.111	57.452	15
Infraestrutura	117.511	114.315	101.583	96.806	112.491	108.541	8
Mineração	129.517	67.320	98.584	84.858	178.163	111.688	6
Papel e celulose	104.227	96.273	79.716	68.050	77.457	85.145	9
Química e petroquímica	149.946	171.240	171.321	116.415	124.286	146.642	3
Serviços	59.411	54.610	69.540	326.903	70.705	116.234	4
Serviços de saúde	45.749	48.385	39.662	39.403	49.068	44.453	17
Siderurgia e metalurgia	87.245	98.932	80.430	55.272	82.577	80.891	10
Telecomunicações	89.496	441.884	96.952	70.810	106.035	161.035	2
Têxtil	21.710	20.179	21.172	23.494	24.095	22.130	20
Transporte	62.792	45.057	54.653	109.934	87.752	72.038	11
Varejo	27.771	26.835	25.596	23.614	25.741	25.911	19

Fonte: elaboração dos autores a partir de dados do anuário “Melhores e Maiores” da Revista Exame.

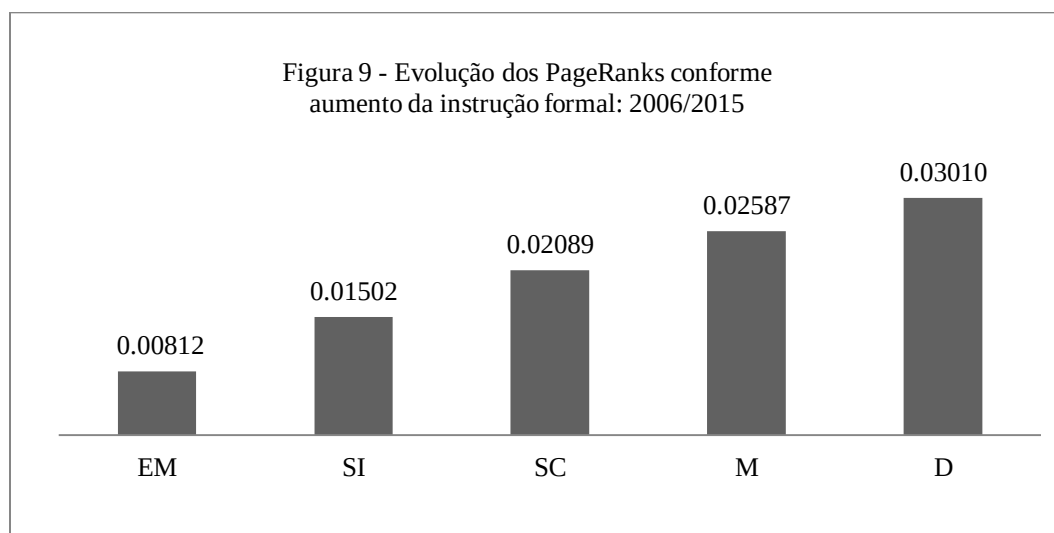
Percebe-se que há uma correlação entre os setores que mais geram riqueza e o grau de incorporação de capital físico, humano e tecnologia. Assim, setores como, em ordem

decrecente, energia, telecomunicações, química e petroquímica, farmacêutico, mineração, infraestrutura, siderurgia/metalurgia e transporte costumam investir na contratação de capital humano avançado e com formação em ciência e tecnologia. No ranking da consultoria Universum, preparado para a revista Exame, algumas empresas desses setores figuram entre as mais desejadas pelos estudantes de engenharia. São empresas como: Petrobrás, Odebrecht, Vale, Eletrobrás, Camargo Corrêa, Embraer, Gerdau, Andrade Gutierrez, Rede Globo, Bayer, Usiminas, 3M, Braskem, Shell, CSN entre outras²⁰.

5.2.3. PageRanks

Como vimos, o PageRank é um algoritmo de avaliação de links que atribui um valor ponderado numérico a cada vértice de uma rede complexa (o Google o utiliza para atribuir uma ponderação numérica a cada elemento de um conjunto de documentos com hiperlinks, como a World Wide Web). O foco recai sobre a mensuração da importância relativa de cada elemento da rede. Esse algoritmo pode ser aplicado a qualquer conjunto de vértices com arestas pesadas. O peso numérico que ele atribui a qualquer dado elemento E é referido como o PageRank de E denotado por $PR(E)$. O PageRank é uma variante da centralidade do autovetor, cujo algoritmo é baseado em autovalores e autovetores de matrizes simétricas (adjacência), cujo objetivo é medir a importância de um vértice em função da importância de seus vizinhos.

Na prática, é possível verificar que, mesmo se um vértice está conectado somente a alguns outros vértices da rede (tendo assim uma baixa centralidade de grau), estes vizinhos podem ser importantes e, conseqüentemente, o vértice será também o será, obtendo uma centralidade de autovetor elevada. A diferença é que o PageRank também leva em consideração a direção dos links e o peso das arestas em consideração – fazendo com que o vértice em questão influencie com um grau de força diferente cada um dos outros vértices no dígrafo. Além disso, é calculada a influência não apenas nos nós diretamente ligado a ele, mas também nos que não tem ligação alguma. Pode-se verificar na figura 7 que, em termos médios e mais agregados, os PageRanks crescem com o aumento dos graus de instrução formal, conforme já pudemos avaliar nos graus ponderados.



Fonte: estatísticas dos autores a partir das redes complexas elaboradas no Gephi 0.9.2.

²⁰ Disponível em: <<https://exame.abril.com.br/carreira/as-empresas-dos-sonhos-dos-alunos-de-engenharia-no-brasil>>. Acesso em 22/02/2019.

Com relação ao EM (ensino médio), é fácil ver que tem o menor PageRank de todos, indicando que ele terá, em termos relativos, pouca influência ou capacidade de capturar variações salariais. Por contraste, o D (doutorado) tem o maior PageRank, indicando que ele é responsável pelas maiores capturas de variações salariais. Novamente, o resultado é o previsto pela literatura de capital humano: os investimentos em capital humano traduzem-se em maior capacidade de obter salários maiores.

Assim, os valores encontrados demonstram que, em média e em ordem crescente de nível instrucional, a variação salarial “capturada” avança da seguinte forma: aumento do grau de instrução, em ordem crescente de grau de instrução, 8,12% (EM/SI), 15,02% (SI/SC), 20,89% (SC/M), 25,87% (M/D) e 30,10% (D), respectivamente.

Em termos mais desagregados, a tabela 8 mostra algumas variações positivas e negativas de PageRanks que merecem destaque. A tabela exclui as ocupações de EM (que, por definição, possuem PRs nulos).

Na tabela 8, destacam-se positivamente, no nível SI, as ocupações de BiotGen (15,72%), além dos cargos de DP&D (16,26%) e Ger (15,66%). Os destaques negativos são Mecat (-21,62%) e Mat (-13,04%). No nível SC, a ocupação que apresentou maior variação entre PR1 e PR4 foi Mecat (12,96%). A maior variação negativa ocorreu no cargo de Ger (-10,16%).

Tabela 8 – PageRanks (PRs) e variação entre o primeiro e o último períodos nos triênios de 2006-08, 2009-11, 2012-14 e no biênio de 2015-16.

