

Análise de Redes Sociais Científicas:

Modelagem Multi-relacional

Victor Ströele¹, Geraldo Zimbrão¹, Jano M. Souza¹

¹COPPE/UFRJ – Graduate School of Computer Sciences – Federal University of Rio de Janeiro – Caixa Postal 68.513, 21945-970 - Rio de Janeiro, RJ, Brazil

{stroele,zimbrao,jano}@cos.ufrj.br

***Abstract.** Social Networks are dynamic social structures formed by individuals or organizations. Usually, these networks are represented by nodes connected by one or more relationships types. Although they are extremely complex structures, analyze it allows us to detect many types of connections between people within and outside their institutions.*

***Resumo.** As Redes Sociais são estruturas sociais dinâmicas formadas por indivíduos ou organizações. Geralmente, essas redes são representadas por nós ligados por um ou mais tipos de relacionamentos. Embora sejam estruturas extremamente complexas, analisá-las nos permite detectar diversos tipos de conexões entre as pessoas dentro e fora de suas instituições.*

1. Introdução

Atualmente, existem muitos dados com conceitos diversos distribuídos pela Web. Alguns desses dados possuem características políticas, econômicas, ou sociais. Tais dados podem ser extraídos de páginas pessoais, ferramentas sociais e colaborativas, dentre outras formas de comunicação que produzam esses tipos de dados.

Muitas vezes um pesquisador usa a Web para encontrar artigos, livros, ou mesmo outro pesquisador que domina o assunto para auxiliar no desenvolvimento de suas pesquisas. Assim, os pesquisadores criam ligações entre eles pela colaboração direta ou indireta, citação e avaliação do trabalho desenvolvido em conjunto.

Esse fenômeno permite que os pesquisadores estudem como as conexões entre as pessoas são estabelecidas e como elas evoluem ao longo do tempo. Vários esforços têm sido feitos para analisar as redes sociais a fim de ajudar a compreender as estruturas sociais.

Uma rede social é um conjunto de objetos, onde cada um deles está conectado a outro objeto. Uma rede social pode ser representada por um grafo no qual os nós estão relacionados ou não por arestas. Uma rede social reflete uma estrutura social que pode ser representada por indivíduos ou organizações e suas relações. Em geral, as relações representam um ou mais tipos de interdependência (como idéia e religião) ou relacionamentos mais específicos (como troca de conhecimento, informação e amizade).

Os atores e seus relacionamentos são os elementos fundamentais na constituição de uma rede social. Os atores podem ser representados por uma pessoa individualmente ou por um grupo de pessoas, como grupos informais ou organizações formais.

A análise das redes sociais proporciona três resultados principais (Knoke and Song 2008). Primeiro, as estruturas das relações são mais importantes para compreender um relacionamento entre duas entidades do que os atributos de perfil dessas entidades, tais como: idade, sexo, ideologia, etc. Assim, como as relações influenciam as entidades sociais, sem considerar seus atributos individuais, a análise da rede anseia por oferecer uma melhor compreensão das ações sociais produzidas por essas relações, avaliando a estrutura da rede como um todo.

Segundo, as redes sociais afetam percepções, crenças e ações através de uma série de estruturas que são socialmente construídas pelas relações entre as entidades. As relações sociais diretas facilitam a interação entre as entidades favorecendo a troca de informação e aumentando as chances da entidade influenciar e ser influenciada por outras entidades. As relações indiretas, embora com menor presença, também expõem as entidades às novas influências através de transitividade por uma entidade que esteja relacionada diretamente.

Terceiro, os relacionamentos estruturais podem ser vistos como processos dinâmicos. As redes sociais não são estáticas, elas sofrem constantes alterações através das interações entre as pessoas, grupos, ou organizações. As entidades podem transformar as estruturas relacionais, às quais elas fazem parte direta ou indiretamente, de forma intencional ou não, através da influência dos seus conhecimentos. Assim, com a análise da rede social, podem ser localizados os principais influenciadores da rede social.

Em síntese, pode ser dito que analisar uma rede social científica permite identificar comunidades de pesquisas, pesquisadores que detêm maior influência (centralizadores), compreender a evolução social dos pesquisadores ao longo do tempo, sugerir novos relacionamentos para aprimorar as pesquisas dos pesquisadores ou melhorar o fluxo de conhecimento da rede, etc.

