

Atividade: Desenvolvimento de competências e ferramentas em prospecção, avaliação estratégica, gestão de informação e do conhecimento

Exploração de dados e visualização de informação

Atividade: Desenvolvimento de competências e ferramentas em prospecção, avaliação estratégica, gestão de informação e do conhecimento

Exploração de dados e visualização de informação



Brasília, DF
dezembro, 2018

Centro de Gestão e Estudos Estratégicos (CGEE)

Organização social supervisionada pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC)

Presidente

Marcio de Miranda Santos

Diretores

Regina Maria Silverio

Joaquim Aparecido Machado

Desenvolvimento de competências e ferramentas em prospecção, avaliação estratégica, gestão de informação e do conhecimento. Exploração de dados e visualização de informação. Brasília: Centro de Gestão e Estudos Estratégicos, 2018.

199p. : il.

1. Ciência de dados. 2. Análise de redes complexas. 3. Ciência e Tecnologia
I. CGEE. II. Título.

Centro de Gestão e Estudos Estratégicos (CGEE), SCS Qd 9, Torre C, 4º andar, Ed. Parque Cidade Corporate, CEP: 70308-200 - Brasília, DF, Telefone: (61) 3424 9600, <http://www.cgее.org.br>

Todos os direitos reservados pelo Centro de Gestão e Estudos Estratégicos (CGEE). Os textos contidos nesta publicação poderão ser reproduzidos, armazenados ou transmitidos, desde que seja citada a fonte.

Referência bibliográfica:

Centro de Gestão e Estudos Estratégicos- CGEE. Desenvolvimento de competências e ferramentas em prospecção, avaliação estratégica, gestão da informação e do conhecimento. Exploração de dados e visualização de informação. Brasília, DF: 2018. 199p.

Este relatório é parte integrante das atividades desenvolvidas no âmbito do Contrato de Gestão. Programa: (Exploração de dados e visualização de informação). Projeto – (8.10.56.01.51.01).

Atividade: Desenvolvimento de competências e ferramentas em prospecção, avaliação estratégica, gestão de informação e do conhecimento

Exploração de dados e visualização de informação

Supervisão

Joaquim Aparecido Machado

Consultor

Jackson Max Furtunato Maia

Equipe técnica do CGEE

Eduardo Amadeu Dutra Moresi

Flávia Montandon Fagundes Pinto

Rogério da Silva Castro

SUMÁRIO

1. INTRODUÇÃO.....	2
2. ANÁLISE DE REDES COMPLEXAS.....	2
2.1 INSIGHT NET	3
2.2 INSIGHT NET BROWSER.....	4
3. MINERAÇÃO DE TEXTOS.....	5
4. TESTES DE ALGORITMOS E PROTÓTIPOS	5
5. INTERAÇÃO COM OUTROS PROJETOS DO CENTRO	7
6. JOURNAL CLUB.....	8
7. PARTICIPAÇÃO EM EVENTO.....	9
8. ALTERAÇÕES NO PROJETO E PERSPECTIVAS PARA 2019	9
Anexo I (índice no próprio anexo)	16
Anexo II (Índice no próprio anexo).....	128

1. INTRODUÇÃO

O reconhecimento e análise de padrões nas grandes massas de dados atualmente acessíveis permitem multiplicar a capacidade de atuação do CGEE, desde que as técnicas adequadas de extração, tratamento e carga de dados sejam empregadas para obter essa informação. Nesse sentido, o projeto "Exploração de Dados e Visualização de Informações" visa fortalecer as competências do Centro, desenvolvendo e validando fundamentos, metodologias e ferramentas de análise de dados de CTI disponíveis, ampliando seu portfólio de serviços e auxiliando o embasamento metodológico das suas demais atividades e ações.

Este relatório sumariza os principais resultados e contribuições obtidos em 2018 no Projeto de Atividade. Os dois principais resultados, versões com várias novas funcionalidades das ferramentas de análise e navegação de redes, os softwares Insight Net e Insight Net Browser, são objeto dos documentos "Manual CGEE Insight Net: Release 3.1.0" e "InsightNet Browser - Guia do usuário: Release 1.6.2", anexos I e II deste relatório, onde são descritas em detalhes todas as funcionalidades dos programas. Nas seções seguintes são relatadas as principais atividades e avanços realizados em 2018 no âmbito deste Projeto.

2. ANÁLISE DE REDES COMPLEXAS

Ao longo dos últimos anos o CGEE vem consolidando a área de análise de redes complexas aplicada a bases de dados de interesse em CTI. Ao longo das implementações em software dos processos de análise, foi percebida a conveniência da separação entre a fase de análises exploratórias de dados, voltada para a mineração de dados e reconhecimento de padrões, e a fase de comunicação dos dados minerados e padrões reconhecidos. A separação definiu duas ferramentas com funções correspondentes. No Insight Net (iN), o usuário realiza o trabalho de análise e

preparação dos arquivos de rede a serem exportados para posterior exploração visual na outra ferramenta, o Insight Net Browser (iNB). Posteriormente, algoritmos necessários foram incorporados à ferramenta de mineração textual Insight Data (iD), de modo que seu acervo também pode ser analisado em redes de similaridade semântica de documentos no Insight Net e seus resultados exportados para o Insight Net Browser. Com esta conexão entre suas três principais ferramentas de análise de dados, o Centro tem a possibilidade de realizar análises de redes com documentos do acervo próprio (integrado ao Insight Data), de currículos Lattes e de metadados das principais bases de dados de produção acadêmica existentes.

2.1 INSIGHT NET

Na linha de desenvolvimento de análises de redes de documentos e currículos Lattes, centrada nas funcionalidades da ferramenta iN, foram desenvolvidas novas funcionalidades em 2018, ressaltando-se:

- a) A geração de redes de coocorrências de palavras-chave com opção de fusão de conceitos sinônimos para facilitar a análises de contextos e de áreas de conhecimento. Esta funcionalidade replica no iN redes semelhantes às do software VosViewer, amplamente utilizado na área de cientometria;
- b) A versão 3.1 da ferramenta de software Insight Net, resultado da migração de todas as funcionalidades da versão anterior, a iN 2, para a versão 0.9.2 do sistema Gephi, agora baseada em Java 8 (a iN 2 dependia do Java 7, que foi descontinuado). No lançamento dessa versão foi também testada a redação do manual na ferramenta Sphinx, que, além de permitir a geração do mesmo texto em formatos PDF e HTML, facilita muito a redação de novas versões;
- c) Foram testados três critérios matemáticos propostos internamente no CGEE para a determinação do corte inferior dos pesos de redes de similaridade semântica, com resultados promissores no sentido de automatizar a geração desse tipo de redes;

- d) Foram realizados testes da conversão do banco de dados para estrutura NoSQL, de acordo com o padrão atualmente utilizado na ferramenta Insight Data, com o propósito de viabilizar a manipulação de metadados extraídos mas não totalmente aproveitados de currículos e artigos, além de facilitar futuras integrações entre as ferramentas;
- e) Foram implementadas melhorias de desempenho que reduziram o tempo de importação de dados de arquivos grandes por um fator entre 5 e 10; e
- f) Nova opção de importação de dados textuais, incluindo todos os formatos disponibilizados pelas principais bases de dados científicos e tecnológicos e que permite, também, a importação de dados de planilhas Excel ou formatos CSV genéricos;

2.2 INSIGHT NET BROWSER

No que diz respeito a atividades de consolidação e ampliação de metodologias para exploração e visualização de dados, foi entregue a versão 1.6 da ferramenta Insight Net Browser (também preparada no Sphinx, para facilitar as edições posteriores), com novas funcionalidades, a saber:

- a) Exibição de metadados das redes visualizadas;
- b) Definição por cores de quaisquer partições baseadas em atributos categóricos existentes na rede;
- c) Janela com tabelas de nós e arestas com capacidade de busca, seleção, pesquisa, ordenação e exportação de dados tabulados;
- d) Ambiente de apresentação, onde visualizações de dados da rede selecionados pelo usuário podem ser exibidas em sequência de slides, incluindo a possibilidade de explorações interativas dos dados com a plateia e salvamento de apresentações diferentes para uma mesma rede;
- e) Foram realizadas melhorias significativas no módulo de pesquisa avançada, que agora permite buscas com operadores lógicos incluindo a possibilidade de buscas em mais de um atributo dos dados (termo no título e ano, por exemplo);

- f) Funcionalidade de backup e restauração da rede, que permite a recuperação de todo o processo de exploração dos dados, incluindo configurações, customizações, pesquisas, partições, visualizações e apresentações;
- g) Foi testada a criação de um conversor do formato de dados GML para o GEXF nativo do iNB, viabilizando assim a exibição de redes geradas por outros programas além do Gephi; e
- h) Criação de um novo módulo de exploração de palavras-chave, com geração de tabela exportável de ocorrências de palavras-chave por ano de publicação dos artigos que as contém, exibição em gráfico de linha da evolução dessas ocorrências. O módulo também permite a seleção e pesquisa de nós tanto na tabela quanto no gráfico e foi implementada ainda a geração de gráficos para visualização de coocorrências e para rápida visualização das frequências das palavras-chave;

3. MINERAÇÃO DE TEXTOS

Na linha de desenvolvimento da ferramenta de mineração textual Insight Data, foram implementadas melhorias na interface de carga de documentos (Insight Data Client), que agora permite a inserção de documentos em lote por meio de planilhas Excel, incluindo uma funcionalidade de inserção do banco de teses e dissertações da CAPES.

4. TESTES DE ALGORITMOS E PROTÓTIPOS

Como forma de testar possibilidades de futuras implementações em ferramentas próprias do Centro, foram iniciados desenvolvimentos colaborativos com estudantes e/ou pesquisadores de quatro instituições de ensino superior, que resultaram, respectivamente, em testes de métodos e protótipos funcionais. Os principais resultados foram:

- a) Com o Dr. Ricardo Sampaio, pesquisador da Fundação Oswaldo Cruz e

professor da Universidade de Brasília, juntamente com alunos do seu curso sobre Ciência de Dados, foi desenvolvido um protótipo em software para a extração, tratamento e normalização de dados de produção tecnológica (patentes, registros de software, cultivares, topologias de circuitos integrados e marcas) registrados na plataforma Lattes. A ferramenta foi desenvolvida no modelo *open source*, foi implementada no ambiente R, testada com sucesso em outro projeto do CGEE e está disponível para download no endereço https://github.com/matheusbsilva/lattes_patents. Os resultados iniciais mostram forte potencial de incorporação dos seus algoritmos na ferramenta Insight Net;

- b) Com o Prof. Jesús Mena-Chalco e o estudante de doutorado Luciano Rossi, ambos da Universidade Federal do ABC, em São Paulo, foi testado um algoritmo em Python de modelagem de tópicos para caracterização de documentos que pode complementar o uso de palavras-chave para a caracterização de documentos em redes de similaridade semântica;
- c) Em continuação à interação iniciada em 2017 com o grupo do Prof. Antônio Brasil, da Universidade Federal do Rio de Janeiro, foram testadas várias técnicas de análise em bases de dados específicos de Ciências Humanas e Sociais. Essa relação foi iniciada a partir do desenvolvimento de aplicações de análises de redes a dados do Lattes da área de políticas culturais que foram incorporadas no mestrado de Marcelo Paiva, posteriormente contratado pelo CGEE. A especificidade dos dados das áreas de ciências humanas e sociais é um desafio em si que abre novas possibilidades de atuação para o Centro;
- d) Como tema de desenvolvimento de trabalho de final de curso e estágio acadêmico do estudante de Engenharia Mecatrônica Otto Glass, foram desenvolvidos e testados algoritmos de aprendizado de máquina, particularmente redes neurais, para a previsão de arestas em redes de coautorias. Os primeiros resultados foram promissores no sentido da estimativa de tendências de evolução de colaboração de grupos de pesquisa. Além disso, considerando que a análise de redes e o processamento de linguagem natural

(cujas técnicas já são fortemente empregadas nas ferramentas iN e iD) fazem parte, junto com o aprendizado de máquina, do conjunto de técnicas que compreendem a área emergente da Ciência de Dados, pode-se afirmar que a incorporação de técnicas de redes neurais a seu portfólio de técnicas aproximará o CGEE ainda mais da fronteira do conhecimento, no que diz respeito ao apoio à tomada de decisão com base em evidências.

5. INTERAÇÃO COM OUTROS PROJETOS DO CENTRO

Houve intensa participação da equipe no projeto “Apoio Técnico ao CNPq na Utilização de Métodos e Ferramentas Modernas de Inteligência em CTI.” Esse projeto, executado em forte colaboração com equipes do Conselho Nacional de Desenvolvimento Científico e Tecnológico, consistiu de 3 subprojetos encomendados para:

- a) Subsidiar o planejamento estratégico daquela Agência por meio da elaboração de cenários para a sua atuação futura;
- b) Realizar análises exploratórias de dados para o desenvolvimento de indicadores de potencial de inovação dos Institutos Nacionais de Ciência e Tecnologia; e
- c) Realizar análises de resultados do programa temático SISBIOTA e do programa permanente associado ao Comitê Assessor de Ciências Ambientais.

A colaboração com o CNPq foi fundamental para o desenvolvimento de novas técnicas de identificação de padrões informacionais relacionados à inovação tecnológica, e deverão ser incorporados ao portfólio de ferramentas metodológicas do Centro. Em particular, o protótipo da ferramenta de extração de produções tecnológicas a partir de dados do currículo Lattes, mencionada acima, foi intensamente testado no ambiente do projeto com o CNPq.

Ocorreu, ainda, cooperação interna com os projetos “Diagnóstico da Situação Atual nas CHSSA Brasileiras” e “Avaliação do Programa de Educação Tutorial – PET”, nos

quais foram testados, em bases de resumos de teses e resumos de projetos, respectivamente, as novas funcionalidades de importação de textos genéricos e os algoritmos para cálculo de pesos ótimos para a similaridade semântica entre documentos do iN, além de testes das novas opções de navegação do iNB. Adicionalmente, deve-se notar que a base de dados de resumos de teses empregada no “Diagnóstico da Situação Atual nas CHSSA brasileiras” é um subconjunto dos dados obtidos para a operação da nova versão do Insight Data Client, mencionada acima.

Por fim, em uma colaboração que só tende a crescer ao longo de 2019, vários métodos e novas funcionalidades listadas nas seções anteriores foram testados ou discutidos no contexto da Atividade “Observatório em Ciência, Tecnologia e Inovação”, particularmente, testes de candidatas a métricas de inovação.

6. JOURNAL CLUB

Ao longo do ano, foi dada continuidade às apresentações do Journal Club, espaço interno do CGEE dedicado à discussão e ao aprofundamento de aspectos metodológicos (conceitos, métodos e ferramentas) de potencial interesse para uso no Centro, tendo sido apresentados e debatidos os seguintes artigos:

- a) On the role of design in information visualization;
- b) Tracking researchers and their outputs: new insights from ORCID;
- c) The current state of fake news: challenges and opportunities,
- d) Getting Things Done: The Science behind Stress-Free Productivity;
- e) From Data Quality to Big Data Quality;
- f) Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature;
- g) Data-driven, networked urbanism;
- h) Identificação do grafo de genealogia acadêmica de pesquisadores: Uma abordagem baseada na Plataforma Lattes;

- i) Governança digital como vetor para uma nova geração de tecnologias de participação social no Brasil; e
- j) O método Científico como mediação esclarecida em Negociação Multilateral sobre o Acesso a Recursos Genéticos e Repartição de Benefícios.

7. PARTICIPAÇÃO EM EVENTO

"FTA 2018 - Future in the making" em junho de 2018, em Bruxelas, onde foi apresentada uma técnica desenvolvida no CGEE na área de mineração e organização de dados bibliométricos, sob o título "Taxonomy from Keywords: A methodology for understanding new concepts and design future-oriented and evidence-based STI policies".

8. ALTERAÇÕES NO PROJETO E PERSPECTIVAS PARA 2019

Esforços substancialmente acima do que tinha sido antecipado na fase de planejamento dos itens 5.b e 5.c do projeto "Apoio Técnico ao CNPq na Utilização de Métodos e Ferramentas Modernas de Inteligência em CTI." foram exigidos da equipe deste projeto, desencadeando atrasos em algumas das suas metas. Neste sentido, foram realizadas alterações nos prazos de execução de algumas de suas metas.

O prazo previsto para a elaboração de documento de especificações da ferramenta Insign Tech foi adiado para junho de 2019.

As metas de elaboração de documento contendo propostas de modelo de transferência de metodologias de exploração e visualização de dados, bem como o estudo de alternativas de proteção à propriedade intelectual das tecnologias desenvolvidas foram fundidas em uma meta única com prazo adiado para junho de 2019. A fusão de metas considerou a dependência entre os dois resultados.

Ao longo da sua execução, a Atividade “Exploração de Dados e Visualização de Informação” produziu um aumento substancial de complexidade em métodos e ferramentas que, por sua vez, multiplicaram o potencial de serviços que o Centro pode prestar no auxílio à tomada de decisão com base em evidências. Por outro lado, apenas uma fração da informação contida nas importantes bases de dados disponíveis pode realmente ser explorada com as técnicas existentes. Além da organização do legado de ferramentas e métodos desenvolvidos em um corpo metodológico consistente e, eventualmente, em uma plataforma de software integrada, percebeu-se a necessidade da elaboração e da formalização de estratégias de desenvolvimentos futuros que considerem o estado da arte da Ciência de Dados na sua concepção. Assim, com o objetivo de subsidiar a atuação futura do CGEE no que diz respeito à análise e visualização de dados, informação e conhecimento, foi proposta uma nova meta que consiste em promover um estudo contínuo sobre conceitos e técnicas inovadores em Ciência de Dados e Representação do Conhecimento que apresentem potencial impacto estratégico para o Centro. Essa meta prevê a entrega de relatórios periódicos que tenham foco, no mínimo, em:

- a) Levantamentos do estado da arte nas áreas de conhecimento abordadas e uma seleção de conceitos métodos, técnicas e algoritmos que possam ser empregados pelo CGEE;
- b) Avaliações de viabilidade de implementação no Centro das técnicas, conceitos, métodos e algoritmos levantados;
- c) Relatos de resultados de protótipos e provas de conceito eventualmente testados, considerando rotas de implementação, nos casos considerados promissores.

Anexo I

Manual CGEE Insight Net

Release 3.1.0

30 de dezembro de 2018

1	Introdução	1
1.1	Contexto e Visão Geral	1
1.2	Funcionalidades experimentais	2
2	Instalação do CGEE Insight Net	3
2.1	Pré-requisitos	3
2.2	Instalação do software Gephi	4
2.3	Configuração da central de atualizações	5
2.4	Instalação do <i>CGEE Insight Net</i>	7
2.5	Atualização do <i>CGEE Insight Net</i>	13
3	Configuração do CGEE Insight Net	15
3.1	Configuração do banco de dados	16
3.2	Configuração do usuário para acessar o banco de dados de Currículos Lattes do CGEE .	17
3.3	Configuração das colunas exibidas	18
3.4	Exibição da relevância das palavras-chave	20
3.5	Parâmetros da pesquisa por similaridade	21
3.6	Licenças	23
3.7	Protocolos de execução	24
3.8	Memória <i>cache</i> de Currículos Lattes	25
4	Conceitos gerais do uso do <i>CGEE Insight Net</i>	27
4.1	Fluxo de trabalho	27
5	Uso do <i>CGEE Insight Net</i> para analisar Currículos Lattes	31
5.1	Importação dos Currículos Lattes	33
5.2	Formação da rede	40
5.3	Visualização de atributos dos pesquisadores	45
6	Criação de redes de referências bibliográficas BibTeX	49
6.1	Importação dos dados bibliográficos	51
6.2	Formação da rede	55
7	Criação de redes de referências bibliográficas genéricas	59
7.1	Importação dos dados bibliográficos	61
7.2	Formação da rede	66

8	Análise das redes criadas	69
8.1	Filtragem dos resultados	69
8.2	Análise de clusters	73
8.3	Análise de assortatividade	74
8.4	Análise das palavras-chave	77
8.5	Criação de redes de co-ocorrências de palavras-chave	90
8.6	Eliminação interativa de nós da rede e do banco de dados	91
8.7	Criação de uma nova rede a partir do subconjunto de nós selecionados	93
8.8	Seleção interativa de nós vizinhos na rede	94
8.9	Visualização interativa do currículo de pesquisadores no browser	95
9	Funcionalidades comuns de apoio	99
9.1	Recuperação do grafo a partir das informações que constam no banco de dados	100
9.2	Cópia e recuperação do banco de dados	102
9.3	Estatísticas do banco de dados	104
9.4	Protocolos de execução	105
	Referências Bibliográficas	107

1.1 Contexto e Visão Geral

O *plugin CGEE insight Net* foi concebido para operar junto com o software *Gephi*¹ para viabilizar a análise de grandes volumes de dados disponíveis para o CGEE de modo a organizá-los como redes complexas manipuláveis por usuários com pouca experiência ou treinamento em programação. Essa ferramenta vem sendo continuamente desenvolvida no CGEE desde 2013. A aplicação tem se mostrado eficaz para visualizar redes de coautorias e de similaridade temática entre currículos disponibilizados na Plataforma Lattes do CNPq e de similaridade temática entre artigos disponibilizados em grandes bases de dados como *Scopus* e *Web of Science*, embora outras fontes textuais também possam ser exploradas. Este guia de usuário descreve a versão 3 do *plug-in*, com detalhes sobre a instalação dos seus componentes e suas principais funcionalidades.

¹ <http://www.gephi.org>

1.2 Funcionalidades experimentais

O CGEE Insight Net está em um processo de aprimoramento constante e a grande maioria das funcionalidades se encontra em um estado robusto e estável. Outros métodos e algoritmos foram acrescentados apenas recentemente e podem apresentar instabilidades ou deficiências no processamento de dados específicos. Essas funcionalidades estão marcadas no *plugin* com a palavra `EXPERIMENTAL` ou o símbolo



Instalação do CGEE Insight Net

2.1 Pré-requisitos

O *CGEE Insight Net* usa a versão 0.9.2 do software Gephi. Na data da atualização do presente manual de sistemas. A versão 0.9.2 do Gephi depende da instalação do ambiente Java na versão 1.8 nos ambientes de Windows, Linux e macOS.

2.2 Instalação do software Gephi

O software Gephi pode ser baixado na versão 0.9.2 pela página da ferramenta na internet:

<https://github.com/gephi/gephi/releases>

O software vem com um instalador automático que suporta os principais sistemas operacionais. O processo é documentado no site do Gephi¹

As principais características do Gephi podem ser revisadas na página <https://gephi.org/features/>.

Caso sejam instaladas diversas versões do Java no computador do usuário, o Gephi permite a configuração de qual delas deve ser usada. Para isso, existe uma variável `jdkhome` no arquivo `gephi.conf` que consta no subdiretório `etc` da instalação do Gephi. **Recomenda-se tornar este arquivo “gephi.conf” gravável para o usuário comum.**

¹ <https://gephi.org>

2.3 Configuração da central de atualizações

Esse passo é necessário apenas uma vez para cada computador. O *CGEE Insight Net* é disponibilizado online, no seguinte endereço:

<http://projeto.cgee.org.br/InsightNet3/release/updates.xml>

Para instalar o módulo, esse endereço deve ser especificado na tela *Tools > Plugins*, na aba *Settings* do Gephi, clicando no botão “Add”:

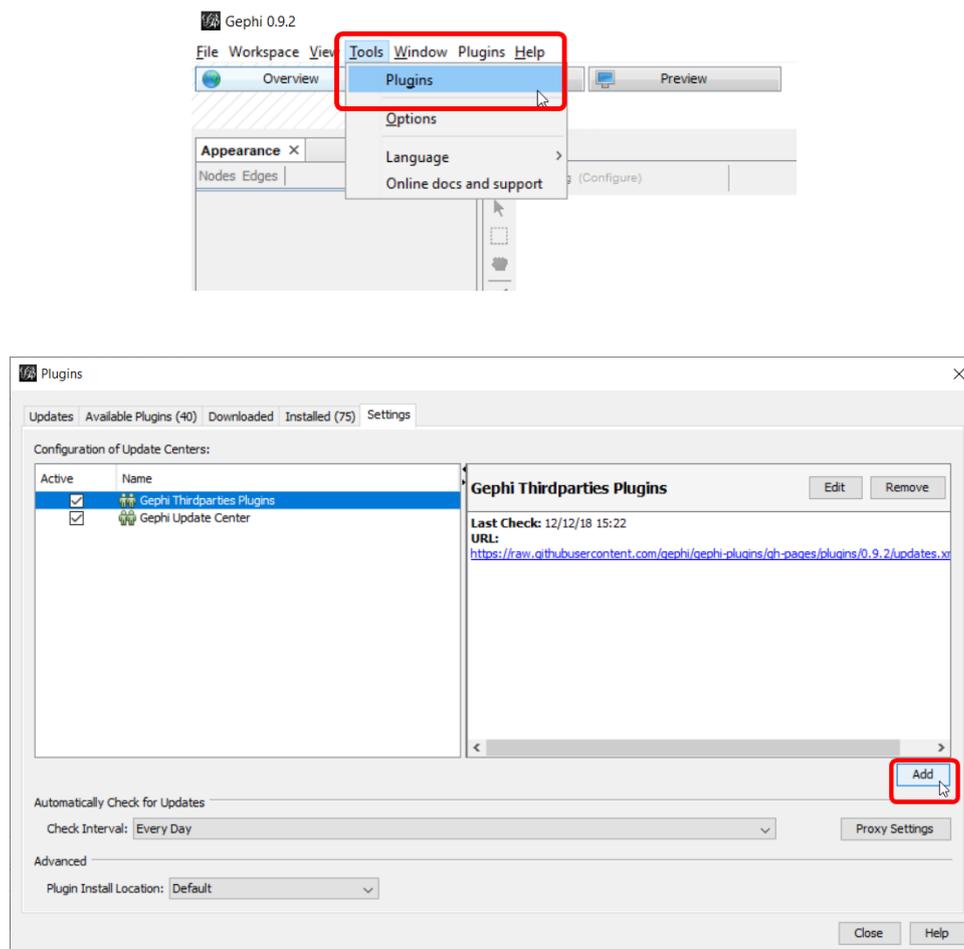


Figura 2.1: Configuração dos centrais de atualização Gephi

No diálogo que aparece, os seguintes dados devem ser especificados:

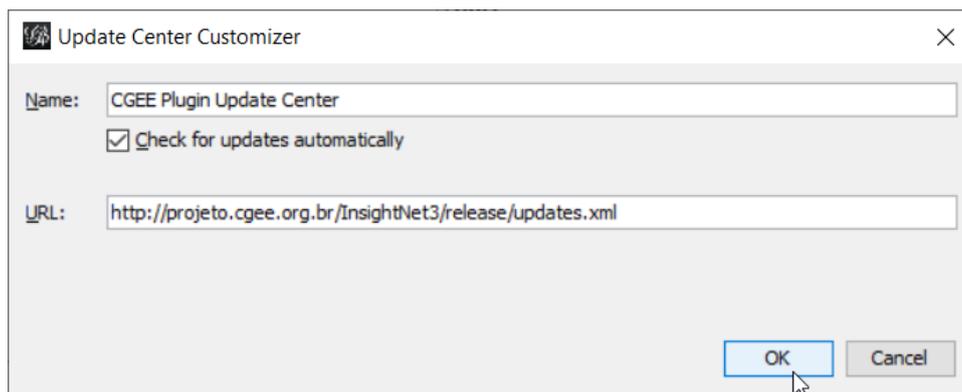


Figura 2.2: Configuração da central de atualizações do CGEE Insight Net

Com isso, a central de atualização aparecerá na lista de configuração. Caso seja necessário, o usuário deve configurar o *proxy* da conexão com a internet (botão *Proxy Settings*).

2.4 Instalação do *CGEE Insight Net*

Com a central de atualização configurada, o usuário pode selecionar e instalar o *CGEE Insight Net*, também pela tela de *plug-ins* (*Tool > Plugins*). Clicando na aba “*Available Plugins*”, a ferramenta mostra uma lista de todos os *plug-ins* disponíveis, entre eles o “*CGEE Analysis plugin*”, que deve ser selecionado pelo usuário, seguido por um clique no botão “*Install*”:

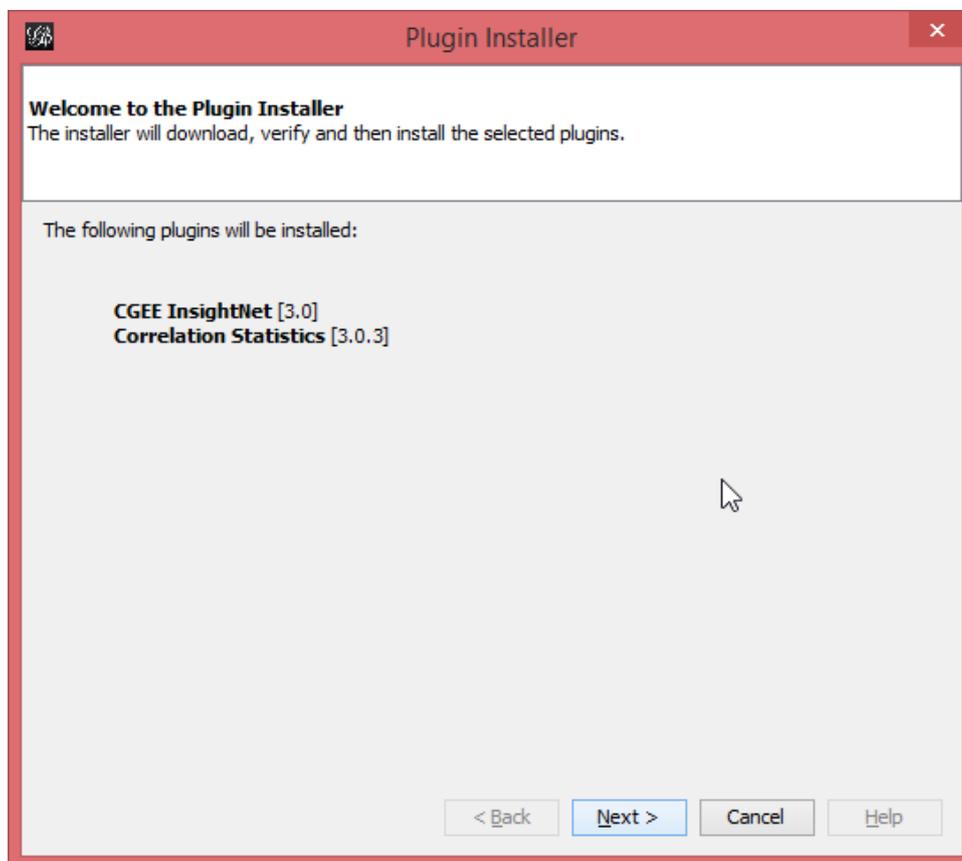
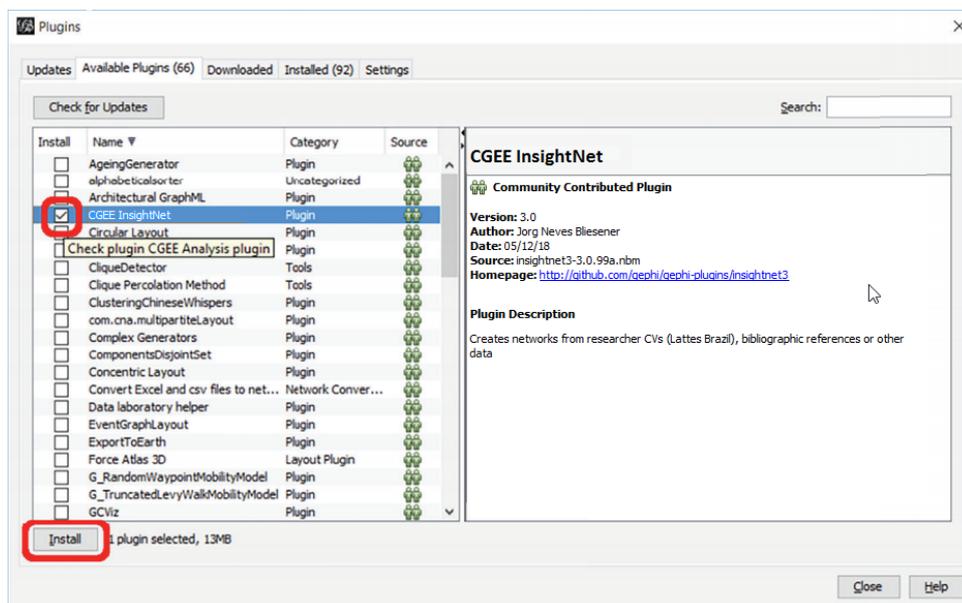


Figura 2.3: Instalação do CGEE Insight Net

Ao chegar à tela de licença, o usuário deve aceitar o texto da(s) licenças exibida(s) para concluir a instalação² :

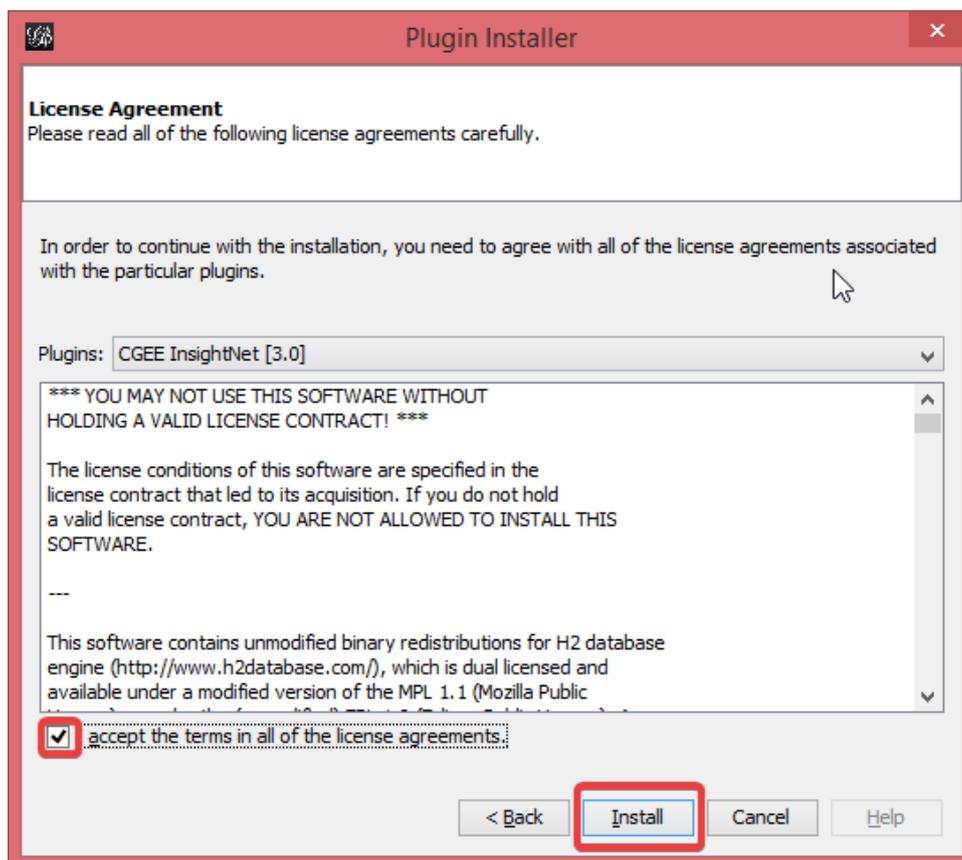


Figura 2.4: Concordância com a(s) licenças(s) do produto

Depois disso, o *CGEE Insight Net* é baixado pela internet e um aviso de falta de assinatura digital é exibido, e pode ser ignorado:

² As licenças exibidas incluem as licenças das bibliotecas usadas no produto.

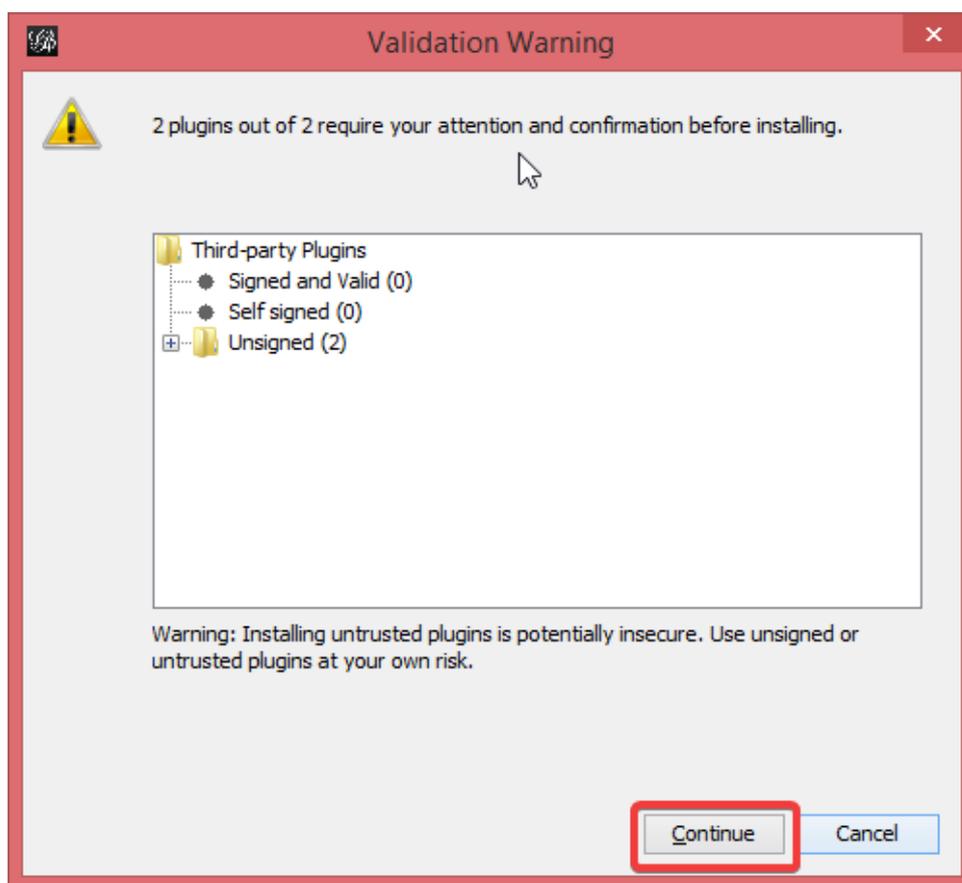
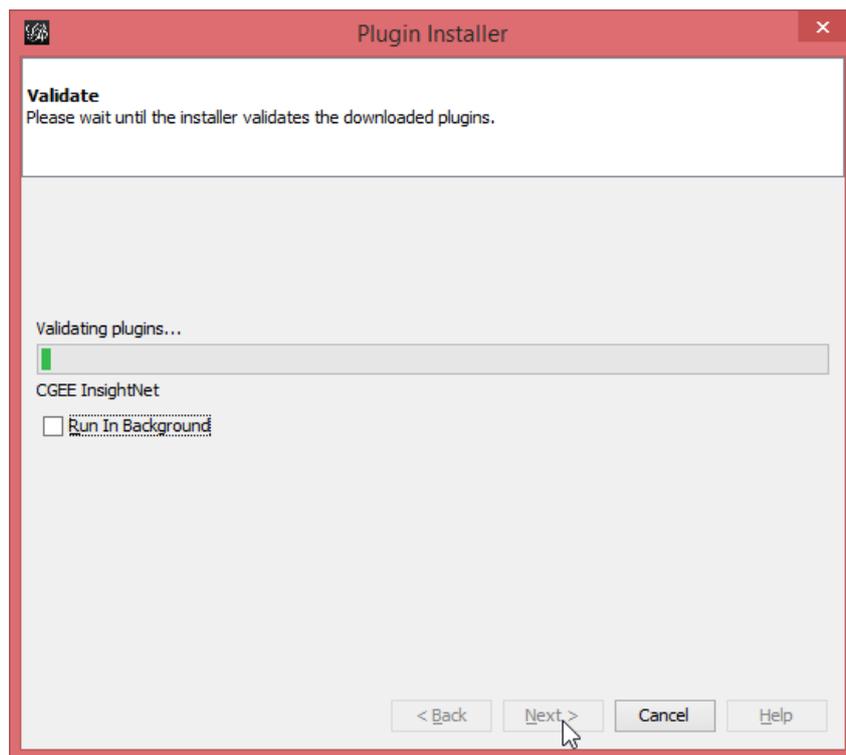


Figura 2.5: Download e aviso de falta de assinatura

Em seguida, o Gephi deve ser reiniciado e a instalação do CGEE Insight Net é concluída:

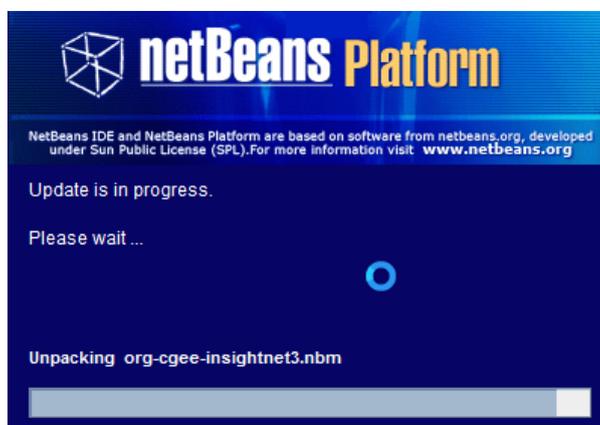
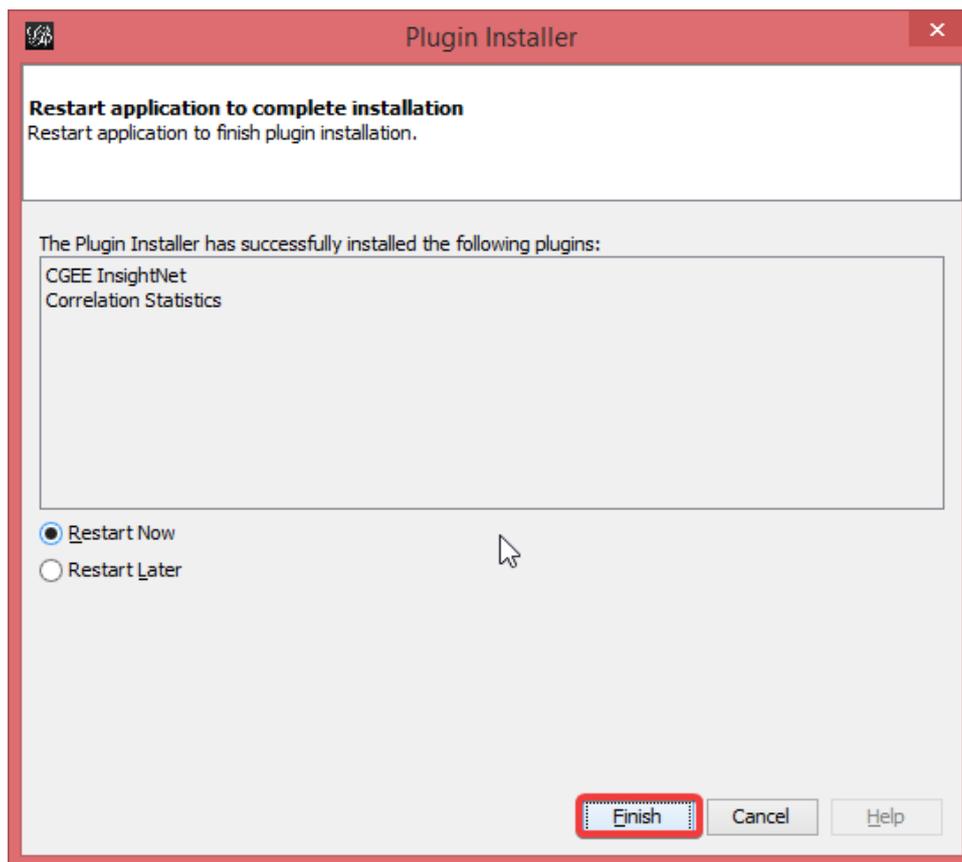


Figura 2.6: Instalação do CGEE Insight Net

Depois de reiniciar, o *GEE Insight Net* tenta alterar o arquivo `gephi.conf` que consta no subdiretório `etc` da instalação do Gephi³. Caso essa tentativa tenha êxito, a seguinte mensagem é exibida e o sistema é reiniciado outra vez:

³ Conforme relatado na Seção 2.2, recomenda-se que o administrador do Sistema torne esse arquivo gravável para o usuário



Figura 2.7: Mensagem de alteração bem-sucedida do arquivo `gephi.conf`

Caso o arquivo `gephi.conf` não possa ser alterado, a seguinte mensagem é exibida:

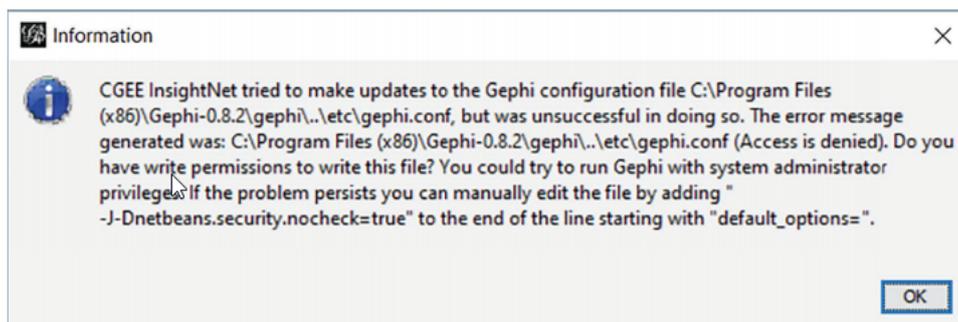
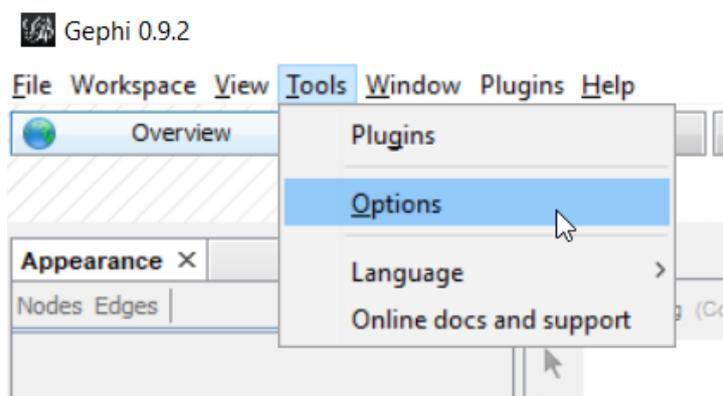


Figura 2.8: Mensagem de alteração malsucedida do arquivo `gephi.conf`

Nesse caso, a gestão e o uso de licenças adicionais (ver [Seção 3.6](#)) fica indisponível e uma das alterações sugeridas no diálogo deve ser realizada por um administrador do sistema. A solução de menor complexidade é a concessão dos privilégios de gravação do arquivo `gephi.conf` para o usuário final, conforme descrito na [Seção 2.2](#).

Depois desse procedimento, o *CGEE Insight Net* aparece na tela de opções do Gephi. Para usar efetivamente, as licenças correspondentes aos módulos do contratados ainda devem ser instaladas conforme descrito na [Seção 3.6](#).



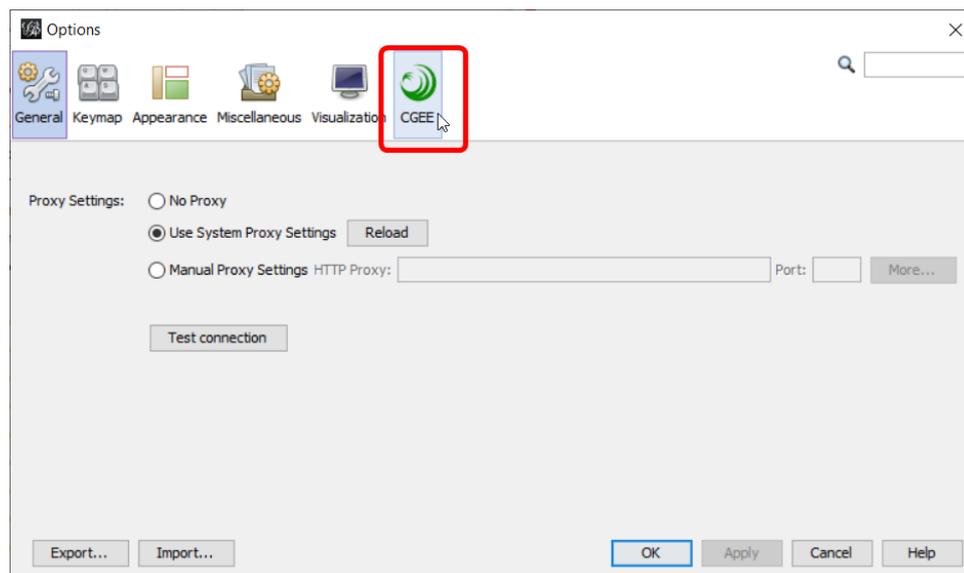


Figura 2.9: CGEE Insight net instalado

2.5 Atualização do *CGEE Insight Net*

O *CGEE Insight Net* é atualizado automaticamente, de acordo com a configuração de atualizações na configuração dos *plug-ins*:

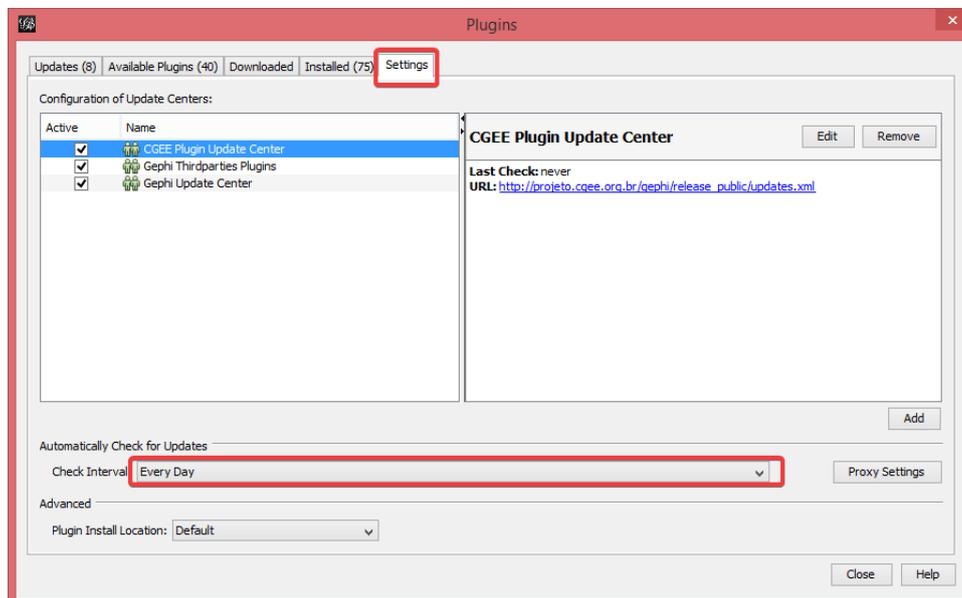


Figura 2.10: Configuração da atualização automática

Caso haja atualizações do *CGEE Insight Net*, um aviso aparecerá no Gephi:

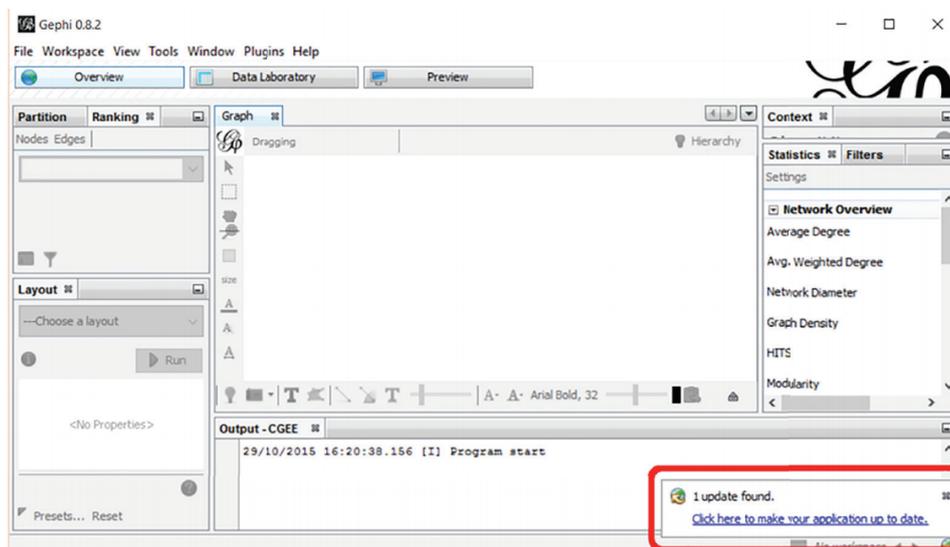


Figura 2.11: Aviso de atualização

Clicando na mensagem de atualização, o Gephi exibe a lista de atualizações disponíveis e, clicando em *Next*, inicia o processo de atualização:

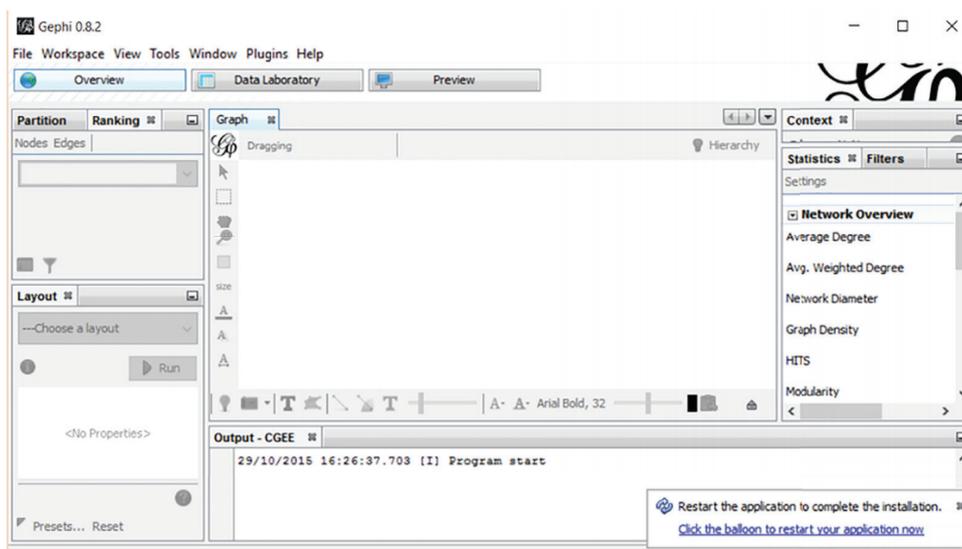
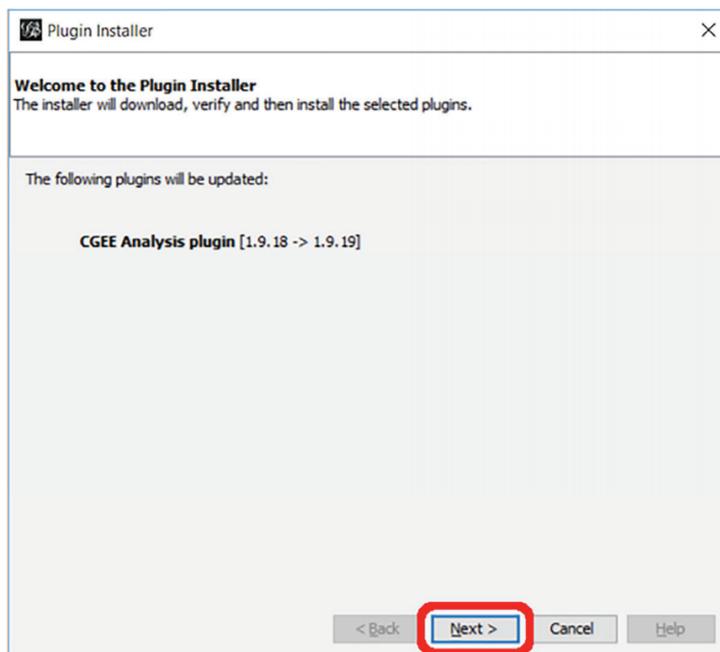


Figura 2.12: Aviso de atualização

Clicando no aviso, o Gephi conclui a atualização e reinicia o programa.

