Identificando correlações entre métricas de Análise de Redes Sociais e o *h-index* de pesquisadores de Ciência da Computação

Ayslânya J. Wanderley[1], Alexandre N. Duarte[1], Alisson V. de Brito[1], Mateus A. S. Prestes[1], Felipe C. Fragoso[1]

[1] Centro de Informática – Universidade Federal da Paraíba (UFPB)
Programa de Pós-Graduação em Informática – PPGI
João Pessoa – PB – Brasil

{ayslanya.jw, mateus.prestes, felipe.cfragoso}@gmail.com, {alexandre, alisson}@ci.ufpb.br

Abstract *Evidences show that the way researchers collaborate has a strong impact on their productivity. This study analyzed the correlation between Social Network Analysis metrics applied to a Computer Science scientific collaboration network and the researchers scientific performance. The correlation was studied using the Spearman correlation coefficient and the results show that researchers who establish strong bonds of collaboration and act as a mediator within a network, in general, have a higher scientific rank.*

Resumo. *Evidências apontam que a forma como os pesquisadores colaboram tem forte impacto sobre a sua produtividade. Neste sentido, este estudo analisou a correlação existente entre as métricas de Análise de Redes Sociais das redes de colaboração de pesquisadores de Ciência da Computação e o seu desempenho científico. A correlação foi estudada usando o Coeficiente de Correlação de Postos de Spearman e os resultados encontrados demonstram que pesquisadores que estabelecem laços de colaboração fortes e possuem um papel de intermediador dentro da rede, em geral, possuem um índice de desempenho mais elevado.*

1. Introdução

É muito difícil avaliar a curto prazo a relevância científica de um pesquisador. Muitas vezes, resultados levam anos ou décadas para demonstrar seu real valor. Porém, a comunidade científica atual, especificamente, órgãos de fomento e instituições de pesquisa, demandam formas mais imediatistas para aferir o impacto ou relevância de um pesquisador. Por essa razão, diferentes índices de relevância científica têm sido propostos ao longo dos anos (Wainer e Vieira, 2013).

De forma geral, tais índices tentam auferir a relevância de um pesquisador através do impacto de suas publicações. Como medir precisamente o impacto de uma publicação também tende a ser uma tarefa extremamente complexa, tal medida acaba sendo aproximada pelo número de citações recebidas.

O mais conhecido destes índices é o *h-index*, que combina a quantidade de artigos publicados com o número de citações obtidas por cada artigo, para caracterizar a produção científica e relevância de um pesquisador (Hirsch, 2005).

No entanto, Martins *et al* (2013) defendem que esses indicadores são insuficientes, uma vez que desconsideram os padrões de relação nos quais os pesquisadores se inserem, bem como a relação entre estes padrões e os atributos e comportamentos individuais.

Nesta perspectiva, em busca de novos fatores que possam servir de complemento para a avaliação da relevância científica, estudos recentes (Araújo *et al* 2013, Maia *et al* 2008, Faria *et al* 2012) apontam que o modo como os cientistas colaboram entre si tem um forte impacto sobre sua produtividade. Essa relação de colaboração científica é representada, computacionalmente, através de uma rede (grafo), emergindo assim as chamadas redes de colaboração científica.

Para Freire (2011), este tipo de rede pode ser definida como uma rede social onde os pesquisadores são representados por nós e as relações de colaboração entre eles, sejam elas publicações, orientações, participações em bancas, entre outras, formam arestas entre estes nós.

Ao assumir esse caráter, as redes de colaboração científica podem ser analisadas através da aplicação de técnicas de Análise de Redes Sociais (ARS), de modo a fornecer informações relevantes sobre a dinâmica da interação entre os pesquisadores.

Diante deste contexto, este estudo aplicou métricas de ARS à rede de colaboração de pesquisadores bolsistas de produtividade (PQ) do CNPq na área de Ciência da Computação, com o intuito de responder a seguinte questão de pesquisa: *Existe alguma correlação entre métricas de ARS aplicadas sobre uma rede de colaboração científica e métricas utilizadas para avaliar a relevância científica*[1] *dos pesquisadores participantes da rede?*

Tal estudo faz parte de uma pesquisa em andamento que visa analisar as associações existentes entre as métricas de ARS aplicadas às redes de colaboração científica e os índices de desempenho dos pesquisadores participantes das redes, de modo a utilizá-las na construção de um Modelo de Regressão capaz de prever o sucesso acadêmico de pesquisadores de acordo com sua participação e relevância na rede.

