

Ambiente Analítico Web para Análise da Colaboração Científica no Cenário Médico

Hector N. Melo¹, Ruben A. Perorazio², Jonice Oliveira^{1,2}

¹Departamento de Ciência da Computação – Universidade Federal do Rio de Janeiro (UFRJ)

²Programa de Pós-Graduação em Informática – Universidade Federal do Rio de Janeiro (UFRJ)

rrperero@ppgi.ufrj.br, {hectornm, jonice}@dcc.ufrj.br

Abstract. *The Brazilian government implemented the National Institute of Science and Technology for Cancer Control (INCTCC) with the purpose of strategically invest in research in this area. Aiming to increase the national collaboration between scientists of the area, a Social Network Analysis approach is important. However, the typical handling and visualization mechanisms of softwares of this type are limited. The aim of this paper is to describe an analytical tool to support decision making that allow a dynamic analysis of the scientific production and collaboration through INCTCC.*

Resumo. *O governo brasileiro implantou o Instituto Nacional de Ciência e Tecnologia para o Controle do Câncer (INCTCC) com o objetivo de investir estrategicamente em pesquisa nesta área. Visando aumentar a colaboração nacional entre os cientistas da área, uma abordagem fundamentada em Análise de Redes Sociais mostra-se importante. Entretanto, os mecanismos de tratamento de dados e visualização típicos de softwares desse tipo são limitados. O objetivo deste trabalho é descrever um ambiente analítico para tomada de decisão que permita analisar dinamicamente a produção científica e colaborações realizadas através do INCTCC.*

1. Introdução

O governo brasileiro, tendo como principais objetivos amenizar as disparidades científicas entre diferentes áreas do território nacional e aumentar os índices de colaboração entre seus pesquisadores, criou os Institutos Nacionais de Ciência e Tecnologia (INCT)¹. O INCT para o Controle do Câncer (INCTCC)² faz parte de uma fração deste projeto, com foco na comunidade científica cujas pesquisas são focadas em câncer (a partir de agora chamaremos apenas de "comunidade médica"). Com isto torna-se necessária uma gestão estratégica da comunidade: identificando grupos, analisando a interação entre eles, apontando interações que deveriam existir entre grupos de subáreas similares ou complementares, apoiando o crescimento e fortalecimento de interações,

¹ INCT. http://estatico.cnpq.br/programas/inct/_apresentacao/apresentacao.html.

² INCTCC. http://www1.inca.gov.br/conteudo_view.asp?id=2630

diminuindo vícios ou comportamentos prejudiciais e estimulando a produção científica na área.

Para atingir tais objetivos, um ambiente de análise e suporte à tomada de decisão mostra-se importante para o direcionamento dos investimentos pelos gestores do projeto. Os gestores devem conseguir visualizar a interação científica entre pesquisadores de diferentes regiões do país e subáreas, tanto em termos de (i) poder entender a estrutura da rede formada, quanto (ii) poder analisar as disparidades quantitativas no que se refere à produção científica na área do câncer.

O primeiro objetivo é alcançado através de uma abordagem com base em Análise de Redes Sociais (ARS), através de uma rede de coautoria. Tal rede consiste de um grafo cujos nós são os autores dentro da comunidade médica e arestas são ligações que possuem os autores que publicaram mutuamente o mesmo artigo. Através de softwares próprios, tais como Gephi³ ou Pajek⁴, é possível visualizar facilmente a disposição de componentes da rede, a conectividade dos nós, os nós com mais ligações, ou seja, a disposição estrutural dos pesquisadores de forma geral.

Entretanto, os mesmos softwares não possuem facilidades para uma análise temporal e não conseguem prover mecanismos de visualização adequados para o segundo objetivo. Para uma análise mais abrangente, o gestor precisa de um ambiente onde possa ver a quantidade de produções acadêmicas (por exemplo, artigos publicados, orientações, projetos, etc), podendo filtrar as informações de diferentes formas – por região, estado, subárea de pesquisa, titulação dos autores, dentre outros. Outro problema se deve ao fato de que esses softwares possuem um propósito muito abrangente de prover visualização por grafo, mas sem um foco específico. A nossa ferramenta, da forma que nós construímos, supre as necessidades explicitadas anteriormente que os softwares anteriormente citados não supririam independentemente.

Neste sentido, criamos um ambiente analítico que permite analisar a rede social científica e colaborações realizadas no INCTCC. O resultado pode ser consultado on-line⁵. No mesmo site, também é possível assistir a um vídeo mostrando o funcionamento da ferramenta.

