Passos para a reprodução dos resultados da pesquisa

Mesas separadas: divisões temáticas e metodológicas da Ciência Política Brasileira

Nilton Sainz

Adriano Codato

Rodrigo da Silva

Augusto Clemente

ABSTRACT:

Introdução: Descrevemos a configuração temática e metodológica da Ciência Política brasileira. Através do estudo da dinâmica relacional entre diferentes agendas de pesquisa e abordagens específicas a cada uma delas, identificamos a centralidade de determinadas temáticas e a densidade metodológica da Ciência Política brasileira. Materiais/Métodos: Foram examinadas 1.849 teses e dissertações defendidas entre 2013 e 2020 em onze Programas de Pós-Graduação. Analisamos títulos, resumos e palavras-chave desses trabalhos. Os dados foram coletados no Catálogo de Teses e Dissertações da CAPES. Análises: A rede temática mostrou a emancipação da área e a consolidação de cinco comunidades diferentes. Elas cobrem os grandes temas canônicos da Ciência Política mainstream: 1) teoria política democrática, 2) Estado, governo e relações internacionais, 3) atores sociais, participação e políticas públicas, 4) instituições políticas e poderes do Estado e 5) partidos, eleições e estudos legislativos. A rede metodológica revelou uma estrutura semântica organizada a partir dessa divisão onde cada comunidade tende a possuir seus próprios estilos, métodos e técnicas de pesquisa. Discussão: Este achado é consistente com avaliações que enfatizam a institucionalização da Ciência Política brasileira. Essa institucionalização se expressa na predominância de temas de pesquisa ligados especificamente à política (e não à economia, à sociedade etc.), e maior aprimoramento metodológico.

Keywords: Ciência Política, Ciência Política Brasileira, Teses e dissertações, Pós-graduação, Análise de redes sociais.

Replicação dos procedimentos de Análises de Redes

O mesmo procedimento deverá ser realizado para ambas as redes (Temática e Metodológica)

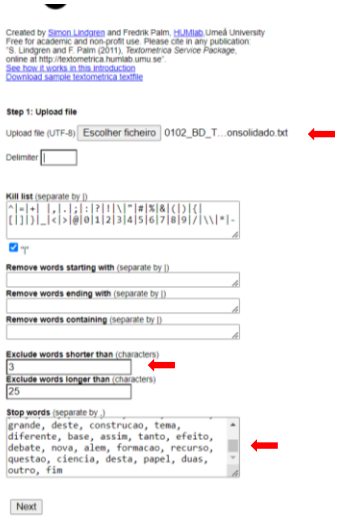
Lematização e construção das redes

1) No site http://textometrica.humlab.umu.se/ realiza-se o upload do arquivo "0102_BD_T&D_T-PC-R_2013-2020_Consolidado.txt”

2) No espaço para “stop words” deverá ser inserido a seguinte lista de palavras:

a, e, o, do, da, de, em, na, no, e, ne, para, pra, um, uma, uns, umas, as, os, das, dos, nas, nos, por, com, ao, aos, naquele, naquela, naquilo, outro, outra, outros, outras, naqueles, naquelas, isso, esse, essa, esses, essas, este, esta, estes, estas, ele, ela, eles, elas, nos, pelo, pelos, pela, pelas, mas, esta, estao, seu, sua, seus, suas, teus, tuas, teu, tua, ou, sao, se, sobre, sem, nao, sim, que, entao, acho, acha, ante, apos, ate, contra, desde, entre, per, perante, sob, tras, num, numa, nuns, numas, so, a, nem, tambem, como, ainda, porem, todavia, contudo, antes, depois, obstante, apesar, desse, dessa, disso, desses, dessas, quer, logo, pois, porque, qual, quais, mais, menos, quanto, quando, quanta, quantos, quantas, tanto, tanta, tantos, tantas, embora, mesmo, ali, aqui, la, as, deste, desta, destes, destas, disto, meu, meus, minha, minhas, cujo, cujas, cuja, cujos, ai, aquele, aquela, aqueles, aquelas, aquilo, as, quem, outrem, onde, algures, alhures, nenhures, algo, algum, alguma, alguns, algumas, varios, varias, qualquer, quaisquer, cada, como, eu, se, ai, as, assim, voce, agora, me, te, mim, a, p, c, nem, n, eh, pq, de, da, que, politica, em, na, os, se, dos, para, como, uma, um, com, por, ao, entre, sobre, brasil, politico, analise, sua, processo, trabalho, pesquisa, sao, nos, partir, seu, objetivo, mai, esta, pelo, nao, pela, brasileiro, foi, tem, brasileira, ou, tese, sistema, est, nacion, relacao, forma, presente, foram, resultado, relaco, essa, periodo, ano, analisar, meio, dado, esse, paise, tambem, ser, contexto, mudanca, dessa, dissertacao, quais, programa, sociedad, maior, desse, atuacao, compreende, dois, pai, qual, principais, busca, atraves, grande, deste, construcao, tema, diferente, base, assim, tanto, efeito, debate, nova, alem, formacao, recurso, questao, ciencia, desta, papel, duas, outro, fim

3) No espaço "Exclude words shorter than" se determina a exclusão de palavras com menos de 3 letras



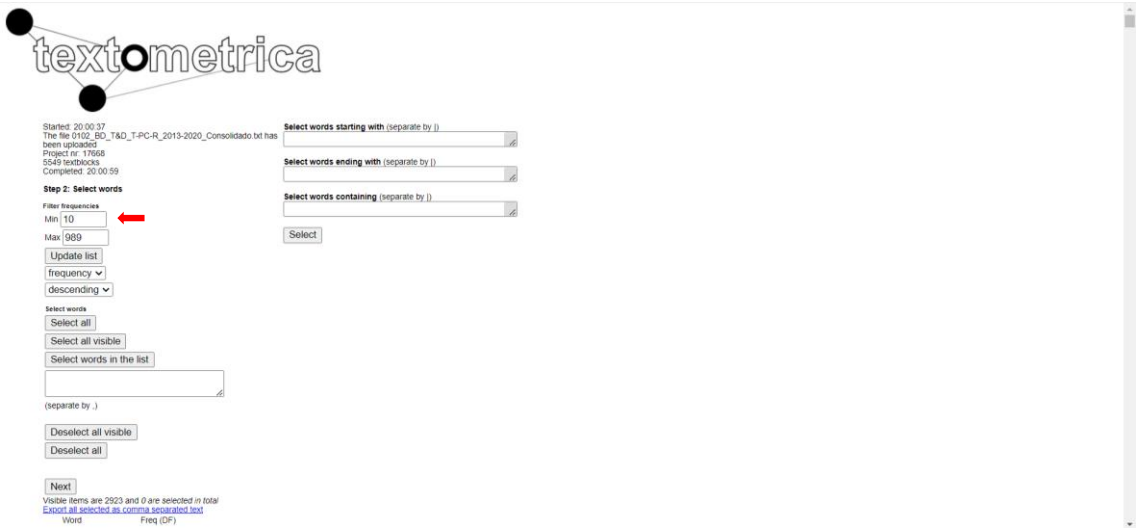
Para realizar a lematização da Rede Temática

Em Escolher ficheiro, selecione o banco de dados "0102_BD_T&D_T-PC-R_2013-2020_Consolidado.txt" > Em "Exclude words shorter than" digite 3 > Cole as palavras acima em "stop words" no espaço indicado

4) Em "next" seguirá para a etapa de "tokenization".

