

Análise da Rede de Colaboração e Abrangência do Simpósio Brasileiro de Sistemas de Informação

Alternative Title: Analysis of the Collaboration Network and Scope of the Brazilian Symposium on Information Systems

Patrícia Gavazzi Morais de Oliveira
Universidade Federal de Itajubá
Av. BPS, 1303, Pinheirinho
Itajubá, MG, Brasil
patgavazzi@gmail.com

Melise M. V. de Paula
Universidade Federal de Itajubá
Av. BPS, 1303, Pinheirinho
Itajubá, MG, Brasil
melisepaula@unifei.edu.br

RESUMO

Os eventos científicos proporcionam a integração entre diferentes comunidades formando redes de colaboração que podem ser caracterizadas por diferentes propriedades e padrões. A análise dessas informações pode contribuir de forma significativa para o aperfeiçoamento dessas redes e a disseminação do conhecimento criado nesses cenários. Aspectos relacionados à distribuição do conteúdo das publicações, à extensão do alcance da rede e ao aumento da colaboração podem ser aprimorados a partir dessas discussões. O objetivo deste trabalho foi analisar como é definida a colaboração entre as diversas comunidades no Simpósio Brasileiro de Sistemas de Informação através da utilização de um sistema gerenciador de banco de dados orientado a grafos. Os dados coletados foram analisados em um primeiro momento para a definição do modelo de dados a ser utilizado. Tendo sido especificado o modelo, os aspectos analisados foram a rede de colaboração, a distribuição geográfica e a abrangência dos temas abordados. Os resultados indicam que há uma evolução no comportamento das redes de colaboração formadas ao longo dos anos e que os padrões encontrados a partir desta evolução estão principalmente ligados à distribuição geográfica dos autores e instituições.

Palavras-Chave

Rede de Colaboração; Banco de Dados Orientado a Grafos.

ABSTRACT

Scientific events provide the integration between different communities forming collaborative networks that can be characterized by different properties and patterns. The analysis of this information can contribute significantly to the improvement of these networks and the dissemination of the knowledge created in these scenarios.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2017, June 5th–8th, 2017, Lavras, Minas Gerais, Brazil.
Copyright SBC 2017.

Aspects related to the distribution of the content of the publications, the extension of the reach of the network and the increase of the collaboration can be improved from these discussions. The purpose of this work was to analyze how the collaboration between the different communities in the Brazilian Symposium on Information Systems is defined through the use of a graph-driven database management system. The collected data were analyzed in a first moment for the definition of the data model to be used. Having specified the model, the analyzed aspects were the collaboration network, the geographic distribution and the comprehensiveness of the topics covered. The results indicate that there has been an evolution in the behavior of collaboration networks formed over the years and that the pattern seen on this evolution is mainly related to the geographical distribution of authors and institutions.

CCS Concepts

• Information systems→Information retrieval • Information systems→Data management systems→Database design models→Graph-based database models

Keywords

Collaboration Network; Graph Database.

1. INTRODUÇÃO

O conhecimento em suas diversas formas só existe através da relação do ser humano com a natureza e com outros indivíduos. A construção do conhecimento é uma atividade movida pelo instinto e curiosidade humana e só é alcançada através da socialização dos indivíduos [3].

Na ciência, a percepção de que o conhecimento é obtido através da interação do homem é ainda mais clara quando se observa que a figura do pesquisador isolado não é mais tão comum [7]. Cada vez mais as instituições e os pesquisadores estão colaborando entre si e aumentando a rede de colaboração científica; esta começa a existir a partir do momento em que há a interação de dois ou mais pesquisadores em um projeto de pesquisa, compartilhando recursos de diversos tipos a fim de se atingir o objetivo final que é a construção de conhecimento [9].

Além da tecnologia, os eventos científicos também são oportunidades para integração e socialização do conhecimento científico. No Brasil, a Sociedade Brasileira de Computação

realiza diversos eventos como o Simpósio Brasileiro de Banco de Dados, Simpósio Brasileiro em Engenharia de Software, Congresso Brasileiro de Informática na Educação e o Simpósio Brasileiro de Sistemas de Informação (SBSI)¹. Analisar como vem sendo definida a rede de colaboração e o alcance desses eventos pode estimular discussões interessantes e o aprimoramento da comunidade científica envolvida.

Desta forma, o objetivo principal deste trabalho foi analisar a rede de colaboração estabelecida no SBSI nos nove últimos anos (2008-2016) e a abrangência do evento no que tange à distribuição geográfica, os temas abordados e a participação das instituições de ensino brasileiras. O caminho definido para a realização dessa análise foi a utilização de um banco de dados orientado a grafos, o Neo4j. Para isso, os dados foram coletados e analisados para que fosse possível obter uma modelagem que favorecesse as consultas necessárias para responder os pontos de análise definidos.

O artigo está organizado da seguinte maneira: na Seção 2, serão apresentados os trabalhos relacionados com o tema abordado. Na Seção 3, será apresentada a metodologia adotada. Já a Seção 4 traz as análises realizadas a partir dos indicadores definidos e seus resultados. Por fim, a quinta e última seção traz a as considerações finais.

2. TRABALHOS RELACIONADOS

Em [5], foi feito um estudo a partir da mineração de textos das publicações do SBSI nos anos de 2008 a 2011. O trabalho consiste no desenvolvimento de um sistema com dois módulos principais: geração de grafos e classificação automática de textos.

