

Análise de redes e FTA para uma avaliação estratégica dos Institutos Nacionais de Ciência e Tecnologia

Jackson Max Furtunato Maia¹, Ana Victória Gruginski de Carvalho Ladeira²,
Cristiano Hugo Cagnin³ e Adriana Badaró de Carvalho Villela⁴

Resumo

Este artigo descreve resultados preliminares da prova de conceito de uma aplicação metodológica em *Future-Oriented Technology Analysis* (FTA), fundamentada em análise de redes sociais, para auxiliar na avaliação do Programa Institutos Nacionais de Ciência e Tecnologia (INCT). Usando bibliometria de coautorias e similaridade contextual como base para determinar arestas ponderadas entre diferentes pesquisadores e para representar a rede correspondente, aplicamos análises de centralidade e de agrupamento para identificar grupos de pesquisa e seus atores mais relevantes,

Abstract

This article describes the preliminary results of a proof of concept of a new methodological application on Future-Oriented Technology Analysis (FTA) based on social network analysis. This approach was used to support an evaluation of the National Institutes for S&T (NIST) program. Using coauthorship and contextual similarity bibliometry as basis for determining weighted edges between different researchers and representing the corresponding network, we applied centrality and community detection analysis to identify research groups and its most relevant actors, besides

- 1 Analista em Ciência e Tecnologia Sênior do Inpe e assessor do CGEE. Foi coordenador técnico da Coordenação de Apoio à Pesquisa, Desenvolvimento e Aplicações (COAPD) do CNPq e da Coordenação de Pesquisa, Desenvolvimento e Inovação da AEB. É doutor em Ciências (Física) pela Universidade de São Paulo.
- 2 Aluna de graduação de Engenharia de Computação na Universidade de Brasília (UnB) e estagiária no Centro de Gestão e Estudos Estratégicos.
- 3 Assessor do CGEE, com PhD pela Universidade de Manchester, Reino Unido, bem como mestrado e graduação em Engenharia de Produção pela UFSC. Atua nas áreas de inovação, estudos de futuro (foresight) e sustentabilidade, com vasta experiência em projetos nacionais e internacionais no âmbito da Comissão Europeia e de parcerias com instituições como UNIDO, UNEP e OCDE.
- 4 Graduada em História pela Universidade Federal de Minas Gerais (UFMG) e Mestre em Desenvolvimento Sustentável pela UnB. Atualmente, trabalha no CGEE em projetos de acompanhamento e avaliação (A&A) de programas em CT&I, estudos de futuro, desenvolvimento de metodologias e disseminação da informação.

além de caracterizar sua estrutura hierárquica. Também avaliamos a aplicação de diferentes métricas de redes para estudar a evolução temporal dos INCT como redes de pesquisa. Este é um passo intermediário para a identificação de especialidades que possam ajudar a suprir futuras demandas nacionais por conhecimento ou para enfrentar grandes desafios nacionais.

Palavras-chave: Análise de redes sociais. Avaliação de CT&I. Problemas Sociais Complexos.

characterizing its hierarchical structure. We also evaluated the application of different network metrics to study the temporal development of the NIST's as research networks. This is an intermediary step to identify competencies that may help supplement future national demands for knowledge or to face major national challenges. We conclude with a discussion of the implications of using social network analysis in FTA practice and propose a FTA methodological approach to help identify collaboration opportunities between different NIST research networks to face complex problems.

Keywords: Social network analysis. S&T evaluation. Complex social problems.

1. Introdução

A reorientação dos sistemas de ciência, tecnologia e inovação (CT&I) para enfrentar 'grandes' desafios que afetam as sociedades contemporâneas tem se tornado uma preocupação crescente dos atores desse sistema (CAGNIN *et al.*, 2012). Esse esforço é parte de um novo enfoque de políticas de inovação, que é mais global em escala e orientado a objetivos sociais, ao invés da prática anterior (anos 1940 e 1950), que priorizava nações e era orientada ao suporte do complexo militar-industrial (GASSLER *et al.*, 2008).

Os temas cobertos pelo termo 'grandes' desafios são formados por '*wicked problems*' (RITTEL; WEBER 1973), que são difíceis ou impossíveis de serem resolvidos por agências individualmente ou apenas por enfoques de planejamento racionais. Até mesmo o significado de tais desafios tende a ser contestado por diferentes atores. Isso é, há bastante tempo, compreendido por acadêmicos e ativistas, e a articulação de tais desafios não é uma novidade. A novidade se insere, portanto, na atenção crescente dada a esses temas para a formulação de novas missões para políticas de CT&I. As razões para isso são complexas: em parte, refletem a percepção de urgência crescente para se enfrentar problemas que, se negligenciados, podem ter consequências negativas em escala global nas próximas décadas, mas também refletem esforços de direcionamento de CT&I para alcançar objetivos estratégicos explícitos.

Nesse contexto, Cagnin *et al.* (2012) afirmam que qualquer tentativa de se compreender e atuar sistemicamente deve abranger uma diversidade de barreiras estabelecidas, tanto organizacionais como epistemológicas e setoriais, requerendo:

- interdisciplinaridade que transcenda os limites de comunidades epistemológicas tradicionais, conjugando diferentes bases de conhecimento e, em particular, fomentando uma colaboração próxima entre as ciências ‘duras’, sociais e humanas.
- coordenação extradepartamental e coerência além de silos tradicionais que caracterizam o processo decisório e político, pois grandes desafios exigem respostas de múltiplas agências e é fundamental que as respostas sejam coerentes para que sejam efetivas.
- enfoques de governança envolvendo múltiplos níveis e que reconheçam o princípio da subsidiariedade, enquanto garantem coerência entre agendas e atividades globais, regionais, nacionais e subnacionais.
- convergência ou fusão tecnológica que abra novas possibilidades de se gerenciar, mitigar ou mesmo eliminar algumas das causas e dos sintomas prejudiciais associados com os ‘grandes’ desafios.
- colaboração entre distintas indústrias que ofereçam complementaridade de recursos.
- inserção, explícita, de horizontes de longo prazo nas agendas políticas e práticas de gestão de curto prazo.

Com base no exposto, o presente artigo busca expandir a compreensão de como os INCT podem ser utilizados como um instrumento de política de CT&I capaz de ajudar o Brasil a enfrentar grandes desafios nacionais e globais. Tal análise será feita a partir de um ponto de vista metodológico. Dessa forma, a próxima seção coloca a análise de redes em contexto para verificar sua potencial contribuição dentro de enfoques de FTA⁵. A seção seguinte explica o que são os INCT, contextualizando o programa dentro do sistema de inovação brasileiro. As seções subsequentes descrevem detalhes metodológicos de análise de redes empregados no trabalho, abordam as etapas de preparação e processamento dos dados extraídos de currículos da Plataforma Lattes, mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e apresentam alguns resultados da análise preliminar realizada. A última seção discute possíveis implicações da análise de redes no apoio à tomada de decisão no (re) direcionamento de programas como o dos INCT.

