

Identificando a Reputação de Pesquisadores Usando um Modelo de Perfil Adaptativo

Cristiano R. Cervi^{1,2}, Renata Galante², José Palazzo M. de Oliveira²

¹Instituto de Ciências Exatas e Geociências – Universidade de Passo Fundo (UPF)
Caixa Postal 611 – 99.001-970 – Porto Alegre – RS – Brasil

²Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil

cervi@upf.br, {galante,palazzo}@inf.ufrgs.br

Abstract. *In this paper we address the problem of identifying the profile of researchers and measure their reputation in the area. We approached the process of specifying the profile using elements of the academic life of the researcher and the researcher classified by levels of reputation. For classification, some metrics were defined based on the balance of the trajectory of a researcher, involving the key elements of their curriculum. The elements do not only focus on production and citation of papers, but in the breadth of life science researcher's built over his career. The proposed model can be used in different contexts and different areas of knowledge, because it is extensive, flexible and adaptive.*

Resumo. *Neste trabalho abordamos o problema de identificar o perfil de pesquisadores e medir sua reputação na área de atuação. Abordamos o processo de especificação do perfil utilizando elementos da vida acadêmica do pesquisador, bem como classificamos o pesquisador por meio de níveis de reputação. Para a classificação, algumas métricas foram definidas baseadas no equilíbrio da trajetória de um pesquisador, envolvendo os principais elementos de seu currículo. Os elementos não focam apenas em produção e citação de artigos, mas sim na abrangência da vida científica do pesquisador construída ao longo de sua carreira. O modelo proposto neste trabalho pode ser utilizado em diferentes contextos e em diferentes áreas do conhecimento, pois é abrangente, flexível e adaptativo.*

1. Introdução

A área de modelagem de usuários (*User Modeling*) tem como objetivo descobrir conhecimento sobre o usuário em determinado assunto (*expert finding*) e de que forma representar esse conhecimento. Diversos sistemas e métodos (Fink e Kobsa, 2002; Middleton, Shadbolt e De Roure, 2004; Gauch et al. 2007; Maia, Almeida e Almeida, 2008) tem sido propostos para se modelar o perfil de um usuário, quer seja de forma explícita ou implícita e até mesmo da união das duas técnicas. A utilização desses sistemas e métodos se dá nas mais variadas áreas de aplicação, como por exemplo: (i) detecção de clientes em sites de vendas on-line; (ii) recomendação de produtos; (iii) recomendação de especialistas para participarem de defesas de mestrado e de doutorado; (iv) detecção de pesquisadores para formarem rede de colaboração científica; (v) organização de comitê de programa de conferência científica; (vi) organização de corpo editorial e revisores de periódicos, dentre outros.

A área de métricas científicas tem um papel importante na comunidade acadêmica, pois podem auxiliar no processo de medição da qualidade da produção científica e na identificação de especialistas em determinada área. As métricas recentes se baseiam fortemente nas citações de artigos de pesquisadores (Hirsch, 2005; Jin et al., 2007; Jin, 2007; Hirsch, 2009; Zang, 2009; Jin, 2007). Entretanto, essas formas de avaliação não consideram a trajetória do pesquisador, nem mesmo o cenário onde ele está inserido. As citações a artigos deveriam ser apenas um dos elementos a serem analisados e não o elemento exclusivo.

No contexto acadêmico, a identificação de um perfil pode ocorrer por meio de um processo de análise da trajetória de toda a carreira de um pesquisador. Tal processo envolve não somente aspectos relacionados a produção científica, como artigos e livros publicados, mas também por outros elementos inerentes a atividade de um pesquisador, como por exemplo: orientações de trabalhos de mestrado e de doutorado; participação em defesas de mestrado e de doutorado; trabalhos apresentados em conferências; participação em projetos de pesquisa, dentre outros.

Diante desse contexto, existe um campo de pesquisa em aberto, pois é necessária uma métrica que englobe toda a vida do pesquisador e que proporcione uma análise abrangente. Nesse trabalho tratamos o problema de identificar e medir a reputação de um pesquisador em sua área de atuação. Isto não é uma tarefa simples, uma vez que pode envolver visões diferentes, quando pesquisadores atuam em áreas de pesquisa diferentes, porém a principal vantagem é tornar nossa proposta adaptativa. Primeiramente é necessário definir quais dados são utilizados para avaliar a reputação de um pesquisador, como representar estes dados em um modelo (perfil) e onde encontrá-los. Após, é necessário definir uma métrica para se chegar a um índice para a reputação.

Este trabalho tem como objetivo especificar um modelo de perfil de pesquisador com dados obtidos mediante análise de sua trajetória científica e classificá-lo em um nível de reputação. O processo de modelagem do perfil envolve a identificação de quais dados são relevantes para a definição do perfil e os apresenta por meio de uma ontologia de domínio. O processo de classificação do pesquisador o enquadra em um nível de reputação definido através de um índice gerado por uma nova métrica proposta neste artigo. O modelo de perfil e a métrica foram definidos baseados no equilíbrio da trajetória de um pesquisador, envolvendo os principais elementos de seu currículo. Cabe ressaltar que tanto o perfil quanto as métricas não focam apenas em publicação de artigos ou citação, mas sim na abrangência da vida científica do pesquisador construída ao longo de sua carreira.

O restante deste artigo está dividido como segue. A seção 2 descreve os trabalhos relacionados. O perfil proposto através de uma ontologia de domínio é apresentado na seção 3. A seção 4 especifica a métrica para identificar a reputação e classificar os pesquisadores em níveis. Experimentos e análises de resultados são apresentados e discutidos na seção 5. Por fim, a seção 6 apresenta as considerações finais e perspectivas de trabalhos futuros.

2. Trabalhos Relacionados

Diversos trabalhos existentes na literatura abordam questões sobre identificação de perfis acadêmicos, produção científica de pesquisadores, identificação de especialistas baseados em dados científicos ou simplesmente análise de produção para obter estatísticas relevantes. Como nosso trabalho envolve tanto a modelagem de um perfil quanto o seu uso para o cálculo da métrica de reputação, esta seção apresenta primeiramente os trabalhos relacionados a modelagem de usuário e em seguida os trabalhos relacionados com métricas de reputação. A

seção é finalizada com uma discussão sobre o relacionamento dos dois temas de pesquisa e as contribuições da nossa proposta.

