

CONHECENDO A REDE DE COAUTORIA DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA DA ÁREA DE ENGENHARIA DE PRODUÇÃO E A SUA INFLUÊNCIA NO NÍVEL DE PRODUTIVIDADE

Ricardo Lopes de Andrade

Universidade Federal de Pernambuco - UFPE
Av. Prof. Moraes Rego, 1235 – Cidade Universitária, Recife – PE
ricardolopesa@gmail.com

Leandro Chaves Rêgo

Universidade Federal de Pernambuco - UFPE
Av. Prof. Luis Freire s/n – Cidade Universitária, Recife – PE
leandro@de.ufpe.br

RESUMO

Estudos apontam uma forte correlação entre produtividade científica e colaboração acadêmica, outros mostram correlações positivas entre algumas métricas de centralidades de Análise de Rede Social (ARS) com índices de desempenho, tais como o índice h e índice g. Neste artigo, construímos uma rede de coautoria entre os bolsistas de produtividade do CNPq da área de engenharia de produção e analisamos quais métricas de ARS impactam o nível de produtividade deles. Nossos resultados indicam que o nível de produtividade apresenta correlação de Spearman positiva significativa com cinco de nove métricas analisadas. Concluimos que os pesquisadores que assumem um papel de “intermediador”, que apresentam maiores números de ligações ou têm maiores facilidades de estabelecer parcerias de publicações, tendem a ter maior nível de produtividade. Além disso, constatou-se também que colaborar com frequência com os mesmos pesquisadores não impacta no nível de produtividade.

PALAVARAS CHAVE. Rede de Coautoria. Nível de Produtividade. Correlação.

Área principal (Teoria e Algoritmos em Grafos)

ABSTRACT

Studies show a strong correlation between scientific productivity and academic collaboration, others show positive correlations between some centrality metrics of Social Network Analysis (SNA) with performance levels, such as the h and the g index. In this paper, we build a co-authorship network among CNPq's research productivity fellows in the area of Industrial Engineering and analyze which SNA metrics impact their productivity level. Our results indicate that the level of productivity has significant positive Spearman correlation with five out of nine metrics analyzed. We conclude that the researchers who assume the role of “intermediary”, who have higher numbers of connections or are able to more easily establish publications partnerships, tend to have higher levels of productivity. In addition, it also found that to collaborate frequently with the same researchers does not impact in the productivity level.

KEYWORDS. Co-authorship Network. Productivity Level. Correlation.

Main area (Theory and Algorithms on Graphs)

1. Introdução

A avaliação de desempenho é uma função inevitável de gestão em qualquer nível, pois promove o progresso do desenvolvimento (Abbasi *et al.*, 2011). Nesse sentido, dentro de um ambiente de investigação como em universidades e institutos de pesquisa, deve haver também uma avaliação de desempenho para os pesquisadores, que geralmente é realizada com base na produtividade deles. Ressalta-se que o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) é um dos principais órgãos de financiamento da pesquisa científica no Brasil. Inclui-se, entre as formas de financiamento, a Bolsa de Produtividade em Pesquisa (PQ), que é atribuída a pesquisadores de todas as áreas. A escolha do pesquisador a ser contemplado pela bolsa é baseada não só na qualidade de um projeto submetido, mas, principalmente, na “qualidade” do pesquisador (WAINER & VIEIRA, 2013).

O desempenho ou a “qualidade” do pesquisador pode ser avaliada por meio da quantificação das atividades de publicação. Para ABBASI *et al.* (2011), o pesquisador terá uma alta apreciação se os trabalhos publicados por ele são citados por outros. Desta forma, o número de citações qualifica as publicações (LEHMANN *et al.*, 2006).

Para auferir o impacto das publicações vários índices vêm sendo propostos. O índice h, índice mais conhecido, foi proposto por Hirsch (2005) como uma medida que combina de uma forma simples, a quantidade de publicações e o impacto das publicações. O índice h, apesar de algumas deficiências – por exemplo, não considera uma quantidade de citações mínima para se obter um valor específico, como relatado por Egghe (2006) – é o índice geralmente usado por muitos bancos de dados acadêmicos (por exemplo, *Web of Science* e *Scopus*) para medir o desempenho dos pesquisadores.

No entanto, as diferentes medidas de desempenho científico: (como, número de publicações, número de citações) e os vários índices disponíveis na literatura observam somente o passado dos pesquisadores, o que pode levar a uma percepção subjetiva do resultado futuro da pesquisa, por parte dos avaliadores. O meio onde os pesquisadores estão inseridos não está ao alcance dos indicadores. Dessa forma, não são percebidos os padrões relacionais desse meio, as estratégias de sociabilidade a partir do modo como se mobilizam e são mobilizados, assim como a dependência entre esses padrões relacionais e atributos e comportamentos individuais (MARTINS *et al.*, 2013). Por isso, torna-se necessário agregar mais fatores no processo de avaliação de desempenho que possam preencher lacunas deixadas pelos indicadores.

No entanto, quais outros fatores poderiam ser incorporados no processo de avaliação de desempenho? Vários estudos, como os de Eaton *et al.* (1999) e Lee & Bozeman (2005), apontam uma forte relação entre produtividade e colaboração. A pesquisa de Hart (2000) mostra que a colaboração leva à “melhoria da qualidade das publicações”. Logo, um dos fatores a serem considerados como forma de elucidar os resultados de produtividade científica dos pesquisadores está relacionado ao modo como eles colaboram entre si. Neste estudo, a relação de colaboração é a de coautoria (desenvolvimento de um mesmo artigo por dois ou mais autores) e é representada por uma rede, ou grafo, terminando por constituir as chamadas redes de coautoria.