Nesse passo, será gerado o "project" do textométrica.
É possível acessá-lo automaticamente através de desse link: http://textometrica.humlab.umu.se/step2.php?project_id=16316

5) Na etapa "selection" será realizada a seleção dos termos a serem analisados.

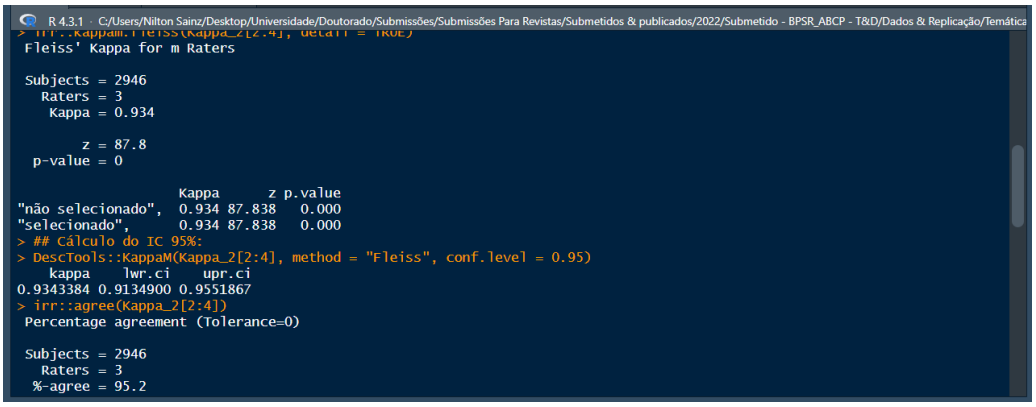


No espaço de frequências de palavras determinou-se a frequência mínima = 10

Foi realizado um processo de revisão e seleção dos termos por três avaliadores independentes, que poderá ser verificado no arquivo "0103_Seleção_Palavras_Avaliadores_Temática.xlsx" na aba "Seleção_avaliadores"

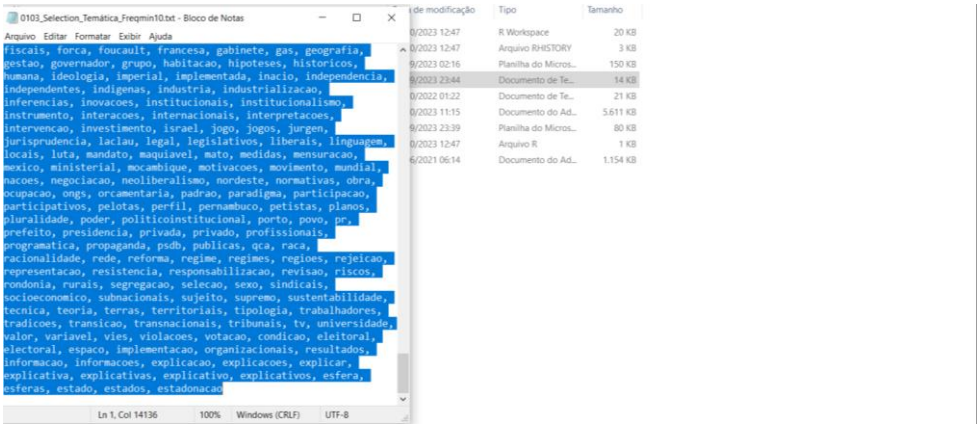
Nessa etapa realizamos o teste de Kappa de Fleiss que poderá ser reproduzido através do Script in R no arquivo "Kappa_Fleiss_RedeTemática_ScriptR.R" com a utilização do banco de dados "Kappa_2.xlsx"

O resultado do teste Kappa de Fleiss pode ser observado abaixo:

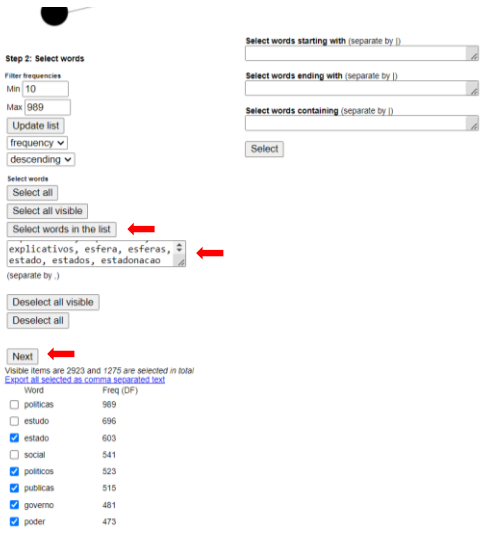


Interpretação do teste estatístico: O teste Kappa varia entre -1 e 1. Quanto mais próximo de 1 o valor de Kappa, maior a concordância entre os avaliadores. Nesse caso, Kappa = 0,93 indica que há uma concordância "quase perfeita" entre os avaliadores (Landis and Koch, 1977, p. 165).

Os termos selecionados na pesquisa podem ser acessados no formato .txt por meio do arquivo "0103_Selection_Temática_Freqmin10.txt".



