

Apoiando a Composição de Serviços em Ecossistemas de Software Científico

Phillipe Marques¹, José Maria N. David¹, Victor Ströele¹, Regina Braga¹,
Fernanda Campos¹, Marco Antônio P. Araújo¹

¹Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora
(UFJF) CEP: 36036-900 – Juiz de Fora – MG – Brazil

phillipeisrael@ice.ufjf.br,

{jose.david,victor.stroele,regina.braga,fernanda.campos,marco.araujo}
@ufjf.edu.br

Abstract. *Social Networks allow detecting relationships between people in collaborative environments. The analysis of these relationships arises as an important mechanism to enhance collaboration among research groups, through the development of services, aiming to support the conduct of scientific experiments. In a Scientific Software Ecosystem context, there is a need to compose services obtained from geographically distributed repositories. Therefore, the creation of services is a complex activity and needs adequate support. This work aims to present a solution to support the composition of services from the use of visualization of multi-relational social networks.*

Resumo. *Redes Sociais permitem detectar relações entre pessoas em ambientes colaborativos. A análise dessas relações surge como um importante mecanismo para potencializar a colaboração entre os grupos de pesquisa, através da criação de serviços, visando apoiar a realização de experimentos científicos. No contexto de um Ecossistema de Software Científico, é necessário compor serviços obtidos de repositórios geograficamente distribuídos. Portanto, a criação de serviços é uma atividade complexa e necessita de um suporte adequado. Este trabalho tem como objetivo apresentar uma solução para apoiar a composição de serviços a partir da utilização de visualização de redes sociais multi-relacionais.*

1. Introdução

A área de *e-Science* tem como objetivo auxiliar cientistas na realização de experimentos científicos frequentemente conduzidos por equipes geograficamente distribuídas. Essas equipes podem reutilizar serviços web, ou outros tipos de serviços, quando trabalham em domínios de aplicação similares. O suporte à colaboração é necessário, por permitir a coordenação de atividades na execução de tarefas que compõem a experimentação. Além disso, oportunidades de colaboração podem ser úteis por permitir que cientistas interajam com potenciais parceiros em suas pesquisas. Neste contexto, elementos que criam oportunidades de colaboração na área científica podem ser úteis também para o apoio à decisão. Consideremos, por exemplo, um contexto no qual cientistas de diferentes institutos de pesquisa estão geograficamente distribuídos. Se estes cientistas possuírem uma relação de coautoria ou trabalharem em áreas similares, uma oportunidade de colaboração pode surgir. Isto permite a eles reutilizarem os mesmos serviços científicos, ou serviços semelhantes.

Diante da necessidade de auxiliar cientistas na realização de experimentos científicos, a plataforma E-SECO (do inglês *E-Science Software Ecosystem*) foi proposta (Freitas et al., 2015). Esta plataforma se baseia na abordagem de um Ecossistema de Software Científico (ECOSC). Manikas (2016) define um Ecossistema de Software (ECOS) como: “a interação entre software e ator em relação a uma infraestrutura tecnológica comum, que resulta em um conjunto de contribuições e influências direta ou indireta em relação ao ecossistema”. Freitas et al. (2015) caracterizam um ECOSC por suas relações com fornecedores de software científico, institutos de pesquisa, órgãos de fomento, instituições financiadoras e partes interessadas nos resultados de pesquisa.

A plataforma E-SECO possui um processo de desenvolvimento que permite construir serviços e adicioná-los à plataforma (Arakaki et al., 2016). O processo de desenvolvimento inclui as seguintes atividades: (i) Planejamento inicial, (ii) Busca e recuperação de serviços, (iii) Criação de serviços, e (iv) Composição de serviços. Esta última permite que serviços que foram utilizados em experimentos anteriores, possam ser reutilizados e compostos. Para tanto, é necessário prover suporte à colaboração entre diferentes grupos de pesquisa. Neste contexto, a identificação de relações entre diferentes cientistas, geograficamente distribuídos, é importante para auxiliar a avaliação destas novas composições. Acreditamos que uma solução que permita analisar e visualizar redes sociais científicas seja capaz de prover informações para auxiliar o desenvolvimento de serviços, especialmente, a atividade de composição de serviços. Além disto, um suporte à interoperabilidade entre serviços e a possibilidade de identificar dependências entre serviços, em uma composição, auxilia na avaliação das composições realizadas. A interoperabilidade pode ser definida como a capacidade de diferentes sistemas usarem os serviços uns dos outros efetivamente (Pokraev, 2009).

Ströele et al. (2017) apresentam uma ferramenta que permite realizar a análise de uma rede social científica baseada em visualizações. Para realizar esta visualização, são utilizadas técnicas de agrupamento, baseadas em mineração de dados. Embora esta ferramenta permita apresentar uma visualização de redes científicas, a sua utilização não está integrada a um ECOSC, e não foi proposta para apoiar o desenvolvimento de serviços integrados a uma plataforma de ecossistema de software.

O objetivo deste trabalho é apresentar um serviço, denominado SCView, cujo objetivo é apoiar a composição de outros serviços em uma plataforma de ECOSC. Esta solução é baseada na visualização de serviços e na análise de redes sociais. Através desta análise e de elementos de visualização, pesquisadores podem interagir para avaliar as composições realizadas. No contexto do mundo real, as redes sociais são principalmente multi-relacionais, isto é, as pessoas ou instituições estão relacionadas através de diferentes tipos de relacionamentos. Esta visualização busca identificar redes de pesquisadores e, através da comunicação entre estes cientistas, apoiar a tomada de decisões sobre as composições de serviços realizadas na plataforma E-SECO. Através dos resultados obtidos, buscamos melhorar o processo de desenvolvimento de serviços na plataforma potencializando a colaboração na realização de experimentos científicos.

Como contribuição, este trabalho apresenta uma solução integrada a uma plataforma de ecossistemas para apoiar a colaboração no desenvolvimento de serviços. Esta solução é baseada na análise de redes sociais científicas e na visualização da composição de outros serviços.