Os grafos foram gerados através da utilização da ferramenta Pajek que requer como entrada um arquivo de texto formatado, incluindo especificação dos vértices, arestas e pesos a serem gerados [5]. No trabalho foram criados dois grafos; um referente à colaboração entre autores no ano de 2008 e outro referente à colaboração entre universidade nos anos de 2008 a 2010. Porém não foi feito um estudo detalhado da evolução destas colaborações ao longo dos anos do evento.

O foco principal do estudo foi o segundo módulo da aplicação que se refere à classificação automática de textos. Tendo como base o resumo de cada artigo, o sistema analisa, a partir da mineração de textos, quais seriam as categorias nas quais o artigo melhor se enquadraria. Posteriormente, foram analisadas as categorias sugeridas pelo sistema com as categorias definidas pelos autores no momento da submissão do artigo. Como resultado desta análise, foi observado que há alguma dificuldade na correta classificação dos artigos pelos autores justificada pelas divergências encontradas entre as categorias sugeridas e aquelas realmente utilizadas [5].

Em [2], foi realizada a análise da rede de colaboração do SBSI para os anos de 2006, 2008, 2009 e 2010. Foram utilizados dois critérios para análise dos autores participantes: (1) autores com centralidade de grau elevada, ou seja, autores que possuem um alto número de colaboração com outros autores (2) quantidade de participações dos autores no evento. Desta forma, foram identificados os autores com maior grau de colaboração e analisada a rede destes autores ao longo dos anos.

Posteriormente, foram identificados os autores com maior número de participação dentre as quatro edições analisadas e criadas as redes com base nestes autores. Além da análise de autores, foi feita uma análise da evolução da colaboração entre as instituições.

Como resultado da análise de [2], com exceção da UFRJ que teve um aumento significativo no intervalo de 2006 a 2011, as instituições tiveram um aumento razoavelmente gradual de parcerias ao longo das edições. O trabalho também utilizou corte de vértices a fim de identificar autores e instituições importantes na rede que, caso fossem removidos, impactariam a rede, desconectando o grafo.

Ambos os trabalhos abordados nesta seção realizaram, de algum modo, a análise da rede de colaboração do SBSI para autores e instituições. Os resultados apresentados pelos trabalhos acima relacionados serão complementados com os dados das edições de 2012 a 2016.

3. METODOLOGIA

Para alcançar o objetivo proposto, as seguintes etapas foram realizadas: definição dos indicadores a serem avaliados; coleta e processamento dos dados dos anais do SBSI considerando os anos de 2008 a 2016; análise dos resultados obtidos através dos indicadores aferidos.

3.1 Definição dos Indicadores

Para atingir os dois principais objetivos gerais, análise da abrangência do SBSI e da rede de colaboração, foram estabelecidos alguns indicadores que podem ser considerados objetivos específicos tendo sido definidos através de alguns pontos de análise.

Para a análise da abrangência do simpósio, foram considerados os seguintes aspectos: distribuição geográfica, categorias abordadas e participação das instituições de ensino brasileiras. Os indicadores definidos foram:

- Distribuição geográfica: quantidade total de artigos publicados por região e série histórica da quantidade de artigo por região nos de 2008 a 2016;

Para esta análise, foram considerados apenas artigos publicados por autores pertencentes às instituições nacionais, uma vez que a análise se baseou nas diferentes regiões do Brasil. É importante ressaltar que a região do artigo é a combinação das regiões dos autores; para artigos em que há colaboração de autores de diferentes regiões, ambas foram consideradas como região do artigo, porém para casos em que há colaboração de autores da mesma região, o artigo foi contabilizado apenas uma vez para aquela região.

- Distribuição por categoria: quantidade de artigos por categoria considerando a seção *Categories e Subject Descriptors* dos artigos submetidos;

A partir de 2015, os artigos publicados no SBSI passaram a conter um indicador de categoria. Este indicador aparece no artigo na seção “Categories and Subject Descriptors” e deve conter as categorias nas quais o artigo melhor se enquadra dentre as categorias do ACM Computing Classification System².

¹ <http://www.sbc.org.br/>

² <http://sbsi2016.ufsc.br/>

Este sistema de classificação trata-se de uma lista de categorias fornecida pela ACM, desde 1998, que é periodicamente atualizada. As categorias são distribuídas em quatro níveis, sendo o quarto composto pelo “Subject Descriptor” [1]. Considerando que um artigo pode se enquadrar em diferentes categorias de diferentes níveis, totalizam 1447 opções de categorias e descritores de assunto a serem selecionados pelos autores em suas publicações. Nas análises realizadas para os dois últimos anos do evento, cada artigo classificado pelos autores em determinada categoria foi considerado uma ocorrência daquela categoria.

- Distribuição por instituições: relação entre a quantidade de instituições brasileiras participantes do simpósio e o total de instituições de ensino brasileiras que possuem os cursos de Sistemas de Informação e Ciência da Computação (Bacharelado e Licenciatura) e Engenharia da Computação.