2. Contribuição da análise de redes para FTA

Análise de redes e FTA compartilham a mesma base filosófica com a teoria de sistemas e a teoria de complexidade (NUGROHO; SARITAS, 2009). Segundo os autores, estudos explícitos combinando enfoques de análise de redes e de FTA datam do final dos anos 1980. Inicialmente, a análise de redes sociais era utilizada para mapear estruturas de distintas comunidades de pesquisa que eram acessadas e, por meio de um processo de conominação e estudo das relações

5 Foresight, forecasting e technology assessment.

entre os atores mapeados, geravam uma lista de *stakeholders* para potencial participação em distintas etapas de estudos de FTA. Mais recentemente, distintas formas de usar a análise de redes em estudos de FTA incluem (NUGROHO; SARITAS, 2009):

- análise dos dados gerados por um exercício de FTA, contendo a opinião de especialistas, estatísticas e respostas a consultas, trazendo como benefício principal uma mudança, do foco de análise nos atributos do indivíduo para um foco nas relações e alianças entre indivíduos dentro de uma rede.
- geração de subsídios para a definição dos limites e do escopo de um exercício de FTA, ajudando na decisão de que temas são relevantes e como os mesmos se relacionam.
- identificação e seleção de atores-chave, por meio da compreensão de sua posição ou importância relativa nas suas redes de relacionamento ao longo do tempo.
- captura de sinais fracos de mudança que podem enriquecer a compreensão de sistemas sob análise, além de ajudar a estruturar uma agenda comum para facilitar transições de tais sistemas no futuro.
- fomentar colaborações e ações interdisciplinares entre atores relevantes para fomentar transições do sistema sob análise.

As potenciais combinações entre FTA e análise de redes aqui descritas foram anteriormente mapeadas por Scapolo e Porter (in CAGNIN *et al.*, 2008), que sugerem que a análise de redes pode apoiar três, entre seis famílias de métodos usados em FTA: monitoramento, análise de tendências e simulação e modelagem (as outras três famílias seriam métodos criativos, opinião de especialistas e cenários). Nessa análise, os autores concluem que três temas mereceriam atenção para o desenvolvimento de novos métodos para uso em atividades de FTA: i) uso de ferramentas avançadas para lidar com todo tipo de dado e informação, incluindo busca, mineração, organização, visualização e interpretação de recursos de informação disponíveis em meio eletrônico; ii) métodos capazes de extrair, organizar, comparar e combinar uma diversidade de julgamentos humanos, incluindo interesses e opiniões, dando, assim, maior robustez, qualidade e validade científica a enfoques participativos; e iii) ferramentas de colaboração em rede que permitam a contribuição ampla de qualquer um, em qualquer lugar. Nesse âmbito, enfoques de análise de redes têm o potencial de impactar positivamente nos dois primeiros temas.

Já no contexto de 'grandes' desafios, três funções principais podem ser alocadas para enfoques de FTA (CAGNIN *et al.*, 2012):

- Identificação de um desafio.
- Articulação de um desafio já identificado em nível geral.
- Auxílio ao alinhamento de atores para lidar com o desafio.

Das três funções acima descritas, provavelmente, a mais difícil de implementar seria a identificação de novos desafios, uma vez que políticos e tomadores de decisão possuem, em teoria, um interesse intrínseco de saber sobre o novo ou sobre imprevistos que podem ser disruptivos para suas atividades (GEORGIOU; CASSINGENA HARPER, 2012).

Entretanto há uma distância grande entre querer identificar novos desafios e fazê-lo de forma efetiva. Experiências de *horizon scanning*, nesse cenário, convergem na necessidade de se usar informação de maneira holística ao invés de atômica. Da mesma forma, existe uma necessidade de se construir narrativas plausíveis, temas abrangentes ou guarda-chuva e *clusters*, ao invés de simples listas (por exemplo, de temas ou prioridades), além de se desenvolver habilidades para perceber interconexões (GEORGIOU; CASSINGENA HARPER, 2012). É justamente nesse contexto que o uso de enfoques e de ferramentas de análise de redes pode ser extremamente relevante.

Partindo de uma abordagem de sistemas de inovação e tratando dos aspectos estruturais e funcionais de tais sistemas para considerar os papéis relevantes de FTA, são identificados três benefícios que podem ser gerados na reorientação do foco de CT&I para grandes desafios: informar, estruturar e desenvolver habilidades. Com base nisso, CAGNIN *et al.* (2012) apresenta os papéis que a FTA pode ter na reorientação de sistemas de CT&I, para que estes possam melhor endereçar 'grandes' desafios. Cabe ressaltar que, nesse contexto, a análise de redes pode alavancar tais papéis ao lhes dar maior embasamento, seja por meio de novas evidências, ou por meio de conexões anteriormente não percebidas. O objetivo é que o debate, a combinação de ideias e a geração de conhecimento coletivo, dentro de um 'espaço seguro', gerem maior aprendizado, orientando, assim, de forma mais efetiva, tanto prioridades como novos mercados a serem fomentados.

3. Os Institutos Nacionais de Ciência e Tecnologia

O Programa Institutos Nacionais de Ciência e Tecnologia (INCT) foi criado em 2008 como instrumento estratégico para enfrentar o desafio de promover e consolidar a excelência nas atividades de ciência e tecnologia (C&T) no País, aprofundar sua internacionalização e estimular a participação balanceada de diferentes regiões do Brasil na produção de conhecimento. Para tanto, o Programa teve como premissa que grupos de pesquisa fossem mobilizados e agregados para a atuação em rede.

O Art. 20. da Portaria 429, de 17/7/2008, que instituiu o Programa INCT, estabelece que:

Os Institutos Nacionais serão formados por uma instituição sede, caracterizada pela excelência de sua produção científica e/ou tecnológica, alta qualificação na formação de recursos humanos e com capacidade de alavancar recursos de outras fontes, e por um conjunto de laboratórios ou grupos associados de outras instituições, articulados na forma de redes científico-tecnológicas.

Conforme as orientações estratégicas⁶ contidas em seu Documento Básico, o Programa INCT foi concebido para fortalecer os instrumentos de apoio à pesquisa, ao avanço tecnológico e à inovação, com vistas a contribuir para os objetivos maiores do Plano de Ação em Ciência, Tecnologia & Inovação (PACTI), que refletia a política de CT&I à época de seu lançamento. Assim, as principais vertentes do programa são: pesquisa, formação de recursos humanos, internacionalização, transferência do conhecimento para o setor empresarial e para o setor público, difusão e disseminação do conhecimento para a sociedade.

As primeiras 122 redes formadas a partir do edital de 2008 foram classificadas pelo CNPq em oito grupos temáticos (GT), conforme a Tabela 1. A área estratégica Ciências do Mar foi contemplada na segunda chamada pública (Edital 71/2010), a partir da qual foram formadas três redes, classificadas no GT Ecologia e Meio Ambiente, totalizando 125 INCT, compostos por cerca de 7.000 pesquisadores.