A proposta desta ferramenta vai ao encontro com a proposta da formação de um Ecosistema Digital e Estratégico. Um ecossistema digital pode ser definido como “um conjunto de espécies digitais que possuem relação entre si análogas à de um ecossistema biológico, tais como o reforço mútuo, dependência mútua ou parasitismo” [Bustamante 2010]. Neste sentido, a nossa ferramenta, como elemento de um ecossistema digital e estratégico, desempenha um papel de reforço na análise da produção científica, que é um dos pilares para o desenvolvimento tecnológico e social.

O restante deste artigo está organizado da maneira a seguir. Na seção 2, apresentamos trabalhos que se relacionam com o nosso. Na seção 3, descrevemos a nossa ferramenta detalhadamente, contemplando sua arquitetura. Na seção 4, discutimos

³ Gephi: <https://gephi.org/>

⁴ Pajek: <http://pajek.imfm.si/doku.php>

⁵ Ambiente de Análise para o INCT para o Controle do Câncer. <http://inctcc.meximas.com/>

os resultados de uma possível utilização da ferramenta pelo gestor, mostrando suas potencialidades e resultados enriquecedores desta análise. Por fim, na seção 5 apresentamos nossas considerações finais, trabalhos futuros e agradecimentos.

2. Trabalhos Relacionados

O conhecimento científico é criado e transmitido através das publicações científicas. A comunidade científica existe, dentre outros motivos, devido à necessidade de troca de ideias e conhecimento, típico da profissão [Vans 2013].

Sendo a publicação o principal meio de comunicação do conhecimento de um cientista para o mundo, ela diz muito sobre ele - principalmente sobre seus pares (coautores). Desta forma, a maneira mais prática de avaliar a troca de ideias em grande escala por um computador utilizando algoritmos é coletando tais informações de coautoria.

Existem diversos trabalhos que utilizam redes de coautoria com as mais diversas finalidades. Newman [2004], por exemplo, utilizou dados das áreas de Biologia, Física e Matemática para achar padrões de colaboração entre os pesquisadores. Outro trabalho amplo que utiliza a mesma abordagem é o de [Menezes et al 2009], que procura analisar a colaboração científica da Ciência da Computação no Brasil, América do Norte e Europa.

A análise de redes de coautoria também se mostra apropriada para o caso presente. É possível visualizar facilmente no seu grafo os componentes (i.e. grupos isolados de nós interconectados entre si), os nós com muitas ou poucas ligações, articulações (nós que fazem a interconexão entre diferentes grupos e se retirados, dividem a componente em duas), etc. Tais observações podem permitir a um gestor que ele entenda a dinâmica da pesquisa científica numa área entre grupos de pesquisa, e com isso possa elaborar uma estratégia de investimento focada nelas.

Neste sentido, Morel et al [2009] fazem uma análise focando na análise de componentes e de articulações em uma rede de pesquisadores em doenças tropicais negligenciadas dentro de regiões brasileiras onde essa pesquisa é tradicionalmente baixa (Nordeste, Centro-Oeste e Norte). Parte de um programa governamental, a intenção de seu trabalho é achar pontos cruciais da rede onde o investimento seja feito e traga maior retorno possível para a sociedade. Assim, eles priorizam, por exemplo, pesquisadores ou instituições articulações como foco do investimento. Existe outro trabalho semelhante, porém focando nas redes de pesquisadores de tuberculose [Morel and Vasconcellos 2012]. Entretanto, em seus trabalhos ele foca apenas na utilização da Análise de Redes Sociais para visualização do grafo, de forma que métricas de centralidade de grafo, como *betweenness* e *closeness*⁶, não sejam levadas em conta.

Outro trabalho fortemente relacionado trabalha também com redes de coautoria para uma comunidade médica dada dentro do INCTCC [Perorazio et al 2012]. Por um lado, este trabalho possui uma análise mais acurada do que apenas analisar coautoria de artigos científicos, pois também faz uso de participação conjunta em projetos, patentes e diferentes tipos de produção científica. Por outro lado, o foco deste trabalho não é na

⁶ Mais detalhes das métricas (*betweenness*, *closeness*, *clustering* e grau) são explorados na Seção 4.2.

produção de uma ferramenta, mas sim em uma análise pontual para prover resultados ao governo.

A Plataforma Lattes⁷ disponibiliza aos pesquisadores um grafo de colaboração. Entretanto, este grafo é restrito ao próprio usuário, e portanto, não pode ser usado para fins de análise por outra pessoa, que é o caso desta ferramenta.

Outros trabalhos utilizam uma abordagem semelhante à nossa para análise de colaboração científica em saúde, com destaque para: alzheimer [Sorensen et al 2010], doenças infecciosas e microbiologia clínica [González-Alcaide et al 2010] e cardiologia [Valderrama-Zurián 2007], entretanto nenhum deles possui foco na criação de uma ferramenta para a tomada de decisão, que é o foco do presente artigo.

