

**Universidade Federal da Paraíba – Campus I**

**Curso de Pós-Graduação em Informática**

**André Calisto Souza Medeiros Guedes**

**Um Estudo sobre os Impactos dos Relacionamentos  
Sociais na Avaliação do Mérito Científico**

**João Pessoa**

**2014**

André Calisto Souza Medeiros Guedes

**Um Estudo sobre os Impactos dos Relacionamentos  
Sociais na Avaliação do Mérito Científico**

Trabalho de Dissertação apresentado  
à Universidade Federal da Paraíba – Campus I  
como requisito à obtenção do título de Mestre no  
curso de Pós-Graduação em Informática.

Orientador: *Prof. Dr. Alexandre  
Nóbrega Duarte.*

**João Pessoa**

**2014**

*G924u Guedes, André Calisto Souza Medeiros.*

*Um estudo sobre os impactos dos relacionamentos sociais na avaliação do mérito científico / André Calisto Souza Medeiros Guedes.- João Pessoa, 2014.*

*86f. : il.*

*Orientador: Alexandre Nóbrega Duarte*

*Dissertação (Mestrado) - UFPB/CI*

## **AGRADECIMENTOS**

Primeiramente, agradeço a Deus por ter me proporcionado saúde e condições para que até aqui eu chegasse. Agradeço também a toda minha família que, de uma forma ou de outra, me incentivaram e me apoiaram para o término dessa primeira jornada.

Agradecimento de forma especial a Professor DR. Alexandre Nóbrega, que teve paciência em me orientar da melhor forma possível, mostrando quais os caminhos eram importante ser traçados para que o trabalho chegasse a seu objetivo.

Não poderia deixar também de agradecer aos meus amigos que sempre estavam ali para dar uma palavra de apoio, à eles, meu muito obrigado: Ana Paola, Severino Félix, Hidelberg de Oliveira, Douglas dos Santos, Saskya Gurgel, Renata Mendonça, Luciano Fernandes, Danilo Dantas, Pablo Andrey, enfim a todos que contribuíram direta e indiretamente para que essa monografia pudesse ter sua conclusão de forma exitosa.

Agradeço também de forma muito especial e sentimental ao meu tio DR. Floriano Camelo de Souza Neto, que mostrou ser um homem de caráter e dignidade. Soube de forma aguerrida lutar contra uma doença covarde e traiçoeira. E mesmo que ela tenha tirado meu tio do meu convívio e de minha família, ela não conseguiu destruir todo o legado que ele nos deixou e que ficará em nossas memórias, corações, como também por meio de seus vários textos que foram escritos de forma excepcional.

Enfim, meu muito obrigado à todos que de alguma forma contribuíram para que eu pudesse ter o sucesso alcançado nesta monografia.

## **RESUMO**

Técnicas de análise de redes sociais podem ser utilizadas para extrair conhecimento da rede por meio da análise das interações entre seus membros. Neste trabalho pretendemos aplicar técnicas de análise de redes sociais a redes formadas pela colaborações científicas entre pesquisadores. Pretendemos com isso avaliar se as relações de colaboração entre pesquisadores possuem alguma influência na avaliação do mérito científico. Utilizaremos como estudos de caso o processo de concessão das chamadas bolsas de produtividade em pesquisa do CNPQ e a avaliação dos programas de pós-graduação pela CAPES, ambos focados na área de Ciência da Computação.

**PALAVRAS CHAVES:** Análise de Redes Sociais, Redes de Colaboração Científica, Avaliação do Mérito Científico

## **ABSTRACT**

Social network analysis techniques can be used to extract knowledge from the network by analyzing the interactions between its members. In this work we intend to apply techniques of social network analysis on networks formed through scientific collaborations among researchers. Our objective with this study is to analyze if these collaboration connections may have any influence in the evaluation of scientific merit. We will focus our work on two specific scenarios. The first scenario targets the evaluation of individual merit for granting scholarships to researcher and is conducted by CNPQ, the Brazilian Research Council. The second scenario focus on the evaluation of graduate programs in the country and is conducted by CAPES, the governmental body entitled to evaluate and coordinate the graduate education program in Brazil. On both scenarios we will restrict our analysis to the Computer Science area.

**KEYWORDS:** Social Network Analysis, Scientific Collaboration Networks, Evaluation of Scientific Merit

## SUMÁRIO

<b>CAPÍTULO 1 – INTRODUÇÃO .....</b>	<b>9</b>
1.2 DEFINIÇÃO DO PROBLEMA.....	12
1.3 OBJETIVOS.....	13
1.3.1 OBJETIVO GERAL.....	13
1.3.2 OBJETIVOS ESPECÍFICOS.....	13
1.4 QUESTÕES DE PESQUISA .....	13
1.5 ESTRUTURA DO TRABALHO DE DISSERTAÇÃO .....	14
<b>CAPÍTULO 2 – FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>15</b>
2.1 TEORIA DOS GRAFOS.....	15
2.1.1 DEFINIÇÕES BÁSICAS.....	16
2.1.2 REPRESENTAÇÃO DE GRAFOS .....	21
2.2 ANÁLISE DE REDES SOCIAIS .....	23
2.2.1 COMO SURTIRAM AS REDES SOCIAIS.....	23
2.2.2 MÉTODOS DE ANÁLISE DE REDES SOCIAIS .....	24
2.2.3 MÉTRICAS DE ANÁLISE.....	26
2.3 AVALIAÇÃO DO MÉRITO CIENTÍFICO NO BRASIL.....	29
2.3.1 Bolsas de Produtividade em Pesquisa.....	29
2.4 CONCLUSÃO DO CAPITULO.....	31
<b>CAPÍTULO 3 – TRABALHOS RELACIONADOS .....</b>	<b>32</b>
3.1 ESTUDO DA REDE DE COAUTORIA E DA INTERDISCIPLINARIDADE NA PRODUÇÃO CIENTÍFICA COM BASE NOS MÉTODOS DE ANÁLISE DE REDES SOCIAIS: AVALIAÇÃO DO CASO DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÕES – PPGCI/UFMG .....	32
3.2 FORMAÇÃO DE REDES SOCIAIS DE COAUTORIA NA ÁREA DE GESTÃO DE PESSOAS: UMA ANÁLISE BIBLIOMÉTRICA EM PERIÓDICOS BRASILEIROS NO TRIÊNIO DE 2007 A 2009 .....	33
3.3 REDE SOCIAL DE COAUTORIA EM CIÊNCIA DA INFORMAÇÃO: ESTUDO SOBRE A ÁREA TEMÁTICA DE “ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO” .....	34
3.4 REDES SOCIAIS NA PRODUÇÃO CIENTÍFICA DOS PROGRAMAS DE PÓS- GRADUAÇÃO DE CIÊNCIAS CONTÁBEIS DO BRASIL.....	35
3.5 REDES SOCIAIS E INTELECTUAIS EM ADMINISTRAÇÃO DA INFORMAÇÃO: UMA ANÁLISE DO PERÍODO 1997-2006 .....	36
3.6 ANÁLISE DAS REDES DE COLABORAÇÃO CIENTÍFICA ATRAVÉS DO ESTUDO DAS CO- AUTORIAS, NOS CURSOS DE PÓS-GRADUAÇÃO DO BRASIL NO TEMA TRATAMENTO TEMÁTICO DA INFORMAÇÃO .....	37

3.7 ANÁLISE TEMPORAL DA RELAÇÃO ORIENTADOR-ORIENTADO: UM ESTUDO DE CASO SOBRE A PRODUTIVIDADE DOS PESQUISADORES DOUTORES DA ÁREA DE CIÊNCIA DA COMPUTAÇÃO.....	38
3.8 COAUTORIA COMO INDICADOR DE REDES DE COLABORAÇÃO CIENTÍFICA.....	39
3.9 COMPARAÇÃO DOS TRABALHOS RELACIONADOS .....	40
<b>CAPÍTULO 4 – METODOLOGIA .....</b>	<b>43</b>
4.1 INTRODUÇÃO .....	43
4.1.1 <i>Seleção da Amostra</i> .....	43
4.1.2 <i>Geração das Redes</i> .....	43
4.1.3 <i>Métricas e Consultas</i> .....	45
<b>CAPÍTULO 5 – RESULTADOS.....</b>	<b>48</b>
5.1 INTRODUÇÃO .....	48
5.2 AS REDES SOCIAIS DE PESQUISA.....	48
5.3 ANÁLISE DE VARIÂNCIA .....	49
5.3.1 <i>Periódicos</i> .....	50
5.3.2 <i>Conferências</i> .....	52
5.3.3 <i>Orientações de Mestrado</i> .....	53
5.3.4 <i>Orientações de Doutorado</i> .....	54
5.3.5 <i>Bancas de Mestrado</i> .....	56
5.3.6 <i>Bancas de Doutorado</i> .....	57
5.3.7 <i>Centralidade de Grau</i> .....	58
5.3.8 <i>Centralidade de Proximidade</i> .....	65
5.3.9 <i>Centralidade de Intermediação</i> .....	66
5.3.10 <i>Centralidade de AutoVetor</i> .....	69
5.3.11 <i>PageRank</i> .....	70
5.3.12 <i>Número de Triângulos</i> .....	72
5.5 CONCLUSÃO DO CAPÍTULO.....	75
<b>CAPÍTULO 6 – CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>82</b>
6.1 CONCLUSÕES.....	82
6.2 TRABALHOS FUTUROS.....	83
<b>REFERÊNCIAS .....</b>	<b>84</b>

## Capítulo 1 – Introdução

Uma rede social é uma estrutura social entre atores, em sua maioria, indivíduos ou organizações. A rede indica as formas pelas quais tais atores estão conectados por meio de vários relacionamentos sociais (Lacombe et al., 2003). A World Wide Web (WWW) facilita o crescimento de comunidades virtuais, onde os indivíduos com interesses comuns podem se comunicar de forma síncrona ou assíncrona por vários meios, tais como e-mail, mensagens e compartilhamento de arquivos. Enquanto as pessoas interagem, elas desenvolvem suas relações diretamente. Além destas comunicações diretas, quando os indivíduos contribuem acessando outras informações eles também desenvolvem relações indiretamente. Estas relações indiretas se tornam significativas com a quantidade e frequência das interações (Hanneman et al., 2005).

A análise de redes sociais é o mapeamento e medição de relações e fluxos entre pessoas, grupos, organizações, animais, computadores ou outras entidades. Os nós da rede representam tais entidades enquanto os links mostram relações ou fluxos entre os nós. A análise de redes sociais proporciona tanto uma abstração do que acontece entre as entidades como uma análise matemática das relações de interação entre os participantes da rede (Lacombe et al., 2003).

Técnicas de análise de redes sociais podem ser utilizadas para extrair conhecimento da rede pela análise das interações entre seus membros. Esta análise muitas vezes pode ser feita sem considerar as propriedades individuais de cada membro, considerando apenas como e o quanto cada membro se relaciona com outros membros da rede.

Neste trabalho pretendemos aplicar técnicas de análise de redes sociais a redes formadas por meio das colaborações científicas entre pesquisadores. Tais relações podem assumir diversas formas, como orientação em programas de pós-graduação, coautorias em artigos científicos, participações em bancas ou comissões de avaliação, etc.

Esperamos com isso identificar se os relacionamentos entre pesquisadores de uma determinada área têm algum efeito, direto ou indireto, na avaliação do mérito científico de seus pares.

Particularmente, pretendemos nos focar na análise dos relacionamentos entre pesquisadores na área de Ciência da Computação visando identificar possíveis efeitos desses relacionamentos na concessão das chamadas bolsas de produtividade em pesquisa do CNPQ.

As próximas seções abordarão o escopo do trabalho, apresentando em que consiste o trabalho, a metodologia utilizada para o seu desenvolvimento, os objetivos e a estrutura do restante deste documento.

## **1.1 CONTEXTUALIZAÇÃO DO TRABALHO**

Na trajetória histórica da humanidade a ciência vem assumindo múltiplas definições, envolvendo referências metodológicas, ideológicas, filosóficas e as mais variadas técnicas.

A ciência, entendida como conhecimento da natureza e exploração desse mesmo conhecimento, envolve três aspectos básicos: uma história, um método de investigação e uma comunidade de investigadores (Kneller, 1980). A ciência apresenta-se, então, como uma sucessão de ações dentro do movimento histórico mais amplo da própria civilização.

A atividade básica da ciência é a pesquisa, que pode ser conceituada como o processo formal e sistemático de desenvolvimento do método científico. E seu objetivo fundamental é descobrir respostas para problemas mediante o emprego de distintos procedimentos metodológicos (Texeira, 2003).

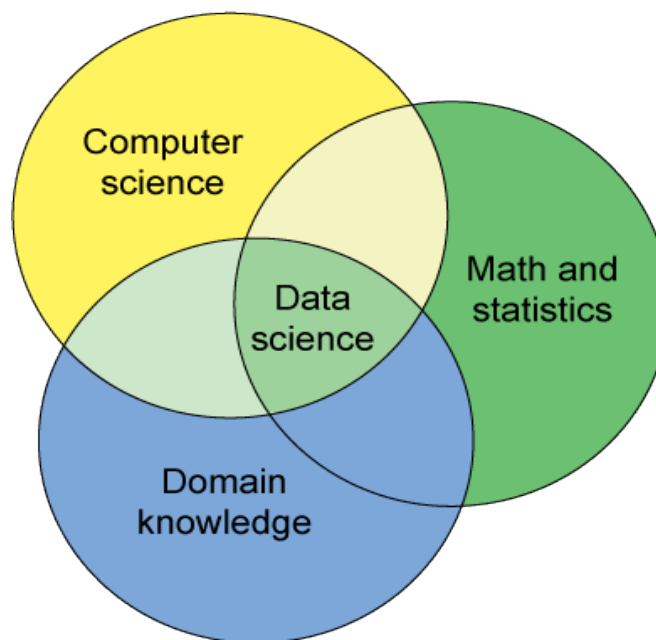
A atividade de pesquisa é empreendida no intuito de descobrir e construir novos conhecimentos, para tal é necessário desenhar ou projetar o caminho a ser seguido, uma vez que cada caminho poderá levar o investigador a alcançar diferentes resultados, devendo assim avaliar as restrições e oportunidades colocadas pelo contexto dentro do qual pretende trabalhar (Texeira, 2003).

Desta maneira, a pesquisa compõe-se de três momentos: fase exploratória da pesquisa, trabalho de campo e tratamento do material. E é na fase de tratamento do material que o pesquisador analisa e interpreta os dados coletados na pesquisa. A análise tem como objetivo organizar e resumir os dados de tal forma que possibilitem o fornecimento de respostas ao problema proposto para investigação (Minayo, 1994).

Para essa análise e interpretação dos dados é necessário da estatística que é uma ciência que se dedica à coleta, análise e interpretação de dados e utiliza de teorias probabilísticas para explicar a frequência de fenômenos e para possibilitar a previsão desses fenômenos com base na sumarização e a interpretação de observações (Triola, 2008).

Mas com o avanço da tecnologia os dados hoje ficaram maiores e mais heterogêneo, saindo de informações só numéricas e textuais. Assim, Ciência de dados (Data Science) está surgindo para enfrentar os desafios do processamento de grandes conjuntos de

dados, ou seja, "Big Data", que consiste de dados estruturados, não estruturados ou semiestruturados que as grandes empresas produzem. Buscando entender a explosão de novos dados gerados a partir de dispositivos inteligentes, web, mobile e mídias sociais. Portanto, Ciência dos Dados é o estudo da extração generalizável de conhecimento a partir de dados. Essa ciência incorpora elementos diferentes e baseia-se em técnicas e teorias de diversas áreas, incluindo processamento de sinal, matemática, modelos de probabilidade, aprendizagem de máquina, aprendizado estatístico, programação de computadores, engenharia de dados, reconhecimento de padrões e aprendizagem, visualização, modelagem de incerteza, armazenamento de dados e computação de alto desempenho, com o objetivo de extrair significado a partir de dados e criação de produtos de dados, ampliando o domínio da estatística no avanço da computação com os dados (Cleveland, 2001). A figura 1.1 mostra a interseção das disciplinas chaves da Data Science: Ciências da Computação (Computer Science), Matemática e Estatística (Math and Statistics) e Conhecimento Interdisciplinaridade (Domain Knowledge).



**Figura 1.1.** Interseção das disciplinas chaves para Data Science.

O avanço da ciência da informação e das ciências em geral se dá pela constante elaboração de novas pesquisas e pela concretização e divulgação de seus resultados que se processam em diferentes tipos de suportes. Publicar os resultados de suas pesquisas é um compromisso que os cientistas são compelidos a cumprir. O avanço do conhecimento produzido pelos pesquisadores tem de ser transformado em informação acessível para a comunidade científica (Chapula, 1998).

Portanto, os estudos quantitativos da produção científica (Cienciometria) têm permitido entender melhor a amplitude e a natureza das atividades de pesquisa desenvolvidas nas diferentes áreas do conhecimento, de diversos países, instituições e pesquisadores (NORONHA *et al.*, 2000). Bem como, mede a difusão do conhecimento científico e o fluxo da informação sob enfoques diversos (VANTI, 2002). A Cienciometria trás indicadores que podemos analisar e estudar para poder compreender essas atividades científicas, como por exemplo, o **número de trabalhos** (livros, artigos, publicações científicas, relatórios etc.) publicados, **número de citações** (impacto dos artigos ou assuntos citados), **Co-autoria** (grau de colaboração na ciência em nível nacional e internacional), entre outras (Chapula, 1998).

A análise dos dados e cienciométricos oferece informações sobre a orientação e a dinâmica científica de um país, bem como sobre sua participação na ciência e na tecnologia mundial. Análises cooperativas tornam possível identificar redes científicas e revelar os elos entre países, instituições e pesquisadores, assim como permitem conhecer o impacto dos principais programas e organizações. A cienciometria também traz à luz a estrutura das disciplinas científicas e as conexões entre elas (Spinak, 1998).

Desta maneira, se mostra de grande importância o estudo proposto pelo trabalho, pois está totalmente inserido na Ciência dos Dados, utilizando da computação para obtenção dos dados, análise e inferência estatísticas das atividades acadêmicas que o pesquisador realiza e que a cienciometria afirma ser importante. Portanto, o trabalho se mostra bastante relevante, pois será possível estudar e entender um pouco melhor a rede dos pesquisadores científicos bolsistas de produtividade do CNPQ na área de computação e se essa rede de colaboração têm algum impacto no mérito científico, por intermédio da análise de rede social.

### **1.3 DEFINIÇÃO DO PROBLEMA**

A definição do problema a ser pesquisado neste trabalho está mais bem apresentada a seguir, utilizando-se o modelo *Goal, Question, Metric* (CALDIERA; ROMBACH, 1994):

**Analisar** os relacionamentos entre pesquisadores brasileiros na área de computação **com a intenção** de avaliar a influência desses relacionamentos **com respeito a** avaliação do mérito científico **do ponto de vista da** análise de redes sociais **no contexto** dos programas de concessão de bolsas de produtividade em pesquisa do CNPQ.

## 1.4 OBJETIVOS

### 1.4.1 OBJETIVO GERAL

O objetivo geral deste trabalho é avaliar se é possível identificar algum impacto dos relacionamentos sociais entre pesquisadores na avaliação do mérito científico de seus pares.

### 1.4.2 OBJETIVOS ESPECÍFICOS

Com a finalidade de alcançar o objetivo geral, os seguintes objetivos específicos foram estabelecidos:

1. Analisar se existem métricas de análise de redes sociais que diferenciem os bolsistas de produtividade em pesquisa dos demais pesquisadores na área de Ciência da Computação.
2. Analisar quais os bolsistas diferenciam em relação ao outro bolsista para cada variável estudada.

## 1.5 Questões de Pesquisa

Elencamos a seguir as questões de pesquisas que serão avaliadas neste trabalho:

QP1: Bolsistas de produtividade nível 1 têm mais colaboradores que a média dos bolsistas de produtividade nível 2 e DT na área?

QP2: Bolsistas de produtividade nível 1 têm mais colaboradores bolsistas de produtividade do que a média dos bolsistas de produtividade nível 2 e DT?

QP3: A proporção de pesquisadores que têm mais colaboradores que seus colaboradores é maior entre os bolsistas de produtividade nível 1 do que entre os demais pesquisadores bolsistas de produtividade?

QP4: Existem diferenças nas métricas de centralidade dos bolsistas de produtividade nível 1 dos bolsistas de produtividade nível 2 e DT?

## **1.6 ESTRUTURA DO TRABALHO DE DISSERTAÇÃO**

Este trabalho possui mais 5 capítulos, além deste capítulo introdutório. São eles:

### **Capítulo 2 – Fundamentação Teórica**

Este capítulo aborda todo o conteúdo que serve como alicerce para a realização do trabalho. Serão apresentados os principais conceitos relacionados à análise de redes sociais e a avaliação do mérito científico no Brasil, focando no processo de concessão de bolsas de produtividade em pesquisa pelo CNPQ.

### **Capítulo 3 – Trabalhos Relacionados**

Este capítulo apresenta os principais trabalhos relacionados a esta dissertação. São discutidos 8 (oito) trabalhos relacionados a análise de redes sociais, principalmente de redes de coautoria e cocitação.

### **Capítulo 4 – Metodologia**

Este capítulo detalha todo o caminho e etapas que o trabalho vai percorrer para alcançar os objetivos propostos.

### **Capítulo 5 – Resultados**

Este capítulo apresenta os resultados obtidos no trabalho e que serviram para atingir os objetivos do trabalho, como responder as questões de pesquisa que foram elaboradas neste trabalho.

### **Capítulo 6 – Conclusões e Trabalhos Futuros**

Este capítulo apresenta as conclusões obtidas no trabalho e quais os trabalhos futuros a serem realizados.

## Capítulo 2 – Fundamentação Teórica

Este capítulo aborda todo o conteúdo que serve como alicerce para a realização do trabalho. Serão apresentados os principais conceitos relacionados à análise de redes sociais e a avaliação do mérito científico no Brasil, focando no processo de concessão de bolsas de produtividade em pesquisa pelo CNPQ e na avaliação dos programas de pós-graduação pelas CAPES.

### 2.1 TEORIA DOS GRAFOS

A teoria dos grafos é um ramo da matemática que estuda as relações entre os objetos de um determinado conjunto, empregando estruturas chamadas de grafos,  $G(V,A)$ , onde  $V$  é um conjunto não vazio de objetos denominados vértices e  $A$  é um conjunto de pares não ordenados de  $V$ , chamado arestas (Biggs, 1986).

Essa teoria surgiu em 1736 por meio de um problema famoso da matemática denominado Sete pontes de Königsberg e solucionado por Leonhard Euler. O problema é baseado na cidade de Königsberg, que é cortada pelo Rio Prególia, onde há duas grandes ilhas que na época continha sete pontes, conforme mostra a figura 2.1 (Euler, 1741).

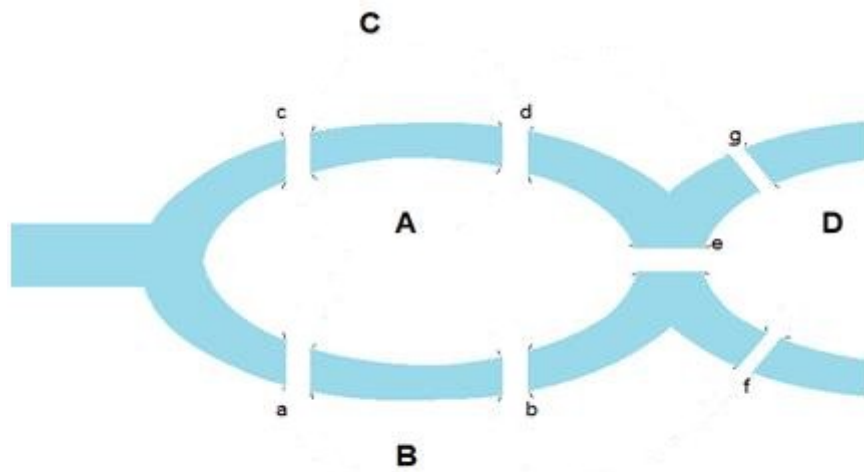


Figura 2.2. As setes pontes de Königsberg

O problema consistia em percorrer todas as pontes passando uma única vez em cada uma delas. Desta maneira Leonhard Euler utilizou-se de um raciocínio simples, os caminhos (que seria as 4 regiões de terra (A à D) ele transformou em retas e suas intersecções em pontos (que seria as pontes (a à g)), gerando desta maneira o primeiro grafo da história como mostra a figura 2.2.