A representação em forma de grafo apresenta a vantagem, segundo Wasserman & Faust (1994), de permitir a utilização da Teoria dos Grafos para analisar as interações sociais, os padrões e as implicações. Tal abordagem é conhecida como Análise de Rede Social (ARS). De maneira individual, as métricas da ARS, aplicadas à rede de colaboração, permitem avaliar o *status* de um pesquisador. E de forma global, para Barabasi *et al.* (2002), permitem determinar como o campo de pesquisa está evoluindo e fazer previsões sobre a direção desse campo e onde os avanços terão maiores probabilidades de ocorrerem. O mesmo é relatado por Newman (2004), onde, segundo ele, a estrutura dessas redes revela muitas características e padrões interessantes das comunidades acadêmicas.

Os trabalhos realizados por Abbasi & Altman (2011) e Abbasi *et al.* (2011) estudaram a correlação existente entre métricas de centralidade de ARS, obtidas por meio de uma rede de coautoria, e os índices de desempenho (índice h e índice g, respectivamente) dos pesquisadores participantes dessa rede. Os resultados encontrados nesses trabalhos apontaram uma correlação

positiva entre algumas métricas de centralidade de ARS e os índices de desempenho dos pesquisadores analisados.

Neste contexto, com as métricas de ARS aplicadas à rede de coautoria dos pesquisadores bolsistas de produtividade em pesquisa (PQ) do CNPq da área de Engenharia de Produção, nosso objetivo, de modo geral, é verificar quais métricas de ARS impactam o nível de produtividade dos pesquisadores. Especificamente, será desenvolvida uma rede de coautoria entre os pesquisadores bolsistas de produtividade, e o *status* dos pesquisadores será avaliado por meio das métricas de ARS. Por fim, uma análise da correlação entre as métricas de ARS com o índice *h* e com o nível de produtividade dos pesquisadores será realizada.

O trabalho está organizado da seguinte forma: a Seção 2 apresenta uma fundamentação teórica sobre rede de colaboração e as métricas de ARS; a Seção 3 expõe de forma sucinta os critérios de distribuições de bolsas de produtividades e o índice *h*; a Seção 4 descreve o processo metodológico aplicado no desenvolvimento do estudo; a Seção 5 exhibe os resultados encontrados; por fim, a Seção 6 apresenta as considerações finais do estudo e propostas para trabalhos futuros.

2. Fundamentação Teórica

2.1 Rede de Coautoria

A coautoria, elaboração de um mesmo artigo por dois ou mais autores, é uma forma de colaboração. Coautoria implica numa relação temporal e acadêmica, onde autores compartilham ideias e recursos. Ela é, por exemplo, diferente das citações que podem ocorrer sem que os autores se conheçam ou tenham se comunicado e estendem-se ao longo do tempo.

Para Kumar (2015), os estudos sobre coautoria ganharam novos interesses após Newman (2001) (2004) ter usado métodos de análise de rede social para investigar as características e padrões interessantes das comunidades acadêmicas. Kempe & Kleinberg (2005) também relataram o surgimento de muitas pesquisas de análise de rede de coautoria e na identificação dos autores influentes na mesma. Observa-se também que os trabalhos de Huang *et al.* (2013) e de Liu *et al.* (2014) buscaram estudar a rede de coautoria na possibilidade de avaliar o *status* do autor em um determinado campo e, assim, melhorar as relações para chegar mais perto do núcleo da comunidade por meio da identificação dos pesquisadores mais influentes.

Diferentes motivos podem levar dois ou mais pesquisadores a produzir um artigo conjuntamente. Para Gibbons *et al.* (1994), a colaboração científica é engrenada pelo aumento da complexidade do conhecimento e pela demanda de competências mais especializadas e interdisciplinares nas pesquisas. A pesquisa de Hart (2000), dos autores da literatura de biblioteconomia acadêmica, mostrou que os autores consideram "a melhoria da qualidade do artigo" como um benefício da coautoria. Outra motivação que leva os pesquisadores a colaborarem é a pressão que sentem de instituições de fomento para aumentar a produção de publicações. Assim, eles veem a coautoria como um meio para publicar mais artigos. Destaca-se que uma forte relação entre produtividade e colaboração foi encontrada nos estudos de Eaton *et al.* (1999).

2.2 Análise de Redes Sociais

A análise de rede social (ARS) tem uma história que remota, pelo menos, meio século e produziu muitos resultados a respeito da influência social, grupos sociais, a desigualdade, a propagação de doenças, a comunicação de informações e, na verdade, quase todos os tópicos que têm interessado a sociologia do século XX (NEWMAN, 2001).

A ARS é baseada na premissa de que as relações entre os atores sociais podem ser descritas por um grafo. Os nós (vértices) do grafo representam os atores sociais e as arestas conectam pares de nós e, portanto, representam as interações sociais. Essa representação permite aos pesquisadores a análise do que seria considerado um problema inerentemente evasivo e pouco compreendido: um emaranhado de interações sociais (WASSERMAN & FAUST, 1994). Ainda segundo Wasserman & Faust (1994) um conjunto desenvolvido de algoritmos matemáticos são aplicados para a análise e visualização de rede.

Matematicamente, temos que um **grafo** G é um conjunto finito e não vazio de **nós**, $V(G) = \{v_1, v_2 \dots v_n\}$, e um conjunto finito de **arestas** $A(G)$, onde cada aresta é um par de

elementos (não necessariamente ordenados ou distintos) de $V(G)$. Se todos os pares de $A(G)$ são de elementos distintos, diz-se que o grafo não possui **laços**.

Caminho é uma sucessão de vértices e arestas, onde cada aresta liga o vértice que a precede ao vértice que a segue, não repetindo arestas. O comprimento de um caminho que contenha dois determinados vértices será dado pelo número de arestas presentes no caminho. Se existir mais de um caminho entre dois vértices, então a **distância geodésica** entre tais vértices será dada pelo menor comprimento entre todos esses caminhos.