O restante deste trabalho encontra-se estruturado da seguinte forma: a Seção 2 descreve os trabalhos relacionados. A Seção 3 apresenta uma breve fundamentação teórica. Na Seção 4 é apresentada a metodologia utilizada no desenvolvimento do estudo. Já a Seção 5 expõe os resultados encontrados. Concluindo o trabalho, a Seção 6 apresenta as considerações finais da pesquisa e propostas para trabalhos futuros.

2. Trabalhos Relacionados

O impacto das relações de colaboração de um cientista sobre sua produtividade tem sido foco de diversas pesquisas que buscam encontrar novos indicadores de avaliação do sucesso científico de pesquisadores. Faria (2011) analisou a rede social de Ciência da Computação no Brasil em relação a parâmetros de produtividade (número de artigos publicados e número de projetos de pesquisa) e métricas de análise de redes sociais (grau, *betweenness*, coeficiente de clusterização, *closeness*), encontrando correlação direta entre o aumento da produção e as métricas avaliadas.

---

[1] Este trabalho utiliza o *h-index* para auferir a relevância científica de um pesquisador

A relação entre os padrões de colaboração e de produtividade também é analisada na pesquisa de Araujo *et al* (2013), que demonstrou que o perfil de colaboração característico de pesquisadores premiados com bolsas de estudos governamentais sugere um forte impacto na sua produtividade.

Outros trabalhos recentes (Abassi e Altman 2011; Abassi *et al* 2011) estudaram a correlação existente entre algumas métricas de ARS, coletadas a partir de uma rede de coautoria, e os índices de produção dos pesquisadores participantes dessa rede. Os resultados encontrados nestes trabalhos apontam uma correlação positiva entre algumas métricas de ARS e o *h-index* e o *g-index* dos pesquisadores analisados.

A proposta deste artigo difere das demais pelo fato de correlacionar métricas extraídas de redes individuais de cada pesquisador, permitindo posteriormente gerar um Modelo de Regressão capaz de predizer o sucesso científico de um pesquisador isolado. Além disso, a abordagem utilizada nesta pesquisa considera como relação de colaboração científica, além das coautorias, que são o foco dos trabalhos supracitados, qualquer relação que envolva mais de um pesquisador, incluindo orientações de mestrado e/ou doutorado, participações em bancas de defesa e em projetos de pesquisa. Todas essas relações são consideradas com mesmo grau de importância tendo em vista que representam interações sociais entre os pequisadores.

3. Fundamentação Teórica

3.1. Colaboração Científica e Análise de Redes Sociais

A colaboração científica pode ser definida como o trabalho conjunto de pesquisadores para atingir um objetivo comum que consiste em produzir novos conhecimentos científicos (Guerra, 2012). Sendo assim, as redes de colaboração científica evidenciam as interações que ocorrem entre tais pesquisadores, possibilitando uma análise mais profunda das mesmas. Estas redes vêm sendo estudadas por diversos pesquisadores ao longo dos anos. Newman (2001) analisou as redes de Computação, Física e Medicina no período de 1995 a 1999, encontrando diversas características sociais comuns a essas redes. Outro trabalho relevante foi o de Barabási *et al* (2002), que analisou o comportamento temporal das redes de colaboração, concluindo que as métricas que caracterizam as mesmas evoluem de acordo com o tempo.

À medida que aumenta o interesse de pesquisa neste tipo de rede, observa-se um crescimento da aplicação da metodologia de ARS para compreender a dinâmica das relações de cooperação entre os pesquisadores. O foco principal da ARS é a análise das relações que os indivíduos estabelecem entre si, não se preocupando com os atributos individuais dos mesmos (Marteleto, 2010).