Este trabalho está organizado da seguinte maneira: a seção 2 apresenta trabalhos relacionados. A seção 3 apresenta a plataforma E-SECO. A seção 4 detalha o modelo adotado para análise de redes sociais científicas multi-relacionais, destacando seus conceitos. A seção 5 apresenta a ferramenta de visualização para redes sociais, realizando uma análise sobre a rede. A seção 6 apresenta uma avaliação da solução para apoiar a composição de serviços, destacando como equipes distribuídas, identificadas a partir das redes sociais, colaborem na avaliação de serviços compostos na plataforma. Por fim, a seção 7 apresenta as conclusões obtidas e trabalhos futuros.

2. Trabalhos Relacionados

O crescimento das redes sociais deve-se à evolução da Web. Em muitos problemas de redes sociais científicas, apenas a coautoria é analisada (Evans et al., 2011), porém, redes sociais podem ter outros tipos de relacionamentos. Utilizar visualização em redes sociais permite uma análise mais eficaz das informações extraídas. Isto porque a visualização fornece um modo natural para expressar a conectividade entre os elementos.

Manikas (2016) apresenta uma pesquisa através da qual identifica propostas relacionadas a redes sociais e colaborativas, no contexto de ECOS (Santos et al., 2012) (Sadi e Yu, 2014). Santos et al. (2012) apresentam uma rede sociotécnica, envolvendo as dimensões técnica e social de um ECOS. A dimensão social permite analisar como os *stakeholders* contribuem ao ECOS, através da colaboração entre eles, enquanto a dimensão técnica está focada na plataforma (Campbell e Ahmed, 2010). A evolução do processo de desenvolvimento de software é abordada por Sadi e Yu (2014). Na solução proposta pelos autores, as redes de atores são utilizadas para modelar o contexto sociotécnico de desenvolvimento. Como resultado, permite analisar todas as forças envolvidas no desenvolvimento de software, considerando aspectos técnicos e sociais. Entretanto, esses trabalhos não analisam as especificidades de uma rede científica, focando principalmente em aspectos de desenvolvimento de uma plataforma de ECOS. Além disso, não exploram a forma pela qual cientistas podem participar do processo de desenvolvimento de serviços. Especificamente, estamos interessados em apoiar o processo de composição de serviços. Esse processo deve ser capaz de ajudar equipes de cientistas a reutilizar os serviços para a composição de novos serviços. No contexto de um ECOS, a composição de serviços obtida de diferentes fontes é fundamental, tendo em vista a complexidade e a diversidade dos requisitos que necessitam ser atendidos. Desta forma, as equipes distribuídas geograficamente, identificadas pelo algoritmo de agrupamento, podem colaborar entre si na avaliação das composições realizadas.

A utilização de redes sociais em ECOS é abordada por Santos et al., (2014), porém, questões relacionadas à colaboração entre os pesquisadores identificados, visando apoiar a composição de serviços em uma plataforma de ECOS não são tratadas.

A solução proposta no presente trabalho permite realizar o agrupamento de comunidades científicas, criando diferentes níveis de abstração. A partir dos grupos identificados, é possível descobrir aqueles com interesses comuns que possam avaliar e tomar decisões em relação à composição de serviços.

Uma tendência crescente na utilização de serviços distribuídos na web, que podem ser compostos, é analisada por Autili et al. (2016). Neste trabalho, é destacada uma tendência da internet a possuir uma rede de serviços com baixo acoplamento e

processos de negócios dinâmicos e flexíveis. Nesta rede, a composição de serviços terá um papel importante. Para realizar a composição de serviços, a literatura apresenta trabalhos que utilizam uma análise de entradas e saídas de serviços de maneira sintática e semântica (Mier et al. 2015) (Chen et al. 2017). Porém, a análise de interoperabilidade entre serviços disponíveis não é realizada. Considerando-se a relação entre interoperabilidade e a capacidade de composição de serviços, a análise da influência da interoperabilidade ao selecionar serviços para o processo de composição é fundamental.

Baseando-se na análise da literatura realizada, propomos apresentar uma solução que considere aspectos relacionados às dimensões técnica e social de um ECOS e que ofereça um suporte à composição de serviços. Os aspectos que influenciam a dimensão técnica na plataforma são a interoperabilidade entre serviços e a possibilidade de reutilização de serviços web. O suporte à interoperabilidade oferecido pela plataforma permite obter serviços que interoperam não apenas a nível sintático e semântico, mas também a nível pragmático (Neiva et al., 2015). As redes sociais científicas, por sua vez, influenciam a dimensão social. Esta influência se relaciona à capacidade de gerar conhecimento quando são identificados diferentes grupos de pesquisa, geograficamente distribuídos, e que possuem interesses em comum. O conhecimento é gerado a partir das discussões entre esses grupos, e das avaliações dos serviços compostos que foram desenvolvidos.

3. Plataforma E-SECO

A plataforma E-SECO foi desenvolvida seguindo a abordagem de um ECOSC (Freitas et al., 2015). Nela, é possível conduzir e gerenciar experimentos científicos, que são frequentemente atividades colaborativas. Como ilustrado na Figura 1, o “Ambiente de Desenvolvimento do E-SECO” é um componente web no qual o código da plataforma é disponibilizado como *open source*¹. O ecossistema é composto de artefatos fornecidos por diferentes nós situados em diferentes instituições. A plataforma utiliza uma rede ponto-a-ponto (P2P) (Freitas et al., 2015) através da qual diferentes nós se comunicam. Esta rede fornece uma Camada Cliente que permite a busca por pontos, busca por artefatos e *download* de artefatos. Os artefatos incluem serviços, workflows, documentos de revisões sistemáticas da literatura, conjuntos de dados, modelos, entre outros. As APIs permitem auxiliar o desenvolvimento de workflows científicos em diferentes passos, em um ambiente de desenvolvimento *open source*. Através das APIs, os usuários consomem recursos para criar produtos e artefatos que adicionam valor à plataforma. A partir destas características, a comunidade de desenvolvedores pode estender, adaptar e utilizar as funcionalidades da plataforma com outras aplicações em diferentes domínios.