1.1 Trabalhos Relacionados

Muito trabalho tem sido feito na mineração de comunidades em páginas Web e em e-mails (Tyler, Wilkinson et al. 2003; Bird, Gourley et al. 2006). Outros trabalhos incluem: mineração de grupos (Ichise, Takeda et al. 2005; Han and Yan 2011) e previsão de relacionamentos (Liben-Nowell and Kleinberg 2007; Huang and Lin 2009; Lü and Zhou 2009); tecnologia das redes sociais e web semântica (Dietrich, Dietrich et al. 2008); balanceamento das redes sociais (Monclar 2007).

Existem muitos tipos de aplicações baseadas em análise de redes sociais, como redes escuras (Raab and Milward 2003; Pioch, Barlos et al. 2005), sistemas de recomendação baseado no conteúdo (Golbeck 2005; Huang, Li et al. 2005), as Redes Econômicas (Jackson 2010), o uso da análise de redes sociais (SNA) para examinar as interações formais e informais em departamentos (Ryan and O'Connor 2009), etc.

Durante a modelagem foi importante o estudo do uso do tempo na análise dos relacionamentos. Nesse contexto alguns trabalhos avaliam as mudanças estruturais da rede social (Leskovec, Kleinberg et al. 2005; Potgieter, April et al. 2009) e outros

aplicam funções de penalização com base no ano do relacionamento para diminuir o peso do mesmo (Acar, Dunlavy et al. 2009).

Em muitos problemas apenas o relacionamento de co-autoria é analisado, mas não é difícil ver que essas redes sociais podem ter outros tipos de relações científicas. Embora a análise de redes sociais multi-relacionais seja bastante interessante, poucos estudos têm sido feitos nesta área (Cai, Shao et al. 2005; Jung, Juszczyszyn et al. 2007; Ströele, Oliveira et al. 2009).

1.2 Contribuições

No mundo real, as redes sociais são em sua maioria multi-relacionais, ou seja, pessoas ou instituições estão relacionadas através de tipos de relacionamentos diferentes. Uma das propostas apresentadas neste trabalho é a construção de uma rede social multi-relacional que permita que as análises feitas através dela tenham a influência de todos os relacionamentos que cada indivíduo possui.

Está sendo proposto neste trabalho um modelo para a construção de uma rede social científica multi-relacional. Nesse modelo são considerados vários fatores que influenciam a análise das redes sociais, tais como: tipos de relacionamentos diferentes, peso do relacionamento, idade do relacionamento, perda de informação na transferência de conhecimento entre os pesquisadores, etc.

Baseado em uma rede social científica multi-relacional formulada através do modelo proposto será utilizado um método de mineração de dados para encontrar comunidades de pesquisa. O objetivo é identificar os grupos de pesquisadores que têm interesses comuns no desenvolvimento de suas pesquisas.

Embora o estudo das redes sociais científicas multi-relacionais identifique pesquisadores que possuem o mesmo objetivo de pesquisa, também podem ser identificados grupos interdisciplinares. Esses grupos são construídos por pesquisadores pertencentes a áreas distintas.

Em geral, estudando a formação dessas redes sociais científicas é possível identificar como os pesquisadores e organizações estão desenvolvendo seus trabalhos. A análise das redes sociais científicas indica o grau de envolvimento entre os pesquisadores, entre as áreas de pesquisas e, até mesmo, entre as instituições de ensino. Podem ser identificados também alguns padrões de colaboração inter e intra-universidades. Todas essas informações podem proporcionar uma melhora na comunicação da rede social e, conseqüentemente, melhorar a colaboração entre os pesquisadores.

Na seção 2 será descrita a rede social científica utilizada neste trabalho. Na seção 3 serão apresentadas as etapas da modelagem dessa rede. Na seção 4 são obtidos os resultados e, finalmente, na seção 5 estão descritas as conclusões e os trabalhos futuros.

2. Redes Sociais Científicas

As redes sociais Científicas são tipos específicos de redes sociais que representam as interações sociais oriundas do meio acadêmico. Anualmente, a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), uma instituição que recebe suporte do Ministério da Educação do Governo Federal Brasileiro, avalia os programas de pós-graduação e atribui a eles notas que variam de 1 a 7.