Um grafo é **conexo** quando entre qualquer par de vértices existe sempre um caminho que o une, caso contrário, dizemos que o grafo é **desconexo**. Um **Subgrafo** de um grafo G é qualquer grafo H tal que $V(H) \subseteq V(G)$ e $A(H) \subseteq A(G)$. Um subgrafo H de G é **próprio** se $V(H) \neq V(G)$ ou $A(H) \neq A(G)$. Um subgrafo conexo H de um grafo G é **maximal** se H não é subgrafo próprio de algum subgrafo conexo de G . Dado um grafo desconexo, uma **componente conexa** é um subgrafo conexo e maximal com respeito à inclusão, ou seja, não existe caminho entre um nó pertencente ao subgrafo e outro nó não pertencente ao subgrafo.

A **componente gigante** refere-se à componente conexa com maior número de nós. No caso de rede de coautoria permite investigar o comportamento do maior grupo colaborativo. A presença de mais de um componente na rede indica a existência de grupos que publicam isoladamente. Em rede de coautoria, essa medida indica a quantidade de autores que participa do maior grupo colaborativo da rede (MENA-CHALCO *et al.*, 2012).

2.3 Métricas de ARS

As métricas na ARS podem ser divididas em **globais**, que descrevem as características sobre todo o grafo e **individuais**, que estão relacionadas com a análise das propriedades individuais de atores da rede.

O **número de arestas**, como o próprio nome diz, refere-se à cardinalidade do conjunto das arestas ($A(G)$), denotada por $\#A(G)$. Medida simples que, no caso de redes de coautoria, reflete o número de ligações entre os pesquisadores.

A **densidade** descreve o nível geral de ligações entre os pontos de um grafo. Um grafo “completo” é aquele em que todos os pontos são adjacentes uns aos outros, ou seja, cada nó é ligado diretamente a todos os outros nós, densidade 1. Para o grafo não direcionado com n nós e sem laços, a densidade é definida como:

$$Dens(G) = \frac{2 \times (\#A(G))}{n \times (n - 1)} \quad (1)$$

Diâmetro refere-se ao tamanho da maior distância geodésica entre qualquer par de vértices. O diâmetro de um grafo pode variar de um mínimo de 1, se o grafo for completo, a um máximo de $n - 1$, onde n é a ordem do grafo. Seja $d_G(v_i, v_j)$ a distância geodésica entre os vértices v_i e vértice v_j , então o diâmetro do grafo conexo G será dado por:

$$Dim(G) = \max_{\{v_i, v_j \in V(G)\}} d_G(v_i, v_j) \quad (2)$$

No caso de grafos desconexos, o diâmetro da rede será dado pelo maior valor encontrado entre os diâmetros das componentes conexas.

Um dos principais interesses deste artigo é com as métricas individuais, as quais levam a identificar os pesquisadores mais importantes entre os membros de uma comunidade acadêmica. Eles podem ser a chave para melhorar o desempenho científico e aumentar os resultados da pesquisa (ANASTASIOS *et al.*, 2012).

Freeman (1979) constatou que centralidade é um importante atributo estrutural na rede social e está relacionado com a eficiência, liderança e satisfação. As medidas de centralidade mais utilizadas são a centralidade de grau, centralidade de proximidade e a centralidade de intermediação. Uma outra medida clássica de centralidade é a centralidade de autovetor.

O grau do vértice é calculado em termo do número de vértices adjacentes, o grau pode ser considerado como uma medida de centralidade local. Portanto, a **centralidade de grau** de um vértice i em um grafo G não ponderado é:

$$C_d(i) = \sum_{j=1}^n M_{ij} \quad (3)$$

em que i é o nó focal, j representa todos os outros nós, n é o número total de nós e M é a matriz adjacente, no qual $M_{ij} = 1$ se o nó i está conectado ao nó j e $M_{ij} = 0$, caso contrário.

O **grau ponderado** do vértice é definido por:

$$C_d^w(i) = \sum_{j=1}^n w_{ij}, \quad (4)$$

em que w é a matriz adjacente ponderada e w_{ij} recebe um valor real maior que 0 (zero) se i está conectado a j , esse valor representa o peso dessa ligação. Apesar de ser uma medida simples, o grau mede a influência de um nó. Em muitos contextos sociais, as pessoas com mais conexões tendem a ter mais poder. Em uma rede sem laços, a centralidade de grau poderá variar de 0, no caso de nós isolados, até $n - 1$, quando o nó possui ligações com todos os demais nós da rede. Em uma rede de coautoria, essa medida identifica os autores mais ativos e populares (ABBASI *et al.*, 2011) (ANASTASIOS *et al.*, 2012).

Para uma medida mais genérica, a fim de comparação de centralidade de grau em redes diferentes com diferentes tamanhos, foi proposto por Freeman (1979) uma medida relativa. Essa medida normaliza o número real de ligações pelo número máximo de ligações que se poderia ter. Assim, a **centralidade relativa de grau**, em um grafo sem laços, de um vértice i é dada por:

$$C'_d(i) = \frac{C_d(i)}{n - 1} \quad (5)$$

Outra medida para analisar um nó na rede partiu da teoria da “força das ligações fortes” de Krackhardt (1992). Para o cálculo da força da ligação de um nó dividimos o grau ponderado, $C_d^w(i)$, pela centralidade de grau, $C_d(i)$.

$$TS(i) = \frac{C_d^w(i)}{C_d(i)} \quad (6)$$

Sendo assim, $TS(i)$ representa o **peso médio** das ligações de um nó. A teoria da “força das ligações fortes” é uma contradição a teoria da “força das ligações fracas”, de Granovetter (1973). As ligações fracas proporcionam aos indivíduos da rede acesso à informação e recursos além daqueles disponíveis em seus próprios círculos sociais (vizinhos), ou seja, funcionam como uma ponte; porém as ligações fortes conduzem ao recebimento de conhecimentos úteis e são geralmente mais acessíveis.

