

# Inferência estatística clássica para a confiabilidade de rede de coautoria com enfoque nos vértices

**Sandra Cristina de Oliveira**

**Professora Doutora do curso de Administração da  
UNESP - Câmpus de Tupã**

**Taiane de Paula Ferreira**

**Estudante do curso de Administração da UNESP -  
Câmpus de Tupã**

**Beatriz Barbero Brigantini**

**Estudante do curso de Administração da UNESP -  
Câmpus de Tupã**

**Jéssica Katty Uehara**

**Estudante do curso de Administração da UNESP -  
Câmpus de Tupã**

<http://dx.doi.org/10.1590/1981-5344/1928>

*Um grupo de pesquisa pode ser considerado uma rede social, a qual pode ser modelada por um grafo  $G$  composto por  $k$  vértices e  $m$  arestas. Os pesquisadores que compõem essa rede podem ser interpretados como seus vértices, e as conexões ou ligações entre esses pesquisadores (representadas por trabalhos em coautoria) podem ser consideradas suas arestas. O objetivo deste trabalho foi estudar a medida de confiabilidade de redes considerando os vértices ou pesquisadores não confiáveis e as arestas ou conexões perfeitamente confiáveis. Especificamente, foi proposta uma abordagem inferencial clássica para a confiabilidade de redes, obtendo os estimadores de máxima verossimilhança e os respectivos intervalos de confiança para os componentes individuais (pesquisadores) e para a rede de coautoria; foi aplicada a metodologia a um grupo de pesquisa da UNESP cadastrado no CNPq; e, foram obtidas medidas de centralidade de vértices para auxiliar na identificação de situações onde a inserção de uma aresta ou ligação entre dois pesquisadores do grupo poderia aumentar significativamente a confiabilidade da rede de coautoria. Os resultados mostraram a utilidade da*

*inferência estatística no contexto de confiabilidade de redes sociais, ressaltando que a contribuição de cada pesquisador é de extrema importância para a manutenção de um grupo de pesquisa. Verificou-se ainda que o cálculo da confiabilidade de uma rede de coautoria pode ser bastante desgastante de ser executado e que as medidas de centralidade são uma ferramenta viável quando se intenciona aumentar a confiabilidade dessa rede.*

**Palavras-chave:** *redes sociais; grupos de pesquisa; teoria dos grafos; inferência estatística; medidas de centralidade.*

## **Classical statistical inference for the reliability of co-authorship network with emphasis in nodes**

*A research group may be considered a social network, which may be modeled by a graph  $G$  with  $k$  nodes and  $m$  edges. Researchers that make up this network can be interpreted as its nodes or actors, and the connections or links between those researchers (represented by co-authored papers) can be considered as its edges. The aim of this study was to measure the reliability of networks considering unreliable nodes or researchers and perfectly reliable edges or connections. Specifically, a statistical analysis based on classical inference to the network reliability was proposed, obtaining the maximum likelihood estimators and confidence intervals for the individual components (researchers) and the co-authorship network; the methodology was applied to a research group of UNESP registered in CNPq; and measures of centrality of nodes were obtained to assist in identifying situations where the insertion of an edge or connection between two researchers of the group could significantly increase the reliability of this co-authorship network. The results showed the usefulness of statistical inference in the context of social networks reliability, noting that the contribution of each researcher is of extreme importance for the maintenance of a research group. It was also found that calculating the reliability of a co-authorship network can be quite exhausting to be executed and that the centrality measures are a viable tool when it intends to increase the reliability of this network.*

**Keywords:** *social networks; researcher groups; graph theory; statistical inference; centrality measures.*

Recebido em 26.11.2013 Aceito em 10.11.2014

## 1 Introdução

*Confiabilidade é a probabilidade de um item desempenhar satisfatoriamente uma função requerida sob condições específicas de operação; e o termo confiabilidade de rede está ligado ao cálculo da confiabilidade de qualquer configuração geral de itens (ou componentes), dada à confiabilidade de cada item individual (BARLOW; PROSCHAN, 1981).*

*Redes são sistemas físicos, biológicos ou sociais caracterizados por um conjunto grande de entidades bem definidas que interagem dinamicamente entre si. Para a manutenção da funcionalidade de uma rede é extremamente importante compreender a sua estrutura, bem como suas funções e propriedades. A estrutura de uma rede pode ser representada por um grafo. Para isso, a utilização de ferramentas da teoria dos grafos é fundamental na determinação de propriedades referentes aos aspectos topológicos de uma rede.*

*Intuitivamente, confiabilidade de rede é a probabilidade de uma rede continuar funcionando mesmo quando uma falha acarretar na remoção de um ou mais subconjuntos de componentes (arestas e/ou vértices). Portanto, redes altamente confiáveis são estruturas fortes e, pode-se dizer que uma rede é mais confiável que a outra, se a probabilidade de uma delas ser desconectada for menor que a da outra.*

*Redes sociais são estruturas compostas por pessoas, organizações, territórios, etc, que estão conectadas entre si por um ou vários tipos de relações e que compartilham informações, conhecimentos, interesses, valores e objetivos em comum. Estudos sobre redes sociais analisam o desenvolvimento da atividade de um grupo, indicando tanto os esforços do grupo quanto os esforços individuais. Estas redes podem funcionar em diferentes níveis, tais como, redes de relacionamento, redes de profissionais, redes comunitárias, redes políticas, etc. (LYRA; OLIVEIRA, 2011). Neste contexto, encontram-se as redes de coautoria científica.*

*Redes de coautoria são formadas por pesquisadores e por trabalhos compartilhados entre estes. Tais redes são simétricas, ou seja, o pesquisador um é colaborador do pesquisador dois, em um dado tempo  $t$ , o mesmo número de vezes que o pesquisador dois é colaborador do pesquisador um. A coautoria de produtos gerados pela atividade científica, particularmente de publicações, é um indicador de colaboração.*

*A institucionalização dos grupos de pesquisa no Brasil, por meio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, com a ampliação e atualização constante destes, é uma prática que vem consolidando a pesquisa no país. Desta forma, grupos de pesquisa altamente confiáveis, ou seja, com uma estrutura intensa de colaboração*

*dos pesquisadores, podem contribuir ampla e fortemente para o surgimento e/ou a concretização de ideias, uma vez que estes grupos são responsáveis por grande parte das investigações realizadas na atualidade e também pela formação de inúmeros pesquisadores (MARAFON, 2008).*

*As redes de coautoria têm sido amplamente exploradas por meio da metodologia de análise de redes sociais (ARS), a qual está especificamente interessada na inter-relação dos atores (pesquisadores) e busca entender os mecanismos sociais por trás das conexões (coautoria, citação, etc) e como estes facilitam o fluxo de informação e de conhecimento entre os atores. Segundo Abbasi; Altmann; Hwang (2010), a análise de redes de coautoria é uma das maneiras mais visíveis e acessíveis para identificar as relações de colaboração científica no ambiente acadêmico.*

*Os resultados de diversos estudos sobre coautoria usando ARS revelam que as colaborações entre pesquisadores têm aumentado em todas as áreas do conhecimento e mostram a importância da coautoria na solidificação da produção científica (YOSHIKANE; KAGEURA, 2004, NEWMAN, 2004, LEE; BOZEMAN, 2005, MAIA; CAREGNATO, 2008, HAYASHI; HAYASHI; LIMA, 2008; YAN; DING, 2009, ABBASI; ALTMANN; HWANG, 2010, GOLDENBERG et al., 2010; SOUZA; BARBASTEFANO, 2011, ABBASI; ALTMANN; HOSSAIN, 2011, BARBASTEFANO et al., 2013, ABBASI; WIGAND; HOSSAIN, 2014; dentre outros).*

*Por outro lado, existe a análise estrutural de redes que, por sua vez, estuda mais precisamente os papéis e posições dos atores na rede e utiliza conceitos da teoria dos grafos que fornecem a base matemática para realizar cálculos computacionais e utilizar abordagens mais específicas na análise de dados. Neste contexto, podem ser incluídos alguns poucos estudos sobre confiabilidade de redes de coautoria, os quais têm enfatizado basicamente as distribuições combinatoriais (LYRA; OLIVEIRA, 2011, FERREIRA; OLIVEIRA, 2013, OLIVEIRA; BRIGANTINI; UEHARA, 2013) e que, apenas recentemente, têm considerado o modelo estatístico e os aspectos inferenciais do problema (BRIGANTINI; OLIVEIRA; BRAGA JUNIOR, 2014).*

*Existem diversos estudos sobre confiabilidade de redes que destacam a modelagem e a inferência, no entanto, concentram-se em casos simples de sistemas em série e em paralelo e, especificamente, na resolução de problemas relacionados a redes físicas (OLIVEIRA; ACHCAR, 2000, SHPUNGIN, 2006, CRISTINO; CORDEIRO; LINS, 2007, CORDEIRO et al., 2010, SILVA, 2010, dentre outros).*

*Neste trabalho propõe-se uma interação entre conceitos da teoria dos grafos, da teoria da confiabilidade e da análise de redes sociais, associados à proposta de inferência estatística clássica (ou frequentista) para a confiabilidade de uma rede de coautoria representada por um grupo de pesquisa cadastrado no CNPq.*

*Alguns problemas têm sido apontados na literatura no que diz respeito à estruturação de redes sociais baseadas em coautoria, tais como, erros de grafia dos nomes sem a remoção de ambiguidades e que*

*levam à identificação incorreta de autores, artigos que constam no currículo de um autor, mas não no currículo dos demais, artigos que, por esquecimento, não são cadastrados por nenhum dos autores, etc (SMALHEISER; TORVIK, 2009, WANG, et al., 2012, BARBASTEFANO et al., 2013). Uma vez que a produção científica é cadastrada pelos próprios autores ou pesquisadores e tornada pública na Plataforma Lattes do CNPq (Currículo Lattes), os dados estão sujeitos a possíveis imprecisões de preenchimento. Tais imprecisões impactam na confiabilidade de um pesquisador dentro de uma rede de coautoria e, conseqüentemente, na confiabilidade desta rede. Portanto, justifica-se a relevância de uma abordagem inferencial para a avaliação da medida de confiabilidade de redes de coautoria.*

