Evolução da Pesquisa em Análise de Redes Sociais no Brasil: Uma Visão no Domínio de Computação

Pedro Pinto[1], Fabrício Firmino[1], Tiago Cruz de França[1,2], Jonice Oliveira[1]

1-Universidade Federal do Rio de Janeiro(UFRJ)

2-Universidade Federal Rural do Rio de Janeiro(UFRRJ)

{pedropinto@ufrj.br, tcruz.franca@ufrj.br, jonice@dcc.ufrj.br}

Resumo. *A área de Análise de Redes Sociais (ARS) é uma área multidisciplinar. Recentemente, temos presenciado a grande inserção e contribuição da comunidade de Ciência da Computação devido ao crescente número de dados sociais disponíveis e o aumento do uso de ferramentas de interação. Apesar disto, ainda há poucos registros quantitativos desse crescimento. O objetivo deste trabalho é mostrar esse movimento nas pesquisas científicas brasileiras, tendo como base a comunidade científica brasileira no domínio de Computação. Com esta análise é possível enxergar o crescimento da ARS no Brasil, assim como os principais pesquisadores e os eventos reconhecidos pelos pesquisadores da área.*

1. Introdução

A Análise de Redes Sociais (ARS) tem despertado o interesse de várias áreas desde a Educação até áreas de Marketing e segurança. Apesar de ser uma área de estudos antiga (seus primeiros estudos no século XVII) [Spinoza, 1663], [Althusius, 1603], [Oliveira, 2014], as contribuições em pesquisas e soluções por parte da comunidade de Computação são recentes. Muitos eventos de computação têm adicionado esta área (e seus temas correlatos) em suas chamadas, criando trilhas (chegando a alterar seu nome) e incluindo explicitamente o interesse pelo estudo das interações sociais [Benevenuto, 2014]. Além disso, novos eventos têm surgido.

O interesse da comunidade de computação surgiu com o aumento do número de interações Web, criação e uso de ferramentas de comunicação (e-mail, mídias sociais, etc.) e maior mobilidade e conectividade dos usuários. Os dados se tornaram abundantes e disponíveis e novos métodos e infraestrutura de processamento de grandes volumes de dados surgiram e estão sendo aperfeiçoados.

2. Metodologia

Esta seção contém a descrição das etapas da metodologia adotada.

2.1. Coleta dos Artigos

A Figura 1 apresenta diagrama com os passos da etapa de coleta dos dados. Foram coletados os artigos dos eventos brasileiros na área de Ciência da Computação, apoiados pela SBC[2], selecionados por especialistas da área e que possuíam em suas chamadas o tópico de interesse "Análise de Redes Sociais" ou temas correlatos. Os eventos

[1] Agredecimentos a CAPES, CNPq e FAPERJ.

[2] SBC – Sociedade Brasileira de Computação

analisados foram: BraSNAM; IHC; SBBD; SBIE; SBRC; SBSC; SBSI; SEMISH; Webmedia; WIVA. O WIVA (versão única em 2008) foi o primeiro evento nacional focado na área.

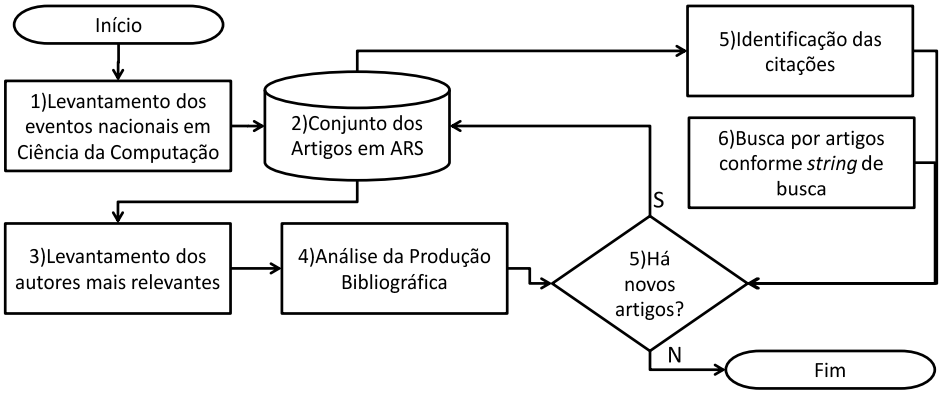

Figura 1 – Processo para coleta dos dados utilizados na análise da evolução da área de ARS no Brasil

Os artigos publicados em eventos nacionais foram coletados e armazenados em uma base. Utilizando este conjunto inicial de dados foram levantados os 10 autores mais relevantes. Tal relevância foi definida com base na maior quantidade de publicações, citações e índices H dos autores ao longo do período analisado.