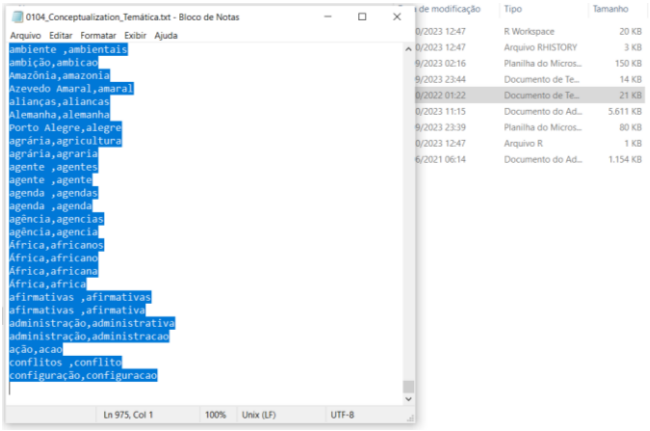
Selecione todos os termos da pasta



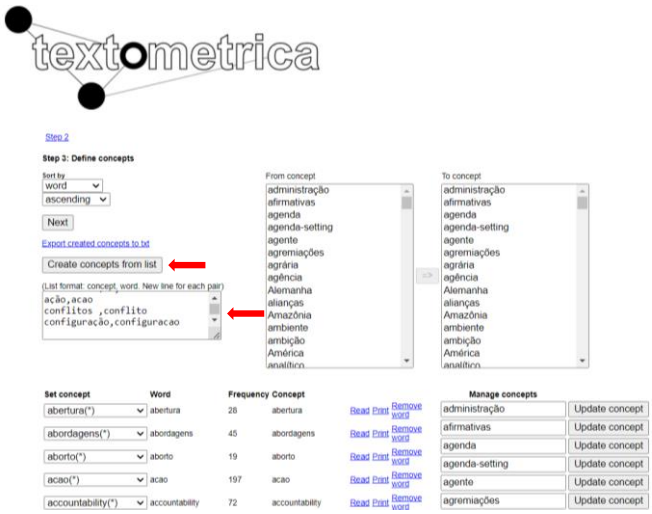
Cole as palavras separadas por vírgulas no espaço indicado > clique em “Select words in the list” (as palavras selecionadas deverão ficar marcadas) > clique em “Next”.

6) Na etapa “**conceptualization**” se realiza a lematização do banco de dados. Essa etapa da pesquisa é qualitativa, onde formamos conceitos através das palavras de radical semelhante[1].

É possível acessar o resultado final desse processo através do arquivo "0104_Conceptualization_Temática.txt



Para reproduzir essa etapa, selecione e copie os termos presentes no arquivo “.txt”



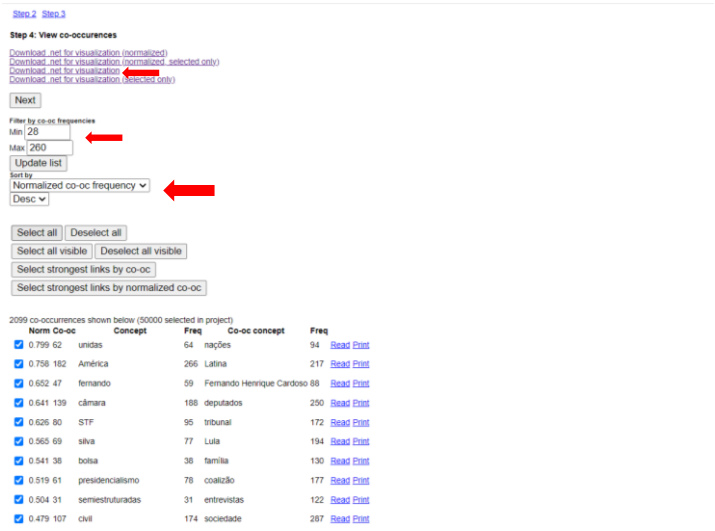
Cole no espaço indicado > Clique em “Create concepts from list” > Clique em “Next”.[2]

[1] Para exemplificar esse processo, nessa etapa note que as palavras “administracao” e “administrativa” se unem para ser representadas pelo conceito “administração”.

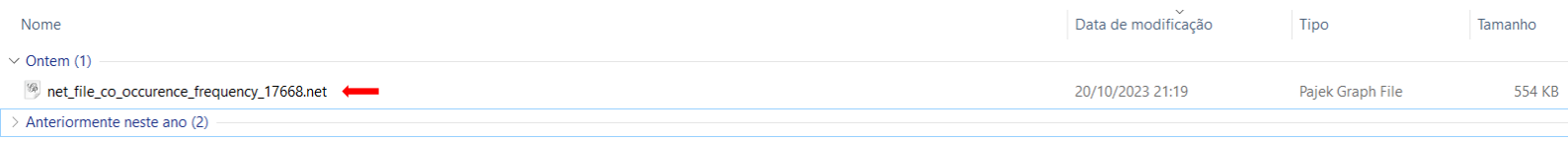
[2] Em caso de demora do software nesse último passo, sugerimos alterar na url o “Step3” para “Step4”.

7) Na etapa "**connection**" se chega ao resultado do processo de coocorrência de termos e conceitos. Para mais detalhes e uma melhor compreensão do método de CCA e sobre o procedimento de construção das redes de coocorrência através do pacote textometrica, sugerimos a leitura do artigo *Introducing Connected Concept Analysis: A network approach to big text datasets* (Lindgren, 2016).

Para realizar a exportação das redes de coocorrência:



Em “filter by co-oc frequencies” preserva-se a determinação do software onde Min 28 e Max 260 > Em “sort by” aplicamos "Normalized co-oc frequency" e "Desc" > Em seguida, clicamos em "Select all" para selecionar todas as co-ocorrências > Na sequência, realiza-se o download da rede através da opção "Download .net for visualization".



É gerado o arquivo “.net” para ser analisado no *software Gephi*.

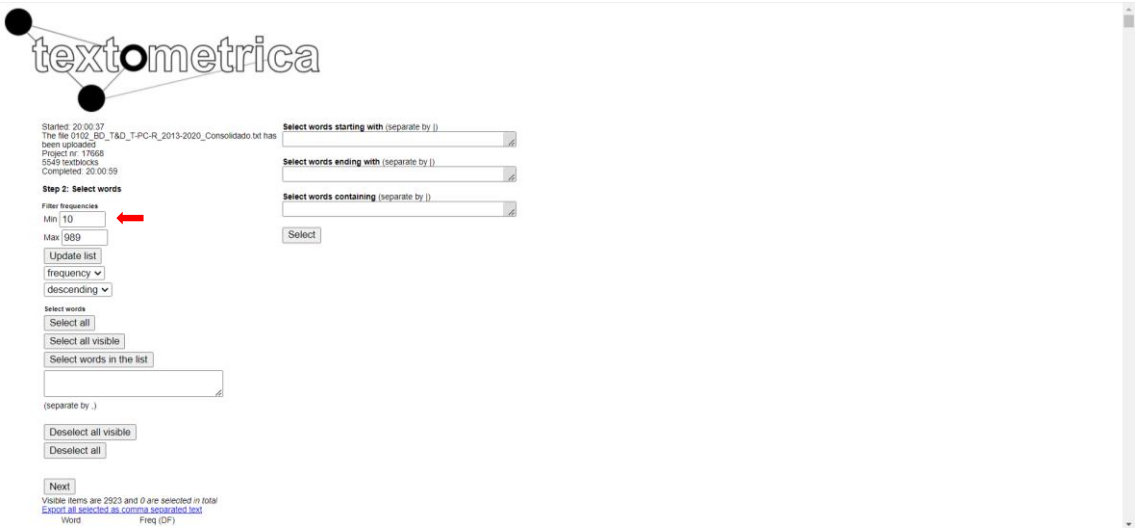
Pare realizar a lematização da Rede Metodológica

