



cgée

Relatório de evolução de desenvolvimento de ferramentas de monitoramento, análise e visualização de dados do CGEE.

Brasília, DF
dezembro, 2019

Relatório de evolução de desenvolvimento de ferramentas de monitoramento, análise e visualização de dados do CGEE.



Brasília, DF
dezembro, 2019

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

Luiz Arnaldo Pereira da Cunha Junior

Relatório de evolução de desenvolvimento de ferramentas de monitoramento, análise e visualização de dados do CGEE. Exploração de dados e visualização de informação. Brasília: Centro de Gestão e Estudos Estratégicos, 2019.

247p. : 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.cgee.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. Relatório de evolução de desenvolvimento de ferramentas de monitoramento, análise e visualização de dados do CGEE. Exploração de dados e visualização de informação. Brasília, DF: 2019. 247 p.

Este relatório é parte integrante das atividades desenvolvidas no âmbito do Contrato de Gestão. Atividade: Desenvolvimento de competências e ferramentas em prospecção, avaliação estratégica, gestão de informação e do conhecimento. Projeto: (Exploração de dados e visualização de informação). (8.10.56.01.51.01).

Relatório de evolução de desenvolvimento de ferramentas de monitoramento, análise e visualização de dados do CGEE.

Supervisão

Marcio de Miranda Santos

Coordenador

Jackson Max Furtunato Maia

Equipe técnica do CGEE

Alberto Akira Okata

Cesar Augusto Costa

Eduardo Amadeu Dutra Moresi

Flávia Montandon Fagundes Pinto

Icaro Lorrán Lopes Costa

Kleber de Barros Alcanfor

Rogério da Silva Castro

Consultor

Jorg Bliesener

SUMÁRIO

1. INTRODUÇÃO.....	2
2. ANÁLISE DE REDES COMPLEXAS.....	3
2.1 INSIGHT NET	3
2.2 INSIGHT NET BROWSER.....	5
3. MINERAÇÃO DE TEXTOS.....	6
4. NOVOS ALGORITMOS E PROTÓTIPOS	6
5. INTERAÇÃO COM OUTROS PROJETOS DO CENTRO E FUTUROS DESENVOLVIMENTOS.....	8
6. JOURNAL CLUB.....	9
7. PARTICIPAÇÃO EM EVENTO.....	10
8. ALTERAÇÕES NO PROJETO E PERSPECTIVAS PARA 2020	10
9. ANEXO I	18
10. ANEXO II	177

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 2019 no Projeto de Atividade. Dois dos 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 3.2.3" e "InsightNet Browser - Guia do usuário: Release 1.6.7", anexos I e II deste relatório, onde são descritas em detalhes todas as funcionalidades dos programas.

Também devem ser mencionados dois avanços importantes nas atividades do Projeto. Com o propósito de subsidiar a formulação das estratégias futuras de atuação do CGEE na análise e visualização de dados, informação e conhecimento foi instituída uma nova meta para organizar discussões técnicas sobre técnicas de mineração e análise de dados. Como resultado dessas discussões, novos algoritmos têm sido experimentados no Centro, incluindo o uso mais frequente de redes neurais e algoritmos genéticos. Nas seções seguintes são resumidas as principais atividades e avanços realizados em 2019 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 e mineração de dados com documentos do acervo próprio (integrado ao Insight Data), de patentes do escritório de patentes dos Estados Unidos da América (USPTO), de currículos Lattes e de metadados das principais bases de dados de produção acadêmica existentes.

2.1 INSIGHT NET

Esta linha é centrada no desenvolvimento de novas funcionalidades da ferramenta insightNet, dentre as quais ressaltam-se:

- a) Implantação e tratamento customizado de idiomas diferentes no CGEE Insight Net, passando dos anteriores 2 para aproximadamente 30. Essa funcionalidade, que consiste basicamente em eliminar dos cálculos os termos poucos relevantes de cada idioma, melhora a acurácia do cálculo das redes de similaridade e aumenta substancialmente os potenciais casos de uso da

ferramenta, particularmente para bases de dados como o Lattes, que têm conteúdos com mais de um idioma;

- b) Exibição da produção acadêmica detalhada de cada pesquisador prospectado nas análises de redes de currículos Lattes, com a explicitação dos idiomas declarados e detectados pela funcionalidade anterior. A funcionalidade também permite agrupar a produção total por partição escolhida pelo usuário (por exemplo, todos os pesquisadores de uma instituição ou agrupamento temático), exportar os dados selecionados para análise mais profundas e ainda mostra os metadados das produções com links para as páginas web das editoras, se houver;
- c) Realização dos primeiros testes da exibição de estatísticas de relevância de termos por nó ou por agrupamento em todos os tipos de rede (Lattes, Scopus, Web of Science ou planilhas preparadas pelo usuário), com a expectativa de melhorias na identificação de redes temáticas, anteriormente baseadas em palavras-chave;
- d) Implementação e testes de rotinas de envio de relatórios de execução de atividades realizadas por um usuário do insightNet para o servidor de CGEE, incluindo o registro de metadados das bases de dados importadas. Com essa funcionalidade, será possível melhorar o rastreamento de atividades do usuário para eventuais suportes técnicos, facilitar a reprodutibilidade das análises realizadas, uma vez que os metadados de cada análise específica são automaticamente armazenados e mitigar dificuldades de comunicação com usuários externos que foram incorporados à base de usuários do insightNet a partir do segundo semestre de 2019;
- e) Elaboração e execução de plano de transferência de conhecimento da empresa Intangibles Management para a equipe de TI do CGEE, o que reduz a dependência do Centro com relação à empresa;
- f) Incorporação de versão HTML do Manual na própria interface de usuário do sistema, também para facilitar sua operação por parte de usuários externos;

- g) Incorporação de melhorias nos cálculos de estatísticas, na identificação de dados duplicados e na seleção de produções específicas obtidas dos currículos Lattes;
- h) Preparação de material de treinamento para usuários externos ao CGEE;
- i) Entrega da versão 3.2.3 do programa (Manual em anexo).

2.2 INSIGHT NET BROWSER

No que diz respeito a atividades de consolidação e ampliação de metodologias para análise e visualização de dados, as principais atividades desenvolvidas foram:

- a) Entrega para testes e posterior consolidação de duas versões (baseadas em abordagens de desenvolvimento diferentes) de um futuro aplicativo similar ao InsightNet Browser com funcionalidades de navegação com visualizações em 3 dimensões;
- b) Correção no módulo "Evolução de palavras-chave" do InsightNet Browser, modificado para permitir a geração de gráficos complementares (histograma e coocorrências) mesmo na ausência do atributo "ano";
- c) Ajustes na arquitetura da aplicação para adequação a novas versões do Firefox;
- d) Criação de novos filtros de busca na "Tabela de nós e arestas" que possibilita a identificação de partições de nós na tabela de arestas;
- e) Melhoria no tratamento e exibição de erros de carga, muito comuns nas extrações de dados de fontes externas, como os dados abertos da Plataforma Sucupira, da CAPES, que geravam arquivos de redes mal formatados. Essa funcionalidade permite ao usuário encontrar e solucionar erros de formatação muito mais rapidamente pela má formatação de arquivos GEXF; e
- f) entrega da versão 1.6.6 do software, com correções de bugs e refatoração do código.

3. MINERAÇÃO DE TEXTOS

Na linha de desenvolvimento da ferramenta de mineração textual, com foco na ferramenta insightData, destacam-se:

- a) A implementação de ajustes no sistema para alinhamento do InsightData Client às especificações do Projeto FINEP;
- b) A migração do ambiente de produção para o protocolo HTTPS com aumento de segurança de acesso de usuários externos;
- c) A atualização das versões do mecanismo de busca (ElasticSearch) e do banco de dados (PostgreSQL) do sistema;
- d) A atualização da API de inclusão e indexação de documentos;
- e) A adaptação do extrator de patentes para capturar os campos de referências bibliográficas, além da implementação de funcionalidade de importação da API para permitir a atualização de patentes já importadas, além da atualização do banco de dados de patentes (214 pacotes, aproximadamente 4.7 milhões de patentes) para incorporar referências bibliográficas;
- f) A implementação da funcionalidade de "Exportar Patentes" em formato TXT do Web of Science, para permitir a análise no InsightNet; e
- g) Seguindo o esforço de internalização crescente no Centro do desenvolvimento das ferramentas de análise de dados, foi finalizada a migração do insightData para o novo ambiente de servidores do CGEE.

4. NOVOS ALGORITMOS E PROTÓTIPOS

Na linha de desenvolvimento e testes de algoritmos para aplicações futuras em demandas de análises de dados do CGEE, foram desenvolvidos protótipos de programas de extração de novos conjuntos de dados, incluindo:

- a) Adaptações do protótipo de extrator de dados de produção tecnológica de currículos Lattes na linguagem R iniciado em colaboração com estudantes do curso

de Engenharia de Software da UnB;

b) O desenvolvimento de protótipo de extrator de dados da base Diretório de Grupos de Pesquisa do CNPq, com interface simplificada, para geração de planilhas Excel com dados selecionados por usuários sem treinamento em tecnologias da informação. Infelizmente, o uso desse programa é limitado pela base de dados desatualizada (2016), mas quando o CNPq disponibilizar uma versão do Diretório, uma nova versão da ferramenta poderá ser desenvolvida com base na atual;

c) Testes de protótipos de três algoritmos tradicionais de ciência de dados para análise de sentimentos (naive Bayes, k nearest neighbors e regressão logística) com o propósito de realizar possíveis análises de consultas do insightSurvey, com supervisão automatizada de desempenho a partir de algoritmo genético elaborado especificamente para essa finalidade;

d) O desenvolvimento e primeiros testes de técnicas de armazenamento de dados de matrizes esparsas que permite o tratamento de redes com 10 vezes mais textos do que o desempenho atual do insightNet;

e) O desenvolvimento, com base da mesma técnica do item anterior, do protótipo de um sumarizador semiautomático de textos para otimizar a análise de textos longos e degravações de reuniões e entrevistas realizadas nos diversos projetos em andamento no CGEE;

f) Teste do algoritmo *Breadth First Search* para otimizar a esparsificação de redes, com resultados promissores no sentido de automatizar a geração de redes de similaridade semântica, hoje calibradas manualmente pelos usuários do InsightNet;

g) Realização dos primeiros testes de algoritmo para a geração de palavras-chave de um dado texto, para facilitar a classificação de clusters nas redes de similaridade semântica;

h) Elaboração de documento de requisitos para a ferramenta InsightTech; e

i) Criação de método de análise de indicadores compostos, baseado em redes de correlações, aplicado no estudo do índice global de inovação (GII – *Global*

Innovation Index) e seus 80 indicadores únicos. Este método foi aplicado com sucesso para identificar vantagens e desvantagens do Brasil no que diz respeito a cada indicador do GII com relação a qualquer outro país.

5. INTERAÇÃO COM OUTROS PROJETOS DO CENTRO E FUTUROS DESENVOLVIMENTOS

Ao longo de 2019 foi elaborado um documento de diretrizes para a implementação prática de um ambiente de discussão de conceitos e técnicas inovadores em Ciência de Dados e Representação do Conhecimento com possíveis impactos estratégicos para o Centro. Além de um documento de base para a operacionalização inicial desse ambiente de discussão, foram realizadas diversas reuniões técnicas das quais foram propostos, concebidos ou originados alguns dos algoritmos relacionados no item 4.

A equipe do projeto participou ativamente na elaboração de algoritmos e novos métodos de análise demandas do Centro e de seus parceiros. Dentre esses destacam-se:

- Métodos para abordar os dados da consulta pública inicial do Programa “Future-se”, realizada no contexto da ferramenta insightSurvey, do CGEE. Foi requisitado que parte da metodologia fosse de explicação simples para público leigo, portanto a parte mais importante foi desenvolvida com a ferramenta Excel. Entretanto, também foram estudadas alternativas de uso de algoritmos de aprendizagem de máquina e algoritmos genéticos para análises de sentimentos das contribuições dos participantes da consulta.
- Métodos semelhantes ao do item anterior, também baseados nas funcionalidades do Excel, foram empregados para minerar os temas de investimentos públicos em pesquisa e desenvolvimento na área de energia, de acordo com requisitos da iniciativa “Energy Big Push” da qual o CGEE faz parte.
- Nova metodologia de análise de indicadores compostos, aplicada aos 80

indicadores únicos do Índice Global de Inovação. A metodologia foi desenvolvida para a etapa “benchmarking” da elaboração de proposta de Política Nacional de Inovação (PNI). Posteriormente, a equipe também trabalhou na integração da metodologia, por meio de uma análise de “gaps” positivos e negativos entre os valores dos indicadores para o Brasil e seleções de outros países, que foi agregada a um conjunto mais abrangente de propostas de prioridades discutidas em entrevistas, oficinas e, posteriormente, na consulta pública sobre a Política, realizada com o insightSurvey. Além disso, a equipe também trabalhou no desenho e tratamento de dados da consulta pública, com resultados a serem entregues em 2020. O aprendizado com a integração de métodos qualitativos e quantitativos desse trabalho abre oportunidades de aplicações futuras importantes para outros projetos do CGEE.

- Continuidade da cooperação interna com os projetos “Diagnóstico da Situação Atual nas CHSSA Brasileiras” e “Observatório em Ciência, Tecnologia e Inovação”.

6. JOURNAL CLUB

Evento mensal já tradicional no CGEE, foi dada continuidade às das apresentações realizadas no 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. Em 2019, foi ultrapassada a marca de 60 apresentações, tendo sido apresentados e debatidos os seguintes artigos:

- a) "Interbank network and regulation policies: an analysis through agent-based simulations with adaptive learning";
- b) "Developing science, technology and innovation indicators: What we can learn from the past";
- c) "Knowledge combination modeling: The measurement of knowledge similarity between different technological domains";

- d) "Towards the Thousandth CubeSat: A Statistical Overview";
- e) "The concept of instrument constituencies: accounting for dynamics and practices of knowing governance";
- f) "A bibliometric method for assessing technological maturity: the case of additive manufacturing";
- g) "Applying Blockchain Solutions to Address Research Reproducibility and Enable Scientometric Analysis";
- h) "Design para a Competitividade no Brasil - O caso do projeto Design Export";
- i) "Eleições Presidenciais no Brasil: Como o índice de Tração indicou a ascensão de Bolsonaro"; e
- j) Edição extra do Journal Club, sobre o artigo " Astronomers Capture First Image of a Black Hole", realizada no primeiro semestre.

7. PARTICIPAÇÃO EM EVENTO

"Complex Networks 2019" em dezembro de 2019, em Lisboa. Embora sem apresentação de trabalho, a participação nesse congresso foi fundamental para avaliar a posição relativa do CGEE com relação ao estado da arte na análise de redes complexas. Essa área, pelo menos por enquanto, tem sido a principal fonte de novos métodos quantitativos para a atuação do Centro em projetos cujos resultados têm bases em dados. O contato com pesquisadores e seus trabalhos, foram extremamente úteis para indicar novos caminhos de desenvolvimento dos métodos de análise. Foi particularmente relevante observar que há novas linhas de pesquisa baseadas em inteligência artificial, "embeddings" e visualização e análises de redes com muitas camadas que podem ser adaptadas para os objetivos do CGEE.

8. ALTERAÇÕES NO PROJETO E PERSPECTIVAS PARA 2020

Na proposta para 2019 para o projeto estava previsto o planejamento e integração de

uma ferramenta integrada com as funcionalidades contidas nas diversas ferramentas em uso no CGEE. Foi avaliado que esse esforço desproporcional aos recursos humanos existentes no Centro e que ocorreria em um prazo muito dilatado, o que comprometeria capacidade do Centro se manter atualizado com relação aos algoritmos mais recentes da ciência de dados. Essa tarefa foi alterada, portanto para a elaboração de um documento descritivo de funcionalidades, bases de dados e algoritmos desenvolvidos no CGEE. Para não perder o foco da necessária integração ao menos no nível metodológico, foi também planejada preparação de:

- Um documento de requisitos para a padronização do trabalho no desenvolvimento de protótipos de ferramentas de análises de dados;
- Um documento técnico contendo proposta de padrões para desenvolvimento de protótipos de ferramentas de análise de dados do CGEE. Essa proposta incluirá formatos de dados definidos para serem intercambiados entre as ferramentas em uso no Centro e também abordará boas práticas de registro de processos, particularmente para garantir a reprodutibilidade dos resultados, bem como estruturar padrões de qualidade dos dados à disposição.

A meta de produzir um documento contendo proposta de modelo de transferência de metodologias de exploração e visualização de dados incluindo alternativas de proteção à propriedade intelectual das tecnologias desenvolvidas foi adiada para 2020.

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 para 2019 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 previa a entrega de relatórios consolidando:

- a) Levantamentos do estado da arte nas áreas de conhecimento abordadas e seleções 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.

Essas metas foram parcialmente cumpridas, no sentido de que as discussões foram realizadas e foram produtivas para a elaboração de novas ideias e protótipos, mas o envolvimento da equipe nas diversas tarefas de outros projetos ao longo do segundo semestre, particularmente na implementação das atividades da PNI baseadas em dados, comprometeu elaboração do sumário executivo. Os principais resultados dessas discussões foram listados basicamente na Seção 4 acima, mas alguns resultados das Seções 2.1 e 2.2 também foram produtos das discussões.

Por fim, dada a importância de um desenvolvimento mais focado na área de visualização da informação, está prevista para 2020 a organização de discussões

estruturadas e a redação de um documento técnico específico para a visualização de dados contendo um levantamento de alternativas de algoritmos, ferramentas e técnicas e a identificação de oportunidades para futuros usos no CGEE.

Os esforços empregados nas atividades listadas foram proporcionais à importância dos desenvolvimentos metodológicos realizados, mas comprometeram, particularmente ao longo do segundo semestre, a capacidade da equipe executar parcial ou inteiramente algumas das metas do Projeto. Essa participação em outros projetos é não apenas desejável, dada a estrutura matricial de trabalho do Centro, mas também necessária para este projeto, uma vez que são justamente os problemas tratados nessas interações que orientam as prioridades do projeto para desenvolvimentos futuros. Para mitigar os impactos dessas colaborações na execução futura de atividades do projeto, o Centro contratou novos membros para a equipe. Espera-se, com isso, que a capacidade da equipe cumprir todas as metas seja bastante ampliada.



cgEE

Centro de Gestão e Estudos Estratégicos
Ciência, Tecnologia e Inovação

Manual CGEE Insight Net

Release 3.2.3

1	Introdução	1
1.1	Contexto e Visão Geral	1
1.2	Ajuda online	2
1.3	Funcionalidades experimentais	3
1.4	Envio do protocolo de execução	4
2	Instalação do CGEE Insight Net	7
2.1	Pré-requisitos	7
2.2	Instalação do software Gephi	8
2.3	Configuração da central de atualizações	9
2.4	Instalação do <i>CGEE Insight Net</i>	11
2.5	Atualização do <i>CGEE Insight Net</i>	17
3	Configuração do CGEE Insight Net	19
3.1	Configuração do banco de dados	20
3.2	Configuração do usuário para acessar o banco de dados de Currículos Lattes do CGEE.	21
3.3	Configuração das colunas exibidas	22
3.4	Exibição da lista de palavras-chave	24
3.5	Parâmetros da pesquisa por similaridade	25
3.6	Detecção de idiomas	28
3.7	Licenças	29
3.8	Protocolos de execução	30
3.9	Memória <i>cache</i> de Currículos Lattes	31
4	Conceitos gerais do uso do <i>CGEE Insight Net</i>	33
4.1	Fluxo de trabalho	33
5	Uso do <i>CGEE Insight Net</i> para analisar Currículos Lattes	37
5.1	Importação dos Currículos Lattes	39
5.2	Formação da rede	46
5.3	Visualização de atributos dos pesquisadores	53
5.4	Visualização e edição das contribuições Lattes	57
6	Criação de redes de referências bibliográficas BibTeX	61
6.1	Importação dos dados bibliográficos	63
6.2	Formação da rede	66

7	Criação de redes de referências bibliográficas genéricas	71
7.1	Importação dos dados bibliográficos	73
7.2	Formação da rede.....	78
8	Análise das redes criadas	83
8.1	Filtragem dos resultados	83
8.2	Análise de clusters	87
8.3	Análise de assortatividade.....	88
8.4	Análise das palavras-chave	91
8.5	Criação de redes de co-ocorrências de palavras-chave.....	103
8.6	Eliminação interativa de nós da rede e do banco de dados	104
8.7	Criação de uma nova rede a partir do subconjunto de nós selecionados.....	106
8.8	Seleção interativa de nós vizinhos na rede	107
8.9	Visualização interativa do currículo de pesquisadores no browser.....	108
8.10	Visualização interativa de contribuições bibliográficas por DOI no browser	111
9	Funcionalidades comuns de apoio	115
9.1	Recuperação do grafo a partir das informações que constam no banco de dados	116
9.2	Cópia e recuperação do banco de dados.....	118
9.3	Estatísticas do banco de dados	120
9.4	Protocolos de execução.....	121
9.5	Envio de protocolo de execução.....	122
10	Informações adicionais	125
10.1	Especificação de formatos para o módulo de referências bibliográficas genéricas	125
10.2	Compilação do <i>CGEE Insight Net</i>	136
10.3	Uso do <i>CGEE Insight Net</i> a partir da linha de comando	138
	Referências Bibliográficas	155

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 Ajuda online

Este manual de usuário está disponível online no **CGEE Insight Net**. No menu *Help* existe a opção *CGEE Insight Net Help*:

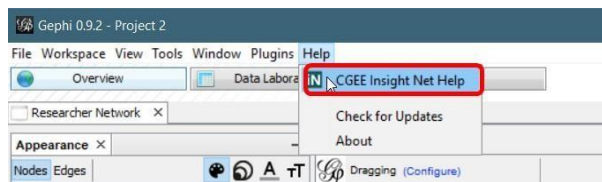



Figura 1.1: Opção para abrir a ajuda on-line

1.3 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 .

1.4 Envio do protocolo de execução

Em caso de erros inesperados no **CGEE Insight Net**, o usuário pode enviar um relatório de erros ao CGEE. Quando acontecer um erro inesperado, a seguinte mensagem é exibida:

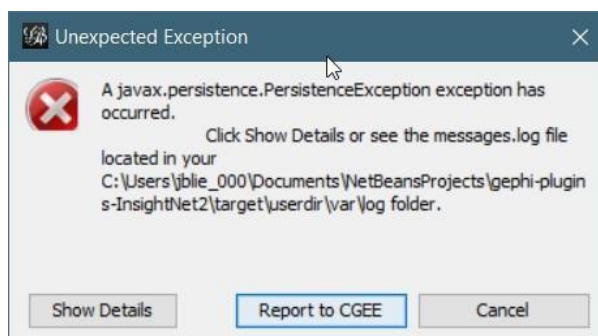


Figura 1.2: Mensagem de erros inesperados

Clicando em *Report to CGEE*, o **CGEE Insight Net** mostra o seguinte diálogo:

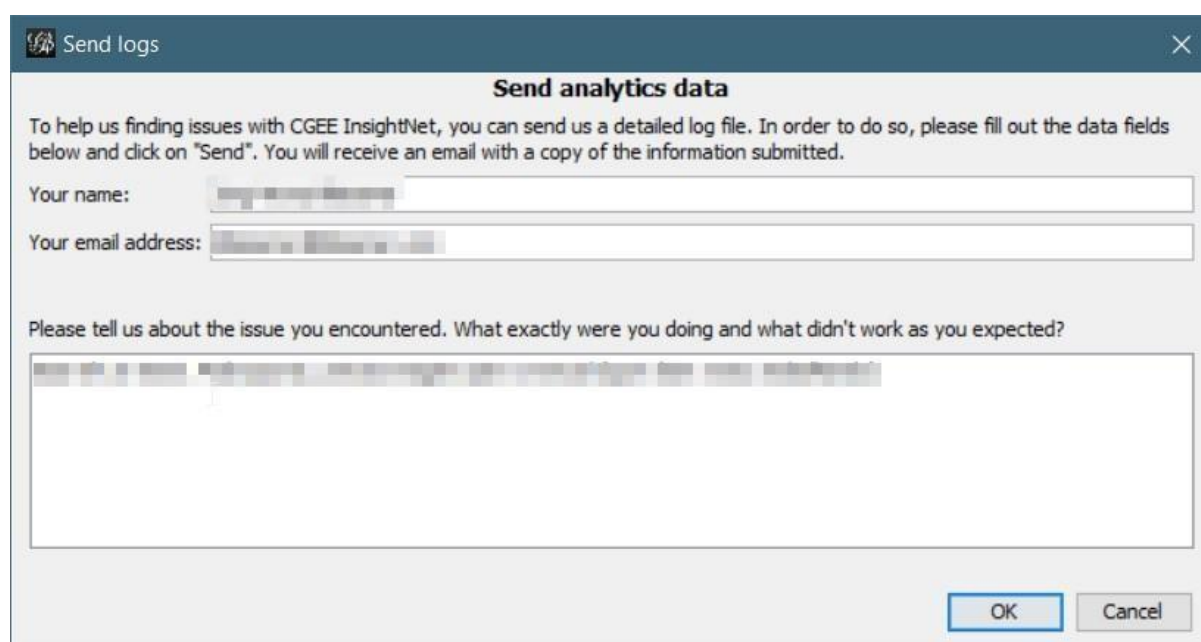


Figura 1.3: Diálogo de envio do protocolo de execução

Recomenda-se preencher todos os campos desse diálogo para que o CGEE possa reproduzir e eliminar possíveis problemas. Depois, o usuário deve clicar em “**OK**” e os dados são enviados. No final deste processo, o resultado do envio é exibido:

Send logs [X]

Send analytics data

To help us finding issues with CGEE InsightNet, you can send us a detailed log file. In order to do so, please fill out the data fields below and click on "Send". You will receive an email with a copy of the information submitted.

Your name:

Your email address:

Please tell us about the issue you encountered. What exactly were you doing and what didn't work as you expected?

Your file was successfully received and an email notification was sent

OK

Figura 1.4: Conclusão do envio do protocolo de execução

O **CGEE Insight Net** também detecta se o *Gephi* for encerrado de forma irregular. Neste caso, a seguinte mensagem é exibida quando o programa for reiniciado:

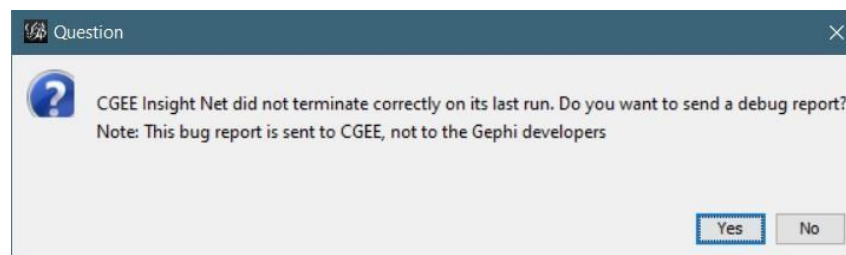


Figura 1.5: Mensagem após o encerramento irregular do Gephi

Clicando em “Yes”, o diálogo de envio do protocolo de execução é exibido e deve ser preenchido conforme descrito anteriormente.

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”:

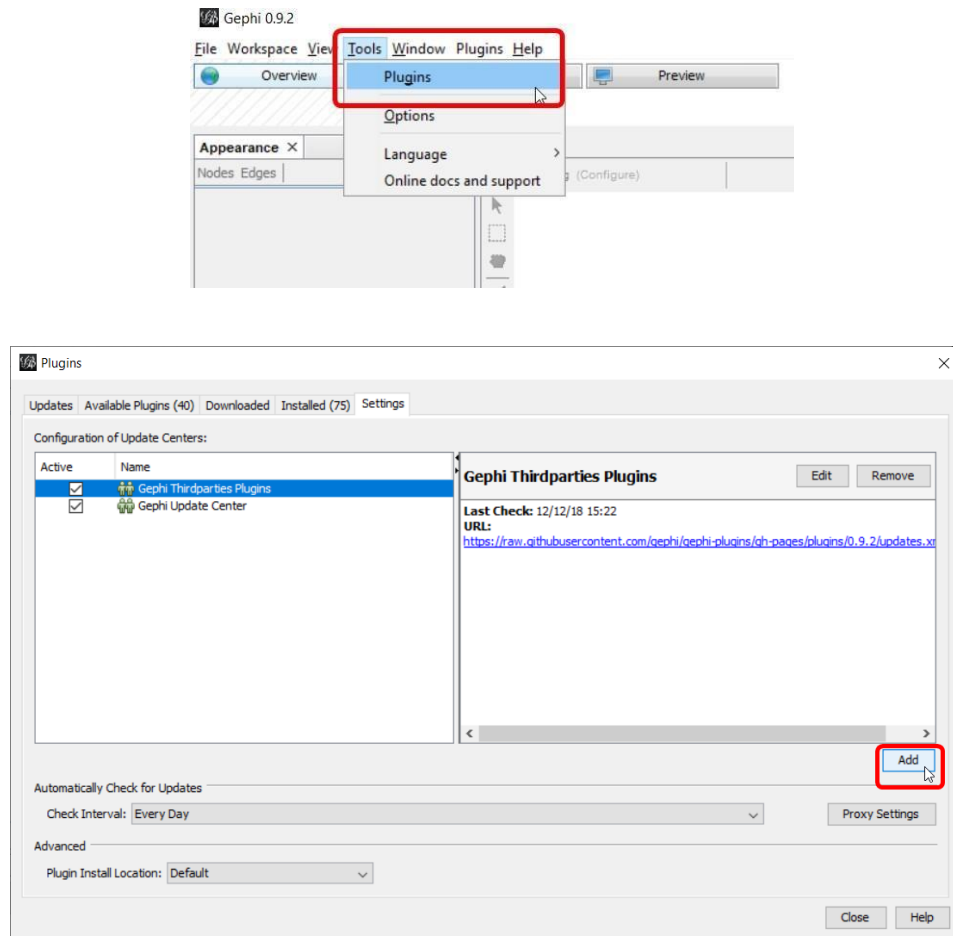


Figura 2.1: Configuração dos centrais de atualização Gephi

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

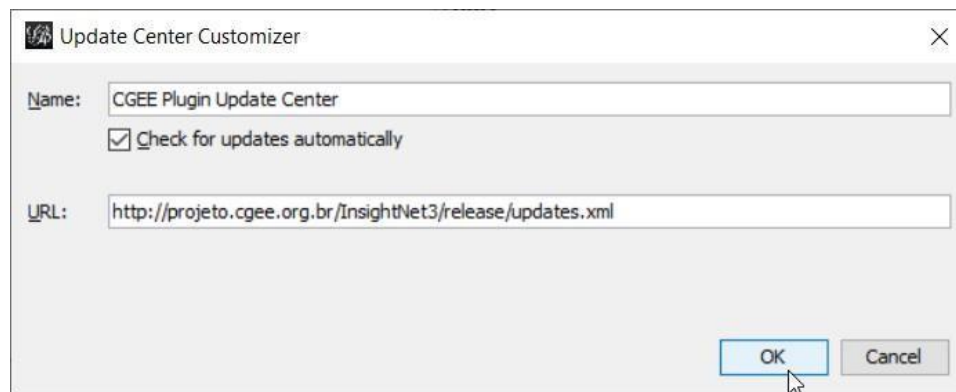


Figura 2.2: Configuração da central de atualizações do CGEE Insight Net

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*”:

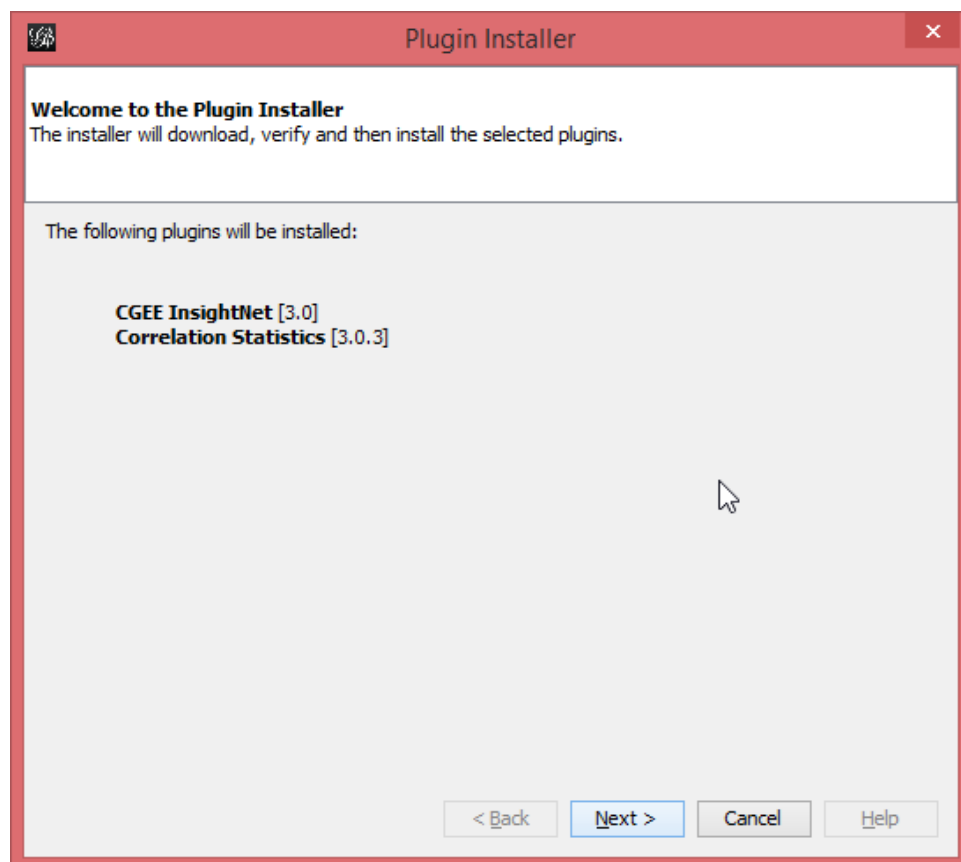
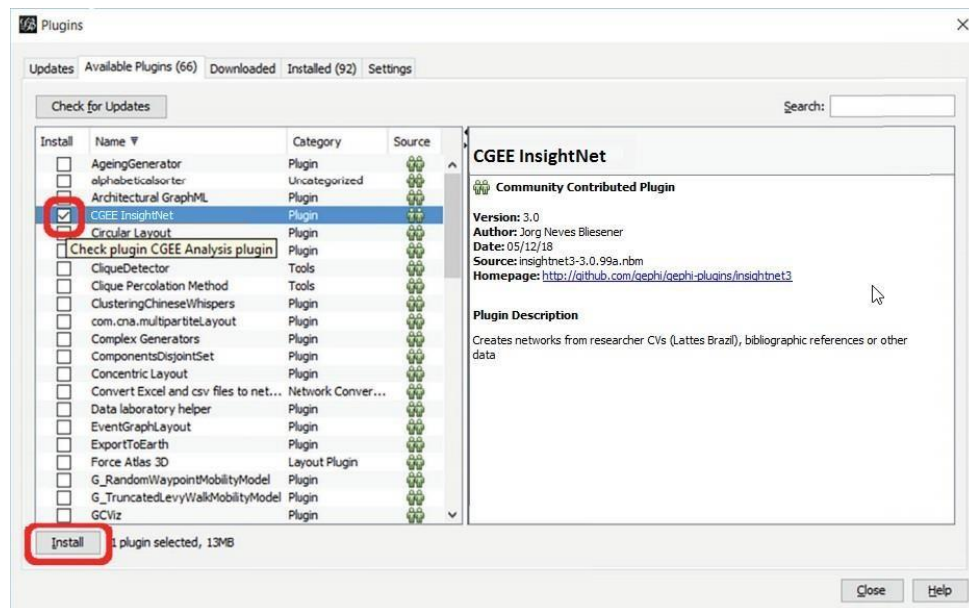


Figura 2.3: Instalação do CGEE Insight Net

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

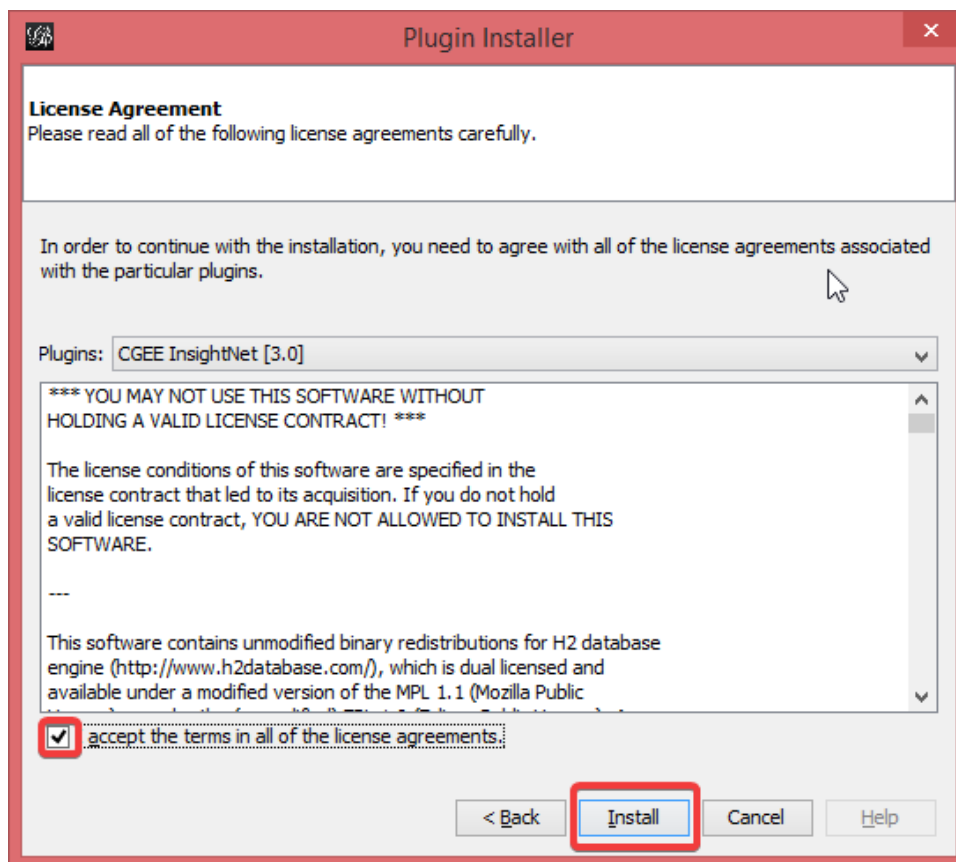


Figura 2.4: Concordância com a(s) licenças(s) do produto

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.

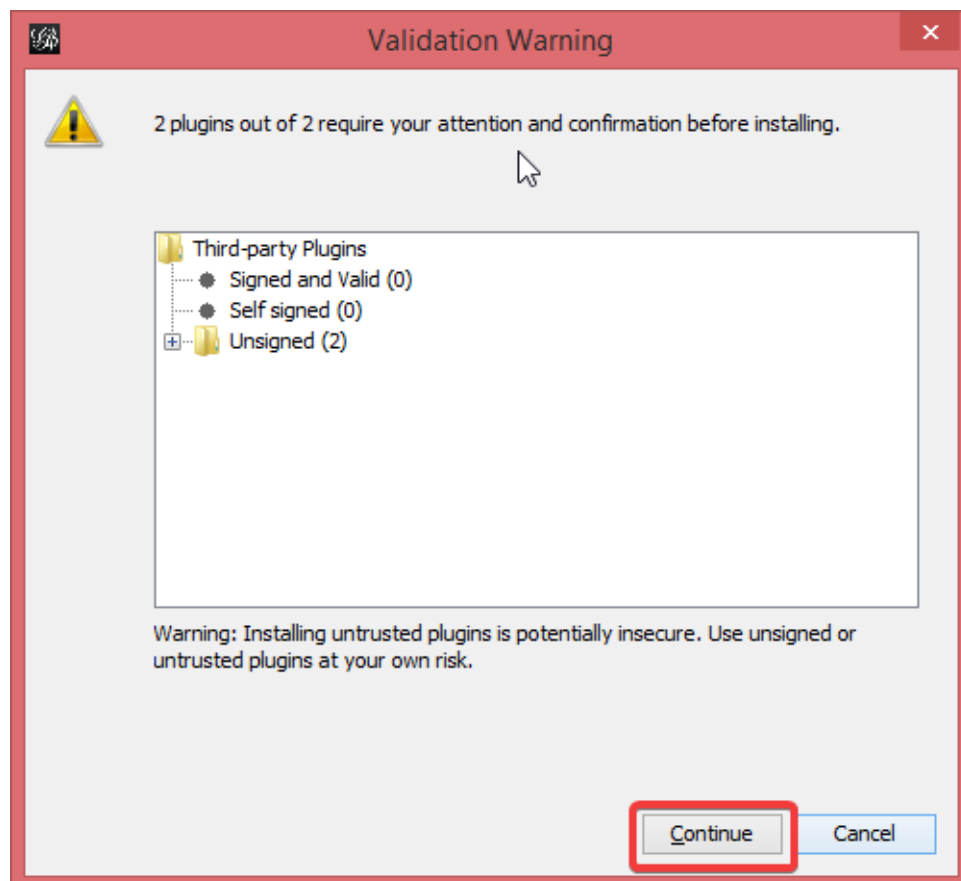
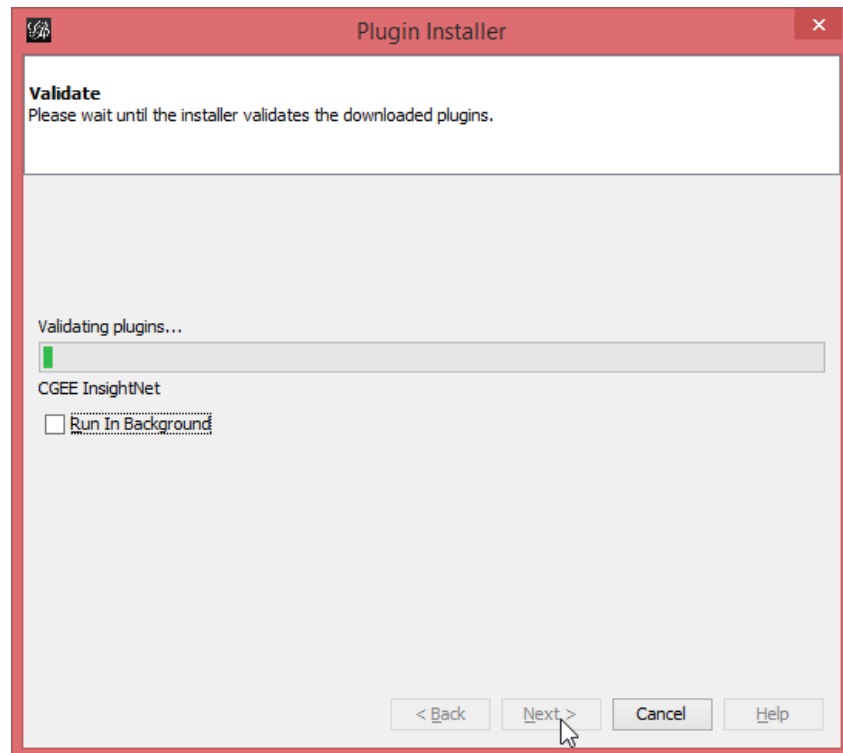


Figura 2.5: Download e aviso de falta de assinatura

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

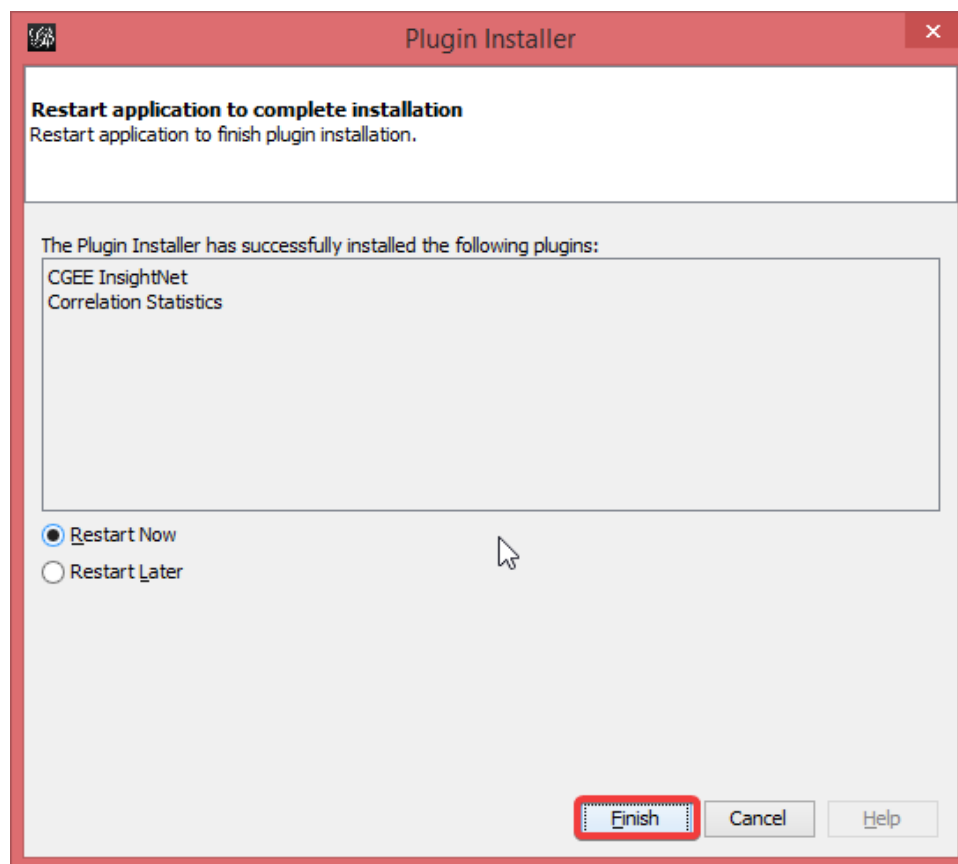


Figura 2.6: Instalação do CGEE Insight Net

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:

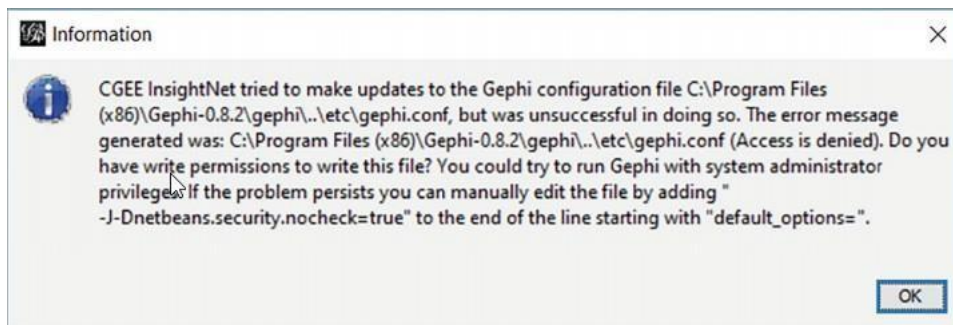
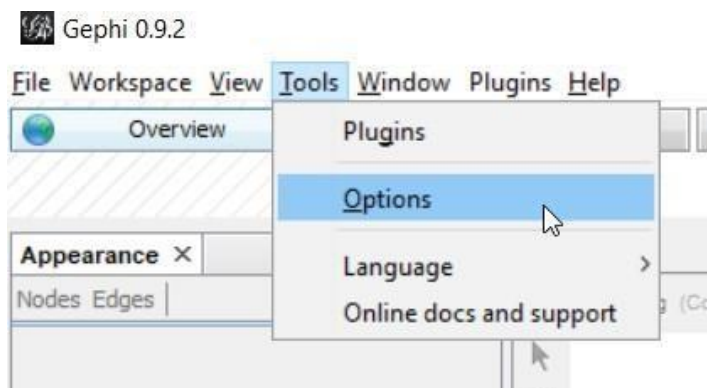


Figura 2.8: Mensagem de alteração malsucedida do arquivo gephi.conf

Nesse caso, a gestão e o uso de licenças adicionais (ver [Seção 3.7](#)) 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 3.7](#).



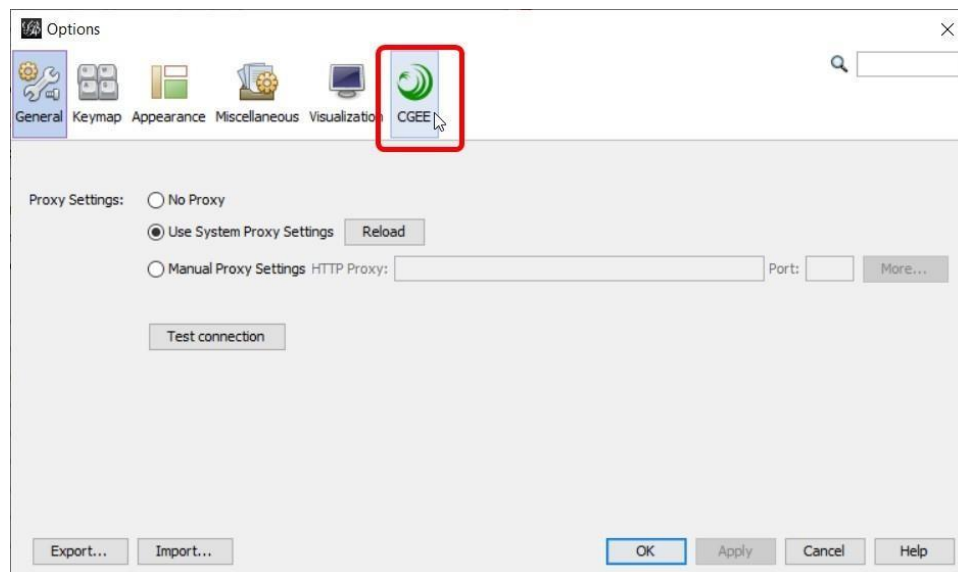


Figura 2.9: CGEE Insight net instalado

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*:

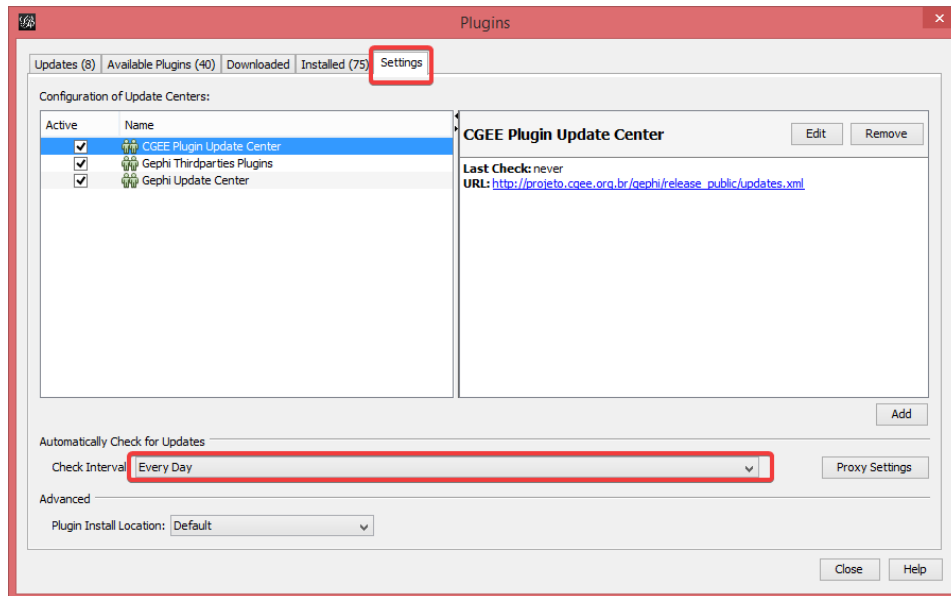


Figura 2.10: Configuração da atualização automática

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

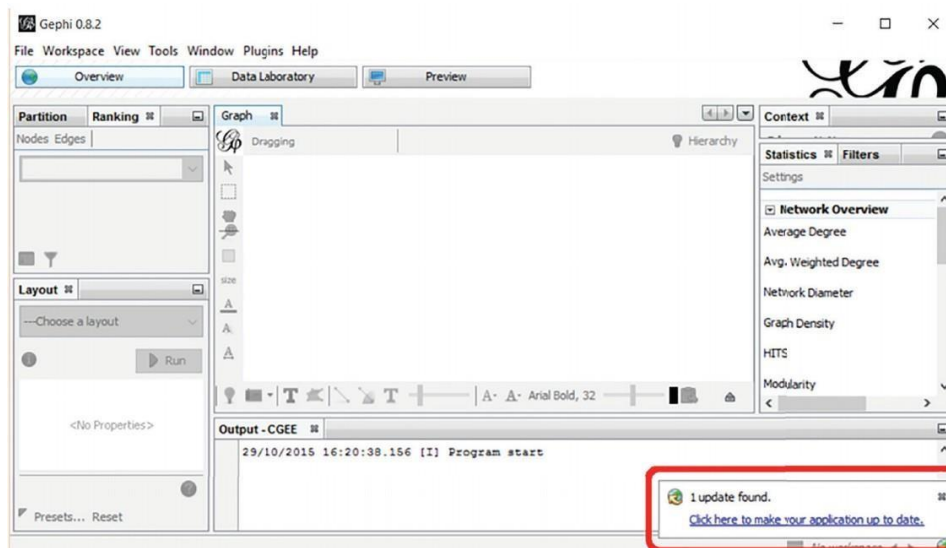


Figura 2.11: Aviso de atualização

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

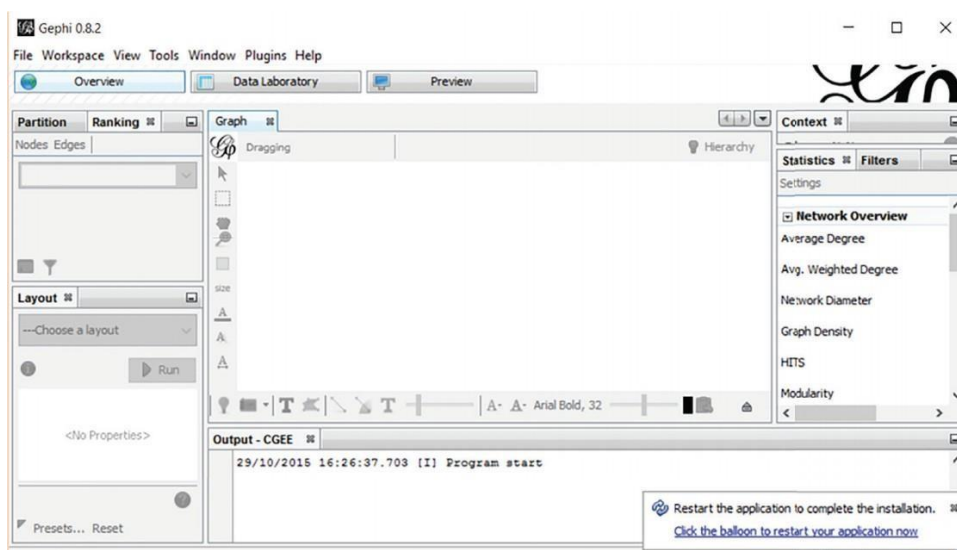
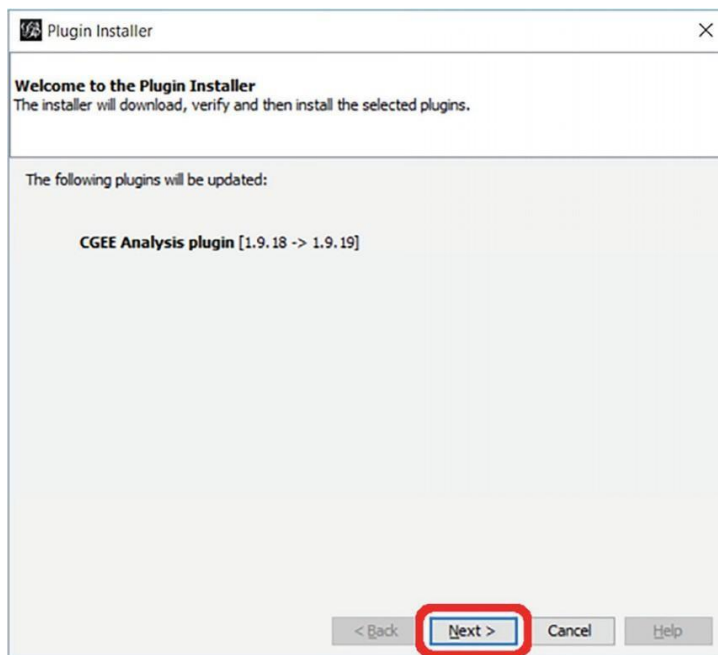


Figura 2.12: Aviso de atualização

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*.