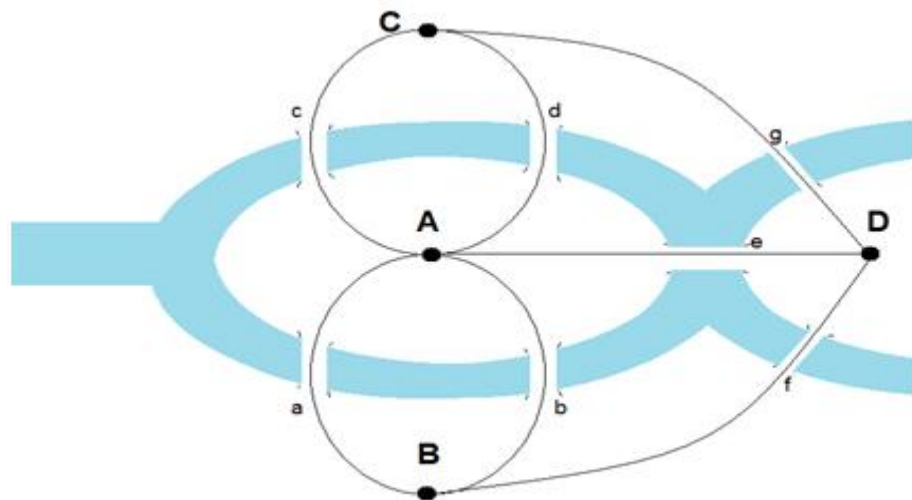


Figura 2.3. Grafo do problema das pontes de Königsberg

Leonhard Euler percebeu que só seria possível atravessar o caminho inteiro passando uma única vez em cada ponte se houvesse exatamente zero ou dois pontos de onde saísse um número ímpar de caminhos, porque é necessário que cada ponto deve haver um número par de caminhos, pois será preciso um caminho para "entrar" e outro para "sair".

### 2.1.1 DEFINIÇÕES BÁSICAS

Nesta subseção será descrito algumas definições básicas de grafos com base nas referências (Bondy & Murty, 1979).

**Definição 1.** Um grafo  $G$  é uma tripla ordenada de conjuntos finitos disjuntos  $(V, E, V \times E)$ , satisfazendo:

$$1 \leq |\{v \in V : (v, e) \in V \times E\}| \leq 2$$

para todo  $e \in E$ .

**Notação:** Os elementos do conjunto  $V := V(G)$  são chamados de vértices de  $G$  e os elementos de  $E := E(G)$  são chamados de arestas de  $G$ . Denomina-se  $V \times E := (V \times E)(G)$  como a lei de incidência do grafo  $G$ .

**Definição 2.** É dito que a aresta  $e$  é incidente ao vértice  $v$ , quando  $(v, e) \in V \times E$ .

**Notação:** Quando a aresta  $(e)$  for incidente a vértices distintos  $x$  e  $y$ , denota-se tal aresta por  $e = xy$ . Assim,  $x$  e  $y$  são chamados extremos desta aresta. É dito que a aresta  $e$  é um laço, quando  $e$  seja incidente a um único vértice  $x$ , denota-se tal aresta por  $e = xx$ , ou seja, uma aresta cujos extremos são representados pelo mesmo vértice, como por exemplo as arestas  $c$  ilustrado na figura 2.3. Outra possibilidade, é que dois vértices sejam unidos por

mais de uma aresta, chamadas de arestas em paralelo, como por exemplo as arestas a e b, demonstrado na figura 2.3.

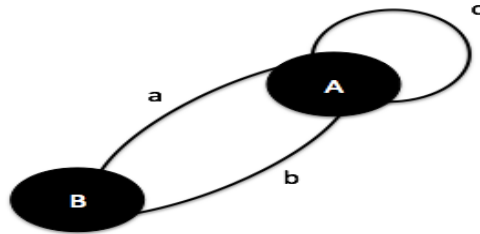


Figura 2.4. Exemplo de arestas em laço e paralelo.

**Definição 3.** Um grafo é dito vazio, quando contém exclusivamente vértices.

**Definição 4.** Um grafo é dito nulo, quando não possui vértices.

**Definição 5.** Um grafo é dito simples, quando não possui laços ou arestas paralelas.

**Definição 6.** Um grafo H é chamado subgrafo de G quando existe  $X \subseteq E(G)$  tal que  $H = G \setminus X$ . Como mostra a figura 2.4.

**Lembrete:** Por definição, todo subgrafo de G tem o mesmo conjunto de vértices de G.

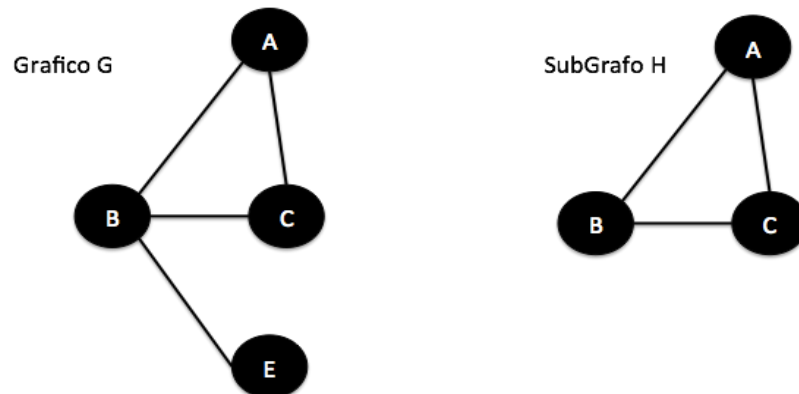


Figura 2.5. Exemplo de um subgrafo.

**Definição 7.** Um grafo é dito conexo se para todo par  $\{x,y\}$  de vértices distintos existe um caminho de x para y. Como mostra a figura 2.5.

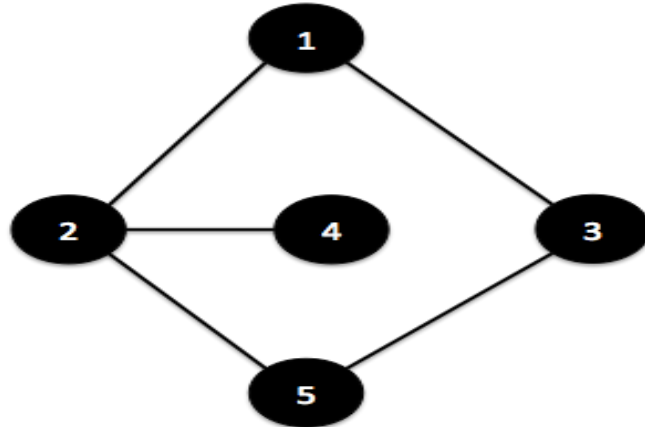


Figura 2.6. Exemplo de Grafo conexo.

**Definição 8.** Um grafo  $G=(V, E)$  é dito ponderado, quando existem valores numéricos (pesos) associados às suas arestas ou vértices. Como mostra a figura 2.6.

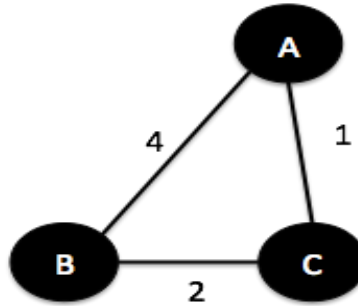


Figura 2.7. Exemplo de um grafo ponderado.

**Definição 9.** Um grafo  $G=(V, E)$  é dito direcionado ou orientado, quando o sentido das ligações entre os vértices é importante. Nesse caso, as arestas possuem um sentido marcado por uma seta denominada de **Arcos**. Como mostra a figura 2.7.

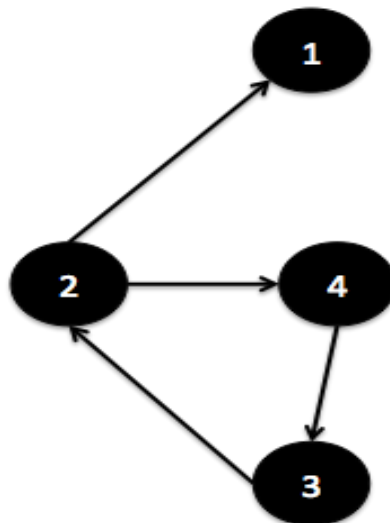


Figura 2.8. Exemplo de um grafo direcionado ou orientado.

**Definição 10 (a).** Dois vértices  $i$  e  $j$  são vizinhos ou adjacentes quando existe uma aresta que liga  $i$  a  $j$  ou vice-versa.

**Notação:** O conjunto de vértices vizinhos do vértice  $i$  será denominado  $\Gamma(i)$  ou  $N(i)$ . Como por exemplo na figura 2.5 os vizinhos  $\Gamma(B) = \{A, C \text{ e } E\}$ .

**Definição 10 (b).** Duas arestas  $a_i$  e  $a_j$  são adjacentes quando compartilham um vértice.

**Definição 11 (a).** Um **passeio** ou **percurso** é uma sequência finita de vértices e arestas  $x_0, a_1, x_1, a_2, \dots, x_{k-1}, a_k, x_k$ , começando e terminando com vértices, tais que,  $x_{i-1}$  e  $x_i$  são os vértices terminais da aresta  $a_i$ ,  $1 \leq i \leq k$ .

**Definição 11 (b).** Uma **cadeia** ou **trilha** é um passeio sem repetições de arestas. E um **caminho** é uma cadeia sem repetições de vértices.

**Observações:** Em um grafo  $G=(V, E)$ , não ponderado, o **comprimento de um caminho** é o número de arestas desse caminho.

Em um grafo  $G=(V, E)$ , ponderado, o **comprimento de um caminho** é a soma dos pesos das arestas desse caminho.

**Definição 12.** A distância entre um par de vértices  $x_i$  e  $x_j$ , denotada por  $d(x_i, x_j)$ , corresponde ao **caminho de menor comprimento** capaz de ligar  $x_i$  a  $x_j$ . No hipótese de o caminho não existir, então  $d(x_i, x_j) = \infty$ .

**Definição 13.** A **Excentricidade** de um vértice  $v \in N$ ,  $Ex(v)$ , é a maior distância entre  $v$  e  $w$ , para todo  $w \in N$ . Como mostra a figura 2.8,  $Ex(v) = 4$ .

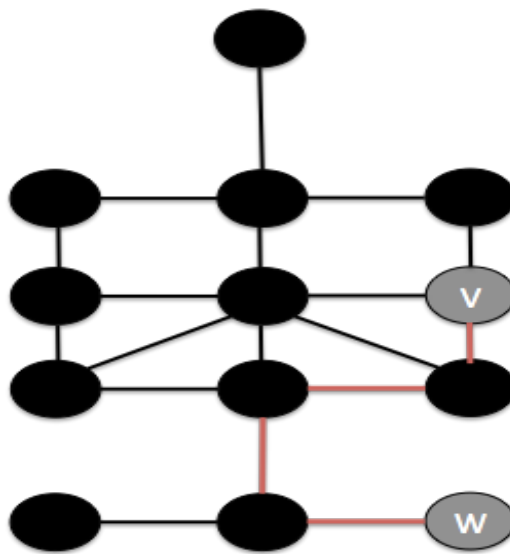


Figura 2.9. Grafo com  $Ex(v) = 4$ .

**Definição 14.** O **Raio** de um grafo  $G$ ,  $\text{Rad}(G)$ , é o **menor valor de excentricidade** para todo vértice  $v \in N$ . Como mostra a figura 2.9,  $\text{Rad}(G) = 3$ .

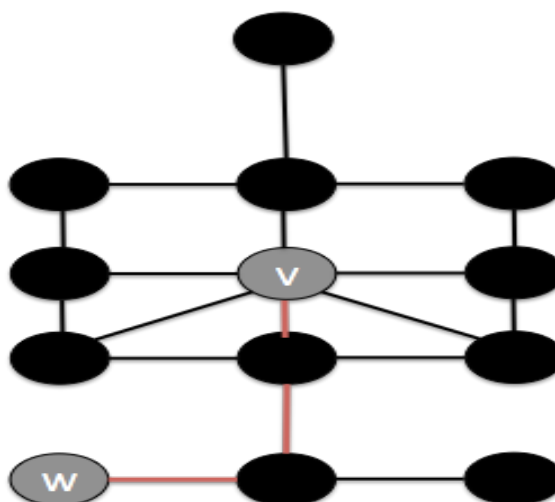


Figura 2.10. Grafo com  $\text{Rad}(G) = 3$ .

**Definição 15.** O **Diâmetro** de um grafo  $G$ ,  $\text{Diam}(G)$ , é o **maior valor de excentricidade** para todo vértice  $v \in N$ . Como mostra a figura 2.10,  $\text{Diam}(G) = 5$ .

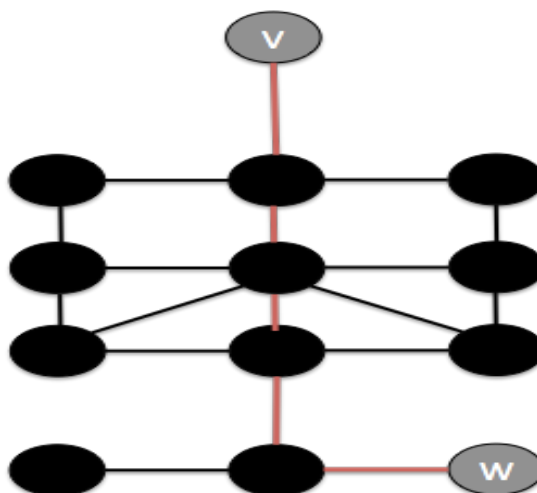


Figura 2.11. Grafo com  $\text{Diam}(G) = 5$

**Definição 16.** O **Centro** de um grafo  $G$ ,  $\text{Centro}(G)$ , é o **subconjunto dos vértices de excentricidade mínima**. Como mostra a figura 2.11.

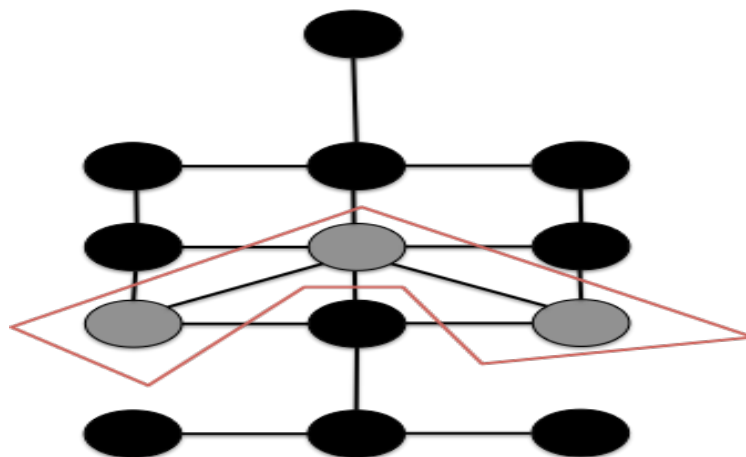


Figura 2.12. Exemplo de Centro (G)

## 2.1.2 REPRESENTAÇÃO DE GRAFOS

Segundo (Cormen, 2002) existem duas maneiras padrão de representar um grafo  $G=(V, E)$ : com uma matriz de adjacência ou com uma lista de adjacência. Essas representações serão mais detalhadas e exemplificadas nesta subseção do trabalho.

### 2.1.2.1 Matriz de Adjacência

Supondo que os vértices são numerados  $1, 2, \dots, |V|$  de alguma maneira arbitrária, desta maneira, a representação de uma matriz de adjacência de grafo  $G=(V, E)$  consiste em uma matriz  $|V| \times |V|$   $A=(a_{ij})$  tal que:

$$a_{ij} = \begin{cases} 1, & \text{se } (i, j) \in E, \\ 0, & \text{em caso contrario.} \end{cases}$$

A matriz 1 representa uma matriz de adjacência do grafo da figura 2.5.

		Matriz 1				
		1	2	3	4	5
1	0	1	1	0	0	
2	1	0	0	1	1	
3	1	0	0	0	1	
4	0	1	0	0	0	
5	0	1	1	0	0	

A matriz 2 representa uma matriz de adjacência do grafo orientado da figura 2.7.

		Matriz 2			
		1	2	3	4
1		0	0	0	0
2		1	0	0	1
3		0	1	0	0
4		0	0	1	0

### 2.1.2.2 Lista de Adjacência

A representação da lista de adjacência em um grafo  $G=(V, E)$ , consiste em um arranjo (ou array) Adj. de  $|V|$  listas, uma para cada vértice em  $V$ . Para cada  $u \in V$ , a lista de adjacência Adj[u] contém todos os vértices  $v$  tais que existe uma aresta  $(u, v) \in E$ . A figura 2.12 representa uma lista de adjacência do grafo não orientado da figura 2.5.

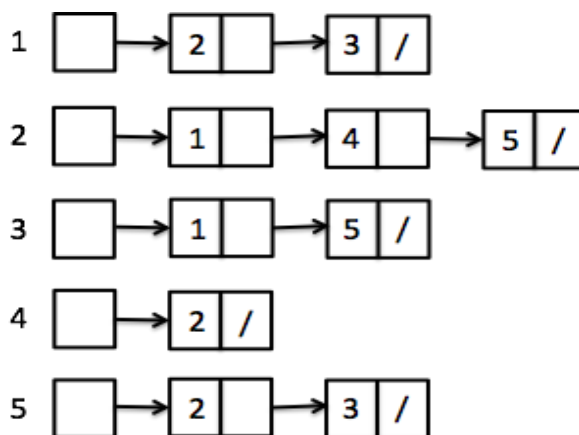


Figura 2.13. Representação de uma lista de adjacência.

A figura 2.13 representa uma lista de adjacência do grafo orientado da figura 2.7.

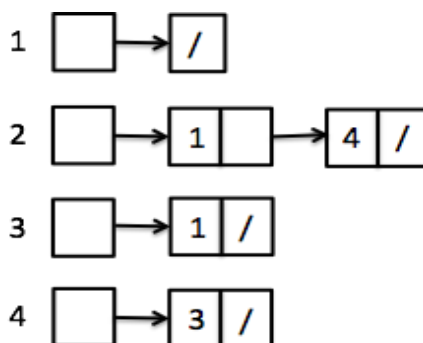


Figura 2.14. Representação de uma lista de Adjacência.

**Observação:** Os vértices em cada lista estão armazenados em uma ordem arbitrária.

## 2.2 ANÁLISE DE REDES SOCIAIS

Análise de Redes Sociais (SNA) representa um conjunto de métodos de análise que os pesquisadores desenvolveram para analisar as redes nas ciências sociais, estudos de comunicação, economia, ciência política, redes de computadores e outros. SNA é uma forma muito interessante de enfatizar que a análise da rede é focada prioritariamente no estudo das informações estruturais contidas nas interações e relações entre as entidades (Yusof et al., 2009).

A Análise de Redes Sociais fornece um visual e uma análise matemática das relações e fluxos entre entidades. Os nós da rede representam tais entidades enquanto os links mostram relações ou fluxos entre os nós. As técnicas de visualização de redes sociais podem mostrar se as interações ocorrem entre todos os membros de um grupo ou se alguns membros do grupo estão se comunicando mais (ou menos) com outros indivíduos específicos. Desta forma, pode-se examinar a natureza das interações e como os indivíduos são posicionados dentro da rede, para desempenhar papéis mais centrais ou mais periféricos nas interações do grupo. Além disso, técnicas de visualização são importantes para ajudar os pesquisadores a entender padrões sociais e de conversação em interações online. E as formas de representação que os analistas de redes sociais utilizam são duas ferramentas matemáticas: Grafos e Matrizes (mostrado na seção 2.1 do trabalho que fala sobre Teoria dos Grafos).

### 2.2.1 COMO SURGIRAM AS REDES SOCIAIS

Segundo (Borgatti, 2013) o surgimento da ideia das relações sociais compoendo uma rede que condiciona a ação dos indivíduos nela inseridas, acontece no século XX. Essa metáfora associada a rede foi inicialmente utilizada pela sociologia, para associar, observar e contextualizar o comportamento individual na estrutura a qual o individuo pertence. Desta maneira, criou-se a metodologia denominada de sociometria, cuja a forma de análise é por meio de um gráfico, conseguindo explicitar os laços de influência e preferência que existe nessa rede e que foi denominado de sociograma.

Cada rede possui sua própria dinâmica, seu próprio comportamento, ainda de acordo com (Borgatti, 2013) em 1951, os nortes americanos Ray Solomonoff e Anatol Rapoport publicaram o trabalho Connectivity of Random Net e introduziram o conceito de redes randômicas, para tentar responder questões como: “Como se formam as redes?” “Quais

são as leis que governam o surgimento de cada um dos seus elementos e estrutura?” Mas só nos anos entre 1958 e 1968, os matemáticos húngaros Paul Erdős e Alfréd Rényi, enfrentaram o desafio de propor uma formulação matemática para descrever todos os grafos complexos em um único esquema, o chamado “mundo randômico” ou teoria randômica dos grafos. Assim surgiu o modelo de **Rede Aleatória**, que estima a quantidade média de arestas presentes no grafo em função do número de vértices e a probabilidade de se relacionarem.

Mas Stanley Milgram em 1967 idealizou uma pesquisa por questionar se sempre estaríamos formando ligações da mesma maneira. Assim ele iniciou um experimento social com o objetivo de encontrar a distância entre duas pessoas quaisquer nos EUA, sendo que a pesquisa consistiu em enviar cartas (na carta havia foto, nome e endereço de uma pessoa, um professor da universidade de Harvard) a várias pessoas de 2 cidades. Como condição, as pessoas que o conhecessem deveriam enviar a carta diretamente para ele, caso contrario deveriam passar a carta para alguém que conhecessem pelo primeiro nome, e que poderiam conhecer o professor, e anotasse seu nome no final de uma das folhas, para que a pesquisa pudesse levantar quantas pessoas (conexões) esta carta passou. O resultado da pesquisa foi 5,5, onde mais de 25% das cartas chegaram ao seu destino, concluindo que as pessoas estavam a 6 graus de separação. Essa pesquisa contribuiu com a criação do modelo de rede denominado “**Mundo Pequeno**”.

Esse modelo de rede é caracterizado por um elevado Coeficiente de Clustering e pequenas distâncias entre os nós. Para Watts e Strogatz segundo (De Sá, 2011), bastavam algumas poucas conexões aleatórias para que os clusters de nós em um grafo fossem conectados e formassem um “mundo pequeno”. Esse modelo ajuda a explicar não só as redes sociais, mas também fenômenos como pandemias e globalização.

Em 2002, Ibert-László Barabási dá o passo seguinte nas descobertas do mundo das redes, criando o modelo de **Rede Livre de Escala**. Esse modelo Segundo (De Sá, 2011) é uma crítica aos modelos de Rede Aleatória e Mundo Pequeno. Para os criadores existia algo como uma ordem ou predisposição na dinâmica da estruturação das redes. Assim, as redes seriam distintas uma das outras, alguns nós teriam poucas conexões e outros teriam várias conexões.

## 2.2.2 MÉTODOS DE ANÁLISE DE REDES SOCIAIS

A perspectiva de redes implica ter vários níveis de análise. As diferenças entre os atores são interpretadas com base nas limitações, oportunidades que surgem e da forma em que eles estão imersos nas redes. A estrutura e o comportamento das redes está baseado nas interações locais entre os atores. Muitas das ideias para os métodos de análise de redes sociais são oriundos da teoria dos grafos, tema já descrito na seção anterior (Borgatti, 2013).

Nesta subseção examinaremos os métodos formais de análise de redes. Apesar da simplicidade das ideias e definições, há boas razões teóricas para achar as propriedades básicas das redes sociais importantes. Por exemplo, até que ponto os indivíduos estão conectados uns com outros e até que ponto a rede em seu conjunto está integrada. As diferenças como os indivíduos estão conectados pode ser extremamente útil para entender seus atributos e comportamentos. Muitas conexões geralmente significa que os indivíduos tem uma alta exposição e mais informação para transmitir. Indivíduos bem conectados podem ser mais influentes e podem estar mais influenciados por outros.