Configuração do CGEE Insight Net

Antes de usar o *CGEE Insight Net*, este deve ser configurado de acordo com os requisitos do usuário, clicando em *Tools > Options* e selecionando a tela *CGEE*.

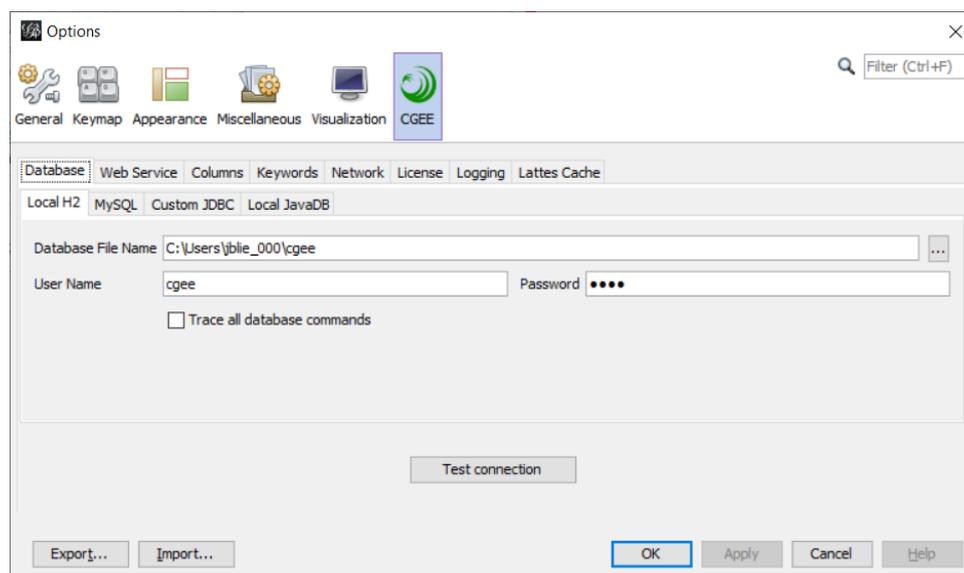


Figura 3.1: Opções de configuração

No caso mais simples de uso local do *CGEE Insight Net*, o usuário apenas confirma as informações pré-configuradas, sem necessidade de alterar qualquer um dos valores. Para atender casos específicos (banco de dados centralizados, acompanhamento detalhado ou depuração de problemas, redução da carga do computador, etc.), os itens de configuração serão explicados em seguida.

3.1 Configuração do banco de dados

A primeira aba da configuração refere-se ao tipo de banco de dados e aos parâmetros de configuração da conexão.

Para o uso local do *CGEE Insight Net*, recomenda-se o uso do banco “*Local H2*”. O nome do arquivo pode ser selecionado pelo usuário. Para conectar, deve ser especificado um nome de usuário e uma senha, sendo que os valores pré-configurados devem atender a maioria dos casos. A customização do usuário e da senha permite certo nível de proteção de acesso, mas não envolve nenhum tipo de criptografia no nível binário da base.

O uso do *CGEE Insight Net* em ambientes centralizados geralmente envolve um banco de dados em outro servidor. No caso do banco “*MySQL*”, o caminho do módulo de conexão (o *Driver JDBC*) e os parâmetros de conexão (servidor, porta, usuário, senha e nome do banco de dados) precisam ser definidos na aba correspondente e serão fornecidos pelo administrador do servidor.

Outros bancos de dados podem ser configurados na aba “*Custom JDBC*”, o que, geralmente, ainda requer a customização dos comandos SQL envolvidos. Caso necessário, sugere-se o envolvimento de especialistas do departamento de TI.

A aba “*Local JavaDB*” permite o uso de um banco legado, implementado em Java, caso os métodos anteriores não produzam os efeitos desejados. Ressalta-se que o banco JavaDB possui desempenho inferior em relação ao banco H2 sugerido na configuração padrão.

Depois da configuração dos parâmetros do banco de dados, a conexão deve ser verificada, clicando no botão “*Test connection*”.

3.2 Configuração do usuário para acessar o banco de dados de Currículos Lattes do CGEE

O acesso ao banco de dados de Currículos Lattes do CGEE é restrito por usuário e senha. A aba “*Web service*” permite a configuração do nome do usuário e da respectiva senha anteriormente criados no *web service* do CGEE:

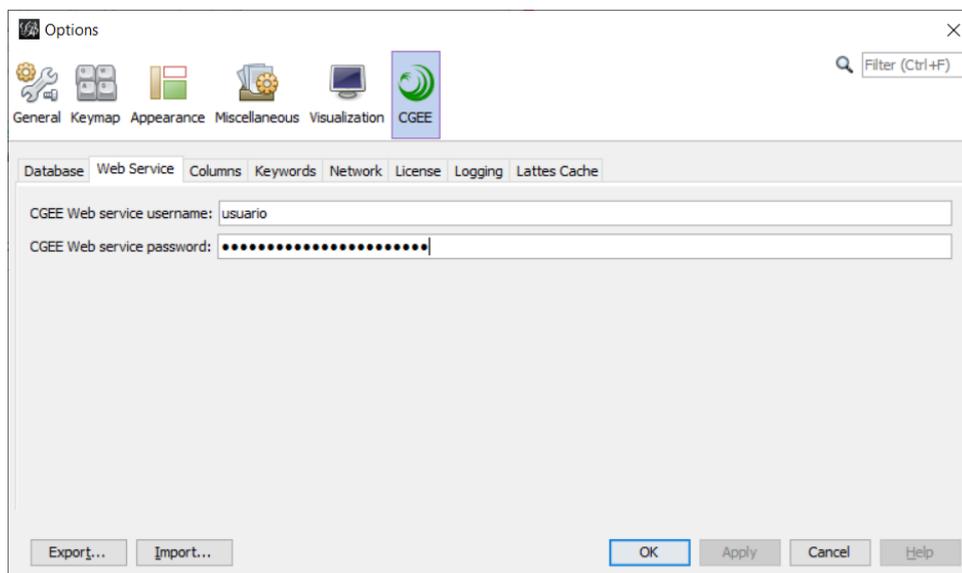


Figura 3.2: Configuração do usuário e da senha para acessar o banco de dados de Currículos Lattes do CGEE

3.3 Configuração das colunas exibidas

As diversas fontes de dados trazem uma grande quantidade de informações, cuja exibição completa pode tornar a operação do programa ineficiente. A relevância dessas informações depende do projeto específico.

Para não sobrecarregar a tela e permitir a exibição dos dados mais relevantes, o *CGEE Insight Net* permite a seleção de atributos dos pesquisadores e das referências bibliográficas e das colunas correspondentes no laboratório de dados. A aba “Columns” configura, para cada tipo de dados, as colunas que serão exibidas por padrão, sem a necessidade de personalizações do usuário para cada análise:

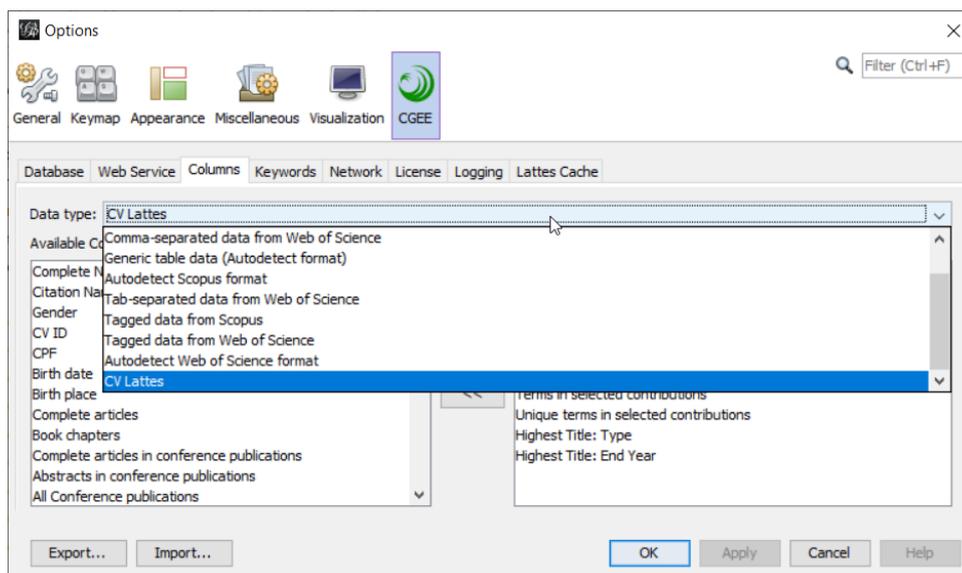


Figura 3.3: Seleção do tipo de dados para configurar as colunas que serão exibidas por padrão

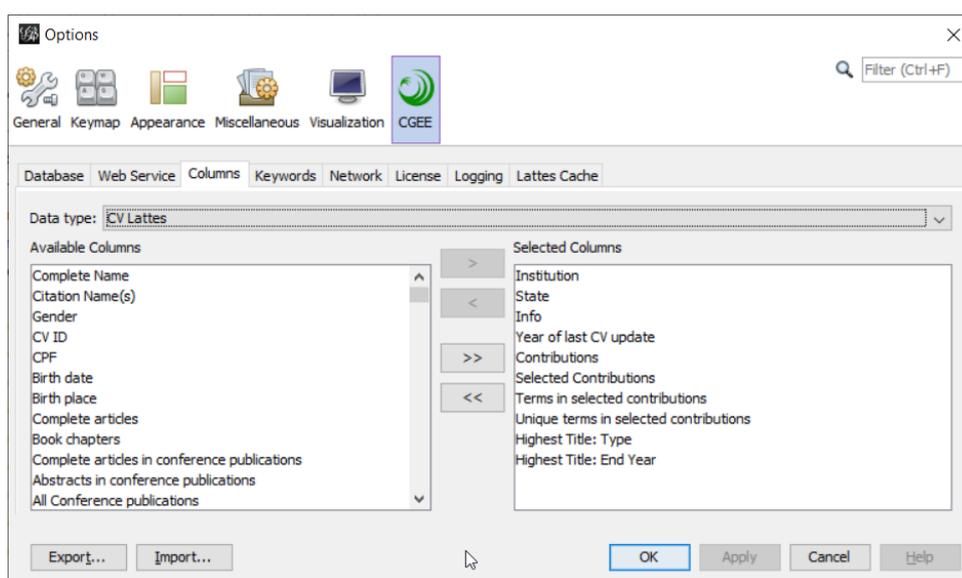


Figura 3.4: Configuração das colunas dos currículos Lattes que serão exibidas por padrão

As colunas na lista da direita são aquelas que aparecem no laboratório de dados. As colunas na lista da esquerda não serão exibidas. O usuário pode clicar em uma ou mais colunas em ambas as janelas

segurando a tecla *Ctrl* ou *Shift* e clicar nos botões  ou  para levar essas colunas para a outra lista. Os botões  e  levam todos os elementos de uma lista para outra.

3.4 Exibição da relevância das palavras-chave

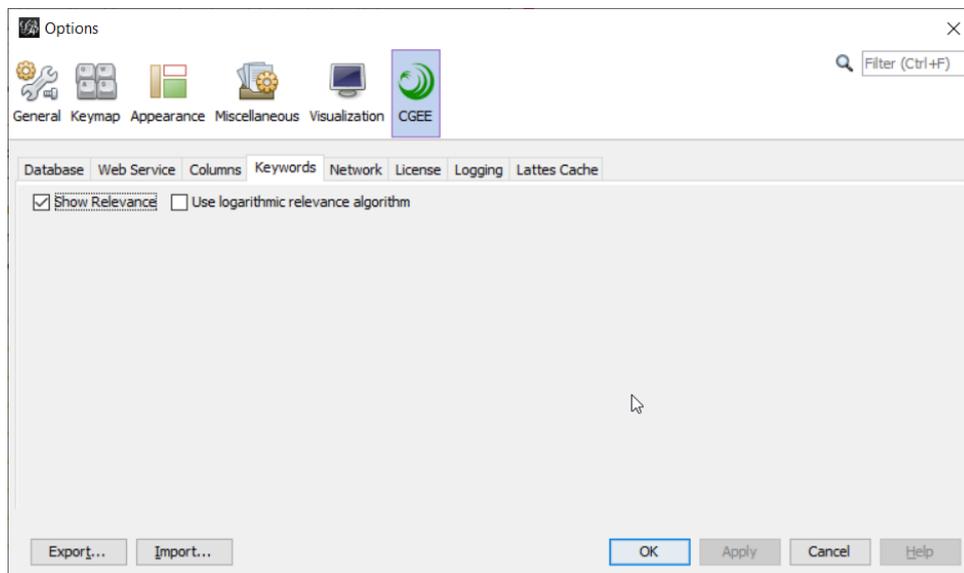


Figura 3.5: Configuração da exibição das palavras-chave

O *CGEE Insight Net* permite a exibição opcional das relevâncias das palavras-chave. Geralmente, a janela de palavras-chave (ver seção 3.4.5) mostra, para cada palavra-chave, a sua frequência dentro do conjunto de dados selecionados (pesquisadores ou *clusters*). A opção “*Show Relevance*” permite o uso experimental de um algoritmo que calcula a relevância das palavras-chave a partir do algoritmo “*tf. idf*”.

Caso a opção “*Show relevance*” for selecionada, o algoritmo pode usar pesos lineares ou pesos logarítmicos para a frequências das palavras, dependendo da configuração da opção “*Use logarithmic relevance algorithm*”.

3.5 Parâmetros da pesquisa por similaridade

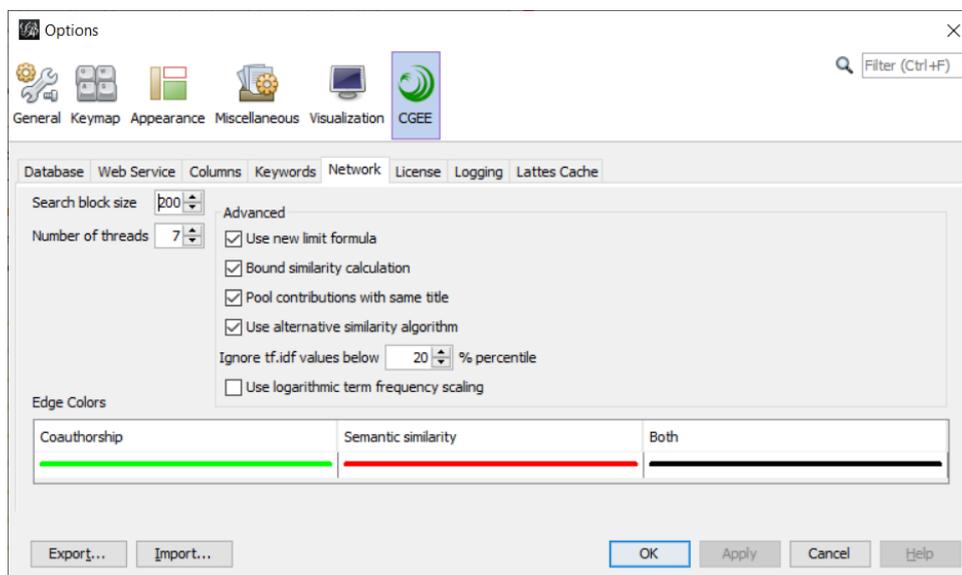


Figura 3.6: Configuração do cálculo e da exibição das redes

Os parâmetros “*Search block size*”, “*Number of threads*” e os cinco parâmetros avançados (“*Advanced*”) permitem configurar detalhes do processo de pesquisa por coautoria.

“*Search block size*” determina quantas contribuições (artigos, capítulos de livros e trabalhos em eventos) serão agrupados em um bloco de pesquisa, que é alocado a um núcleo de processador do computador. Deve ser considerado que cada bloco gera certo “*overhead*”, um processamento adicional. Assim, sugere-se minimizar a quantidade de blocos. Por outro lado, em computadores com vários núcleos, o processamento dos blocos pode ser paralelizado, o que favorece a escolha de uma quantidade maior de blocos. O valor padrão de 200 representa um equilíbrio entre os dois objetivos, mas pode ser alterado pelo usuário.

O parâmetro “*Number of threads*” indica quantos blocos de processamento serão analisados em paralelo. O valor padrão varia de computador para computador. Esse valor é igual à quantidade de processadores (ou núcleos) disponíveis na linguagem Java exceto um, para ainda disponibilizar capacidade de processamento para as tarefas de visualização. Caso necessário, esse valor pode ser reduzido para diminuir a carga do computador.

Os parâmetros avançados (“*Advanced*”) configuram otimizações dos algoritmos de cálculo de similaridade. Recomenda-se deixar todos eles ligados para obter o melhor desempenho:

- “*Use new limit formula*”: Otimização do cálculo da distância *Levenshtein* máxima a partir da similaridade específico cada pelo usuário
- “*Bound similarity calculation*”: Otimização do critério de conclusão de busca.
- “*Pool contributions with same title*”: Contribuições com o mesmo título são agrupadas. Assim, a busca por similaridade precisa ser executada apenas uma única vez para todos os títulos iguais.
- “*Use alternative similarity algorithm*”: Uso de um algoritmo otimizado de cálculo de similaridade.
- “*Ignore tf.idf values below x% percentile*”: Os termos cujo valor de relevância (tf.idf) são abaixo do percentil configurado serão eliminados da busca de similaridade. Se esse valor for diferente

de zero, um aviso é exibido no diálogo de busca.

- “*Use logarithmic term frequency scaling*”: Nos algoritmos de busca por similaridade semântica, a relevância de um termo é calculado a partir do algoritmo “tf.idf”, que considera como um dos seus dois elementos a frequência com que um termo ocorre em um documento. Esta configuração permite selecionar se a frequência será utilizada de forma original ou se o logaritmo dessa frequência será usado, que pode ser mais adequado para documentos com tamanhos diferentes. Entretanto, o uso dessa opção deve ser avaliado caso por caso.

3.5.1 Coloração das arestas do grafo

A tabela na parte inferior do diálogo permite a configuração das cores das arestas que aparecem no grafo. As três colunas “*Coauthorship*”, “*Contextual similarity*” e “*Both*” mostram as cores em que são exibidas as arestas que possuem apenas coautorias, apenas similaridade contextual ou ambas. Clicando no campo que mostra a linha, uma tela de seleção de cores é exibida:

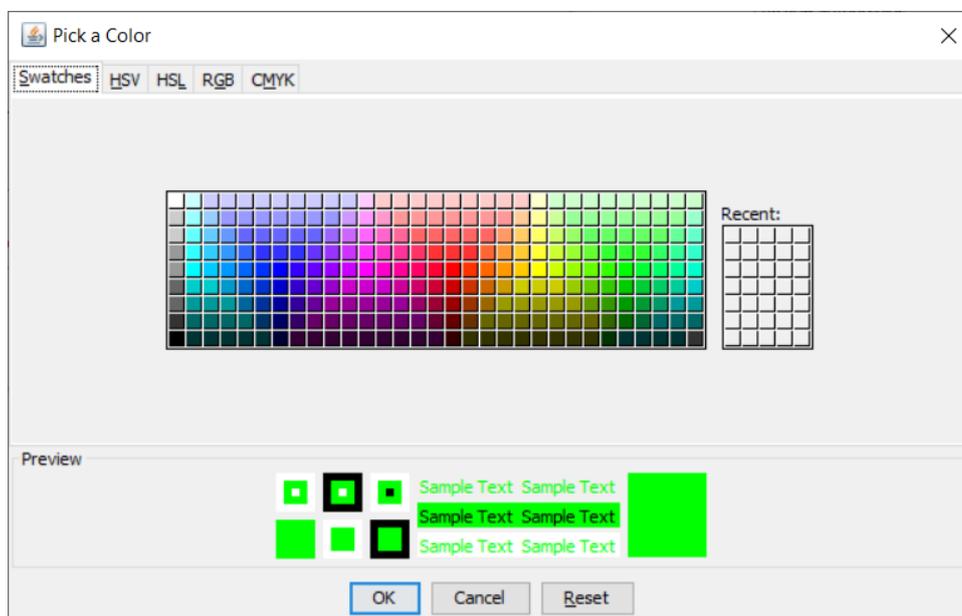


Figura 3.7: Seleção de cores

3.6 Licenças

O acesso às funcionalidades do *CGEE Insight Net* é restrito por licenças opcionais. Cada módulo possui uma licença específica:

- Currículos Lattes (ver Seção 5)
- Referências bibliográficas BibTeX(ver Seção 6)
- Referências bibliográficas genéricas (var Seção 7)

Essas licenças podem ser instaladas na aba “*License*”:

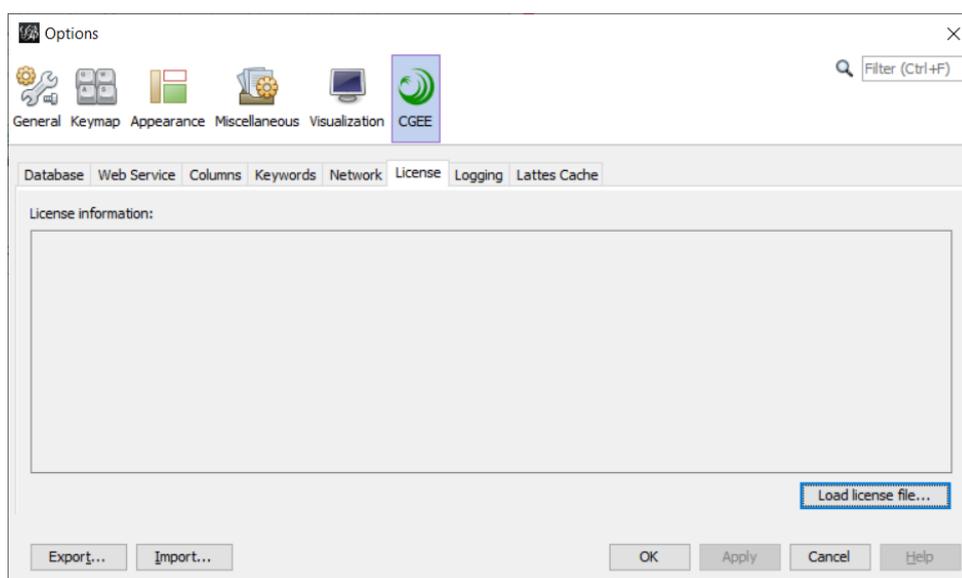


Figura 3.8: Instalação de licenças

As licenças são disponibilizadas em forma de arquivos criptografados que podem ser carregados com o botão “*Load license file*”. Depois da validação da licença, a disponibilidade é exibida no diálogo:

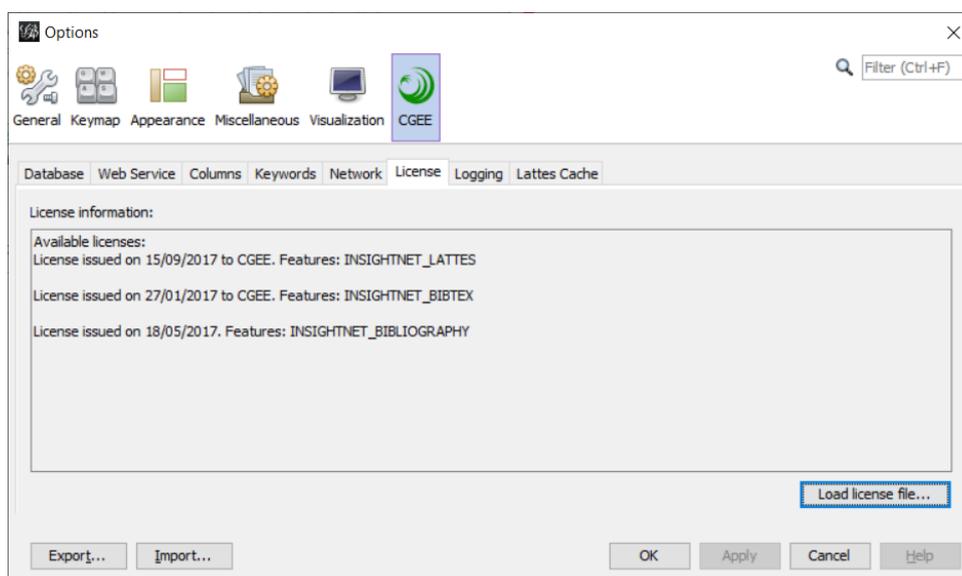


Figura 3.9: Indicação das licenças disponíveis

3.7 Protocolos de execução

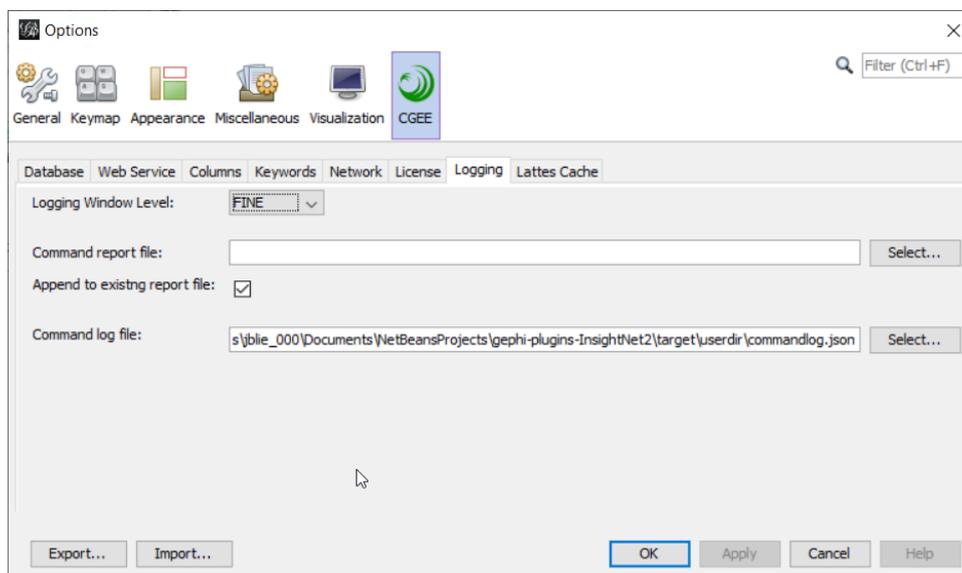


Figura 3.10: Configuração do protocolo de execução

Durante a sua execução, o *CGEE Insight Net* gera informações e avisos que podem ser exibidos na tela e que permitem um acompanhamento e mesmo uma depuração em caso de problemas.

O grau de detalhamento dessas mensagens pode ser configurado com o item “*Logging level*”. A configuração inicial (*INFO*) gera registros que permitem o acompanhamento da execução no nível de um usuário com pouca experiência. Os níveis *WARNING* e *SEVERE* mostram apenas erros e avisos mais graves e os níveis *FINE*, *FINER* e *FINEST* geram registros de depuração que, geralmente, não são relevantes para os usuários, mas podem ser úteis para a análise de eventuais erros de carga ou de processamento.

Ainda existem dois relatórios de execução, que protocolam as atividades do *plugin*. Enquanto o *Command report file* demonstra as informações da forma legível para o usuário, o *Command log file* é um protocolo mais adequado para o processamento automático.

3.8 Memória *cache* de Currículos Lattes

O módulo “Currículo Lattes” - caso for habilitado por licença - mantém uma memória local (“*cache*”) para agilizar a carga de currículos recentemente usados. Essa memória é limitada em termos da quantidade de currículos, do tamanho total dos currículos e da idade máxima. Esses limites podem ser configurados nesta aba:

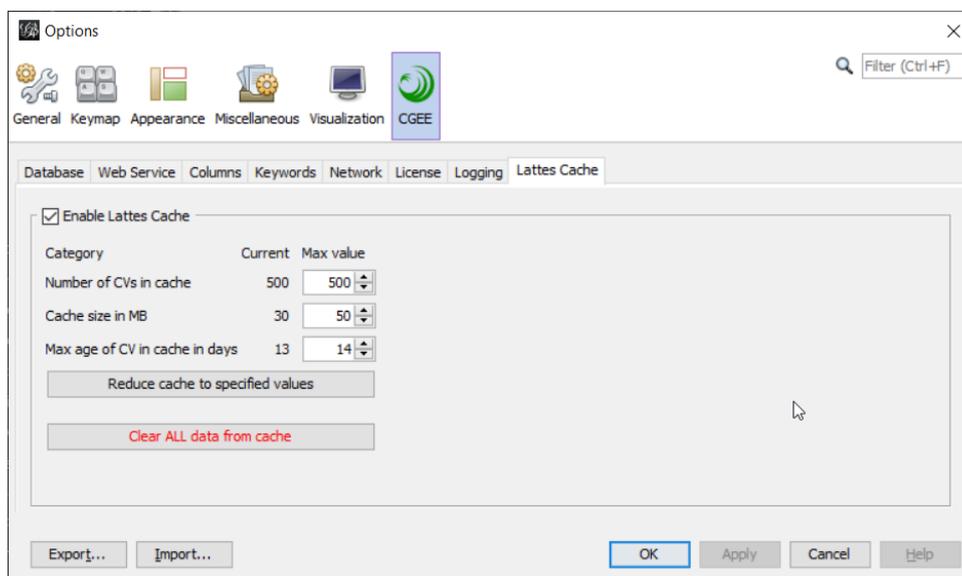


Figura 3.11: Configuração da memória *cache* do módulo Lattes

Nesta aba, a memória *cache* ainda pode ser desabilitada completamente. Ainda podem ser eliminados da memória *cache* todos os currículos ou apenas aqueles que excedem os limites configurados (caso estes valores foram ajustados).

Conceitos gerais do uso do *CGEE Insight Net*

4.1 Fluxo de trabalho

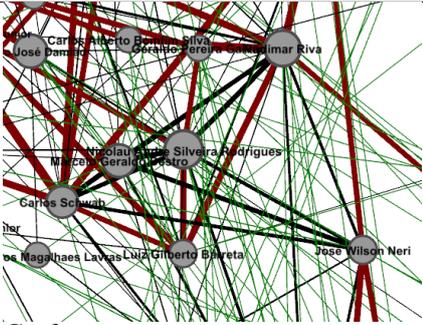
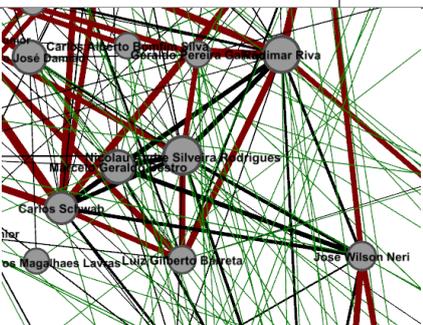
Ao utilizar o *CGEE insight Net*, deve ser considerado que ele trabalha com três repositórios de informações:

- Os dados de entrada, dependendo do módulo de processamento: * Currículos *Lattes* em formato XML * Referências bibliográficas dos sistemas serviços Web of Science® e Scopus® em formato BibTeX * Referências bibliográficas genéricas em formato textual ou planilha Excel®
- O banco de dados contendo todas as informações, incluindo detalhes sobre as contribuições, palavras-chave e graus de similaridade.
- O grafo composto de nós e arestas, ambos com certos atributos.

As informações do grafo podem ser visualizadas e manipuladas diretamente na ferramenta *Gephi*, através das funções de manipulação de nós e arestas. O acesso ao banco de dados é realizado exclusivamente pelo *CGEE Insight Net*. Essa diferença é importante, pois certas operações feitas no *Gephi* podem não ser registradas no *CGEE Insight Net* e vice-versa.

A tabela em seguida mostra o fluxo típico de informações do processamento no caso dos Currículos *Lattes* que gera uma rede de pesquisadores pelo *CGEE Insight Net*. Para os casos de referência bibliográfica, o processo é similar, mas a rede gerada representa contribuições bibliográficas, tais como artigos ou trabalhos em eventos.

Tabela 4.1: Fluxo de informações do *plugin*

Passo	Origem	Operação	Destino
1	<pre><?xml version="1.0" encoding="UTF-8" <CURRICULO-VITAE SISTEMA-ORIGEM DATA-ATUALIZACAO="10082011" HOR NUMERO-IDENTIFICADOR="854972258 xmlns:lattes="http://www.cnpq.br <DADOS-GERAIS NOME-COMPLETO=" NOME-EM-CITACOES-BIBLIOGRAFIC NACIONALIDADE="B" CPF="147015 UF-NASCIMENTO="MG" CIDADE-NAS</pre> <p>Currículo Lattes XML</p>	<p>→</p> <p>Importação</p>	<ul style="list-style-type: none"> • Pesquisadores • Artigos • Capítulos Livros • Trabalhos em Eventos • Palavras-chave <p>Banco de dados</p>
2	<ul style="list-style-type: none"> • Pesquisadores • Artigos • Capítulos Livros • Trabalhos em Eventos • Palavras-chave <p>Banco de dados</p>	<p>→</p> <p>Seleção de contribuições</p>	<ul style="list-style-type: none"> • Pesquisadores • Contribuições selecionadas: <ul style="list-style-type: none"> ○ Artigos ○ Capítulos Livros ○ Trabalhos em Eventos • Palavras-chave <p>Banco de dados</p>
3	<ul style="list-style-type: none"> • Pesquisadores • Contribuições selecionadas: <ul style="list-style-type: none"> ○ Artigos ○ Capítulos Livros ○ Trabalhos em Eventos • Palavras-chave <p>Banco de dados</p>	<p>→</p> <p>Pesquisa de similaridade</p>	<ul style="list-style-type: none"> • Pesquisadores • Contribuições selecionadas: <ul style="list-style-type: none"> ○ Artigos ○ Capítulos Livros ○ Trabalhos em Eventos • Palavras-chave • Similaridades • Colaborações <p>Banco de dados</p>
4	<ul style="list-style-type: none"> • Pesquisadores • Contribuições selecionadas: <ul style="list-style-type: none"> ○ Artigos ○ Capítulos Livros ○ Trabalhos em Eventos • Palavras-chave • Similaridades • Colaborações <p>Banco de dados</p>	<p>→</p> <p>Visualização</p>	 <p>Grafo</p>
5		<ul style="list-style-type: none"> • Pesquisadores • Contribuições selecionadas: <ul style="list-style-type: none"> ○ Artigos ○ Capítulos Livros ○ Trabalhos em Eventos • Palavras-chave • Similaridades • Colaborações <p>Banco de dados</p>	<p>Análise do grafo e pesquisas de palavra-chave</p>

Esta sequência demonstra que o grafo é gerado apenas no último passo de visualização. É relevante mencionar que manipulações no grafo, que são operações do *Gephi*, não se refletem dentro do banco de dados. Por outro lado, podem ser geradas várias visualizações do mesmo banco de dados, permitindo

análises visuais diferentes a partir do mesmo banco de dados. A separação do grafo do banco de dados também permite o compartilhamento de dados no nível de rede (nós e arestas) sem divulgar dados potencialmente sigilosos que constam nos currículos importados.

As seções a seguir detalham os passos descritos.

Uso do *CGEE Insight Net* para analisar Currículos Lattes

O *CGEE Insight Net* permite a criação de redes de pesquisadores por co-autorias e similaridade semântica das publicações.

Para habilitar essa funcionalidade do *CGEE Insight Net*, a licença `INSIGHTNET_LATTES` deve ser instalada, conforme descrito na [Seção 3.6](#). Essa licença é exibida assim no diálogo *Tools > Options > CGEE > License*:

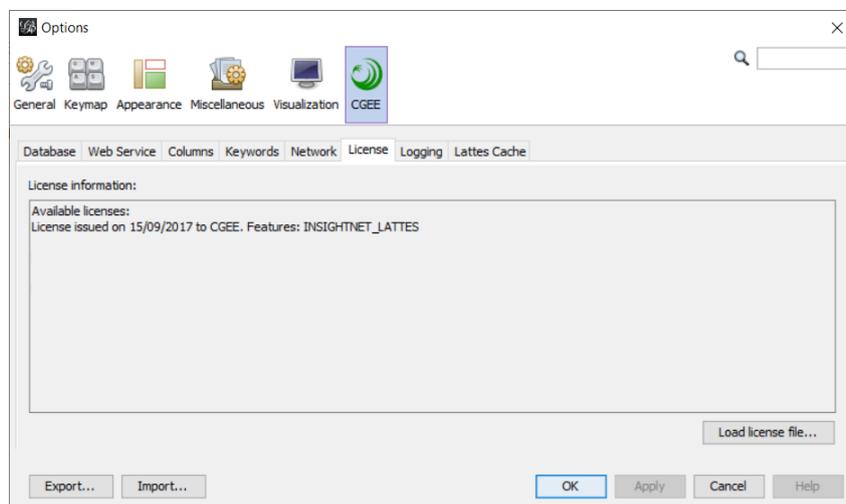


Figura 5.1: Licença requerida para o módulo de redes de Currículos Lattes

Caso a licença esteja habilitada, aparece o sub-menu “*CGEE Insight Net Lattes*”:

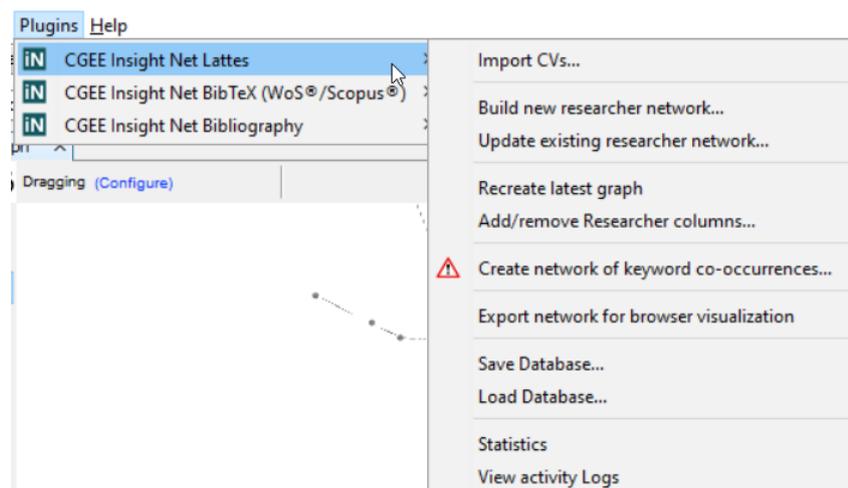


Figura 5.2: Sub-menu CGEE Insight Net Lattes

5.1 Importação dos Currículos Lattes

Para processar as informações dos Currículos Lattes em formato XML, estes devem ser importados no banco de dados a partir da função *Plugins > CGEE insight Net Lattes > Import*, que exibe o seguinte diálogo, cujo formato depende da aba selecionada na parte superior:

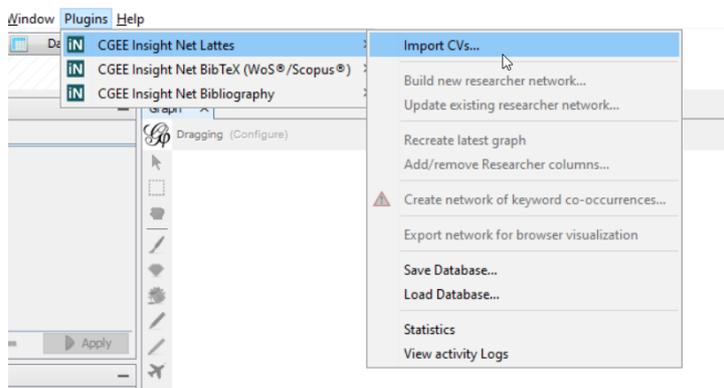


Figura 5.3: Menu de importação de Currículos

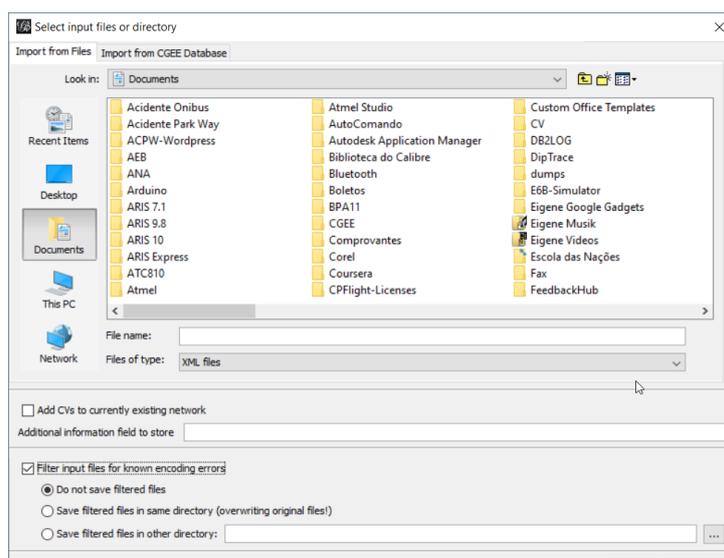


Figura 5.4: Diálogo de importação de Currículos em Arquivos

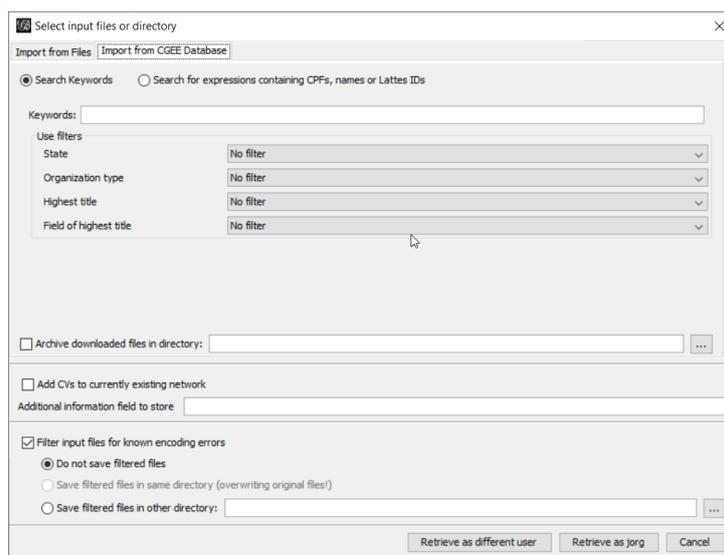


Figura 5.5: Diálogo de importação de Currículos do Banco de Dados

Esse diálogo permite a importação de currículos Lattes em arquivos XML ou por conexão direta com o banco de dados do CGEE.

5.1.1 Importação de arquivos XML

Selecionando a aba “*Import from files*”, o usuário pode importar currículos gravados no formato XML no computador local, em algum diretório compartilhado em rede ou mesmo um dispositivo móvel de armazenamento. O escopo da importação depende da seleção dos arquivos na lista apresentada:

- Clicando em um arquivo XML, este será importado;
- Vários arquivos XML podem ser selecionados com “*Shift-Clique*” ou “*Ctrl-Clique*”, de acordo com os padrões de uso do sistema operacional;
- O usuário também pode selecionar um ou mais diretórios. Nesse caso, todos os arquivos XML nesses(s) diretório(s) serão importados.

Os arquivos importados devem seguir o padrão XML dos Currículos Lattes do CNPq¹.

5.1.2 Acessando o banco de dados do CGEE

Selecionando a aba “*Import from CGEE database*”, o software pode acessar diretamente o banco de dados de currículos do CGEE. Neste caso, o diálogo oferece duas funcionalidades para recuperar currículos Lattes. Essas funcionalidades serão descritas em seguida.

Recuperação por palavras-chave

A opção “*Search Keywords*” permite que o usuário especifique palavras-chave que serão aplicadas na pesquisa de especialistas por competência, seguindo padrões de uso do Portal da Inovação². Adicionalmente, os currículos obtidos podem ser filtrados por “Unidade da Federação”, “Tipo de organização”, “Maior titulação” e “Área da maior titulação”:

¹ Verificar em <http://lattes.cnpq.br/web/plataforma-lattes/extracao-de-dados>

² Verificar em <http://www.portalinovacao.mcti.gov.br/pi/#/pi>

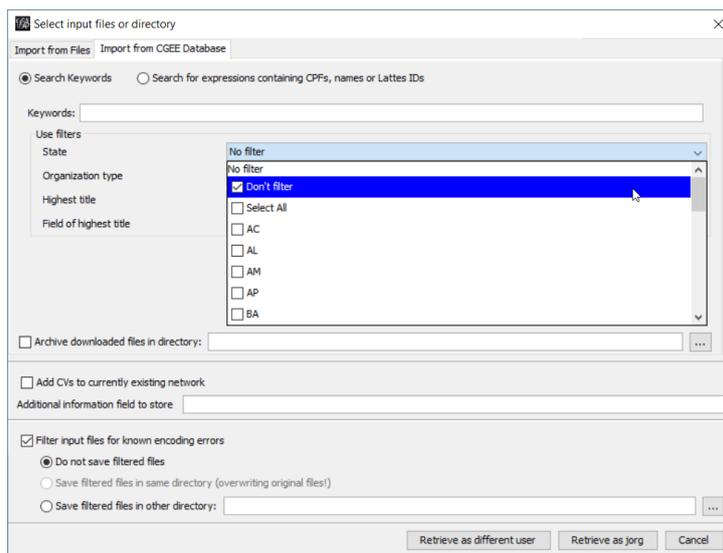


Figura 5.6: Importação de dados do Banco de dados do CGEE

Destaca-se a diferença entre as primeiras duas opções de cada filtro:

- “*Don't filter*” não aplica nenhum filtro nos currículos
- “*Select all*” elimina aqueles currículos cujo critério não consta na lista de valores válidos

Como exemplo, é possível citar o pesquisador estrangeiro que não preencheu o campo “UF” no seu currículo. Se o usuário escolher “*Select all*”, este currículo não fará parte da importação, pois “*Select all*” considera apenas currículos que possuem um dos valores definidos na lista de estados. Para incluir o pesquisador estrangeiro, o usuário teria que selecionar “*Don't filter*”.

Durante a digitação da palavra-chave, uma busca prévia é iniciada no servidor e a quantidade de currículos que atendem aos critérios selecionados é exibida no campo “*No. of found CVs*”. Para isso, é necessário que o usuário tenha digitado, no mínimo, três letras no campo “*Keywords*”, seguido por um intervalo sem digitação de, no mínimo, dois segundos.

As palavras-chave digitadas são automaticamente copiadas para o campo “*Additional information field to store*” e serão exibidas no campo “*info*” do Laboratório de dados do Gephi.

Pesquisa por filtro

A opção “*Search for expressions containing CPFs, names or Lattes IDs*” permite a especificação de filtros usados pelo software “*WebExtractor*” do CGEE:

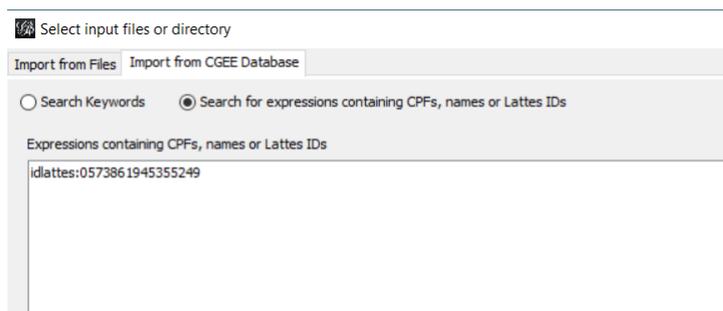


Figura 5.7: Recuperação de currículos por expressão de busca

A expressão do filtro deve usar o seguinte formato:

```
<critério de busca>:<item1>,<item2>,...
```

O critério de busca determina o campo de informação usado para realizar as buscas e pode ser um dos seguintes valores:

- CPF: realiza uma busca pelos CPFs dos pesquisadores
- Nome: realiza uma busca pelos nomes dos pesquisadores, sem acentos e caracteres especiais
- Idlattes: realiza uma busca pelo número identificador do currículo na base Lattes

Seguem alguns exemplos de expressões de filtros:

- Extração dos currículos dos pesquisadores que possuem os CPFs 123.456.789-12 ou 987.654.321-00: `cpf:12345678912,98765432100`
- Extração do currículo do pesquisador “Pesquisador 1”: `nome: Pesquisador 1`
- Extração dos currículos dos pesquisadores “Pesquisador 1”, “Pesquisador 2” e “Pesquisador 3”: `nome:Pesquisador 1,Pesquisador 2,Pesquisador 3`
- Extração dos currículos Lattes com os identificadores “0000000000000000” e “1111222233334444”: `idlattes:0000000000000000,1111222233334444`

Arquivamento dos dados originais

Os currículos Lattes recuperados podem ser arquivados, junto com um texto descritivo da operação. Essa funcionalidade, importante para evitar que atualizações do Lattes inviabilizem a reprodução e validação de uma análise realizada, é ativada com a opção “*Archive downloaded files in directory*”:



Figura 5.8: Arquivamento dos dados originais

Se essa opção estiver ligada e um diretório válido for especificado, cada importação gerará um arquivo `Import _ <data> _ <hora>.zip` que contém

- Todos os currículos baixados, bem como
- Um arquivo `SUMMARY.TXT`, que descreve os insumos e o resultado da operação

```

Data import on Aug 10, 2017 2:07:41 PM

Import description:
Data source: CGEE Web Service, using filter expression:
nome:joao silva

Info String: Test Silva
Add CVs to existing network: No
Filter data for known encoding errors: Yes
Do not save filtered files

CV Id          Name          Last update  Result  Time
-----
[redacted]     Joao Silva    '2014       NEW     3 ms
[redacted]     João Silva    '2013       NEW     4 ms
[redacted]     João Silva    '2016       NEW     6 ms
[redacted]     Joao Silva    '2011       NEW     1 ms

Import result:
NEW           : 4
UPDATED      : 0
IGNORED      : 0
NOTFOUND     : 0
ERROR        : 0

```

Figura 5.9: Arquivo SUMMARY.TXT, descrevendo os insumos e o resultado da importação

5.1.3 Opções comuns

As opções descritas em seguida permitem controlar o processo da importação, independentemente da fonte de dados.

Apagar ou manter os dados do banco antes da importação

A opção “Add CVs to currently existing network” está disponível se, na hora da importação, já houver um banco de dados com currículos Lattes e diferencia entre uma importação inicial e uma importação incremental (que não elimina dados anteriores).



Figura 5.10: Opção de importação inicial ou incremental

Se essa opção for selecionada, os currículos importados serão acrescentados às informações já existentes na base. Se um currículo importado já existe na base e a versão importada é mais recente do que a versão na base, o currículo na base é substituído pela versão importada.

Se a opção não for selecionada, todos os dados que já existem no banco de dados serão apagados antes da importação. Desta forma, os dados importados substituem os dados existentes.

Campo adicional de informação

Cada pesquisador importado é representado como um nó no grafo criado. Esses nós possuem atributos, tais como o número do Currículo Lattes (atributo “id”), o nome do pesquisador (atributo “label”) e outros. O atributo “info” dos nós é preenchido com o valor especificado no campo “Additional information field to store” durante a importação.

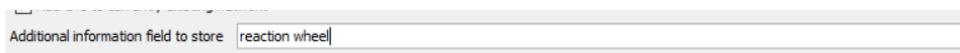


Figura 5.11: Opção do campo adicional de informação

Essa funcionalidade permite que durante a importação incremental de vários Currículos Lattes em vários passos os pesquisadores sejam categorizados em grupos com identificadores distintos. Se o mesmo currículo é importado várias vezes com valores diferentes no campo “info”, esses valores serão adicionados e separados com o caractere “/”.

Limpeza dos dados

Para permitir o processamento de arquivos contendo alguns tipos de erros identificados na base de currículos, foi desenvolvida uma correção automática, da seguinte forma:

- Caracteres CTRL-Z são substituídos por símbolos de interrogação (“?”).
- A codificação dos arquivos é determinada automaticamente e, caso não confira com a codificação declarada na linha inicial, a declaração é corrigida.



Figura 5.12: Limpeza dos dados e gravação dos arquivos corrigidos

Se a limpeza dos dados for habilitada com a opção “*Filter input files for known encoding errors*”, os currículos corrigidos podem opcionalmente ser gravados na mesma pasta (sobrescrevendo os arquivos originais) ou em outra pasta.

5.1.4 Processo de importação

Durante a importação, o CGEE Insight Net mostra uma barra de progresso e informa sobre o andamento da importação.