Código	Ocupação	2006-08 PR1	2009-11 PR2	2012-14 PR3	2015-16 PR4	Var. (%) PR1 a PR4
201SI	BiotGen	0,01190	0,01306	0,01369	0,01377	15,72%
202SI	Mecat	0,01685	0,01791	0,01436	0,01321	-21,62%
203SI	Pesq	0,01280	0,01209	0,01211	0,01182	-7,64%
211SI	Mat	0,01832	0,01622	0,01654	0,01593	-13,04%
212SI	Info	0,01315	0,01251	0,01283	0,01280	-2,69%
213SI	FiQm	0,01392	0,01304	0,01307	0,01271	-8,71%
214SI	ECiv	0,01335	0,01577	0,01500	0,01451	8,68%
221SI	Bio	0,01120	0,01074	0,01075	0,01056	-5,76%
1237SI	DP&D	0,02135	0,02208	0,02253	0,02482	16,26%
1426SI	Ger	0,01738	0,01681	0,01935	0,02010	15,66%
201SC	BiotGen	0,01879	0,02196	0,02040	0,02049	9,03%
202SC	Mecat	0,01639	0,01603	0,01791	0,01851	12,96%
203SC	Pesq	0,01898	0,01906	0,01924	0,01960	3,28%
211SC	Mat	0,01916	0,01869	0,01781	0,01766	-7,85%
212SC	Info	0,01773	0,01655	0,01617	0,01596	-9,95%
213SC	FiQm	0,01989	0,02037	0,02042	0,02075	4,33%
214SC	ECiv	0,01987	0,02008	0,02049	0,02031	2,22%
221SC	Bio	0,01322	0,01316	0,01323	0,01308	-1,02%
1237SC	DP&D	0,03678	0,03658	0,03701	0,03729	1,39%
1426SC	Ger	0,02809	0,02640	0,02620	0,02524	-10,16%
201M	BiotGen	0,01899	0,02046	0,01952	0,01993	4,92%
202M	Mecat	0,01652	0,02143	0,02546	0,02292	38,77%
203M	Pesq	0,02269	0,02267	0,02079	0,02050	-9,65%
211M	Mat	0,02275	0,02376	0,02242	0,02433	6,95%
212M	Info	0,02281	0,02194	0,02075	0,02072	-9,15%
213M	FiQm	0,02283	0,02889	0,03143	0,03099	35,76%
214M	ECiv	0,02434	0,02707	0,02790	0,02759	13,36%

221M	Bio	0,01815	0,01852	0,01755	0,01712	-5,65%
1237M	DP&D	0,05701	0,04232	0,04423	0,04646	-18,50%
1426M	Ger	0,03264	0,03165	0,02867	0,02816	-13,75%
201D	BiotGen	0,02535	0,02681	0,02506	0,02483	-2,05%
202D	Mecat	0,01613	0,01253	0,02003	0,02112	30,94%
203D	Pesq	0,02823	0,02834	0,02717	0,02501	-11,42%
211D	Mat	0,03283	0,03047	0,02930	0,03002	-8,56%
212D	Info	0,02855	0,02498	0,02412	0,02345	-17,84%
213D	FiQm	0,02705	0,03248	0,03621	0,03468	28,22%
214D	ECiv	0,02827	0,03018	0,03305	0,03183	12,60%
221D	Bio	0,02818	0,03079	0,02918	0,02733	-3,02%
1237D	DP&D	0,04834	0,04725	0,04253	0,04901	1,38%
1426D	Ger	0,03807	0,03716	0,03435	0,03371	-11,43%

Fonte: estatísticas dos autores a partir das redes complexas elaboradas no Gephi 0.9.2.

Obs: as ocupações com maiores variações positivas estão realçadas em negrito, sendo as realçadas em itálico e na cor cinza aquelas com maiores variações negativas.

As maiores variações positivas, entre todos os níveis de formação, ocorreram no nível de M, tendo destaque novamente Mecat (38,77%), FiQm (35,76%), ECiv (13,36%). As maiores variações negativas em M registraram-se nos cargos DP&D (-18,50%) e Ger (-13,75%).

No nível instrucional mais alto de D (doutorado), as principais variações positivas foram Mecat (30,94%) – ocupação que destacou-se em todos os níveis – FiQm (28,22%), ocupação que destacou-se apenas em M (mestrado) e em D (doutorado) e ECiv, engenheiros civis (12,60%), que registrou variações positivas em todos os níveis, mas seus maiores aumentos foram em M e D. As ocupações que tiveram reduções fortes de PRs foram: Pesq (-11,42%), Info (-17,84%) e o cargo de Ger (-11,43%).

Consideramos ser útil comparar as variações de PRs entre a amostra com os cargos DP&D e Ger e entre uma e outra sem estes cargos. Isso ajuda a analisar os efeitos de *learning-by-doing* e a testar se a rede possui resiliência necessária para caracterizá-la como uma rede SF. A caracterização de uma rede complexa como sendo livre de escala (*scale-free* ou SF) é, por si só, um objeto de estudo na network Science e na economia computacional. Todavia, uma análise mais aprofundada sobre uma rede SF transcende o escopo desse trabalho, razão pela qual a deixaremos para trabalhos futuros. Contudo, não podemos nos furtar a, diante dos dados de PR, identificar a rede como sendo do tipo livre de escala.

Como as maiores remunerações dos cargos DP&D e Ger são decorrentes não tanto do grau de instrução, mas da experiência acumulada pelo *learning-by-doing*, observamos na tabela 9 que as variações de PRs sem contar os referidos cargos, são bem maiores para todos os níveis (à exceção de SI, superior incompleto).

Tabela 9 – Variações dos PageRanks (PRs) entre o 1º triênio de 2006-08 e o último biênio de 2015-16, incluindo e excluindo os cargos de DP&D e Ger.

Níveis	Com DP&D e Ger.	Sem DP&D e Ger.
SI (superior incompleto)	-0,31%	-4,38%
SC (superior completo)	0,42%	1,62%
M (mestrado)	4,31%	9,41%
D (doutorado)	1,88%	3,61%
Total	1,57%	2,57%

Fonte: estatísticas dos autores a partir das redes complexas elaboradas no Gephi 0.9.2.

Tal constatação é uma evidência da interpretação estabelecida por Acemoglu e Autor (2014) de que os indivíduos que cursaram a mesma quantidade de anos em uma mesma escola podem ter níveis assimétricos de capital humano. E que isso ocorre devido aos incentivos salariais, os quais aumentam a atração do *learning-by-doing* em face do acúmulo de capital humano formal avançado obtido pelos anos de estudo nos níveis de mestrado e doutorado.