Antes de estabelecer os indicadores para análise da rede de colaboração, foi necessário investigar as relações entre autores e instituições neste tipo de evento. Ao se fazer essa análise, considerou-se como colaboração a publicação em conjunto de dois ou mais autores e/ou duas ou mais instituições. As colaborações foram analisadas em pares de autores; cada artigo publicado em conjunto foi considerado uma colaboração entre autores e cada artigo publicado por autores de diferentes instituições foi considerado uma colaboração entre instituições.

Na Figura 1, é ilustrado um grafo que representa parte dos dados coletados no qual um artigo foi publicado por quatro autores. No grafo, os nós verde, azul e cinza representam os autores, suas instituições e os artigos publicados, respectivamente.

Ao analisar a colaboração em pares, pode-se considerar que o ato de colaborar é bidirecional, ou seja, se o autor 22 colaborou com o autor 18, o autor 18 também colaborou com o autor 22. Contudo, para efeito de quantificação, foi considerada somente uma colaboração. Logo, para os dados da Figura 1, foram contabilizadas 6 colaborações entre autores e 1 colaboração entre instituições: o autor 22 colaborou com os autores 18, 21 e 216 (3 colaborações); o autor 18 colaborou com o autor 21 e com o 216 (2 colaborações); e o autor 21 colaborou com o autor 216 (1 colaboração).

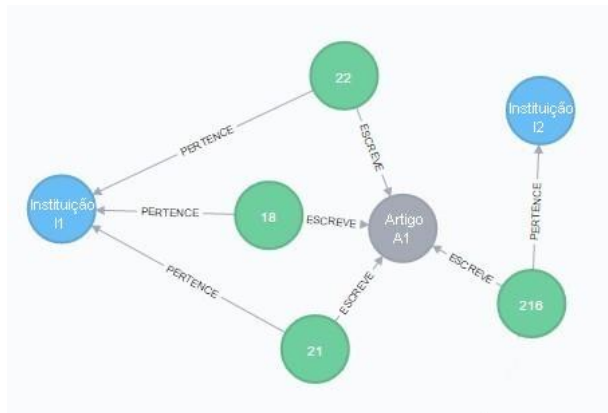


Figura 1. Exemplo de colaboração entre autores.

Depois de estabelecida a estratégia de quantificação da colaboração, foram definidos os seguintes indicadores:

- Série histórica da rede colaboração do SBSI: quantidade de colaborações entre autores e entre instituições ao longo dos nove anos considerados;
- Série histórica da rede de colaboração entre instituições públicas e privadas: quantidade de colaborações entre instituições públicas e privadas ao longo dos nove anos considerados;
- Distribuição geográfica da rede de colaboração: quantidade de colaborações entre instituições de diferentes estados e regiões do país;

3.2 Coleta e Processamento dos Dados

A coleta foi realizada recuperando os anais do evento que são disponibilizados em arquivos no formato pdf na Biblioteca Digital Brasileira de Computação³. A extração dos dados foi feita de forma manual para um arquivo de extensão .csv. Após a extração dos dados relevantes, os mesmos foram tratados para garantir integridade e unicidade das informações. Em função do volume de dados a ser analisado, não foram utilizadas técnicas computacionais como desambiguação de nomes. Desta forma, foi feita uma análise do currículo Lattes dos autores, de forma a certificar que os mesmos não aparecessem em duplicidade na base de dados devido às possíveis variações do nome, fato que poderia impactar no resultado das análises.

Depois de coletados os dados, foi necessário investigar qual estratégia de persistência poderia beneficiar a execução das consultas necessárias para medição dos indicadores. Como resultado desta análise foi possível constatar algumas características do domínio:

- Foco em relacionamentos (instituições-autores-artigos-categorias): todos os indicadores envolvem a manipulação dos relacionamentos
- Dados fortemente ligados
- Necessidade de consultas complexas

De acordo com [4], o modelo orientado a grafos pode ser útil nesse cenário. Para [6], a organização do grafo facilita a interpretação dos dados de formas diferentes baseadas nos relacionamentos. Desta forma, optou-se por usar um sistema gerenciador de banco de dados baseado nesse modelo, o Neo4j.

O próximo passo foi definir a estratégia de modelagem lógica/física dos dados. Nessa etapa, o objetivo foi analisar qual a melhor abordagem para representar os dados no modelo orientado a grafos de forma que as consultas definidas a partir dos indicadores pudessem ser realizadas de maneira adequada.

Os nós do grafo representam as instituições (azul), artigo (cinza), autor (verde) e as categorias (amarelo). Os significados das arestas podem ser facilmente entendidos na observação da Figura 2: autor **escreve** um artigo, autor **pertence** a uma instituição e artigo **possui** uma categoria. Por uma limitação de espaço, as propriedades dos nós não serão detalhadas.

³<http://www.lbd.dcc.ufmg.br/dbdcomp/servlet/PesquisaEvento?evento=sbsi>

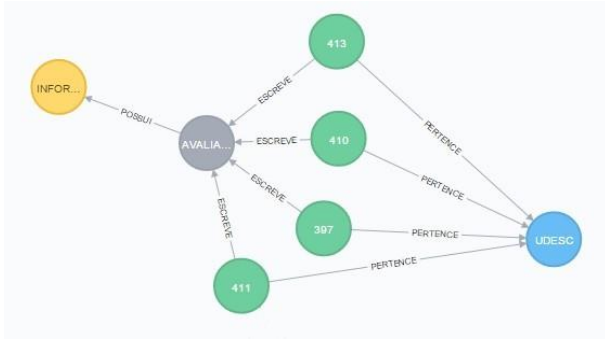


Figura 2. Estrutura dos dados no Neo4j dos anais do SBSI.