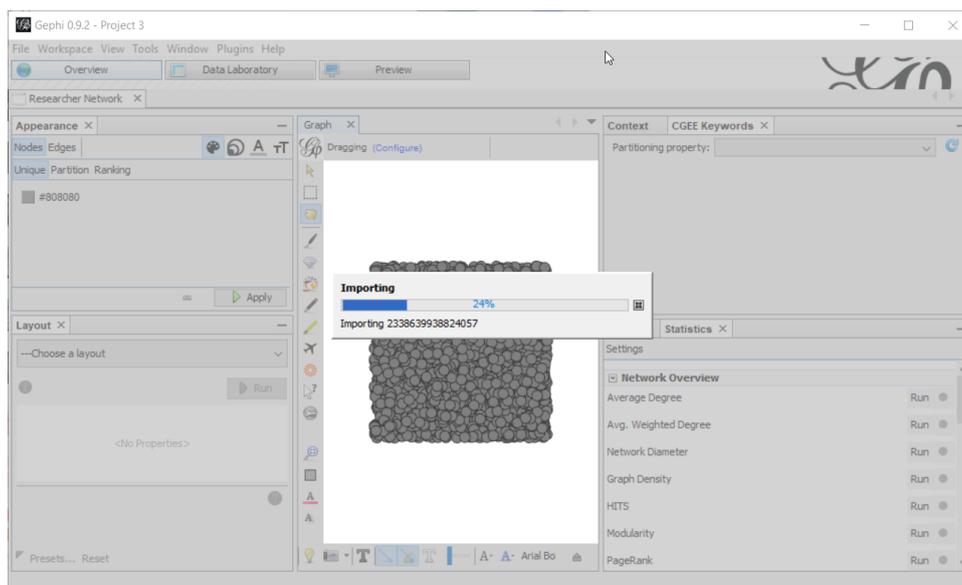


Figura 5.13: Importação dos currículos Lattes

No final da importação, a quantidade de pesquisadores importados, atualizados, ignorados e não importados por erros nos dados é exibida:

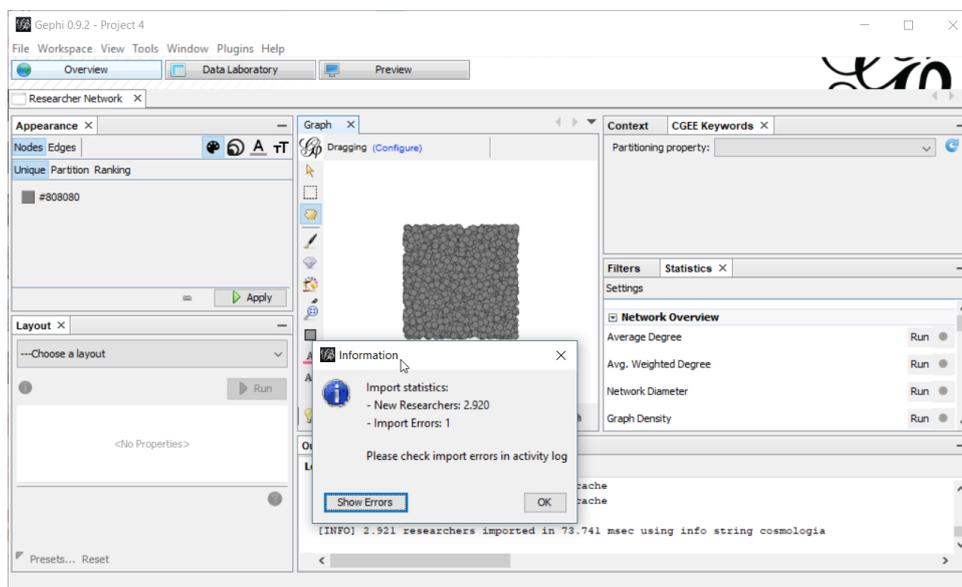


Figura 5.14: Resultado da importação dos currículos Lattes

Recomenda-se verificar essa quantidade de pesquisadores com a quantidade esperada para identificar possíveis divergências.

Caso forem identificados erros na importação, esses podem ser exibidos e detalhados. Adicionalmente, os currículos correspondentes podem ser baixados diretamente do site do CNPq em formato XML e importados manualmente (ver seção [Seção 5.1.1](#)).

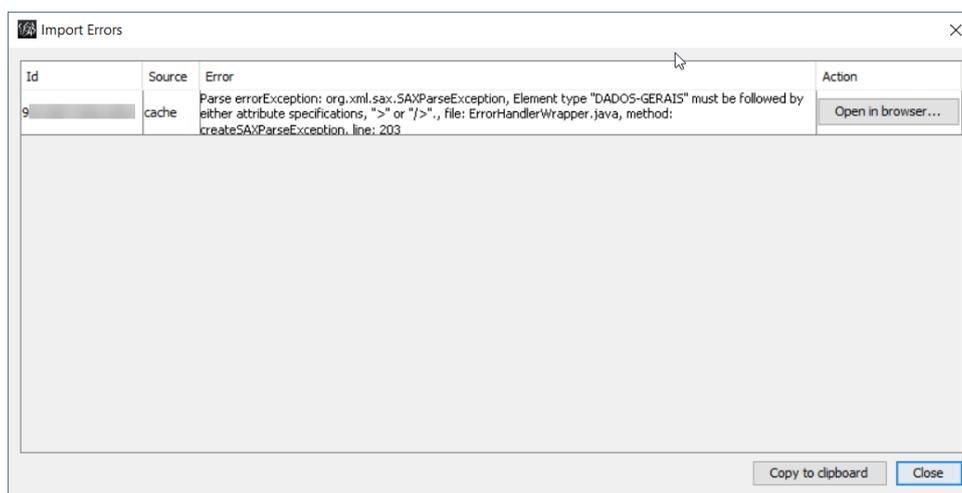


Figura 5.15: Diálogo de erros de importação

Adicionalmente, o protocolo de execução (ver seção [Protocolos de execução](#)) registra informações sobre o andamento da importação, de acordo com o grau de detalhe especificado na tela de configuração (ver seção [:options-log](#)).

O relatório de execução (ver seção [Protocolos de execução](#)) reúne todas as informações detalhadas da importação, no mesmo formato do arquivo SUMMARY.TXT (veja [Arquivamento dos dados originais](#)).

5.2 Formação da rede

Depois da importação dos currículos na base de dados, a rede é formada a partir das pesquisas por coautoria e por similaridade contextual. Os passos 2-4 da Tabela [Tabela 4.1](#) são realizados em uma única operação, transformando o conteúdo do banco de dados em um grafo. Para formar a rede, o usuário deve clicar em *Plugins > CGEE Insight Net Lattes > Build new researcher network* e preencher ou confirmar os dados do diálogo que é exibido:

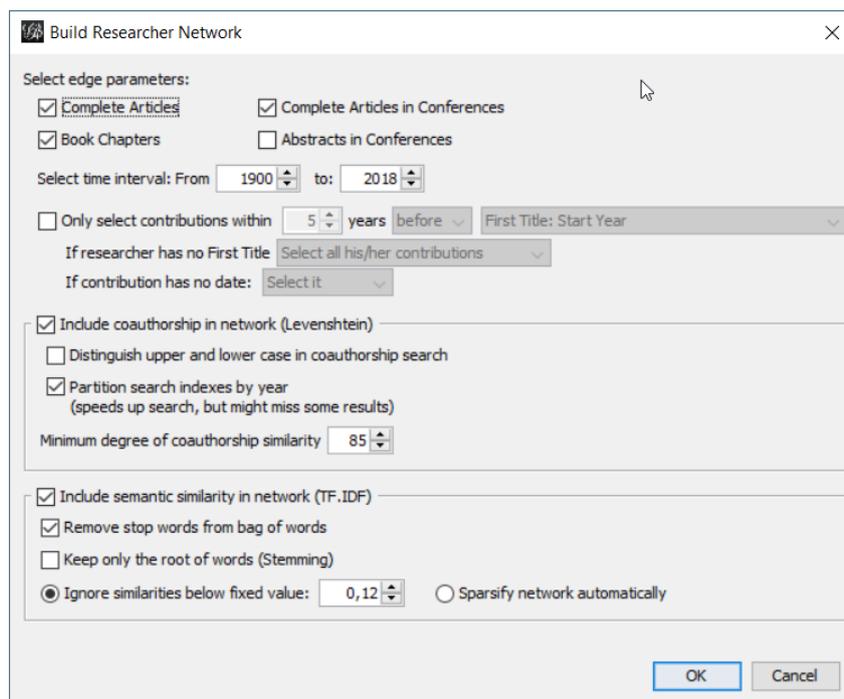
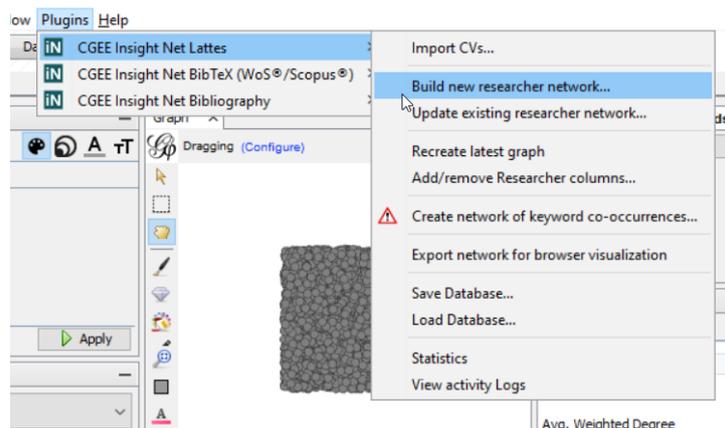


Figura 5.16: Menu e diálogo da formação da rede de currículos Lattes

As opções do diálogo serão explicadas em seguida.

5.2.1 Escopo da rede formada

Na parte superior do diálogo o usuário especifica quais tipos de contribuições farão parte do escopo da formação da rede:

- Artigos científicos completos (desconsiderando artigos de resumo em eventos)
- Capítulos em livros
- Trabalhos em eventos (Artigos completos ou apenas Resumos)
- As contribuições selecionadas podem ainda ser limitadas por período de publicação – o diálogo mostra o ano mínimo e o ano máximo de todas as contribuições importadas.
- Outra possível limitação do escopo temporal é em relação às titulações dos pesquisadores. Se a caixa “*Only select contributions within ___ years*” for selecionada, apenas as contribuições dentro da faixa temporal selecionada farão parte da rede construída. Nesse caso, é necessário definir o tratamento das contribuições dos pesquisadores que não obtiveram a titulação selecionada e também das contribuições que não possuem data.

Um fato relevante é que a rede dos pesquisadores será montada **apenas** pelas contribuições aqui selecionadas.

5.2.2 Opções da pesquisa por coautoria

Na parte do meio do diálogo, existem algumas opções que permitem controlar o processo de pesquisa por coautoria:

- A caixa “*Include coauthorship in network (Levenshtein)*” permite determinar se a pesquisa por coautoria é realizada e habilita os outros campos desta seção. Se essa caixa não for selecionada, a rede formada não terá arestas de coautorias.
- Geralmente, a pesquisa não distingue entre letras minúsculas e letras maiúsculas e considera palavras como “Contribuição” e “CONTRrIBUIÇÃO” como iguais. Marcando a opção “*Distinguish upper and lower case in coauthorship search*”, as duas palavras são consideradas diferentes e os resultados dos cálculos de coautoria serão ser diferentes
- Para encontrar contribuições com nome semelhantes, cada tipo de contribuição só é procurada dentro do conjunto de contribuições do mesmo tipo (por exemplo, para achar um artigo com título semelhante, são apenas analisados os títulos dos outros artigos e não dos trabalhos em eventos ou dos capítulos de livros). Esse particionamento reduz significativamente o tempo de busca. Adicionalmente, o usuário pode especificar que as contribuições também devem ter sido publicadas no mesmo ano. Dessa forma, a similaridade de um artigo publicado em 2005 será apenas procurada nos artigos publicados em 2005 (e não nos artigos de todos os anos). Essa opção agiliza significativamente a pesquisa por similaridade, mas pode levar à uma situação onde contribuições inseridas com o ano errado não serão encontradas.
- A similaridade mínima a partir da qual os títulos de duas contribuições são considerados iguais também pode ser alterada pelo usuário. Valores entre 85% e 90% se mostraram adequados para minimizar a incidência de falsos positivos e falsos negativos no que se refere ao número de coautorias.

5.2.3 Opções da pesquisa por similaridade semântica

A parte inferior do diálogo permite a seleção das opções de pesquisa por similaridade semântica:

- A caixa “*Include semantic similarity in network (TF.IDF)*” determina se a pesquisa por similaridade semântica (também conhecida como “Similaridade contextual”) é realizada e habilita os outros campos dessa seção. Se essa caixa não for selecionada, a rede formada não terá arestas de similaridade semântica.

- O usuário pode selecionar se os pré-processamentos dos termos “*Stop words*” e “*Stemming*” serão realizados ou não.
 - “*Stop Words*” são as palavras mais frequentes de cada idioma, que não agregam informação aos termos identificados e serão eliminados da pesquisa. Os *stop words* são implementados apenas para os títulos em Inglês e Português.
 - O “*Stemming*” reduz, em um algoritmo específico por idioma, cada palavra a uma raiz que desconsidera flexões gramaticais. Nesse momento, apenas os idiomas Português e Inglês são tratados pelo stemming. Títulos em outros idiomas permanecem na forma original.
- O cálculo da similaridade semântica gera arestas entre praticamente qualquer par de pesquisadores. A grande maioria deles com baixos valores de similaridade que não agregam informações relevantes ao conteúdo do gráfico. Por esse motivo, existem três métodos para reduzir a quantidade de arestas na rede:
 - “*Ignore similarities below fixed value*”: valores abaixo de um limite especificado podem ser desconsiderados, produzindo o valor final zero como similaridade contextual.
 - “*Sparsify network automatically*”: Um algoritmo automático [7] é utilizado para reduzir a quantidade de arestas na rede. Observe-se que testes realizados com esse algoritmo não levaram a resultados conclusivos quanto à sua eficácia.
 - Para o cálculo de similaridade podem ser considerados apenas os termos mais relevantes das contribuições. Essa configuração é realizada no diálogo de opções do *CGEE Insight Net* (ver seção *Configuração do CGEE Insight Net*), advertindo-se que, na grande maioria dos casos, essa opção só deva ser empregada por usuários experientes. Se um percentil de relevância dos termos for definido, uma mensagem correspondente é exibida:

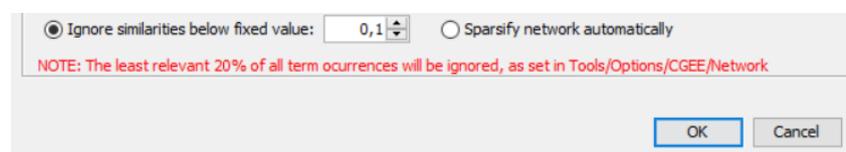


Figura 5.17: Aviso sobre configuração de limite inferior de relevância

Clicando em “OK”, o *CGEE Insight Net* inicia a sequência de processamento: No primeiro passo, as contribuições dentro do escopo especificado são identificadas, selecionadas e pré-processadas. O processo pode ser interrompido clicando no símbolo  do indicador de progresso:

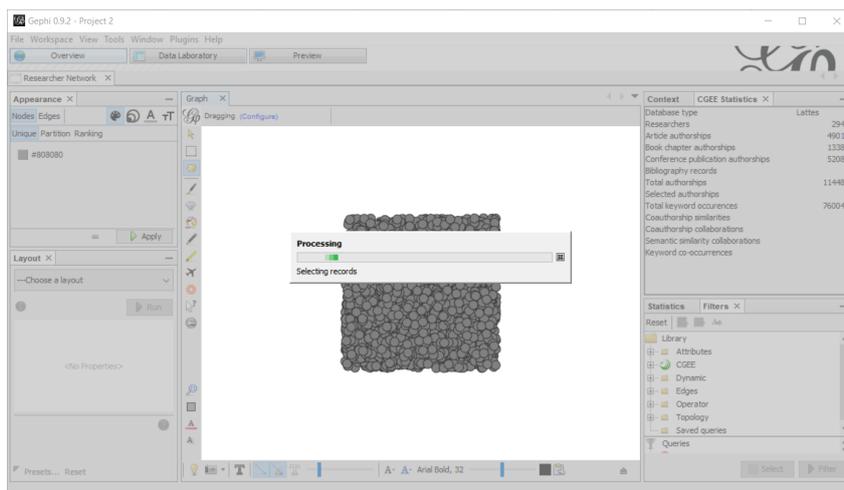


Figura 5.18: Seleção das contribuições e pré-processamento

Depois desta fase, o *CGEE Insight Net* inicia a formação da rede (passo 3 da Tabela 4.1). Esse passo pode levar um tempo considerável, dependendo da quantidade de contribuições selecionadas, da capacidade do computador de paralelizar a pesquisa (quantidade de processadores e núcleos), do tipo do banco de dados, da velocidade de conexão e dos parâmetros especificados pelo usuário. O *CGEE Insight Net* mostra uma barra de progresso que indica o percentual dos dados já processados. Novamente, o processo pode ser interrompido clicando no símbolo  desta barra:

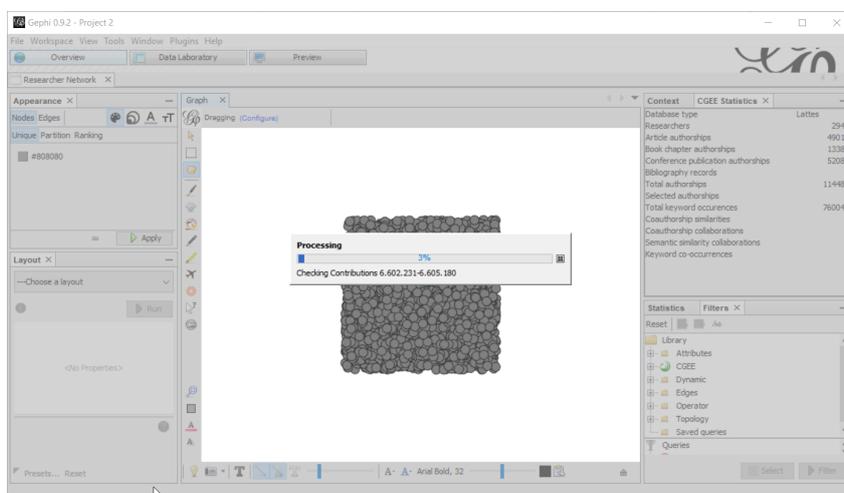


Figura 5.19: Processamento da pesquisa por similaridade

Depois da conclusão deste passo, os dados são pós-processados e a rede de colaboração é montada visualmente na tela:

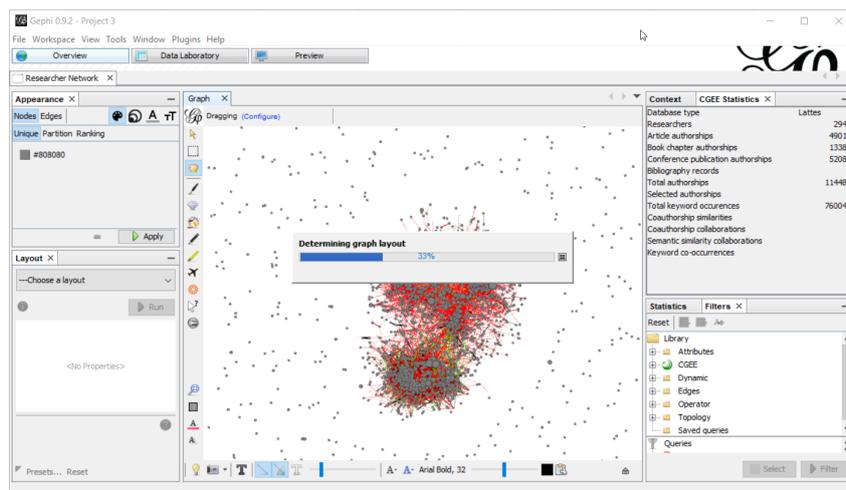


Figura 5.20: Pós-processamento e montagem da rede na tela

Após a conclusão desta etapa, a tela é liberada pelo *CGEE Insight Net* e o usuário pode analisar a rede com as ferramentas disponíveis do Gephi, tais como análise de *clusters*, particionamentos ou estatísticas da rede, usando as ferramentas disponibilizadas pelo Gephi.

5.2.4 Atualização da pesquisa

Para atualizar os cálculos de uma rede formada, o usuário pode selecionar o item “*Update existing researcher network...*”:

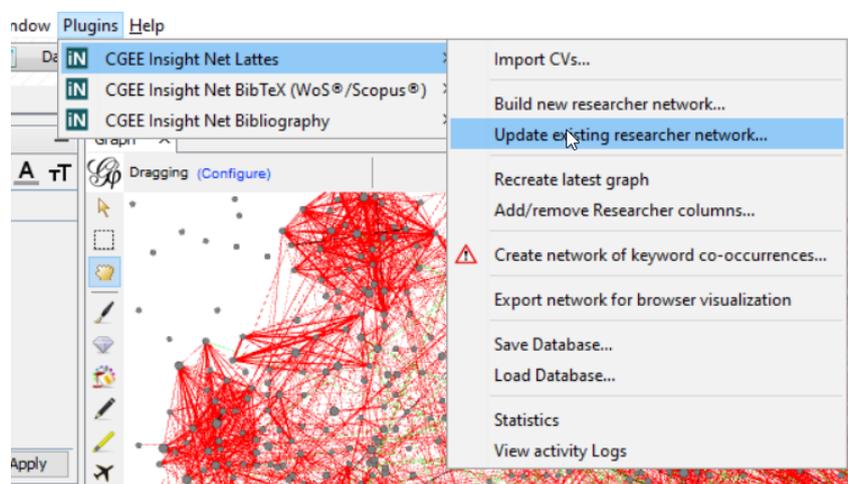


Figura 5.21: Atualização da rede já calculada

Diferentemente da opção “*Build new researcher network...*”, que elimina qualquer rede pré-existente, o item “*Update existing researcher network...*” mantém os dados dos cálculos que não são selecionados pelo usuário. Dessa forma, é possível atualizar apenas as arestas de similaridade contextual, sem recalcular toda a rede de coautorias.

5.3 Visualização de atributos dos pesquisadores

Cada pesquisador possui uma grande quantidade de atributos que são obtidos através do seu Currículo Lattes:

- Nome completo;
- Nome em citações;
- Instituição;
- Estado da Instituição;
- Quantidade de artigos completos, capítulos de livros e publicações completas e resumos em eventos;
- Ano da última atualização do currículo;
- Campo adicional de informação, definido pelo usuário durante a importação dos dados;
- Quantidade total, bem como de cada tipo de contribuição bibliográfica;
- Data de nascimento;
- Local de nascimento;
- Dados da primeira, da última e da mais alta titulação que constam no currículo: ano de início e de fim, ano da titulação, tipo, instituição e assunto da titulação.
- Os mesmos dados são registrados para a primeira e última titulação de cada tipo de titulação: - Ensino fundamental; - Ensino médio; - Curso técnico profissionalizante; - Graduação; - Especialização; - Residência Médica; - Mestrado profissionalizante; - Mestrado; - Doutorado; e - Livre Docência.

Os seguintes atributos estão disponíveis apenas por pessoas com autorização explícita para processar dados pessoais:

- CPF;
- Sexo.

Os seguintes atributos são definidos para cada pesquisador durante o cálculo de uma rede de co-autorias ou de similaridade semântica:

- Quantidade total, bem como de cada tipo de contribuição bibliográfica **selecionada para o cálculo da rede**;
- Quantidade média de palavras chave por contribuição bibliográfica **selecionada para o cálculo da rede**;
- Porcentagem das contribuições bibliográficas selecionadas para o cálculo de rede que possuem, no mínimo uma palavra-chave.

O cálculo de uma rede de similaridade semântica ainda fornece os seguintes atributos:

- Quantidade total de *termos* em todas as contribuições bibliográficas selecionadas do pesquisador;
- Quantidade de **termos diferentes** em todas as contribuições bibliográficas selecionadas do pesquisador.

Um *termo* no sentido do parágrafo anterior corresponde, basicamente a uma palavra. Entretanto *stop-words* (ver Seção 5.2.3) não fazem parte dessa contagem e o *stemming* pode reduzir a quantidade de termos diferentes.

A relevância dessas informações depende do projeto específico do usuário, que deve escolher o subconjunto que melhor atende aos seus requisitos.

Os dados relevantes podem ser selecionados com a função “Add/remove Researcher columns... ”:

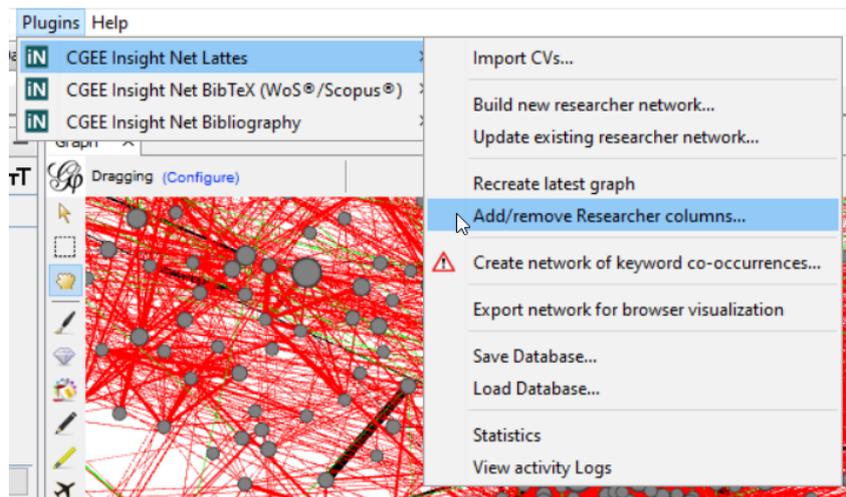


Figura 5.22: Funcionalidade para selecionar dados relevantes dos pesquisadores

Na seleção dessa funcionalidade, o seguinte diálogo é exibido:

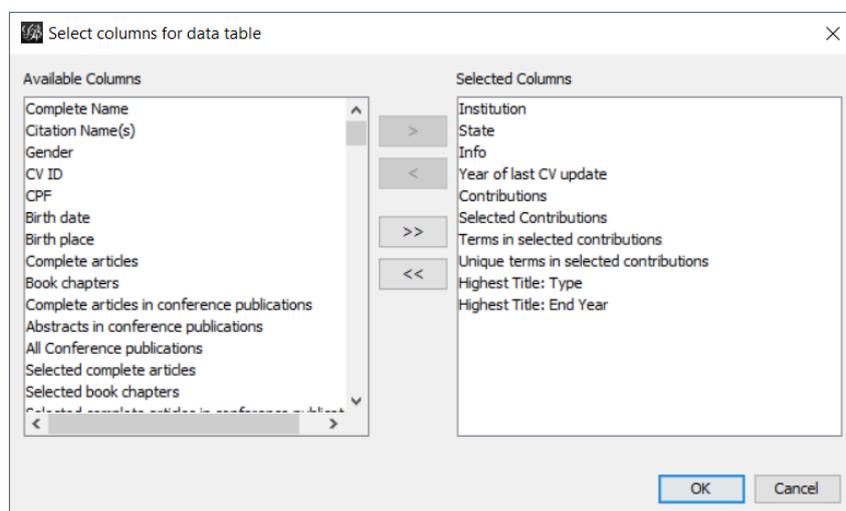


Figura 5.23: Seleção dos atributos dos pesquisadores

O diálogo corresponde à tela de opções descrita na Seção 3.3. A lista à direita mostra os atributos atualmente exibidos e a do lado esquerdo contém os atributos disponíveis (não exibidos). O usuário pode selecionar atributos nas listas com clique e *Shift/Ctrl-clique* e usar os botões , ,  e  para movê-los entre as duas listas, conforme descrito na Seção 3.3.

Destaca-se que o Gephi possui duas funcionalidades de visualização dos atributos. A primeira é na

exibição do grafo da rede. Selecionando a ferramenta  e clicando em um dos nós do grafo, todos os atributos desse nó serão exibidos, desde que eles tenham sido selecionados com a função “Add/ Remove Researcher column” ou no diálogo “Tools/Options/CGEE/Columns” do Insight Net Browser:

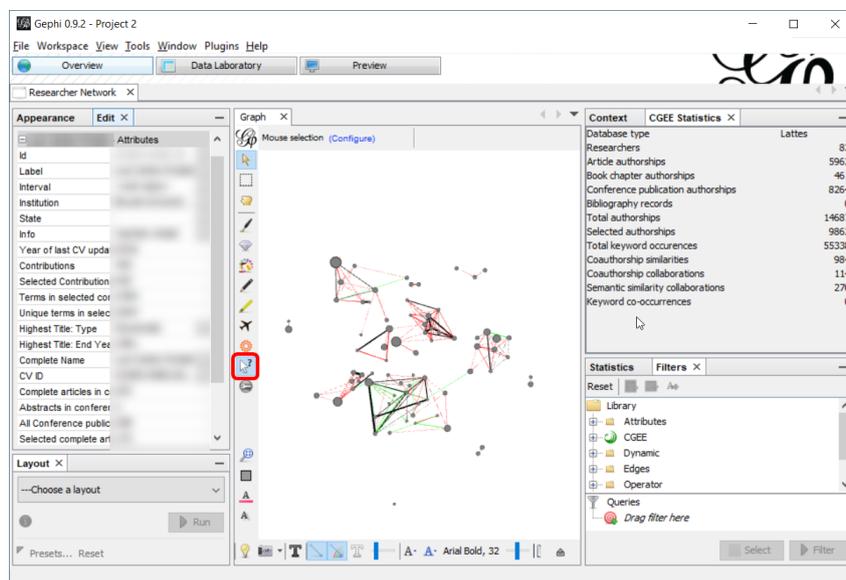


Figura 5.24: Exibição no grafo de todos os atributos habilitados do pesquisador

Na alternativa de visualização do laboratório de dados (“Data Laboratory”), o Gephi limita a quantidade de colunas exibidas em 20. Se os nós da rede apresentarem mais atributos, os 20 mais relevantes devem ser selecionados com um clique no símbolo :

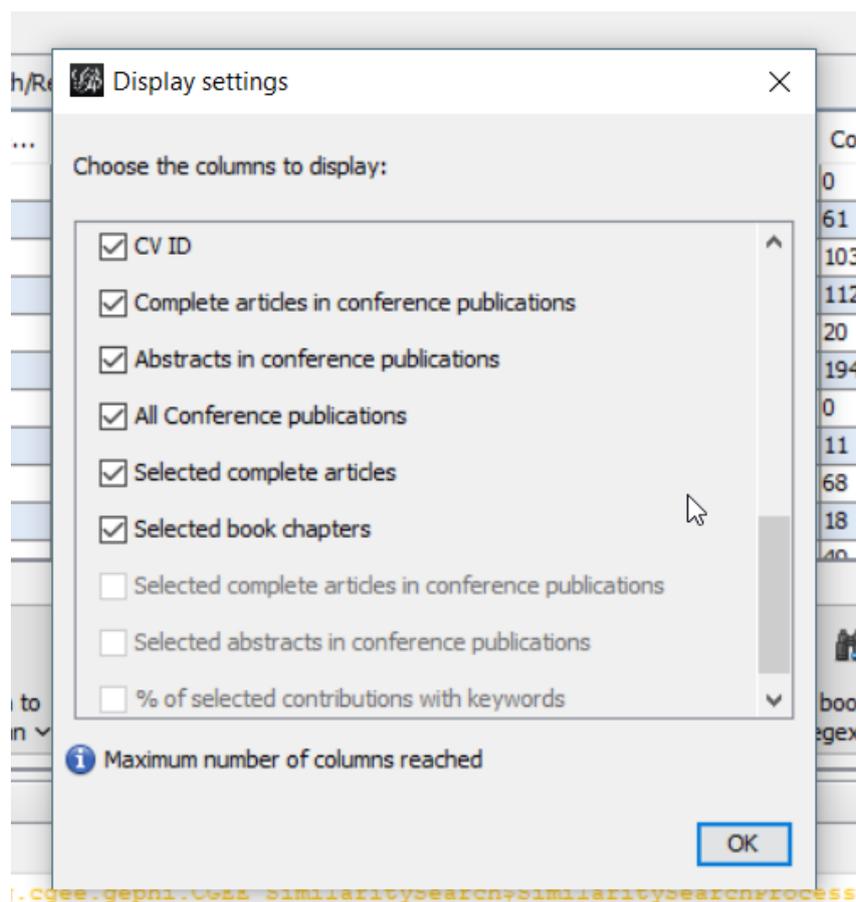


Figura 5.25: Seleção de colunas de atributos no laboratório de dados

Criação de redes de referências bibliográficas BibTeX

O CGEE Insight Net oferece, com licença adicional, a criação de redes de referências bibliográficas a partir de arquivos BibTeX dos serviços Web of Science® e Scopus®. Devido a divergências na formatação de arquivos BibTeX, o funcionamento com outras fontes de dados não pode ser garantido.

As redes bibliográficas são criadas a partir da similaridade semântica entre títulos e/ou resumos (“*abstracts*”) das publicações.

Para habilitar essa funcionalidade do *CGEE Insight Net*, a licença `INSIGHTNET_BIBTEX` deve ser instalada, conforme descrito na [Seção 3.6](#). Essa licença é exibida assim no diálogo *Tools > Options > CGEE > License*:

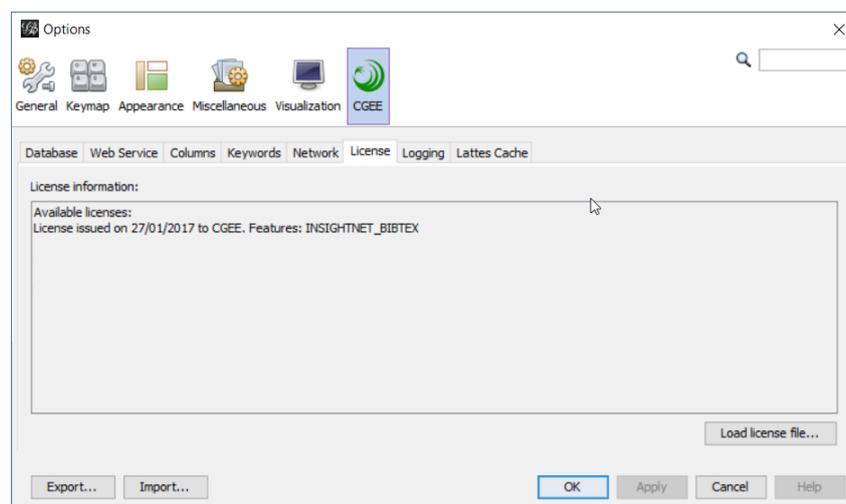


Figura 6.1: Licença requerida para o módulo de redes bibliográficas

Caso a licença esteja habilitada, aparece o sub-menu “*CGEE Insight Net BibTeX (WoS®/Scopus®)*”:

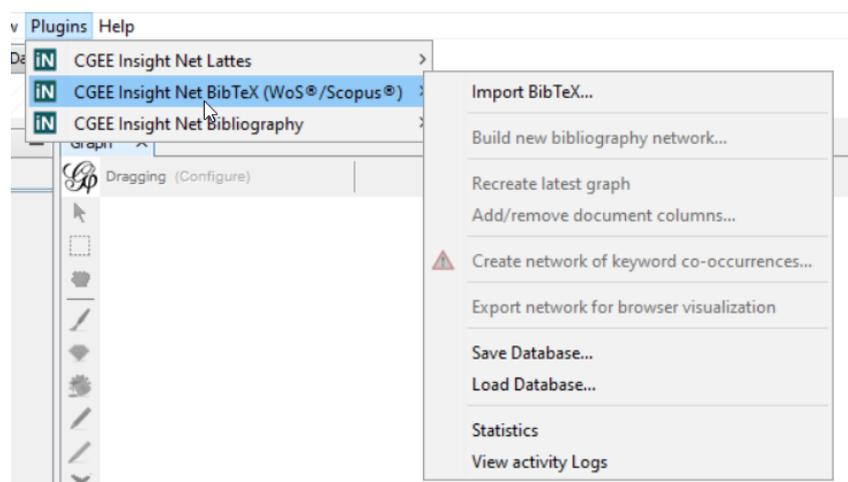
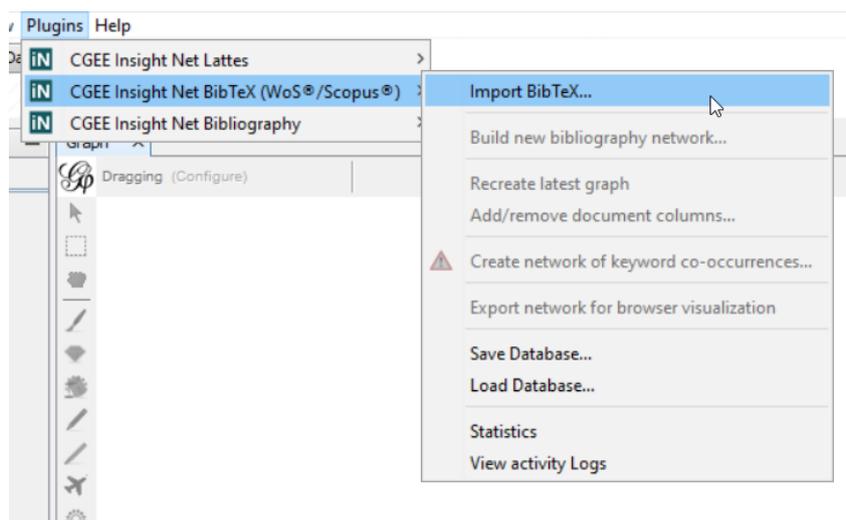


Figura 6.2: Sub-menu CGEE Insight Net BibTeX (WoS®/Scopus®)

6.1 Importação dos dados bibliográficos

Para importar dados bibliográficos, o usuário deve clicar em *Plugins* > *CGEE Insight Net BibTeX (WoS®/Scopus®)* > *Import BibTeX* e escolher os arquivos a serem importados:

- Clicando em um arquivo, este será importado;
- Vários arquivos podem ser selecionados com “*Shift-Clique*” ou “*Ctrl-Clique*”, de acordo com os padrões de uso do sistema operacional;
- O usuário também pode selecionar um ou mais diretórios. Nesse caso, todos os arquivos nesse(s) diretório(s) serão importados.



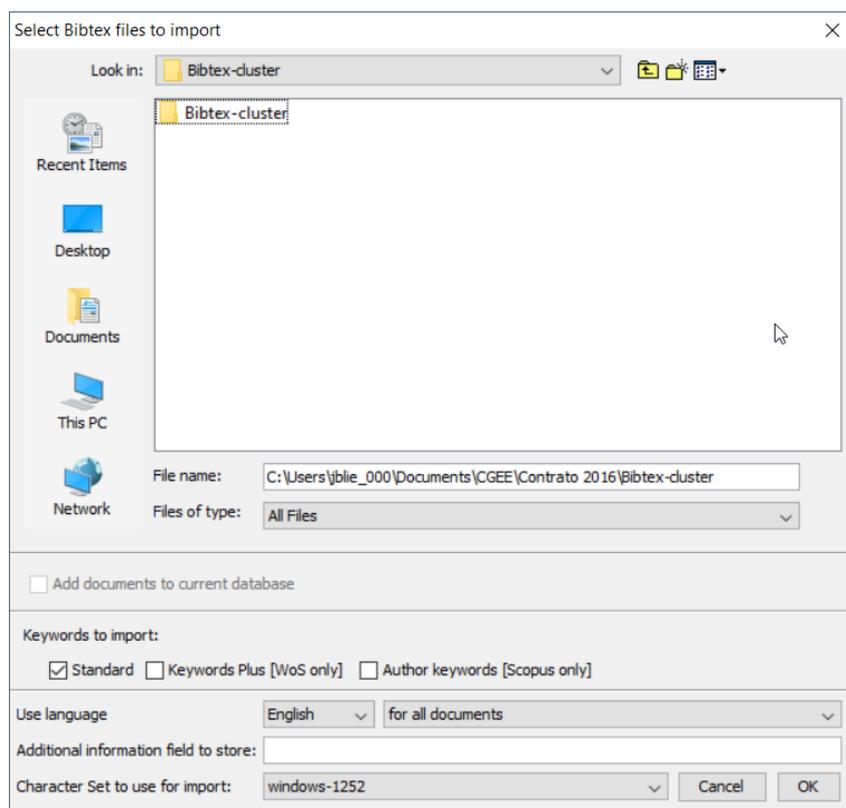


Figura 6.3: Importação de arquivos BibTeX

É importante mencionar novamente que o módulo foi desenvolvido especificamente para dados originados dos serviços *Web of Science*® e *Scopus*® e que não é garantido o funcionamento com arquivos BibTeX gerados por outros serviços. Para arquivos genéricos, recomenda-se o módulo de *Bibliografia Genérica* (ver Seção 7).

6.1.1 Apagar ou manter os dados do banco antes da importação

A opção “*Add documents to current database*” diferencia entre uma importação inicial e uma importação incremental.

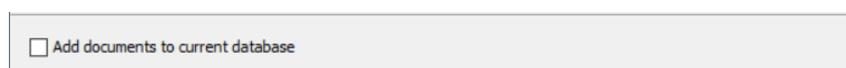


Figura 6.4: Opção de importação inicial ou incremental

Se essa opção for selecionada, as referências bibliográficas importadas serão acrescentadas às informações já existentes na base. Caso uma referência bibliográfica já exista na base, a versão importada complementa as informações existentes.

Se a opção não for selecionada, todos os dados que já existem no banco de dados serão apagados antes da importação. Desta forma, os dados importados substituem os dados atuais.

6.1.2 Palavras-chave a serem importadas

Os serviços *Web of Science*® e *Scopus*® publicam as palavras-chave definidas pelo autor, bem como palavras-chave adicionais, extraídas pelas equipes das duas empresas a partir dos dados da publicação e do seu contexto. A opção “*Keywords to import*” permite a definição de quais delas serão importadas:

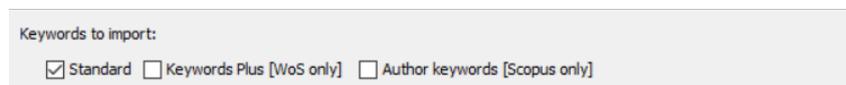


Figura 6.5: Seleção de palavras-chave a serem importadas

6.1.3 Idioma

Na parte inferior do diálogo de importação aparecem as definições que determinam o tratamento das referências bibliográficas em relação ao idioma.

Geralmente, as publicações possuem um campo que determina o idioma da publicação. Entretanto, na prática, a maioria dos dados de títulos e resumos das duas plataformas é disponibilizada em inglês, mesmo que o idioma original da publicação seja outro.

Para tratar esses casos, o diálogo de importação permite especificar o idioma padrão para as publicações sem idioma definido ou até para todos os documentos importados:

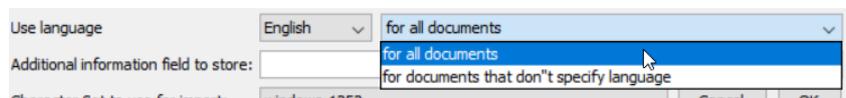


Figura 6.6: Definição do idioma padrão

6.1.4 Campo adicional de informação

Cada referência bibliográfica importada é representada como um nó no grafo criado. Esses nós possuem atributos, tais como o identificador DOI (*Digital Object Identifier* – vide <http://www.doi.org/>) da publicação (atributo “*DOI*”), o seu título e outros. O atributo “*info*” dos nós é preenchido com o valor especificado no campo “*Additional information field to store*” durante a importação.



Figura 6.7: Campo adicional de informação

6.1.5 Conjunto de caracteres

A última opção do diálogo especifica o conjunto de caracteres (character set) do(s) arquivo(s) a ser(em) importado(s). Contrário aos arquivos de Currículos Lattes, essa informação não está contida dentro de arquivos BibTex e precisa ser fornecida pelo usuário.

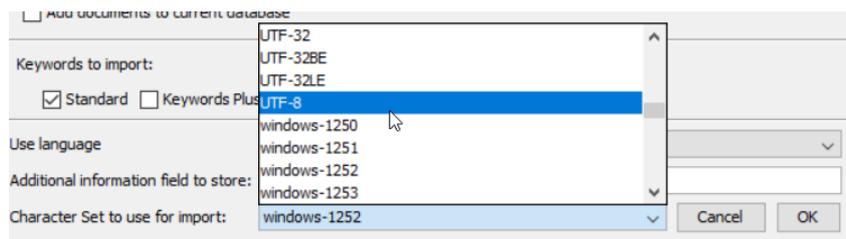


Figura 6.8: Definição do conjunto de caracteres

Os charsets mais usados são “Windows-1252”, para arquivos nativos do Windows e “UTF-8”, para arquivos que foram baixados da internet. A seleção do *charset* errado se manifesta em erros nas letras acentuadas durante a visualização do título e do resumo da referência bibliográfica.

6.1.6 Processo de importação

Durante a importação, o *CGEE Insight Net* mostra uma barra de progresso. Recomenda-se verificar a quantidade de referências bibliográficas no final da importação a partir da estatística do banco de dados (ver Seção 9.3).

Adicionalmente, o protocolo de execução (ver Seção 9.4) registra informações sobre o andamento da importação, de acordo com o grau de detalhe especificado na tela de configuração (ver Seção 3).

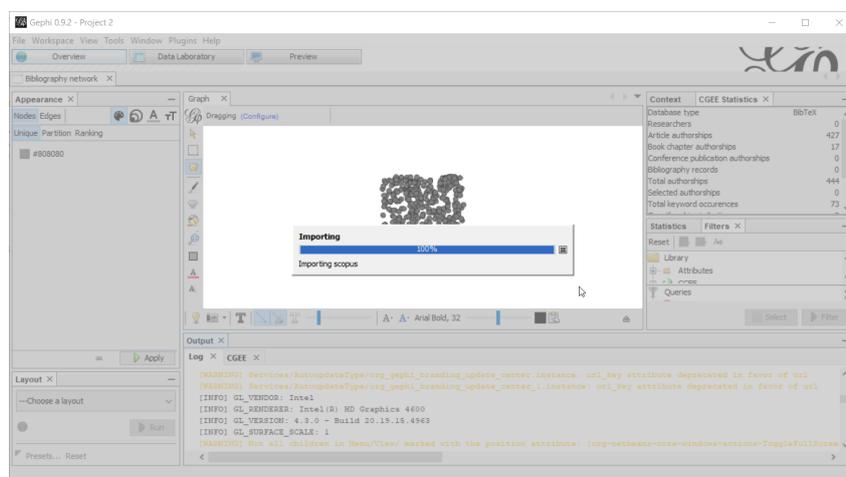


Figura 6.9: Importação de referências bibliográficas

6.2 Formação da rede

Depois da importação das referências bibliográficas na base de dados, a rede é formada a partir das pesquisas por similaridade contextual. Os passos 2-4 da [Tabela 4.1](#) são realizados em uma única operação, transformando o conteúdo do banco de dados em um grafo.

Para formar a rede, o usuário deve clicar em *Plugins > CGEE Insight Net BibTeX (WoS®/Scopus®) > Build new bibliography network* e preencher ou confirmar os dados do diálogo que é exibido:

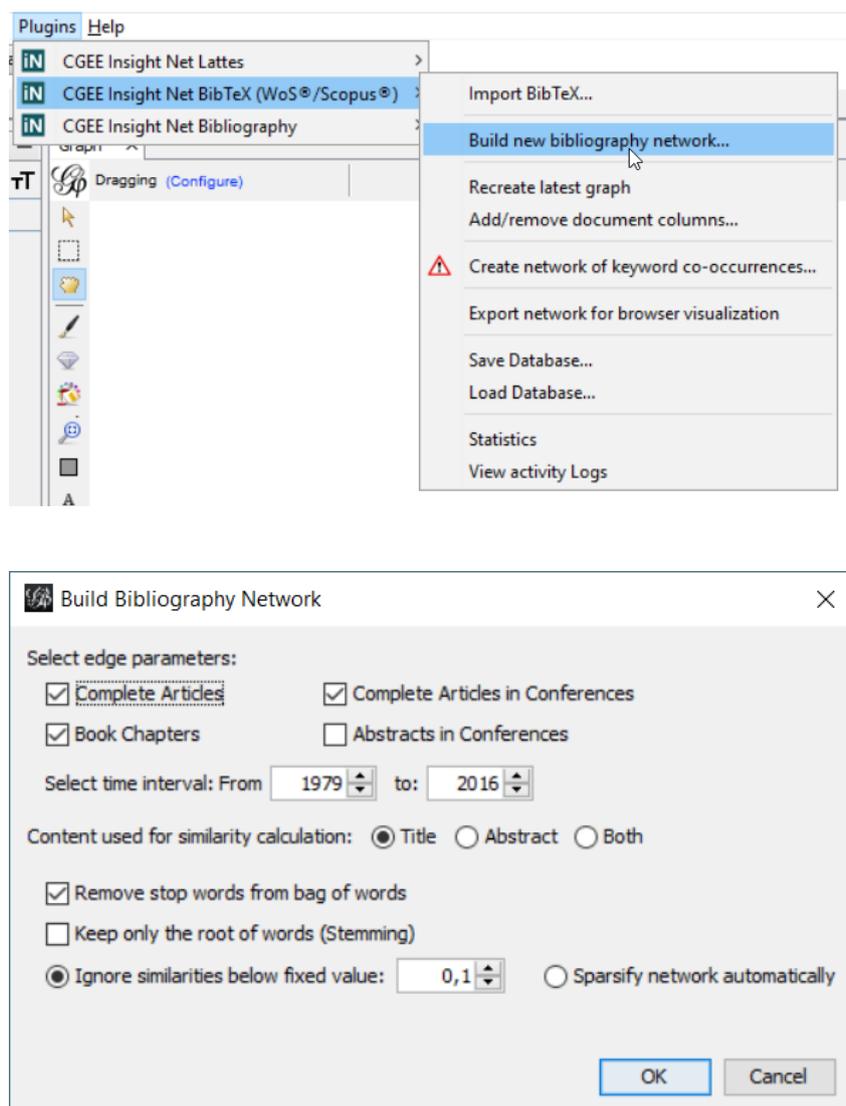


Figura 6.10: Menu e diálogo da formação da rede de referências bibliográficas

As opções do diálogo correspondem são explicadas em seguida.

6.2.1 Escopo da rede formada

Na parte superior do diálogo o usuário especifica quais tipos de publicações farão parte do escopo da formação da rede:

- Artigos científicos completos (“*Complete Articles*”). Nesta categoria se encaixam referências BibTex dos seguintes tipos: ** Article * Review * Letter * Chronology * Reprint * Meeting Abstract * Note * Editorial*
- Capítulos em livros (“*Book Chapters*”). Para esse tipo de entidade são mapeadas as referências BibTex que contêm a palavra “*Proceedings*” no tipo.
- Trabalhos em eventos (“*Conference publications*”): São as referências bibliográficas com a palavra “*Conference*” no tipo.
- As contribuições selecionadas podem ainda ser limitadas por uma faixa de anos de publicação – o diálogo mostra a data mínima e a data máxima de todas as contribuições importadas.

Destaca-se novamente que a rede de referências bibliográficas será montada apenas para as contribuições selecionadas.

6.2.2 Conteúdo considerado para o cálculo de similaridade contextual

A parte inferior do diálogo permite a seleção das opções da pesquisa por similaridade semântica.

- A seleção “*Content used for similarity calculation*” determina se a pesquisa por similaridade semântica considera apenas o título da referência bibliográfica (“*Title*”), apenas o resumo (“*Abstract*”) ou ambos (“*Both*”).
- O usuário pode selecionar se os pré-processamentos dos termos “*Stop words*” e “*Stemming*” serão realizados ou não. Esses dois algoritmos dependem da definição correta do idioma da referência bibliográfica.
 - *Stop Words* são as palavras mais frequentes de cada idioma, que não agregam informação aos termos identificados e serão eliminados da pesquisa. Os stop words são implementados apenas para as referências bibliográficas em Inglês e Português.
 - O *Stemming* reduz, em um algoritmo específico por idioma, cada palavra a uma raiz que desconsidera flexões gramaticais. Nesse momento, apenas os idiomas Português e Inglês são tratados pelo stemming. Referências bibliográficas em outros idiomas permanecem na forma original.
- O cálculo da similaridade semântica gera arestas entre praticamente qualquer par de referências bibliográficas, a maioria com baixos valores de similaridade, que não agregam informações relevantes ao conteúdo do gráfico. Por esse motivo, existem três métodos para reduzir a quantidade de arestas na rede:
 - “*Ignore similarities below fixed value*”: valores abaixo de um limite especificado podem ser desconsiderados, produzindo o valor final zero como similaridade semântica.
 - “*Sparsify network automatically*”: Um algoritmo automático [7] ainda experimental é utilizado para reduzir a quantidade de arestas na rede.
 - Para o cálculo de similaridade podem ser considerados apenas os termos mais relevantes das contribuições. Esta configuração é feita no diálogo de opções do CGEE Insight Net (ver [Seção 3](#)). Se um percentil de relevância dos termos for definido, uma mensagem correspondente é exibida:

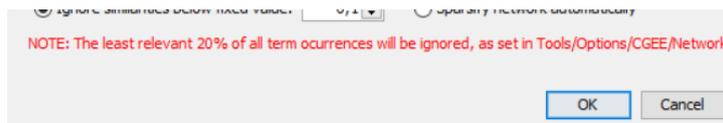


Figura 6.11: Aviso sobre configuração de limite inferior de relevância

Criação de redes de referências bibliográficas genéricas

O CGEE Insight Net oferece, com licença adicional, a criação de redes de referências bibliográficas a partir de arquivos dos serviços Web of Science® e Scopus®, bem como usando planilhas Excel® ou até qualquer outro formato estruturado de dados.

As redes bibliográficas são criadas a partir da similaridade semântica entre títulos e/ou resumos (“*abstracts*”) das publicações. A funcionalidade de criação de redes de co-ocorrências de palavras-chaves (ver Seção 8.5) complementa essa análise.