Os pesquisadores utilizados para a construção da rede social científica modelada neste trabalho foram selecionados das instituições nível seis e nível sete dos programas de Ciência da Computação. Os pesquisadores são todas as pessoas cadastradas no currículo Lattes que possuem o título de doutorado. As instituições nível sete são: COPPE/UFRJ, PUC-RIO e UFMG. Já as instituições nível seis são: UFPE e UFRGS.

Existem várias maneiras de identificar um relacionamento científico entre dois pesquisadores. Em geral, essas relações podem ser: participações em Projetos; relacionamentos de co-autoria; orientações em dissertações e teses; participação em banca de defesa de dissertações e teses; produções técnicas; participação em comissões examinadoras; dentre outros tipos de relações científicas.

Neste trabalho foram utilizados quatro tipos de relacionamentos diferentes na etapa de modelagem da rede social científica. São eles: participações em Projetos; relacionamentos de co-autoria; participação em banca de defesa de dissertações e teses; e produções técnicas.

A associação entre os pesquisadores pode ser mais forte ou mais fraca, segundo o grau do relacionamento entre eles. Pesquisadores que, por exemplo, têm publicações em comum, que trabalham em áreas semelhantes e que participaram no desenvolvimento de um mesmo projeto, podem ser considerados como tendo um relacionamento forte. Por outro lado, se dois pesquisadores participaram de apenas uma banca examinadora, o relacionamento entre eles é considerado fraco. Além disso, há casos em que os pesquisadores não estão diretamente conectados, nesse caso, a ligação entre eles se dará por intermédio de outros pesquisadores.

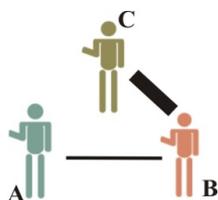


Figura 1 - Relacionamento Direto e Indireto

Tabela 1 - Matriz de pesos correspondente a Rede Social da Figura 1

	A	B	C
A	-	1	0
B	1	-	2
C	0	2	-

A pequena rede social apresentada na Figura 1 ilustra a situação descrita anteriormente. O pesquisador B está diretamente relacionado aos pesquisadores A e C. Os pesquisadores A e C não estão diretamente relacionados, porém eles estão ligados indiretamente, por transitividade, pelo pesquisador B.

Além disso, o grau do relacionamento entre os pesquisadores B e C é maior que o grau do relacionamento entre os pesquisadores A e B. O grau do relacionamento é representado pela aresta mais grossa entre os elementos mais fortemente relacionados e por arestas mais finas para os elementos menos fortemente relacionados. Esse fato pode ser confirmado através da matriz de pesos representada na Tabela 1, na qual pode ser observado que o peso do relacionamento entre B e C é o dobro do peso do relacionamento entre A e B.

3. Modelagem da Rede Social Científica Multi-relacional

Na modelagem do relacionamento de uma rede social o peso do mesmo representa o quão fortemente dois elementos estão conectados. O peso de uma ligação em uma rede social multi-relacional deve considerar o peso de todos os tipos de relacionamentos existentes entre os dois elementos.

O processo da modelagem da rede social multi-relacional foi dividido em três etapas: número de relacionamentos comuns entre os pesquisadores; a idade dos relacionamentos que ligam esses pesquisadores; e perda de conhecimento quando o relacionamento entre os pesquisadores é indireto, ou seja, quando a conexão entre eles ocorre por intermédio de outros pesquisadores.

3.1 Número de Relacionamentos em Comum

Foi necessário estabelecer uma medida que pudesse diferenciar os relacionamentos mais fracos dos mais fortes. A primeira opção foi contar os totais de relacionamentos (participação em projetos, publicações de co-autoria, participação em banca de trabalhos científicos, e produções técnicas) entre os pesquisadores. Através desse critério, definiu-se que pesquisadores com maior número de interações possuem relacionamentos mais fortes. A equação (1) define essa etapa da modelagem.

$$R_i = \frac{CR_i}{P1+P2}, \quad i = 1, \dots, t. \quad (1)$$

Onde R_i representa o grau de relacionamento 'i', CR_i é o número de relacionamentos comuns do tipo 'i' entre os pesquisadores 1 e 2, P1 representa o total de relacionamentos do tipo 'i' do pesquisador 1, P2 é o total de relacionamentos do tipo 'i' do pesquisador 2, e 't' é o total de tipos de relacionamento.