2.1. Modelagem de usuários

Os trabalhos de Tang et al. (2008a), Tang et al. (2008b) e Tang, Zhang e Yao (2007), por exemplo, utilizam dados da produção científica de pesquisadores no contexto de redes sociais acadêmicas para analisar as relações que pesquisadores possuem entre si. Redes de colaboração científica entre pesquisadores são apresentadas e discutidas em Liu et al (2005), Chen et al. (2008), Huang et al. (2008) e Menezes et al. (2009). Basicamente, os trabalhos analisam redes de co-autoria desses pesquisadores através da definição de grafos e suas relações. Os dados são obtidos da DBLP (DBLP 2010).

Outros trabalhos utilizam dados de produção científica para encontrar especialistas em determinada área do conhecimento (ZHANG, TANG e LI, 2007; MIMNO e MCCALLUM, 2007; FU et al., 2007; LI et al., 2007). Esses trabalhos utilizam informações de páginas pessoais dos pesquisadores e de diferentes bibliotecas digitais, como DBLP (DBLP 2010) e CiteSeer (CITeseer 2010).

Em uma área de aplicação de mercado, o trabalho de Park e Chang (2009) apresenta um modelo de perfil de consumidores baseado em comportamento individual ou em grupo, construídos por meio do comportamento do usuário. Esse comportamento é identificado em sua navegação em páginas *web*, inserção de itens em carrinhos de compras (*basket insertion purchase*) e áreas de interesse.

A Microsoft, por meio do site Microsoft Academic Search¹, disponibiliza consultas sobre publicações, autores, conferências e periódicos, mediante um mecanismo de busca que coleta informações na *web* e armazena em um banco de dados. Sistemas semelhantes são utilizados em Tang et al. (2008a), Tang et al. (2008b) e na Digital Bibliography & Library Project (DBLP 2010). Tais sistemas permitem a visualização de dados de pesquisadores, bem como apresentam as redes de colaboração científica entre os pesquisadores.

2.2. Métricas Para Analisar Pesquisadores

O trabalho de Krapivin et al. (2009) afirma que a função de se utilizar métricas para avaliar a produção de pesquisadores está fundamentada em dois contextos: medir a qualidade da produção científica e medir a contribuição e a reputação relacionada a pesquisadores. Nesse sentido, os trabalhos de Small (1973), White & McCain (1998), Hirsch (2005), Jin et al. (2007) e Jin (2007), abordam temas como a citação para medir a reputação de pesquisadores.

A popularização do *PageRank* (Page, 1998) fez com que diversos trabalhos no âmbito acadêmico se utilizassem de seus algoritmos para a análise da reputação de pesquisadores. Trabalhos como os de Ding et al. (2009) e Chen et al. (2006) aplicam o *PageRank* a conjuntos de dados a fim de ordenar autores e artigos, respectivamente. O trabalho de Ding et al. (2009) ordena autores por meio de redes de co-citação. Já o trabalho de Chen et al. (2006) utiliza o conceito de *damping factor*, que se baseia no fato de que um leitor tende a procurar um caminho de ordem igual a dois links nas referências do artigo original antes de sair em busca de um artigo novo.

O algoritmo PaperRank (KRAPIVIN et al., 2009) modifica o *PageRank* original. A alteração considera, além das citações entre artigos, o *ranking* do artigo citante e a densidade de citações feitas pelo artigo citante. Outras métricas também foram desenvolvidas, como o

¹ Disponível em: <http://academic.research.microsoft.com>

h-index (HIRSCH, 2005), o AR-index (JIN, 2007) e o G-Index (TOL, 2008), todos calculando a importância individual de pesquisadores.

Dentre os trabalhos citados tanto para modelagem do perfil como para cálculo de reputação, nossa proposta se difere por especificar um modelo de perfil de pesquisadores flexível e adaptativo, que pode ser utilizado em diferentes áreas de aplicação. A adaptabilidade do modelo envolve desde a inserção e eliminação de categorias e elementos, bem como o ajuste de seus pesos. Outro ponto importante em nossa abordagem é que o modelo e a métrica se propõem a identificar a reputação de um pesquisador de forma abrangente, buscando um equilíbrio entre diferentes elementos que fazem parte da trajetória científica de um pesquisador, ao invés de levar em consideração apenas métricas como o H-index ou o G-index, que são baseados apenas em citações bibliográficas.

3. Modelo de Perfil Proposto

O objetivo desta seção é especificar o modelo de perfil do pesquisador. O modelo é baseado em elementos que constroem a trajetória de um pesquisador ao longo de sua carreira científica, facilmente identificados em seu currículo. Tais elementos não focam apenas em produção de artigos, mas sim na abrangência da vida científica do pesquisador.

O modelo de perfil proposto nesse trabalho foi definido através de uma ontologia que visa representar o comportamento do pesquisador, identificado por meio de um conjunto de elementos e de seus atributos. As próximas seções apresentam, respectivamente, as categorias definidas no modelo, o conjunto de elementos de cada categoria e suas representações na ontologia (Figura 1).

3.1. Categorias e Elementos

Em nosso estudo, identificamos diversos elementos importantes para a definição do perfil de um pesquisador. Alguns destes elementos são utilizados amplamente nas comunidades científicas das diversas áreas do conhecimento, conforme pode ser observado em trabalhos como os de Chen et al. (2007), Chen et al. (2008), Ding et al. (2009), Fu et al. (2007), Krapivin, Marchese e Cassati (2009), Li et al. (2007), Menezes et al. (2009), Tang, Zhang e Yao (2007), Tang et al. (2008a), Tang et al. (2008b) e Zhang, Tang e Li (2007). A Tabela 1 apresenta as categorias definidas, seus elementos e a sigla de cada um deles.