#### 2.2.2.1 Conexões

Desde que as redes estão definidas por seus atores e as conexões entre eles, é importante examinar o número de atores, o número de conexões que são possíveis e o número de conexões efetivamente existentes. Os atores individuais podem ter poucos ou muitos laços. Os indivíduos podem receber, mas não emitir relações ou tanto podem receber quanto emitir. Essas diferenças básicas entre as conexões dos atores podem ser crítica, explicando o poder que tem na rede ou há que ponto sua inclusão na rede é influente (Borgatti, 2013).

#### 2.2.2.2 Tamanho e Densidade

O tamanho de uma rede é muito importante na análise de redes sociais. Idealizando um grupo de 10 pessoas de uma rua de uma cidade. Suponhamos que não seja difícil que as 10 pessoas se conhecessem e que eles se relacionam e que por exemplo compartilhem os nomes de seus filhos. Agora vamos considerar a população do bairro da cidade, supondo que seria extremamente difícil para qualquer pessoa conhecer todos as outras pessoas e seria praticamente impossível em uma rede única compartilhar todos os nomes dos filhos (Borgatti, 2013).

O tamanho é crítico para a estrutura das relações sociais, a medida que o grupo cresce, a proporção que todos os laços poça estejam conectados diminui e muito provavelmente surgirá grupos diferentes.

### 2.2.2.3 Distância

As propriedades da rede que até agora descrevemos tem relações diretas de um ator para o outro, mas a maneira que os atores estão imersos nas redes é mais complexa que isso. Por exemplo, dois atores A e B, podem cada um ter cinco amigos. Mas pode acontecer que nenhum amigo de A tenha outros amigos, exceto o próprio ator A. Em contra partida, cada amigo de B tem também cinco amigos. As informações disponíveis para B e o seu potencial de influenciar é muito maior que A. Desta maneira, as vezes ser “amigo do amigo” pode ter bastante consequências (Borgatti, 2013).

Quantos atores estão a diferentes distancias uns dos outros pode ser importante para entender as diferenças entre atores nas limitações e oportunidades que tem como resultado de sua posição. Também é interessante saber de quantas formas é possível conectar dois atores dada uma distancia, isto é, pode o ator A alcançar o ator B em mais de uma maneira? As vezes conexões múltiplas podem indicar uma conexão mais forte entre os atores que uma só relação.

## 2.2.3 MÉTRICAS DE ANÁLISE

Existem várias formas de medir os conceitos relacionados aos métodos de análise de redes sociais: conexão, tamanho, densidade e distância, descritos na subseção anterior. Desta maneira, nesta subseção iremos descrever algumas métricas de análise de redes sociais.

### 2.2.3.1 Centralidade de Grau

Os Atores que têm mais vínculos com outros atores podem ter posições vantajosas. Porque eles têm muitos laços ou conexões e podem ter formas alternativas de satisfazer as necessidades e, portanto, depender menos dos outros indivíduos. Além disso, já que tem muitas relações, pode acessar e pode obter um número maior de informações que passa na rede (Borgatti, 2013).

Este tipo de medida permite medir o grau de cada nó dependendo das suas relações. É um método de avaliar o papel central de uma relação direta entre um ator para outros atores. No caso de uma rede dirigida (onde os laços têm direção), costuma-se definir duas medidas separadas de centralidade de grau, ou seja, grau de entrada e grau de saída. Assim, grau de entrada é uma contagem do número de ligações direcionadas ao nó e grau de saída é o número de ligações que o nó dirige aos outros. Quando os laços estão associados a alguns aspectos positivos, como a amizade ou colaboração, o grau de entrada é muitas vezes interpretado como uma forma de popularidade e o grau de saída como uma forma de agrupamento.

Segundo (De Sá, 2011) Phillip Bonacich propôs uma alteração da abordagem da centralidade de grau e que foi amplamente aceite como superior a medida original (o **Índice de Poder de Bonacich ou a Centralidade de Bonacich**). O enfoque original da centralidade argumenta que os atores que tem mais conexões são provavelmente mais poderosos ou influentes, porque podem influenciar diretamente a mais atores. Isto faz sentido pois ter o mesmo grau não necessariamente faz os atores igualmente importantes. Por exemplo, imaginamos dois atores A e B cada um com 5 amigos próximos. Os amigos de A são bastante isolados e não tem muitos outros amigos além de A. Já os amigos de B têm muitos amigos que por sua vez também têm muitos amigos e assim por diante. Neste caso, o ator B seria mais central que A, porque os amigos de B estão melhor conectados que os amigos de A, mas A seria mais influente que B, pois como seus amigos tinham poucas conexões eles eram mais dependente de A, diferentemente dos amigos de B que não dependiam tanto dele por possuir várias outras conexões importantes.

### 2.2.3.2 Centralidade de Proximidade

As medidas de centralidade de grau podem ser criticadas porque só consideram as relações imediatas que o ator possui, sem levar em conta as relações indiretas com os demais atores. Um ator pode estar conectado com muitos outros nós, mas esses nós podem estar distantes da maioria do conjunto da rede. Em um caso como este, o ator pode ser bastante central, mas só em uma área local (Borgatti, 2013).

Desta maneira, a centralidade de proximidade permite definir qual a distância métrica de determinado nó por meio da medição do comprimento dos seus caminhos mais curtos (distância geodésica). O distanciamento de um nó é definido como a soma das distâncias a todos os outros nós e a sua proximidade é definida como sendo o inverso do

distanciamento. Assim, quanto mais central o nó menor a sua distância total a todos os outros nós. A proximidade pode ser considerado como uma medida de quanto tempo vai demorar para se propagar uma informação do nó para todos os outros nós sequencialmente.

Mas se a análise acontece em uma rede muito grande e complexa essa medida pode ser distorcida. Por exemplo, dois atores A e B, sendo A mais próximo de um grupo menor e fechado de uma rede maior, mas bastante distante de todos os membros da “população” da rede. O ator B está a moderada distancia de todos os atores da rede global. A medição de proximidade dos dois atores A e B poderia ser bastante similar em magnitude, embora o ator B seja mais central que A neste exemplo, porque B pode alcançar mais nós da rede com o mesmo esforço. Desta maneira, a medida **Centralidade de AutoVetor** tem como objetivo encontrar os atores mais centrais em termos de estrutura global ou geral e ter menos enfoque em padrões mais locais.

### 2.2.3.3 Centralidade de Intermediação

A centralidade de intermediação é uma medida para avaliar a importância de um nó na rede, quantificando o número de vezes que esse nó funciona como uma ponte ao longo do caminho mais curto entre dois outros nós. Na sua concepção, os vértices que têm uma alta probabilidade de ocorrência entre dois vértices escolhidos aleatoriamente, tem uma alta intermediação (Borgatti, 2013).

Por exemplo, suponhamos que uma pessoa denominada A e outra denominada B queiram enviar uma mensagem para uma pessoa que não eles conhecem denominada E. Mas A conhece uma pessoa denominada C que conhece E e manda a mensagem por C para que chegue até E. Mas B conhece duas pessoas (C e D) que também conhecem E e envia por eles a mensagem para que chegue até E. Provavelmente a mensagem de B chegará ou terá uma possibilidade maior de chegar até E antes da mensagem que A enviou. Pois A depende exclusivamente de uma única pessoa para que a mensagem chegue ao seu destino. Diferentemente de B que divide a responsabilidade entre duas pessoas para que a mensagem chegue ao seu destino. Por isso, ter mais de um caminho faz um nó menos dependente e mais “poderoso”.

Já a **centralidade de fluxo** expande a noção de centralidade de intermediação e supõe que os atores usaram todos os caminhos possíveis que os conectam e não somente o caminho mais curto.

#### 2.2.3.4 Densidade

A densidade é uma relação entre o número de ligações diretas existentes e o número total de ligações possíveis. Isto dá uma indicação do grau de envolvimento da rede, mostrando quão ativamente os atores estão envolvidos. A densidade de uma rede é definida como o número de ligações de uma rede dividida pelo número máximo de possíveis ligações. Quando a densidade é 0, a rede não tem qualquer ligação, e quando a densidade é de 100%, todos os atores de uma rede são conectados uns aos outros (Borgatti, 2013).

### 2.3 AVALIAÇÃO DO MÉRITO CIENTÍFICO NO BRASIL

Nesta seção detalhamos o processo de avaliação do mérito científico do CNPq realizado periodicamente no Brasil.

O processo de avaliação realizado anualmente pelo CNPQ, visa avaliar o mérito científico individual dos pesquisadores com o intuito de conceder as chamadas Bolsas de Produtividade em Pesquisa, que valorizam e “premiam” os pesquisadores com reconhecida produção científica em sua área específica do conhecimento. Desta maneira, detalharemos como funciona a avaliação das Bolsas de Produtividade em Pesquisa, por ser o estudo de caso do trabalho.

#### 2.3.1 Bolsas de Produtividade em Pesquisa

Criado em 1951, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), agência do Ministério da Ciência, Tecnologia e Inovação (MCTI), tem um papel fundamental na formulação e na condução das políticas de ciência, tecnologia e inovação, colaborando para o desenvolvimento nacional e o reconhecimento das instituições de pesquisa e pesquisadores brasileiros pela comunidade científica internacional (CNPQ, 2013).

O CNPQ oferece várias modalidades de bolsas no país, destacando-se as bolsas de produtividade em pesquisa, que tem como finalidade premiar pesquisadores que se destaquem entre seus pares, valorizando sua produção científica segundo critérios normativos aprovados pelo Decreto nº 4.728 de 9 de junho de 2003 (CNPQ, 2013).

Para concorrer a este tipo de bolsa existem algumas normas gerais e específicas. Em normas gerais para realizar a solicitação, o pesquisador precisa preencher um formulário

em Formulário de Propostas Online (Efomento, 2013), levando em consideração as datas e as normas de cada modalidade. A classificação e o julgamento das propostas seguem as seguintes fases: análise pela área técnica; análise por consultores ad hoc; análise comparativa de mérito e classificação das propostas por Comitês de Assessoramento (esse parecer deve levar em consideração, os pareceres da área técnica e dos consultores ad hoc e as especificidades das modalidades); e decisão final pela Diretoria, em função da disponibilidade financeira do CNPq.

Em normas específicas o pesquisador precisa possuir alguns requisitos e critérios necessário para a concessão, são eles:

- Possuir o título de doutor ou perfil científico equivalente;
- Ser brasileiro ou estrangeiro com situação regular no País;
- Dedicar-se às atividades constantes de seu pedido de bolsa, e
- Poderá ser aposentado, desde que mantenha atividades acadêmico-científicas oficialmente vinculadas a instituições de pesquisa e ensino.

O enquadramento, classificação e a progressão do bolsista PQ e DT, por categoria (Pesquisador 1 (nível A, B, C e D), Pesquisador 2 e Pesquisador em Desenvolvimento Tecnológico e Extensão Inovadora – DT nível 1 (A, B, C e D) e nível 2), como também recomendações de rebaixamento de nível ou exclusão do quadro de bolsistas são atribuições dos Comitês de Assessoramento, sendo atualizados a cada 3 (três) anos. Esses critérios (item 1.3.4. do estatuto), independentemente do CA, devem seguir os subsequentes itens: méritos científicos do projeto; relevância, originalidade e repercussão da produção científica do candidato; formação de recursos humanos em nível de Pós-Graduação; contribuição científica, tecnológica e de inovação, incluindo patentes; coordenação ou participação em projetos e / ou redes de pesquisa; inserção internacional do proponente; participação como editor científico; participação em atividades de gestão científica e acadêmica. Além de levarem em consideração quando preciso os seguintes itens: foco nos grandes problemas nacionais; abordagens multi e transdisciplinares; impacto social; comunicação com a sociedade; interação com o parque produtivo; e conservação ambiental.

Os requisitos e critérios mínimos para o enquadramento e classificação são: por categoria o pesquisador 1 necessita possuir no mínimo, 8 (oito) anos de doutorado por ocasião da implementação da bolsa. O pesquisador 2 precisa ter, no mínimo, 3 (três) anos de doutorado por ocasião da implementação da bolsa. O pesquisador 1 será enquadrado nos níveis (A, B, C e D), seguindo os critérios do item 1.3.4. do estatuto, tendo peso maior, para

cada CA que julgue importante para sua área de pesquisa. Para o pesquisador 2 em que não há especificação de nível, é avaliado a produtividade do pesquisador, com ênfase nos trabalhos publicados e orientações, ambos referente ao últimos 5 (cinco) anos.

As quotas por categorias e níveis por progressão são analisadas pelo Comitê de Assessoramento, sem que haja uma solicitação, por ocasião do julgamento da proposta do bolsista para uma nova vigência da bolsa PQ e DT. A partir do momento em que todas as quotas de bolsas da categoria 1 são contempladas, o ingresso da categoria 2 para a 1, só ocorre por meio de uma substituição de um pesquisador 1 que tenha seus programas de pesquisas finalizados ou que tenha sido movido para categoria 2, excluído do sistema ou promovido a bolsista Sênior. A progressão ou rebaixamento entre os níveis de categoria 1 fica a critério específico dos Comitês de Assessoramento, não dependendo de quotas e respeitadas uma quota máxima de 10% (dez por cento) do total de bolsas de cada Programa Básico para categoria/nível 1A.

## **2.4 CONCLUSÃO DO CAPITULO**

Nesta seção, foram descritos alguns dos vários conceitos da Teoria dos Grafos que servem como auxílio e base para iniciar um estudo de Análise de Rede Social. Descrevemos um breve histórico de como surgiu a ARS, métodos de análise e algumas das mais importantes métricas da ARS. Desta maneira, podemos concluir que a análise de rede social dispõe de um conjunto de métricas que auxiliam no entendimento da dinâmica dos relacionamentos existentes nas redes estudadas. Essas métricas podem medir as posições vantajosas que os atores envolvidos nas redes podem ter, a importância do ator na rede, a dependência entre atores, etc.

Também foi abordado o funcionamento do processo de avaliação do mérito científico para a possibilidade de se tornar um bolsista de produtividade. E pudemos contatar que o CNPq não detalha tanto o processo dessa análise. Pretende-se por meio da ARS analisar se a avaliação vai além do que foi descrito pelos termos do CNPq ou se existe alguma influência do seu comportamento em uma rede de colaboração científica.

## Capítulo 3 – Trabalhos Relacionados

Este capítulo apresenta os principais trabalhos relacionados a esta dissertação. São discutidos 8 (oito) trabalhos relacionados a análise de redes sociais, principalmente de redes de coautoria e cocitação.

### 3.1 ESTUDO DA REDE DE COAUTORIA E DA INTERDISCIPLINARIDADE NA PRODUÇÃO CIENTÍFICA COM BASE NOS MÉTODOS DE ANÁLISE DE REDES SOCIAIS: AVALIAÇÃO DO CASO DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI/UFMG

Este artigo apresenta um estudo de coautoria e da interdisciplinaridade na produção científica dos professores do Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFMG), utilizando como método de estudo a ser aplicado na Ciência da Informação (CI) a Análise de Redes Sociais (SNA). Tendo como objetivo mostrar como as técnicas SNA se aplicam aos problemas da CI, identificando, se essa rede de coautoria dos professores do PPGCI/UFMG é densa (nesse caso está associado a quantidade de artigos em coautoria, indicando a execução de um programa de pesquisa sólido, com resultados obtidos a partir da colaboração dos professores) e se existe colaboração entre os professores das diferentes linhas de pesquisa (neste trabalho as linhas de pesquisa foram: Informação, Cultura e Sociedade (ICS); Organização e Uso da Informação (OUI) e Gestão da Informação e do Conhecimento (GIC)) (Silva et al., 2007).

As métricas que foram utilizadas na SNA neste artigo foram: *grau nodal* (levando em consideração um grafo não direcionado, é o número de linhas incidentes em um nó ou autor), *distância geodésica* (número de laços que existe no caminho mais curto entre qualquer par de nós ou atores), *centralidade* (conceito associado aos laços dos quais o nó ou ator participa, e se relaciona com a importância ou projeção dos atores de uma rede social), *centro-periferia* (construído a partir das medidas de centralidade, na qual o centro se caracteriza por terem laços entre si e estarem próximos, em termos das distâncias geodésicas, ou seja, o centro é mais coeso), *núcleo (core)*, um subgrafo no qual cada nó é adjacente a no mínimo  $k$  outros nós no subgrafo, isto é o conceito baseia-se no grau dos nós que pertencem ao subgrafo, definindo assim, um núcleo com um número mínimo de  $k$  laços entre os seus componentes ou um  $k$ -núcleo (*K-core*) e *cluster* (é a divisão de grupos da rede em subconjunto de atores, construído a partir dos laços e a posição ocupada por eles na rede).

A análise da rede de coautoria teve como base dois tipos de produção bibliográfica, trabalhos completos em eventos e artigos completos em periódicos. Além disso os autores tomaram algumas preocupações e criaram algumas regras de validação dos resultados, como por exemplo: a atualização do quadro de professores, tanto dos professores que não apareciam na lista por serem novatos, como também dos docentes aposentados; a correção da grafia dos nomes e sua unificação para facilitar a criação da relação de coautoria; a verificação se os mesmos artigos constavam na lista de todos os autores.

Primeiramente foram feitas análises estatísticas: comparação do número de publicações nos períodos antes e após a realização do trabalho; produção por autor e sua distribuição; participação de cada autor no total da produção do programa; número médio de autores por artigo; e número total de colaboradores.

A SNA foi feita com o auxílio do software UCINET a partir dos dados da matriz binária, contendo apenas informações que indicam a existência de laços de coautoria e a matriz valorada, contendo informações sobre o número de colaborações para o período proposto pelo trabalho. Com essas informações analisou-se a estrutura da rede obtida, consequentemente foram feitas análises com algumas medidas de centralidade, centro-periferia, centralidade de grau para um autor, centralidade de intermediação (analisa o quanto um nó está no caminho geodésico entre outros nós) e a análise de quantos clusters se formaram.

A realização deste trabalho permitiu realizar uma análise maior e mais aprofundada da colaboração entre os professores, enfatizando informações que não seriam capazes de ser perceptível apenas com métodos estatísticos. O trabalho também buscou destacar a colaboração entre os professores nas coautorias em artigos, fazendo com que essas colaborações sejam mais eficazes e mais produtivas.

### **3.2 FORMAÇÃO DE REDES SOCIAIS DE COAUTORIA NA ÁREA DE GESTÃO DE PESSOAS: UMA ANÁLISE BIBLIOMÉTRICA EM PERIÓDICOS BRASILEIROS NO TRIÊNIO DE 2007 A 2009**

Este artigo mostra um estudo quantitativo de coautoria na área de gestão de pessoas, no triênio de 2007 a 2009, por meio dos periódicos nacionais. Esse trabalho tem como objetivo analisar a estrutura dessa rede social de coautoria (SCHEFFER et al., 2012).

O método quantitativo assinala-se pelo uso da quantificação tanto nas modalidades de coleta de informações quanto no tratamento por meio de técnicas estatísticas. A realização da coleta de dados incidiu em 14 periódicos brasileiros que possuíam alto fator de impacto (levou-se em consideração os artigos classificados A ou B nacional, pelo Sistema Qualis-Capes, na área de Administração, envolvendo a temática de Gestão de Pessoas e de Recursos Humanos).

Após o término da coleta de dados, iniciou-se o estabelecimento das variáveis para a análise da estrutura da rede social. Foram considerados as seguintes variáveis: autores e coautores dos artigos, instituições às quais os mesmos estavam vinculados quando da publicação dos trabalhos e temas que nortearam as pesquisas.

Na construção da rede social levaram em conta as relações como não direcionais, pois como os autores publicam em coautoria, considera-se a existência de reciprocidade nas respectivas publicações. As métricas utilizadas pelo trabalho foram: a densidade da rede, permitindo a visualização configuração da intensidade da cooperação dentro da rede analisada; a centralidade de grau (degree centrality), indicando o número de laços que um ator possui com outros atores; e por fim a centralidade de intermediação (betweenness centrality) identificando os atores intermediários mais relevantes, indicando aqueles que possuem maior grau de interação na rede. Para a tabulação dos dados foi utilizado o software Excel e para o tratamento dessas informações usaram o UCINET 6 versão 6.186.

A realização desse trabalho concluiu que a rede de coautoria apresenta baixa densidade e heterogeneidade, pois a maioria dos artigos publicados foram produzidos por autores diferentes, baixo número de recorrência de publicação dos autores e poucos laços de intermediação entre as publicações, o que indica baixa existência de colaboração para publicação na área de Gestão de Pessoas e Recursos Humanos no Brasil.

### **3.3 REDE SOCIAL DE COAUTORIA EM CIÊNCIA DA INFORMAÇÃO: ESTUDO SOBRE A ÁREA TEMÁTICA DE “ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO”**

Este trabalho expõe uma rede social de coautoria em ciência da informação, realizando um estudo de análise da dinâmica das redes na produção científica do grupo de trabalho GT2 do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), que

se destina à investigação da área “Organização e Representação do Conhecimento” (ORC) (da Silva et al., 2012).

A coleta de dados incidiu em uma pesquisa documental do tipo exploratória, tendo como base os anais do ENANCIB realizados no período de 1994 e 2011 e dos currículos dos autores pesquisados e que possuíam registros na Plataforma Lattes. Levando em consideração os autores com quatro ou mais publicações nesse encontro. As informações retiradas dessa coleta foram: a formação acadêmica; o vínculo institucional; a região geográfica; os atores mais produtivos e as redes de coautoria do GT2/ENANCIB.

O trabalho realizou estudos quantitativos, por meio da média e variação da produção científica, além da análise das microrredes, coesão social, periferia e clusters. Para formar os grafos utilizou-se o software Pajek e para organizar as informações coletadas em um texto compatível com Pajek, utilizou-se o programa Bloco de Notas.

Com o estudo conclui-se que houve um aumento na produção científica, devido ao crescimento das redes de coautoria, apesar dos atores mais influentes estarem concentrados em uma região.

### **3.4 REDES SOCIAIS NA PRODUÇÃO CIENTÍFICA DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE CIÊNCIAS CONTÁBEIS DO BRASIL**

Este trabalho mostra a rede social de coautoria dos programas de Pós-Graduação de Ciências Contábeis do Brasil. Através do questionamento “Quais as redes sociais formadas a partir da produção científica definitiva dos programas de pós-graduação em ciências contábeis no Brasil?”, criou-se o objetivo de identificar a formação dessas redes sociais no triênio de 2007-2009 (NASCIMENTO et al., 2011).

Para que o objetivo fosse alcançado primeiramente foi realizado um estudo descritivo para identificar a formação das redes sociais da produção científica definitiva no período exposto, em seguida uma pesquisa documental.

A realização da coleta de dados incidiu uma população de 21 cursos de mestrado em ciências contábeis sugerido pela CAPES (2010), distribuídos em 17 instituições, 199 professores que faziam parte do corpo docente destes programas. A coleta de dados referente a produção científica dos docentes foi realizada na Plataforma Lattes, aceitando os artigos publicados e aceitos em periódicos.

O trabalho analisou a evolução da produção científica dos programas de pós-graduação, a formação de redes sociais através da qualificação dos periódicos (A1 com 100 pontos, A2 com 80 pontos, B1 com 60 pontos, B2 com 50 pontos, B3 com 30 pontos, B4 com 20 pontos, B5 com 10 pontos e C com zero ponto) de acordo com classificação da CAPES (2010) e a formação da rede social dos programas de pós graduação de ciências contábeis. A variável utilizada para ser analisada nas redes sociais foi a centralidade da rede, medindo o número de laços que um ator tem com outros atores. O programa utilizado para a efetuação da análise da rede social foi o UNICET.

Com o estudo deste trabalho conclui-se que as redes formadas pelos programas de pós-graduação em ciências contábeis são esparsas, pouco densas e apresentam ligações fracas.