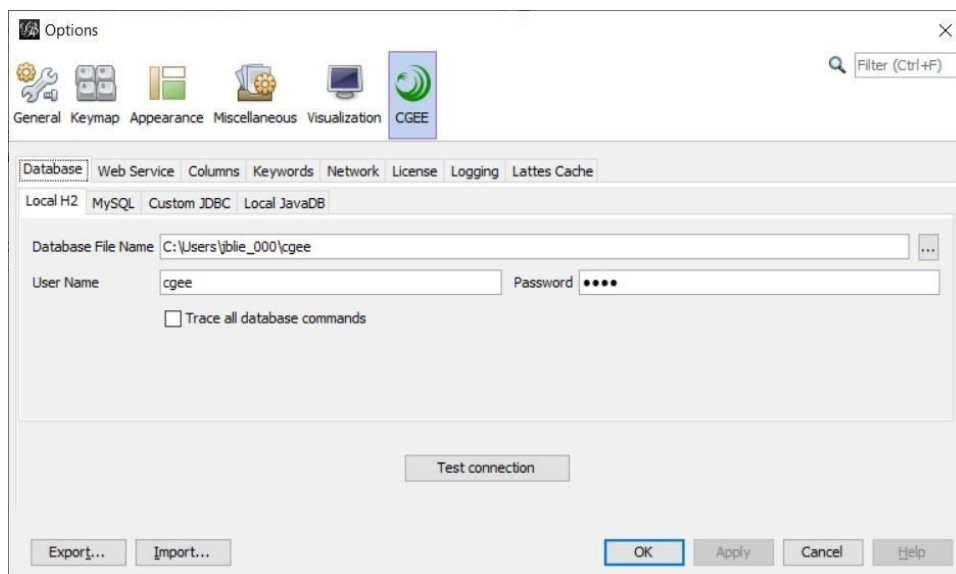


Figura 3.1: Opções de configuração

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:

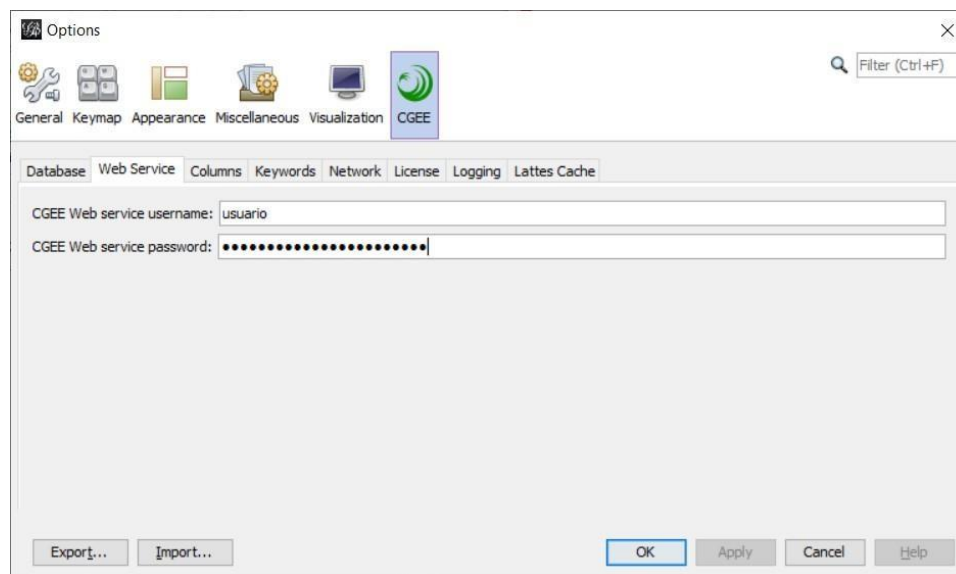


Figura 3.2: Configuração do usuário e da senha para acessar o banco de dados de Currículos Lattes do CGEE

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:

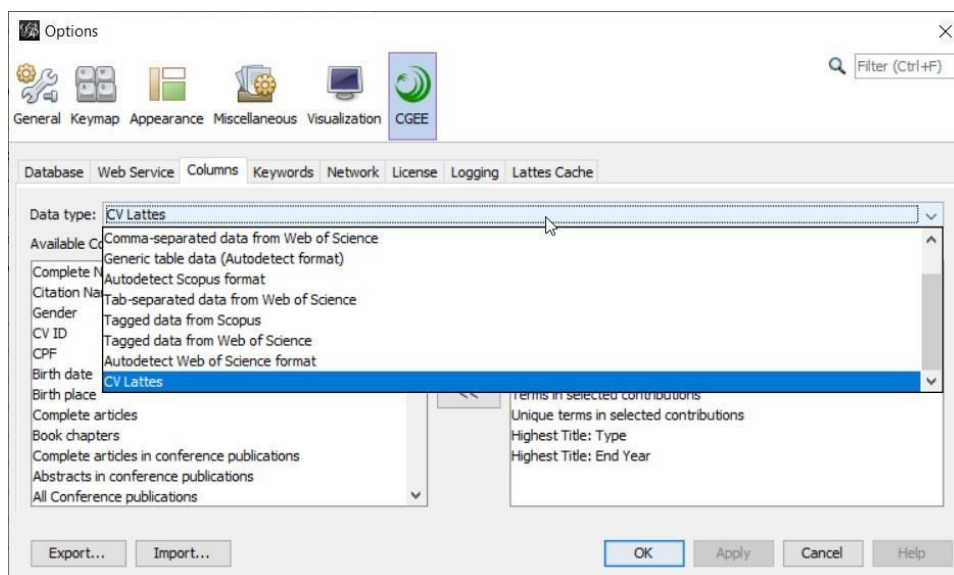


Figura 3.3: Seleção do tipo de dados para configurar as colunas que serão exibidas por padrão

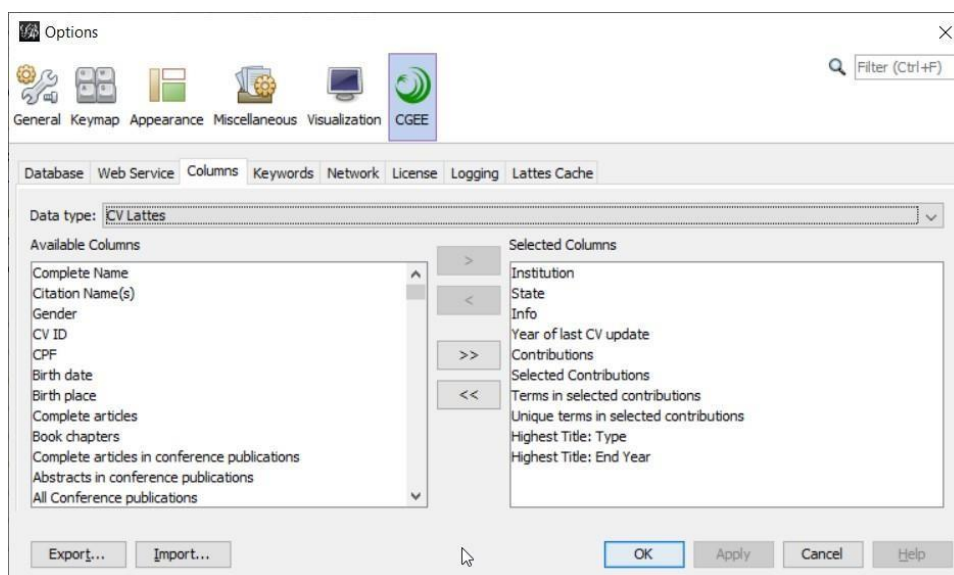






Figura 3.4: Configuração das colunas dos currículos Lattes que serão exibidas por padrão

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 lista de palavras-chave

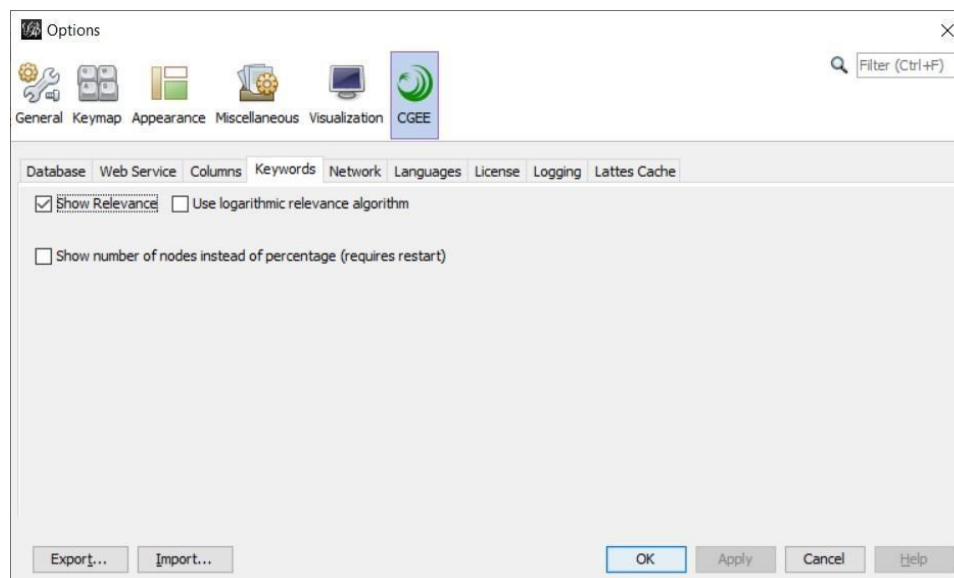


Figura 3.5: Configuração da exibição das palavras-chave

O *CGEE Insight Net* permite a exibição opcional das relevâncias das palavras-chave. Geralmente, a janela de palavras-chave (ver Seção 8.4) 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*”.

Na janela de palavras-chave (ver Seção 8.4), a quantidade de nós que referenciam uma palavra-chave pode ser exibida como percentual da quantidade total de nós ou como número absoluto. A opção “*Show number of nodes instead of percentage*” permite alternar entre essas duas opções. Depois de alterar essa configuração, o *Gephi* precisa ser reiniciado.

3.5 Parâmetros da pesquisa por similaridade

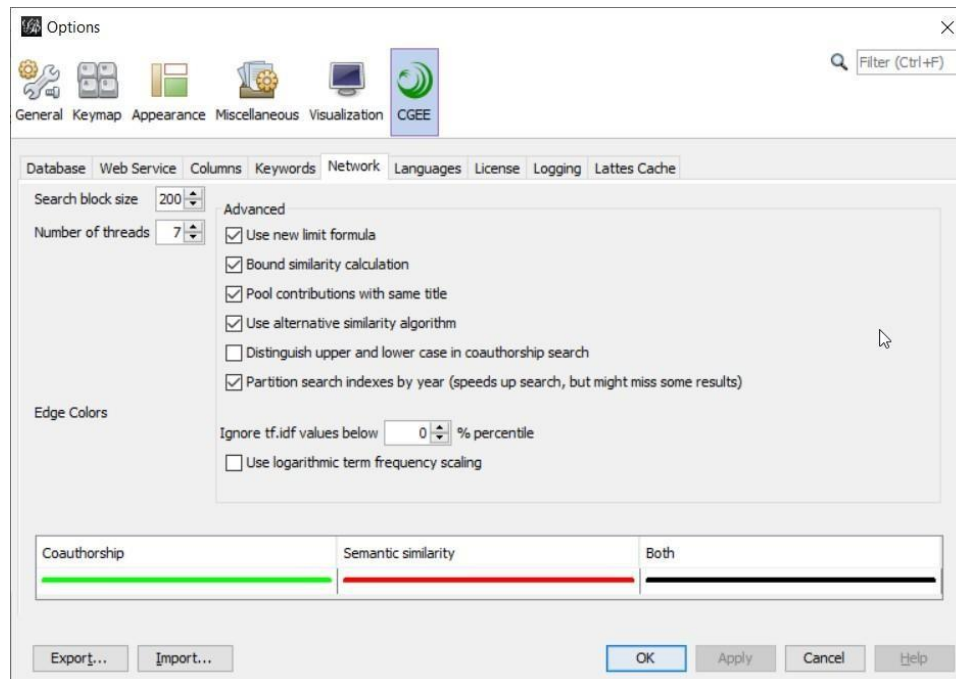


Figura 3.6: Configuração do cálculo e da exibição das redes

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 especificada 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.
- 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.
- “*Ignore tf.idf values below x% percentile*”: Os termos cujo valor de relevância (tf.idf) são abaixo do percentil confi gurado 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:

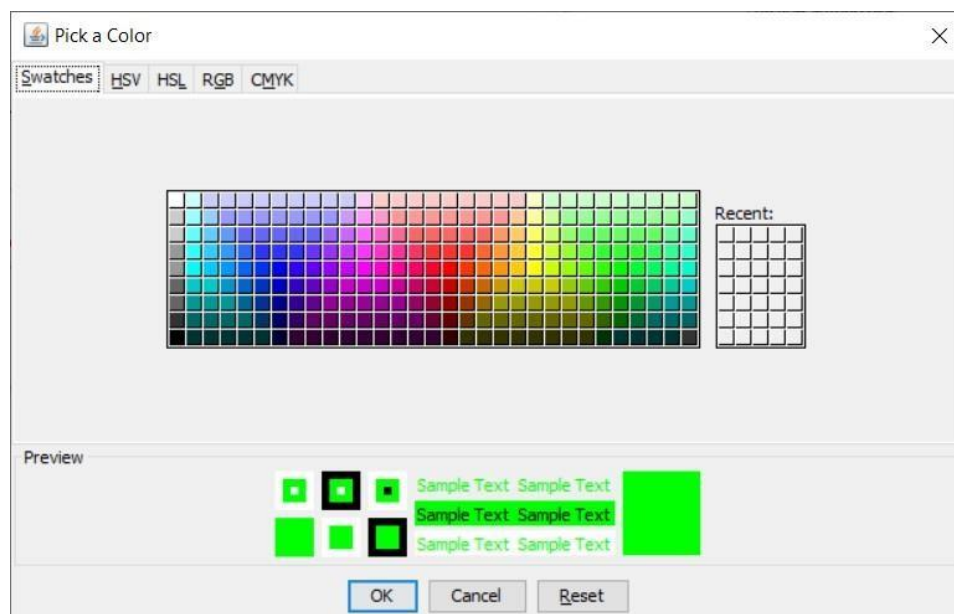


Figura 3.7: Seleção de cores

3.6 Detecção de idiomas

A partir da versão 3.1 do *CGEE Insight Net*, o tratamento de textos em vários idiomas foi reformulado. Para cada documento cuja similaridade será analisada, o *CGEE Insight Net* tenta determinar o idioma do título e do resumo, para poder realizar a análise com os parâmetros corretos do idioma. Essa detecção de idioma pode ser configurado na tela “*Languages*”.

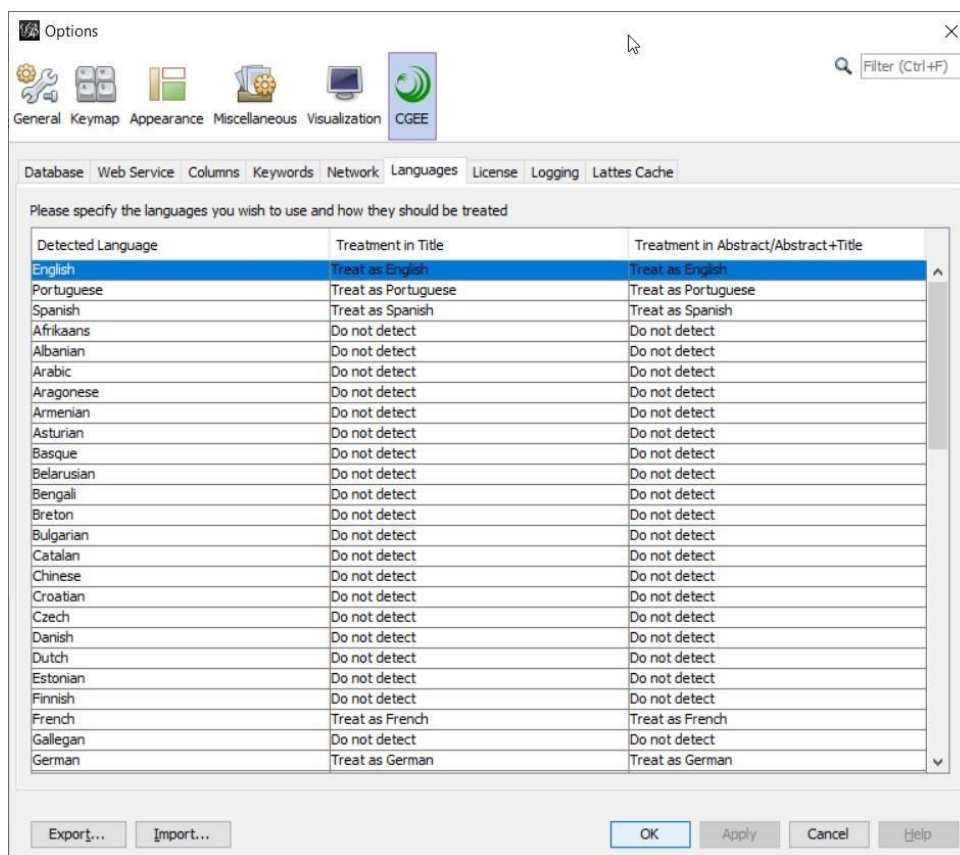


Figura 3.8: Configuração da detecção de idiomas

Deve ser observado que não existem analisadores de texto para todos os idiomas que podem ser detectados. Recomenda-se limitar a quantidade de idiomas detectados para manter a fidelidade dessa detecção. Na configuração básica, são apenas reconhecidos documentos em Inglês, Português, Espanhol, Francês e Alemão. A detecção e o tratamento podem ser diferenciados por título ou por resumo.

3.7 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*”:

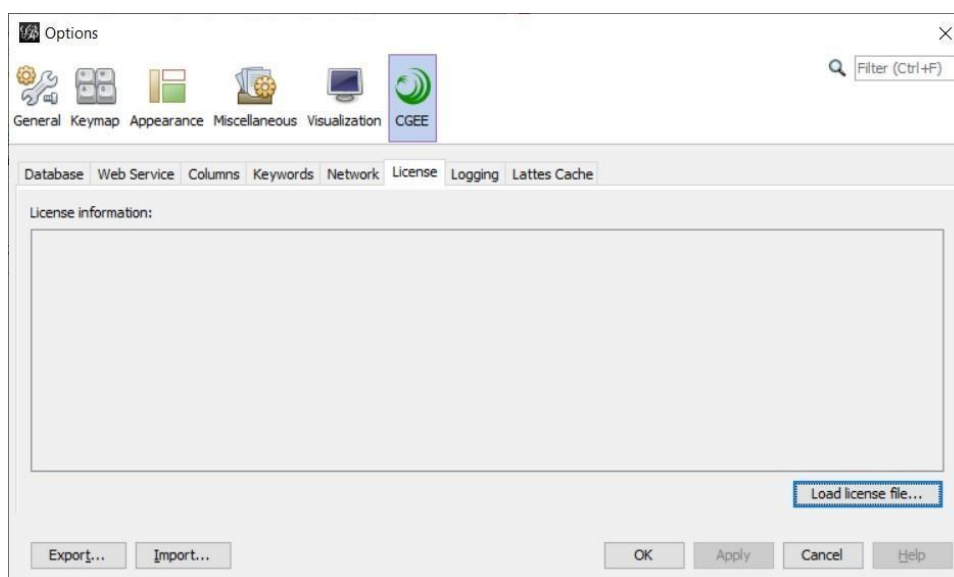


Figura 3.9: Instalação de licenças

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:

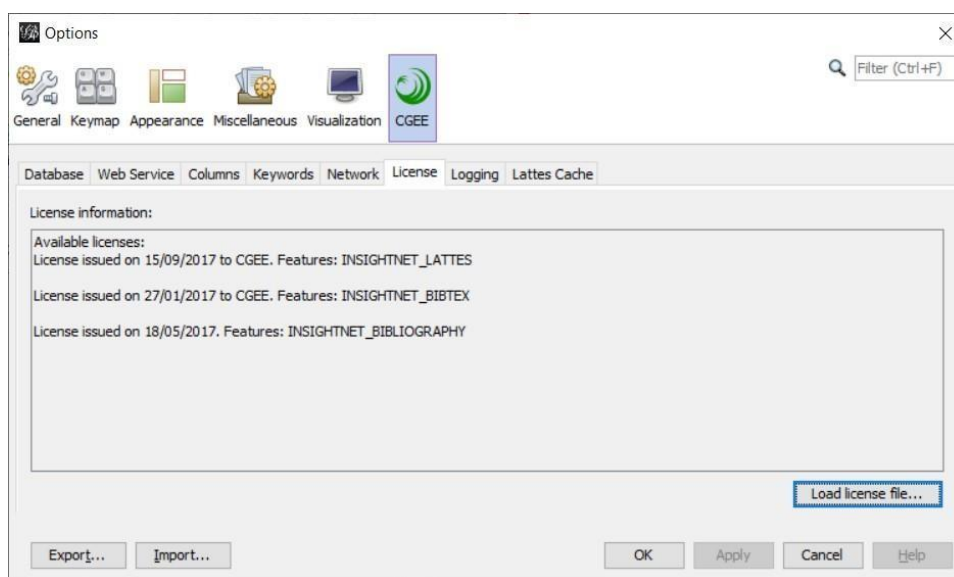


Figura 3.10: Indicação das licenças disponíveis

3.8 Protocolos de execução

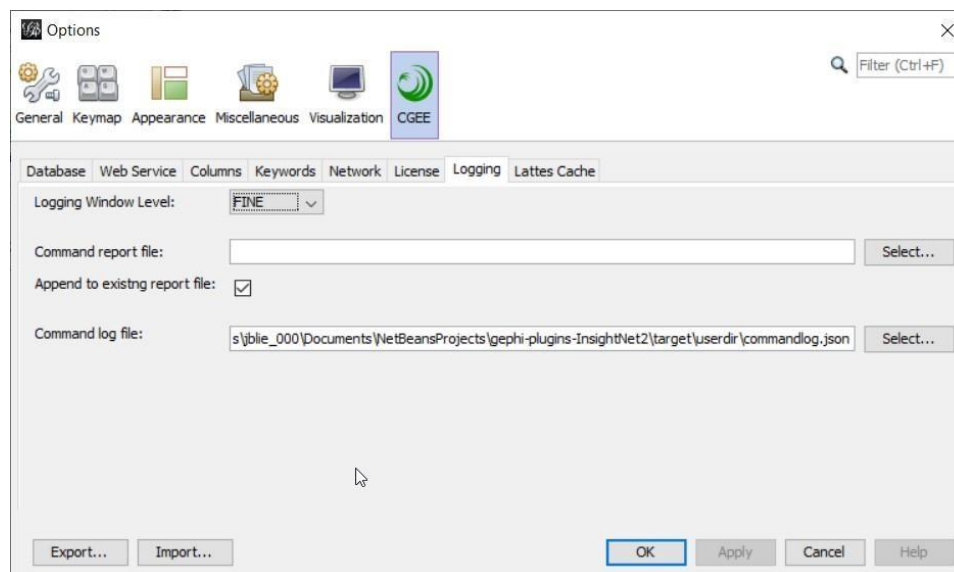


Figura 3.11: Configuração do protocolo de execução

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.

3.9 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:

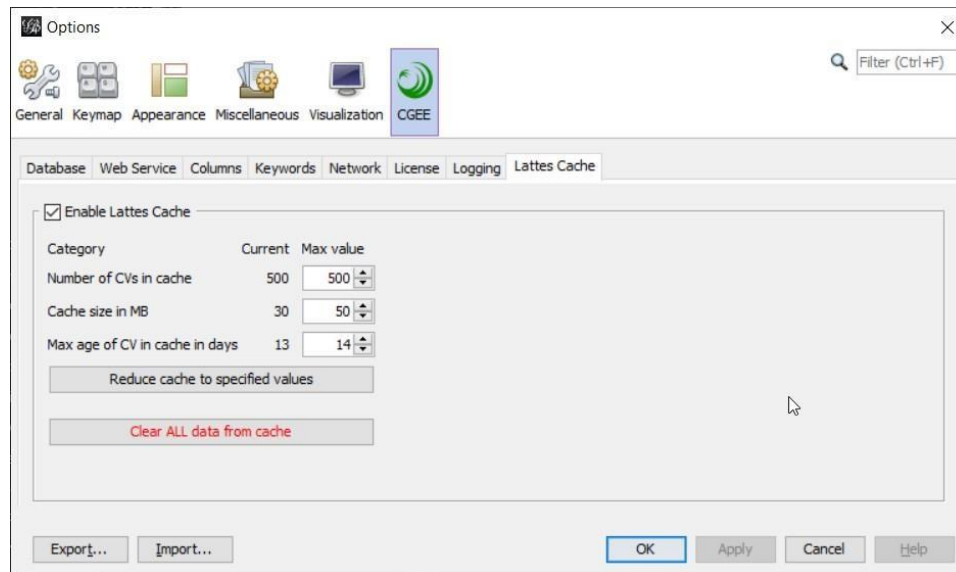


Figura 3.12: Configuração da memória *cache* do módulo Lattes

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

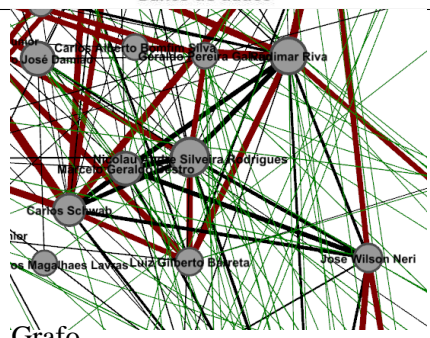
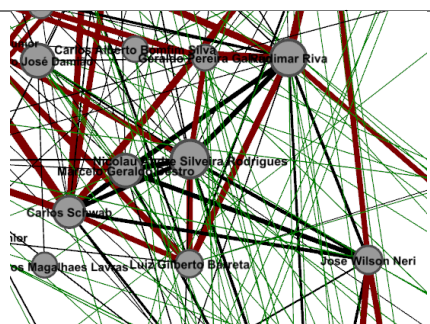
Ao utilizar o *CGEE insight Net*, deve ser considerado que ele trabalha com três repositórios de informações:

- 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>	Importação	<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>	Seleção de contribuições	<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>	Pesquisa de similaridade	<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>	Visualização	 <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 3.7](#). Essa licença é exibida assim no diálogo *Tools > Options > CGEE > License*:

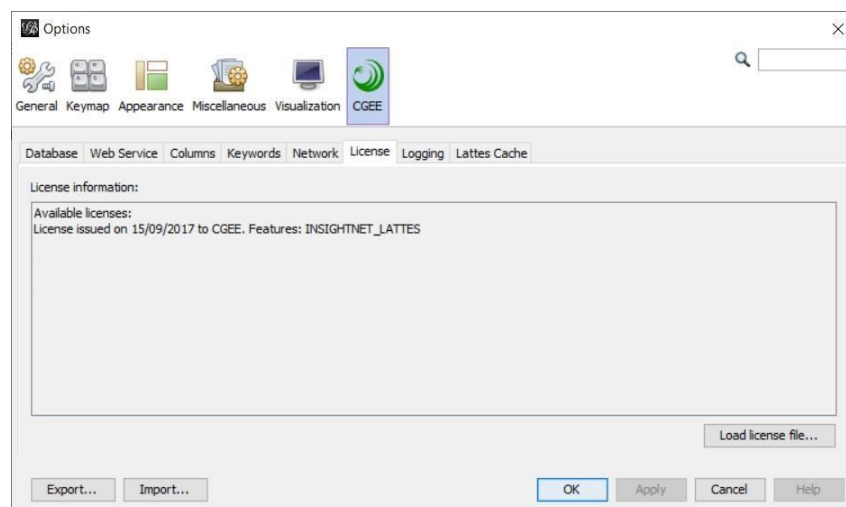


Figura 5.1: Licença requerida para o módulo de redes de Currículos Lattes

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

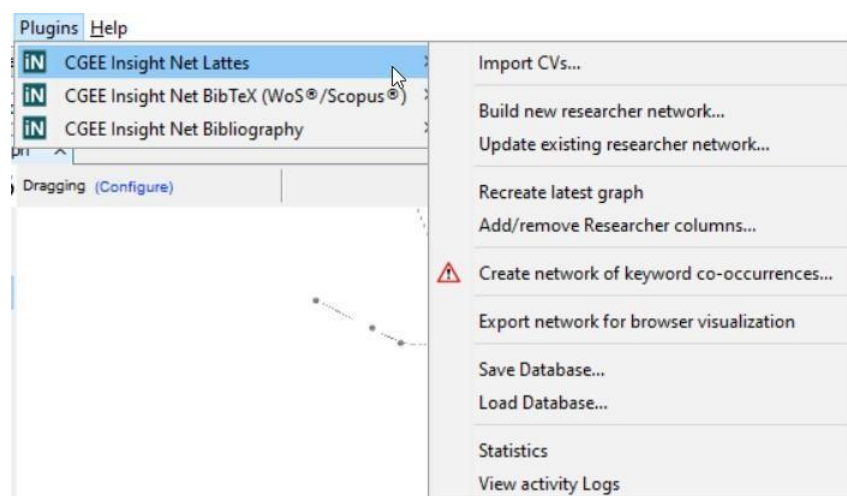


Figura 5.2: Sub-menu CGEE Insight Net Lattes

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:

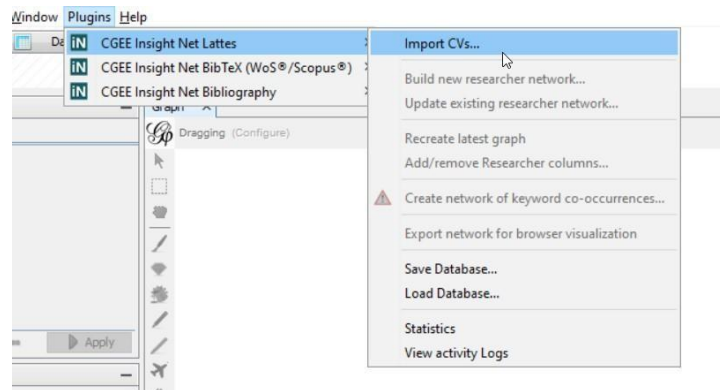


Figura 5.3: Menu de importação de Currículos

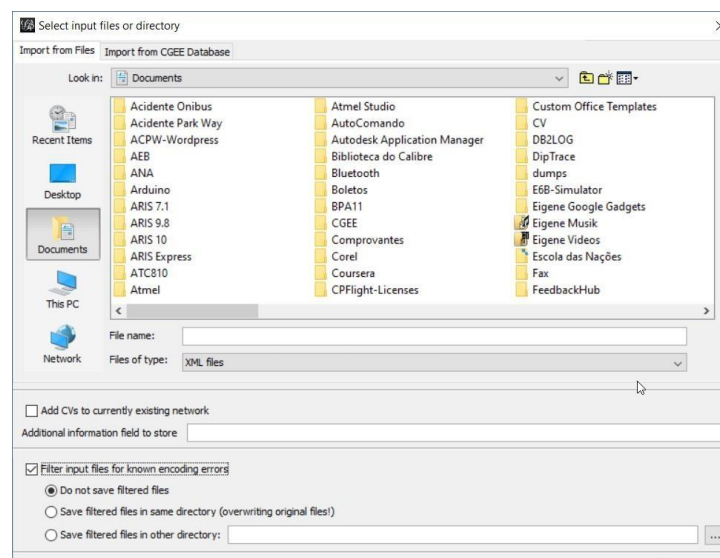


Figura 5.4: Diálogo de importação de Currículos em Arquivos

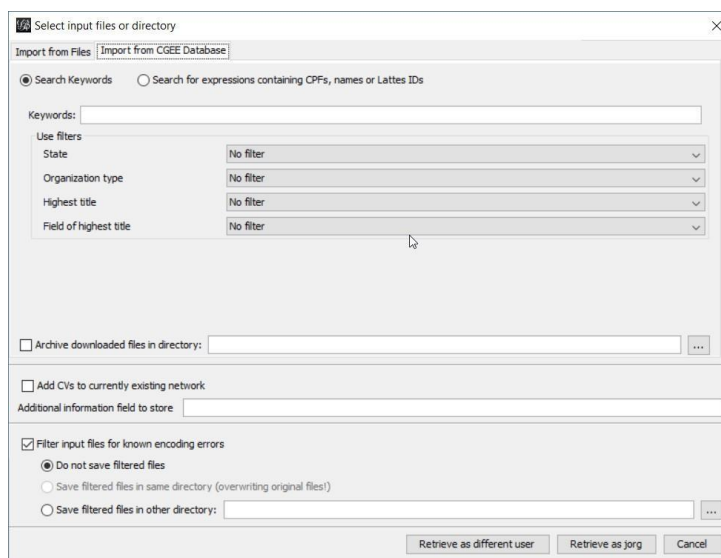


Figura 5.5: Diálogo de importação de Currículos do Banco de Dados

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>

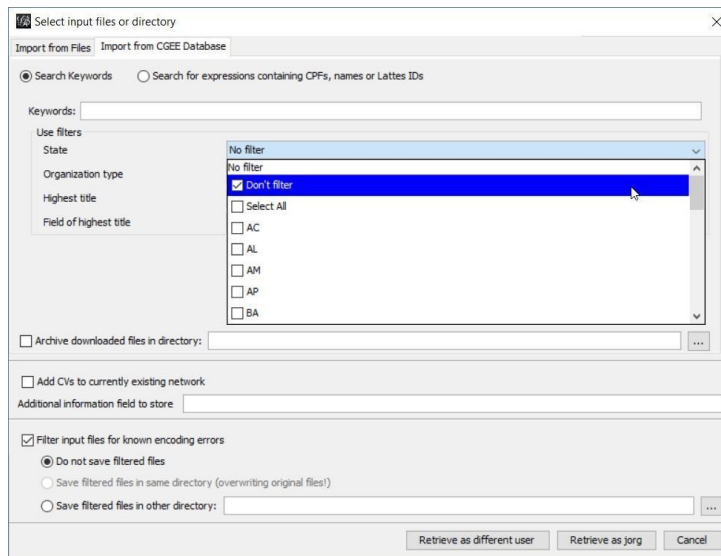


Figura 5.6: Importação de dados do Banco de dados do CGEE

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:

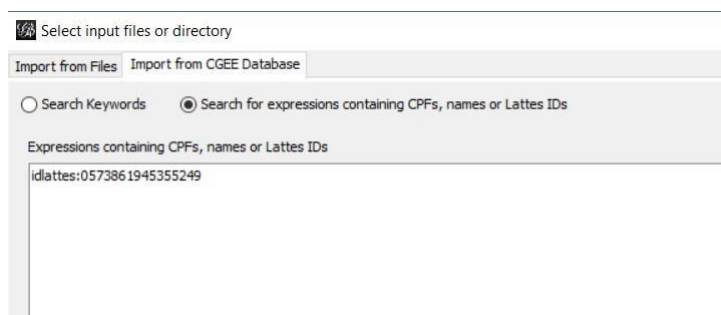


Figura 5.7: Recuperação de currículos por expressão de busca

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.

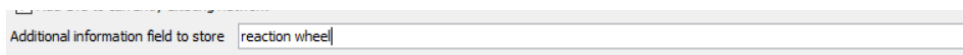


Figura 5.11: Opção do campo adicional de informação

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.

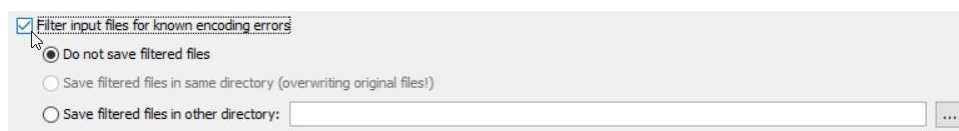


Figura 5.12: Limpeza dos dados e gravação dos arquivos corrigidos

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.

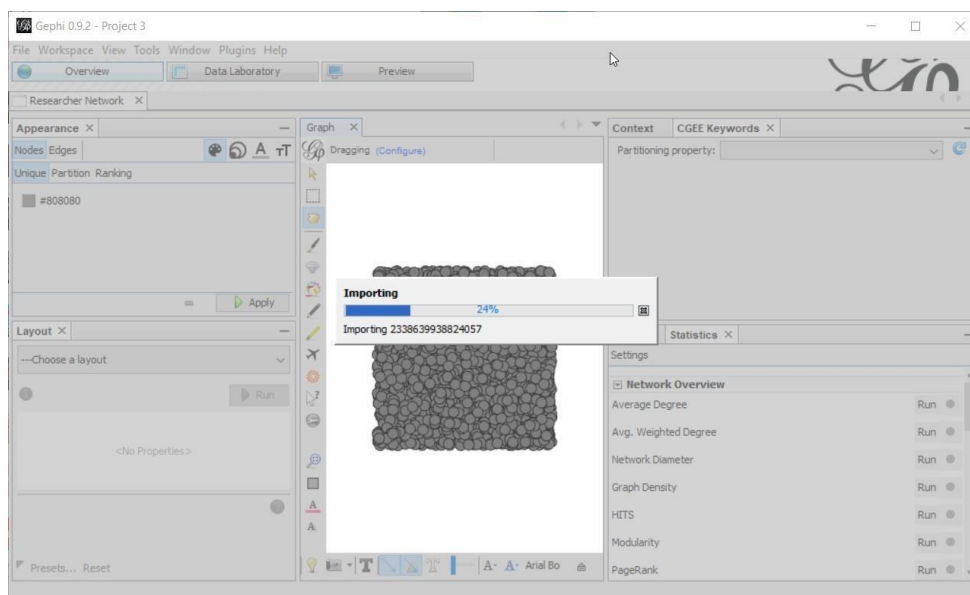


Figura 5.13: Importação dos currículos Lattes

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

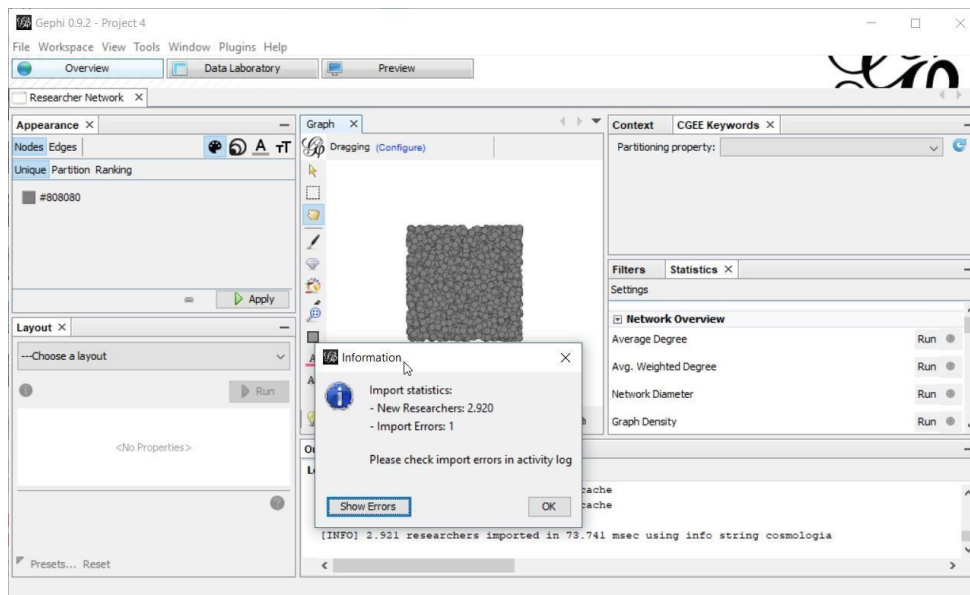


Figura 5.14: Resultado da importação dos currículos Lattes

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](#)).

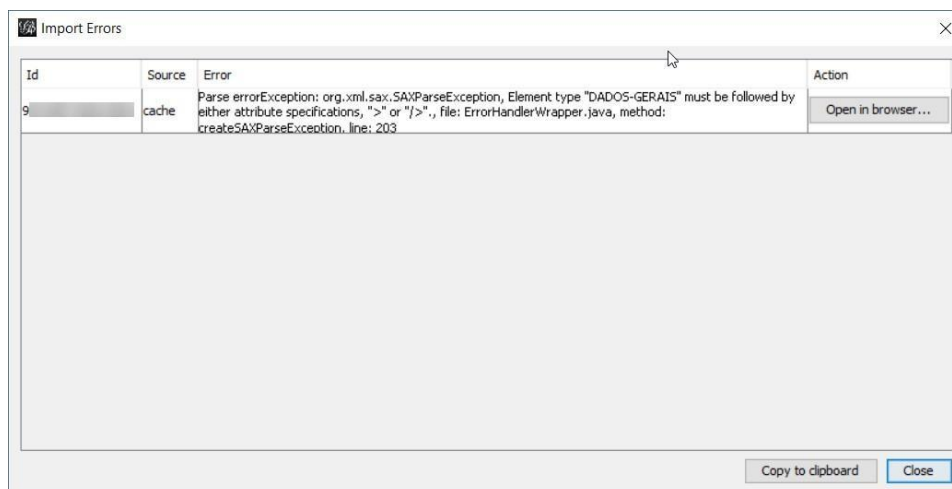


Figura 5.15: Diálogo de erros de importação

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:

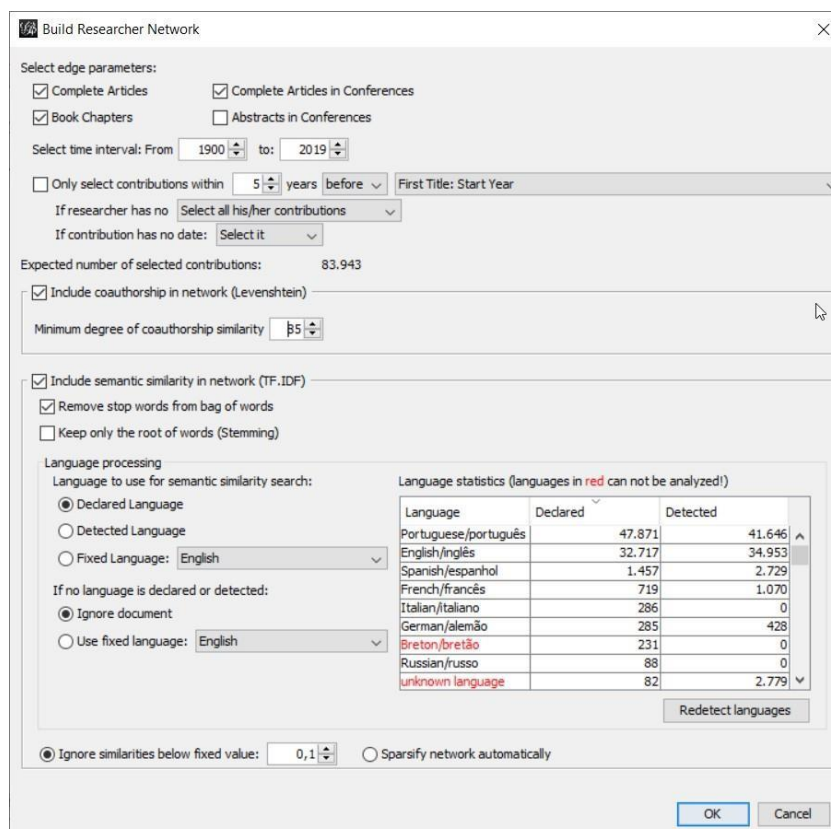
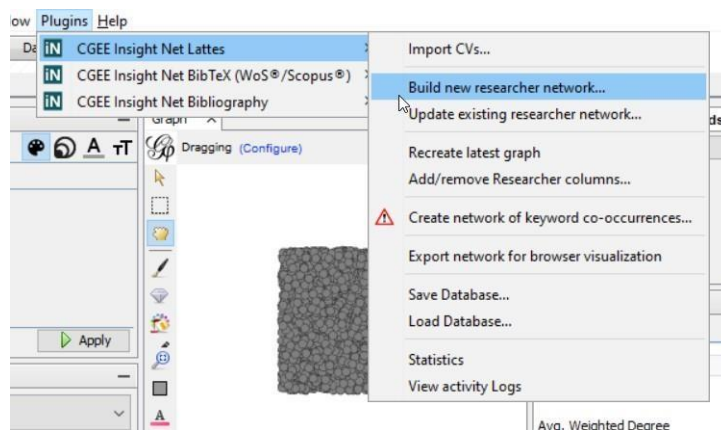


Figura 5.16: Menu e diálogo da formação da rede de currículos Lattes

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.

A visualização dos detalhes das contribuições Lattes (ver [Seção 5.3](#)) permite a inclusão e exclusão manual de contribuições bibliográficas no escopo de cálculo da rede. Caso for realizada alguma mudança de seleções desta forma, a seguinte informação é exibida pelo *plugin*:

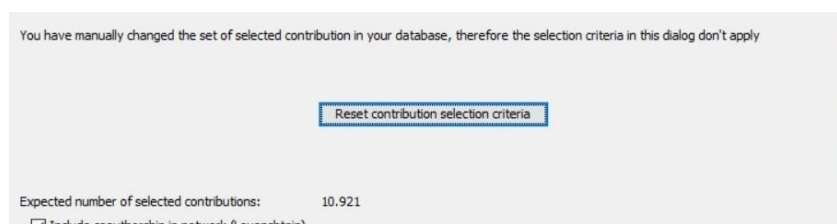


Figura 5.17: Informação sobre a alteração manual do escopo de cálculo da rede

Para desfazer as alterações manuais do escopo de rede, o usuário pode clicar no botão ‘*Reset contribution selection criteria*’. Com esta ação, o escopo de cálculo da rede volta aos critérios algorítmicos mencionados em cima.

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

A quantidade de contribuições selecionadas é calculada quando o diálogo for aberto e cada vez quando uma das opções mencionadas é alterada. Durante o tempo desse cálculo, as opções de seleção permanecem desabilitadas:

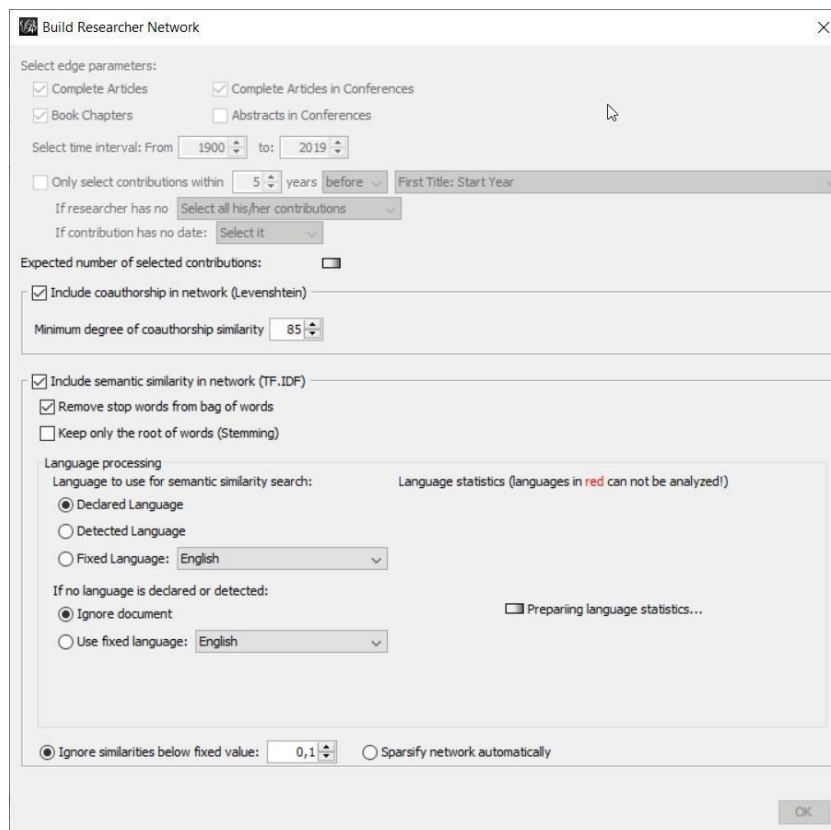


Figura 5.18: Recálculo da quantidade de contribuições selecionadas

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.
- 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.
- Se qualquer uma dessas opções for selecionada, o idioma do texto se torna relevante. Neste caso, aparece no diálogo a estatística de idiomas declarados e detectados nas contribuições.