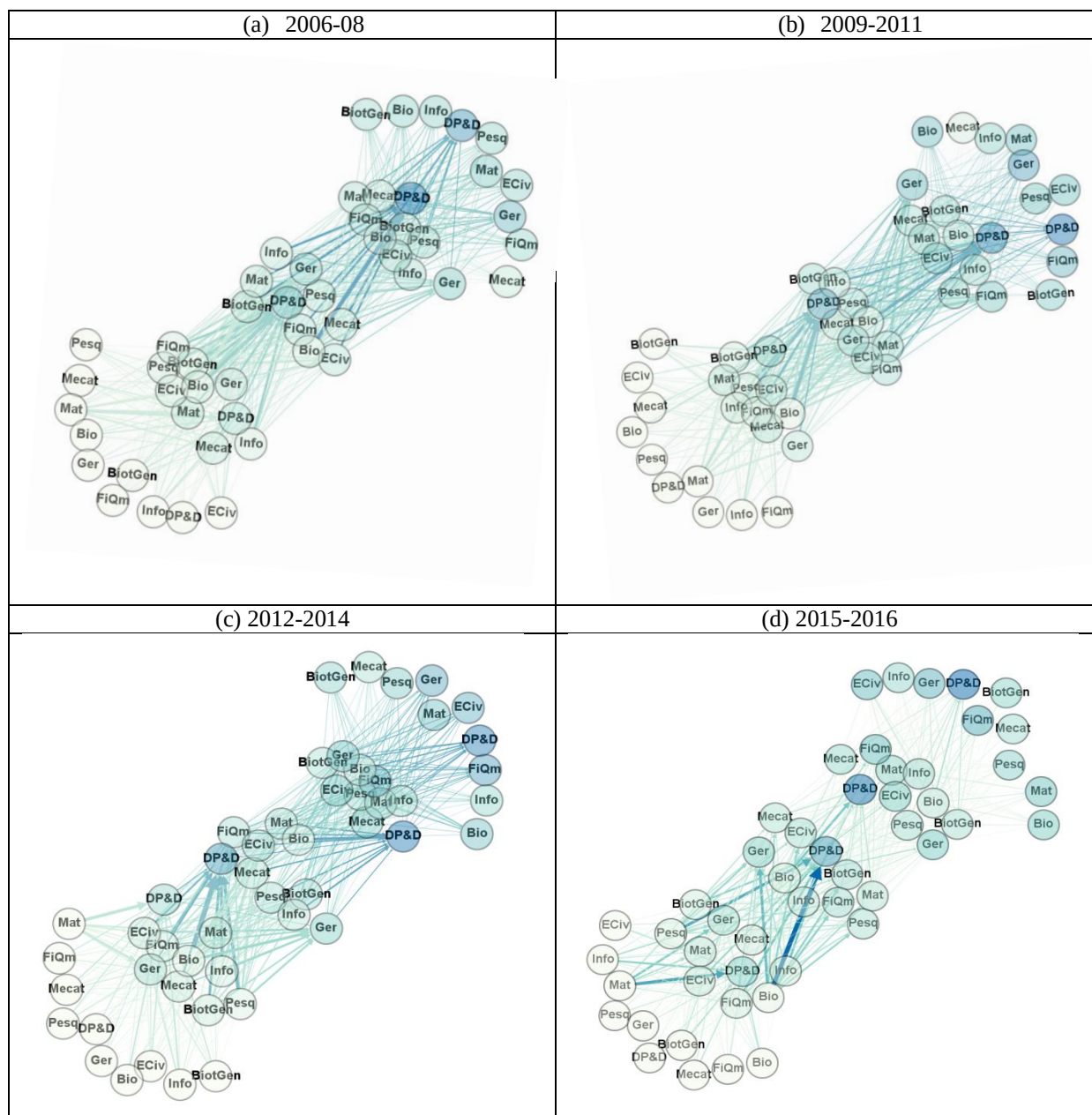
E também corrobora a relação positiva entre formação escolar e maiores remunerações que foi explorada por vários autores, tendo sido descrita por Sahn e Alderman, 1988, os quais utilizaram o procedimento de Heckman para corrigir o viés de seleção da amostra na estimativa da oferta de salários e nas equações de oferta líquida do mercado de trabalho e para homens e mulheres do Sri Lanka usando dados de 1980-1981. Eles descobriram que os salários foram afetados pela educação e experiência. E que a elasticidade da oferta de trabalho em relação aos salários é positiva e menor em áreas urbanas do que rurais (0,14 para homens rurais, 0,07 para homens urbanos, 0,15 para mulheres rurais e 0,03 para mulheres urbanas).

Veum (1999) usou dados da Pesquisa Nacional Longitudinal da Juventude (*National Longitudinal Survey of Youth – NLSY*) referentes à 1990 para testar a validade das previsões do modelo de capital humano sobre a relação entre treinamento, salários iniciais e crescimento salarial. Tal como o modelo deixa implícito, o treinamento, particularmente o treinamento financiado pelo empregador, está positivamente relacionado ao crescimento salarial. O treinamento financiado pela empresa também parece ser útil para todos os empregos ou ter um componente geral. Além disso, há algumas evidências de que os trabalhadores pagam pelo treinamento inicial através de um salário inicial reduzido. Os resultados fornecem apoio parcial ao modelo de capital humano.

Pergamit e Veum (1999), também confirmaram essa relação utilizando os dados da mesma NLSY para o mesmo ano de 1990. Eles analisaram os determinantes e as consequências de uma promoção entre os jovens trabalhadores. A maioria dos eventos que os trabalhadores chamavam de "promoções" não envolviam mudança de posição ou deveres, mas eram simplesmente uma atualização da posição atual. Normalmente, apenas uma pessoa foi considerada para a promoção. Os homens eram mais propensos a serem promovidos do que as mulheres e os brancos, mais propensos do que os negros ou hispânicos. A aquisição de treinamento da empresa e o recebimento de uma promoção prévia foram dois dos mais importantes indicadores de promoção. As consequências da promoção incluíram aumento de salários, recebimento de treinamento, responsabilidades de supervisão e aumento da satisfação no trabalho.

Schömborg (2007), utilizou a mesma pesquisa citada nos dois parágrafos anteriores para comparar as fontes de crescimento salarial de jovens trabalhadores do sexo masculino em dois mercados de trabalho muito diferentes: os Estados Unidos e a Alemanha. O autor primeiro desenvolveu um método simples para decompor o crescimento salarial em componentes devidos à acumulação de capital humano geral, à acumulação de capital humano específico da empresa (afetado pelo *learning-by-doing*) e à procura de emprego. A análise empírica deles utilizou os dados de registros administrativos (Alemanha) e do já mencionado NLSY (Estados Unidos) para coortes que entraram no mercado de trabalho no final dos anos 1970 e início dos anos 80. Apesar de as taxas de mobilidade dos dois países diferirem substancialmente, ambos exibiram resultados semelhantes nas fontes de crescimento salarial, com a acumulação geral de capital humano sendo a fonte mais importante e a procura por respondendo por um acréscimo de 25% ou mais no crescimento total dos salários. Não há provas de que retornos à acumulação de capital humano específico da firma foram maiores para aprendizes alemães do que para os norte-americanos que concluíram ou desistiram do ensino médio.

Figura 10 – Evolução temporal das redes conforme seu PageRank e pesos das arestas



Fonte: Redes complexas elaboradas pelos autores no Gephi 0.9.2.

Legenda 10 - Cores dos PageRanks

	$PR \geq 0,42$
	$0,20 \leq GPent < 0,42$
	$0,16 \leq GPent < 0,20$
	$0 < GPent < 0,16$

Guisan e Aguayo (2007) também exploraram a referida relação ao afirmarem que a União Europeia ficou atrás dos Estados Unidos, tanto em taxas de emprego como em salários. As autoras investigaram as relações entre salários, produtividade e desenvolvimento de capital humano em 5 países da União Europeia: França, Alemanha, Itália, Espanha e Reino Unido e os compararam com os Estados Unidos. Iniciaram pela análise do papel da produtividade na

explicação dos salários reais. Em seguida, apresentaram um estudo comparativo da evolução dos salários e da produtividade durante o período 1985-2005 e, finalmente, estimaram um modelo econométrico que relacionou salários reais com produtividade e produtividade com acúmulo de capital humano. Como medidas de capital humano, utilizaram a média de anos de escolaridade, com base em estimativas de Barro e Lee e as despesas de pesquisa por habitante. Concluíram que a União Europeia deveria desenvolver políticas, a fim de aumentar o apoio ao capital humano, promovendo a educação e aumentando as despesas para atingir níveis mais elevados de salários reais e taxas mais elevadas de emprego, para, ao fim e ao cabo, convergir para os níveis dos Estados Unidos.