Neste sentido, a ARS dispõe de um conjunto de métricas que auxiliam na compreensão da dinâmica dos relacionamentos existentes nas redes estudadas. Dentre elas, destacam-se as métricas de Centralidade (*Centrality*), que permitem a avaliação da importância de um nó dentro da rede.

Neste trabalho serão utilizadas três métricas de centralidade propostas por Freeman (1978), a saber: Centralidade de Grau Normalizada, Centralidade de Intermediação Normalizada e Centralidade de Proximidade Normalizada. Além das métricas de centralidade, serão utilizadas também as métricas de Autoridade e Grau Ponderado. As definições das mesmas podem ser observadas a seguir:

A. Centralidade de Grau Normalizada (*Normalized Degree Centrality*): representa o número de ligações que um nó possui, dividido pelo número de ligações possíveis. A centralidade de grau de um nó *i* é expressa pela Equação 1:

$$C_G(i) = \frac{\sum_{j=1}^{n} aij}{n-1} \tag{1}$$

onde *aij* indica se existe ligação entre o nó *i* e o nó *j* (se existir, então *aij* = 1, caso contrário, *aij* = 0) e *n* representa o número de nós dentro da rede. Esta métrica é usada para identificar os nós que têm o maior número de ligações na rede (Chelmis e Prasanna, 2011).

B. Centralidade de Intermediação Normalizada (*Normalized Betweenness Centrality*): representa a quantidade de vezes que um determinado nó aparece no caminho geodésico entre dois nós da rede, sendo expressa pela Equação 2:

$$C_I(i) = \frac{\sum_{j,k \wedge i \neq j \neq k} \frac{g_{jik}}{g_{jk}}}{\frac{(n-1)(n-2)}{2}} \tag{2}$$

onde *n* é o número de nós, g_{jk} é o número de caminhos mais curtos do nó j para o nó k, e g_{jik} é o número de caminhos mais curtos de nó j para o nó k que passam pelo nó i. Esta métrica permite analisar o potencial de comunicação de um ator dentro da rede (Silva *et al* 2008).

C. Centralidade de Proximidade Normalizada (*Normalized Closeness Centrality*): mede o comprimento médio dos caminhos mais curtos de um vértice para cada um dos outros vértices de um grafo. A centralidade de proximidade de um vértice i é calculada pela Equação 3.

$$C_C(i) = \frac{n-1}{\sum_{j=1}^{n} eij} \tag{3}$$

onde n é o número de nós e *eij* é o número de arestas existentes no caminho mais curto do nó *i* para o nó *j*. Esta métrica indica a capacidade de alcance de um nó dentro da rede (Chelmis e Prasanna, 2011).

D. Grau Ponderado (*Weighted Degree*): É definida por Abbasi e Altmann (2011) como a soma de todos os pesos das arestas ligadas a um nó, sendo expressa pela Equação 4.

$$G_P(i) = \sum_{j=1}^{n} w_{ij} \tag{4}$$

onde *n* é o número de nós, w_{ij} representa o peso da aresta entre o nó i e o nó j, ou seja, representa a quantidade de vezes que os dois nós se relacionaram. Tal métrica evidencia a força das relações entre os pesquisadores.

E. Autoridade (*Authority*): o valor de autoridade é calculado somando-se a quantidade de *hubs* com o qual o nó em questão está conectado. Seu cálculo é feito através do Algoritmo HITS, proposto por Kleinberg (1998). A Equação 5 descreve como se obtém a autoridade de um nó dentro da rede, onde *h* representa o número de hubs.

$$A(i) = \sum_{j=1}^{n} h(j) \quad (5)$$

No contexto das redes de colaboração, as métricas mencionadas permitem entender a influência de um pesquisador dentro de sua rede social de pesquisa. Apesar de existirem diversas outras métricas de ARS, os conceitos previamente descritos serão suficientes para a compreensão dos resultados que serão apresentados. As escolhas destas métricas, em detrimento de outras métricas utilizadas nos trabalhos relacionados, se deve ao fato das mesmas terem forte ligação com o desempenho de um nó dentro da rede e serem utilizadas para medir a importância de um nó individual, como foi revelado nos trabalhos de Freeman (1978) e Abassi e Altmann (2011).