Além disso, a plataforma E-SECO oferece um ambiente colaborativo para apoiar o desenvolvimento e a execução de workflows científicos. Apoia a análise de proveniência de dados durante a execução de workflows através do “Módulo de Dados de Proveniência”, ilustrado na Figura 1 (Sirqueira et al., 2016). Cheney et al. (2012) classificam proveniência de dados como a informação sobre a origem, contexto, derivação, propriedade ou histórico de algum artefato. Convém ressaltar que, embora o serviço SCView utilize o módulo de proveniência, não é objetivo deste trabalho abordar sobre este módulo. A plataforma permite também conectar serviços de diferentes fontes.

¹ <https://github.com/pgcc/plscience-ecos>

Esses serviços devem ser capazes de interoperar em alto nível de abstração, a fim de alcançar uma colaboração significativa para apoiar o desenvolvimento de workflows científicos.

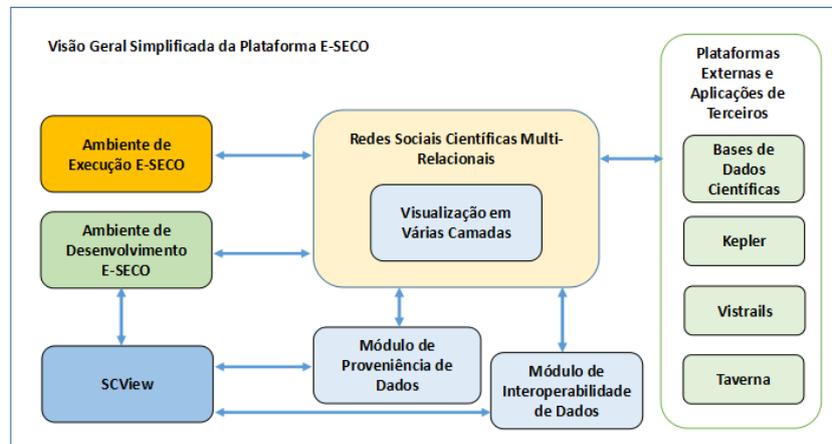


Figura 1. Visão Geral Simplificada da plataforma E-SECO

3.1. Composição de Serviços na Plataforma E-SECO

Para participar de um experimento, cientistas necessitam do suporte de serviços oferecidos pela plataforma E-SECO por fontes externas (MyExperiment², BioCatalogue³, entre outras). A plataforma E-SECO oferece um serviço para apoiar a identificação de redes de colaboração e coautoria. Essa identificação permite descobrir diferentes redes de pesquisadores que possam validar as composições realizadas na plataforma, que devem satisfazer os requisitos das partes interessadas no experimento.

Para registrar os serviços, é realizada a anotação semântica de serviços obtidos de diferentes repositórios. Inicialmente, é realizada uma pesquisa baseada em palavras-chave ou termos relacionados ao contexto de uso de um serviço. Como resultado, os serviços são listados de forma padronizada. Assim, é possível obter informações sobre o repositório de origem e os metadados relacionados à descrição do serviço. Neste contexto, serviços atômicos e compostos podem ser registrados. Serviços atômicos realizam uma ação específica, enquanto os serviços compostos são formados por múltiplos serviços atômicos. Informações adicionais relacionadas ao contexto de uso do serviço também podem ser registradas.

Visando auxiliar a atividade de composição, a plataforma oferece o serviço SCView, conforme ilustrado na Figura 1. Visualizar as relações da composição permite que *stakeholders* (desenvolvedores e cientistas) possam analisar as dependências funcionais e aquelas relacionadas à interoperabilidade entre os serviços. Para tanto, a abordagem PRIME (Neiva et al., 2015) é utilizada pelo serviço SCView, para realizar o ranqueamento de serviços interoperáveis, de acordo com os parâmetros informados a partir de uma requisição. A partir da seleção de um serviço entre os ranqueados, desenvolvedores podem analisar visualmente quais serviços estão associados ao que foi selecionado.

² <http://www.myexperiment.org/home>

³ <https://www.biocatalogue.org/>

Para realizar a representação visual, o SCView utiliza um grafo $G = (V, E)$. Onde V é um conjunto de vértices que representa todos os serviços associados à seleção. Cada serviço $s \in V$ é apresentado como um nó. E por sua vez, representa um conjunto com todas as arestas presentes no grafo, onde $E = D \cup I$. D é um conjunto de arestas direcionadas que indicam dependências entre serviços. Cada aresta $d \in D$ indica que um serviço s_1 depende de um serviço s_2 . I , por sua vez, representa um conjunto de arestas não direcionadas que indicam interoperabilidade entre serviços. Cada aresta $i \in I$ indica que dois serviços são capazes de interoperar. Este grafo possui arestas direcionadas, indicando dependências funcionais, e não direcionadas, indicando interoperabilidade.

Conforme mostrado na Figura 2, os serviços compostos são representados por nós em cor vermelha. Os serviços recuperados do repositório “BioCatalogue” são representados por nós em cor preta, enquanto os serviços recuperados de instâncias da plataforma E-SECO são representados por nós em cor verde. Os serviços que interoperam são ilustrados por linhas tracejadas.