Category	Elements	Acronym
Identification (ID)	Education Degree	ED
Advisory (ADV)	Master Dissertation Advisor	MDA
	Phd Thesis Advisor	PTA
	Postdoctoral Advisor	PA
Examining Board (EB)	Participation in Examination Boards Master Dissertation	PEBMD
	Participation in Examination Boards Phd Thesis	PEBPT
Membership (MS)	Conference Committee Coordinator	CCC
	Conference Committee Member	CCM
	Editorial Board Member	EBM
	Reviewer of Journals	RJ
Production (PROD)	Articles in Scientific Journals	ASJ
	Book Chapter Published	BCP
	Books Published	BP
	Complete Work Published in Conference Proceedings	CWPCP
	H-Index	HI
	Network Co-autorship	NC
	Research Projects	RP
	Software With Register	SWR
Software Without Register	SWOR	

Tabela 1. Categorias, elementos e siglas que representam o perfil do pesquisador.

3.2. Ontologia

A ontologia para representar o perfil do pesquisador é dividida em cinco categorias: *Advisory*, *Examining Board*, *Identification*, *Membership* and *Production*. A divisão foi necessária para melhor visualizar o conjunto de elementos da trajetória do pesquisador.

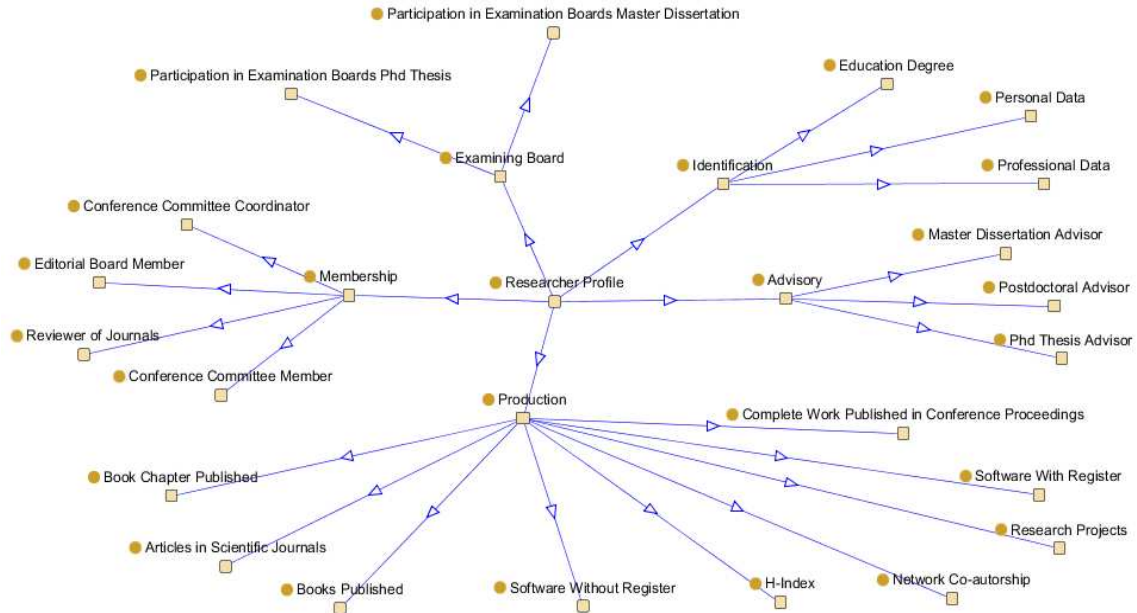


Figura. 1. Representação gráfica da ontologia.

O objetivo da ontologia apresentada na Figura 1 é aumentar a expressividade na representação, tornando mais clara a apresentação do modelo. A ontologia representa as cinco categorias e todos os seus elementos. Para uma melhor compreensão do modelo, apresentamos uma descrição de todos os elementos utilizados:

- *Identification*: por meio do elemento Education Degree, apresenta a titulação do pesquisador. As opções deste elemento são mestrado, doutorado ou pós-doutorado;
- *Advisory*: representa as orientações do pesquisador. Master Dissertation Advisor representa as orientações de mestrado. Phd Thesis Advisor representa as orientações de doutorado. Por fim, as orientações de pós doutorado são representadas pelo elemento Postdoctoral Advisor;
- *Examining Board*: especifica a participação em bancas do pesquisador. O elemento Participation in Examination Boards Master Dissertation representa a participação em defesas de mestrado. As participações em defesas de doutorado são representadas pelo elemento Participation in Examination Boards Phd Thesis;
- *Membership*: apresenta as conferências onde o pesquisador foi coordenador ou membro de comitê de programa, bem como se foi ou ainda é revisor de periódico ou membro de corpo editorial. O elemento Conference Committee Coordinator representa as conferências onde o pesquisador foi coordenador de comitê de programa. Conference Committee Member representa as conferências onde o pesquisador teve participação como revisor de trabalhos. O elemento Editorial Board Member diz respeito à atuação do pesquisador como membro de corpo editorial de periódico. Por fim, o elemento Reviewer of Journals apresenta os periódicos onde o pesquisador é ou foi revisor;

- *Production*: especifica a produção científica e outros elementos relacionados a produção científica do pesquisador. *Articles in Scientific Journals* representa os artigos publicados em periódicos. *Book Chapter Published* e *Books Published* representam, respectivamente, os capítulos de livros e os livros publicados pelo pesquisador. Os artigos publicados em conferências são representados pelo elemento *Complete Work Published in Conference Proceedings*. O elemento *H-Index* diz respeito às citações dos artigos do pesquisador por outros pesquisadores. *Network Co-authorship* está associado ao número de pesquisadores que possuem trabalhos publicados em conjunto com o pesquisador em questão. O elemento *Research Projects* representa o número de projetos de pesquisa do pesquisador. Os elementos *Software With Register* e *Software Without Register* representam os softwares em que o pesquisador teve participação. O primeiro refere-se aos softwares com registro e o segundo aos softwares sem registro.

4. Métrica Proposta Para Identificar a Reputação de Pesquisadores

O objetivo da métrica proposta neste trabalho é classificar o pesquisador em um nível de reputação definido através de um índice entre 1 e 5. Esta normalização foi proposta com o intuito de não distanciar com valores numéricos pesquisadores iniciantes, intermediários e experientes, bem como tornar mais compreensível a identificação da reputação. A métrica para identificar a reputação é definida por meio de um somatório entre os vários elementos que compõem o modelo de perfil do pesquisador. Para isto, são definidos pesos para as cinco categorias especificadas no modelo. O somatório dos pesos das categorias equivale a 100. Como dentro de cada categoria existem elementos, estes também possuem pesos individuais. O somatório dos pesos dos elementos é limitado ao peso máximo de cada categoria e seu somatório também está limitado ao valor 100.