Para habilitar essa funcionalidade do *CGEE Insight Net*, a licença `INSIGHTNET_BIBLIOGRAPHY` deve ser instalada, conforme descrito na Seção 3.6. Essa licença é exibida assim no diálogo *Tools > Options > CGEE > License*:

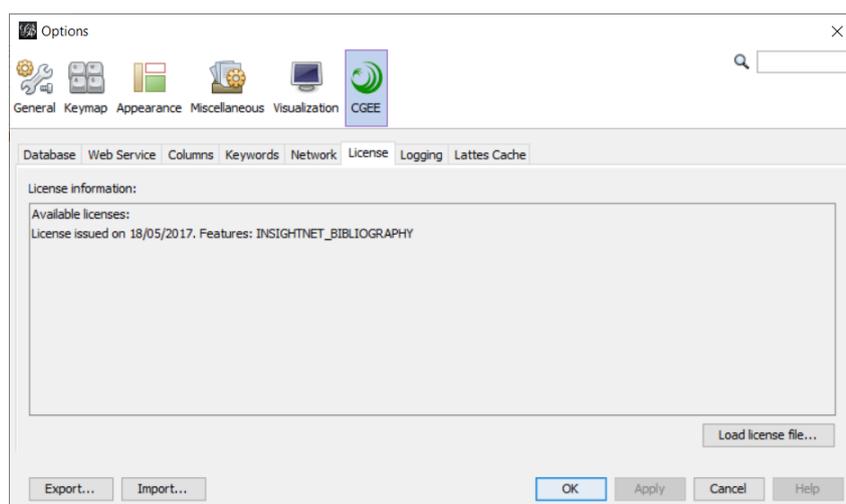


Figura 7.1: Licença requerida para o módulo de redes bibliográficas

Caso a licença esteja habilitada, aparece o sub-menu “*CGEE Insight Net Bibliography*”:

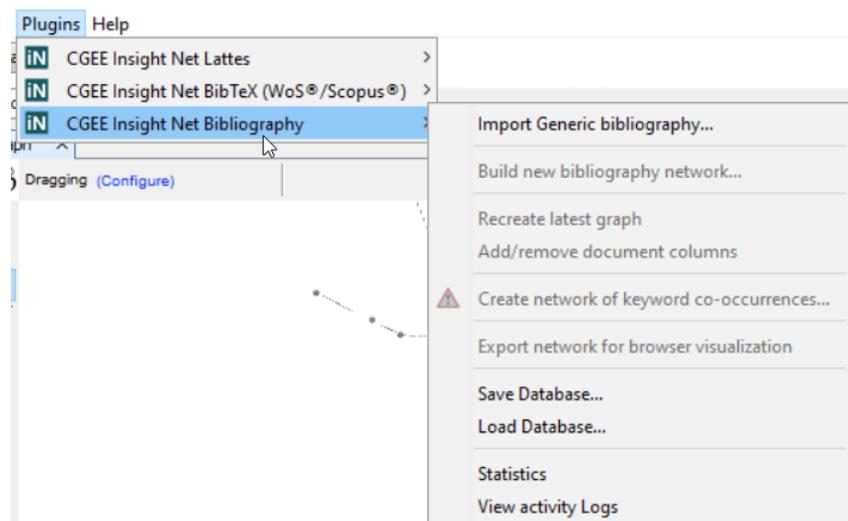


Figura 7.2: Sub-menu *CGEE Insight Net Bibliography*

7.1 Importação dos dados bibliográficos

Para importar dados bibliográficos, o usuário deve clicar em *Plugins > CGEE Insight Net Bibliography > Import Generic Bibliography* e escolher os arquivos a serem importados:

- Clicando em um arquivo, este será importado;
- Vários arquivos podem ser selecionados com “*Shift-Clique*” ou “*Ctrl-Clique*”, de acordo com os padrões de uso do sistema operacional;
- O usuário também pode selecionar um ou mais diretórios. Nesse caso, todos os arquivos nesse(s) diretório(s) serão importados.

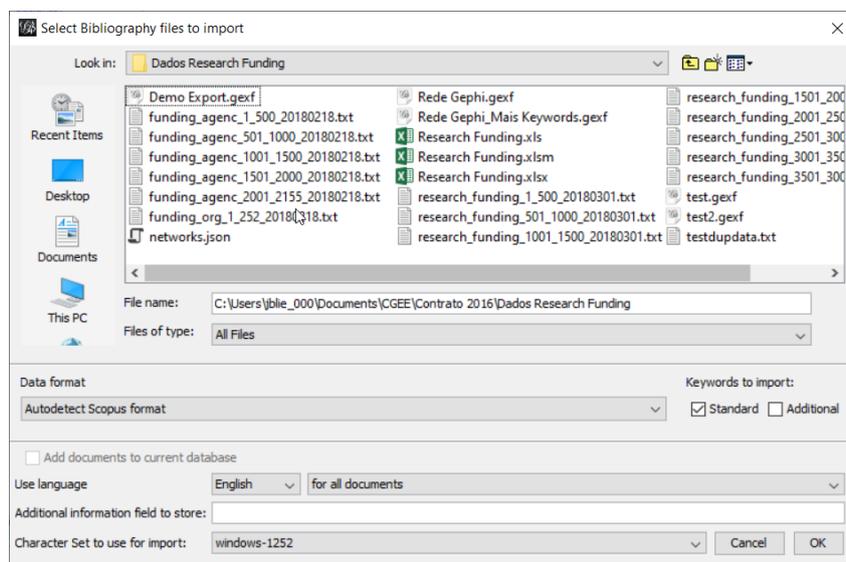
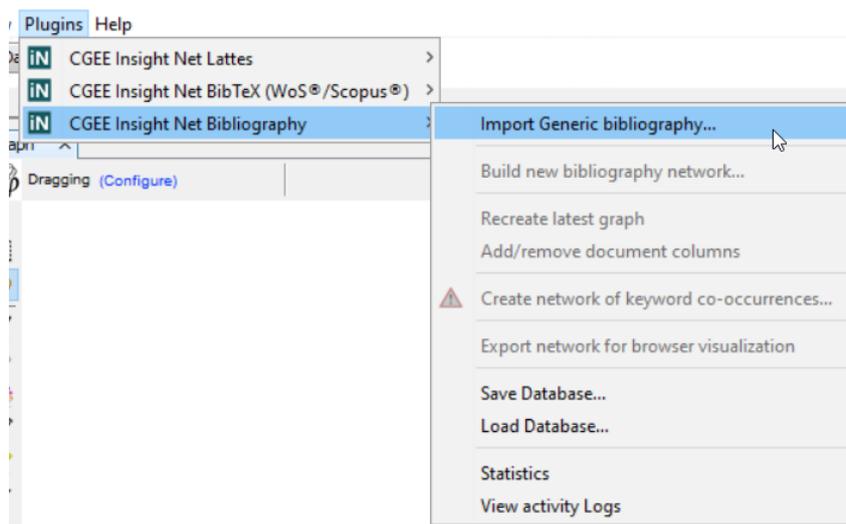


Figura 7.3: Importação de arquivos de bibliografia

7.1.1 Formato de dados

O módulo de bibliografia genérica traz uma lista de formatos de dados pré-definidos, que atendem a demandas específicas. A escolha do formato apropriado na importação dos dados é essencial. Esta

seção descreve os formatos disponíveis e orienta sobre os seus usos.

Formatos *Web of Science*[®] e *Scopus*[®]

Os serviços *Web of Science*[®] *Scopus*[®] disponibilizam dados em vários formatos que o *CGEE Insight Net* pode importar:

- *Comma separated values (CSV)*
- *Tab separated values (TSV)*
- *RIS* ou *Tagged*

O *Web of Science*[®] e o *Scopus* ainda permitem a disponibilização dos dados em formato “BibTeX”, que pode ser importado pelo módulo específico do *CGEE Insight Net* (ver Seção 6). Entretanto, o módulo de referências BibTeX carrega apenas um subconjunto das informações disponibilizadas.

A lista de formatos *Web of Science*[®] e *Scopus*[®] mostra, para cada um dos serviços as três opções mencionadas acima, bem como a possibilidade de detectar o formato automaticamente, o que, geralmente, é a forma mais indicada:

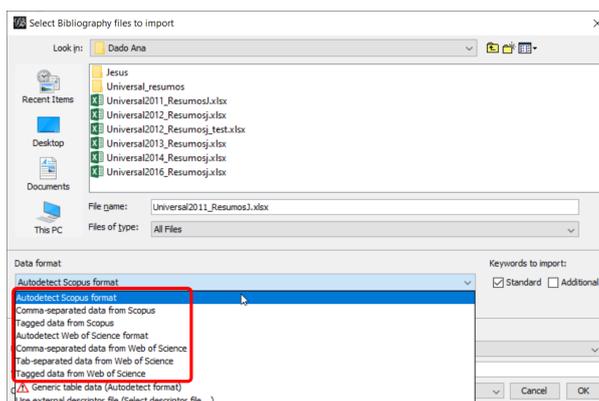


Figura 7.4: Formatos de arquivos de bibliografia dos serviços *Web of Science*[®] e *Scopus*[®]

As opções *Autodetect Scopus format* e *Autodetect Web of Science Format* permitem a importação de dados textuais desses serviços sem a necessidade de saber se o formato é *CSV*, *TSV* ou *Tagged*.

Arquivos no formato BibTeX devem ser importados pelo módulo de referências bibliográficas BibTeX (ver :numref:‘bibtex’):

Formato “*Generic table data (Autodetect format)*”

O formato *Generic table data (Autodetect format)* usa planilhas Excel como arquivos de entrada. Neste caso, recomenda-se a importação de apenas um único arquivo, pois o usuário precisa especificar a aba da planilha a ser importada:

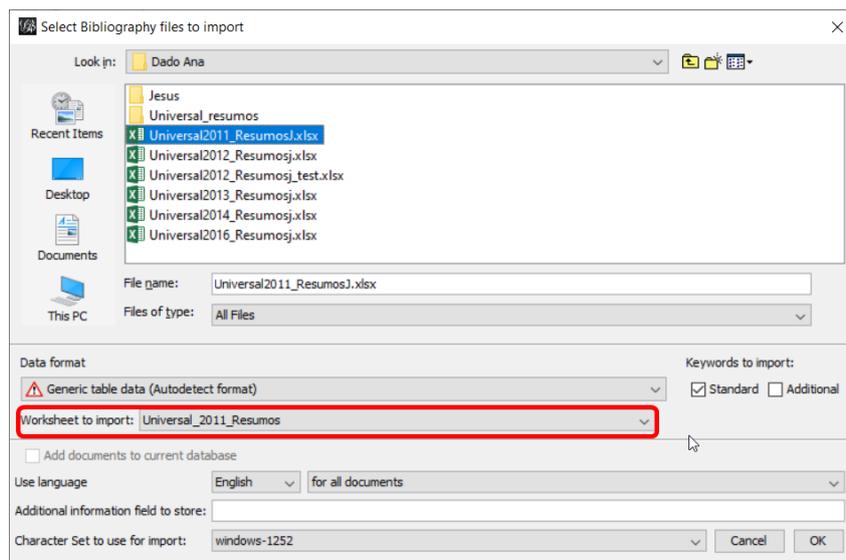


Figura 7.5: Seleção da aba da planilha excel no formato “*Generic Table data (Autodetect format)*”

A importação de dados genéricos presume os seguintes pré-requisitos:

- Os dados a serem importados constam em uma única aba da planilha
- A primeira linha da planilha contém os cabeçalhos que descrevem o conteúdo das colunas
- A partir da segunda linha da planilha, cada linha contém exatamente uma referência bibliográfica
- O nome da coluna, como especificado na primeira linha, determina o comportamento da importação, de acordo com a tabela em seguida. O nome deve ser escrito exatamente como especificado em seguida (observando letras maiúsculas e minúsculas) para obter o comportamento especificado.
- As colunas adicionais, que possuem um nome que não consta na tabela em seguida serão importadas como atributos dos nós no Gephi
- Colunas da planilha que não possuem nome na primeira linha são ignoradas

Nome da coluna	Comportamento da importação
<i>Title</i>	Título da publicação, usado para a criação da rede de similaridade semântica
<i>Abstract</i>	Resumo da publicação, usado para a criação da rede de similaridade semântica
<i>Keywords</i>	Contém a lista de palavras-chave da publicação, separadas por ponto-e-vírgula
<i>ID</i>	Contém um identificador único da publicação, usado para desduplicar referências bibliográficas durante a importação
<i>DOI</i>	Contém o <i>Document Object Identifier</i> , também usado para a desduplicação
<i>Language</i>	Contém o idioma da publicação.
<i>Authors</i>	Contem os autores da publicação, separados por ponto-e-vírgula
<i>Year</i>	Contém o ano da publicação, usado para filtrar as publicações na criação da rede de similaridade semântica
<i>Document type</i>	Contém o tipo de documento
<i>Source title</i>	Contém o título da revista, do livro ou do evento em que foi publicado o elemento

Formato “*Use external descriptor file*”

Para importar dados em formatos que não estão cobertos por uma das opções anteriores, o *CGEE Insight Net* oferece com esta opção a possibilidade de especificar um formato em um arquivo externo. A especificação deste arquivo consta na `descriptors`.

7.1.2 Palavras-chave a serem importadas

Os serviços *Web of Science*® e *Scopus*® publicam as palavras-chave definidas pelo autor, bem como palavras-chave adicionais, extraídas pelas equipes das duas empresas a partir dos dados da publicação e do seu contexto. A opção “*Keywords to import*” permite a definição de quais delas serão importadas:

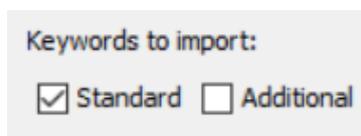


Figura 7.6: Seleção de palavras-chave a serem importadas

7.1.3 Apagar ou manter os dados do banco antes da importação

A opção “*Add documents to current database*” diferencia entre uma importação inicial e uma importação incremental.

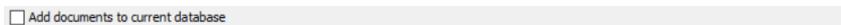


Figura 7.7: Opção de importação inicial ou incremental

Se essa opção for selecionada, as referências bibliográficas importadas serão acrescentadas às informações já existentes na base. Caso uma referência bibliográfica já exista na base, a versão importada complementa as informações existentes.

Se a opção não for selecionada, todos os dados que já existem no banco de dados serão apagados antes da importação. Desta forma, os dados importados substituem os dados atuais.

7.1.4 Idioma

Na parte inferior do diálogo de importação aparecem as definições que determinam o tratamento das referências bibliográficas em relação ao idioma.

Geralmente, as publicações possuem um campo que determina o idioma da publicação. Entretanto, na prática, a maioria dos dados de títulos e resumos das duas plataformas é disponibilizada em inglês, mesmo que o idioma original da publicação seja outro.

Para tratar esses casos, o diálogo de importação permite especificar o idioma padrão para as publicações sem idioma definido ou até para todos os documentos importados:



Figura 7.8: Definição do idioma padrão

7.1.5 Campo adicional de informação

Cada referência bibliográfica importada é representada como um nó no grafo criado. Esses nós possuem atributos, tais como o identificador DOI (*Digital Object Identifier* – vide <http://www.doi.org/>) da publicação (atributo “DOI”), o seu título e outros. O atributo “info” dos nós é preenchido com o valor especificado no campo “Additional information field to store” durante a importação.

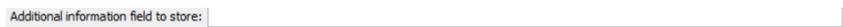


Figura 7.9: Campo adicional de informação

7.1.6 Conjunto de caracteres

A última opção do diálogo especifica o conjunto de caracteres (character set) do(s) arquivo(s) a ser(em) importado(s). Contrário aos arquivos de Currículos Lattes, essa informação não está contida dentro de arquivos BibTex e precisa ser fornecida pelo usuário.

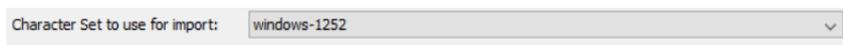


Figura 7.10: Definição do conjunto de caracteres

Os character sets mais usados são “Windows-1252”, para arquivos nativos do Windows e “UTF-8”, para arquivos que foram baixados da internet. A seleção do *character set* errado se manifesta em erros nas letras acentuadas durante a visualização do título e do resumo da referência bibliográfica.

7.1.7 Processo de importação

Durante a importação, o *CGEE Insight Net* mostra uma barra de progresso. Recomenda-se verificar a quantidade de referências bibliográficas no final da importação a partir da estatística do banco de dados (ver Seção 9.3).

Adicionalmente, o protocolo de execução (ver Seção 9.4) registra informações sobre o andamento da importação, de acordo com o grau de detalhe especificado na tela de configuração (ver Seção 3).



Figura 7.11: Importação de referências bibliográficas

7.2 Formação da rede

Depois da importação das referências bibliográficas na base de dados, a rede é formada a partir das pesquisas por similaridade contextual. Os passos 2-4 da [Tabela 4.1](#) são realizados em uma única operação, transformando o conteúdo do banco de dados em um grafo.

Para formar a rede, o usuário deve clicar em *Plugins > CGEE Insight Net Bibliography > Build new bibliography network* e preencher ou confirmar os dados do diálogo que é exibido:

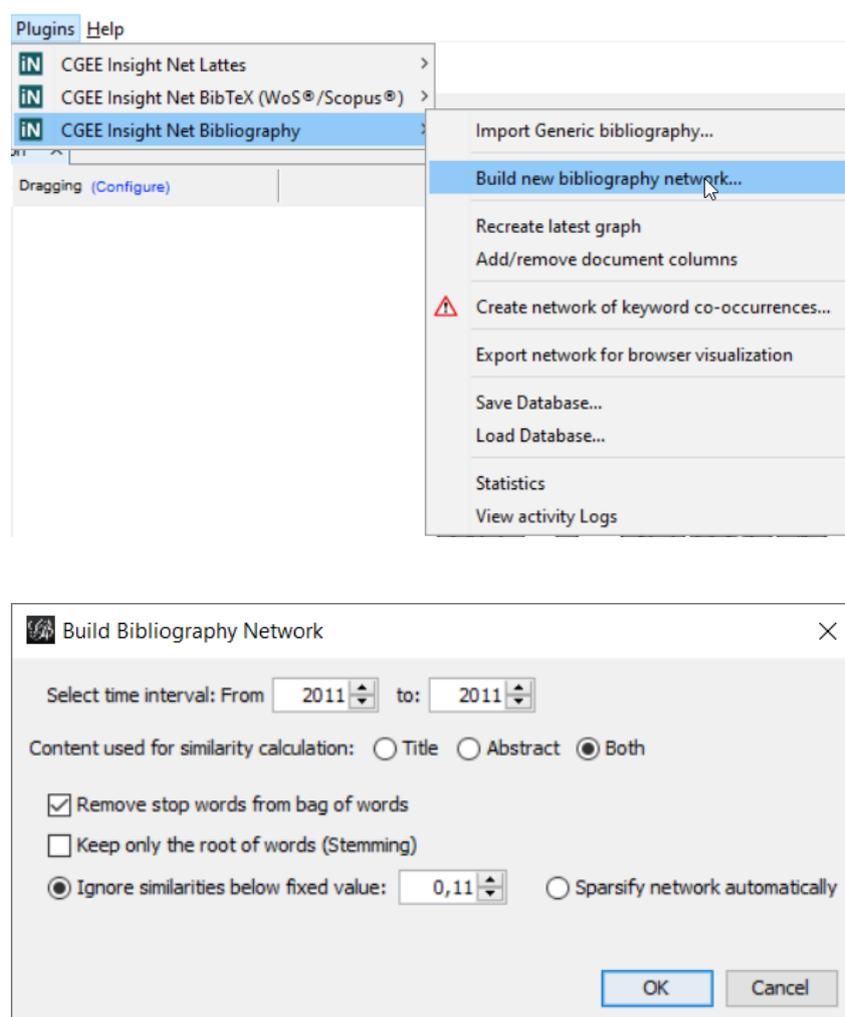


Figura 7.12: Menu e diálogo da formação da rede de referências bibliográficas

As opções do diálogo são explicadas em seguida.

7.2.1 Escopo da rede formada

Na parte superior do diálogo o usuário especifica quais publicações farão parte do escopo da formação da rede. No módulo de referências bibliográficas genéricas, o único critério é o intervalo de anos de publicação.

Destaca-se novamente que a rede de referências bibliográficas será montada apenas para as contribuições selecionadas.

7.2.2 Conteúdo considerado para o cálculo de similaridade contextual

A parte inferior do diálogo permite a seleção das opções da pesquisa por similaridade semântica.

- A seleção “*Content used for similarity calculation*” determina se a pesquisa por similaridade semântica considera apenas o título da referência bibliográfica (“*Title*”), apenas o resumo (“*Abstract*”) ou ambos (“*Both*”).
- O usuário pode selecionar se os pré-processamentos dos termos “*Stop words*” e “*Stemming*” serão realizados ou não. Esses dois algoritmos dependem da definição correta do idioma da referência bibliográfica.
 - *Stop Words* são as palavras mais frequentes de cada idioma, que não agregam informação aos termos identificados e serão eliminados da pesquisa. Os stop words são implementados apenas para as referências bibliográficas em Inglês e Português.
 - O *Stemming* reduz, em um algoritmo específico por idioma, cada palavra a uma raiz que desconsidera flexões gramaticais. Nesse momento, apenas os idiomas Português e Inglês são tratados pelo stemming. Referências bibliográficas em outros idiomas permanecem na forma original.
- O cálculo da similaridade semântica gera arestas entre praticamente qualquer par de referências bibliográficas, a maioria com baixos valores de similaridade, que não agregam informações relevantes ao conteúdo do gráfico. Por esse motivo, existem três métodos para reduzir a quantidade de arestas na rede:
 - “*Ignore similarities below fixed value*”: valores abaixo de um limite especificado podem ser desconsiderados, produzindo o valor final zero como similaridade semântica.
 - “*Sparsify network automatically*”: Um algoritmo automático [7] ainda experimental é utilizado para reduzir a quantidade de arestas na rede.
 - Para o cálculo de similaridade podem ser considerados apenas os termos mais relevantes das contribuições. Esta configuração é feita no diálogo de opções do CGEE Insight Net (ver [Seção 3](#)). Se um percentil de relevância dos termos for definido, uma mensagem correspondente é exibida:

NOTE: The least relevant 20% of all term occurrences will be ignored, as set in Tools/Options/CGEE/Network

Figura 7.13: Aviso sobre configuração de limite inferior de relevância

8.1 Filtragem dos resultados

O *CGEE Insight Net* define quatro filtros que modificam a exibição do grafo, eliminando informações específicas determinadas pelo usuário. Esses filtros são exibidos na aba “*Filters*” do Gephi, na categoria “CGEE”:

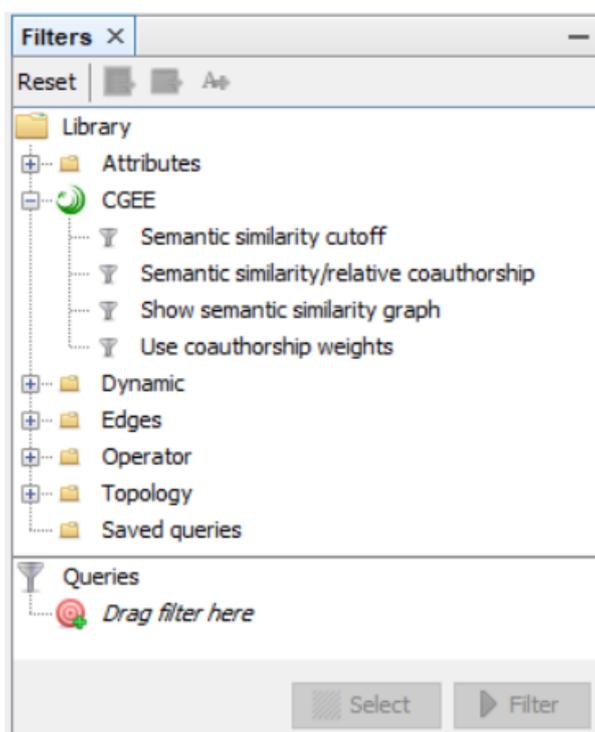


Figura 8.1: Filtros da categoria “CGEE”

Alguns desses filtros são relevantes apenas para redes que possuem arestas de coautoria e também de similaridade semântica. Outros podem ser usados em ambos os tipos de redes.

Para escolher um dos filtros, o mesmo deve ser selecionado pelo usuário com clique duplo. Existem dois tipos de aplicar o filtro:

1. O filtro é aplicado com um clique no botão “*Filter*”. Desta forma, a rede muda de acordo com o filtro selecionado.
2. Existe ainda o botão “*Select*”, que **destaca** os elementos da rede que atendem ao critério do filtro. Entretanto, esta funcionalidade é usada, principalmente, com os filtros padrão do Gephi. Para os filtros específicos do CGEE, este botão não agrega valor.

8.1.1 Filtro “*Show semantic similarity graph*”

Esse filtro não possui parâmetros de configuração e tem efeito apenas em redes que possuem ambos os tipos de arestas (coautoria e similaridade semântica). Ele elimina todas as arestas verdes e transforma as arestas pretas em vermelhas ao exibir o grafo. Como peso das arestas, nesse caso, é usada exclusivamente a similaridade semântica.

8.1.2 Filtro “*Semantic similarity/relative coauthorship*”

Esse filtro também traz resultados apenas em redes que possuem arestas dos dois tipos (coautoria e similaridade semântica). Ele permite a definição de um intervalo de exibição das arestas, controlado pelo valor do atributo “*Semantic similarity/relative coauthorship*”. Arestas que representam colaborações baseadas na coautoria com similaridade semântica zero (“arestas verdes”) sempre carregam o valor zero nesse atributo. Arestas com zero coautorias, que apenas possuem similaridade semântica (“arestas vermelhas”) são consideradas com um valor alto, além do valor de qualquer aresta preta. O intervalo de exibição pode ser escolhido com dois marcadores. O gráfico abaixo desses marcadores representa um histograma de valores encontrados no grafo:

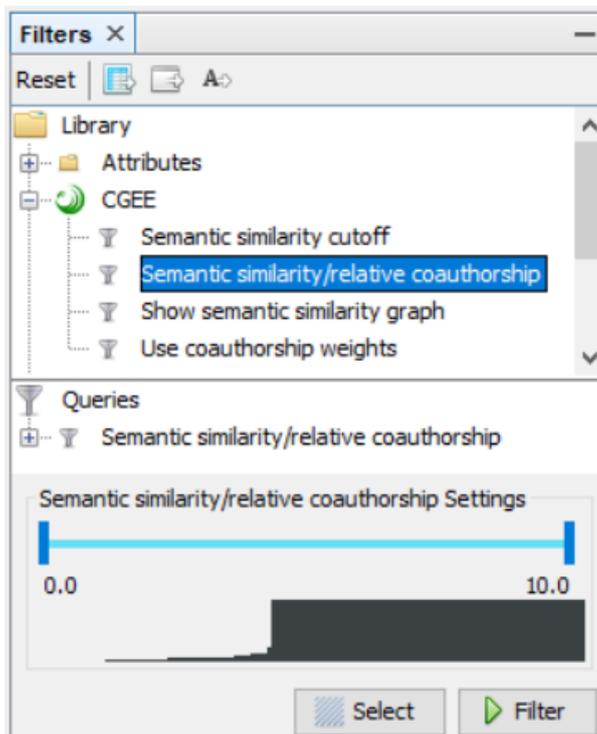


Figura 8.2: Parâmetros do filtro *Semantic similarity/relative coauthorship*

8.1.3 Filtro “Semantic similarity cutoff”

Esse filtro funciona em redes que possuem arestas do tipo “similaridade semântica” e desconsidera na exibição qualquer similaridade semântica abaixo do limite especificado pelo usuário.

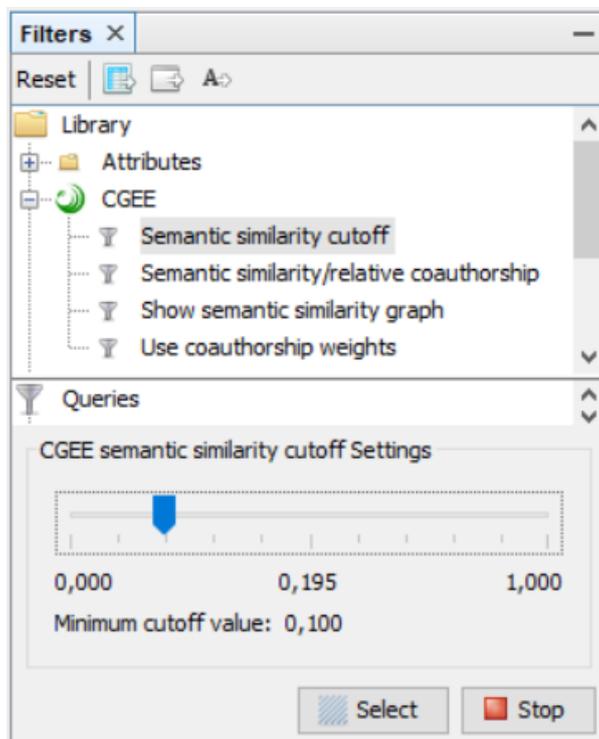


Figura 8.3: Parâmetro do filtro *Semantic similarity cutoff*

Aumentando o valor mínimo da similaridade semântica para o valor máximo possível (1.0), todas as arestas de similaridade semântica serão eliminadas e o grafo exibirá apenas as arestas de coautorias, caso existam.

8.1.4 Filtro “Use coauthorship weights”

Em grafos que possuem arestas com valores de similaridade contextual bem como de coautorias (“arestas pretas”), normalmente se usa como peso a similaridade contextual/semântica. O filtro “*Use coauthorship weights*” permite, nesses casos, a aplicação das coautorias relativas como peso da aresta, conforme demonstram os seguintes diagramas:

Weight	Coauthorships	Relative coauthors...	Semantic similarity
0,177	1	0,104	0,177
0,308	1	0,104	0,308
0,337	1	0,104	0,337
0,479			0,479
0,585	5	0,269	0,585
0,104	1	0,104	0,108
0,108	1	0,104	0,108
0,794	22	0,479	0,794

Figura 8.4: Sem filtro “*Use coauthorship weights*”

Weight	Coauthorships	Relative coauthors...	Semantic similarity
0,104	1	0,104	0,177
0,104	1	0,104	0,308
0,104	1	0,104	0,337
0,479			0,479
0,269	5	0,269	0,585
0,104	1	0,104	
0,104	1	0,104	0,108
0,471	22	0,471	0,794

Figura 8.5: Com filtro “Use coauthorship weights”

8.2 Análise de clusters

As funcionalidades do Gephi e os *plug-ins* existentes permitem a determinação de “clusters”, grupos de pesquisadores que entre si possuem mais ligações do que com pesquisadores externos. O algoritmo mais utilizado no Gephi bem é representado pela estatística “Modularity”.

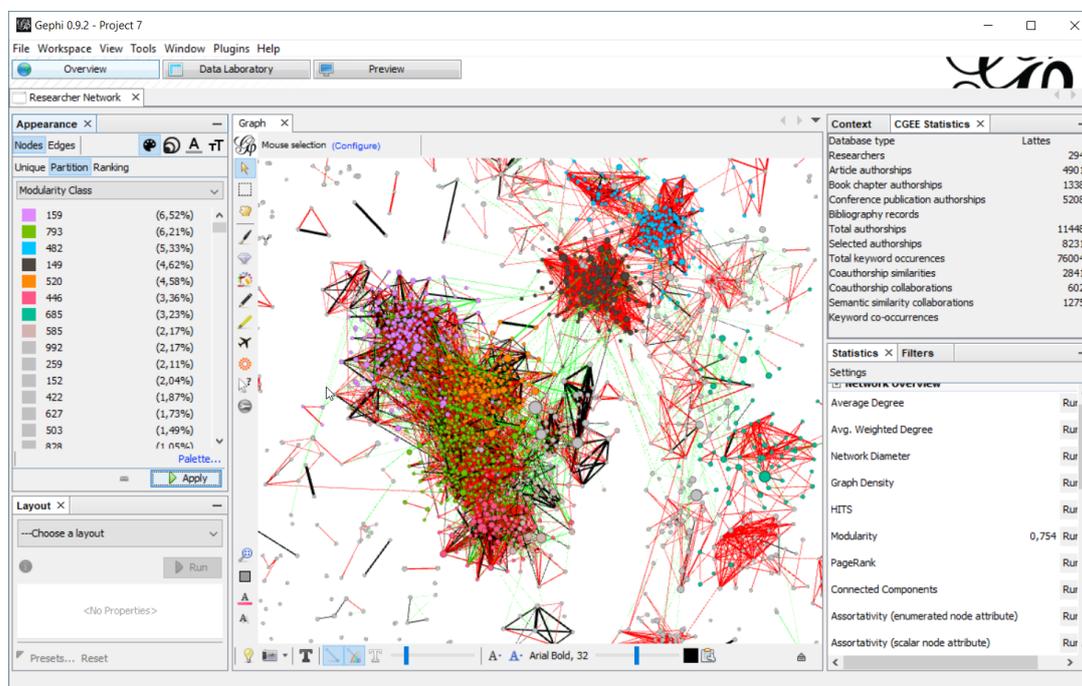


Figura 8.6: Clusters na rede gerada

8.3 Análise de assortatividade

A assortatividade (também conhecida como “homofilia”) de uma rede descreve a tendência da existência de uma aresta entre dois nós que possuem valores semelhantes em um atributo selecionado [Assort].

Por exemplo, em redes sociais, existem tendências fortes de estabelecer amizades dentro do mesmo nível educacional ou da mesma nacionalidade. Nesses casos, a assortatividade de uma rede social teria um valor alto positivo com relação aos atributos “nível educacional” ou “nacionalidade”.

Já em redes de relacionamento sexual, a preferência geralmente é pelo gênero oposto, levando a assortatividade com relação ao atributo “gênero” para um valor negativo.

Em redes não-assortativas, a existência da aresta não é correlacionada ao atributo selecionado e o valor da assortatividade em relação ao atributo escolhido é próximo de zero.

O *CGEE Insight Net* permite a análise da assortatividade em relação a um atributo selecionado a partir de duas estatísticas do Gephi no painel “Statistics”:

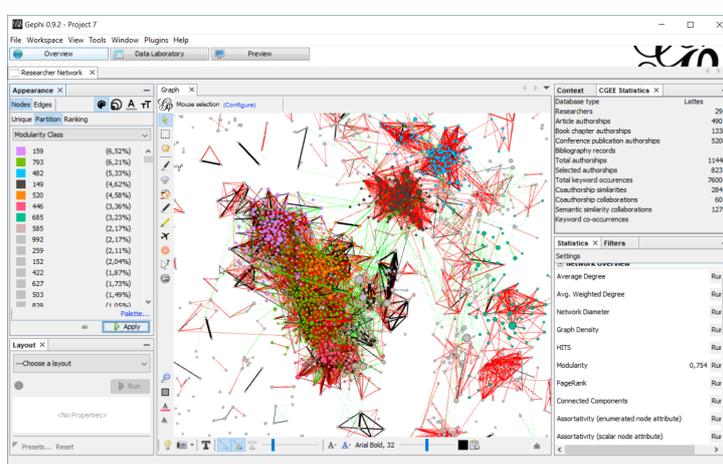


Figura 8.7: Estatísticas de Assortatividade

Dependendo do tipo de atributo, uma das duas estatísticas deve ser escolhida

- Para atributos enumerados (ou categóricos) deve ser escolhida a estatística “Assortativity (*enumerated node attribute*)”. Atributos enumerados são aqueles em que os valores não possuem ordem, tais como valores textuais ou categorias numeradas.
- Atributos numéricos que possuem uma ordem entre si podem ser avaliados com a estatística “Assortativity (*scalar node attribute*)”. São aqueles onde o valor representa uma medida numérica.

Para calcular a assortatividade, o usuário deve clicar no botão  da estatística selecionada e escolher o atributo com relação ao qual a assortatividade deve ser calculada:

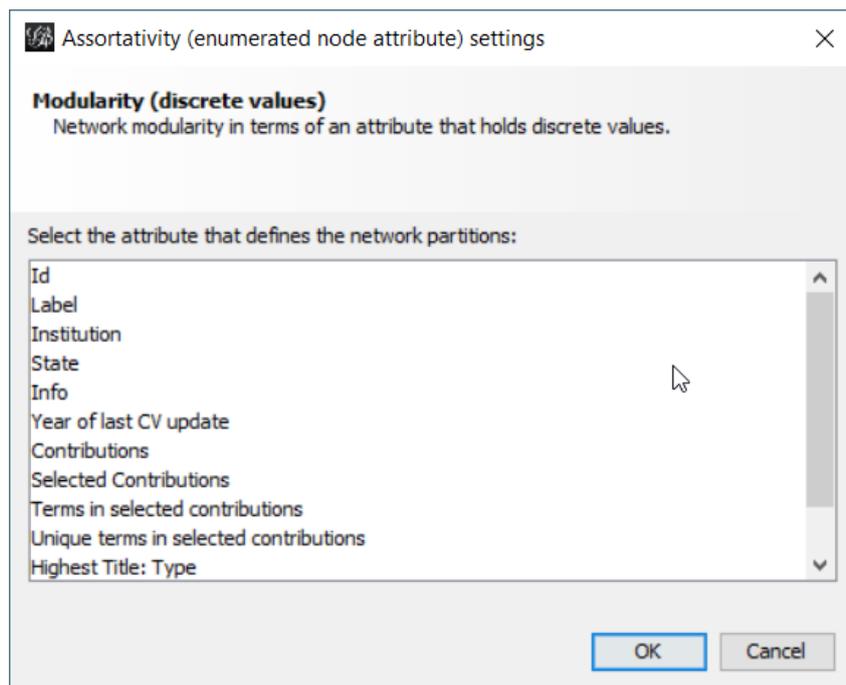
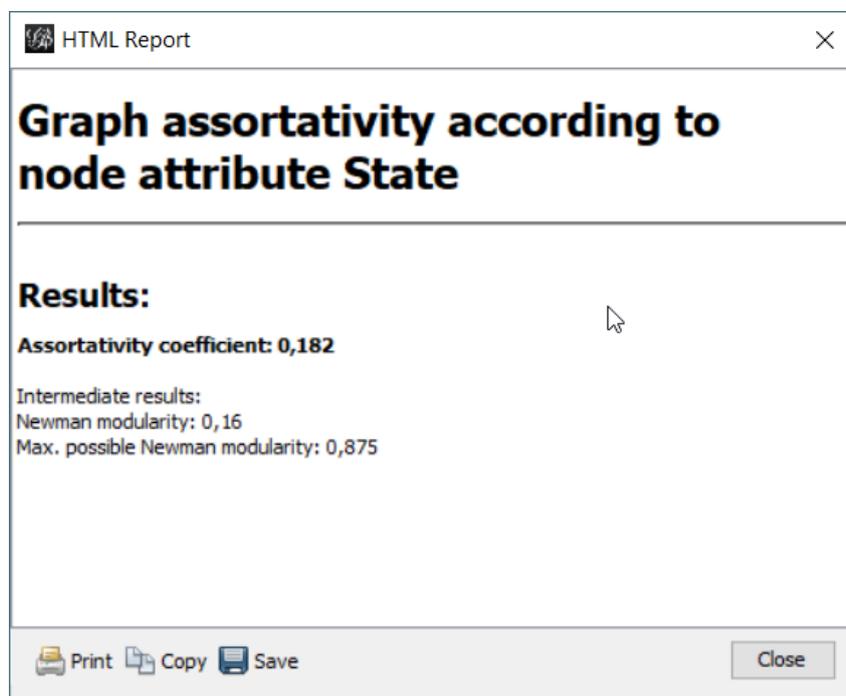


Figura 8.8: Seleção do atributo de assortatividade

Clicando em “OK”, a assortatividade é calculada e aparece no diálogo de resultado, bem como na lista de cálculos estatísticos do Gephi:



Metric	Value	Run	Status
Average Degree		Run	●
Avg. Weighted Degree		Run	●
Network Diameter		Run	●
Graph Density		Run	●
HITS		Run	●
Modularity	0,754	Run	ⓘ
PageRank		Run	●
Connected Components		Run	●
Assortativity (enumerated node attribute)	0,18	Run	ⓘ
Assortativity (scalar node attribute)		Run	●
Percolation estimators (EXPERIMENTAL)		Run	●

Figura 8.9: Resultados do cálculo de assortatividade

8.3.1 Estimador de percolação

Essa funcionalidade avançada é atualmente usada apenas para testes internos do CGEE. Os fundamentos matemáticos podem ser encontrados na literatura especializada [Percolation].

8.4 Análise das palavras-chave

Para aprofundar a análise das redes, as palavras-chave dos Currículos Lattes dos pesquisadores e das contribuições são importadas no banco de dados. Para redes bibliográficas, as palavras-chave são extraídas dos arquivos carregados.

Essas palavras-chave permitem, quando examinadas em conjunto, a identificação das áreas de trabalho dos pesquisadores representados pelo conjunto de seus currículos.

Para redes de referências bibliográficas, as palavras-chave caracterizam o conteúdo da publicação e permitem, em conjunto, uma estimativa geral dos conteúdos das publicações.

O CGEE Insight Net permite a visualização dessas palavras-chave, que constam **apenas** no seu banco de dados e não no grafo do Gephi. Desta forma, é essencial que o banco de dados seja **coerente** com o grafo. Incoerências podem surgir se um grafo Gephi for carregado por um arquivo “.gephi” ou “.gexf” e se o banco de dados não possuir o mesmo conteúdo que esse arquivo. Neste caso, sugere-se nova importação dos Currículos Lattes ou das referências bibliográficas no banco de dados, nova geração do grafo ou a recuperação do grafo a partir do banco de dados, conforme descrito na [Seção 9.1](#).

Para realizar a pesquisa de palavras-chave, o usuário deve clicar no símbolo  na barra lateral da tela “Graph” do Gephi:

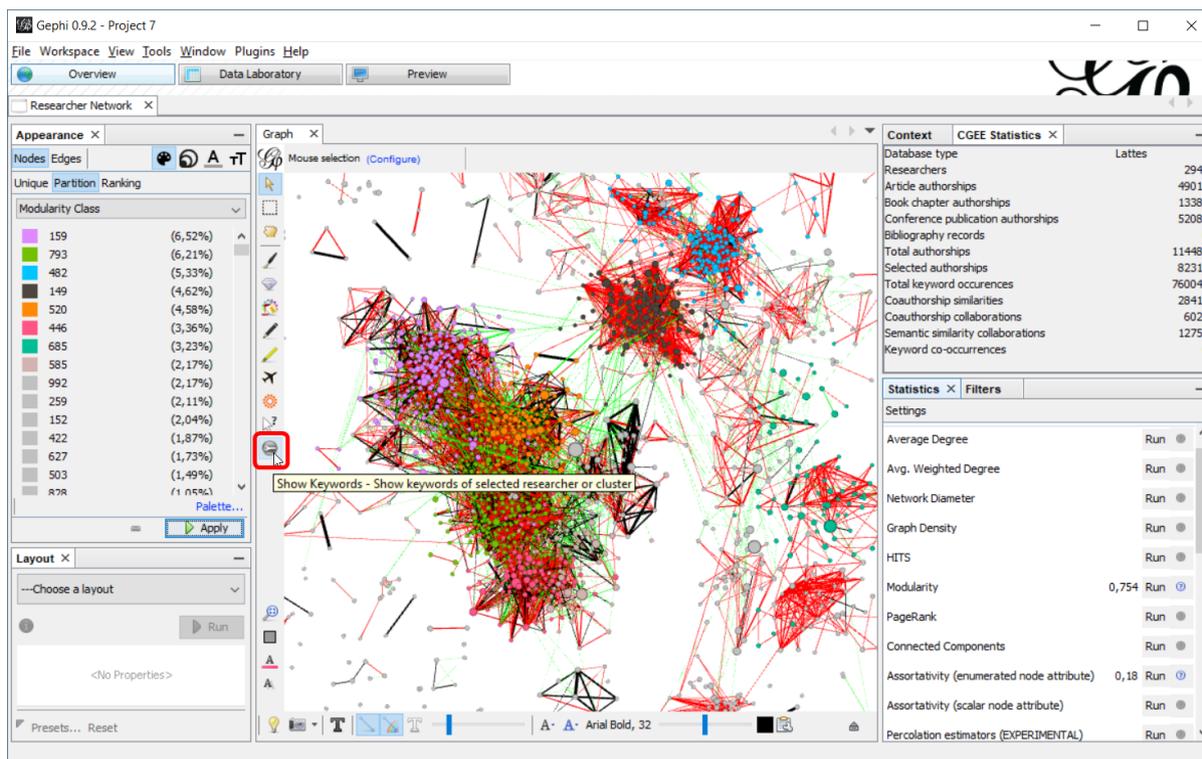


Figura 8.10: Seleção da funcionalidade “Palavras-chave”

Clicando nesse símbolo e selecionando nós no grafo, aparece a janela “CGEE Keywords”, que mostra as palavras-chave do nó selecionado e as frequências (quantidades de ocorrências) de cada uma:

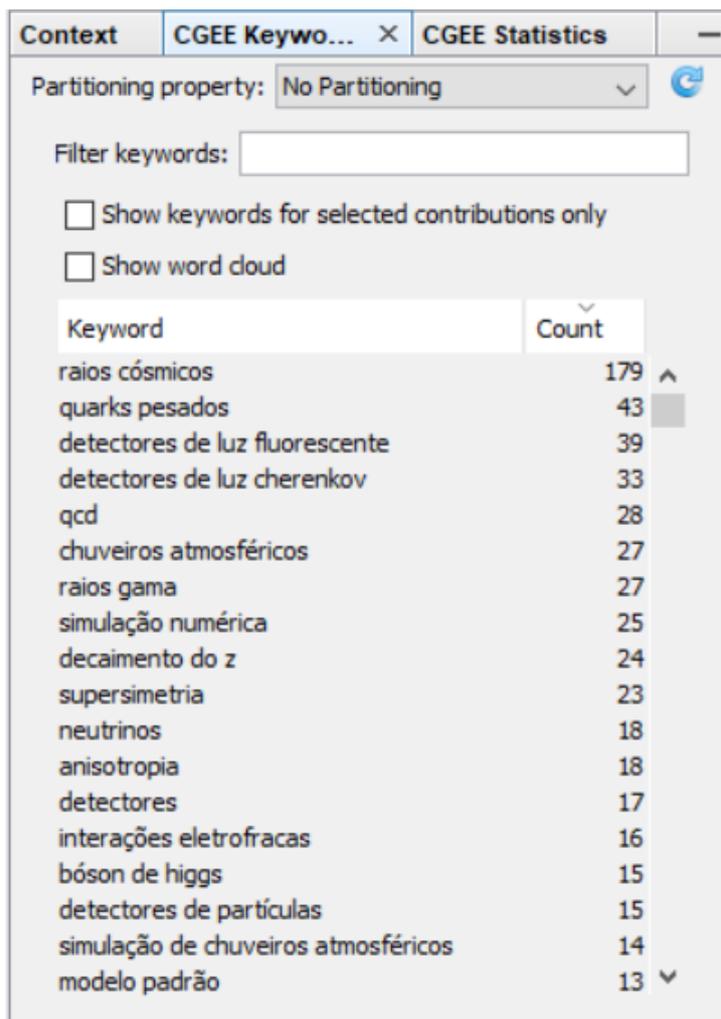


Figura 8.11: Janela de palavras-chave com frequências

Essa janela permite algumas configurações que serão explicadas em seguida.

8.4.1 Filtragem das palavras-chave

A lista de palavras-chave pode ser filtrada para exibir apenas alguns dos resultados. Existem dois tipos de filtros, descritos em seguida. Em ambos os casos, o texto exibido na lista muda de cor e aparece na tela o número de palavras-chave que são eliminadas da lista pelo filtro.

Filtragem por termo digitado

O usuário pode digitar um texto na caixa em cima da lista. Nesse caso, apenas as palavras-chave que contêm o texto digitado aparecem na lista:



Figura 8.12: Filtragem de palavras-chave por termo

Filtragem por contribuições selecionadas

Essa função se aplica apenas em redes de Currículos Lattes, em que existem vários elementos que permitem especificar palavras-chave:

- Na descrição da formação,
- Nas atividades de pesquisa e desenvolvimento,
- Nas produções científicas (artigos, capítulos de livros, trabalhos em eventos),
- Nas orientações de graduação, mestrado e doutorados
- Outros

A exibição de todas as palavras-chave de um único pesquisador permite uma visão global das áreas de atuação contribuindo para uma avaliação rápida do conteúdo semântico integrado de toda a produção do pesquisador. Por outro lado, considerando que a identificação das coautorias, da similaridade contextual e o agrupamento em clusters utilizam apenas as contribuições selecionadas durante a pesquisa de similaridade, é razoável exibir apenas as palavras-chave dessas contribuições selecionadas.

A opção “*Show only keywords for selected contributions*” permite alternar entre as duas formas de exibição descritas:

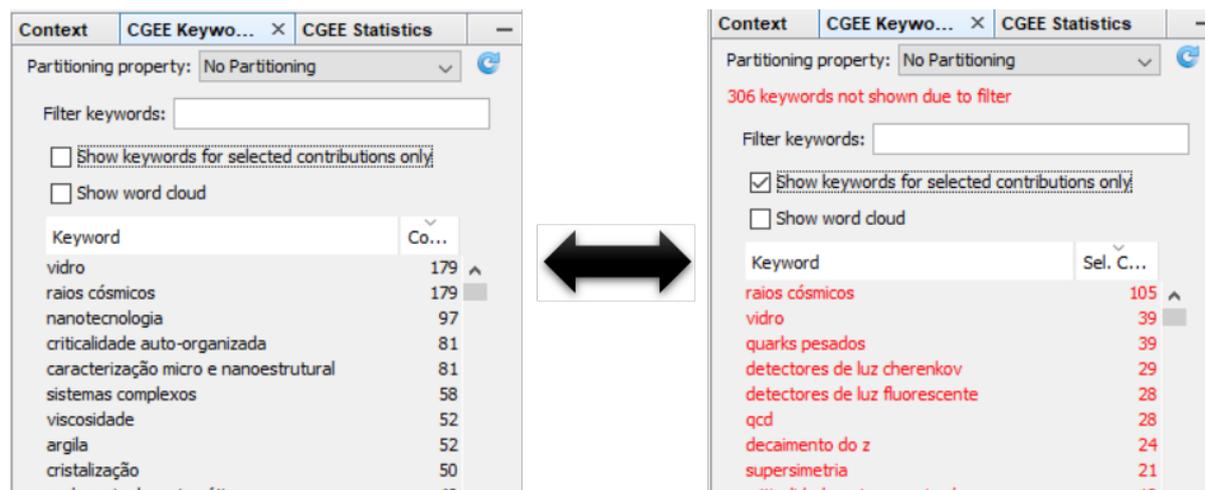


Figura 8.13: Filtragem das palavras-chave pelas contribuições selecionadas

8.4.2 Palavras-chave por nó ou por cluster de nós

A lista “*Partitioning property*” na parte superior permite selecionar se a janela exibe apenas as palavras-chave do nó selecionado (“*No clustering*”) ou as palavras-chave de todos os nós de um grupo definido (partição). Para determinar a partição, um atributo numérico, ou os atributos “*Info*” ou “*Institution*”, pré-definidos nos nós, devem possuir o mesmo valor para todos os membros do grupo. Casos particulares de partições importantes são os agrupamentos calculados por diferentes métodos no Gephi. Nesses casos, a partição é definida a partir das arestas entre os nós, sendo, portanto um atributo pós-processado. Os algoritmos de agrupamento (*clustering*) do Gephi descritos na Seção 8.2 usam atributos diferentes para especificar o número do cluster e o usuário precisa selecionar aquele mais adequado à sua análise.

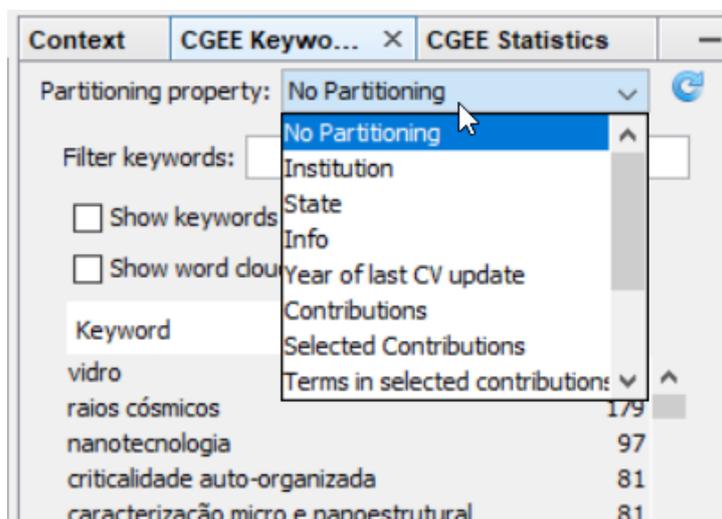


Figura 8.14: Seleção do atributo que define a partição dos nós

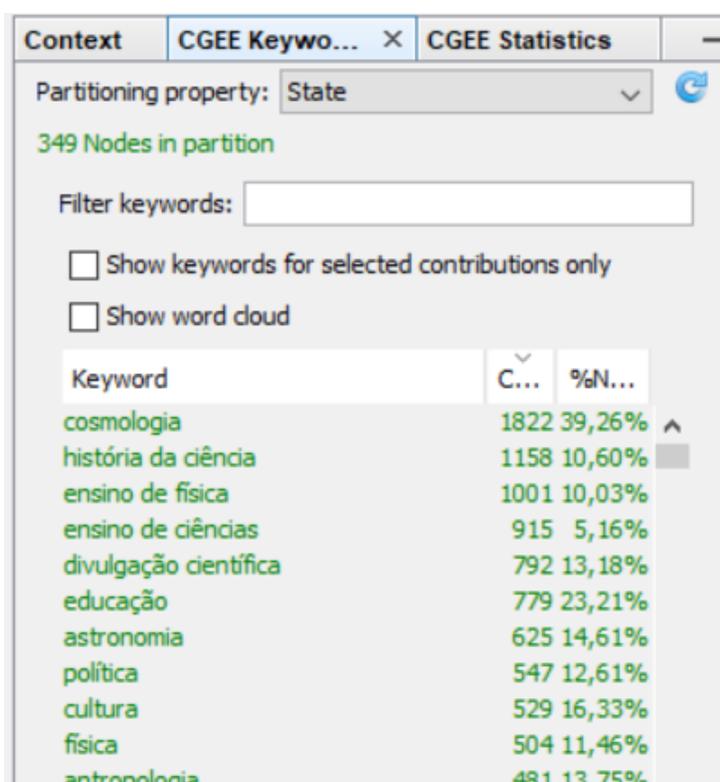
Depois de executar um algoritmo de *clustering*, a lista precisa ser atualizada manualmente, clicando no símbolo . Além de todos os valores numéricos integrais, a lista de possíveis atributos de particionamento mostra os seguintes atributos em redes de Currículos Lattes:

- Todos os atributos numéricos integrais

- *Info*
- *Institution*
- *State*
- *Gender*
- Todas as informações sobre a formação dos pesquisadores

Alguns dos valores numéricos não representam clusters, no sentido de agrupamento, como por exemplo o número de contribuições, mas que ainda são partições.

Se a opção “*Partitioning property*” for ativada, a lista mostra todas as palavras-chave da partição da qual o nó selecionado pertence, junto com sua quantidade de nós. Para destacar o fato que a lista mostra as palavras-chave de uma partição e não do nó individual, as palavras-chave são exibidas em verde. A lista mostra ainda a porcentagem de nós em que cada palavra-chave é encontrada:



Keyword	C...	%N...
cosmologia	1822	39,26%
história da ciência	1158	10,60%
ensino de física	1001	10,03%
ensino de ciências	915	5,16%
divulgação científica	792	13,18%
educação	779	23,21%
astronomia	625	14,61%
política	547	12,61%
cultura	529	16,33%
física	504	11,46%
antropologia	481	13,75%

Figura 8.15: Palavras-chave de um cluster de 349 nós

Nesse exemplo, a palavra-chave mais frequente (“cosmologia”) tem 1.822 ocorrências em 137 nós (39,26% de 349 nós).