*Se, por um lado tem-se a manutenção de grupos de pesquisa e de suas atividades como um fator de extrema importância para o desenvolvimento científico e tecnológico do país, por outro, tem-se a necessidade de recursos de modelagem e de inferência para a confiabilidade de redes de coautoria que, aliados aos estudos existentes usando ARS, poderão gerar resultados relevantes na implicação de estratégias de sobrevivência e de competitividade, de forma que grupos de pesquisa mantenham a sua funcionalidade.*

## **2 Objetivos**

*Um grupo de pesquisa pode ser considerado uma rede social, a qual pode ser modelada por um grafo. Os pesquisadores que compõem essa rede podem ser interpretados como seus vértices, e os laços ou ligações entre esses agentes (representadas por publicações em comum ou trabalhos em coautoria) podem ser consideradas suas arestas.*

*O objetivo geral deste trabalho foi estudar a medida de confiabilidade de redes de coautoria com enfoque nos vértices ou pesquisadores, ou seja, considerando os vértices não confiáveis ou propensos a falhas (ou seja, um ou mais pesquisadores pode(m) deixar de pertencer à rede) e as arestas ou relações de coautoria perfeitamente confiáveis. Especificamente, pretendeu-se: propor uma análise estatística baseada em inferência clássica (ou frequentista) para a confiabilidade de uma rede de coautoria, obtendo os estimadores de máxima verossimilhança, bem como os respectivos intervalos de confiança para a confiabilidade dos vértices ou pesquisadores (individualmente) e para a confiabilidade da rede, ou seja, a probabilidade do grupo de pesquisa permanecer em atividade em um dado tempo  $t$ ; desenvolver a análise inferencial proposta para um grupo de pesquisa da UNESP, cadastrado no CNPq; e, finalmente, obter medidas de centralidade de vértices para auxiliar na identificação de situações onde a inserção de uma aresta (ou de uma ligação de coautoria entre dois pesquisadores do grupo) pode aumentar significativamente a confiabilidade desta rede.*

## 3 Revisão bibliográfica

### 3.1 Conceitos básicos da teoria dos grafos

*Nesta seção são apresentados alguns conceitos básicos da teoria dos grafos, os quais se encontram em Boaventura Netto e Jurkiewicz (2009).*

*Grafo é uma noção simples, abstrata e intuitiva usada para representar a ideia de alguma espécie de relação entre itens. É representado por uma figura com nós ou vértices, significando os itens, unidos por traços denominados arestas, configurando a relação imaginada.*

*A representação matemática de um grafo simples não orientado é  $G=(V,E)$ , onde  $V$  é um conjunto finito e não vazio cujos elementos são os vértices; e  $E$  é um conjunto de subconjuntos de dois elementos de  $V$  cujos elementos são as arestas. Além disso, o conjunto de vértices  $V$  tem cardinalidade (número de elementos)  $|V|=k$ , o conjunto de arestas  $E$  tem cardinalidade  $|E|=m$  e cada aresta é denotada por  $\{v_i, v_j\}$ , onde  $v_i, v_j \in V$ .*

*O grau de um vértice  $v_i$ , denotado por  $d(v_i)$ , é o número de arestas incidentes neste, e dois vértices são denominados adjacentes se existe uma aresta entre eles. Um percurso ou caminho, é uma família de ligações sucessivamente adjacentes. Quando a última ligação da sequência for adjacente à primeira, o percurso é considerado fechado e será denominado ciclo. Do contrário, este será aberto. Só ocorre um caminho quando todos os pontos e linhas do grafo forem distintos e, neste caso, tem-se uma caminhada ou trilha.*

*Um grafo é conexo quando, a partir de qualquer um de seus vértices, é possível chegar aos demais. Caso contrário, ele é denominado desconexo. A conectividade de aresta, denotada por  $\chi(G)$ , é o menor número de arestas cuja remoção torna o grafo  $G$  desconexo. A conectividade de vértice, denotada por  $\kappa(G)$ , é o menor número de vértices cuja remoção (juntamente com as arestas ligadas a ele) torna o grafo  $G$  desconexo. Um subgrafo gerador de  $G$  é um grafo obtido de  $G$  apenas pela eliminação de algumas de suas arestas (sem torná-lo desconexo).*

*Um grafo  $G$  com  $k$  vértices pode ser representado por uma matriz, denotada por  $A_{(G)}$  de ordem  $k$ , denominada matriz de adjacência de  $G$ , onde a entrada  $a_{ij}$  da matriz é igual a um, se  $v_i$  e  $v_j$  forem adjacentes, ou igual a zero, caso contrário, para todo  $i, j = 1, 2, \dots, k$ .*

*Dois grafos  $G=(V_1, E_1)$  e  $H=(V_2, E_2)$  são iguais quando  $V_1=V_2$  e  $E_1=E_2$ . São grafos isomorfos aqueles que possuem a mesma estrutura, ou seja, o mesmo número de vértices e de arestas, porém, possuem configuração diferente.*

### 3.2 Conceitos básicos da análise de redes sociais

*Os estudos sobre redes sociais são interdisciplinares e, por isso, apresentam diversas metodologias de análise, que têm como base a estrutura em forma de redes. Além da teoria dos grafos, outra metodologia utilizada nesse estudo é a análise de redes sociais (ARS). A linguagem matemática da teoria dos grafos é utilizada como base para a ARS e os conceitos são bastante semelhantes.*

*A seguir, apresentam-se alguns conceitos relevantes ao entendimento da ARS, os quais podem ser observados em Newman (2010) e Silva (2010).*

*Os atores, elementos ou vértices podem ser unidades sociais individuais (pessoas ou empresas) ou unidades sociais coletivas (instituições, organizações, nações). Os laços relacionais estabelecem a relação entre os atores. Estes laços podem ser classificados de três maneiras: laços ausentes, laços fracos ou laços fortes, e podem ocorrer devido a qualquer tipo de vínculo, tais como parentesco, amizade, profissional, etc.*

*A relação é um conjunto de laços que apresentam o mesmo critério de ligação. As relações possuem duas propriedades importantes que condicionam os métodos de análise de dados disponíveis: direcionamento e valoração.*

*Quanto ao direcionamento, uma relação pode ser direcional, ou seja, um ator é transmissor e o outro é receptor (amizade, citação, etc); ou não direcional, isto é, a relação é recíproca (conhecimento, coautoria, etc).*

*Quanto à valoração, uma relação pode ser dicotômica (implica na presença ou ausência de um determinado laço entre dois atores); ou valorada, com valores discretos ou contínuos (atribui-se peso à relação ou, por ex., quantos artigos foram produzidos em coautoria por um determinado número de pesquisadores). Os atributos de um ator são suas características individuais, tais como nome, sexo e idade.*

*Dentre as ferramentas mais utilizadas em ARS, têm-se: a estatística descritiva (gráficos, tabelas, distribuição de frequências, medidas descritivas, etc), as medidas de centralidade (grau de informação, de proximidade e de intermediação) e a análise de clusters (partições da rede em subconjunto de atores, construído a partir dos laços e da posição ocupada por eles).*

### **3.3 Cálculo da confiabilidade de redes**

*Dada uma rede modelada por um grafo simples não orientado  $G=(V,E)$  com  $k$  vértices e  $m$  arestas. Para esta rede estar em funcionamento (ou em atividade), em um dado tempo  $t$ , todo par de vértices deve estar conectado por pelo menos um caminho.*

*Considere que as arestas são perfeitamente confiáveis e somente os vértices estão propensos a falhas. Assim, cada vértice  $i$  ( $i=1,2,\dots,k$ ) possui uma probabilidade de funcionar (ou de não falhar), denotada por  $p_i$ . Existem casos onde todos os vértices do grafo que modela a rede possuem a mesma probabilidade de funcionar, podendo ser denotada*

*simplesmente por  $p$ . Além disso, os vértices são dois a dois independentes, isto é, a falha de um não implica na falha do outro.*

*Para calcular a confiabilidade de uma rede (probabilidade do grafo  $G$  que modela esta rede permanecer conexo, dada a falha de um ou mais vértices) é necessário, primeiramente, determinar a probabilidade de cada estado de funcionamento da rede, ou seja:*

$$\prod_{i \in V'} p_i \prod_{i \in (V \setminus V')} (1 - p_i) \quad (1)$$

*onde  $V$  é o conjunto de vértices do grafo  $G$  e  $V'$  é o conjunto formado pelos vértices do grafo  $G$  que funcionam. Quando os vértices do grafo  $G$  que modela a rede possuem a mesma probabilidade de funcionar  $p$ , a confiabilidade da rede é dada por:*

$$p_{R_G} = \sum_{i=2}^k S_i p^i (1-p)^{k-i} \quad (2)$$

*onde  $G$  é o grafo que modela a rede contendo  $k$  vértices e  $m$  arestas, e  $S_i$  é o número de subgrafos conexos de  $G$  contendo  $i$  vértices (GOLDSCHMIDT; JAILLET; LASOTA, 2006).*

*Quando os vértices do grafo que modela a rede possuem probabilidades de funcionar distintas  $p_i$ , a confiabilidade da rede  $p_{R_G}$  é calculada de forma similar a expressão (2), ou seja, obtidos os subgrafos conexos de  $G$  contendo  $i$  vértices, deve-se calcular a probabilidade de cada estado de funcionamento da rede e somar tais resultados.*

## **4 Material e métodos**

### **4.1 Coleta de dados e construção da rede de coautoria**

*Com o propósito de construir uma rede de coautoria científica formada por pesquisadores de um grupo de pesquisa da UNESP<sup>1</sup>, devidamente cadastrado no CNPq, considerou-se o grupo intitulado Centro de Pesquisa em Administração e Agronegócio (CEPEAGRO) da área de Ciências Sociais Aplicadas.*

*Se cada pesquisador é representado por um vértice e dois vértices possuem uma aresta entre si se, e somente se, os respectivos pesquisadores possuem pelo menos uma publicação em comum, então, a confiabilidade do grupo de pesquisa no tempo  $t$  (agosto de 2012), representado por um grafo não direcionado que modela a rede de coautoria entre os pesquisadores deste grupo, é a probabilidade do referido grupo permanecer em atividade no tempo  $t$ , mesmo se uma ou mais falhas (alteração no número de pesquisadores do grupo, ou seja, um*

<sup>1</sup> No Ranking Universitário Folha (RUF) de 2013, a UNESP apareceu em sexto lugar dentre as universidades públicas e privadas brasileiras avaliadas quanto às dimensões ensino, pesquisa, mercado, inovação e internacionalização. Nesse mesmo ano, o mais respeitado ranking internacional de universidades – Ranking BRICS e Economias Emergentes – publicado pela tradicional revista britânica *Times Higher Education* (THE), avaliou as 100 melhores universidades de países emergentes. Apenas quatro instituições brasileiras foram classificadas, das quais a UNESP ocupou o 87º lugar. A escolha foi baseada em treze indicadores, que vão de investimento à pesquisa, passando por publicações científicas, número de doutorados e de estudantes estrangeiros.