3. Ferramenta

A ideia de usar uma abordagem em que os dados fiquem disponíveis online tem o objetivo de tornar o uso fácil e possível de se fazer em qualquer máquina pelos coordenadores do INCT para o Controle do Câncer. A base de dados é atualizada quinzenalmente, para permitir uma análise mais acurada.

Vamos mostrar um exemplo de utilização da ferramenta. Suponha que o coordenador deseja visualizar o *grau médio* dos pesquisadores – isto é, o número médio de pessoas com quem um pesquisador publica um trabalho científico – por titulação, dentro de um estado. Como mostra a Figura 1, para fazer isso, o coordenador posiciona "Maior Formação" e "Estado" nas Colunas e "Grau" nas Linhas. É possível observar, por exemplo, que em média, os pesquisadores de nível de Graduação possuem um grau maior que todos os outros, exceto para doutores no Rio de Janeiro e pós-doutores na Bahia, um padrão que mostra que a pesquisa nas titulações mais altas não está muito articulada. Dentre algumas possibilidades, o coordenador pode (1) salvar essa consulta e (2) abri-la em uma análise posterior ou (3) alternar para o modo gráfico rapidamente, ou até mesmo (4) exportar os dados das consultas para os formatos .XSL, .CSV e .PDF. Seguindo este raciocínio, o coordenador pode fazer e armazenar variadas consultas neste ambiente.

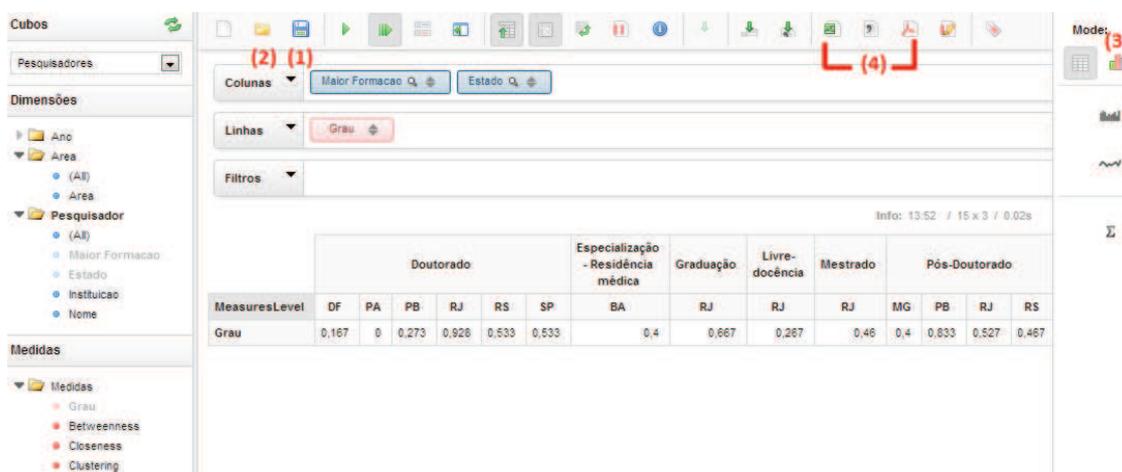


Figura 1. Exemplo de utilização da ferramenta e algumas de suas funções

⁷ Plataforma Lattes. <http://lattes.cnpq.br/>

Nas seguintes subseções, apresentamos detalhes da arquitetura da nossa ferramenta, dos dados utilizados, da modelagem multidimensional do Data Warehouse, e da constituição do site em si.

3.1. Arquitetura

A arquitetura da nossa ferramenta pode ser observada na Figura 2. A imagem baseia-se na representação de [Kimball 1998].