Tabela 1. Grupos Temáticos

Grupos Temáticos	
Agrárias e Agronegócios	12
Ciências Sociais e Humanas Aplicadas	10
Ecologia e Meio Ambiente	21
Engenharia e Tecnologia da Informação	12
Energia	10
Exatas	12
Nanotecnologia	10
Saúde	38

Fonte: Programa do 2º Seminário de Acompanhamento e Avaliação dos INCT. CNPq, 2013.

6 Programa Institutos Nacionais de C&T. Documento de Orientação aprovado pelo Comitê de Coordenação, em 29 de julho de 2008. Disponível em: <http://estatico.cnpq.br/programas/inct/apresentacao/pdf/015_anexo.pdf>. Acesso em: 31 mar. 2015.

Para a primeira fase, o Programa envolveu recursos financeiros em torno de R\$ 900 milhões, cuja origem ressalta a importante mobilização e parceria dos principais agentes de fomento à CT&I no País. Além das parcerias com os Estados, por meio das Fundações de amparo à pesquisa (FAP), houve aporte de recursos de outros parceiros como Petrobrás, Banco Nacional de Desenvolvimento Econômico e Social (BNDES), Ministério da Saúde, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Ministério da Educação (MEC), Ministério da Integração Nacional, Secretaria de Políticas de Promoção da Igualdade Racial (Seppir) e Ministério da Cultura.

Em junho de 2014, houve o lançamento da nova chamada pública, iniciando a segunda fase do Programa INCT. Foram mantidos: as vertentes que orientaram a primeira fase; os objetivos de mobilizar e articular, para a atuação em redes, os “melhores grupos de pesquisa em áreas de fronteira da ciência e em áreas estratégicas”; a “redução de desequilíbrios regionais do desenvolvimento científico e tecnológico”, por meio da descentralização geográfica; e a “efetiva integração entre as instituições ou grupos participantes”⁷. Esses novos objetivos do programa permitem, de forma mais explícita que anteriormente, que se fomente uma maior articulação entre INCT existentes e novos, que por ventura venham a ser selecionados, para atuarem no enfrentamento de grandes desafios nacionais.

O aporte financeiro inicial previsto foi de aproximadamente R\$ 640 milhões, dos quais R\$ 300 milhões virão do governo federal e R\$ 340 milhões das FAP. A chamada 16/2014 previu o apoio a projetos em temas induzidos e de livre escolha (demanda espontânea), tanto para a formação de novas redes quanto de INCT preexistentes.

Embora muitos de seus atributos e alguns dos seus objetivos gerais sejam semelhantes a programas anteriores, como o Programa de Apoio a Núcleos de Excelência (Pronex) e o Instituto do Milênio, o INCT é considerado estratégico no Sistema Nacional de Ciência e Tecnologia, por agregar em um só instrumento características como: relação direta com o ambiente político mais amplo, devido à associação com políticas como a Estratégia Nacional de Ciência, Tecnologia e Inovação (ENCTI) e o Plano Brasil Maior (PBM); pesquisa científica básica e fundamental, competitiva internacionalmente; foco em áreas de fronteira e estratégicas; articulação do sistema de C&T com o sistema empresarial; maior articulação interinstitucional para implementação do Programa; diversificação de fontes de recursos; flexibilidade no financiamento à pesquisa; e difusão da ciência para a sociedade.

⁷ Conselho Nacional de Desenvolvimento Científico e Tecnológico. Chamada INCT – MCTI/CNPq/CAPES/FAPs nº 16/2014. Brasília, DF: CNPq, 2014, 38 pp. Disponível em: <http://inct.cnpq.br/documents/10180/124986/Chamada+INCT_16-2014.pdf/3d511440-8d6f-413c-ac64-176b7aco2902>. Acesso em 18 de maio de 2015.

Passados seis anos da sua criação e com o lançamento da segunda fase, apresenta-se um momento estratégico para refletir sobre o Programa. Neste sentido, o estímulo para a formação das redes que caracterizam os INCT pode ir além da mera referência a áreas estratégicas, dando origem a institutos com equipes numerosas que atuam sinergicamente em diversos assuntos ou desafios. Assim, futuras propostas a novas chamadas poderiam ser submetidas não somente com o objetivo de formar novas redes, mas de maneira a articular a competência distribuída nas redes INCT existentes, com vistas a resolver de modo inter-relacionado questões complexas. No estabelecimento de metas objetivas, a existência de foco temático serviria como elemento básico na composição das redes, justificando tanto as parcerias entre grupos consolidados quanto aquelas que incorporem grupos emergentes.

Essa ênfase nas relações entre os atores, em detrimento do foco mais tradicional no pesquisador ou no grupo de pesquisa, junto com seus números expressivos, torna o Programa INCT um bom caso de estudo de viabilidade para o uso de ferramentas de análise de redes, como proposto aqui. Além disso, é esperado que os INCT sejam capazes de abordar desafios complexos presentes e futuros da sociedade brasileira, em diferentes áreas, como saúde pública, envelhecimento da população, violência urbana e mudança climática. Dada a natureza interdisciplinar dos problemas a serem abordados e a quantidade massiva de dados disponíveis, é muito importante projetar métodos sistemáticos e automatizados ou semiautomatizados para identificar competências complementares em diferentes redes de pesquisa, de modo que sejam feitas abordagens multitarefas capazes de encontrar soluções inovadoras para problemas complexos.

4. Análise de redes

A análise de redes estuda propriedades de indivíduos considerando tanto contextos locais quanto contextos globais, buscando ressaltar como o indivíduo se relaciona com o todo e como o todo afeta o indivíduo (M. NEWMAN, 2010). Esse conjunto de técnicas difere da análise estatística usual, por não focar em atores ou em reduções a partir de modelos, mas sim na informação relacional entre os objetos de estudo, permitindo análises tanto qualitativas quanto quantitativas de diversos temas. Qualquer conjunto de entidades que tenha, entre pares de seus elementos, relações, simétricas ou não, que possam ser explicitadas e quantificadas como características, significados, atividades ou origens, entre outras, pode ser definido como uma rede.

A análise de redes tem como objetos básicos de estudo nós, que representam os elementos da rede e arestas, que representam as relações entre os nós. No nó, atributos individuais quantitativos ou qualitativos dos elementos da rede podem ser representados por características como cor, tamanho ou formato. Nas arestas, pode-se quantificar a intensidade da relação (ou peso, representado pela espessura da linha), se esta é dirigida ou simétrica e o tipo de relação,

usando, por exemplo, cor. Das configurações de arestas, podem-se obter informações sobre características globais da rede e características contextuais dos nós, como as métricas de centralidade explicadas a seguir.