3.2. Relevância Científica

A relevância científica de um pesquisador tem sido comumente mensurada através de métricas de produção e impacto. As medidas de produção consideram o número de artigos publicados por um pesquisador durante sua carreira acadêmica, enquanto as métricas de impacto avaliam o quanto a produção do cientista teve importância para sua área de pesquisa (Wainer e Vieira, 2013).

Desta forma, há uma crescente tendência em se avaliar pesquisadores por esses indicadores bibliométricos. O índice mais utilizado para essa finalidade, o *h-index*, foi proposto por Hirsch em 2005, e é definido da seguinte forma: "um pesquisador possui um índice h se h dos seus Np artigos tem pelo menos h citações cada e os outros (Np-h) não tem mais do que h citações cada." Em outras palavras, um cientista tem um índice h = 10, se tiver 10 artigos publicados e cada um deles tiver sido citado pelo menos 10 vezes por outros trabalhos.

Essa métrica é a mais utilizada por serviços de contagem de citações, como o *Google Scholar*[2], para medir a relevância acadêmica de pesquisadores. Através dela tenta-se identificar tanto o volume de publicação como o impacto da produção de um pesquisador na sua área de atuação.

Além do *h-index*, existem outras métricas, a exemplo do *g-index* (Egghe, 2006) , *RP-index* e *CP-index* (Altmann *et al,* 2009), que também avaliam o desempenho de um pesquisador de forma análoga, porém acrescentando novos parâmetros que buscam aperfeiçoar os valores estimados pelo *h-index*.

4. Metodologia

4.1. Seleção da amostra

Dentre os 404 pesquisadores apontados em setembro de 2013 pelo site[3] do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) como bolsistas de

---

[2] Disponível em: http://scholar.google.com/citations

[3] Disponível em: http://efomento.cnpq.br/efomento/distribuicaoGeografica/distribuicaoGeografica.do?metodo=apresentar

produtividade na área de Ciência da Computação, foram sorteados aleatoriamente 202 pesquisadores para fazer parte do conjunto de treinamento. Deste conjunto, foi preciso excluir 28 pesquisadores cujo *h-index* não foi identificado com precisão pelo serviço de contagem de citações do Google Scholar, devido à problemas de ambiguidade no seu nome. Assim, o conjunto de treinamento final foi composto por 174 pesquisadores. A escolha de pesquisadores PQs se justifica pelo fato de serem reconhecidos como casos de sucessos entre os membros de sua área. Os outros 202 pesquisadores integram o conjunto de validação que será utilizado em trabalhos futuros.

4.2. Obtenção dos Dados

As redes de colaboração utilizadas nesta pesquisa foram geradas a partir das informações extraídas dos Currículos Lattes dos pesquisadores selecionados na amostra. A ferramenta *LattesCrawler*[4] foi utilizada para extrair as relações de colaboração de cada pesquisador, representadas por co-autoria em artigos publicados em periódicos e conferências, orientações de mestrado e/ou doutorado e participações em banca de mestrado e/ou doutorado e em projetos de pesquisa. A ferramenta mencionada gerou um arquivo *gml (Graph Modeling Language),* representando cada rede individual extraída, com o intuito de possibilitar a visualização das mesmas através da ferramenta *Gephi*[5]. Na Figura 1, é possível observar um exemplo de uma destas redes de colaboração, representada através de um grafo, extraída de um Currículo Lattes. Os vértices representam os pesquisadores e as arestas representam as colaborações científicas.