Em Escolher ficheiro, selecione o banco de dados "0102_BD_T&D_T-PC-R_2013-2020_Consolidado.txt" > Em "Exclude words shorter than" digite 3 > Cole as palavras acima em "stop words" no espaço indicado

8) Em "next" seguirá para a etapa de "tokenization".

Nesse passo, será gerado o "project" do textométrica.
É possível acessá-lo automaticamente através de desse link: http://textometrica.humlab.umu.se/step2.php?project_id=16350

9) Na etapa "selection" será realizada a seleção dos termos a serem analisados.

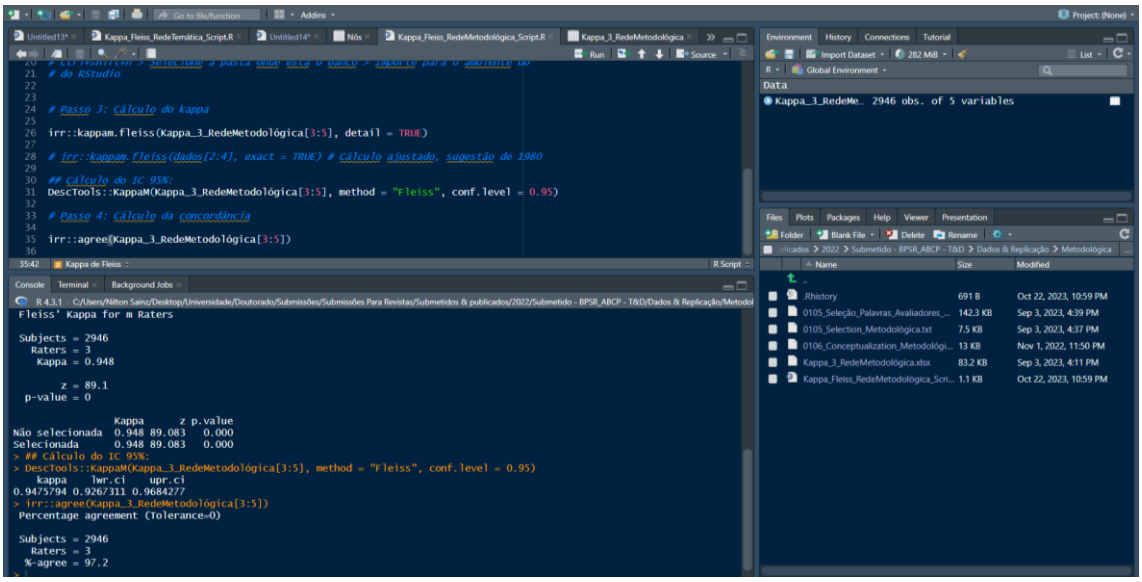


No espaço de frequências de palavras determinou-se a frequência mínima = 10

Foi realizado um processo de revisão e seleção dos termos por três avaliadores independentes, que poderá ser verificado no arquivo "0105_Seleção_Palavras_Avaliadores_Metodológica.xlsx" na aba "Seleção_avaliadores"

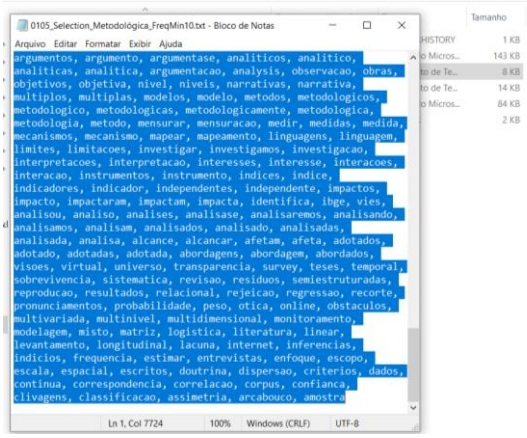
Nessa etapa realizamos o teste de Kappa de Fleiss que poderá ser reproduzido através do Script in R no arquivo "Kappa_Fleiss_RedeMetodológica_Script.R" com a utilização do banco de dados "Kappa_3_RedeMetodológica.xlsx"

O resultado do teste Kappa de Fleiss pode ser observado abaixo:

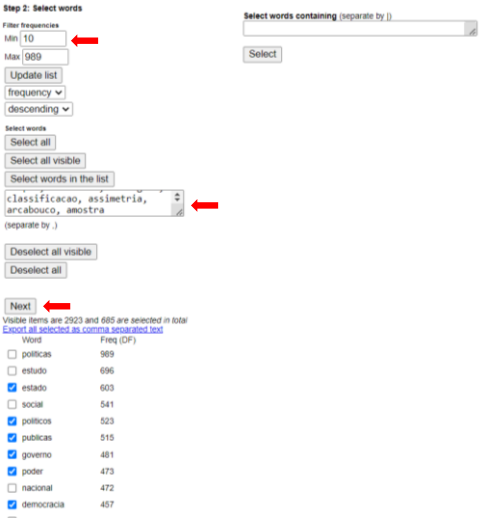


Interpretação do teste estatístico: O teste Kappa varia entre -1 e 1. Quanto mais próximo de 1 o valor de Kappa, maior a concordância entre os avaliadores. Nesse caso, Kappa = 0,94 indica que há uma concordância "quase perfeita" entre os avaliadores (Landis and Koch, 1977, p. 165).

Os termos selecionados na pesquisa podem ser acessados no formato .txt por meio do arquivo "0105_Selection_Metodológica_Freqmin10.txt".