O número de relacionamentos comuns foi dividido pela soma total dos relacionamentos de cada pesquisador para que a força do relacionamento fosse relativa ao total de relacionamentos. O objetivo dessa medida é evitar que os relacionamentos com a mesma frequência tenham a mesma força.

Depois de aplicar a equação (1) para os quatro tipos de relacionamentos, foram somados todos os graus desses relacionamentos (R_i) e atribuídos pesos para cada um deles. A equação (2) representa essa etapa:

$$TR_{AB} = \sum_{i=1}^t \alpha_i R_i, \quad (2)$$

onde TR_{AB} significa o número total de relacionamentos entre os pesquisadores A e B, α_i é o peso dado ao relacionamento 'i', R_i é o resultado da equação (1).

Neste trabalho foi adotado $\alpha = 1$ para todos os tipos de relacionamentos. Entretanto, a modelagem da rede social científica multi-relacional foi feita de forma que ela possa ser utilizada em qualquer problema de análise de rede social.

Assim, se um novo problema tiver que ser modelado com pesos diferentes para cada tipo de relacionamento ($\alpha \neq 1$), a modelagem da rede social multi-relacional proposta neste trabalho ainda pode ser utilizada.

3.2 Idade do Relacionamento

Outro fator importante a ser considerado na definição do grau de relacionamento é a idade do mesmo, ou seja, é importante saber o ano em que o relacionamento foi criado. A idade do relacionamento é útil para indicar se o relacionamento reflete uma conexão atual, ou se é apenas uma conexão que existia no passado e que talvez nem exista nos dias atuais.

Para ilustrar a importância de analisar a idade de relacionamento, suponha que dois pesquisadores A e B tenham publicado três trabalhos há vinte anos, e dois outros pesquisadores C e D publicaram um artigo em conjunto no ano passado. Se for considerado apenas o número de publicações em comum, será concluído que A e B têm uma relação mais forte do que os pesquisadores C e D. No entanto, as relações entre A e B são muito antigas e, provavelmente, esses pesquisadores podem não estar trabalhando juntos nos dias atuais. Por outro lado, as conexões entre C e D são recentes, o que indica que eles atualmente têm interesses comuns.

Para considerar a idade dos relacionamentos na modelagem da rede social científica foi adicionado um peso do ano para os relacionamentos na equação (2), obtendo a seguinte equação:

$$TR_{AB} = \sum_{i=1}^t \sum_{j=1}^d \rho_j \alpha_i R_i, \quad (3)$$

onde d é a duração do relacionamento em anos e ρ_j é uma função de penalização com relação ao ano do relacionamento. Assume-se que os relacionamentos exatos possuem duração de um ano e os relacionamentos contínuos têm duração igual ao número de anos que a conexão existiu.

A definição da função de penalização com base no ano do relacionamento foi feita a partir da análise dos resultados gerados por três tipos de funções diferentes, são elas: função potência (Acar, Dunlavy et al. 2009), função exponencial e função sigmóide.

O objetivo das funções de penalização, como o próprio nome já diz, é penalizar os relacionamentos muito antigos de tal forma que, caso não surjam novos relacionamentos entre os pesquisadores a ligação entre eles receberá uma penalização maior a cada ano até ser completamente eliminada da rede social científica.

Com o uso da função de penalização, se dois autores publicarem em conjunto no ano X e no ano X+2, a publicação do ano X+2 terá um peso maior que a publicação do ano X, mesmo as duas publicações tendo influência no peso final dos relacionamentos entre os autores.