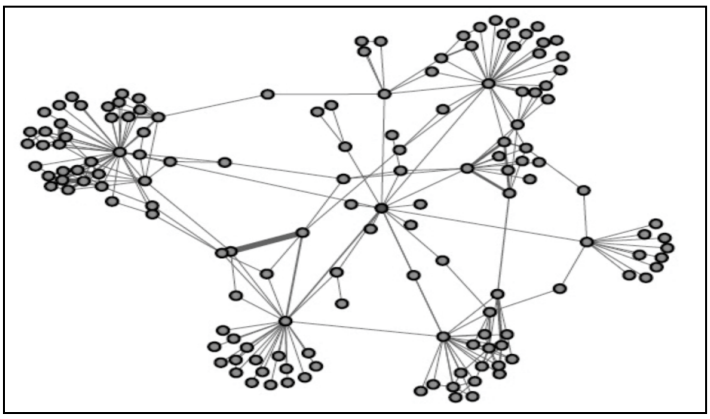

Figura 1. Rede de colaboração científica extraída do Currículo Lattes

4.3 Aplicação das Métricas de ARS

Para cada pesquisador selecionado na amostra foram calculados valores para as seguintes métricas de ARS: Centralidade de Grau Normalizada, Centralidade de Intermediação Normalizada, Centralidade de Proximidade Normalizada, Grau Ponderado e Autoridade. Os cálculos dessas métricas foram feitos com o auxílio da ferramenta *Gephi*, escolhida pelo fato de ser uma ferramenta gratuita e de atender às

[4] Disponível em: https://github.com/marcilioLemos/LABSNA/tree/master/LattesCrawler/src/br/ufpb/ci/labsna/lattescrawler
[5] Disponível em: https://gephi.org/

necessidades desta pesquisa. Para fins de padronização, todas as métricas calculadas foram arredondadas para três casas decimais depois da vírgula.

Já para a coleta do *h-index* dos pesquisadores estudados utilizou-se a ferramenta *Publish or Perish* (Harzing, 2007), que permite extrair métricas relativas ao impacto de produção de um pesquisador, utilizando como fonte de dados o Google Scholar, que segundo Bar-Ilan (2008) é o serviço de contagem de citações que melhor retrata o *h-index* dos pesquisadores de Ciência da Computação, pois considera as publicações em conferências e não somente em periódicos.

4.4 Metodologia para a Análise Estatística

Antes do tratamento estatístico das variáveis coletadas na pesquisa, realizou-se um teste W de Shapiro-Wilk, com nível de significância de 5%, para avaliar a normalidade dos dados. Este teste investiga se uma amostra aleatória provém de uma distribuição normal (Scudino, 2008). Diante disso, constatou-se que apenas uma das seis variáveis (cinco métricas de ARS e o *h-index*) estudadas possuía distribuição normal.

Os resultados desse teste podem ser vistos na Tabela 1, que apresenta na primeira coluna todas as variáveis analisadas e nas colunas subsequentes os testes estatísticos efetuados segundo o teste de Shapiro-Wilk.

Analisando a Tabela 1 verifica-se que a maioria das variáveis estudadas não apresenta distribuição normal, já que apenas a Centralidade de Intermediação alcançou um nível de significância superior a 0,05, que foi o nível estabelecido neste estudo.

Uma vez que, na maioria das variáveis, não foi verificado o pressuposto de normalidade, exigido para a utilização de técnicas estatísticas paramétricas, utilizou-se uma medida de correlação não-paramétrica, o Coeficiente de Correlação de Postos de Spearman (ρ)[6]. Este coeficiente não é sensível a assimetrias na distribuição nem à presença de *outliers*[7], não exigindo, portanto, que os dados provenham de populações normais (Pestana e Gageiro, 2000). Uma fórmula simplificada para entender como é feito o cálculo desse coeficiente é vista na Equação 6.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \qquad (6)$$

onde n é o número de pares (x_i, y_i) e d_i = (postos de x_i dentre os valores de x)- (postos de y_i dentre os valores de y).

Para testar a significância da correlação dos dados utilizou-se o teste da hipótese nula bicaudal, com um nível de significância de 5%. Sendo assim, foram testadas as seguintes hipóteses:

- **H_0:** $\rho = 0$ (não existe correlação entre as variáveis)
- **H_1:** $\rho \neq 0$ (hipótese alternativa)

---

[6] O Coeficiente de Correlação de Postos de Spearman mede a intensidade da relação entre variáveis ordinais, utilizando os valores de ordem das observações em vez de seu valor observado.

[7] Observações que apresentam um grande afastamento das restantes ou são inconsistentes com elas.

Por fim, o tratamento dos dados referentes foi feito utilizando o programa estatístico *Statistical Package for Social Sciences* (SPSS), versão 22.