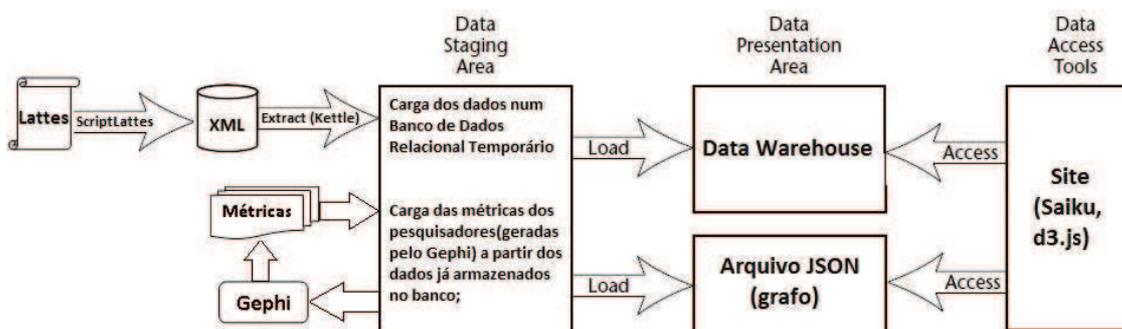


Figura 2. Arquitetura do Ambiente Analítico

Primeiramente, obtivemos a lista de pesquisadores envolvidos através do coordenador do INCA para o Controle do Câncer. Usamos os pesquisadores desta lista como entrada para o ScriptLattes [Mena-Chalco and Cesar-Jr 2009], o qual extrai dados da Plataforma Lattes e nos fornece como saída um arquivo XML estruturado com as informações presentes nos currículos dos pesquisadores. O XML passa então por um processo ETL (*Extract, Transform, Load*), onde são extraídas suas informações e inseridas em um banco de dados (utilizamos o Pentaho Kettle⁸ para o processo ETL e o Microsoft SQL Server⁹ para armazenar o banco de dados). Com as informações presentes no banco de dados, fomos capazes de gerar uma consulta de grafo no Gephi, onde obtivemos métricas de Análise de Redes Sociais (*betweenness, closeness, clustering* e grau) relacionadas a esses pesquisadores, e as exportamos para uma planilha .CSV. Tais métricas foram carregadas no mesmo banco de dados, cujas informações são usadas para a carga em um Data Warehouse (DW) e para um arquivo .json para a exibição do grafo no site. Este DW (também armazenado no Microsoft SQL Server) é utilizado para as análises multidimensionais pela ferramenta no nosso site. Detalhes da implementação do site serão mais explorados na Seção 3.4.

3.2 Dados Extraídos

A Plataforma Lattes possui uma grande quantidade de dados dos pesquisadores, inclusive as suas publicações e coautorias destas. Desta forma, extraímos dados referentes às publicações e também outros importantes para uma filtragem de dados que possibilitem a análise do gestor, tais como: estado, instituição, grupo de pesquisa, linha de pesquisa e área de atuação, maior titulação (Graduação, Especialização, Mestrado, Doutorado e Pós-doutorado).

⁸ Pentaho Kettle (open source), mais informações em: <http://www.pentaho.com>

⁹ Microsoft SQL Server. <https://www.microsoft.com/sqlserver/pt/br/default.aspx>

A comunidade médica conta com 118 pesquisadores envolvendo apenas 5 estados brasileiros e em sua maioria Doutores (101 pesquisadores ou 85,6%), sendo a maioria também do Rio de Janeiro (92 pesquisadores ou 78%). Tendo isso em mente, o modelo multidimensional foi modelado da forma que é descrito na seção 3.3.

3.3 Modelo Multidimensional

Criamos um modelo multidimensional para o nosso Data Warehouse com o intuito de uma análise dinâmica, de forma que o coordenador do projeto seja capaz de cruzar diferentes dimensões e não ficar restrito à análise do grafo. O modelo pode ser observado na Figura 3.

O nosso modelo conta com quatro Dimensões – Publicação ("publicacao"), Ano ("ano", Dimensão do Tempo no nosso modelo), Área ("area") e Pesquisador ("pesquisador").

O modelo possui também duas tabelas Fatos ("fato_publicou" e "fato_metricas"). Isso se deve à natureza das diferentes informações com que trabalham as duas tabelas. No primeiro caso, estamos apenas contabilizando a publicação usando diferentes critérios (área, ano, etc), e conseqüentemente, as dimensões levam apenas eles em conta. No segundo caso, queremos avaliar apenas os pesquisadores através de métricas de Análise de Redes Sociais (*betweenness*, *closeness*, *clustering* e grau), então as únicas dimensões envolvidas para a tabela de Fatos de Métricas são as Dimensões do Pesquisador, Área e Tempo. Desta forma, é possível analisar a evolução das métricas ao longo do tempo e por área, assim como usando outros atributos do pesquisador (filtro por titulação, por estado e por instituição).



Figura 3. Modelo Multidimensional do Data Warehouse

3.4 Implementação do Site

A estrutura do site possui duas aplicações primordiais: uma para possibilitar um ambiente OLAP e outra para exibição de um grafo de pesquisadores. Para a interface OLAP, utilizamos o projeto open source Saiku¹⁰, que funciona como uma aplicação em Java¹¹ no servidor. Utilizamos sua base para o DW e customizamos ele para suprir as nossas necessidades.

Para o grafo personalizado, utilizamos o *Force-Directed Graph* disponível na biblioteca D3.js¹², também open source. Feito em Javascript, ele utiliza o algoritmo de *Force-Directed Graph*, um tipo de exibição dinâmica de um grafo que utiliza física para simular atração entre os nós ligados entre si. Esse tipo de visualização permite que o usuário arraste o grafo e veja os nós movendo-se de forma agradável. O algoritmo para construção deste tipo de layout também permite a exibição de simetrias entre os nós e tende a criar layouts livres de cruzamentos quando isso é possível [Kobourov 2012].