Após aplicar a equação (3) em todos os relacionamentos foi construída uma matriz de pesos MxM que representa o grau de relacionamento entre cada par de pesquisadores de uma rede social científica multi-relacional, onde M é o número de pesquisadores do conjunto de dados. Como o grau do relacionamento representa a similaridade entre os pesquisadores essa matriz é chamada de *matriz de similaridade* e está representada na equação (4).

$$MS = \begin{cases} TR_{AB} & \text{se } A \text{ se relaciona com } B \\ 0 & \text{caso contrário.} \end{cases} \quad (4)$$

3.3 Perda de Informação em Relacionamentos Longos

Outro conceito introduzido na modelagem da rede social científica multi-relacional é a perda de informações quando o caminho entre o pesquisador de origem e o pesquisador de destino passa por pesquisadores intermediários. Acredita-se que todo conhecimento que é passado de um indivíduo para outro tem alguma perda de conteúdo.

Há muitas razões para explicar a perda de conhecimento durante a transferência da informação, tais como: erros nas informações transferidas; transferência incompleta de informações; interpretação errada do conhecimento passado; desejo de reter parte do conhecimento adquirido para autoproteção; disputa por conhecimento, etc.

Tentando refletir a perda de conteúdo durante a troca de informações, foi considerado que o nó receptor recebe a informação com uma perda de $N\%$ do total do conhecimento de que ele poderia receber, onde N é o número de nós intermediários entre a fonte e o receptor.

Assumindo que o conhecimento máximo (grau do relacionamento), que pode ser transmitido entre dois pesquisadores A e B, seja dado por $MaxFlow_{AB}$, então o novo grau do relacionamento entre eles será dado como segue:

$$\overline{TR} = MaxFlow_{AB} - \frac{N * MaxFlow_{AB}}{100} \quad (5)$$

onde N é o número de elementos intermediários entre A e B, e $MaxFlow_{AB}$ é o fluxo máximo calculado utilizando a matriz de similaridade (equação (4)).

No final do processo de modelagem da rede social científica multi-relacional é obtida uma matriz de fluxo máximo com resistência, a qual será utilizada pelo método de mineração de dados para encontrar as comunidades científicas.

4. Estudo de Caso

O estudo de caso deste trabalho visa identificar as comunidades de pesquisa inter e intra-universidades brasileiras através de uma rede social científica multi-relacional ponderada formada por pesquisadores e quatro tipos de relacionamentos diferentes que existem nas instituições científicas. O conjunto de dados e o processo de modelagem dessa rede social foram descritos detalhadamente na seção 2.

Neste trabalho foi utilizado o algoritmo de agrupamento *k-means* (Han and Kamber 2006) para identificar grupos de pesquisadores com interesses comuns. Esse algoritmo utiliza a matriz de fluxo máximo definida na equação (5) e, tem como objetivo, agrupar os pesquisadores que possuem o maior fluxo de conhecimento entre si.

A Figura 2 mostra os resultados obtidos pela técnica de mineração de dados. As regiões maiores mostram as instituições brasileiras, e os retângulos menores mostram os grupos dentro de uma organização. Cada retângulo possui dois números, o primeiro identifica de maneira única o pesquisador e o segundo número, que está entre parênteses, identifica um grupo gerado pelo método de detecção de grupo.

Para facilitar a visualização da rede social, na Figura 2 as arestas representam apenas os relacionamentos mais fortes, ou seja, nessa figura é exibida a árvore geradora mínima do grafo social.

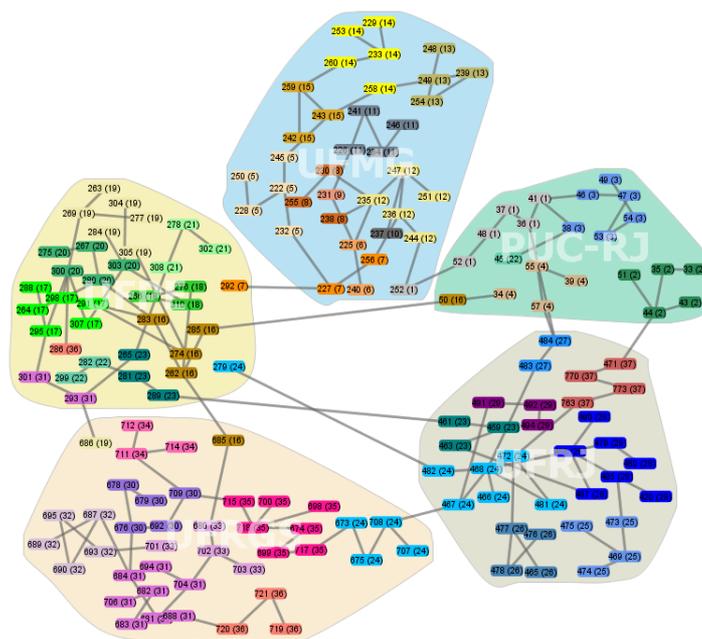


Figura 2 - Relacionamentos Internos e Externos (Árvore Geradora Mínima)