### **3.5 REDES SOCIAIS E INTELLECTUAIS EM ADMINISTRAÇÃO DA INFORMAÇÃO: UMA ANÁLISE DO PERÍODO 1997-2006**

Este trabalho tem como objetivo identificar e analisar as redes sociais e intelectuais da área de Administração da Informação, procurando entender a dependência social na construção do conhecimento da área. Foram analisados os artigos publicados nos anais do Enanpad (por ser um dos principais meios de comunicação da área) no período de 1997 a 2006. Avaliaram-se as redes de autoria e cocitação que foram geradas a partir das referências adotadas nestes artigos (Graeml et al., 2010).

A realização dessa coleta de dados incidiu em uma população de 339 artigos, 358 pesquisadores com produção de artigos na área e 9287 referências, correspondendo a 6652 citações de autores diferentes utilizados na fundação dos trabalhos.

Parte da análise foi realizada em comparação entre dois períodos de cinco anos (1997 a 2001 e 2002 a 2006), por entender que eram momentos distintos da área que, ao longo do tempo, modificou-se não apenas socialmente, no que diz respeito ao número de pesquisadores vinculados, mas também na temática e no desenvolvimento teórico em curso. O método de pesquisa utilizado foi o quantitativo com o auxílio das redes sociais. A análise foi dividida em três etapas: análise da estrutura social da área, investigando os relacionamentos de cooperação entre os pesquisadores; compreensão da fundamentação teórica da área a partir da análise de cocitação de autores; e por último, analisou-se a relação dessas duas primeiras etapas, com o objetivo de examinar a importância dos relacionamentos sociais na estrutura intelectual da área de Administração da Informação.

As métricas utilizadas para análise de rede social foram: média de laços por autor, número de componentes, tamanho do componente principal, tamanho do segundo maior componente, tamanho do terceiro maior componente, autores isolados, densidade, distância média, distância máxima (diâmetro), centralização e coeficiente de agrupamento.

Os dados foram organizados no software Microsoft Excel, para a construção das redes foram utilizados os programas UNICET 6 e o PAJEK 1.22 e para o procedimento estatístico utilizou-se o programa SPSS 16.

Com o estudo deste trabalho conclui-se que a análise das redes sociais na área de Administração da Informação trouxe distintos aspectos que eram pouco apreciados em artigos sobre a análise da produção científica da área. Desta maneira, aparece muito outros focos de análise como possíveis temas para futuras pesquisas.

### **3.6 ANÁLISE DAS REDES DE COLABORAÇÃO CIENTÍFICA ATRAVÉS DO ESTUDO DAS CO-AUTORIAS, NOS CURSOS DE PÓS-GRADUAÇÃO DO BRASIL NO TEMA TRATAMENTO TEMÁTICO DA INFORMAÇÃO**

Este trabalho tem como objetivo analisar e descrever a situação das redes de colaboração de coautoria científica existentes, no que diz respeito à participação dos docentes dos programas de Pós graduação em Ciência da Informação do Brasil, no tema Tratamento Temático da Informação (Oliveira et al., 2009).

A metodologia empregada pelo trabalho foi utilizar como base um estudo de 2007, que coletou os pesquisadores mais produtivos através de produção de artigos publicados em periódicos na área de Tratamento Temático da Informação dos Programas de Pós-Graduação do Brasil, usando, como instrumental de busca, a Plataforma Lattes.

A realização dessa coleta de dados incidiu em uma população de 19 pesquisadores que foram colocados em ordem alfabética, numerados e nominados de acordo com a instituição de origem, na produção científica de artigos publicados em periódicos, livros e capítulos de livros.

A análise foi realizada com base em métodos quantitativos e qualitativos, procurando avaliar as redes de coautorias dentro do grupo, denominado intragrupo e com pesquisadores fora do grupo, denominado extra grupo, com o total da produção em colaboração, em cada uma das três variáveis. Depois levantou-se a frequência de coautoria intragrupo, entre os 19 pesquisadores, relativa à produção de artigos, produção de livros e em

capítulos de livros. O programa utilizado para uma melhor visualização e análise da rede social foi o Pajek.

Com o estudo deste trabalho conclui-se que as redes de colaboração científicas mais forte, acontece nas publicações de capítulos de livros, sendo um pouco mais frágil para as publicações de artigos e livros. O trabalho também conclui que o programa Pajek, utilizado para visualizar e analisar as redes de coautoria é bastante recomendado e apropriado.

### **3.7 ANÁLISE TEMPORAL DA RELAÇÃO ORIENTADOR-ORIENTADO: UM ESTUDO DE CASO SOBRE A PRODUTIVIDADE DOS PESQUISADORES DOUTORES DA ÁREA DE CIÊNCIA DA COMPUTAÇÃO**

Este trabalho apresenta uma rede social de coautoria entre orientador e orientado, sobre a produtividade dos pesquisadores doutores da área de Ciência da Computação. Tem como objetivo analisar as principais características e relações desse grupo, examinando, por exemplo, quão longo é o tempo de duração da coautoria entre o orientador e orientado (TUESTA et al., 2012).

A consumação da coleta de dados sucedeu em uma população de 1746 doutores formados na área, entre os anos de 1981 e 2009. Identificando 889 currículos dos docentes que fazem parte do corpo permanente dos 45 programas de Pós-Graduação acadêmica (que possuem doutorado ou mestrado acadêmico) presentes na última avaliação no triênio 2007 à 2009, da CAPES. Essas informações foram baixadas pelo scriptLattes, armazenados e tabulados em um banco de dados por um *parser*. Neste trabalho os currículos foram divididos nas formações do pesquisador, orientação e publicação em periódicos.

Na análise descritiva foi calculado a moda, média e desvio padrão das métricas: Número de coautores de todas as publicações; Número de coautores em publicações orientador-orientado; Número de publicações orientador-orientado; Duração do doutorado; Duração do relacionamento orientador-orientado; Tempo transcorrido até a primeira publicação orientador-orientado desde o início do doutorado; Tempo transcorrido até a primeira publicação orientador-orientado desde o fim do doutorado; Tempo transcorrido até a primeira publicação sem a presença do orientador desde o fim do doutorado; Número de publicações dos orientadores; Número de publicações dos orientados.

Na visualização da rede construiu-se duas figuras para representar as relações de orientação, coautoria e *AuthorRank*, modificação feita do PageRank. O *AuthorRank* é uma

medida que indica o grau de coautoria de um pesquisador em relação aos outros pesquisadores.

Na análise de correlação foi utilizado para mensurar a eficácia de uma associação entre duas variáveis, uma medida não paramétrica, o coeficiente de Spearman. Este coeficiente de correlação foi empregado para calcular a relação entre os Anos de Relacionamento, que é o tempo de início do doutorado, até a última publicação e o número total de publicações do orientado.

Com o fim deste trabalho conclui-se que através da análise de rede social foi possível identificar que o tempo de duração média do período de colaboração é superior em aproximadamente 3 anos à data da primeira publicação e à data do fim do doutorado.

### **3.8 COAUTORIA COMO INDICADOR DE REDES DE COLABORAÇÃO CIENTÍFICA**

Este trabalho mostra um estudo bibliométrico sobre redes de colaboração científica entre os professores dos programas de Pós-Graduação em Epidemiologia da Universidade Federal de Pelotas (PPGE/UFPel). Desta maneira o objetivo do trabalho é saber quanto colaborativo são os professores do PPGE/UFPel, se as colaborações aumentaram ou diminuíram ao longo do ciclo estudado, se o número de colaboração aumenta a produtividade, se os professores colaboram mais com colegas ou com pesquisadores externos ao programa e se os docentes mais produtivos são os mais colaborativos (MAIA et al., 2008).

A realização da coleta de dados incidiu em uma população de 237 artigos em 87 diferentes periódicos, envolvendo 369 autores e coautores, no período de 1991 e 2002.

Neste trabalho foi utilizado primeiramente o método de quantificação denominado completo, que indica que a colaboração é maior quanto maior é o número de colaborador. Este método foi aplicado para poder verificar o número de autores por artigo e sua relação com a produtividade dos professores, através do cálculo do coeficiente de colaboração, que é um método para avaliar se há crescimento do número de autores por documento e também de trabalhos de múltipla autoria. Em seguida utilizou-se a análise de redes sociais, calculando o grau de centralidade (centrality), intermediação (betweenness) e proximidade (closeness) para cada professor.

A coleta de dados foi armazenada em um banco de dados do programa EndNote, que permitiu importar registros bibliográficos de bases de dados como a PubMed, facilitando

este processo. A tabulação das informações como: autores, ano de publicação, título dos periódicos e número de autores do artigo, foi realizada através dos programas Microsoft Excel e SPSS. Para a análise de redes sociais utilizou-se o programa UNICET.

Com o termino deste trabalho conclui-se que o número de colaboradores não influência no número de produtividade e que os professores publicam mais artigos em autoria compartilhada que individual.

### 3.9 COMPARAÇÃO DOS TRABALHOS RELACIONADOS

Esta seção mostra a tabela com as comparações dos trabalhos relacionados que foram abordado em toda seção 3.

**Tabela 3.3. Tabela de comparações dos trabalhos relacionados.**

Trabalhos Relacionados	Finalidades	Métricas	Tecnologias	Resultados
Trabalho 1	Mostrar como as técnicas SNA se aplicam aos problemas da Ciência da Informação para os professores do PPGCI/UFMG.	Centralidade de Grau, Centralidade de autovetor e coeficiente de clusterização	UCINET.	O trabalho destacou a colaboração entre os professores nas coautorias em artigos, fazendo com que essas colaborações sejam mais eficazes e mais produtivas.
Trabalho 2	Analisar a estrutura da rede social de coautoria na área de gestão de pessoas, no triênio de 2007 a 2009, através dos periódicos nacionais.	Densidade da rede, centralidade de grau, centralidade de intermediação.	UCINET e Microsoft Excel.	A rede de coautoria apresentou baixa densidade e heterogeneidade, pois a maioria dos artigos publicados foram produzidos por autores diferentes, baixo número de recorrência de publicação dos autores e poucos laços de intermediação entre as publicações.
Trabalho 3	Analisar a dinâmica	Média e variação da	Pajek e Bloco de	Houve um aumento

	das redes na produção científica do grupo de trabalho GT2 do ENANCIB, que se destina à investigação da área “Organização e Representação do Conhecimento” (ORC).	produção científica, além da análise das micro redes, coesão social, periferia e clusters	Notas.	na produção científica, devido ao crescimento das redes de coautoria, apesar dos atores mais influentes estarem concentrados em uma região.
Trabalho 4	Identificar a formação das redes sociais coautoria dos programas de Pós-Graduação de Ciências Contábeis do Brasil no triênio de 2007-2009.	Centralidade da rede, medindo o número de laços que um ator tem com outros atores.	UCINET	As redes formadas se mostraram esparsas, pouco densas e apresentam ligações fracas.
Trabalho 5	Procurar entender a dependência social na construção do conhecimento da área de Administração da Informação, no período de 1997 a 2006.	Média de laços por autor, densidade, centralidade e coeficiente de clusterização.	UNICET, PAJEK e SPSS.	O trabalho concluiu que a análise das redes sociais na área de Administração da Informação trouxe distintos aspectos que eram pouco apreciados.
Trabalho 6	Analisar e descrever a situação das redes de colaboração de coautoria científica existentes, no que diz respeito à participação dos docentes dos programas de Pós graduação em CI, no tema Tratamento Temático da Informação	Total da produção em colaboração e a frequência de coautoria.	Pajek.	Concluiu que as redes de colaboração científicas mais forte, acontece nas publicações de capítulos de livros, sendo um pouco mais frágil para as publicações de artigos e livros.
Trabalho 7	Analisar as principais características e relações dos pesquisadores doutores da área de	Moda, média e desvio padrão das métricas	Informações foram baixadas pelo script Lattes, armazenados e tabulados em um	Concluiu que através da análise de rede social foi possível identificar que o tempo de

	Ciência da Computação, examinando por exemplo quão longo é o tempo de duração da coautoria entre o orientador e orientado.		banco de dados por um <i>parser</i> .	duração média do período de colaboração é superior em aproximadamente 3 anos à data da primeira publicação e à data do fim do doutorado.
Trabalho 8	Saber quão colaborativo são os professores do PPGE/UFPel, se as colaborações aumentaram ou diminuíram ao longo do ciclo estudado, se o número de colaboração aumenta a produtividade, se os professores colaboram mais com colegas ou com pesquisadores externos ao programa e se os docentes mais produtivos são os mais colaborativos.	Cálculo do coeficiente de colaboração, centralidade de grau centralidade de proximidade e de intermediação	Microsoft Excel, SPSS e UNICET.	Concluiu que o número de colaboradores não influencia no número de produtividade e que os professores publicam mais artigos em autoria compartilhada que individual.

A comparação dos trabalhos relacionados mostrada na tabela 3.1, nos dá base para iniciar os estudos de ARS, por meio dos conceitos de análise, das métricas e como devemos iniciar as pesquisas.

Concluimos também que a análise de redes sociais é muito abrangente, sendo utilizada nas mais diversas áreas de pesquisa e tem um leque de opções para que se possa realizar pesquisas. Como também apresenta uma quantidade de métricas interessantes, incluindo algumas das que serão utilizadas no trabalho, como a centralidade de grau, centralidade de proximidade, intermediação e autovetor.

## Capítulo 4 – Metodologia

Este capítulo detalha todo o caminho e etapas que o trabalho vai percorrer para alcançar os objetivos propostos.

### 4.1 INTRODUÇÃO

Nesta seção, serão descritas, todas as etapas realizadas para a conclusão do trabalho. Será especificado como os dados foram coletados, como as redes de análise foram geradas e como foram analisadas as métricas quantitativas de Análise de Rede Social.

#### 4.1.1 Seleção da Amostra

A primeira etapa do trabalho será pesquisar e saber quem são os pesquisadores bolsistas de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Essas informações serão extraídas no próprio site da instituição. (CNPq2, 2014).

#### 4.1.2 Geração das Redes

Com a amostra coletada a próxima etapa do trabalho será extrair o Currículo Lattes (Lattes, 2013) de cada pesquisador. A figura 4.1 mostra o currículo lattes do autor deste trabalho.

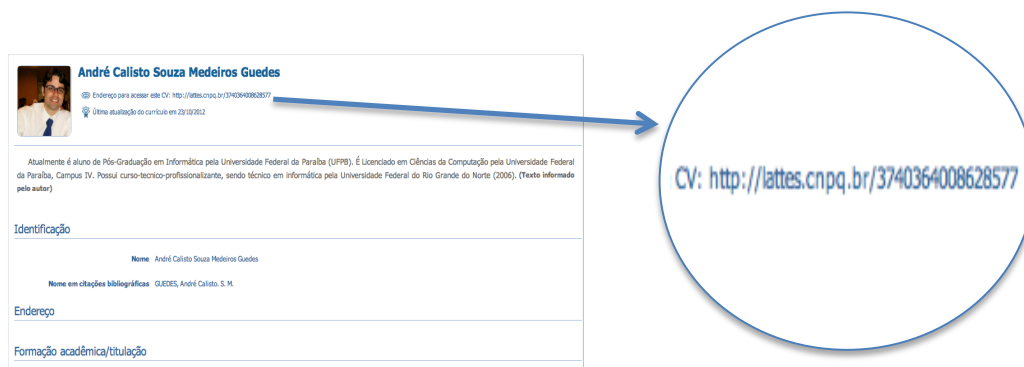


Figura 4.1. Exemplo do Currículo Lattes.

A partir do Currículo Lattes, extrairemos a parte que está ampliada na figura (<http://lattes.cnpq.br/3740364008628577>) e pegaremos os últimos números do link, neste caso (3740364008628577).

De posse destes códigos de identificação, utilizaremos a ferramenta LattesCrawler (LattesCrawler, 2014) para extrair as informações de colaboração para cada um dos pesquisadores. A figura 4.2 ilustra os links (tom de azul mais claro) das pessoas que estão em colaboração. As conexões representam relações de: co-autoria em artigos publicados em periódicos e conferências, orientações de mestrado e/ou doutorado e participações em banca de mestrado e/ou doutorado e em projetos de pesquisa. A ferramenta de coleta gera um arquivo com a extensão gml (Graph Modeling Language), que pode ser utilizado por diferentes ferramentas de análise e visualização de redes, como o Gephi. Esta é uma ferramenta que permite explorar e compreender os gráficos gerados a partir das conexões entre os nós da rede, além de permitir o cálculo de valores referentes às diferentes métricas de ARS. O usuário interage com a representação, manipulação de estruturas, formas e cores para revelar informações que estão ocultas. O objetivo da ferramenta é ajudar os analistas de dados a criar hipóteses e intuitivamente descobrir padrões. É uma ferramenta complementar para as estatísticas tradicionais, como o pensamento visual com interfaces interativas (Gephi, 2013).



**Figura 4.2. Figura que mostra as referências com os links para outros colaboradores.**

Além da geração do arquivo gml, as informações dos pesquisadores também serão armazenadas em um banco de dados para facilitar consultas que serão necessárias para o trabalho. Serão criadas duas tabelas, uma tabela com o nome "nó" e outra com o nome "arestas". A tabela "nó" possuirá os campos: id, como chave primaria, idlattes, label, bolsa, periódicos, conferencias, orientacoesmestrado, orientacoesdoutorado, bancasmestrado, bancasdoutorado e seed do tipo text.. A tabela "arestas" possuirá os campos: id, como chave

primaria; origem e destino, como chave secundaria da tabela nó do campo idlattes e valor. A figura 4.3 mostra o modelo entidade e relacionamento das duas tabelas do banco de dados que serão criadas.

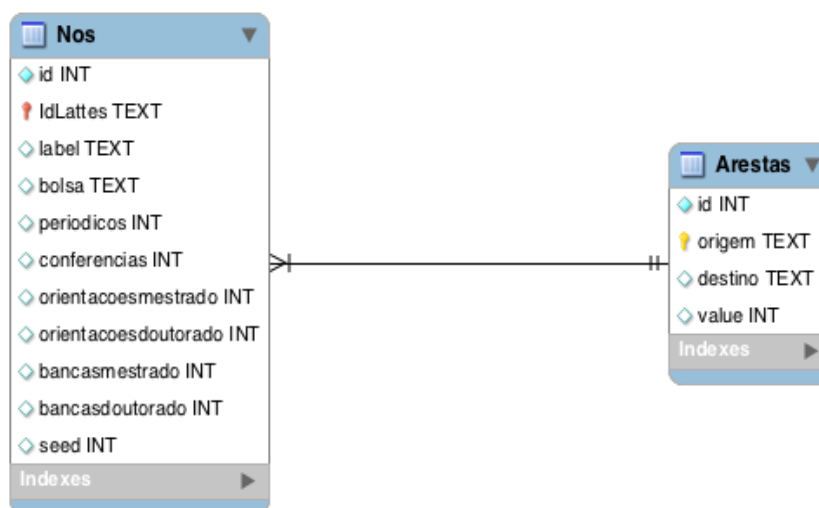


Figura 4.3. Modelo Entidade Relacionamento do Banco de Dados que será criado.

Desta maneira, poderemos realizar consultas para responder às questões de pesquisa propostas no Capítulo 1 do trabalho e que será descrita logo abaixo.

#### 4.1.3 Métricas e Consultas

As métricas geradas servirão para mensurar como se comporta os pesquisadores de produtividade na sua rede em relação com seus colaboradores. Se o pesquisador é mais central, se tem uma maior interação com seus colaboradores, etc. E também para responder duas das 4 questões de pesquisa. São elas:

- **QP1:** Bolsistas de produtividade nível 1 têm mais colaboradores que a média dos bolsistas de produtividade nível 2 e DT na área?
- **QP4:** Existem diferenças nas métricas de centralidade dos bolsistas de produtividade nível 1 dos bolsistas de produtividade nível 2 e DT?

As métricas utilizadas neste trabalho, cujos valores serão calculados pela ferramenta Gephi, serão: centralidade de grau, centralidade de proximidade, centralidade de intermediação, centralidade de autovetor, PageRank e o número de triângulos. Estas métricas são as medidas que julgamos necessárias para chegar ao resultado que o trabalho propôs, que

é saber a importância, influência e dependência de um pesquisador na rede dos bolsistas de produtividade do CNPq.

Além dessas métricas, criaremos consultas no banco de dados que serão capazes de nos ajudar a responder outras questões de pesquisa:

- **QP2** - Bolsistas de produtividade nível 1 têm mais colaboradores bolsistas de produtividade do que a média dos bolsistas de produtividade nível 2 e DT?
- **QP3** - A proporção de pesquisadores que têm mais colaboradores que seus colaboradores é maior entre os bolsistas de produtividade nível 1 do que entre os demais pesquisadores bolsistas de produtividade?

Para poder responder essas questões de pesquisa será necessário criar um software na linguagem Java, que permita descer mais de um nível do grafo, pois a ferramenta Gephi não permite esse tipo de consulta.

Então serão criadas 4 (quatro) classes: Main, Metricas, Vertices e Connect. A classe Main é a classe principal que chama os métodos que fazem as consultas e que geram as métricas. Esses métodos se encontram na classe Metrica. A classe Vertices tem todos os atributos dos "nós". E a classe Connect é a classe que faz a conexão e consultas com o banco de dados que já foi descrito a subseção anterior. A figura 4.4 mostra o diagrama de classe do programa que será criado para fazer essas consultas.

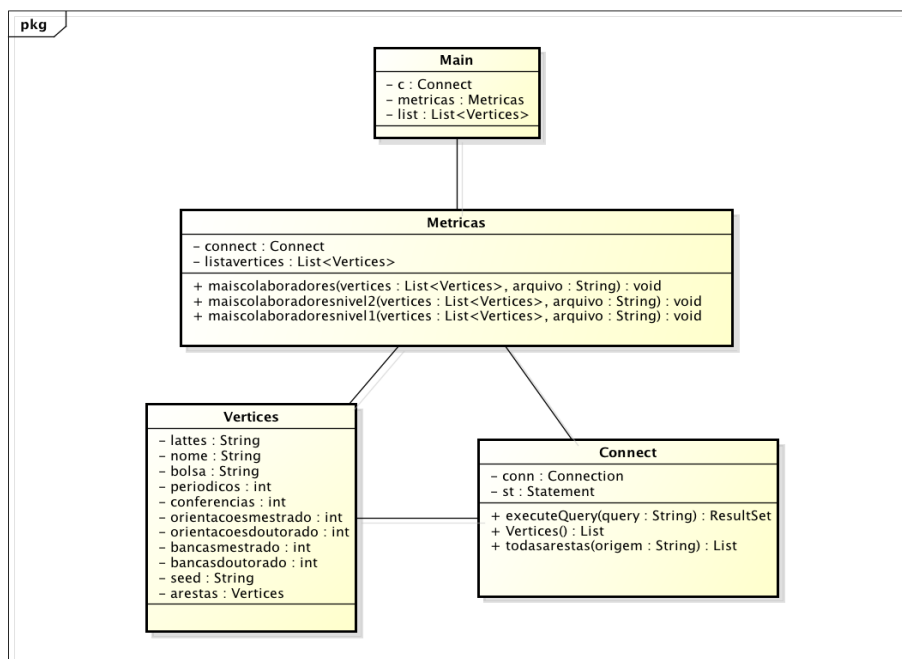


Figura 4.4. Diagrama de Classe do software que será criado para fazer consultas.

As medidas que serão criadas são: a porcentagem da quantidade de colaboradores que tem um grau maior que o colaborador principal (COLA1), a porcentagem de colaboração

maior de bolsistas de produtividade nível 1 que um pesquisador têm em relação a seus colaboradores (COLA2) e a porcentagem de colaboração maior de bolsistas de produtividade nível 2 que um pesquisador têm em relação a seus colaboradores (COLA3).

Com todas esses dados, métricas e resultados gerados das consultas, uma análise estatística será feita, por meio da análise de variância do tipo one-way. Esse tipo de análise visa, essencialmente, verificar se existe uma diferença significativa entre as médias e se os fatores exercem influência em alguma variável dependente (Morettin, 2013).

Além dessa análise, será feito um teste de post-hoc para ~~poder~~ identificar ~~aonde~~ acontecem diferenças que, por ventura, venham aparecer na análise de variância. O teste que será feito é o de Tukey. O método de Tukey consiste na construção de intervalos de confiança para todos os pares de médias, de tal forma, que o conjunto de todos os intervalos tenha um determinado grau de confiança  $\gamma$  (por exemplo, se  $\gamma = 0.95$ , temos 95% de confiança de que os intervalos obtidos contêm todas as respectivas diferenças) (Morettin, 2013).