*ou mais vértices deixar(em) de pertencer à rede) acarretar na remoção de um ou mais subconjuntos de vértices deste grafo.*

*I Neste estudo foram adotados os seguintes procedimentos metodológicos:*

*Levantamento da produção científica – artigos em periódicos, livros e trabalhos completos em eventos – publicados e cadastrados pelos pesquisadores do grupo de pesquisa na Plataforma Lattes, desde o momento da inserção destes no grupo de pesquisa<sup>2</sup>. Dada a produção científica de cada pesquisador, o conjunto de dados necessário às análises propostas foi composto por:*

*a) número total de publicações de cada pesquisador (isto é, soma das publicações que cada pesquisador possui, sejam para o referido grupo de pesquisa ou para outros fins), desde o seu ingresso no grupo de pesquisa estudado;*

*b) número de publicações ou de colaborações de cada pesquisador para o referido grupo de pesquisa;*

*II Organização e sistematização dos dados coletados, permitindo representar e analisar por meio de um grafo<sup>3</sup> as características de colaboração entre os pesquisadores investigados (coautorias em publicações científicas);*

*III Cálculo de duas medidas de centralidade de vértices: 1) medida de proximidade; e, 2) medida de grau de informação. Tais medidas auxiliarão na identificação de situações onde a inserção de uma aresta (ligação entre dois pesquisadores deste grupo) pode aumentar significativamente a confiabilidade da rede.*

#### *4.1 Cálculo da confiabilidade da rede de coautoria*

*Como mencionado anteriormente, um grupo de pesquisa pode ser considerado como uma rede social, a qual pode ser modelada por um grafo simples não orientado  $G=(V,E)$  com  $k$  vértices (pesquisadores que compõem o grupo de pesquisa) e  $m$  arestas (ligações de coautoria).*

*Dado que as arestas são perfeitamente confiáveis e somente os vértices estão propensos a falhas, então, a confiabilidade da rede ou a probabilidade do referido grupo permanecer em atividade no tempo  $t$ , mesmo se uma ou mais falhas acarretar na remoção de um ou mais subconjuntos de vértices deste grafo, é dada por (vide Seção 3.3):*

*Vértices possuem a mesma probabilidade de funcionar  $p$ :*

$$p_{R_G} = \sum_{i=2}^k S_i p^i (1-p)^{k-i}$$

*Vértices possuem probabilidades de funcionar distintas  $p_i$ : a confiabilidade da rede  $p_{R_G}$  é calculada de forma similar a expressão anterior, ou seja, obtidos os subgrafos conexos de  $G$  contendo  $i$  vértices, deve-se calcular a probabilidade de cada estado de funcionamento da rede e somar tais resultados.*

<sup>2</sup> Como critério para a abordagem inferencial, foram considerados apenas os pesquisadores que tinham no mínimo trinta publicações no item I (a).

<sup>3</sup> Ressalta-se que o grafo foi gerado automaticamente por meio do arquivo *scriptLattes V7.02*, o qual foi desenvolvido no CCSL-IME/USP por Jesús P. Mena-Chalco e Roberto M. Cesar-Jr.

### 4.3 Inferência estatística

Assumindo uma rede de coautoria modelada por um grafo simples não orientado  $G=(V,E)$  com  $k$  vértices e  $m$  arestas, define-se  $Y_i$  como uma variável indicadora para o funcionamento do  $i$ -ésimo vértice (ou seja, do  $i$ -ésimo pesquisador),  $i=1,2,\dots,k$ , e  $Y$  como uma variável indicadora para o funcionamento da rede. Assim, tem-se:

$$Y_i = \begin{cases} 1 & , \text{ se o } i\text{-ésimo vértice está funcionando} \\ 0 & , \text{ se o } i\text{-ésimo vértice não está funcionando} \end{cases}$$

$$Y = \begin{cases} 1 & , \text{ se a rede está funcionando} \\ 0 & , \text{ se a rede não está funcionando} \end{cases}$$

Seja  $p_i = P\{Y_i = 1\}$ ,  $i=1,2,\dots,k$ , a confiabilidade do  $i$ -ésimo vértice e  $p_{R_G} = P\{Y = 1\}$  a confiabilidade da rede. Então, sob certas condições de independência condicional dos  $Y_i$ 's e também dos  $p_i$ 's,  $p_{R_G}$  depende de  $h(p_1, p_2, \dots, p_k)$ , onde  $h$  é alguma função das confiabilidades dos componentes individuais  $p_i$ 's e depende da estrutura da rede (série, paralelo ou qualquer outra configuração geral) (BARLOW; PROSCHAN, 1981).

O processo de estimação dos  $p_i$ 's e de  $p_{R_G}$  pode ser feito pelo método de máxima verossimilhança, que é uma técnica de estimação bastante usual na inferência estatística. O princípio de verossimilhança sustenta que, se o modelo estiver corretamente identificado, todas as informações que os dados fornecem sobre os parâmetros estão contidas na função de verossimilhança. Assim, o método consiste em selecionar os estimadores dos parâmetros do modelo que maximizam a probabilidade de se obter a amostra realmente observada.

Neste caso, a função de verossimilhança para  $\mathbf{p} = (p_1, p_2, \dots, p_k)'$  é dada por:

$$L(D|\mathbf{p}) = \prod_{i=1}^k \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i} \quad (3)$$

onde  $D = \{(n_i, x_i); i=1,2,\dots,k\}$  é um conjunto de dados onde  $n_i$  é o número total de publicações do pesquisador  $i$  e  $x_i$  é o número de publicações do pesquisador  $i$  para o grupo de pesquisa, em um certo tempo  $t$ , sendo  $i=1,2,\dots,k$  e  $k$  o número de vértices do grafo  $G$ , ou seja, de pesquisadores que compõem a rede (OLIVEIRA; ACHCAR, 2000).

Seja  $l(\mathbf{p})$  o logaritmo natural da função de verossimilhança  $L(D|\mathbf{p})$ . Então, tem-se:

$$l(\mathbf{p}) = \ln L(D|\mathbf{p}) = \sum_{i=1}^k [\ln n_i! - \ln x_i! - \ln(n_i - x_i)! + x_i \ln p_i + (n_i - x_i) \ln(1 - p_i)] \quad (4)$$

Derivando a expressão (4) em relação a  $p_i, i=1,2,\dots,k$ , obtêm-se as seguintes equações de verossimilhança:

$$\frac{\partial l(\mathbf{p})}{\partial p_i} = \frac{x_i}{p_i} - \frac{(n_i - x_i)}{(1 - p_i)} \quad (5)$$

Igualando as equações (6) a zero, obtêm-se os estimadores de máxima verossimilhança (EMV) de  $p_i, i = 1, 2, \dots, k$ , dados por:

$$\hat{p}_i = \frac{x_i}{n_i} \quad (6)$$

As distribuições assintóticas para os estimadores  $\hat{p}_i, i = 1, 2, \dots, k$  são expressas por  $\hat{p}_i \stackrel{a}{\sim} Normal(p_i, I_{ii}^{-1}(\mathbf{p})), i = 1, 2, \dots, k$ , onde  $I_{ii}^{-1}(\mathbf{p})$  é o  $i$ -ésimo termo da diagonal da inversa da matriz de informação de Fisher (CASELLA; BERGER, 2010).

Seja  $I_{ii}(\mathbf{p})$  a matriz de informação de Fisher dada por:

$$I_{ii}(\mathbf{p}) = \begin{pmatrix} -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_1^2}\right) & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_1 \partial p_2}\right) & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_1 \partial p_3}\right) & \dots & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_1 \partial p_k}\right) \\ -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_2 \partial p_1}\right) & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_2^2}\right) & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_2 \partial p_3}\right) & \dots & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_2 \partial p_k}\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_k \partial p_1}\right) & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_k \partial p_2}\right) & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_k \partial p_3}\right) & \dots & -E\left(\frac{\partial^2 l(\mathbf{p})}{\partial p_k^2}\right) \end{pmatrix} \quad (7)$$

onde  $l(\mathbf{p})$  é o logaritmo natural da função de verossimilhança  $L(D|\mathbf{p})$ , definida em (4), e  $\mathbf{p} = (p_1, p_2, \dots, p_k)'$ .

Derivando novamente as equações (5) em relação a  $p_i$ , obtêm-se as derivadas de segunda ordem que correspondem às diagonais da matriz expressa por (7), as quais são dadas por:

$$\frac{\partial^2 l(\mathbf{p})}{\partial p_i^2} = -\frac{x_i}{p_i^2} - \frac{(n_i - x_i)}{(1 - p_i)^2} \quad (8)$$

Observa-se ainda que as derivadas de segunda ordem (na referida matriz) correspondentes a  $\frac{\partial^2 l(\mathbf{p})}{\partial p_i \partial p_j}, i \neq j, i = 1, \dots, k; j = 1, \dots, k$ , são iguais a zero.