Foram encontrados casos em que um autor pertencia a mais de uma instituição e publicava diferentes artigos com filiações diferentes. Por exemplo, um autor A, filiado às instituições I1 e I2, publicou um artigo AR1 pela I1 e outro artigo AR2 pela I2. Este tipo de relacionamento foi identificado para mais de um autor. Por este motivo, o relacionamento **escreve** foi qualificado através de uma propriedade referente à Instituição pela qual ele publicou aquele artigo.

4. ANÁLISE DOS RESULTADOS

4.1 Análise da Abrangência

A Figura 3 ilustra a participação das cinco regiões do Brasil. Já a Figura 4 ilustra a série histórica desta distribuição. De um total de 519 artigos publicados por autores de instituições brasileiras, a região que mais contribuiu para o evento foi a região Sudeste participando de 245 publicações nas nove últimas edições do evento; o estado da região que mais se destaca é o estado do Rio de Janeiro com 119 publicações. No Nordeste, segunda região com maior índice de participação, o estado com mais participações foi Pernambuco, com 64 publicações. Os estados do Amapá, Roraima, Rondônia e Tocantins foram os únicos estados do Brasil que não participaram do SBSI nas últimas nove edições.

É de se esperar que a distribuição das instituições participantes por região seja semelhante à distribuição da quantidade de artigos por regiões e isto de fato ocorreu, uma vez que as regiões Sudeste e Nordeste foram as duas regiões com maior número de artigos e também as regiões com maior índice de instituições participantes do evento.

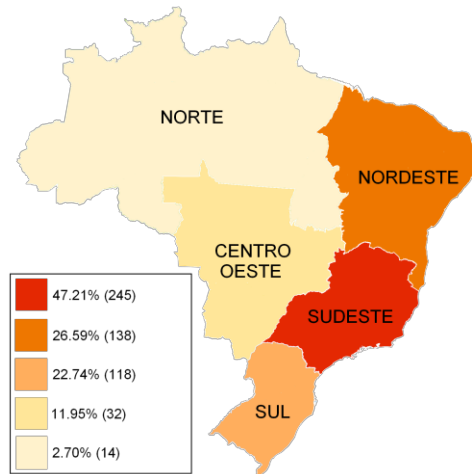


Figura 3. Quantidade de artigos do SBSI por região.

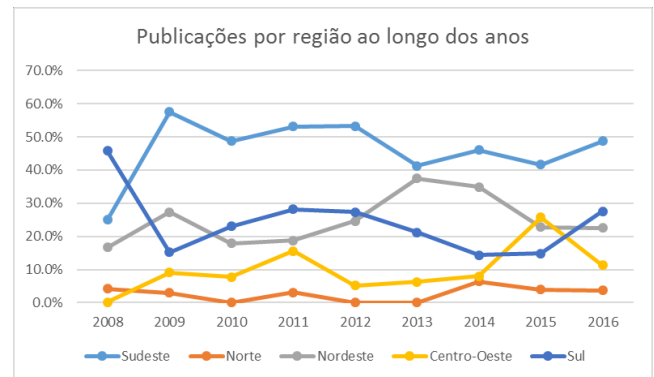


Figura 4. Série histórica da distribuição de artigos por regiões.

É possível observar que a participação das regiões, quando analisada ano a ano, não apresenta grande variação em relação ao cenário geral do simpósio ilustrado na Figura 3. A região Sudeste tem sido a protagonista em termos de publicação de artigos e a região Norte a região com menor percentual de publicação em todos os anos do evento desde 2009. Já a região Nordeste, que no cenário geral aparece como a segunda maior colaboradora do simpósio, reverte esta posição com a região Sul quando analisado ano a ano.

Ao associar esses dados com a análise dos locais de realização do evento, é possível fazer algumas considerações: as edições dos eventos anteriores a 2008 foram sediadas na região Sul o que pode justificar a quantidade significativa de contribuições em 2008. Neste mesmo ano (2008), o evento teve a região Sudeste como sede sendo possível observar que, comparada a 2008, a contribuição da região Sudeste aumentou nos anos seguintes. Todas as cinco regiões do país já sediaram o evento, porém a região Norte foi a única que sediou apenas uma vez. Além disso, muitas regiões tiveram um aumento em seu índice de participação no evento no ano em que foram sede ou no ano seguinte.

Portanto, uma possível maneira de aumentar a distribuição da participação por região e incentivar regiões com menos contribuições, como a região Norte, seria estabelecer uma estratégia em que a escolha da sede considerasse esses dados.

Em relação às categorias, dos 179 artigos publicados nas duas últimas edições do evento, três não possuem a seção referente à categoria e um não possuía categoria válida, portanto a análise baseia-se nos outros 175 artigos.

Das 1447 categorias, 237 foram utilizadas pelos autores. Na tabela 1, estão descritas as categorias que apresentaram um número de ocorrências maior que o quartil superior da amostra (maior que 4) e o nível a que pertencem [1]. A coluna ocorrência indica a quantidade de vezes que a (s) categoria (s) foi utilizada.