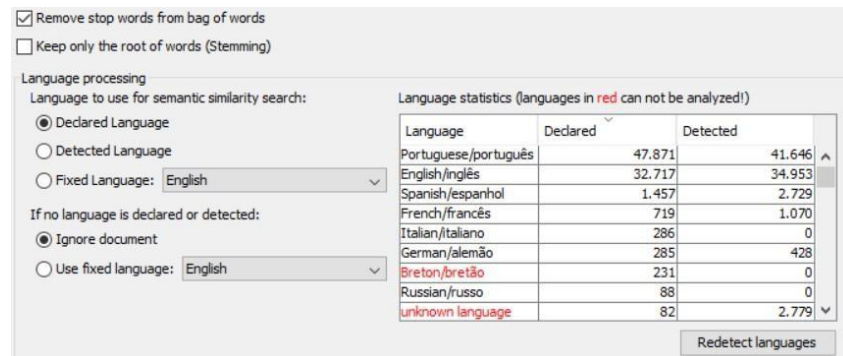


Figura 5.19: Estatística de idiomas detectados e declarados

- O botão “*Redetect languages*” permite realizar uma nova detecção de idiomas com parâmetros diferentes daqueles configurados na tela “*Languages*” da configuração do plugin (ver Seção 3.6):

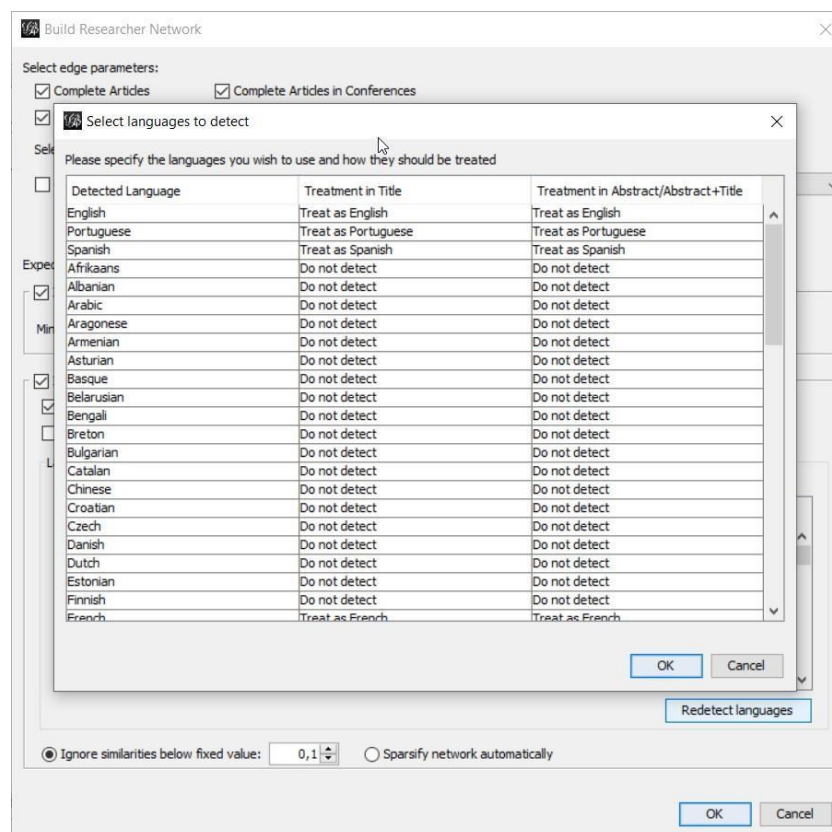


Figura 5.20: Nova detecção de idiomas

- 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:

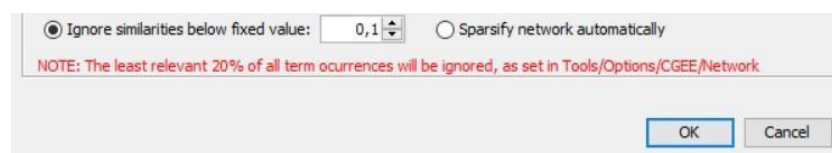


Figura 5.21: Aviso sobre configuração de limite inferior de relevância

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 de indicador de progresso:

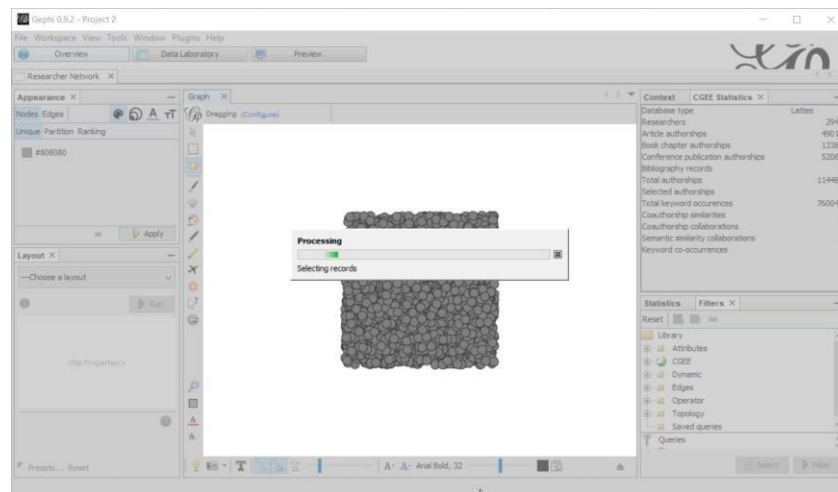


Figura 5.22: Seleção das contribuições e pré-processamento

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:

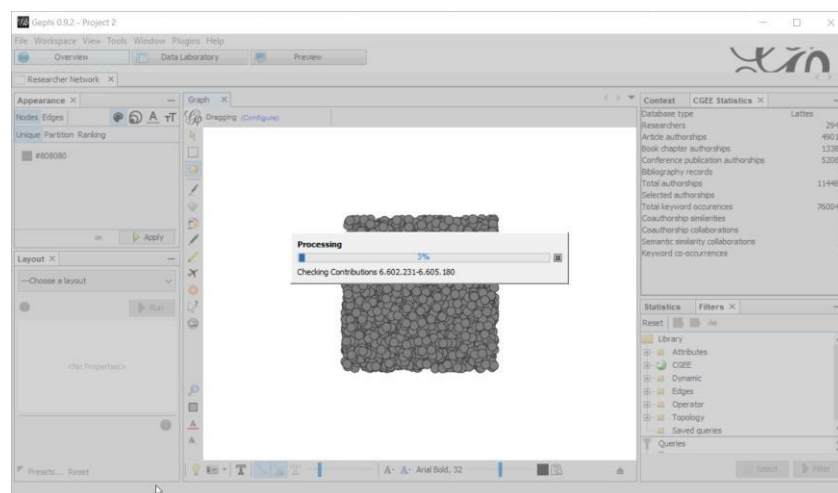


Figura 5.23: Processamento da pesquisa por similaridade

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

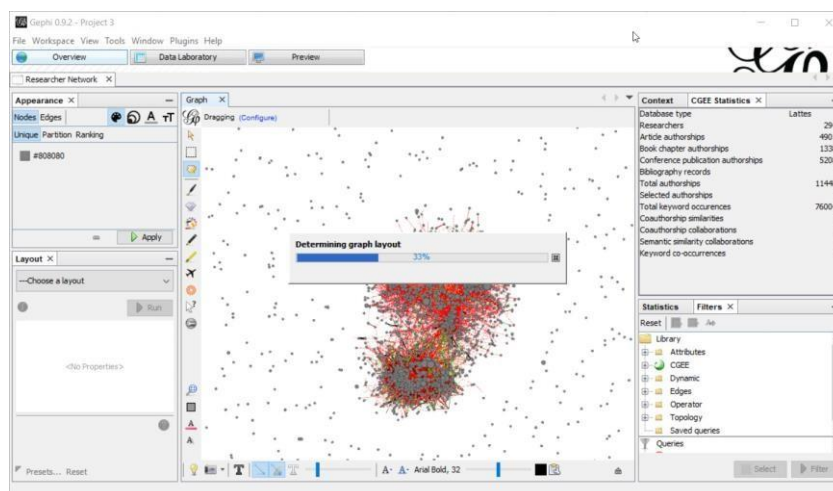


Figura 5.24: Pós-processamento e montagem da rede na tela

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. . .*”:

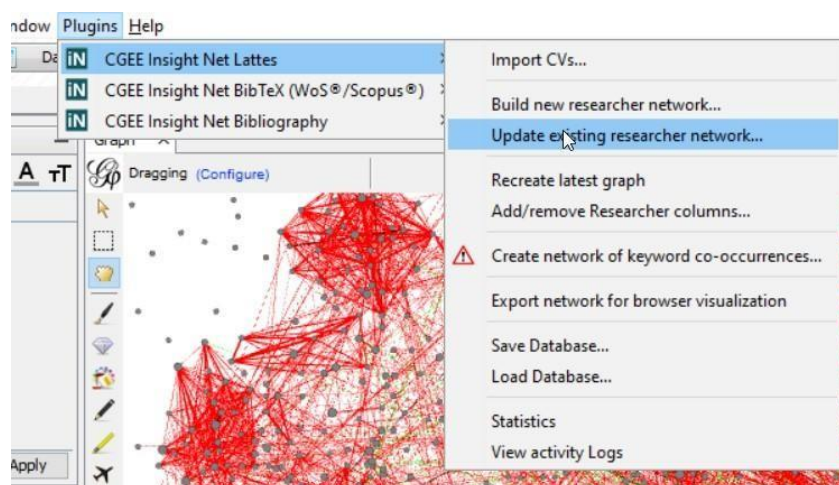


Figura 5.25: Atualização da rede já calculada

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. . .”:

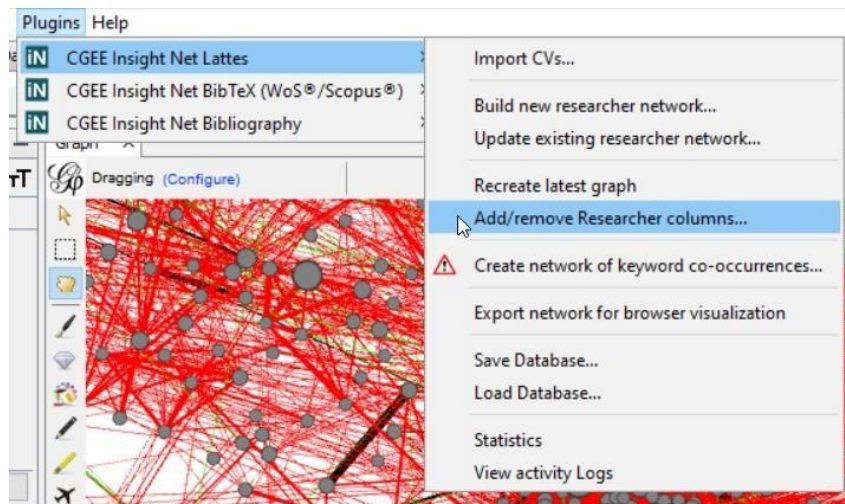


Figura 5.26: Funcionalidade para selecionar dados relevantes dos pesquisadores

Na seleção dessa funcionalidade, o seguinte diálogo é exibido:

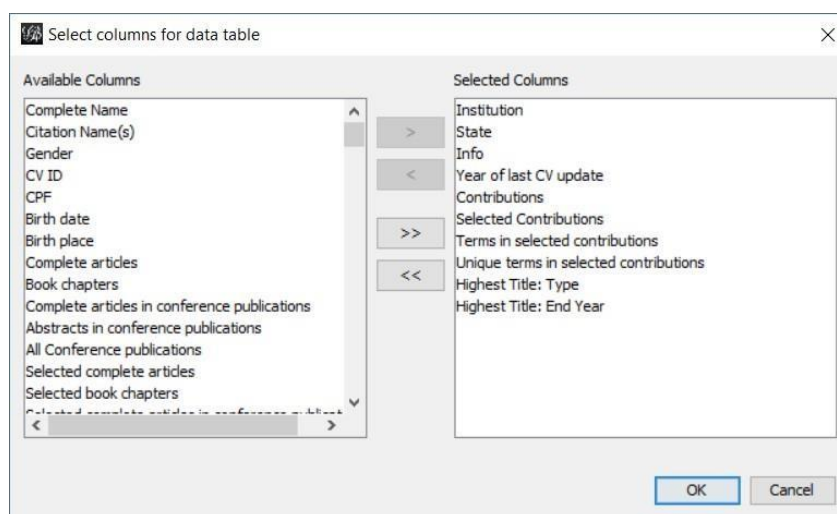

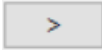
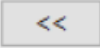




Figura 5.27: Seleção dos atributos dos pesquisadores

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:

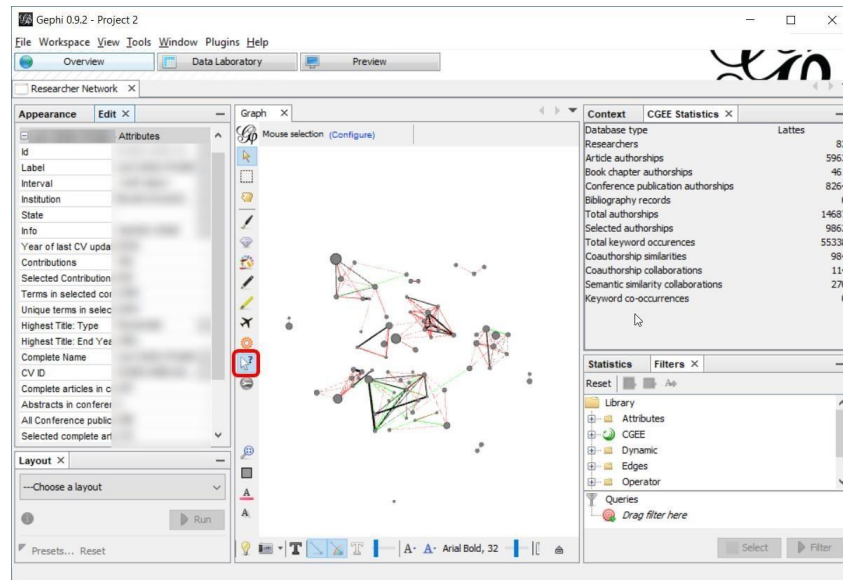



Figura 5.28: Exibição no grafo de todos os atributos habilitados do pesquisador

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 :

5.4 Visualização e edição das contribuições Lattes

A rede de coautorias e de similaridade semântica é calculada a partir das contribuições bibliográficas **selecionadas** que constam nos currículos Lattes dos pesquisadores importados. Os detalhes dessas contribuições são visualizados na aba do Laboratório de Dados do Gephi:

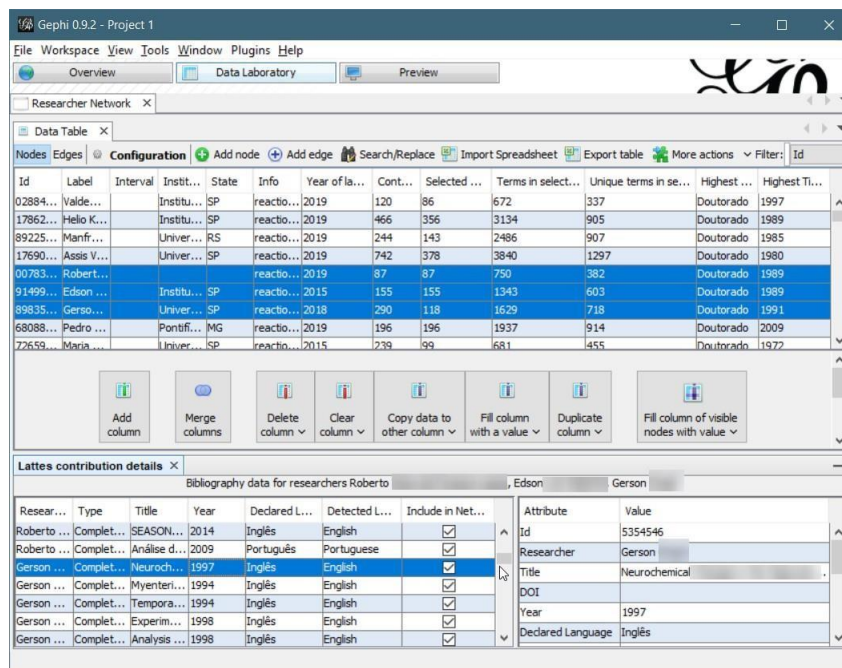
The screenshot shows the Gephi 0.9.2 interface. The main window displays a 'Data Table' with columns for researcher information and contribution statistics. Below the table is a toolbar with various data manipulation tools. A secondary window titled 'Lattes contribution details' is open, showing a list of publications for a selected researcher and a detailed view of one specific entry.

Id	Label	Interval	Instit...	State	Info	Year of la...	Cont...	Selected ...	Terms in select...	Unique terms in se...	Highest ...	Highest Ti...
02884...	Valde...		Institu...	SP	reactio...	2019	120	86	672	337	Doutorado	1997
17862...	Helio K...		Institu...	SP	reactio...	2019	466	356	3134	905	Doutorado	1989
89225...	Manfr...		Univer...	RS	reactio...	2019	244	143	2486	907	Doutorado	1985
17690...	Assis V...		Univer...	SP	reactio...	2019	742	378	3840	1297	Doutorado	1980
00783...	Robert...				reactio...	2019	87	87	750	382	Doutorado	1989
91499...	Edson ...		Institu...	SP	reactio...	2015	155	155	1343	603	Doutorado	1989
89835...	Gerso...		Univer...	SP	reactio...	2018	290	118	1629	718	Doutorado	1991
68088...	Pedro ...		Pontifi...	MG	reactio...	2019	196	196	1937	914	Doutorado	2009
72659...	Maria ...		Univer...	SP	reactio...	2015	239	99	681	455	Doutorado	1972

Type	Title	Year	Declared Lan...	Detected Lan...	Include in Networ...	Attribute	Value
Complete ...	Attitude D...	1997	Inglês	English	<input checked="" type="checkbox"/>	Id	5352308
Complete ...	ORBEST - ...	1997	Inglês	English	<input checked="" type="checkbox"/>	Researcher	Roberto
Complete ...	Reducing ...	1997	Inglês	English	<input checked="" type="checkbox"/>	Title	Reducing T...
Complete ...	Parameter ...	1994	Inglês	English	<input checked="" type="checkbox"/>	DOI	
Complete ...	Optimal Es...	1988	Inglês	English	<input checked="" type="checkbox"/>	Year	1997
Complete ...	Chaos in S...	1999	Inglês	English	<input checked="" type="checkbox"/>	Declared Language	Inglês
Complete ...	Rigid Body...	1999	Inglês	English	<input checked="" type="checkbox"/>		

Figura 5.30: Exibição dos detalhes das contribuições bibliográfica de um Currículo Lattes

O *CGEE Insight Net* permite a seleção de vários pesquisadores com as funcionalidades de *Shift-Click* e *Ctrl-Click* e mostra a produção para a união deles, exibindo adicionalmente uma coluna com os nomes. Se na lista das contribuições for selecionada exatamente uma contribuição bibliográfica, os dados dela são exibidos no lado direito:



The screenshot shows the Gephi 0.9.2 interface. The top window is titled 'Gephi 0.9.2 - Project 1' and contains a menu bar (File, Workspace, View, Tools, Window, Plugins, Help) and a toolbar with buttons for Overview, Data Laboratory, and Preview. Below this is a 'Researcher Network' tab. The main area is a 'Data Table' with columns: Id, Label, Interval, Instit..., State, Info, Year of la..., Cont..., Selected ..., Terms in select..., Unique terms in se..., Highest ..., Highest Ti... The table lists several researchers with their respective data. Below the table is a toolbar with buttons: Add column, Merge columns, Delete column, Clear column, Copy data to other column, Fill column with a value, Duplicate column, and Fill column of visible nodes with value. At the bottom, a 'Lattes contribution details' window is open, showing 'Bibliography data for researchers Roberto, Edson, Gerson'. It contains a table with columns: Resear..., Type, Title, Year, Declared L..., Detected L..., Include in Net..., Attribute, and Value. The table lists contributions for Roberto, Edson, and Gerson. The 'Include in Net...' column has checkboxes, some of which are checked. The 'Attribute' column shows details for the selected contribution, such as Id (5354546), Researcher (Gerson), Title (Neurochemical), DOI, Year (1997), and Declared Language (Inglês).

Figura 5.31: Exibição dos detalhes das contribuições bibliográfica de vários Currículos Lattes

A tabela de contribuições fornece várias funcionalidades de edição e cópia.

- O idioma **declarado** das publicações individuais pode ser alterado de acordo com a lista de idiomas que podem ser analisados.
- Contribuições individuais podem ser incluídas ou excluídas do escopo de cálculo da rede de co-autorias e similaridades semânticas, clicando na caixa de seleção na coluna “Include in Network build”. Neste caso, a seleção por critérios do diálogo *Build new Researcher Network* não estará disponível e exibirá apenas o botão ‘Reset contribution selection criteria’
- Uma lista de contribuições pode ser incluída ou excluída do escopo de cálculo da rede de co-autorias e similaridade semântica, selecionando as contribuições com *Shift-Click* ou *Ctrl-Click*. Depois deve ser clicado o botão direito do mouse em cima da tabela de contribuições. No menu *popup* que é exibido, o usuário pode selecionar *Include all selected contributions in network build* ou *Exclude all selected contributions from network build*:

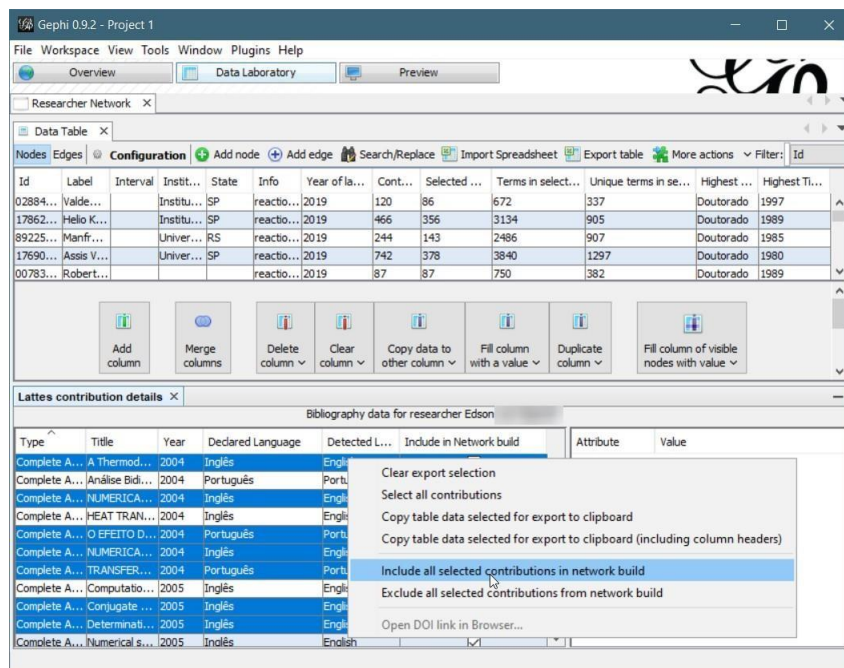


Figura 5.32: Inclusão ou exclusão de várias contribuições bibliográficas de Currículos Lattes no escopo do cálculo de rede

- Um clique com botão direito na lista de contribuições permite selecionar ou deselecionar todas as contribuições a partir das opções *Select all contributions* e *Clear exportSelections*
- As contribuições selecionadas por *Shift-Click*, *Ctrl-Click* e pela opção *Select all contributions* podem ser exportadas para a área de transferência do computador e posteriormente inseridas em ferramentas como *Word®* ou *Excel®*. Para isso existem as opções *Copy table data selected for export to clipboard* e *Copy table data selected for export to clipboard (including column headers)* no menu de *popup* que é exibido com clique do botão direito na tabela de contribuições.
- Para contribuições que possuem um *Document Object Identifier (DOI)*, a referência bibliográfica pode ser aberta na internet com a opção *Open DOI link in Browser...*

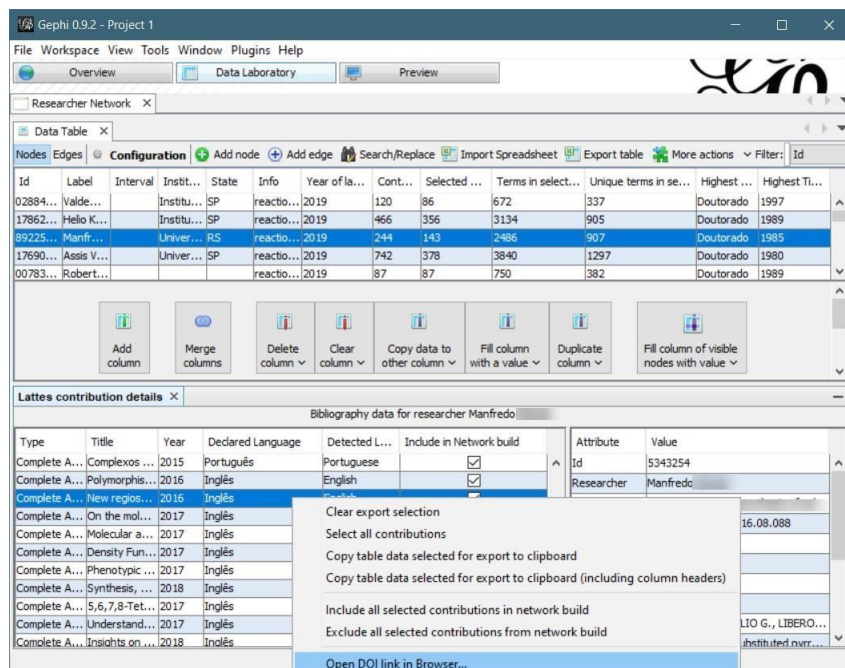


Figura 5.33: Exibir o Document Object Identifier no Browser

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 3.7](#). Essa licença é exibida assim no diálogo *Tools > Options > CGEE > License*:

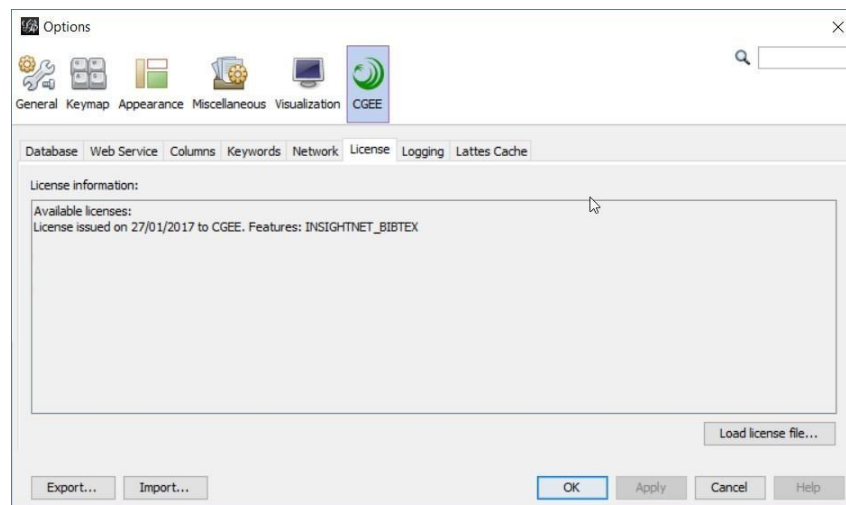


Figura 6.1: Licença requerida para o módulo de redes bibliográficas

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

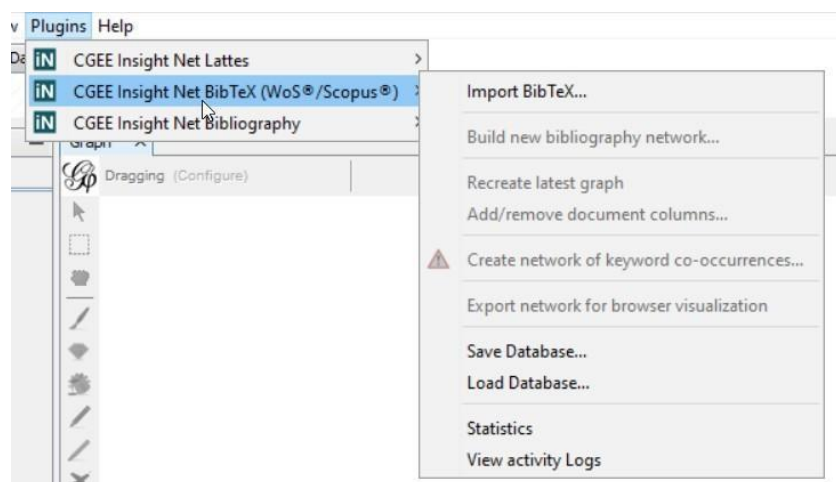


Figura 6.2: Sub-menu CGEE Insight Net BibTeX (WoS®/Scopus®)

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.

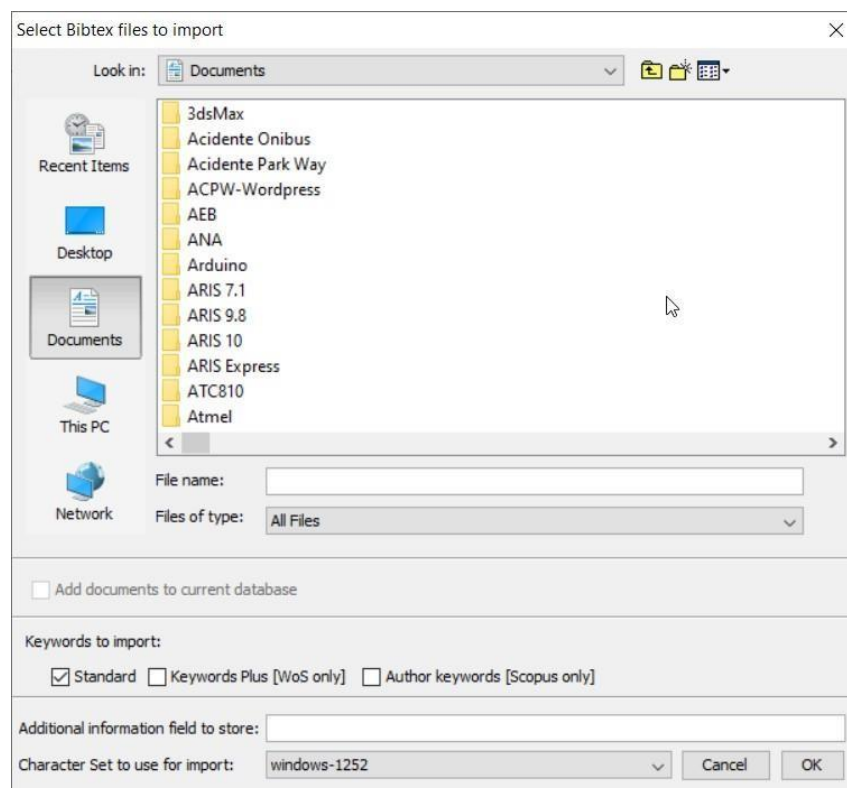
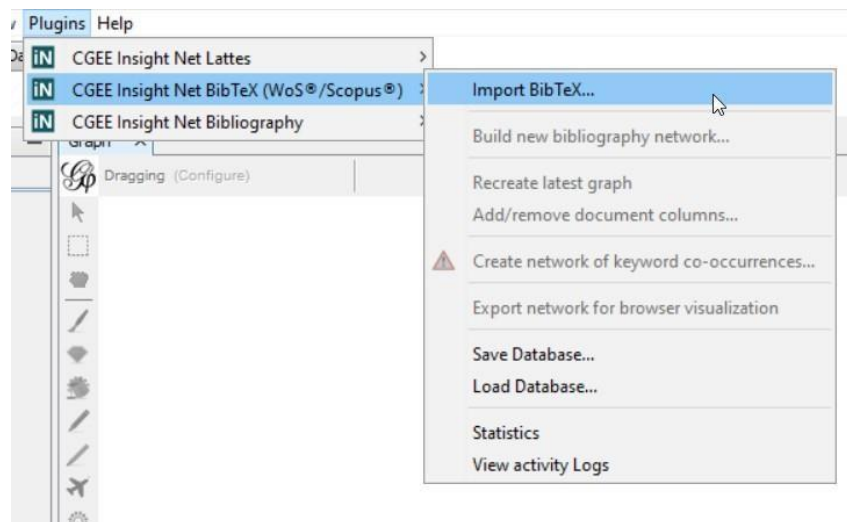


Figura 6.3: Importação de arquivos BibTeX

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.

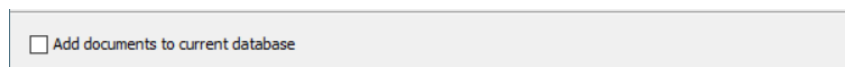


Figura 6.4: Opção de importação inicial ou incremental

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 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 “*DOP*”), 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 6.6: Campo adicional de informação

6.1.4 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.

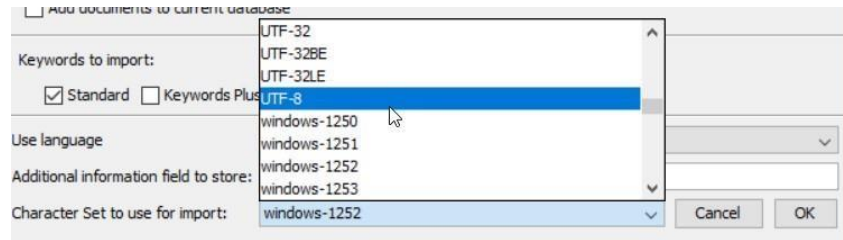


Figura 6.7: Definição do conjunto de caracteres

Os charactersetts 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 referênciabibliográfica.

6.1.5 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).

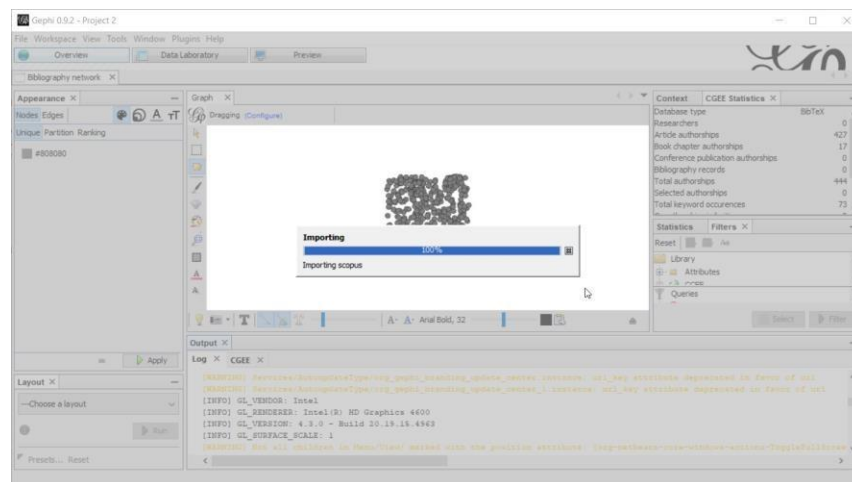


Figura 6.8: Importação de referências bibliográficas

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:

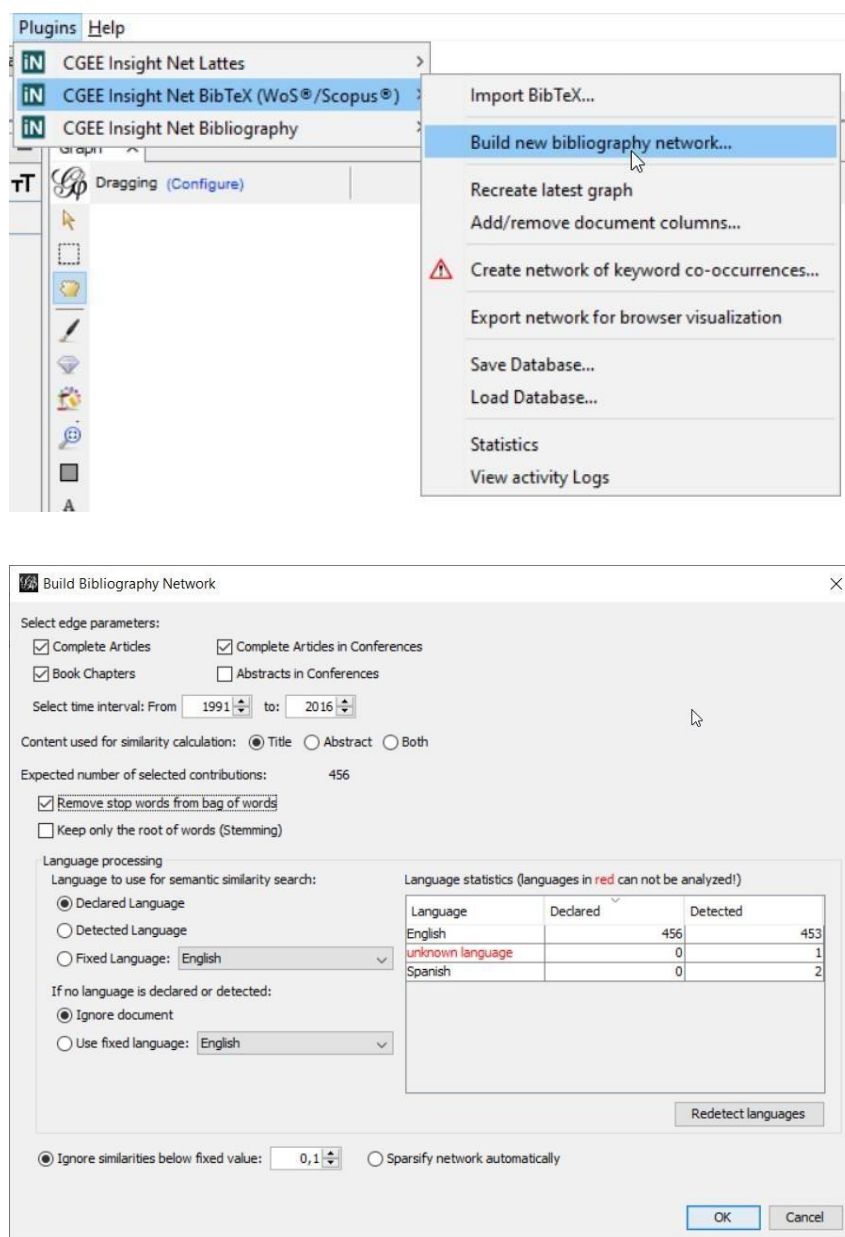


Figura 6.9: Menu e diálogo da formação da rede de referências bibliográficas

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.
 - Se qualquer uma dessas opções for selecionada, o idioma do texto se tornarelevante. Neste caso, aparece no diálogo a estatística de idiomas declarados e detectados, de acordo com a seleção no campo “*Context used for similarity calculation*”.

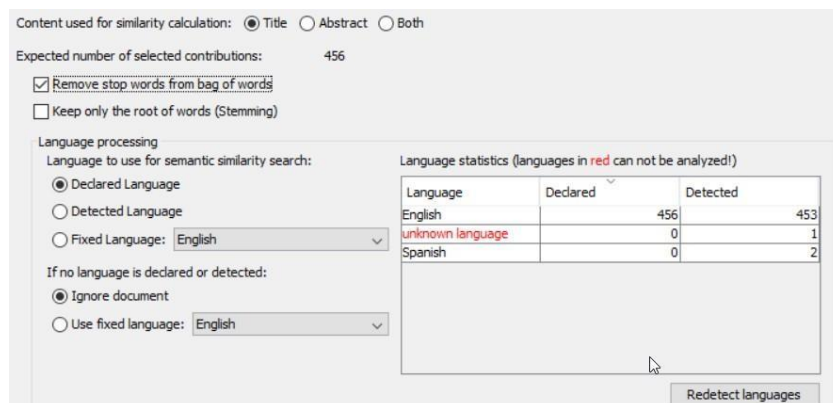


Figura 6.10: Estatística de idiomas detectados e declarados

- O botão “Redetect languages” permite realizar uma nova detecção de idiomas com parâmetros diferentes daqueles configurados na tela “Languages” da configuração do plugin (ver Seção 3.6):

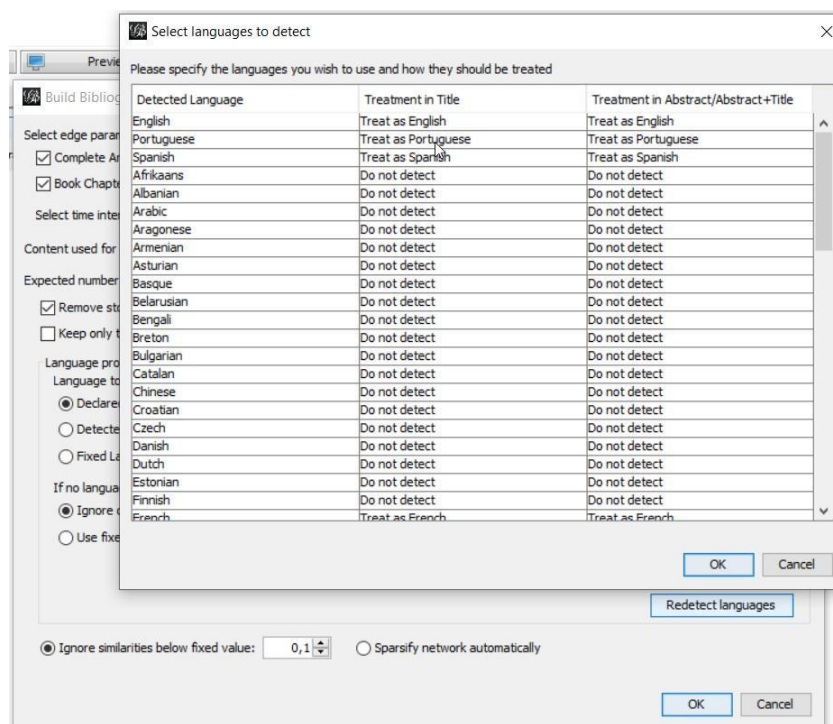


Figura 6.11: Nova detecção de idiomas

- Adicionalmente, o usuário pode alterar os idiomas declarados e detectados no laboratório de dados do Gephi, para corrigir possíveis erros nesses dados:

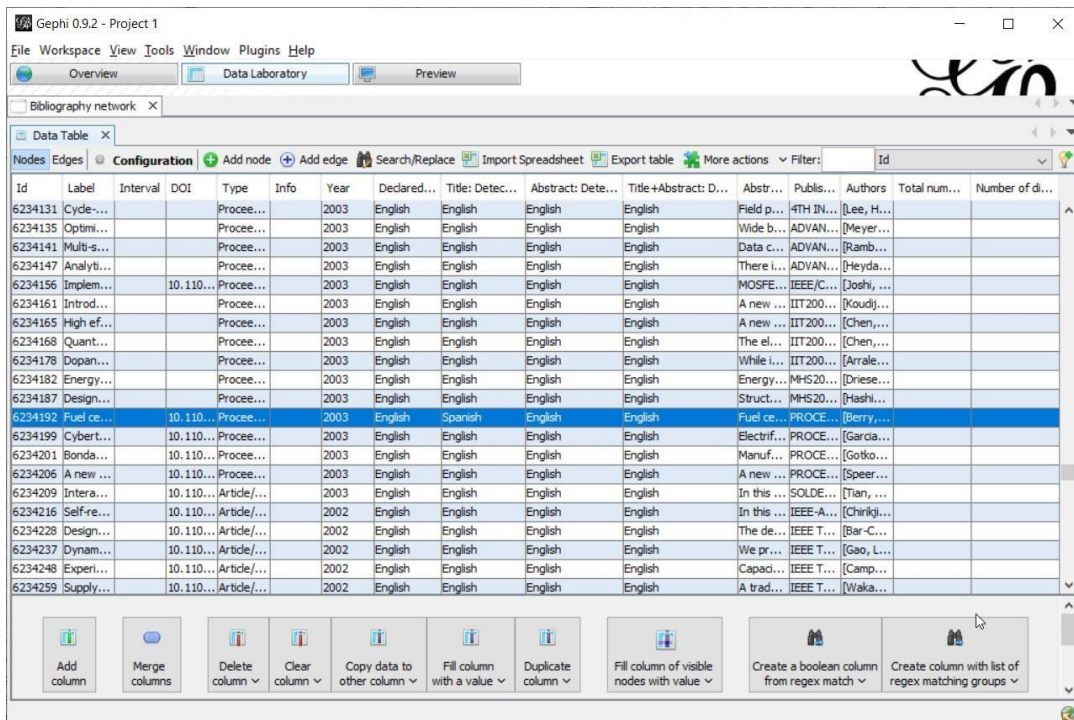


Figura 6.12: Correção manual dos idiomas das contribuições no laboratório de dados do Gephi

- 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:

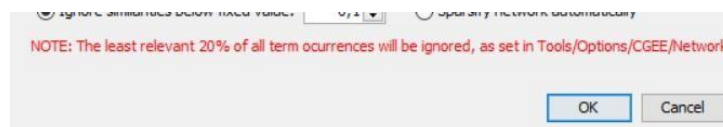


Figura 6.13: Aviso sobre configuração de limite inferior de relevância

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 3.7. Essa licença é exibida assim no diálogo *Tools > Options > CGEE > License*:

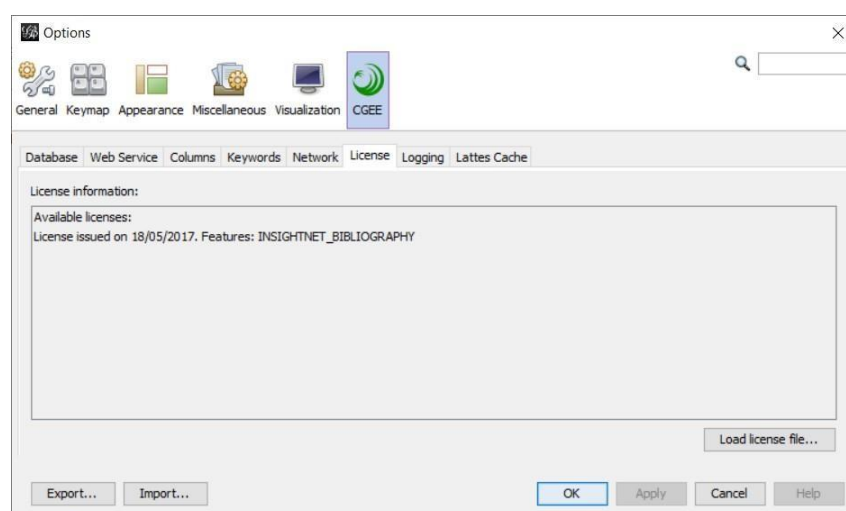


Figura 7.1: Licença requerida para o módulo de redes bibliográficas

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

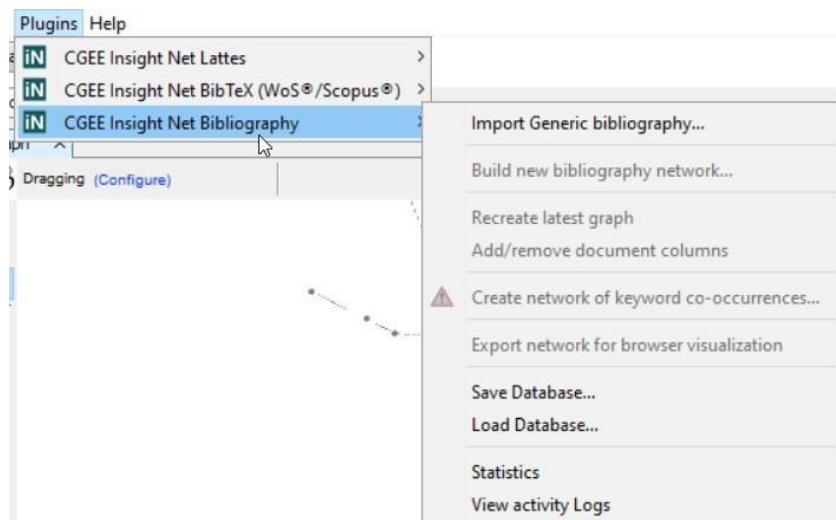


Figura 7.2: Sub-menu *CGEE Insight Net Bibliography*

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.

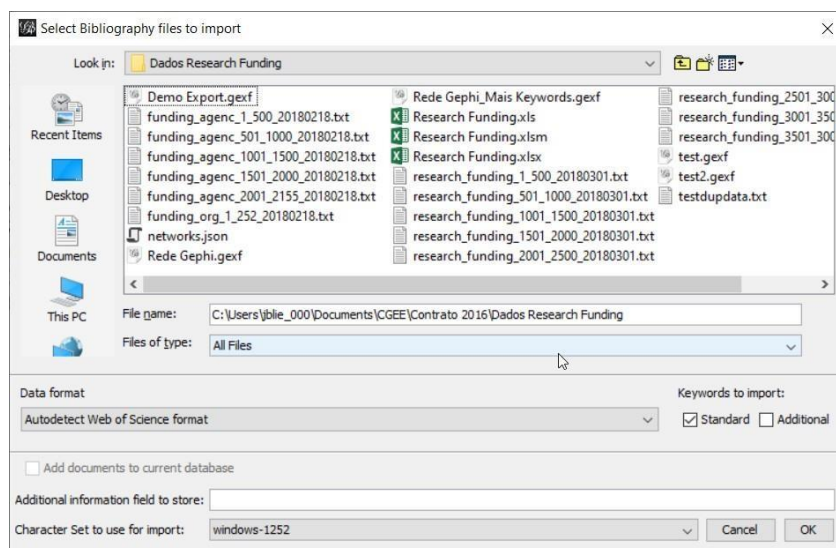
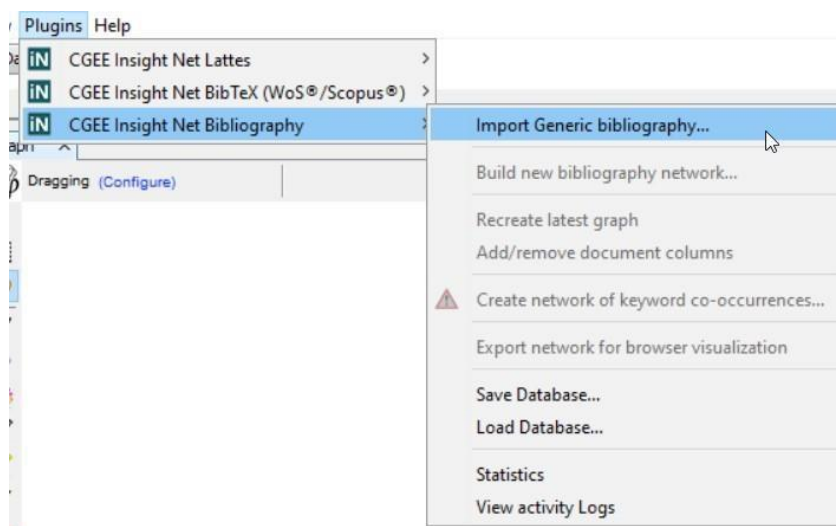


Figura 7.3: Importação de arquivos de bibliografia

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*
- Dados genéricos em tabelas *Excel* (funcionalidade experimental)

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:

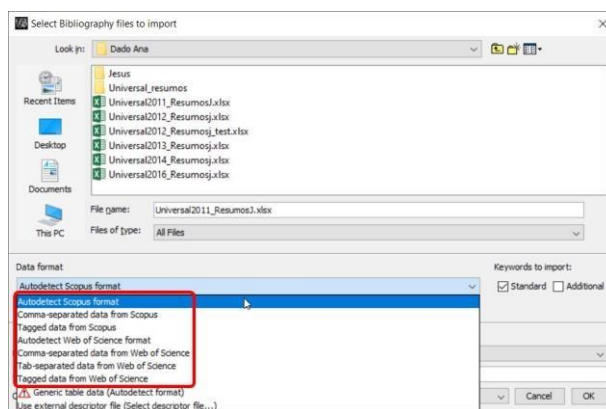


Figura 7.4: Formatos de arquivos de bibliografia dos serviços *Web of Science*[®] e *Scopus*[®]

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:

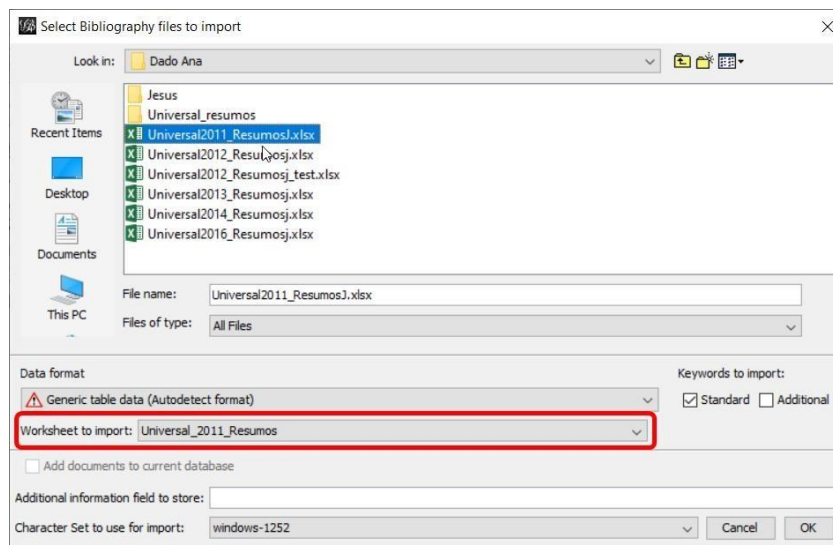


Figura 7.5: Seleção da aba da planilha excel no formato “Generic Table data (Autodetect format)”

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 deduplicar referências bibliográficas durante a importação
<i>DOI</i>	Contém o <i>Document Object Identifier</i> , também usado para a deduplicação
<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 [Seção 10.1](#).

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.

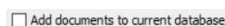


Figura 7.7: Opção de importação inicial ou incremental

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 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 7.8: Campo adicional de informação

7.1.5 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.

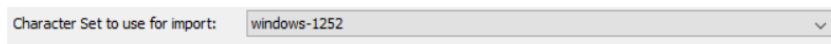


Figura 7.9: Definição do conjunto de caracteres

Os charactersetts 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 referênciabibliográfica.

7.1.6 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).