O *Force-Directed Graph* do D3.js também possui a possibilidade de criação de grupos de nós através do uso de cores. Na nossa ferramenta, colocamos cores para representar pesquisadores de diferentes estados.

4. Resultados

Nesta seção, pretendemos mostrar os resultados de uma análise feita com o uso da nossa ferramenta, utilizando seus diferentes recursos (grafo de colaboração, tabelas e gráficos estatísticos). O foco desta análise é mostrar as potencialidades da nossa ferramenta aplicada a um exemplo real, simulando uma possível utilização pelos gestores do INCTCC.

4.1 Grafo de Colaboração

O grafo de colaboração é mostrado na Figura 4. Podemos notar algumas informações a respeito do grupo formado pelo INCTCC: em maioria é formado por pesquisadores do Rio de Janeiro e em minoria pelos demais. O maior componente possui em sua maioria pesquisadores do Rio de Janeiro, mas também alguns pesquisadores de outros estados em posição de destaque na rede social. Por exemplo, o nó “Pesquisador”¹³ (nó em laranja destacado no grafo) revela ser do Rio Grande do Sul e fortemente ligado a outros pesquisadores de não apenas do Rio de Janeiro, mas também da Bahia e Distrito Federal.

Ainda podemos notar também que existem 14 nós que não possuem ligação alguma entre si, além de diversos nós com 3 ou menos ligações entre si, demonstrando uma fraca interação entre diversos pesquisadores da comunidade. Ao mesmo tempo, vemos a formação de pequenos grupos, podendo identificar três diferentes tipos de caso em (1), (2) e (3). Em (1), o grupo de pesquisadores do Rio de Janeiro tem uma

¹⁰ Saiku (open source): <http://meteorite.bi/saiku>

¹¹ Java: http://www.java.com/pt_BR/

¹² D3js: <http://d3js.org/>, e o Force-Directed Graph: <http://bl.ocks.org/mbostock/4062045>

¹³ O nome real do pesquisador foi ocultado por questões de privacidade.

colaboração forte restrita aos seus membros. Em (2), temos um grupo que participa do maior componente do grafo, mas se restringe a colaborar com poucos nós fora de seu grupo. Em (3), temos algo semelhante, entretanto os pesquisadores são da Bahia. Dada a distância física que este estado tem com o Rio de Janeiro, onde estão a maioria dos pesquisadores, este é um fator que pode dificultar a colaboração.