Assim, os intervalos de confiança  $100(1 - \alpha)\%$  para  $p_i, i = 1, 2, \dots, k$ , são dados por:

$$p_i : \left[ \hat{p}_i - z_{\alpha/2} \sqrt{\hat{I}_{ii}^{-1}(\mathbf{p})}; \hat{p}_i + z_{\alpha/2} \sqrt{\hat{I}_{ii}^{-1}(\mathbf{p})} \right] \quad (9)$$

onde  $1 - \alpha$  é o nível de confiança do intervalo, o ponto  $z_{\alpha/2}$  é determinado de uma distribuição normal padronizada e  $\hat{I}_{ii}^{-1}(\mathbf{p})$  é o  $i$ -ésimo termo da diagonal da inversa da matriz de informação de Fisher amostral (CASELLA; BERGER, 2010).

Além dos intervalos de confiança assintóticos dados pela expressão (9), podem ser obtidos intervalos de confiança exatos para  $p_i, i = 1, 2, \dots, k$ . Um intervalo de confiança exato para proporções baseado na relação

entre a distribuição binomial e a distribuição F (MORISSETTE; KHORRAM, 1998) e, neste caso, é dado por:

$$p_i : \left[ \frac{1}{1 + \frac{n_i - x_i + 1}{x_i} F_{\alpha/2; 2(n_i - x_i + 1); 2x_i}}; \frac{\frac{x_i + 1}{n_i - x_i} F_{\alpha/2; 2(x_i + 1); 2(n_i - x_i)}}{1 + \frac{x_i + 1}{n_i - x_i} F_{\alpha/2; 2(x_i + 1); 2(n_i - x_i)}} \right], i = 1, 2, \dots, k \quad (10)$$

onde  $(1-\alpha)$  é o nível de confiança do intervalo e o ponto  $F_{\alpha/2; v_1; v_2}$  é determinado de uma distribuição F com  $v_1$  e  $v_2$  graus de liberdade.

Como a confiabilidade da rede  $p_{R_G}$  é uma função da confiabilidade dos componentes individuais  $p_i, i = 1, 2, \dots, k$ , para obter o EMV de  $p_{R_G}$  utiliza-se a propriedade de invariância dos estimadores de máxima verossimilhança, ou seja, basta tomar os estimadores de máxima verossimilhança de  $p_i$ , expressos em (6), e substituir em  $p_{R_G}$ , obtendo  $\hat{p}_{R_G} = h(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$ .

Seja  $g^*(\mathbf{p}) = \left[ \frac{\partial p_{R_G}}{\partial p_1} \quad \frac{\partial p_{R_G}}{\partial p_2} \quad \frac{\partial p_{R_G}}{\partial p_3} \quad \dots \quad \frac{\partial p_{R_G}}{\partial p_k} \right]$  um vetor de ordem  $1 \times k$  e

seja  $\Sigma = I^{-1}(\mathbf{p})$  a inversa da matriz de informação de Fisher expressa por (7). Assim, pelo método Delta<sup>4</sup> (SEN; SINGER; PEDROSO-DE-LIMA, 2009), o intervalo de confiança  $100(1-\alpha)\%$  (assintótico) para  $p_{R_G}$  será dado por:

$$p_{R_G} : \left[ \hat{p}_{R_G} \pm z_{\alpha/2} \sqrt{\hat{g}^*(\mathbf{p}) \hat{\Sigma} \hat{g}^*(\mathbf{p})'} \right] \quad (11)$$

onde  $\hat{g}^*(\mathbf{p})$  e  $\hat{\Sigma}$  são matrizes amostrais de  $g^*(\mathbf{p})$  e de  $\Sigma$ , respectivamente.

#### 4.4 Medidas de centralidade de vértices

As medidas de centralidade são usadas na ARS para verificar a importância de um vértice em relação aos demais vértices de uma rede. Por meio das medidas de centralidade, tais vértices podem ser ordenados em função de sua importância relativa. Como o poder é uma característica que surge de relações, este pode ser associado às medidas de centralidade, mostrando a distribuição do poder dentro de uma rede e a influência dos atores (ou vértices) em dominar ou influenciar outros atores.

Diferentes medidas de centralidade servem para diferentes tipos de importância (posição, fluxo, influência, etc). Dentre as medidas existentes, foram utilizadas:

**Medida de proximidade:** Relaciona a distância total de um vértice aos demais vértices da rede, ou seja, indica a velocidade de acesso de um

<sup>4</sup> Se a convergência em distribuição de um parâmetro é conhecida, então, pelo método Delta, a convergência em distribuição de uma função desse parâmetro também é conhecida e, essa função deve satisfazer algumas condições, tais como, ser diferenciável e contínua.

*vértice a outro da rede, apontando os vértices que precisam de melhorias. O cálculo da medida de proximidade do vértice  $i$  ( $v_i$ ) é dado por:*

$$C_p(v_i) = \sum_{j=1}^k d_{v_i v_j} \quad (12)$$

*onde  $d_{v_i v_j}$  representa a menor distância entre o vértice  $i$  ( $v_i$ ) e o vértice  $j$  ( $v_j$ ) e  $k$  é o número de vértices da rede. O elemento mais central da rede é aquele com menor valor de  $C_p(v_i)$ , ou seja, o elemento que se comunica com maior agilidade aos demais elementos da rede, devido a sua posição estrutural.*

**Medida de grau de informação:** *Atribui relevância a um vértice em função do número de ligações diretas que este estabelece com os demais vértices da rede, ou seja, avalia a interferência direta (ou efeito imediato para um tempo  $t+1$ ) de um vértice no outro por meio do número de caminhos de comprimento unitário que tem origem em um vértice. O cálculo da medida de grau de informação do vértice  $i$  ( $v_i$ ) é dado por:*

$$C_g(v_i) = d(v_i), 1 \leq i \leq k \quad (13)$$

*onde  $k$  é o número de vértices da rede.*

## 5 Aplicação e resultados

*O grafo que modela a rede de coautoria científica considerada (CEPEAGRO) foi gerado automaticamente por meio do arquivo scriptLattes V7.02 conforme as características de colaboração entre os pesquisadores deste grupo (coautorias em publicações científicas). Especificamente, foram considerados artigos em periódicos, livros e trabalhos completos em eventos, cadastrados e publicados pelos pesquisadores do CEPEAGRO na Plataforma Lattes do momento da inclusão destes no referido grupo até agosto de 2012 (tempo  $t$ ). Ressalta-se que os pesquisadores que se desligaram do grupo (em algum momento desde a sua formação) foram desconsiderados do estudo. Obteve-se, então, a rede de coautoria científica modelada pelo grafo  $G$  simples não orientado e conexo da Figura 1, com  $m = 8$  arestas ou relações de coautoria e  $k = 7$  vértices ou pesquisadores (cada vértice  $i$  denotado por  $i=1,2,\dots,7$ ), sendo: 1) Oliveira; 2) Pereira; 3) Scalco; 4) Pigatto; 5) Gabriel Filho; 6) Queiroz e 7) Machado, respectivamente.*

*A partir do grafo  $G$  da Figura 1, é possível formar quarenta e três subgrafos conexos, ou seja, um subgrafo com sete vértices, cinco subgrafos com seis vértices, nove subgrafos com cinco vértices, onze subgrafos com quatro vértices, nove subgrafos com três vértices e oito subgrafos com dois vértices. Observa-se ainda que, dada a configuração do grafo  $G$ , não é possível formar subgrafos com apenas um vértice (ou pesquisador), uma vez que o limite mínimo estabelecido pelo CNPq para a formação de um grupo de pesquisa é de dois pesquisadores.*

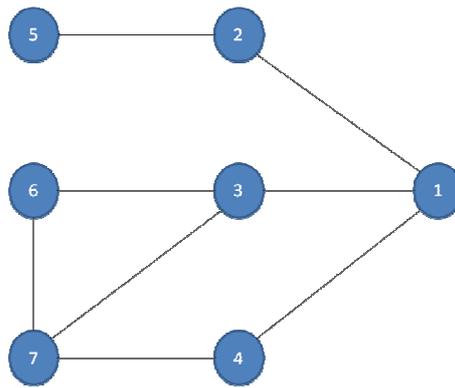


Figura 1 – Grafo G que modela a rede de coautoria científica.

Considere a situação (fictícia) onde todos os vértices do grafo G têm a mesma confiabilidade  $p_i$ ,  $i=1,2,\dots,7$  (denotada apenas por  $p$ ), ou seja, todos os pesquisadores contribuem igualmente para o grupo. Assim, a confiabilidade da referida rede é dada por  $p_{R_G} = \sum_{i=2}^7 S_i p^i (1-p)^{7-i}$ , onde  $S_i$  é número de subgrafos conexos de G contendo  $i$  vértices.

Como  $S_2 = 8$ ,  $S_3 = 9$ ,  $S_4 = 11$ ,  $S_5 = 9$ ,  $S_6 = 5$  e  $S_7 = 1$ , então, a confiabilidade desta rede é dada por:

$$p_{R_G} = \sum_{i=2}^7 S_i p^i (1-p)^{7-i} = S_2 p^2 (1-p)^5 + S_3 p^3 (1-p)^4 + S_4 p^4 (1-p)^3 + S_5 p^5 (1-p)^2 + S_6 p^6 (1-p) + S_7 p^7 (1-p)^0 = 8p^2(1-p)^5 + 9p^3(1-p)^4 + 11p^4(1-p)^3 + 9p^5(1-p)^2 + 5p^6(1-p) + p^7$$

A Tabela 1 a seguir mostra os resultados de simulações feitas para diferentes valores de  $p$ , sendo possível observar o comportamento da confiabilidade da rede de coautoria à medida que se aumenta a confiabilidade de cada vértice ou pesquisador, isto é, o valor de  $p$ . Portanto, dada a configuração deste grupo e as relações de coautoria existentes, verifica-se que a probabilidade de falha do vértice acima de 0,8 ( $p \leq 0,2$ ) faz com que a confiabilidade da rede (probabilidade do grupo permanecer em atividade no tempo  $t$ , ou seja, em agosto de 2012) fique mais próxima de zero.