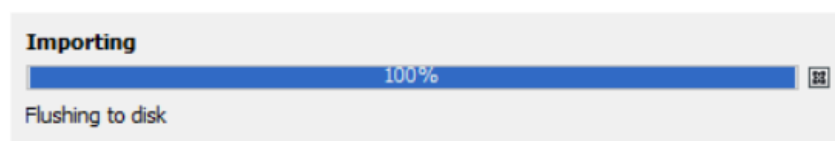


Figura 7.10: Importação de referências bibliográficas

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:

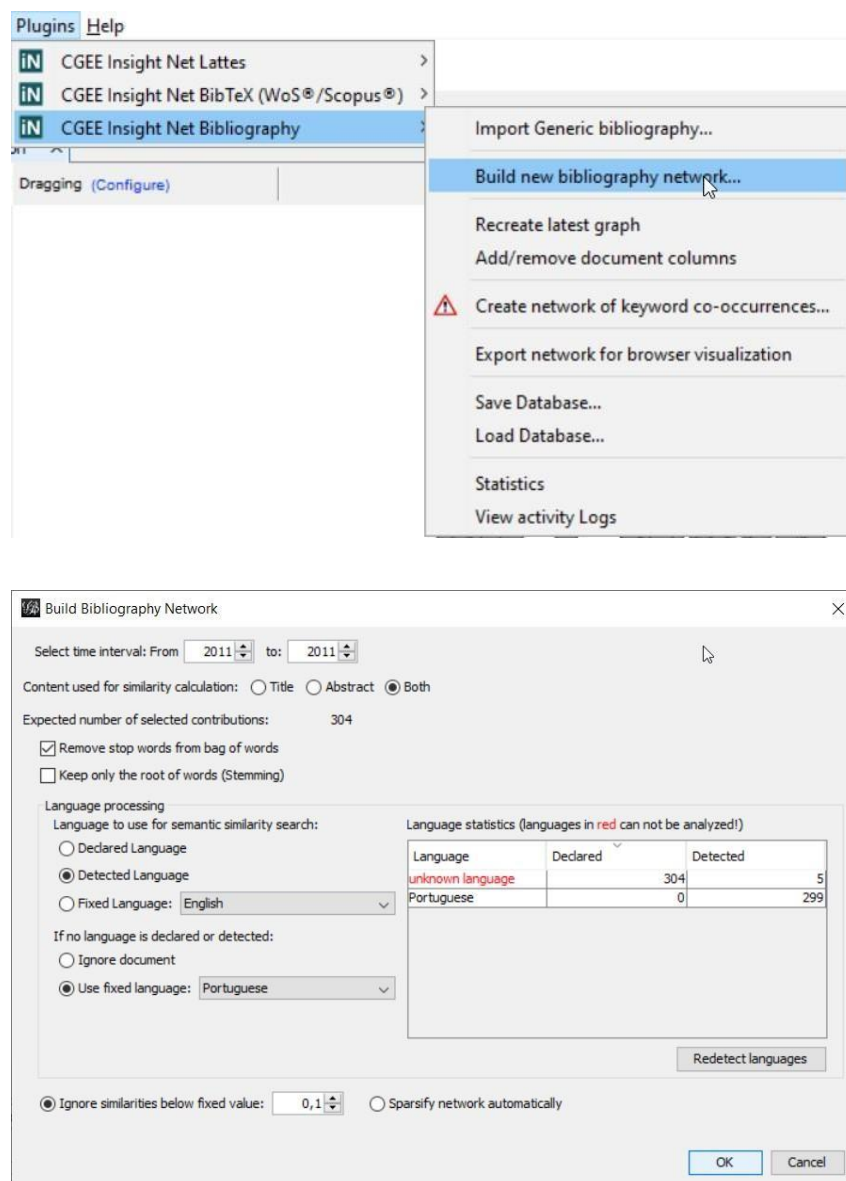


Figura 7.11: Menu e diálogo da formação da rede de referências bibliográficas

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.
 - Se qualquer uma dessas opções for selecionada, o idioma do texto se torna relevante. Neste caso, aparece no diálogo a estatística de idiomas declarados e detectados, de acordo com a seleção no campo “*Context used for similarity calculation*”.

Content used for similarity calculation: Title Abstract Both

Expected number of selected contributions: 304

Remove stop words from bag of words
 Keep only the root of words (Stemming)

Language processing

Language to use for semantic similarity search:

Declared Language
 Detected Language
 Fixed Language: English

If no language is declared or detected:

Ignore document
 Use fixed language: English

Language statistics (languages in red can not be analyzed!)

Language	Declared	Detected
unknown language	304	57
Spanish	0	8
English	0	1
Portuguese	0	238

Redetect languages

Figura 7.12: Estatística de idiomas detectados e declarados

- O botão “*Redetect languages*” permite realizar uma nova detecção de idiomas com parâmetros diferentes daqueles configurados na tela “*Languages*” da configuração do plugin (ver Seção 3.6):

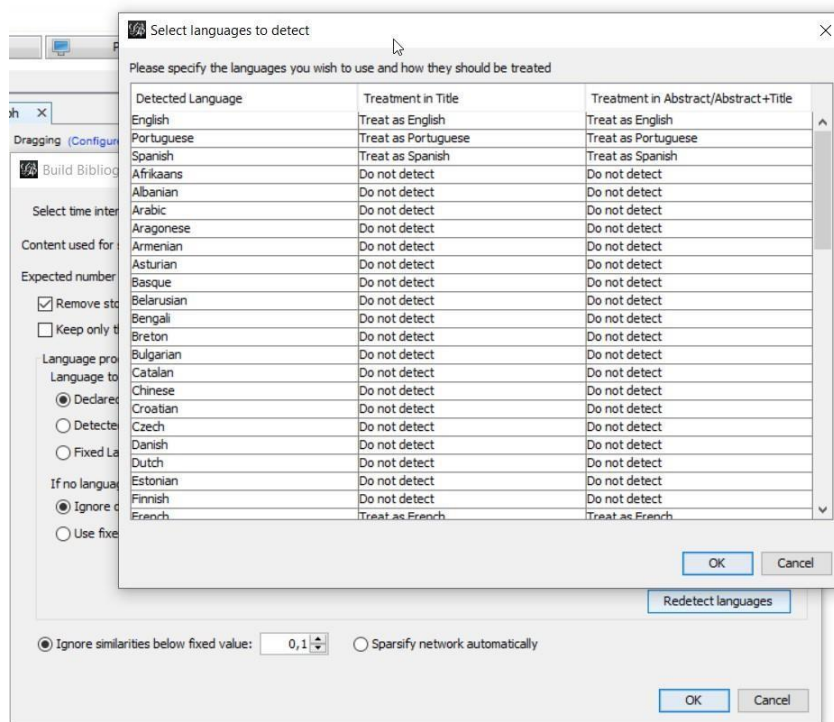


Figura 7.13: Nova detecção de idiomas

- Adicionalmente, o usuário pode alterar os idiomas declarados e detectados no laboratório de dados do Gephii, para corrigir possíveis erros nesses dados:

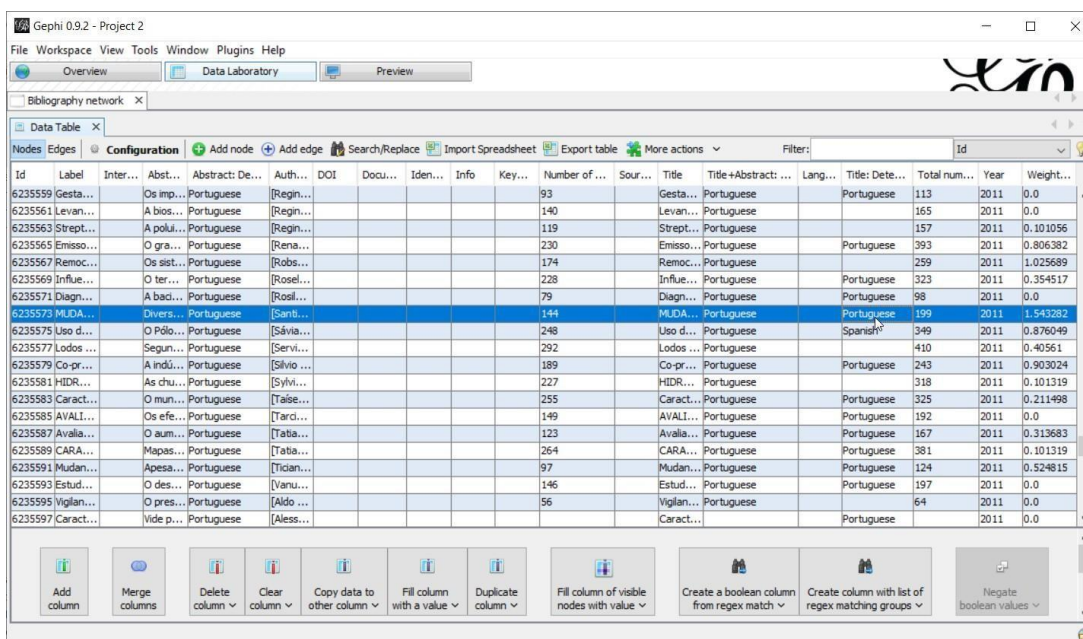


Figura 7.14: Correção manual dos idiomas das contribuições no laboratório de dados do Gephii

- 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 quantidades 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 7.15: 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”:

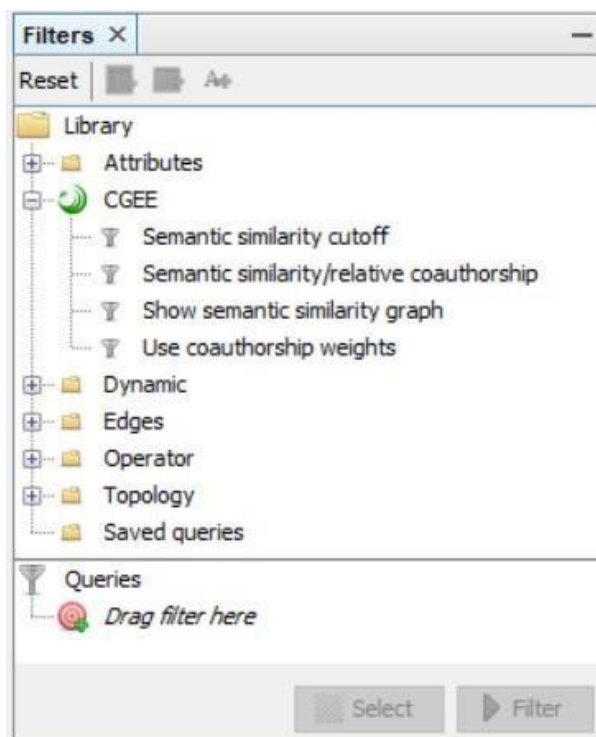


Figura 8.1: Filtros da categoria “CGEE”

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:

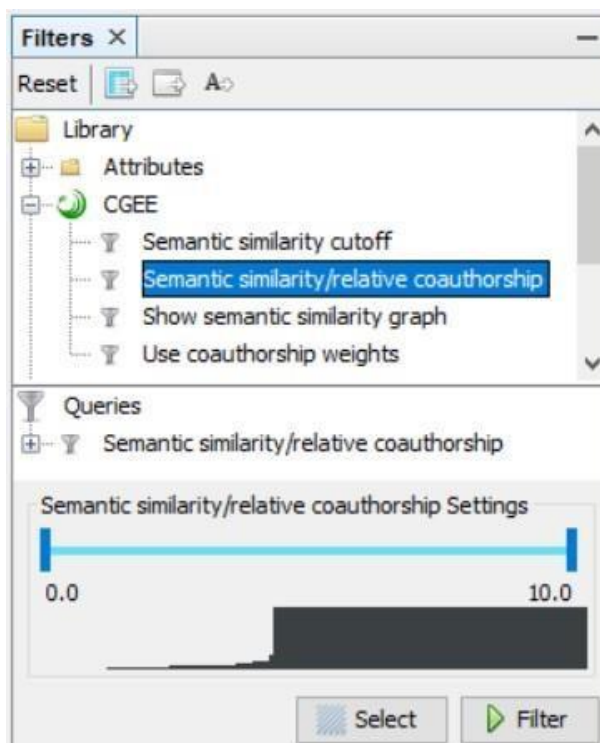


Figura 8.2: Parâmetros do filtro *Semantic similarity/relative coauthorship*

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.

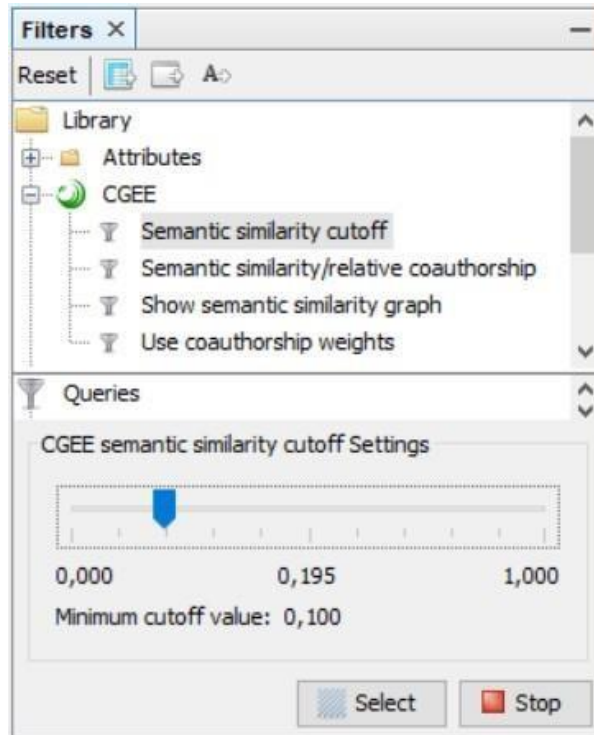


Figura 8.3: Parâmetro do filtro *Semantic similarity cutoff*

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	1	0,104	0,479
0,585	5	0,269	0,585
0,104	1	0,104	0,104
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”.

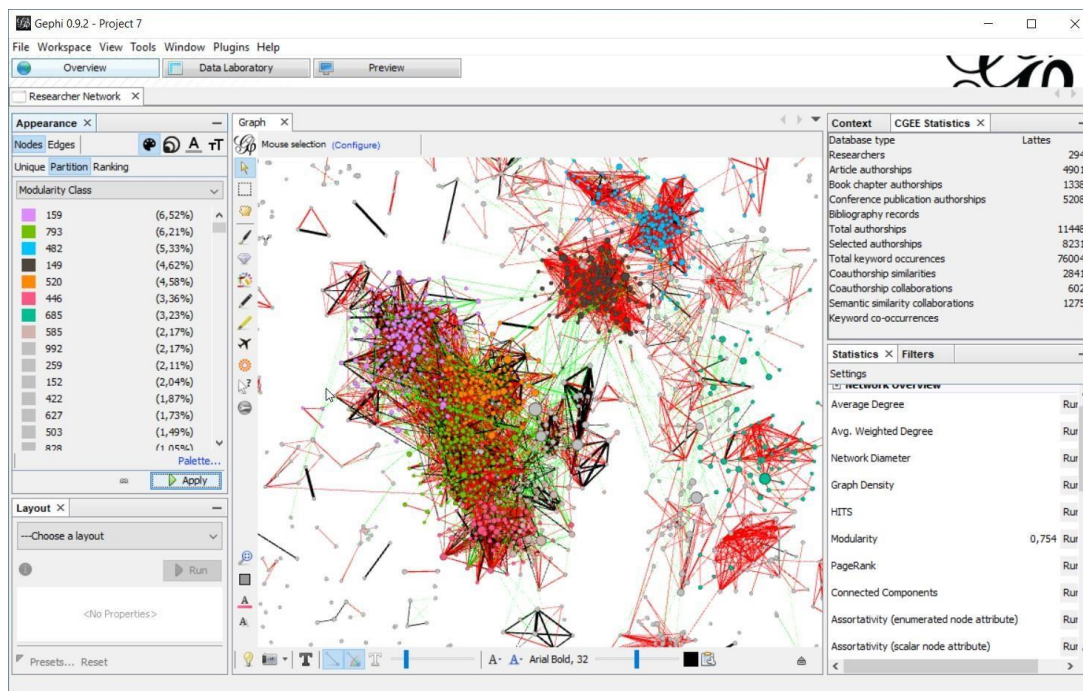


Figura 8.6: Clusters na rede gerada

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”:

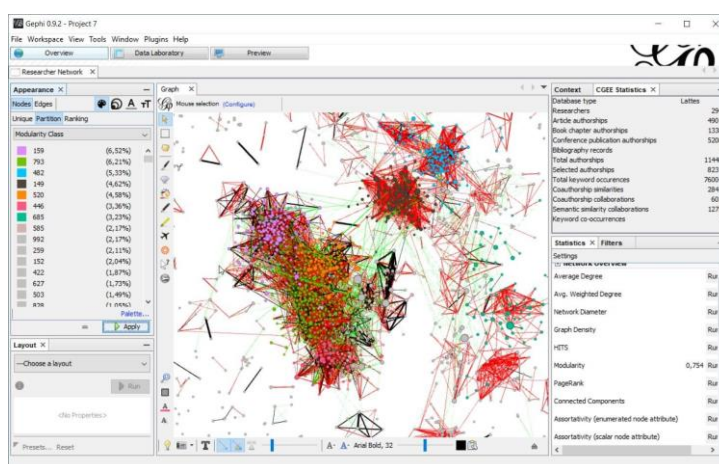



Figura 8.7: Estatísticas de Assortatividade

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(s) atributo(s) com relação ao qual(is) a assortatividade deve ser calculada:

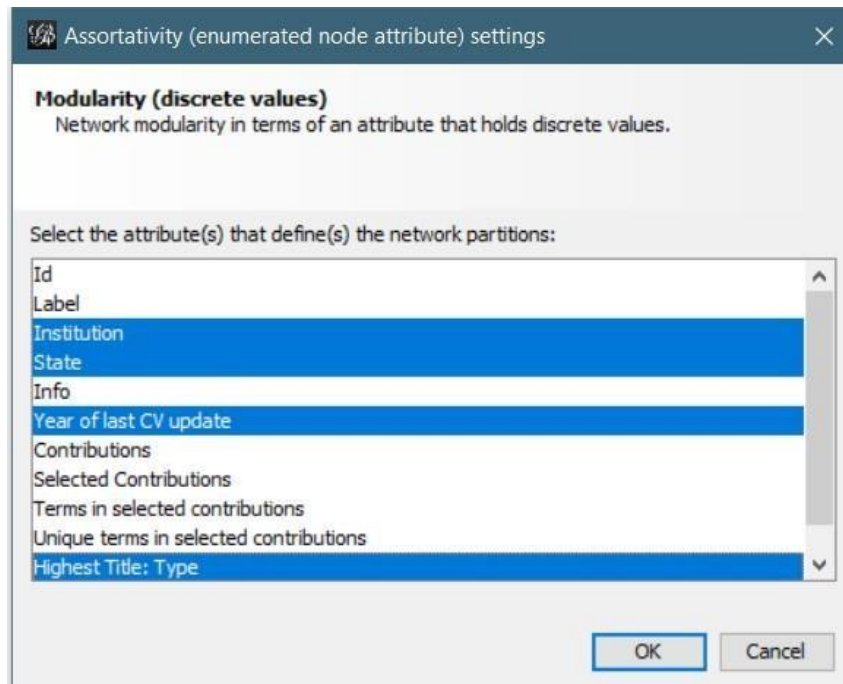


Figura 8.8: Seleção do(s) atributo(s) de assortatividade

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

HTML Report

Graph assortativity according to node attributes

Results:

Attribute	Assortativity coefficient	Intermediate results	
		Newman modularity	Max. possible Newman modularity
Institution	0,125	0,115	0,921
State	0,133	0,089	0,672
Year of last CV update	0,166	0,085	0,514
Highest Title: Type	0,019	0,002	0,121

Print Copy Save Close

Filters	Statistics ×
Settings	
Graph Density	Run ●
HITS	Run ●
Modularity	Run ●
PageRank	Run ●
Connected Components	Run ●
Assortativity (enumerated node attribute)	Year of last CV update: 0,17 Run ?
Assortativity (scalar node attribute)	Run ●
Percolation estimators (EXPERIMENTAL)	Run ●
Max entropy cutoff (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:

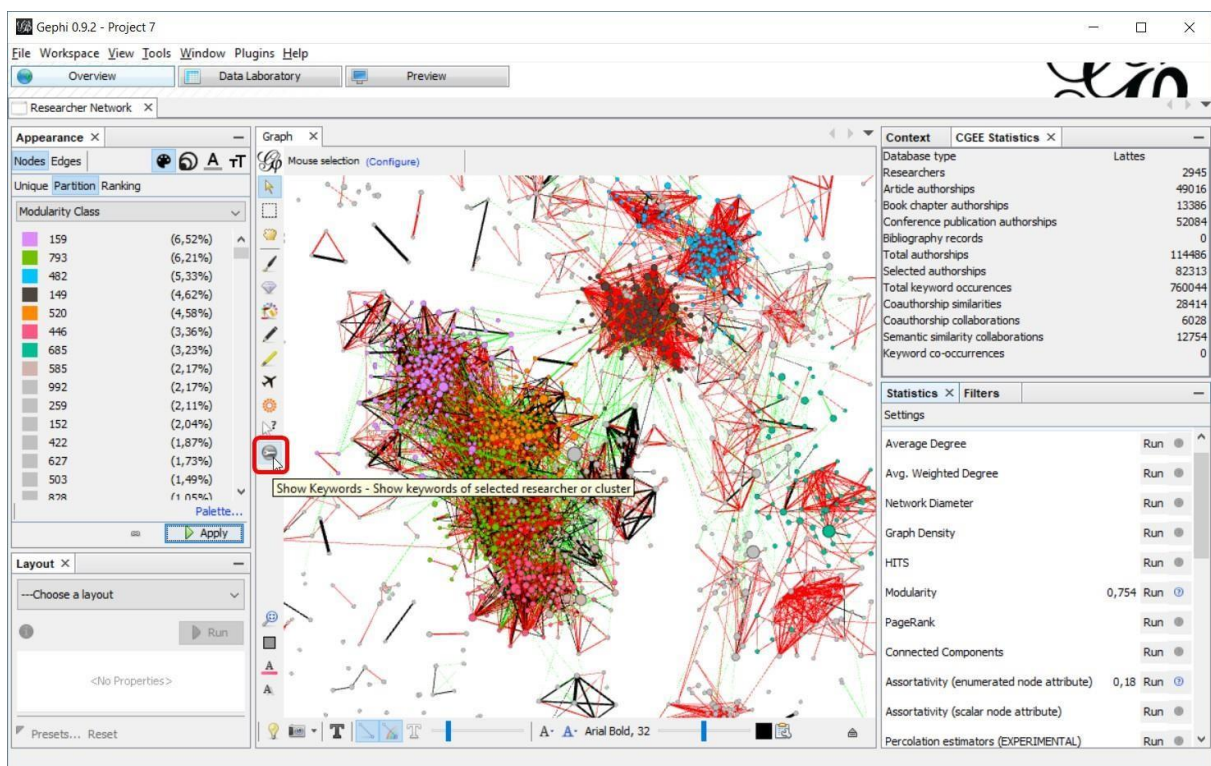
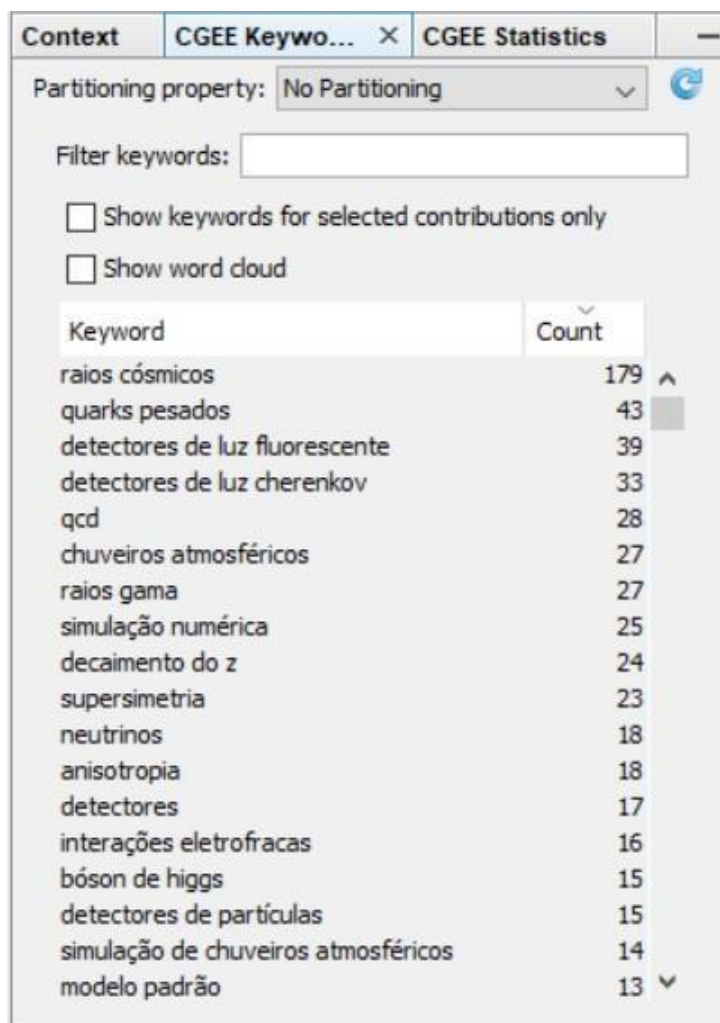


Figura 8.10: Seleção da funcionalidade “Palavras-chave”

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:



Keyword	Count
raios cósmicos	179
quarks pesados	43
detectores de luz fluorescente	39
detectores de luz cherenkov	33
qcd	28
chuveiros atmosféricos	27
raios gama	27
simulação numérica	25
decaimento do z	24
supersimetria	23
neutrinos	18
anisotropia	18
detectores	17
interações eletrofracas	16
bóson de higgs	15
detectores de partículas	15
simulação de chuveiros atmosféricos	14
modelo padrão	13

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:

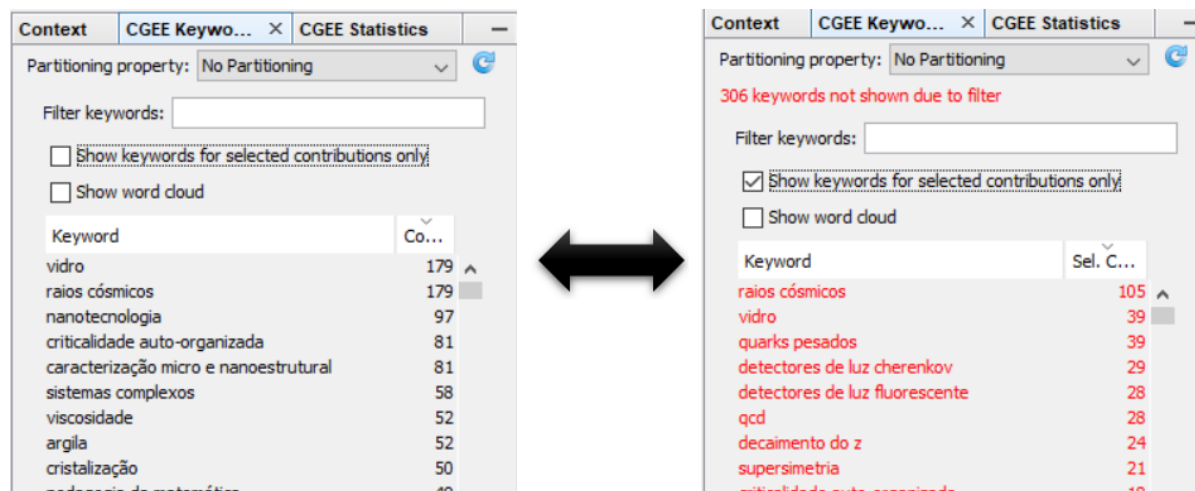


Figura 8.13: Filtragem das palavras-chave pelas contribuições selecionadas

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.

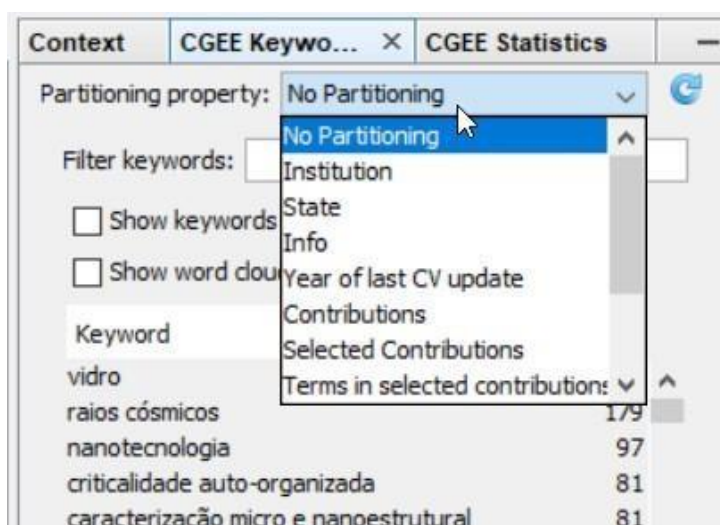


Figura 8.14: Seleção do atributo que define a partição dos nós

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. Dependendo da configuração (ver [Seção 3.4](#)), a lista mostra ainda a porcentagem ou a quantidade 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. Observe-se que no exemplo em baixo a lista foi configurada para mostrar a quantidade absoluta de nós (ver [Seção 3.4](#)):

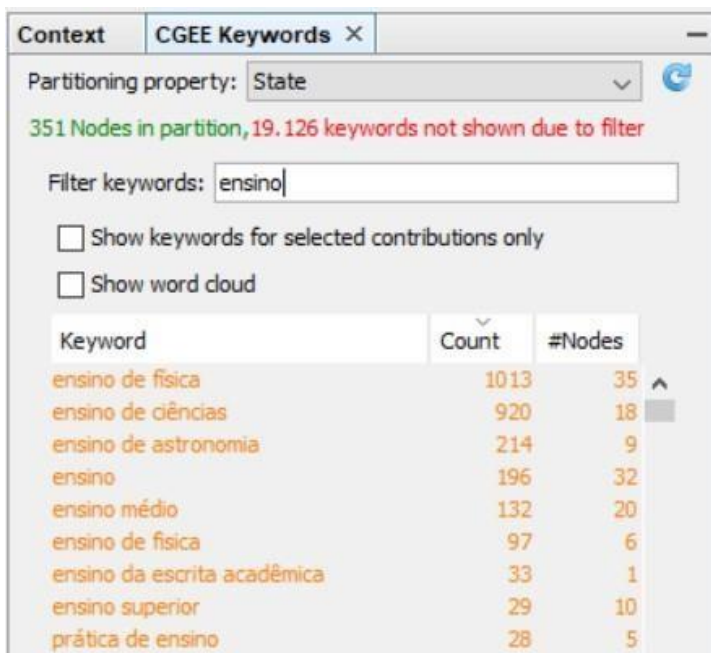


Figura 8.16: Lista de palavras-chave por partição, filtrada

Conforme descrito na Seção 3.4, 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:

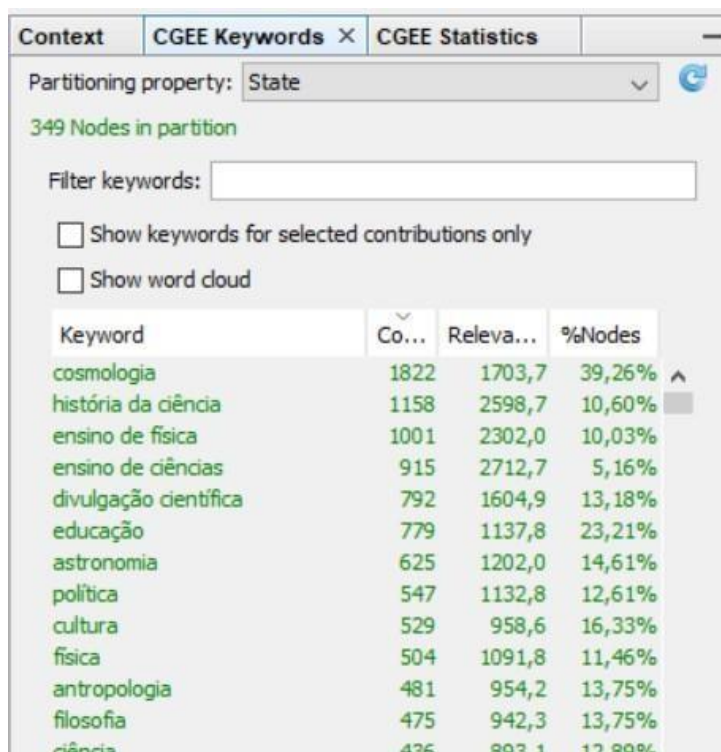


Figura 8.17: Relevância das palavras-chave exibidas na lista

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:

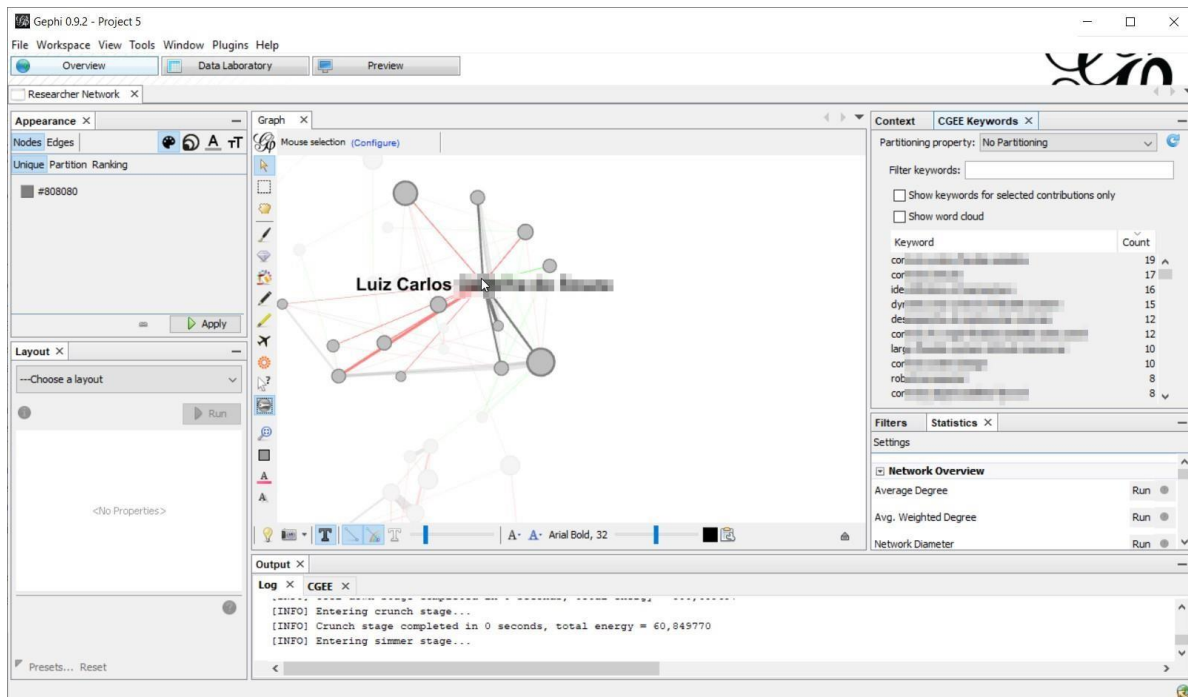


Figura 8.18: Exibição das palavras-chave do nó selecionado

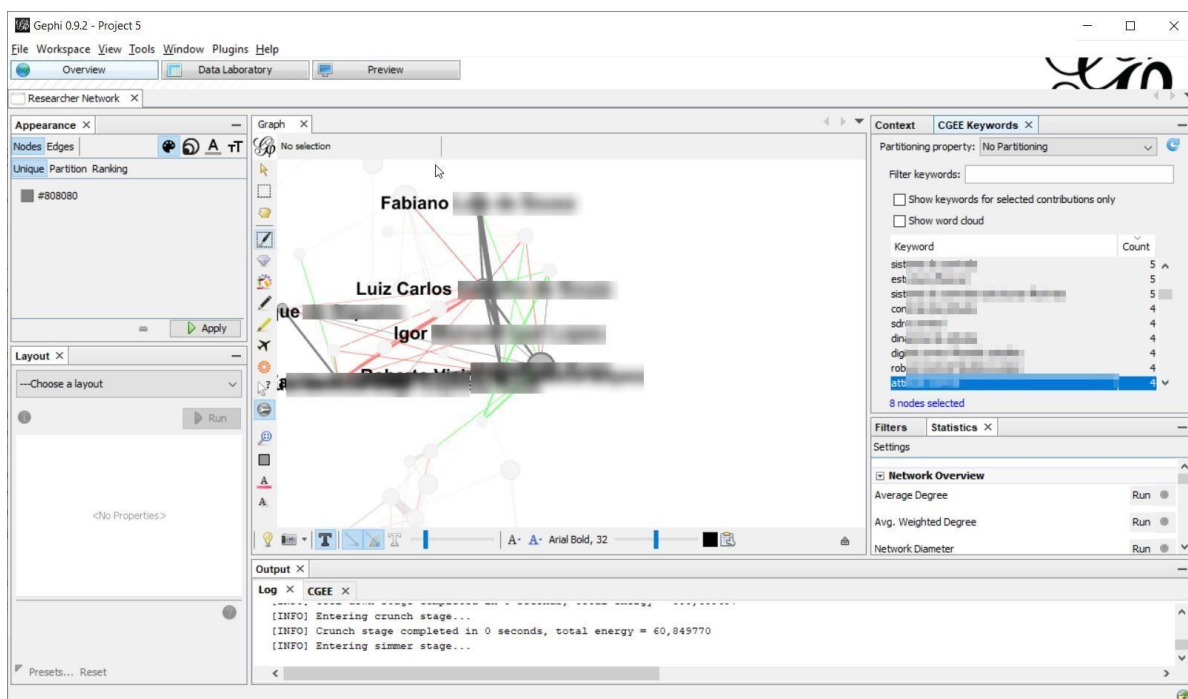


Figura 8.19: Exibição dos nós que usam a palavra-chave selecionada

A lista permite a seleção de uma única palavra-chave com um clique do botão esquerdo do mouse. Segurando 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 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:

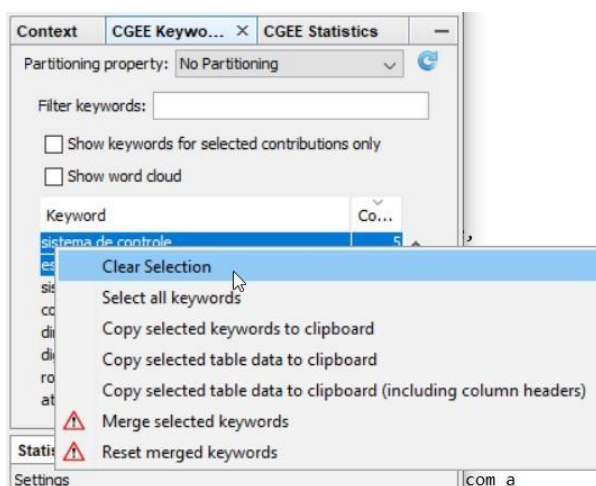


Figura 8.20: Funcionalidades adicionais na lista de palavras-chave

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:

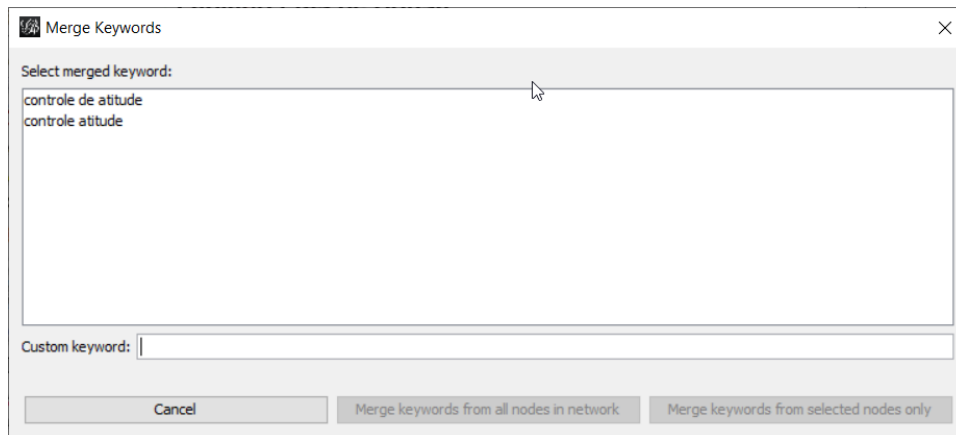


Figura 8.21: Diálogo de junção de palavras-chave

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

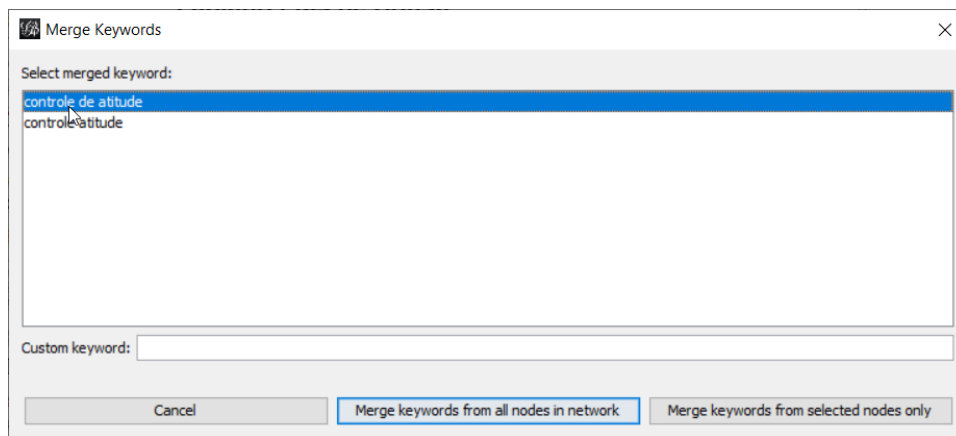


Figura 8.22: Seleção da palavra-chave juntada

Alternativamente, é possível digitar uma palavra-chave nova que substitui todas as palavras-chave selecionadas:

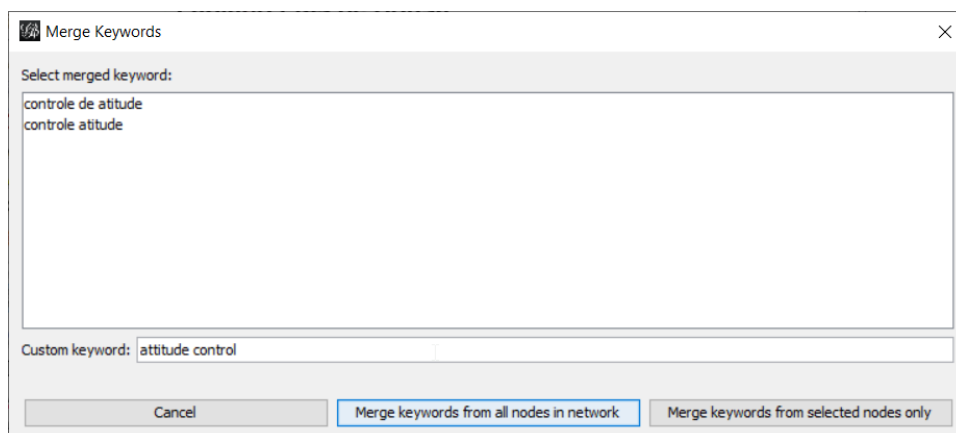


Figura 8.23: Especificação de uma palavra-chave nova

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:

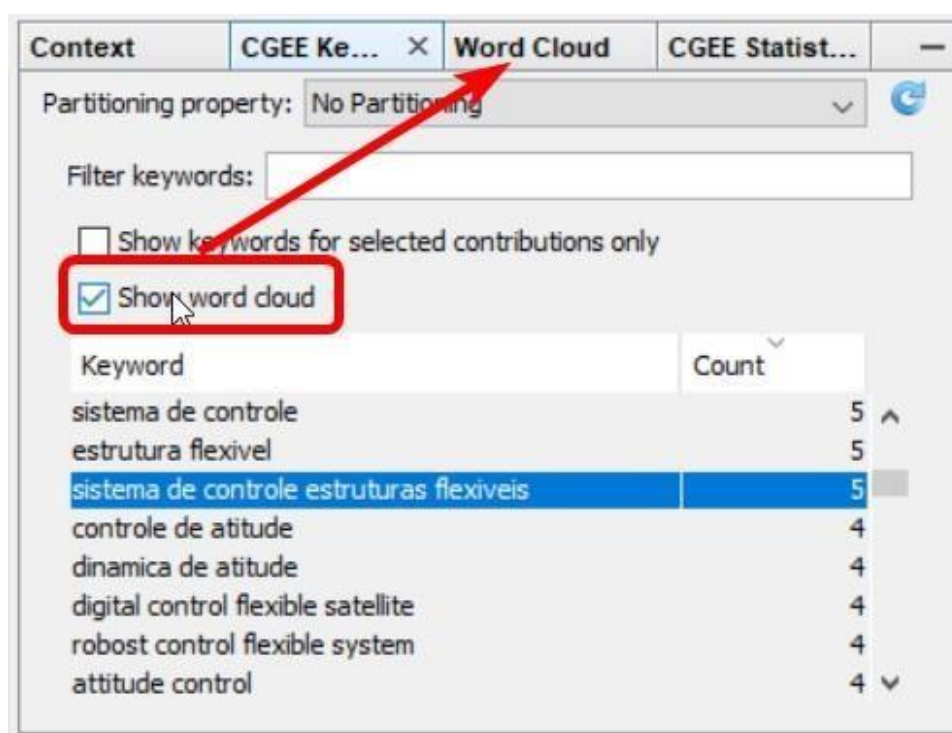


Figura 8.24: Habilitação da nuvem de palavras-chave

Clicando na aba “*Word Cloud*”, a nuvem de palavras é exibida, junto com eventuais avisos sobre a falta de precisão:

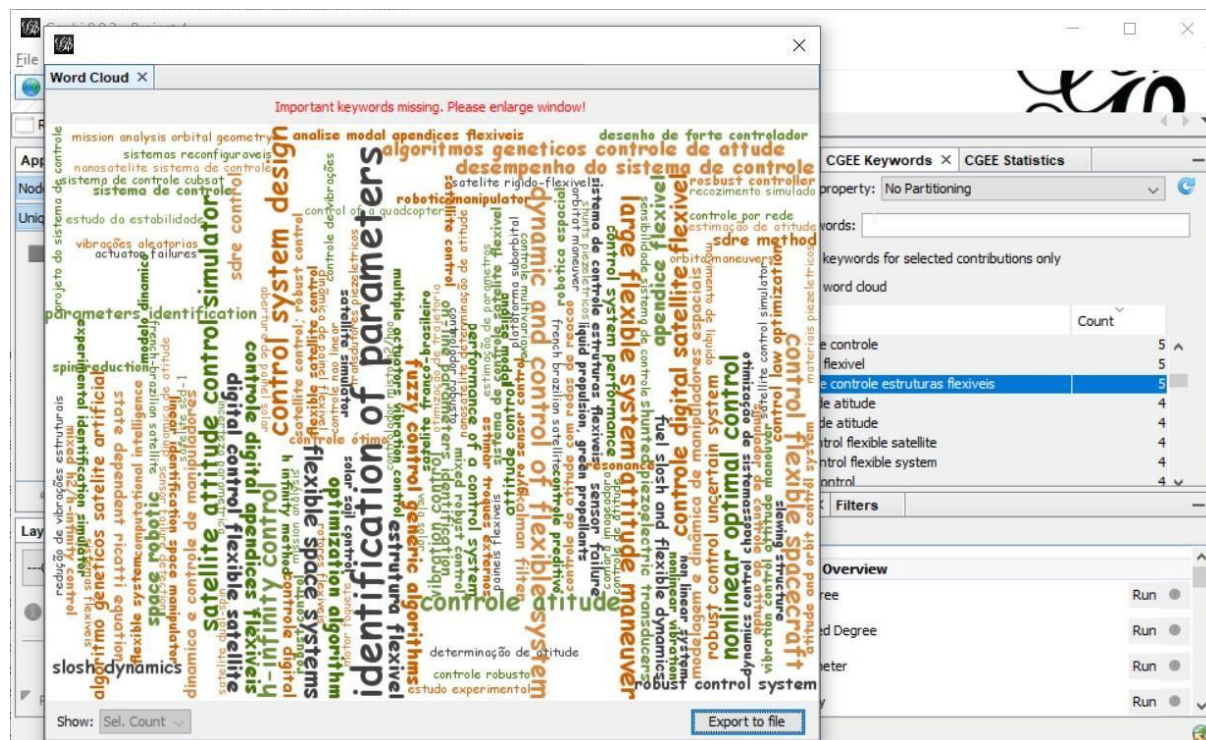


Figura 8.26: Separação da janela de nuvem de palavras

Se o espaço disponível na janela não permitir a exibição de, no mínimo, uma das 50 palavras-chave mais importantes de acordo com o critério selecionado, é exibida a mensagem “Important keywords missing. 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:

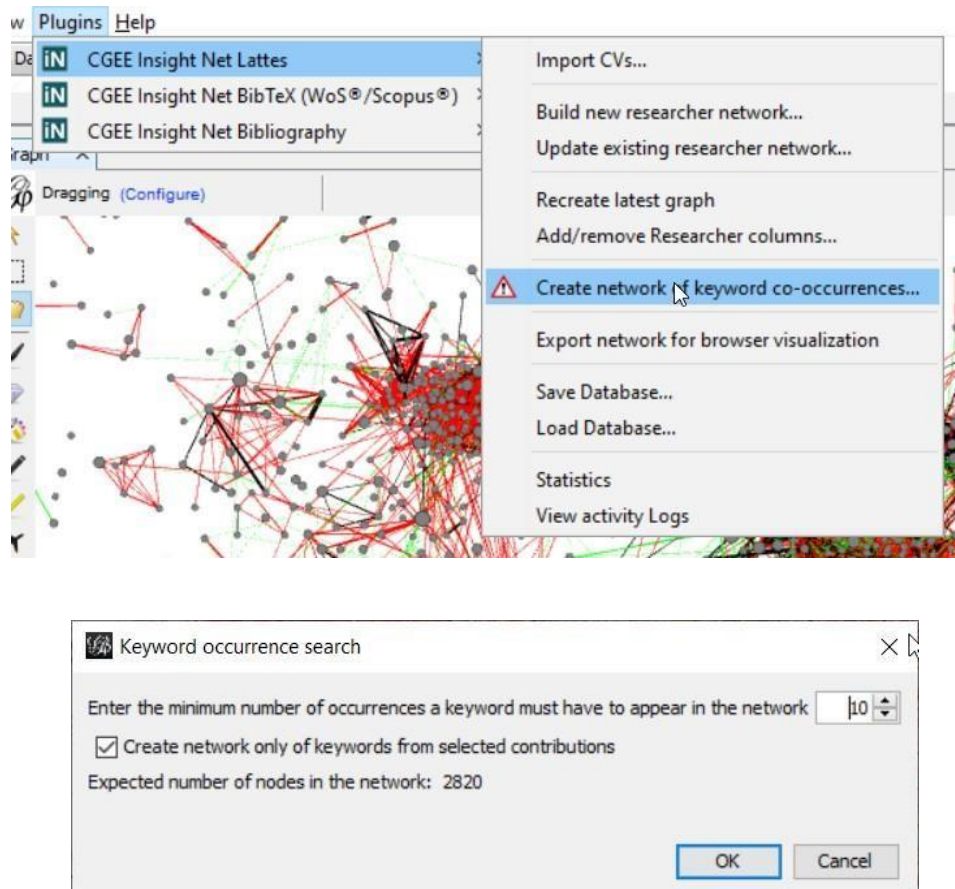


Figura 8.27: Funcionalidade para criar redes de palavras-chave

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 (clizando 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:

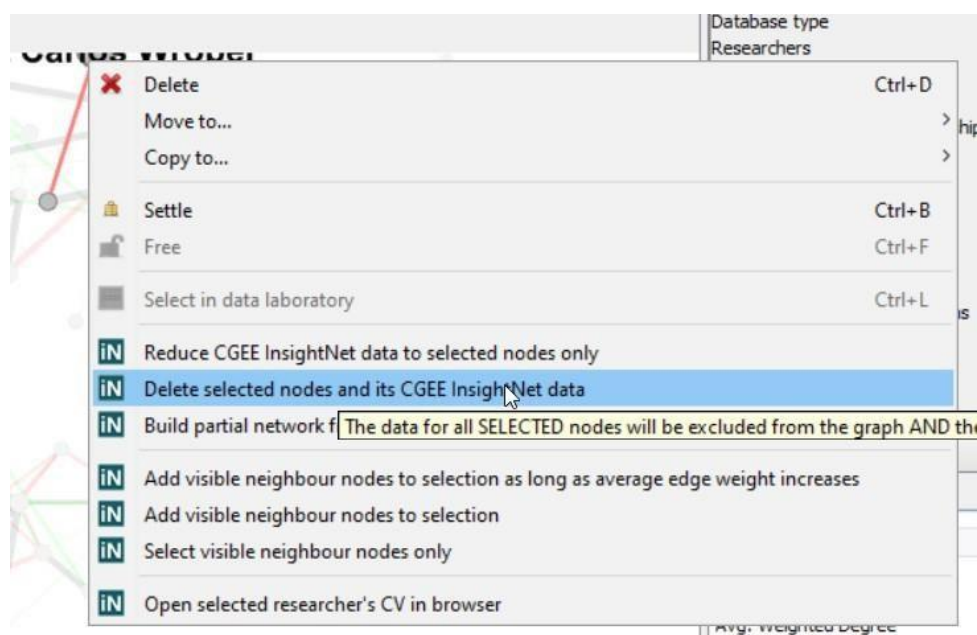


Figura 8.28: Eliminação interativa de nós do banco de dados na visualização do grafo

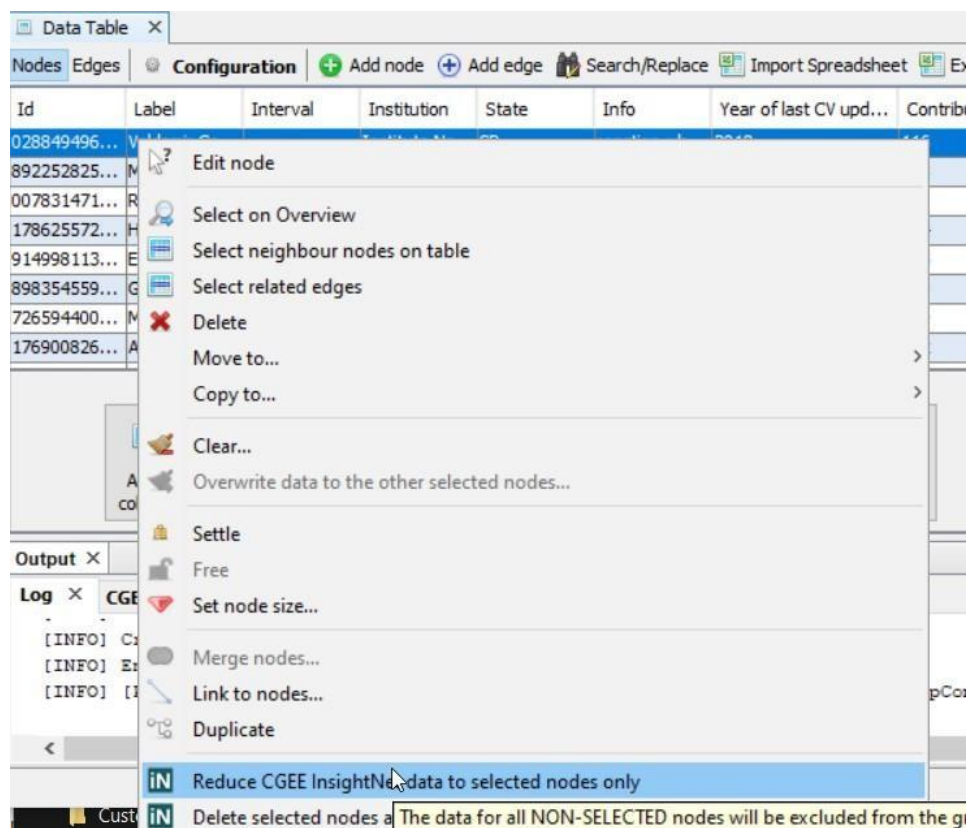


Figura 8.29: Eliminação interativa de nós do banco de dados no laboratório de dados

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.