A fim de se obter a produção bibliográfica dos autores selecionados e identificar outros trabalhos de ARS, foram utilizadas as páginas Google Scholar Citations desses autores. Foram adicionados à base os trabalhos de ARS publicados por esses autores em outros eventos bem como os trabalhos que os citam (e que também possuem no mínimo um autor brasileiro). Em seguida foram selecionados os veículos de publicação (eventos e periódicos) onde os trabalhos foram publicados. Novos artigos de autores brasileiros publicados em tais veículos também foram adicionados à base. Os passos de verificação de produção do autor, identificação de novos trabalhos e eventos relacionados à área de ARS e checagem das referências dos artigos para se identificar outros trabalhos de brasileiros foram repetidos iterativamente até não se encontrar novos trabalhos.

Listagem 1 - Termos utilizados na busca de novos trabalhos

Análise de Redes Sociais; Redes Sociais; Redes Sociais Online; Análise de Sentimentos; Análise de Redes Complexas; Mídias sociais; Facebook; Twitter; Orkut; (ou qualquer outra mídia social.).

Além das etapas de inclusão de trabalhos à base de artigo, foi realizada uma busca no Google Scholar[3] por artigos que contivessem no título, resumo ou palavras-chaves alguns dos termos (em português) descritos na Listagem 1. Os mesmos termos em inglês também foram pesquisados. Novos trabalhos com participação de brasileiros foram recuperados e adicionados à base.

Resumidamente, os filtros para inclusão de artigos na base foram: (i) possuir pelo menos um autor brasileiro; (ii) ter sido publicado entre os anos de 2001 a 2014; e (iii) possuir no título, resumo ou nas palavras-chaves termos relacionados a ARS. Em 2001

[3] https://scholar.google.com.br/

foi publicada a primeira referência nacional na área. Com base nesta observação foi determinada a data inicial como sendo o ano de 2001.

Ao final da coleta e após se aplicar todos os filtros a base contém 556 trabalhos produzidos por 1.117 autores distintos. A lista completa de todas as referências encontra-se em: *https://www.dropbox.com/s/p59rhy36a5rv88p/Link.xlsx?dl=0.*

2.2 Tratamento e Padronização dos Dados

Os artigos da base foram agrupados por ano. Seus metadados – como título, resumo, palavras-chave, veículo, ano de publicação, número de citações que o trabalho obteve até a data de captura dos dados, nome dos autores, respectivos índices H e instituições dos autores- foram processados para que fosse possível realizar as análises desejadas. Foram obtidos: dados quantitativos (organizados em uma planilha) e o grafo da rede de coautoria (para análise topológica) no qual os vértices são os pesquisadores e a aresta definida pela interação de coautoria nos artigos.

2.3 Definição das Análises

Analisar o crescimento quantitativo da ARS utilizando: i) o número de publicações ao longo do tempo, ii) o número de autores ao longo do tempo e iii) o número de eventos e periódicos que publicaram trabalhos relacionados a ARS no intervalo definido. Para identificar a contribuição dos pesquisadores da área de computação foram realizadas dois tipos de análises: pesquisadores como maior número de publicações, e verificação da centralidade (grau de intermediação ou *betweenness*, PageRank e centralidade de autovetor) no grafo de representação da rede de coautoria [Newman, 2010].

Além dessas análises, também foram levantados os tópicos em destaque na área de ARS. Os tópicos foram obtidos dos títulos, palavras-chaves e resumos dos artigos. Finalmente, para identificar os periódicos e eventos mais relevantes para o desenvolvimento da área no Brasil, buscou-se comparar e classificar os meios de publicação identificados na base em um ranque.

3. Análise dos Resultados

Tomando como base os artigos avaliados, percebeu-se que a quantidade de publicações cresce a cada ano. A Figura 2 apresenta o gráfico desse crescimento. No intervalo total de tempo analisado, o número de trabalhos publicados cresceu mais de 100 vezes.

A quantidade de autores também cresceu do primeiro ao último intervalo observado como pode ser visto no gráfico da Figura 3. A Figura 4 apresenta o crescimento do número de eventos e períodos no intervalo analisado. Nesta mesma figura também é possível observar a quantidade de novos eventos a cada ano que incluíram ARS nos tópicos de seu interesse.