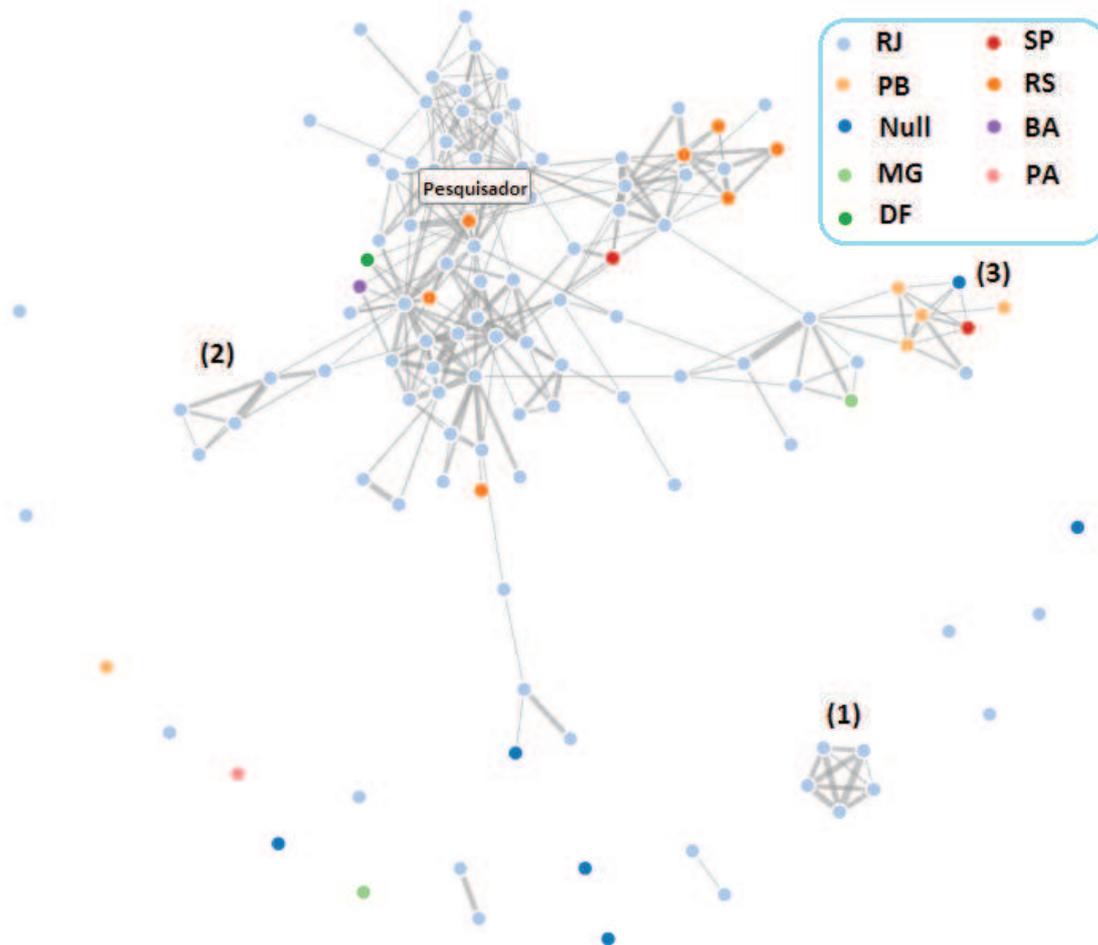


Figura 4. Modelo Multidimensional do Data Warehouse

Tendo em mente este cenário e que o principal objetivo é tornar a rede com o maior número de relacionamentos possível, podemos sugerir medidas que possam aumentar a quantidade de relacionamentos, principalmente focando nos (i) nós desconexos do maior componente e (ii) nos grupos fechados. Pensando na possível dificuldade de colaboração científica entre pesquisadores de diferentes estados, a primeira medida seria incentivar a colaboração entre pesquisadores presentes no maior componente e outros desconexos deste, desde que sejam do mesmo estado. Com uma prioridade menor, os grupos fechados também receberiam o mesmo incentivo. O incentivo pode se dar através do investimento em seminários, simpósios, co-orientações ou até mesmo programas que ofereçam bolsas para a colaboração.

Entretanto, não são todos os pesquisadores que são do mesmo estado, de forma que o investimento também poderia focar em oferecer bolsas para estimular a produção entre pesquisadores de diferentes estados.

4.2 Ambiente OLAP

Uma visualização que podemos obter – cruzando os dados das áreas das publicações dos pesquisadores, o ano e a quantidade de publicações – é mostrada na Figura 5. Para simplificar o conteúdo e não poluir a imagem com informações, filtramos os resultados cujo Ano é maior ou igual a 2006 e apenas as áreas que são mostradas no gráfico. Os nomes das áreas foram editados e ampliados na imagem a fim de ficarem mais visíveis.

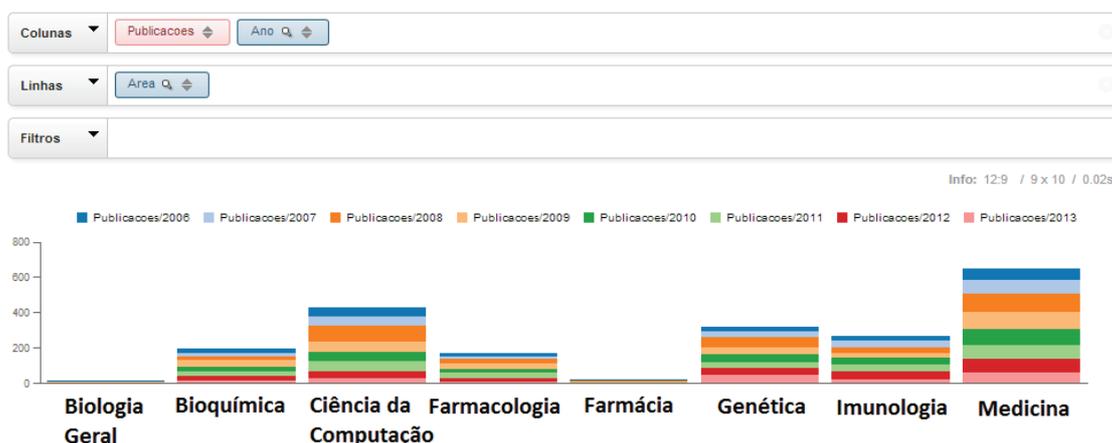


Figura 5. Distribuição de Pesquisa por Área

Como podemos ver, a pesquisa na área do câncer tem alta interdisciplinaridade, envolvendo diversas áreas da saúde e até mesmo Ciência da Computação. Entretanto, podemos notar algumas disparidades neste gráfico. A área da farmacologia, com o foco na produção de medicamentos, tem pesquisa baixa se comparada com outras áreas – tais como Imunologia, Medicina e Ciência da Computação. Uma vez que este artigo não procura identificar quais são as áreas de pesquisa prioritárias para o governo – uma vez que somos profissionais da Computação –, cabe aos coordenadores utilizar tais informações para definir prioridades.