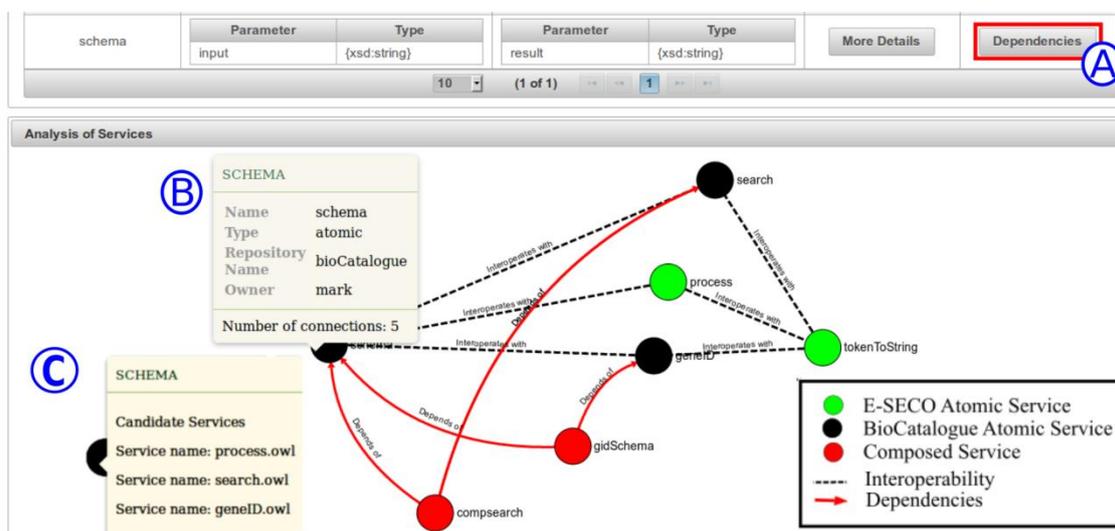


Figura 2. Grafo gerado pelo SCView com relações de dependência e interoperabilidade

No SCView, desenvolvedores podem selecionar o botão “Dependencies” (Figura 2-A). Como resultado, é apresentado um grafo (Figura 2-B), e o conjunto de dados do desenvolvedor de serviços é exibido, bem como o repositório de origem e o tipo de serviço. A partir dessas informações, a importância do serviço e os impactos de modificações, ou remoções, podem ser verificados. Assim, os pesquisadores interessados neste serviço, identificados pela ferramenta de visualização de redes sociais, podem participar da avaliação dessa composição. Além disso, o grupo de desenvolvedores responsáveis pela construção tem a informação dos requisitos e os cientistas interessados na construção do serviço.

A partir do grafo gerado os *stakeholders* identificados podem avaliar as dependências entre os serviços e verificar quais serviços podem se comunicar. Se o serviço não atender à requisição, o desenvolvedor pode selecionar outro serviço mais adequado. Para isso, ele necessita selecionar um nó (serviço) com o cursor e, clicar com o botão direito para visualizar candidatos para substituição (Figura 2-C). Quando esta ação é realizada, são mostrados serviços semanticamente compatíveis com as entradas e

saídas do serviço atual. O desenvolvedor pode então selecionar um serviço e gerar uma nova composição. O serviço composto é então anotado semanticamente e registrado. Neste ponto, a solução considera não apenas as métricas de dependência funcional, mas a capacidade de interoperar com outros serviços (locais e geograficamente distribuídos). Além disso, é possível realizar uma nova busca adicionando parâmetros relacionados à interface de serviço e novas informações contextuais, para obter serviços pragmaticamente interoperáveis. Finalmente, o serviço composto pode ser validado pelos cientistas identificados e, geograficamente distribuídos, representados na visualização da rede social.

O módulo de visualização da plataforma E-SECO, denominado “Multi-Layer Visualization”, suporta a extração e análise das relações estabelecidas nas redes sociais científicas. Ele implementa todos os aspectos relacionados à Rede Social Científica Multi-Relacional. Este módulo está relacionado ao ambiente de desenvolvimento de software científico, bem como “E-SECO Execution Environment”. Além disso, o módulo de visualização recebe dados da execução de workflows científicos, bem como das plataformas externas, como Kepler, Tavern, VisTrails e outras bases de dados científicas (DBLP, plataforma Lattes, ResearchGate, entre outros). O serviço SCView interage com o módulo de visualização utilizando o módulo de interoperabilidade. Isto permite analisar visualmente serviços que são interoperáveis.

Visando apresentar como as redes complexas são utilizadas na plataforma, a próxima seção apresenta as estratégias utilizadas para gerar as redes complexas, permitindo identificar grupos de pesquisa que colaboram entre si. Assim, é possível observar como as bases de dados científicas apoiam a plataforma, permitindo agrupar pesquisadores de acordo com relações de colaboração.

4. Estratégias de Análise de Redes Complexas

A utilização de um modelo de rede social científica multi-relacional é importante por auxiliar a colaboração entre pesquisadores na plataforma E-SECO. Este modelo permite a identificação de grupos de pesquisa que podem colaborar em experimentos científicos realizados na plataforma. A colaboração entre *stakeholders* associa-se ao conhecimento necessário para apoiar a dimensão social em um ECOS.

Para apoiar a dimensão técnica de um ECOS, utiliza-se o serviço SCView. Este serviço é utilizado para visualizar composições de serviços, obtidos de diferentes instituições de pesquisas. A visualização de grupos de pesquisadores, quando associada ao SCView, permite que redes sociais contribuam na identificação de instituições e pesquisadores, agrupando-os de acordo com relações de colaboração. Desta forma, os grupos identificados podem avaliar as composições realizadas de maneira colaborativa, descrevendo o que pode ser alterado ou opinando sobre detalhes do serviço que podem ser úteis para sua execução. Em redes sociais científicas, dois participantes são considerados conectados se ambos são coautores em um artigo (Newman, 2004), colaboraram em um experimento ou compartilham os mesmos interesses científicos. Estas redes são mais complexas, pois suas relações envolvem diferentes tipos de colaboração ou interação científica. Assim, podemos considerar que as Redes Sociais Científicas são uma espécie de Rede Social Multi-Relacional. Em redes multi-relacionais, a troca de conhecimento ocorre através de diferentes tipos de relacionamentos. A análise destas redes pressupõe que os elementos estão trocando

diferentes tipos de conhecimento, dependendo dos tipos de relacionamentos a eles associados. Para analisar a estrutura da rede social científica, foram adotadas duas estratégias: análise de centralidade de elementos e agrupamento.