A distância média de um determinado nó inicial para todos os demais nós da rede é calculada pela **centralidade de proximidade**. Seja d_{ij} a distância geodésica entre os nós i e j . Define-se a centralidade de proximidade do nó i , $C_c(i)$, como:

$$C_c(i) = \frac{1}{\sum_j d_{ij}} \quad (7)$$

A medida pode ser normalizada utilizando a maior distância possível entre quaisquer dois nós de uma rede de n nós. Este valor é $n - 1$. Então, a **centralidade relativa de proximidade** do nó i , é dada por:

$$C'_c(i) = C_c(i)(n - 1) \quad (8)$$

A centralidade de proximidade é uma medida inversa, ou seja, os vértices mais centrais da rede, de acordo com essa medida, são aqueles que possuem uma distância menor dos outros vértices. Um nó que está, em média, numa posição mais próxima dos outros nós pode obter informações de maneira mais eficiente, ou seja, a medida de proximidade está relacionada à independência e eficiência na comunicação com outros nós (FREEMAN, 1979). Nesse sentido, na rede de coautoria, um autor com uma centralidade de proximidade alta pode indicar uma maior possibilidade de estabelecer parcerias de publicação (SOUZA *et al.*, 2012).

A **centralidade de intermediação** é o número de caminhos mais curtos que passam por meio de um dado nó. Considerando $j \neq k \neq i$, seja g_{jk} o número de caminhos mais curtos do nó j para o nó k e g_{jik} é o número de caminhos mais curtos do nó j para o nó k passando por i . Determinamos a centralidade de intermediação $C_b(i)$ do vértice i como (BRANDES, 2001):

$$C_b(i) = \sum_{j,k} \frac{g_{jik}}{g_{jk}} \quad (9)$$

Da mesma forma que as centralidades anteriores, existe a necessidade de normalizar essa medida. A centralidade de intermediação do vértice i é normalizada pelo número máximo possível de caminhos mais curtos, excluindo o vértice i . Dado uma rede não direcionada, o máximo é: $[(n-1)(n-2)]/2 = (n^2 - 3n + 2)/2$. Assim, a **centralidade relativa de intermediação** é dada por:

$$C'_b(i) = \frac{2 \times C_b(i)}{(n^2 - 3n + 2)} \quad (10)$$

A intermediação é um indicador do potencial de um nó de desempenhar um papel de “intermediador” ou “porteiro”, podendo controlar com maior frequência o fluxo de informação na rede FREEMAN (1979). Nesse sentido, em uma rede de coautoria, um autor com alto valor de centralidade de intermediação indica que um número significativo das parcerias estabelecidas na rede envolve, de forma direta ou indireta, as publicações relacionadas a esse autor (SOUZA, 2012).

Uma medida de importância do nó na rede baseada nas conexões, a **centralidade de autovetor** é fundamentada na ideia que um determinado nó terá centralidade alta se estiver conectado a nós com posições centrais na rede (BONACICH, 1987), ou seja, a centralidade de um nó não depende apenas do número de nós adjacente, mas também da centralidade desses nós. Seja λ uma constante, então, a centralidade de autovetor do vértice i , $C_e(i)$ será dada por:

$$C_e(i) = \frac{1}{\lambda} \sum_{j=1}^n M_{i,j} C_e(j) \quad (11)$$

Usando a notação vetorial, seja $X = (C_e(1), C_e(2) \dots C_e(n))$ o vetor de centralidade, podemos reescrever a equação acima como, $\lambda X = MX$.

Ao assumir que a centralidade são valores não negativos (usando o teorema de Perron-Frobenius), podemos mostrar que λ é o maior autovalor da matriz de adjacência, sendo X o autovetor correspondente (JACKSON, 2008).

Para normalizar a centralidade de autovetor é usada a norma Euclidiana (ABBASI *et al.*, 2011). Então, a **centralidade relativa de autovetor** é dada por:

$$C'_e(i) = \frac{C_e(i)}{\sqrt{\sum_{j=1}^n C_e(j)^2}} \quad (12)$$

O **PageRank** trabalha contando o número e a qualidade dos links de uma página, a fim de determinar uma estimativa aproximada de quão influente uma página web é (LIU *et al.*, 2014). Seja T uma página web. O T_i representa uma das páginas da web. Assim, o PageRank é definido como função:

$$PR(T_A) = (1 - \delta) + \delta \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right) \quad (13)$$

Onde $PR(T_A)$ é o PageRank da página T_A , $PR(T_i)$, $i=1, \dots, n$, são os PageRank das páginas T_i , $i=1, \dots, n$, que se conectam a página T_A , $C(T_i)$ é o número de links de saída na página T_i e δ é um fator de amortecimento, que pode estar situado entre 0 e 1.

O **AuthorRank** é um PageRank modificado, o qual considera o peso da ligação e revela o *status* de um autor em uma rede de coautoria. Em rede de coautoria, o peso está relacionado à frequência de colaboração, ou seja, o número de produções feitas em coautoria (LIU *et al.*, 2005). O AuthorRank de um autor i , $AR(i)$ é definido como uma função:

$$AR(i) = (1 - \delta) + \delta \sum_{j=0}^n AR(j) \times w_{j,i}, \quad (14)$$

em que $w_{j,i}$ é o peso da aresta entre os autores j e i .

A **excentricidade**, $e(v_i)$, de um vértice é a distância geodésica máxima a partir dele para qualquer outro vértice do grafo. Quanto menor a excentricidade de um vértice, melhor é o relacionamento com os outros vértices. A excentricidade de v_i será dada por:

$$e(v_i) = \max_{v_j \in V} d_G(v_i, v_j) \quad (15)$$

A excentricidade máxima da rede é chamada de **diâmetro**, enquanto a excentricidade mínima é chamada de **raio**. Os vértices com excentricidade mínima formam o centro de G ,

enquanto os vértices com excentricidade máxima são chamados vértices periféricos. No caso de redes desconexas, a excentricidade de cada vértice é calculada levando em consideração o componente ao qual pertence.