Outro fator para se levar em conta ao investir em uma dessas áreas são as métricas específicas para os pesquisadores destas áreas, como definimos a seguir. O *grau* de um nó é o número de ligações que ele tem com outros nós. O *betweenness* de um nó, por exemplo, mede quantitativamente o seu papel como intermediário na rede. Em outras palavras, o *betweenness* é maior quando um nó faz parte de vários caminhos entre nós do grafo. O *closeness*, por sua vez, está interessado em medir a "proximidade" que um nó tem em relação ao restante da rede. Para este cálculo, é necessário calcular a distância do nó em questão para todos os outros nós da rede. Um valor de *closeness* alto indica que o nó possui uma alta influência na rede, já que consegue alcançar outros nós com poucas arestas. Por fim, o *clustering* indica a tendência que um nó tem de formar um grupo, podendo-se isolar ou não do resto da rede [Wasserman, 1994].

Na Figura 6, temos uma consulta cruzando informações da Área (com os mesmos filtros anteriores), Maior Formação, *Closeness* e *Betweenness* médio dos pesquisadores para aquela área/formação (células vazias indicam a inexistência de pesquisadores com tal formação para aquela área). Notamos um *betweenness* e *closeness* relativamente alto para a área de Farmacologia, muito embora ela tenha uma produção relativamente baixa, como vimos anteriormente. Isso indica que esta área,

embora não tenha alta produtividade, tenha um papel fundamental na estrutura da rede, servindo como intermediário importante entre outras colaborações em artigos. Por isso, mostra-se extremamente conveniente investir em pesquisadores envolvidos nela para mantê-la como intermediadora e ao mesmo tempo aumentar a sua produtividade. O mesmo vale para a área de Medicina.

Maior Formacao	MeasuresLevel	Biologia Geral	Bioquímica	Ciência da Computação	Farmacologia	Farmácia	Genética	Imunologia	Medicina
Doutorado	Betweenness		0		10,507	0	3,208	2,582	10,238
	Closeness		0		0,206	0	0,204	0,709	0,314
Especialização - Residência médica	Betweenness								0,151
	Closeness								0,201
Livre-docência	Betweenness								18,2
	Closeness								0,274
Mestrado	Betweenness					2,133			0
	Closeness					0,636			0,073
Pós-Doutorado	Betweenness	0	11,533	0	15,905	4,136	5,462	9,043	9,135
	Closeness	0,267	0,293	0,067	0,438	0,782	0,185	0,358	0,229

Figura 6. Métricas por Área de Pesquisa e Área de Formação

Entretanto, é importante investir em outras áreas com comportamentos opostos também ou ao contrário estaríamos aumentando as desigualdades na rede. Por isso, cabe ao coordenador saber em qual área com métricas pequenas ele irá investir.

É importante notar também a diferença na maior formação. A área da Medicina e da Farmácia possuem alguns profissionais com *betweenness* e *closeness* baixo. Entretanto, se programas de formação (mestrado, doutorado e pós-doutorado) forem realizados intencionalmente para estes pesquisadores, isso os incentivará dentro da rede, uma vez que eles se beneficiarão com uma titulação mais alta ao mesmo tempo em que contribuirão para uma rede mais conectada. É importante, portanto, também investir em profissionais de titulação mais baixa.

4.3 Aliança entre as abordagens anteriores

Vimos na primeira abordagem uma possível solução para melhorar a estrutura da rede social, aumentando o índice de colaboração entre os nós. Entretanto, essa abordagem sozinha não leva em conta outros detalhes, tais como as Áreas de Pesquisa abordadas e a titulação dos profissionais da área. Desta forma, a aliança entre os resultados das duas análises é capaz de nortear o coordenador para um investimento mais focado em resolver os problemas relatados. Além disso, com a nossa ferramenta, o coordenador pode fazer outros diversos tipos de consultas que não as anteriores.

5. Considerações Finais e Trabalhos Futuros

Pudemos observar que a análise estrutural da rede de colaboração aliada à análise quantitativa das produções científicas trazem um panorama claro da rede científica objeto de estudo, orientando um investimento estratégico. A nossa ferramenta cumpre o papel de possibilitar tais análises de maneira simples e intuitiva.

Dado que esta ferramenta é relativamente nova, a intenção é de que incrementemos ela para uma análise com mais parâmetros no futuro. Por isso, pretendemos colocar novas Dimensões no nosso modelo multidimensional que possam oferecer mais consultas. Tais dimensões se referem à titulação do pesquisador

(Graduação, Especialização, Mestrado e Doutorado), ao estado onde ele mora no Brasil (ou se é pesquisador de fora do país), tipos de publicação científica (artigo completo em periódico, artigo completo em conferência, resumos expandidos e outras produções bibliográficas). Além disso, outra evidência importante de colaboração entre pesquisadores e/ou instituições é a participação em projetos em conjunto. Este dado pode ser levado em conta em projetos futuros também.