4.1. Importância Local e Importância Global

Várias métricas de centralidades caracterizam a estrutura de uma rede complexa, tais como: *Closeness*, *Betweenness*, *Node Degree*, entre outras (Wasserman e Faust, 1994). Estas métricas podem definir os nós mais importantes da rede. A importância do nó é dada com base na semântica da rede complexa. Ströele et al. (2017) caracterizam o grau de importância dos pesquisadores por seu potencial de colaboração. Dessa forma, duas métricas são utilizadas, uma para analisar a centralidade global dos pesquisadores e outra para avaliar a centralidade local. A centralidade global baseia-se no conceito de distância, para identificar os elementos mais centrais de uma rede complexa. Elementos globalmente centrais permitem espalhar informações para a rede mais rapidamente do que outros. Por isso é importante identificá-los. Neste trabalho, a métrica *Closeness* será adotada como métrica global, analisando o potencial de colaboração dos pesquisadores, considerando a rede como um todo. A centralidade local, por sua vez, analisa as conexões dos vizinhos de um nó. Essa métrica tem como objetivo caracterizar se o nó considerado está significativamente conectado a seus vizinhos (se possui influência local). Por fim, a métrica *Node Degree* é utilizada para analisar a colaboração local dos pesquisadores, identificando seu potencial de colaboração com seus vizinhos.

A figura 3-A apresenta a métrica *Closeness*, onde V é o conjunto de pesquisadores, “ n ” é o número de pesquisadores na rede social científica, “ $d(x, y)$ ” é uma função que calcula a colaboração entre os pesquisadores “ x ” e “ y ” e “ X ” é o pesquisador para quem a proximidade (*closeness*) está sendo calculada. A figura 3-B, por sua vez, apresenta a métrica de *Node Degree*, onde “ $g(x)$ ” é o número de links que o pesquisador “ x ” tem, “ n ” é o número de pesquisadores na rede social científica e “ x ” é o pesquisador para quem a Centralidade Local foi calculada.

The figure shows two mathematical formulas side-by-side, each enclosed in a box with a circled letter above it. Formula A is $C_x = \frac{\sum_{y \in V - \{x\}} d(x, y)}{n - 1}$. Formula B is $LC_x = \frac{2 \times g(x)}{n(n - 1)}$.

Figura 3. Métricas de *Closeness* (A) e *Node Degree* (B), utilizadas para caracterizar o potencial de colaboração entre pesquisadores

Elementos com alta centralidade local representam pesquisadores com alta colaboração entre seus pares. Geralmente, esses pesquisadores representam um grupo de pesquisa em suas instituições. Por outro lado, os pesquisadores com alta centralidade global têm alta colaboração com a maioria dos pesquisadores de redes sociais, ou seja, são capazes de propagar suas ideias mais facilmente para vários pesquisadores da rede científica. Além da análise estrutural da rede científica, foram também utilizadas estratégias de agrupamento para identificar as comunidades de pesquisa científica. Tais estratégias foram utilizadas na visualização da rede científica. Esta visualização tem como objetivo identificar diferentes grupos de pesquisadores no grafo, a partir das relações entre eles. A seção 4.2 apresenta o algoritmo de agrupamento utilizado.

4.1.1. Importância Global para a Composição de Serviços

Buscando caracterizar os serviços mais importantes em uma rede que inclui relações de dependências e interoperabilidade entre serviços, utilizou-se a métrica de *Closeness*. Analisar a importância destes serviços permite avaliar o impacto causado na rede, caso ocorra alguma falha em um dos serviços. Esta métrica permite capturar o quão central é o vértice (serviço) em relação à rede de composição como um todo. Assim, são consideradas todas as arestas, tanto as que indicam interoperabilidade, quanto aquelas que indicam dependências associadas à composição de serviços, desconsiderando o sentido das arestas que indicam dependência.

A métrica de *Closeness*, ilustrada na Figura 3-A, será utilizada para apoiar a composição de serviços. Para tanto, V representa o conjunto de serviços, “ n ” representa o número de serviços $d(x,y)$ de uma função que calcula a distância entre serviços “ x ” e “ y ” e “ x ” é o serviço para o qual a métrica de *Closeness* está sendo calculada. Assim, é possível analisar a importância de um serviço a partir de sua centralidade na rede. Esta métrica é aplicada aos serviços atômicos que são utilizados em composições. No grafo gerado, os nós (serviços), cujo valor de centralidade calculado seja inferior a 2, recebem arestas com maior espessura, indicando que são serviços mais centrais na rede, possuindo, portanto, maior importância para a rede. Para caracterizar as dependências utilizou-se as métricas de *fan-in* e *fan-out*. Marin et al., (2004) definem *fan-in* de um método m como o número de métodos distintos que podem invocar m e *fan-out* de um método m como o número de métodos distintos que podem ser invocados por m . No caso deste trabalho, onde consideramos *web services*, *fan-in* se relaciona ao número de serviços que chamam o serviço analisado, enquanto *fan-out* representa o número de serviços atômicos chamados por um serviço composto.

4.2. Algoritmo de Agrupamento: Fluxo Máximo

O algoritmo de agrupamento tem como objetivo identificar grupos de pessoas no grafo social que possuem um relacionamento forte entre eles. Para realizar esta identificação, o algoritmo segue a estratégia de analisar fluxos de conhecimento em rede social. Assim, pessoas com um grande fluxo de informações entre elas possuem tendência a pertencer a um mesmo grupo.

O Grafo Social com fluxo será representado por $G = (X, U, f)$, onde temos X como o conjunto de pesquisadores, U como o conjunto de relações entre pesquisadores e $f(m + 1$ vetor dimensional) como o conjunto de fluxos máximos entre cada par de pesquisadores. Assim, para todo $x_i, x_j \in X$, o fluxo máximo entre estes dois pesquisadores será f_w , onde $0 \leq w \leq m$. O cálculo do fluxo máximo entre os elementos foi feito usando o algoritmo de Edmonds-Karp (Edmonds e Karp, 1972). Mais detalhes sobre o algoritmo utilizado podem ser encontrados em (Ströele et al. 2017).