4.1. Métrica

A seguir, apresentamos as fórmulas que definem a métrica para calcular a reputação de pesquisadores. Levando-se em consideração as Fórmulas 1, 2, 3, 4, 5 e 6 apresentadas a seguir, algumas observações devem ser destacadas:

- a letra "R" após a abreviação da categoria significa "Researcher", fazendo referência ao pesquisador que se deseja descobrir a reputação;
- a letra "W" em frente ao nome dos elementos do modelo significa "Weight", ou seja, o peso daquele elemento;
- a sigla "NL" significa "Number of Levels" e representa o número de níveis desejados no modelo. Na seção 5.3, em nossos experimentos, definimos 5 níveis;
- a sigla "MVR" significa "Maximum Value of Range", ou seja, o máximo valor do intervalo. Este valor representa o número mais expressivo de cada elemento no modelo. Na seção 5.3, em nossos experimentos, definimos o valor mais expressivo de cada elemento levando-se em consideração um grupo de 52 pesquisadores.

A Fórmula 1 apresenta os elementos da categoria *Identification (ID)* que fazem referência ao nível de titulação do pesquisador, sendo considerado no cálculo apenas um dos valores, ou seja, a titulação máxima do pesquisador: ou *Mestrado*, ou *Doutorado* ou *Pós-doutorado*.

$$ID_{(R)} = (ED * WED) \quad (1)$$

Os elementos da categoria *Advisory* (ADV) são apresentados na Fórmula 2 e correspondem a *Master Dissertation Advisor*, *Phd Thesis Advisor* e *Postdoctoral Advisor*.

$$ADV_{(R)} = ((MDA * (WMDA / (NL - 1))) / (MVR / (NL - 1))) + ((PTA * (WPPTA / (NL - 1))) / (MVR / (NL - 1))) + ((PA * (WPA / (NL - 1))) / (MVR / (NL - 1))) \quad (2)$$

A Fórmula 3 apresenta os elementos da categoria *Examination Board* (EB), onde dois elementos são definidos: *Participation in Examination Boards Master Dissertation* and *Participation in Examination Boards Phd Thesis*.

$$EB_{(R)} = ((PEBMD * (WPEBMD / (NL - 1))) / (MVR / (NL - 1))) + ((PEBPT * (WPEBPT / (NL - 1))) / (MVR / (NL - 1))) \quad (3)$$

A Fórmula 4 apresenta os elementos da categoria *Membership* (MS). Esta categoria contempla quatro elementos: *Conference Committee Coordinator*, *Conference Committee Member*, *Editorial Board Member* and *Reviewer of Journals*.

$$MS_{(R)} = ((RJ * (WRJ / (NL - 1))) / (MVR / (NL - 1))) + ((EBM * (WEBM / (NL - 1))) / (MVR / (NL - 1))) + ((CCM * (WCCM / (NL - 1))) / (MVR / (NL - 1))) + ((CCC * (WCCC / (NL - 1))) / (MVR / (NL - 1))) \quad (4)$$

Os elementos da categoria *Production* (PROD) são apresentados na Fórmula 5. No total, são nove elementos: *Articles in Scientific Journals*, *Book Chapter Published*, *Books Published*, *Complete Work Published in Conference Proceedings*, *H-Index*, *Network Co-authorship*, *Research Projects*, *Software With Register* and *Software Without Register*.

$$PROD_{(R)} = ((HI * (WHI / (NL - 1))) / (MVR / (NL - 1))) + ((CWPCP * (WCWPCP / (NL - 1))) / (MVR / (NL - 1))) + ((SWIR * (WSWIR / (NL - 1))) / (MVR / (NL - 1))) + ((SWOR * (WSWOR / (NL - 1))) / (MVR / (NL - 1))) + ((RP * (WRP / (NL - 1))) / (MVR / (NL - 1))) + ((ASJ * (WASJ / (NL - 1))) / (MVR / (NL - 1))) + ((BP * (WBP / (NL - 1))) / (MVR / (NL - 1))) + ((BCP * (WBCP / (NL - 1))) / (MVR / (NL - 1))) + ((NC * (WNC / (NL - 1))) / (MVR / (NL - 1))) \quad (5)$$

Finalmente, a Fórmula 6 apresenta o índice de reputação do pesquisador (Rep-Index) envolvendo o somatório das cinco categorias definidas no modelo.

$$Rep-Index_{(R)} = \sum ID + \sum MS + \sum EB + \sum ADV + \sum PROD \quad (6)$$

O cálculo final do índice de reputação do pesquisador é obtido pelo somatório das cinco categorias: *Identification* (ID), *Membership* (MS), *Advisory* (ADV), *Examination Board* (EB) e *Production* (PROD). Já o resultado de cada categoria é obtido por meio da multiplicação do número de ocorrências de determinado elemento pelo peso do próprio elemento dividido pelo número de intervalos menos um. O resultado desta operação é dividido pelo valor máximo do intervalo, também dividido pelo número de intervalos menos um.

4.2. Níveis de Reputação

O modelo proposto trabalha com níveis de reputação adaptativos. Dependendo do contexto ou da área de utilização, pode-se definir quantos níveis de reputação desejar. Nos experimentos, que são apresentados na seção 5, trabalhamos com a reputação do pesquisador em cinco níveis. A partir do valor máximo de cada elemento (com exceção da categoria *Identification*) os cinco níveis foram gerados. Os níveis são a classificação do pesquisador e identificam sua reputação, calculada por meio do modelo proposto.

A escala de valores do índice de reputação varia de 0 a 100, onde um valor mais próximo de 0 representa uma baixa reputação (geralmente para pesquisadores em início de carreira) e um valor próximo de 100 representa uma alta reputação (geralmente pesquisadores mais experientes, com uma longa trajetória científica).

Como o modelo é abrangente, flexível e adaptativo, ele pode ser utilizado em diferentes áreas do conhecimento e em diferentes contextos, uma vez que os critérios de avaliação ou necessidades podem ser distintos. Da mesma forma, os pesos dos elementos do modelo ou ainda os próprios elementos, também podem ser adaptados ou suprimidos, se não são importantes dentro do contexto de utilização.