Tabela 1 - Confiabilidade da rede de coautoria científica modelada pelo Grafo G (vértices com confiabilidades iguais), considerando diferentes valores para  $p$ .

Valores de $p$	Confiabilidade da rede $p_{R_G}$	Valores de $p$	Confiabilidade da rede $p_{R_G}$
0,9	0,805100	0,4	0,282112
0,8	0,634880	0,3	0,223403
0,7	0,500805	0,2	0,145472
0,6	0,403776	0,1	0,054024
0,5	0,335938		

## Inferência estatística

Considere que cada vértice ou pesquisador  $i$ ,  $i=1,2,\dots,7$ , do grafo  $G$  que modela a rede de coautoria possui uma confiabilidade denotada por  $p_i$ . Conforme os quarenta e três subgrafos conexos obtidos a partir de  $G$ , a expressão genérica da confiabilidade da rede é dada por:

$$\begin{aligned}
 p_{R_G} = & p_1 p_2 p_3 p_4 p_5 p_6 p_7 + p_1 p_2 (1 - p_3) p_4 p_5 p_6 p_7 + p_1 p_2 p_3 (1 - p_4) p_5 p_6 p_7 + \\
 & + p_1 p_2 p_3 p_4 (1 - p_5) p_6 p_7 + p_1 p_2 p_3 p_4 p_5 (1 - p_6) p_7 + p_1 p_2 p_3 p_4 p_5 p_6 (1 - p_7) + \\
 & + p_1 (1 - p_2) p_3 p_4 (1 - p_5) p_6 p_7 + p_1 p_2 (1 - p_3) p_4 (1 - p_5) p_6 p_7 + p_1 p_2 (1 - p_3) p_4 p_5 (1 - p_6) p_7 + \\
 & + p_1 p_2 p_3 (1 - p_4) (1 - p_5) p_6 p_7 + p_1 p_2 p_3 (1 - p_4) p_5 (1 - p_6) p_7 + p_1 p_2 p_3 (1 - p_4) p_5 p_6 (1 - p_7) + \\
 & + p_1 p_2 p_3 p_4 (1 - p_5) (1 - p_6) p_7 + p_1 p_2 p_3 p_4 (1 - p_5) p_6 (1 - p_7) + p_1 p_2 p_3 p_4 p_5 (1 - p_6) (1 - p_7) + \\
 & + (1 - p_1) (1 - p_2) p_3 p_4 (1 - p_5) p_6 p_7 + p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_5) p_6 p_7 + \\
 & + p_1 (1 - p_2) p_3 (1 - p_4) (1 - p_5) p_6 p_7 + p_1 (1 - p_2) p_3 p_4 (1 - p_5) (1 - p_6) p_7 + \\
 & + p_1 (1 - p_2) p_3 p_4 (1 - p_5) p_6 (1 - p_7) + p_1 p_2 (1 - p_3) p_4 (1 - p_5) (1 - p_6) p_7 + \\
 & + p_1 p_2 (1 - p_3) p_4 p_5 (1 - p_6) (1 - p_7) + p_1 p_2 p_3 (1 - p_4) (1 - p_5) (1 - p_6) p_7 + \\
 & + p_1 p_2 p_3 (1 - p_4) (1 - p_5) p_6 (1 - p_7) + p_1 p_2 p_3 (1 - p_4) p_5 (1 - p_6) (1 - p_7) + \\
 & + p_1 p_2 p_3 p_4 (1 - p_5) (1 - p_6) (1 - p_7) + (1 - p_1) (1 - p_2) (1 - p_3) p_4 (1 - p_5) p_6 p_7 + \\
 & + (1 - p_1) (1 - p_2) p_3 (1 - p_4) (1 - p_5) p_6 p_7 + (1 - p_1) (1 - p_2) p_3 p_4 (1 - p_5) (1 - p_6) p_7 + \\
 & + p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_5) (1 - p_6) p_7 + p_1 (1 - p_2) p_3 (1 - p_4) (1 - p_5) (1 - p_6) p_7 + \\
 & + p_1 (1 - p_2) p_3 (1 - p_4) (1 - p_5) p_6 (1 - p_7) + p_1 (1 - p_2) p_3 p_4 (1 - p_5) (1 - p_6) (1 - p_7) + \\
 & + p_1 p_2 (1 - p_3) (1 - p_4) p_5 (1 - p_6) (1 - p_7) + p_1 p_2 (1 - p_3) p_4 (1 - p_5) (1 - p_6) (1 - p_7) + \\
 & + (1 - p_1) (1 - p_2) (1 - p_3) (1 - p_4) (1 - p_5) p_6 p_7 + (1 - p_1) (1 - p_2) (1 - p_3) p_4 (1 - p_5) (1 - p_6) p_7 + \\
 & + (1 - p_1) (1 - p_2) p_3 (1 - p_4) (1 - p_5) (1 - p_6) p_7 + (1 - p_1) (1 - p_2) p_3 (1 - p_4) (1 - p_5) p_6 (1 - p_7) + \\
 & + (1 - p_1) p_2 (1 - p_3) (1 - p_4) p_5 (1 - p_6) (1 - p_7) + p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_5) (1 - p_6) (1 - p_7) + \\
 & + p_1 (1 - p_2) p_3 (1 - p_4) (1 - p_5) (1 - p_6) (1 - p_7) + p_1 p_2 (1 - p_3) (1 - p_4) (1 - p_5) (1 - p_6) (1 - p_7)
 \end{aligned}$$

(14)

O processo de estimação da confiabilidade de cada vértice ou pesquisador  $i$  ( $p_i, i=1,2,\dots,7$ ) e da confiabilidade da rede de coautoria  $p_{R_G}$  (probabilidade do grupo de pesquisa permanecer em atividade em agosto de 2012) foi feito pelo método de máxima verossimilhança. Assim, conforme as informações de publicações científicas (artigos em periódicos, livros e trabalhos completos em eventos) dos pesquisadores do grupo CEPEAGRO obtidas por meio do arquivo scriptLattes V7.02, o conjunto de dados  $D = \{(n_i, x_i); i=1,2,\dots,8\}$  para o processo de estimação é dado por:

Tabela 2 – Conjunto de dados  $D = \{(n_i, x_i); i=1,2,\dots,7\}$  referente aos pesquisadores do grupo de pesquisa CEPEAGRO.

Vértice ou pesquisador $i$	$n_i$	$x_i$
1	30	20
2	73	23
3	32	25

4	39	16
5	88	14
6	37	18
7	57	27

onde  $n_i$  é o número total de publicações do pesquisador  $i$  (seja para o referido grupo de pesquisa ou para outros fins) e  $x_i$  é o número de colaborações do pesquisador  $i$  apenas para o grupo de pesquisa considerado,  $i=1,2,\dots,k$ , sendo que  $k = 7$  é o número de vértices (pesquisadores) do grafo  $G$  da Figura 1. Assim, a função de verossimilhança para  $\mathbf{p} = (p_1, p_2, \dots, p_7)'$  é dada por:

$$L(D|\mathbf{p}) = \prod_{i=1}^7 \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i-x_i}$$

Os estimadores de máxima verossimilhança (EMV) de  $p_i$  são dados por  $\hat{p}_i = \frac{x_i}{n_i}$ ,  $i=1,2,\dots,7$ , e os respectivos intervalos de confiança  $100(1-\alpha)\%$  (assintóticos) são expressos por  $p_i : \left[ \hat{p}_i \pm z_{\alpha/2} \sqrt{\hat{I}_{ii}^{-1}(\mathbf{p})} \right]$ ,  $i=1,2,\dots,7$ . Para um nível de confiança de 95% tem-se que  $z_{\alpha/2} = 1,96$  e  $\hat{I}_{ii}^{-1}(\mathbf{p})$  é o  $i$ -ésimo termo da diagonal da inversa da matriz de informação de Fisher amostral, da expressão (7), dada por:

$$\hat{I}^{-1}(\mathbf{p}) = \begin{pmatrix} 135,00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 338,28 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 187,25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 161,19 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 657,79 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 148,11 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 228,63 \end{pmatrix}^{-1} \tag{15}$$

Tem-se ainda que os intervalos de confiança  $100(1-\alpha)\%$  (exatos) são expressos por:

$$p_i : \left[ \frac{1}{1 + \frac{n_i - x_i + 1}{x_i} F_{\alpha/2; 2(n_i - x_i + 1); 2x_i}}; \frac{\frac{x_i + 1}{n_i - x_i} F_{\alpha/2; 2(x_i + 1); 2(n_i - x_i)}}{1 + \frac{x_i + 1}{n_i - x_i} F_{\alpha/2; 2(x_i + 1); 2(n_i - x_i)}} \right]$$

onde os pontos  $F_{\alpha/2; v_1; v_2}$  e  $F_{1-\alpha/2; v_1; v_2}$  são determinados de uma distribuição F com  $v_1$  e  $v_2$  graus de liberdade, respectivamente. Assim, tem-se:

Tabela 3 – Cálculos preliminares para a determinação dos intervalos de confiança de 95% (exatos) para as confiabilidades  $p_i, i=1,2,\dots,7$ .