A filtragem dessa lista de palavras-chave por partição, de acordo com a [Seção 8.4.1](#) exibe as palavras-chave na cor laranja:

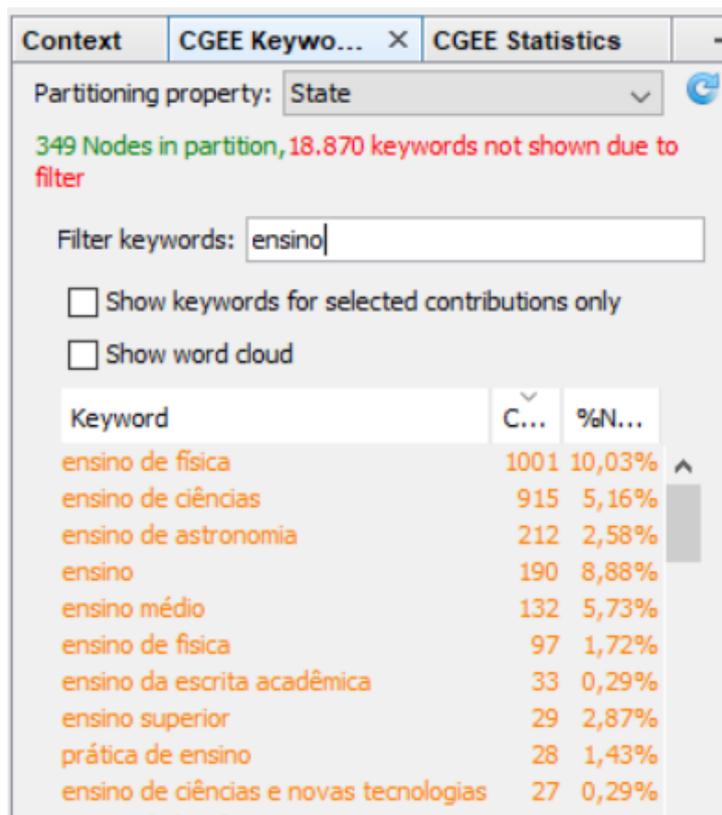


Figura 8.16: Lista de palavras-chave por partição, filtrada

Conforme descrito na Seção 3, pode ser calculada a relevância das palavras-chave por partição. Caso a referida opção tenha sido selecionada na tela de configuração, sempre ressaltando tratar-se de uma funcionalidade experimental, a relevância aparece como coluna adicional na lista de palavras-chave:

Keyword	Co...	Releva...	%Nodes
cosmologia	1822	1703,7	39,26%
história da ciência	1158	2598,7	10,60%
ensino de física	1001	2302,0	10,03%
ensino de ciências	915	2712,7	5,16%
divulgação científica	792	1604,9	13,18%
educação	779	1137,8	23,21%
astronomia	625	1202,0	14,61%
política	547	1132,8	12,61%
cultura	529	958,6	16,33%
física	504	1091,8	11,46%
antropologia	481	954,2	13,75%
filosofia	475	942,3	13,75%
ciência	436	893,1	12,89%

Figura 8.17: Relevância das palavras-chave exibidas na lista

8.4.3 Seleção de nós a partir das palavras-chave

Conforme descrito na Seção 8.4, a lista de palavras-chave é exibida a partir da seleção de um ou mais nós e o conteúdo mostrado depende da configuração. Como funcionalidade adicional, itens da lista de palavras-chave podem ser selecionados e, nesse caso, todos os nós que usaram esses itens são destacados no grafo:

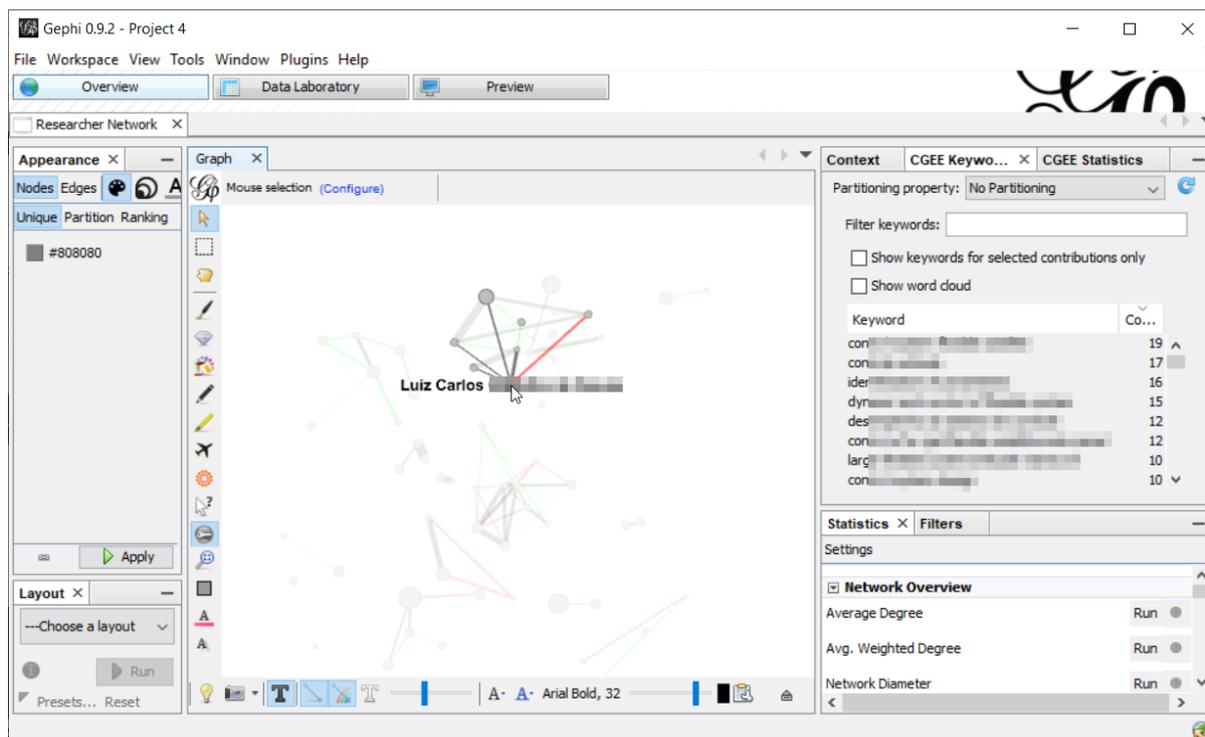


Figura 8.18: Exibição das palavras-chave do nó selecionado

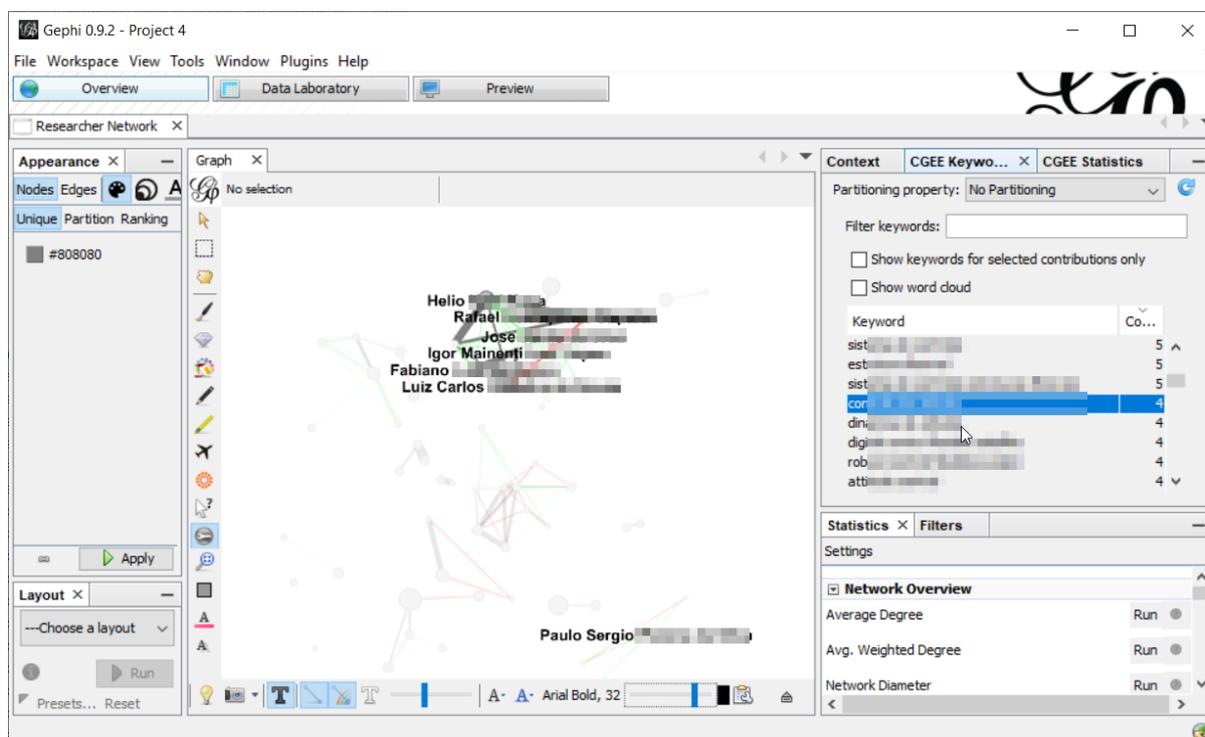


Figura 8.19: Exibição dos nós que usam a palavra-chave selecionada

A lista permite a seleção de uma única palavra-chave com um clique do botão esquerdo do mouse. Seguindo o botão esquerdo, várias palavras-chave podem ser selecionadas, passando o mouse em cima dos nós. A mesma funcionalidade é obtida com um clique na primeira e na última palavra-chave, segurando a tecla Shift. Finalmente, várias palavras-chave podem ser selecionadas e desselecionadas independentemente.

mente uma da outra, clicando nelas e segurando a tecla Ctrl.

8.4.4 Funcionalidades adicionais da lista de palavras-chave

Um clique com botão direito na lista de palavras-chave mostra a seguinte lista de funcionalidades adicionais:

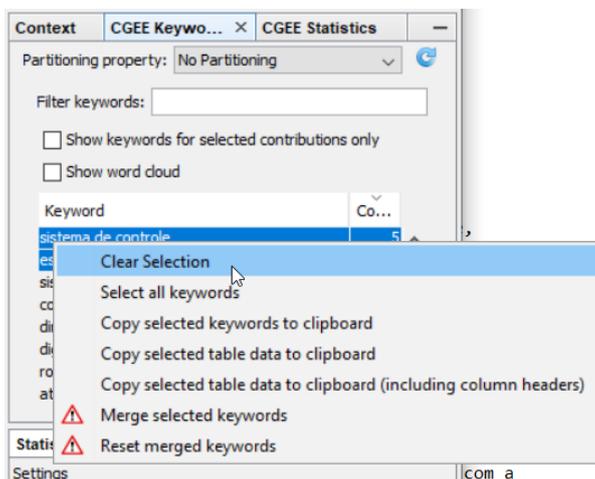


Figura 8.20: Funcionalidades adicionais na lista de palavras-chave

Os primeiros dois itens (“*Clear selection*” e “*Select all keywords*”) permitem retirar seleções prévias ou selecionar todos os itens da lista, considerando a funcionalidade descrita na seção anterior.

O terceiro item “*Copy selected keywords to clipboard*” copia as palavras-chave selecionadas para a área de transferência do sistema operacional, da qual elas podem ser coladas em programas de edição de textos e tabelas. O quarto item “*Copy selected table data to clipboard*” acrescenta as frequências e os outros valores numéricos exibidos. O próximo item “*Copy selected table data to clipboard (including column headers)*”, além da funcionalidade anterior, grava uma primeira linha com os nomes das colunas.

O item “*Merge selected keywords*” aparece apenas se mais que uma palavra-chave for selecionada na lista. Este item permite a junção de várias palavras-chave sinônimas. Selecionando essas palavras-chave e clicando em “*Merge selected keywords*”, o seguinte diálogo aparece na tela:

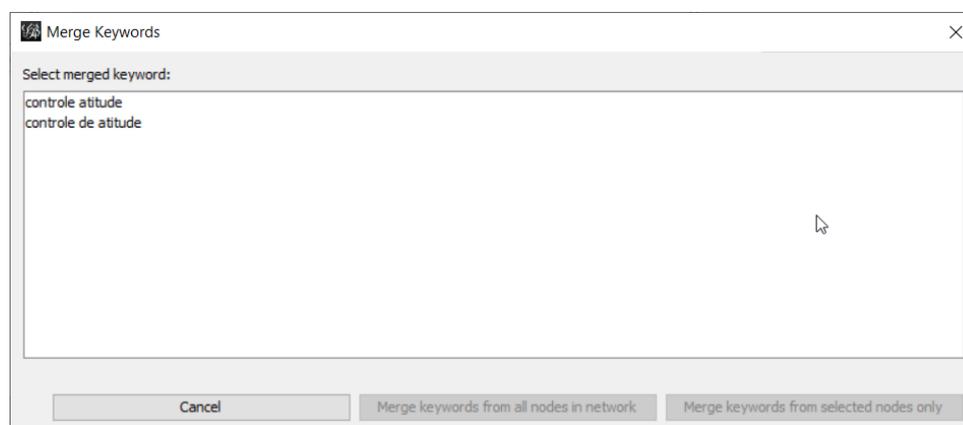


Figura 8.21: Diálogo de junção de palavras-chave

O usuário precisa selecionar qual das palavras-chave será a palavra-chave juntada. Todas as outras palavras-chave serão eliminadas:

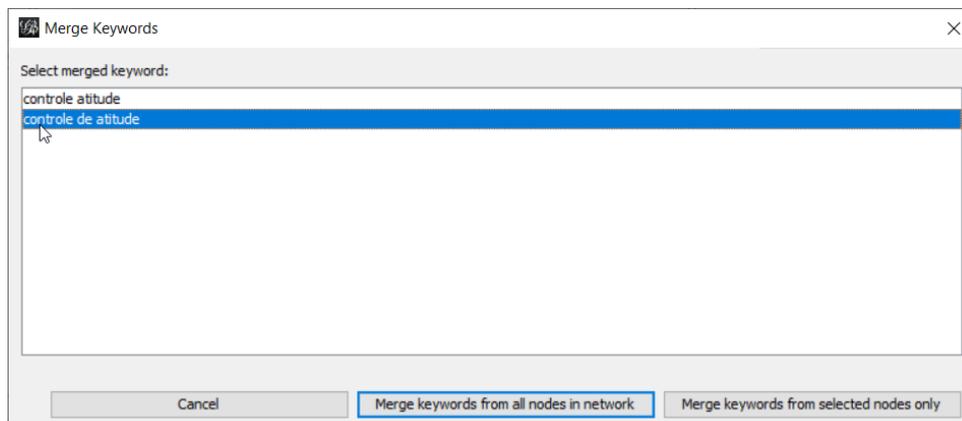


Figura 8.22: Seleção da palavra-chave juntada

Finalmente, para realizar a junção das palavras-chave, o usuário precisa determinar se serão juntadas apenas as palavras-chave que ocorrem nos nós selecionados (“*Merge keywords from selected nodes only*”) ou se a operação deve ser realizada em todos os nós da rede (“*Merge keywords from all nodes in network*”).

O item “*Reset merged keywords*” desfaz **todas** as operações de palavras-chaves juntadas e repõe a lista no estado original.

8.4.5 Nuvem de palavras-chave

A nuvem de palavras-chave exibe o conjunto de palavras-chave em uma única visualização em que o tamanho da palavra corresponde à frequência, à relevância ou à porcentagem de nós que contêm essa palavra.

Para exibir a nuvem de palavras, o usuário deve selecionar o item “*Show Word Cloud*” na janela de palavras-chave. Neste caso, aparece uma nova aba que mostra a nuvem de palavras:

Please enlarge window.” Caso contrário, aparece a mensagem “*Complete. xx% of relevant keywords shown*”. É importante notar que a nuvem sempre é montada na sequência decrescente dos valores do critério selecionado (frequência, relevância ou porcentagem de nós). Entretanto, palavras-chave de maior valor usam mais espaço e podem não mais caber na área disponível da nuvem, enquanto palavras de menor valor ocupam menos espaço e assim podem ser incluídas.

8.5 Criação de redes de co-ocorrências de palavras-chave

Como funcionalidade adicional, o *CGEE Insight Net* permite a criação de redes de co-ocorrências de palavras-chave. Isso significa que existem arestas entre palavras-chaves que ocorrem juntos em um ou mais contribuições bibliográficas¹. Cada contribuição em que as ambas as palavras-chave ocorrem aumenta o peso da aresta entre as duas palavras-chave em um.

As redes de palavras-chave podem ser criadas por qualquer tipo de rede previamente criada no Gephi. A opção “*Create network of keyword co-occurrences*” abre o seguinte diálogo:

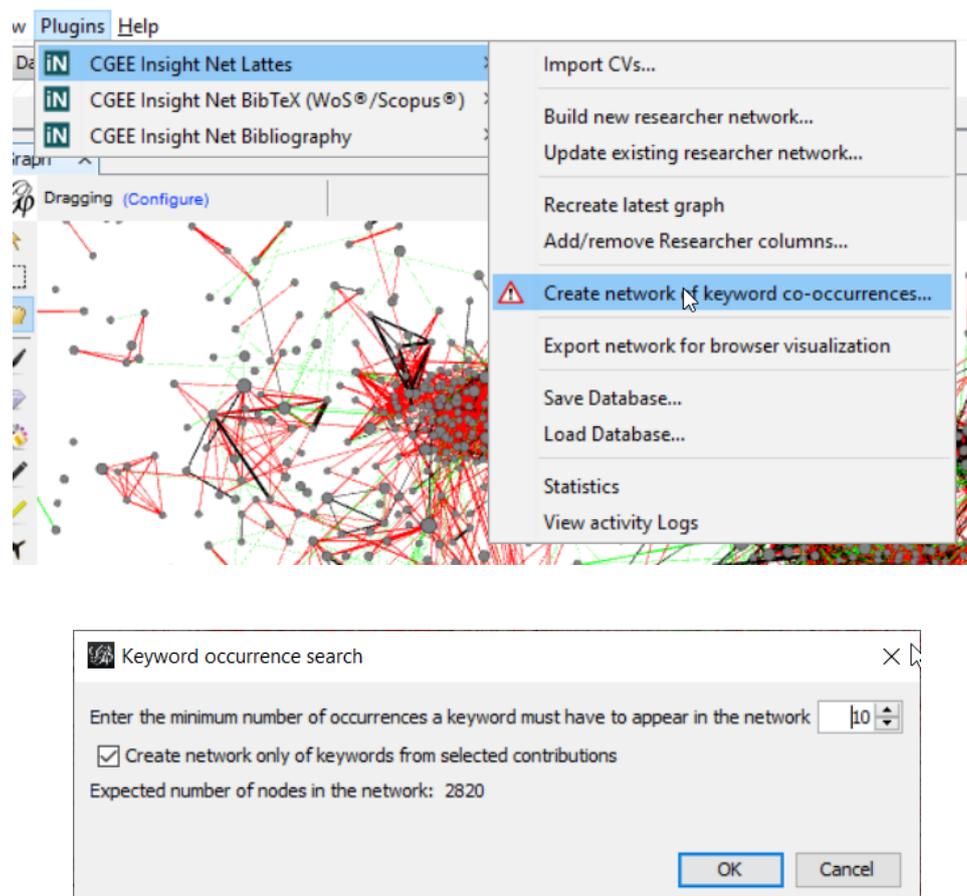


Figura 8.26: Funcionalidade para criar redes de palavras-chave

Como a quantidade de palavras-chave costuma ser alta, apenas as palavras-chaves com o maior número de ocorrências podem ser consideradas. Essa quantidade mínima de ocorrências de cada palavra-chave pode ser especificada.

A segunda opção refere-se ao escopo de extração de palavras-chave. Podem ser consideradas as palavras-chaves de **todas** as contribuições bibliográficas ou apenas aquelas que foram consideradas na criação da rede anterior (de pesquisadores ou de referências bibliográficas).

¹ *Contribuições bibliográficas* são as referências bibliográficas importadas pelos módulos “BibTeX” e “Bibliografia genérica”, bem como os artigos, capítulos de livros e trabalhos em eventos que constam nos Currículos Lattes.

8.6 Eliminação interativa de nós da rede e do banco de dados

Conforme descrito na seção [Seção 4](#), a rede de pesquisadores ou de referências bibliográficas é criada a partir do conteúdo de banco de dados. Isso significa que a eliminação de um nó com as funcionalidades do Gephi (clcando no nó com botão direito e selecionando “Delete”), o exclui apenas do grafo, mas **não** do banco de dados. Se a rede for calculada novamente ou recuperada a partir da função “Recreate latest graph” (ver [Seção 9.1](#)), o nó reaparece.

O *CGEE Insight Net* possui duas funcionalidades que permitem a exclusão interativa de nós do banco de dados, que podem ser executadas com um clique do botão direito no nó dentro da visualização do grafo ou no laboratório de dados:

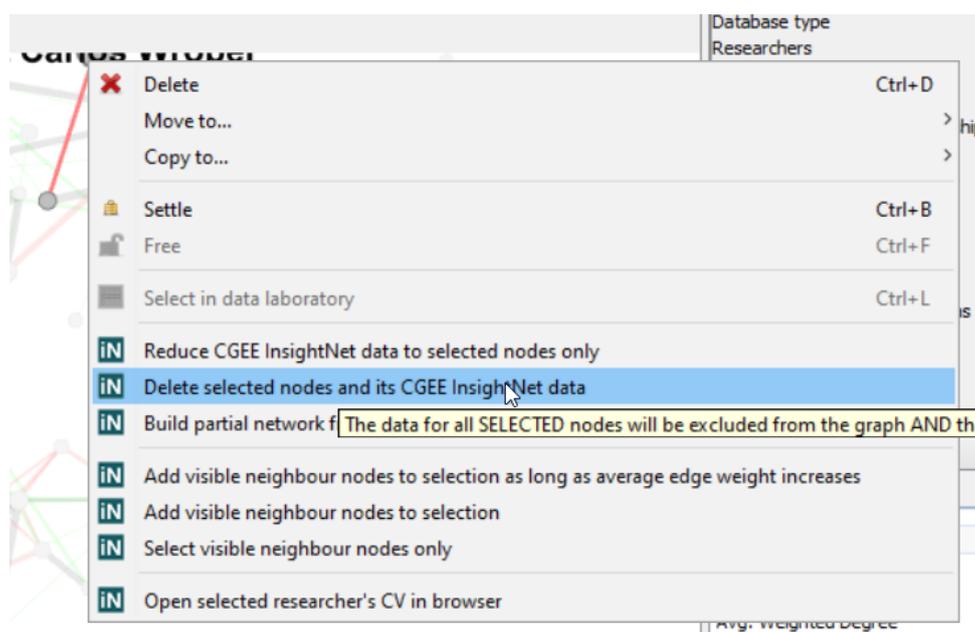


Figura 8.27: Eliminação interativa de nós do banco de dados na visualização do grafo

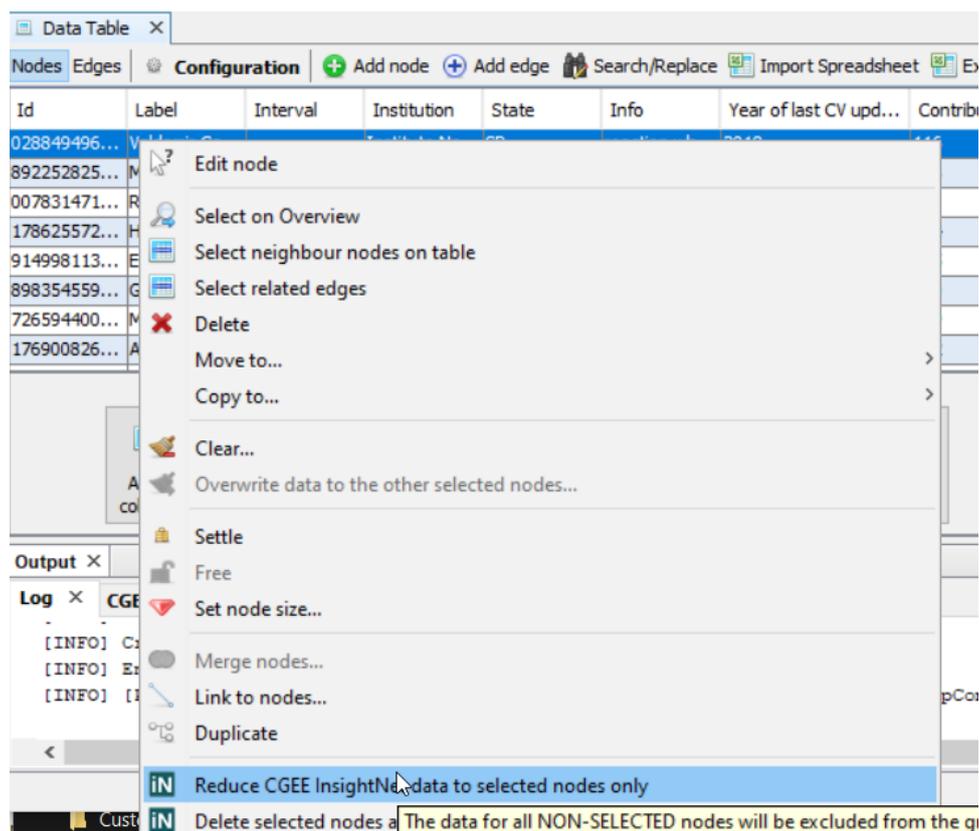


Figura 8.28: Eliminação interativa de nós do banco de dados no laboratório de dados

As duas funções “Reduce CGEE InsightNet data to selected Nodes only” e “Delete selected nodes and its CGEE InsightNet data” possuem objetivos complementares. Ambas requerem um conjunto de nós selecionados. Com um clique na função “Reduce CGEE InsightNet data to selected Nodes only”, apenas esses nós selecionados permanecem no banco de dados, todos os nós não selecionados serão eliminados do grafo e do banco de dados.

Já um clique na função “Delete selected nodes and its CGEE InsightNet data” elimina apenas os nós selecionados e os não selecionados permanecem no grafo e no banco de dados.

A eliminação de um ou mais nós da rede altera os pesos das arestas entre os nós e a rede de similaridade semântica precisa ser recalculada. Por esse motivo, todas as arestas da rede são eliminadas e a rede precisa ser recalculada.

8.7 Criação de uma nova rede a partir do subconjunto de nós selecionados

A funcionalidade “*Build partial network from selected nodes*”, também disponível com clique com botão direito após a seleção de vários nós na visualização do grafo ou no laboratório de dados, permite a criação de uma nova rede em que constam apenas os nós selecionados.

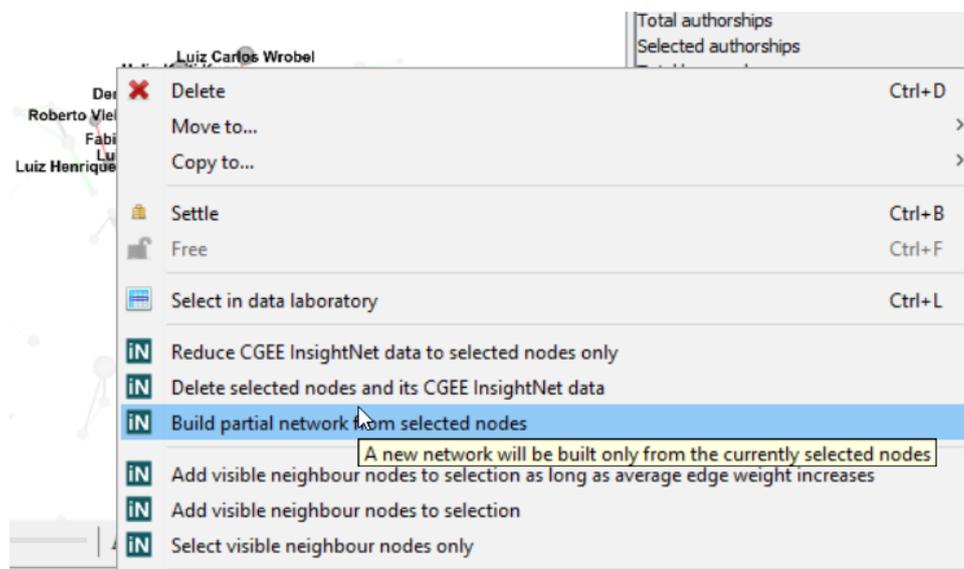


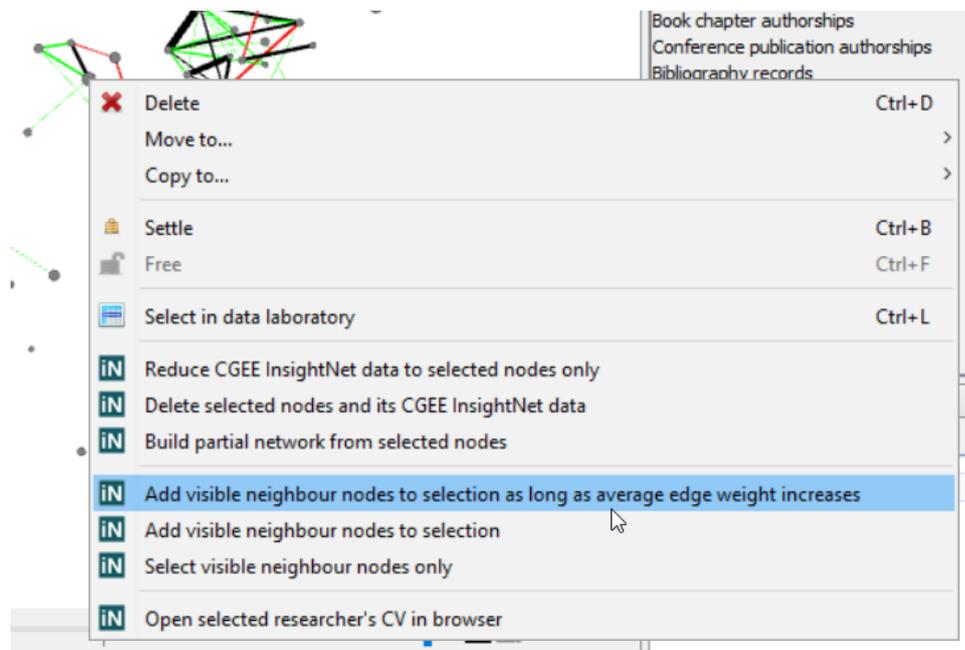
Figura 8.29: Criação de uma rede a partir do subconjunto de nós selecionados

Os nós que não foram selecionados vão permanecer no banco de dados, mas não farão parte da rede criada.

Esta funcionalidade permite a análise de várias sub-redes sem precisar realizar novas importações de dados.

8.8 Seleção interativa de nós vizinhos na rede

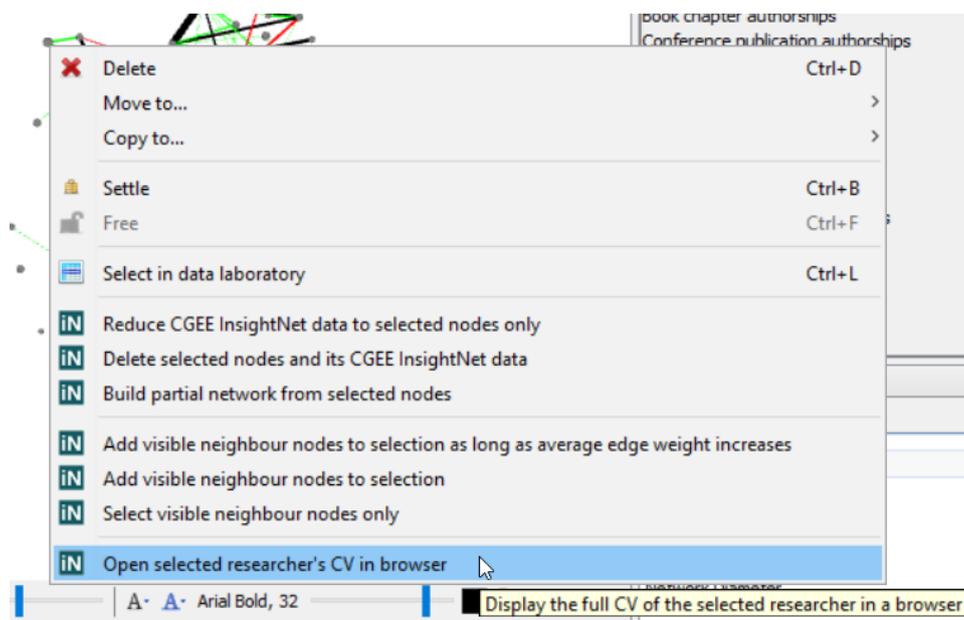
A partir de um ou mais nós selecionados, o *CGEE Insight Net* permite a seleção dos nós vizinhos com um clique do botão direito do mouse no nó dentro da visualização do grafo ou no laboratório de dados. A função “*Add visible neighbour nodes to selection*” acrescenta ao conjunto de nós selecionados todos os vizinhos visíveis desses nós. Já a função “*Select visible neighbour nodes only*” substitui o conjunto de nós selecionado pelos nós vizinhos.

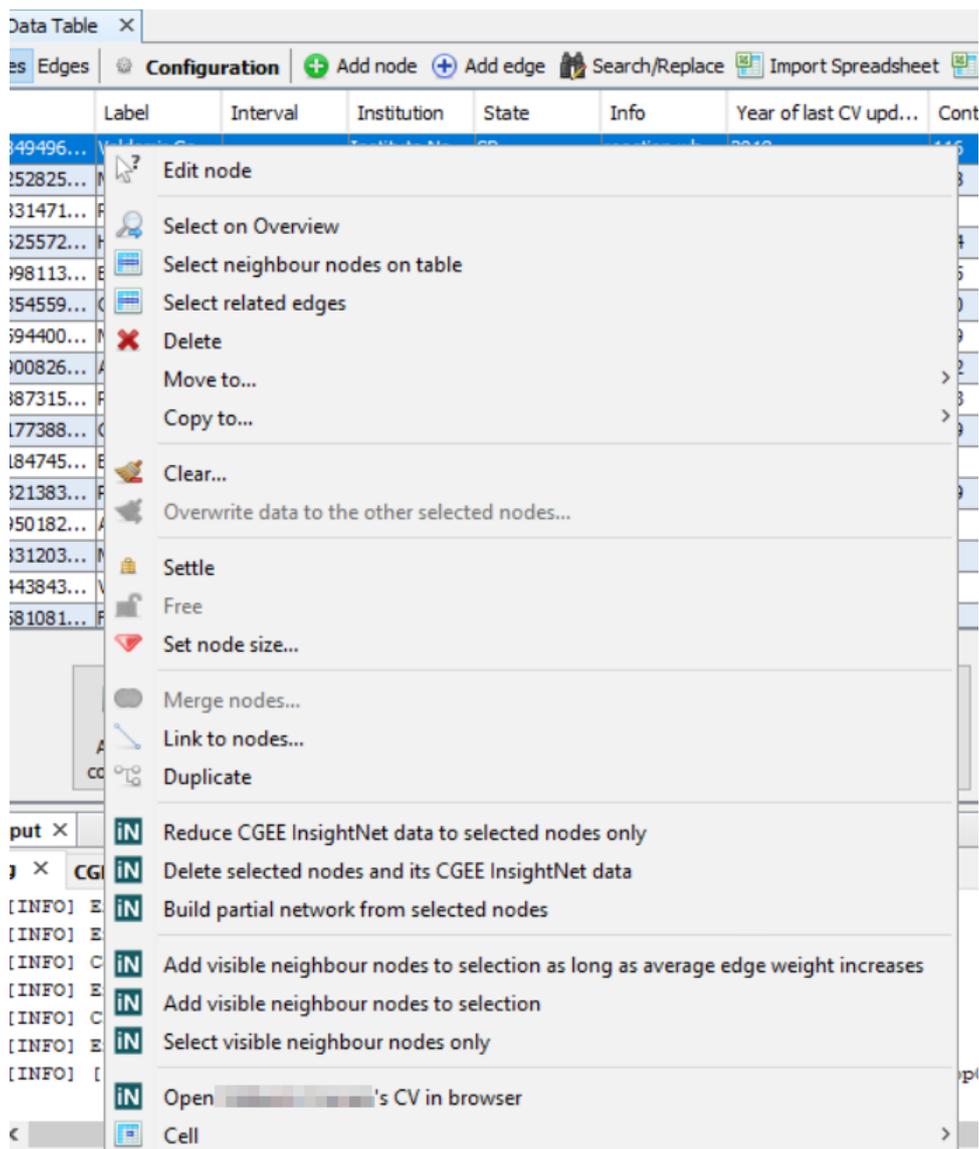


Já a funcionalidade “*Add visible neighbour nodes to selection as long as average edge weight increases*” repete o processo de adicionar vizinhos enquanto o peso média das arestas não diminua. Desta forma, o processo termina quando os nós adicionados não agregam mais informações relevantes ao subconjunto de nós selecionados.

8.9 Visualização interativa do currículo de pesquisadores no browser

Em redes de pesquisadores, o Currículo Lattes de um pesquisador pode ser aberto no browser, clicando com o botão direito do mouse no nó da rede e selecionando a funcionalidade “*Open selected researcher’s CV in browser*” na visualização do grafo ou “*Open <name>’s CV in browser*” no laboratório de dados. Essa funcionalidade abre o site do CNPq e, portanto, depende do acesso à internet.





The screenshot shows a web browser window with the following elements:

- Browser Tab:** Currículo do Sistema de Currículo
- Address Bar:** Not secure | buscatextual.cnpq.br/buscatextual/visualizacv.do?metodo...
- Page Header:** CNPq logo, Currículo Lattes logo, and navigation links (English, Print, Zoom, XML, etc.).
- Navigation Menu:** Dados gerais | Formação | Atuação | Produções | Inovação | Educação e Popularização de C & T | Orientações | Bancas | +
- Profile Section:** A profile picture of a man and a bio section with text in Portuguese.
- Background:** A network graph interface with a table of nodes and an output log.

Figura 8.30: Visualização do currículo do pesquisador no Browser

Funcionalidades comuns de apoio

As funcionalidades descritas nessa seção se aplicam a todos os tipos de rede e são replicadas em todos os sub-menus do CGEE Insight Net. Entretanto, essa seção se limita à demonstração dos diagramas dos itens apenas no sub-menu “*CGEE Insight Net Lattes*”.

9.1 Recuperação do grafo a partir das informações que constam no banco de dados

A seleção das contribuições e a pesquisa de similaridade são realizadas exclusivamente no banco de dados, de acordo com a tabela na [Seção 4](#). O último passo (passo 4 da tabela) gera o grafo a partir das informações que constam no banco de dados.

Esse passo pode ser executado isoladamente e permite a recuperação do grafo sem precisar executar a demorada pesquisa por similaridade, uma vez que a ferramenta Gephi não permite desfazer algumas ações realizadas. O item *Plugins > CGEE Insight Net ... > Recreate latest graph* extrai as informações consolidadas do banco de dados e cria um novo grafo no Gephi:

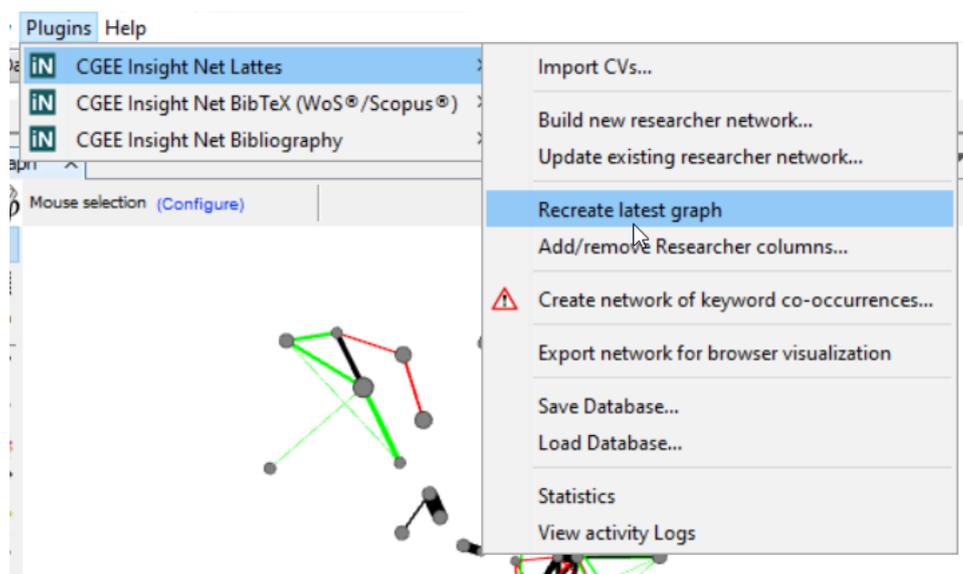


Figura 9.1: Recuperação do grafo

Selecionando essa opção, o grafo é criado a partir das informações no banco de dados, da mesma forma em que o passo 4 da tabela da [Seção 4](#) é executado:

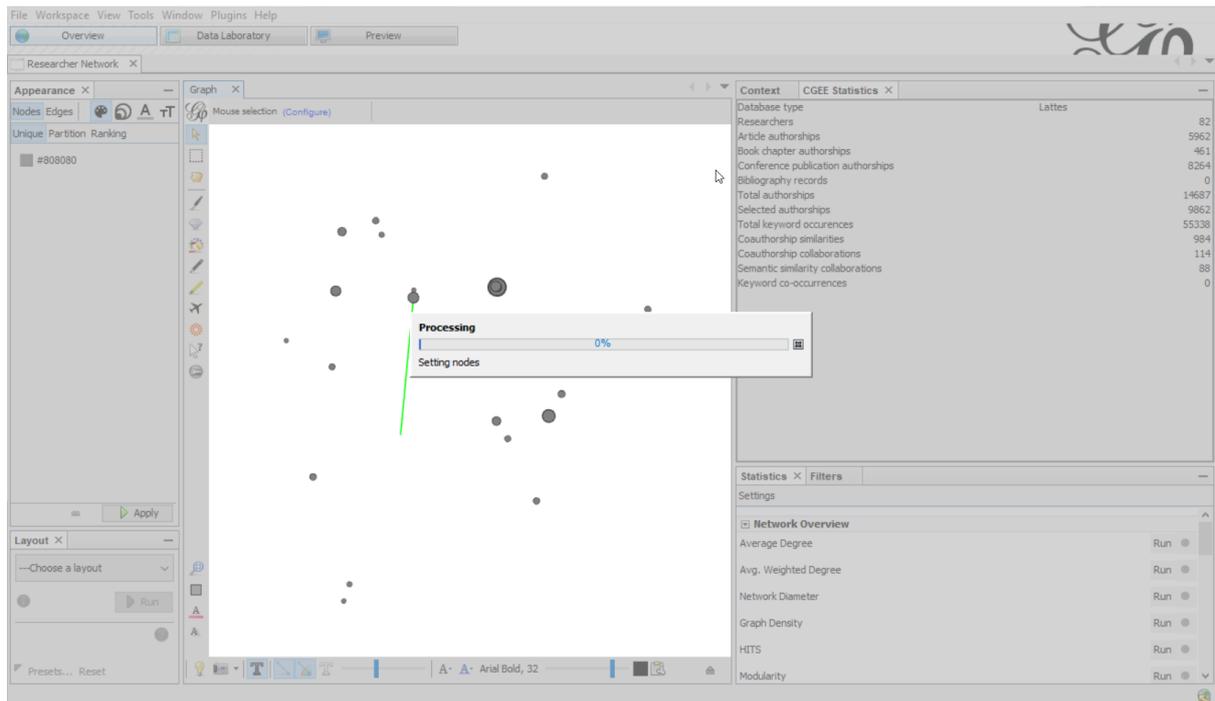


Figura 9.2: Recuperação do grafo

9.2 Cópia e recuperação do banco de dados

Conforme mencionado na Seção 4, todos os dados relevantes selecionados para análise dos currículos constam no banco de dados e apenas no final do processamento são transformados em um grafo visível. Assim, o banco de dados é o repositório principal das informações – o grafo gravado no arquivo `.gephi` é apenas uma representação visual das informações computadas.

O *CGEE Insight Net* permite a gravação do banco de dados em um arquivo do tipo `.cge` e a respectiva recuperação a partir dos itens “*Save Database*” e “*Load Database*” do menu *Plugins > CGEE Insight Net* . . . :

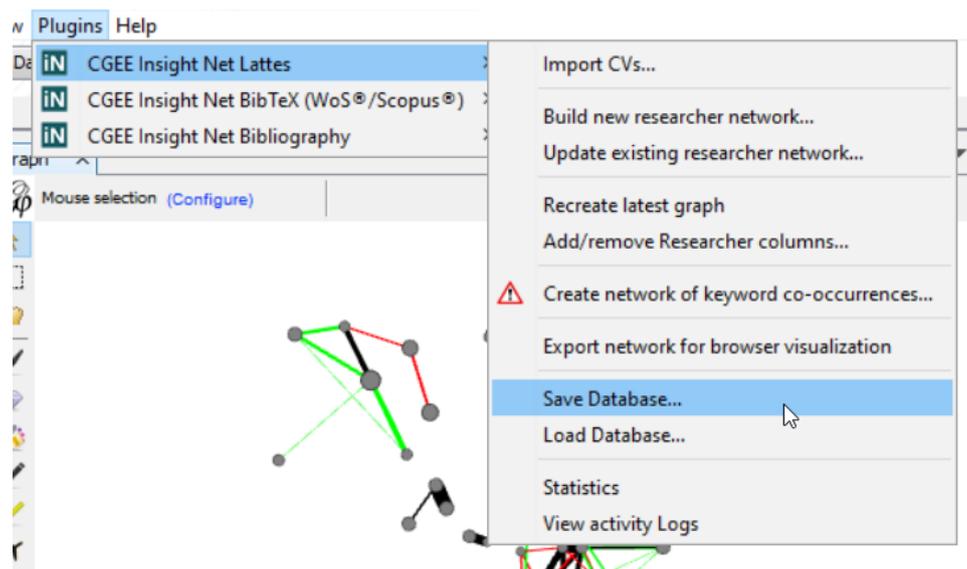


Figura 9.3: Funcionalidades de gravação e recuperação do banco de dados

Selecionando o item “*Save Database*”, o sistema apresenta um diálogo e solicita a definição do arquivo a ser gravado:

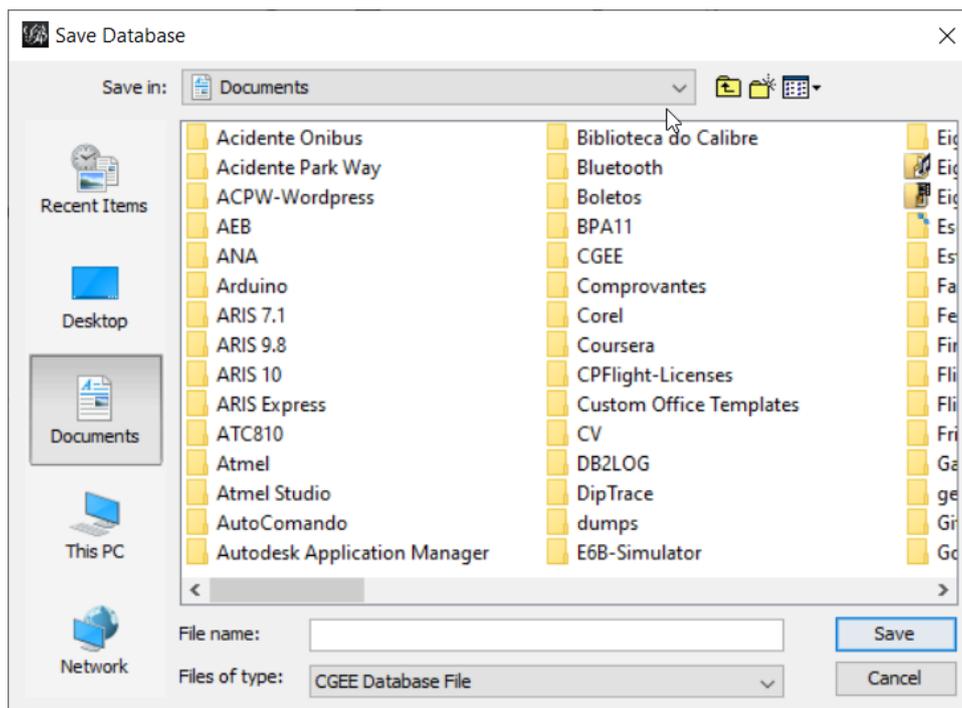


Figura 9.4: Especificação do arquivo de backup do banco de dados

Se o usuário especificar um arquivo válido e clicar em “Save”, o backup do banco de dados é gerado:

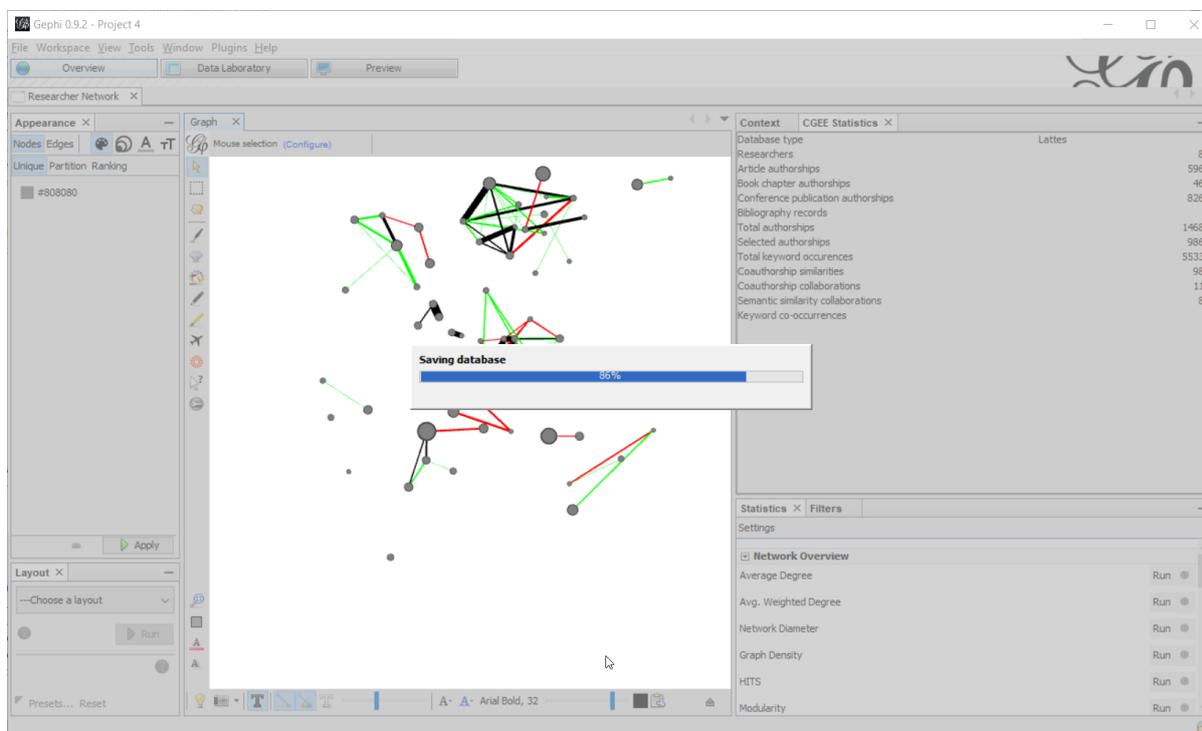
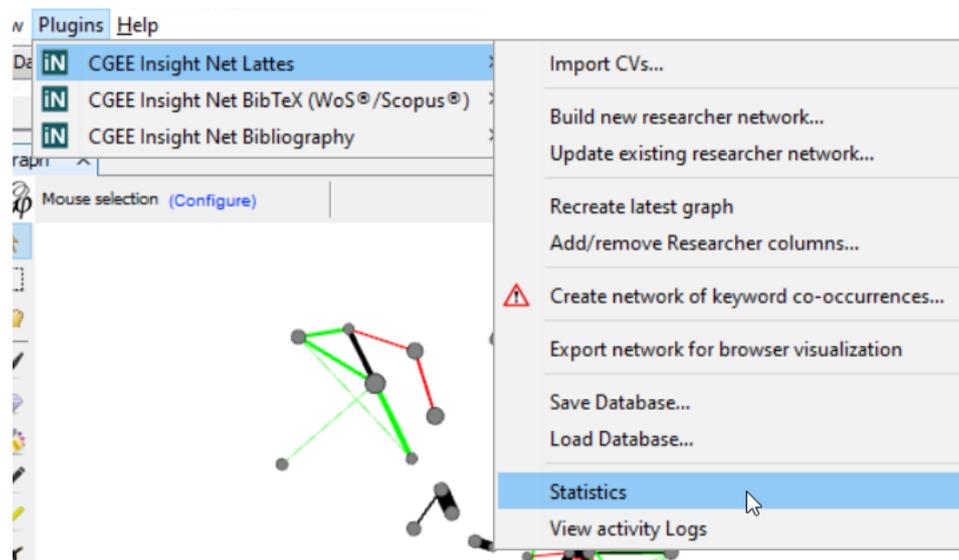


Figura 9.5: Geração do backup da base de dados

9.3 Estatísticas do banco de dados

O *CGEE Insight Net* permite exibir uma estatística do banco de dados com a opção *Plugins > CGEE Insight Net ... > Statistics*, que abre a janela de estatística do banco de dados:



Context	CGEE Statistics	
Database type		Lattes
Researchers		82
Article authorships		5962
Book chapter authorships		461
Conference publication authorships		8264
Bibliography records		0
Total authorships		14687
Selected authorships		9862
Total keyword occurrences		55338
Coauthorship similarities		984
Coauthorship collaborations		114
Semantic similarity collaborations		88
Keyword co-occurrences		0

Figura 9.6: Janela de estatística

Essa janela mostra o tipo e a quantidade de registros para vários tipos de dados no banco. Percebe-se que depois da importação e antes da geração da rede, os valores para “*Selected Contributions*”, “*Coauthorship Similarities*”, “*Coauthorship collaborations*”, “*Semantic similarity collaborations*” e “*Keyword co-occurrences*” ficam com o valor zero, já que essas entidades são geradas apenas durante a formação da rede.