Para se verificar a contribuição na divulgação da área de ARS no Brasil foram feitas duas análises: os 10 autores com mais publicações relacionadas à ARS por ano no intervalo analisado (Figura 5); e a rede de coautoria apresentando os grupos de pesquisa formados por esses autores (Figura 6) no qual o tamanho do nó (autor) é determinado pelo PageRank do mesmo e a intensidade (peso) de uma aresta é definida pela quantidade de publicações em coautoria entre dois autores. Observe-se que os nomes em

destaque são semelhantes aos destacados no grafo da Figura 6 são os mesmos da Figura 5.

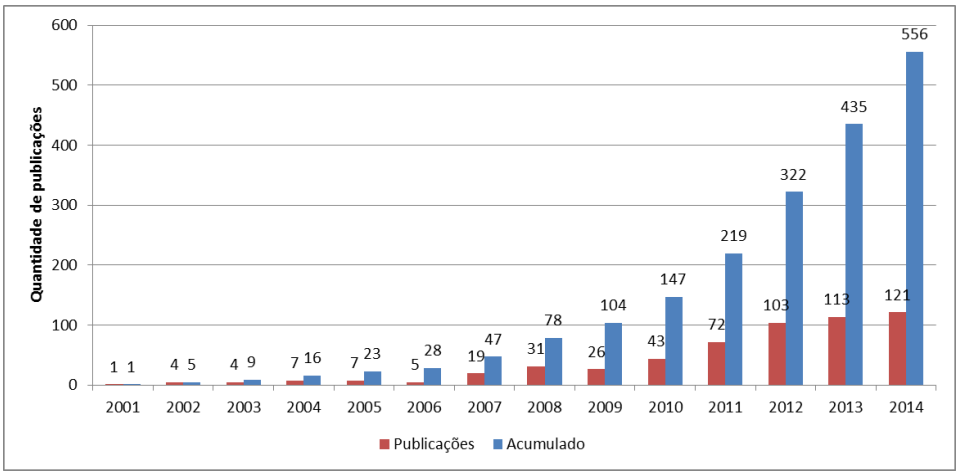

Figura 2 - Quantidade de artigos publicados entre 2001 e 2014.