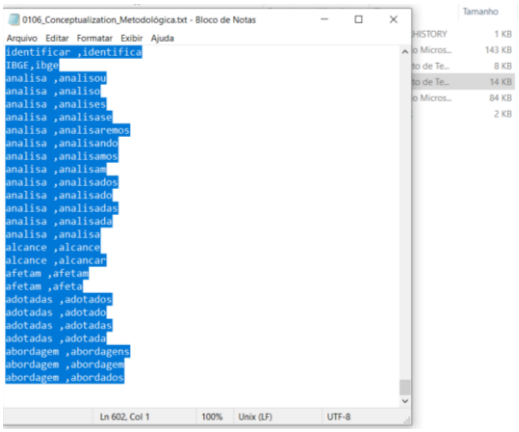
Selecione todos os termos da pasta



Cole as palavras separadas por vírgulas no espaço indicado > clique em “Select words in the list” (as palavras selecionadas deverão ficar marcadas) > clique em “Next”.

10) Na etapa “**conceptualization**” se realiza a lematização do banco de dados. Essa etapa da pesquisa é qualitativa, onde formamos conceitos através das palavras de radical semelhante.[3][4]

É possível acessar o resultado final desse processo através do arquivo "0106_Conceptualization_Metodológica.txt”



Para reproduzir essa etapa, selecione e copie os termos presentes no arquivo “.txt”

---

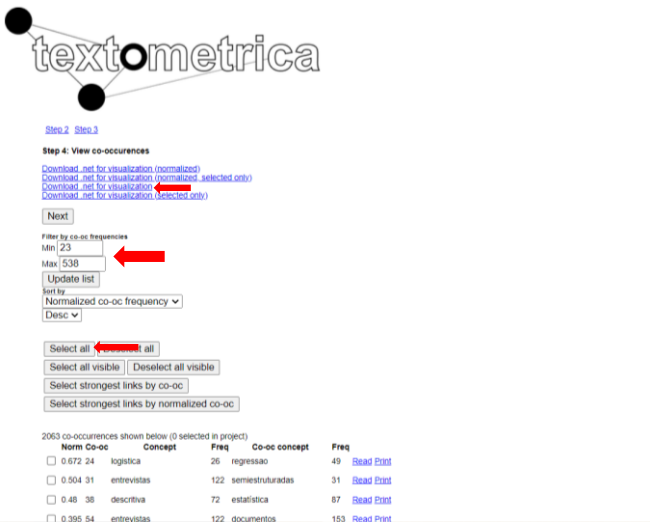
[3] Para exemplificar esse processo, nessa etapa note que as palavras “empirica”, “empíricas, “empiricamente”, “empírico” e “empíricos” se unem para ser representadas pelo conceito “empirica”.

[4] Nesta etapa, criou-se espécies de “hubs” temáticos na rede Metodológica agregando os nós mais frequentes de cada área identificada na rede Temática, Foram eles: Estado, Gov. & RI, Instituições, Atores & PP, Partidos e eleições e Teoria política. Ver nota de rodapé 7, página 10 do artigo original.

Cole no espaço indicado > Clique em “Create concepts from list” > Clique em “Next”.[5]

11) Na etapa "**connection**" se chega ao resultado do processo de coocorrência de termos e conceitos. Para mais detalhes e uma melhor compreensão do método de CCA e sobre o procedimento de construção das redes de coocorrência através do pacote textometrica, sugerimos a leitura do artigo *Introducing Connected Concept Analysis: A network approach to big text datasets* (Lindgren, 2016).

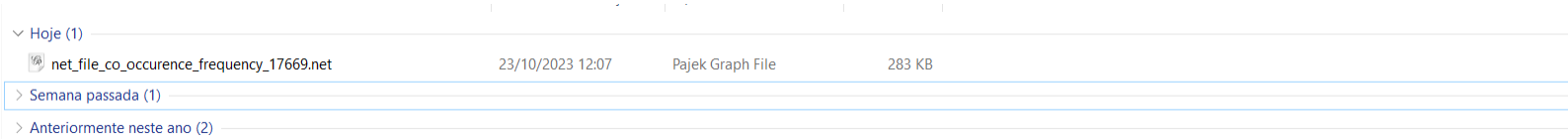
Para realizar a exportação das redes de coocorrência:



Em “filter by co-oc frequencies” preserva-se a determinação do software onde Min 23 e Max 538 > Em “sort by” aplicamos "Normalized co-oc frequency" e "Desc" > Em seguida, clicamos em "Select all" para selecionar todas as co-

[5] Em caso de demora do software nesse último passo, sugerimos alterar na url o “Step3” para “Step4”.

ocorrências > Na sequência, realiza-se o download da rede através da opção "Download .net for visualization".

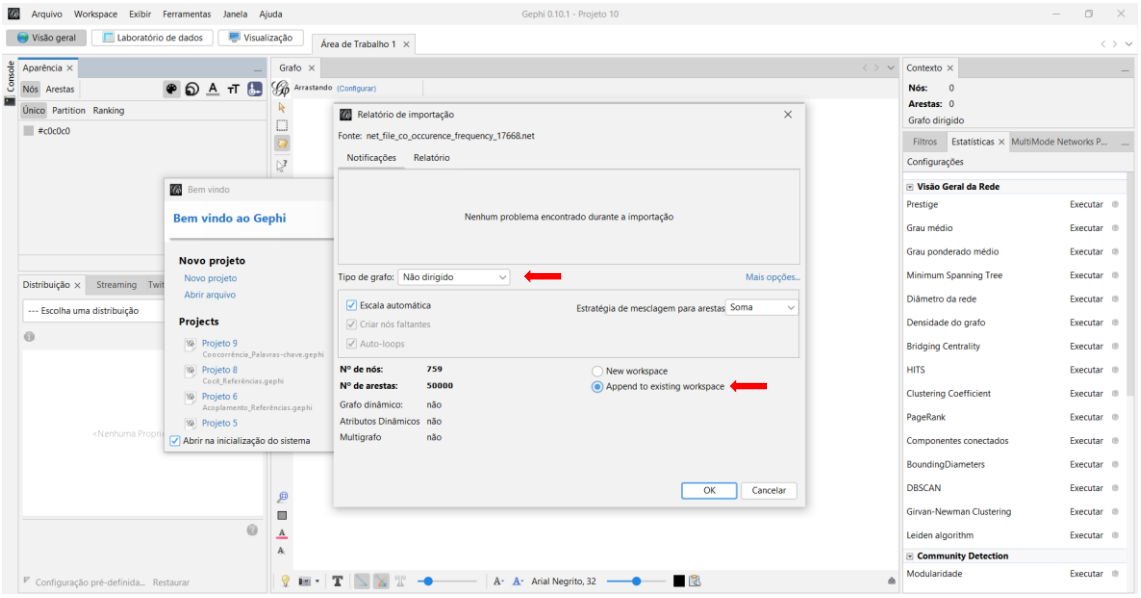


É gerado o arquivo “.net” para ser analisado no *software Gephi*.