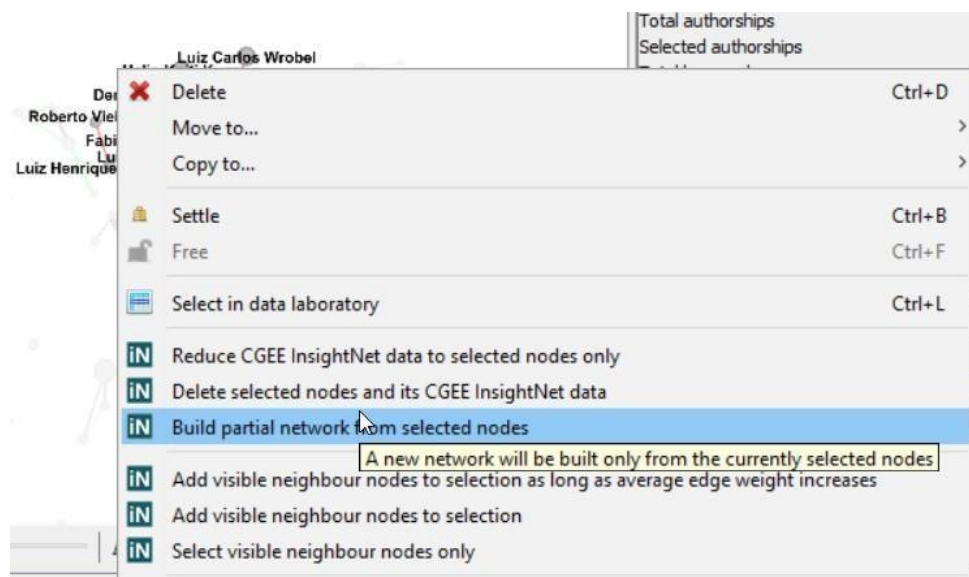


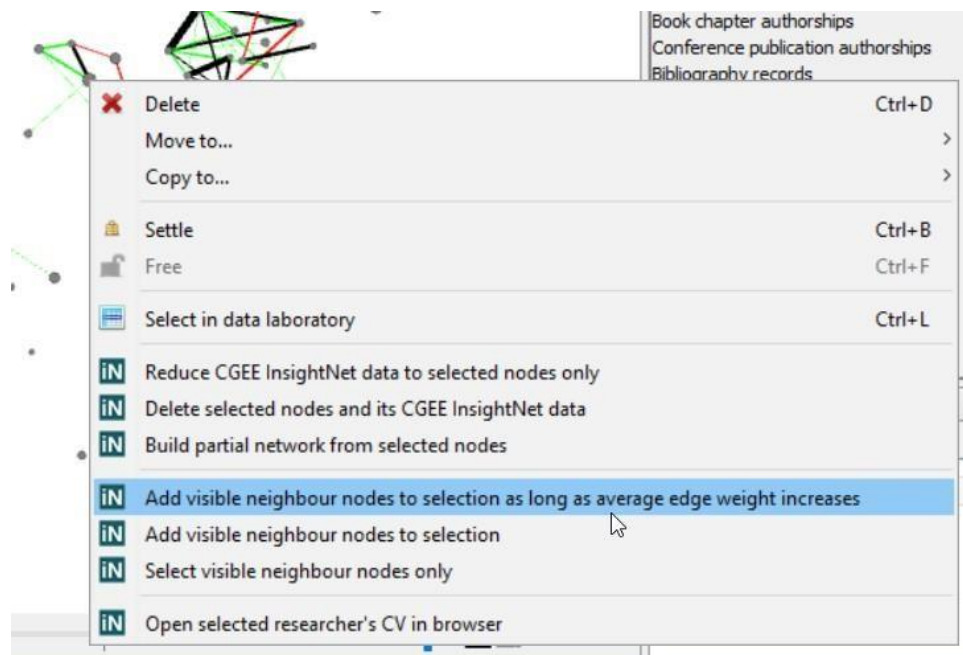
Figura 8.30: Criação de uma rede a partir do subconjunto de nós selecionados

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

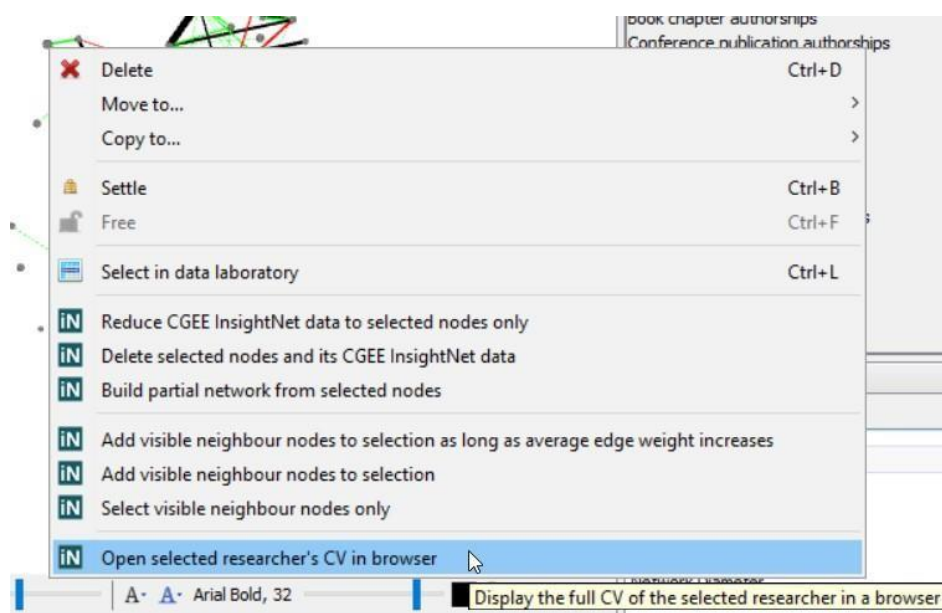
A partir de um ou mais nós selecionados, o *CGEE Insight Net* permite a seleção dos nós vizinhos com um clique do botão direito do mouse no nó dentro da visualização do grafo ou no laboratório de dados. A função “*Add visible neighbour nodes to selection*” acrescenta ao conjunto de nós selecionados todos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

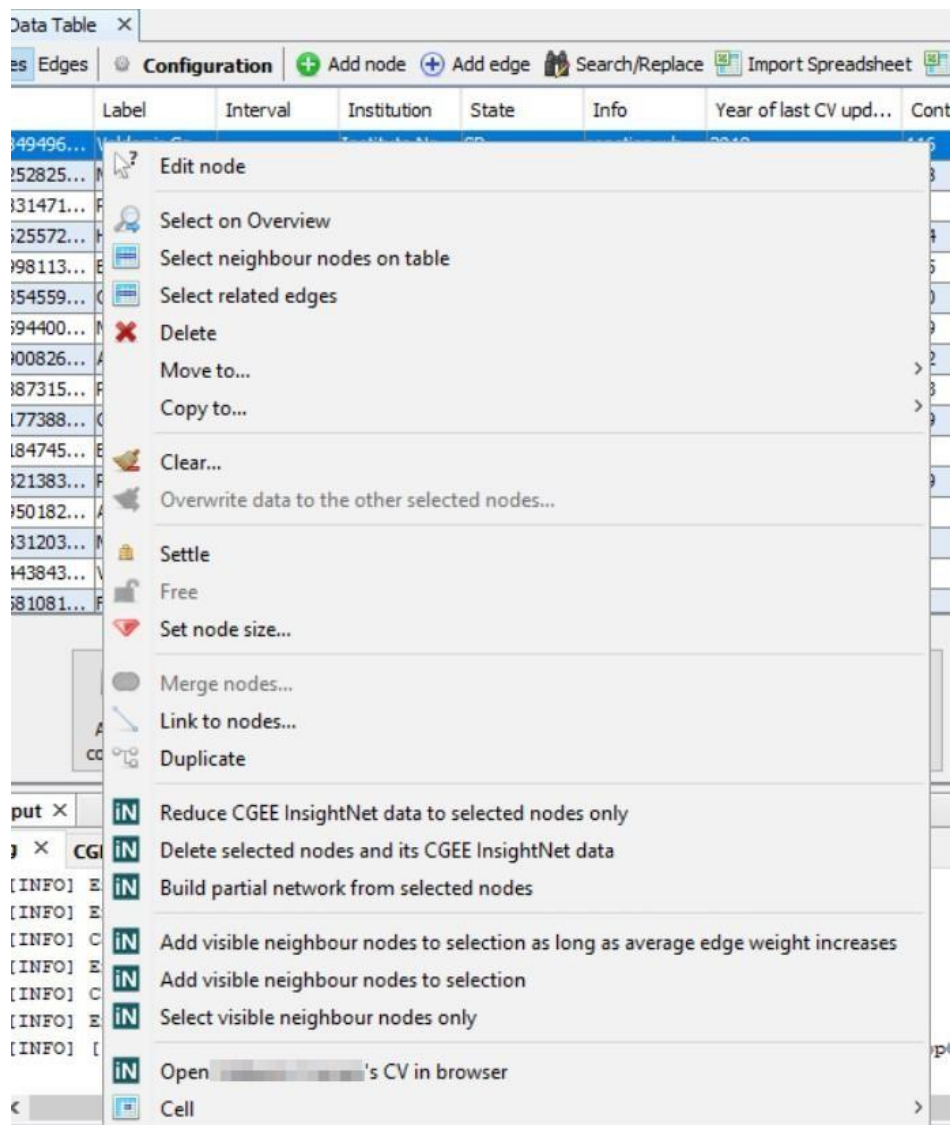


Já a funcionalidade “*Add visible neighbour nodes to selection as long as average edge weight increases*” repete o processo de adicionar vizinhos enquanto o peso média das arestas não diminua. Desta forma, o processo termina quando os nós adicionados não agregam mais informações relevantes ao subconjunto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





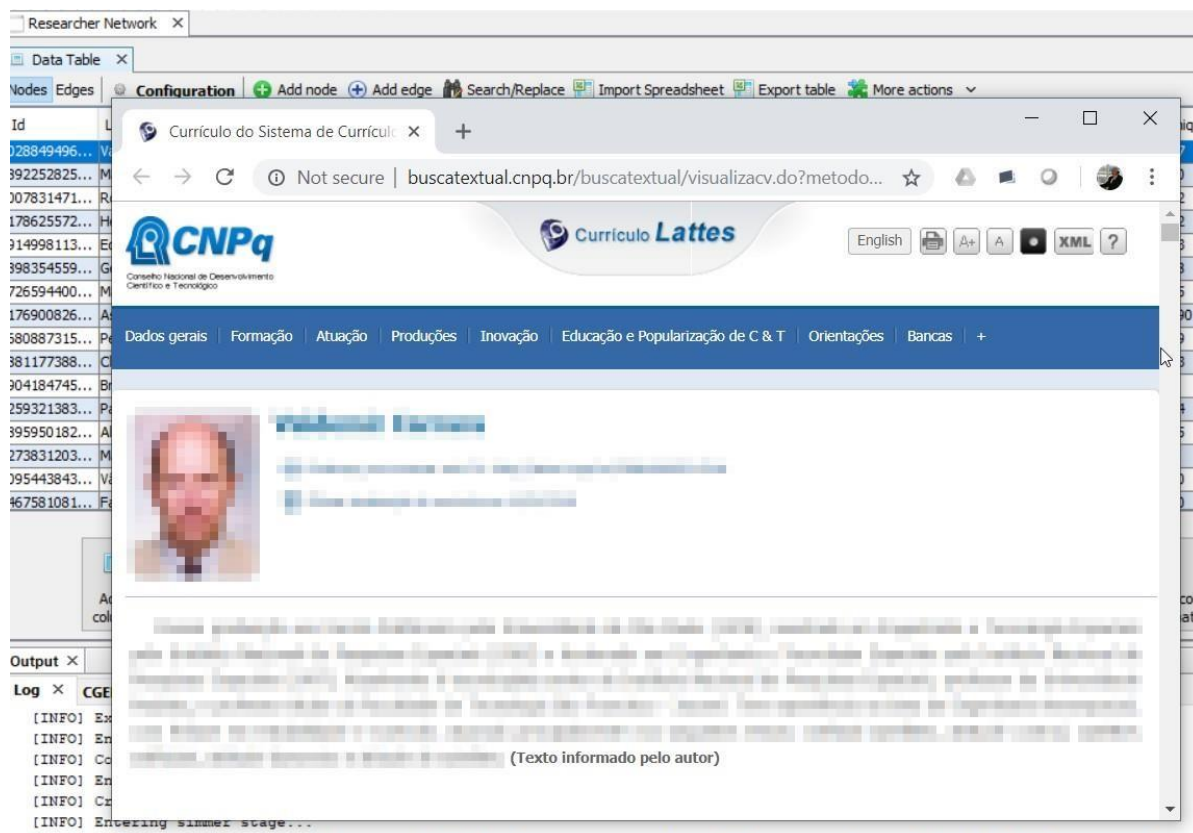
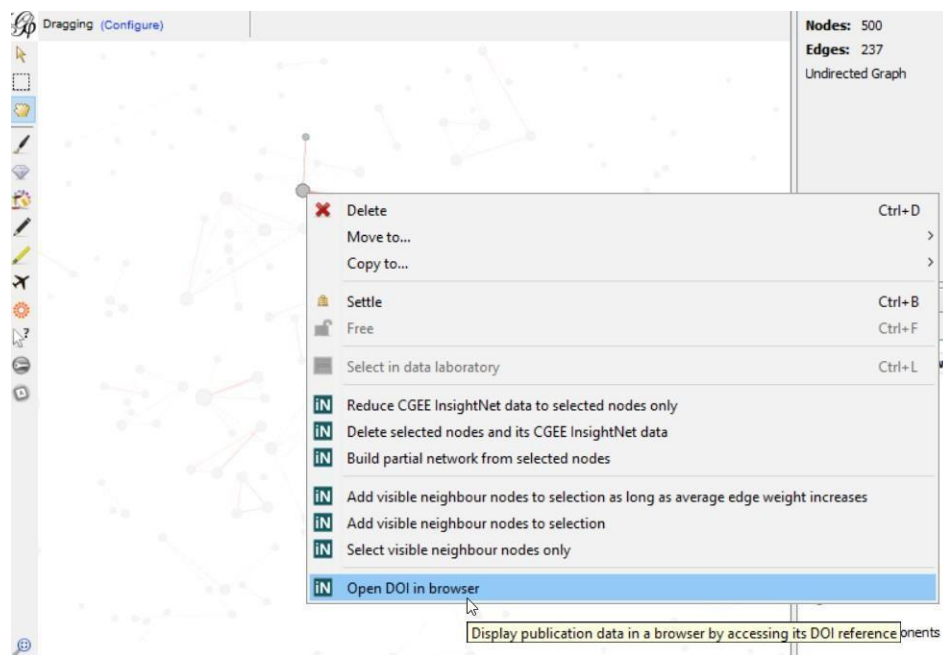
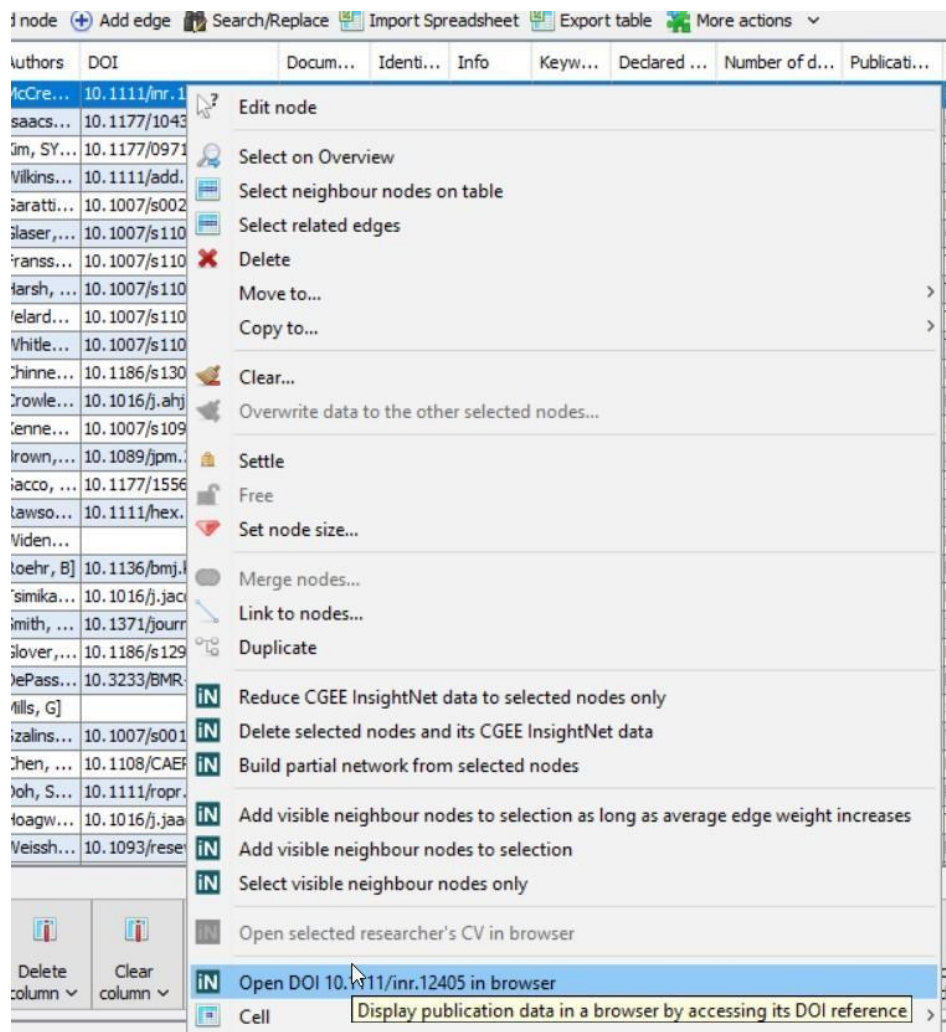


Figura 8.31: Visualização do currículo do pesquisador no Browser

8.10 Visualização interativa de contribuições bibliográficas por DOI no browser

Em redes de contribuições bibliográficas, o *Document Object Identifier* (DOI) permite acesso direto à referência bibliográfica e, dependendo da editora, também ao conteúdo. Um clique com botão direito do mouse na publicação no grafo ou no laboratório de dados mostrará o menu de *popup* com a opção “*Open DOI <doi> in browser*”. Clicando nessa opção, a página do DOI é exibido no browser do usuário.





The screenshot shows the Gephi 0.9.2 interface with a 'Bibliography network' data table. The table lists various bibliographic entries with columns for ID, Label, Abstract, Authors, DOI, Document Type, and Language. A browser window is overlaid on the right, displaying the Wiley Online Library page for the article 'Developing nursing research in the United Arab Emirates: a narrative review' by M. McCreaddie et al. The article page includes the journal title 'International Nursing Review / Volume 65, Issue 1', the publication date '11 October 2017', and the DOI 'https://doi.org/10.1111/inr.12405'. A 'PDF' button is visible on the article page.

ID	Label	Abstract	Abstract...	Authors	DOI	Docum...	Identi...	Info	Keyw...	Declared...
5553153	Develo...	AimThas...	English	[McCre...	10.1111/inr.12405	Review	WOS/0...		[capad...	English
5553167	Cultural...	Purpos...	English	[Isaacs...	10.1177/10436596177...	Review	WOS/0...		[end-of...	English
5553174	The Eth...	Merton...	English	[Kim, S...	10.1177/09717218177...	Article	WOS/0...			English
5553177	Addict...	Establis...	English	[Wilkins...	10.1111/add.14075	Article	WOS/0...		[alcoho...	English
5553194	Benefit...	Consid...	English	[Garatti...	10.1007/s00228-017-2...	Article	WOS/0...		[benef...	English
5553203	Changi...	With thi...	English	[Glaser...	10.1007/s11024-018-9...	Editorial ...	WOS/0...		[resear...	English
5553213	The Dr...	Over th...	English	[Frans...	10.1007/s11024-017-9...	Article	WOS/0...		[resear...	English
5553223	The Ris...	The em...	English	[Harsh...	10.1007/s11024-017-9...	Article	WOS/0...		[compu...	English
5553234	The Wa...	Althoug...	English	[Velard...	10.1007/s11024-018-9...	Article	WOS/0...		[resear...	English
5553240	The Im...	The pa...	English	[White...	10.1007/s11024-018-9...	Article	WOS/0...		[scienc...	English
5553249	Nationa...	Backgr...	English	[Chinne...	10.1186/s13063-018-2...	Article	WOS/0...		[burde...	English
5553261	How w...	Backgr...	English	[Crowle...	10.1016/j.ahj.2017.09...	Article	WOS/0...			English
5553273	Sub-na...	In the l...	English	[Kenne...	10.1007/s10961-017-9...	Article	WOS/0...		[stem c...	English
5553282	An Upd...	Backgr...	English	[Brown...	10.1089/jpm.2017.0287	Article	WOS/0...		[nh, p...	English
5553289	In Defe...	Nationa...	English	[Sacco...	10.1177/15562646177...	Article	WOS/0...		[quest...	English
5553298	Involvi...	Backgr...	English	[Rawso...	10.1111/hex.12604	Article	WOS/0...		[infect...	English
5553309	RESEAR...			[Widen...		News Item	WOS/0...			English
5553311	Mathild...			[Roehr, B]	10.1136/bmj.k403	Biographi...	WOS/0...			English
5553313	NHLBI ...	Pathop...	English	[Tsmika...	10.1016/j.jacc.2017.1...	Review	WOS/0...		[aortic ...	English
5553336	Inciden...	Backgr...	English	[Smith...	10.1371/journal.pone...	Review	WOS/0...			English
5553346	Estimat...	Backgr...	English	[Glover...	10.1186/s12961-017-0...	Article	WOS/0...		[medic...	English
5553362	Factors...	OBJEC...	English	[DePass...	10.3233/BMR-169628	Article	WOS/0...		[clinical...	English
5553371	Surrey ...			[Mills, G]		News Item	WOS/0...			English
5553373	Water ...	Poland i...	English	[Szalins...	10.1007/s00128-017-2...	Review	WOS/0...		[water ...	English
5553379	Review...	Purpos...	English	[Chen...	10.1108/CAER-07-201...	Review	WOS/0...		[bibliom...	English
5553387	Resear...	This stu...	English	[Doh, S...	10.1111/ropr.12261	Article	WOS/0...		[resear...	English
5553398	Trends ...			[Hoagw...	10.1016/j.jaac.2017.0...	Editorial ...	WOS/0...			English
5553407	Resear...	Resear...	English	[Weissh...	10.1093/reseval/rvx034	Review	WOS/0...		[social ...	English

Figura 8.32: Visualização da referência bibliográfica por DOI 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:

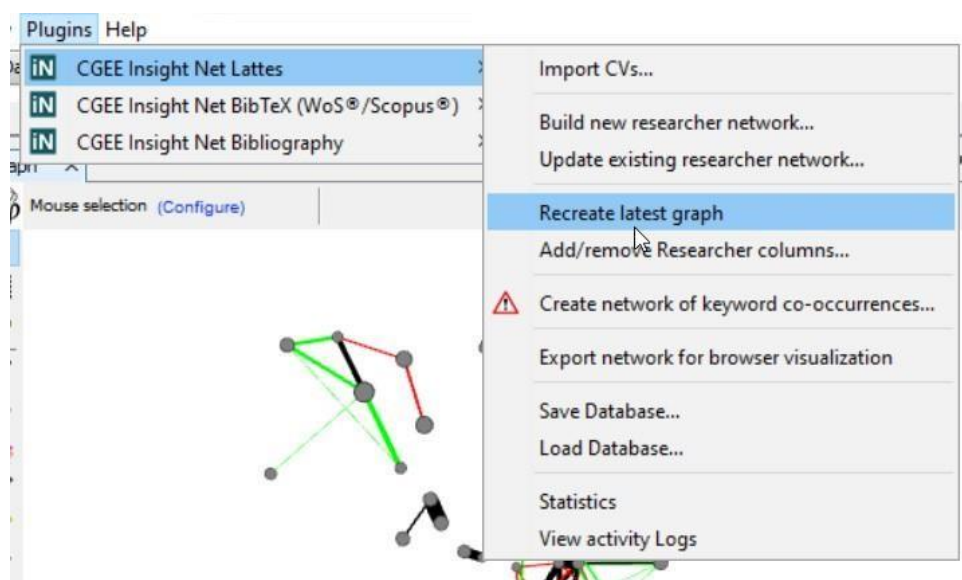


Figura 9.1: Recuperação do grafo

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:

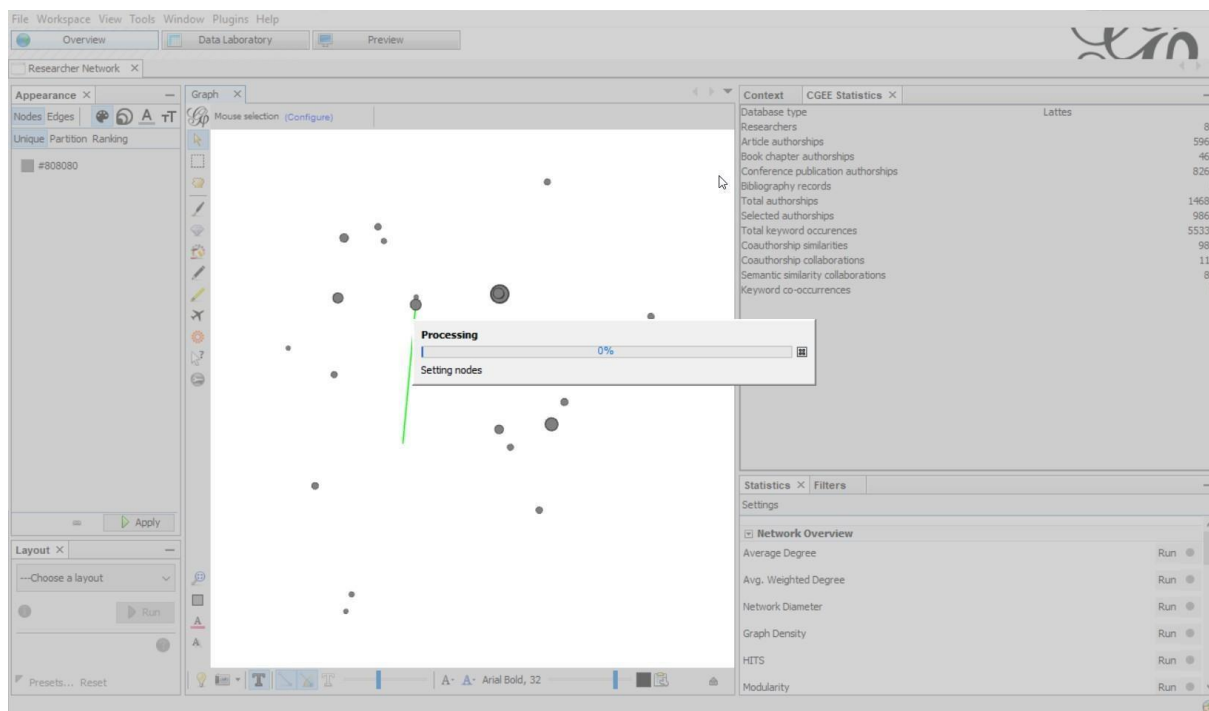


Figura 9.2: Recuperação do grafo

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* ...:

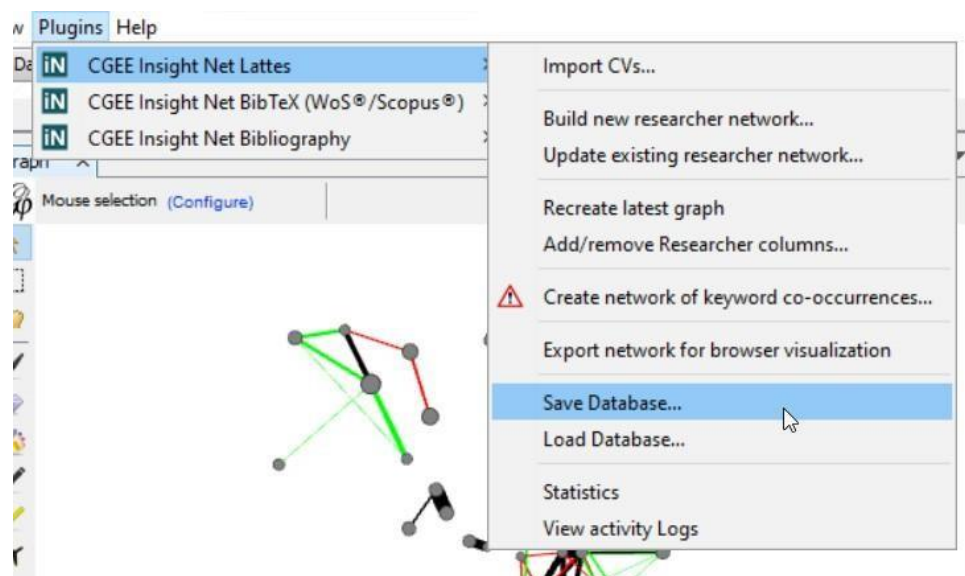


Figura 9.3: Funcionalidades de gravação e recuperação do banco de dados

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

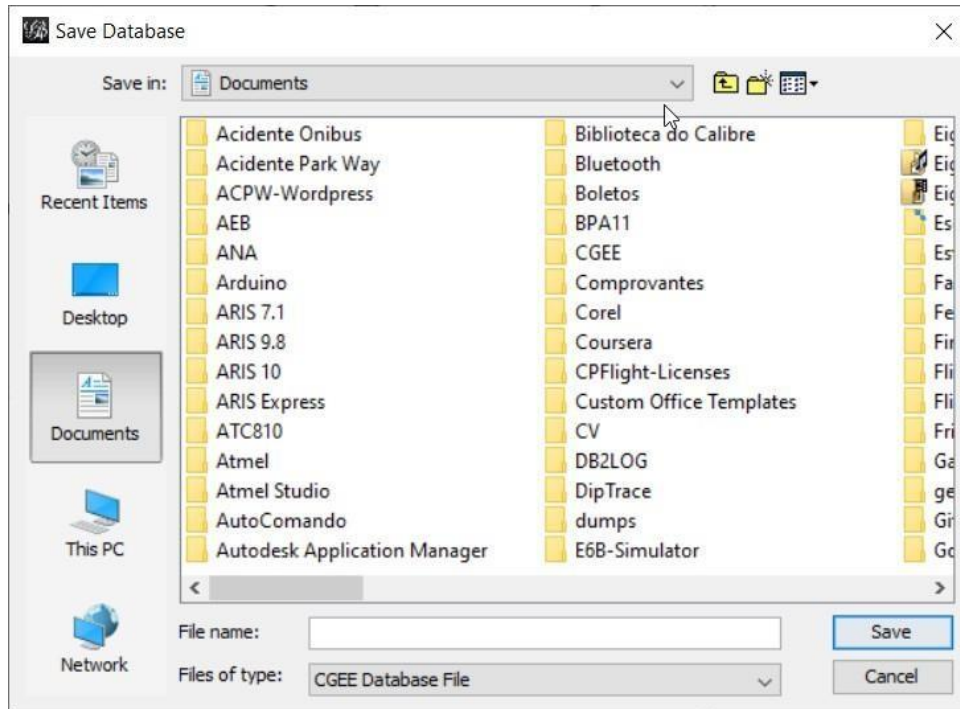


Figura 9.4: Especificação do arquivo de backup do banco de dados

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

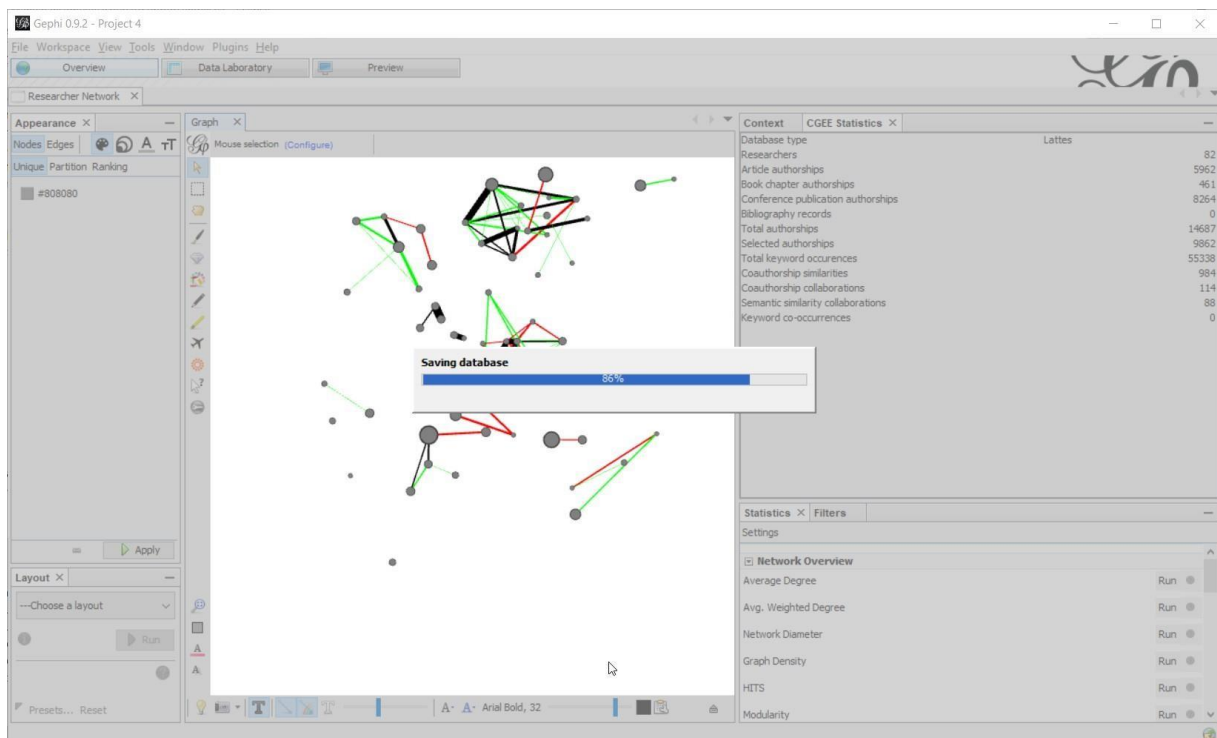
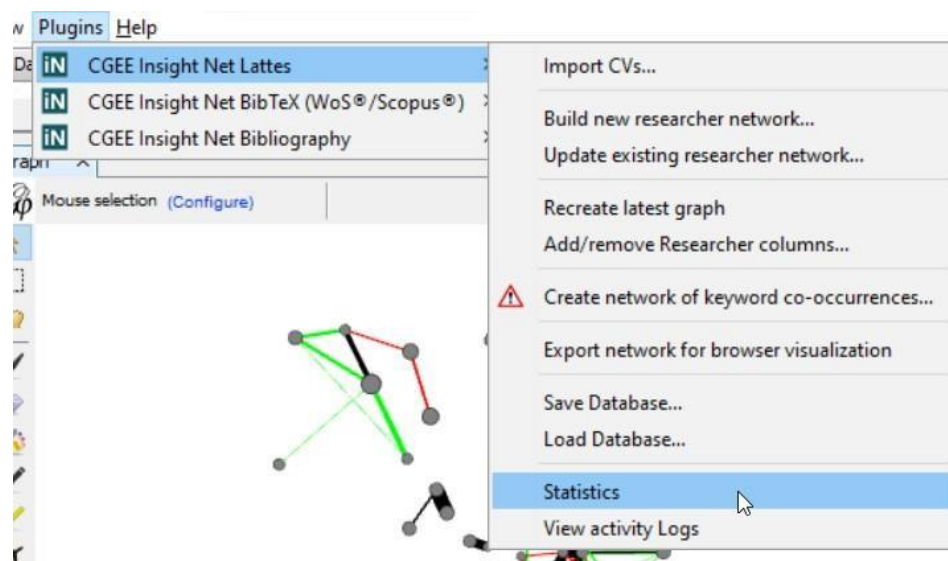


Figura 9.5: Geração do backup da base de dados

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:

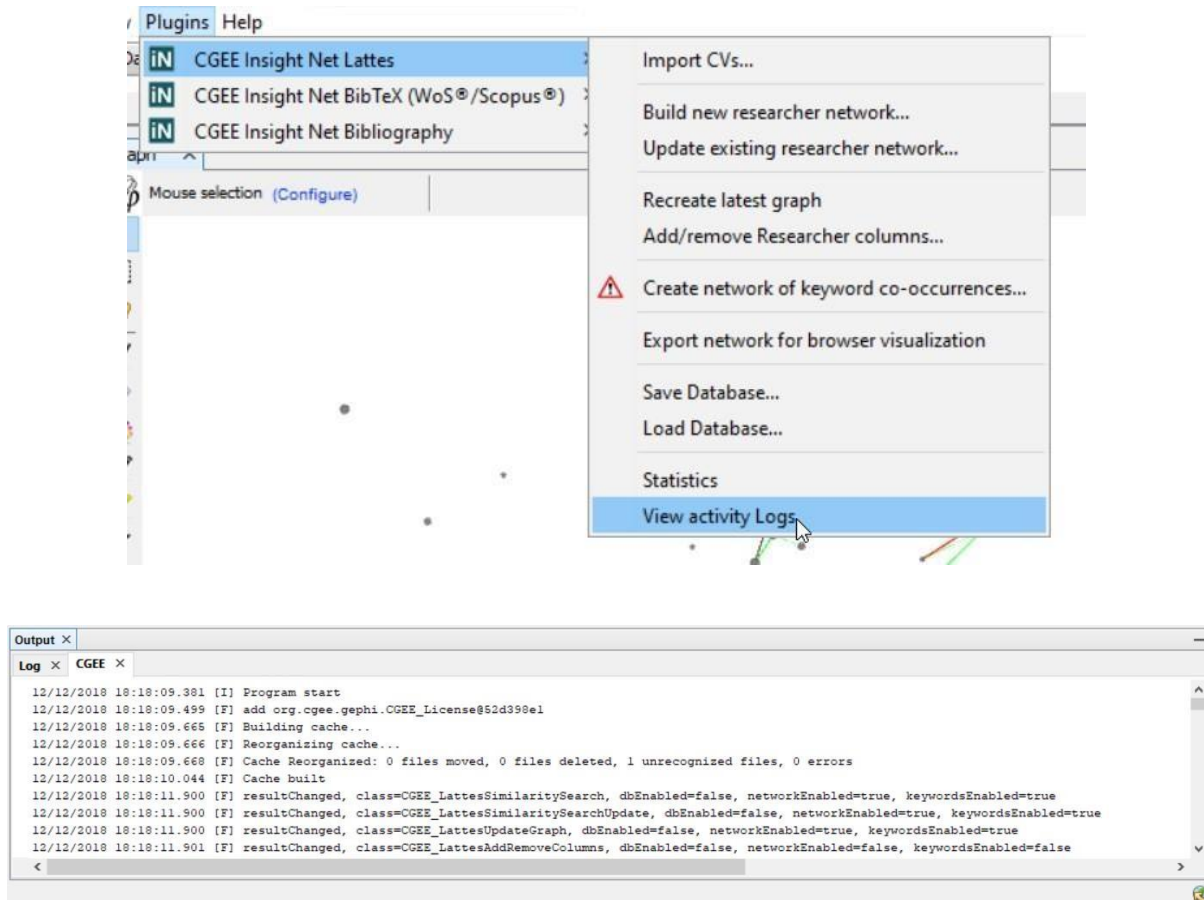


Figura 9.7: Protocolo de execução

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” exhibe 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

9.5 Envio de protocolo de execução

O *CGEE Insight Net* permite enviar dados sobre a execução do *Gephi* e do *plugin* ao CGEE para facilitar a análise de possíveis problemas. A opção “*Send debug information to CGEE*” está disponível em todos os menus de *plugin*:

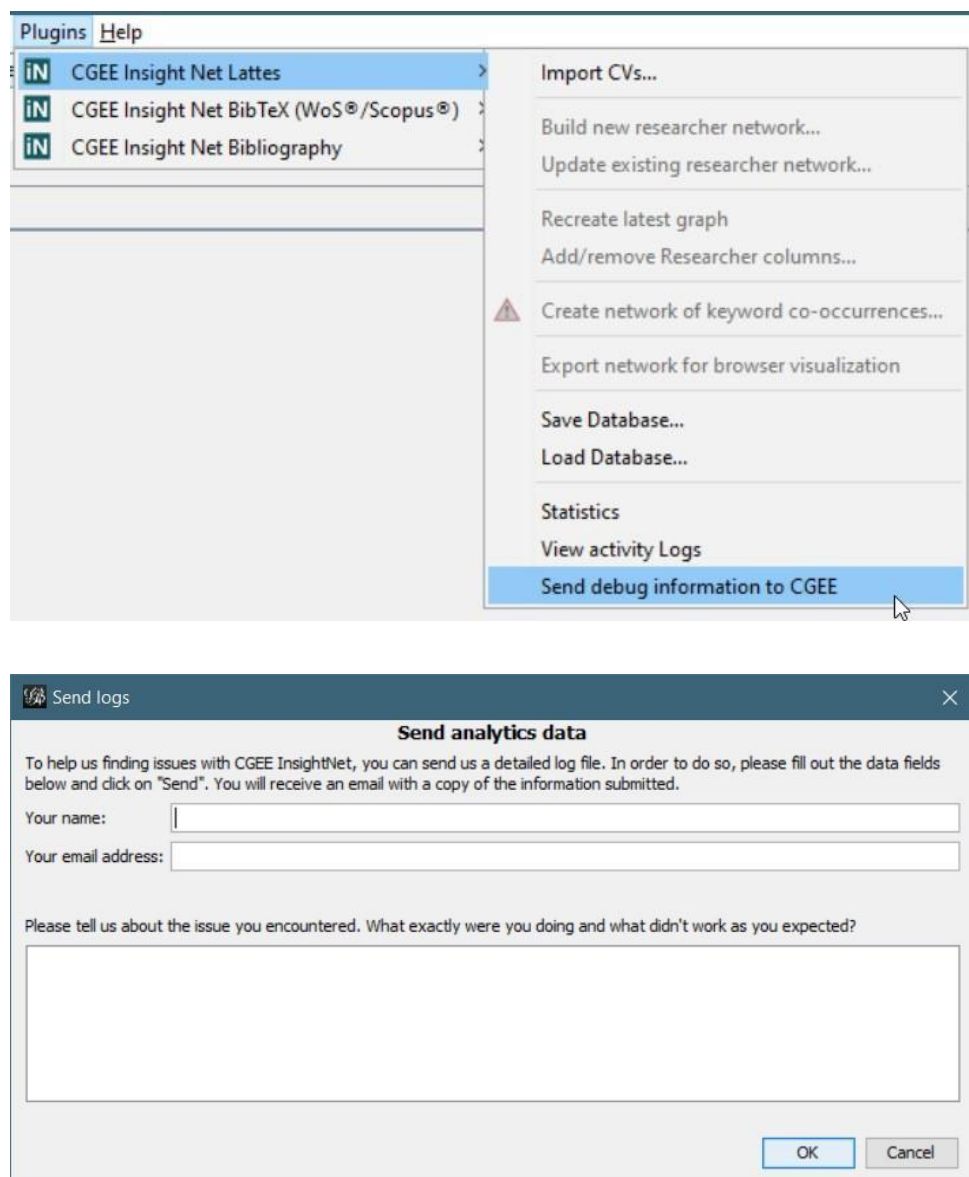


Figura 9.8: Enviar protocolos de execução

Caso a execução do *Gephi* for interrompida sem fechar o programa corretamente, o seguinte diálogo é exibido na próxima vez que o *Gephi* for iniciado, que oferece ao usuário enviar os protocolos da última execução:

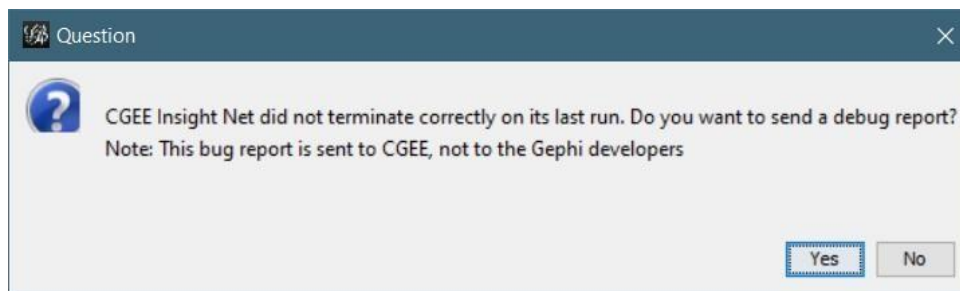


Figura 9.9: Diálogo que aparece depois da terminação forçada do *Gephi*

10.1 Especificação de formatos para o módulo de referências bibliográficas genéricas

O módulo de referências bibliográficas genéricas permite a importação de dados estruturados em diversos formatos a partir de arquivos descritivos desses formatos.

Os arquivos estão em formato JSON, em forma estruturada. Existem, três tipos de entidades nessa estrutura:

- *Descriptor* é a entidade raiz que descreve características globais dos arquivos a serem importadas. O *Descriptor* contém, entre outros, vários *Fields*.
- *Field* é a entidade que descreve um atributo do conjunto de dados a ser carregado, como título, ano de publicação ou palavras-chave. A definição de um *Field* pode incluir vários *Transformers*.
- Os dados de um *Field* podem ser tratados com um ou mais *Transformers*, que realizam operações genéricas.

Como exemplo, segue o formato de importação de dados genéricos de planilhas Excel. Os elementos específicos serão descritos nas seguintes seções:

```
{
  "name": "Generic",
  "description": "Generic [autodetect]",
  "lineFormat": "AUTO",
  "hasHeaderLine": "true",
  "addUnknownFieldsAsAttributes": "true",
  "createFields": "true",
  "fieldDefs": [
    {
      "name": "Identifier",
      "tag": "ID",
      "fieldType": "ID"
    }, {
```

(continues on next page)

(continuação da página anterior)

```

    "name": "DOI",
    "tag": "DOI",
    "transformers": [
      {
        "transformerType": "LOWERCASE"
      }
    ],
    "fieldType": "DOI"
  }, {
    "name": "Title",
    "tag": "Title",
    "fieldType": "TITLE"
  }, {
    "name": "Abstract",
    "tag": "Abstract",
    "fieldType": "ABSTRACT"
  }, {
    "name": "Authors",
    "tag": "Authors",
    "transformers": [
      {
        "transformerType": "SPLIT",
        "ifFormat": ["CSV", "TSV", "EXCEL"],
        "matchRegex": ";\\s+"
      }
    ],
    "fieldType": "AUTHORS"
  }, {
    "name": "Language",
    "tag": "Language",
    "fieldType": "LANGUAGE"
  }, {
    "name": "Keywords",
    "tag": "Keywords",
    "transformers": [
      {
        "transformerType": "SPLIT",
        "ifFormat": ["CSV", "TSV", "EXCEL"],
        "matchRegex": ";\\s+"
      }
    ],
    "fieldType": "KEYWORDS"
  }, {
    "name": "Year",
    "fieldType": "YEAR",
    "tag": "Year",
    "transformers": [
      {
        "transformerType": "NUMBER"
      }
    ]
  }, {
    "name": "Document type",
    "tag": "Document Type",
    "fieldType": "TYPE"
  }, {

```

(continues on next page)

(continuação da página anterior)

```

    "name": "Source Title",
    "tag": "Source title",
    "fieldType": "SOURCE"
  }
]
}

```

10.1.1 Descriptor

A entidade *descriptor* é a entidade raiz do arquivo de descrição de formato de dados e possui os seguintes campos:

Tabela 10.1: Campos do *Descriptor*

Campo	Significado
name	Nome do formato, utilizado para referências internas
description	Breve descrição do formato, utilização interna
lineFormat	Formato das linhas. Aceita os valores AUTO (identificação automática), CSV (separação por vírgula ou outro caractere definido pelo campo separationChar), TSV (separação por caractere TAB 0x09), TAGGED (formato RIS) ou EXCEL (planilha Excel®).
separationChar	Caso especificado, é o caractere que separa os campos no formato CSV. Se não for especificado, é utilizada a vírgula.
lineFormatType	Determina o tipo de arquivo para o formato TAGGED. Este formato possui variações para arquivos do “Web of Science” e do “Scopus”. Assim, o campo aceita os valores Scopus e Web of Science.
hasHeaderLine	Campo booleano (valores true ou false), que especifica se a primeira linha do arquivo contém os nomes das colunas. Relevante para arquivos nos formatos TSV, CSV e EXCEL.
columnList	Para formatos TSV, CSV e EXCEL, este campo permite especificar uma lista com nomes das colunas, caso o campo hasHeaderLine tiver o valor false.
addUnknownFieldsAsAttributes	Campo booleano que define que qualquer coluna de dados que possui um nome, mas que não foi definida explicitamente no <i>descriptor</i> será importada como atributo com o nome da coluna.
include	Lista de arquivos de <i>descriptors</i> que serão incluídos no processamento
fieldDefs	Lista de <i>Fields</i> , que descrevem os atributos da importação. Ver Seção 10.1.2
ignoreRisTags	Lista com nomes dos <i>Fields</i> que serão ignorados na importação de dados no formato TAGGED

10.1. Especificação de formatos para o módulo de referências bibliográficas

10.1.2 Field

A entidade *Field* descreve o conteúdo e o tratamento dos campos individuais de dados durante e depois da importação. A lista de *Field*s* consta no atributo ‘*fieldDefs*’ da entidade **descriptor*.