Coefficiente de cluster (agrupamento) juntamente com o valor médio de caminho mais curto, pode identificar um efeito “*small-world*”. Indica como os nós estão inseridos na vizinhança. Dado um nó i o coeficiente de agrupamento é a razão entre o número de triângulos que contém um determinado nó i e o número de arestas possíveis entre os vizinhos. Sendo NT_i o número de triângulos contendo o nó i então temos o coeficiente de *cluster* como:

$$CL_i = \frac{NT_i}{\binom{C_d(i)}{2}} \quad (16)$$

O coeficiente médio de agrupamento é o valor médio dos coeficientes individuais. Na rede de coautoria o coeficiente de agrupamento de um determinado autor indica o quanto os colaboradores dele também colaboram entre si.

Todas as métricas individuais acima apresentadas nos fornecem informações a respeito dos pesquisadores na rede sobre diferentes perspectivas posicionais e a partir delas serão estudadas as relações com as características individuais dos pesquisadores.

3. Avaliação de Desempenho Acadêmico

A avaliação de desempenho é uma função inevitável de gestão em qualquer nível, pois promove o progresso do desenvolvimento (ABBASI *et al.*, 2011). Neste sentido, dentro de um ambiente de pesquisa como em universidades e institutos de pesquisa, deve-se também haver uma avaliação de desempenho para os acadêmicos, que geralmente é realizada com base na produtividade do pesquisador.

3.1 Sistema de Avaliação da Produtividade em Pesquisa

No Brasil, o CNPQ é um dos principais órgãos de financiamento da pesquisa científica. Entre as formas de financiamento do CNPq inclui-se a Bolsa de Produtividade em Pesquisa (PQ), que é atribuída a pesquisadores de todas as áreas, baseado não só na qualidade de um projeto submetido, mas principalmente na “qualidade” do pesquisador (WAINER & VIEIRA, 2013). A PQ é organizada em níveis, em ordem crescente: 2, 1D, 1C, 1B, 1A, sendo que os últimos quatro níveis são coletivamente chamados de “níveis 1”. Para o pesquisador se enquadrar na categoria 1, exige-se que ele tenha, no mínimo, 8 (oito) anos de doutorado por ocasião da implementação da bolsa. O enquadramento do pesquisador na categoria 2 exige que ele tenha, no mínimo, 3 (três) anos de doutorado por ocasião da implementação da bolsa. O desempenho do pesquisador é avaliado por meio de indicadores referentes ao quinquênio anterior, no caso da categoria 2, e do decênio anterior, no caso da categoria 1.

3.2 Índice h

Uma boa medida para avaliar o desempenho dos acadêmicos é por meio da quantificação das atividades de publicação, sendo que o pesquisador terá uma alta apreciação se os trabalhos publicados por ele são citados por outros (ABBASI *et al.*, 2011). Para Lehmann *et al.* (2006), o número de citações qualifica as publicações. Portanto, o desempenho do pesquisador não está atribuído apenas à quantidade de publicações, mas também ao impacto, ou seja, o quanto essas publicações tiveram consequências para a área de pesquisa em que o pesquisador se situa, ou até mesmo para outras áreas.

O índice h foi proposto por Hirsch (2005) como uma medida que combina, de uma forma simples, a quantidade de publicações e o impacto das publicações (ou seja, número de citações). O índice h é definido da seguinte maneira: “Um pesquisador possui um índice h de k, se k de N trabalhos dele possuem ao menos k citações cada, e os outros (N – k) trabalhos tenham no máximo k citações cada” (Hirsch, 2005). O índice h como relatado por Egghe (2006) não considera uma quantidade de citações mínima para se obter um valor específico. Apesar dessa deficiência, esse índice é geralmente usado por muitos bancos de dados acadêmicos (por exemplo, *Web of Science* e *Scopus*) para medir o desempenho dos pesquisadores.

4. Metodologia

A rede de coautoria desenvolvida neste artigo foi formada por pesquisadores bolsistas de produtividades do CNPq da área de engenharia de produção. As informações sobre quais pesquisadores são bolsistas foram retiradas do site do CNPq, onde, na busca, utilizamos os seguintes filtros: grande área – engenharias; área – engenharia de produção; modalidade – produtividade em pesquisa e; níveis – todos. No total foram apontados, em 16 de fevereiro de 2015, 140 pesquisadores bolsistas de produtividade, sendo esses utilizados na construção da rede.

As relações de coautoria foram obtidas pelo software scriptLattes que utiliza os IDs Lattes (código que identifica o pesquisador registrado na Plataforma Lattes) para extrair as publicações dos pesquisadores e detectar automaticamente as produções iguais (Mena-Chalco & Cesar-Jr., 2009). Foram consideradas, para as relações de coautorias, somente artigos publicados em periódicos e artigos aceitos para publicações, entre 2005 a 2014.

A ferramenta scriptLattes gera um arquivo em formato *.gdf*. Com esse formato é possível utilizar o software Gephi para visualizar a rede de coautoria e calcular as métricas de ARS. O software Gephi é de código aberto, usado para exploração e compressão de redes, oferecendo um conjunto de métricas para análise de redes divididas em três níveis: visão geral da rede, visão geral dos nós e visão geral das arestas, sendo possível calcular as medidas mais comuns de análise de redes.

O índice h dos pesquisadores foi obtido na aba “Indicadores da Produção”, no site do CNPq, quando utilizamos a ferramenta de busca por currículos Lattes e clicamos sobre o nome do pesquisador. Nessa aba, está disponível o índice h calculada pela *Web of Science* e *Scopus*. Consideramos o índice h desses bancos de dados, pois são usados no cálculo, somente, os artigos publicados em periódicos e registrados no currículo Lattes, os quais foram utilizados para a formação da rede de coautoria.