Visualização da Análise de Redes no software *Gephi*

Para a construção de ambas as redes (Temática e Metodológica) o procedimento deverá ser o mesmo.

12) Na etapa "**Data Vizualization**", para reproduzir resultados semelhantes aos do artigo deverão ser realizados os procedimentos descritos abaixo no *software Gephi*:

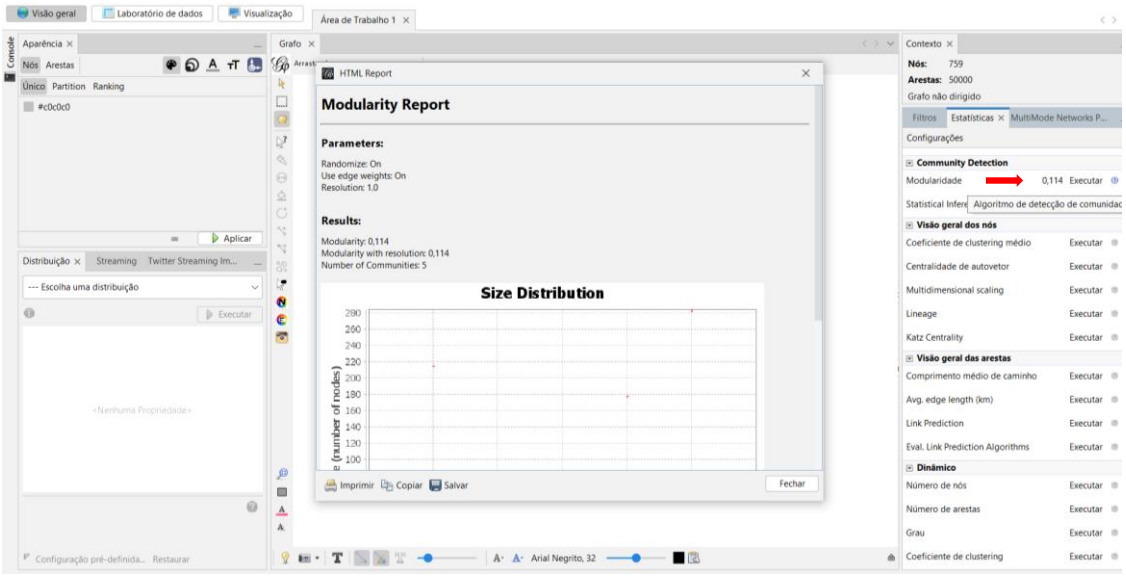


Abra o arquivo ".net" que você realizou o download conforme indicado nos passos anteriores. No "Tipo de grafo" selecione a opção "Não dirigido" > Selecione "Append to existing workspace".

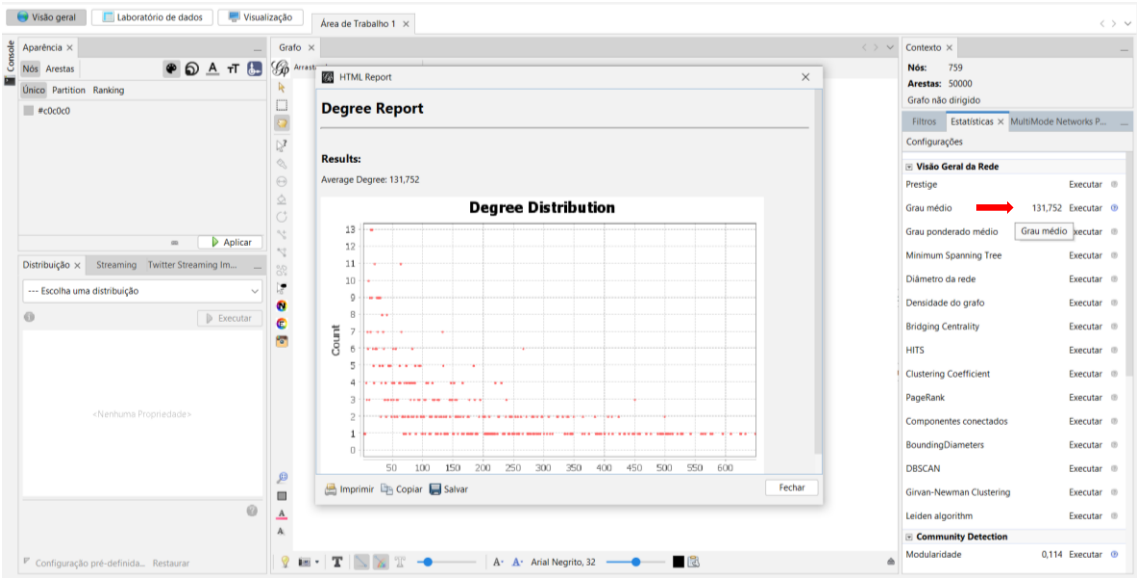
i) Realize o cálculo das seguintes métricas citadas na seção de materiais e métodos páginas 13 e 14:

Métricas	Descrição
Modularidade	Medida de estrutura da rede. Mensura a força da divisão da rede em clusters ou comunidades. Detecta clusters de palavras significativas a partir de suas equivalências estruturais e coesões intragrupo, captando "sub-redes", isto é, domínios temáticos da rede semântica.
Grau	Número absoluto de conexões entre um determinado vértice e os demais vértices da rede, constituindo indicador de influência de um termo. Uma palavra com centralidade elevada de grau indica que ela está muito presente no corpus
Intermediação	É a frequência com que um vértice se coloca nos caminhos mais curtos entre os demais vértices, mostrando quais são os termos mais estratégicos da rede. Em uma rede semântica expressa palavras "pontes" situadas nas fronteiras entre os clusters