Vértice $i$	$n_i$	$x_i$	$n_i - x_i + 1/x_i$	$F_{\alpha/2; 2(n_i - x_i + 1); 2x_i}$	$x_i + 1/n_i - x_i$	$F_{\alpha/2; 2(x_i + 1); 2(n_i - x_i)}$
-------------	-------	-------	---------------------	--	---------------------	--

1	30	20	1,550	$F_{0,025;22;40} = 2,03488$	2,100	$F_{0,025;42;20} = 2,27836$
2	73	23	3,217	$F_{0,025;102;46} = 1,68338$	0,480	$F_{0,025;48;100} = 1,59998$
3	32	25	1,320	$F_{0,025;16;50} = 2,08098$	3,714	$F_{0,025;52;14} = 2,63285$
4	39	16	2,500	$F_{0,025;48;32} = 1,94088$	0,739	$F_{0,025;34;46} = 1,86071$
5	88	14	6,357	$F_{0,025;150;28} = 1,89210$	0,203	$F_{0,025;30;148} = 1,66643$
6	37	18	2,111	$F_{0,025;40;36} = 1,91942$	1,000	$F_{0,025;38;38} = 1,90700$
7	57	27	2,148	$F_{0,025;62;54} = 1,69185$	0,933	$F_{0,025;56;60} = 1,67828$

Portanto, de acordo com os cálculos da expressão (15) e da Tabela 3, as EMV e os respectivos intervalos de confiança de 95% para  $p_i, i = 1, 2, \dots, 7$ , são dados na Tabela 4 a seguir.

Tabela 4 – Estimativas de máxima verossimilhança (EMV) e intervalos de confiança (IC) de 95% para as confiabilidades  $p_i, i = 1, 2, \dots, 7$ .

	EMV	IC assintótico (95%)	IC exato (95%)
$p_1$	0,6667	[0,4980 0,8354]	[0,4719 0,8271]
$p_2$	0,3151	[0,2085 0,4216]	[0,2113 0,4344]
$p_3$	0,7813	[0,6380 0,9245]	[0,6003 0,9072]
$p_4$	0,4103	[0,2559 0,5646]	[0,2557 0,5790]
$p_5$	0,1591	[0,0827 0,2355]	[0,0898 0,2525]
$p_6$	0,4865	[0,3254 0,6475]	[0,3192 0,6560]
$p_7$	0,4737	[0,3441 0,6033]	[0,3398 0,6103]

Como  $n_i, i = 1, 2, \dots, 7$ , é maior ou igual a 30, a abordagem assintótica na estimação intervalar pode ser considerada como um procedimento apropriado. Na Tabela 4 pode-se observar que os intervalos assintóticos de 95% de confiança são mais robustos do que os intervalos exatos, uma vez que a diferença entre o limite inferior e o limite superior é menor para todo  $p_i, i = 1, 2, \dots, 7$ .

Uma vez que a confiabilidade da rede  $p_{R_G}$  é uma função da confiabilidade dos componentes individuais  $p_i, i = 1, 2, \dots, 7$ , para obter a EMV de  $\hat{p}_{R_G}$  deve-se tomar a EMV de cada  $p_i, i = 1, 2, \dots, 7$  e substituir em  $p_{R_G}$ . O intervalo de confiança  $100(1-\alpha)\%$  (assintótico) para  $p_{R_G}$  é dado por

$$p_{R_G} : \left[ \hat{p}_{R_G} \pm z_{\alpha/2} \sqrt{\hat{g}^*(\mathbf{p}) \hat{\Sigma} \hat{g}^*(\mathbf{p})'} \right], \text{ onde } \mathbf{g}^*(\mathbf{p}) = \begin{bmatrix} \frac{\partial p_{R_G}}{\partial p_1} & \frac{\partial p_{R_G}}{\partial p_2} & \frac{\partial p_{R_G}}{\partial p_3} & \dots & \frac{\partial p_{R_G}}{\partial p_7} \end{bmatrix} \text{ e}$$

$\Sigma = I^{-1}(\mathbf{p})$ . Tem-se que:

$$\begin{aligned} \frac{\partial p_{R_G}}{\partial p_1} &= p_2 p_3 p_4 p_5 p_6 p_7 + p_2(1-p_3)p_4 p_5 p_6 p_7 + p_2 p_3(1-p_4)p_5 p_6 p_7 + p_2 p_3 p_4(1-p_5)p_6 p_7 + \\ &+ p_2 p_3 p_4 p_5(1-p_6)p_7 + p_2 p_3 p_4 p_5 p_6(1-p_7) + p_2(1-p_3)p_4(1-p_5)p_6 p_7 + p_2(1-p_3)p_4 p_5(1-p_6)p_7 + \\ &+ p_2 p_3(1-p_4)(1-p_5)p_6 p_7 + p_2 p_3(1-p_4)p_5(1-p_6)p_7 + p_2 p_3(1-p_4)p_5 p_6(1-p_7) + \\ &+ p_2 p_3 p_4(1-p_5)(1-p_6)p_7 + p_2 p_3 p_4(1-p_5)p_6(1-p_7) + p_2 p_3 p_4 p_5(1-p_6)(1-p_7) + \\ &+ (1-p_2)p_3 p_4(1-p_5)p_6(1-p_7) + p_2(1-p_3)p_4(1-p_5)(1-p_6)p_7 + p_2(1-p_3)p_4 p_5(1-p_6)(1-p_7) + \\ &+ p_2 p_3(1-p_4)(1-p_5)(1-p_6)p_7 + p_2 p_3(1-p_4)(1-p_5)p_6(1-p_7) + p_2 p_3(1-p_4)p_5(1-p_6)(1-p_7) + \\ &+ p_2 p_3 p_4(1-p_5)(1-p_6)(1-p_7) + (1-p_2)p_3 p_4(1-p_5)(1-p_6)(1-p_7) + \\ &+ p_2(1-p_3)p_4(1-p_5)(1-p_6)(1-p_7) - (1-p_2)(1-p_3)(1-p_4)(1-p_5)p_6 p_7 + \\ &+ (1-p_2)(1-p_3)p_4(1-p_5)(1-p_6)(1-p_7) + (1-p_2)p_3(1-p_4)(1-p_5)(1-p_6)(1-p_7) + \\ &+ p_2(1-p_3)(1-p_4)(1-p_5)(1-p_6)(1-p_7) \end{aligned}$$

$$\begin{aligned} \frac{\partial p_{R_G}}{\partial p_2} &= p_1 p_3 p_4 p_5 p_6 p_7 + p_1(1-p_3)p_4 p_5 p_6 p_7 + p_1 p_3(1-p_4)p_5 p_6 p_7 + \\ &+ p_1 p_3 p_4 p_5(1-p_6)p_7 + p_1 p_3 p_4 p_5 p_6(1-p_7) + p_1(1-p_3)p_4 p_5(1-p_6)p_7 + \\ &+ p_1 p_3(1-p_4)p_5(1-p_6)p_7 + p_1 p_3(1-p_4)p_5 p_6(1-p_7) + p_1 p_3 p_4 p_5(1-p_6)(1-p_7) - \\ &- (1-p_1)p_3 p_4(1-p_5)p_6 p_7 + p_1(1-p_3)p_4 p_5(1-p_6)(1-p_7) + \\ &+ p_1 p_3(1-p_4)p_5(1-p_6)(1-p_7) - (1-p_1)(1-p_3)p_4(1-p_5)p_6 p_7 - \\ &- (1-p_1)p_3(1-p_4)(1-p_5)p_6 p_7 - (1-p_1)p_3 p_4(1-p_5)(1-p_6)p_7 + \\ &+ p_1(1-p_3)(1-p_4)p_5(1-p_6)(1-p_7) - (1-p_1)(1-p_3)(1-p_4)(1-p_5)p_6 p_7 - \\ &- (1-p_1)(1-p_3)p_4(1-p_5)(1-p_6)p_7 - (1-p_1)p_3(1-p_4)(1-p_5)(1-p_6)p_7 - \\ &- (1-p_1)p_3(1-p_4)(1-p_5)p_6(1-p_7) + (1-p_1)(1-p_3)(1-p_4)p_5(1-p_6)(1-p_7) - \\ &- p_1 p_3(1-p_4)(1-p_5)(1-p_6)(1-p_7) + p_1(1-p_3)(1-p_4)(1-p_5)(1-p_6)(1-p_7) \end{aligned}$$

$$\begin{aligned} \frac{\partial p_{R_G}}{\partial p_3} &= p_1 p_2(1-p_4)p_5 p_6 p_7 + p_1 p_2 p_4 p_5 p_6(1-p_7) + p_1 p_2(1-p_4)(1-p_5)p_6 p_7 + \\ &+ p_1 p_2(1-p_4)p_5(1-p_6)p_7 + p_1 p_2(1-p_4)p_5 p_6(1-p_7) + p_1 p_2 p_4(1-p_5)p_6(1-p_7) + \\ &+ p_1(1-p_2)(1-p_4)(1-p_5)p_6 p_7 + p_1(1-p_2)p_4(1-p_5)p_6(1-p_7) + \\ &+ p_1 p_2(1-p_4)(1-p_5)(1-p_6)p_7 + p_1 p_2(1-p_4)(1-p_5)p_6(1-p_7) + \\ &+ p_1(1-p_2)(1-p_4)(1-p_5)(1-p_6)p_7 + p_1(1-p_2)(1-p_4)(1-p_5)p_6(1-p_7) + \\ &+ (1-p_1)(1-p_2)(1-p_4)(1-p_5)(1-p_6)p_7 + (1-p_1)(1-p_2)(1-p_4)(1-p_5)p_6(1-p_7) - \\ &- (1-p_1)p_2(1-p_4)p_5(1-p_6)(1-p_7) + p_1(1-p_2)(1-p_4)(1-p_5)(1-p_6)(1-p_7) - \\ &- p_1 p_2(1-p_4)(1-p_5)(1-p_6)(1-p_7) \end{aligned}$$