Outra intenção futura é agregar diferentes mecanismos de visualização de informação para grandes quantidades de dados à essa ferramenta. No momento ela apenas explora duas abordagens (grafos e OLAP), mas é possível obter maior entendimento da informação uma vez que expandirmos a quantidade de visualizações da ferramenta.

Além disso, pretendemos explorar mais possibilidades e capacidades do *Force-Directed Graph*. Pretendemos deixar a cargo do usuário a personalização de propriedades dos nós, tais como o uso de cores para categorizar o que ele quiser (não só o estado, mas outros parâmetros, como titulação), ou o uso de diferentes tamanhos dos nós para denotar alguma informação numérica (por exemplo, a quantidade de publicações, ou alguma métrica de análise de redes sociais).

Por fim, vale ressaltar que todos os sites incluídos nas notas de rodapé foram acessados em dezembro de 2013.

Agradecimentos. Gostaríamos de agradecer ao Instituto Nacional de Ciência e Tecnologia para Controle do Câncer e ao CNPq pelo apoio e concessão de bolsas (processo: 573806/2008-0) e também à FAPERJ (processo: E26/170.026/2008). Também agradecemos à FAPERJ através do Jovem Cientista do Nosso Estado (processo: E_23/2013) e ao CNPq pela Bolsa PQ2 (processo: 308219/2010-4) e pelo Edital Universal (processo: 487239/2012-1)

6. Referências

- Bustamante, J. (2010). Poder comunicativo, ecossistemas digitais e cidadania digital. Cidadania e redes digitais. São Paulo: Comitê Gestor da Internet. Available at http://www.cidadaniaeredesdigitais.com.br/_pages/artigos_01.htm.
- González-Alcaide, G., Valderrama-Zurián, J. C., & Ramos-Rincón, J. M. (2010). Producción científica, colaboración y ámbitos de investigación en Enfermedades Infecciosas y Microbiología Clínica (2003–2007). *Enfermedades Infecciosas y Microbiología Clínica*, 28(8), 509-516.
- Kimball, R. (Ed.). (1998). *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*. New York: Wiley.
- Kobourov, S. G. (2012). Spring Embedders and Force Directed Graph Drawing Algorithms. arXiv preprint arXiv:1201.3011.
- Mena-Chalco, J. P., Cesar Junior, R. M. scriptLattes: An open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, vol. 15, n. 4, páginas 31--39, 2009.

- Menezes, G. V., Ziviani, N., Laender, A. H., & Almeida, V. (2009, April). A geographical analysis of knowledge production in computer science. In *Proceedings of the 18th international conference on World wide web* (pp. 1041-1050). ACM.
- Morel, C. M., Serruya, S. J., Penna, G. O., Guimaraes, R. (2009) "Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research", *Development and Capacity Building Programmes on Neglected Diseases*. In: *PLoS Neglected Tropical Diseases*, 3(8): e501.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National academy of Sciences of the United States of America*, 101(Suppl 1), 5200-5205.
- Perorazio, R. A., Faria, F. F., Monclar, R., Oliveira, J., Souza, J. M. (2012) "Estudando Dinâmicas de Grupo Através da Utilização da Análise de Redes Sociais em uma Comunidade Médica". In: *Brazilian Workshop on Social Network Analysis and Mining*; 2012; Curitiba, PR.
- Sorensen, A. A., Seary, A., & Riopelle, K. (2010). Alzheimer's Disease Research: A COIN Study Using Co-authorship Network Analytics. *Procedia-Social and Behavioral Sciences*, 2(4), 6582-6586.
- Valderrama-Zurián, J. C., González-Alcaide, G., Valderrama-Zurián, F. J., Aleixandre-Benavent, R., & Miguel-Dasit, A. (2007). Coauthorship Networks and Institutional Collaboration in *Revista Española de Cardiología Publications*. *Revista Española de Cardiología (English Edition)*, 60(2), 117-130.
- Vanz, S. A. D. S. (2013). *Redes Colaborativas nos Estudos Métricos de Ciência e Tecnologia | Collaborative Networks in Metric Studies of Science and Technology*. *Liinc em Revista*, 9(1).
- Vasconcellos, A. G., & Morel, C. M. (2012). Enabling Policy Planning and Innovation Management through Patent Information and Co-Authorship Network Analyses: A Study of Tuberculosis in Brazil. *PloS one*, 7(10), e45569.
- Wasserman, S. (1994). *Social network analysis: Methods and applications*(Vol. 8). Cambridge university press.