Tabela 10.2: Campos do *Field*

Campo	Significado
name	Nome do campo. Este nome será o nome do atributo do nó no Gephi, ou seja, o nome da coluna no laboratório de dados.
fieldType	<p>Tipo do campo de dados, definindo o tratamento no Gephi. Os seguintes tipos de campos são definidos:</p> <ul style="list-style-type: none"> • IGNORE: O campo não será importado no Gephi • ID: O campo será importado como identificador do nó no Gephi • DOI: Document Object Identifier, usado para desduplicar documentos • TITLE: Título da publicação, usado na busca de similaridade • ABSTRACT: Resumo da publicação, usado na busca de similaridade • AUTHORS: Lista de autores • KEYWORDS: Lista de palavras-chave da publicação • KEYWORDS_ADDITIONAL: Lista de palavras-chave adicionais da publicação • YEAR: Ano de publicação, usado no filtro de busca de similaridade • LANGUAGE: Idioma da publicação. Valores permitidos: “english”, “portuguese”, “inglês” e “português” • SOURCE: Nome da fonte da publicação (revista, livro etc.) • TYPE: Tipo de publicação (texto livre) • ATTRIBUTE: O campo será importado como atributo genérico, sem tratamento específico
tag	Este campo contém o nome do <i>Field</i> como ele aparece no cabeçalho dos arquivos CSV e TSV e é usado para identificar o <i>field</i> na lista de dados durante a importação de um desses formatos.
risTag	Este campo contém o nome do <i>Field</i> como ele aparece nos arquivos RIS (formato TAGGED). O valor é usado para identificar o <i>Field</i> durante a importação neste formato.
multipleItems	Campo booleano (valores true e false) que especifica que o conteúdo do <i>Field</i> é uma lista de valores.
transformers	Lista de <i>transformers</i> , que definem um tratamento do(s) valor(es) durante a importação dos dados. Ver Seção 10.1.3
insertTransformers	Este campo pode ter os valores BEFORE ou AFTER e define se na os <i>Transformers</i> deste campo serão executados antes ou depois de <i>Transformers</i> já existentes.

10.1.3 Transformer

A entidade *Transformer* define tratamentos dos dados durante a importação. Cada *Field* pode ter uma lista de *Transformers* em seu campo `transformers`.

Tabela 10.3: Campos do *Transformer*

Campo	Significado
<code>transformerType</code>	<p>Tipo do Transformer. Os seguintes tipos são definidos e descritos em seguida:</p> <ul style="list-style-type: none"> • ALWAYS: Aplicar expressão regular no valor do campo • CONDITIONAL: Aplicar expressão regular se o valor do campo confere com outra expressão regular • SPLIT: Dividir o valor de campo em uma lista, usando expressão regular como critério de divisão • SELECT: Selecionar um item específico da lista de valores • FILTER: Eliminar valores que não conferem com uma expressão regular • DEDUP: Eliminar valores duplicados da lista de valores • NUMBER: Transforma valor do campo em um valor numérico • LOWERCASE: Transformar texto do valor em letras minúsculas • FORMATINFO: <i>Tag</i> adicional usado em arquivos do formato TAGGED do Scopus®
<code>ifFormat</code>	<p>O <i>Transformer</i> é executado apenas se o formato do arquivo consta na lista de formatos especificado neste campo. Exemplo: <code>ifFormat: ["CSV", "TSV", "EXCEL"]</code></p>
<code>conditionRegex</code> <code>matchRegex</code> <code>substitutionRegex</code>	<p>Expressões regulares usados no transformador. O conteúdo depende do tipo de transformador e é descrito nas seções seguintes.</p>

Transformer ALWAYS

Este *transformer* substitui todas as ocorrências da expressão regular `matchRegex` no(s) valor(es) do campo pela expressão regular `substitutionRegex`. Grupos da primeira expressão regular são referenciados na segunda com o símbolo `$`. O seguinte exemplo demonstra parte do tratamento do país nos dados do Scopus:

```
[...]
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "^(.*) ([.])$",
      "substitutionRegex": "$1"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "(.*) ([Cc] [Hh] [Ii] [Nn] [Aa]) (.*)",
```

(continues on next page)

(continuação da página anterior)

```

      "substitutionRegex": "$2"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "(.*) ([Cc] [Aa] [Nn] [Aa] [Dd] [Aa]) (.*)",
      "substitutionRegex": "$2"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "(.*?) (\\s+) (.+) (\\s+) (.+) (\\s+) (.*)",
      "substitutionRegex": "$7"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "^ [A-Za-z] [A-Za-z] \\s [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] $",
      "substitutionRegex": "United States"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "^ [A-Za-z] [A-Za-z] \\s [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] [0-9] $",
      "substitutionRegex": "United States"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "^ (^0-9)* [0-9] [0-9] [0-9] [0-9] [0-9] $",
      "substitutionRegex": "United States"
    }, {
      "transformerType": "ALWAYS",
      "matchRegex": "^ [A-Za-z] [A-Za-z] $",
      "substitutionRegex": "United States"
    }, {
    [...]
  
```

Transformer CONDITIONAL

Este *transformer* funciona análogo ao *transformer* ALWAYS, mas permite condicionar a substituição. Ele é executado apenas se o valor de entrada confere com a expressão regular `conditionRegex`. Se este campo não foi definido no *transformer*, ele se comporta igual ao *transformer* ALWAYS.

Transformer SPLIT

Este *transformer* quebra um único valor em uma lista (um vetor) de valores. Como critério de divisão, a expressão regular `matchRegex` é utilizado. Segue, como exemplo, o tratamento de autores nos dados do Scopus®, apenas para dados nos formatos CSV, “TSV” e EXCEL:

```

[... ]
  }, {
    "name": "Authors",
    "fieldType": "AUTHORS",
    "tag": "Authors",
    "risTag": "AU",
    "transformers": [
      {
        "transformerType": "SPLIT",
        "ifFormat": ["CSV", "TSV", "EXCEL"],
      }
    ]
  }

```

(continues on next page)

(continuação da página anterior)

```

                "matchRegex": ",\\s+"
            }
        ],{
    },{
[...]
```

Transformer SELECT

Este *transformer* reduz uma lista (um vetor) de valores para um único elemento (escalar), definido pela posição na lista. A posição do elemento selecionado é definido como valor numérico no campo `matchRegex`. O valor 0 seleciona o primeiro elemento da lista, o valor 1 o segundo etc. Valores negativos selecionam elementos a partir do fim da lista: -1 é o último elemento da lista, -2 o penúltimo etc:

```

[...]
```

```

}, {
  "name": "Organizations",
  "fieldType": "ATTRIBUTE",
  "tag": "C1",
  "multipleItems": "true",
  "transformers": [
    {
      "transformerType": "ALWAYS",
      "matchRegex": "\\[[.+?\\]\\]\\s+",
      "substitutionRegex": ""
    }, {
      "transformerType": "SPLIT",
      "ifFormat": ["CSV", "TSV", "EXCEL"],
      "matchRegex": ";\\s+"
    }, {
      "transformerType": "SPLIT",
      "matchRegex": ",\\s*"
    }, {
      "transformerType": "SELECT",
      "matchRegex": "0"
    }, {
      "transformerType": "DEDUP"
    }
  ]
}, {
[...]
```

Transformer FILTER

O *transformer* FILTER elimina todos os valores de uma lista de valores que não conferem com a expressão regular `matchPattern`:

```

[...]
```

```

}, {
  "name": "ISSN",
  "fieldType": "ATTRIBUTE",
  "tag": "ISSN",

```

(continues on next page)

(continuação da página anterior)

```

"risTag": "SN",
"transformers": [
  {
    "transformerType": "SPLIT",
    "matchRegex": ";\\s+"
  }, {
    "transformerType": "FILTER",
    "ifFormat": [ "TAGGED" ],
    "matchRegex": "(.*)\\s+\\(ISSN\\)"
  }, {
    "transformerType": "CONDITIONAL",
    "matchRegex": "(.*)\\s+\\(ISSN\\)",
    "substitutionRegex": "$1"
  }
]
}, {
[...]
```

Transformer DEDUP

O *transformer* DEDUP elimina os valores duplicados de uma lista de valores:

```

[...]
```

```

}, {
  "name": "Organizations",
  "fieldType": "ATTRIBUTE",
  "tag": "C1",
  "multipleItems": "true",
  "transformers": [
    {
      "transformerType": "ALWAYS",
      "matchRegex": "\\[[.+?\\]\\]\\s+",
      "substitutionRegex": ""
    }, {
      "transformerType": "SPLIT",
      "ifFormat": [ "CSV", "TSV", "EXCEL" ],
      "matchRegex": ";\\s+"
    }, {
      "transformerType": "SPLIT",
      "matchRegex": ",\\s*"
    }, {
      "transformerType": "SELECT",
      "matchRegex": "0"
    }, {
      "transformerType": "DEDUP"
    }
  ]
}, {
[...]
```

Transformer NUMBER

O *transformer* NUMBER interpreta o conteúdo do *field* como valor numérico:

```
[...]
}, {
  "name": "Year",
  "fieldType": "YEAR",
  "tag": "PY",
  "transformers": [
    {
      "transformerType": "NUMBER"
    }
  ]
}, {
[...]
```

Transformer LOWERCASE

O *transformer* LOWERCASE substitui todas as letras maiúsculas no valor do *field* por letras minúsculas:

```
[...]
}, {
  "name": "DOI",
  "fieldType": "DOI",
  "tag": "DOI",
  "risTag": "DO",
  "transformers": [
    {
      "transformerType": "LOWERCASE"
    }
  ]
}, {
[...]
```

Transformer FORMATINFO

Este *transformer* é utilizado apenas em dados do formato TAGGED com `lineFormatType SCOPUS`. Neste formato, existem campos de dados que utilizam o mesmo *tag* N1. O valor do campo inicia com um texto específico que determina de que campo efetivamente se trata:

```
N1 - Conference code: 121263
N1 - Export Date: 15 March 2017
N1 - Funding details: MOE, Ministry of Education
N1 - Funding details: NRF-2013R1A1A2011601, MOE, Ministry of Education
N1 - Funding details: NRF, National Research Foundation of Korea
N1 - Funding text: This research was supported by Basic Science Research_
->Program through
the National Research Foundation of Korea (NRF) funded by the Ministry of_
->Education (NRF-2013R1A1A2011601)
and the Human Resource Training Program for Regional Innovation and_
->Creativity through the Ministry of
Education and National Research Foundation of Korea (NRF-2015H1C1A1035548).
```

Para poder distinguir esses casos, o *transformer* FORMATINFO permite especificar o texto inicial que determina o tipo de campo:

```
[...]
},{
  "name": "CODEN",
  "fieldType": "ATTRIBUTE",
  "tag": "CODEN",
  "risTag": "N1",
  "transformers": [
    {
      "transformerType": "FORMATINFO",
      "matchRegex": "CODEN:"
    }
  ]
},{
  "name": "Export date",
  "fieldType": "ATTRIBUTE",
  "risTag": "N1",
  "transformers": [
    {
      "transformerType": "FORMATINFO",
      "matchRegex": "Export Date:"
    }
  ]
},{
  "name": "Funding Details",
  "fieldType": "ATTRIBUTE",
  "tag": "Funding Details",
  "risTag": "N1",
  "transformers": [
    {
      "transformerType": "FORMATINFO",
      "matchRegex": "Funding details:"
    },
    {
      "transformerType": "SPLIT",
      "ifFormat": ["CSV", "TSV", "EXCEL"],
      "matchRegex": ";\s+"
    }
  ]
},{
  "name": "Funding Text",
  "fieldType": "ATTRIBUTE",
  "tag": "Funding Text",
  "risTag": "N1",
  "transformers": [
    {
      "transformerType": "FORMATINFO",
      "matchRegex": "Funding text:"
    }
  ]
},{
  "name": "Tradenames",
  "fieldType": "ATTRIBUTE",
  "tag": "Tradenames",
  "risTag": "N1",
  "transformers": [
    {
      "transformerType": "FORMATINFO",
      "matchRegex": "Tradenames:"
    }
  ]
}
```

(continues on next page)

(continuação da página anterior)

```
    }  
  ]  
}, {  
[...]
```

10.2 Compilação do *CGEE Insight Net*

O *CGEE Insight Net* é composto de dois componentes básicos:

- A biblioteca `InsightNetLibrary`, que permite executar as operações a partir da linha de comando (veja [Seção 10.3](#))
- O *plugin* `CGEE Insight Net 3.0`, que disponibiliza a interface de usuário para realizar as operações da linha de comando no *Gephi* e acrescenta componentes de filtros, estatísticas e análise de palavras-chave.

Os dois componentes são realizadas como módulos *Java* e utilizam a ferramenta *Maven* para facilitar o processo de compilação.

O código fonte para ambos os módulos consta no repositório *SVN* do CGEE, junto com arquivos de configuração da plataforma de desenvolvimento *Netbeans*, que permitem abrir os projetos diretamente neste ambiente de desenvolvimento.

10.2.1 Biblioteca `InsightNetLibrary`

A biblioteca `InsightNetLibrary` possui um arquivo `pom.xml` que descreve todo o processo de compilação. O comando `mvn install` não requer configurações adicionais e gera o arquivo `InsightNetLibrary-<versão>.jar`.

Deve ser observado que o número de versão da biblioteca `InsightNetLibrary` deve ser o mesmo do *plugin* `CGEE Insight Net 3.0`.

10.2.2 *Plugin* `CGEE Insight Net 3.0`

A compilação do *Plugin* `CGEE Insight Net 3.0` ocorre de forma indireta, já que este *plugin* faz parte da infra-estrutura do *Gephi 0.9*. Conforme descrito no site de desenvolvimento do *Gephi* (<https://github.com/gephi/gephi-plugins>), existem três projetos:

- o projeto `Gephi Plugins` que funciona como base e interface com a plataforma *Gephi*
- o projeto `CGEE Insight Net 3.0` que contém o próprio código, é referenciado pelo projeto `Gephi Plugins` e referencia o projeto `InsightNetLibrary`
- o projeto `Correlation Statistics`, que fornece estatísticas adicionais.

Deve ser observado que o número de versão do *plugin* `CGEE Insight Net 3.0` deve ser o mesmo da biblioteca `InsightNetLibrary`.

Os projetos `CGEE Insight Net 3.0` e `Correlation Statistics` não podem ser compilados ou executados diretamente. Essas operações sempre devem ser realizadas a partir do projeto `Gephi Plugins`. Informações adicionais constam na página já mencionada.

10.2.3 Documentação do *CGEE Insight Net*

Esta documentação foi criada com a ferramenta “Sphinx”¹. O código-fonte consta no repositório *SVN* do CGEE e pode ser editado e transformado em documentos PDF e HTML com o seguinte conjunto de ferramentas:

¹<https://www.sphinx-doc.org>

- Eclipse (a partir da versão Oxygen)
- Eclipse plugin ReST editor do Eclipse Marketplace
- Python, a partir da versão 3.6
- Sphinx 1.8.2
- Miktex www.miktex.org
- Perl www.perl.org

Ainda devem ser instalado o módulo Python `sphinx_bootstrap_theme` com o programa `pip`:

```
pip install sphinx_bootstrap_theme
```

A partir disso, pode ser criada uma configuração de execução no Eclipse:

1. Criar *Run configuration* no Eclipse: Usar o “make.exe” do próprio projeto Manual InsightNet e marcar “html” e “latexpdf” como formatos de saída
2. Compilar pelo *Run project* e verificar os arquivos gerados na pasta `build`. Destaca-se que a documentação no formato PDF consta dentro da pasta `build/latex`.

Existe ainda uma integração com a ferramenta *Gitlab* que executa todo o processo em um *container Docker* após um *commit* no repositório:

```

image: jbliesener/sphinx-doc-18-portuguese

pages:
  script:
    - make html
    - make latexpdf
    # - make docx
    # - make text
    - mkdir public
    - mkdir public/html
    - cp -r build/html/* public/html
    - mkdir public/pdf
    - cp build/latex/*.pdf public/pdf
    # - mkdir public/docx
    # - cp build/docx/*.docx public/docx
    # - mkdir public/text
    # - cp -r build/text/* public/text
  artifacts:
    paths:
      - public
  only:
    - master
  tags:
    - sphinx
  
```

10.3 Uso do CGEE Insight Net a partir da linha de comando

10.3.1 Configuração Inicial

A linha de comando é executada a partir dos arquivos `.jar` que constam no diretório `%APPDATA%\gephi\0.9.2\dev\modules\`. Este caminho precisa ser informado ao Java a partir do parâmetro `-cp`. Como o caminho especificado pelo variável `%APPDATA%` geralmente contém espaços, é necessário incluir todo este caminho em aspas duplas.

O nome da classe que executa os comandos é `com.bliesener.cgee.cmdline.CommandLineParser`. Depois do nome da classe seguem os parâmetros do banco de dados:

- `-d` <nome da classe do driver jdbc> [para bancos H2 locais é `org.h2.Driver`]
- `-U` <nome do usuário do banco de dados> [para bancos H2 locais, geralmente é `cgee`]
- `-p` <senha do usuário do banco de dados> [para bancos H2 locais, geralmente é `cgee`]
- `-u` <url do banco de dados> [para bancos H2 locais, geralmente é `jdbc:h2:<caminho do banco>`]

Assim, os parâmetros básicos para executar um comando são, geralmente:

```
java -cp "%APPDATA%/gephi/0.9.2/dev/modules/*" com.bliesener.cgee.cmdline.
↳CommandLineParser
  -d org.h2.Driver -U cgee -p cgee -u jdbc:h2:<caminho do banco> <comando ____
↳a ser executado>
```

Para simplificar este comando, sugere-se gravar os parâmetros básicos em variáveis de ambiente:

```
SET CP="%APPDATA%/gephi/0.9.2/dev/modules/*"
SET DBPARAMS=-d org.h2.Driver -U cgee -p cgee -u "jdbc:h2:<caminho do ____
↳banco>"
SET CGEEPARAMS=-cp %CP% com.bliesener.cgee.cmdline.CommandLineParser
↳%DBPARAMS%
```

A partir desta configuração, a linha de comando pode ser executada com o comando

```
java %CGEEPARAMS% <comando a ser executado>
```

10.3.2 Ajuda

O comando

```
java %CGEEPARAMS% --help
```

mostra todos os comandos e parâmetros que a ferramenta reconhece.

10.3.3 Protocolo de execução

Os seguintes parâmetros podem ser especificados ANTES do comando:

- `-l` <logfile>: Grava o protocolo de execução no arquivo especificado

- `-v <logLevel>`: Especifica o nível de protocolo. `<logLevel>` pode ser um dos seguintes: - OFF, - SEVERE - WARNING - INFO - CONFIG - FINE - FINER - FINEST -ALL
- `-r <reportfile>`: Grava o relatório de execução no arquivo especificado
- `--createreport`: Se o parâmetro `-r` for especificado e o arquivo referenciado já existir, os dados existentes são apagados.

10.3.4 Comandos

A ferramenta disponibiliza os seguintes comandos:

- *newimport*
- *importbibtex*
- *importgeneric*
- *select/unselect*
- *newjob*
- *similaritysearch/tfidfsearch/keywordcooccurrence*
- *consolidate*
- *backup/restore*
- *shell*
- *batch*
- *statistics*
- *reduce*
- *exportgexf*
- *version*
- *query*

10.3.5 newimport

O comando `newimport` permite a importação de currículos Lattes na ferramenta.

```
newimport      Read CVs from directory or web service
Usage: newimport [options] Files or directories to import
Options:
  --delete
    Delete all tables in database before starting (CAUTION!)
    Default: false
  --downloadarchivedir
    Download compressed archive of all CVs to this directory
  --filter
    Use filter expression to select input data from web service
  --info
    Information to store in the 'info' field for each imported CV
  --keyword
```

(continues on next page)

(continuação da página anterior)

```
Use keyword to select input data from web service
--password
  Password to access web service
--repair
  Repair known encoding errors
  Default: false
--username
  User name to access web service
--ws
  Import from Webservice instead of using individual files
  Default: false
```

Parâmetros

Como parâmetro(s) obrigatório(s), o comando espera o nome do(s) diretório(s) ou arquivo(s) a serem importados. Se o nome especificado representa um diretório, todos os arquivos nele serão importados.

O identificador único usado para o currículo importado é o nome do arquivo, já que esta informação nem sempre consta nos dados XML.

--delete: Apagar dados antes da importação. Se este parâmetro for especificado, todos os dados do banco de dados serão apagados antes da importação.

Sem este parâmetro, os dados lidos serão adicionados ao banco existente ou mesclados com ele. Para identificar se um currículo já existe no banco, é usado o identificador (nome do arquivo sem anexo .xml). Se o currículo já existir, a data da última atualização é comparada. Se o currículo importado for mais novo do que o já existente, os dados deste currículo são eliminados do banco e o arquivo é importado.

--info: Gravar campo de informação. Este parâmetro permite a especificação de um texto que será gravado no campo "info" do banco de dados. Textos com espaços devem ser especificados entre aspas duplas.

--repair: Pré-processar os currículos importados para eliminar caracteres inválidos.

--ws: Recuperar dados do *web service*. Este parâmetro permite a importação dos dados via *web service* do CGEE. Mesmo especificando o parâmetro **--ws**, ainda é necessário definir um diretório de importação, **que não será usado**.

--downloadarchivedir: Baixar currículos para um diretório específico. Se esse parâmetro for especificado, os currículos baixados pelo *web service* (usando a opção **--ws**) serão salvos no diretório especificado, em uma pasta compactada. O caminho do diretório deve estar especificado com aspas duplas, e caso este não exista, será criada uma pasta nova no caminho especificado.

--keyword e **--filter:** Selecionar dados do *web service*. Estes parâmetros selecionam dados do *web service* por palavra-chave (**--keyword**) ou por filtro (**--filtro**) e devem ser usados em conjunto com o parâmetro **--ws**. A palavra-chave ou filtro utilizado deve estar separado com aspas duplas.

--username: Definir nome de usuário usado para acessar o *web service*. Usado em conjunto com a opção **--ws**.

--password: Definir senha para o nome de usuário definido com a opção **--username**. Usado em conjunto com a opção **--ws**.

Exemplos

```

JAVA %CGEEPARAMS% newimport --delete --ws --repair --info "INCT para
→Mudancas Climaticas" --downloadarchivedir "C:/Users/Documents/Diretorio_
→CVs" --username "user" --password "admin" --keyword "mudanças climáticas"
JAVA %CGEEPARAMS% newimport -info Demo "C:\users\jorg_
→bliesener\documents\cgEE\Teste100"
JAVA %CGEEPARAMS% newimport --ws --username "user" --password "password" --
→filter "cpf: cpf1,cpf2,cpf3"
  
```

10.3.6 importbibtex

O comando `importbibtex` permite a importação de arquivos do formato bibtex na ferramenta:

```

importbibtex      Read Bibtex files from directory
Usage: importbibtex [options] Files or directories to import
Options:
  --charset
    Character set to use on import
  --delete
    Delete all tables in database before starting (CAUTION!)
    Default: false
  --forcelanguage
    Set language for all records
    Default: false
  --info
    Information to store in the 'info' field for each imported CV
  --keywords
    Type of keywords to import, (standard, author [scopus only],
→plus
    [WoS only]
  --language
    Language to use for records with unknown language or for all
    records (see --replacelanguage)
    Default: English
  
```

Parâmetros

Como parâmetro(s) obrigatório(s), o comando espera o nome do(s) diretório(s) ou arquivo(s) a serem importados. Se o nome especificado representa um diretório, todos os arquivos nele serão importados.

`--charset`: Especifica a codificação de caracteres usada nos arquivos a serem importados. Os valores sugeridos são UTF-8 e windows-1252.

`--delete`: Apagar dados antes da importação. Se este parâmetro for especificado, todos os dados do banco de dados serão apagados antes da importação. Sem este parâmetro, os dados lidos serão adicionados ao banco existente ou mesclados com ele.

`--info`: Gravar campo de informação. Este parâmetro permite a especificação de um texto que será gravado no campo *info* do banco de dados. Textos com espaços devem ser especificados entre aspas duplas.

`--keywords`: Determina quais palavras-chave serão importadas do arquivo BibTeX. Este parâmetro pode ser especificado mais de uma vez para importar tipos de palavras-chave diversos. Valores válidos:

- `standard`: Palavras-chave padrão
- `author keywords`: Palavras-chave especificados pelo autor (apenas para arquivos *Scopus*®),
- `keywords plus`: Palavras-chave adicionais (apenas para arquivos *Web of Science*)

`--forcelanguage`: Trata todas as entradas contidas nos arquivos importados como sendo de uma língua específica.

`--language`:

- Se este parâmetro for especificado junto com `--forcelanguage`: Trata *todas* as referências bibliográficas importadas como o da linguagem determinada nessa opção.
- Se este parâmetro for especificado *sem* `--forcelanguage`: Trata as referências bibliográficas importadas *que não especificam um idioma* como o da linguagem determinada nessa opção.

Opções:

- English
- Portuguese

Exemplo

```
importbibtex --delete "c:\users\jblie_000\Documents\CGEE\Contrato_
→2016\Bibtex com erro\Klionsky2012445_short.bib"
```

10.3.7 importgeneric

O comando `importgeneric` permite a importação de dados bibliográficos genéricos, usando arquivos JSON para definir o formato dos arquivos importados (ver [Seção 10.1](#)):

```
importgeneric      Read generic bibliography files from directory, using
                   format descriptor
Usage: importgeneric [options] Files or directories to import
Options:
  --charset
    Character set to use on import
  --delete
    Delete all tables in database before starting (CAUTION!)
    Default: false
  --descriptor
    Custom descriptor file
  --forcelanguage
    Set language for all records
    Default: false
  --info
    Information to store in the 'info' field for each imported CV
  --keywords
    Type of keywords to import, (author, database]
  --language
    Language to use for records with unknown language or for all
    records (see --replacelanguage)
    Default: English
  --saveDescriptor
```

(continues on next page)

(continuação da página anterior)

```
Export effective descriptor to file
--sheet
  Sheet name or index when importing from Excel files. Enclose
  number in double quotes to force using sheet name instead of sheet
  index
--type
  Generic descriptor type
  Possible Values: [Autodetect Scopus format, Comma-separated data
→from Scopus,
  Tagged data from Scopus, Autodetect Web of Science format,
  Comma-separated data from Web of Science, Tab-separated data
→from Web of Science,
  Tagged data from Web of Science, Generic table data (Autodetect
→format),
  Use external descriptor file]
```

Parâmetros

Como parâmetro(s) obrigatório(s), o comando espera o nome do(s) diretório(s) ou arquivo(s) a serem importados. Se o nome especificado representa um diretório, todos os arquivos nele serão importados.

--charset: Especifica a codificação de caracteres usada nos arquivos a serem importados. Os valores sugeridos são UTF-8 e windows-1252.

--delete: Apagar dados antes da importação. Se este parâmetro for especificado, todos os dados do banco de dados serão apagados antes da importação. Sem este parâmetro, os dados lidos serão adicionados ao banco existente ou mesclados com ele.

--info: Gravar campo de informação. Este parâmetro permite a especificação de um texto que será gravado no campo *info* do banco de dados. Textos com espaços devem ser especificados entre aspas duplas.

--sheet: Especificar nome ou número da aba em planilhas Excel®. Para importações de planilhas Excel®, este parâmetro permite definir o nome ou o número da aba em que constam os dados a serem importados.

--type: Formato de dados. O parâmetro **--type** é obrigatório e determina o formato dos dados de entrada. Os seguintes valores são aceitos e devem ser especificados em aspas duplas:

- Autodetect Scopus format
- Comma-separated data from Scopus
- Tagged data from Scopus
- Autodetect Web of Science format
- Comma-separated data from Web of Science
- Tab-separated data from Web of Science
- Tagged data from Web of Science
- Generic table data (Autodetect format)
- Use external descriptor file

--descriptor: Nome do arquivo que especifica o formato do arquivo a ser importado, caso for definido type "Use external descriptor file"

--saveDescriptor: Exportar o arquivo de especificação do formato para o arquivo especificado

--keywords: Determina quais palavras-chave serão importadas do arquivo BibTeX. Este parâmetro pode ser especificado mais que uma vez para importar tipos de palavras-chave diversos. Valores válidos:

- standard: Palavras-chave padrão
- author keywords: Palavras-chave especificados pelo autor (apenas para arquivos *Scopus*®),
- keywords plus: Palavras-chave adicionais (apenas para arquivos *Web of Science*)

--forcelanguage: Trata todas as entradas contidas nos arquivos importados como sendo de uma língua específica.

--language:

- Se este parâmetro for especificado junto com --forcelanguage: Trata *todas* as referências bibliográficas importadas como o da linguagem determinada nessa opção.
- Se este parâmetro for especificado *sem* --forcelanguage: Trata as referências bibliográficas importadas *que não especificam um idioma* como o da linguagem determinada nessa opção.

Opções:

- English
- Portuguese

10.3.8 select/unselect

Os comandos `select` e `unselect` definem quais contribuições científicas dos pesquisadores farão parte da análise de coautoria e de similaridade, e se o leitor de contribuições `SelectedContributionReader` será usado.

```

select      Add contributions to/Remove contributions from subset to
            process
Usage: select [options]
Options:
  --after
    use with --title: select/unselect only contributions within this
    many years after the researcher has obtained the specified title
    Default: 0
  --all
    add/remove all contributions
    Default: false
  --defaultContributionSelect
    use with --title: select/unselect all contributions that do not
    have a publication date
    Default: false
  --defaultResearcherSelect
    use with --title: select/unselect all contributions for researchers
    that don't have the specified title
    Default: false
  --title
    select/unselect contributions after this title
  --where

```

(continues on next page)

(continuação da página anterior)

```

filter expression

unselect      Add contributions to/Remove contributions from subset
              to process
Usage: unselect [options]
Options:
  --after
    use with --title: select/unselect only contributions within this
    many years after the researcher has obtained the specified title
    Default: 0
  --all
    add/remove all contributions
    Default: false
  --defaultContributionSelect
    use with --title: select/unselect all contributions that do not
    have a publication date
    Default: false
  --defaultResearcherSelect
    use with --title: select/unselect all contributions for researchers
    that don't have the specified title
    Default: false
  --title
    select/unselect contributions after this title
  --where
    filter expression
    
```

Parâmetros

--where: Permite a especificação de uma cláusula de escopo. Esta cláusula segue a sintaxe do JPA (ver <http://www.objectdb.com/java/jpa/query/jpql/where>), sendo que o comando completo começa com `SELECT bc FROM BASECONTRIBUTION`

Os campos e relacionamentos devem seguir o modelo de banco de dados da ferramenta (ver documentação e diagrama de entidades e relacionamentos).

--all: Seleciona/Desseleciona todas as contribuições.

--title: Seleciona/Desseleciona contribuições depois de um título específico (Ex: Mestrado, Doutorado). O programa só levará em consideração as publicações feitas depois do título especificado.

--after: Seleciona/Desseleciona contribuições contidas no período definido de anos depois do título. Deve ser usada com a opção **--title**.

--defaultContributionSelect: Seleciona/Desseleciona todas as publicações sem data especificada. Deve ser usada com a opção **--title**.

Exemplos

```

java %CGEEPARAMS% select --where "TYPE (bc)=Artigo"
java %CGEEPARAMS% select --where "TYPE (bc)=CapituloLivro AND bc.ano>=1990
→AND bc.ano<=2015"
java %CGEEPARAMS% select --where "TYPE (bc)=TrabalhoEmEventos AND TREAT (bc
→as TrabalhoEmEventos).workType=com.bliesener.cgee.entities.
→TrabalhoEmEventos.WorkType.COMPLETO AND bc.ano>=1965 AND bc.ano<=2015"
    
```

10.3.9 newjob

O comando `newjob` prepara o banco de dados para a execução de uma pesquisa de similaridade e define o escopo da mesma.

Parâmetros

Os parâmetros se dividem em três grupos:

- Parâmetros que constam no diálogo da pesquisa por similaridade da ferramenta:
 - `--casesensitive`
 - `--partitionyear`
 - `--dolevenshtein`
 - `--minsimilarity` e `--minsimilaritylevenshtein` (sinónimos)
 - `--dotfidf`
 - `--content`
 - `--dostemming`
 - `--usestopwords`
 - `--minsimilaritytfidf`
 - `--sparse`
 - `--dokeywordcooccurrence`
 - `--minkeywordcooccurrence`
 - `--selectedcontributiononly`
- Parâmetros que constam na tela principal de configurações da ferramenta:
 - `--blocksize`
 - `--useoldlimitformula`
 - `--usealternativelevenshtein`
 - `--poolsametitle`
 - `--usebktreelimit`
 - `--tflog`
 - `--tfidfpercentile`
 - `--roundtfidf`
- Parâmetros de seleção de escopo - Ver próximo parágrafo:
 - `--contributionreader`
 - `--entitytype`
 - `--whereclause`
 - `--type`

--update

Seleção de Escopo

Durante a fase de desenvolvimento da ferramenta, o escopo de análise de similaridade variou e foi necessária uma forma flexível para defini-lo. Hoje existem quatro métodos para especificar quais contribuições serão incluídas na pesquisa de similaridade:

1. Todas as contribuições: Nenhum dos parâmetros `--contributionreader`, `--entitytype`, `--whereclause` será especificado.
2. Contribuições de apenas um tipo específico. O parâmetro `--entitytype` especifica este tipo:
 1. `java %CGEEPARAMS% newjob --dolevenshtein --entitytype Artigo`
 2. `java %CGEEPARAMS% newjob --dolevenshtein --entitytype CapituloLivro`
 3. `java %CGEEPARAMS% newjob --dolevenshtein --entitytype TrabalhoEmEventos`
3. Contribuições que atendem uma cláusula `WHERE` específico, de acordo com a sintaxe do JPA (veja <http://www.objectdb.com/java/jpa/query/jpql/where>):
 1. `java %CGEEPARAMS% newjob --dolevenshtein --whereclause "TYPE(bc)=CapituloLivro AND bc.ano>=1990 AND bc.ano<=2015"`
4. Para escopos mais complexos, como aquele usado frequentemente no CGEE (todos os Artigos, todos os capítulos de livros e apenas trabalhos completos em eventos, pode ser especificada uma classe java específica que fornece estas contribuições. Para isso, existe a classe `com.bliesener.cgee.similarity.SelectedContributionReader`, que retorna todas as contribuições selecionadas com os comandos `select/unselect`:
 1. `java %CGEEPARAMS% newjob --dolevenshtein--contributionreader com.bliesener.cgee.similarity.SelectedContributionReader`

Exemplos

```
JAVA %CGEEPARAMS% newjob --minsimilaritytfidf 0.05 --partitionyear
JAVA %CGEEPARAMS% newjob --contributionreader SelectedContributionReader --
->dolevenshtein false --dostemming false --minsimilaritytfidf 0.15
```

10.3.10 `similaritysearch/tfidfsearch/keywordcooccurrence`

O comando `similaritysearch` executa a pesquisa por co-autoria definida com `newjob`. O comando `tfidfsearch` executa a pesquisa definina com `newjob`, realizando a busca por similaridade semântica. O comando `keywordcooccurrence` cria redes de co-ocorrências de palavras-chaves.

```
similaritysearch      Start similarity search on database
Usage: similaritysearch [options]
Options:
  --threads
      Set number of concurrent threads
```

(continues on next page)

(continuação da página anterior)

Default: 1

tfidfsearch Start tf.idf similarity search on database

Usage: tfidfsearch [options]

Options:

--threads

Set number of concurrent threads

Default: 1

keywordcooccurrence Create network of keyword co-occurrences

Usage: keywordcooccurrence [options]

Options:

--threads

Set number of concurrent threads

Default: 1

Parâmetros

--threads: Define a quantidade de execuções em paralelo e corresponde ao respectivo parâmetro definido na tela de opções da ferramenta visual. Recomenda-se um valor que fica abaixo da quantidade de núcleos do computador usado.

Exemplos

```
JAVA %CGEEPARAMS% tfidfsearch --threads 7
JAVA %CGEEPARAMS% similaritysearch --threads 7
JAVA %CGEEPARAMS% keywordcooccurrence --threads 7
```

10.3.11 consolidate

O comando `consolidate` processa os resultados da pesquisa **por coautorias** e gera as arestas do grafo no banco de dados.

consolidate Consolidate similarity search results

Usage: consolidate [options]

Options:

--checksymmetry

Check graph symmetry

Default: false

--checksymmetryonly

Check graph symmetry only, do not consolidate

Default: false

Parâmetros

Os parâmetros servem apenas para a depuração do procedimento.

Exemplos

```
JAVA %CGEEPARAMS% consolidate
```

10.3.12 backup/restore

Os comandos `backup` e `restore` permitem a gravação do banco de dados em arquivos externos e a sua recuperação.

```
backup      Backup internal database
Usage: backup [options]
Options:
  --filename
      Output file name
  -z, --zip
      Zip compress output file
      Default: false

restore     Restore internal database from backup
Usage: restore [options]
Options:
  --filename
      Input file name
  -z, --zip
      Input file is zip compressed
      Default: false
```

Parâmetros

`--filename`: Nome do arquivo a ser carregado/salvo

`--zip`: Define se o arquivo de input/output é uma pasta comprimida ou não.

Exemplos

```
JAVA %CGEEPARAMS% backup --filename "Test"
JAVA %CGEEPARAMS% restore --filename "Test" --zip
```

10.3.13 shell

O comando `shell` abre um interpretador de linha de comando que permite a digitação de comandos interativos.

```
shell      Enter interactive shell
Usage: shell [options]
```

10.3.14 batch

O comando `batch` permite a execução de vários comandos gravados em arquivo.

```
batch      Run commands from batch file
Usage: batch [options] Batch file to run
```

Parâmetros

Como parâmetro do comando deve ser especificado o nome do arquivo que contém os comandos a serem executados.

Exemplo

```
JAVA %CGEEPARAMS% -v FINE -l c:\temp\cgee.log batch c:\temp\cgeeCommands.
->txt
```

10.3.15 reduce

O comando `reduceremove` todos os nós do banco de dados exceto aqueles selecionados.

```
reduce      Remove all but a specific set of nodes from network
Usage: reduce [options] List of remaining ids in network
```

Parâmetros

Como parâmetro do comando devem ser especificados os ids dos nós que deverão permanecer na rede.

Exemplo

```
JAVA %CGEEPARAMS% reduce "id1, id2, id3"
```

10.3.16 statistics

O comando `statistics` mostra as informações gerais sobre a base.

```
statistics  Provide database statistics
Usage: statistics [options]
```

10.3.17 exportgexf

O comando `exportgexf` exporta a rede contida no banco de dados como um arquivo do tipo `.gexf`.

```
exportgexf  Export Graph to GEXF file
Usage: exportgexf [options] Name of file to export
Options:
  --withinfo
    Export the 'info' field for each Pesquisador
  Default: false
```

Parâmetros

Como parâmetro obrigatório do comando deve ser fornecido o nome do arquivo de output.

--withinfo: Define se o campo 'info' dos nós também será exportado.

Exemplo

```
JAVA %CGEEPARAMS% exportgexf --withinfo "Test.gexf"
```

10.3.18 version

O comando `version` mostra a versão atual da biblioteca.

```
version      Print library version  
Usage: version [options]
```

10.3.19 query

O comando `query` permite realizar pesquisas JPA diretamente na base de dados do *CGEE Insight Net*. Esta funcionalidade serve, principalmente, para depurar o módulo.

Parâmetros

O parâmetro obrigatório é a pesquisa a ser realizada.

--limit: Limita a quantidade de resultados.

Exemplo

```
java %cgeeparams% query "SELECT COUNT(bc) FROM BaseContribution bc"
```

10.3.20 Fluxo de Trabalho

Geralmente, um fluxo de criação de redes segue a seguinte sequência de comandos:

1. Importação dos CVs/Arquivos Bibtex/Referências bibliográficas - comando *newimport* ou *import-bibtex* ou *importgeneric*
2. Seleção das publicações desejadas - comandos *select/unselect* OBS: Nessa etapa são escolhidos os tipos de publicação e o período de publicações desejado
3. Determinação dos parâmetros de busca - comando *newjob* OBS: Nessa etapa são definidas a maior parte das opções contidas na tela usada para rodar a rede.
4. Execução da busca por coautorias e/ou similaridade semântica - comandos *similaritysearch/tfidfsearch/keywordcooccurrence*
5. Consolidação dos resultados - comando *consolidate*

6. Exportação dos resultados - comandos *exportgexf* e/ou *backup/restore*

Segue um exemplo de um fluxo de trabalho de análise de rede de pesquisadores.

```
java %CGEEPARAMS% newimport --delete --info "Test100" "c:\users\jorg_
->bliesener\documents\cgee\teste100"
java %CGEEPARAMS% select --where "TYPE(bc)=Artigo"
java %CGEEPARAMS% select --where "TYPE(bc)=CapituloLivro"
java %CGEEPARAMS% select --where "TYPE(bc)=TrabalhoEmEventos AND TREAT(bc_
->as TrabalhoEmEventos).workType=com.bliesener.cgee.entities.
->TrabalhoEmEventos.WorkType.COMPLETO"
java %CGEEPARAMS% newjob --dolevenshtein --partitionyear --
->contributionreader com.bliesener.cgee.similarity.
->SelectedContributionReader
java %CGEEPARAMS% similaritysearch --threads 7
java %CGEEPARAMS% consolidate
```

Depois da execução destes comandos, a rede consta no banco de dados e pode ser extraída a partir de comandos SQL.

O mesmo fluxo pode ser realizado em arquivo batch, de acordo com o exemplo em seguida. A vantagem é que a conexão com o banco de dados precisa ser feita apenas uma única vez. Neste exemplo, o arquivo terá o nome `cmdline.batch`.

```
# importar os curriculos
import --delete --info "Test100" "c:\users\jorg_
->bliesener\documents\cgee\teste100"
# selecionar todos os Artigos, todos os capítulos de livros e apenas os_
->trabalhos em eventos COMPLETOS
select --where "TYPE(bc)=Artigo"
select --where "TYPE(bc)=CapituloLivro"
select --where "TYPE(bc)=TrabalhoEmEventos AND TREAT(bc as
->TrabalhoEmEventos).workType=com.bliesener.cgee.entities.
->TrabalhoEmEventos.WorkType.COMPLETO"
# realizar pesquisa levenshtein e consolidar os resultados
newjob --dolevenshtein --partitionyear --contributionreader com.bliesener.
->cgee.similarity.SelectedContributionReader
similaritysearch --threads 7
consolidate
```

Para executar o fluxo e gravar os resultados com protocolo fino em arquivo de protocolo `cgeeLog.log` pode ser usado o seguinte comando:

```
java %CGEEPARAMS% -l cgeeLog.log -v FINE batch cmdline.txt
```

No exemplo abaixo, é realizada uma extração de currículos usando um filtro de CPF através do webservice. Depois, são selecionadas as publicações do tipo Artigo Completo, Capítulo de Livro, e Participações em Eventos Completas, no período de 1999 a 2007. A rede é rodada com similaridade semântica mínima de 0.05, e tanto a busca por coautorias quanto por similaridade semântica são realizadas. Depois, a rede gerada é exportada para um arquivo `.gexf`:

```
JAVA %CGEEPARAMS% newimport --delete --ws --repair --info "Física Quântica
->" --downloadarchivedir "C:/Users/Documents/minhas_redes" --username "user
->" --password "password" --filter "cpf:cpf1, cpf2, cpf3;"
JAVA %CGEEPARAMS% unselect --all
JAVA %CGEEPARAMS% select --where "TYPE(bc)=Artigo AND TREAT(bc as Artigo).
->workType=com.bliesener.cgee.entities.Artigo.WorkType.COMPLETO AND bc.ano>
->=1999 AND bc.ano<=2007" (continues on next page)
```

(continuação da página anterior)

```
JAVA %CGEEPARAMS% select --where "TYPE(bc)=CapituloLivro AND bc.ano>=1999  
→AND bc.ano<=2007"  
JAVA %CGEEPARAMS% select --where "TYPE(bc)=TrabalhoEmEventos AND TREAT(bc  
→as TrabalhoEmEventos).workType=com.bliesener.cgee.entities.  
→TrabalhoEmEventos.WorkType.COMPLETO AND bc.ano>=1999 AND bc.ano<=2007"  
JAVA %CGEEPARAMS% newjob --minsimilaritytfidf 0.05 --partitionyear  
JAVA %CGEEPARAMS% similaritysearch --threads 7  
JAVA %CGEEPARAMS% tfidfsearch --threads 7  
JAVA %CGEEPARAMS% consolidate  
JAVA %CGEEPARAMS% exportgexf --withinfo "C:/Users /Documents/Minhas Redes/  
→rede_antes.gexf"
```

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.

insightNet Browser - Guia do usuário

Release 1.6.7

CGEE

1	Instalação	3
1.1	Requisitos	3
1.2	Dispositivos e telas	3
1.3	Instalação	4
1.4	Estrutura de pastas	4
1.5	Arquitetura.....	6
2	Execução	7
2.1	A escolha do Firefox como navegador padrão.....	7
2.2	Criando o perfil Firefox para o insightNet Browser	8
2.3	Abrindo o insightnet Browser	13
3	Interface	15
3.1	Barra de título	16
3.1.1	Pesquisa	16
3.1.2	Menu de redes	16
3.1.3	Números básicos da rede.....	16
3.1.4	Menu principal	16
3.2	Área de detalhamento	18
3.2.1	Ferramentas da área de detalhamento	19
3.2.2	Exibindo/ocultando arestas na área de detalhameto.....	20
3.3	Área de navegação	22
3.3.1	Janela de dica do nó	22
3.3.2	Controles de navegação	22
3.3.3	Visão geral da rede (minitela)	23
3.3.4	Menu de partições	24
4	Configuração	25
4.1	Configurações de exibição da rede	26
4.1.1	Transparência das arestas.....	26
4.1.2	Espessura das arestas	26
4.1.3	Fator para cálculo do tamanho dos labels	26
4.1.4	Tamanho mínimo dos labels.....	26
4.1.5	Fator para cálculo do tamanho dos nós	26
4.1.6	Arestas curvas	27
4.1.7	Exibir arestas enquanto o gráfico é arrastado.....	27
4.1.8	Exibir dicas dos nós	27

4.2	Configurações das nuvens de palavras.....	27
4.2.1	Atributo para as Nuvens de palavras.....	27
4.2.2	Idioma da rede.....	27
4.2.3	Tipo de espiral.....	27
4.2.4	Tamanho mínimo do texto.....	28
4.2.5	Tamanho máximo do texto.....	28
4.2.6	Número máximo de palavras.....	28
4.2.7	Personalizar Stopwords.....	28
4.2.8	Número de termos extraídos nos campos textuais.....	29
4.3	Reinicialização, backup e restauração.....	29
4.3.1	Limpar dados.....	29
4.3.2	Reiniciar rede.....	29
4.3.3	Restaurar backup.....	29
4.3.4	Backup da rede.....	29
5	Utilização	31
5.1	Visualizando uma rede existente.....	31
5.2	Acrescentando novas redes.....	32
5.2.1	Via URL.....	32
5.2.2	Via Menu de redes.....	32
5.3	Interação e navegabilidade.....	33
5.4	Selecionando nós.....	33
5.4.1	Selecionando com um clique.....	33
5.4.2	Selecionando com área de seleção.....	34
5.4.3	Selecionando em gráficos.....	35
5.5	Ferramenta lente de aumento.....	35
6	Módulos	37
6.1	Metadados da rede.....	37
6.2	Tabelas de nós e arestas.....	39
6.2.1	Selecionando registros.....	39
6.2.2	Exportação.....	40
6.2.3	Exibir/Ocultar colunas.....	40
6.2.4	Tabela de nós.....	41
6.2.5	Tabela de arestas.....	42
6.3	Pesquisa avançada.....	43
6.3.1	Termos de busca.....	43
6.3.2	Gerenciamento de pesquisas.....	46
6.4	Nuvem de palavras.....	47
6.4.1	Informações gerais da nuvem.....	48
6.4.2	Seleção de palavras e consultas.....	49
6.4.3	Filtro de frequência.....	51
6.4.4	Tabela de palavras e frequências.....	51
6.4.5	Outras funcionalidades.....	52
6.5	Gráficos da rede.....	53
6.5.1	Scatterplot.....	55
6.5.2	Histograma.....	57
6.5.3	Palavras-chave.....	60
6.6	Visualizações.....	65
6.6.1	Gerenciando visualizações.....	65
6.6.2	Modo de apresentação.....	65

6.7	Partições.....	65
-----	----------------	----

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.

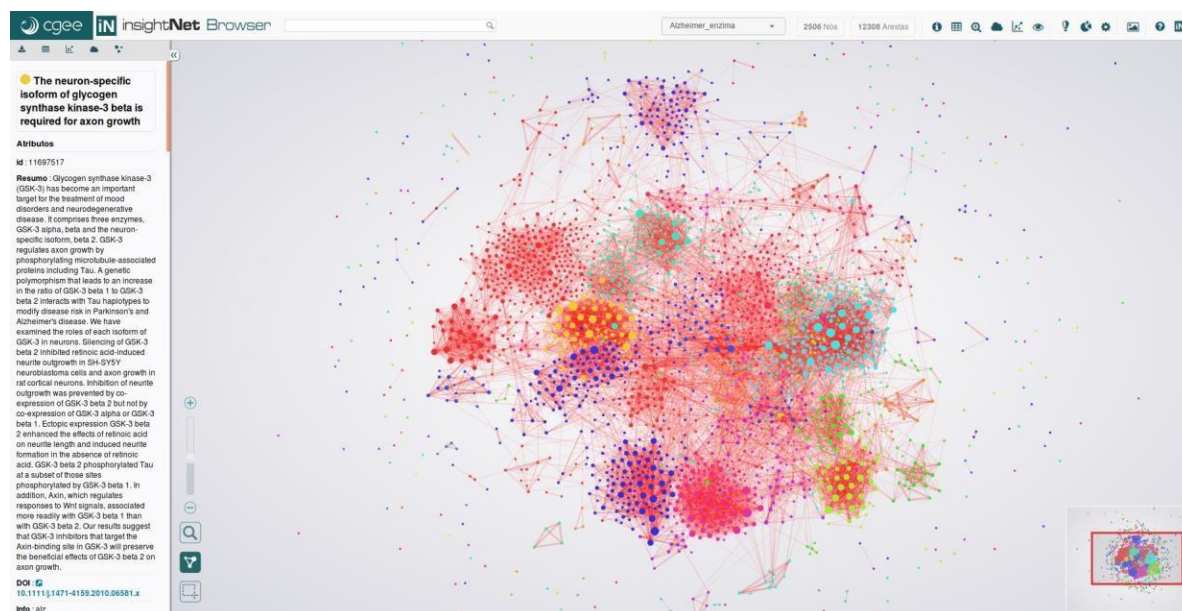


Figura 1: Interface do usuário insightNet Browser

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 textos 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.



Figura 2: 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.

Atenção: Antes de executar a aplicação leia a seção *Executando o insightNet Browser*.

1.4 Estrutura de pastas

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

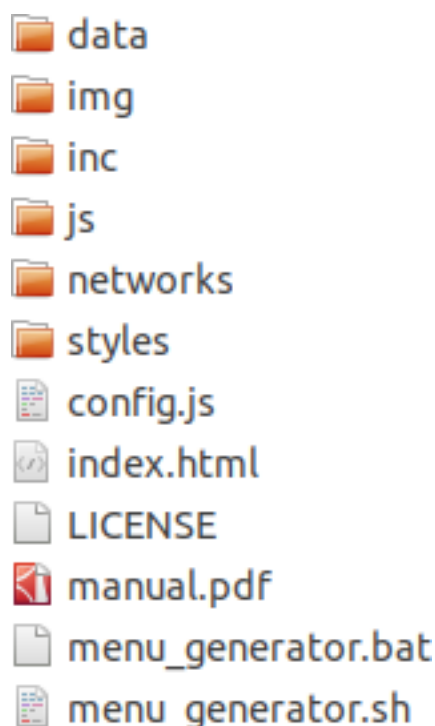


Figura 3: Estrutura de pastas do insightNet Browser

Tabela 1: 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 no navegador (*browser*), sem a necessidade de instalação no sistema operacional.



Figura 4: 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 haja 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/>.

Antes de executar o insightNet Browser é preciso entender algumas particularidades da aplicação.

2.1 A escolha do Firefox como navegador padrão

Há um mecanismo de segurança implementado em todos os navegadores web chamado **CORS**, *Cross-Origin Resource Sharing* (Compartilhamento de recursos com origens diferentes), o qual impede que aplicações web em um determinado protocolo/domínio acessem recursos de outro protocolo/domínio. Na prática, isso obriga as aplicações web a estarem hospedadas em um servidor para, por exemplo, terem acesso a arquivos (XML ou JSON) de dados em seus diretórios internos.

Saiba mais

Para mais informações sobre CORS acesse https://developer.mozilla.org/pt-BR/docs/Web/HTTP/Controle_Acesso_CORS.

Esse mecanismo impossibilitaria o funcionamento desktop do insightNet Browser. No entanto, o navegador Firefox sempre seguiu uma política diferente para esse mecanismo e, até a atualização 68, ele não bloqueava o acesso a arquivos XML e JSON no mesmo diretório da aplicação web. Como o insightNet Browser precisa ler um arquivo GEXF da rede (que é um tipo de arquivo XML), o Firefox se tornou a única opção de navegador que possibilitava o normal funcionamento da ferramenta.

No entanto, depois da atualização do Firefox para a versão 68 (em julho de 2019), a política de CORS do navegador mudou e passou a trazer como padrão uma configuração mais restrita, impedindo o acesso a arquivos XML ou JSON, mesmo que esses se encontrem dentro do diretório da aplicação. Isso bloqueia o acesso do insightNet Browser aos arquivos de rede (GEXF/XML), por exemplo.

A simples alteração dessa configuração resolve o problema de funcionamento do insightNet Browser, porém pode criar uma vulnerabilidade no navegador caso o mesmo seja utilizado para navegação em sites não confiáveis ou na execução de outras aplicações web de origens desconhecidas.