5. Ferramenta de Visualização de Redes Sociais na Plataforma E-SECO

O principal objetivo da ferramenta de visualização é apoiar a interação e análise de redes aos usuários da plataforma E-SECO sob diversas perspectivas. Assim, é possível buscar parceiros científicos, propor novos grupos semânticos na plataforma E-SECO, e promover a disseminação do conhecimento na plataforma.

são importantes para o desenvolvimento do trabalho em seus grupos de pesquisa considerando seu potencial de colaboração local. A rede social permitiu observar também que os pesquisadores com alta centralidade global, são candidatos para a difusão da informação, devido ao grande potencial de colaboração com a maioria dos pesquisadores.

A análise de agrupamento, por sua vez, permite avaliar como ocorre a colaboração entre pesquisadores. Além disso, podemos estudar a troca de conhecimento entre grupos e instituições educacionais. Em nossos estudos consideramos apenas a relação de coautoria (Evans et al., 2011). Comparando os resultados obtidos em estudos anteriores percebemos que, com a adição de novas relações na rede social, a estrutura do grupo mudou. Em alguns casos, pesquisadores que estavam em diferentes grupos se mudaram para o mesmo grupo. Isso ocorreu porque esses pesquisadores têm múltiplas relações uns com os outros. Consequentemente, as relações destes pesquisadores são mais fortes do que aquelas entre pesquisadores que têm apenas relações de coautoria. Ainda, considerando a estrutura do agrupamento, também analisamos as centralidades locais e globais dos grupos de pesquisa considerando cada grupo como um único elemento que consolida todas as relações de seus membros na rede social científica.

Uma análise mais detalhada sobre rede social científica é realizada em (Ströele et al., 2017). A próxima seção apresenta um cenário de utilização que caracteriza como as redes sociais científicas podem apoiar o processo de composição de serviços na plataforma E-SECO.

6. SCView em ação

Para analisar as composições de serviço que podem ser realizadas, um cenário de uso para SCView é apresentado no contexto da plataforma E-SECO. Através deste cenário estamos interessados em analisar como a utilização de redes sociais permite identificar potenciais grupos que podem colaborar entre si. Esta atividade está relacionada ao suporte à composição de serviços no contexto de um ecossistema de software científico. O objetivo deste cenário é caracterizar a reutilização de serviços obtidos de diferentes repositórios. A reutilização está relacionada à capacidade de composição, na perspectiva dos desenvolvedores, no contexto de um ECOSC. Para isso, esses desenvolvedores utilizam o serviço de visualização para analisarem as dependências (funcionais e interoperabilidade) entre os serviços.

Cenário: inicialmente, a partir dos requisitos especificados, o desenvolvedor acessou a interface da plataforma E-SECO para fazer uma solicitação e localizar os serviços registrados. Consideramos a situação na qual o desenvolvedor não informa parâmetros de entrada e saída na requisição, mas o termo “Protein Database” como uma descrição de serviço e “Public” como o tipo de licença para o serviço. Ao executar a pesquisa, o sistema classifica os serviços disponíveis no repositório por relevância. O serviço de suporte à interoperabilidade permite realizar esta classificação considerando aspectos sintáticos, semânticos e pragmáticos. O serviço “schema” foi então selecionado. O desenvolvedor seleciona este serviço através do botão “Dependencies” (Figura 5-A). Assim, um grafo é gerado conforme ilustrado na Figura 5. Este grafo apresenta serviços que interoperam com o serviço selecionado e as dependências relacionadas aos serviços (Figura 5-B). Neste exemplo, os cientistas interessados são aqueles que pertencem ao grupo UFRJ. Eles irão participar da validação da composição. Caso seja necessário analisar a rede social entre pesquisadores de instituições que colaboram com o grupo

UFRJ, os desenvolvedores precisam selecionar o botão “Exibir Rede Social” (Figura 5-C). Uma rede com múltiplas relações entre pesquisadores será apresentada em uma nova janela (Figura 5-D). Conforme observado na Figura 5-D, o grupo UFRJ foi identificado pelo serviço de análise de redes sociais, possuindo relações de coautoria com outras instituições. Isto possibilita que os grupos identificados possam colaborar para avaliar a composição de serviços realizada. Nesta rede é apresentada uma visualização de segundo nível, através da qual é possível analisar diferentes grupos (*clusters*) produzidos por um algoritmo de mineração de dados. Através dessa visualização foi possível identificar outros cientistas que podem participar da validação dessa composição.

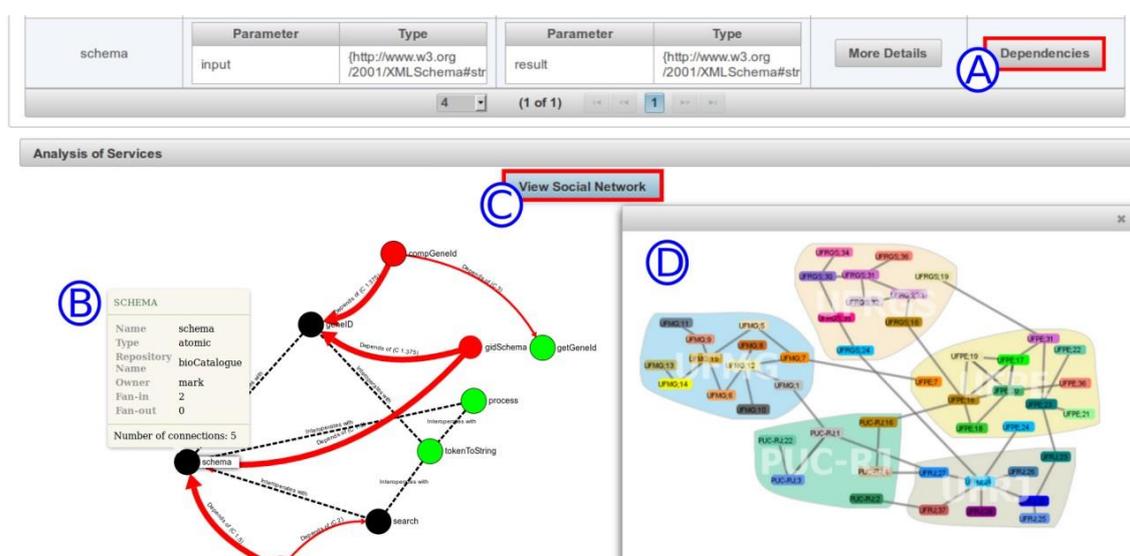


Figura 5. Suporte à Composição de Serviços na Plataforma E-SECO

A partir da rede social gerada, identificou-se o grupo de pesquisa da UFJF como potencial colaborador, devido aos interesses de pesquisa e trabalhos já realizados em parceria.