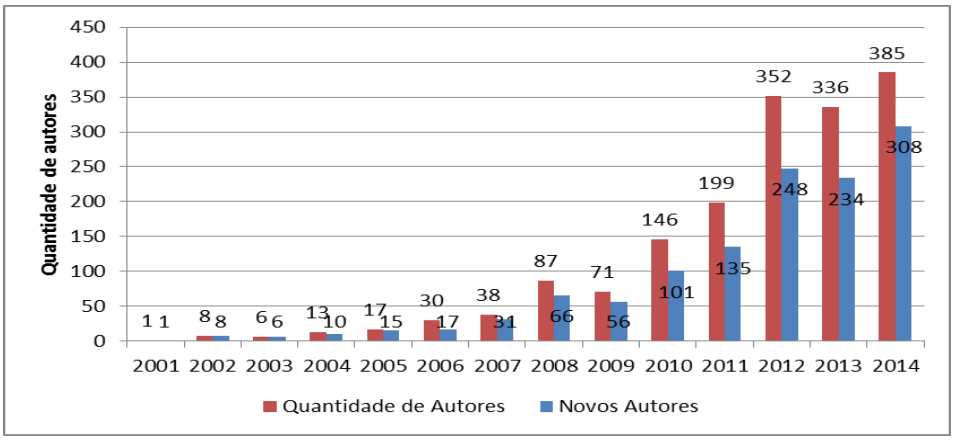

Figura 3 - Quantidade de autores entre 2001 e 2014

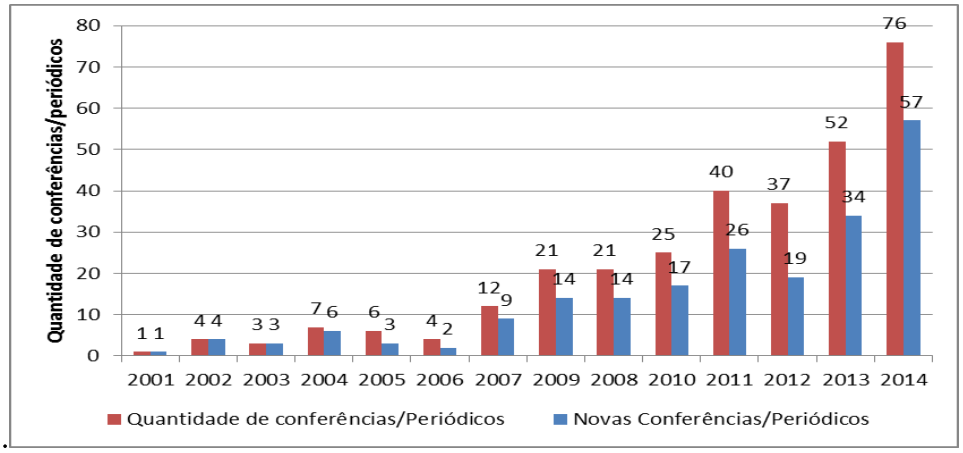

Figura 4 - Quantidade de eventos entre 2001 e 2014

A Tabela 2 apresenta as mídias sociais *online* mais utilizadas nos artigos coletados. Disparadamente, a mídia social mais utilizada foi o Twitter. A

Figura 7 apresenta o gráfico com os veículos de publicação que aceitaram mais trabalhos de ARS agrupados por ano no intervalo analisado. Foram incluídos apenas veículos com no mínimo 10 publicações.

A primeira revista da base a publicar trabalhos de ARS foi a Ciência da Informação em 2001. Alguns veículos que não são estritamente de computação (como Ciência da Informação e Informação & informação) aparecem na base porque pesquisadores de Computação publicaram ou foram referenciados por artigos publicados nesses veículos. A maioria dos eventos possuem conceito Qualis A1, B4 e C. Os 59 eventos contidos na base foram avaliados.