Tchernis (2010) investigou os efeitos da acumulação de capital humano nos salários. Utilizou um modelo de seleção para especificar a dependência entre mobilidade de trabalho e salários. Esse modelo foi estimado usando uma estrutura totalmente bayesiana que permitiu ao autor amostrar a partir da exata distribuição de uma amostra pequena dos parâmetros do modelo. Os parâmetros do modelo, segundo o autor, não informaram direta ou totalmente os efeitos da mobilidade do trabalho sobre os salários. Isto, segundo ele, ocorreu devido à natureza sequencial da dependência entre mobilidade e salários. Todavia, ele inferiu esses efeitos prevendo salários futuros que corresponderam a cenários alternativos de mobilidade e, portanto, padrões diferentes de acumulação de capital humano. Assim sendo, ele simulou distribuições preditivas posteriores de mobilidade e salários. Esta abordagem permitiu levar em consideração duas fontes de incerteza: a incerteza do modelo e a incerteza de estimativa/previsão. Tchernis estimou o modelo usando trabalhadores do PSID (*Panel Study of Income Dynamics*²¹). O modelo foi estimado separadamente para graduados do ensino médio e superior no intuito de permitir efeitos diferenciais da acumulação de capital humano pós-escolar nos salários. Para cada grupo, foram escolhidos dois indivíduos "representativos" hipotéticos, correspondentes a diferentes estágios da carreira, e simulados os caminhos de mobilidade-salário para esses indivíduos. Os efeitos da mobilidade do trabalho foram inferidos comparando-se os resultados salariais em períodos posteriores, correspondendo a trajetórias de mobilidade passadas alternativas. O autor descobriu que a mobilidade da mão-de-obra teve forte efeito sobre a distribuição de salários em períodos posteriores, tanto para grupos educacionais quanto para os dois estágios da carreira de um indivíduo. Também encontrou evidências que sugerem que esses efeitos foram muito maiores para os graduados em faculdades do que para os graduados do ensino médio. Para todas as suas simulações, somente um cenário de mobilidade levou a uma perda salarial (em comparação com a alternativa sem mobilidade). E mostrou que tanto a magnitude quanto a direção dos efeitos da mobilidade da mão-de-obra nos salários dependem fortemente do momento da mudança de emprego. As suas estimativas apontaram que uma mudança de emprego em um momento ideal pode aumentar os salários de um jovem graduado do ensino médio em 9% e os salários de um jovem graduado da faculdade em até 26%. Estes resultados indicaram que é necessária uma investigação mais aprofundada dos efeitos do capital humano no crescimento salarial, bem como a construção de novas medidas de capital humano que reflitam competências transferíveis entre alguns, mas não todos, os postos de trabalho.

Phillipon e Resheff (2012) utilizaram informações detalhadas sobre salários, educação e ocupações para esclarecer a evolução do setor financeiro dos EUA no último século. Descobriram um conjunto de novos fatos estilizados inter-relacionados: os empregos financeiros eram relativamente intensivos em e muito bem pago até a década de 1930 e depois da década de 1980, mas não no período intermediário. Os autores investigaram os determinantes dessa evolução e afirmaram que a desregulamentação financeira e atividades corporativas ligadas a *IPOs* e risco de crédito aumentaram a demanda por habilidades relacionadas à capital humano avançado pelos empregadores das instituições financeiras. Curiosamente, o capital físico, isto é, os computadores e a tecnologia da informação

²¹ PSID é a mais antiga pesquisa longitudinal domiciliar do mundo. O estudo começou em 1968.

desempenharam um papel mais limitado. A análise também mostrou que os salários em finanças eram excessivamente altos por volta de 1930 e de meados dos anos 90 até 2006. Para o período recente, estimaram que os rendimentos extras, associados à bônus e às remunerações por produtividade, representaram 30% a 50% do diferencial salarial entre o setor financeiro e o resto do setor privado.

Alon (2018) construiu um modelo de longo prazo de acumulação de capital humano impulsionado pelo aumento especialização da força de trabalho. Nele, os indivíduos aumentam a eficiência do tempo dedicado para a aquisição de capital humano, concentrando os investimentos em conjuntos mais restritos de habilidades. Mostrou que a evolução dos currículos secundários e pós-secundários nos Estados Unidos a partir de 1870 até 2000 confirma a presença destas mudanças no âmbito da especialização. Seus exercícios quantitativos evidenciaram que a especialização foi responsável por cerca de 29% do aumento do prêmio pago pela habilidade e por 25-30% do aumento do nível educacional no período de 1965-2005. O efeito sobre o prêmio pago pela habilidade é em grande parte devido a um declínio na especialização no ensino médio, onde a formação profissional foi substituída por requisitos de graduação acadêmica. As previsões do seu modelo também são consistentes com as alterações internacionais nos prêmios pagos pelas habilidades, com a variação dos níveis de ensino e com as modificações na organização das instituições de ensino. O autor considera que uma política importante resultante da análise é que a abertura de espaço para treinamento profissional especializado em escolas secundárias pode ser uma ferramenta eficaz para combater a desigualdade de renda.

Outras duas conclusões finais estão relacionadas à estrutura da rede complexa e dizem respeito à teoria das redes. Ambas deverão ser exploradas em artigos futuros, uma vez que a sua análise extrapola os objetivos dessa dissertação.

A primeira delas diz respeito ao fato de que a rede complexa apresenta resiliência, pois mesmo com a exclusão de hubs importantes relativos aos cargos mencionados, as conclusões sobre os PRs se mantêm.

A segunda, que na verdade é um corolário da primeira, é a constatação de que, uma vez que a rede é resiliente, isto reforça sua característica de uma rede SF (*scale-free* ou livre de escala). A identificação de redes SF constitui-se, por si só, em uma linha de pesquisa nas áreas de teoria dos grafos, economia computacional e *network Science*. Portanto, acredita-se que há interesse na comunidade científica pela divulgação e maior exploração desse resultado empírico, razão pela qual pretende-se pesquisa-lo melhor no futuro.

5.2.4. Centralidade do autovetor

Como já descrevemos no item 1.3.1, a centralidade do autovetor de um agente é definida pela centralidade dos agentes com os quais se relaciona, isto é:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j \quad (7)$$

sendo que x_i/x_j a centralidade do agente i/j ; a_{ij} denota a matriz de adjacência A ($a_{ij} = 1$ se os agentes i e j estão conectados por uma aresta e $a_{ij} = 0$ se não estão) e λ indica o maior autovetor da matriz A .

Assim, a variação salarial de um agente que melhora seu capital humano por meio de instrução formal ou *learning by doing* define-se pela variação salarial de seus *alters* (outros agentes diretamente relacionados ao agente central).

A centralidade do autovetor é nula no nível EM (ensino médio), pois os graus de entrada dos agentes são nulos, como já mencionamos. Já nos níveis educacionais seguintes,

temos: superior incompleto (SI com 0.003104); superior completo (SC com 0.038817); mestrado (M com 0.241605) e doutorado (D igual a 1, pois os graus de saída são nulos).

Novamente, percebe-se a importância do mestrado para o aumento da variação salarial dos agentes: os *alters* dos mestres são agentes com superior completo (SC com $=0,038817$, o segundo mais alto de todos) ou doutorado (x_i = igual a 1, o mais elevado). Tal evidência acrescenta maior centralidade às ocupações que possuem mestrado.