$$\begin{aligned} \frac{\partial p_{R_G}}{\partial p_4} &= p_1 p_2 p_3 p_5 p_6 p_7 + p_1 p_2(1-p_3)p_5 p_6 p_7 - p_1 p_2 p_3 p_5 p_6 p_7 + p_1 p_2(1-p_3)(1-p_5)p_6 p_7 + \\ &+ p_1 p_2(1-p_3)p_5(1-p_6)p_7 + p_1(1-p_2)(1-p_3)(1-p_5)p_6 p_7 + p_1 p_2(1-p_3)(1-p_5)(1-p_6)p_7 + \\ &+ p_1 p_2 p_3(1-p_5)(1-p_6)(1-p_7) + p_1(1-p_2)(1-p_3)(1-p_5)(1-p_6)p_7 + \\ &+ (1-p_1)(1-p_2)(1-p_3)(1-p_5)(1-p_6)p_7 - (1-p_1)(1-p_2)p_3(1-p_5)p_6(1-p_7) - \\ &- (1-p_1)p_2(1-p_3)p_5(1-p_6)(1-p_7) + p_1(1-p_2)(1-p_3)(1-p_5)(1-p_6)(1-p_7) \end{aligned}$$

$$\begin{aligned} \frac{\partial p_{R_G}}{\partial p_5} &= p_1 p_2 p_3 p_4 p_6 p_7 + p_1 (1 - p_2) p_3 p_4 p_6 p_7 - (1 - p_1) (1 - p_2) p_3 p_4 p_6 p_7 - \\ &- p_1 (1 - p_2) (1 - p_3) p_4 p_6 p_7 - p_1 (1 - p_2) p_3 (1 - p_4) p_6 p_7 - p_1 (1 - p_2) p_3 p_4 (1 - p_6) p_7 - \\ &- p_1 (1 - p_2) p_3 p_4 p_6 (1 - p_7) + p_1 p_2 p_3 (1 - p_4) (1 - p_6) (1 - p_7) - (1 - p_1) (1 - p_2) (1 - p_3) p_4 p_6 p_7 - \\ &- (1 - p_1) (1 - p_2) p_3 (1 - p_4) p_6 p_7 - (1 - p_1) (1 - p_2) p_3 p_4 (1 - p_6) p_7 - \\ &- p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_6) p_7 - p_1 (1 - p_2) p_3 (1 - p_4) (1 - p_6) p_7 - \\ &- p_1 (1 - p_2) p_3 (1 - p_4) p_6 (1 - p_7) - p_1 (1 - p_2) p_3 p_4 (1 - p_6) (1 - p_7) - \\ &- (1 - p_1) (1 - p_2) (1 - p_3) (1 - p_4) p_6 p_7 - (1 - p_1) (1 - p_2) (1 - p_3) p_4 (1 - p_6) p_7 - \\ &- (1 - p_1) (1 - p_2) p_3 (1 - p_4) (1 - p_6) p_7 - (1 - p_1) (1 - p_2) p_3 (1 - p_4) p_6 (1 - p_7) + \\ &+ (1 - p_1) p_2 (1 - p_3) (1 - p_4) (1 - p_6) (1 - p_7) - p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_6) (1 - p_7) - \\ &- p_1 (1 - p_2) p_3 (1 - p_4) (1 - p_6) (1 - p_7) \\ \frac{\partial p_{R_G}}{\partial p_6} &= -p_1 p_2 (1 - p_3) p_4 p_5 (1 - p_7) + p_1 p_2 p_3 (1 - p_4) (1 - p_5) (1 - p_7) - \\ &- p_1 p_2 (1 - p_3) (1 - p_4) p_5 (1 - p_7) - p_1 p_2 (1 - p_3) p_4 (1 - p_5) (1 - p_7) + \\ &+ (1 - p_1) (1 - p_2) (1 - p_3) (1 - p_4) (1 - p_5) p_7 - (1 - p_1) (1 - p_2) p_3 (1 - p_4) (1 - p_5) (1 - p_7) - \\ &- (1 - p_1) p_2 (1 - p_3) (1 - p_4) p_5 (1 - p_7) - p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_5) (1 - p_7) - \\ &- p_1 p_2 (1 - p_3) (1 - p_4) (1 - p_5) (1 - p_7) \\ \frac{\partial p_{R_G}}{\partial p_7} &= p_1 p_2 (1 - p_3) p_4 p_5 p_6 + p_1 p_2 (1 - p_3) p_4 (1 - p_5) p_6 + (1 - p_1) (1 - p_2) p_3 p_4 (1 - p_5) p_6 + \\ &+ p_1 (1 - p_2) (1 - p_3) p_4 (1 - p_5) p_6 + p_1 p_2 p_3 (1 - p_4) (1 - p_5) (1 - p_6) + \\ &+ (1 - p_1) (1 - p_2) (1 - p_3) p_4 (1 - p_5) p_6 + (1 - p_1) (1 - p_2) p_3 p_4 (1 - p_5) (1 - p_6) - \\ &- p_1 p_2 (1 - p_3) (1 - p_4) p_5 (1 - p_6) + (1 - p_1) (1 - p_2) (1 - p_3) (1 - p_4) (1 - p_5) p_6 + \\ &+ (1 - p_1) (1 - p_2) (1 - p_3) p_4 (1 - p_5) (1 - p_6) + (1 - p_1) (1 - p_2) p_3 (1 - p_4) (1 - p_5) (1 - p_6) - \\ &- (1 - p_1) p_2 (1 - p_3) (1 - p_4) p_5 (1 - p_6) - p_1 p_2 (1 - p_3) (1 - p_4) (1 - p_5) (1 - p_6) \end{aligned}$$

Substituindo a EMV de cada  $p_i, i=1,2,\dots,7$ , nas expressões das derivadas  $\frac{\partial p_{R_G}}{\partial p_1} \frac{\partial p_{R_G}}{\partial p_2} \frac{\partial p_{R_G}}{\partial p_3} \dots \frac{\partial p_{R_G}}{\partial p_7}$ , obtém-se  $\hat{g}^*(\mathbf{p})$ . Fazendo  $\hat{g}^*(\mathbf{p}) \hat{\Sigma} \hat{g}^*(\mathbf{p})'$ , onde  $\hat{\Sigma} = \hat{I}^{-1}(\mathbf{p})$  da expressão (15), extraíndo a sua raiz quadrada e multiplicando o resultado desta por  $z_{\alpha/2} = 1,96$ , obtém-se o erro de estimativa para o intervalo de confiança  $100(1-\alpha)\%$  para  $p_{R_G}$ . A estimativa de máxima verossimilhança de  $p_{R_G}$  e o respectivo intervalo de confiança de 95% encontram-se na Tabela 5 a seguir.

Tabela 5 - Estimativa de máxima verossimilhança (EMV) e intervalo de confiança (IC) de 95% para a confiabilidade da rede de coautoria científica  $p_{R_G}$ .

	EMV	IC assintótico (95%)
$p_{R_G}$	0,5972	[0,4583 0,7268]

Observa-se que, de acordo com a configuração do grupo de pesquisa, número de pesquisadores e de relações de coautoria existentes, conjuntamente com os dados de produção científica disponibilizados, a confiabilidade da rede com enfoque nos vértices ou nos pesquisadores é moderada, pois as EMV das confiabilidades dos pesquisadores têm valor mínimo igual a 15,91% e valor máximo de 78,13% (Tabela 4). Observa-se ainda que os pesquisadores "3" e "1" são os mais confiáveis, ou seja, aqueles que têm as maiores contribuições relativas para o grupo de pesquisa, respectivamente. Por outro lado, o pesquisador "5" é o menos confiável nesse sentido.

Portanto, pode-se ressaltar que a contribuição de cada pesquisador é de extrema importância para a manutenção do grupo de pesquisa, pois à medida que a colaboração deste pesquisador para o grupo aumenta, a sua confiabilidade também aumenta e, conseqüentemente, verifica-se um acréscimo na confiabilidade da rede de coautoria.

#### Medidas de Centralidade

Como mencionado anteriormente, as medidas de centralidade auxiliam tanto na mensuração da importância relativa de um vértice em relação aos demais, como na identificação da inserção de novas arestas de forma a aumentar a confiabilidade de uma rede. Dentre todas as opções possíveis de inserções de uma nova aresta na rede de coautoria modelada pelo grafo G da Figura 1, podem ser gerados treze grafos não isomorfos a partir de G (Tabela 6).

Tabela 6 - Grafos não isomorfos gerados a partir de G (Figura 1) com a inserção de uma aresta.

Grafo	Ligação	Grafo	Ligação	Grafo	Ligação
G1	Pesquisadores 1 e 5	G6	Pesquisadores 2 e 6	G11	Pesquisadores 4 e 6
G2	Pesquisadores 1 e 6	G7	Pesquisadores 2 e 7	G12	Pesquisadores 5 e 6
G3	Pesquisadores 1 e 7	G8	Pesquisadores 3 e 4	G13	Pesquisadores 5 e 7
G4	Pesquisadores 2 e 3	G9	Pesquisadores 3 e 5		
G5	Pesquisadores 2 e 4	G10	Pesquisadores 4 e 5		

Assim, foram calculadas as medidas de centralidade de proximidade e de grau de informação para os vértices do Grafo G (Figura 1), as quais se encontram na Tabela 7.

Segundo as medidas obtidas, os vértices ou pesquisadores mais centrais do grafo G são "1" e "3", respectivamente, ou seja, são os pesquisadores com maiores velocidades de acesso e maiores efeitos de influência nos demais. Além dos vértices "1" e "3" serem os que apresentam mais relações na rede, o processo de inferência estatística mostra ainda que estes são os mais confiáveis. Portanto, se tais vértices ou pesquisadores fossem retirados do grafo (por alguma razão), a rede de coautoria científica ficaria menos conectada e, conseqüentemente, com a confiabilidade sensivelmente diminuída, visto que alguns caminhos deixariam de existir.

Tabela 7 - Medidas de centralidade dos vértices do grafo G da Figura 1.