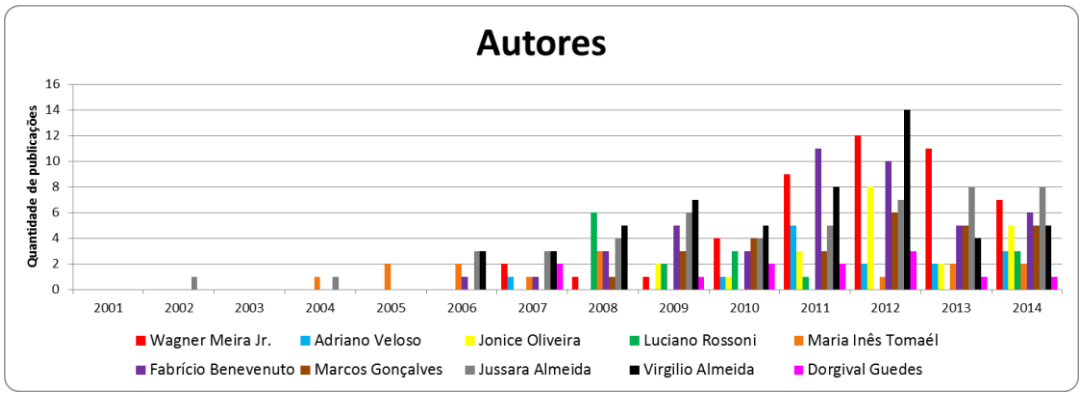

Figura 5 - Publicações anuais dos 10 principais autores

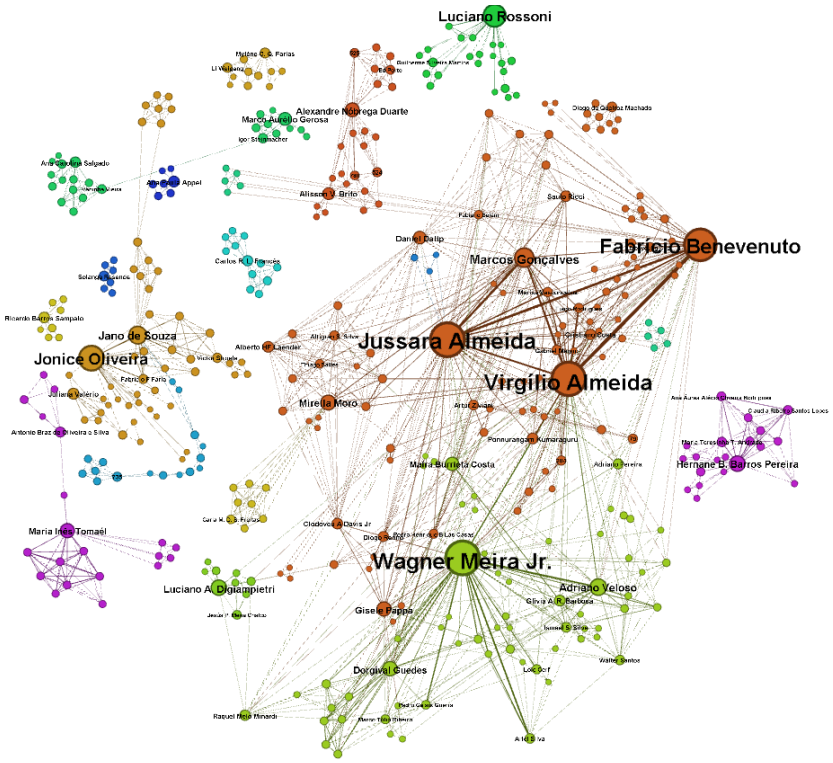

Figura 6 - Grafo de coautoria

Tabela 2 - Quantidade de trabalhos pelas mídias sociais *online* mais citadas

Mídia	Twitter	Facebook	YouTube	Foursquare	Google+	Orkut	Linkedin
Quantidade	56	19	5	4	4	4	1

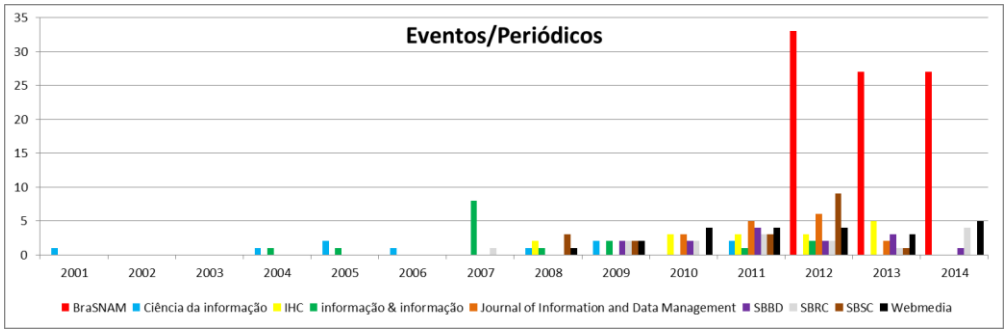

Figura 7 - Quantidade de publicações por veículo.

4. Considerações Finais

Este trabalho analisou a evolução da pesquisa em ARS no Brasil na comunidade de Ciência da Computação. Foi verificado o crescimento quantitativo da área de ARS utilizando os artigos, eventos e publicações ao longo do tempo, os principais pesquisadores que auxiliaram na divulgação da área no Brasil, além os temas, subáreas, eventos e periódicos nacionais mais comuns. Conclui-se que a ARS ainda é uma área em crescimento, com novos adeptos a cada ano.

Este trabalho possui algumas limitações. Apenas os artigos que estavam disponíveis *online* foram considerados. As buscas por artigos foram feitas na DBComp[4], Google Scholar e sites dos eventos. Essa abordagem torna difícil sua reprodução. Um procedimento sistemático de recuperação de trabalhos poderia assegurar melhor a reprodução das análises e obtenção dos resultados apresentados. O período analisado iniciou em 2001 até 2014. Em 2001 quando alguns veículos ainda não disponibilizavam os artigos em bases *online*, o que também dificultou a realização das análises apresentadas e pode incluir um viés nos resultados obtidos. A falta de padrão nos artigos (como falta de palavras-chave) também acrescentam vieses à algumas das análises realizadas.

Para trabalhos futuros, espera-se repetir a etapa de coleta adotando uma abordagem sistemática, além de incluir novas análises e identificação de padrões.

5. Referências

Tomaél, M. I., Alcará, A. R., Chiara, I. G. D. (2005), *Das redes sociais a inovação,* Ciência da informação

Spinoza, B. (1663), *Renati Des Cartes principiorum philosophiae pars I&II met cogitara Metaphysica*, Amsterdam: J. Riewerts

Althusius, J. (1603), *Politica Methodice Digesta*, Indianapolis: Liberty Press

Oliveira, J., Benevenuto F. (2014), *Os Desafios da Análise de Redes Sociais* III Brazilian Workshop on Social Network Analysis and Mining

Newman, M. (2010) *Networks: an introduction.* Oxford University Press

[4] http://www.lbd.dcc.ufmg.br/bdbcomp/