6. Considerações finais

A pesquisa desenvolveu-se a partir das medidas estatísticas derivadas de quatro redes complexas baseadas no algoritmo proporcional Yifan Hu Proporcional (Hu, 2006) para os triênios 2006-08, 2009-2011 e 2012-2014 e uma para o biênio 2015-2016. Usando esses dados, comparamos as variações salariais resultantes do investimento em capital humano avançado e do *learning-by-doing* entre as ocupações referidas.

A análise das classes de modularidade delineou os vértices que compõem os *clusters* dentro das redes, bem como mensurou a intensidade das ligações dentro dos *clusters* em questão. Ela mostrou alterações no cluster 1 que registraram uma variação significativa nas remunerações das ocupações de EM (ensino médio). Nos triênios de 2006-08 e 2012-14, tais ocupações fizeram parte desse mesmo *cluster*, o que não ocorreu no triênio de 2009-11 e no biênio 2014-15. No quarto trimestre de 2008 e no primeiro semestre de 2009, os efeitos da Grande Recessão foram mais fortes na economia brasileira. De modo análogo, o PIB brasileiro caiu 9% em 2014-2017, o que pode ter afetado a remuneração das ocupações de EM (ensino médio) desse cluster 1. Outra hipótese, é que pode ter havido um movimento natural de substituição por causa da evolução da carreira desses trabalhadores e também da mobilidade desse mercado trabalho. Assim, os trabalhadores com EM no primeiro triênio podem ter obtido o grau de SC (superior completo) no segundo e/ou trocado de emprego. E isto pode ter impulsionado a demanda por novas contratações no terceiro triênio. Essa conjectura é confirmada pela maior presença de ocupações com SI (superior incompleto) no segundo triênio e no biênio 2015-16, no qual também aumentaram bastante as ocupações com SC (superior completo). As classes de modularidade também mostraram que os graus de instrução M e D (mestrado e doutorado) formaram um único cluster. E esse *cluster* é justamente o mais representativo da rede. Nele também estariam os profissionais com ensino superior completo e que são identificados como “Biólogos e afins - (212SC)”, “Profissionais de Informática – 221(SC)” e “Matemáticos, Estatísticos e afins – (211SC)”. Outro resultado é que nos períodos ímpares (1 e 3), os empregados com EM (ensino médio) apresentaram frágeis evidências de formação de um cluster. Já nos períodos pares (2 e 4), eles se agruparam mais fortemente. O mesmo ocorreu com os Diretores e Gerentes de P&D com superior incompleto (1237SI e 1426SI). E as ocupações com SI (superior incompleto) e SC (superior completo) não fizeram parte de um *cluster* claramente definido. O que pode se dizer é que há uma alternância de entrada para SI (superior incompleto) entre os graus de modularidade 2 e 1, e os de SC (superior completo) entre 0 e 1.

Na análise dos graus ponderados médios, observamos que os valores acima de 28 evidenciam que nas 4 redes há 26 *hubs* definidos. Nesses *hubs* encontram-se todas as ocupações com SI (superior incompleto) e SC (superior completo), metade das ocupações com M (mestrado) e uma única ocupação de EM (ensino médio). A existência desses *hubs* corrobora a influência do acúmulo de capital humano na obtenção de maiores variações salariais nas ocupações (a constatação de Acemoglu e Autor, 2014). Cabe mencionar que o valor de corte do grau médio ponderado adotado como referência para classificar um *hub*, corresponde a um nível equivalente ao valor médio destes graus (que é de 27,976). Os agentes que atingem o nível médio ou têm valores acima da média, foram incluídos em um *hub*. De tal forma que podemos afirmar que quando uma ocupação/nível de instrução faz parte de um

hub, é porque seus trabalhadores receberam uma variação salarial acima da média, o que reflete a influência mencionada. Uma vez que não existe nenhuma ocupação com D (doutorado) que seja um *hub*, reforçamos a conclusão de que a obtenção do mestrado possui maior relevância, em termos de acréscimo salarial, do que o doutoramento nas contratações dos setores da indústria brasileira intensiva em PD&I. Ainda em relação aos graus ponderados, como as arestas estão ponderadas (no sentido de “valoradas” ou “pesadas”) pelo índice proposto na subseção 3.10, os graus mais pesados refletem a maior capacidade remuneratória das ocupações e níveis de instrução. Portanto, verifica-se o crescimento médio dos salários de acordo com o aumento do grau de instrução, evidenciando que o investimento em capital humano proporciona aumentos de salários que vão de 7,88% (quando ocorre o ingresso nos cursos superiores relacionados à PD&I, isto é, no nível SI – superior incompleto) até o pico de 51,58% (quando um trabalhador obtém um doutorado nesses cursos). Também foi possível observar que quanto maiores os crescimentos médios das remunerações, maior tende a ser o aumento dos seus desvios-padrão.

Analisando os graus ponderados *de entrada*, no tocante ao nível superior incompleto, como a nossa análise restringe-se à 4 períodos, o aumento do grau de instrução mostrou um efeito de 13,42 pontos percentuais maior na composição da variação salarial no segundo período em comparação ao primeiro. Algumas ocupações registraram aumentos e outras quedas nas proporções salariais relativas aos aumentos do grau de instrução de EM (ensino médio) para SI (superior incompleto). No 3º período o aumento do grau de instrução passou a impactar 4,09 p.p. a menos do que o período anterior, em média, exceto para Gerentes de P&D (códigos 1426). Por fim, a última comparação indicou novamente um aumento na proporção salarial de 14,32 p.p., e aqui o destaque fica para os Gerentes e Diretores de P&D, que tiveram a maior participação na variação salarial. Também descobrimos que entre os profissionais com superior completo houve aumento médio da participação dos profissionais com SI (superior incompleto) na composição da variação salarial do grau de instrução SC (superior completo) de 9,81, 23,17 e 14,19 p.p em média. Isso mostra que as pessoas que completam o nível superior conseguem obter salários maiores. Quando estes profissionais saíram da graduação para o mestrado, os valores encontrados para a composição da variação salarial foram de 47,21, -21,08 e 1,41 p.p. Para a comparação do segundo período em relação ao primeiro, os valores encontrados foram quase todos negativos (exceto CBO 202), o que indica uma queda da influência dos mestres na composição salarial geral. Por que isso ocorre? Uma hipótese é que há uma parte considerável das profissões com SC (superior completo) com salários maiores do que outras com M (mestrado). Outra hipótese é que a diferença salarial pode ter caído consideravelmente entre os anos do triênio considerado. Em relação aos doutores, houve uma forte retração na composição salarial do aumento do grau de instrução de M (mestrado) para D (doutorado) em 32,91 p.p. em média. Para os demais períodos, encontramos aumentos de 55,87 e 10,55 p.p. A análise do grau de entrada mostrou também que o *learning by doing* afeta a remuneração destas profissões. Entre as ocupações com maior capacidade remuneratória da indústria brasileira baseada em inovação e P&D, no triênio de 2006-2008, encontram-se os diretores de pesquisa e desenvolvimento com superior completo. Eles se beneficiaram de maiores variações salariais do que os profissionais com mestrado (50,509 contra 46,207). Do mesmo modo, os biólogos com superior incompleto (43,177) apareceram em terceiro lugar nesse ranking, acima dos que possuem superior completo (39,654) e bem acima daqueles que possuem mestrado (27,413). Esse padrão repete-se para biotecnologistas, geneticistas, engenheiros civis, profissionais de informática, matemáticos, físicos, químicos e afins. De certo modo, esta assimetria na remuneração dos profissionais mais experientes e com maior grau de *learning by doing* em relação aos com melhor educação formal, porém menos experientes, ajuda a compreender o ponto levantado por Acemoglu e Autor (2014), mencionado no item 2 deste trabalho. E também espelha o fato de que a formação mais intensiva de mestres e doutores no Brasil é um fenômeno recente. É possível levantar a hipótese de que estes profissionais com qualificações mais altas ainda não

tiveram tempo suficiente de trabalho para conquistarem maiores remunerações do que aqueles com maior experiência e nível de *learning by doing*. Isso somado à escassez de mestres e doutores nesse mercado de trabalho e à importância da experiência nessas ocupações e temos um ambiente institucional que possibilita o fenômeno mencionado.