A análise de variância e o teste de Tukey foram realizados pelo *software estatístico* Statistica (versão 7.0). Statistica é um software estatístico que provê de um conjunto de ferramentas para análise, gestão e visualização de bases de dados (Statistica, 2014).

## **Capítulo 5 – Resultados**

Este capítulo apresenta todos os resultados obtidos no trabalho e que serviram para alcançar os objetivos do trabalho.

### **5.1 INTRODUÇÃO**

Depois de ter realizado a revisão bibliográfica da literatura, percebe-se que o assunto proposto pelo trabalho, análise de redes sociais, aplica-se em vários segmentos, mostrando ser um tema bastante atual mesmo não sendo tão novo.

Sabendo dessa flexibilidade e da atualidade do assunto, propomos a utilização da análise de redes sociais para estudar como se comporta as redes dos professores bolsistas de produtividade do CNPQ do Brasil, na área de Computação, para compreender os impactos dos relacionamentos na avaliação do mérito científico no Brasil.

As seções a seguir apresentam algumas redes sociais dos docentes bolsistas de produtividade.

### **5.2 AS REDES SOCIAIS DE PESQUISA**

As redes sociais de pesquisa do trabalho apresentado, nesta seção, foram construídas utilizando o software Gephi, tomando como base os dados coletados dos currículos Lattes dos docentes bolsistas de produtividade do CNPQ, gerando uma rede individual dos pesquisadores bolsistas com os seus colaboradores diretos e com os colaboradores de seus colaboradores (nível 2 da rede).

A rede dos pesquisadores constituem em 404 pesquisadores da categoria PQs e 63 da categoria DT, e todas essas informações foram retiradas por meio do site do CNPq. A rede dos bolsistas é composta por 404 pesquisadores bolsistas de produtividade níveis 1 e 2 e 64 pesquisadores bolsistas de produtividade nível DT, tendo um total de 467 pesquisadores.

A rede social dos pesquisadores bolsistas nível 1 e 2 possui 49.671 nós e 211.509 arestas. Os bolsistas de produtividade representam apenas 1,86% dos nós. A rede social dos bolsistas DTs possui 19.196 nós e 64.264 arestas e os pesquisadores DTs representam 0,53% dos nós.

A Figura 5.1 mostra um exemplo de uma rede de um dos pesquisadores bolsistas de produtividade gerada pela ferramenta Gephi.

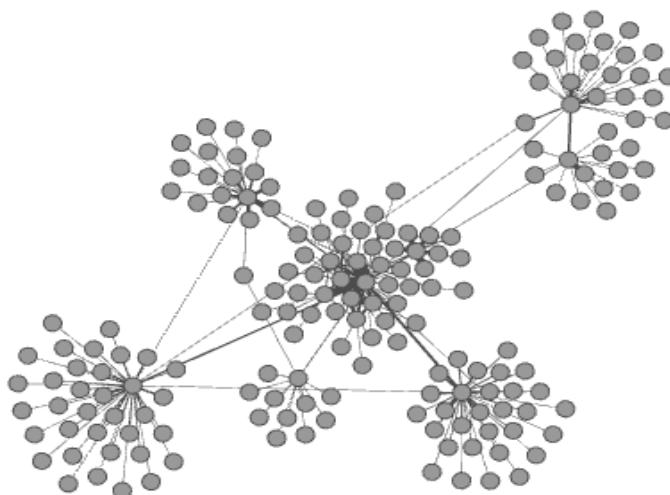


Figura 5.1. Rede de um pesquisador gerada pelo Gephi.

### 5.3 ANÁLISE DE VARIÂNCIA

Nesta seção, analisamos os resultados obtidos por meio das métricas de redes sociais aplicadas às redes construídas durante este trabalho e pelas variáveis quantitativas que são extraídas por meio dos dados obtidos dos Currículos Lattes dos pesquisadores. Serão analisados a Centralidade de Grau, a Centralidade de Proximidade, a Centralidade de Intermediação, a PageRank, o Número de Triângulos feitos na rede, a Centralidade de AutoVetor, os periódicos submetidos, a participação em conferências, as orientações em mestrado e doutorado e a participação em bancas de mestrado e doutorado.

A análise de variância foi realizada por meio do *software estatístico* Statistica (versão 7.0). Foi realizada uma análise de variância simples (one way ANOVA), com o objetivo em identificar se havia alguma diferença entre as categorias de pesquisador bolsista para cada variável e métrica citada acima.

A tabela 5.1 mostra o resultado obtido pela análise ANOVA.

**Tabela 5.1. Resultado da análise de variância por meio do software estatístico Statistica.**

	SS	Degr. of	MS	F	p-valor
Bolsa/Periódicos	82944.5	5	16588.9	66.710	<b>0.000000</b>
Bolsa/Conferências	413830	5	82766	22.3200	<b>0.000000</b>
Bolsa/Orientação-Mestrado	25639.8	5	5128.0	26.8421	<b>0.000000</b>
Bolsa/Orientação-Doutorado	6570.83	5	1314.17	50.4845	<b>0.000000</b>
Bolsa/Bancas-Mestrado	15477.3	5	3095.5	4.8746	<b>0.000235</b>
Bolsa/Bancas-Doutorado	13328.95	5	2665.79	18.6796	<b>0.000000</b>
Bolsa/Grau	41018.2	5	8203.6	11.0476	<b>0.000000</b>
Bolsa/C.Proximização	0.00348	5	0.00070	2.1	0.064966
Bolsa/C.Intermediação	0.89434	5	0.17887	5.056	<b>0.000160</b>
Bolsa/PageRank	0.015494	5	0.003099	2.7857	<b>0.017170</b>
Bolsa/N.Triangulos	158653	5	31731	8.1005	<b>0.000000</b>
Bolsa/Autovetor	4.3748	5	0.8750	15.337	<b>0.000000</b>
Bolsa/COLA1	2903.2	5	580.65	1.7296	0.126334
Bolsa/COLA2	9979.1	5	1995.8	2.1329	0.060450
Bolsa/COLA3	24231.0	5	4846.2	3.5186	<b>0.003932</b>

Em destaque vermelho estão as métricas e variáveis que se diferenciam em relação as categorias de bolsas. Observando, principalmente, a coluna p-valor (valor de significância de  $p < ,05000$ ) que indica a probabilidade dos valores dessas variáveis serem parecidos. Em relação a cada categoria de bolsas, percebemos um baixo valor na maioria das variáveis e métricas, sendo que em apenas 3 (com a cor preta) observamos um valor que indica que não há diferença entre os grupos de bolsas.

Para identificar onde estão essas diferenças foi realizado o teste de Tukey HSD (Honestly Significant Difference) para cada variável e medida. Para as métricas foram realizados dos testes de Tukey. O primeiro separando os bolsistas nível 1 em suas subcategorias (A, B, C e D). Se não é possível afirmar se há alguma diferença do pesquisador nível 1 (de forma geral) aos demais bolsistas, o segundo teste é realizado sem a divisão das subcategorias, para possibilitar responder as questões de pesquisas.

### 5.3.1 Periódicos

Como foi mostrado na tabela 5.1, na seção anterior, a variável periódicos têm diferenças para as categorias de bolsas. Com o teste de Tukey podemos detalhar em qual categorias houve diferença. A tabela 5.2 mostra o resultado do teste de Tukey.

Tabela 5.2. Teste de Tukey para a variável Periódicos.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
<b>PQDT</b>		<b>0.000020</b>	<b>0.000020</b>	<b>0.000020</b>	<b>0.000020</b>	0.996345
<b>PQ1A</b>	<b>0.000020</b>		<b>0.000021</b>	<b>0.000020</b>	<b>0.000020</b>	<b>0.000020</b>
<b>PQ1B</b>	<b>0.000020</b>	<b>0.000021</b>		0.989723	<b>0.048668</b>	<b>0.000020</b>
<b>PQ1C</b>	<b>0.000020</b>	<b>0.000020</b>	0.989723		0.118207	<b>0.000020</b>
<b>PQ1D</b>	<b>0.000020</b>	<b>0.000020</b>	<b>0.048668</b>	0.118207		<b>0.000020</b>
<b>PQ2</b>	0.996345	<b>0.000020</b>	<b>0.000020</b>	<b>0.000020</b>	<b>0.000020</b>	

Neste caso, os resultados mostram, em destaque vermelho, que por meio do teste de Tukey na variável periódicos, os bolsistas PQ1A se diferem de todos os outros bolsistas. Os bolsistas PQ1B e PQ1D só tem comportamento semelhante aos PQ1C. Já os bolsistas PQDT e PQ2 também têm o mesmo comportamento, em relação a variável periódicos. A tabela 5.3 e a figura 5.2 mostram a média, desvio padrão e o intervalo de confiança (95% de confiança) para cada categoria de bolsa, em relação a essa variável.

Tabela 5.3 Tabela com a média, desvio padrão e intervalo de confiança da variável periódicos.

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		<b>-95.00%</b>	<b>95.00%</b>	
<b>PQDT</b>	15.30	12.41	18.19	11.49
<b>PQ1A</b>	68.39	48.36	88.42	46.33
<b>PQ1B</b>	42.24	34.36	50.12	17.32
<b>PQ1C</b>	39.85	36.01	43.68	10.81
<b>PQ1D</b>	31.04	27.29	34.80	15.86
<b>PQ2</b>	16.55	15.17	17.92	11.12

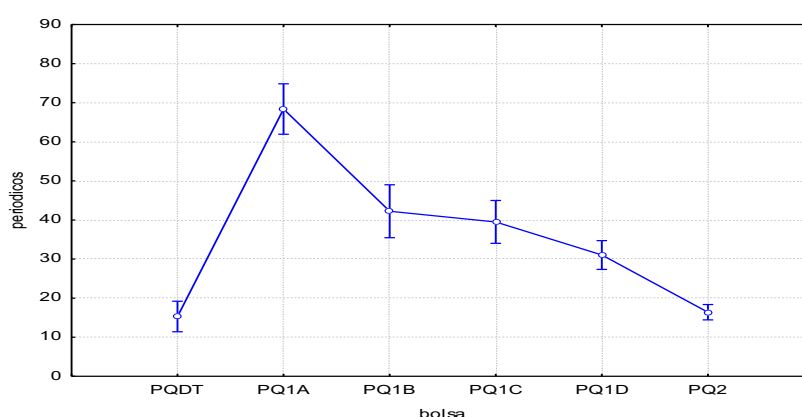


Figura 5.2. Gráfico com a média e o intervalo de confiança da variável periódicos.

Percebemos através da figura 5.2 e a tabela 5.3 a diferença que foi mostrada pelo teste de Tukey e está representada na figura pelos intervalos de confiança, que são os traços verticais.

Desta maneira, podemos concluir que a variável periódicos têm fortes indícios que possa ser uma boa variável para classificar um bolsista pois, como mostra a figura 5.2 o

bolsista PQ1A, por exemplo, com 95% de confiança, vai estar entre a faixa de 48 a 88 periódicos publicados, assim, tendo um comportamento totalmente diferente e mostrando que publica mais periódicos que os demais bolsistas.

### 5.3.2 Conferências

A variável conferências também possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.4 mostra o resultado do teste de Tukey em relação a essa variável, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

Tabela 5.4. Teste de Tukey para a variável conferências.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		<b>0.001488</b>	<b>0.000021</b>	<b>0.005652</b>	<b>0.016257</b>	0.487549
PQ1A	<b>0.001488</b>		0.656606	0.986679	0.588902	<b>0.000021</b>
PQ1B	<b>0.000021</b>	0.656606		0.209741	<b>0.008945</b>	<b>0.000020</b>
PQ1C	<b>0.005652</b>	0.986679	0.209741		0.924859	<b>0.000021</b>
PQ1D	<b>0.016257</b>	0.588902	<b>0.008945</b>	0.924859		<b>0.000020</b>
PQ2	0.487549	<b>0.000021</b>	<b>0.000020</b>	<b>0.000021</b>	<b>0.000020</b>	

A tabela 5.4 mostra que o comportamento dos bolsistas nível 1 (PQ1A, PQ1B, PQ1C e PQ1D) têm um comportamento parecido, em relação a participação de conferências. Nessa categoria de bolsa, as únicas categorias que se diferem são os bolsistas PQ1B com os bolsistas PQ1D. A tabela também mostra que os bolsistas DTs tem o mesmo comportamento dos bolsistas PQ2.

A tabela 5.5 e a figura 5.3 mostram a média, desvio padrão e o intervalo de confiança (95% de confiança) para variável de conferências.

Tabela 5.5 Tabela com a médias, desvio padrão e intervalo de confiança para a variável conferências.

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		<b>-95.00%</b>	<b>95.00%</b>	
PQDT	83.65	69.25	98.05	57.18
PQ1A	141.26	95.17	187.35	106.59
PQ1B	169.10	132.81	205.38	79.72
PQ1C	129.15	97.95	160.35	88.00
PQ1D	117.63	99.96	135.30	74.65
PQ2	68.69	63.28	74.10	43.68

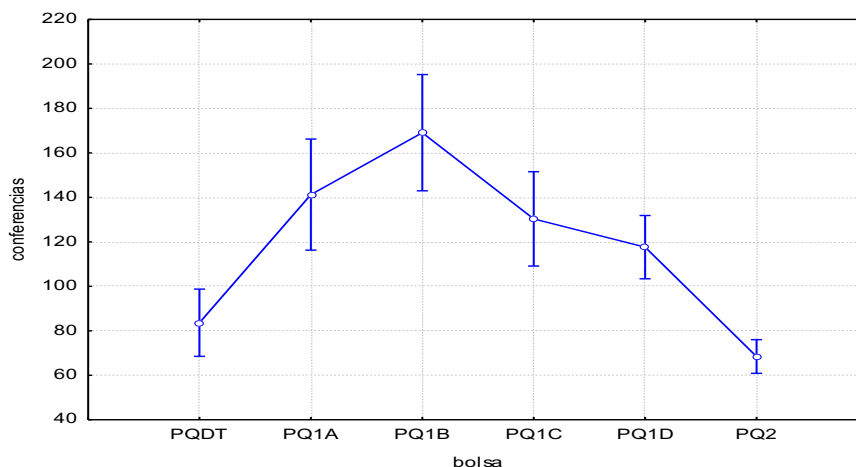


Figura 5.3. Gráfico com a média e intervalo de confiança da variável conferências.

Por meio da tabela 5.5 e da figura 5.2, podemos afirmar com 95% de confiança, pelos intervalos de confiança que são representados pelos traços verticais na figura, que os bolsistas PQ1A, PQ1B, PQ1C e PQ1D têm comportamentos diferentes dos demais pesquisadores, no caso, DT e PQ2. Mostrando que possivelmente uma variável pode indicar e classificar os bolsistas de nível 1 e demais.

### 5.3.3 Orientações de Mestrado

A variável orientações de mestrado também possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.6 mostra o resultado do teste de Tukey em relação a essa variável, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

Tabela 5.6. Teste de Tukey para variável orientações de mestrado.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		<b>0.000020</b>	<b>0.000426</b>	<b>0.000075</b>	<b>0.002643</b>	0.737911
PQ1A	<b>0.000020</b>		0.525957	0.299831	<b>0.001647</b>	<b>0.000020</b>
PQ1B	<b>0.000426</b>	0.525957		0.999971	0.569483	<b>0.000021</b>
PQ1C	<b>0.000075</b>	0.299831	0.999971		0.549060	<b>0.000020</b>
PQ1D	<b>0.002643</b>	<b>0.001647</b>	0.569483	0.549060		<b>0.000020</b>
PQ2	0.737911	<b>0.000020</b>	<b>0.000021</b>	<b>0.000020</b>	<b>0.000020</b>	

A tabela 5.6 mostra que o comportamento do bolsista PQ1A só difere entre os bolsista de nível 1 do PQ1D. Os bolsistas PQ1B e PQ1C não diferem entre os bolsistas nível 1, e o bolsista PQ1D difere do bolsista PQ1A. Em relação aos outros níveis de bolsa, esses pesquisadores nível 1 têm comportamentos diferentes dos demais, DT e PQ2, que no caso têm comportamentos parecidos entre si.

A tabela 5.7 e a figura 5.4 mostram a média, desvio padrão e o intervalo de confiança (95% de confiança) da variável orientações de mestrado.

Tabela 5.7. Tabela com a média, desvio padrão e intervalo de confiança da variável orientações de mestrado.

Bolsas	Média	Confiança		Desvio Padrão
		-95.00%	95.00%	
PQDT	15.32	11.73	18.90	14.24
PQ1A	37.04	26.89	47.20	23.48
PQ1B	29.90	23.61	36.20	13.84
PQ1C	28.67	21.93	35.41	19.01
PQ1D	24.25	20.67	27.84	15.14
PQ2	12.61	11.23	13.99	11.16

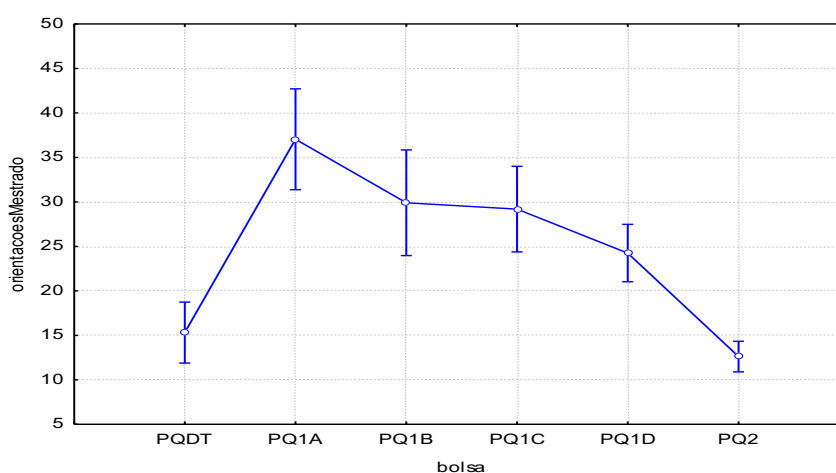


Figura 5.4. Gráfico com a média e intervalo de confiança para a variável orientações de mestrado.

Por meio da tabela 5.7 e da figura 5.4, podemos afirmar com 95% de confiança, pelos intervalos de confiança que são representados pelos traços verticais na figura, que os bolsistas nível 1 tem comportamentos diferentes dos demais pesquisadores. Por exemplo, o intervalo de confiança do bolsista PQ1D está entre 20 à 27 orientações de mestrado, em quanto o do bolsista PQ2 está entre 11 à 13 orientações de mestrado.

### 5.3.4 Orientações de Doutorado

A variável orientações de doutorado também possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.8 mostra o resultado do teste de Tukey em relação a essa variável, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

Tabela 5.8. Tabela com o teste de Tukey para a variável orientações de doutorado.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.000020	0.000020	0.000020	0.000038	1.000000
PQ1A	0.000020		0.095992	0.001714	0.000020	0.000020
PQ1B	0.000020	0.095992		0.936361	0.001868	0.000020
PQ1C	0.000020	0.001714	0.936361		0.016362	0.000020
PQ1D	0.000038	0.000020	0.001868	0.016362		0.000020
PQ2	1.000000	0.000020	0.000020	0.000020	0.000020	

A tabela 5.8 mostra uma grande diferença entre os bolsistas. Entre os bolsistas nível 1, os bolsistas PQ1A só tem um comportamento parecido com o bolsista PQ1B. O Bolsista PQ1B além dos bolsistas PQ1A, eles não diferem dos bolsistas PQ1C. Já os bolsistas PQ1D diferem de todos os bolsistas. Entre os outros bolsistas, só possuem comportamentos parecido entre PQ2 e DT.

A tabela 5.9 e a figura 5.5 mostram a média, desvio padrão e o intervalo de confiança (95% de confiança).

Table 5.9. Tabela com o teste de Tukey para a variável orientações de doutorado.

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		-95.00%	95.00%	
PQDT	2.16	1.04	3.28	4.46
PQ1A	15.30	10.50	20.11	11.12
PQ1B	11.29	9.34	13.23	4.27
PQ1C	9.79	7.59	11.98	6.19
PQ1D	6.44	4.98	7.89	6.14
PQ2	2.16	1.68	2.64	3.85

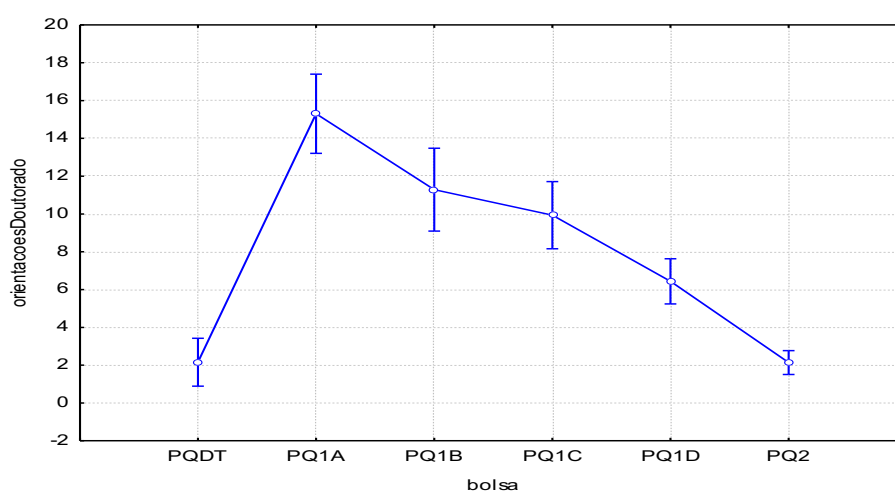


Figura 5.5. Gráfico com a média, desvio padrão e intervalo de confiança da variável orientações de doutorado.

Por meio da tabela 5.9 e da figura 5.5, podemos afirmar com 95% de confiança pelos intervalos de confiança que são representados pelos traços verticais na figura que os

bolsistas nível 1 tem comportamentos diferentes dos demais pesquisadores. Por exemplo, o intervalo de confiança do PQ1A está entre 10 à 20 orientações de doutorado, já o pesquisador DT têm um intervalo de confiança entre 1 à 3 orientações de doutorado.

### 5.3.5 Bancas de Mestrado

A variável bancas de mestrado também possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.10 mostra o resultado do teste de Tukey em relação a essa variável, em destaque vermelho são os valores que indicam diferença entre os bolsistas

**Table 5.10. Tabela com o teste de Tukey para a variável bancas de mestrado.**

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.999723	<b>0.020886</b>	0.946256	0.107781	0.999769
PQ1A	0.999723		0.160440	0.997516	0.631394	0.995961
PQ1B	<b>0.020886</b>	0.160440		0.276057	0.724874	<b>0.003472</b>
PQ1C	0.946256	0.997516	0.276057		0.855584	0.810542
PQ1D	0.107781	0.631394	0.724874	0.855584		<b>0.004523</b>
PQ2	0.999769	0.995961	<b>0.003472</b>	0.810542	<b>0.004523</b>	

A tabela 5.10 mostra que há grande paridade entre as categorias de bolsistas, existindo só alguns casos separados. Por exemplo, os bolsistas PQ1A e PQ1C não diferem de nenhum outro bolsista. Os bolsistas PQ1B diferem dos bolsistas DT e PQ2. O bolsista PQ1D só diferem dos bolsistas PQ2. Os DTs só diferem dos PQ1B e os bolsistas PQ2 só diferem dos bolsistas PQ1B e PQ1D.

A tabela 5.11 e a figura 5.6 mostram a média, desvio padrão e o intervalo de confiança (95% de confiança).