Como os níveis de produtividade em pesquisas são em ordem crescente: 2, 1D, 1C, 1B e 1A, atribuímos valores a esses níveis, seguindo a ordem de importância para poder realizar as correlações de Spearman.

5. Análises e Resultados

Entre artigos publicados e aceitos, os pesquisadores possuem em média 31,6 artigos, desses em média 10,1 foram elaborados em coautoria, resultando em 147 ligações na rede. A rede está dividida em 30 **componentes**, sendo a componente gigante composta por 92 vértices, representando aproximadamente 65,71% dos nós da rede; a segunda possui 8 vértices (5,71%) e 23 pesquisadores possuem componente tamanho 1, ou seja, cerca de 16,42% dos bolsistas não possuem colaboradores na rede.

Ao analisar a **densidade** vemos que a rede é dispersa, pois somente possui densidade igual a 0,016, ou seja, apenas estão ocorrendo 1,16% das ligações possíveis na rede. A densidade é considerada muito baixa se comparada com a densidade máxima que é igual a 1, porém pode ser justificada pelo fato de a rede ser formada somente por um grupo restrito da produção dos pesquisadores (apenas artigos publicados em periódicos e artigos aceitos para publicações, entre 2005 a 2014) e que são bolsistas de produtividade da área de engenharia de produção, desconsiderando a colaboração com outros pesquisadores.

O **diâmetro** da rede é igual a 13 e o **raio** 0, representando a excentricidade máxima e a mínima respectivamente, sendo o raio da componente gigante igual a 7. O coeficiente de **cluster** médio é igual a 0,48, sabendo que esse coeficiente pode variar de 0 a 1, temos então que aproximadamente metade das possíveis coautorias entre os coautores de um dado autor estão presentes na rede.

A **distância média de um caminho** entre um par de vértices é aproximadamente 6,03. Esse valor refere-se à componente gigante, e significa que, em média, 6,03 ligações separam dois pesquisadores quaisquer na componente. O número de caminhos mais curtos é de 8.642.

A Figura 1 ilustra a rede de coautoria dos bolsistas gerada pelo software Gephi, em que a espessura das arestas é proporcional ao peso (total de artigos em coautoria) e o diâmetro do nó é proporcional ao grau.

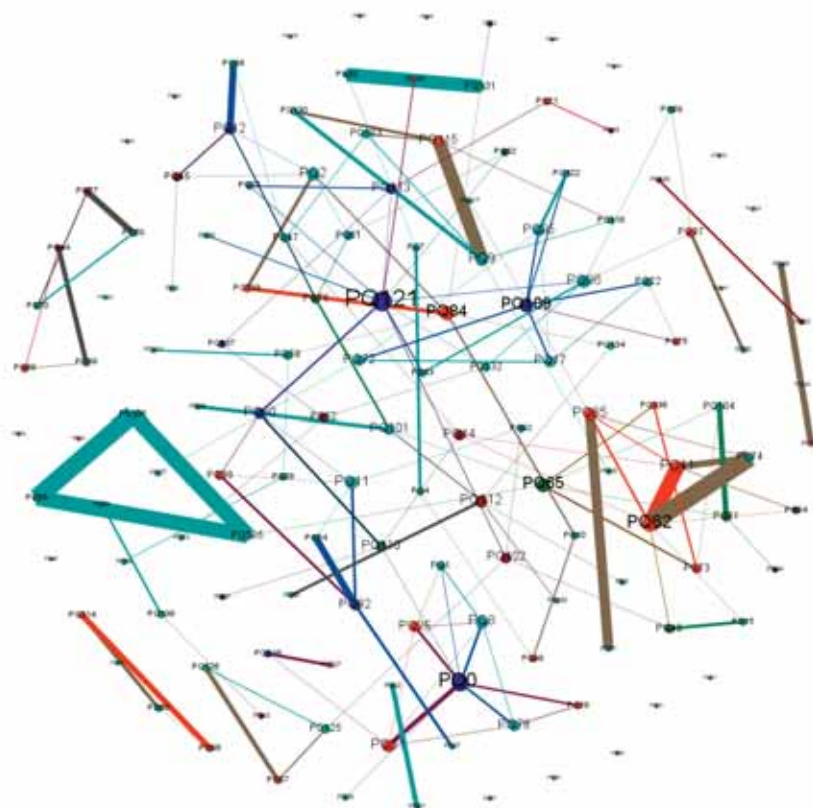


Figura 1- Rede de coautoria entre os PQ da área de engenharia de produção.

Optamos por não revelar a identidade dos bolsistas, portanto, representamos cada bolsista por PQ seguido por números que variam de 0 a 139.

5.1 Correlação de Spearman

Para avaliar quais métricas de ARS impactam o nível de produtividade dos bolsistas usamos a correlação de Spearman. O resultado é mostrado na Tabela 1, que apresenta, além das correlações entre nove métricas de ARS e o nível de produtividade, a correlação entre essas métricas e o índice h. As métricas de centralidades utilizadas foram normalizadas.

Em relação ao índice h, observa-se que o índice calculado pela *Web of Science* apresentou correlações diferentes ao índice calculado pela *Scopus* com as métricas de ARS. Enquanto o primeiro tem correlações significativas positivas com a centralidade de grau, a centralidade de intermediação, o grau ponderado e o AuthorRank, o segundo apresentou correlação significativa com a maioria das métricas, exceto excentricidade e coeficiente de *cluster*. Em geral, as correlações entre as métricas de ARS foram mais fortes com o índice h da *Scopus* que com o índice h da *Web of Science*. A exemplo do grau ponderado, mostrando que quanto mais um pesquisador elabora artigos em coautoria maior é o índice h. Em outras palavras, pesquisadores que mantêm ligações fortes tendem a ter melhor desempenho. Porém essa conclusão é mais direcionada ao índice h calculado pela *Scopus*.