O exame dos PageRanks (PRs) comparou as variações dessa métrica estatística entre uma amostra com os cargos DP&D e Ger e entre uma e outra sem estes cargos. Isso foi feito no intuito de verificar duas hipóteses: (i) atestar os efeitos de *learning-by-doing*, e (ii) testar se a rede possui resiliência necessária para caracterizá-la como uma rede SF (rede livre de escala ou scale free, cuja análise fica para trabalhos futuros). Observamos que as maiores remunerações dos cargos DP&D e Gerentes derivam não tanto do grau de instrução, mas da experiência acumulada pelo *learning-by-doing*, uma vez que as variações de PRs sem contar os referidos cargos, são bem maiores para os níveis (à exceção de SI). Também confirmamos a relação positiva entre formação escolar avançada e maiores remunerações, o que reforça os resultados encontrados por Sahn e Alderman (1988), Veum (1999), Schömborg (2007), Guisan e Aguayo (2007), Tchernis (2010) e Phillipon e Resheff (2010). Também confirmamos a afirmação de Acemoglu e Autor (2014) de que os indivíduos que cursaram a mesma quantidade de anos em uma mesma escola podem ter níveis assimétricos de capital humano por causa de incentivos salariais que tornam atraente o processo de *learning-by-doing*. E observamos que pelo fato de que a rede complexa apresentar resiliência, mesmo após a exclusão de hubs importantes relativos aos cargos mencionados, isto reforça sua característica de uma rede SF (*scale-free* ou livre de escala), o que é um resultado que pretendemos verificar em trabalhos futuros.

Finalmente, na análise da centralidade dos autovetores, pudemos investigar o quanto a variação salarial de um agente que melhora seu capital humano por meio de instrução formal ou *learning by doing* pode estar conectada à variação salarial de seus *alters* (outros agentes diretamente relacionados ao agente central). Com efeito, a centralidade do autovetor é nula no nível EM (ensino médio), o que é óbvio, dado que os graus de entrada dos agentes são nulos. Já nos níveis educacionais seguintes, temos: SI (0.003104), SC (0.038817), M (0.241605) e D (igual a 1, pois os graus de saída são nulos). Então, novamente concluímos que o nível de mestrado é o mais importante para o aumento da variação salarial dos agentes. Basta verificar que os *alters* dos mestres são agentes com superior completo (SC, com valor igual a 0,038817, o segundo mais alto de todos) ou doutorado (igual a 1, o valor mais elevado).

7. Referências

- ALBERT, R. e BARABÁSI, A-L. “Statistical mechanics of complex networks”. **Reviews of Modern Physics**. 74: 47–97, 2002. doi:10.1103/RevModPhys.74.47. Disponível em: <<https://arXiv:cond-mat/0106096>>. Acesso em 15/02/2019.
- ALON, T. **Earning more by doing less: human capital specialization and the college wage premium**. Evanston, Illinois: Northwestern University, Jan. 2018. Mimeo. Disponível em: <https://economics.yale.edu/sites/default/files/january_jmp_draft-1n9gm2w.pdf>. Acesso em 08/02/2019.
- AMORIM, J. B. P. **O problema das quatro cores**. Tese de doutoramento. Coimbra: Universidade de Coimbra, 1953.
- ARBACHE, J. S.; DE NEGRI, J. A. **Diferenciais de salários interindustriais no Brasil: evidências e implicações**. 2002. (Texto para Discussão, nº 918).
- ARBACHE, J. S.; DE NEGRI, J. A. Filiação industrial e diferencial de salários no Brasil. **Revista Brasileira de Economia**, v. 58, n. 2, p. 159-184, 2004.

- ARELLANO, M.; BOVER, O. "Another look at the instrumental-variable estimation of error components model". **Journal of Econometrics**, v. 68, p. 29-52, 1995.
- AUSTIN, D. "How Google finds your needle in the web's haystack", **The American Mathematical Society**, 2019. <<http://www.ams.org/publicoutreach/feature-column/fcarc-pagerank>>. Acesso em 16/02/2019.
- BANERJEE, A. et al. The diffusion of microfinance. **Science**, v. 341, n. 6144, p. 1236498, 2013.
- BARABÁSI, A.-L. **Linked: How everything is connected to everything else and what it means for business, science and everyday life**. Plume Books, New York, 2003, 294p.
- _____ e BONABEAU, E. "Scale-free networks". **Scientific American**. 288 (5): 60-69. May, 2003. doi:10.1038/scientificamerican0503-60. Disponível em: <<http://eaton.math.rpi.edu/csums/papers/FoodWebs/barabasisciam.pdf>>. Acesso em 06/02/2019.
- _____ e ALBERT, R. "Emergence of scaling in random networks". **Science**, 286(5439): 509-512, 1999. Disponível em: <<https://arxiv.org/pdf/0906.0612.pdf>>. Acesso em 06/02/2019.
- _____. "Scale-free networks: a decade and beyond". **Science**, 325: 412-413. July, 2009. Disponível em: <<http://barabasi.com/f/303.pdf>> Acesso em 09/02/2019.
- BARRO, Robert J. "Economic growth in a cross section of countries". **Quarterly Journal of Economics**, 106, maio, pp. 407-443, 1991.
- _____. "Human capital and economic growth". In: **Policies for long-run economic growth**. Kansas City, Missouri: Federal Reserve Bank of Kansas City, 1992.
- _____. "Determinants of economic growth: a cross-country empirical study". **NBER Working Paper** 5698, August, 1996
- _____ e LEE, J. W. "A new data set of educational attainment in the world, 1950–2010". **Journal of development economics**, v. 104, p. 184-198, 2013.
- _____ e LEE, J. W. "International comparisons of educational attainment". **Journal of monetary economics**, v. 32, n. 3, p. 363-394, 1993.
- BECKER, G.S. "Investment in human capital: a theoretical analysis". **The Journal of Political Economy**, 70(5): 9-49, 1962.
- BLUNDELL, R.; BOND, S. "Initial conditions and moment restrictions in dynamic panel data models". **Journal of econometrics**, v. 87, n. 1, p. 115-143, 1998.
- BONACICH, P. "Power and centrality: a family of measures". **The American Journal of Sociology**, 92 (5): 1170-1182, 1987.
- BRIN, S. e PAGE, L. "The anatomy of a large-scale hypertextual web search engine". In: **Proceedings of the seventh International Conference on the World Wide Web**, 107-117, 1998.
- BUSACKER, R. G. e SAATY, T. L. **Finite graphs and networks: an introduction with applications**. New York: McGraw-Hill, 1965.
- CALLAWAY, D. S.; NEWMAN, M. E. J.; STROGATZ, S. H. e WATTS, D. J. 2000). "Network robustness and fragility: percolation on random graphs". **Physical Review Letter**, 85: 5468–71, 2000. doi:10.1103/PhysRevLett.85.5468. Disponível em: <<https://arXiv:cond-mat/0007300>>. Acesso em 06/02/2019.