Tabela 1 – Teste de Normalidade de Shapiro-Wilk

	Shapiro-Wilk		
	Estatística	df	Sig.
h_index	0,949	174	0,000
Centralidade de Grau N.	0,787	174	0,000
Centralidade de Intermediação N.	0,985	174	0,065
Centralidade de Proximidade N.	0,735	174	0,000
Grau Ponderado	0,841	174	0,000
Autoridade	0,554	174	0,000

5. Resultados e Discussões

Para responder à questão de pesquisa elencada no início desse trabalho, foi necessário realizar testes de associações entre as variáveis coletadas. Para utilizar o método de correlação correto, realizou-se um teste de normalidade com as variáveis envolvidas no estudo. As correlações entre essas variáveis foram calculadas utilizando o Coeficiente de Correlação de Postos de Spearman, que é o mais adequado para variáveis que não apresentam distribuição normal. Os resultados obtidos com os testes de correlação podem ser observados na Tabela 2. Nela, é possível verificar as cinco métricas de ARS dispostas na primeira coluna e na segunda coluna observa-se o coeficiente de correlação destas com o *h-index*, além do nível de significância da correlação (Sig.) e a quantidade da amostra (N).

Tabela 2 – Matriz de Coeficientes de Correlação de Postos de Spearman (rhô de Spearman)

VARIÁVEIS[a]		*H_INDEX*
CGN	Coeficiente de Correlação	-0,061
	Sig. (2 extremidades)	0,426
	N	174
CIN	Coeficiente de Correlação	0,223*
	Sig. (2 extremidades)	0,003
	N	174
CPN	Coeficiente de Correlação	-0,061
	Sig. (2 extremidades)	0,426
	N	174
GP	Coeficiente de Correlação	0,485*
	Sig. (2 extremidades)	0,000
	N	174
AUT	Coeficiente de Correlação	-0,151*
	Sig. (2 extremidades)	0,047
	N	174

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

[a].CGN; Centralidade de Grau Normalizada; CIN: Centralidade de Intermediação Normalizada; CPN: Centralidade de Proximidade Normalizada; GP: Grau Ponderado; AUT: Autoridade.

Os resultados apresentados na Tabela 2 apontam que a métrica de centralidade que apresenta correlação positiva significativa (cc = 0,223) com o *h-index*, considerando um nível de significância de 5%, é a métrica de Intermediação Normalizada. Isso sugere que o desempenho de um pesquisador tende a crescer à medida que ele assume um papel de intermediador dentro da rede, possibilitando a comunicação entre grupos distintos.

Ainda de acordo com os resultados da Tabela 2, verifica-se que o Grau Ponderado também apresenta uma correlação positiva significativa (cc = 0,485) com o índice de desempenho (*h-index*), com um nível de significância de 5%. Sendo assim, pode-se afirmar que um pesquisador com um alto grau ponderado tem um desempenho científico maior. Isso pode ser explicado pela teoria proposta por [8], que afirma que pesquisadores que mantém laços fortes tendem a ter melhor desempenho do que pesquisadores que estabelecem laços fracos com seus colaboradores. Em outras palavras, cientistas que colaboram frequentemente com os mesmos pesquisadores demonstram maior desempenho do que aqueles que colaboram com muitos pesquisadores diferentes.

Outra métrica que apresentou correlação significativa a um nível de 5% com o *h-index*, porém negativa, foi a métrica de Autoridade (cc = -0,151). Esse fato corrobora a afirmação anterior, já que pesquisadores com métricas de autoridade alta colaboram com o maior número de *hubs* dentro da rede, estabelecendo relações com diversos colaboradores diferentes, apresentando assim muitos laços fracos.

Como as métricas normalizadas de Proximidade e Grau não apresentaram correlação significativa, entende-se, que nesse contexto de pesquisa, as mesmas não têm impacto significativo no desempenho de um pesquisador.

Dessa forma, pode-se inferir que o desempenho acadêmico de um pesquisador está diretamente associado com seu poder de comunicação dentro da rede, assim como, tem forte associação com o peso das relações que ele estabelece com seus colaboradores.