Felizmente, há como utilizar o navegador Firefox por perfil e manter configurações distintas para cada perfil, além de poder rodar várias instâncias do Firefox ao mesmo tempo, cada uma com um perfil diferente. Na prática, isso significa que podemos criar um perfil do Firefox, com uma configuração específica, apenas para executar o insightNet Browser e outro perfil, com configuração mais restrita, para a navegação normal na web.

Siga o passo-a-passo abaixo para criar e configurar um perfil para a utilização do insightNet Browser:

2.2 Criando o perfil Firefox para o insightNet Browser

Passo 1. Esse primeiro procedimento habilitará a opção de escolha de perfil na execução do Firefox e a execução de uma nova instância sempre que o navegador é executado.

- Abra o programa Executar do Windows, utilizando as tecla de atalho: Windows + R



. Ou no campo de pesquisa do Windows digite “Executar” para encontrar o programa e clique no mesmo.

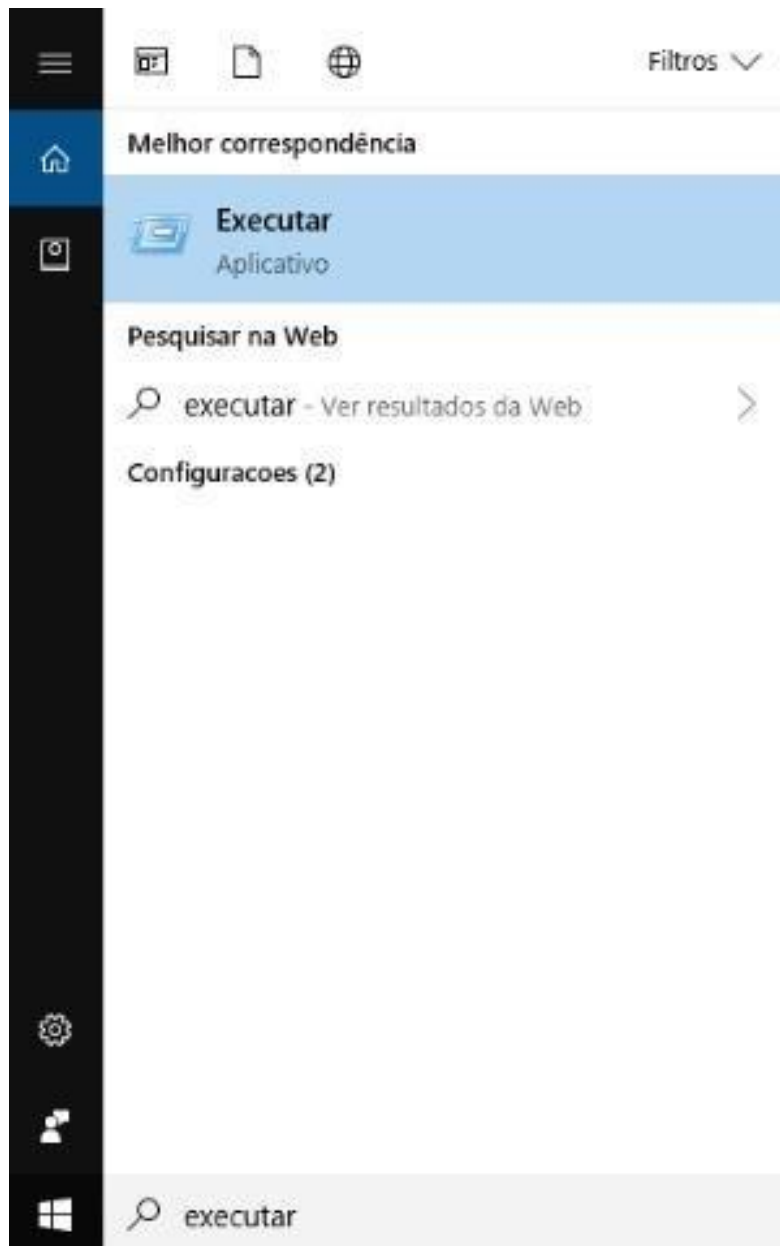


Figura 5: Buscando a aplicação Executar do Windows

- No campo **Abrir** (da janela Executar), digite o comando `firefox -P -new-instance` e clique em **OK**.

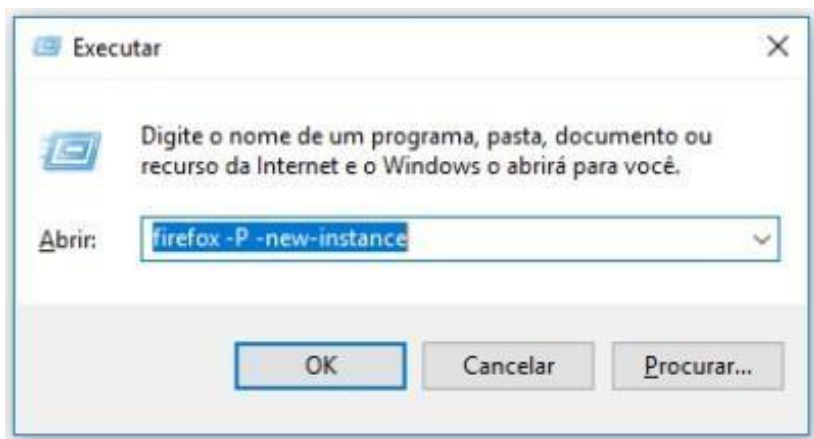


Figura 6: Executando o Firefox com escolha de perfis ativada

Dica: No Linux: Abra o terminal, digite o comando `firefox -P -new-instance` e clique em **ENTER**.

- Na janela **Firefox - Perfil de usuário**, desmarque a opção **Usar o perfil selecionado sem perguntar ao iniciar**.

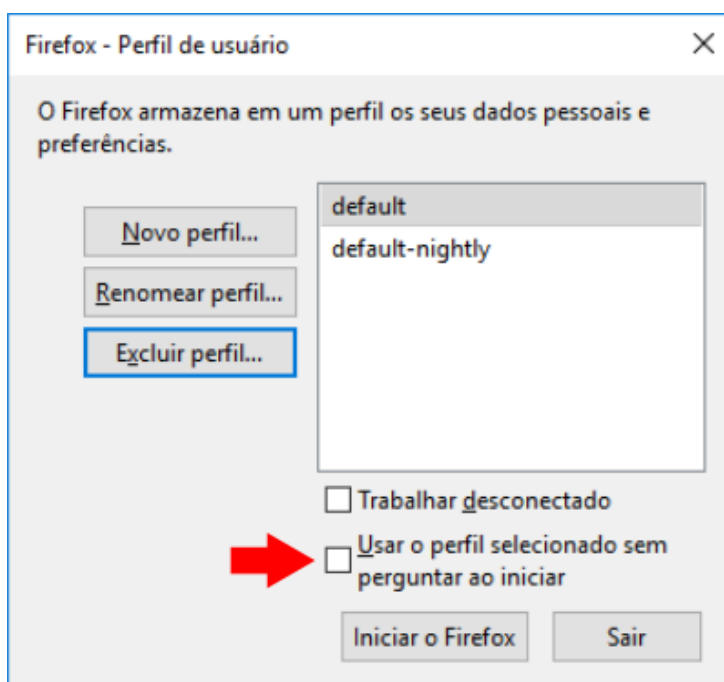


Figura 7: Janela Firefox - Perfil de usuário

Isso fará com que o Firefox sempre exiba a janela de perfis ao abrir. Mantenha essa janela aberta e siga o próximo passo.

Passo 2. Vamos agora criar um novo perfil para o insightNet Browser e fazer as configurações necessárias:

- Ainda na janela **Firefox – Perfil de usuário**, clique no botão **Novo perfil...** e, em seguida, clique

no botão **Avançar**. Na janela seguinte digite um nome para o novo perfil (ex.: insightNet Browser) e clique em **Concluir**.

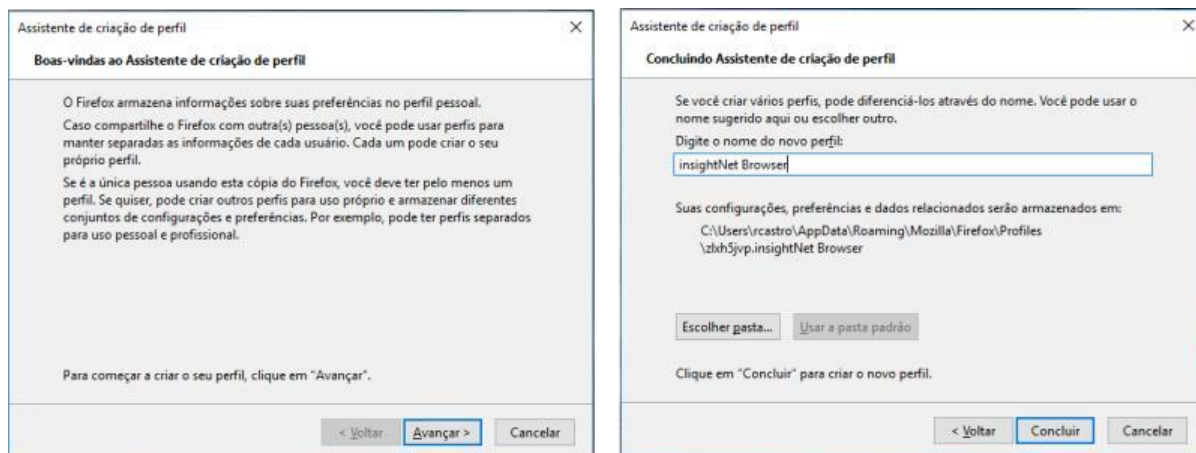


Figura 8: Criando um novo perfil no Firefox

- Ao clicar em **Concluir** o novo perfil será criado e exibido na lista da janela de perfis. Selecione o novo perfil e clique em **Iniciar o Firefox**. Será aberta uma instância do Firefox para o perfil escolhido.

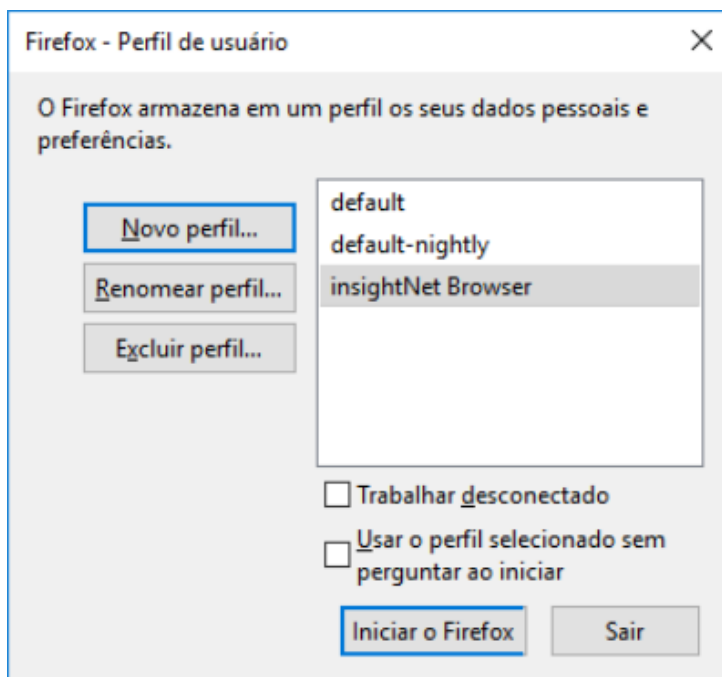


Figura 9: Escolhendo o perfil do Firefox

Passo 3. Agora vamos configurar o novo perfil do Firefox para executar o insightNet Browser:

- Com o Firefox aberto no novo perfil criado, digite `about:config` na barra de endereços do navegador e clique em **ENTER**.

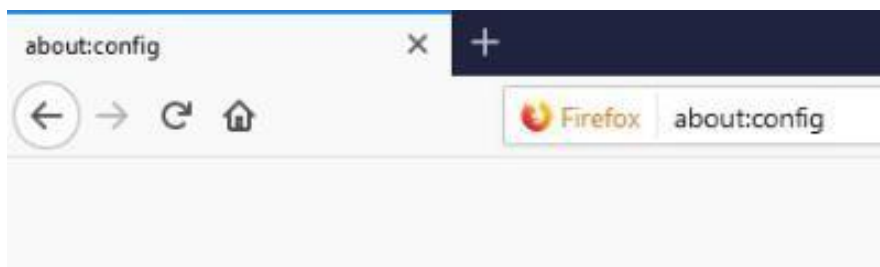


Figura 10: Entrando nas configurações avançadas do Firefox

- Na página que será exibida clique em **Eu aceito o risco!**

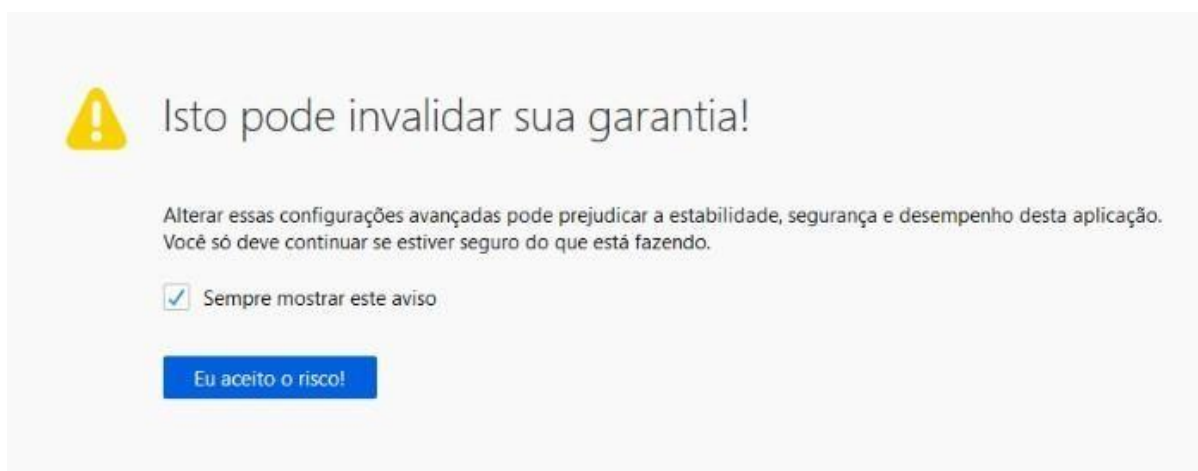


Figura 11: Acessando as configurações avançadas do Firefox

- Na página de configurações avançadas do Firefox, no campo **Procurar**, faça uma busca pelo item de configuração desejado, digitando: `strict_origin`.



Figura 12: Buscando a configuração avançada do Firefox

- Dê dois cliques na opção `security.fileuri.strict_origin_policy` alterando-a de **true** para **false**.

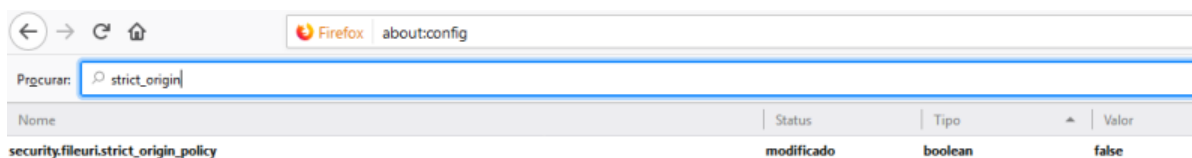


Figura 13: Alterando a configuração avançada do Firefox

- Pronto! Feche o Firefox e siga as instruções a seguir para abrir a aplicação. Todas as vezes que o Firefox for aberto você será poderá escolher o perfil a ser utilizado.

Dica: Mesmo que você esteja utilizando o Firefox no perfil do insightNet Browser, você pode abrir outras instâncias do navegador com outros perfis (ex.: default) e navegar pela web normalmente.

2.3 Abrindo 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**.

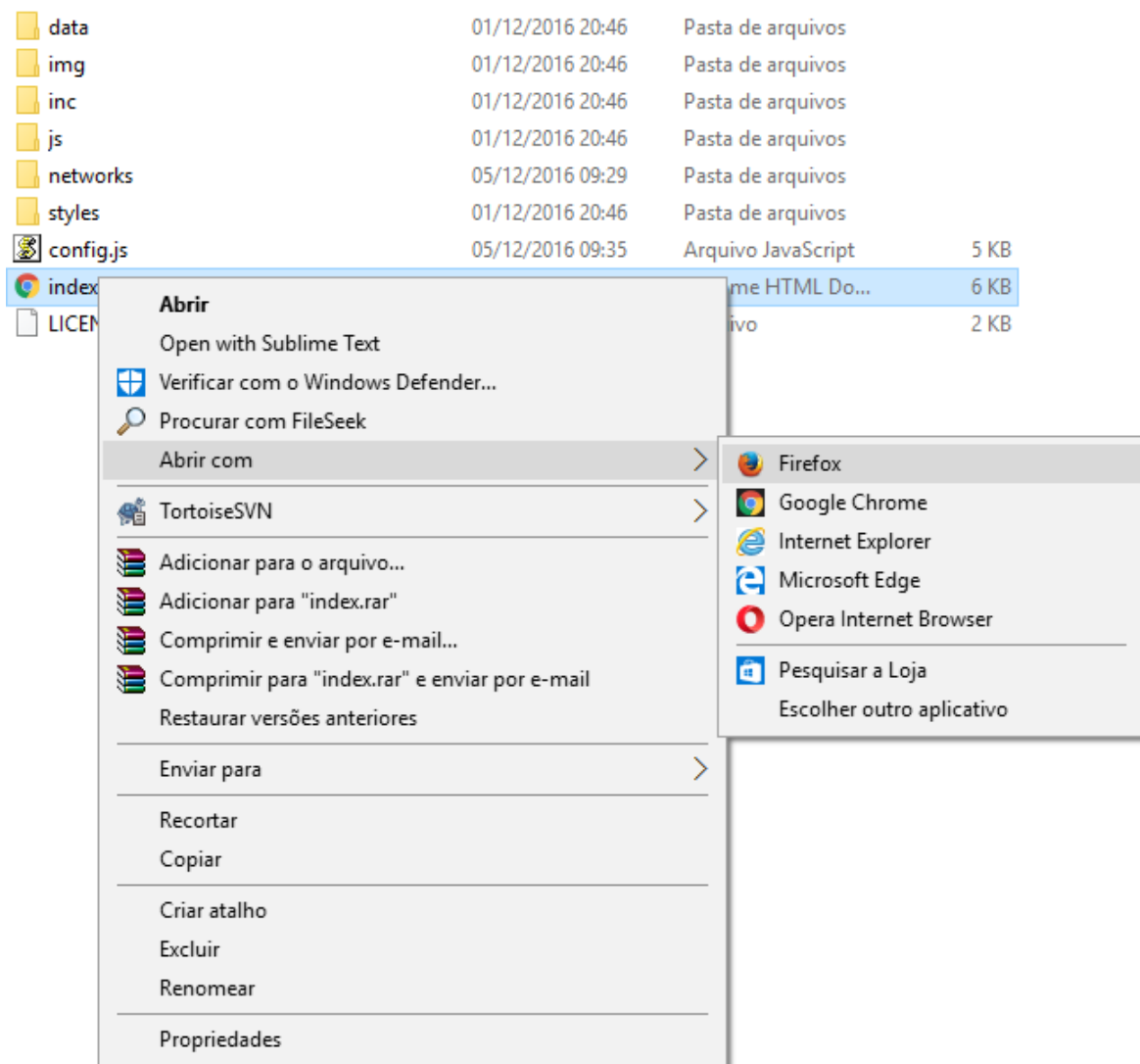


Figura 14: Abrindo a aplicação com o navegador Firefox

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:

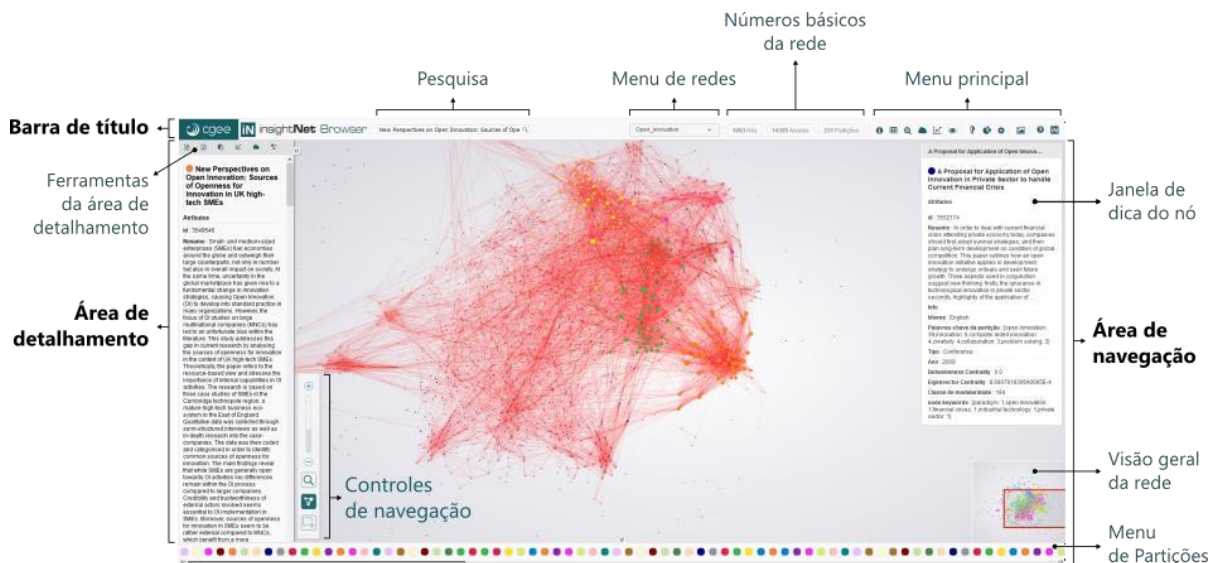


Figura 15: Elementos de interface do insightNet Browser

3.1 Barra de título

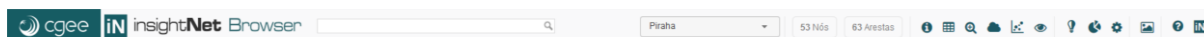


Figura 16: Barra de título insightNet Browser

3.1.1 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*.

3.1.2 Menu de redes

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

3.1.3 Números básicos da rede













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

3.1.4 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:

Tabela 2: 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.

3.2 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]

Nodes 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

Figura 17: 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.

3.2.1 Ferramentas da área de detalhamento

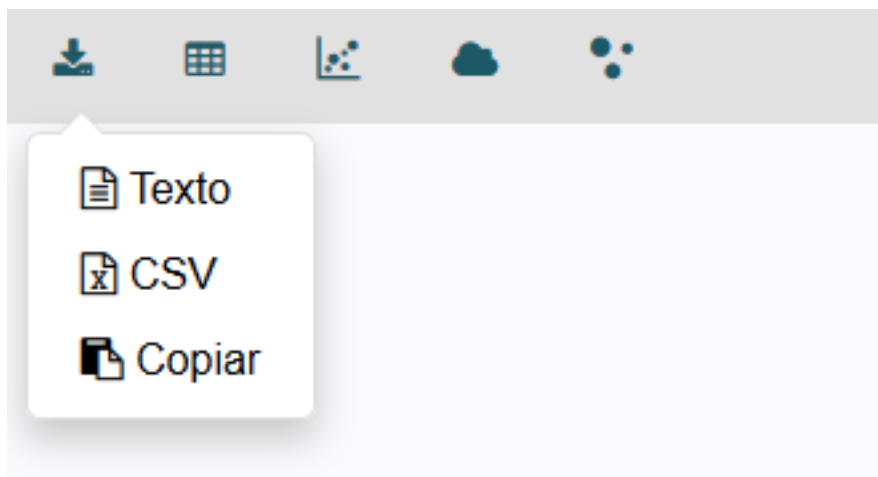










Figura 18: Ferramentas da área de detalhamento

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*.

Tabela 3: 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).

3.2.2 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

Figura 19: 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.

3.3 Área de navegação

3.3.1 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.

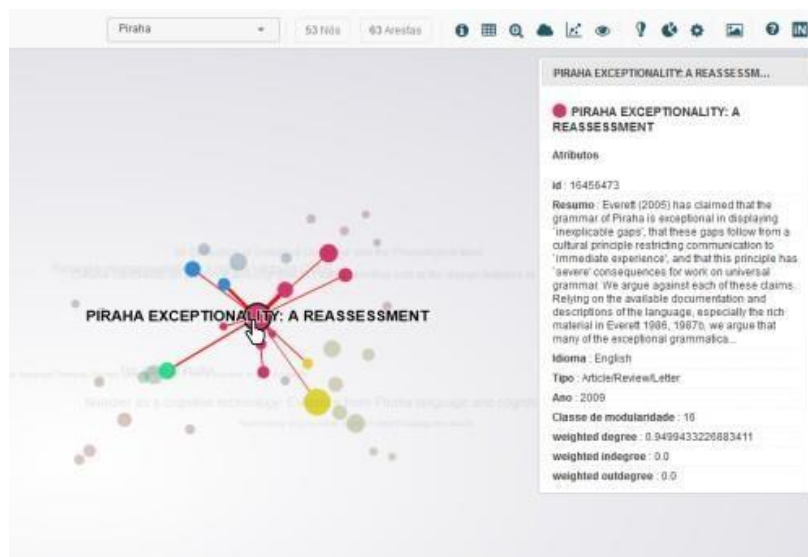


Figura 20: Exemplo de exibição da janela de dica do nó





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

3.3.2 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.



Figura 21: 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).

3.3.3 Visão geral da rede (minitela)

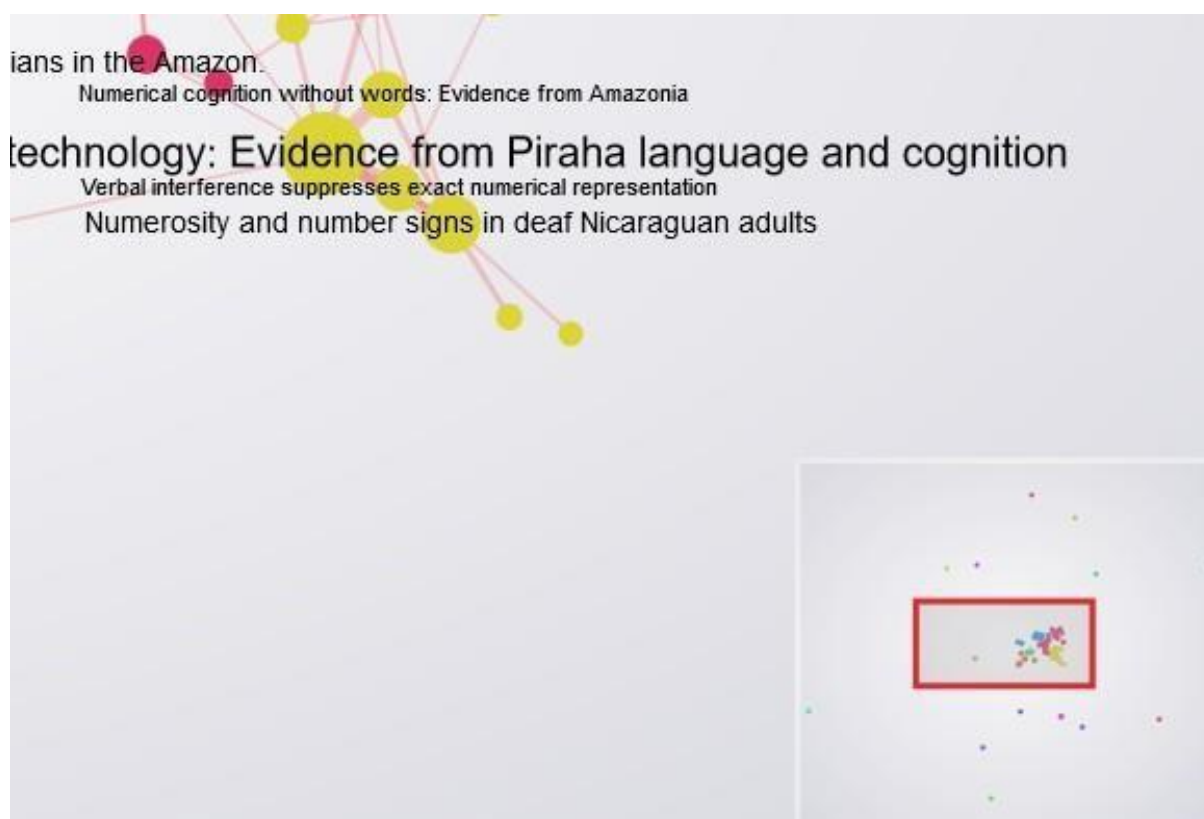



Figura 22: Ferramenta de visão geral da rede

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 (clitando, arrastando e soltando). Nesse caso a ferramenta de visão geral pode auxiliar na movimentação da rede para a área desejada.

3.3.4 Menu de partições

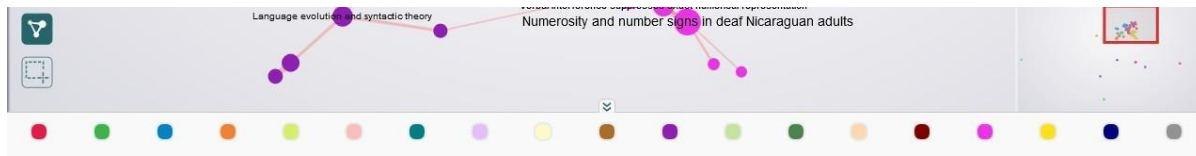




Figura 23: Menu de partições

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.

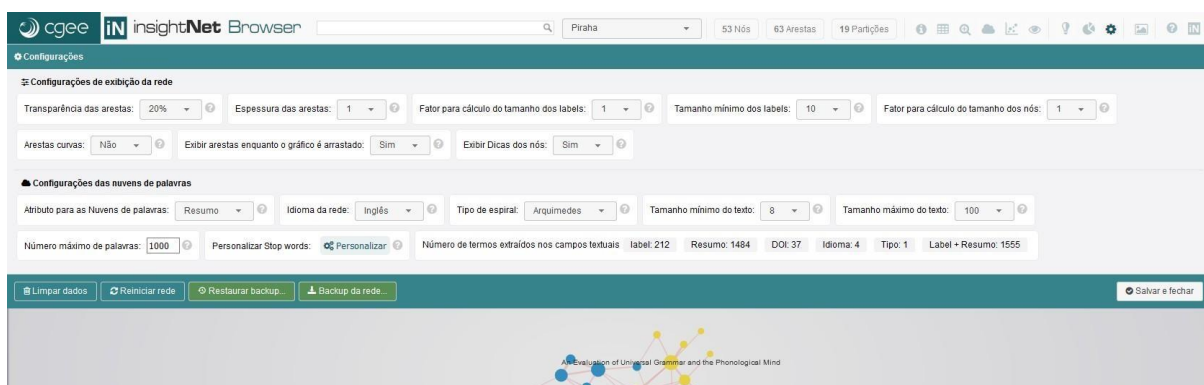



Figura 24: Painel de configurações do insightNet Browser

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.

4.1 Configurações de exibição da rede

4.1.1 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%.

4.1.2 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.

4.1.3 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.

4.1.4 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.

4.1.5 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.

4.1.6 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.

4.1.7 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.

4.1.8 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.

4.2 Configurações das nuvens de palavras

4.2.1 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.

4.2.2 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.

<p>Atenção: Ao alterar o idioma da rede você perderá as personalizações feitas na lista de <i>stopwords</i> anterior.</p>
--

4.2.3 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.

4.2.4 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.

4.2.5 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.

4.2.6 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.

4.2.7 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*.

Menu de exportação:

cópia Excel CSV PDF

#	Texto	Excluir
1	a	<input type="checkbox"/>
2	a's	<input type="checkbox"/>
3	able	<input type="checkbox"/>
4	about	<input type="checkbox"/>
5	above	<input type="checkbox"/>
6	according	<input type="checkbox"/>
7	accordingly	<input type="checkbox"/>
8	across	<input type="checkbox"/>
9	actually	<input type="checkbox"/>
10	after	<input type="checkbox"/>

Exibindo 1 a 10 de 544 palavras

« 1 2 3 4 5 ... 55 »

Adicionar stop word: +

Restaurar stop words Fechar

Pesquisar na tabela

Paginação da tabela

Incluir stopword

Restaurar a listagem padrão de stopwords

Figura 25: 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.

4.2.8 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.

4.3 Reinicialização, backup e restauração

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

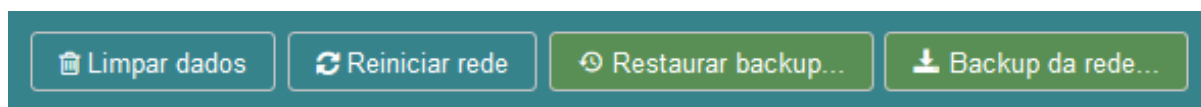


Figura 26: Botões de manutenção da aplicação e da rede

4.3.1 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).

4.3.2 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.

4.3.3 Restaurar backup

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

4.3.4 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.

5.1 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.

5.2 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.

5.2.1 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 (Veja *Abrindo 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.



Figura 27: Carregando da rede via URL - Barra de endereços do navegador

5.2.2 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 umarede.

5.3 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.

5.4 Selecionando nós

5.4.1 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*.

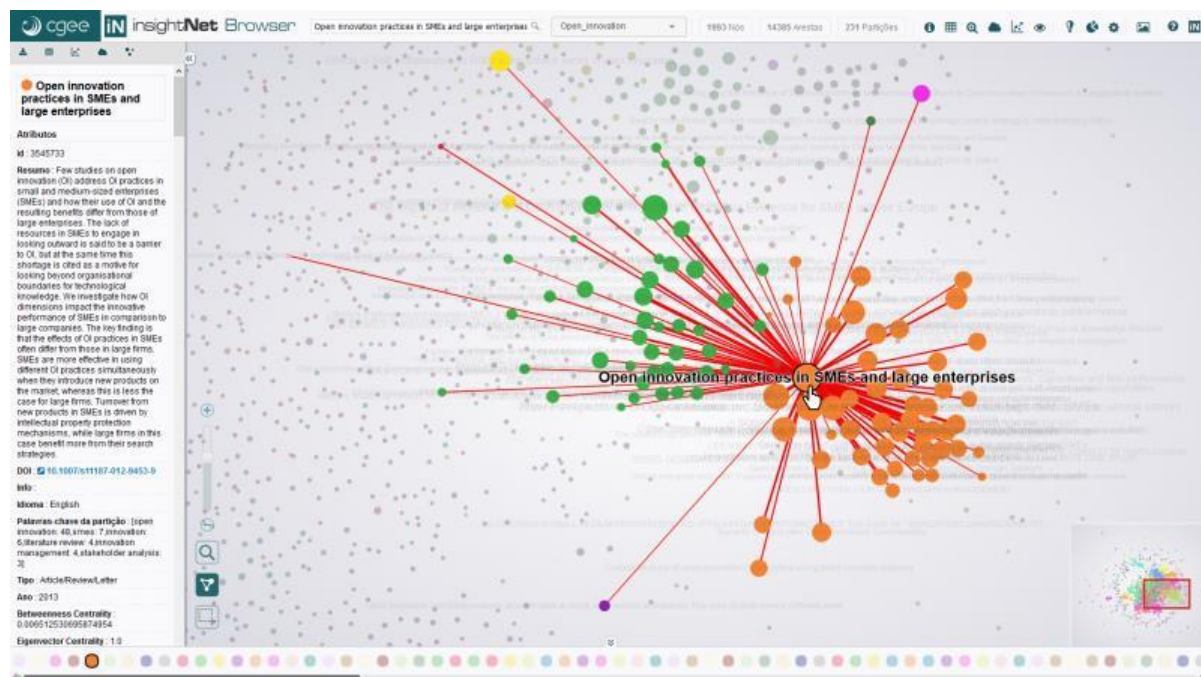




Figura 28: Selecionando um nó e visualizando seus dados na área de detalhamento

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 .

5.4.2 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*.

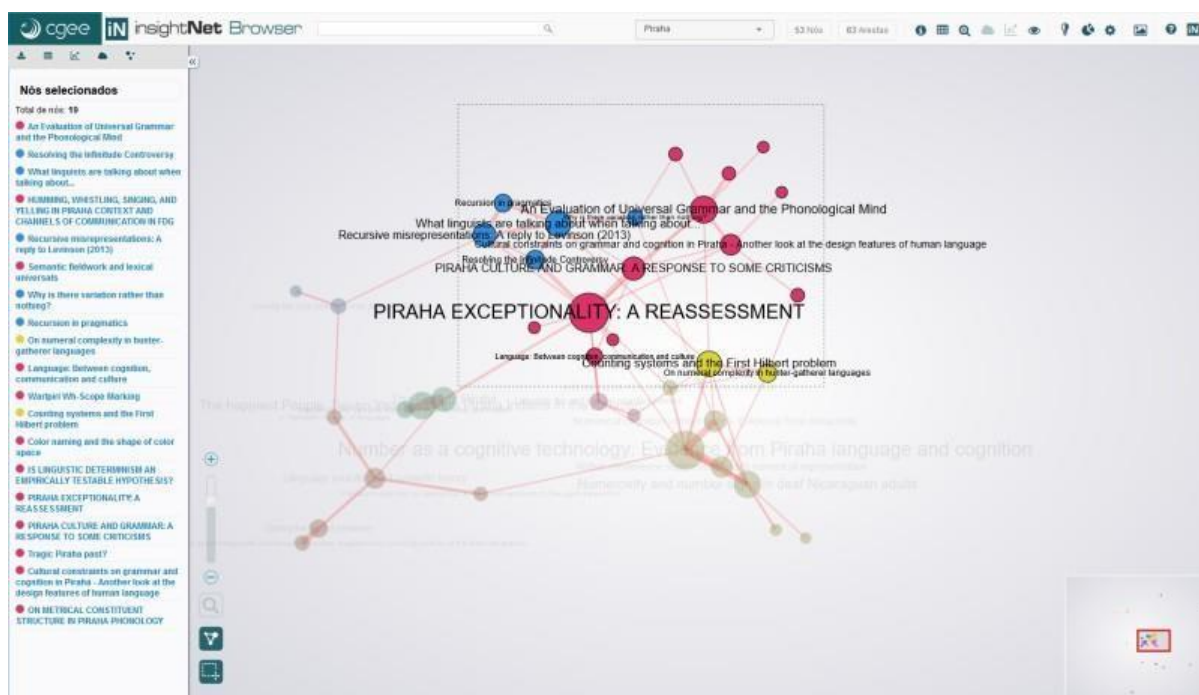


Figura 29: Selecionando um conjunto de nós com a ferramenta de seleção

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

5.4.3 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.

5.5 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*.

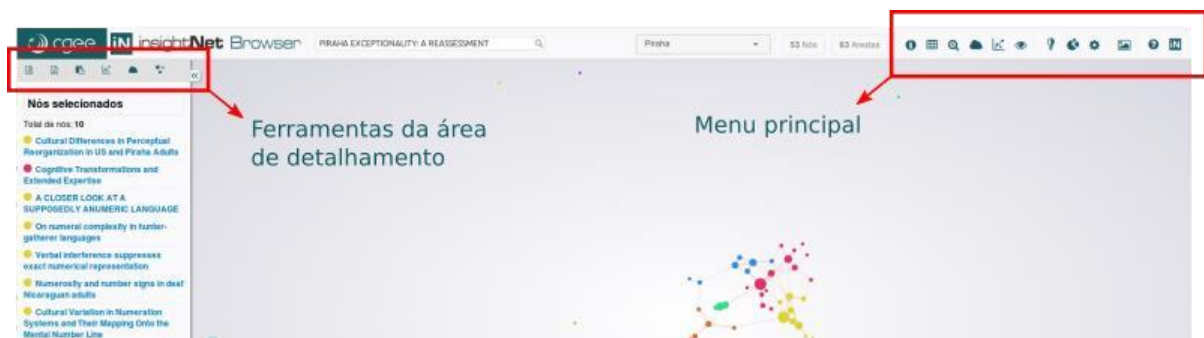



Figura 30: Acessando módulos pelos menus da aplicação

6.1 Metadados da rede

Esse 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.

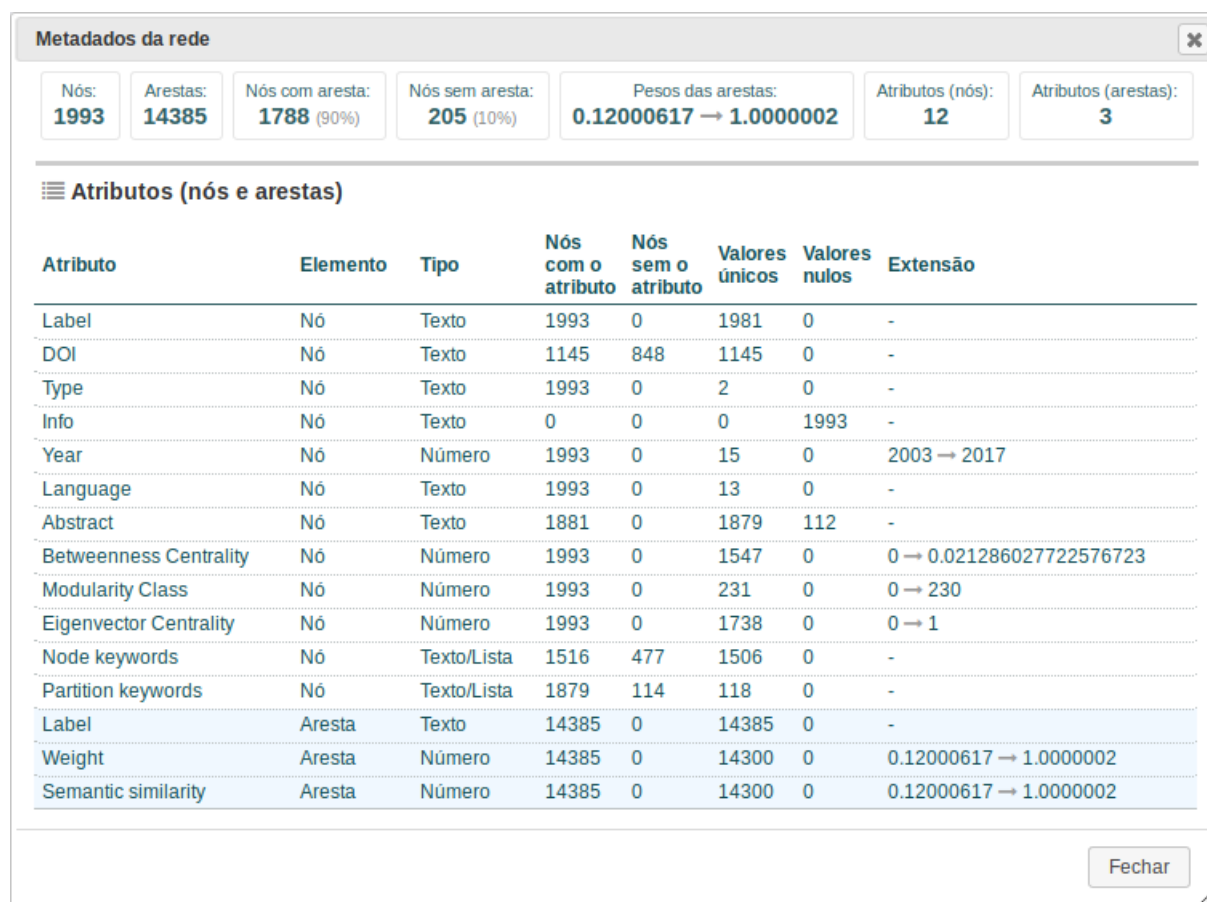


Figura 31: Janela do módulo de Metadados da rede

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.

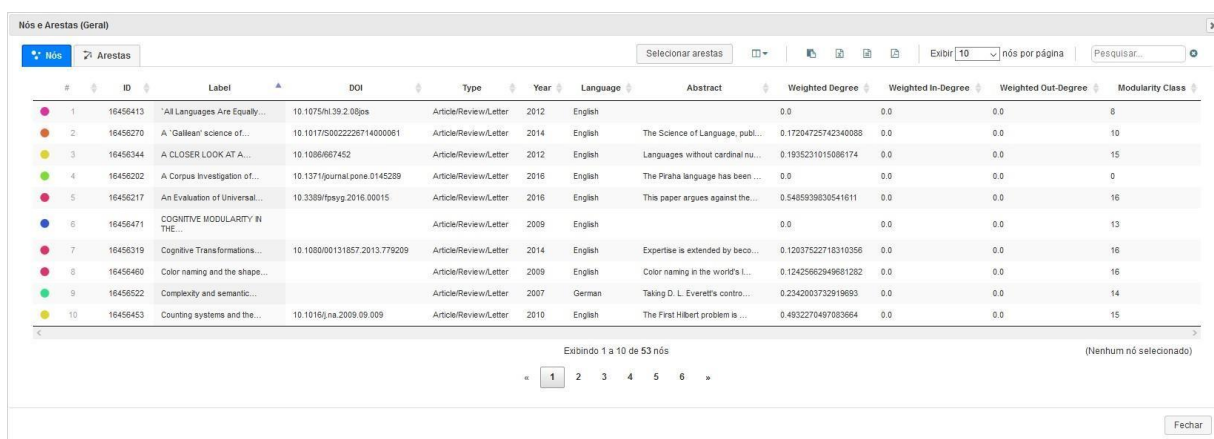
¹ Em alguns casos, alguns nós podem conter um determinado atributo com um valor válido, NULO ou em branco, enquanto em outros nós o mesmo atributo pode ser inexistente.

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.

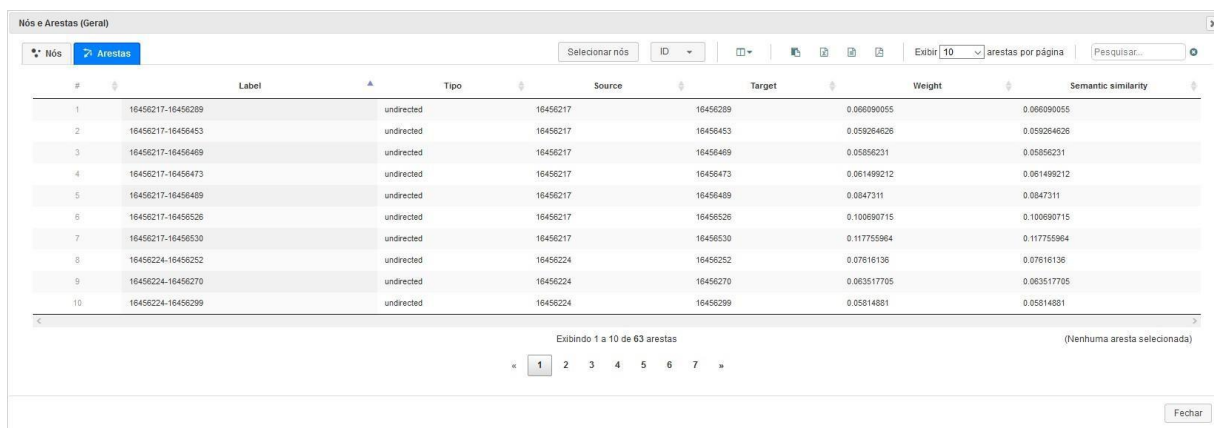
6.2 Tabelas de nós e arestas

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/rl.39.2.08ps	Article/Review/Letter	2012	English		0.0	0.0	0.0	8
2	16456270	A 'Galleian' science of...	10.1017/S0022226714000061	Article/Review/Letter	2014	English	The Science of Language, publ...	0.17204725742340088	0.0	0.0	10
3	16456344	A CLOSER LOOK AT A...	10.1086/667452	Article/Review/Letter	2012	English	Languages without cardinal nu...	0.1935231015086174	0.0	0.0	15
4	16456202	A Corpus Investigation of...	10.1371/journal.pone.0145289	Article/Review/Letter	2016	English	The Piraha language has been ...	0.0	0.0	0.0	0
5	16456217	An Evaluation of Universals...	10.3389/psyg.2016.00015	Article/Review/Letter	2016	English	This paper argues against the...	0.5485939830541611	0.0	0.0	16
6	16456471	COGNITIVE MODULARITY IN THE...		Article/Review/Letter	2009	English		0.0	0.0	0.0	13
7	16456319	Cognitive Transformations...	10.1080/00131857.2013.779209	Article/Review/Letter	2014	English	Expertise is extended by beco...	0.12037522718310358	0.0	0.0	16
8	16456460	Color naming and the shape...		Article/Review/Letter	2009	English	Color naming in the world's L...	0.12425662949681282	0.0	0.0	16
9	16456522	Complexity and semantic...		Article/Review/Letter	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	Article/Review/Letter	2010	English	The First Hilbert problem is ...	0.4932270497083664	0.0	0.0	15

Figura 32: Tabela geral de nós da rede



#	Label	Tipo	Source	Target	Weight	Semantic similarity
1	16456217-16456289	undirected	16456217	16456289	0.086090055	0.086090055
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	0.05614861	0.05614861

Figura 33: Tabela geral de arestas da rede

6.2.1 Selecionando registros

A seleção de nós e arestas nas tabelas irá exibir dados da seleção na *Área de detalhamento*. Para selecionar um registro basta clicar sobre o mesmo e para retirar a seleção basta clicar novamente sobre o

registro. É possível selecionar mais de um registro, consecutivos ou não, seguindo os padrões do sistema opcional: utilizando o clique em conjunto com a tecla CTRL (control) para seleções não consecutivas ou o clique em conjunto com tecla SHIFT para seleções consecutivas.

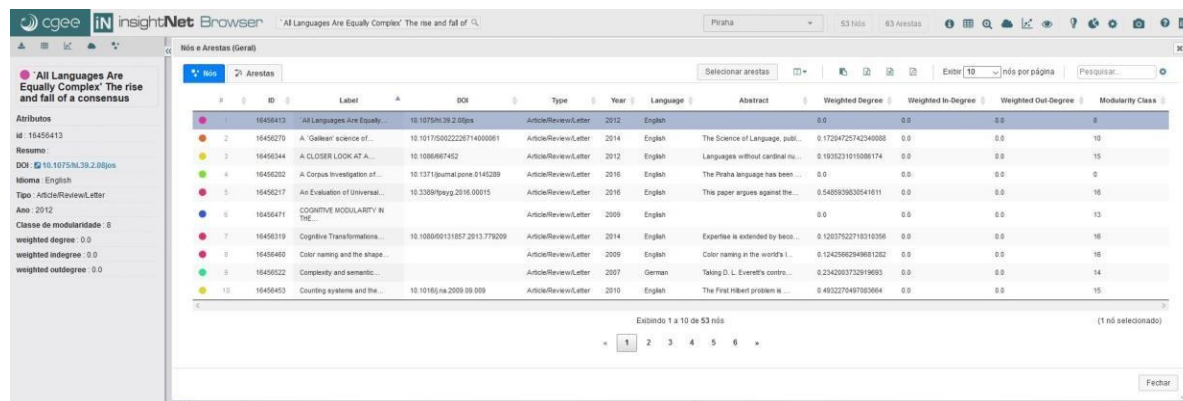


Figura 34: Selecionando um registro na Tabela geral de nós

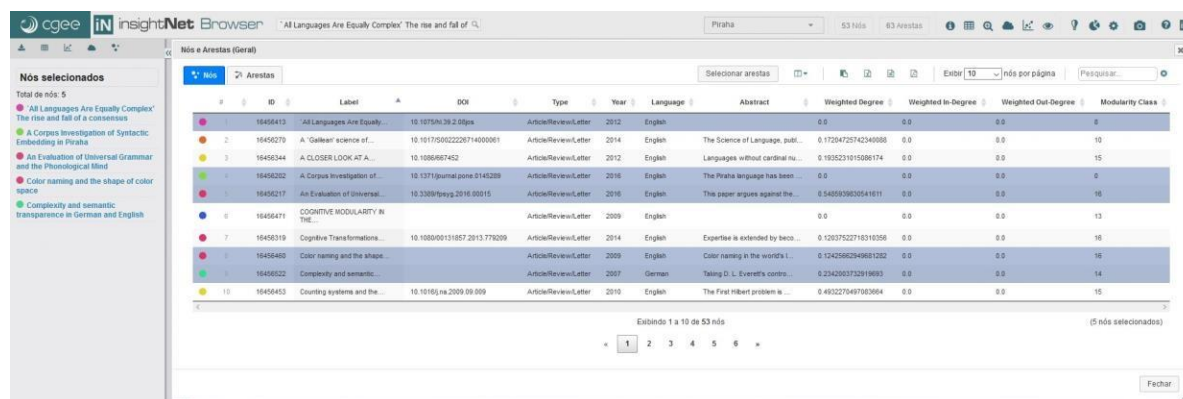


Figura 35: Selecionando vários registros na Tabela geral de nós

6.2.2 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.




Figura 36: Menu de exportação da tabela geral de nós e arestas

Atenção: Apenas colunas visíveis serão incluídas na exportação.

6.2.3 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 .