- CLAUSET, A., NEWMAN, M. E. J. e MOORE, C. “Finding community structure in very large networks”. **Physical Review**, 70(6): 066111, 2004. Disponível em: <<https://arxiv.org/pdf/cond-mat/0408187.pdf>>. Acesso em 07/02/2019.
- CLAUSET, A.; SHALIZI, C. R. e NEWMAN, M. E. J. “Power-law distributions in empirical data”. **SIAM Review**. June, 2007. doi:10.1137/070710111. Disponível em: <<https://arXiv:0706.1062>>. Acesso em: 08/02/2019.
- COHEN, R. e HAVLIN, S. **Complex Networks: Structure, Robustness and Function**. Cambridge, Massachussets: Cambridge University Press, 2010.
- COHEN, D. e SOTO, M. “Growth and human capital: good data, good results”. **Journal of Economic Growth**, vol. 12, pag. 51–76, 2007.
- COHEN, R.; EREZ, K.; BEN-AVRAHAM, D. e HAVLIN, S. “Resilience of the Internet to Random Breakdowns”. **Physical Review Letter**, 85: 4626–8, 2000. doi:10.1103/PhysRevLett.85.4626. Disponível em: <<https://arXiv:cond-mat/0007048>>. Acesso em: 04/02/2019.
- COHEN, R.; EREZ, K.; BEN-AVRAHAM, D. e HAVLIN, S. “Breakdown of the Internet under Intentional Attack”. **Physical Review Letter**, 86: 3682–5, 2001. Disponível em: <<https://arXiv:cond-mat/0010251>>. Acesso em :07/02/2019.
- DOROGOVTSSEV, S. N. e MENDES, J. F. F. “Evolution of networks”. **Advances in Physics**, 51 (4): 1079-1187. June, 2002. doi:10.1080/00018730110112519. Disponível em: <<https://arxiv.org/abs/cond-mat/0106144>>. Acesso em 06/02/2019.
- EASLEY, D. e KLEINBERG, J. **Networks, crowds and markets: reasoning about a highly connected world**. Cambridge, Massachussets: Cambridge University Press, 2010.
- ENDERS, W. **Applied econometric time series**. 3rd ed. New York: Wiley, 2009. 554p.
- ERDŐS, P. e RÉNYI, A. “On the evolution of random graphs”. **Publications of the Mathematical Institute of the Hungarian Academy of Sciences**, Budapest, 1960
- ERDŐS, P. e RÉNYI, A. “On Random Graphs”. **Publicationes Mathematicae**. 6: 290–297, 1959.
- EXAME MAIORES E MELHORES. São Paulo. Anuário. Várias edições.
- FLAMENT, C. **Applications of graph theory to group structure**. New Jersey: Prentice-Hall, 1963.
- FORTUNATO, S. “Community detection in graphs”. **Physics Report**, 486(3-5): 75-174, 2010. Disponível em: <<https://arxiv.org/pdf/0906.0612.pdf>>. Acesso em: 08/02/2019.
- FRUCHTERMAN, T. M. J. and REINGOLD, E. M. “Graph drawing by force-directed placement”. **Software: Practice and Experience**, 21(11), 1991.
- GAMA, J., CARVALHO, A P. L., FACELLI, K. et al. **Extração de conhecimento de dados – data mining**. 2ª ed. Lisboa: Sílabo, 2015.
- GIRVAN, M. e NEWMAN, M. E. J. “Community structure in social and biological networks”. **Proceedings of NSF Workshop on Next Generation Data Mining**, p. 99-108, 2002.
- GITTLEMAN, M. e WOLFF, E. N. International comparisons of inter-industry wage differentials. **Review of Income and Wealth**, v. 39, n. 3, p. 295-312, 1993.
- GRANGER, C.W.J. Investigating causal relations by econometric models and cross-spectral methods. **Econometrica: Journal of the Econometric Society**, p. 424-438, 1969.

- GRILICHES, Z. "Education, human capital and growth: a personal perspective". **Journal of Labor Economics**, University of Chicago Press, vol. 15(1), pages 330-344, January, 1997.
- GUISAN, M.C. e AGUAYO, E. "Wages, productivity and human capital in the European Union: econometric models and comparison with USA – 1985-2005". **Applied Econometrics and International Development**, vol. 7-1. 2007. Disponível em: <<http://www.usc.es/economet/reviews/aeid714.pdf>>. Acesso em 14/02/2019.
- GUSSO, D. Agentes da inovação: quem os forma, quem os emprega? In: DE NEGRI, J. A.; DE NEGRI, F.; COELHO, D. (org.). **Tecnologia, exportação e emprego**. Brasília: IPEA, 2006
- HANUSHEK, E. A. KINKO D.D. "Schooling, labor-force quality and growth of nations". **American Economic Review**, v. 90, n. 3, 2000.
- HU, YIFAN. "Efficient, high-quality force-directed graph drawing". **The Mathematica Journal**, 10, v. 1, 2006. Disponível em: <http://www.mathematica-journal.com/issue/v10i1/contents/graph_draw/graph_draw.pdf>. Acesso em: 21/02/2019.
- IOANNIDES, Y. M e LOURY, L. D. Job information networks, neighborhood effects, and inequality. **Journal of economic literature**, v. 42, n. 4, p. 1056-1093, 2004.
- JACKSON, M. O. **Social and economic networks**. Princeton University Press, 2008.
- KANNEBLEY JR., S.; SHIMADA, E. e DE NEGRI, F. Efetividade da lei do bem no estímulo aos dispêndios em P&D: uma análise com dados em painel. **Pesquisa e Planejamento Econômico (IPEA)**, v. 46, n. 3, 2016.
- KOSSINET, G. e WATTS, D. J. "Neural network ensembles, cross validation and active learning". In: **Advances in Neural Information Processing Systems**, v. 7, p. 231-238, 1995.
- KRUEGER, A. B. e SUMMERS, Lawrence. H. Reflections on inter-industry wage structure. In Lang, K. & Leonard, J., editor, **Unemployment and the Structure of Labor Markets**. Basil Blackwell, Oxford, 1987.
- _____ e SUMMERS, Lawrence H. Efficiency wages and the inter-industry wage structure. **Econometrica: Journal of the Econometric Society**, p. 259-293, 1988.
- LESKOVEC, J., KLEINBERG, J. e FALOUTSOS, C. "Graphs over time: densification laws, shrinking diameters and possible explanations". In: **Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining**, p. 177-187, New York, NY, USA. ACM, 2005.
- LESSLEY, B. V., JOHNSON, D. M. e HANSON, J. C. **Using the Partial Budget to Analyze Farm Change: 1990-1991**. Maryland: University of Maryland System, pp. 7, 1991.
https://www.arec.umd.edu/sites/arec.umd.edu/files/docs/Using%20the%20Partial%20Budget_0.pdf.
- LUCAS JR, R. E. "On the mechanics of economic development". **Journal of monetary economics**, v. 22, n. 1, p. 3-42, 1988.
- MANKIW, N. G.; ROMER, D.; WEIL, D. N. "A contribution to the empirics of economic growth". **The quarterly journal of economics**, v. 107, n. 2, p. 407-437, 1992.
- MINCER, Jacob. "Investment in human capital and personal income distribution". **Journal of Political Economy**, v. LXVI, n. 4, p. 281-302, 1958.