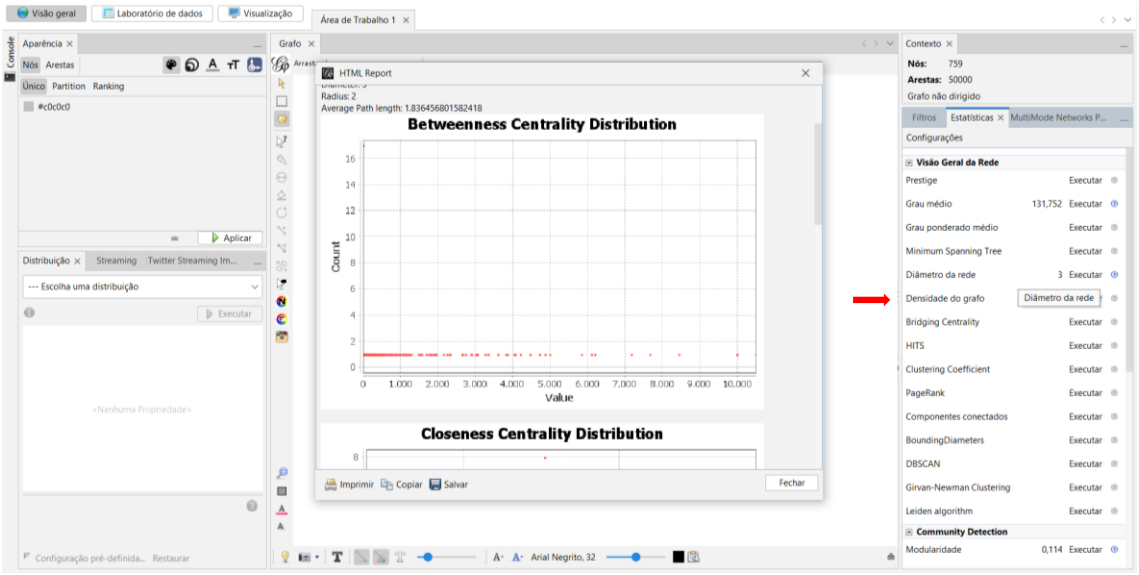
Proximidade	É a distância média de um vértice ou nó de todos os demais vértices da rede. É uma medida de influência global que demonstra o quanto um vértice “viaja” pelos circuitos mais curtos da rede.
Peso de aresta	O peso de aresta entre duas palavras expressa a força da associação entre os termos. Representam as relações semânticas entre os conceitos ou termos.
Grau ponderado	É a soma do peso das arestas conectadas com determinado vértice, indicando os vértices de alta conectividade ao esquadrinhar o peso das arestas.



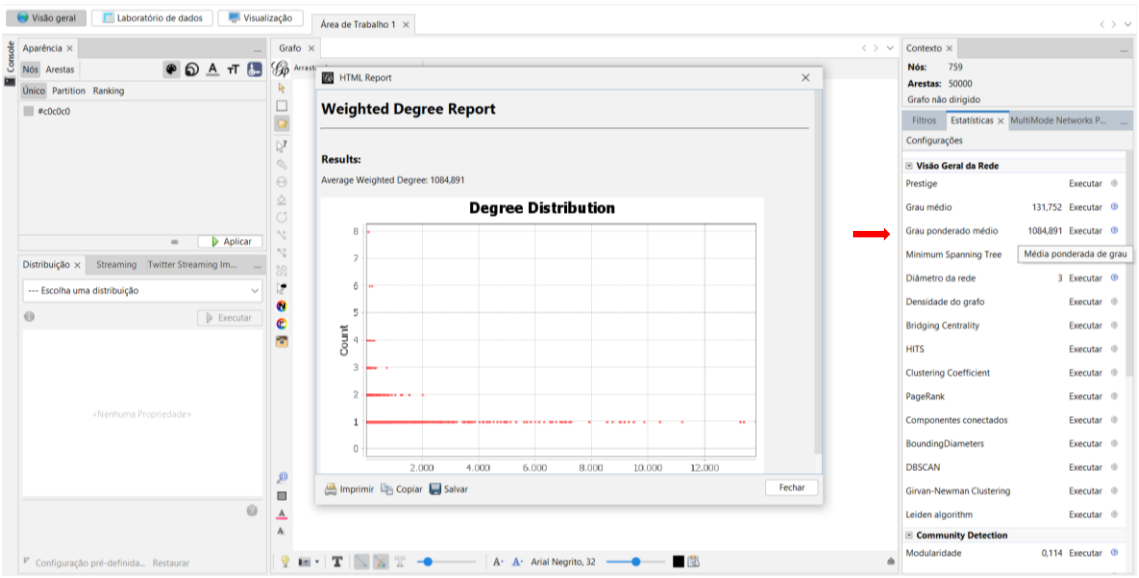
Na aba Estatísticas > Execute “Modularidade” com resolução 1.0



Na aba Estatísticas > Execute “Grau médio”