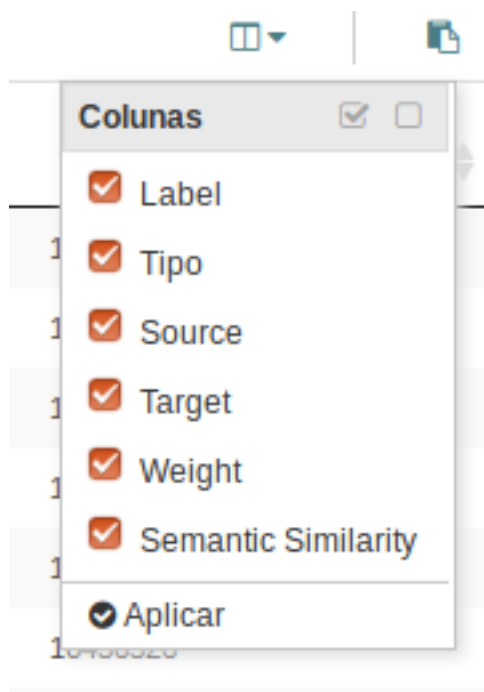



Figura 37: Exibir/Ocultar colunas da tabela geral de nós e arestas

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.

6.2.4 Tabela de nós

Selecionar arestas

Essa funcionalidade irá ajudar a localizar, na tabela de arestas, as arestas relacionadas a um ou mais nós selecionados na tabela de nós.

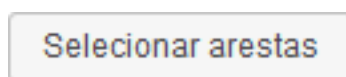


Figura 38: Botão para destacar, na tabela de arestas, as arestas dos nós selecionados na tabela de nós

Basta selecionar um ou mais nós e clicar no botão **Selecionar arestas**. Você será redirecionado para a tabela de arestas e as arestas relacionadas aos nós escolhidos estarão selecionadas e posicionadas nas primeiras posições da tabela.

6.2.5 Tabela de arestas

Selecionar nós

Essa funcionalidade irá ajudar a localizar, na tabela de nós, os nós relacionados a uma ou mais arestas selecionadas na tabela de arestas.

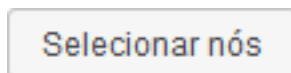


Figura 39: Botão para destacar, na tabela de nós, os nós das arestas selecionadas na tabela de arestas

Basta selecionar uma ou mais arestas e clicar no botão **Selecionar nós**. Você será redirecionado para a tabela de nós e os nós relacionados às arestas escolhidas estarão selecionados e posicionados nas primeiras posições da tabela.

Selecionar atributos para exibição

Essa funcionalidade permite que você escolha qual atributo dos nós será exibido nas colunas **source** e **target** das arestas. Facilitando a identificação dos nós ou provendo uma tabela de relações para um determinado atributo do nó.

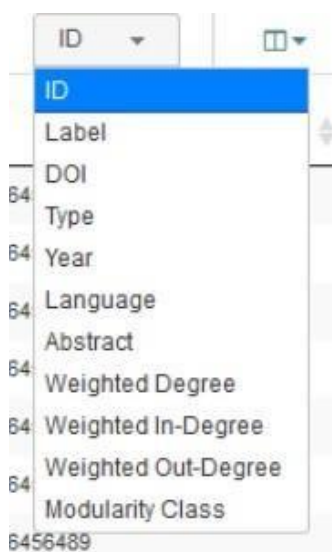



Figura 40: Seleção de atributo dos nós a ser exibido na tabela de arestas

#	Label	Tipo	Source	Target	Weight	Semantic similarity
1	16456353-16456460	undirected	2012	2009	0.05922897	0.05922897
2	16456460-16456526	undirected	2009	2005	0.06442706	0.06442706
3	16456217-16456289	undirected	2016	2014	0.06099055	0.06099055
4	16456217-16456453	undirected	2016	2010	0.05254626	0.05254626
5	16456217-16456469	undirected	2016	2009	0.05856231	0.05856231
6	16456217-16456473	undirected	2016	2009	0.061499212	0.061499212
7	16456217-16456489	undirected	2016	2009	0.0847311	0.0847311
8	16456217-16456526	undirected	2016	2005	0.100890715	0.100890715
9	16456217-16456530	undirected	2016	1988	0.117755964	0.117755964
10	16456224-16456252	undirected	2015	2015	0.07616136	0.07616136

Figura 41: Exemplo de atributo selecionado para exibição na tabela de arestas

6.3 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.

Figura 42: Janela do módulo de Pesquisa avançada

6.3.1 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.

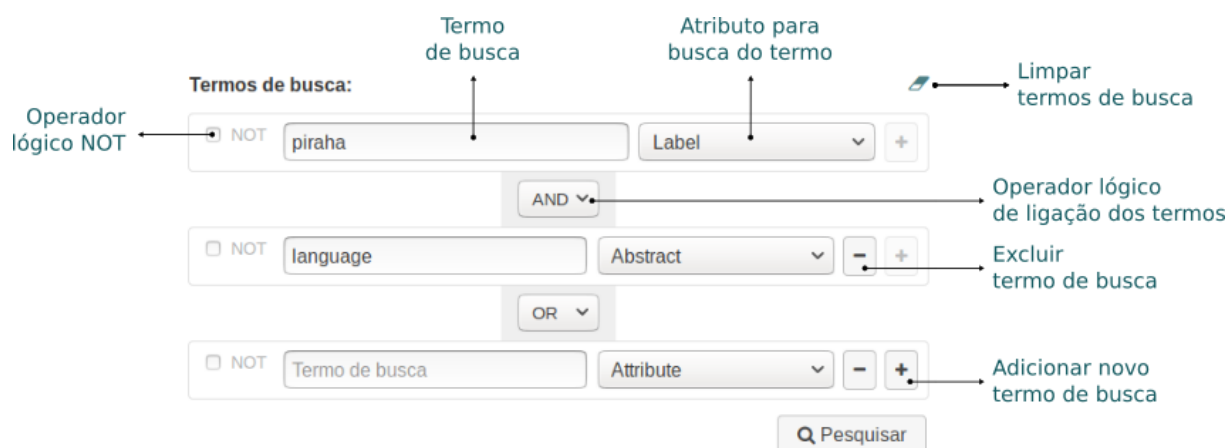


Figura 43: Área dos termos de busca

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:

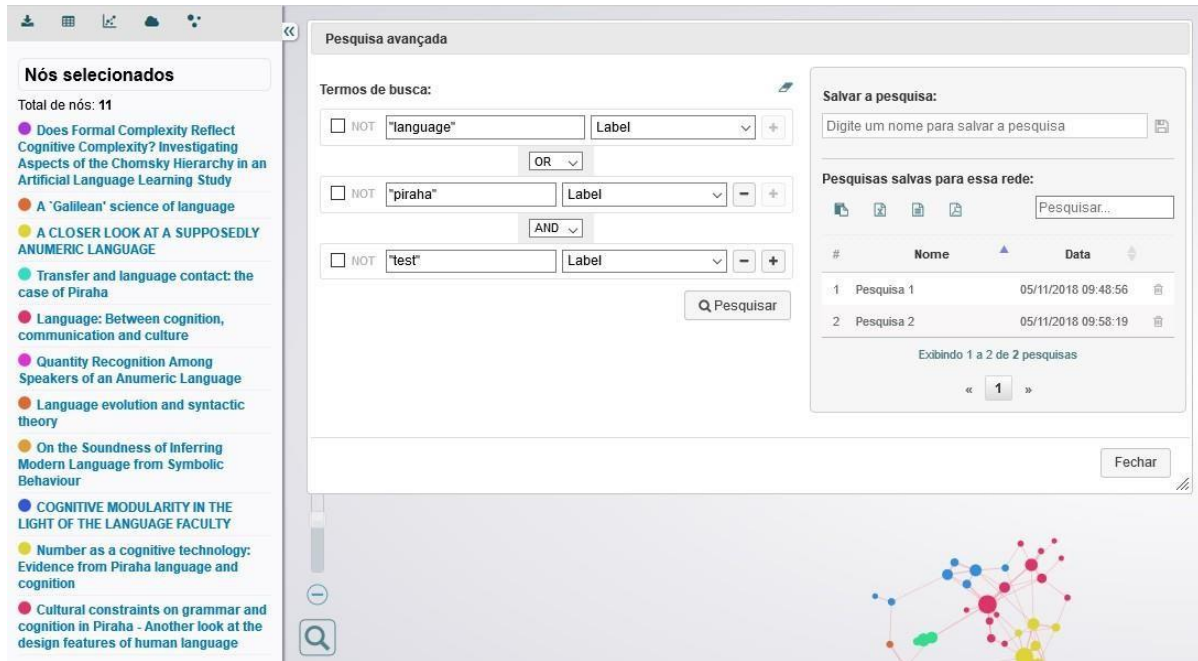


Figura 44: Executando uma busca na Pesquisa Avançada

O resultado dessa busca serão os 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:

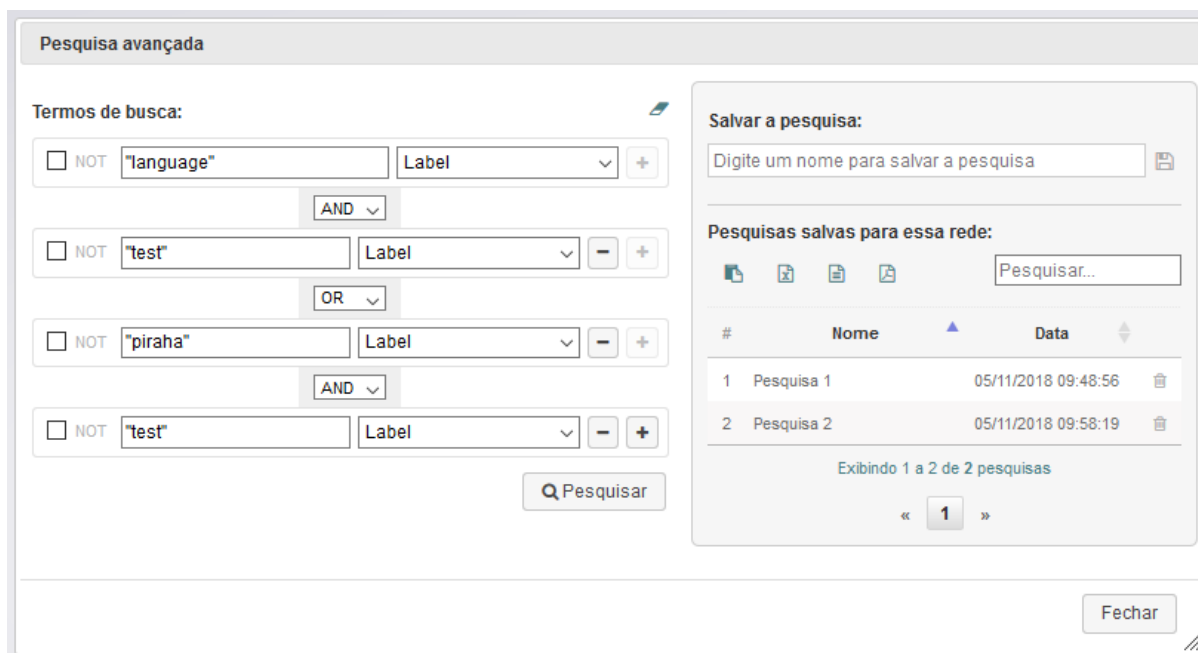


Figura 45: Reestruturando a busca na Pesquisa Avançada

6.3.2 Gerenciamento de pesquisas

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

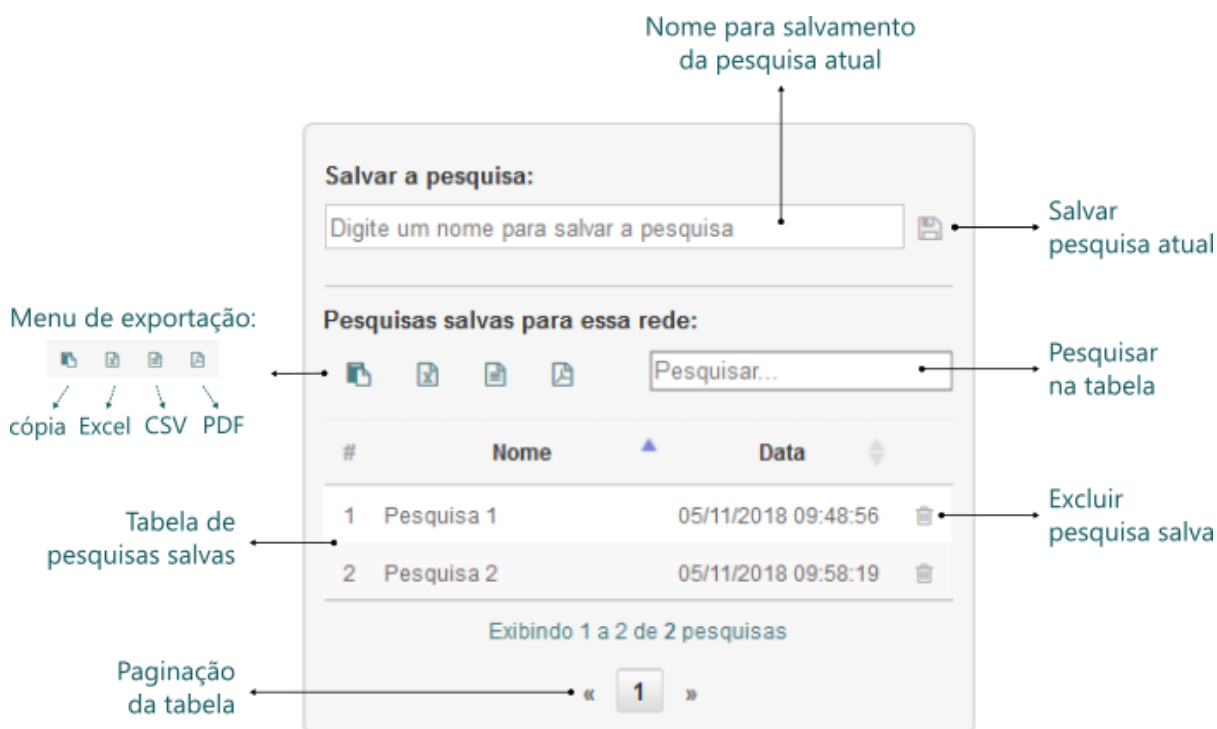




Figura 46: Área de gerenciamento de pesquisas

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*.

6.4 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.

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.

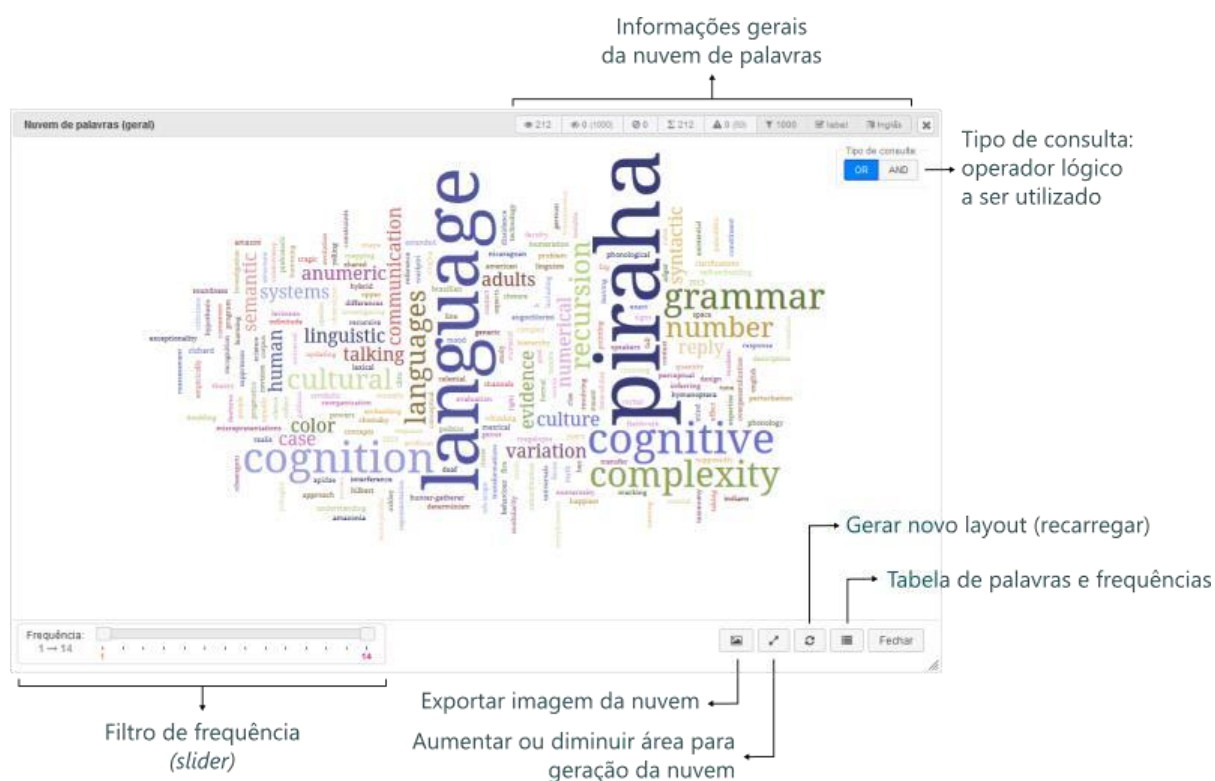


Figura 47: Módulo Nuvem de palavras insightNet Browser

6.4.1 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.

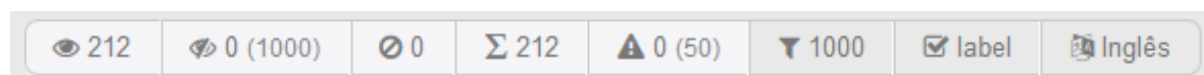











Figura 48: Informações gerais da Nuvem de palavras

	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.

6.4.2 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.

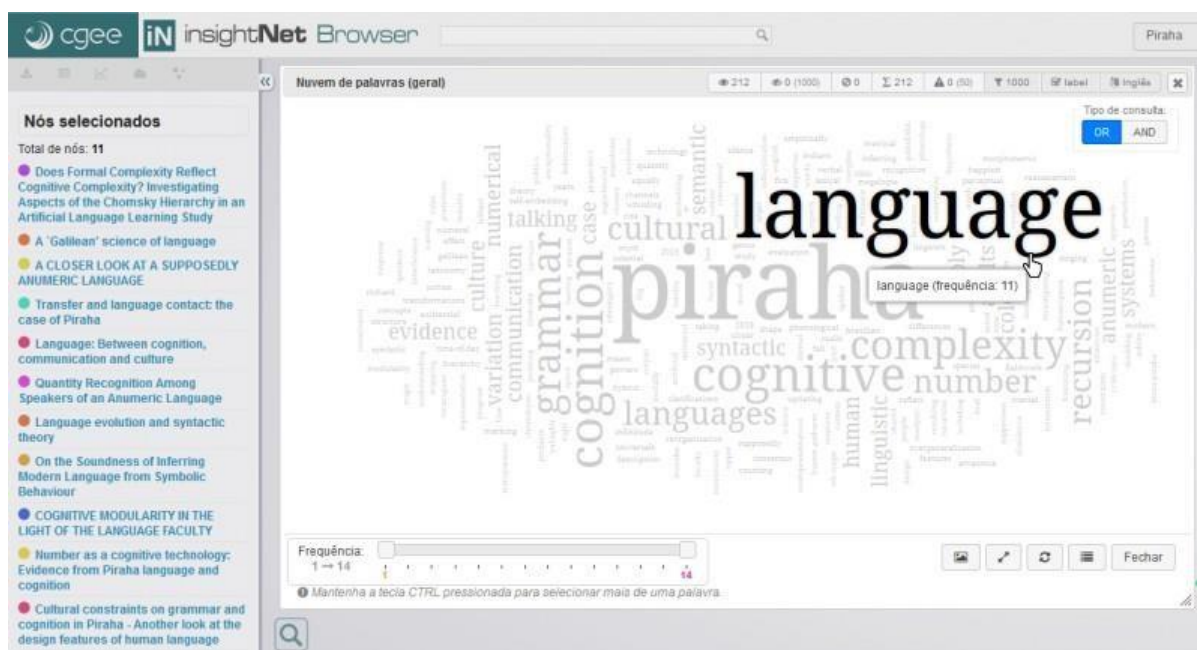


Figura 49: Selecionando uma palavra e consultando os nós relacionados

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:



Figura 50: Utilizando os operadores lógicos nas nuvens de palavras

6.4.3 Filtro de frequência

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

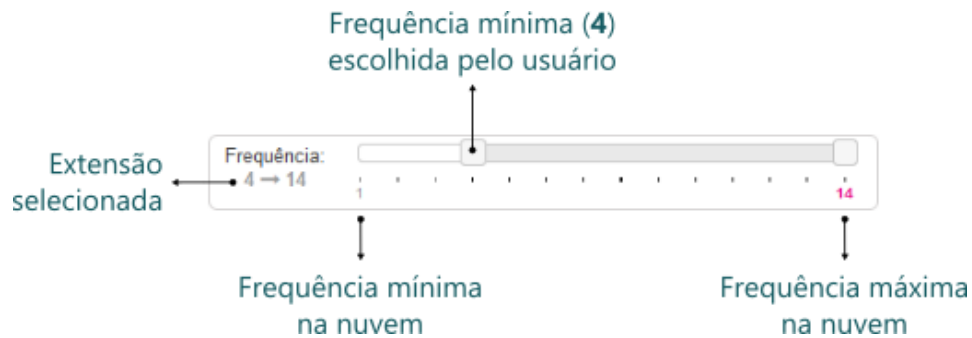



Figura 51: Filtro de frequências da nuvem de palavras

6.4.4 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.

Menu de exportação:

Pesquisa na tabela:
 Menu auxiliar de marcação:
 Ocultar/exibir palavra na nuvem:
 Aplicar alterações:

#	Palavra	Frequência	
1	piraha	14	<input checked="" type="checkbox"/>
2	language	11	<input checked="" type="checkbox"/>
3	cognitive	5	<input checked="" type="checkbox"/>
4	cognition	5	<input checked="" type="checkbox"/>
5	grammar	4	<input checked="" type="checkbox"/>
6	complexity	4	<input checked="" type="checkbox"/>
7	recursion	3	<input checked="" type="checkbox"/>
8	cultural	3	<input checked="" type="checkbox"/>
9	languages	3	<input checked="" type="checkbox"/>
10	number	3	<input checked="" type="checkbox"/>

Exibindo 1 a 10 de 212 palavras

Paginação da tabela: « 1 2 3 4 5 ... 22 »

Figura 52: 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.

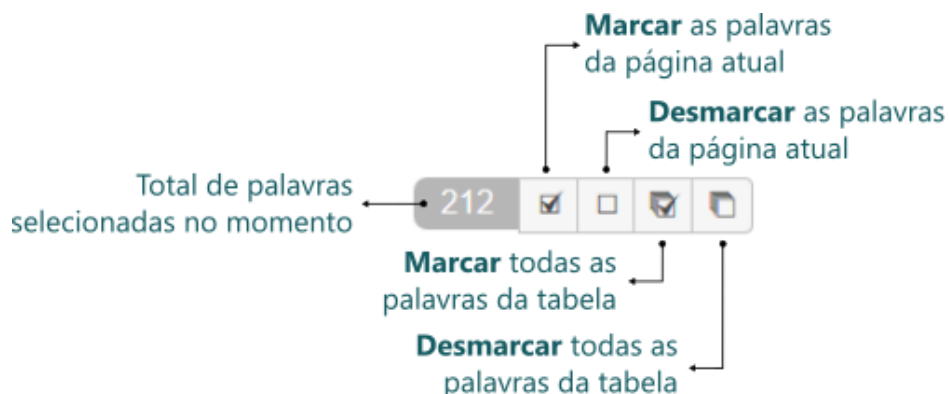



Figura 53: Menu auxiliar de marcação da tabela de palavras e frequências

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.


6.4.5 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](#).

6.5 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*.



Figura 54: 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.

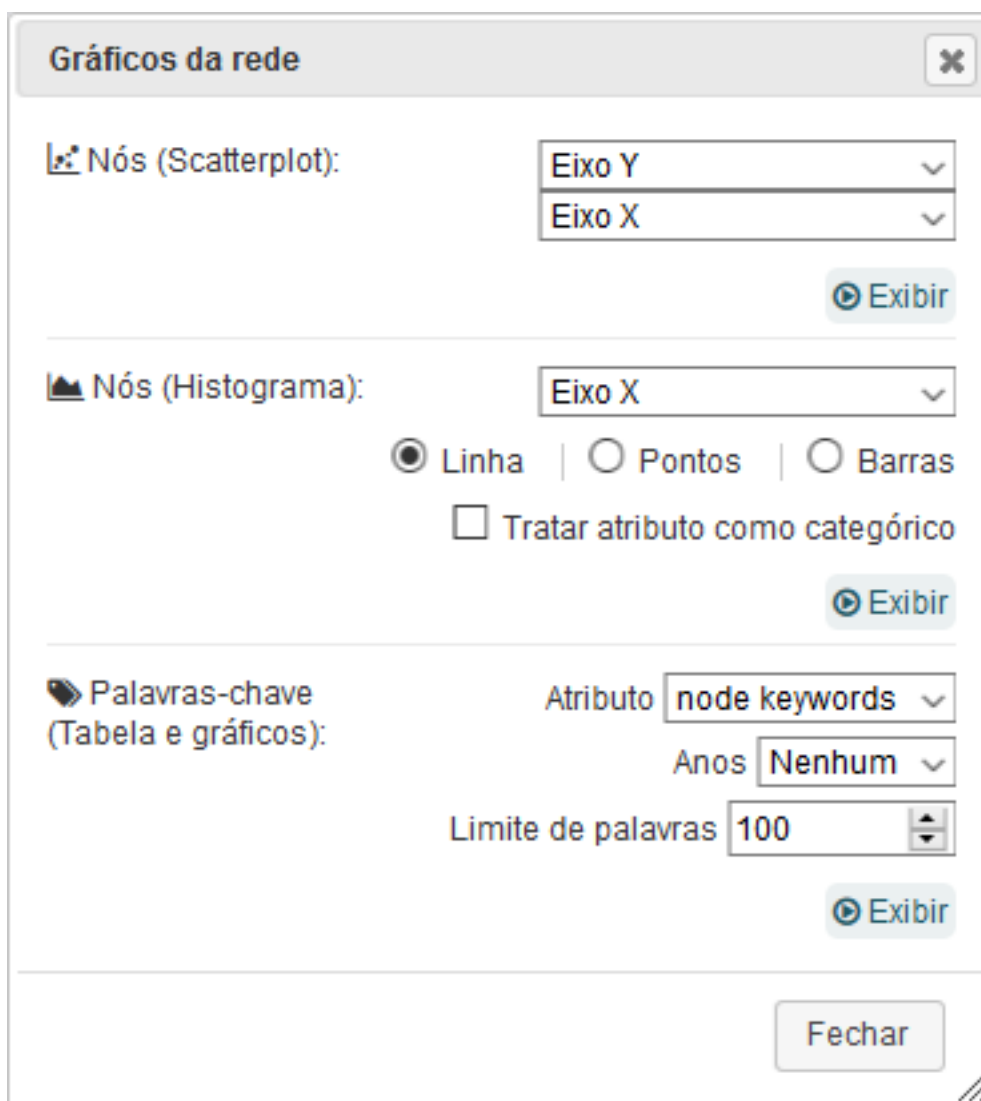


Figura 55: Menu de gráficos da rede

Cada tipo de gráfico ou visualização é gerado por um módulo diferente da aplicação. São 4 módulos de visualização de dados e iremos detalhar cada um deles nesse capítulo. São eles: **Scatterplot**, **Histograma** e **Palavras-chave** (com outras opções de gráficos dentro do módulo).

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**:

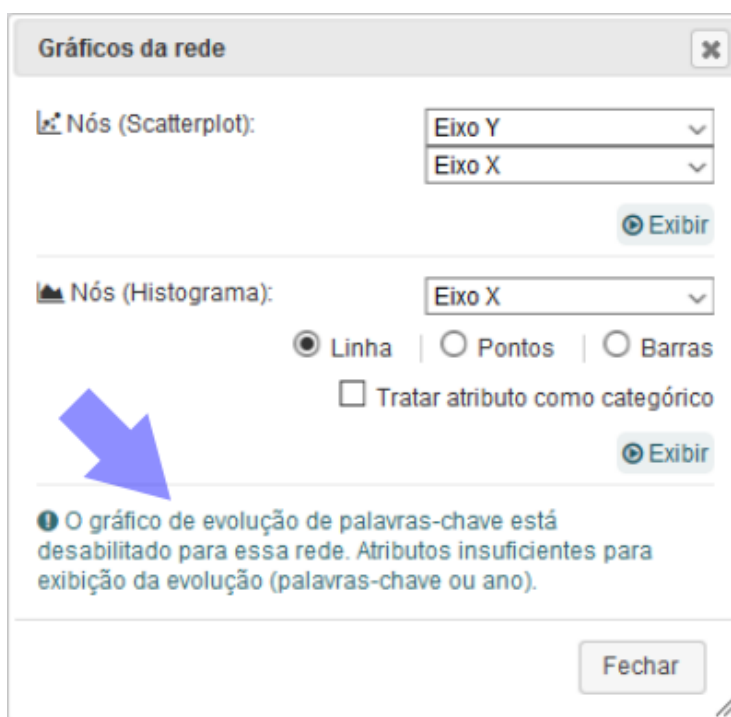


Figura 56: Mensagem quando não há visualizações para palavras-chave

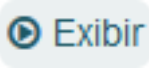
6.5.1 Scatterplot

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.



Figura 57: Escolhendo os eixos do gráfico scatterplot

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

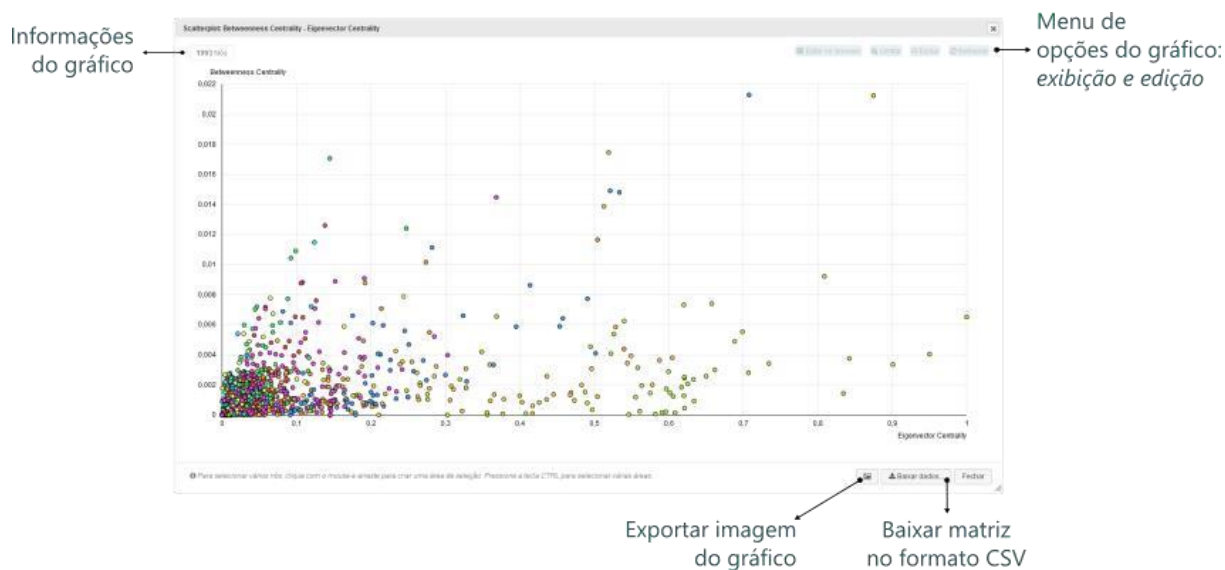


Figura 58: Gráfico scatterplot do insightNet Browser

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.

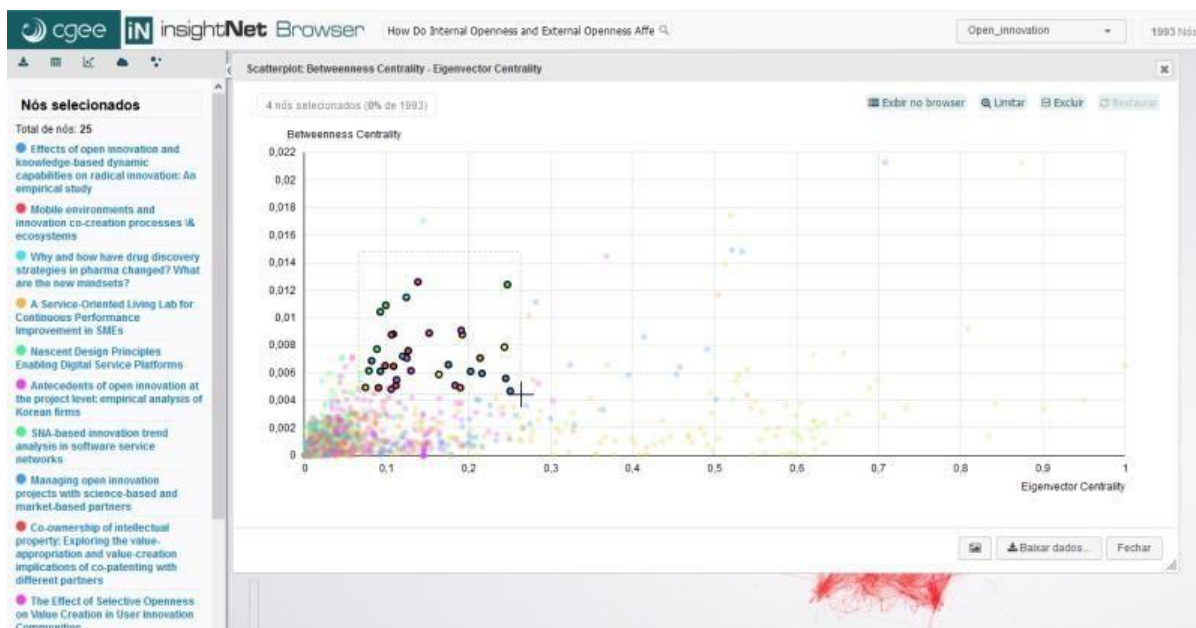


Figura 59: Selecionando pontos no gráfico scatterplot

Menu de opções (exibindo ou editando)

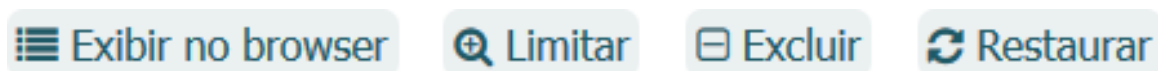


Figura 60: Menu de opções do gráfico scatterplot (exibir ou editar)

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.

6.5.2 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.

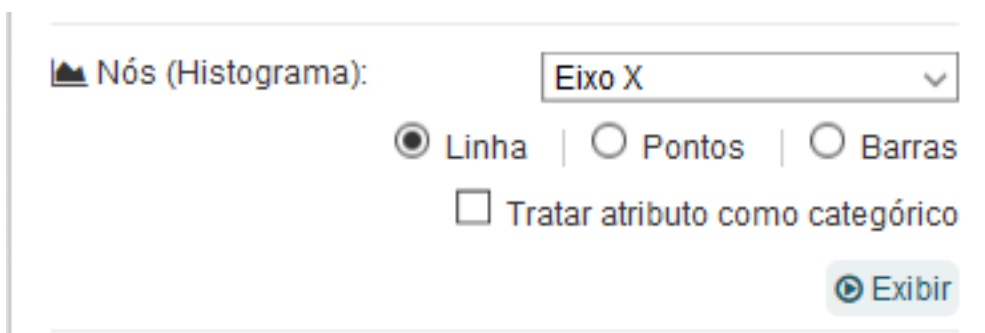



Figura 61: Escolhendo as opções para geração do histograma

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 e obter mais opções de ordenação.

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.

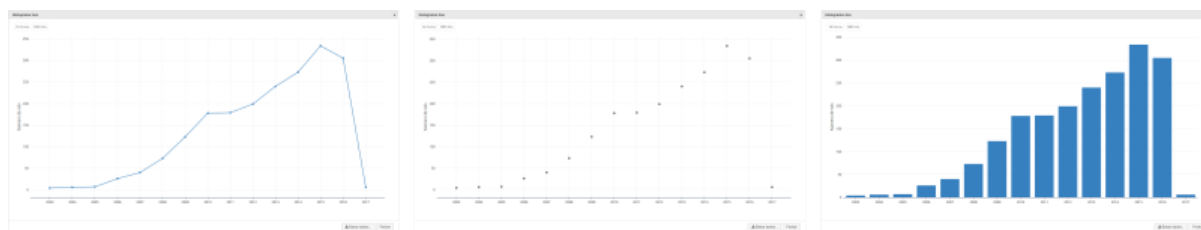


Figura 62: Opções de gráficos histograma: linha, pontos e barras.

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

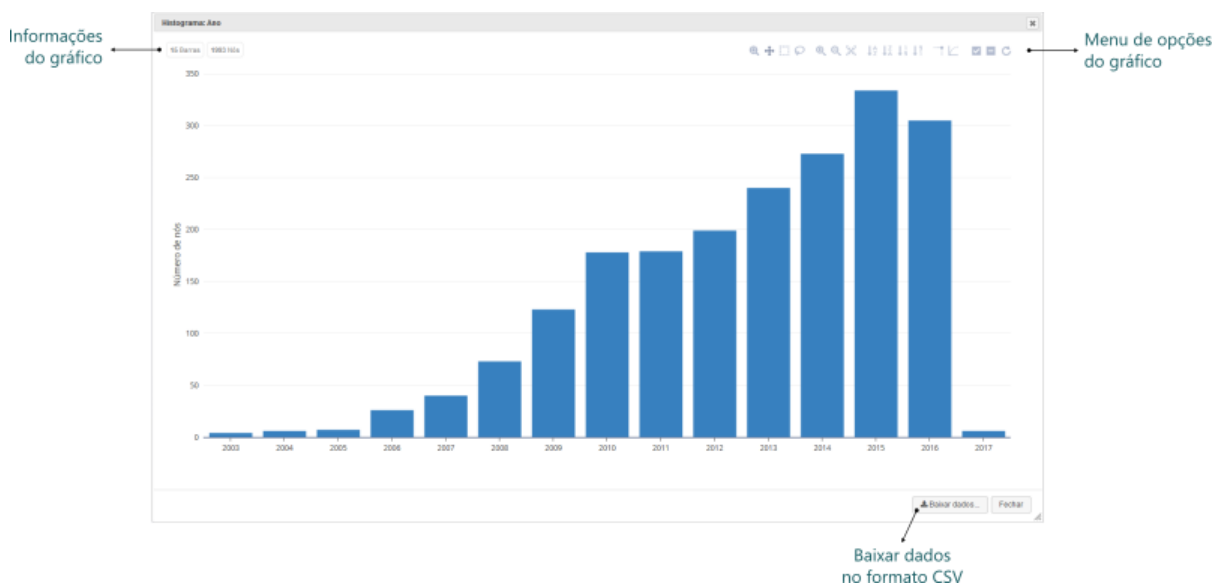


Figura 63: Gráfico histograma do insightNet Browser

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 um mesmo valor para o atributo escolhido (eixo X).

Selecionando pontos

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 opções 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.

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

Ao selecionar um ou mais barras ou pontos, informações quantitativas serão exibidas na **área de informações** do gráfico.

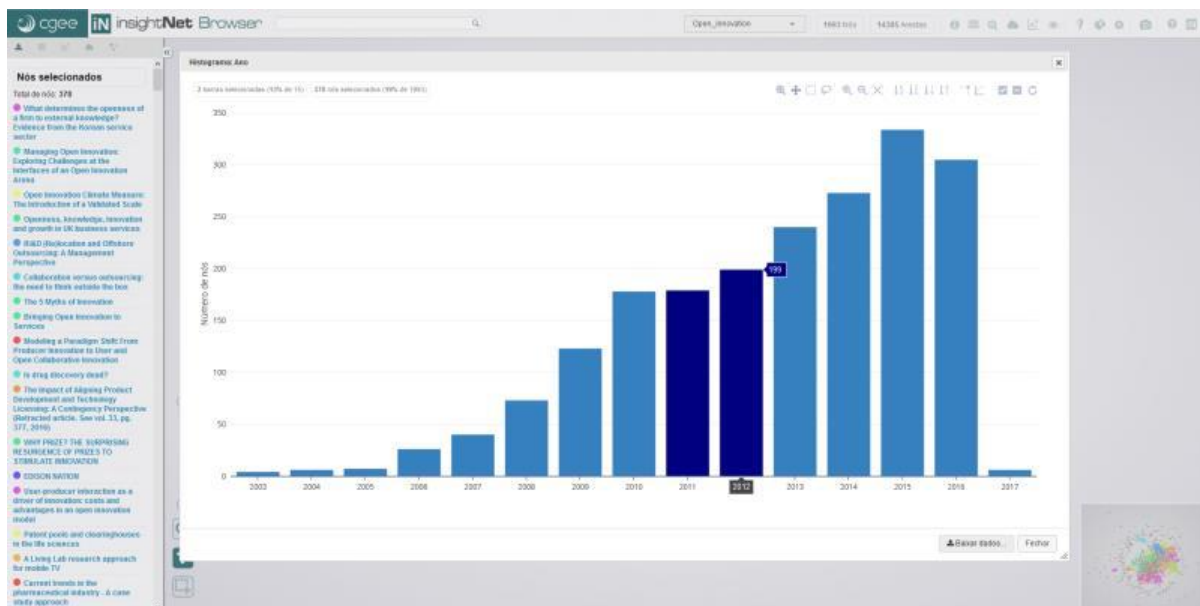


Figura 64: Selecionando barras no gráfico histograma

Menu de opções



Figura 65: Menu de opções do gráfico histograma

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.

6.5.3 Palavras-chave

Esse módulo apresenta uma tabela de ocorrências de palavras-chave (por ano e/ou total) e alguns gráficos: **Histograma**, **Evolução ou Ocorrências** e **Co-ocorrências** de palavras-chave.

Antes de entrar no módulo você deverá informar qual atributo de palavras-chave você irá visualizar, o atributo utilizado para o ano, caso queira visualizar as tabela e gráfico de evolução e um limite de palavras-chave a exibir. Um número muito grande de palavras-chave irá tornar a geração de tabela e gráficos mais lenta.

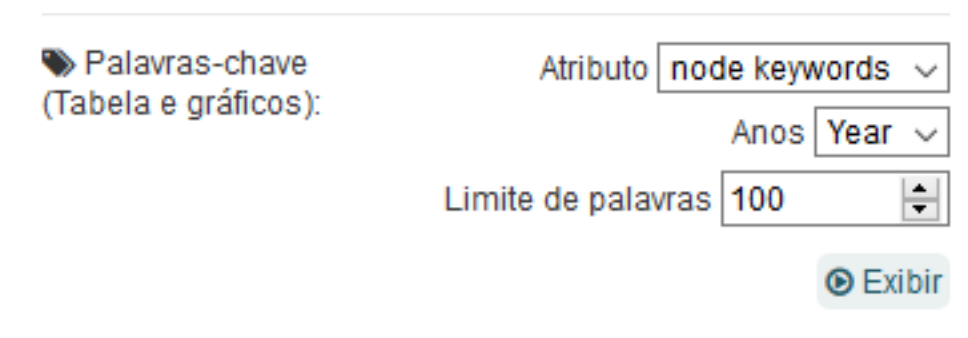


Figura 66: Escolhendo as opções para geração de gráficos de palavras-chaves

Atenção: Essa funcionalidade depende da existência de atributos específicos para palavras-chave, com formatos pré-definidos. Além disso esses atributos devem ter nome e estrutura reconhecidos pelo insightNet Browser. Os nomes utilizados são: **Node keywords** ou **node keywords** e o formato desse atributo, no arquivo GEXF, deve obedecer o padrão do exemplo abaixo:

```
<attvalue for="node keywords" value="[open: 2,museums: 1,user and smart innovation: 1,smart innovation: 1]"/>
```

Se você tem em sua rede atributos com palavras-chave, mas esse módulo continua indisponível no *menu de gráficos*, possivelmente o atributo está com nome ou formato não reconhecidos pela aplicação.

Tabela de evolução ou ocorrências

Depois de escolher as opções desejadas e clicar no botão



uma das duas opções de tabelas será exibida:

A **Tabela de evolução de palavras-chave** será exibida sempre que um atributo **Ano** estiver presente.

#	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)

Histograma Gráfico de evolução Co-ocorrências (experimental) Fechar

Figura 67: Tabela de evolução de palavras-chave

Atenção: Para que a **Tabela de evolução de palavras-chave** seja exibida é necessário que os nós tenham um atributo que represente o ano. Esse atributo deve estar nomeado com uma das opções a seguir: “Year”, “year”, “Ano”, “ano” ou “Highest Title: End Year”.

A **Tabela de ocorrências de palavras-chave** será exibida quando não houver um atributo **Ano**.

#	Palavra-chave	total
1	open innovation	923
2	innovation	149
3	crowdsourcing	63
4	absorptive capacity	48
5	collaboration	48
6	knowledge management	43
7	intellectual property	37
8	innovation management	36
9	co-creation	35
10	smes	33

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

Histograma Gráfico de evolução Co-ocorrências (experimental) Fechar

Figura 68: Tabela de ocorrências de palavras-chave

Dica: Ao clicar em uma linha da tabela, além de selecionar a palavra-chave, os nós que contém a palavra-chave serão exibidos na *Área de detalhamento*.

Colunas geradas

Na **Tabela de evolução de palavras-chave** são geradas três colunas para cada ano encontrado nos registros: **ano**, **ano/freq** e **ano/total**. As colunas com os sufixos **/freq** e **/total** iniciam ocultas por

padrão, mas podem ser exibidas, permitindo a exportação dos dados e geração de outros gráficos em outras aplicações.

#	Palavra-chave	2007	2007/freq	2007/total	2008	2008/freq	2008/total	2009
1	alzheimer's disease	40	40	124	53	53	158	79
2	alzheimers-disease	13	13	124	21	21	158	32
3	oxidative stress	8	8	124	13	18	158	17
4	amyloid precursor protein	16	17	124	24	24	158	30
5	a-beta	10	10	124	17	17	158	21
6	transgenic mice	15	15	124	17	17	158	19
7	in-vivo	10	10	124	20	20	158	16

Figura 69: Colunas geradas na Tabela de evolução de palavras-chave

A primeira coluna, com o **ano** no título, é o total de nós onde ocorreram a palavra-chave.

A segunda coluna, com o ano e sufixo **/freq** no título, é uma soma das frequências das palavras-chave, que podem ocorrer mais de uma vez por nó. Esse caso ocorre nas redes geradas de currículos Lattes, onde os termos podem ocorrer em um ou mais artigos do currículo (nó).

A terceira coluna, com o ano e sufixo **/total** no título, é o total de nós no ano, não importando se há ou não a palavra-chave.

Na **Tabela de ocorrências de palavras-chave** é gerada apenas a coluna **total** com a quantidade de nós nos quais o termo foi encontrado.

Atenção: Em ambos os tipos de tabela a última coluna, chamada **total**, representa a quantidade de nós nos quais o termo foi encontrado e não a frequência do termo.

Selecionando registros

A seleção de palavras-chave na tabela irá exibir dados dos nós que contém os termos selecionados na *Área de detalhamento*. Para selecionar um registro basta clicar sobre o mesmo e para retirar a seleção basta clicar novamente sobre o registro. É possível selecionar mais de um registro, consecutivos ou não, seguindo os padrões do sistema opcional: utilizando o clique em conjunto com a tecla CTRL (control) para seleções não consecutivas ou o clique em conjunto com tecla SHIFT para seleções consecutivas.

#	Palavra-chave	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	total
91	performance	0	0	0	0	0	0	1	0	1	0	3	0	2	0	0	7
92	knowledge networks	0	0	0	0	0	1	0	0	0	1	0	2	3	0	0	7
93	virtual communities	0	0	0	0	0	0	3	0	1	0	0	3	0	0	0	7
94	strategic alliances	0	0	0	1	1	0	1	3	1	0	0	0	0	0	0	7
95	high-tech industries	0	0	0	0	0	0	0	0	0	0	0	1	3	2	0	6
96	disruptive innovation	0	0	0	0	0	0	0	0	1	0	2	0	1	2	0	6
97	manufacturing	0	0	0	0	0	0	1	0	0	0	1	1	0	3	0	6
98	appropriability	0	0	0	0	0	0	0	2	1	1	0	0	0	2	0	6
99	services	0	0	0	0	1	1	0	1	1	0	0	1	0	1	0	6
100	corporate venture capital	0	0	0	0	0	0	0	0	1	1	2	0	0	2	0	6

Figura 70: Selecionando um registro na Tabela de evolução de palavras-chave

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.



Figura 71: Menu de exportação da tabela

Atenção: Apenas colunas visíveis serão incluídas na exportação.

Exibir/Ocultar colunas

Como o número de colunas (anos) é variável na **Tabela de evolução de palavras-chave**, você pode ocultar algumas colunas para melhorar a visualização da tabela. Para isso utilize o botão .

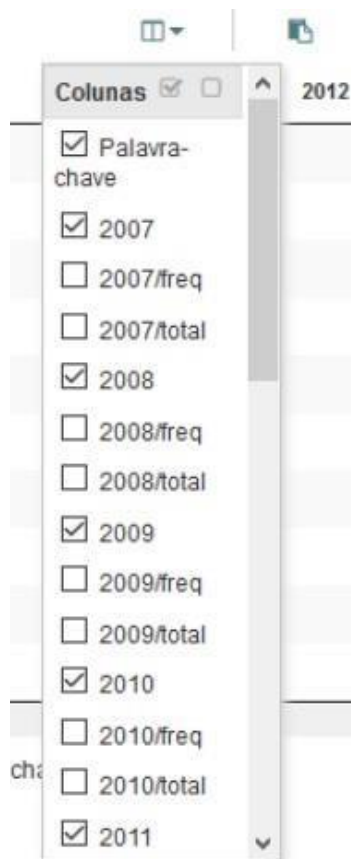


Figura 72: Exibir/Ocultar colunas das tabelas de evolução ou ocorrências

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.

Gráficos

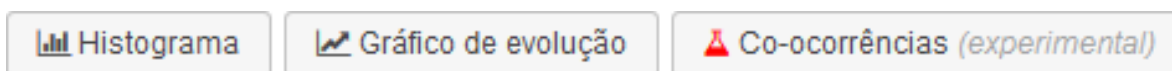



Figura 73: Menu de gráficos das tabelas de evolução ou ocorrências

Histograma

O histograma de palavras-chave exibe um gráfico de barras com os totais dos nós onde as palavras selecionadas ocorreram. O gráfico permite a comparação de várias palavras-chave, onde cada barra representa uma palavra-chave.

Antes de gerar o histograma você deverá selecionar as palavras-chave que ocuparão o eixo X do gráfico e clicar no botão  Histograma.

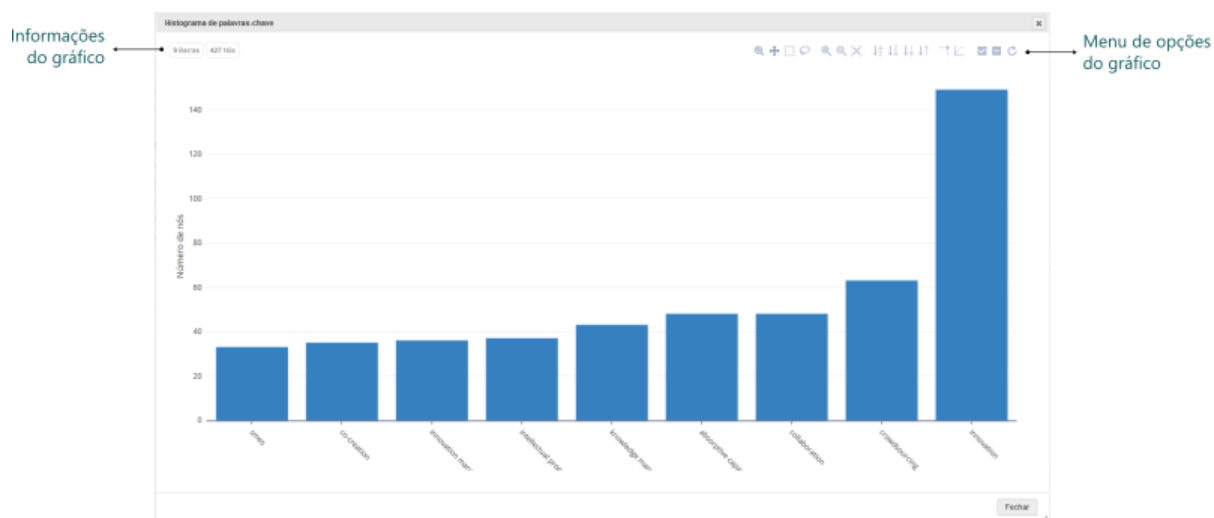


Figura 74: Gráfico histograma de palavras-chave

Evolução

Em desenvolvimento. . .



Co-ocorrências

Em desenvolvimento. . .

6.6 Visualizações

Em desenvolvimento. . .

6.6.1 Gerenciando visualizações

Em desenvolvimento. . .

6.6.2 Modo de apresentação

Em desenvolvimento. . .

6.7 Partições

Em desenvolvimento. . .