A figura 6 destaca o grafo de composição gerado, com o apoio à validação de composições. Visando analisar as dependências do serviço, o cientista João pode também posicionar o cursor sobre o nó que representa o serviço escolhido para obter informações de dependências. Para tanto, são apresentadas métricas de *fan-in* e *fan-out*. Estas informações são importantes para destacar os impactos que podem ser causados caso um serviço falhe, ou caso seja removido da rede.

Neste cenário, o cientista posicionou o cursor sobre o serviço “schema”, o qual apresentou o valor “2” para *fan-in* e “0” para *fan-out* (Figura 6-A). Isto indica que dois serviços utilizam o serviço selecionado e que este serviço não utiliza nenhum outro. Portanto, sua remoção pode causar um impacto para os serviços que o utilizam. Quando aplicamos estas métricas a todos os serviços da rede, é possível avaliar o impacto causado para a rede caso ocorram alterações na versão do serviço ou alguma falha. Além disto, podemos avaliar se determinado serviço possui alta complexidade. Um alto valor de *fan-in* indica um grande impacto para mudanças. Um alto valor de *fan-out*, por sua vez, indica uma maior complexidade do serviço composto.

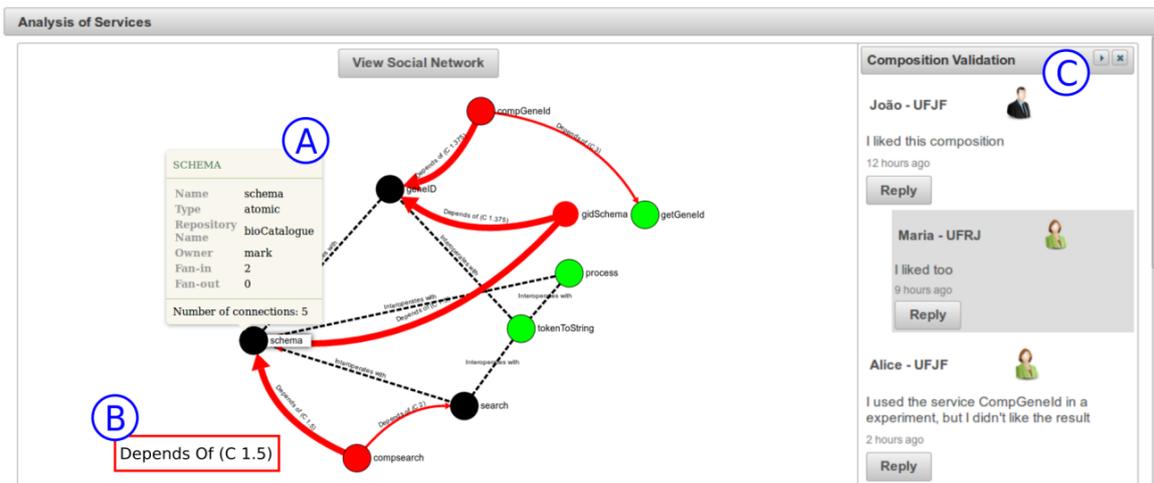


Figura 6. Análise do Grafo Gerado e Validação da Composição

O grafo apresentado possui arestas que indicam dependências com diferentes níveis de espessura, de acordo com o valor de *Closeness* calculado para o serviço que é utilizado em composições. Neste caso, este valor foi calculado para 3 diferentes serviços: “search”, “schema” e “genelD”. Para o serviço “search”, o valor de *Closeness* obtido foi igual a “2”. Por isto, a aresta de dependência que conecta o serviço “compsearch” ao serviço “search” possui menor espessura, pois este serviço não é muito central na rede, não sendo muito importante para a rede como um todo. O valor de *Closeness* obtido é mostrado na aresta que conecta serviços compostos e suas dependências. Para o serviço “schema”, o valor obtido foi igual a “1,5” (apresentado visualmente nas arestas que se conectam ao serviço e destacado na Figura 6-B), enquanto para o serviço genelD, o valor encontrado foi igual a “1,375”. Isto indica que os serviços “schema” e “genelD” são mais centrais na rede e, portanto, possuem maior importância para a rede.

A importância de analisar serviços que interoperam entre si também deve ser destacada. Para tanto, a plataforma utiliza o serviço de suporte à interoperabilidade, descrito anteriormente. É a partir desta análise que um cientista pode descobrir visualmente quais potenciais serviços podem substituir um serviço específico. Assim, em um caso de falha de um serviço identificado pela métrica de *Closeness* como mais central na rede, por exemplo, o cientista pode identificar visualmente outros que podem substituí-lo.

Com base nessas informações, cientistas podem utilizar o serviço de validação de composições (Figura 6-C) para se comunicar com potenciais colaboradores geograficamente distribuídos e, assim, avaliar a composição realizada. Desta forma, colaboram com sua experiência para destacar, por exemplo, a utilidade da composição ou se o resultado da execução do serviço desejado foi satisfatório em relação às expectativas. É importante ressaltar que, devido às limitações de espaço, não é parte do escopo deste artigo apresentar e discutir como essa validação ocorreu através de um serviço de suporte à discussões, ilustrado na Figura 6-C.

7. Conclusão

Ao apoiarmos a interação entre especialistas de diferentes áreas estamos também oferecendo suporte aos aspectos técnicos de uma atividade de composição. O

aperfeiçoamento pode ser alcançado através da interação de especialistas em uma área específica. Como resultado dessa interação, buscamos desenvolver serviços que possam apoiar atividades em experimentos científicos.