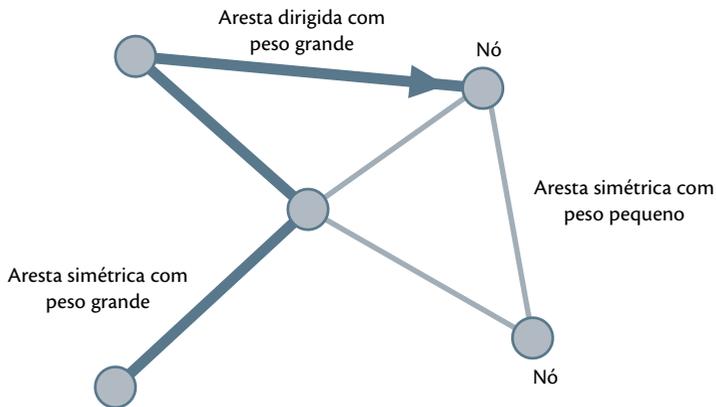


Figura 1. Ilustração de conceitos de análise de redes

Neste artigo, serão discutidas redes em que está presente apenas um tipo de aresta e redes em que existem mais de uma relação possível entre dois nós. Tais redes com mais de um tipo de aresta são chamadas de multiplex (KIVELÄ, *et al.*, 2013) e são, na verdade, duas ou mais redes “tradicionais” sobrepostas. Para que tais redes sejam válidas, as sobrepostas devem possuir características topológicas semelhantes. No caso das redes utilizadas no artigo, as arestas possíveis são relações de coautoria (aresta verde) ou relações de similaridade contextual (aresta vermelha). No caso em que ambas as relações estão presentes, a aresta representada é preta. Tanto a rede de coautorias quanto a rede de similaridade contextual possuem um alto coeficiente de agrupamento, distribuição de grau em lei de potência e estrutura de comunidade, tornando-as semelhantes topologicamente e, logo, elegíveis para formar uma rede multiplex.

4.1. Parâmetros de redes

Centralidade

As métricas de centralidade da rede têm como objetivo definir as relevâncias dos nós no contexto da rede, considerando o conjunto e os pesos das relações entre os nós (BORGATTI *et al.*, 2010). Dentre as métricas de centralidade, duas foram testadas nessa prova de conceito: a “*betweenness*” e a centralidade de autovetor.

A *betweenness* define a relevância de um nó como “ponte” entre os demais. Seu cálculo considera o quão frequentemente um nó intermedeia menores caminhos possíveis entre quaisquer outros pares de nós. No caso de uma rede de coautorias, por exemplo, a alta *betweenness* identifica os autores que tendem a separar subredes de coautores distintos. A exclusão de um nó de alta *betweenness* em uma rede tende a separar subredes que sejam conectadas por aquele nó.

A centralidade de autovetor pode ser interpretada como uma medida da “popularidade” de um nó “[...] no sentido de que um nó com alta centralidade de autovetor é conectado a nós que, por sua vez, são bem conectados [...]” (BORGATTI *et al.*, 2010). No caso de redes sociais, a centralidade de autovetor pode ser usada como indicador de possíveis líderes ou coordenadores de um grupo de pessoas.

Além da centralidade, foram usadas ou testadas outras métricas nesse trabalho. As principais estão listadas a seguir:

- a. Densidade: proporção entre o número de arestas que a rede possui e o número de arestas que uma hipotética rede completa (onde todas as relações possíveis estão presentes), com o mesmo número de nós, possuiria.
- b. Distância geodésica: menor caminho possível entre dois nós, onde o caminho é medido como o número de nós intermediários entre os dois nós.
- c. Comprimento médio de caminho: média das distâncias entre dois nós quaisquer na rede.
- d. Diâmetro: maior distância geodésica da rede, ou seja, o maior dos menores caminhos entre dois nós.
- e. Componente conectado: conjunto completo de nós conectados entre si, formando uma subrede.
- f. Grau: número de arestas conectadas a um nó.
- g. Grau médio: razão entre o número de arestas e o número de nós da rede.

Detecção de Comunidades

Redes reais possuem características muito específicas que as diferem de redes geradas aleatoriamente. Nas redes aleatórias, a probabilidade de haver uma aresta entre dois nós de uma rede é a mesma para qualquer par de nós possível. Redes reais frequentemente possuem uma estrutura de organização diferente, com a maioria dos nós tendo poucas arestas e poucos nós

tendo muitas arestas, em uma distribuição de grau que geralmente segue uma lei de potência. Outra característica que difere as redes reais das redes aleatórias é a presença de uma estrutura de comunidades, o que significa que a densidade de arestas não é homogeneamente distribuída na rede, formando agrupamentos, ou comunidades, de nós mais densamente interconectados.

Comunidades geralmente representam grupos de nós com características em comum. Em redes sociais, comunidades podem, por exemplo, representar grupos de amizade ou de pessoas que trabalham em conjunto. Em redes de páginas da internet, essas comunidades podem representar páginas que tratam de tópicos relacionados (FORTUNATO, 2010). Detectar essas comunidades é uma das aplicações possíveis da análise de redes e o significado da comunidade varia de acordo com os tipos de nós e arestas dos quais a rede é formada.

5. Redes e pesquisadores a partir de dados da plataforma Lattes

5.1. Caracterização das Redes

A disseminação e disponibilidade cada vez maior de grandes bancos de dados bibliográficos permitiram, principalmente a partir dos anos 90, a construção e análise de redes completas ou bastante abrangentes de coautores de artigos científicos, nas diversas áreas de conhecimento (M. NEWMAN 2004). Nessas redes, pesquisadores são representados por nós e as conexões são quantificadas pelo número de publicações em coautoria. A análise de redes de coautoria tem permitido um ganho de entendimento sem precedentes de padrões em redes de pesquisa, evolução desses padrões e identificação de comunidades temáticas e seus principais atores.

Os resultados relevantes obtidos com as análises de redes de conexões diretas entre pesquisadores, como é o caso das redes de coautoria, levaram naturalmente à proposição de redes baseadas em conexões indiretas, como tentativas de extrair mais informação relacional dos dados disponíveis. Exemplos desse último tipo são redes de similaridade entre publicações, onde indicadores indiretos são representados como arestas de redes de pesquisadores ou de publicações (ver, por exemplo, Yan, E ; Ding, Y. 2012). Nesse trabalho, além de testar algoritmos de identificação de coautores, com base em dados de currículos Lattes, foi empregado um algoritmo de identificação de similaridade contextual entre as publicações dos pesquisadores. Para essa rede de similaridade, novamente os nós representam pesquisadores, mas as arestas são calculadas por um algoritmo baseado em técnicas corriqueiramente empregadas na área de recuperação de informação: distância de cosseno e tf.idf (para detalhes sobre esses algoritmos,

ver Manning *et al.*, 2010). Uma expectativa do uso do algoritmo de similaridade é identificar redes formadas por semelhança temática entre currículos, mesmo que os pesquisadores jamais tenham colaborado.

Os dois algoritmos descritos, juntamente com uma rotina de extração, pré-processamento e filtragem dos dados processados, foram integrados em uma ferramenta computacional desenvolvida no CGEE e incorporada como *plugin* do *software* de visualização de redes Gephi (BASTIAN *et al.*, 2009). Os resultados mostrados a seguir devem ser entendidos como uma prova de conceito do uso da ferramenta de análise de redes para volumes de dados em larga escala. Como ocorre usualmente em processos de mineração de dados, a primeira etapa consiste na limpeza e normalização dos dados.