4.1 Análise dos Grupos

A Análise dos Grupos permite que seja avaliado como ocorre a comunicação entre os pesquisadores. Além disso, pode ser estudada como ocorre a troca de conhecimento entre os grupos e entre as instituições de ensino. Nessa etapa do trabalho serão analisadas as comunidades científicas encontradas pelo algoritmo de agrupamento por fluxo máximo.

A Tabela 2, mostra de maneira consolidada a distribuição dos grupos intra e inter instituições. A diagonal principal representa o número de grupos que existem apenas na universidade específica. Já as outras células indicam o número de grupos que são comuns às duas universidades.

Embora a maioria dos grupos seja formada por pesquisadores de apenas uma universidade, alguns deles possuem pesquisadores de universidades diferentes. Nestes casos, pode-se dizer que há uma grande troca de informações entre as instituições através desses grupos.

Tabela 2 - Distribuição dos Grupos

	UFRJ	PUC-RJ	UFMG	UFRGS	UFPE
UFRJ	6	0	0	1	2
PUC-RJ	0	4	1	0	1
UFMG	0	1	10	0	1
UFRGS	1	0	0	8	1
UFPE	2	1	1	1	8
Grupos Comuns	2	2	2	2	4

Ainda com base na Tabela 2 pode-se dizer que não há grandes interesses em se desenvolver pesquisas em conjunto entre os pares de instituições (UFRJ, PUC-RJ) e (UFRJ, UFMG), já que esses pares não possuem nenhum grupo em comum. Assim, o

fluxo de informações entre essas instituições é menor que entre as que possuem grupos em comum.

Além da análise dos grupos interinstitucionais também foram analisadas as áreas de interesse de cada pesquisador dos grupos. A distribuição das áreas pelos grupos está na Tabela 3. Essa tabela mostra que os grupos possuem áreas de atuação semelhantes. Entretanto, existem alguns grupos que possuem pesquisadores de áreas diferentes. Esses grupos são chamados de *grupos interdisciplinares* e são muito importantes para o desenvolvimento dos estudos científicos, pois eles indicam que existem pesquisadores “misturando” soluções de diferentes áreas. Através desses estudos podem surgir idéias bastante inovadoras.

Tabela 3 - Distribuição das áreas de atuação por grupo

	1	2	3	4	5	6	7	13	14	15	16	17	18	19	20	22	23	24	26	28	29	30	31	33	35	36	37
Análise e Complexidade de Algoritmos		X		X													X										
Arquitetura de Sistemas de Computação									X	X	X							X	X						X	X	
Banco de Dados	X				X								X					X					X	X			X
Engenharia de Software		X	X						X	X	X		X	X				X					X	X			
Hardware																	X								X		
Inteligência Artificial	X													X									X				
Linguagem de Programação			X					X				X			X												
Mineração de Dados																X											
Modelos Analíticos e de Simulação																		X			X		X				
Otimização																				X							
Processamento Gráfico							X															X					
Redes de Computadores						X																					
Sistema de Informação		X			X					X	X	X			X			X					X	X			X
Teleinformática			X																		X						X

4.2 Análise dos Relacionamentos

Os relacionamentos entre cada par de universidades foram examinados através de uma matriz simétrica, mostrada na Tabela 4, na qual o par ij representa o número total de relacionamentos entre as universidades ‘ i ’ e ‘ j ’.

Tabela 4 - Totais de relacionamentos inter institucionais

	UFRJ	PUC-RJ	UFMG	UFRGS	UFPE
UFRJ	–	19	6	4	7
PUC-RJ	19	–	24	3	23
UFMG	6	24	–	11	8
UFRGS	4	3	11	–	10
UFPE	7	23	8	10	–
TOTAL	36	69	49	28	48
TOTAL (RP)	1.09	3.14	1.36	0.68	1.30

O total de relacionamentos externos foi analisado relativamente ao número de pesquisadores (Relacionamentos por Pesquisador – RP), pois o número de pesquisadores de cada instituição é diferente. É mais provável que uma instituição com muitos pesquisadores tenha mais relacionamentos externos do que uma instituição com poucos pesquisadores.