Os coeficientes de correlação entre as métricas de ARS (isto é, centralidade de grau, centralidade de intermediação, centralidade de autovetor e AuthorRank) e o nível de produtividade são positivos e significativos, com exceções para o grau ponderado, a média das ligações fortes, a excentricidade e o coeficiente de *cluster*, os quais não obtiveram correlações significativas.

Tabela 1 – Teste de correlação de Spearman

Variáveis	Nível de Produtividade	Índice h (<i>Web of Science</i>)	Índice h (<i>Scopus</i>)
Centralidade de Grau	0,29*	0,21*	0,30*
Centralidade de Intermediação	0,31*	0,20*	0,33*
Centralidade de Proximidade	0,23*	0,10	0,18*
Centralidade de Autovetor	0,25*	0,14	0,22*
Grau Ponderado	0,14	0,17*	0,31*
AuthorRank	0,22*	0,23*	0,33*
Média das Ligações Fortes	0,02	0,10	0,23*
Excentricidade	0,05	-0,04	0,15
Coef. de Cluster	0,06	0,14	0,09

* A correlação é significativa no nível 0,05 (2 extremidades).

A centralidade que apresentou maior correlação com o nível de produtividade foi a centralidade de intermediação, seguido pela centralidade de grau. Dessa forma, os pesquisadores que apresentam as maiores centralidades de grau, maiores números de coautores, tendem a ter maiores níveis de produtividade. Mas essa tendência é maior para os pesquisadores que assumem um papel de “intermediador”, controlando a frequência do fluxo de informação. Por apresentarem uma maior possibilidade de estabelecer parcerias de publicações, os pesquisadores com maiores centralidades de proximidade tendem também a apresentar um maior nível de produtividade, conforme a correlação apresentada.

Não existiu correlação significativa entre o nível de produtividade com o grau ponderado e com a média das ligações fortes, mostrando que colaborar com frequência com os mesmos pesquisadores não impacta o nível de produtividade. As medidas de excentricidade e coeficiente de *cluster* também não apresentaram correlações significativas com o nível de produtividade nem com o índice h.

A Tabela 2 mostra os 5 pesquisadores mais bem posicionados nas métricas que apresentaram correlações significativas com o nível de produtividade. Observa-se que alguns bolsistas nível 2 estão entre os 5 em todas as métricas, exceto, centralidade de grau. Também é notável a participação do nível 1A na primeira posição em todas as métricas.

Tabela 2 – Os 5 pesquisadores mais bem posicionados

Centralidade					
Intermediação	Grau	Proximidade	Autovetor	AuthorRank	
PQ121 (1A)	PQ121 (1A)	PQ121 (1A)	PQ0 (1A)	PQ0 (1A)	
PQ61 (2)	PQ0 (1A)	PQ32 (1B)	PQ121 (1A)	PQ121 (1A)	
PQ32 (1B)	PQ62 (1D)	PQ61 (2)	PQ8 (2)	PQ82 (1A)	
PQ110 (1B)	PQ65 (1B)	PQ60 (1A)	PQ31 (1D)	PQ62 (1D)	
PQ101 (2)	PQ84 (1D)	PQ14 (1C)	PQ25 (1D)	PQ9 (2)	

6. Considerações Finais

Um dos critérios de escolha dos pesquisadores a serem contemplados com bolsas de produtividade em pesquisa do CNPq é a análise do histórico produtivo a fim de avaliar o desempenho deles. Vimos que uma boa medida para avaliar o desempenho, ou a “qualidade” dos pesquisadores é por meio da quantificação das atividades de publicação. O índice h, medida baseada em citações, é o índice mais conhecido e geralmente usado por grandes bancos de dados, como a *Web of Science* e *Scopus*. Relatamos também que tais medidas apresentam deficiências, pois não capturam os padrões relacionais e as interações sociais dos pesquisadores, as quais poderiam nos dar informações sobre o desempenho futuro desses pesquisadores.

Dessa forma, construímos uma rede de coautoria entre os bolsistas de produtividade em pesquisa do CNPq da área de engenharia de produção e verificamos quais métricas de ARS impactam o nível de produtividade. Utilizamos as métricas de centralidades normalizadas: de grau, de intermediação, de proximidade e de autovetor. E as métricas: AuthorRank, grau ponderado, média das ligações fortes, excentricidades e coeficiente de cluster.

Os resultados da análise de correlação de Spearman mostraram que o nível de desempenho dos bolsistas está associado de forma positiva significativa com cinco entre as nove medidas de ARS estudadas. No entanto, o coeficiente de correlação para a centralidade de intermediação, centralidade de grau e centralidade de autovetor foram mais altos e significativos.

Em relação ao grau de intermediação, os pesquisadores que controlam, de maneira direta ou indireta, o fluxo de informação na rede, apresentam maior nível de produtividade do que aqueles que não apresentam essa característica de intermediador. Pesquisadores com maiores ligações ou coautores possuem um maior nível de produtividade do que aqueles que publicam com poucos parceiros. Pela centralidade de autovetor, pesquisadores que estão ligados a pesquisadores mais centrais têm maior centralidade, favorecendo também a terem maior nível de produtividade. A possibilidade de estabelecer mais parcerias de publicações faz o pesquisador que apresenta elevado grau de proximidade obter um maior nível de produtividade.

A não existência de correlação significativa com o grau ponderado e com a média das ligações fortes nos faz concluir que a frequência de colaboração com os mesmos pesquisadores não impacta o nível de produtividade.

Embora as correlações entre o nível de produtividade e algumas métricas de ARS sejam classificadas como fracas a moderadas, foram significantes ao nível de significância de 0,05. Nesse sentido, poderá ser capaz de escrever uma linha de regressão, sendo possível prever o nível de produtividade de um pesquisador quando se conhece a posição dele na rede de coautoria, ficando essa análise de regressão para um trabalho futuro.