5. Experimentos e Análise de Resultados

O objetivo dos experimentos é validar o modelo de perfil especificado na seção 3 por meio das métricas definidas na seção 4. Os experimentos classificam o pesquisador em níveis que identificam sua reputação. Esta seção, além de apresentar os experimentos, também discute os resultados obtidos.

5.1. Aquisição de Dados

Para a validação do modelo foram obtidos dados do currículo de 52 pesquisadores brasileiros da área de Ciência da Computação. Estes pesquisadores estão vinculados a um dos Programas de Pós-graduação em Computação que possui conceito 6 (em uma escala de 3 a 7) na avaliação realizada pela Capes². A escolha por este grupo de pesquisadores foi motivada por agregar pesquisadores de níveis iniciante, intermediário e experiente, bem como representar um programa de pós-graduação de qualidade e com reconhecida inserção internacional.

Os dados foram coletados da Plataforma Lattes³, da DBLP, do Microsoft Academic Search e do Arnetminer⁴. Destaca-se que não foi necessário fazer tratamento de desambiguação de nomes ou de publicações, uma vez que o Arnetminer, a DBLP e o Microsoft Academic Search já incorporam tal recurso. Com relação ao Lattes, a extração foi semiautomatizada e realizada individualmente para cada pesquisador. O resultado da extração individual foi a quantificação, para cada pesquisador, dos elementos do modelo que possuíam relação com o Lattes.

5.2. Métricas de Comparação

Por estarem sendo amplamente utilizados na comunidade acadêmica da área de computação e por terem sido pioneiros no processo de classificação de pesquisadores, utilizamos o H-Index e o G-Index como métricas a serem comparadas com nossa proposta.

² *Capes*: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

³ *Plataforma Lattes*: Base de dados de currículos de pesquisadores brasileiros.

⁴ *Disponível em*: <http://arnetminer.org/>

Destaca-se que o H-Index e o G-Index utilizam apenas citações de artigos. Nossa abordagem, por envolver diversos elementos da trajetória científica de um pesquisador, não pode ser comparável diretamente a estes índices. Nos resultados apresentados na seção 5.4 calculamos o valor de correlação entre eles e o nosso índice.

5.3. Metodologia

Para a realização dos experimentos foram utilizadas as categorias e elementos definidos no modelo apresentado na seção 3. Para cada categoria foi definido um peso, onde o somatório destes pesos é igual a 100. Da mesma forma, para cada elemento foi definido um peso, onde o somatório destes pesos também é igual a 100.

Os pesos das categorias e os pesos dos elementos foram definidos levando-se em consideração as particularidades da área de Ciência da Computação, como por exemplo, a valorização de trabalhos publicados em conferências. Em algumas outras áreas, tais elementos poderiam ser desconsiderados ou ainda pouco valorizados, uma vez que outros elementos podem ser considerados fundamentais.

Os valores dos intervalos foram definidos por meio da análise do currículo dos 52 pesquisadores (seção 5.1). Para cada pesquisador foi analisado todo o conjunto de elementos do modelo, identificando o valor de ocorrência de cada elemento. Tendo-se todos os valores de todos os elementos dos 52 pesquisadores, foi identificado o valor máximo de cada elemento. Este valor máximo originou o elemento Maximum Value of Range (MVR). A Tabela 2 apresenta as categorias e seus pesos, os elementos e seus pesos, bem como o máximo valor de cada elemento.

Category	Weight	Elements	Weight	Maximum Value of Range	
Identification (ID)	15	Education Degree (ED) ⁵	Postdoctoral	15	-
			Doctorate	12	-
			Master	8	-
Advisory (ADV)	15	Postdoctoral Advisor (PA)	6	4	
		Phd Thesis Advisor (PTA)	5	30	
		Master Dissertation Advisor (MDA)	4	66	
Examining Board (EB)	15	Participation in Examination Boards Master Dissertation (PEBMD)	5	72	
		Participation in Examination Boards Phd Thesis (PEBPT)	10	26	
Membership (MS)	15	Conference Committee Coordinator (CCC)	3	9	
		Conference Committee Member (CCM)	2	107	
		Editorial Board Member (EBM)	6	5	
		Reviewer of Journals (RJ)	4	32	
Production (PROD)	40	Articles in Scientific Journals (ASJ)	12	96	
		Book Chapter Published (BCP)	3	35	
		Books Published (BP)	5	15	
		Complete Work Published in Conference Proceedings (CWPCP)	8	437	
		H-Index (HI)	6	24	
		Network Co-autorship (NC)	2	132	
		Research Projects (RP)	1	37	
		Software With Register (SWR)	2	3	
		Software Without Register (SWOR)	1	18	
Total	100	Total	100		

Tabela 2. Categorias, elementos, pesos e valor máximo do intervalo de cada elemento.

⁵ O elemento Education Degree, da categoria Identification, é o único que possui opções, onde apenas uma deve ser utilizada (ou Pós-doutorado ou Doutorado ou Mestrado) e seu peso máximo é 15.

Diante do valor máximo identificado para cada elemento, 5 intervalos foram definidos. Estes intervalos têm por objetivo classificar o pesquisador em níveis de reputação. Os intervalos tem valores entre 0 e 100. Assim, pegamos o valor máximo de cada elemento e dividimos por quatro, gerando os cinco intervalos.

Após os cálculos, o Rep-Index do pesquisador é definido com base nos níveis apresentados na fórmula 7.

$$Rep-Index_{(R)} = \begin{cases} 1, & \text{Se } Rep-Index_{(R)} \geq 0 \text{ e } < 20 \\ 2, & \text{Se } Rep-Index_{(R)} \geq 20 \text{ e } < 40 \\ 3, & \text{Se } Rep-Index_{(R)} \geq 40 \text{ e } < 60 \\ 4, & \text{Se } Rep-Index_{(R)} \geq 60 \text{ e } < 80 \\ 5, & \text{Se } Rep-Index_{(R)} \geq 80 \text{ e } < 100 \end{cases} \quad (7)$$

5.4. Resultados

Experimentos foram realizados com os 52 pesquisadores, sendo identificados de R_1 a R_{52} . Como o grupo de pesquisadores é heterogêneo, no que tange a trajetória científica de cada um, algumas questões devem ser esclarecidas:

- No grupo existem pesquisadores que concluíram o curso de doutorado a pouco tempo, que estão iniciando sua carreira como orientadores de mestrado e de doutorado, que possuem pouca inserção internacional;
- Também existem pesquisadores que concluíram o curso de doutorado a um bom tempo, que realizaram curso de pós-doutorado, que possuem um número de orientações de mestrado e de doutorado significativo, bem como possuem inserção internacional por meio de publicação de artigos e livros;
- Por fim, no grupo definido existem pesquisadores com uma vasta experiência em pesquisa, com muitas orientações de mestrado e de doutorado, bem como possuem vasta inserção internacional, com diversos artigos e livros publicados e projetos em cooperação com instituições internacionais.