Das 46 categorias que aparecem na Tabela 1, 3 pertencem ao primeiro nível, 12 pertencem ao segundo nível, 17 são pertencentes ao terceiro nível e 14 pertencem ao quarto nível da lista de categorias da ACM, porém as pertencentes ao quarto nível são as com menores números de ocorrências. Considerando que o primeiro nível traz uma categoria mais genérica e o quarto nível uma categoria mais específica, observou-se que os autores optam por utilizar termos mais genéricos para classificar suas publicações.

Em [5], foi identificada certa dificuldade dos pesquisadores em classificar suas pesquisas na área de SI. No trabalho, os autores sugerem o uso de uma ferramenta para apoiar essa tarefa. Embora a utilização da classificação analisada nesse trabalho tenha sido posterior ao trabalho apresentado em [5], pode-se afirmar que esta dificuldade ainda é persistente.

É correto afirmar que a dificuldade em classificar os artigos aliada ao uso de categorias mais genéricas pode prejudicar a busca de trabalhos do evento por parte da comunidade científica, uma vez que a categoria pode ser um importante indicador das literaturas relacionadas [1]. Portanto, para ampliar o alcance dos trabalhos, pode ser útil definir uma estratégia que facilite essa categorização, por exemplo, exigir o uso de níveis inferiores da classificação disponibilizada pela ACM.

A análise relacionada à distribuição por instituições de ensino considerou apenas instituições credenciadas pelo MEC. Foram selecionadas as instituições que oferecem os cursos de Ciência da Computação, Engenharia da Computação e Sistemas de Informação.

De acordo com dados do MEC, no Brasil, há 629 instituições que possuem ao menos um dos três cursos⁴; destas, 313 possuem o curso de Ciência da Computação, 198 Engenharia da Computação e 392 o curso de Sistemas de Informação, sendo que há instituições que ofertam mais de um dos cursos.

Tabela 1. Categorias mais utilizadas no SBSI

Categorias	Ocorrências	Nível
Information systems applications	33	2
Software engineering	28	2
Management of computing and information systems	26	2
Information systems	21	1
Administrative data processing	16	2
Artificial intelligence	15	2
Computer applications	14	1
Database applications - Project and people management	12	2-3
Database management	11	2
Data mining	10	4
Government - Information storage and retrieval - Management - Software management	9	4-2-3-3
Computers and education - Life and medical sciences - Medical information systems	7	2-2-4
Applications - Computer and information science education - Human factors - Natural language processing - Requirements/specifications - Software development - Text	6	3-3-4-3-3-4-4-3-3

analysis - Types of systems - user/machine systems		
Database administration - Information interfaces and presentation - Information systems education - learning - Systems - Methodologies - Pattern recognition - Software architectures	5	3-2-4-3-3-4-2-3
Computing methodologies - Decision support - Information search and retrieval - Management techniques - Online information services - Programming teams - Public policy issues - Security and protection - Software maintenance - Staffing - Tools	4	1-4-3-4-3-4-3-3-4-4-4

A Figura 5 ilustra a distribuição destas 629 instituições por estados no Brasil, permitindo observar que grande parte dos cursos de Ciência da Computação, Engenharia da Computação e Sistemas de Informação estão concentrados na região Sudeste. Esta grande concentração de cursos de graduação na região pode ser apontada como um dos motivos de a mesma ser a maior colaboradora do evento.

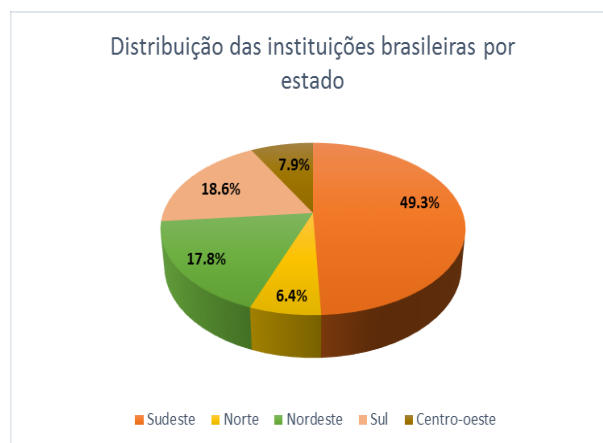


Figura 5. Gráfico da distribuição das instituições.

De um total de 629 instituições brasileiras ofertantes dos cursos relacionados à computação, apenas 114 já participaram do SBSI nas últimas nove edições. A Figura 6 ilustra a relação entre o total de instituições e o número de instituições participantes do evento por região, e permite observar que o SBSI abrange uma parcela pequena do total de instituições em todas as regiões. A consequência desta pequena abrangência do evento pode ser a limitação da rede de colaboração, que possui uma grande oportunidade de crescimento quando analisado o número de instituições que poderiam estar participando do evento e consequentemente colaborando com a ciência e disseminação do conhecimento.

⁴ <http://emec.mec.gov.br/>

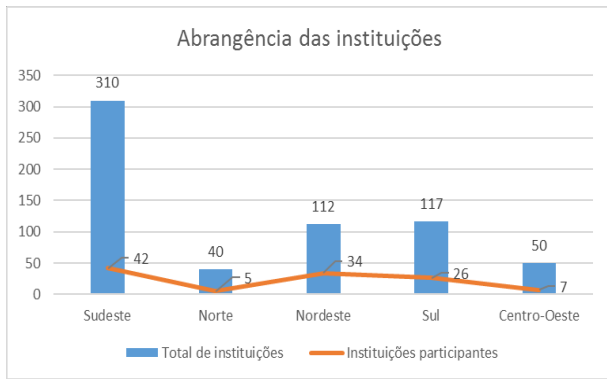


Figura 6. Gráfico da abrangência das instituições.