Cabe ressaltar que embora os coeficientes de correlação encontrados sejam classificados como fracos e moderados em relação a magnitude da correlação, a utilização do teste da hipótese nula bicaudal permitiu considerar os resultados apresentados relevantes, tendo em vista que a significância estatística foi garantida para um nível de 5%.

6. Considerações Finais

O processo de produção científica está cada vez mais pautado na ideia de colaboração. Entretanto, os estudos que relacionam os padrões de colaboração e o desempenho científico de pesquisadores ainda são escassos. Nesse sentido, acredita-se que ainda há muitas questões em aberto nesta área de pesquisa.

Diante deste contexto, este estudo analisou as redes de colaboração científica de pesquisadores classificados como bolsistas de produtividade na área de Ciência da Computação e extraiu métricas referentes às relações de colaboração com o intuito de investigar uma possível correlação destas, com um índice de relevância científica desses pesquisadores.

Os resultados preliminares demonstraram que o *h-index* desses pesquisadores apresenta uma correlação significativa positiva com a Centralidade de Intermediação Normalizada e com o Grau Ponderado e uma correlação significativa negativa com a métrica de Autoridade. Dessa forma, constatou-se que pesquisadores que assumem um papel de intermediador dentro da rede e mantém relações de colaboração fortes, ou seja, colaboram várias vezes com o mesmo pesquisador, tendem a ter desempenhos mais altos[8] do que aqueles que mantêm relacionamentos fracos.

Cabe ressaltar que as correlações entre as variáveis não indica necessariamente relação de causalidade, apenas indicam a existência de uma relação entre as variáveis na qual os valores das mesmas alteram-se simultaneamente. Assim, não é possível afirmar qual variável produz efeito sobre a outra. Essa limitação pode ser contornada com um Teste de Causalidade de Granger (Granger, 1969).

Apesar dos resultados obtidos até o momento sugerirem associações entre algumas métricas de ARS e uma métrica de relevância acadêmica de pesquisadores, não foi possível a construção de um Modelo de Regressão com capacidade de prever o sucesso acadêmico desses pesquisadores através das métricas encontradas, pois os dados ainda são insuficientes.

Diante deste fato, almeja-se como trabalho futuro, aplicar a metodologia supracitada em um conjunto de dados mais amplo que contenha tantos pesquisadores PQ's como pesquisadores não PQ's da área de Ciência da Computação. Isso permitirá uma análise mais aprofundada dos resultados o que consequentemente tornará a pesquisa mais robusta.

Além disso, pretende-se realizar experimentos com outros índices utilizados para auferir a relevância científica de um pesquisador, a exemplo do *g-index*, do *RP-index* e do *CP-index*, com o intuito de verificar se as correlações existentes apresentam semelhanças ou diferem entre si.

Assim, com a incorporação dos resultados dessa nova etapa do estudo pretende-se utilizar técnicas de Regressão Logística, com intuito de conceber um modelo matemático preditivo capaz de apontar qual será a relevância científica de um pesquisador no futuro. Para tanto, haverá a incorporação de aspectos temporais, que permitam correlacionar as variáveis ao longo dos anos.

Após a obtenção da fórmula de Regressão, serão feitos experimentos utilizando as redes dos pesquisadores do conjunto de validação, de modo a verificar a acurácia e a precisão do modelo. Para realização desses experimentos pretende-se regredir as redes dos pesquisadores do conjunto em questão para 5 e 10 anos atrás, de modo a comparar os resultados obtidos através da aplicação do modelo matemático com os resultados das redes atuais.

Ainda como trabalho futuro, sugere-se um estudo comparativo com outras áreas de pesquisa de modo a verificar se as associações das métricas de ARS com os índices

---

[8] Pesquisadores com alto desempenho se caracterizam pelo fato de possuírem um *h-index* elevado, ou seja, publicaram muitos artigos com grande impacto na sua área.

de desempenho científico são semelhantes independente da área de atuação dos pesquisadores ou se apresentam diferenças significativas.

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 of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures", Journal of Informetrics.

Altmann , J., Abbasi, A., Hwang, J. (2009) "Evaluating the productivity of researchers and their communities: The RP-index and the CP-index", International Journal of Computer Science and Applications, vol. 6, no. 2, pp. 104–118.