- _____. “Human capital and economic growth”. **NBER Working Paper** n. 803, Nov. 1981
- NAKABASHI, L. e FIGUEIREDO, L. Capital Humano e Crescimento: impactos diretos e indiretos. **Economia Aplicada**, 12(1):151-171, fevereiro, 2008.
- _____. e SALVATO, M. A. “Human Capital Quality in the Brazilian States”. **Economia**, Brasília (DF), v. 8, n. 2, pp. 211-229, may-aug, 2007.
- NELSON, R. R.; PHELPS, E. S. “Investment in humans, technological Diffusion, and economic growth”. **The American economic review**, v. 56, n. 1/2, p. 69-75, 1966.
- NEWMAN, M. E. J. “Mixing patterns in networks”. **Physical Review**, 67(2): 026126, 2003a.
- NEWMAN, M. E. J. “The structure and function of complex networks”. **SIAM Review**, 45(23): 167-228, 2003b.
- NEWMAN, M. E. J. “Modularity and community structure in networks”. **Proceedings of the National Academy of Sciences of the United States of America**, 103(23): 8577-8528, 2006.
- NEWMAN, M. E. J. e GIRVAN, M. “Finding and evaluating community structure in networks”. **Physical Review**, 69(2): 026113, 2004.
- NONNEMAN, W. e VANHOUDT, P. “A further augmentation of the Solow model and the empirics of economic growth for OECD countries”. **The Quarterly Journal of Economics**, vol.111, Issue 3, 1, Pages 943–953, August 1996, Disponível em: <<https://doi.org/10.2307/2946677>> . Acesso em: 07/02/2019.
- MANKIW, N.G., D. ROMER and D.N. WEIL. “A contribution to the empirics of economic growth”. **Quarterly Journal of Economics**, 107(2): 407-437, 1992.
- PAULA, S. R.; UHR, D.; FIALHO, M.; PASSOS, M. O. e UHR, J.G.Z. “Canais de crescimento do capital humano: uma análise em painel dinâmico para o Brasil”. **Anais do 46º Encontro Nacional de Economia**. Rio de Janeiro, dez. 2018. Disponível em: <https://www.anpec.org.br/encontro/2018/submissao/files_I/i6-91eb78507652be81c311b5179a3e52d3.pdf>. Acesso em 22/02/2019.
- PEREIRA, J. M. S. **Grafos e redes: teoria e algoritmos básicos**. Rio de Janeiro: Interciência, 2013.
- PELINESCU, Elena. “The impact of human capital on economic growth”. **Procedia Economics and Finance**, 22, pp. 184-190, 2015.
- PERGAMIT, M.R. e VEUM, J. R. “What is a promotion?”. **ILR Review**, v. 52, issue: 4, pp. 581-601, July 1, 1999. Disponível em: <<https://journals.sagepub.com/doi/10.1177/001979399905200405>>. Acesso em 18/02/2019.
- PHILLIPON, T. e RESHEFF, A. “Wages and human capital in the US Financial Industry: 1909-2006”. **Quarterly Journal of Economics**, v. 127, Issue 4, pp. 1551-1609, Nov. 2012. Disponível em: <http://pages.stern.nyu.edu/~tphilipp/papers/pr_qje2012.pdf>. Acesso em: 14/02/2019.
- PONS, P. e LATAPY, M. “Computing communities in large networks using random networks”. **Journal of Graph Algorithms and Applications**, v. 10, n.2, p. 191-218, 2006.
- PRICE, D. D. S. “Networks of scientific papers”. **Science**, 149(3683): 510-515, 1965. Disponível em: <<http://garfield.library.upenn.edu/papers/pricenetworks1965.pdf>>. Acesso em: 09/02/2019.

- RABUSKE, M. A. **Introdução à teoria dos grafos**. Florianópolis: Editora da UFSC, 1992.
- ROMER, PAUL M. “Increasing returns and long-run growth”. **Journal of Political Economy**, 94, pp. 500/521, jun. 1986.
- _____. “Human capital and growth: theory and evidence”. **NBER Working Paper** n. 3173, Nov., 1989.
- RUOHONEN, K. **Graph Theory**. Tampere, Finlândia: Tampere University of Technology, 2013. Disponível em: <http://math.tut.fi/~ruohonen/GT_English.pdf> Acesso em: 14/02/2019.
- SAHN, D. and ALDERMAN, H. “The effects of human capital on wages, and the determinants of labor supply in a developing country”, **Journal of Development Economics**, Vol. 29, Issue 2, pp.157-183, 1988. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0304387888900338>>. Acesso em 14/02/2019.
- SALA-I-MARTIN, X. “I Just Ran Two Million Regressions”. **The American Economic Review**, vol. 87, n. 2, pp. 178-183, 1997.
- _____. **Apuntes de crecimiento económico**. 2. Ed. Barcelona: Antoni Bosch, 2000.
- _____. e MULLIGAN, C. B. “Measuring Aggregate Human Capital”. **Center Discussion Paper**, 1995.
- SCHOMBERG, U. “Wage Growth Due to Human Capital Accumulation and Job Search: A Comparison between the United States and Germany”, **Industrial & Labor Relations Review**, Vol. 60, Issue 4, 2007. Disponível em: <<http://discovery.ucl.ac.uk/17154/1/17154.pdf>> Acesso em: 22/02/2019.
- SCHULTZ, T. W. “Capital formation by education”. **The Journal of Political Economy**, 68(6), p. 571–583, 1960.
- _____. “Investment in human capital”. **The American Economic Review**, vol. 51, n. 1, p. 1-17, mar. 1961.
- SCOTT, J. Social network analysis. **Sociology**, v. 22, n. 1, p. 109-127, 1988.
- SOLOW, R. M. “A contribution to the theory of economic growth”. **Quartely Journal of Economics**, n.70, v.1, p. 65-94, fev.1956.
- SORAMÄKI, K. et. al. “The topology of interbank payment flows”. *Physica A: Statistical Mechanics and its Applications*. 379 (1), 2007, p. 317–333. Bibcode: 2007PhyA..379..317S. doi:10.1016/j.physa.2006.11.093
- STEYVERS, M.; TENENBAUM, J. B. “The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth”, **Cognitive Science**, 29 (1), 2005, p. 41–78. Disponível em: <https://doi:10.1207/s15516709cog2901_3>. Acesso em: 16/02/2019.
- TCHERNIS, R. “Measuring human capital and its effects on wage growth”, **Journal of Economic Surveys**, Vol. 24, No. 2, 2010, pp. 362–387. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-6419.2009.00617.x>> Acesso em: 08/02/2019.
- VEUM, J. R. “Training, Wages, and the Human Capital Model”, **Southern Economic Journal**, vol. 65, no. 3, 1999, pp. 526–538. Disponível em: <www.jstor.org/stable/1060813>. Acesso em: 06/02/2019.

- WASSERMAN, S. e FAUST, K. **Social network analysis: methods and applications**. Cambridge, Massachussets: Cambridge University Press, 1994.
- WATTS, D. J. “Networks, dynamics and the small-world phenomenon”. **American Journal of Sociology**, v. 105, Issue 2, 493-527, Sep. 1999.
- _____ e STROGATZ, S. H. “Collective dynamics of small-world networks”. **Nature**, 193(6684): 440-442, 1998.
- WEST, D. B. et al. **Introduction to graph theory**. Upper Saddle River, NJ: Prentice hall, 1996.
- .