Neste trabalho, utilizamos um método de detecção de grupos para identificar comunidades de pesquisa na rede social científica brasileira. Esta identificação permite apoiar a composição e reutilização de serviços em uma plataforma de ECOSC. A estratégia de dados de abstração permite analisar a rede social em diferentes perspectivas: (i) nível de cooperação entre instituições; (ii) relacionamentos fortes e fracos; (iii) pesquisadores que desempenham um papel centralizador na rede social, e assim por diante. Como resultado, nosso objetivo foi apoiar a composição de serviços na plataforma SECO com a adição de elementos de colaboração associados à dimensão social em um ECOS.

Não foram encontrados na literatura trabalhos que consideram a análise de relações múltiplas, nem com a representação evolutiva da rede, quando existe foco na análise de redes sociais científicas. Assim, a solução proposta representa uma contribuição na medida em que ela apoia o desenvolvimento de serviços através da identificação de *stakeholders* que possam auxiliar na composição de serviços em uma plataforma de ECOSC. Ao mesmo tempo, esta solução considera aspectos técnicos (reutilização, interoperabilidade, dependência funcional, entre outros) que podem influenciar na decisão sobre esta composição.

Como trabalhos futuros, cabe explorar a análise de redes sociais científicas, em profundidade, considerando um contexto real de desenvolvimento de software científico. Também pretendemos investigar a extensão na qual o apoio de outros aspectos sociais poderia melhorar o processo de desenvolvimento de software na plataforma do ecossistema, tais como, reputação e confiança.

Agradecimentos

Aos participantes do Núcleo de Pesquisa em Engenharia do Conhecimento (NEnC). Este projeto é parcialmente financiado pela CAPES, CNPq e FAPEMIG. Phillipe Marques é apoiado pela CAPES.

Referências

- Arakaki, M.; Martins, G.; David, J. M.; Braga, R.; Campos, F. and Neiva, F. “Um Processo para o Desenvolvimento de Serviços de Colaboração em um Ecossistema de Software Científico”. 13º Simpósio Brasileiro de Sistemas Colaborativos. 1522–1535. 2016.
- Autili, M., Tivoli, M. and Goldman, A. “Thematic series on service composition for the future internet”. *Journal of Internet Services and Applications* 7, no. 1. 1-4. 2016.
- Campbell, P. R. and Ahmed, F. “A three-dimensional view of software ecosystems”. In *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*. ACM. 81-84. 2010.
- Chen F.; Lu C.; Wu H.; and Li M. “A semantic similarity measure integrating multiple conceptual relationships for web service discovery”. *Expert Systems with Applications*. 19-31. 2017.

- Cheney, J., Finkelstein, A., Ludäscher, B., and Vansummeren, S. “Principles of provenance (dagstuhl seminar 12091)”. In: *Dagstuhl Reports* (Vol. 2, No. 2). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2012.
- Edmonds, J.; Karp, R. M. “Theoretical improvements in algorithmic efficiency for network flow problems”. *Journal of the ACM (JACM)*, v. 19, n. 2. 248-264. 1972.
- Evans, T. S.; Lambiotte, R., Panzarasa, P. “Community structure and patterns of scientific collaboration in business and management”. *Scientometrics*.381-396. 2011.
- Freitas, V.; David, J. M. N.; Braga, R. and Campos, F. “Uma Arquitetura para Ecosystema de Software Científico”. *WDES*. 41- 48. 2015.
- Manikas, K. “Revisiting software ecosystems research: A longitudinal literature study”. *Journal of Systems and Software* 117. 84-103. 2016.
- Marin, Marius, Van Deursen, A. and Moonen, L. "Identifying aspects using fan-in analysis." *Reverse Engineering. Proceedings. 11th Working Conference on. IEEE*. 132-141. 2004.
- Mier, P. R.; Pedrinaci, C.; Lama, M. and Mucientes M. “An integrated semantic Web service discovery and composition framework”. *IEEE Transactions on Services Computing*. 2015.
- Neiva, F. W.; David, J. M. N.; Braga, R.; Campos F. and Freitas, V. “PRIME: Pragmatic interoperability architecture to support collaborative development of scientific workflows”. In IX SBCARS. 50-59. 2015.
- Newman, M. E. J. “Co-authorship networks and patterns of scientific collaboration”. In *Proceedings of the National Academy of Sciences*, 101: 5200-5205. 2004.
- Pokraev, S. V. “Model-driven semantic integration of service-oriented applications”. *Tese de Doutorado. Novay*. 2009.
- Sadi, M. H., and Yu, E. “Analyzing the evolution of software development: from creative chaos to software ecosystems”. *Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on*. 2014.
- Ströele, V.; Campos, F.; David, J. M. N., Braga, R; Abdalla, A.; Lancellotta, P. I.; Zimbrão, G.; Souza, J. “Data Abstraction and Centrality Measures to Scientific Social Network Analysis”. *21st International Conference on Computer Supported Cooperative Work in Design*. 2017.
- Santos, R. P. and Werner, C. M. L. “Treating social dimension in software ecosystems through reuseecos approach”. In *Digital Ecosystems Technologies (DEST), 6th IEEE International Conference on*. 1-6. 2012.
- Santos, R. P.; Esteves, M. G. P.; Freitas, G. D. S. and de Souza, J. M. “Using Social Networks to Support Software Ecosystems Comprehension and Evolution”. *Social Networking*, 2014.
- Sirqueira, T. F. M.; Dalpra, H. L. O.; Braga, R.; Araújo, M. A. P.; David, J. M. N.; Campos, F. “E-SECO ProVersion: An Approach for Scientific Workflows Maintenance and Evolution”. *Procedia Computer Science*. 547-556. 2016.
- Wasserman, S. and K. Faust. “*Social Network Analysis: Methods and Applications*”. Cambridge, UK: Cambridge University Press. 1994.