Analisando a última linha da Tabela 4 pode ser observado que a PUC-RJ é a instituição com o maior número de relacionamentos externos por pesquisador, ou seja,

dentre as instituições analisadas nesse estudo de caso, a PUC-RJ é a que realiza o maior número de trabalhos interinstitucionais. Por outro lado, a UFRGS é a universidade com a menor relação de pesquisadores com ligações externas.

A Tabela 5 também é uma matriz simétrica na qual o par ij representa o total de relacionamentos fortes entre as instituições 'i' e 'j'. Os relacionamentos fortes são aqueles que formam a árvore geradora mínima da rede social científica. Cada relacionamento forte indica uma forte cooperação entre os pesquisadores. A árvore geradora mínima da rede social deste trabalho pode ser visualizada na Figura 2.

Nessa tabela foi analisado o total de relacionamentos externos fortes por pesquisador (RFP). É possível ver que a PUC-RJ e a UFRJ são as instituições mais fortemente relacionadas a outras instituições. Por outro lado, a UFMG e a UFRGS são as menos fortemente ligadas a outras universidades.

Tabela 5 - Totais de relacionamentos fortes

	UFRJ	PUC-RJ	UFMG	UFRGS	UFPE
UFRJ	–	3	0	1	2
PUC-RJ	3	–	1	0	1
UFMG	0	1	–	0	1
UFRGS	1	0	0	–	2
UFPE	2	1	1	2	–
TOTAL	6	5	2	3	6
TOTAL (RFP)	0.18	0.22	0.05	0.07	0.16

Como já era esperado, os relacionamentos intra-institucionais são geralmente mais fortes que os interinstitucionais. Assim, conclui-se que pesquisadores que estão locados na mesma instituição de ensino possuem uma tendência maior de trabalharem juntos do que com pesquisadores de outras instituições.

4.3 Validação dos Resultados

A análise dos relacionamentos e os grupos formados pelo método proposto foram validados utilizando uma avaliação qualitativa. Foram entrevistados – com o auxílio de um questionário – os pesquisadores de uma das universidades, as respostas foram analisadas, e, em seguida, comparadas com os resultados de nossa abordagem.

Nesse formulário os pesquisadores foram questionados sobre suas áreas de interesse, se ele/ela trabalha com pesquisadores de outras áreas, se ele/ela geralmente se relaciona com pesquisadores de outras instituições e outros tipos de perguntas, a fim de mapear o comportamento científico do pesquisador.

A maioria dos pesquisadores que responderam ao questionário afirma estar mais diretamente relacionado a pesquisadores da mesma instituição do que com pesquisadores de instituições externas. Cada pesquisador indicou os nomes dos colegas com os quais ele se considera mais estreitamente relacionado profissionalmente, o que permitiu avaliar os relacionamentos fortes e fracos. Os resultados obtidos com o questionário mostraram que tanto os métodos propostos quanto as análises da rede social científica estão muito próximos da realidade.

5. Conclusão e Trabalhos Futuros

Neste trabalho foi modelada uma rede social científica multi-relacional, onde os elementos dessa rede representam pesquisadores de instituições de ensino brasileiras e as ligações são diferentes tipos de relacionamentos científicos. Nessa modelagem foram abordados diversos conceitos envolvidos na análise das redes sociais científicas, tais como: idade do relacionamento, perda de informação na transferência de conhecimento, tipos de relacionamentos diferentes, etc.

As análises realizadas sobre a rede social científica segundo o modelo proposto neste trabalho mostraram a validade desse modelo. Segundo os resultados obtidos a modelagem da rede social científica multi-relacional está bem próxima da realidade das instituições de ensino.