Uma possível medida a ser tomada para o aumento da abrangência do simpósio seria o levantamento das instituições que não possuem relação com o evento e alguma ação por parte dos organizadores de modo a apresentar o evento às mesmas e incentivá-las a participar. Uma destas ações a serem tomadas seria um estudo sobre a viabilidade de definir os locais de sede anual do evento de acordo com as regiões que menos participam ou que possuem menos colaboradores participantes do evento com o objetivo de ampliar a rede de colaboração e a amplitude do simpósio.

Um ponto a ser questionado nestas sugestões é a relação entre quantidade e qualidade. O aumento da abrangência do evento tanto em relação à localização quanto à quantidade de instituições não implica, necessariamente, em um aumento da qualidade das contribuições. Contudo, o fato de ampliar essas possibilidades pode proporcionar maiores oportunidades de discussão e debates que são recursos catalisadores de novas ideias e soluções.

4.2 Análise da Rede de Colaboração

Em relação à rede de colaboração entre os autores do SBSI, de um total de 1230 autores que já tiveram seus artigos apresentados nas edições analisadas, 132 pares já publicaram juntos em mais de uma edição. Conforme o gráfico da Figura 7, 101 pares colaboraram em duas edições do evento e apenas 1 par colaborou em seis edições. Os dois autores que colaboraram nas 6 edições do evento foram também identificados como os dois autores que mais publicaram artigos no SBSI.

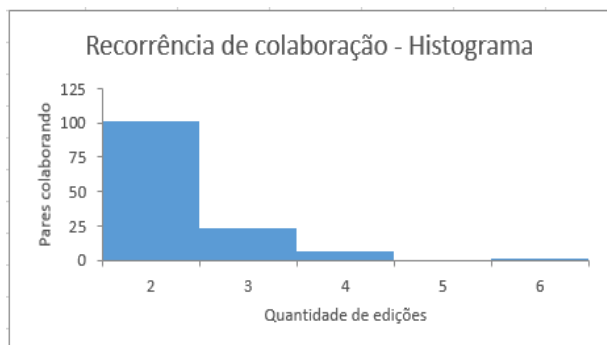


Figura 7. Histograma da recorrência de colaboração entre autores.

A Figura 8 apresenta o gráfico da evolução do número de autores participantes do simpósio e do número de colaborações entre autores por edição. O número de autores e o número de

colaborações apresentam uma variação relativamente semelhante ao longo dos anos. Este comportamento pode ser justificado pelo fato de que a grande maioria dos artigos é publicada em parceria, ou seja, por mais de um autor. Desta forma, quando há um aumento no número de autores em uma determinada edição do evento, consequentemente, pode haver um aumento no número de colaborações.

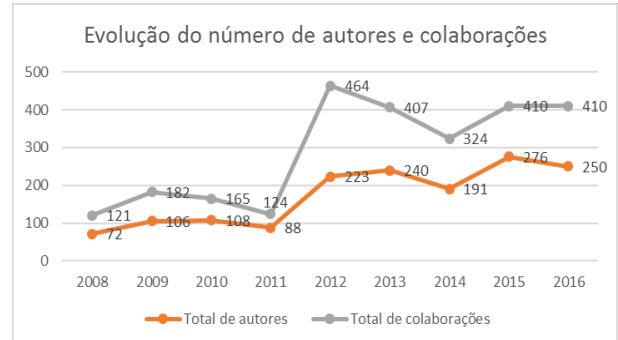


Figura 8. Gráfico da evolução do número de autores e colaborações.

A fim de mostrar a evolução da rede, foram analisadas as colaborações referentes à primeira e última edição do evento analisadas. A Figura 9 representa o grafo para o ano de 2008 e a Figura 10, a rede de colaboração no ano de 2016. Cada grupo de colaboração representa ao menos um artigo, mas pode haver casos que envolvem mais de um artigo.

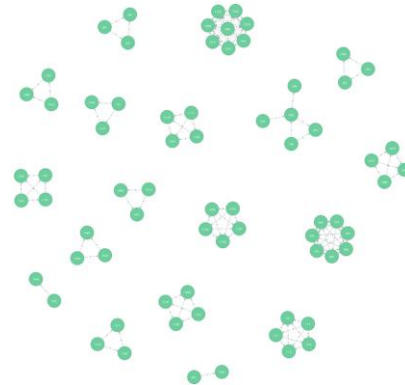


Figura 9. Grafo da rede de colaboração entre autores (2008).

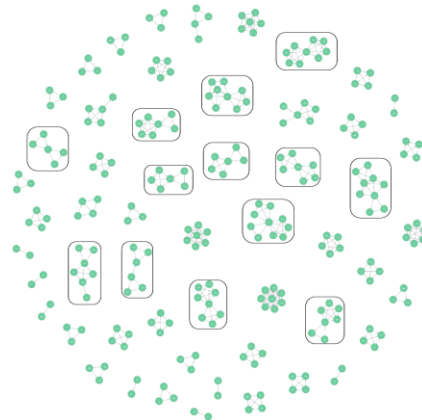


Figura 10. Grafo da rede de colaboração entre autores (2016).