5.2. Limpeza dos Dados

O objetivo da primeira etapa do trabalho foi identificar os pesquisadores a serem extraídos dentre os cerca de 4 milhões de currículos do espelho da Plataforma Lattes existente no CGEE (números de outubro de 2014). Os currículos foram extraídos em formato XML, a partir de uma tabela fornecida pelo CNPq com dados de identificação dos pesquisadores que atuaram nos INCT, no período de 2008 a 2012. O processo de limpeza dos dados envolveu a eliminação de duplicações existentes no espelho do Lattes e a correção de códigos de identificação erroneamente registrados ou inexistentes na tabela de origem. Neste último caso, a extração teve que ser feita por nome, o que exigiu a exclusão manual de homônimos. Ao final do processo, menos de 2% dos currículos não puderam ser recuperados. Por se tratar de uma prova de conceito, essa perda foi considerada aceitável. O número final de currículos recuperados foi 6827, distribuídos em 118 de 125 INCT fomentados. Como o objetivo da prova de conceito é principalmente avaliar a aplicabilidade das técnicas de análise de redes e não avaliar os INCT, esse número foi considerado aceitável.

5.3. Tratamento dos dados

A partir dos dados obtidos, foram geradas as redes individuais de cada um dos INCT e duas redes completas, agrupando todos os pesquisadores dos INCT em um único grafo: uma no período de 2003 a 2007, antes de o programa INCT ser implementado, e outra no período de 2008 a 2013, considerando os cinco primeiros anos do programa. Para cada uma das duas redes completas, foram geradas três sub-redes: uma só com relações de similaridade, uma só com relações de coautoria e outra contendo ambas. Para cada rede, foram calculados agrupamentos de pesquisadores de acordo com o algoritmo de modularidade de Lambiotte *et al.* (2009). Além disso, foram também calculadas estatísticas de nós, como centralidades, grau médio e grau ponderado.

Nas duas redes globais de coautorias, a resolução da modularidade do principal componente conectado foi ajustada até o conjunto total de nós ficar agrupado em oito grupos distintos, de modo a ser possível comparar os elementos dos grupos com os pesquisadores agrupados de acordo com as grandes áreas determinadas pelo CNPq: Agrárias, Energia, Engenharia e Tecnologia da Informação, Exatas e Naturais, Humanas e Sociais, Ecologia e Meio Ambiente, Nanotecnologia, Saúde.

6. Principais resultados do artigo

Um passo natural de validação dos algoritmos de construção de redes propostos é testar se os cálculos de métricas de redes têm correspondência com dados da realidade. Uma pergunta natural nesse sentido é examinar se existe alguma relação entre medidas de relevância de nós da rede, como as métricas de centralidade, e atributos reais desses nós, como a função de gestor na rede. Para testar a hipótese, foram calculadas as centralidade de autovetor e a *betweenness* para cada nó de todas as redes, com a finalidade de identificar as posições nos *rankings* de centralidade dos coordenadores dos INCT. Considerando que a posição de coordenador é relacionada a uma carreira consagrada em suas respectivas áreas, foi empregado o conjunto completo de publicações (ao invés de publicações em um intervalo específico de tempo, por exemplo). Além disso, uma vez que os INCT são ao mesmo tempo redes de colaboração e temáticas, os cálculos foram feitos considerando o conjunto completo de arestas correspondentes a coautorias e similaridade contextual (também chamada de rede multiplex). Os resultados obtidos foram:

- 66% dos coordenadores estão entre as três primeiras posições do *ranking* de centralidade de autovetor.
- 72% dos coordenadores estão entre as três primeiras posições do *ranking* de *betweenness*.

Apesar de haver a necessidade de análises estatísticas mais detalhadas, é interessante notar que apenas com os cálculos de centralidade seja possível isolar, com razoável precisão, o coordenador de cada INCT. Como um indicativo da validade dos dados da rede multiplex obtida, vale notar que as distribuições de escores de centralidade obtidos (ver Figura 2) são consistentes com resultados de distribuições de *betweenness* de redes de coautoria relatados na literatura especializada (NEWMAN, 2004).

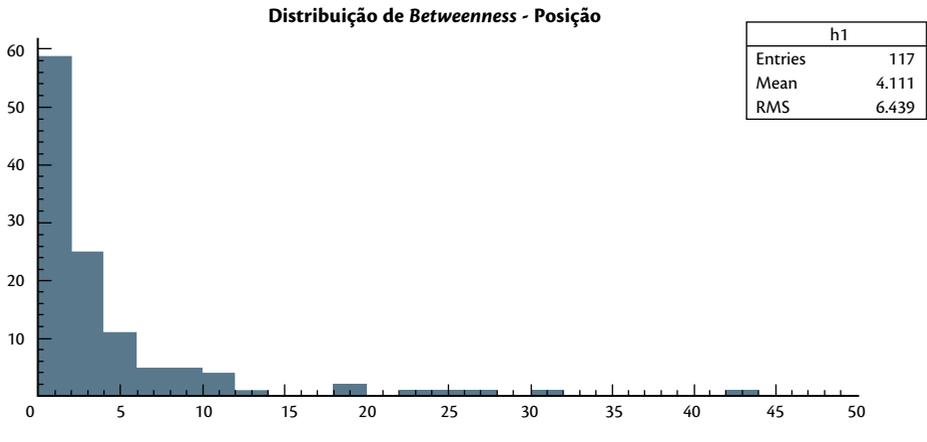


Figura 2. Distribuição da posição do coordenador no *ranking* da *betweenness*.

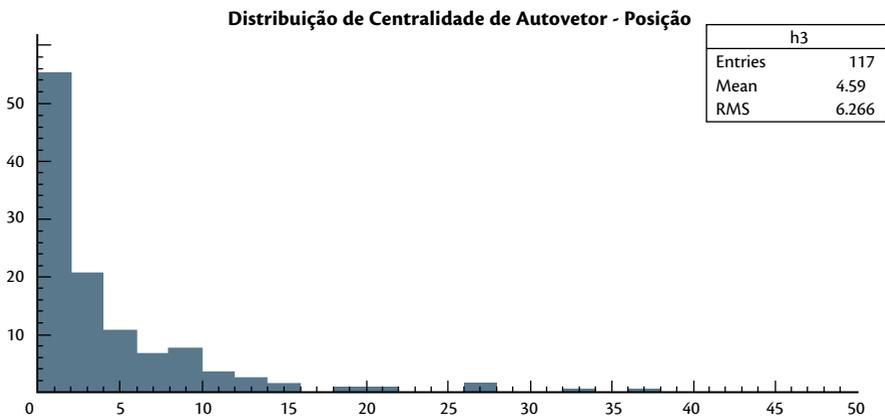


Figura 3. Distribuição da posição do coordenador no *ranking* da centralidade de autovetor.