Ainda para trabalho futuro, poder-se-ia analisar informações como tempo de doutorado dos pesquisadores, pois há restrições em atribuir bolsas de níveis mais altos a pesquisadores mais jovens (do ponto de vista de idade científica): para o pesquisador se enquadrar na categoria 1, exige-se que ele tenha, no mínimo, 8 (oito) anos de doutorado por ocasião da implementação da bolsa; o enquadramento do pesquisador na categoria 2 exige que ele tenha, no mínimo, 3 (três) anos. Com essas informações poderíamos analisar por que pesquisadores bem posicionados na rede possuem níveis de produtividade inferiores a outros com posições abaixo deles, como podemos observar na Tabela 2.

Referências

Abbasi, A., Altmann, J. (2011), "On the correlation between research performance and social network analysis measures applied to research collaboration networks". In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual. Waikoloa, HI: IEEE.

Abbasi, A.; Altmann, J.; Hossain, L. (2011), Identifying the effects the co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, vol. 5, p. 594-607.

Anastasios, T.; Sgoiropoulou, C.; Papageorgiou, E.; Terraz, O.; Miaoulis, G. (2012), Co-authorship networks in academic research communities: the role of network strength. *16th Panhellenic Conference on Informatics*.

Barabasi, A. L.; Jeong, H.; Neda, Z.; Ravasz, E.; Schubert, A; Ticssek, T. (2002), Evolution of the social network of scientific collaboration. *Physica A: Statistical Mechanics and Applications*, 311(3-4):590-614.

Bonacich, P. (1987), Power and centrality: a family of measures. *The American Journal of Sociology*, p. 1170-1182.

Brandes, U. (2001), Um algoritmo mais rápido para betweenness centralidade. *Journal of Mathematical Sociologia*, 25(2), p. 163-177.

- Eaton, J. P.; Ward, J. C.; Kumar, A.** (1999), Structural Analysis of Co-Author Relationships and Author Productivity in Selected Outlets for Consumer Behavior Research. *Journal of Consumer Psychology*, p. 39-59.
- Egghe, L.** (2006), Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.
- Freeman, L. C.** (1979), Centrality in Social Networks Conceptual Clarification. *Social Networks*, 215-239.
- Gibbons, M.; Limoges, C.; Nowotny, H.; Schwartzman, S.; Scott, P.; Trow, M.** (1994), The New Production of Knowledge. *Sage*, London.
- Granovetter, M. S.** (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360.
- Hart, R.** (2000), Co-authorship in the academic library literature: A survey of attitudes and behaviors. *Journal of Academic of Librarianship* p. 339-345.
- Hirsch, J. E.** (2005), An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569.
- Huang, P.-Y., Liu, H.-Y.; Chen, C.-H.; Cheng, P.-J.** (2013), The impact of social diversity and dynamic influence propagation for identifying influencers in social networks, in: *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on*, Vol. 1, pp. 410–416.
- Kempe, D.; Kleinberg, J.** (2005), É Tardos, Influential nodes in a diffusion model for social networks, in: *Automata, Languages and Programming, Springer*, pp. 1127–1138.
- Krackhardt, D.** (1992). The strength of strong ties: The importance of philos in organizations. *Networks and Organizations: Structure, Form, and Action*, 216–239.
- Kumar, S.** (2015), Co-authorship networks: a review of the literature. *Aslib Journal of Information Management*, Kuala Lumpur, Malaysia. Vol. 67 Iss 1 p. 55-73.
- Lee, Sooho; Bozeman, Barry.** (2005), The impact of research collaboration on scientific productivity. *Social Studies of Science*, v. 35, n. 673.
- Lehmann, S.; Jackson, A. D.; Lautrup, B. E.** (2006), Measures for measures. *Nature*, 444(7122), 1003–1004.
- Liu, J.; Li, Y.; Ruan, Z.; Fu, G.; Chen, X.; Sadiq, R.; Deng, Y.** (2014), A new method to construct co-author networks. *Physica A*. Chongqing, China, out.
- Liu, X.; Bollen, J.; Nelson, M. L.; Sompel, H. Van De.** (2005), Co-authorship networks in the digital library research community. *Information Processing and Management*. V. 41, p. 1461-1480.
- Martins, L. D; Ferreira, S. M. S. P.** (2013), Proposta de metodologia de mapeamento e avaliação da produção científica da Universidade de São Paulo com foco na estrutura e dinâmica de suas redes de colaboração científica: em busca de novos modelos causais. *Liinc em Revista*, Rio de Janeiro, v. 9, n. 1, p. 181-195.
- Mena-Chalco, J. P.; Cesar Junior, R. M.** (2009), ScriptLattes: An open-source knowledge extraction system from the Latts platform. *Journal of the Brazilian Computer Society*, vol. 15, p. 31-39.
- Newman, M. E. J.** (2001), Who is the best connected scientist? A study of scientific coauthorship networks. *Complex Networks*, pp. 337-370.
- Newman, M. E. J.** (2004), Coauthorship networks and patterns of scientific collaboration. *Proceeding of the National Academy of Sciences*, vol. 101, p. 5200-5204.
- Souza, C. G.; Barbastefano, R. G.; Lima, L. S.** (2012), Redes de colaboração científica na área de química no Brasil: um estudo baseado nas coautorias dos artigos da revista Química Nova. *Química Nova*, São Paulo, vol. 35, n. 4, p. 671-676.
- Wasserman S.; Faust, K.** (1999), *Social Networks Analysis: Methods and Applications*. Cambridge University Press. *Structural analysis in social the social sciences series*, vol. 8.
- Wainer, J.; Veira, P.** (2013), Correlation between bibliometrics and peer evaluation for all disciplines: the avaluation of Brazilian scientists. *Scientometrics (online)*, vol. 96, p. 395-410.