Ao comparar as redes de colaboração dos anos de 2008 e 2016, além de observar o crescimento da rede como um todo, devido ao aumento do número de autores, observa-se também uma diferença de comportamento nas colaborações. Em 2008, os grupos de colaborações eram isolados, não havendo ligação entre eles. Já em 2016, esta ligação entre os grupos de colaboração é maior, conforme destacado em preto na Figura 10.

O fato de existir ligação entre os grupos de colaboração pode indicar que o autor que representa o elo de ligação da parceria é uma possível ponte para futuras colaborações entre autores destes grupos distintos, ampliando ainda mais a rede de colaboração do simpósio.

Em relação à rede de colaboração entre instituições, conforme o gráfico da Figura 11, a colaboração das instituições participantes apresentou uma evolução relativamente similar à evolução da colaboração entre autores. Contudo é pequena a quantidade de colaboração entre instituições.

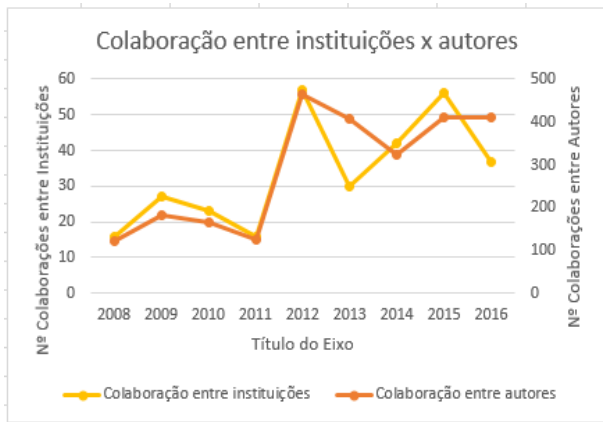


Figura 11. Gráfico de colaboração entre instituições de 2008 a 2016.

Outro ponto de análise foi a relação entre a rede de colaboração das instituições públicas e privadas. Nos últimos nove anos do SBSI, 164 instituições tiveram autores publicando artigos no evento, das quais 108 são instituições públicas e 56 instituições privadas. O número de instituições públicas, no geral, teve um aumento gradual ao longo dos anos, enquanto as instituições privadas não apresentaram um padrão, conforme ilustrado na Figura 12.

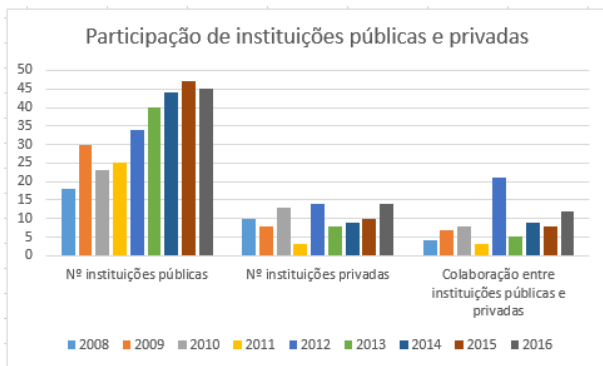


Figura 12. Gráfico da participação de instituições públicas e privadas.

Em relação à rede de colaboração entre os dois tipos de instituições estudados, esta não apresenta um peso significativo. De um total de 304 colaborações entre instituições, as

colaborações entre instituições públicas e privadas representam apenas 25% (77 colaborações). Instituições públicas são as que mais colaboram entre si, representando 41,44% das colaborações. O baixo índice de instituições privadas no SBSI e, conseqüentemente, o baixo índice de colaborações que envolvem este tipo de instituição pode estar relacionado ao fato de que instituições privadas possuem menor incentivo à pesquisa quando comparadas às instituições públicas devido à viabilidade econômica [8].

Por fim, foi analisada a rede de colaboração entre os estados e regiões do país. O gráfico da Figura 13 mostra o número de colaborações inter-regionais para cada uma das cinco regiões do país, tendo como destaque a região Sudeste com 54 colaborações inter-regionais e a região Norte com apenas uma colaboração inter-regional.

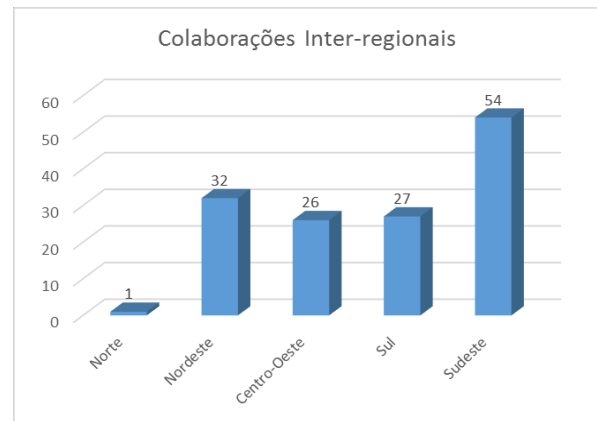


Figura 13. Gráfico de colaborações inter-regionais.

Os valores apresentados no gráfico da Figura 13 são melhores detalhados na Tabela 2 que aponta o número de colaborações entre cada região.