9.4 Protocolos de execução

Conforme descrito na Seção 3, o *CGEE Insight Net* gera diversos registros de protocolo. Eles podem ser visualizados com a opção *Plugins > CGEE Insight Net > View Logs*, que abre a janela de protocolo:

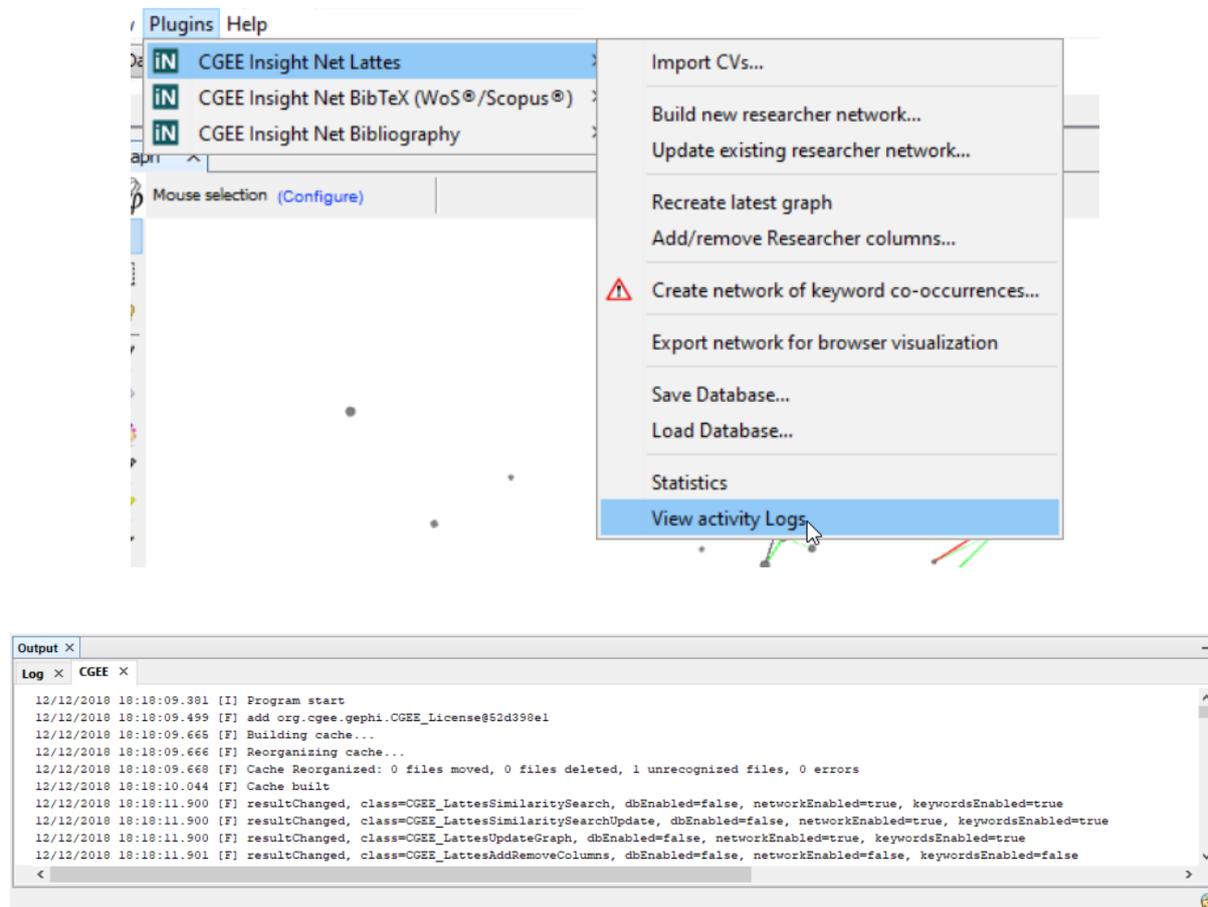


Figura 9.7: Protocolo de execução

A janela possui duas abas:

- Na aba “Log” são exibidas mensagens do próprio ambiente do Gephi, sem relação ao *CGEE Insight Net*
- Já a aba “CGEE” exibe os registros de protocolo de execução do *CGEE Insight Net*.

Cada linha nesta aba “CGEE” é marcada com data e hora e com o tipo de registro:

Típo	Significado
[E]	Erro severo, integridade dos dados não garantida
[!]	Aviso importante
[I]	Informação sobre a execução do CGEE Insight Net
[F]	Informações detalhadas sobre a execução do CGEE Insight Net
[f]	Informação de depuração
[*]	Informação detalhada de depuração

Referências Bibliográficas

- [Assort] M. Newman, “7.13 *HOMOPHILY AND ASSORTATIVE MIXING*,” em *Networks. An Introduction*, New York, Oxford University Press, 2010.
- [Percolation] A.-L. Barabási, “8.2 *Percolation Theory* ff”, em *Network Science*, Cambridge, Cambridge University Press, 2016, p. 273ff.

Anexo II

insightNet Browser - Guia do usuário

Release 1.6.2

CGEE

dez 16, 2018

1	Instalação	3
1.1	Requisitos	3
1.2	Dispositivos e telas	3
1.3	Instalação	4
1.3.1	Executando o insightnet Browser	4
1.4	Estrutura de pastas	5
1.5	Arquitetura	7
2	Interface	9
2.1	Barra de título	10
2.1.1	Pesquisa	10
2.1.2	Menu de redes	10
2.1.3	Números básicos da rede	10
2.1.4	Menu principal	10
2.2	Área de detalhamento	12
2.2.1	Ferramentas da área de detalhamento	13
2.2.2	Exibindo/ocultando arestas na área de detalhamento	14
2.3	Área de navegação	16
2.3.1	Janela de dica do nó	16
2.3.2	Controles de navegação	16
2.3.3	Visão geral da rede (minitel)	17
2.3.4	Menu de partições	18
3	Configuração	19
3.1	Configurações de exibição da rede	20
3.1.1	Transparência das arestas	20
3.1.2	Espessura das arestas	20
3.1.3	Fator para cálculo do tamanho dos labels	20
3.1.4	Tamanho mínimo dos labels	20
3.1.5	Fator para cálculo do tamanho dos nós	21
3.1.6	Arestas curvas	21
3.1.7	Exibir arestas enquanto o gráfico é arrastado	21
3.1.8	Exibir dicas dos nós	21
3.2	Configurações das nuvens de palavras	21
3.2.1	Atributo para as Nuvens de palavras	21
3.2.2	Idioma da rede	22

3.2.3	Tipo de espiral	22
3.2.4	Tamanho mínimo do texto	22
3.2.5	Tamanho máximo do texto	22
3.2.6	Número máximo de palavras	22
3.2.7	Personalizar Stopwords	22
3.2.8	Número de termos extraídos nos campos textuais	23
3.3	Reinicialização, backup e restauração	23
3.3.1	Limpar dados	24
3.3.2	Reiniciar rede	24
3.3.3	Restaurar backup	24
3.3.4	Backup da rede	24
4	Utilização	25
4.1	Visualizando uma rede existente	25
4.2	Acrescentando novas redes	26
4.2.1	Via URL	26
4.2.2	Via Menu de redes	26
4.3	Interação e navegabilidade	27
4.4	Selecionando nós	27
4.4.1	Selecionando com um clique	27
4.4.2	Selecionando com área de seleção	28
4.4.3	Selecionando em gráficos	29
4.5	Ferramenta lente de aumento	29
5	Módulos	31
5.1	Metadados da rede	31
5.2	Tabelas de nós e arestas	33
5.2.1	Selecionando linhas	34
5.2.2	Exportação	34
5.2.3	Exibir/Ocultar colunas	34
5.3	Pesquisa avançada	35
5.3.1	Termos de busca	36
5.3.2	Gerenciamento de pesquisas	38
5.4	Nuvem de palavras	39
5.4.1	Informações gerais da nuvem	40
5.4.2	Seleção de palavras e consultas	41
5.4.3	Filtro de frequência	43
5.4.4	Tabela de palavras e frequências	43
5.4.5	Outras funcionalidades	45
5.5	Gráficos da rede	45
5.5.1	Scatterplot	47
5.5.2	Histograma	50
5.5.3	Palavras-chave	54
5.6	Visualizações	63
5.6.1	Gerenciando visualizações	64
5.6.2	Modo de apresentação	65
5.7	Partições	66

Introdução

O insightNet Browser é uma ferramenta de visualização e mineração de dados mediante a interação e navegação em grafos de redes, previamente gerados e salvos em formato GEXF.

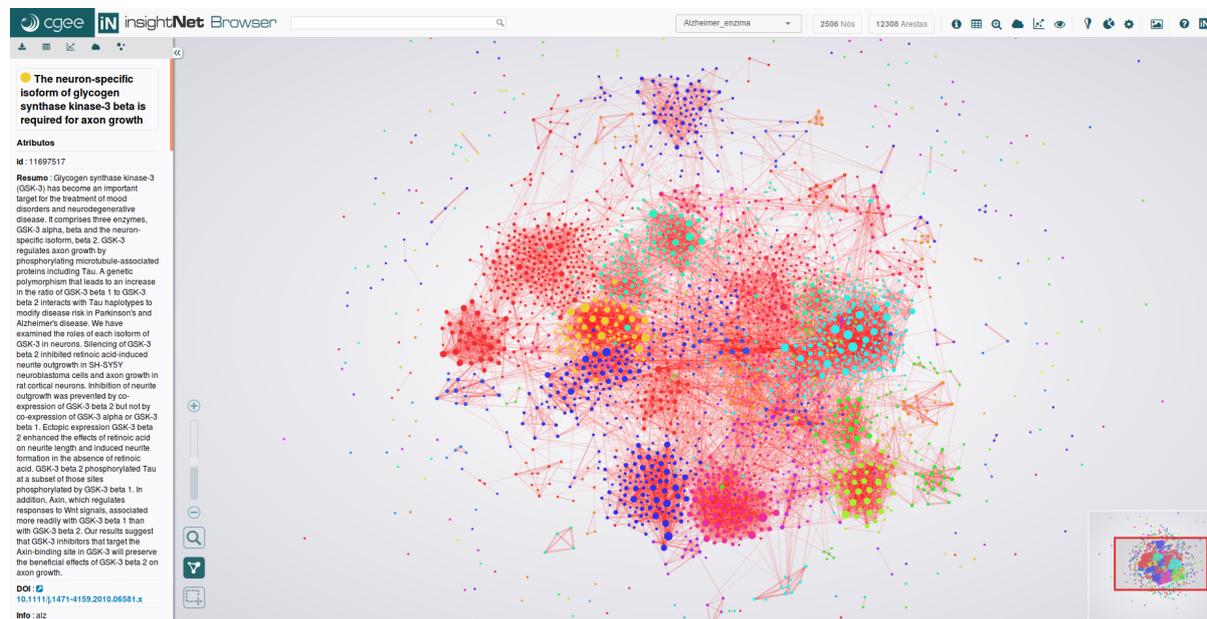


Figura1-: Interface do usuário insightNet Browser

O foco do insightNet Browser está na visualização e exploração interativa do grafo de rede e de outros gráficos gerados a partir de atributos dessa rede. A ferramenta engloba de forma intuitiva várias técnicas de visualização da informação, para que o usuário possa **consultar**, **correlacionar**, **quantificar** e **qualificar** os diversos metadados e atributos que podem estar contidos nos elementos do grafo (**nós** e **arestas**).

As principais características do browser são:

- Exibição de dados dos nós e suas ligações;
- Extração de dados da rede em vários formatos (PDF, Excel, CSV e texto simples);
- Exportação da rede e gráficos em formato imagem;
- Geração de nuvens de palavras da rede ou de um conjunto de nós selecionados;
- Geração de vários tipos de gráficos com atributos dos nós;
- Gráficos de evolução, frequência e co-ocorrências de palavras-chave;
- Customização dos parâmetros de visualização da rede;
- Pesquisa avançada combinando diferentes atributos da rede;
- Armazenamento de visualizações e exibições em modo apresentação;
- Particionamento customizado, mantendo as cores originais dos nós ou atribuindo novas cores.

O insightNet Browser é uma ferramenta que complementa o processo de pesquisa, análise e prospecção, proporcionando insights intuitivos e novas percepções sobre os metadados que envolvem a rede, além

das relações e agrupamentos explícitos. Para tanto, a ferramenta oferece grande facilidade na mineração desses dados da rede por meio da interatividade com elementos visuais, em gráficos como scatterplot, histograma e nuvem de palavras. Interatividade que vai desde a criação de consultas visuais por meio de cliques e seleções, até a edição dos elementos do gráfico para focar nos dados de interesse.

1.1 Requisitos

Para executar o insightNet Browser em qualquer sistema operacional você precisará do navegador Firefox[®]. Algumas restrições à linguagem Javascript impedem que a aplicação execute normalmente em outros navegadores.

Dica: A última versão do navegador Firefox pode ser baixada no link: <https://www.mozilla.org/pt-BR/firefox/new/>

1.2 Dispositivos e telas

Recomendamos que a aplicação seja executada em dispositivos com tamanho de tela mínimo de 768 pixels. Isso significa que a interface da aplicação se comportará bem em dispositivos como tablets (com tela acima de 8 polegadas), notebooks e desktops, não sendo recomendada a utilização do mesmo em celulares e tablets com menos de 8 polegadas.



Figura2-: Dispositivos e tamanhos de tela compatíveis com o insightNet Browser

1.3 Instalação

O insightNet Browser não necessita de um processo de instalação no sistema operacional, como a maioria dos softwares. Toda a aplicação está contida em uma pasta que geralmente é distribuída em um arquivo compactado, no formato ZIP. Basta descompactar a pasta no local desejado e a aplicação estará pronta para ser utilizada.

1.3.1 Executando o insightnet Browser

Para executar a aplicação basta abrir o arquivo **index.html**, na raiz da pasta 'insightnet-browser', com o navegador Firefox.

Caso você utilize outro navegador como padrão, abra o arquivo **index.html** clicando com o **botão direito** do mouse sobre o ele e escolhendo as opções **Abrir com » Firefox**.

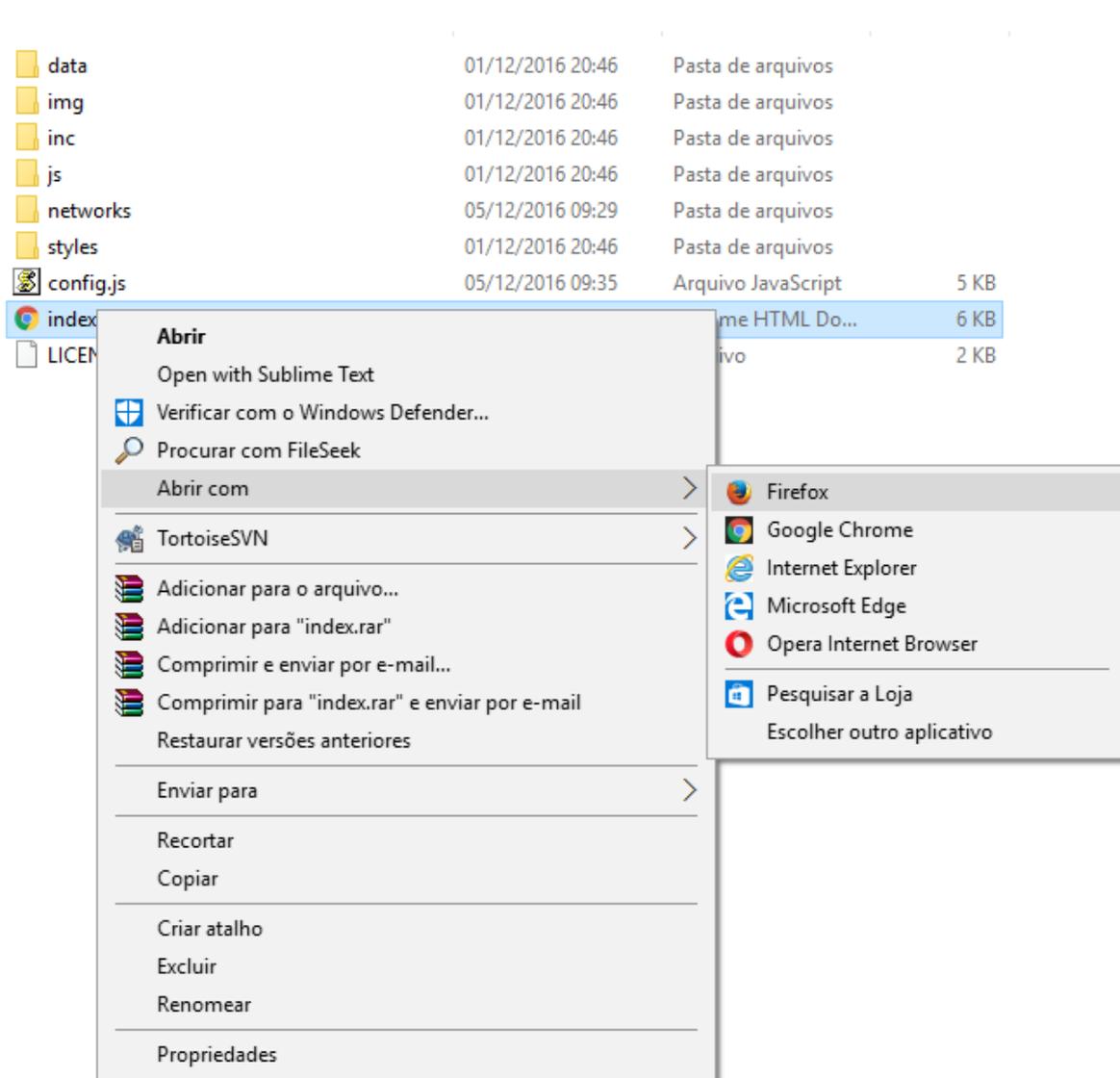


Figura3-: Abrindo a aplicação com o navegador Firefox

1.4 Estrutura de pastas

Ao descompactar a aplicação, você verá a seguinte estrutura de pastas, dentro da pasta principal (insightnet-browser):

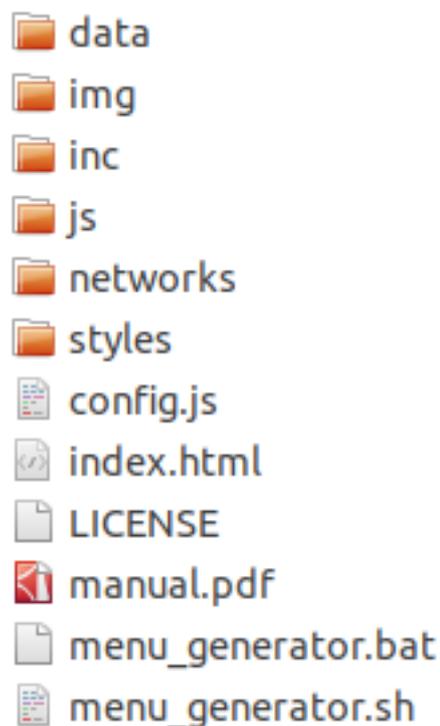


Figura4-: Estrutura de pastas do insightNet Browser

Tabela1-: Estrutura da pasta raiz

Pasta/Arquivo	Descrição
data/	Pasta que armazena dados utilizados internamente pela aplicação, como stopwords por idioma. Os dados se encontram no formato JSON.
img/	Pasta de imagens utilizadas pela aplicação.
inc/	Pasta com arquivos HTML utilizados pela aplicação para inserção de blocos e janelas popup.
js/	Pasta com os arquivos fontes da aplicação e bibliotecas utilizadas (Javascript).
networks/	Pasta onde as redes em formato GEXF são armazenadas.
styles/	Pasta com arquivos de folhas de estilo da aplicação (CSS).
config.js	Arquivo avançado de configurações gerais da aplicação. A alteração do conteúdo desse arquivo só deve ser executada por usuários experientes e com conhecimentos em programação javascript.
index.html	Arquivo HTML de execução da aplicação.
LICENSE	Arquivo com a licença básica do software.
manual.pdf	Guia do usuário insightNet Browser.
menu_generator.bat	Windows - Script para geração do arquivo networks.json , na pasta networks/. Esse arquivo contém dados das redes contidas nesta pasta e é utilizado na geração do menu de redes da aplicação.
menu_generator.sh	Linux - Script para geração do arquivo networks.json , na pasta networks/. Esse arquivo contém dados das redes contidas nesta pasta e é utilizado na geração do menu de redes da aplicação.

1.5 Arquitetura

O insightNet Browser oferece grande praticidade na sua utilização, pois tem uma arquitetura simples e portátil, pois trata-se de uma aplicação web executada apenas em um navegador (*browser*), sem a necessidade de instalação no sistema operacional.



Figura5-: Linguagens e tecnologias no insightNet Browser

As tecnologias utilizadas para o desenvolvimento do insightNet Browser são: HTML, Javascript, SASS/CSS, JSON e XML. A aplicação foi desenvolvida tendo como base inicial o código da aplicação **JavaScript GEXF Viewer for Gephi**, de **Raphaël Velt**, publicada sob licença MIT em <https://github.com/raphv/gexf-js>.

Formato GEXF

O arquivo do tipo GEXF é um arquivo de dados estruturados no formato XML. Nele estão listados todos os nós da rede em um primeiro conjunto de dados e em um segundo estão as arestas da rede.

Embora exista uma especificação conhecida e publicada para a validação de arquivos GEXF, nem todos os programas que lidam com redes implementam essa validação.

Para que o arquivo GEXF seja lido corretamente pelo insightNet Browser é necessário que o mesmo contenha atributos de posicionamento de cada nó e que não hajam arestas com nó de origem ou destino que não estejam relacionados na rede.

Recomendamos a utilização prévia de algum programa de manipulação de redes para a geração de um layout e a aplicação de outras propriedades visuais como particionamento ou modularização, antes de carregar essa rede no insightNet Browser.

Saiba mais

Para mais informações sobre o formato de arquivo GEXF acesse <https://gephi.org/gexf/format/>.

Interface

A interface do insightNet Browser é bastante intuitiva, muitos usuários conseguem tirar o máximo de proveito da aplicação sem necessitar desse guia.

A aplicação é executada em uma única aba do navegador, onde algumas funcionalidades são executadas em pequenas janelas embutidas (*popups*), sem que o usuário precise abrir outras janelas ou abas do navegador.

A seguir, detalharemos os elementos que compõem a interface da aplicação:

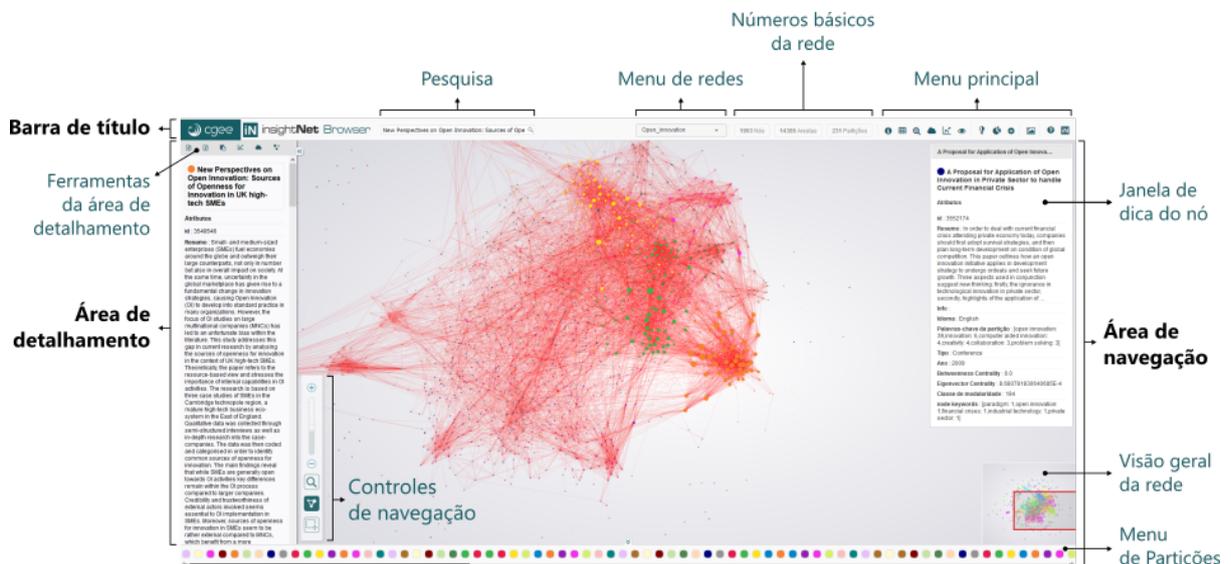


Figura6: Elementos de interface do insightNet Browser

2.1 Barra de título

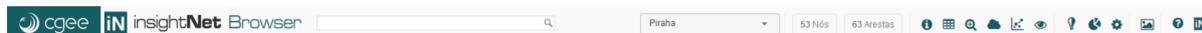


Figura7-: Barra de título insightNet Browser

2.1.1 Pesquisa

No campo de pesquisa você pode executar buscas rápidas nos labels/títulos dos nós que compõem a rede.

Dica: Para pesquisas em outros atributos dos nós veja *Pesquisa avançada*.

2.1.2 Menu de redes

Utilize esse controle para acessar as redes de exemplo que acompanham a aplicação.

2.1.3 Números básicos da rede

Essa área exibirá os **números de nós**, **arestas** e **partições** da rede carregada.

2.1.4 Menu principal

No menu principal você poderá ter acesso às informações da rede, exportação de dados, geração de gráficos, nuvem de palavras e configurações da aplicação.

O menu é composto pelos seguintes botões:

Tabela2-: Menu principal

	Metadados da rede: informações gerais da rede e dos atributos de nós e arestas (mais detalhes em <i>Metadados da rede</i>).
	Tabelas de nós e arestas e seus atributos (mais detalhes em <i>Tabelas de nós e arestas</i>).
	Pesquisa avançada em atributos dos nós da rede (mais detalhes em <i>Pesquisa avançada</i>).
	Gerar a nuvem de palavras extraídas de um atributo da rede (mais detalhes em <i>Nuvem de palavras</i>).
	Gerar gráficos da rede (mais detalhes em <i>Gráficos da rede</i>).
	Gerenciamento de visualizações e modo de apresentação (mais detalhes em <i>Visualizações</i>).
	Alternar cor de fundo para aumento ou diminuição do contraste.
	Atribuir um particionamento de cores para a rede utilizando, por exemplo, um atributo categórico (mais detalhes em <i>Partições</i>).
	Configurações gerais da aplicação (mais detalhes em <i>Configuração</i>).
	Capturar a tela e gerar uma imagem da rede no formato PNG, que será aberta em uma nova aba (tecla de atalho F2).
	Acesso a esse guia, na versão HTML (navegável). O guia será aberto em uma nova aba.
	Informações de desenvolvimento e copyright da aplicação.

Nota: O detalhamento de cada uma das opções do menu principal será apresentado nos próximos capítulos desse guia.

2.2 Área de detalhamento

The image shows two side-by-side panels from the insightNet Browser interface. The left panel displays the details for a specific node, while the right panel shows a list of selected nodes.

Node Detail View (Left Panel):

- Node Title:** Cognitive Transformations and Extended Expertise
- Atributos:**
 - id : 16456319
 - Resumo : Expertise is extended by becoming immersed in cultural practices. We look at an example of mathematical expertise in which immersion in cognitive practices results in the transformation of expert performance.
 - DOI : [10.1080/00131857.2013.779209](https://doi.org/10.1080/00131857.2013.779209)
 - Idioma : English
 - Tipo : Article/Review/Letter
 - Ano : 2014
 - Classe de modularidade : 16
 - weighted degree : 0.12037522718310356
 - weighted indegree : 0.0
 - weighted outdegree : 0.0
- Ligações não direcionadas:** 2
 - Linguistic fire and human cognitive powers – 16456319-16456381 [0.06270186]
 - Counting systems and the First Hilbert problem – 16456319-16456453 [0.05767337]

Node List View (Right Panel):

- Section:** Nós selecionados
- Total de nós:** 12
- List of selected nodes:
 - An Evaluation of Universal Grammar and the Phonological Mind
 - Resolving the Infinitude Controversy
 - What linguists are talking about when talking about...
 - HUMMING, WHISTLING, SINGING, AND YELLING IN PIRAHA CONTEXT AND CHANNELS OF COMMUNICATION IN FDG
 - Recursive misrepresentations: A reply to Levinson (2013)
 - Semantic fieldwork and lexical universals
 - Why is there variation rather than nothing?
 - Recursion in pragmatics
 - Language: Between cognition, communication and culture
 - Warlpiri Wh-Scope Marking
 - PIRAHA EXCEPTIONALITY: A REASSESSMENT
 - PIRAHA CULTURE AND GRAMMAR: A RESPONSE TO SOME CRITICISMS

Figura8-: Área de detalhamento insightNet Browser: exibindo um nó (à esquerda) e uma listagem de nós selecionados (à direita)

A área de detalhamento oferece a visualização dos metadados de um ou mais nós selecionado(s) na área de navegação.

Quando apenas um nó é selecionado, a área irá exibir todos os atributos deste nó e listar as suas ligações com outros nós da rede. Quando vários nós são selecionados, a área irá exibir uma listagem desses nós.

Dica: Quando a rede é formada por artigos a aplicação tentará extrair um DOI dos metadados e exibir um link para o mesmo. Quando a rede é formada por currículos da plataforma Lattes, um link para o currículo será exibido.

2.2.1 Ferramentas da área de detalhamento

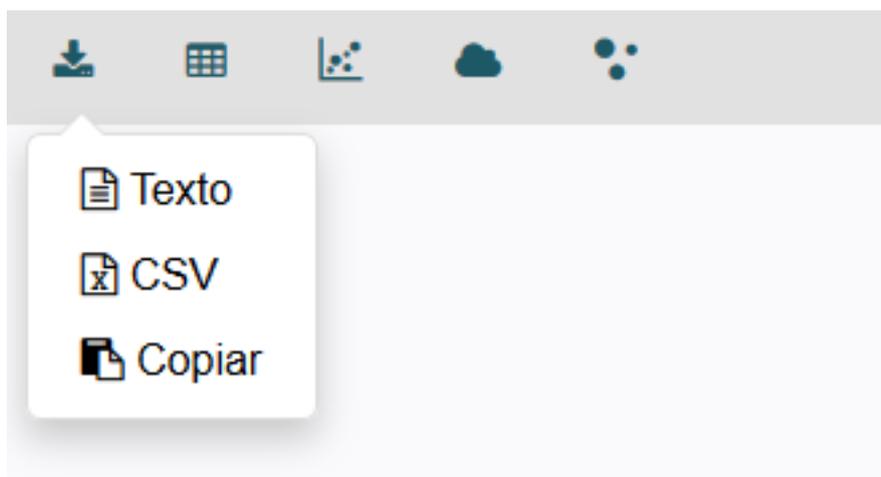


Figura9-: Ferramentas da área de detalhamento

Essas ferramentas permitem a exportação e/ou cópia dos dados dos nós exibidos na área de detalhamento, além da geração de gráficos e nuvens de palavras de atributos escolhidos pelo usuário.

Importante: O menu **Ferramentas da área de detalhamento** tem efeito somente sobre os dados de nós que estão sendo exibidos na área de detalhamento (dados dos nós de uma listagem ou nó em destaque e suas ligações). Para gerar visualizações sobre os dados de toda a rede utilize as opções do *menu principal*.

Tabela3-: Ferramentas da área de detalhamento

 Exportação:	 Texto	Exportar os dados exibidos na área de detalhamento para um arquivo no formato texto.
	 CSV	Exportar os dados exibidos na área de detalhamento para um arquivo no formato csv.
	 Copiar	Copiar os dados exibidos na área de detalhamento para a área de transferência do sistema operacional.
	Visualizar os dados da área de detalhamento em tabelas de nós e arestas.	
	Gerar gráficos da rede (scatterplot, linhas, barras, chords etc.).	
	Gerar a nuvem de palavras extraídas de um atributo dos nós exibidos.	
	Realçar os nós presentes na área de detalhamento, destacando-os na área de navegação.	

Dica: Quando os detalhes de apenas um nó (nó selecionado) são exibidos na área de detalhamento, as exportações, as tabelas, os gráficos e a nuvem de palavras irão abranger dados tanto do nó exibido quanto de suas ligações (nós ligados ao nó selecionado).

2.2.2 Exibindo/ocultando arestas na área de detalhamento

Quando os dados de um único nó estão sendo exibidos na área de detalhamento, você tem a opção de ocultar/exibir a lista de arestas ligadas ao nó em exibição.

Cognitive Transformations and Extended Expertise

Atributos

id : 16456319

Resumo : Expertise is extended by becoming immersed in cultural practices. We look at an example of mathematical expertise in which immersion in cognitive practices results in the transformation of expert performance.

DOI : [10.1080/00131857.2013.779209](https://doi.org/10.1080/00131857.2013.779209)

Idioma : English

Tipo : Article/Review/Letter

Ano : 2014

Classe de modularidade : 16

weighted degree : 0.12037522718310356

weighted indegree : 0.0

weighted outdegree : 0.0

Ligações não direcionadas 2

- Linguistic fire and human cognitive powers** – 16456319-16456381 [0.06270186]
- Counting systems and the First Hilbert problem** – 16456319-16456453 [0.05767337]

Há 2 ligações ocultas

Figura10-: Área de detalhamento com dados de um nó: exibindo as arestas ligadas ao nó (à esquerda) e ocultando essas arestas (à direita)

Para ocultar as arestas basta clicar no ícone , ao lado do título da lista de arestas. Para voltar a exibir a lista de arestas clique no botão exibir, como mostrado na figura acima.

Importante: As exportações, tabelas, gráficos e a nuvem de palavras da área de detalhamento irão ignorar as arestas do nó se essas estiverem ocultas.

2.3 Área de navegação

2.3.1 Janela de dica do nó

Pequena janela (*tooltip*) com uma listagem limitada dos atributos do nó quando o mouse é posicionado sobre ele. De acordo com a posição do nó em relação à tela, essa janela pode ser exibida à direita ou esquerda, na área de navegação.

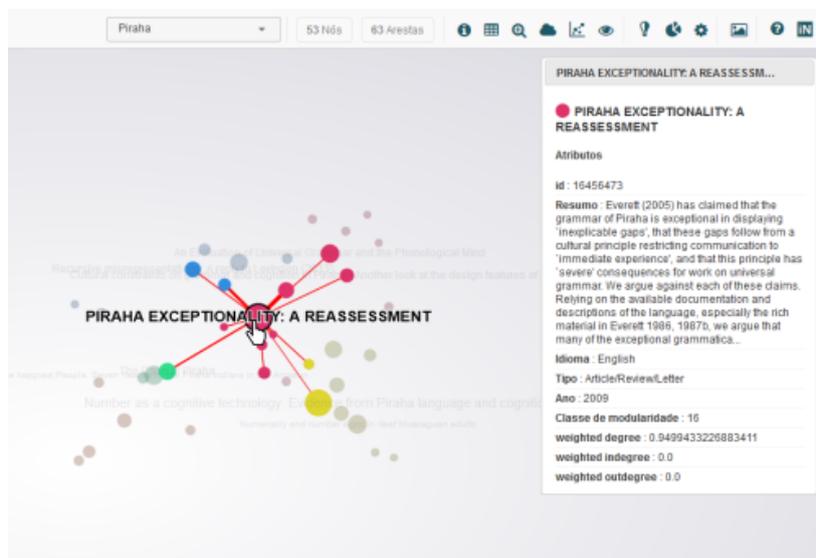


Figura11-: Exemplo de exibição da janela de dica do nó

Dica: Você pode habilitar ou desabilitar a janela de dica do nó nas configurações gerais da aplicação.

2.3.2 Controles de navegação

Esses controles auxiliam na navegação e visualização da rede, controlando os níveis de zoom, selecionando vários nós ao mesmo tempo, exibindo ou ocultando arestas e melhorando a acessibilidade em determinados casos.



Figura12-: Controles de navegação

	Controle do nível de zoom da visualização.
	Ativar/desativar a ferramenta lente de aumento.
	Ativar/desativar a exibição de arestas.
	Ativar/desativar a ferramenta de seleção (selecionar nós desenhando uma área de seleção).

2.3.3 Visão geral da rede (minitela)

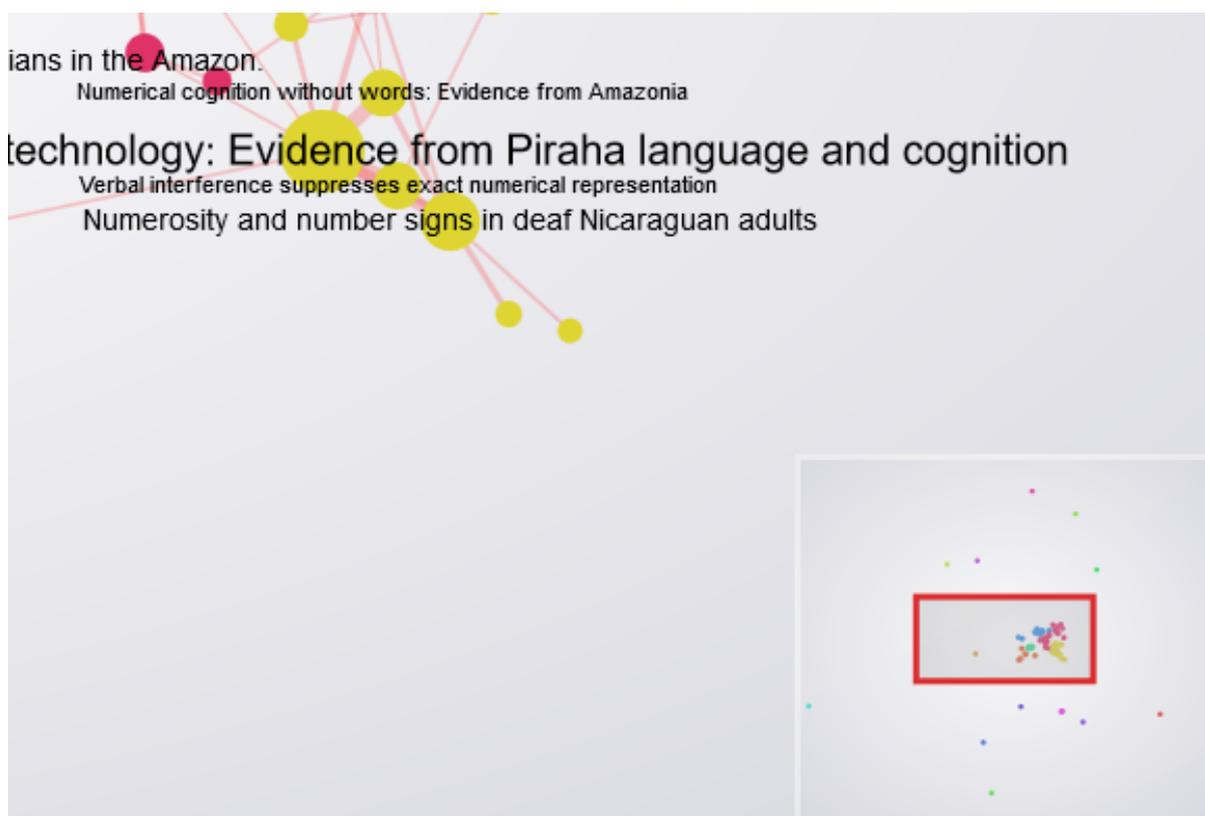


Figura13-: Ferramenta de visão geral da rede

A ferramenta de visão geral é uma miniatura da rede carregada. Por meio dela você pode acompanhar os níveis de zoom e posicionamento na área de navegação.

Com ações do mouse você pode utilizar a ferramenta de visão geral para aumentar ou diminuir o zoom e movimentar a rede pela tela de visualização.

Dica: Quando a ferramenta de seleção é ativada , o cursor perde a função de **movimentar a rede** na área de navegação (clikando, arrastando e soltando). Nesse caso a ferramenta de visão geral pode auxiliar na movimentação da rede para a área desejada.

2.3.4 Menu de partições



Figura14-: Menu de partições

Quando você escolhe um atributo para partição de cores da rede, clicando em , no menu principal, uma barra horizontal com círculos coloridos representando cada partição é exibida na base da área de navegação. Ao clicar em uma partição (cor) os nós com aquele atributo são destacados na área de navegação e listados na área de detalhamento.

Nota: Ao carregar a rede pela primeira vez, nenhuma partição é reconhecida até que o usuário indique uma partição. As cores exibidas quando não há uma partição são as cores originais presentes nos metadados do GEXF.

Configuração

As configurações gerais do insightNet Browser podem ser acessadas no menu principal, clicando no botão . No painel de controles que será exibido, você poderá alterar e personalizar vários parâmetros de visualização da rede e de seus elementos, além de controlar alguns parâmetros de geração das nuvens de palavras.

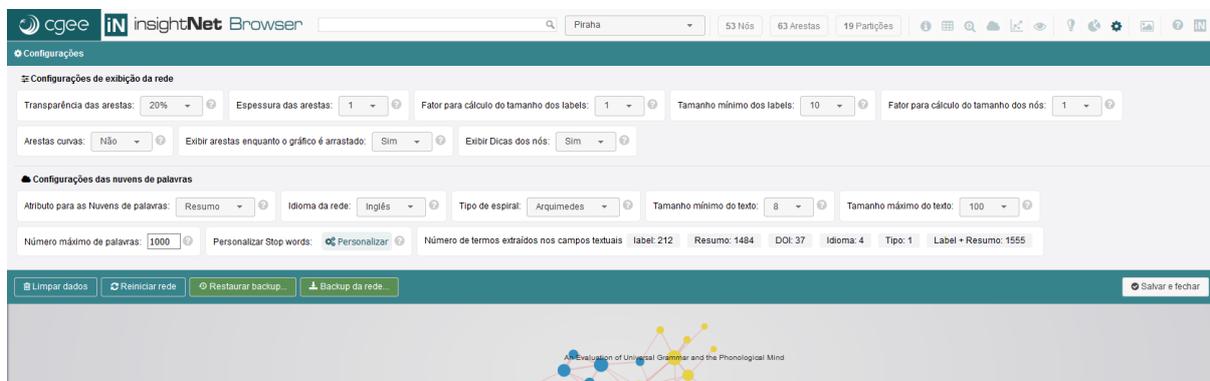


Figura15-: Painel de configurações do insightNet Browser

O painel está dividido em duas seções de controles: **Configurações de exibição da rede** e **Configurações das nuvens de palavras**, e alguns botões de manutenção da aplicação ou rede.

A aplicação utiliza uma API de armazenamento do navegador (*local storage*) para salvar e manter as configurações, separadas por rede, ou seja, quando você alterar alguma configuração, essa customização será armazenada somente para a rede carregada no momento. Sempre que abrir essa rede, as customizações ligadas a ela serão recuperadas.

Dica: Sempre que precisar de uma ajuda rápida sobre qualquer opção do painel de configurações, passe o mouse sobre o ícone , no lado direito de cada controle para que um texto de ajuda seja exibido.

Quando o painel de configurações está aberto algumas funcionalidades e opções da aplicação ficarão desabilitadas.

Atenção: Como o armazenamento de dados é feito no *cache* do navegador, todas as informações serão apagadas quando houver uma limpeza desse *cache*.

A seguir detalharemos cada uma das opções do painel de configurações.

3.1 Configurações de exibição da rede

3.1.1 Transparência das arestas

Permite alterar o nível de visibilidade/transparência das linhas que representam as arestas, utilizando um valor percentual. Escolher o valor 0% fará com que as arestas fiquem invisíveis.

Valor padrão: 20%.

3.1.2 Espessura das arestas

Permite alterar o fator para cálculo da espessura das linhas que representam as arestas. Esse número será multiplicado pela espessura atual das arestas.

Valor padrão: 3.

3.1.3 Fator para cálculo do tamanho dos labels

Permite alterar o fator para cálculo do tamanho dos labels/títulos dos nós. Esse número será multiplicado pelo tamanho atual da fonte.

Repare que nos níveis mais baixos de zoom, os labels dos nós tendem a desaparecer. Aumente o valor dessa opção caso queira que os labels apareçam nesses níveis mais baixos.

Valor padrão: 1.

3.1.4 Tamanho mínimo dos labels

Permite alterar o tamanho mínimo da fonte utilizada para a exibição dos labels ou títulos, utilizando um valor em pixels. Quanto menor o valor informado, mais labels serão exibidos.

Valor padrão: 10 pixels.

3.1.5 Fator para cálculo do tamanho dos nós

Permite alterar o fator para cálculo do tamanho dos nós da rede.

O tamanho do nó já vem atribuído no arquivo GEXF pelo software que o gerou. Caso os menores nós fiquem muito pequenos para a visualização, você poderá aumentar esse fator, que será multiplicado ao tamanho real dos nós.

Valor padrão: 1.

3.1.6 Arestas curvas

Permite alterar a forma das linhas que representam as arestas. Informe **SIM** para exibir as arestas como curvas, ou **NÃO** para exibi-las como retas.

Valor padrão: Não.

3.1.7 Exibir arestas enquanto o gráfico é arrastado

Permite habilitar ou não a exibição das arestas da rede durante a movimentação do gráfico (clique e arrastar). Essa opção pode ser desativada para melhorar a performance da navegação.

Valor padrão: Sim.

3.1.8 Exibir dicas dos nós

Permite ativar ou desativar a janela de pré-visualização (*tooltip*), com informações resumidas do nó. A janela de pré-visualização é exibida sempre que o usuário passa o mouse sobre algum nó da rede. Informe **SIM** para exibir a pré-visualização, ou **NÃO** para não exibi-la.

Essa opção pode prejudicar o desempenho da navegação em redes com grande quantidade de nós e/ou arestas.

Valor padrão: Sim.

3.2 Configurações das nuvens de palavras

3.2.1 Atributo para as Nuvens de palavras

Atributo do nó que será utilizado para a geração das nuvens de palavras. A aplicação irá identificar os atributos do tipo texto e listá-los nesse controle para sua escolha. Tanto a nuvem de palavras da rede quanto a nuvem da área de detalhamento utilizarão esse atributo.

Valor padrão: Título/label.

3.2.2 Idioma da rede

Esse controle permite a especificação do idioma que será utilizado na filtragem de *stopwords* para as nuvens de palavras. Há dois idiomas disponíveis: **português** e **inglês**. Mais adiante, nesse capítulo, você verá como personalizar essa lista de *stopwords*.

Valor padrão: Inglês.

Atenção: Ao alterar o idioma da rede você perderá as personalizações feitas na lista de *stopwords* anterior.

3.2.3 Tipo de espiral

Esse controle permite a escolha do tipo de espiral que será utilizado na geração do layout das nuvens de palavras. Há dois tipos disponíveis: **Arquimedes** e **retangular**.

Valor padrão: Arquimedes.

3.2.4 Tamanho mínimo do texto

Permite especificar, em pixels, o tamanho mínimo da fonte utilizada na nuvem de palavras para as palavras com a menor frequência.

Valor padrão: 8 pixels.

3.2.5 Tamanho máximo do texto

Permite especificar, em pixels, o tamanho máximo da fonte utilizada na nuvem de palavras para as palavras com a maior frequência.

Valor padrão: 100 pixels.

3.2.6 Número máximo de palavras

Permite especificar a quantidade de palavras a serem utilizadas na geração da nuvem. Se a quantidade escolhida é menor que a quantidade real de palavras, a aplicação irá descartar as palavras com as menores frequências.

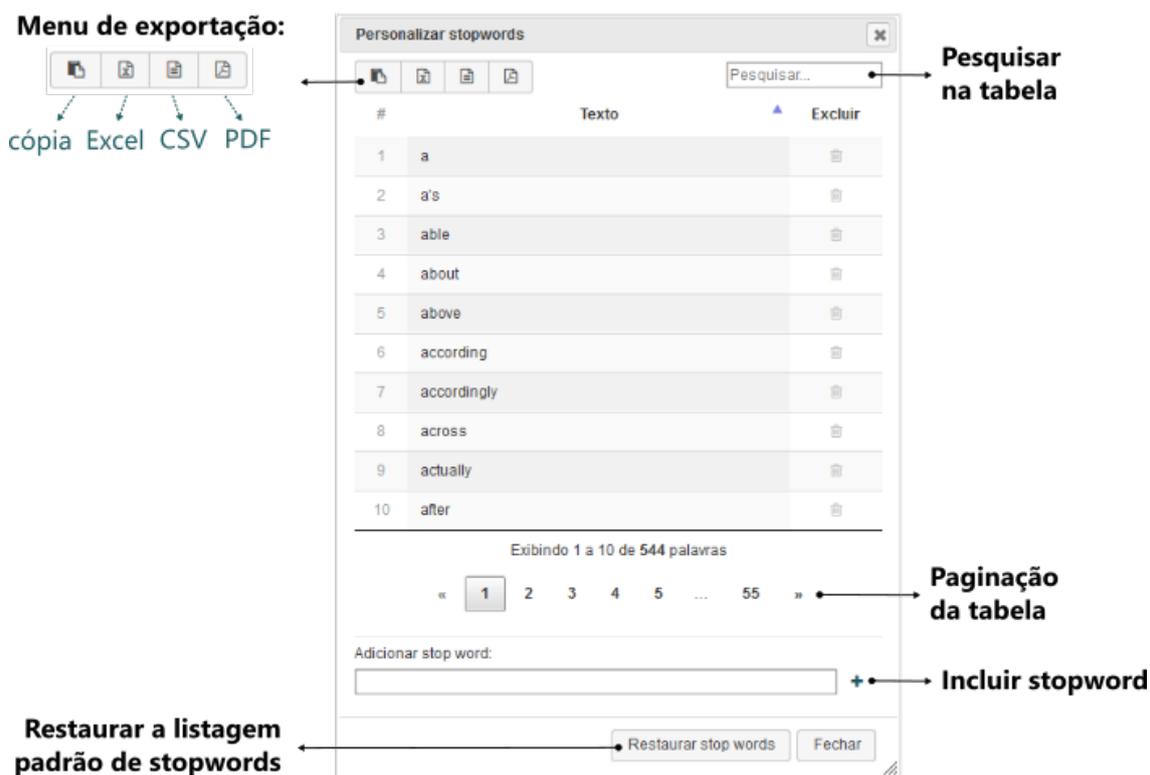
Para não limitar a quantidade de palavras deixe esse campo em branco.

Valor padrão: 1000.

3.2.7 Personalizar Stopwords

As *stopwords* são palavras que devem ser excluídas na contagem de frequência das palavras. Trata-se de pronomes, verbos e outros termos que não são relevantes para a análise. Esse controle permite a exclusão e inclusão de *stopwords*.

Ao clicar no botão **Personalizar** será exibida a tabela de *stopwords*.

Figura16-: Tabela de *stopwords*

Atenção: As customizações feitas na lista de *stopwords* são armazenadas no navegador (*local storage*) e serão perdidas se você limpar os dados de *cache* do navegador.

Essa lista customizada é compartilhada com todas as redes carregadas, diferente do restante das configurações, que são armazenadas por rede.

3.2.8 Número de termos extraídos nos campos textuais

No momento que a rede é carregada a aplicação faz uma leitura inicial (*parsing*) de termos contidos nos atributos textuais dos nós. A contagem desses termos pode auxiliar na escolha do atributo para geração da nuvem de palavras.

3.3 Reinicialização, backup e restauração

Na barra inferior do painel de configurações, à esquerda, estão os botões de manutenção da aplicação e da rede.



Figura17-: Botões de manutenção da aplicação e da rede

3.3.1 Limpar dados

Limpar todos os dados da aplicação que foram armazenados no *cache* do navegador para todas as redes utilizadas (*stopwords*, pesquisas avançadas, visualizações e configurações customizadas).

3.3.2 Reiniciar rede

Reiniciar a rede atual com parâmetros padrões, excluindo do *cache* do navegador todas as pesquisas avançadas, visualizações e configurações customizadas da rede atual.

3.3.3 Restaurar backup

Restaurar todos os dados da rede usando um arquivo de backup (*stopwords*, pesquisas avançadas, visualizações e configurações customizadas)

3.3.4 Backup da rede

Criar uma cópia de segurança dos dados da rede atual em um arquivo JSON que pode ser salvo em uma pasta de sua escolha. Serão salvos as *stopwords*, pesquisas avançadas, visualizações e configurações customizadas da rede atual.

Importante: O arquivo de backup é gerado para a rede que está aberta no momento e só pode ser utilizado para restaurar essa mesma rede. Se for utilizado para restaurar uma outra rede podem ocorrer erros na aplicação.

O insightNet Browser conta com uma interface simples e intuitiva. Todos os botões e controles da aplicação exibem dicas de suas funcionalidades quando passamos o mouse sobre eles.

Vale lembrar que o insightNet Browser não conta com algoritmos de análise de redes e, portanto, não agrega nenhum dado ao arquivo GEXF, armazenando customizações apenas nas memórias física e virtual do navegador. **O arquivo GEXF nunca é alterado pela aplicação.** O objetivo é ser um ambiente de visualização, navegação e consulta dos dados contidos no arquivo GEXF.

Imagine a rede como um grande site onde os nós seriam as páginas e os conteúdos dessas páginas seriam os atributos e metadados de cada nó. A representação gráfica da rede, nesse caso, seria uma espécie de mapa do site ou menu principal. Além de você poder navegar visualmente pelas páginas (nós) desse site (rede), com alguma prática, você também conseguirá montar consultas (*queries*) visuais e obter informações que não estão explícitas nos metadados da rede.

As aplicações de tratamento e análise de redes contam com interfaces complexas e priorizam as funcionalidades de processamento das redes, dando pouca atenção à visualização e interação com metadados da rede.

Atenção: Como a visualização de redes no navegador consome muita memória e processamento, recomendamos a visualização de uma rede por vez para obter um melhor desempenho. Abrir várias redes em abas ou janelas do navegador poderá ocasionar erros no navegador ou interrupção de scripts da aplicação.

4.1 Visualizando uma rede existente

Quando você inicia a utilização do insightNet Browser algumas redes de exemplo já estão incluídas e podem ser acessadas pelo menu de redes, na barra de título da aplicação (veja [Menu de redes](#)). Novas redes podem ser adicionadas à aplicação, como explicamos a seguir.

4.2 Acrescentando novas redes

Para que uma rede seja carregada e visualizada na aplicação, primeiramente, o arquivo GEXF dessa rede deve estar na pasta **networks** da aplicação (Veja *Estrutura de pastas*).

Importante: Os nomes dos arquivos de rede GEXF não podem conter caracteres especiais, acentos ou espaços em branco e devem estar, preferencialmente, em letras minúsculas.