**Table 5.11. Tabela com a média, desvio padrão e intervalo de confiança da variável bancas de mestrado.**

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		<b>-95.00%</b>	<b>95.00%</b>	
PQDT	22.78	17.71	27.85	20.12
PQ1A	24.57	8.37	40.77	37.46
PQ1B	42.76	24.43	61.09	40.27
PQ1C	27.45	16.38	38.53	31.23
PQ1D	33.96	26.83	41.09	30.12
PQ2	21.75	19.19	24.31	20.67

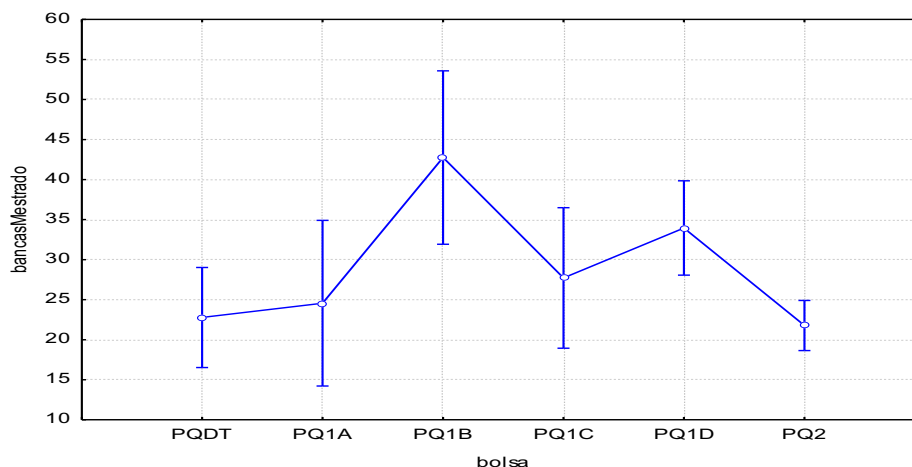


Figura 5.6. Gráfico com a média e intervalo de confiança da variável bancas de mestrado.

Por meio da tabela 5.11 e da figura 5.6, podemos afirmar com 95% de confiança, pelos intervalos de confiança que são representados pelos traços verticais na figura, que só é possível classificar individualmente as categorias de bolsas por meio dessa variável. Pois existem categorias de bolsas nível 1 que têm o mesmo comportamento dos bolsistas nível 2 e DT, como as variáveis PQ1A e PQ1C. Por exemplo, o intervalo de confiança do bolsista PQ1C que está entre 16 à 38 bancas de mestrado se encontra dentro do intervalo de confiança do bolsista PQ2 que têm um intervalo de confiança entre 19 e 24 bancas de mestrado.

### 5.3.6 Bancas de Doutorado

A variável bancas de doutorado também possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.12 mostra o resultado do teste de Tukey em relação a essa variável, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

Tabela 5.12. Tabela com o teste de Tukey para a variável bancas de doutorado.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		<b>0.001412</b>	<b>0.000020</b>	<b>0.000493</b>	<b>0.001298</b>	0.965995
PQ1A	<b>0.001412</b>		<b>0.044900</b>	0.999978	0.869749	<b>0.001814</b>
PQ1B	<b>0.000020</b>	<b>0.044900</b>		<b>0.013634</b>	<b>0.000079</b>	<b>0.000020</b>
PQ1C	<b>0.000493</b>	0.999978	<b>0.013634</b>		0.901288	<b>0.000415</b>
PQ1D	<b>0.001298</b>	0.869749	<b>0.000079</b>	0.901288		<b>0.000416</b>
PQ2	0.965995	<b>0.001814</b>	<b>0.000020</b>	<b>0.000415</b>	<b>0.000416</b>	

Nessa tabela 5.12, podemos perceber que há uma grande diferença entre as categorias de bolsas. Os bolsistas PQ1A, PQ1C e PQ1D, têm um mesmo comportamento em

relação à essa variável. Os bolsistas PQ1B se diferem de todos os outros pesquisadores. Já os pesquisadores PQ2 e DT só têm comportamentos parecidos entre eles.

A tabela 5.13 e a figura 5.7 mostram a média, desvio padrão e o intervalo de confiança (95% de confiança)

Tabela 5.13. Tabela com a média, desvio padrão e intervalo de confiança para a variável bancas de doutorado.

Bolsas	Média	Confiança		Desvio Padrão
		-95.00%	95.00%	
PQDT	5.78	3.83	7.73	7.74
PQ1A	17.13	9.04	25.22	18.71
PQ1B	27.57	14.33	40.82	29.10
PQ1C	16.45	10.47	22.44	16.88
PQ1D	13.89	11.04	16.74	12.04
PQ2	7.11	6.08	8.15	8.39

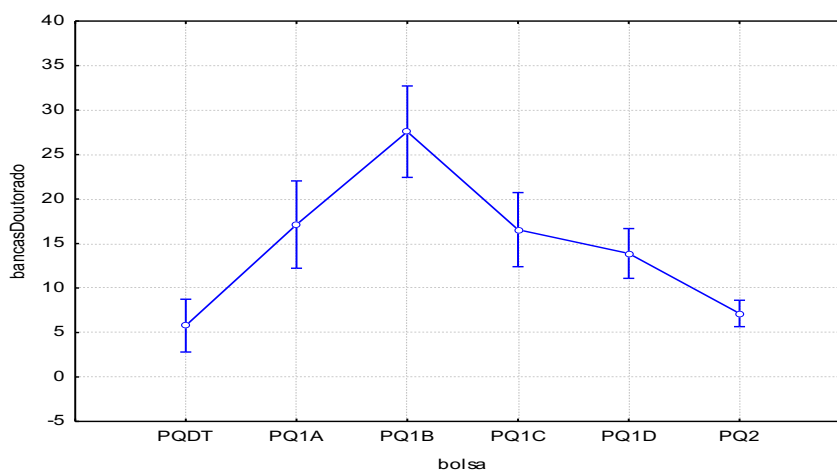


Figura 5.7. Gráfico com a média e intervalo de confiança da variável bancas de doutorado.

Por meio da tabela 5.13 e da figura 5.7, podemos afirmar com 95% de confiança, pelos intervalos de confiança que são representados pelos traços verticais na figura, que os bolsistas nível 1 têm comportamentos diferentes dos demais pesquisadores, permitindo classificar os bolsistas nível 1, nível 2 e DT. Como por exemplo, o pesquisador PQ1B que tem um intervalo de confiança entre 14 a 40 bancas de doutorado, enquanto o bolsista DT tem um intervalo de confiança entre 3 e 7 bancas de doutorado.

### 5.3.7 Centralidade de Grau

Como já foi descrito na seção 2.2.3.1, este tipo de medida permite aferir o grau de cada nó, dependendo das suas relações. É um método de avaliar o papel central de uma relação direta entre um ator para outros atores.

A métrica Centralidade de Grau também possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.14 mostra o resultado do teste de Tukey em relação a essa métrica, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

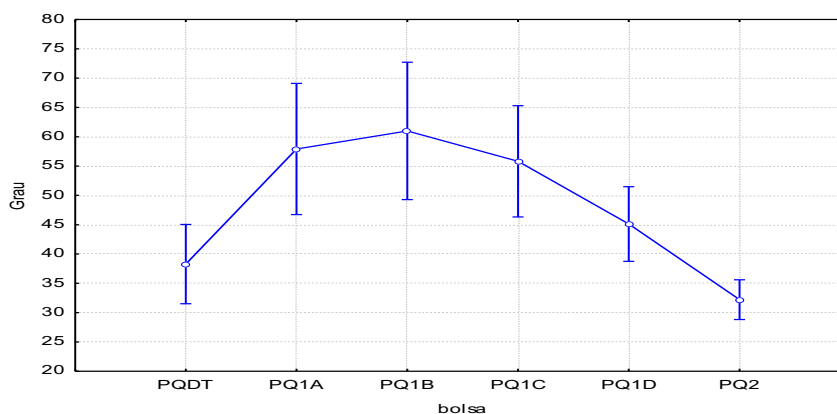
**Tabela 5.14. Tabela com o teste de Tukey para métrica centralidade de grau.**

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		<b>0.037484</b>	<b>0.012402</b>	<b>0.036774</b>	0.697969	0.615669
PQ1A	<b>0.037484</b>		0.999046	0.999764	0.370787	<b>0.000243</b>
PQ1B	<b>0.012402</b>	0.999046		0.984571	0.178194	<b>0.000069</b>
PQ1C	<b>0.036774</b>	0.999764	0.984571		0.441411	<b>0.000079</b>
PQ1D	0.697969	0.370787	0.178194	0.441411		<b>0.005914</b>
PQ2	0.615669	<b>0.000243</b>	<b>0.000069</b>	<b>0.000079</b>	<b>0.005914</b>	

A tabela 5.14 mostra que os bolsistas nível 1 têm os mesmos comportamentos entre si, em relação à métrica centralidade de grau, e comportamentos diferentes em relação aos bolsistas PQ2 e DT, com relação à mesma métrica. Destacando os bolsistas PQ1D que também têm comportamentos parecidos aos bolsistas DTs. Então podemos afirmar com 95% de confiança que os bolsistas nível 1 (A, B e C) tem uma colaboração maior que os demais pesquisadores, como mostra a tabela 5.15 e a figura 5.8 que apresentam a média, desvio padrão e o intervalo de confiança (95% de confiança).

**Tabela 5.15 Tabela com a média, desvio padrão e intervalo de confiança para métrica centralidade de grau.**

Bolsas	Média	Confiança -95.00%	Confiança 95.00%	Desvio Padrão
PQDT	38.29	29.51	47.06	34.83
PQ1A	57.91	43.67	72.16	32.94
PQ1B	61.00	47.43	74.57	29.80
PQ1C	55.03	44.72	65.34	29.09
PQ1D	45.13	39.09	51.16	25.48
PQ2	32.26	29.24	35.29	24.46



**Figura 5.8. Gráfico com a média e o intervalo de confiança da centralidade de grau.**

Então podemos perceber por meio da tabela 5.15 e da figura 5.8 que o pesquisador PQ1A, por exemplo, tem um intervalo de confiança entre 43 à 72 grau de colaboração e o pesquisador PQ2 tem um intervalo de confiança entre 29 à 35 grau de colaboração. Afirmando que o pesquisador PQ1A tem mais colaboradores que o pesquisador PQ2.

A tabela 5.16 mostra agora o teste de Tukey sem a divisão das subcategorias do nível 1 da bolsa de produtividade.

Table 4. Tabela com o teste de Tukey para os bolsistas PQ1.

Bolsas	PQDT	PQ1	PQ2
PQDT		0.003345	0.260955
PQ1	0.003345		0.000022
PQ2	0.260955	0.000022	

A tabela 5.16 mostra que existe diferenças entre os PQ1 com os PQ2 e DT e também expõe que os PQ2 e PQDT tem comportamentos semelhantes.

A tabela 5.17 e a figura 5.9 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) dos bolsistas PQ1, PQ2 e PQDT.

Tabela 5.17. Tabela com a média, desvio padrão e intervalo de confiança dos bolsistas PQ1, PQ2 e DT, para métrica grau.

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		-95,00%	95,00%	
PQDT	38.29	29.51	47.06	34.83
PQ1	51.57	46.93	56.22	28.61
PQ2	32.26	29.24	35.29	24.46

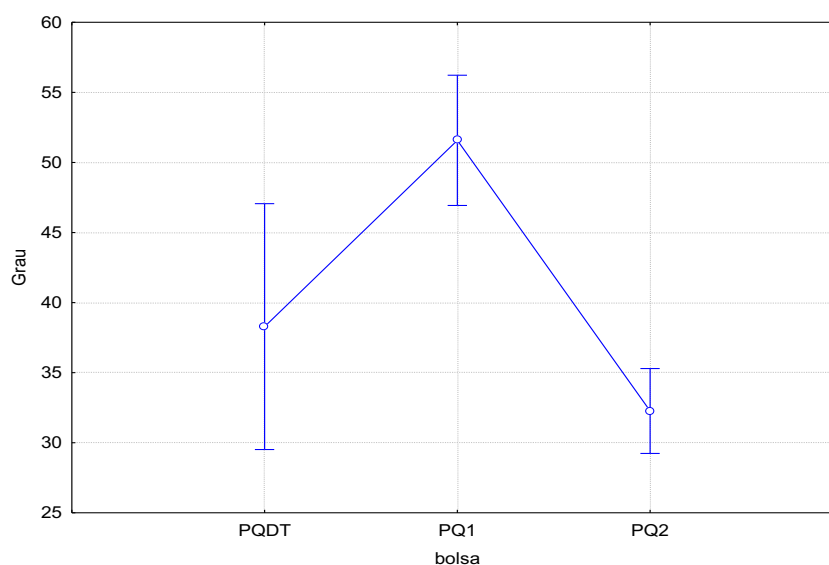


Figura 5.9. Gráfico com a média e intervalo de confiança dos bolsistas PQ1, PQ2 e PQDT, para métrica grau.

Por meio dessas informações da tabela 5.17 e da figura 5.9, notamos que os bolsistas PQ1 se destacam em colaboração em relação aos demais pesquisadores, mesmo o pico do intervalo de confiança do PQDT (47,06) passando de forma anódina do início do intervalo de confiança do PQ1 (46,93). Assim, os pesquisadores têm um maior grau de colaboração que os bolsistas PQ2 e PQDT, respondendo a **QP1** (Bolsistas de produtividade nível 1 têm mais colaboradores que a média dos bolsistas de produtividade nível 2 e DT na área?).

Isso mostra que os pesquisadores PQ1 têm um vínculo maior com os outros atores, podendo ter posições vantajosas na rede, dependendo menos dos outros indivíduos e podendo acessar e obter um número maior de informações que passa na rede.

Além da centralidade de grau, também foi medida a porcentagem da quantidade de colaboradores que têm um grau maior que o colaborador principal (COLA1), a porcentagem de colaboração maior de bolsistas de produtividade nível 1 que um pesquisador tem em relação a seus colaboradores (COLA2) e a porcentagem de colaboração maior de bolsistas de produtividade nível 2 que um pesquisador tem em relação a seus colaboradores (COLA3).

Como mostrou a tabela 5.1, a porcentagem da quantidade de colaboradores que tem um grau maior que o colaborador principal e a porcentagem de bolsistas de produtividade nível 1 que têm mais colaboradores bolsistas de produtividade nível 1 do que bolsistas de produtividade nível 2 e DT, não havendo nenhum tipo de diferença entre as categorias de bolsas.

A tabela 5.18 mostra o resultado do teste de Tukey em relação a medida COLA1.

**Tabela 5.18. Tabela com o teste de Tukey para a medida COLA1.**

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.813475	0.091698	0.937072	0.985434	0.994986
PQ1A	0.813475		0.845084	0.999127	0.970200	0.900162
PQ1B	0.091698	0.845084		0.578845	0.239683	0.101918
PQ1C	0.937072	0.999127	0.578845		0.998485	0.983619
PQ1D	0.985434	0.970200	0.239683	0.998485		0.999653
PQ2	0.994986	0.900162	0.101918	0.983619	0.999653	

Como podemos perceber, não há nenhum destaque em vermelho que mostra onde há diferença entre os bolsistas.

A tabela 5.19 e a figura 5.10 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) da medida.

Tabela 5.19. Tabela com a média, desvio padrão e o intervalo de confiança para medida COLA1.

Bolsas	Média	Confiança -95.00%	Confiança 95.00%	Desvio Padrão
PQDT	23.73	18.37	29.08	21.25
PQ1A	18.26	8.72	27.80	22.06
PQ1B	11.83	6.24	17.42	12.29
PQ1C	20.07	13.71	26.42	17.63
PQ1D	21.65	16.88	26.41	20.11
PQ2	22.68	20.56	24.80	17.11

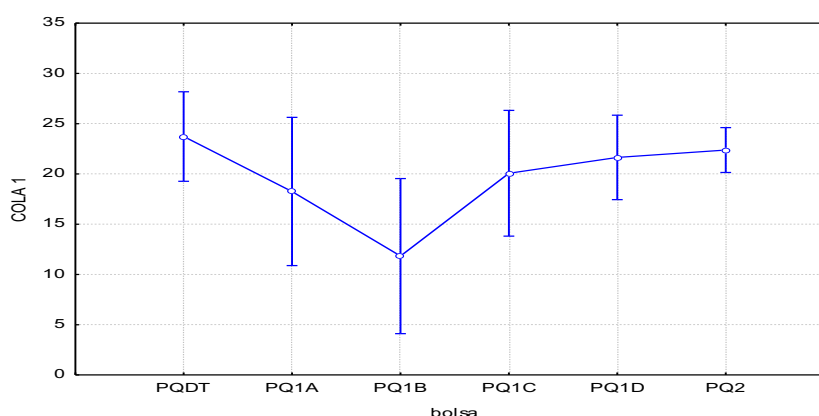


Figure 5.10. Gráfico com a média e intervalo de confiança para a medida COLA1.

Desta maneira, podemos responder a **QP3** com 95% de confiança que a proporção de pesquisadores que têm mais colaboradores que seus colaboradores é igual entre os bolsistas de produtividade nível 1 do que entre os demais pesquisadores bolsistas de produtividade.

A tabela 5.20 mostra o resultado do teste de Tukey em relação a essa medida COLA2.

Tabela 5.20. Tabela com o teste de Tukey para medida COLA2.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.254554	0.977869	0.999991	0.893867	0.996329
PQ1A	0.254554		0.866904	0.456558	0.710657	0.067389
PQ1B	0.977869	0.866904		0.994233	1.000000	0.872880
PQ1C	0.999991	0.456558	0.994233		0.979425	0.994537
PQ1D	0.893867	0.710657	1.000000	0.979425		0.404443
PQ2	0.996329	0.067389	0.872880	0.994537	0.404443	

Como podemos perceber, também não há nenhum destaque em vermelho que mostra onde há diferença entre os bolsistas.

A tabela 5.21 e a figura 5.11 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) da medida COLA2.

Tabela 5.21. Tabela com a média, desvio padrão e intervalo de confiança a medida COLA2.

Bolsas	Média	Confiança		Desvio Padrão
		-95.00%	95.00%	
PQDT	22.97	15.38	30.57	30.16
PQ1A	39.14	21.02	57.26	41.91
PQ1B	28.64	15.63	41.65	28.58
PQ1C	23.21	13.70	32.73	26.83
PQ1D	28.65	20.76	36.53	33.33
PQ2	20.75	17.12	24.37	29.31

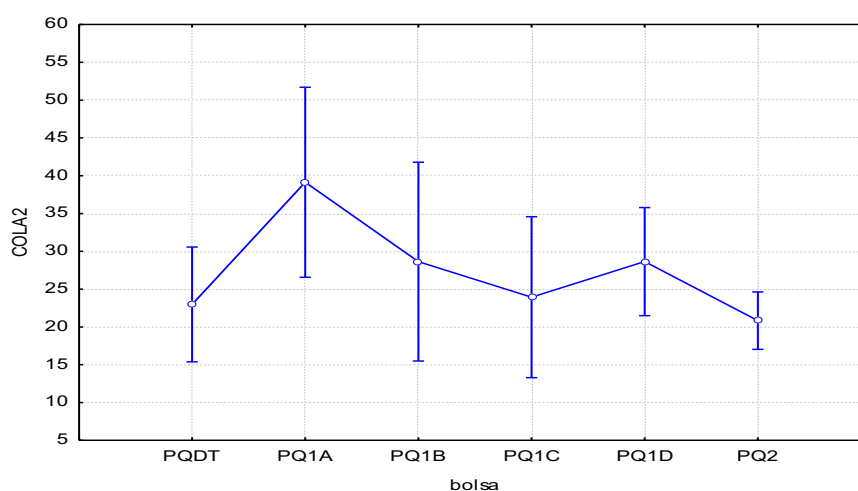


Figura 5.11. Gráfico com a média e intervalo de confiança da medida COLA2.

Por meio desses valores, podemos perceber que não há nenhuma diferença entre os bolsistas em relação as medidas COLA2.

Entre essas medidas, a única medida que houve alguma diferença entre os bolsistas foi a COLA3, como mostra a tabela 5.1. A tabela 5.22 mostra o teste de Tukey em relação a essa medida e é possível notar que existe apenas diferença entre dois tipos de bolsas.

Tabela 5.22. Tabela com o teste de Tukey para a medida COLA3.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.212442	0.564713	0.997243	0.701811	0.977671
PQ1A	0.212442		0.997860	0.162730	0.811435	<b>0.032487</b>
PQ1B	0.564713	0.997860		0.438566	0.985181	0.199375
PQ1C	0.997243	0.162730	0.438566		0.566793	1.000000
PQ1D	0.701811	0.811435	0.985181	0.566793		0.091091
PQ2	0.977671	<b>0.032487</b>	0.199375	1.000000	0.091091	

Como podemos perceber, só existe diferença entre os bolsistas PQ1A e PQ2 para esse tipo de medida.

A tabela 5.23 e a figura 5.12 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) da medida COLA3.

Tabela 5.23. Tabela com a média, desvio padrão e intervalo de confiança para medida COLA3.

Bolsas	Média	Confiança		Desvio Padrão
		-95.00%	95.00%	
PQDT	38.77	28.89	48.65	39.22
PQ1A	59.13	42.68	75.57	38.03
PQ1B	54.18	38.36	69.99	34.74
PQ1C	33.96	20.89	47.03	36.86
PQ1D	48.01	39.61	56.41	35.49
PQ2	35.20	30.59	39.81	37.15

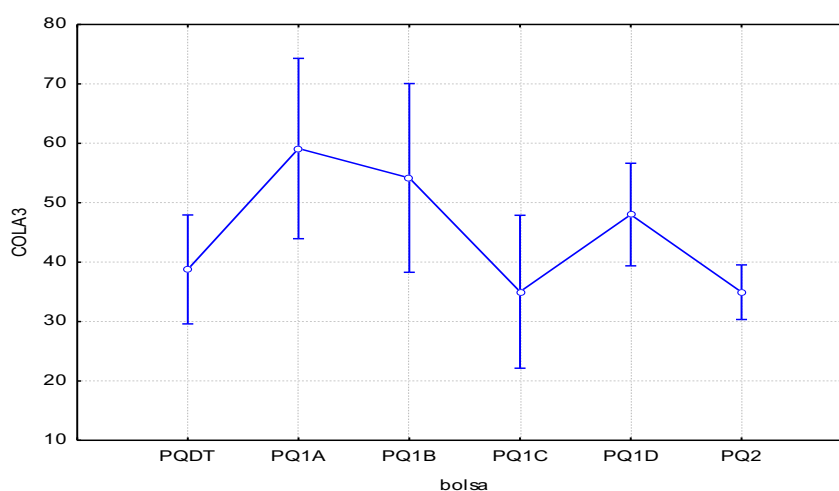


Figura 5.12. Gráfico com a média e intervalo de confiança da medida COLA3.

Assim, realizou-se o teste de Tukey novamente sem a divisão das subcategorias dos bolsistas nível 1 para podermos identificar de forma global como se comporta a diferença entre as bolsas, desta maneira, junto com a medida COLA2, conseguir responder a **QP2** (Bolsistas de produtividade nível 1 têm mais colaboradores bolsistas de produtividade do que a média dos bolsistas de produtividade nível 2 e DT?).

A tabela 5.24 mostra o teste de Tukey para a medida COLA3 sem a divisão dos bolsistas PQ1.

Table 5.24. Tabela com o teste de Tukey para os bolsistas PQ1, PQ2 e PQDT, para medida COLA3.

Bolsas	PQDT	PQ1	PQ2
PQDT		0.240177	0.744583
PQ1	0.240177		<b>0.002499</b>
PQ2	0.744583	<b>0.002499</b>	

Por meio da tabela 5.24, podemos notar que sem a divisão dos bolsistas nível 1, existe uma diferença desse bolsista com o pesquisador PQ2. Assim, com 95% de confiança, por meio da análise estatística realizada nas medidas COLA2 e COLA3, podemos afirmar que não há diferença na proporção dos bolsistas de produtividade nível 1 com mais

colaboradores bolsistas de produtividade nível 1 do que a proporção dos bolsistas de produtividade nível 2. Mas existe diferença na proporção de na proporção dos bolsistas de produtividade nível 1 com mais colaboradores bolsistas de produtividade nível 2 do que a proporção dos bolsistas PQ2.

### 5.3.8 Centralidade de Proximidade

A Centralidade de Proximidade permite definir qual a distância métrica de determinado nó por meio da medição do comprimento dos seus caminhos mais curtos, auxiliando a medida de Centralidade de Grau que podem ser críticas, porque só tomam conta as relações imediatas que o ator possui, sem levar em conta as relações indiretas com os demais atores. Assim, um ator pode estar conectado com muitos outros nós, mas esses nós podem estar distante da maioria do conjunto da rede. Em um caso como este, o ator pode ser bastante central, mas só em uma área local.