Um dos trabalhos futuros é modelar a rede social da base de dados DBLP, que é uma base de dados bastante conhecida no meio científico e pode ser modelada segundo os padrões propostos neste trabalho. Como a modelagem pode ser aplicada a qualquer tipo de rede social, outro trabalho futuro é modelar outros tipos de redes sociais nas quais o fator tempo tenha um impacto importante na análise da rede.

Referências

- Acar, E., D. M. Dunlavy, et al. (2009). Link Prediction on Evolving Data Using Matrix and Tensor Factorizations. IEEE International Conference on Data Mining Workshops.
- Bird, C., A. Gourley, et al. (2006). Mining email social networks. Proceedings of the 2006 international workshop on Mining software repositories.
- Cai, D., Z. Shao, et al. (2005). Community mining from multi-relational networks. Proceedings of the 2005 European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'05), Porto, Portugal.
- Dietrich, J., J. Dietrich, et al. (2008). "Using social networking and semantic web technology in software engineering – Use cases, patterns, and a case study." Journal of Systems and Software **Volume 81**(Issue 12): Pages 2183-2193.
- Golbeck, J. (2005). Semantic Web Interaction through Trust Network Recommender Systems End User Semantic Web Interaction Workshop 4th International Semantic Web Conference.
- Han, J. and M. Kamber (2006). Data Mining: Concepts and techniques, USA, Morgan Kaufmann Publishers.
- Han, L. and H. Yan (2011). "BSN: An automatic generation algorithm of social network data." Journal of Systems and Software **Volume 84**(Issue 8): 1261-1269
- Huang, Z., X. Li, et al. (2005). "Link Prediction Approach to Collaborative Filtering." JCDL: 141-142.
- Huang, Z. and D. K. J. Lin (2009). "The time-series link prediction problem with applications in communication surveillance." INFORMS J. Computing **vol. 21**: 286-303.

- Ichise, R., H. Takeda, et al. (2005). Community Mining Tool using Bibliography Data. Proceedings of the Ninth International Conference on Information Visualisation, IEEE.
- Jackson, M. O. (2010). Social and Economic Networks, Princeton University Press.
- Jung, J. J., K. Juszczyszyn, et al. (2007). "Centrality Measurement on Semantically Multiplex Social Networks: Divide-and-Conquer Approach." International Journal of Intelligent Information and Database Systems Vol. 1, No. 3/4: 277-292.
- Knoke, D. and Y. Song (2008). GSocial Network Analysis.
- Leskovec, J., J. Kleinberg, et al. (2005). Graphs Over Time: Densification Laws, Shrinking Diameters and Possible Explanations. Proceedings of the eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining.
- Liben-Nowell, D. and J. Kleinberg (2007). "The Link Prediction problem for social networks." Journal of the American Society for Information Science and Technology vol. 58: 1019–1031.
- Lü, L. and T. Zhou (2009). Role of weak ties in link prediction of complex networks. In Proceeding of the 1st ACM international Workshop on Complex Networks Meet information & Knowledge Management, Hong Kong.
- Monclar, R. S., et al (2007). A New Approach to Balance Social Networks. Proceedings of UK Social Network Conference.
- Pioch, N., F. Barlos, et al. (2005). A Link and Group Analysis Toolkit (LGAT) for Intelligence Analysis, https://analysis.mitre.org/proceedings/Final_Papers_Files/348_Camera_Ready_Paper.pdf.
- Potgieter, A., K. A. April, et al. (2009). "Temporality in Link Prediction: Understanding Social Complexity." Journal Article E:CO vol. 11.
- Raab, J. and H. Milward (2003). "Dark Networks as Problems." Journal of Public Administration Research and Theory, vol. 13, no. 4. pp 413-439.
- Ryan, S. and R. V. O'Connor (2009). "Development of a team measure for tacit knowledge in software development teams." Journal of Systems and Software Volume 82(Issue 2): 229-240
- Ströele, V., J. Oliveira, et al. (2009). Mining and Analyzing Multirelational Social Networks. 2009 International Conference on Social Computing (SocialCom09), Vancouver, Proceedings of International Conference on Social Computing (IEEE CS).
- Tyler, J. R., D. M. Wilkinson, et al. (2003). Email as Spectroscopy: Automated Discovery of community Structure Within Organizations. Proceedings of the First International Conference on Communities and Technologies, M. Huysman, E. Wenger, V. Wulf