4.2.1 Via URL

As redes são carregadas no browser por meio da adição do nome do arquivo de rede GEXF (sem a extensão ‘.gexf’) ao final da URL da aplicação, precedido do símbolo # (jogo da velha ou *hashtag*).

Um exemplo: suponhamos que você tenha descompactado o insightNet Browser na raiz do seu HD (C:/insightnet-browser) e deseja carregar a rede do arquivo **pihara.gexf** para visualização.

1. Primeiro você deve copiar seu arquivo **pihara.gexf** para a pasta **networks** (C:/insightnet-browser/networks);
2. Após abrir a aplicação no navegador Firefox (Veja *Executando o insightnet Browser*), na barra de endereços do navegador, você deve digitar “#” mais o nome do arquivo “pihara” no final da URL, após o “/index.html”, sem a extensão “.gexf”, e clicar em **ENTER** para carregar a rede, como exibido na imagem a seguir.

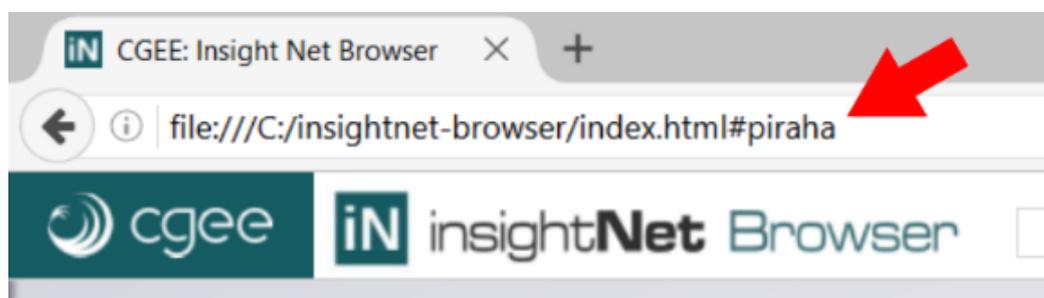


Figura18-: Carregando da rede via URL - Barra de endereços do navegador

4.2.2 Via Menu de redes

Após copiar seu arquivo de rede GEXF para a pasta **networks**, você deverá executar o *script* de geração do menu de redes, clicando no arquivo **menu_generator** específico para o seu sistema operacional:

menu_generator.bat - para Windows.

menu_generator.sh - para Linux.

O *script* será rapidamente executado e exibirá uma mensagem de conclusão. Clique em qualquer tecla para fechar a janela do *script*. Agora o nome da nova rede será exibido no menu de redes.

Importante: O insightNet Browser, por ser uma aplicação predominantemente Javascript e rodar apenas no navegador, não pode ter acesso aos arquivos do seu sistema operacional. Essa é uma restrição de segurança imposta por todos os navegadores aos *scripts* Javascript e que impede que a aplicação leia

automaticamente os arquivos GEXF que são colocados na pasta **networks** para geração do menu de redes. Esse processo deve ser manual.

O que o *script* de geração do menu de redes faz?

O *script* lê os arquivos do tipo GEXF contidos na pasta **networks** e gera, na mesma pasta, um arquivo chamado **networks.json** com os nomes dos arquivos da pasta. Quando a aplicação é acessada ela busca pelo arquivo **networks.json** e o utiliza para gerar o *menu de redes*.

Por sua vez, o menu de redes é apenas uma forma automatizada de incluir o nome do arquivo de rede na URL da aplicação e recarregar a página, no momento em que você escolhe uma rede.

4.3 Interação e navegabilidade

Para interagir com a rede e seus nós você utilizará o mouse e os *controles de navegação*. Você pode alterar o nível de zoom com o botão de rolagem do mouse (*mouse scroll*) ou no controle do nível de zoom e movimentar a rede na tela clicando, arrastando e soltando.

Para ter uma prévia dos dados de um nó basta posicionar o mouse sobre o mesmo e será exibida a janela de pré-visualização (caso esteja habilitada - Veja *Exibir dicas dos nós*) com alguns atributos do nó.

A ferramenta *Visão geral da rede (minitela)* também pode ser utilizada para aumentar ou diminuir o zoom e movimentar a rede pela tela.

4.4 Selecionando nós

4.4.1 Selecionando com um clique

Para selecionar um nó e ver todos os seus dados e ligações na *área de detalhamento*, basta clicar sobre o nó desejado.

Se houver uma partição de cor escolhida pelo usuário, sempre que um nó for selecionado, a partição da qual ele faz parte será destacada no *menu de partições*, no rodapé da tela. Para visualizar todos os nós de uma partição basta clicar sobre seu ícone no *menu de partições* e uma lista dos nós da partição será exibida na *área de detalhamento*.

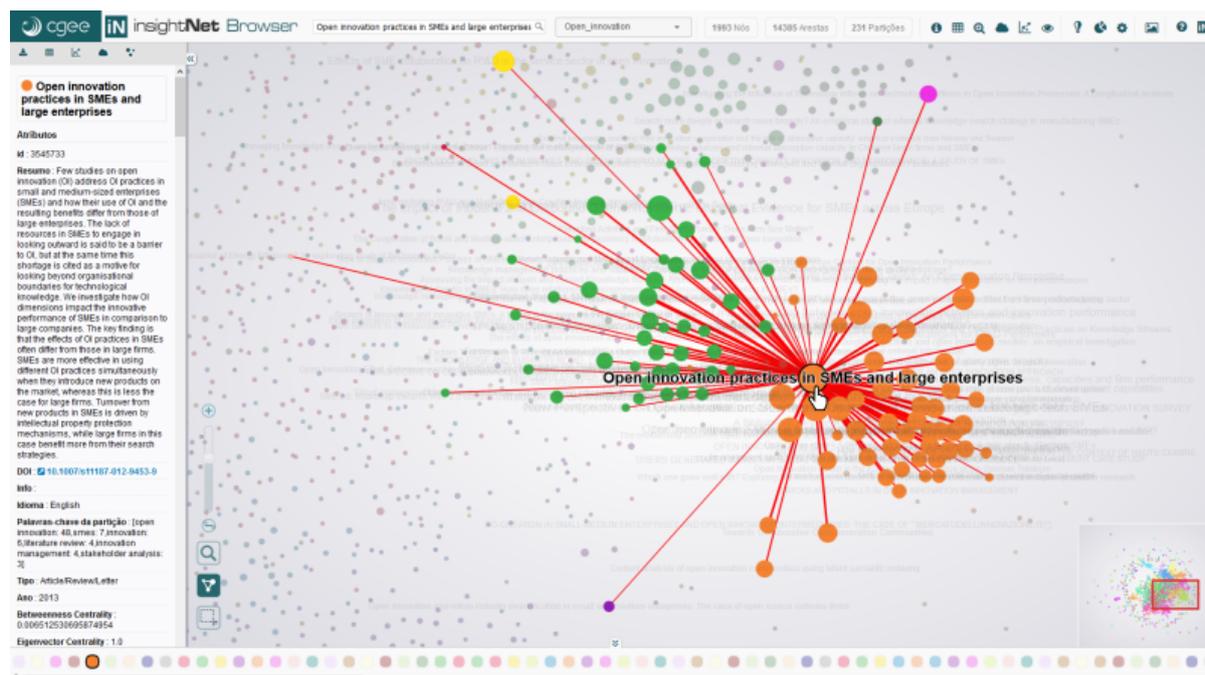


Figura19-: Selecionando um nó e visualizando seus dados na área de detalhamento

Dica: Para isolar a visualização apenas ao nó selecionado e suas ligações na rede, desative a opção de exibição de arestas no controle de navegação .

4.4.2 Selecionando com área de seleção

Você também tem a opção de selecionar um conjunto de nós desenhando uma área de seleção em volta deles.

Clique no controle para ativar ou desativar a ferramenta de seleção . Para desenhar a área de seleção clique e arraste até envolver os nós desejados e então solte para que a lista de nós selecionados seja exibida na área de detalhamento. Quando você seleciona um conjunto de nós a listagem dos mesmos é exibida na área de detalhamento, com isso você pode exportar esses nós e todos os seus atributos ou gerar visualizações (gráficos ou nuvem de palavras) desse conjunto, utilizando as *ferramentas da área de detalhamento*.

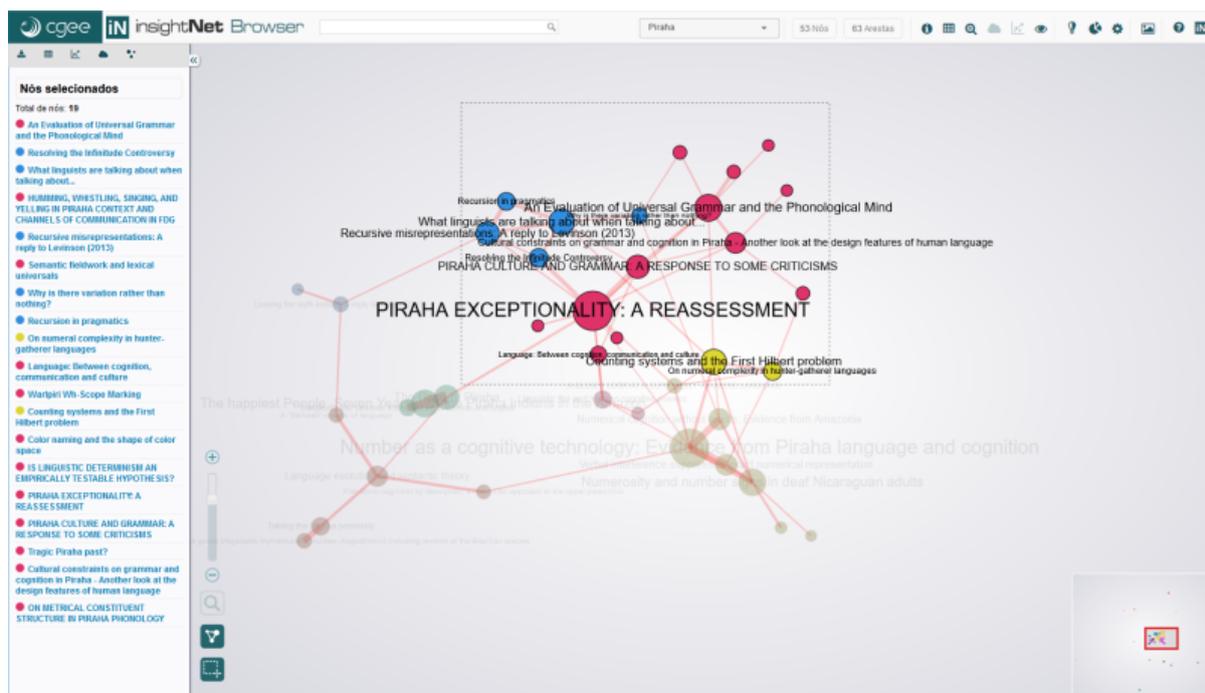


Figura20-: Selecionando um conjunto de nós com a ferramenta de seleção

Importante: Quando a ferramenta de seleção está ativada alguns botões do *menu principal* ficarão desabilitados para evitar conflitos de funcionalidades da aplicação.

4.4.3 Selecionando em gráficos

Outra opção extremamente útil para seleção de nós é por meio dos gráficos, tabelas e visualizações geradas pela aplicação, como: scatterplot, histogramas (barras e linhas), evolução de palavras-chave e nuvem de palavras.

Todos os gráficos e tabelas são interativos e seus elementos ou linhas são clicáveis (selecionáveis). Por exemplo, clicando em um elemento (barra, linha, ponto ou palavra) de qualquer gráfico exibirá uma listagem dos nós relacionados ao elemento clicado, na área de detalhamento. O mesmo ocorre nas tabelas da aplicação quando clicamos sobre uma linha.

Tanto em gráficos quanto em tabelas, sempre que você desejar selecionar mais de um elemento (barra, linha, ponto ou palavra), você deve pressionar a tecla **CTRL** e clicar nos elementos a serem adicionados à seleção.

4.5 Ferramenta lente de aumento

Utilize a ferramenta lente de aumento para ver detalhes na área de navegação. Para ativar ou desativar a lente de aumento, clique no botão , nos controles de navegação.

O insightNet Browser possui alguns módulos de visualização e manipulação dos metadados contidos nos nós da rede.

Além da visualização dos dados, os módulos permitem também a interação com os metadados, possibilitando a construção de consultas visuais por meio de cliques e outras formas de seleção utilizando o *mouse*.

Os módulos que veremos a seguir podem ser acessados pelo *menu principal* da aplicação, quando o alvo da visualização são todos os nós da rede, ou pelas *ferramentas da área de detalhamento*, quando o alvo são nós exibidos na *Área de detalhamento*.

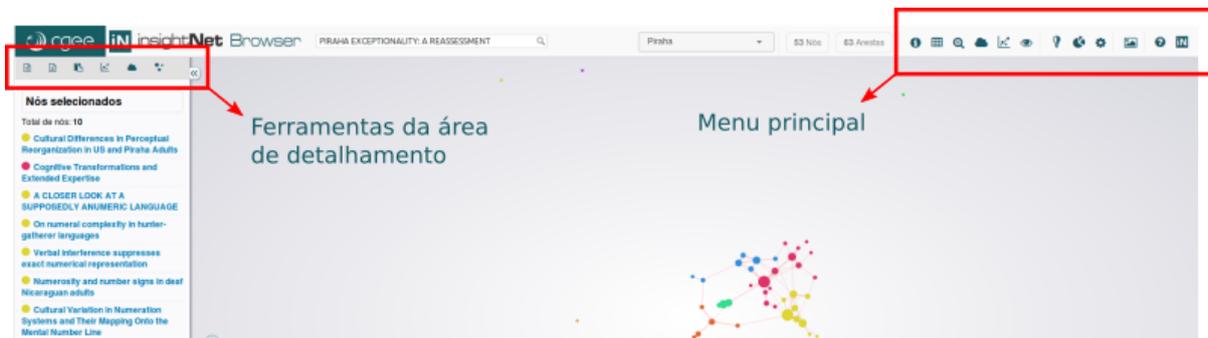


Figura21-: Acessando módulos pelos menus da aplicação

5.1 Metadados da rede

Este módulo pode ser acessado pelo botão , no *menu principal*, e fornece uma visão geral sobre todos os metadados encontrados nos nós e arestas da rede.

Metadados da rede							
Nós: 1993	Arestas: 14385	Nós com aresta: 1788 (90%)	Nós sem aresta: 205 (10%)	Pesos das arestas: 0.12000617 → 1.0000002		Atributos (nós): 12	Atributos (arestas): 3
☰ Atributos (nós e arestas)							
Atributo	Elemento	Tipo	Nós com o atributo	Nós sem o atributo	Valores únicos	Valores nulos	Extensão
Label	Nó	Texto	1993	0	1981	0	-
DOI	Nó	Texto	1145	848	1145	0	-
Type	Nó	Texto	1993	0	2	0	-
Info	Nó	Texto	0	0	0	1993	-
Year	Nó	Número	1993	0	15	0	2003 → 2017
Language	Nó	Texto	1993	0	13	0	-
Abstract	Nó	Texto	1881	0	1879	112	-
Betweenness Centrality	Nó	Número	1993	0	1547	0	0 → 0.021286027722576723
Modularity Class	Nó	Número	1993	0	231	0	0 → 230
Eigenvector Centrality	Nó	Número	1993	0	1738	0	0 → 1
Node keywords	Nó	Texto/Lista	1516	477	1506	0	-
Partition keywords	Nó	Texto/Lista	1879	114	118	0	-
Label	Aresta	Texto	14385	0	14385	0	-
Weight	Aresta	Número	14385	0	14300	0	0.12000617 → 1.0000002
Semantic similarity	Aresta	Número	14385	0	14300	0	0.12000617 → 1.0000002

Fechar

Figura22-: Janela do módulo de Metadados da rede

No topo da janela estão os números da rede: Totais de nós, arestas, nós com e sem arestas, atributos dos nós e arestas e a extensão dos pesos das arestas (pesos mínimo e máximo encontrados).

A principal informação que o módulo exibe é a tabela com os *status* dos atributos de nós e arestas. A tabela contém as seguintes colunas:

Atributo Nome de exibição do atributo, geralmente presente no campo label do atributo.

Elemento Indica se o atributo pertence aos nós ou arestas.

Tipo Tipo de dado do atributo: texto, número ou texto/lista.

Nós com o atributo¹ Número de nós que possuem o atributo, mesmo que com um valor nulo ou em branco.

Nós sem o atributo¹ Número de nós que não possuem o atributo.

Valores únicos Valores únicos encontrados em toda rede para aquele atributo. Útil para análise de atributos categóricos.

Valores nulos Valores nulos ou em branco encontrados em toda rede para aquele atributo.

Extensão Quando o atributo é numérico essa coluna irá exibir os valores mínimo e máximo encontrados.

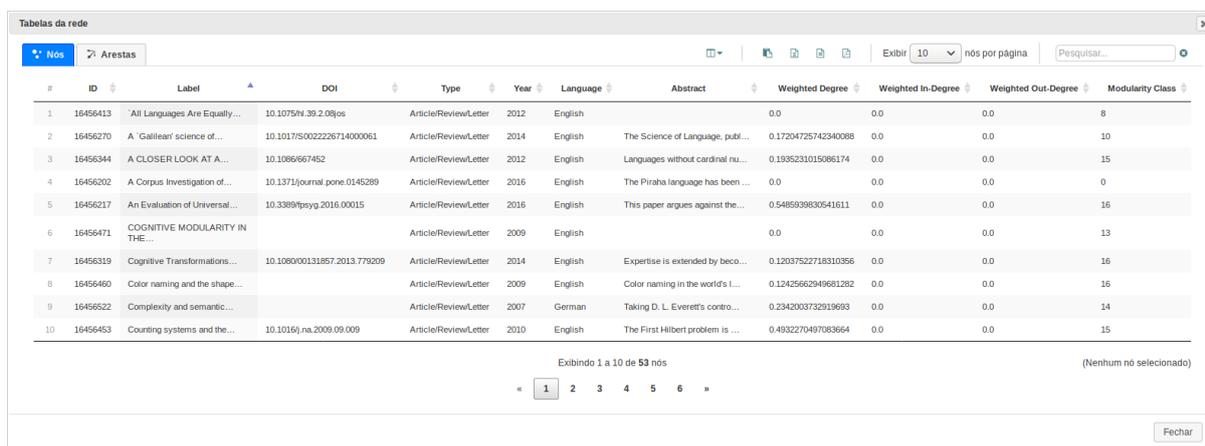
¹ Em alguns casos, alguns nós podem conter um determinado atributo com um valor válido, NULO ou em branco, enquanto em outros nós o mesmo atributo pode ser inexistente.

Importante: O processamento das informações exibidas nos Metadados da rede é executado utilizando a **Web Workers API**. Como o processamento é assíncrono e ocorre independente da aplicação, dependendo do tamanho da rede carregada, esse módulo pode demorar um pouco para carregar e processar todas as informações. Nesse caso o botão de acesso aos Metadados da rede iniciará desabilitado e ficará habilitado assim que o processamento for concluído e as informações estiverem disponíveis.

Web Workers são mecanismos que permitem que uma operação de um dado script seja executado em uma *thread* diferente da *thread* principal da aplicação Web. Saiba mais em https://developer.mozilla.org/pt-BR/docs/Web/API/Web_Workers_API.

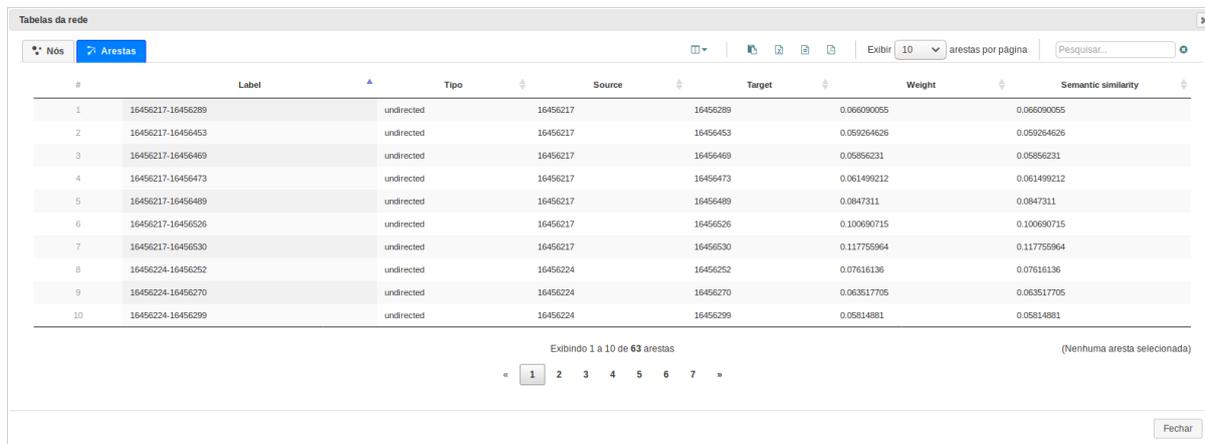
5.2 Tabelas de nós e arestas

Esse módulo pode ser acessado pelo botão , no *menu principal*, e exibe duas tabelas de listagem dos nós e arestas da rede com seus atributos.



#	ID	Label	DOI	Type	Year	Language	Abstract	Weighted Degree	Weighted In-Degree	Weighted Out-Degree	Modularity Class
1	16456413	All Languages Are Equally...	10.1075/nl.39.2.08jos	ArticleReviewLetter	2012	English		0.0	0.0	0.0	8
2	16456270	A 'Galilean' science of...	10.1017/S0022226714000061	ArticleReviewLetter	2014	English	The Science of Language, publ...	0.17204725742340088	0.0	0.0	10
3	16456344	A CLOSER LOOK AT A...	10.1096/667452	ArticleReviewLetter	2012	English	Languages without cardinal nu...	0.1935231015086174	0.0	0.0	15
4	16456202	A Corpus Investigation of...	10.1371/journal.pone.0145289	ArticleReviewLetter	2016	English	The Piraha language has been ...	0.0	0.0	0.0	0
5	16456217	An Evaluation of Universal...	10.3389/psyg.2016.00015	ArticleReviewLetter	2016	English	This paper argues against the...	0.5485939830541611	0.0	0.0	16
6	16456471	COGNITIVE MODULARITY IN THE...		ArticleReviewLetter	2009	English		0.0	0.0	0.0	13
7	16456319	Cognitive Transformations...	10.1080/00131857.2013.779209	ArticleReviewLetter	2014	English	Expertise is extended by beco...	0.12037522718310356	0.0	0.0	16
8	16456460	Color naming and the shape...		ArticleReviewLetter	2009	English	Color naming in the world's l...	0.12425662949681282	0.0	0.0	16
9	16456522	Complexity and semantic...		ArticleReviewLetter	2007	German	Taking D. L. Everett's contro...	0.2342003732919693	0.0	0.0	14
10	16456453	Counting systems and the...	10.1016/j.na.2009.09.009	ArticleReviewLetter	2010	English	The First Hilbert problem is ...	0.4932270497083664	0.0	0.0	15

Figura23-: Tabela geral de nós da rede



#	Label	Tipo	Source	Target	Weight	Semantic similarity
1	16456217-16456289	undirected	16456217	16456289	0.066090055	0.066090055
2	16456217-16456453	undirected	16456217	16456453	0.059264626	0.059264626
3	16456217-16456469	undirected	16456217	16456469	0.05856231	0.05856231
4	16456217-16456473	undirected	16456217	16456473	0.061499212	0.061499212
5	16456217-16456489	undirected	16456217	16456489	0.0847311	0.0847311
6	16456217-16456526	undirected	16456217	16456526	0.100690715	0.100690715
7	16456217-16456530	undirected	16456217	16456530	0.117755964	0.117755964
8	16456224-16456252	undirected	16456224	16456252	0.07616136	0.07616136
9	16456224-16456270	undirected	16456224	16456270	0.063517705	0.063517705
10	16456224-16456299	undirected	16456224	16456299	0.05814881	0.05814881

Figura24-: Tabela geral de arestas da rede

5.2.1 Selecionando linhas

Para selecionar uma linha da tabela basta clicar sobre ela. Clique novamente sobre a linha para retirar a seleção. Para selecionar mais de uma linha você pode usar os recursos de seleção do sistema operacional, com a combinação das teclas **SHIFT** e/ou **CTRL**.

A *Área de detalhamento* irá exibir os nós selecionados na tabela. Na tabela de arestas a área exibirá todos os nós das arestas selecionadas.

5.2.2 Exportação

Como em outras tabelas da aplicação você pode utilizar o menu de exportação para extrair os dados em 4 formatos: cópia para a área de transferência, planilha Excel, arquivo CSV ou PDF.



Figura25-: Menu de exportação da tabela geral de nós e arestas

Atenção: Apenas colunas visíveis serão incluídas na exportação.
--

5.2.3 Exibir/Ocultar colunas

Como o número de atributos (colunas) é variável, algumas tabelas podem ficar com a visualização prejudicada pelo grande número de colunas.

Para facilitar a visualização você pode ocultar algumas colunas que não são importantes. Para isso utilize o botão  .

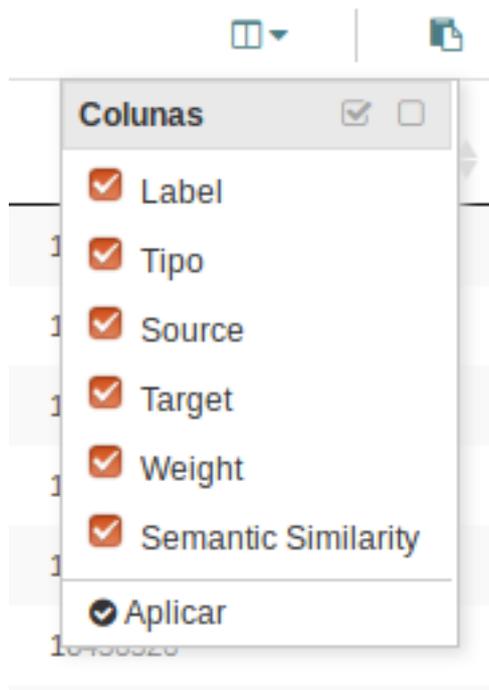


Figura26-: Exibir/Ocultar colunas da tabela geral de nós e arestas

Marque apenas as colunas que você deseja visualizar e clique em **aplicar**.

Dica: Utilize os ícones , no canto superior direito do menu de colunas, para marcar ou desmarcar todas ao mesmo tempo.

5.3 Pesquisa avançada

Esse módulo pode ser acessado pelo botão , no *menu principal*, e oferece uma interface de fácil compreensão e utilização tanto para pesquisas básicas em um atributo específico dos nós da rede, quanto para a aplicação que consultas avançadas envolvendo um ou mais atributos com operadores lógicos.

A janela de **Pesquisa avançada** é dividida em duas áreas: a área de **termos de busca**, onde você monta suas consultas incluindo ou retirando termos e operadores lógicos, e a área de **gerenciamento de pesquisas**, onde você pode salvar suas consultas e executá-las posteriormente com apenas um clique.



Figura27-: Janela do módulo de Pesquisa avançada

5.3.1 Termos de busca

Nessa área você pode fazer pesquisas simples de um ou mais termos em um atributo específico, utilizando ou não o operador lógico NOT (negação) ou utilizar vários termos ligados por operadores lógicos (AND, OR, AND NOT ou OR NOT), pesquisando em um ou vários atributos.

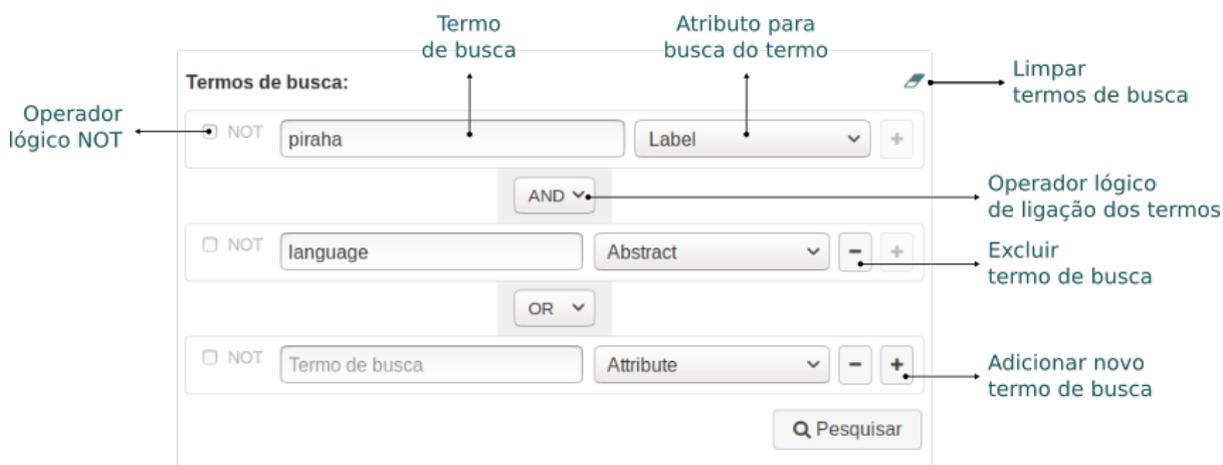


Figura28-: Área dos termos de busca

Sintaxe dos termos

Na utilização do campo **termo de busca**, considere os seguintes comportamentos:

- Quando mais de um termo é informado no campo **termo de busca** a aplicação considera que existe um operador **OR** entre os termos de pesquisa;
- Para cada termo informado a aplicação considera que há operadores curinga antes e depois da palavra, por exemplo: ao buscar pelo termo **set** a aplicação irá retornar resultados com os termos **settings**, **reset**, **setor** etc.

- Para pesquisar um termo ou frase exato, informe o mesmo entre aspas duplas (“”). Exemplo: ao buscar pelo termo “**set**” a aplicação irá retornar somente resultados que contenham o termo exato **set**.

Precedência dos operadores lógicos

Operações com **AND** sempre terão prioridade sobre as operações com **OR**. Como não há uma forma explícita de forçar a precedência, como fazemos em operações matemáticas com o uso dos parênteses (), é preciso reestruturar sua consulta para simular uma precedência arbitrária. Exemplo:

Suponhamos que você faça uma busca como mostrada na figura a seguir:

The screenshot shows the 'Pesquisa avançada' (Advanced Search) interface. On the left, there is a sidebar titled 'Nós selecionados' (Selected Nodes) with a list of 11 items. The main area is titled 'Pesquisa avançada' and contains a search form. The search terms are: NOT 'language' OR NOT 'piraha' AND NOT 'test', all with the attribute 'Label'. The interface includes a search bar, a 'Pesquisar' button, and a 'Salvar a pesquisa' (Save search) section. The 'Pesquisas salvas para essa rede' (Saved searches for this network) section shows a table with 2 saved searches:

#	Nome	Data
1	Pesquisa 1	05/11/2018 09:48:56
2	Pesquisa 2	05/11/2018 09:58:19

The table also shows 'Exibindo 1 a 2 de 2 pesquisas' (Showing 1 to 2 of 2 searches) and a pagination control.

Figura29-: Executando uma busca na Pesquisa Avançada

O resultado dessa busca será o conjunto de nós que contenham os termos **test** e **piraha** no atributo *label* ou o termo **language**, também no mesmo atributo.

Note que o operador **AND** é resolvido primeiro e depois o operador **OR**.

Suponhamos agora que você deseja fazer a mesma busca, porém forçando a precedência dos termos ligados pelo operador **OR**. Para isso você terá que exercitar um pouco seu conhecimento de lógica e reestruturar os termos para obter a precedência desejada.

O objetivo agora será obter os nós que contenham os termos (**language** ou **piraha**) e o termo **test**, todos no atributo *label*. Vamos então reestruturar a busca anterior da seguinte forma:

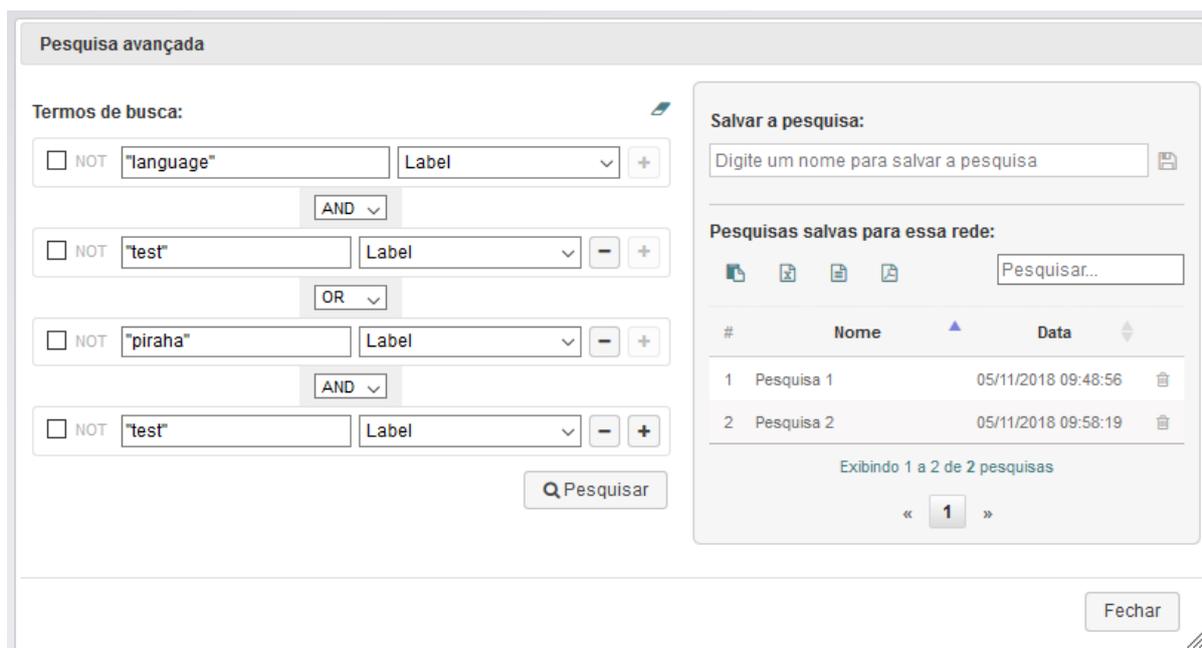


Figura30-: Reestruturando a busca na Pesquisa Avançada

5.3.2 Gerenciamento de pesquisas

Nessa área você pode salvar e gerenciar pesquisas, além de carregar e executar qualquer pesquisa salva com apenas um clique.

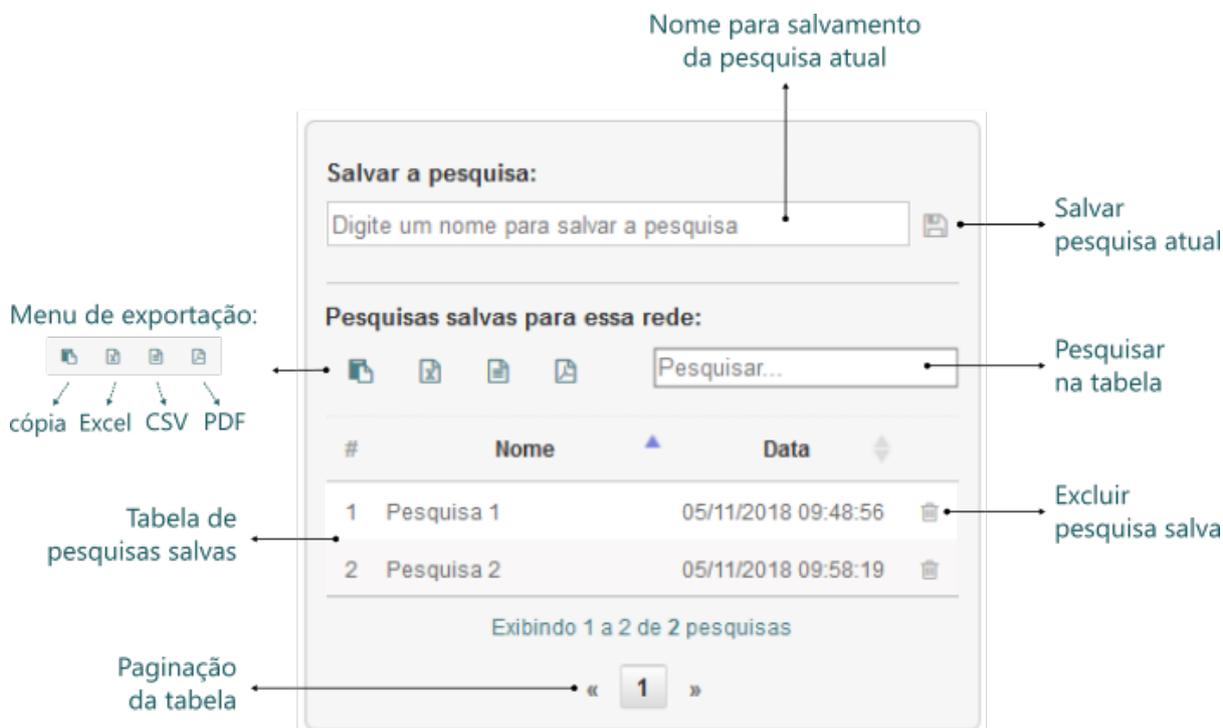


Figura31-: Área de gerenciamento de pesquisas

Para salvar uma pesquisa basta preencher a área dos termos de busca, testar a pesquisa, informar um nome único para a pesquisa no campo **Salvar a pesquisa** e clicar no ícone de salvamento ao lado . Após salva a pesquisa será exibida na tabela de pesquisas para a rede atual.

Dica: No momento do salvamento de uma pesquisa, caso você informe um nome que já exista em alguma pesquisa salva, a aplicação irá exibir uma mensagem de confirmação para sobrescrever ou não a pesquisa existente.

A tabela de pesquisas salvas permite a consulta das pesquisas armazenadas e sua execução imediata quando clicamos sobre o registro da pesquisa na tabela.

Você pode excluir uma pesquisa salva clicando no ícone de exclusão , na tabela de pesquisas, na linha do registro que se deseja excluir.

Atenção: Como o armazenamento de dados é feito no *cache* do navegador, todas as pesquisas salvas serão excluídas quando houver uma limpeza desse *cache*.

5.4 Nuvem de palavras

Esse módulo pode ser acessado pelo botão , no *menu principal*, quando o alvo são todos os nós da rede, ou pelas *ferramentas da área de detalhamento*, quando se deseja uma nuvem de palavras apenas dos nós exibidos na *Área de detalhamento*.

A nuvem de palavras pode ser gerada a partir de qualquer atributo textual dos nós. O atributo a ser utilizado é informado nas *configurações gerais da aplicação*, no item *Atributo para as Nuvens de palavras*.

Atenção: Antes de gerar a nuvem de palavras lembre-se de informar o *atributo* para a geração da nuvem e o *idioma da rede*, nas *configurações gerais da aplicação*. A configuração do idioma é importante pois define qual lista de *stopwords* serão extraídas da nuvem de palavras.

Quando o atributo escolhido for do tipo palavras-chave a aplicação não irá separar todo o conteúdo em palavras, como faz com os demais campos, mantendo os termos como eles foram informados no atributo, seja palavra ou frase.

Com a nuvem gerada você poderá visualizar a frequência de cada palavra passando o mouse sobre ela. Além de poder identificar visualmente as palavras mais frequentes no atributo previamente escolhido, você pode consultar os nós ligados a quaisquer palavras da nuvem com um clique sobre o termo.

Utilizando a tecla CTRL associada ao clique você poderá selecionar mais de uma palavra e consultar os nós relacionados à esse conjunto de palavras, com a opção de aplicar os operadores lógicos **AND** ou **OR** para a busca e exibição dos nós relacionados.

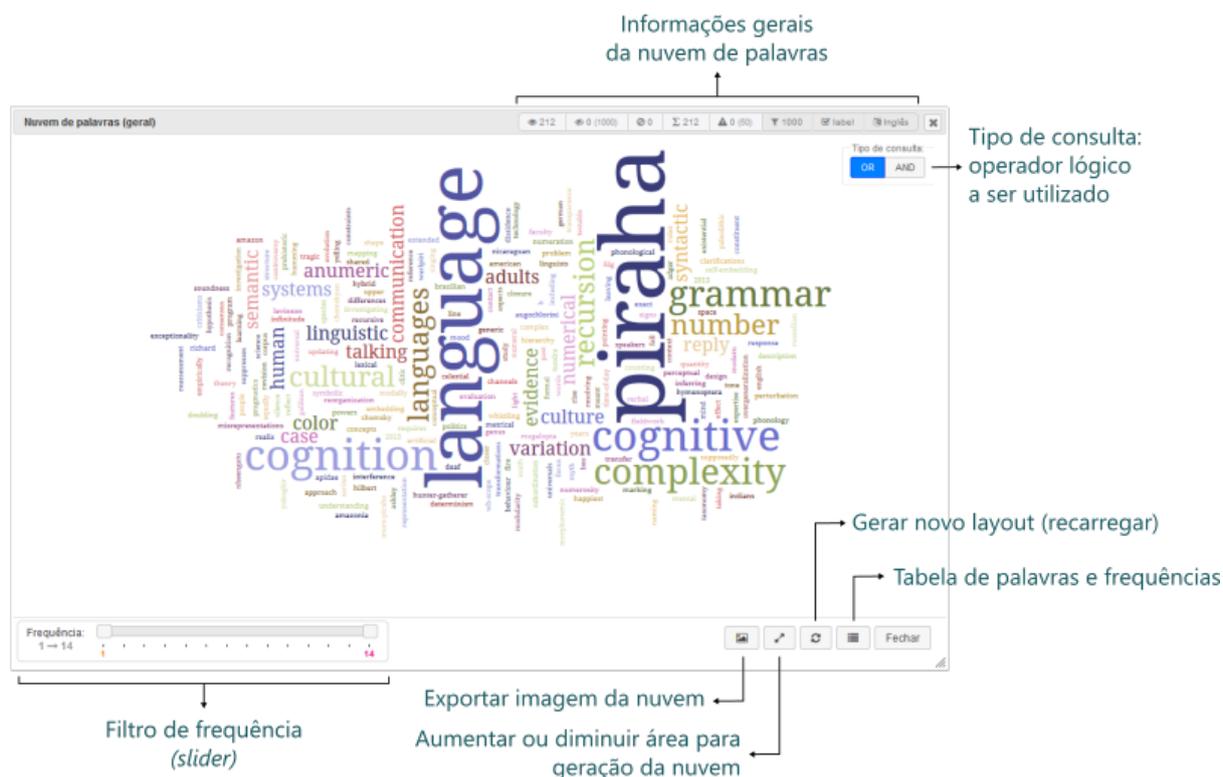


Figura32:- Módulo Nuvem de palavras insightNet Browser

5.4.1 Informações gerais da nuvem

Essa área exibe todos os metadados de geração da nuvem de palavras. Cada informação exibida é simbolizada com um ícone para fácil identificação visual. A seguir detalharemos cada item da área de informação.



Figura33:- Informações gerais da Nuvem de palavras

	Número de palavras que estão sendo exibidas na nuvem.
	Número Total de palavras descartadas pelo algoritmo de geração da nuvem (Veja o aviso Importante , abaixo).
	Número de palavras removidas da nuvem pela configuração de limite, filtro de frequência ou edição do usuário na tabela de frequências.
	Número total de palavras.
	Número de palavras descartadas pelo algoritmo da nuvem nas 50 primeiras posições, em ordem decrescente de frequência (Veja o aviso Importante , abaixo).
	Limite de palavras configurado pelo usuário. Esse limite pode ser alterado nas <i>configurações gerais da aplicação</i> , no item <i>Número máximo de palavras</i> .
	Atributo dos nós que gerou a nuvem de palavras. O atributo pode ser escolhido nas <i>configurações gerais da aplicação</i> , no item <i>Atributo para as Nuvens de palavras</i> .
	Idioma utilizado para remoção das <i>stopwords</i> . O idioma pode ser escolhido nas <i>configurações gerais da aplicação</i> , no item <i>Idioma da rede</i> .

Importante: Durante a geração da nuvem de palavras são feitas inúmeras tentativas de posicionamento dos termos pelo algoritmo de layout. Algumas vezes o processamento dessas tentativas pode ultrapassar o um tempo aceitável e palavras com grande frequência (palavras maiores) podem ser descartadas por não se “encaixarem” em nenhum lugar da nuvem. Então as tentativas são interrompidas para não ocorrer o travamento do navegador e essas palavras são deixadas de fora da nuvem, mesmo que sejam as mais frequentes.

Quando isso ocorre, essa informação é exibida para o usuário nas informações gerais da nuvem, com o ícone  e é necessário o recarregamento da rede até que os termos mais frequentes estejam todos presentes.

5.4.2 Seleção de palavras e consultas

A nuvem de palavras, assim como outros gráficos e visualizações da aplicação, oferece uma forma interativa para criação de consultas visuais (*queries*) por meio da seleção de elementos gráficos (nesse caso as palavras) com ações do mouse (nesse caso o clique). Ainda é possível utilizar os operadores lógicos **AND** ou **OR** na criação da consulta visual.

Os nós relacionados à consulta visual serão listados na *Área de detalhamento*. Para selecionar mais de uma palavra mantenha a tecla CTRL pressionada enquanto clica nas demais palavras. Para retirar a seleção de uma palavra basta clicar novamente sobre a mesma.

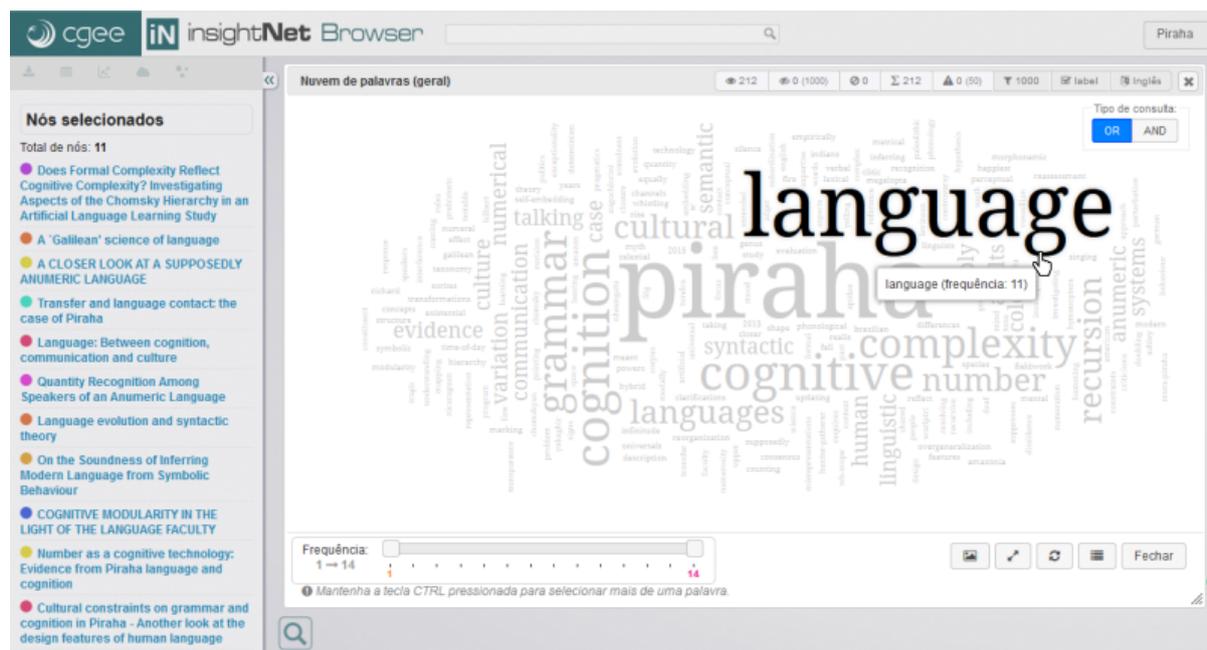


Figura34-: Selecionando uma palavra e consultando os nós relacionados

Nota: Inicialmente o campo **Tipo de consulta** não é exibido, ele só é exibido quando você inicia uma seleção, clicando em alguma palavra.

Tipo de consulta (operadores lógicos)

O campo **Tipo de consulta** será exibido sempre que você iniciar alguma seleção na nuvem de palavras (clicando em algum termo) e só influenciará os resultados quando mais de uma palavra estiverem selecionadas.

Veja, na figura a seguir, um exemplo ilustrativo de uma consulta visual em uma nuvem de palavras extraídas do atributo label dos nós:



Figura35-: Utilizando os operadores lógicos nas nuvens de palavras

5.4.3 Filtro de frequência

Com o filtro de frequências você pode estipular quais as frequências mínima e máxima das palavras que serão exibidas na nuvem.

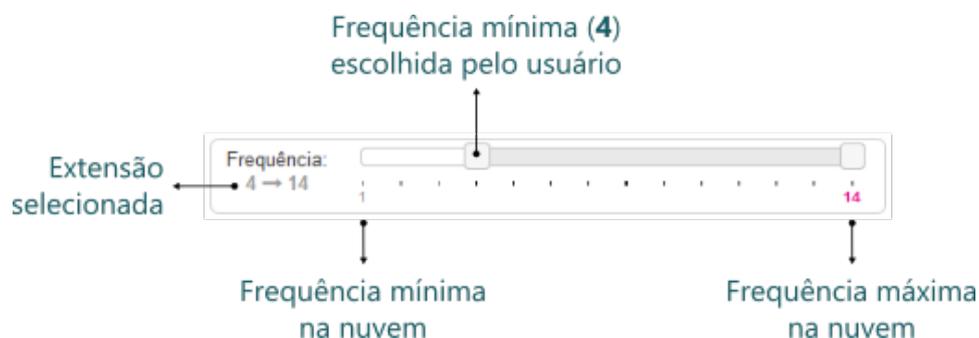


Figura36-: Filtro de frequências da nuvem de palavras

5.4.4 Tabela de palavras e frequências

Clicando no botão  você terá acesso à tabela de palavras e frequências. Por meio dessa tabela você poderá, além de consultar de forma rápida palavras na nuvem, ocultar ou exibir uma ou mais palavras. Essa tabela também exibe o status das palavras em relação à nuvem, se estão presentes na nuvem ou foram descartadas por algoritmo ou filtro.

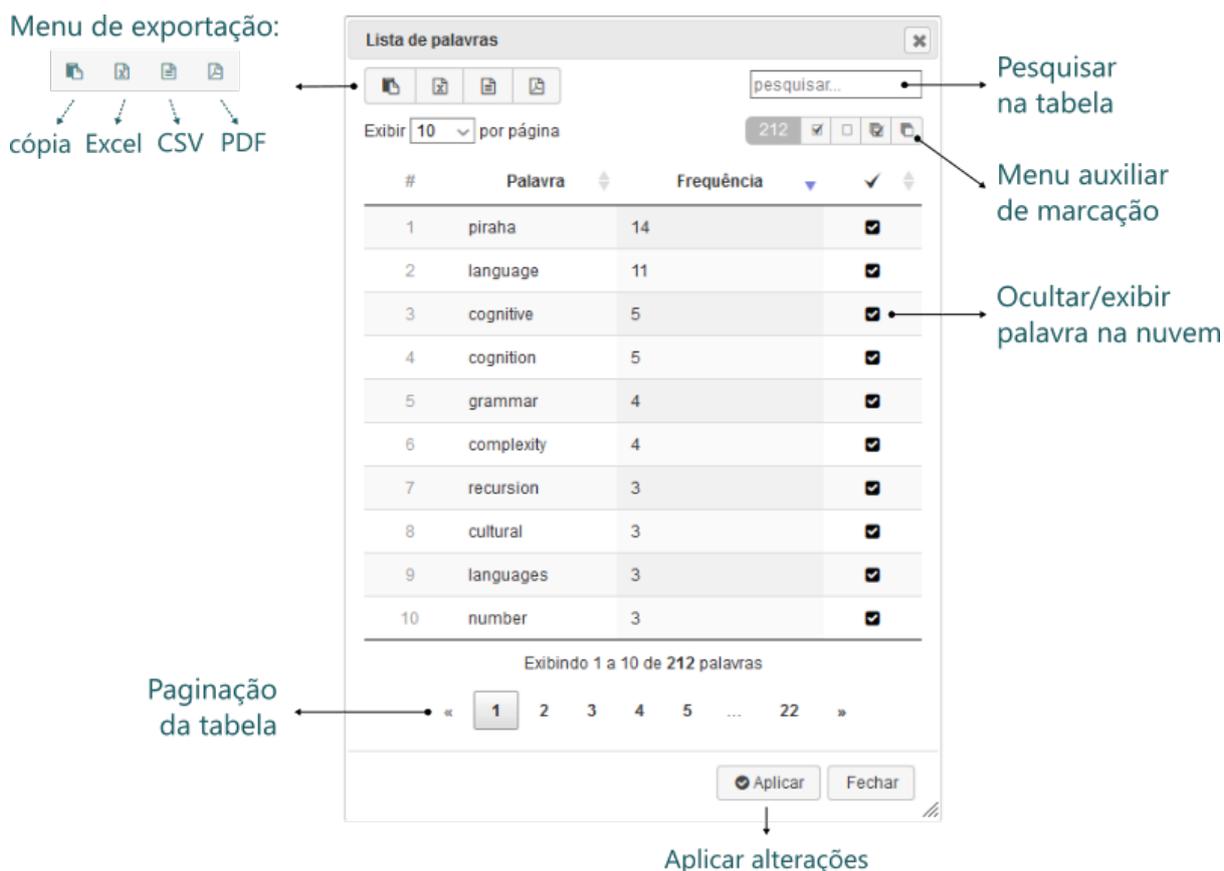


Figura37-: Tabela de palavras e frequências da nuvem de palavras

Assim como em outras tabelas da aplicação o **menu de exportação** possibilita a extração dos dados da tabela nos formatos: cópia para a área de transferência, planilha Excel, arquivo CSV ou PDF.

Para facilitar a marcação/seleção de palavras em lote, o **menu auxiliar de marcação** pode ser usado para marcar ou desmarcar todos os registros da tabela ou somente da página. Além de exibir o total de palavras selecionadas atualmente.



Figura38-: Menu auxiliar de marcação da tabela de palavras e frequências

Para ocultar ou exibir uma ou mais palavras, basta marcar/desmarcar as palavras desejadas e clicar no botão **Aplicar**. A nuvem será gerada novamente com as alterações aplicadas.

Dica: Utilizando o **menu de exportação** você poderá exportar essa tabela para uma planilha ou arquivo CSV e gerar a nuvem de palavras ou outros gráficos em outra aplicação como Excel ou Tableau.

5.4.5 Outras funcionalidades

Exportar nuvem de palavras

Você pode exportar a nuvem de palavras em uma imagem no formato PNG, clicando no botão .