Mas a métrica Centralidade de Proximidade, diferente das outras variáveis e métrica, não possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.25 mostra o resultado do teste de Tukey em relação a essa métrica.

**Tabela 5.25. Tabela com o teste de Tukey para métrica centralidade de proximidade.**

<b>Bolsas</b>	<b>PQDT</b>	<b>PQ1A</b>	<b>PQ1B</b>	<b>PQ1C</b>	<b>PQ1D</b>	<b>PQ2</b>
<b>PQDT</b>		0.702467	0.624584	0.751514	1.000000	0.222039
<b>PQ1A</b>	0.702467		0.999992	0.999952	0.655753	0.999983
<b>PQ1B</b>	0.624584	0.999992		0.999315	0.576876	0.999356
<b>PQ1C</b>	0.751514	0.999952	0.999315		0.700477	0.999998
<b>PQ1D</b>	1.000000	0.655753	0.576876	0.700477		0.144862
<b>PQ2</b>	0.222039	0.999983	0.999356	0.999998	0.144862	

Como podemos perceber não há nenhum valor em destaque vermelho, pois todos as categorias de bolsas têm o mesmo comportamento. Então não podemos afirmar com 95% de confiança quem tem uma melhor centralidade de proximidade, como mostra a tabela com o teste de Tukey e as tabela 5.26 e a figura 5.13 que mostram a média, desvio padrão e intervalo de confiança (95% de confiança) para essa métrica.

**Tabela 5.26. Tabela com a média, desvio padrão e intervalo de confiança da métrica centralidade de proximidade.**

<b>Bolsas</b>	<b>Média</b>	<b>Confiança</b>	<b>Confiança</b>	<b>Desvio Padrão</b>
		<b>-95.00%</b>	<b>95.00%</b>	
<b>PQDT</b>	0.527	0.520	0.535	0.031
<b>PQ1A</b>	0.521	0.518	0.524	0.007
<b>PQ1B</b>	0.520	0.517	0.523	0.007
<b>PQ1C</b>	0.522	0.518	0.525	0.010

<b>PQ1D</b>	0.527	0.524	0.531	0.017
<b>PQ2</b>	0.521	0.519	0.524	0.017

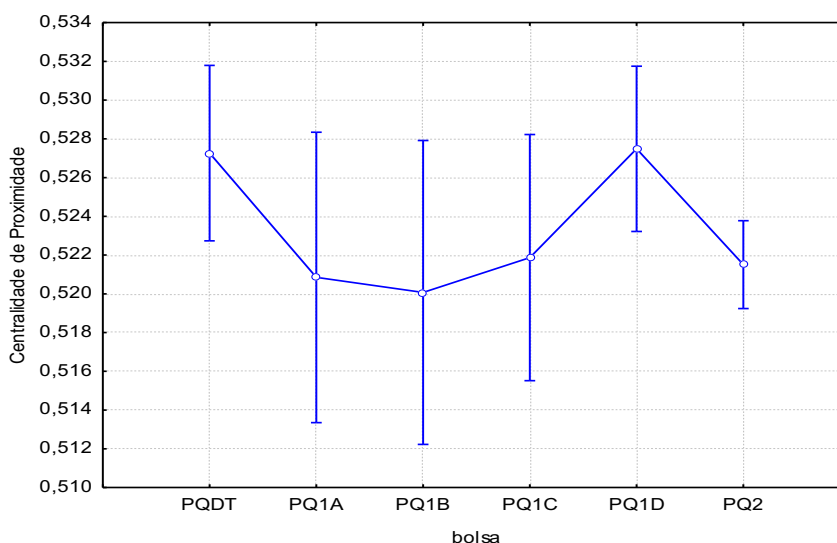


Figura 5.13. Gráfico com a média e intervalo de confiança da métrica centralidade de proximidade.

Desta maneira, os resultados mostraram que a distância métrica de determinado nó, por meio da medição do comprimento dos seus caminhos mais curtos, são bem parecidos entre os bolsistas, não havendo nenhuma diferença entre eles. Assim, essa métrica não nos permite fazer qualquer tipo de classificação de categorias de bolsas.

### 5.3.9 Centralidade de Intermediação

A centralidade de intermediação é uma medida para avaliar a importância de um nó na rede, quantificando o número de vezes que esse nó funciona como uma ponte ao longo do caminho mais curto entre dois outros nós. Na sua concepção, os vértices que têm uma alta probabilidade de ocorrência entre dois vértices escolhidos aleatoriamente, têm uma alta intermediação.

Essa métrica possui diferenças no comportamento dos bolsistas como mostra a tabela 5.1 na seção anterior. A tabela 5.27 mostra o resultado do teste de Tukey em relação a essa métrica, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

Tabela 5.27. Tabela com o teste de Tukey da métrica centralidade de intermediação.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.695065	0.840143	0.663206	0.999942	0.334232
PQ1A	0.695065		0.999977	0.999999	0.773254	<b>0.040179</b>
PQ1B	0.840143	0.999977		0.999999	0.895937	0.103852
PQ1C	0.663206	0.999999	0.999999		0.751663	<b>0.014748</b>

PQ1D	0.999942	0.773254	0.895937	0.751663		0.161805
PQ2	0.334232	<b>0.040179</b>	0.103852	<b>0.014748</b>	0.161805	

A tabela acima mostra que existem poucas diferenças dos bolsistas nessa métrica. Em destaque estão os bolsistas PQ2 que diferenciam dos bolsistas PQ1A e PQ1C.

A tabela 5.28 e a figura 5.14 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) para essa métrica.

Tabela 5.28. Tabela com a média, desvio padrão e intervalo de confiança para métrica centralidade de intermediação.

Bolsas	Média	Confiança		Desvio Padrão
		-95.00%	95.00%	
PQDT	0.526	0.476	0.576	0.199
PQ1A	0.593	0.527	0.658	0.151
PQ1B	0.583	0.516	0.650	0.147
PQ1C	0.589	0.538	0.641	0.144
PQ1D	0.533	0.494	0.572	0.164
PQ2	0.472	0.447	0.497	0.202

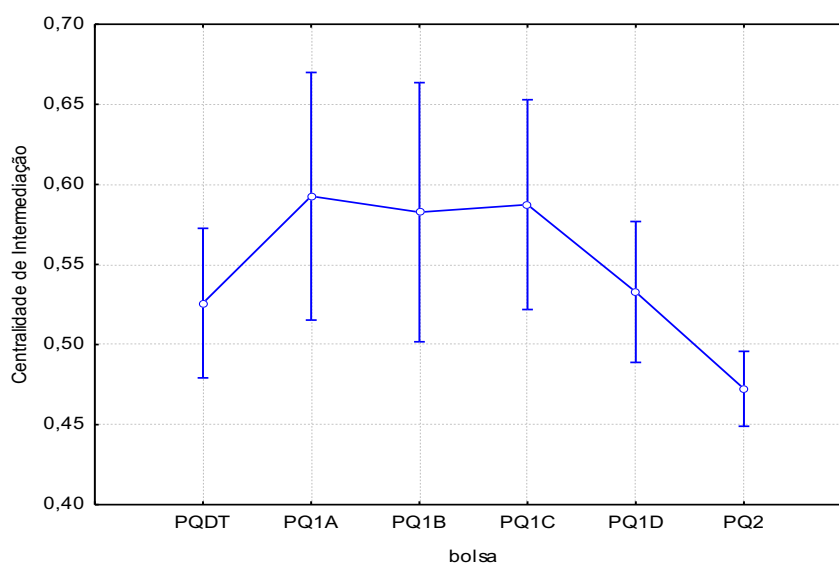


Figura 5.14. Gráfico com a média e intervalo de confiança para métrica centralidade de intermediação.

Com todas essas informações fica muito difícil poder classificar se um pesquisador nível 1 é diferente de um que é nível 2. Desta maneira, realizamos mais uma vez o teste de Tukey, mas agora sem a divisão das subcategorias dos bolsistas de produtividade nível 1.

A tabela 5.29 mostra o teste de Tukey sem separação das subcategorias do bolsista nível 1 para métrica em questão.

Tabela 5.29. Tabela com o teste de Tukey para os bolsistas PQ1, PQ2 e PQDT para métrica centralidade de intermediação.

Bolsas	PQDT	PQ1	PQ2
PQDT		0.428711	0.108921
PQ1	0.428711		<b>0.000038</b>
PQ2	0.108921	<b>0.000038</b>	

Desta maneira, pelas informações da tabela 5.27, podemos afirmar com 95% de confiança que os pesquisadores nível 1 têm um comportamento diferente com os pesquisadores nível 2.

A tabela 5.30 e a figura 5.15 mostram a média, desvio padrão e intervalo de confiança dos bolsistas nível 1, 2 e DT.

Tabela 5.30. Tabela com a média, desvio padrão e intervalo de confiança para os bolsistas PQ1, PQ2 e PQDT, para métrica centralidade de intermediação.

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		<b>-95,00%</b>	<b>95,00%</b>	
PQDT	0.526	0.476	0.576	0.199
PQ1	0.562	0.536	0.587	0.157
PQ2	0.472	0.447	0.497	0.202

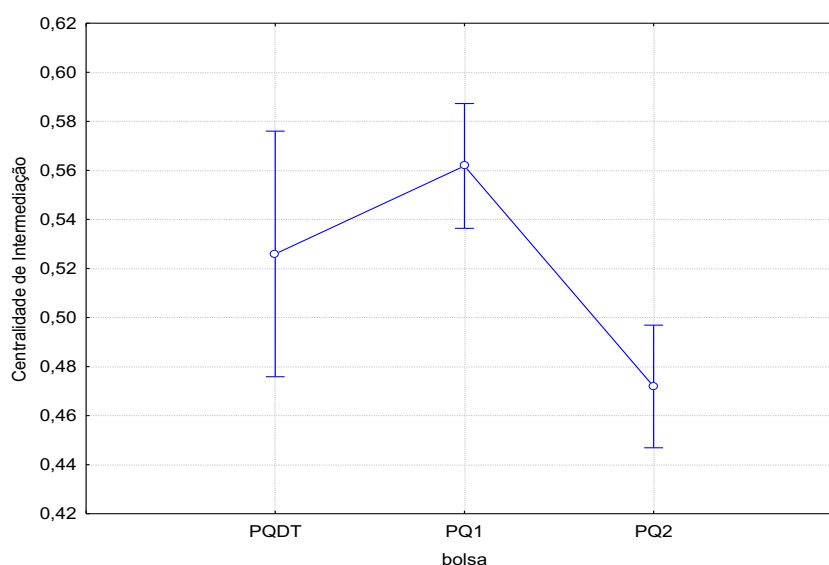


Figura 5.15. Gráfico com a média e intervalo de confiança dos bolsistas PQ1, PQ2 e PQDT, para variável centralidade de intermediação.

Percebemos que os bolsistas PQ1 está quase todo no intervalo de confiança dos pesquisadores PQDT, tendo comportamento bem parecidos, mas diferentes dos bolsistas PQ2. Além disso, podemos dizer que os bolsistas PQ1 tem uma importância maior em sua rede para com os outros colaboradores que os pesquisadores PQ2.

### 5.3.10 Centralidade de AutoVetor

Como já foi descrito na seção 2.2.3.2 deste trabalho, o resultado da centralidade de proximidade pode ser distorcido se a análise acontece em uma rede muito grande e complexa. Por exemplo, dois atores A e B, sendo A mais próximo de um grupo menor e fechado de uma rede maior, mas bastante distante de todos os membros da “população” da rede. O ator B está à moderada distância de todos os atores da rede global. A medição de proximidade dos dois atores A e B poderia ser bastante similar em magnitude, embora o ator B seja mais central que A neste exemplo, porque B pode alcançar mais nós da rede com o mesmo esforço. Desta maneira, a medida centralidade de autovetor tem como objetivo encontrar os atores mais centrais, em termos de estrutura global ou geral, e ter menos enfoque padrões mais locais.

Essa métrica possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.31 mostra o resultado do teste de Tukey em relação à essa métrica, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

**Tabela 5.31. Tabela com o teste de Tukey para métrica centralidade de autovetor.**

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		<b>0.009102</b>	<b>0.002322</b>	<b>0.041071</b>	<b>0.029968</b>	0.614089
PQ1A	<b>0.009102</b>		0.998617	0.981441	0.805920	<b>0.000041</b>
PQ1B	<b>0.002322</b>	0.998617		0.873036	0.525095	<b>0.000024</b>
PQ1C	<b>0.041071</b>	0.981441	0.873036		0.995543	<b>0.000092</b>
PQ1D	<b>0.029968</b>	0.805920	0.525095	0.995543		<b>0.000021</b>
PQ2	0.614089	<b>0.000041</b>	<b>0.000024</b>	<b>0.000092</b>	<b>0.000021</b>	

Por meio da tabela 5.31, podemos perceber que todos os bolsistas nível 1 têm um comportamento, em relação à centralidade de autovetor parecido. E esses bolsistas têm comportamentos diferentes dos demais pesquisadores, no caso, DT e nível 2, que já tem comportamento parecido entre si.

Esses valores demonstram uma melhor compreensão de quanto central é o ator na rede, em relação a centralidade de proximidade, onde os valores ficaram muito próximos, não possibilitando distinguir nenhuma diferença entre os bolsistas.

A tabela 5.32 e a figura 5.16 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) para essa métrica.

Tabela 5.32. Tabela com o teste de Tukey da variável centralidade de autovetor.

Bolsas	Média	Confiança -95.00%	Confiança 95.00%	Desvio Padrão
PQDT	0.607	0.544	0.670	0.250
PQ1A	0.805	0.718	0.892	0.202
PQ1B	0.834	0.735	0.934	0.219
PQ1C	0.761	0.671	0.852	0.254
PQ1D	0.733	0.676	0.790	0.240
PQ2	0.554	0.525	0.584	0.238

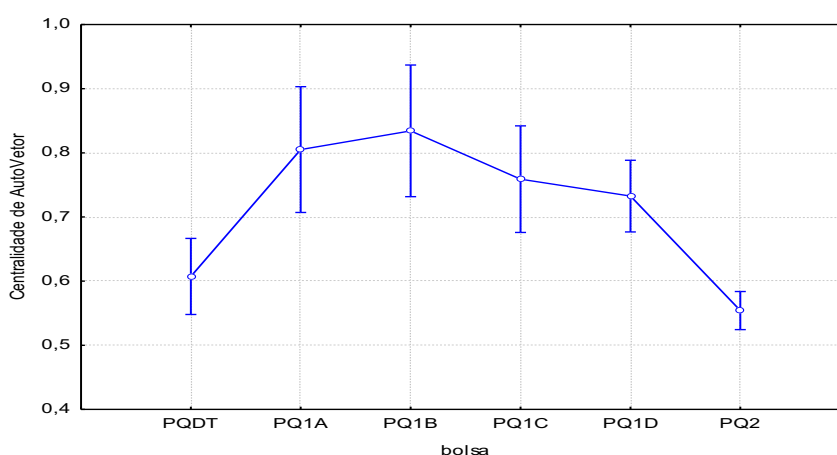


Figura 5.16. Gráfico com a média e intervalo de confiança para métrica autovetor.

Como podemos perceber na tabela 5.32 e figura 5.16 que o bolsista PQ1A com um intervalo de confiança entre 0,718 à 0,892 e o bolsista PQ2 com um intervalo de confiança entre 0,525 à 0,584, então com 95% de confiança, então podemos afirmar que o pesquisador PQ1A, como todos os outros pesquisadores nível 1, são mais centrais em termos de estrutura global ou geral que os pesquisadores bolsistas PQ2 e DT.

### 5.3.11 PageRank

A medida PageRank também permite mensurar a importância dos atores em uma rede, como a medida de centralidade intermediação. Mas por meio do teste de ANOVA, essa métrica tem diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. Mas a tabela 5.33 mostra o resultado do teste de Tukey em relação a essa métrica, que mostram pouca diferença nesse comportamento.

Tabela 5.33. Tabela com o teste de Tukey da métrica PageRank.

Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		0.916467	0.855524	0.817771	0.930564	0.340331
PQ1A	0.916467		0.999987	0.999999	0.523217	0.999979
PQ1B	0.855524	0.999987		0.999999	0.436038	1.000000
PQ1C	0.817771	0.999999	0.999999		0.317835	1.000000
PQ1D	0.930564	0.523217	0.436038	0.317835		<b>0.011017</b>
PQ2	0.340331	0.999979	1.000000	1.000000	<b>0.011017</b>	

Como podemos perceber, existe apenas diferença entre os bolsistas PQ1D e PQ2. Os demais pesquisadores têm comportamento parecido. A tabela 5.34 e a figura 5.17 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) para essa métrica.

Tabela 5.34. Tabela com a média, desvio padrão e intervalo de confiança da métrica PageRank.

Bolsas	Média	Confiança -95.00%	Confiança 95.00%	Desvio Padrão
PQDT	0.0443	0.0340	0.0547	0.0411
PQ1A	0.0361	0.0300	0.0423	0.0142
PQ1B	0.0346	0.0276	0.0416	0.0154
PQ1C	0.0355	0.0290	0.0421	0.0186
PQ1D	0.0499	0.0393	0.0605	0.0447
PQ2	0.0348	0.0310	0.0387	0.0311

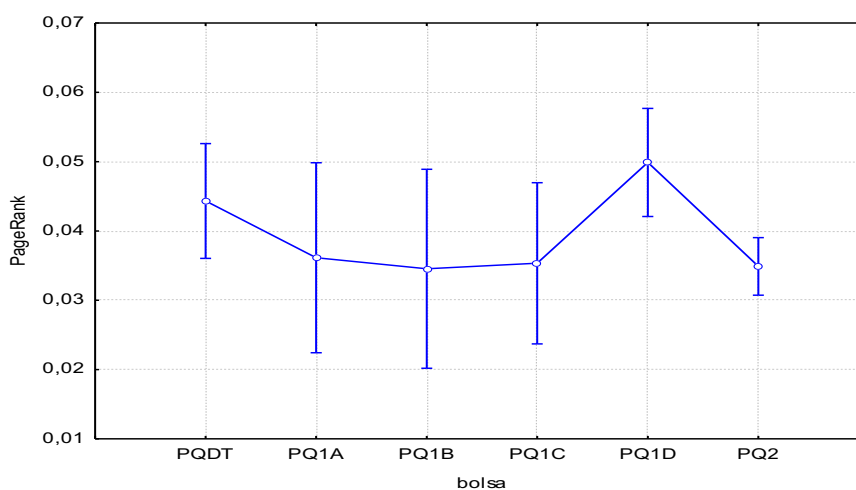


Figura 5.17. Gráfico com a média e intervalo de confiança da métrica PageRank.

Por meio da tabela 5.29 e da figura 5.15, podemos notar, pela informação obtida pelo teste de Tukey, que existe pouca diferença entre os bolsistas. Por exemplo, o intervalo de confiança do pesquisador PQ1A está entre 0,0300 à 0,0423 e o intervalo de confiança do pesquisador PQ2 0,0310 à 0,0387, mostrando que esses pesquisadores estão na mesma faixa de intervalo. Os únicos que diferem são os pesquisadores PQ2 com os pesquisadores PQ1D, que possuem intervalo de confiança entre 0,0393 à 0,0605. Desta maneira, é muito difícil

poder classificar a categoria de bolsa de um pesquisador. Então realizamos mais uma vez o teste de Tukey, mas agora sem a divisão dos bolsistas do nível 1. A tabela 5.35 mostra o teste de Tukey para os bolsistas PQ1, PQ2 e PQDT.

Tabela 5.35. Tabela do teste de Tukey dos bolsistas PQ1, PQ2 e PQDT para métrica PageRank.

Bolsas	PQDT	PQ1	PQ2
PQDT		0.921086	0.113503
PQ1	0.921086		0.080733
PQ2	0.113503	0.080733	

A tabela 5.35 mostra que ao juntar as subcategoria em uma, acaba deixando de existir qualquer tipo de diferença entre os bolsistas. Desta maneira, podemos afirmar com uma confiabilidade de 95% que os pesquisadores PQ1, PQ2 e PQDT, têm a mesma importância e influência em suas redes, em relação a seus pesquisadores, pois possuem comportamentos parecidos.

### 5.3.12 Número de Triângulos

Quantos atores estão à diferentes distâncias uns dos outros, pode ser importante para entender as diferenças entre atores nas limitações e oportunidades que têm como resultado de sua posição. Como por exemplo, que foi descrita em uma das seções do trabalho, dois atores A e B, podem cada um ter cinco amigos, mas supondo que nenhum amigo de A tem amigos, exceto o próprio ator A. Em contra partida cada amigo de B tem também cinco amigos. As informações disponíveis para B e o seu potencial de influenciar é muito maior que A. Desta maneira, as vezes ser “amigo do amigo” pode ter bastantes consequências. Por isso tiramos a medida do número de triângulos ou cliques.

Essa métrica possui diferenças no comportamento dos bolsistas, como mostra a tabela 5.1 na seção anterior. A tabela 5.36 mostra o resultado do teste de Tukey, em relação à essa métrica, em destaque vermelho são os valores que indicam diferença entre os bolsistas.

Tabela 5.36. Tabela com o teste de Tukey na métrica número de triângulos.

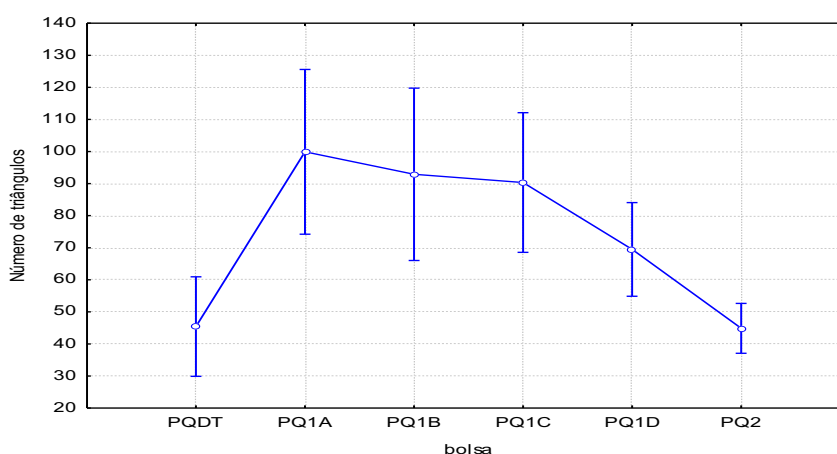
Bolsas	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2
PQDT		<b>0.004823</b>	<b>0.031527</b>	<b>0.012297</b>	0.228863	1.000000
PQ1A	<b>0.004823</b>		0.999094	0.993583	0.328310	<b>0.000798</b>
PQ1B	<b>0.031527</b>	0.999094		0.999991	0.661069	<b>0.009680</b>
PQ1C	<b>0.012297</b>	0.993583	0.999991		0.622534	<b>0.001560</b>
PQ1D	0.228863	0.328310	0.661069	0.622534		<b>0.040909</b>
PQ2	1.000000	<b>0.000798</b>	<b>0.009680</b>	<b>0.001560</b>	<b>0.040909</b>	

Como mostra a tabela 5.36, os pesquisadores nível 1 têm comportamentos parecidos, não havendo nenhum tipo de diferenças entre eles, além do mais, o pesquisador PQ1D possui um comportamento parecido com os pesquisadores PQDT. Os pesquisadores PQ2 só possuem comportamento parecido com os pesquisadores DTs.

A tabela 5.37 e a figura 5.18 mostram a média, desvio padrão e intervalo de confiança (95% de confiança) para essa métrica.