Tabela 2 - Colaborações entre regiões

Região R1	Região R2	Colaborações
Norte	Nordeste	1
Nordeste	Centro-oeste	5
Nordeste	Sudeste	19
Nordeste	Sul	7
Centro-oeste	Sudeste	18
Centro-oeste	Sul	3
Sudeste	Sul	17

Pode-se observar que além de ser a região com maior número de colaborações inter-regionais, a região Sudeste também é a protagonista em número de relações quando analisadas as colaborações por pares de região, apresentando 19 colaborações com a região Nordeste, 18 com a região Centro-oeste e 17 colaborações com a região Sul.

Explorando um pouco mais a rede entre as regiões Sudeste e Nordeste que apresentaram maior número de colaborações (19 colaborações), identificou-se que os estados que participam destes relacionamentos são os estados de Minas Gerais, Rio de Janeiro e São Paulo para região Sudeste e Alagoas, Bahia, Paraíba e Pernambuco na região Nordeste.

Foi possível constatar que as regiões, e até mesmo os estados responsáveis pela colaboração entre as duas regiões com maior grau de colaboração, são regiões e estados com alto índice de participação nas edições do evento. Pode-se supor que a participação de autores e universidades no SBSI, em outros eventos científicos, e a consequente interação entre autores seja um elemento que estimula a criação de novos relacionamentos.

5. CONCLUSÃO

Este trabalho apresentou uma análise da comunidade científica relacionada ao SBSI no que tange à rede de colaboração entre autores, instituições e regiões proveniente das últimas nove edições do evento, assim como dados relacionados à abrangência do evento.

Os resultados das análises realizadas para os anos de 2008 a 2016 têm como objetivo a descoberta de conhecimento relacionado ao evento de modo que os organizadores e as comissões especiais da Sociedade Brasileira de Computação (SBC) possam conhecer um pouco mais sobre o mesmo a fim de estabelecer estratégias para o aumento da qualidade dos trabalhos, no que tange à categorização, e estratégias de ampliação da abrangência do evento de modo a influenciar o aumento da rede de colaboração entre regiões, autores e instituições e, consequentemente, aumentar a propagação de conhecimento através da publicação de artigos.

Como trabalhos futuros, podem-se definir duas perspectivas. Na primeira, o objetivo é explorar o domínio. Uma alternativa é analisar outros aspectos destas redes de colaboração e comparar com outras redes relacionadas aos demais eventos da comunidade científica. Além disso, para ampliar as possibilidades de análise por parte dos organizadores do evento e da Comissão Especial de SI da SBC, poderá ser desenvolvida uma aplicação tendo a base de dados já criada como fonte, definindo estratégias para facilitar a atualização e inserção de novos dados.

A outra perspectiva está relacionada ao aspecto tecnológico. Nessa linha de evolução o objetivo é realizar um estudo comparativo com outros modelos de dados de forma a explorar mais profundamente as vantagens e ou desvantagens do modelo de dados orientado a grafos para a persistência de dados neste contexto.

6. REFERÊNCIAS

- [1] ACM. The 2012 ACM Computing Classification System. Disponível em: <<http://www.acm.org/about-acm/class/how-to-use>>. Acesso em: 11 out. 2016.
- [2] ARTMANN, E.; DIAS, V. Redes sociais do SBSI e o corte de vértices como base para identificar atores importantes na coesão de grupos de pesquisa. In: *Anais VIII Simpósio Brasileiro de Sistemas de Informação*, 8., 2012, São Paulo. Comissão Especial de Sistemas de Informação, 2012. p. 174 - 185.
- [3] BOURDIEU, P. (2004). Os usos sociais da ciência: por uma sociologia clínica do campo científico. São Paulo: UNESP. 86 p.
- [4] HECHT, R.; JABLONSKI, S. NoSQL Evaluation: A Use Case Oriented Survey. In: *Proceeding of the International Conference on Cloud and Service Computing*, 11., 2011, Hong Kong. Washington: Ieee Computer Society, 2011. p. 336 - 341.

- [5] REVOREDO, K.; ARAÚJO R.; SILVEIRA B.; MURAMATSU T. (2012). Minerando publicações científicas para análise da colaboração em comunidades de pesquisa. In: *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, Curitiba- PR.
- [6] SADALAGE J Pramod; FOWLER Martin (2013). NoSQL Essencial - Um Guia Conciso para o Mundo emergente da Persistência Poliglota. São Paulo: Novatec Editora.
- [7] SILVA, E. L. (2002). Rede científica e a construção do conhecimento. *Informação e Sociedade: Estudos*, João Pessoa, v.12, n.1, p. 120-48.
- [8] SOCZEK, D.; ALENCASTRO, M. Pesquisa acadêmica em instituições de ensino superior particulares: desafios e perspectivas. *Revista Intersaberes*, Curitiba, v. 7, n. 13, p.46-66, jun. 2012. Disponível em: <<http://www.grupouninter.com.br/intersaberes/index.php/revista/article/viewFile/247/156>>. Acesso em: 20 out. 2016.
- [9] VANZ, S. A. S.; STUMPF, I. R. C. Colaboração científica: revisão teórica-conceitual. *Perspectivas em Ciência da Informação*, v. 15, n.02, p. 42-55, mai./ago. 2010.