Aumentar área da nuvem

Em nuvens com grande quantidade de termos ou termos mais frequentes muito grandes, o tamanho normal da tela (resolução) pode não ser suficiente para uma boa visualização da nuvem. Nesse caso você tem a opção de simular uma resolução maior e aumentar a área de impressão da nuvem, podendo comportar um número maior de palavras e controlar a visualização com funcionalidade de *zoom* (botão de *scroll* do mouse).

Para aumentar a área de impressão da nuvem e acrescentar mais espaço para os cálculos do algoritmo de layout clique no botão . Para desfazer a ação e retornar a área normal clique no botão novamente.

Gerar novo layout (recarregar nuvem)

Você pode recarregar a nuvem de palavras e gerar um novo layout a qualquer momento, clicando em . Essa opção também é muito útil quando alguma palavra desejada ficou de fora da geração da nuvem. Nesse caso você pode tentar recarregar a nuvem até que a palavra em questão apareça ou o layout mais adequado seja alcançado.

Saiba mais

A módulo Nuvem de palavras utiliza a biblioteca [D3.js](#) e o algoritmo de layout [D3 Cloud](#). Para mais informações sobre o funcionamento do algoritmo de layout e geração da nuvem de palavras acesse [How the Word Cloud Generator Works](#).

5.5 Gráficos da rede

Esse módulo pode ser acessado pelo botão , no *menu principal*, quando o alvo são todos os nós da rede, ou pelas *ferramentas da área de detalhamento*, quando se deseja um gráfico gerado de metadados apenas dos nós exibidos na *Área de detalhamento*.

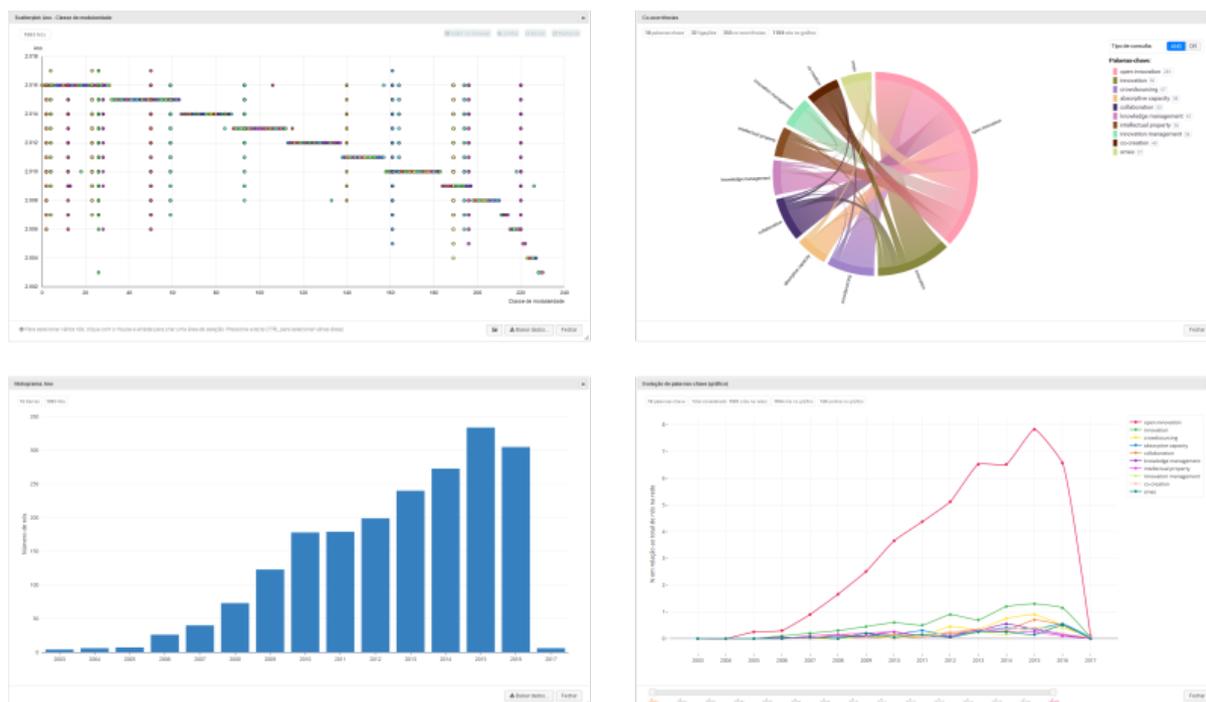


Figura39:- Exemplos de gráficos do insightNet Browser

Ao acessar essa opção a primeira janela exibida é o **menu de gráficos da rede** com os tipos de gráficos disponíveis para a rede e os parâmetros necessários para cada tipo de gráfico.

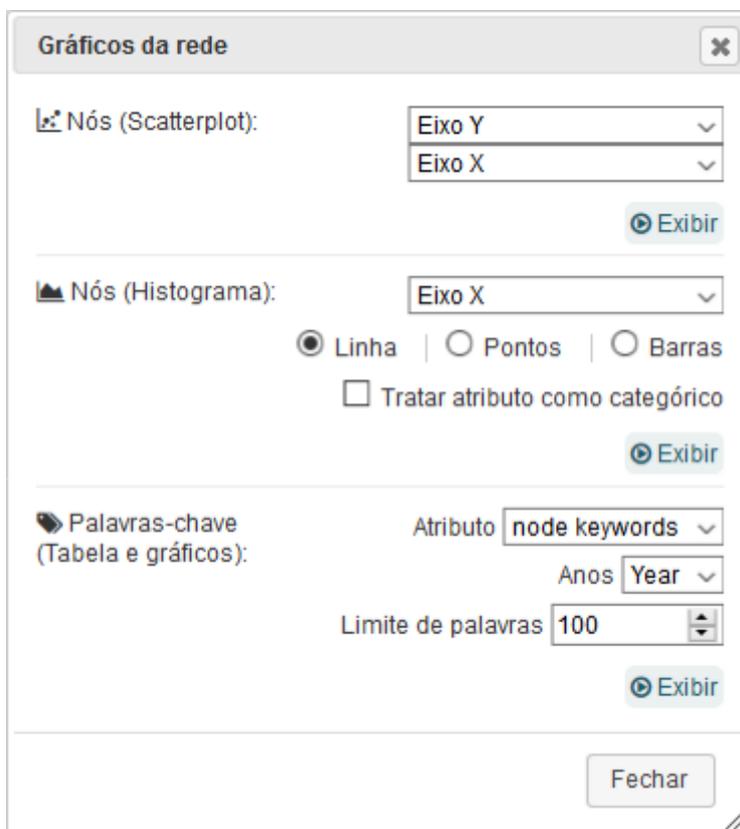


Figura40:- Menu de gráficos da rede

Cada tipo de gráfico ou visualização é gerado por um módulo diferente da aplicação. Iremos detalhar cada um deles nos próximos capítulos.

As visualizações relacionadas às palavras-chave só estarão disponíveis em redes cujos nós contenham atributos específicos de palavras-chave. Este atributo especial só é reconhecido pela aplicação como atributo do tipo palavra-chave quando uma formatação específica é atendida.

Os atributos do tipo palavra-chave são atributos customizados, inseridos nos nós da rede quando utilizamos o *plugin insightNet*, no *Gephi*.

Saiba mais

insightNet é um plugin para o aplicativo Gephi, desenvolvido pelo CGEE, que permite a geração de redes tendo como fonte de dados currículos da plataforma Lattes, artigos científicos de bases como a Web of Science e outros. Podem ser geradas redes de similaridade semântica, co-autorias entre outras.

Quando um atributo específico de palavras-chave não é encontrado na leitura da rede a aplicação irá exibir a seguinte mensagem no **menu de gráficos da rede**:



Figura41-: Mensagem quando não há visualizações para palavras-chave

5.5.1 Scatterplot

Scatterplot é o termo comumente utilizado para os **gráficos de dispersão**, com elementos representados por pontos. É um gráfico que auxilia principalmente na identificação de correlações entre dois atributos numéricos.

Antes de gerar o *scatterplot* você deverá informar quais atributos ocuparão os eixos X e Y do gráfico. A aplicação irá identificar os atributos numéricos dos nós e listá-los nos campos de escolha de eixo.



Figura42-: Escolhendo os eixos do gráfico scatterplot



Depois de escolher os atributos para os eixos do gráfico e clicar no botão **Exibir** o gráfico será gerado com os atributos escolhidos como eixos e os nós como pontos, como na figura a seguir.

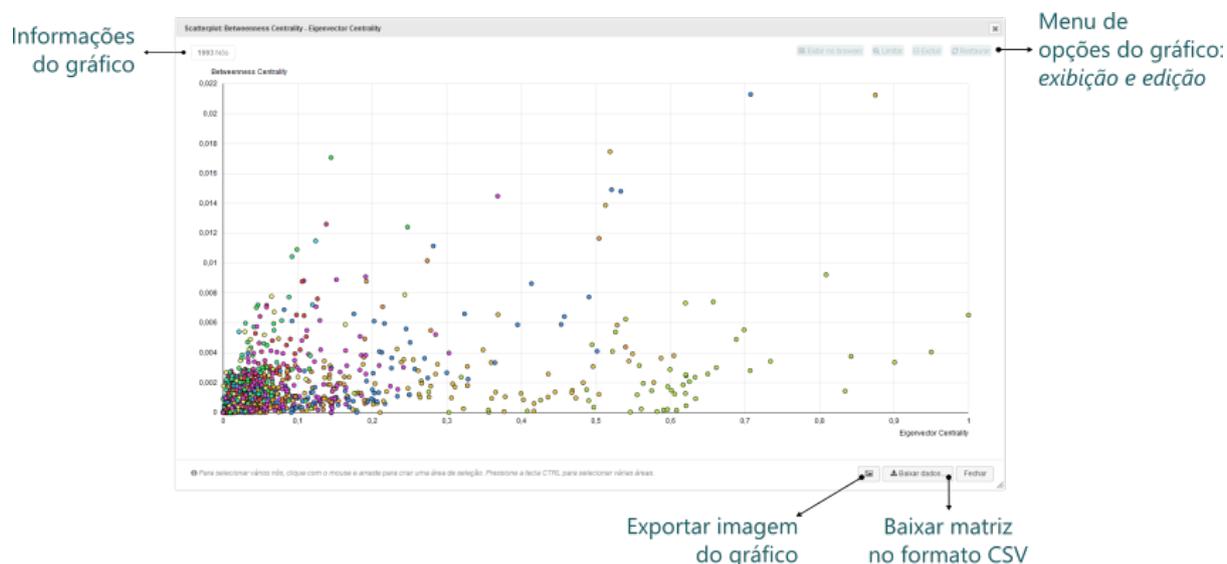


Figura43-: Gráfico scatterplot do insightNet Browser

No gráfico *scatterplot* você pode selecionar pontos tanto para a exibição de seus dados na *Área de detalhamento*, quanto para editar o gráfico, excluindo pontos ou limitando o gráfico aos pontos selecionados.

Para identificar o ponto (nó) sem precisar selecioná-lo, você pode posicionar o mouse sobre o ponto para visualizar seus ID, *label* e dados relacionados aos atributos escolhidos para os eixos do gráfico.

Selecionando pontos

Há duas formas de selecionar pontos no gráfico: clicando sobre o ponto ou desenhando uma área de seleção em volta dos pontos (clicando, arrastando e soltando). Em ambas as formas você pode manter a tecla **CTRL** pressionada para selecionar mais de um ponto ou grupo de pontos.

O **menu de opções** só é habilitado quando algum ponto é selecionado. Ao selecionar um ou mais pontos, informações quantitativas serão exibidas na **área de informações** do gráfico e o **menu de opções** será habilitado.

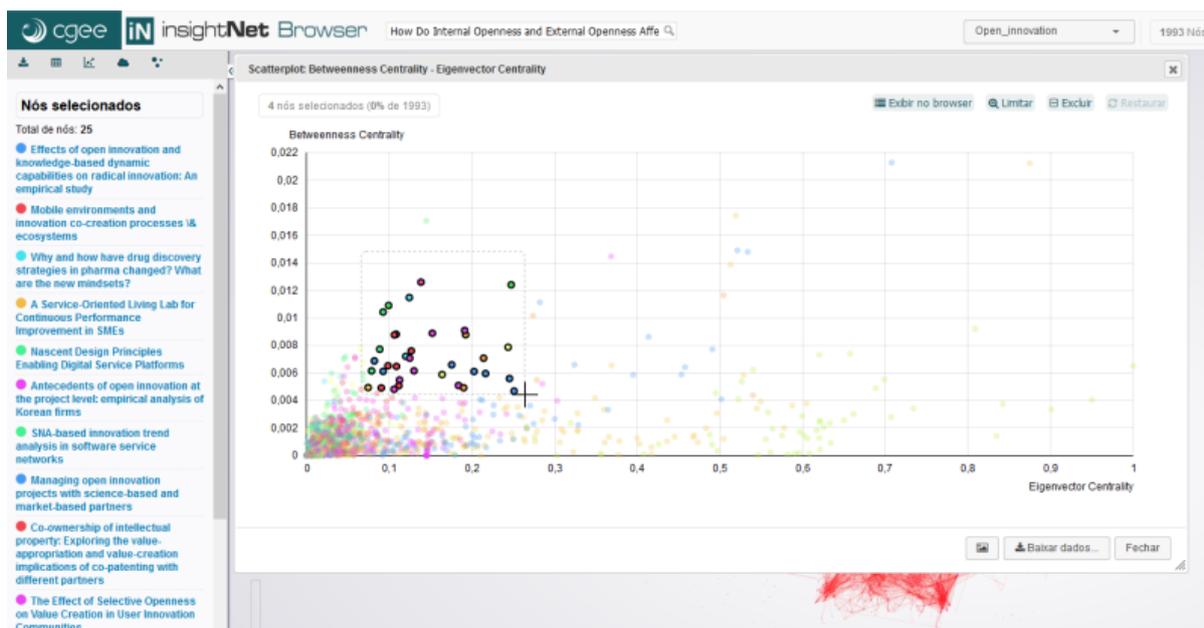


Figura44-: Selecionando pontos no gráfico scatterplot

Menu de opções (exibindo ou editando)



Figura45-: Menu de opções do gráfico scatterplot (exibir ou editar)

Exibir no browser Exibe, na *Área de detalhamento*, os dados do nó (ponto) selecionado ou uma lista de nós quando vários pontos estão selecionados.

Limitar Gera um novo gráfico apenas com os pontos selecionados.

Excluir Gera um novo gráfico sem os pontos selecionados.

Restaurar Desfaz as alterações e retorna ao gráfico inicial, com todos os nós iniciais.

Geram imagem do gráfico

Você pode gerar e salvar em seu computador uma imagem do gráfico no formato PNG. Basta clicar no

botão  no rodapé da janela do gráfico.

Baixar dados

Os dados exibidos no gráfico *scatterplot* podem ser baixados em uma matriz no formato CSV, clicando

no botão  no rodapé da janela do gráfico.

5.5.2 Histograma

O histograma, também conhecido como **distribuição de frequências**, é a representação gráfica em colunas ou em barras de um conjunto de dados previamente tabulado e dividido em classes uniformes ou não uniformes. É um gráfico que auxilia principalmente na quantificação de um atributo numérico ou categórico.

Antes de gerar o histograma você deverá informar qual atributo ocupará o eixo X do gráfico, o tipo de gráfico: linha, pontos ou barras e informar se o atributo deve ser tratado como um dado do tipo categórico.



Figura46-: Escolhendo as opções para geração do histograma

Dica: Ao marcar a opção **Tratar atributo como categórico** você forçará a aplicação a formatar o atributo como se fosse um dado do tipo texto. Essa opção pode ser útil em alguns atributos numéricos onde você queira evitar grandes espaços em branco entre as barras ou obter mais opções de ordenação dos eixos.

Depois de escolher as opções para a geração do histograma e clicar no botão  uma das três opções de gráfico será gerada com os parâmetros escolhidos, como na figura a seguir.

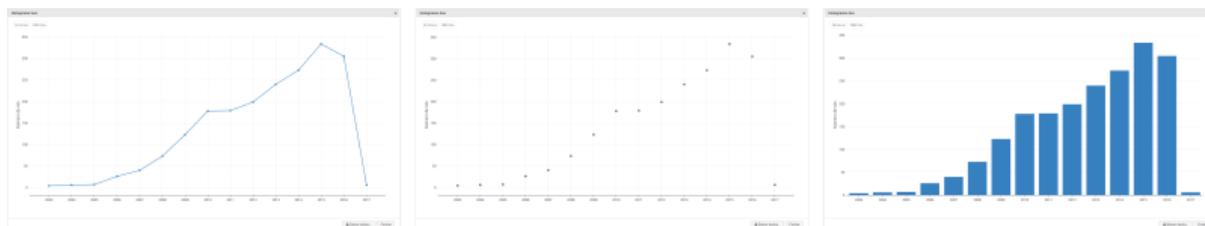


Figura47-: Opções de gráficos histograma: linha, pontos e barras.

Veja na figura a seguir os elementos que compõem a janela do gráfico:

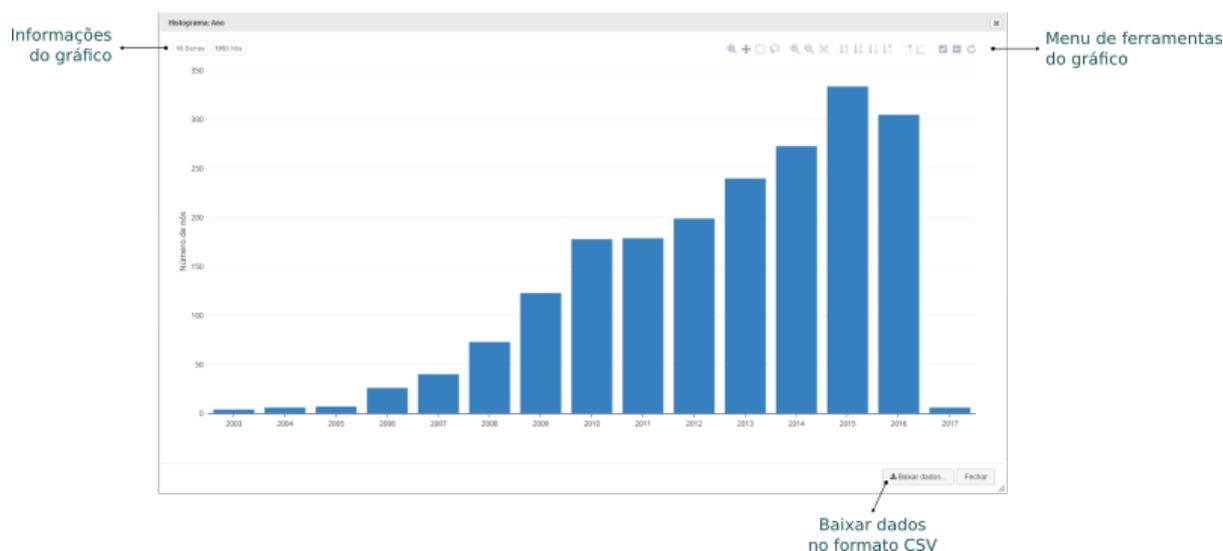


Figura48-: Gráfico histograma do insightNet Browser

No gráfico histograma você pode selecionar barras ou pontos tanto para a exibição de seus dados na *Área de detalhamento*, quanto para editar o gráfico, excluindo barras ou pontos ou limitando o gráfico às barras ou pontos selecionados.

Ao posicionar o mouse sobre a barra ou ponto você irá visualizar a quantidade de nós (eixo Y) que contém o mesmo valor para o atributo escolhido (eixo X).

Selecionando elementos

Há duas formas de selecionar barras ou pontos no gráfico: clicando sobre a barra ou ponto desejado ou uma das ferramentas de seleção do menu de ferramentas do gráfico. Independente da forma de seleção escolhida você pode manter a tecla **CTRL** pressionada para selecionar mais de uma barra ou ponto ou grupo de barras ou pontos.

Nota: O menu de ferramentas só é exibido quando o mouse é posicionado sobre a área do gráfico.

Ao selecionar uma ou mais barras ou pontos, informações quantitativas serão exibidas na **área de informações** do gráfico. Quando nenhuma barra ou ponto está selecionada essa área exibirá o número total de barras ou pontos no gráfico e o total de nós contidos nessas barras ou pontos.

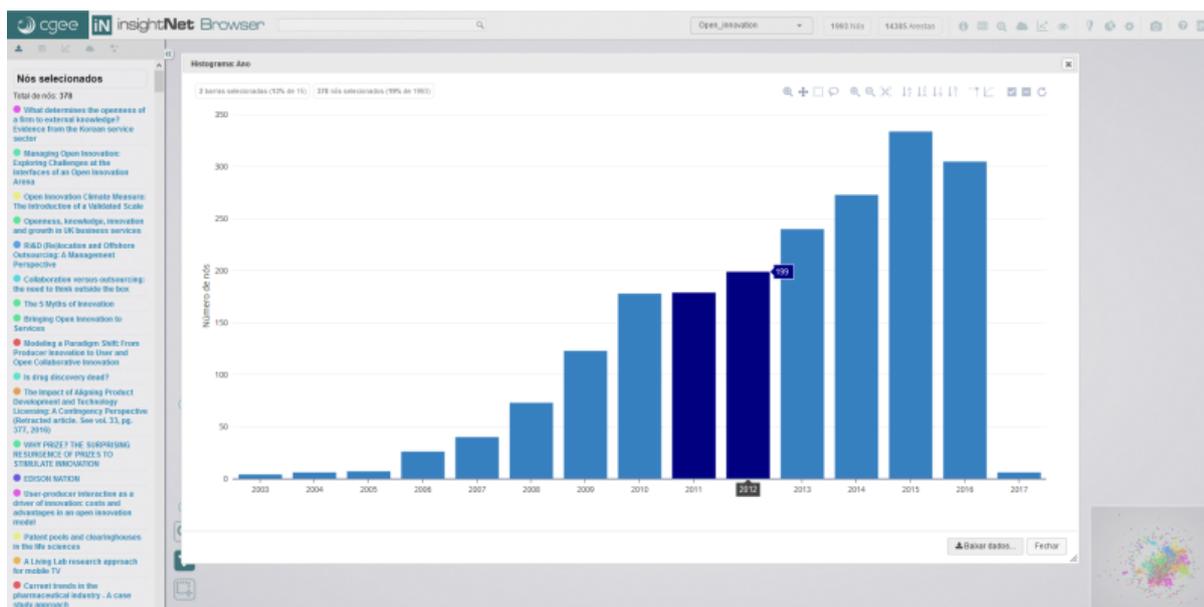


Figura49-: Selecionando barras no gráfico histograma

Menu de ferramentas



Figura50-: Menu de ferramentas do gráfico histograma

Com excessão aos *scatterplot* e nuvem de palavras, esse é o menu de ferramentas padrão dos gráficos da aplicação. As ferramentas disponíveis no menu podem mudar conforme o tipo de gráfico escolhido.

O menu está organizado em cinco grupos de ferramentas: **seleção**, **zoom**, **ordem**, **guias** e **edição**. Em seguida detalharemos cada uma das ferramentas:

Tabela4-: Menu de ferramentas

	Seleção de barras com clique ou aplicação de zoom em uma determinada área, clicando e arrastando na área desejada.
	Permite a movimentação do gráfico horizontalmente.
	Seleção de barras desenhando uma área retangular em volta dos elementos desejados.
	Seleção de barras desenhando uma área de forma livre em volta dos elementos desejados.
	Aproxima uma determinada área, ampliando o gráfico horizontalmente.
	Distancia uma determinada área, diminuindo o gráfico horizontalmente.
	Restaura a escala do gráfico, desfazendo as ações das ferramentas de zoom e movimentação do gráfico.
	Ordena o eixo X do gráfico de forma crescente.
	Ordena o eixo X do gráfico de forma decrescente.
	Ordena o eixo X do gráfico pelos valores do eixo Y de forma crescente.
	Ordena o eixo X do gráfico pelos valores do eixo Y de forma decrescente.
	Ativa ou desativa a exibição de linhas guias na barra ou ponto quando o mouse é posicionado sobre o mesmo.
	Ativa ou desativa a aplicação de logaritmo no eixo Y para facilitar a visualização de valores muito dispersos.
	Limita o gráfico aos elementos selecionados (barras).
	Retira do gráfico os elementos selecionados (barras).
	Restaura todos os elementos do gráfico, desfazendo qualquer edição.

Baixar dados

Os dados exibidos no gráfico podem ser baixados no formato CSV, clicando no botão

 **Baixar dados...**

no rodapé da janela do gráfico.

5.5.3 Palavras-chave

Nos arquivos de rede (GEXF) gerados pelo plugin **CGEE insightNet** para o Gephi, podem existir campos/atributos que contenham listas de palavras-chave relacionadas ao currículo, artigo etc. Essas palavras-chave são extraídas e processadas pelo plugin e inseridas nos arquivos de rede, em um formato bem específico, que também pode ser reconhecido pelo insightNet Browser como um atributo de palavras-chave.

Para obter melhor aproveitamento deste dado, partindo do acompanhamento da evolução das palavras, há uma tabela de evolução e três gráficos de quantificação e qualificação das palavras-chave.

Nota: Enquanto a *nuvem de palavras* separa qualquer texto em palavras simples, os atributos do tipo palavras-chave contém termos pré-estabelecidos que, em alguns casos, podem ser frases com muitas palavras cada.

Antes de gerar a tabela de palavras-chaves você deverá informar o atributo dos nós que contém as palavras-chave (campo com formato especial reconhecido pela aplicação), o atributo que contém o dado ano no formato numérico (inteiro) e a quantidade de palavras-chave a exibir na tabela. Um número grande de palavras pode prejudicar o desempenho da tabela que será exibida.

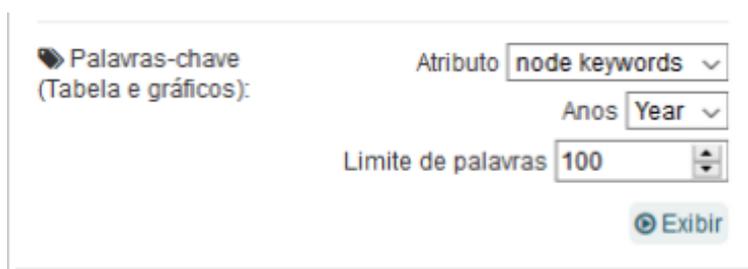


Figura51-: Escolhendo as opções para visualização das palavras-chave

Depois de escolher as opções para visualização das palavras-chave e clicar no botão



a tabela de evolução será exibida com o número de palavras escolhido.

Atenção: Lembre-se que se a tabela e gráficos foram chamados da *área de detalhamento*, todos os dados e números exibidos se referem apenas aos nós contidos na área.

Tabela de evolução

#	Palavra-chave	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	total
1	open innovation	0	0	5	6	18	33	50	73	87	102	130	130	156	131	2	923
2	innovation	0	0	0	2	4	6	9	12	10	18	14	24	26	23	1	149
3	crowdsourcing	0	0	0	0	0	0	2	0	2	9	7	15	18	10	0	63
4	absorptive capacity	0	0	0	0	0	2	2	3	6	2	6	8	7	11	1	48
5	collaboration	0	0	0	0	0	3	1	3	2	4	5	6	14	10	0	48
6	knowledge management	0	0	0	1	0	1	2	5	2	5	6	11	7	3	0	43
7	intellectual property	0	0	0	0	2	3	4	5	2	2	7	5	5	2	0	37
8	innovation management	0	0	0	0	1	0	1	2	2	5	5	3	8	8	1	36
9	co-creation	0	0	0	0	0	1	1	0	1	5	7	7	8	4	1	35
10	smes	0	0	0	0	1	0	4	1	3	1	5	5	3	10	0	33

Exibindo 1 a 10 de 100 palavras-chave (Nenhuma palavra-chave selecionada)

« 1 2 3 4 5 ... 10 »

Histograma Gráfico de evolução Co-ocorrências (experimental) Fechar

Figura52-: Tabela de evolução de palavras-chave

Selecionando linhas

Para selecionar uma linha da tabela basta clicar sobre ela. Clique novamente sobre a linha para retirar a seleção. Para selecionar mais de uma linha você pode usar os recursos de seleção do sistema operacional, com a combinação das teclas **SHIFT** e/ou **CTRL**.

A *Área de detalhamento* irá exibir os nós que contenham pelo menos uma das palavras selecionadas.

Exportação

Como em outras tabelas da aplicação você pode utilizar o menu de exportação para extrair os dados em 4 formatos: cópia para a área de transferência, planilha Excel, arquivo CSV ou PDF.



Figura53-: Menu de exportação da tabela de evolução de palavras-chave

Atenção: Todas as colunas (visíveis ou ocultas) serão incluídas na exportação.

Exibir/Ocultar colunas

Como o número de atributos (colunas) é variável, algumas tabelas podem ficar com a visualização prejudicada pelo grande número de colunas.

Para facilitar a visualização você pode ocultar algumas colunas que não são importantes. Para isso utilize o botão  .

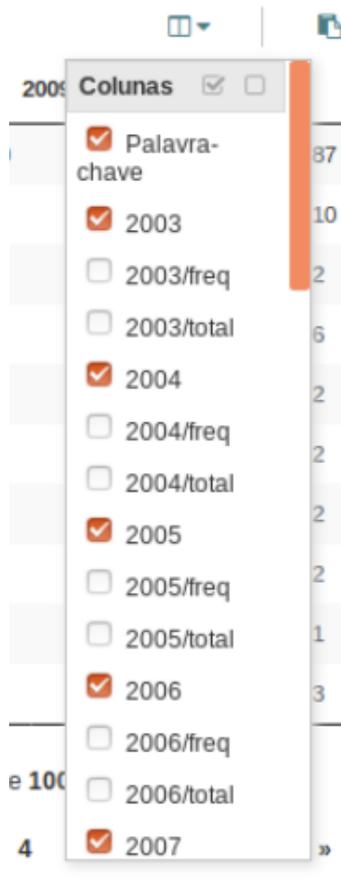


Figura54-: Exibir/Ocultar colunas da tabela de evolução de palavras-chave

Marque apenas as colunas que você deseja visualizar e clique em **aplicar**.

Dica: Utilize os ícones , no canto superior direito do menu de colunas, para marcar ou desmarcar todas ao mesmo tempo.

Opções de gráficos



Figura55-: Opções de gráficos de palavras-chave

No rodapé da janela da tabela de evolução de palavras-chave você encontra o menu de opções de gráficos de palavras-chave. A seguir iremos detalhar cada opção.

Histograma

O histograma de palavras-chave é um gráfico de barras com as frequências das palavras na rede ou de um conjunto de nós da *área de detalhamento*.

Antes de gerar o histograma você deverá selecionar as palavras na tabela e clicar no botão

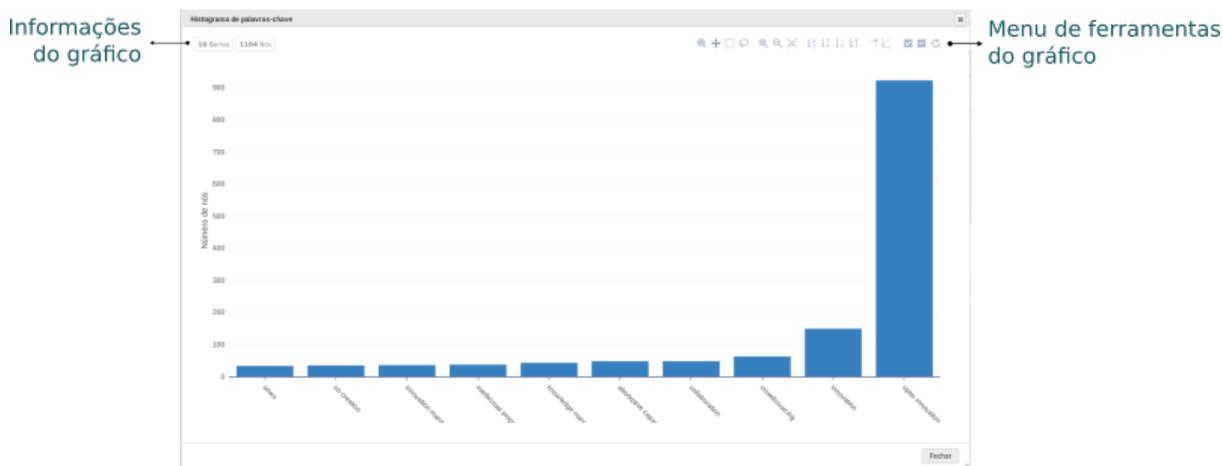


Figura56-: Gráfico histograma das palavras-chave

No gráfico histograma você pode selecionar barras tanto para a exibição de seus dados na *área de detalhamento*, quanto para editar o gráfico, excluindo barras ou limitando o gráfico às barras selecionadas.

Ao posicionar o mouse sobre a barra você irá visualizar a frequência (eixo Y) da palavra-chave (eixo X).

Selecionando elementos

Há duas formas de selecionar barras no gráfico: clicando sobre a barra desejada ou uma das ferramentas de seleção do menu de ferramentas do gráfico. Independente da forma de seleção escolhida você pode manter a tecla **CTRL** pressionada para selecionar mais de uma barra ou grupo de barras.

Nota: O menu de ferramentas só é exibido quando o mouse é posicionado sobre a área do gráfico.

Ao selecionar uma ou mais barras, informações quantitativas serão exibidas na **área de informações** do gráfico. Quando nenhuma barra está selecionada essa área exibirá o número total de barras no gráfico e o total de nós contidos nessas barras.

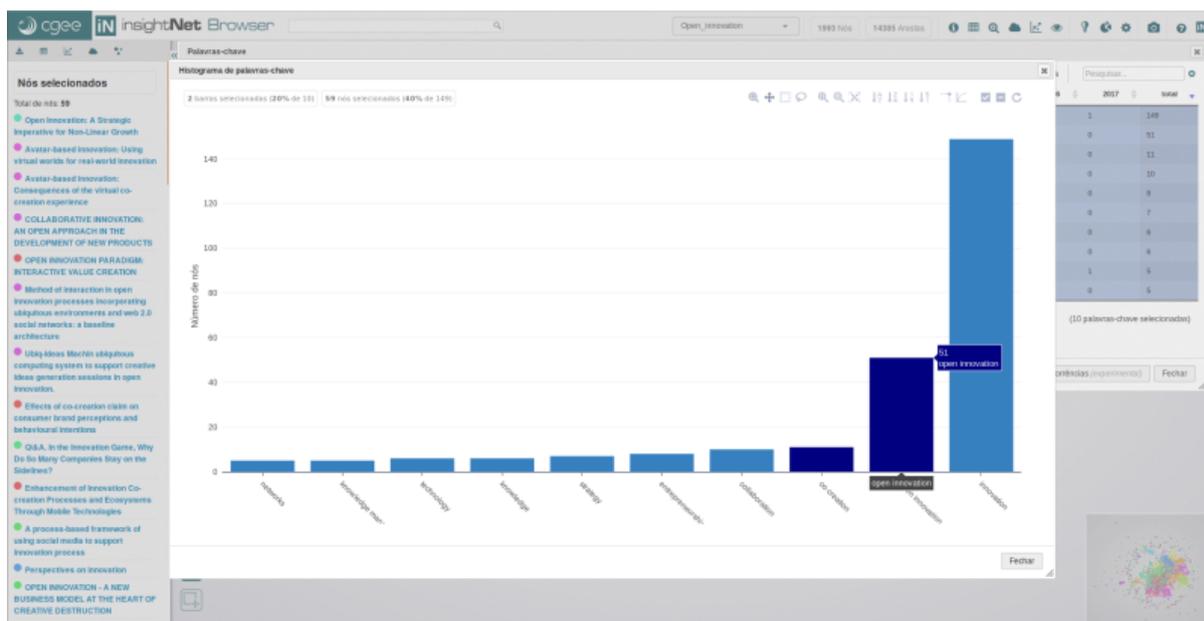


Figura57-: Selecionando barras no gráfico histograma de palavras-chave

Menu de ferramentas



Figura58-: Menu de ferramentas do gráfico histograma

Com exceção aos *scatterplot* e nuvem de palavras, esse é o menu de ferramentas padrão dos gráficos da aplicação. As ferramentas disponíveis no menu podem mudar conforme o tipo de gráfico escolhido.

O menu está organizado em cinco grupos de ferramentas: **seleção**, **zoom**, **ordem**, **guias** e **edição**. Em seguida detalharemos cada uma das ferramentas:

Tabela5-: Menu de ferramentas

	Seleção de barras com clique ou aplicação de zoom em uma determinada área, clicando e arrastando na área desejada.
	Permite a movimentação do gráfico horizontalmente.
	Seleção de barras desenhando uma área retangular em volta dos elementos desejados.
	Seleção de barras desenhando uma área de forma livre em volta dos elementos desejados.
	Aproxima uma determinada área, ampliando o gráfico horizontalmente.
	Distancia uma determinada área, diminuindo o gráfico horizontalmente.
	Restaura a escala do gráfico, desfazendo as ações das ferramentas de zoom e movimentação do gráfico.
	Ordena o eixo X do gráfico de forma crescente.
	Ordena o eixo X do gráfico de forma decrescente.
	Ordena o eixo X do gráfico pelos valores do eixo Y de forma crescente.
	Ordena o eixo X do gráfico pelos valores do eixo Y de forma decrescente.
	Ativa ou desativa a exibição de linhas guias na barra ou ponto quando o mouse é posicionado sobre o mesmo.
	Ativa ou desativa a aplicação de logaritmo no eixo Y para facilitar a visualização de valores muito dispersos.
	Limita o gráfico aos elementos selecionados (barras).
	Retira do gráfico os elementos selecionados (barras).
	Restaura todos os elementos do gráfico, desfazendo qualquer edição.

Evolução

O gráfico de evolução de palavras-chave é um gráfico de linhas com os percentuais das frequências das palavras em relação a um total escolhido pelo usuário, agrupados por ano.

Antes de gerar o gráfico você deverá selecionar as palavras na tabela e clicar no botão

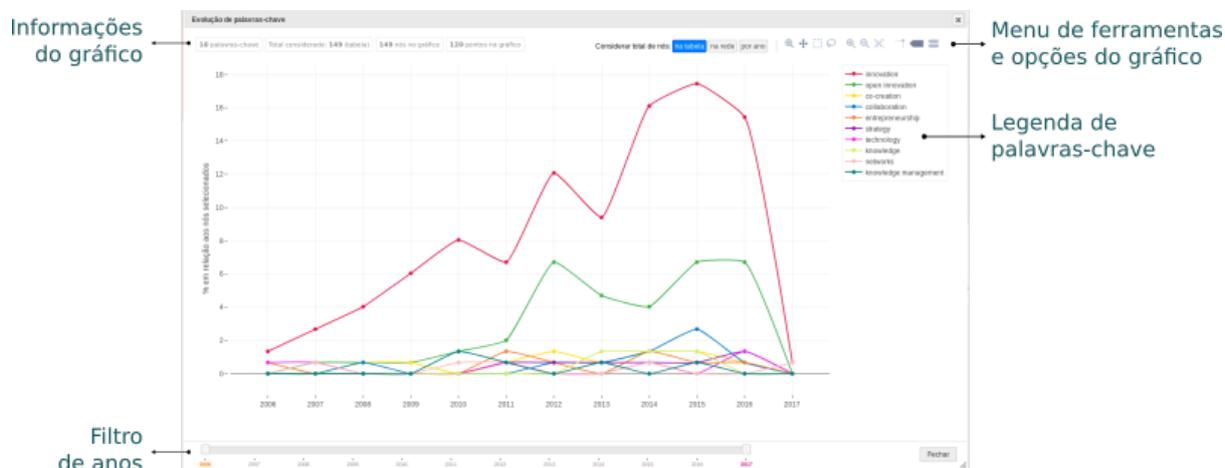


Figura59-: Gráfico de evolução de palavras-chave

No gráfico de evolução você pode clicar nos pontos para a exibição de seus dados na *área de detalhamento*. Cada palavra é representada por uma linha que pode ser ocultada ou exibida com um clique sobre sua legenda (à direita).

Ao posicionar o mouse sobre o ponto você irá visualizar as ocorrências (eixo Y) da palavra-chave (linha) em um determinado ano (eixo X).

Selecionando elementos

Há duas formas de selecionar pontos no gráfico: clicando sobre o ponto desejado ou utilizando uma das ferramentas de seleção do menu de ferramentas do gráfico. Independente da forma de seleção escolhida você pode manter a tecla **CTRL** pressionada para selecionar mais de um ponto ou grupo de pontos.

Nota: O **menu de ferramentas e opções** só é exibido quando o mouse é posicionado sobre a área do gráfico.

Ao selecionar um ou mais pontos, informações quantitativas serão exibidas na **área de informações** do gráfico. Quando nenhum ponto está selecionado essa área exibirá o número total de palavras-chave no gráfico, total de nós considerados (tabela, rede ou por ano), total de nós no gráfico e total de pontos.

Menu de ferramentas e opções



Figura60-: Menu de ferramentas e opções do gráfico de evolução

Esse menu está organizado em quatro grupos: **considerar total**, **seleção**, **zoom** e **guias**. Em seguida detalharemos cada uma das opções e ferramentas:

Total a considerar

O eixo Y do gráfico irá exibir os percentuais de frequências das palavras em relação ao total que você escolher na opção **considerar total de nós**. As opções são:

na tabela Considerar o total de nós contidos na tabela de evolução inicial.

na rede Considerar o total de nós na rede.

por ano Considerar o total de nós em cada ano.

A seguir detalharemos as demais ferramentas:

Tabela6-: Menu de ferramentas

	Seleção de barras com clique ou aplicação de zoom em uma determinada área, clicando e arrastando na área desejada.
	Permite a movimentação do gráfico horizontalmente.
	Seleção de barras desenhando uma área retangular em volta dos elementos desejados.
	Seleção de barras desenhando uma área de forma livre em volta dos elementos desejados.
	Aproxima uma determinada área, ampliando o gráfico horizontalmente.
	Distancia uma determinada área, diminuindo o gráfico horizontalmente.
	Restaura a escala do gráfico, desfazendo as ações das ferramentas de zoom e movimentação do gráfico.
	Ativa ou desativa a exibição de linhas guias na barra ou ponto quando o mouse é posicionado sobre o mesmo.
	Exibe apenas a informação de um ponto quando passamos o mouse sobre.
	Exibe as informações de todos os pontos no mesmo alinhamento vertical quando passamos o mouse sobre um ponto.

Co-ocorrências

O gráfico de co-ocorrências de palavras-chave é um gráfico do tipo *Chords*, onde os blocos na extremidade do círculo representam as palavras-chave e as “cordas” ou ligações no centro do círculo repre-

sentam agrupamentos de co-ocorrências entre as palavras. A espessura de cada “corda” representa a quantidade de co-ocorrências nela contidas.

Antes de gerar o gráfico você deverá selecionar as palavras na tabela e clicar no botão

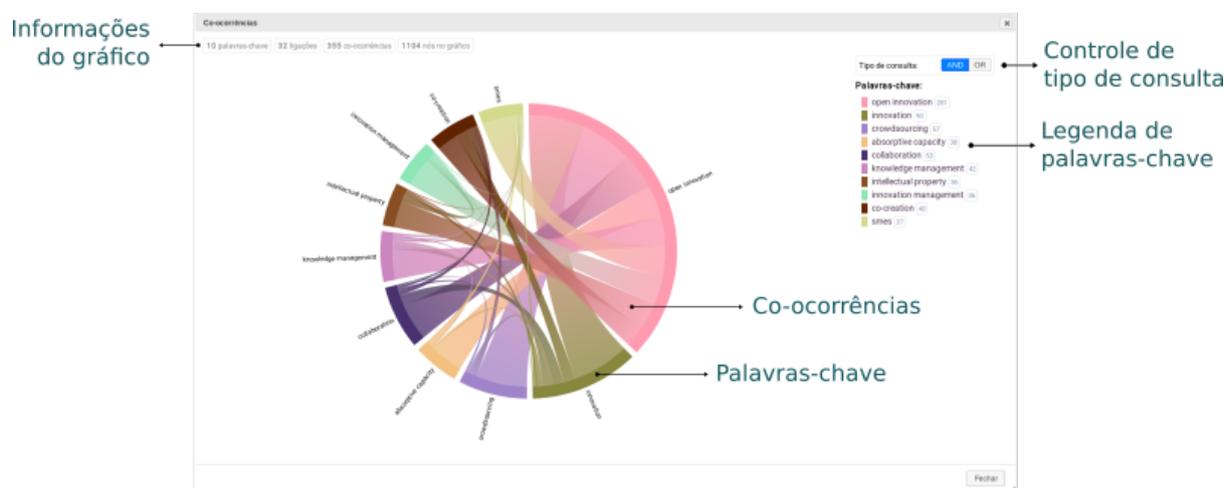


Figura61-: Gráfico de co-ocorrências de palavras-chave

No gráfico de co-ocorrências você pode clicar tanto nos blocos de palavras-chaves, quanto nas “cordas” para a exibição de seus dados na *área de detalhamento*.

Ao posicionar o mouse sobre um bloco de palavra-chave ou sobre uma “corda” você irá visualizar, no topo da janela, as co-ocorrências da palavra ou contidas na “corda”.

Quando o mouse é posicionado sobre um bloco de palavra-chave ou sua legenda, o total de co-ocorrências daquela palavra será exibido.

Selecionando elementos

Há dois tipos de elementos que podem ser selecionados no gráfico de co-ocorrências: palavras-chave (clitando nos blocos ou legendas) e grupos de co-ocorrências (clitando nas “cordas” ou ligações). Para selecionar mais de um elemento (do mesmo tipo) mantenha a tecla **CTRL** pressionada enquanto clica nos elementos desejados.

Ao selecionar um ou mais elementos, informações quantitativas serão exibidas na **área de informações** do gráfico. Quando nenhum elemento está selecionado essa área exibirá o número total de palavras-chave no gráfico, o total de ligações (cordas), o total de co-ocorrências e o total de nós com as palavras-chave do gráfico.

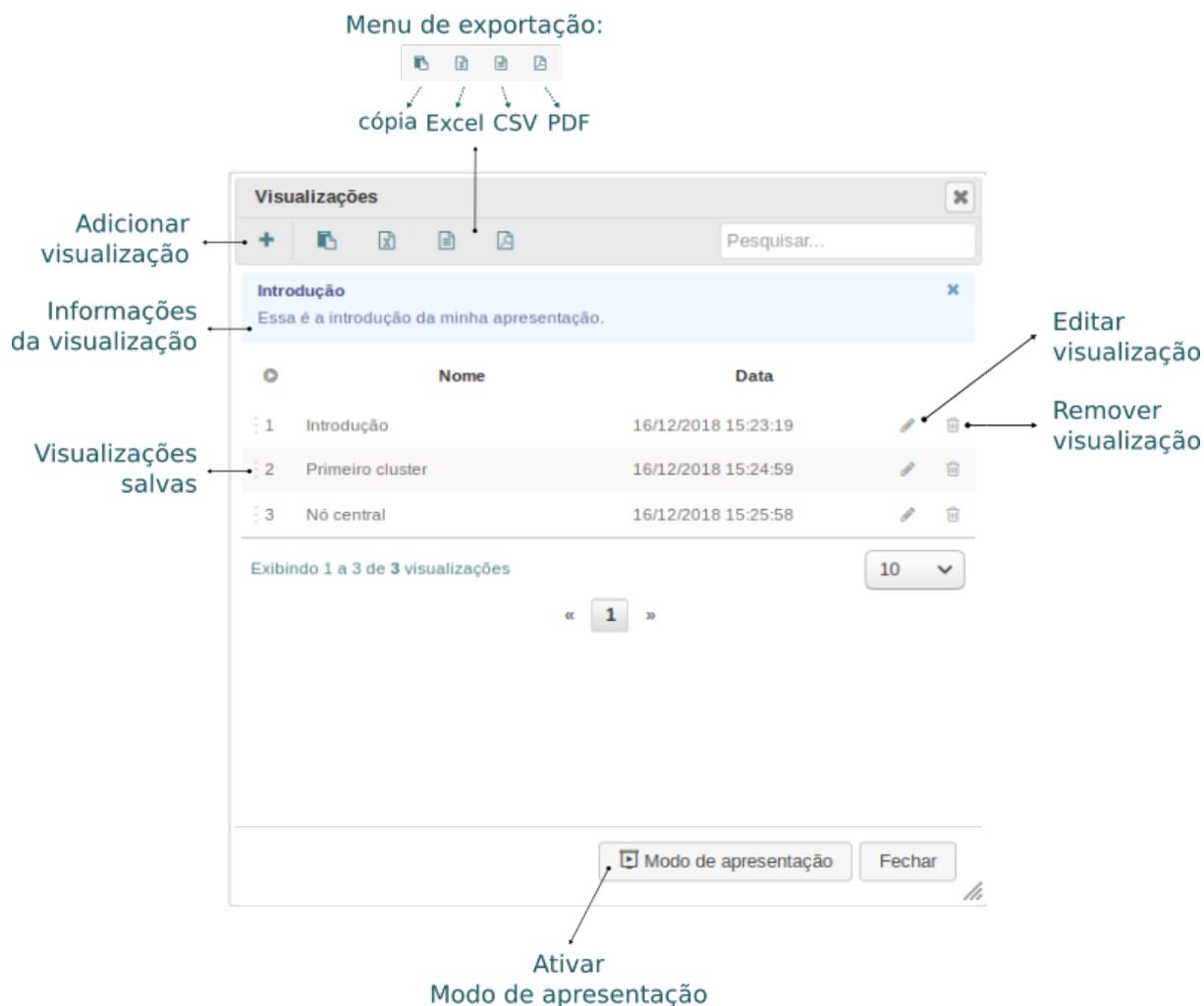


Figura64-: Janela de gerenciamento do módulo de visualizações

Nota: Você pode carregar qualquer visualização clicando sobre ela na tabela de visualizações salvas ou ativando o *modo apresentação*.

5.6.1 Gerenciando visualizações

Adicionar

Para salvar uma visualização, após posicionar a tela e outros elementos da forma desejada, clique no botão e a seguinte janela será exibida para o preenchimento dos dados a serem salvos.

Figura65-: Janela de adição/edição e edição de visualizações

Tabela7-: Campos de armazenamento das visualizações

Campo	Descrição
Nome	Nome único para a visualização a ser salva.
Descrição	Uma descrição ou anotação para a visualização.
Sequência	Posição arbitrária da visualização (para o modo de apresentação).

Editar

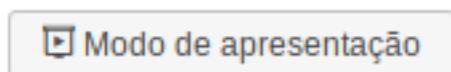
Para editar uma visualização salva, clique no ícone de edição , na linha da visualização que você quer alterar. A janela de edição que será exibida é a mesma mostrada anteriormente em **Adicionar**.

Remover

Para remover uma visualização, clique no ícone de exclusão , na linha da visualização.

5.6.2 Modo de apresentação

Após salvar e ordenar suas visualizações, para ativar o modo de apresentação, clique no botão



, no rodapé da janela de visualizações.

No modo de apresentação a aplicação irá ativar o modo de tela inteira do navegador e adicionar ao rodapé da aplicação um controle de apresentação.

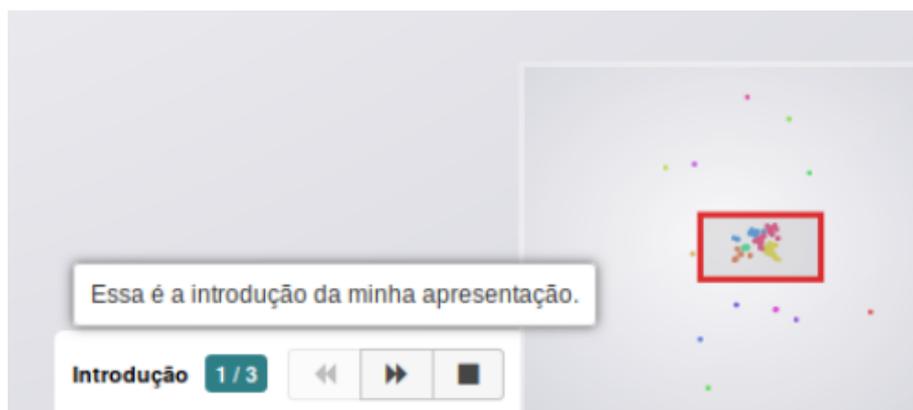


Figura66-: Controle do modo de apresentação

Você pode usar esses controles para navegar pelas visualizações da apresentação ou utilizar as setas do teclado. Passando o mouse sobre os controles você poderá visualizar dicas da visualização atual (descrição) ou os títulos das visualizações anterior e posterior.

Para desativar o modo de apresentação, clique no botão  (*stop*) do controle.

5.7 Partições

Esse módulo pode ser acessado pelo botão , no *menu principal*, e permite a customização das cores dos nós da rede, baseada em um atributo. Essa opção pode ser utilizada para criar partições visuais na rede, facilitando a visualização e seleção de conjuntos de nós por partição.

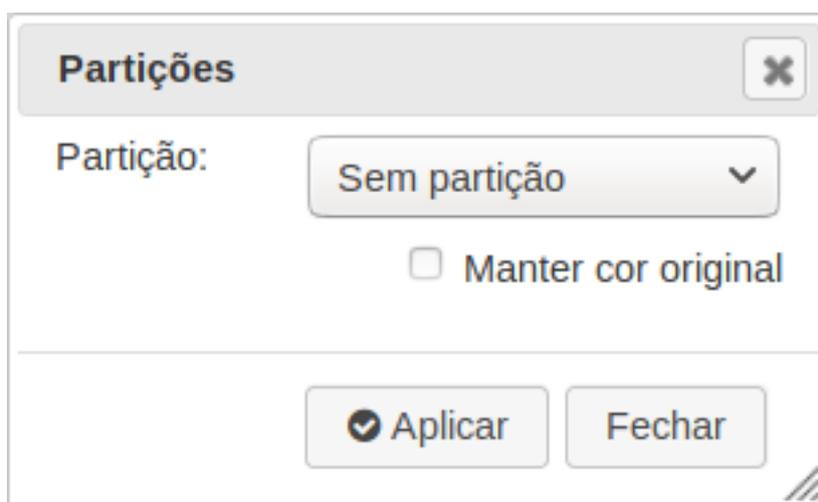


Figura67-: Tabela geral de nós da rede

Você pode manter as cores originais dos nós quando escolher uma partição. Isso pode ser útil quando queremos que a reconheça a partição original da rede e mantenha as mesmas cores.

Quando uma partição é escolhida a aplicação cria o *menu de partições* no rodapé da *área de navegação* da aplicação.



Figura68-: Menu de partições