6.1. Detecção de comunidades - algoritmo de modularidade

Como mencionado anteriormente, a detecção de comunidades é uma das possíveis aplicações da análise de redes. Em redes geradas utilizando somente as arestas de coautoria, as comunidades identificadas representam grupos de pesquisadores que publicam entre si, o que pode definir não apenas um grupo de pessoas que interagem, mas também pode ser usado para identificar, rapidamente e sem qualquer informação externa à rede, tópicos de pesquisa associados ao agrupamento observado sem a necessidade de leitura das listas de publicações, e isso já tem uma longa tradição na literatura (NEWMAN, 2004). No caso do acompanhamento da evolução dos INCT, existe outra vantagem na análise de agrupamentos: verificar se seu cálculo pode ajudar a determinar o grau de aderência entre a realidade das interações das redes e o objetivo expresso no Programa de fomentar a integração acadêmica dos grupos que apresentaram propostas.

Evolução de comunidades temáticas

No *software* utilizado nas análises desse trabalho, o Gephi, o algoritmo de detecção de comunidades utilizado é, na verdade, a combinação de dois outros: o método Louvain (BLONDEL *et al.*, 2008) e um método baseado na medida de estabilidade da rede (LAMBIOTTE *et al.*, 2009).

Uma das vantagens desses algoritmos é a possibilidade de aumentar ou diminuir a resolução dos agrupamentos. A partir das duas redes de coautorias, contendo todos os pesquisadores dos INCT (a primeira contendo as coautorias do período de 2003 a 2007 e a segunda, as de 2008 a 2013), o algoritmo de modularidade foi utilizado no maior componente conectado das redes, até que se encontrassem oito grupos, o mesmo número de grandes áreas definidas pelo CNPq (Ciências Agrárias, Energia, Engenharia e Tecnologia de Informação, Exatas e Naturais, Humanas e Sociais, Ecologia e Meio Ambiente, Nanotecnologia e Saúde). Essas oito comunidades encontradas foram comparadas com os grandes grupos, para que se pudesse identificar as oito grandes comunidades de pesquisadores que publicam entre si, nos INCT, e compará-las com as grandes áreas predefinidas.

A composição das comunidades por pesquisadores de cada uma das grandes áreas foi a seguinte:

Tabela 2. Composição das comunidades por grandes áreas dos INCT (2003-2007)

Comunidade	Agrárias	Energia	Engenharia e TI	Exatas	Humanas e Sociais	Ecologia	Nanotecnologia	Saúde	Total
0	3,9%	5,2%	6,1%	4,0%	2,8%	7,7%	6,6%	63,7%	100,0%
1	0,5%	13,5%	46,6%	8,6%	4,0%	2,9%	9,9%	14,1%	100,0%
2	10,3%	12,9%	4,5%	3,0%	2,1%	61,5%	2,2%	3,6%	100,0%
3	0,7%	4,8%	10,6%	30,2%	0,4%	3,9%	40,8%	8,7%	100,0%
4	0,4%	1,8%	3,1%	0,2%	52,4%	10,1%	0,9%	31,2%	100,0%
5	17,4%	3,7%	0,8%	1,1%	0,5%	9,0%	3,2%	64,3%	100,0%
6	0,5%	0,9%	24,1%	67,1%	0,0%	3,7%	1,4%	2,3%	100,0%
7	82,9%	0,0%	0,5%	0,5%	0,0%	1,4%	0,9%	13,8%	100,0%
%*	8,9%	6,5%	11,0%	10,1%	6,2%	15,4%	10,6%	31,2%	100,0%

Nota: * Percentual de pesquisadores da grande área no total de pesquisadores.

Tabela 3. Composição das comunidades por grandes áreas dos INCT (2008-2013)

Comunidade	Agrárias	Energia	Engenharia e TI	Exatas	Humanas e Sociais	Ecologia	Nanotecnologia	Saúde	Total
0	86,7%	1,3%	0,0%	0,0%	0,0%	12,0%	0,0%	0,0%	100,0%
1	0,0%	0,7%	52,2%	3,0%	21,4%	3,6%	1,1%	18,0%	100,0%
2	0,1%	16,8%	8,6%	4,5%	0,7%	64,2%	0,6%	4,4%	100,0%
3	0,0%	90,9%	0,0%	0,0%	0,0%	0,0%	0,0%	9,1%	100,0%
4	0,0%	0,0%	23,8%	47,9%	0,0%	2,5%	23,5%	2,3%	100,0%
5	0,3%	1,8%	13,1%	29,3%	0,0%	4,7%	43,9%	6,9%	100,0%
6	34,3%	21,8%	11,0%	3,0%	0,0%	21,2%	3,0%	5,7%	100,0%
7	10,3%	3,8%	4,2%	2,6%	2,9%	5,5%	7,2%	63,5%	100,0%
%*	9,5%	6,5%	11,2%	9,0%	3,2%	14,8%	10,9%	35,0%	100,0%

Nota: * Percentual de pesquisadores da grande área no total de pesquisadores.

A distribuição dos pesquisadores de cada grande área por comunidades foi a seguinte:

Tabela 4. Distribuição dos pesquisadores de cada grande área por comunidades obtidas (2003-2007)

	Comunidade								Total
	0	1	2	3	4	5	6	7	
Agrárias	8,1%	0,7%	19,1%	1,3%	0,4%	37,2%	0,2%	33,0%	100,0%
Energia	14,7%	26,9%	32,3%	12,4%	2,5%	10,7%	0,5%	0,0%	100,0%
Engenharia e TI	10,3%	55,0%	6,6%	16,4%	2,5%	1,3%	7,7%	0,1%	100,0%
Exatas	7,4%	11,1%	4,8%	51,0%	0,2%	2,1%	23,3%	0,2%	100,0%
Humanas e Sociais	8,3%	8,3%	5,5%	1,0%	75,3%	1,6%	0,0%	0,0%	100,0%
Ecologia	9,3%	2,4%	65,7%	4,3%	5,9%	11,1%	0,8%	0,3%	100,0%
Nanotecnologia	11,5%	12,1%	3,4%	65,7%	0,8%	5,7%	0,5%	0,3%	100,0%
Saúde	37,8%	5,9%	1,9%	4,7%	9,0%	38,9%	0,3%	1,6%	100,0%
%*	18,5%	13,0%	16,4%	17,1%	9,0%	18,9%	3,5%	3,5%	100,0%

Nota: * Percentual de pesquisadores da comunidade no total de pesquisadores.

Tabela 5. Distribuição dos pesquisadores de cada grande área por comunidades obtidas (2008-2013)

	Comunidade								Total
	0	1	2	3	4	5	6	7	
Agrárias	24,2%	0,0%	0,2%	0,0%	0,0%	0,4%	21,4%	53,8%	100,0%
Energia	0,5%	0,8%	38,3%	8,2%	0,0%	3,5%	19,8%	28,8%	100,0%
Engenharia e TI	0,0%	36,1%	11,4%	0,0%	13,2%	14,7%	5,8%	18,8%	100,0%
Exatas	0,0%	2,5%	7,5%	0,0%	33,1%	40,8%	2,0%	14,1%	100,0%
Humanas e Sociais	0,0%	51,4%	3,3%	0,0%	0,0%	0,0%	0,0%	45,4%	100,0%
Ecologia	2,1%	1,9%	64,0%	0,0%	1,1%	3,9%	8,4%	18,5%	100,0%
Nanotecnologia	0,0%	0,8%	0,8%	0,0%	13,5%	50,5%	1,6%	32,8%	100,0%
Saúde	0,0%	4,0%	1,9%	0,2%	0,4%	2,5%	1,0%	90,2%	100,0%
%*	2,6%	7,7%	14,8%	0,6%	6,2%	12,5%	5,9%	49,6%	100,0%

Nota: * Percentual de pesquisadores da comunidade no total de pesquisadores.