As Figuras 2, 3 e 4 apresentam os resultados dos experimentos realizados para cada um dos 52 pesquisadores, identificando-se o valor do H-Index, do G-Index e do Rep-Index.

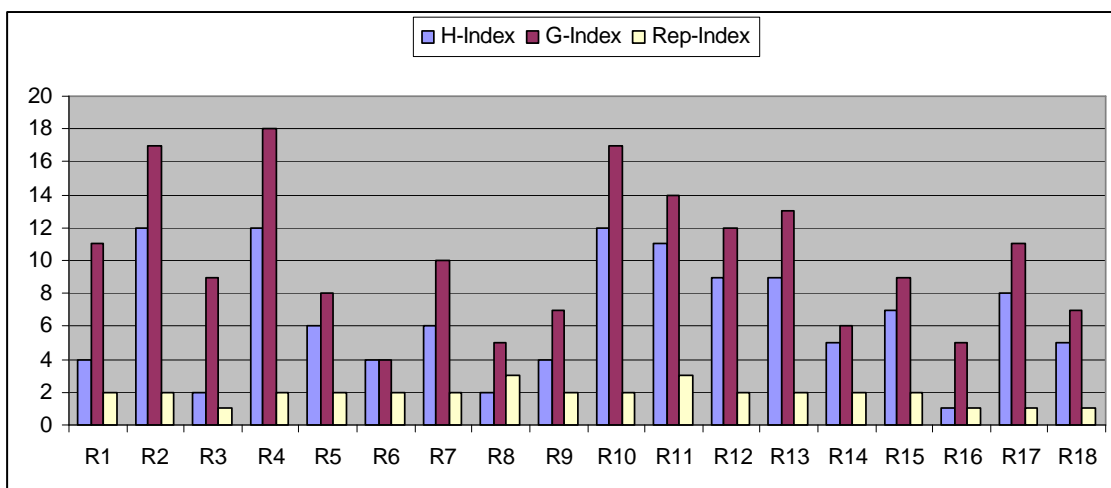


Figura 2. Dados dos pesquisadores R_1 a R_{18}

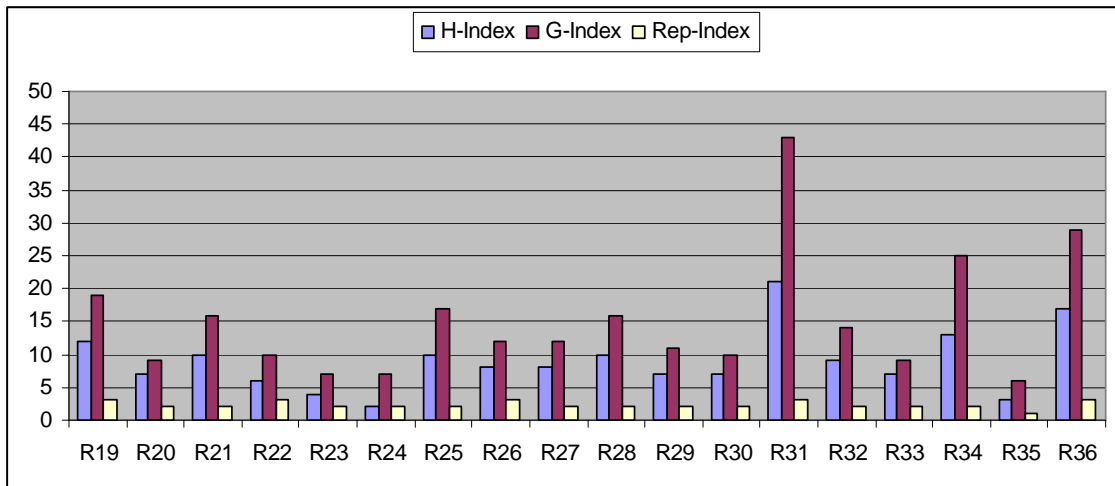


Figura 3. Dados dos pesquisadores R₁₉ a R₃₆.

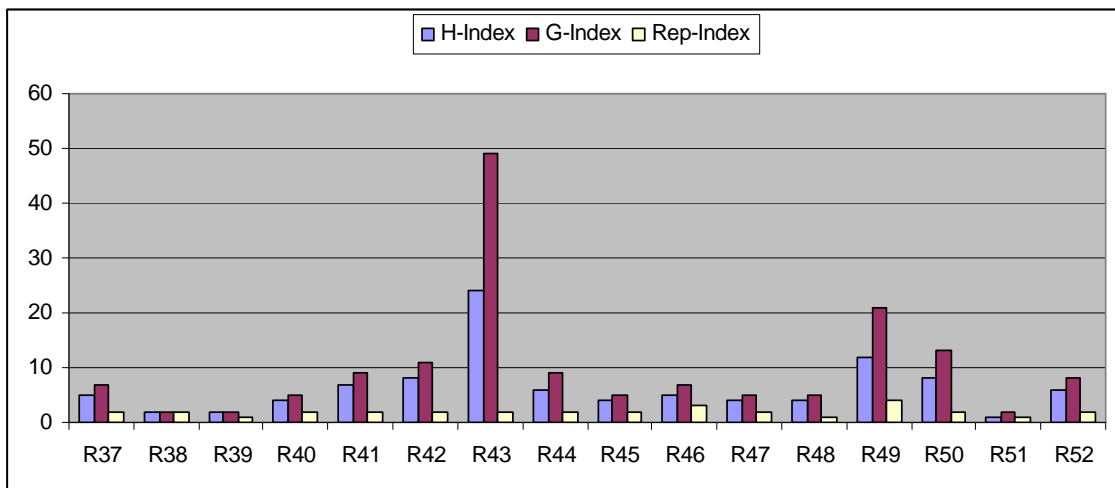


Figura 4. Dados dos pesquisadores R₃₇ a R₅₂.