**Tabela 5.37. Tabela de média e intervalo de confiança da métrica número de triângulos.**

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		-95.00%	95.00%	
<b>PQDT</b>	45.41	32.67	58.16	50.61
<b>PQ1A</b>	99.91	54.45	145.38	105.14
<b>PQ1B</b>	92.90	62.34	123.46	67.14
<b>PQ1C</b>	88.52	60.14	116.89	80.02
<b>PQ1D</b>	69.48	52.38	86.58	72.25
<b>PQ2</b>	44.89	38.21	51.57	53.84



**Figura 5.18. Gráfico da média e intervalo de confiança da métrica número de triângulos.**

A tabela 5.37 e a figura 5.18 mostram que os pesquisadores nível 1 estão no mesmo intervalo de confiança, também nos permite afirmar que os bolsistas PQ1A, PQ1B, PQ1C têm um número de vizinhos maior que os demais pesquisadores. Como, por exemplo, o intervalo de confiança do pesquisador PQ1C está entre 60 à 116 número de triângulos, enquanto o do pesquisador PQ2 está entre 38 à 51 número de triângulos. Mas os pesquisadores PQ1D têm um procedimento parecido com os pesquisadores PQDT. Então realizamos outro teste de Tukey sem a separação dos níveis dos bolsistas nível 1, para podermos saber se os pesquisadores PQ1 são diferentes dos pesquisadores PQDT.

A tabela 5.38 mostra o teste de Tukey para os bolsistas PQ1, PQ2 e PQDT.

Tabela 5.38. Tabela com o teste de Tukey para os bolsistas PQ1, PQ2 e PQDT, para métrica número de triângulos.

Bolsas	PQDT	PQ1	PQ2
PQDT		0.000324	0.998043
PQ1	0.000324		0.000022
PQ2	0.998043	0.000022	

A tabela 5.38 mostra que quando unida as subcategorias dos bolsistas nível 1, eles passam a ter diferenças com todas as outras categorias de bolsas, PQ2 e PQDT.

A tabela 5.39 e a figura 5.19 mostram a média, desvio padrão e intervalo de confiança dos bolsistas para essa métrica.

Tabela 5.39. Tabela com a média, desvio padrão e intervalo de confiança dos bolsistas PQ1, PQ2 e PQDT, para métrica número de triângulos.

Bolsas	Média	Confiança	Confiança	Desvio Padrão
		-95,00%	95,00%	
PQDT	45.41	32.67	58.16	50.61
PQ1	81.78	68.88	94.67	79.37
PQ2	44.89	38.21	51.57	53.84

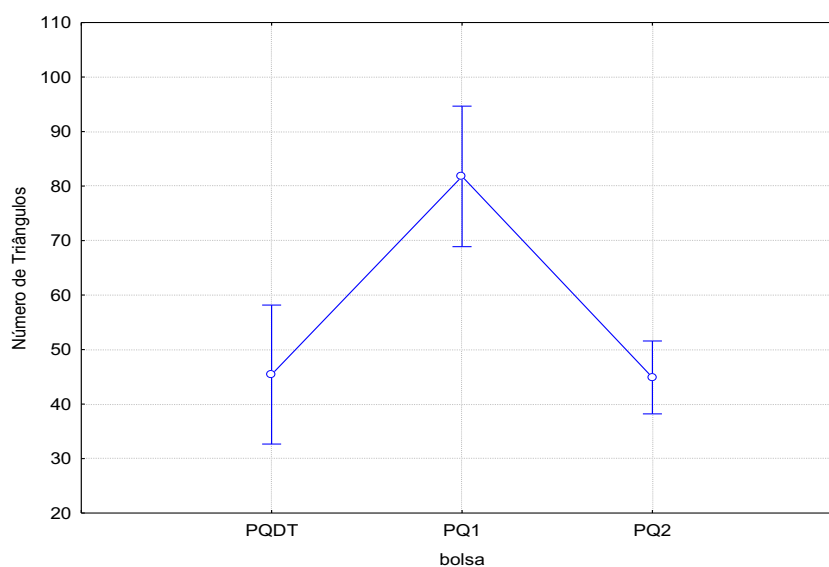


Figura 5.19. Gráfico com a média e intervalo de confiança dos bolsistas PQ1, PQ2 e PQDT, para métrica número de triângulos.

Desta maneira, podemos dizer com um grau de confiança de 95%, que os bolsistas PQ1, têm uma quantidade maior de pesquisadores em comum com seus colaboradores que os pesquisadores PQ2 e PQDT. Mostrando que os pesquisadores PQ1 têm uma probabilidade de colaboração e interação maior entre esses colaboradores, além de uma conexão múltipla mais forte que entre os outros pesquisadores.

## 5.5 CONCLUSÃO DO CAPÍTULO

Neste capítulo foram expostos todos os resultados obtidos pelo trabalho e descrito cada um deles, podemos perceber que em relação as variáveis, métricas e medidas que foram analisadas, em apenas 3 (Centralidade de Proximidade, COLA1 e COLA2) não houveram nenhum tipo de diferença entre os bolsistas.

A tabela 5.40 mostra a codificação utilizada na tabela 5.41 que mostra um resumo dos resultados das participações dos pesquisadores em relação a periódicos, conferências, bancas de mestrado e doutorado e orientações de mestrado e doutorado que foram descritos na seção anterior.

**Tabela 5.40. Codificação para as diferentes modalidades de bolsa do CNPq.**

<b>Código</b>	<b>Modalidades de Bolsas</b>
<b>1A</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nivel 1A
<b>1B</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nivel 1B
<b>1C</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nivel 1C
<b>1D</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nivel 1D
<b>2</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nivel 2
<b>DT</b>	Bolsista de Produtividade Desen. Tec. e Extensao Inovadora do CNPq

Tabela 5.41. Tabela com o resumo dos resultados das variáveis da seção 5.4.

Variáveis	Desempenho Diferente						Desempenho Similar						Conclusões
	PQ1A	PQ1B	PQ1C	PQ1D	PQ2	PQDT	PQ1A	PQ1B	PQ1C	PQ1D	PQ2	PQDT	
Periódicos	1B, 1C, 1D, DT e 2	1A, 1D, DT e 2	1A, DT e 2	1A, 1B, DT e 2	1A, 1B, 1C e 1D	1A, 1B, 1C e 1D	-	1C	1B e 1C	1C	DT	2	PQ1 publicam mais periódicos que os DT e 2.
Conferência	DT e 2	1D, DT e 2	DT e 2	1B, DT e 2	1A, 1B, 1C e 1D	1A, 1B, 1C e 1D	1B, 1C e 1C	1A e 1C	1A, 1B e 1D	1A e 1C	DT	2	PQ1 participam mais de conferências que os DT e 2.
O.Mestrado	1D, DT e 2	DT e 2	DT e 2	1A, DT e 2	1A, 1B, 1C e 1D	1A, 1B, 1C e 1D	1B, 1C	1A, 1C e 1D	1A, 1B e 1D	1B e 1C	DT	2	PQ1 participam mais em orientações de mestrado que os DT e 2.
O.Doutorado	1C, 1D, DT e 2	1D, DT e 2	1A, 1D, DT e 2	1A, 1B, 1C, DT e 2	1A, 1B, 1C e 1D	1A, 1B, 1C e 1D	1B	1A e 1C	1B	-	DT	2	PQ1 participam mais em orientações de doutorado que os DT e 2.
B.Mestrado	-	DT e 2	-	2	1B e 1D	1B	1B, 1C, 1D, DT e 2	1A, 1C e 1D	1A, 1B, 1D, DT e 2	1A, 1B, 1C e DT	1A, 1C e DT	1A, 1C, 1D e 2	PQ1B e PQ1D participam mais em bancas de mestrado que os 2.
B.Doutorado	1B, DT e 2	1A, 1C, 1D, DT e 2	1B, DT e 2	1B, DT e 2	1A, 1B, 1C e 1D	1A, 1B, 1C e 1D	1C e 1D	-	1A e 1D	1A e 1C	DT	2	PQ1 participam mais em bancas de doutorado que os DT e 2.

Podemos analisar e concluir, por meio dos resultados e pela tabela 5.41, com um grau de confiança de 95%, que os bolsistas nível 1 (PQ1A, PQ1B, PQ1C e PQ1D) tem um comportamento diferentes na maioria das variáveis acadêmicas, em relação aos outros pesquisadores. Por exemplo, os bolsistas nível 1 publicam mais periódicos, participam de mais conferências, têm um número maior de orientações de mestrado e doutorado e tem uma participação maior em bancas de doutorado em relação aos bolsistas DT e PQ2.

Em relação a variável bancas de mestrado, os únicos que tiveram desempenho diferentes foram os bolsistas PQ1C e PQ1D que têm uma participação maior em bancas de mestrado em relação aos bolsistas PQ2.

A tabela 5.42 mostra a codificação utilizada para a tabela 5.43 que mostra um resumo dos resultados das métricas de análise de redes sociais que foram medidas no trabalho, são elas: Centralidade de Grau, Centralidade de Proximidade, Centralidade de Intermediação, Centralidade de AutoVetor, PageRank, Número de Triângulos, além das medidas. Nessa tabela será mostrado os bolsistas sem a divisão da categoria dos bolsistas nível 1, pois fica melhor o entendimento, lembrando que para as métricas, realizamos um teste Tukey sem essa divisão. A tabela 5.42 mostra a codificação utilizada na tabela 5.44.

**Tabela 5.42. Codificação para as diferentes modalidades de bolsa do CNPq.**

<b>Código</b>	<b>Modalidades de Bolsas</b>
<b>1</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nível 1
<b>2</b>	Bolsista de Produtividade em Pesquisa do CNPq - Nível 2
<b>DT</b>	Bolsista de Produtividade Desen. Tec. e Extensao Inovadora do CNPq

Tabela 5.43. Tabela com o resumo dos resultados das métricas da seção 5.4.

Métricas e Medidas	Desempenho Diferente			Desempenho Similar			Conclusões
	PQ1	PQ2	DT	PQ1	PQ2	DT	
Grau	DT e 2	1	1	-	DT	2	PQ1 têm mais colaboradores que os bolsistas DT e 2. Eles dependem menos dos outros e obter um número maior de informações que passa na rede.
Proximidade	-	-	-	DT e 2	1 e DT	1 e 2	Todos têm comportamentos similares Não sendo uma boa medida de classificação.
Intermediação	2	1	-	DT	DT	1 e 2	PQ1 têm desempenho diferente em relação ao PQ2. Eles são mais importantes para sua rede interna que os PQ2.
AutoVetor	DT e 2	1	1	-	DT	2	PQ1 é mais central que os PQ2 e DT.
PageRank	-	-	-	DT e 2	DT e 1	1 e 2	Todos têm comportamentos similares. Não sendo uma boa métrica para classificação.
Triângulos	DT e 2	1	1	-	DT	2	PQ1 têm um número de vizinhos maior que os DT e 2. Tendo uma interação maior entre com os colaboradores.

Com os resultados das métricas de análise de rede sociais obtidos e resumidos na tabela 5.43. Podemos afirmar que os pesquisadores PQ1, têm uma maior colaboração que os pesquisadores DT e PQ2, possuindo posições vantajosas na rede, impondo uma maior influência para com os demais colaboradores de sua rede.

Desta maneira, é possível responder a **QP1** e afirmar, com 95% de confiança, que os bolsistas de produtividade nível 1 têm mais colaboradores que a média dos bolsistas de produtividade nível 2 e DT.

Esses pesquisadores também se diferenciam dos demais pesquisadores em relação ao número de triângulos formados e em relação a centralidade de autovetor, tendo uma interação maior com seus colaboradores e sendo mais centrais e mais requisitados em sua rede.

Os pesquisadores nível 1, 2 e DT tiveram comportamentos parecidos na centralidade proximidade e PageRank. Assim, essas métricas não nos permitem fazer algum de tipo de classificação entre os bolsistas.

Os pesquisadores nível 1 e 2 têm comportamentos diferentes em relação a centralidade de intermediação, sendo o primeiro mais importante para sua rede interna que o segundo.

Portanto, podemos responder a **QP4** e afirmar, com 95% de confiança, que existem diferenças nas métricas de centralidade (centralidade de grau, centralidade de autovetor, número de triângulos) com uma maior diferença dos bolsistas de produtividade nível 1 dos bolsistas de produtividade nível 2 e DT. E na métrica de centralidade intermediação há diferença apenas entre os PQ1 e PQ2.

A tabela 5.42 mostra a codificação utilizada para a tabela 5.44 que mostra um resumo dos resultados das medidas criadas que possibilitassem responder as **QP2** e **QP3**. A tabela 5.43 mostra a codificação utilizada na tabela 5.45.

Tabela 5.44. Tabela com o resumo dos resultados das medidas da seção 5.4.

Medidas	Desempenho Diferente			Desempenho Similar			Conclusões
	PQ1	PQ2	PQDT	PQ1	PQ2	PQDT	
COLA1	-	-	-	DT e 2	DT e 1	1 e 2	Todos os bolsistas têm a mesma proporção de mais colaboradores que seus colaboradores.
COLA2	-	-	-	DT e 2	DT e 1	1 e 2	Todos os bolsistas têm a mesma proporção de colaboradores que são PQ1.
COLA3	PQ2	PQ1	-	DT	DT	1 e 2	PQ1 têm uma proporção maior de colaboração com bolsistas de produtividade nível 2, que a proporção dos PQ2.

Medimos a porcentagem de quantos colaboradores têm um grau maior que o seu colaborador principal, como também a porcentagem de colaboração maior com bolsistas de produtividade nível 1 e 2 que um pesquisador têm, em relação a seus colaboradores, como mostra a tabela 5.44. Desta maneira, podemos responder a **QP2**, com 95% de confiança, que a proporção de pesquisadores que têm mais colaboradores que seus colaboradores é igual entre todos os bolsistas de produtividade, não havendo nenhum tipo de diferença entre eles. E também podemos responder a **QP3**, com 95% de confiança, que a proporção colaboração maior de colaboradores bolsistas de produtividade nível 1 é igual para todos os bolsistas. Mas a proporção de colaboradores com colaboração maior de bolsistas de produtividade 2 é maior que a proporção dos pesquisadores PQ2.

## Capítulo 6 – Conclusões e Trabalhos Futuros

Este capítulo apresenta as conclusões obtidas no trabalho e quais os futuros trabalhos a serem realizados.

### 6.1 CONCLUSÕES

Com a realização da análise dos dados dos pesquisadores do CNPQ da área de Ciências da Computação, por meio da utilização de diferentes métricas de análise de redes sociais, podemos compreender como se comporta a rede social dos pesquisadores bolsistas de produtividade, e como está acontecendo as colaborações entre eles, de forma a responder todas as questões de pesquisa elencadas nas seção 1.4.

Avaliamos os pesquisadores de produtividade do CNPq por meio das métricas de centralidade de grau, proximidade, intermediação e autovetor, pagerank e número de triângulos que existiam em cada rede individualmente. Das variáveis tinham a quantidade de periódicos, conferências, orientações de mestrado e doutorado e participações em bancas de mestrado e doutorado, além das medidas COLA1, COLA2 e COLA3 que foram criadas, realizando uma análise de variância para cada métrica, variável e medida citadas.

Por meio dos resultados, podemos perceber que existem métricas de centralidade ( Centralidade de Grau, AutoVetor e Número de Triângulos) que diferenciam os bolsistas de produtividade nível 1 dos bolsistas de produtividade nível 2 e DT **QP4**. Onde geralmente o pesquisador de nível 1 (PQ1A, PQ1B, PQ1C e PQ1D) tem o mesmo comportamento entre si e os melhores resultados em relação aos demais pesquisadores, possuindo posições vantajosas na rede, impondo uma maior influência para com os demais colaboradores de sua rede. Constatamos também que a os bolsistas de produtividade nível 1 têm mais colaboradores que a média dos bolsistas de produtividade nível 2 e DT **QP1**.

Em relação a porcentagem de quantos colaboradores têm um grau maior que o seu colaborador principal, podemos afirmar com 95% de confiança, que a proporção de pesquisadores que têm mais colaboradores que seus colaboradores é igual entre todos os bolsistas de produtividade **QP2**. Já em relação a porcentagem de colaboração maior com bolsistas de produtividade nível 1 em relação a seus colaboradores, podemos afirmar com 95% de confiança, que têm a mesma porcentagem. Em relação a porcentagem de colaboração maior com bolsistas de produtividade nível 2 em relação a seus colaboradores, podemos

afirmar com 95% de confiança, que o PQ1 têm uma proporção maior que a dos pesquisadores PQ2 QP3.

Além disso, os bolsistas de produtividade nível 1 publicam mais periódicos, participam de mais conferências, têm um número de orientações de mestrado e doutorado maior e tem uma participação maior em bancas de doutorado em relação aos demais pesquisadores. Mostrando que essas atividades são muito importantes para classificação do mérito científico.

## **6.2 Trabalhos Futuros**

Por meio da experiência obtida com a realização deste trabalho, sabemos que ainda há muito o que ser feito, pois a análise de redes sociais é um assunto muito desbravador e que permite uma expansão muito maior do que é planejado inicialmente.

É preciso criar uma ferramenta online que possa ajudar os pesquisadores que queiram criar sua rede de colaboração e que possa medir as métricas e comparar com os resultados obtidos nesse trabalho. Realizar um estudo mais detalhado dos pesquisadores em relação à porcentagem do gênero (masculino e feminino) na rede dos pesquisadores e até se aprofundar sobre as regiões e quais instituições esses pesquisadores são e estão no momento.

Além disso, as redes de pesquisadores podem ser expandido em outras áreas, tais como: Física, engenharias, matemática, humanas e analisar se, nessas áreas, as redes de pesquisadores seguem a mesma lógica mostrada na rede de pesquisadores dos bolsistas de produtividade do CNPq em Computação.

## REFERÊNCIAS

BIGGS, N.; LLOYD, E. and WILSON, R. (1986), *Graph Theory, 1736-1936*, Oxford University Press

BONDY, J. A.; MURTY, U. S. R. Título: Graph theory and related topics. Ano de publicação: 1979

BORGATTI, S.P., EVERETT, M.G. & JOHNSON, J.C. (2013). *Analyzing Social Networks*. London: Sage Publications. ISBN-13: 978-1446247419.

CORMEN, Thomas H. (2002). *Algorithmic Complexity*. CRC Press.

CHAPULA, Cesar A. Macias-Chapula. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. *Ci. Inf., Brasília*, v. 27, n. 2, p. 134-140, maio/ago. 1998

CLEVELAND, W. S. Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review / Revue Internationale de Statistique*, 21-26. 2001

CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico. <<http://www.cnpq.br>>. Acesso em: 14 de junho de 2013.

CNPQ2 – Conselho Nacional de Desenvolvimento Científico e Tecnológico. <<http://efomento.cnpq.br/efomento/distribuicaoGeografica/distribuicaoGeografica.do?metodo=apresentar>>. Acesso em: 3 de julho de 2014.

CURRÍCULO LATTES. <<http://lattes.cnpq.br>>. Acesso em: 14 de junho de 2013.

DA SILVA, Alzira Karla Araújo; BARBOSA, Ricardo Rodrigues; DUARTE, Emeide Nóbrega. Rede social de coautoria em Ciência da Informação: estudo sobre a área temática de Organização e Representação do Conhecimento. **Informação & Sociedade: Estudos** 22.2 (2012).

DE SÁ, Hially Rodrigues. Predição de Relacionamentos em Redes Sociais. Dissertação de Mestrado. Recife, 2011

EFOMENTO. <<http://efomento.cnpq.br/efomento/>>. Acesso em: 14 de junho de 2013.

EULER, Leonhard. Solutio problematis ad geometriam situs pertinentes. Pag. 128-140. 1741.

GRAEML, Alexandre Reis, MACADA, Marie Anne; ROSSONI, Luciano. Redes sociais e intelectuais em administração da informação: uma análise cientométrica do período 1997-2006. **Informação e Sociedade** 20.1 (2010): 95-110.

GEPHI. <<http://gephi.org>>. Acesso em: 14 de junho de 2013.

HAIR, Joseph F. Jr.; ANDERSON, Rolph E.; TATANHAM, Ronald L.; BLACK, William C. **Análise Multivariada de Dados**. 5 ed. Porto Alegre: Bookman, 2005.

HANNEMANAND, M. Riddle, **Introduction to social network methods**. Acesso em <<http://www.faculty.ucr.edu/hanneman/nettext/>>. 2005.

KNELLER, G. F. A Ciência como atividade humana. Rio de Janeiro: Zahar, 1980.

LACOMBE, Francisco; HEILBORN, Gilberto. Administração: Princípios e Tendências. São Paulo: **Saraiva**, 2003.

LATTESCRAWLER.

<<https://github.com/marcilioLemos/LABSNA/tree/master/LattesCrawler/src/br/ufpb/ci/labsna/lattescrawler>>. Acesso em: 03 de julho de 2014.

MAIA, M. F. S. ; CAREGNATO, S. E. . **Co-autoria como indicador de redes de colaboração científica. Perspectivas em Ciência da Informação**, v. 13, p. 18-31, 2008.

MINAYO, M. C. S. et al. Pesquisa social: Teoria, Método e Criatividade. Petrópolis: Vozes, 1994.

MORETTIN, P. A. e Toloi, C. M. C. **Estatística Básica. 8ª Ed.** Editora Saraiva. 2013.

NASCIMENTO, S. Do; BEUREN, Ilse Maria. Redes sociais na produção científica dos programas de pós-graduação de Ciências Contábeis do Brasil. **Revista de Administração Contemporânea** 15.1 (2011): 47-66.

NORONHA, D.P. *et al.* Produção científica: análise cienciométrica das comunicações apresentadas nos SNBUs 1978-1998. In: XI Seminário Nacional de Bibliotecas Universitárias, 2000, Florianópolis, SC. **Anais...** Florianópolis: UFSC-BU, 2000. p. 1- 12.

OLIVEIRA, Ely Francina Tannuri; SANTAREM, Luciana Garcia da Silva; SEGUNDO, José Eduardo Santarem. **Análise das redes de colaboração científica através do estudo das co-autorias, nos cursos de pós-graduação do Brasil no tematratoamento temático da informação.** *Nuevas perspectivas para la difusión y organización del conocimiento: actas del congreso.* Servicio de Publicaciones, 2009.

PAULEEN, D.;YOONG, P. **Facilitating virtual team relationships via Internet and conventional communication channels,** Internet Research: Electronic Networking Application sand Policies (11:3), 2001, pp. 190-202.

PORTAL DA CAPES. <<http://www.capes.gov.br>>. Acesso em: 14 de junho de 2013.

SCHEFFER, A. B. B.; OLIVEIRA, Josiane Silva de. ; ULLRICH, Danielle Regina. . Formação de Redes Sociais de Co-Autoria na área de Gestão de Pessoas: Uma Análise Bibliométrica em Periódicos Brasileiros no Triênio de 2007 a 2009. **REGE. Revista de Gestão USP**, v. 19, p. 553-570, 2012.

SILVA, Antonio Braz de Oliveira, et al. Estudo da rede de co-autoria e da interdisciplinaridade na produção científica com base nos métodos de análise de redes sociais: avaliação do caso do programa de pós-graduação em ciência da informação-PPGCI/UFMG 10.5007/1518-2924.2006 v11nesp1p179. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação** 11.1 (2007): 179-194.

SPINAK, Ernesto. Indicadores cienciometricos. *Ci. Inf., Brasília*, v. 27, n. 2, p. 141-148, maio/ago. 1998.

STATISTICA. <<http://www.statsoft.com/Resources/Support/Software-Updates>>. Acesso em: 23 de julho de 2014.

TEXEIRA, Enise Barth. *A Análise de Dados na Pesquisa Científica importância e desafios em estudos organizacionais.* Desenvolvimento em Questão. Editora Unijuí, jul./dez. p. 177-201. 2003.

TRIOLA, Mário F. *Introdução à Estatística.* LTC. 10a edição 2008. 722p. ISBN 85-216-1586-8

TUESTA, E. F. ; DELGADO, K.V. ; DIGIAMPIETRI, L. ; ALCAZAR, J. J. P. ; MUGNAINI, R. ; MENA-CHALCO, J. . Análise temporal da relação orientador-orientado: um estudo de caso sobre a produtividade dos pesquisadores doutores da área de Ciência da Computação. In: **Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)...** Curitiba- Pr. Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2012.

VANTI, N.A.P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento, Brasília, **Ciência da Informação**, v. 31, n. 2, p. 152-162, maio/ago. 2002.

WEBQUALIS. <<http://qualis.capes.gov.br/webqualis/principal.seam>>. Acesso em: 14 de junho de 2013.

YUSOF, Norazah; RAHMAN, Azizah Abdul. Students' interactions in online asynchronous discussion forum: A Social Network Analysis. **Education Technology and Computer... ICETC'09. International Conference on.** IEEE, 2009.