Vértices ou pesquisadores	Medida de proximidade	Medida de grau de informação
---------------------------	-----------------------	------------------------------

1	9	3
2	12	2
3	10	3
4	11	2
5	17	1
6	13	2
7	12	3

Observa-se que os vértices menos centrais do grafo  $G$  são os vértices "5" e "6", respectivamente. Levando em conta testes realizados por alguns autores (LYRA; OLIVEIRA, 2011; FERREIRA; OLIVEIRA, 2013), se o intuito é tornar o grupo de pesquisa mais confiável com a inserção de uma nova aresta ou relação de coautoria, as medidas de centralidade apontam que a ligação entre o pesquisador "5" e o pesquisador "6" poderia trazer este benefício.

Fixando uma aresta entre os vértices "5" e "6", considerando os vértices com confiabilidades  $p_i, i=1,2,\dots,7$  iguais às suas respectivas estimativas de máxima verossimilhança da Tabela 4 (Figura 2), e recalculando a expressão (14), observa-se que, de fato, a confiabilidade da rede aumenta cerca de 8,5% ( $p_{R_G} = 0,6477$ ) em relação à confiabilidade da rede sem a inserção da referida aresta ( $p_{R_G} = 0,5972$ ).

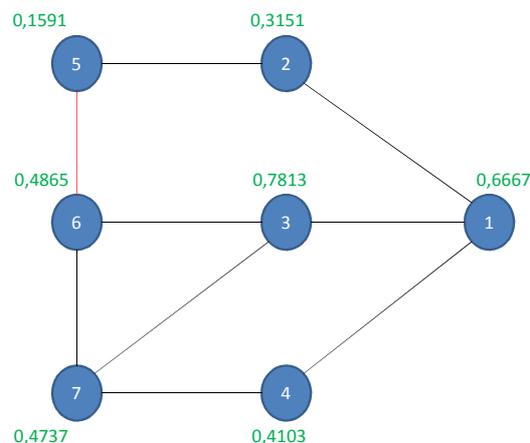


Figura 2 – Grafo  $G_{12}$  (inserção de uma aresta entre os vértices "5" e "6") com as confiabilidades de vértices estimadas  $\hat{p}_i, i=1,2,\dots,7$ .

## 6 Considerações finais

O estudo da confiabilidade de redes de coautoria científica permite identificar quais redes são confiáveis, sob diferentes enfoques (arestas e/ou vértices), de acordo com a participação dos pesquisadores e a intensidade das relações de coautoria existentes.

Neste trabalho foi proposta uma abordagem inferencial clássica para a confiabilidade de uma rede de coautoria com enfoque especificamente nos vértices ou atores (pesquisadores), ou seja, considerando os laços ou arestas (relações de coautoria) perfeitamente confiáveis. Os resultados mostraram a utilidade da inferência estatística no contexto de

*confiabilidade de redes de coautoria, ressaltando que a contribuição de cada pesquisador é de extrema importância para a manutenção de um grupo de pesquisa.*

*Além disso, foram obtidas medidas de centralidade de vértices que auxiliaram na identificação da situação onde a inserção de uma aresta entre dois pesquisadores proporcionou um aumento significativo da confiabilidade desta rede ou do grupo de pesquisa permanecer em atividade em um dado tempo t.*

*Observou-se ainda que o cálculo da confiabilidade de uma rede de coautoria pode ser bastante desgastante de ser executado, quer seja manualmente ou computacionalmente, e o uso de medidas de centralidade pode ser considerada uma alternativa viável quando se intenciona aumentar a confiabilidade desta rede. No entanto, alguns estudos têm mostrado que tais medidas se mostram como uma alternativa auxiliar, mas não inteiramente confiável na análise do aumento da confiabilidade de uma rede (LYRA; OLIVEIRA, 2011; FERREIRA; OLIVEIRA, 2013). Portanto, adicionalmente ao uso destas medidas, recomenda-se a utilização de outras medidas de centralidade, bem como a realização de simulações de forma a obter resultados ainda mais fidedignos.*

## **Agradecimentos**

*Ao CNPq pelo auxílio financeiro (Edital CNPq 18/2012; Processo 406626/2012-0), ao PIBIC/CNPq e a FAPESP pelas bolsas de IC (Processos 22862/2012, 2012/04444-8 e 2012/01690-8, respectivamente).*

## **Referências**

ABBASI, A.; ALTMANN, J.; HOSSAIN, L. Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network measures. *Journal of Informetrics*, v. 5, p. 594-607, 2011.

ABBASI, A.; ALTMANN, J.; HWANG, J. Evaluating scholars based on their academic collaboration activities: two indices, the RC-index and the CC-index, for quantifying collaboration activities of researchers and scientific communities. *Scientometrics*, v. 83, p. 01-13, 2010.

ABBASI, A.; WIGAND, R. T.; HOSSAIN, L. Measuring social capital through network analysis and its influence on individual performance. *Library & Information Science Research*, v. 36, p. 66-73, 2014.

BARBASTEFANO, R. G. *et al.* Impactos dos nomes nas propriedades de redes sociais: Um estudo em rede de coautoria sobre sustentabilidade. *Perspectivas em Ciência da Informação*, v. 18, n. 3, p. 78-95, 2013.

BARLOW, R. E.; PROSCHAN, F. *Statistical theory of reliability and life testing*. New York: Holt, Rinehart and Winston, 1981.

BOAVENTURA NETTO, P. O.; JURKIEWICZ, S. *Grafos: Introdução e Prática*. São Paulo: Edgard Blucher, 2009.

BRIGANTINI, B. B.; OLIVEIRA, S. C.; BRAGA JUNIOR; S. S. Classical statistical inference for the reliability of a co-authorship network with emphasis on edges. *American International Journal of Contemporary Research*, v. 4, p. 19-31, 2014.

CASELLA, G.; BERGER, R. L. *Inferência Estatística*. São Paulo: Cengage Learning, 2010.

CORDEIRO, G. M. *et al.* Modelagem probabilística de setores de um sistema elétrico. *Pesquisa Operacional*, v. 30, p. 563-581, 2010.

CRISTINO, C. T.; CORDEIRO, G. M.; LINS, S. Confiabilidade e avaliação de risco no sistema elétrico. *Revista de Matemática e Estatística*, v. 25, p. 87-109, 2007.

FERREIRA, T. P.; OLIVEIRA, S. C. Análise da confiabilidade de uma rede social fictícia com enfoque nos vértices e do uso de medidas de centralidade. In: VII CONGRESSO DE INICIAÇÃO CIENTÍFICA DAS FACULDADES ADAMANTINENSES INTEGRADAS, 2013, Adamantina. *Anais...*, novembro de 2013.

GOLDENBERG, J. *et al.* The evolving social network of marketing scholars. *Marketing Science*, v. 29, p. 561-567, 2010.

GOLDSCHMIDT, O.; JAILLET, P.; LASOTA, R. On reliability of graphs with node failures. *Networks*, v. 24, p. 251-259, 2006.

HAYASHI, M. C. P. I.; HAYASHI, C. R. M.; LIMA, M. Y. Análise de redes de coautoria na produção científica em educação especial. *Liinc em Revista*, v. 4, n. 1, p. 84-103, 2008.

LEE, S.; BOZEMAN, B. The impact of research collaboration on scientific productivity. *Social Studies of Science*, v. 35, n. 5, p. 673-702, 2005.

LYRA, T. F.; OLIVEIRA, C. S. Um estudo sobre confiabilidade de redes e medidas de centralidade em uma rede de co-autoria. *Revista Eletronica Pesquisa Operacional para o Desenvolvimento*, v. 3, n. 2, p. 160-172, 2011.

MAIA, M. F. S.; CAREGNATO, S. E. Coautoria como indicador de redes de colaboração científica. *Perspectivas em Ciência da Informação*, v. 13, n. 2, p. 18-31, 2008.

MARAFON, G. J. A importância dos grupos de pesquisa na formação de profissionais de Geografia Agrária: A experiência do NEGEF. *Campo-Território: Revista de Geografia Agrária*, v. 3, n. 5, p. 284-290, 2008.

MORISSETTE, J. T.; KHORRAM, S. Exact binomial confidence interval for proportions. *Photogrammetric Engineering & Remote Sensing*, v. 64, p. 281-283, 1998.

NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. *PNAS*, v. 101, p. 5200-5205, 2004.

NEWMAN, M. E. J. *Networks: An introduction*. New York: Oxford University Press, 2010.

OLIVEIRA, S. C.; ACHCAR, J. A. Confiabilidade de redes: Um enfoque Bayesiano. *Revista de Matemática e Estatística*, v. 18, p. 167-194, 2000.

OLIVEIRA, S. C.; BRIGANTINI, B. B.; UEHARA, J. K. Análise da confiabilidade de uma rede social fictícia com enfoque nas arestas e do uso de medidas de centralidade. In: XX SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 2013, Bauru. *Anais...*, novembro de 2013.

SEN, P. K.; SINGER, J. M.; PEDROSO-DE-LIMA, A. C. *From finite sample to asymptotic methods in statistics*. Cambridge: Cambridge University Press, 2009.

SILVA, T. S. A. *Um estudo de medidas de centralidade e confiabilidade de redes*. 2010. (Dissertação de Mestrado em Tecnologia) – Centro Federal de Educação Tecnológica Celso Suckow da Fonseca do Rio de Janeiro.

SOUZA, C. G.; BARBASTEFANO, R. G. Knowledge diffusion and collaboration networks on life cycle assessment. *The International Journal of Life Cycle Assessment*, v. 16, n. 6, p. 561-568, 2011.

SHPUNGIN, Y. Combinatorial approach to reliability evaluation of network with unreliable nodes and unreliable edges. *International Journal of Computer Science*, v. 1, p. 177-183, 2006.

SMALHEISER, N. R.; TORVIK, V. I. Author name disambiguation. *Annual Review of Information Science and Technology*, v. 43, p. 287-313, 2009.

YAN, E.; DING, Y. Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*. v. 60, n. 10, p. 2107-2118, 2009.

YOSHIKANE, F.; KAGEURA, K. Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, v. 60, n. 3, p. 435-446, 2004.

WANG, D. J. *et al.* Measurement error in network data: a re-classification. *Social Networks*, v. 34, p. 396-409, 2012.