Para verificar a existência ou não de correlação entre os índices de cada pesquisador, foi realizado um experimento onde medimos o coeficiente de correlação⁶ dos valores do H-Index com o Rep-Index e, posteriormente, do G-Index com o Rep-Index. Os resultados são apresentados a seguir:

- na relação do H-Index com o Rep-Index foi encontrado um coeficiente de correlação igual a 0,309327, o que indica uma correlação fraca⁷;
- na relação do G-Index com o Rep-Index foi encontrado um coeficiente de correlação igual a 0,258245, o que indica uma correlação muito fraca.

⁶ *Coefficiente de correlação*: indica a relação entre a produção do pesquisador e a soma dos diferentes pesos das variáveis independentes, podendo variar de -1 a +1, o que representa uma escala que vai de menos significativo até mais significativo.

⁷ *Interpretação do coeficiente de correlação*: de 0.00 a 0.19 indica uma correlação muito fraca, de 0.20 a 0.39 correlação fraca, 0.40 a 0.69 correlação moderada, de 0.70 a 0.89 correlação forte e de 0.90 a 1.00 correlação muito forte.

Analisando-se os valores dos dois coeficientes de correlação identificados, pode-se perceber que nossa métrica (Rep-Index) não é comparável estatisticamente aos demais índices. A justificativa para isto é devido às outras métricas serem definidas levando-se em consideração, exclusivamente, o número de citações de artigos. Nossa proposta se difere pela abrangência do modelo, levando em consideração um grupo de elementos que fazem parte da trajetória científica de um pesquisador, construída ao longo de sua carreira acadêmica.

A figura 5 apresenta o resultado do Rep-Index dos 52 pesquisadores estudados, mostrando os percentuais em cada um dos níveis.

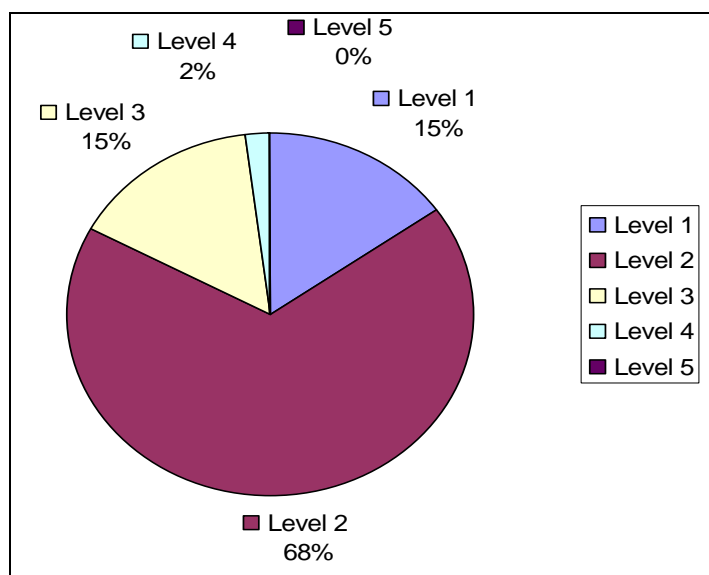


Figura 5. Resultado da classificação dos pesquisadores pelo Rep-Index.

Observa-se que o nível 2 foi o que mais teve concentração de pesquisadores em questão, chegando a 68% do total da amostra.

Um dos problemas identificados na análise dos currículos dos 52 pesquisadores é que nem todos possuem dados para todos os elementos do modelo proposto, o que acabou determinando um índice de reputação abaixo do esperado, principalmente para os pesquisadores classificados nos níveis 1 e 2.

Outra questão relevante é que como a população utilizada nos experimentos representa um dos grupos de excelência da área de computação na pesquisa brasileira, e os valores dos intervalos dos elementos do modelo são definidos pelos valores máximos de ocorrência de cada elemento, o resultado final tende a baixar. Isto se justifica pela maioria dos pesquisadores analisados não manterem um equilíbrio em sua trajetória de pesquisa, pois alguns elementos ficam bem pontuados e outros não. Isto faz com que a média final para o cálculo do índice de reputação tenda a baixar.

Com base nos resultados obtidos por meio dos experimentos conseguimos demonstrar que nossa proposta é viável, uma vez que pode ser utilizada em diferentes áreas do conhecimento, além de levar em consideração as características e peculiaridades de cada área. Levando-se em consideração a abrangência, flexibilidade e adaptabilidade do modelo, basta que ajustes sejam realizados. Isto envolve desde a alteração de pesos e valores dos elementos, bem como a supressão de elementos e até de categorias.

6. Considerações Finais e Trabalhos Futuros

Este trabalho teve como objetivo especificar um modelo de perfil de pesquisadores e apresentar métricas para avaliar a reputação de pesquisadores em sua área de atuação. O modelo levou em consideração os dados obtidos por meio da análise da trajetória científica do pesquisador, construída ao longo de sua carreira.

O processo de modelagem do perfil envolveu a identificação de quais dados foram considerados relevantes para sua definição, sendo apresentados por meio de uma ontologia de domínio. A reputação dos pesquisadores foi identificada por uma classificação, em uma escala de 1 a 5, onde 1 indica baixa reputação e 5 indica alta reputação.

Para a validação do modelo, métricas foram propostas. As métricas foram definidas baseadas no equilíbrio da trajetória do pesquisador, envolvendo os principais elementos de seu currículo. Por meio das métricas definidas, foi possível validar o modelo proposto. O modelo demonstrou equilíbrio entre os diversos elementos que formam o perfil de um pesquisador, utilizando dados quantitativos. Apesar disso, novos experimentos devem ser realizados, principalmente com uma amostra maior de pesquisadores.