Histogramas a seguir mostram a distribuição dos pesquisadores das grandes áreas nos grupos identificados pelo algoritmo de detecção de comunidades:

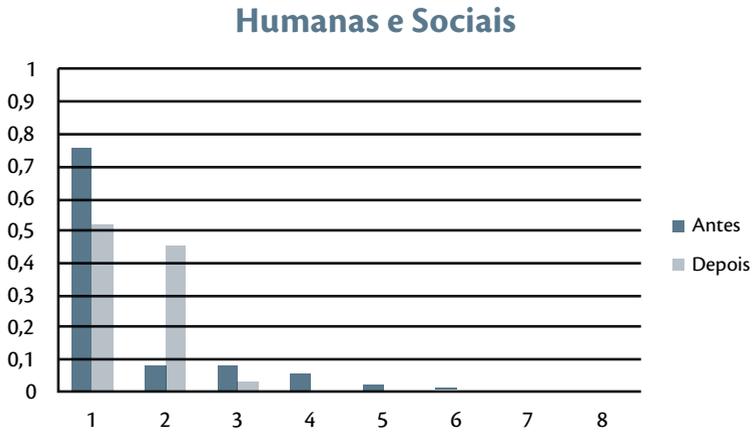


Figura 4. Distribuição dos pesquisadores de Ciências Humanas e Sociais antes e depois do programa INCT

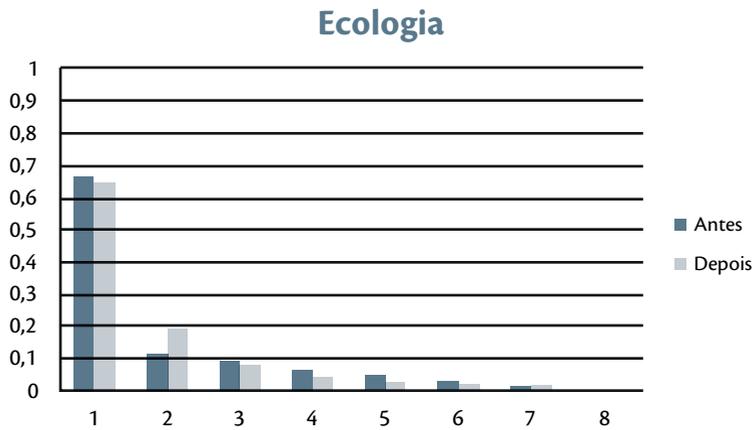


Figura 5. Distribuição dos pesquisadores de Ecologia antes e depois do programa INCT

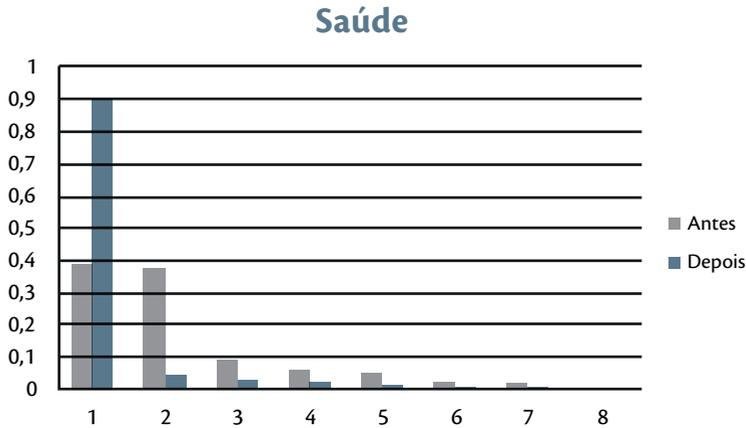


Figura 6. Distribuição dos pesquisadores de Saúde antes e depois do programa INCT

Os dados obtidos sugerem que :

- Houve uma concentração clara dos pesquisadores da grande área de Saúde. Pesquisadores identificados antes do programa como pertencentes a grupos diferentes se concentraram em um grupo principal, com quase 90% dos pesquisadores da área, o que sugere que eles passaram a trabalhar mais em conjunto ou então a tratar de temas semelhantes. Essa concentração, embora mais amenizada, também foi observada nas grandes áreas de Agrárias e Energia.
- Na grande área de Ciências Humanas e Sociais ocorreu o movimento oposto. Aproximadamente 75% dos pesquisadores que foram identificados como pertencentes a um mesmo grupo no programa, agora, encontram-se distribuídos entre dois grupos principais, cada um contendo aproximadamente 40% a 50% dos pesquisadores da grande área. Esse movimento também foi observado, embora de forma mais amenizada, nas grandes áreas de Engenharia e Tecnologia da Informação.
- Na grande área de Ecologia, a separação dos pesquisadores por grupo, depois do programa INCT, permaneceu muito parecida com o que ela era antes.
- Nas outras grandes áreas, não foi possível identificar um movimento tão claro, tendo havido concentração em alguns grupos, desconcentração em outros e, ainda, outros que permaneceram iguais.

6.2. Pertinência do financiamento x multidisciplinaridade

Também foi efetuado um teste em pequena escala, utilizando três INCT (INCT 1, INCT 2 e INCT 3) que tratam de temas similares no período de 2008 a 2013, gerando-se uma rede com as coautorias dos pesquisadores desses INCT. A partir do maior componente conectado da rede, foi utilizado o algoritmo de detecção de comunidades do Gephi, para verificar o nível de correspondência entre as comunidades encontradas e os INCT originais. A resolução do algoritmo foi ajustada, até que fossem encontradas três comunidades, ou seja, até que o número de comunidades fosse igual ao número de INCT contidos na rede.

Os resultados mostram que existem três grandes comunidades de pesquisadores que publicam em colaboração:

- A Comunidade A: é composta por 109 membros do INCT 1 e 39 membros do INCT 3. A maior parte dos membros do INCT 1 está na comunidade A, enquanto aproximadamente metade dos membros do INCT 3 está na comunidade B.
- A Comunidade B: contém membros dos três INCT, porém, é composta majoritariamente por 33 membros do INCT 3, com uma pequena parcela de dez membros do INCT 2 e 23 membros do INCT 1.
- A Comunidade C: que contém 42 membros do INCT 2 e não tem participação de mais nenhum INCT.