Harzing, A.W. (2007) "Publish or Perish", available from <http://www.harzing.com/pop.htm>.

Araújo, E. B., Moreira, A. A., Furtado, V., Pequeno, T. H. C., Andrade JR, J. S. (2013) "Collaboration networks from a large CV database: dynamics, topology and bonus impact", arXiv:1310.1366.

Barabási A. L., Jeong H., Néda Z., Ravasz E., Schubert A., Vicsek T. (2002) "Evolution of the social network of scientific collaborations", Physica A 311(3–4):590–614.

Bar-Ilan, J. (2008) "Which *h-index*? A comparison of WoS, Scopus and Google Scholar", Scientometrics, v. 74, n. 2, p. 257-271, 2008.

Chelmis C., Prasanna V. K. (2011) "Social networking analysis: A state of the art and the effect of semantics", In: Proceedings of the IEEE Third International Conference on Social Computing (SocialCom).

Egghe, L. (2006) "Theory and practise of the g-index", Scientometrics, vol. 69, no. 1, pp. 131-152.

Faria, F. F., Albuquerque, R. P.; Oliveira, J. (2012) "Análise da Produtividade da Rede Social de Computação do Brasil", In: Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2012, Curitiba. Anais do 32º. Congresso da SBC.

Freeman, L. C. (1978) "Centrality in social networks: Conceptual clarification", Social Networks, 1, 215–239.

Freire, V. P., Figueiredo, D. R. (2011) "Ranking in collaboration networks using a group based metric", J. Braz. Comp. Soc. 17 (4): 255–266, 2011.

Granger, C.W.J. (1969) "Investigating causal relations by econometric models and cross-spectral models". Econometrica 34, 541-51

Guerra, G. N. (2012) "Modelo de reputação e ontologia aplicados à rede social científica do observeunb". Master's thesis, Universidade de Brasília - UnB.

Newman, M. E. J. (2001) "The structure of scientific collaboration networks". Proc Natl Acad Sci USA 98(2):404–409.

Hirsch, J. (2005) “An index to quantify an individual’s scientific research output”, Proceedings of the National Academy of Sciences of the United states of America, National Academy of Sciences, v. 102, n. 46, p. 16569.

Kleinberg, J. (1998) “Authoritative sources in a hyperlinked environment”, Proc. of the 9th ACM-SIAM Symposium on Discrete Algorithms, 668--677.

Maia, M. F. S., Caregnato, S. E. (2008) “Co-autoria como indicador de redes de colaboração científica”, Perspectivas em Ciência da Informação, Belo Horizonte, v. 13, n. 2, p. 18-31, maio/ago. 2008.

Marteleto, R. M. (2010) “Redes sociais, mediação e apropriação de informações: situando campos, objetos e conceitos na pesquisa em Ciência da Informação”, Tendências da Pesquisa Brasileira em Ciência da Informação, v.3, n.1, p.27-46.

Martins, D. L., Ferreira, S. M. S. P. (2013) “Mapeamento e avaliação da produção científica da Universidade de São Paulo com foco na estrutura e dinâmica de suas redes de colaboração científica”. Liinc em Revista, v. 9, n. 1.

Pestana, M. H., Gageiro, J. N. (2000) “Análise de Dados para Ciências Sociais: A complementaridade do SPSS”. 2. ed. Sílabo: Lisboa.

Scudino, P. A. (2008) “A Utilização de Alguns Testes Estatísticos para Análise da Variabilidade do Preço do Mel nos Municípios de Angra dos Reis e Mangaratiba”, Estado do Rio de Janeiro.

Silva, M. R., Ma, H.; Zeng, A.-P. (2008) “Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks”, Proceedings of the IEEE, 96:1411:1420.

Wainer, J., Vieira, P. (2013) “Avaliação de bolsas de produtividade em pesquisa do CNPq e medidas bibliométricas:correlações para todas as grandes areas”. Perspect. ciênc. inf. [online]. 2013, vol.18, n.2, pp. 60-78. ISSN 1413-9936. Disponível em: <http://dx.doi.org/10.1590/S1413- 99362013000200005>.