Considerando a proposta apresentada neste artigo, experimentos exaustivos serão realizados com o intuito de calibrar a métrica proposta e apresentá-la por meio de formalismo. Pretende-se, ainda, realizar inferências utilizando-se da ontologia, bem como recomendar a um pesquisador a mudança de nível e evoluir na carreira; identificar pesquisadores com comportamento semelhante; cruzar o fator de impacto dos periódicos onde o pesquisador tem publicação gerando um novo elemento para o perfil e incorporar aspectos temporais no modelo.

Referências

- Chen, P., Xie, H., Maslov, S., and Redner, S. (2007). Finding scientific gems with Google. *Journal of Informetrics*, 1, 8–15.
- Chen, C., Song, I., Yuan, X., and Zhang, J. (2008). The Thematic and Citation Landscape of Data and Knowledge Engineering (1985-2007). *Data & Knowledge Engineering*. v.67, n.2 (Nov. 2008), p. 234-259.
- CiteSeer (Scientific Literature Digital Library). Disponível em <<http://citeseer.ist.psu.edu>>. Acesso em julho , 2010.
- DBLP (Digital Bibliography & Library Project). University of Trier. Disponível em <<http://dblp.uni-trier.de>>. Acesso em julho, 2010.
- Ding, Y., Yan, E., Frazho, A. and Caverlee, J. (2009). PageRank for Ranking Authors in Co-citation Networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229-2243.
- Fink, J.; Kobsa, Alfred. 2002. User Modeling for Personalized City Tours. *Journal of Artificial Intelligence*. Rev.18, 1 (September 2002), 33-74.
- Fu, Y.; Xiang, R.; Liu, Y.; Zhang, M.; Ma, S. Finding Experts Using Social Network Analysis. *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. Silicon Valley, USA, 2007.
- Gauch, S; Aravid, Speretta, M.; Chandramouli, A.; e Micarelli A.. 2007. User profiles for personalized information access. In *The adaptive web*. Lecture Notes In Computer Science, Vol. 4321. Springer-Verlag, Berlin, Heidelberg 54-89.
- Hirsch, J. E. An Index to Quantify an Individual's Scientific Research Output. In *Proceedings of the National Academy of Science*. V.102, n.46, p.16569-16572, 2005.

- Hope, T.; Nishimura, T.; Takeda, H. An Integrated Method for Social Network Extraction. Proceedings of the 15th International Conference on World Wide Web. Edinburgh, Scotland, 2006.
- Huang, J., Zhuang, Z., Li, J., and Giles, C. L. (2008). Collaboration Over Time: Characterizing and Modeling Network Evolution. In Proceedings of the International Conference on Web Search and Web Data Mining. Palo Alto, California, USA, 2008.
- Jin, B. (2007). The AR-Index: Complementing the H-Index. ISSI Newsletter, 3(1) p. 6.
- Jin, B., LiMing, L., Rousseau, R. and Egghe, L. (2007). The R- and AR-indices: Complementing the H-Index. Chinese Science Bulletin. 52(6), 855– 863.
- Krapivin, M., Marchese, M., Cassati, F. (2009). Exploring and Understanding Scientific Metrics in Citation Networks. Complex Sciences, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg. V.5, p.1550, 2009.
- Li, J.; Tang, J.; Zhang, J.; Luo, Q.; Liu, Y.; Hong, M. EOS - Expertise Oriented Search Using Social Networks. In Proceedings of the 16th International Conference on World Wide Web. Banff, Canada, 2007.
- Liu, X., Bollen, J., Nelson, M. L., and Van de Sompel, H. (2005). Co-authorship Networks in the Digital Library Research Community. Information Processing and Management: An International Journal. V.41, n.6 , p.1462-1480, 2005.
- Maia, M.; Almeida, J.; Almeida, V. Identifying User Behavior in Online Social Networks. In Proceedings of the 1st Workshop on Social Network Systems. New York, NY, USA, 2008.
- Menezes, G. V., Ziviani, N., Laender, A. H., and Almeida, V. (2009). A Geographical Analysis of Knowledge Production in Computer Science. In Proceedings of the 18th International Conference on World Wide Web. Madrid, Spain, 2009.
- Middleton, S. E.; Shadbolt, N. R.; De Roure, D. C. 2004. Ontological user profiling in recommender systems. *ACM Transactions Information Systems*. 22, 1 (January 2004), 54-88.
- Mimno, D.; McCallum, A. Expertise Modeling for Matching Papers With Reviewers. In Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining. San Jose, USA, 2007.
- Page, L.; Brin, S.; Motwani, R.; Winograd, T. (1998) The PageRank citation ranking: Bringing order to the Web. Tech. report, Stanford University.
- Park, Y. and Chang, K. Individual and Group Behavior-based Customer Profile Model for Personalized Product Recommendation. *Expert System. Appl.* 36, 2 (Mar. 2009), 1932-1939.
- Small, H. (1973). Cocitation in science literature: New measures of relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
- Tang, J., Zhang, J., Yao, L., Li, J. Extraction and Mining of an Academic Social Network. In Proceedings of the 17th International Conference on World Wide Web. Beijing, China, 2008.
- Tang, J.; Zhang, D.; Yao, L. Social Network Extraction of Academic Researchers. In Proceedings of 7th International Conference on Data Mining. Omaha, USA, 2007.
- Tang, J.; Zhang, J.; Yao, L.; Li, J.; Zhang, L.; Su, Z. ArnetMiner – Extraction and Mining of Academic Social Networks. In Proceeding of the 14th International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA, 2008.

- Tol, R. S. J. A Rational, Successive G-Index Applied to Economics Departments in Ireland. *Journal of Informetrics*, V.2, n.2, p.149-155, 2008.
- Trajkova, J.; Gauch, S. Improving Ontology-Based User Profiles. M.S. Thesis. EECS, University of Kansas. Lawrence, 2003.
- White, H.D., McCain, K.W. (1998). Visualizing a discipline: An author cocitation analysis of information science 1972-1995. *Journal of the American Society for Information Science*, 49, 327–355.
- Zhang, C. T. (2009) The *e*-Index, Complementing the *h*-Index for Excess Citations. *PLoS ONE* 4(5): e5429.
- Zhang, J.; Tang, J.; Li, J. Expert Finding in a Social Network. Proceedings of 12th International Conference on Database Systems for Advanced Applications. Bangkok, Thailand, 2007.