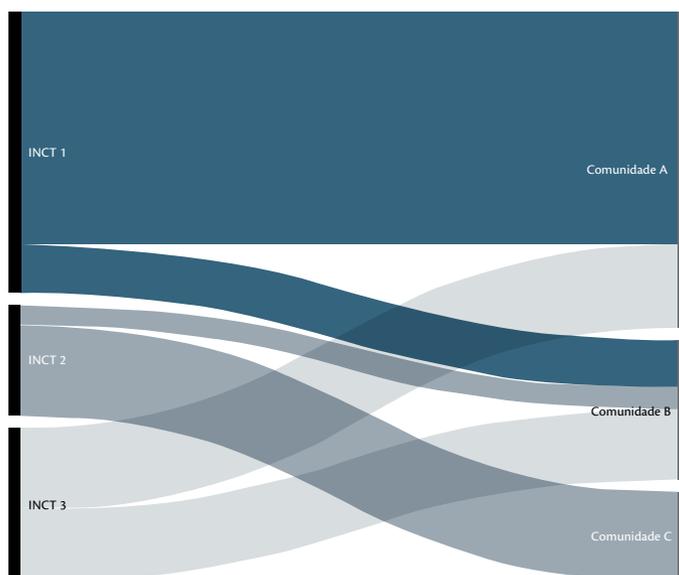


Figura 7. Fluxo dos pesquisadores dos INCT 1, 2 e 3 para as Comunidades A, B, e C.

Os resultados obtidos não permitem, por si, definir um juízo de valor sobre os INCT estudados, pois a divisão dos INCT 1 e 3 em duas grandes comunidades pode indicar tanto que o fomento não foi eficaz na indução do trabalho em rede como pode ser entendido com uma rede com um forte foco na interdisciplinaridade das comunidades A e B. Em qualquer caso, entretanto, os resultados sugerem que o INCT 2 trata-se de uma comunidade focada em um grande tema. Os dados obtidos sobre os INCT 1 e 3 podem, contudo, ajudar as instituições responsáveis pelo Programa a interpretar o nível de aderência dessas redes aos objetivos previstos. Em qualquer caso, o uso dessas técnicas no acompanhamento da implementação de futuras ações similares permitirá possibilidades de correções de curso, com base nas evidências coletadas.

7. Conclusão

Neste artigo, foram reportados os principais resultados de uma prova de conceito da aplicação de ferramentas de análise de redes, como passo intermediário de uma metodologia potencialmente eficiente para explorar a *expertise* demandada por políticas públicas por meio do financiamento de grandes grupos de pesquisa. Foram mostradas evidências de que esses métodos podem contribuir para a avaliação e o acompanhamento das redes existentes e para o planejamento e a implementação de futuras ações no mesmo sentido. Os primeiros resultados foram promissores para alguns parâmetros de análise de redes e inconclusivos para outros. Entretanto, os resultados conclusivos já permitem antever um papel relevante da análise de redes como instrumento de apoio à análise de políticas públicas, tanto para tornar mais efetiva a atuação de cada INCT, de acordo com os objetivos explícitos do programa, como para fomentar possíveis articulações ou trabalhos conjuntos entre INCT, ou subconjuntos de INCT, para responder a grandes desafios nacionais.

Agradecimento

Os autores agradecem a Eduardo do Couto e Silva, por suas importantes sugestões no uso de algumas das ferramentas computacionais e análises estatísticas empregadas neste trabalho e pela leitura crítica do manuscrito. Agradecem, ainda, a Marcelo Paiva, por contribuições nas etapas iniciais do trabalho.

Referências

- BASTIAN M.; HEYMANN S.; JACOMY M. Gephi: an open source *software* for exploring and manipulating networks. **International AAAI Conference on Weblogs and Social Media**, 2009.
- BLONDEL, D.; GUILLAUME, J.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks, **J. Stat. Mech**, P10008, 2008.
- BORGATTI, S. P.; EVERETT, M. G.; JOHNSON, J. C. **Analyzing social networks**. SAGE Publications Limited, Londres, UK, 2013.
- CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO. **Chamada INCT – MCTI/CNPq/ CAPES/FAPs nº 16/2014**. Brasília, DF: CNPq, 2014, 38 p. Disponível em: <http://inct.cnpq.br/documents/10180/124986/Chamada+INCT_16-2014.pdf/3d511440-8d6f-413c-ac64-176b7aco2902>. Acesso em: 18 mai. 2015.
- CAGNIN, C.; AMANATIDOU, E.; KEENAN, M. Orienting European innovation systems towards grand challenges and the roles that FTA can play, **Science and Public Policy**, n. 39, p.140-152, 2012.
- CAGNIN, C.; KEENAN, M. JOHNSTON, R.; SCAPOLO, F.; BARRÉ, R. Future-oriented technology analysis – strategic intelligence for an innovative economy. **Springer-Verlag Berlin Heidelberg**, 2008.
- FORTUNATO, S. Community detection in graphs, **Physics Reports**. n.486, p. 75-174, 2010.
- GASSLER, H.; POLT, W.; RAMMER, C. Priority setting in technology policy – historical developments and recent trends. NAUWELAERS, C. AND WINTJES, R. (eds) **Innovation Policy in Europe**. Cheltenham: Edward Elgar, 2008.
- GEORGIU, L.; CASSINGENA HARPER, J. Policy transfer and learning, **The Handbook of Technology Foresight: Concepts and Practice**. Edward Elgar Publishing Limited, Cheltenham, UK, p.309-341, 2008.
- KIVELÄ, M.; ARENAS, A.; BARTHELEMY, M.; GLEESON, J.; MORENO, Y.; PORTER, M. Multilayer networks, **J. Complex Netw.** v. 2, n. 3, p. 203-271 . 2014.
- LAMBIOTTE, R.; DELVENNE, J.; BARAHONA, M. Laplacian dynamics and multiscale modular structure in networks. **IEEE Transactions on Network Science and Engineering**. v.1, n. 2, p. 76-90, 2009.
- MANNING, C. D.; RAGHAVAN, P.; SCHULTZE, H. **Introduction to information retrieval**, Cambridge University Press, 2010.
- NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration, **Proceedings of the National Academy of Sciences USA**, v. 101, spl.1, p. 5200-5205, 2004.
- _____. **Networks: an introduction**. USA: Oxford University Press, 2010.

NUGROHO, Y.; SARITAS, O. Incorporating network perspectives in foresight: a methodological proposal, **Foresight**, v. 11, n. 6, p. 21-41, 2009.

PROGRAMA INSTITUTOS NACIONAIS DE C&T. **Documento de Orientação aprovado pelo Comitê de Coordenação**. 29 jul. 2008. Disponível em: <http://estatico.cnpq.br/programas/inct/_apresentacao/pdf/015_anexo.pdf>. Acesso em: 31 mar. 2015.

RITTEL, H. ; WEBER, M. Dilemmas in a general theory of planning, **Policy Sciences**, n.4, p. 155–69, 1973.

YAN, E.; DING, Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and cword networks relate to each other. **Journal of the Association for Information Science and Technology**, v. 63, n. 7, p. 1313-1327, 2012.