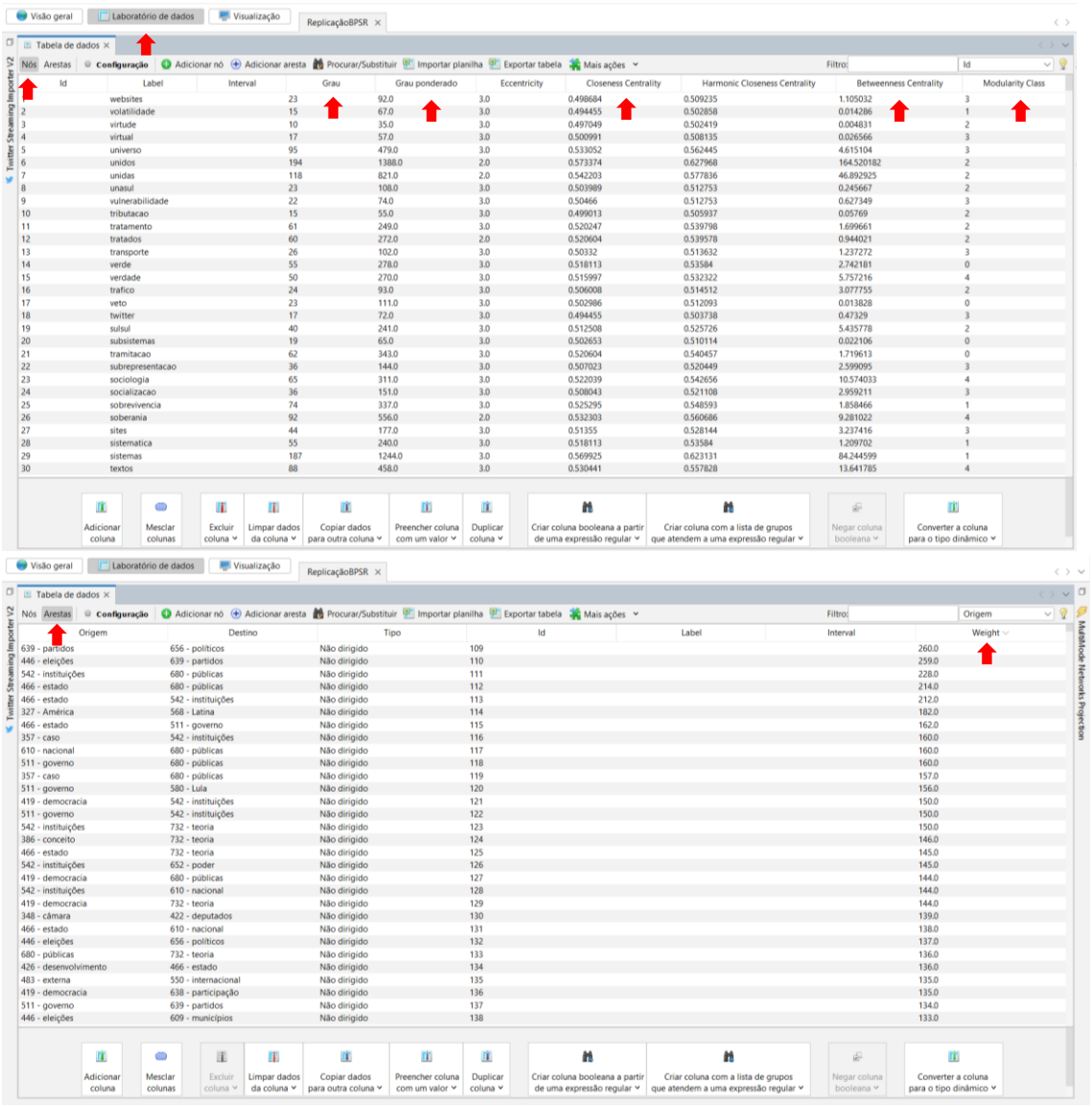


Na aba Estatísticas > Execute Diâmetro da rede > Selecione apenas a opção “Não dirigido”. Essa estatística calcula a Intermediação e Proximidade da rede.



Na aba Estatísticas > Execute o Grau Ponderado.

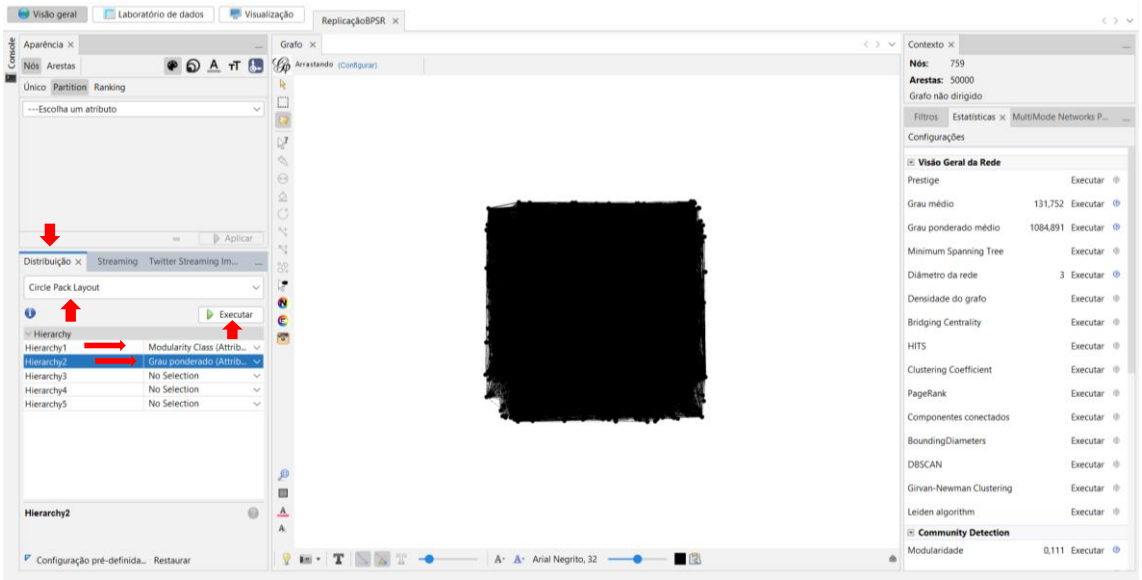
Obs: A estatística de peso de arestas já estará calculada. É possível observar esse resultado em Laboratório de dados na aba de Arestas.



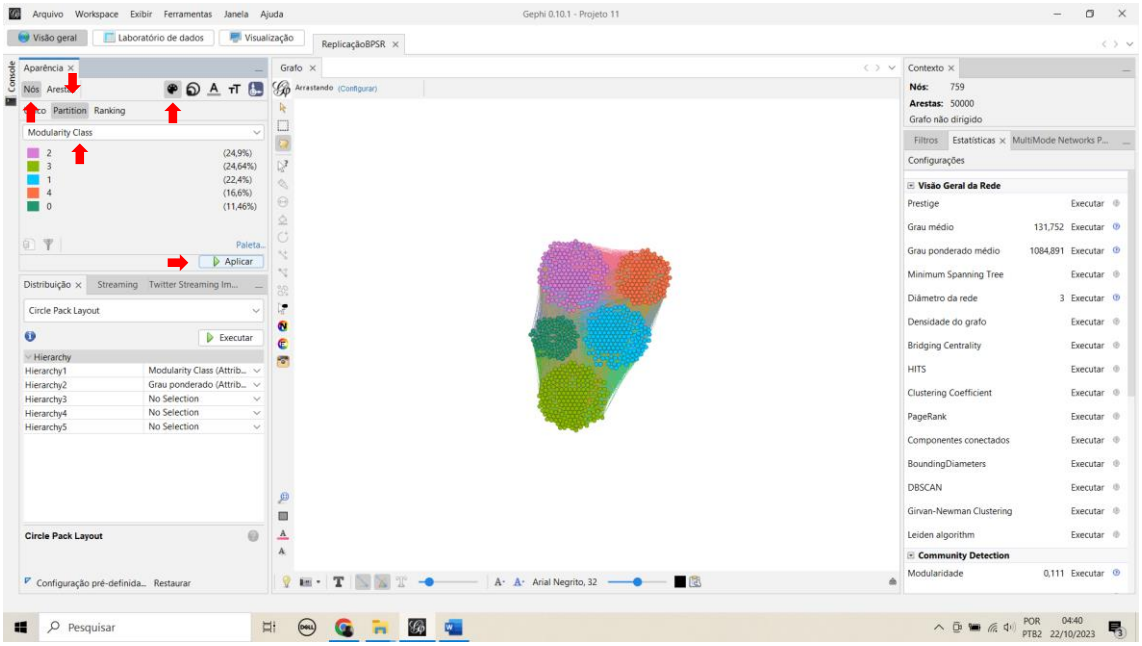
No laboratório de dados será possível conferir os resultados de todas as métricas calculadas anteriormente.

Através desses dados, é possível reproduzir e elaborar da Tabela A1 a A9 presentes nos apêndices do artigo com as métricas de grau ponderado e peso das arestas.

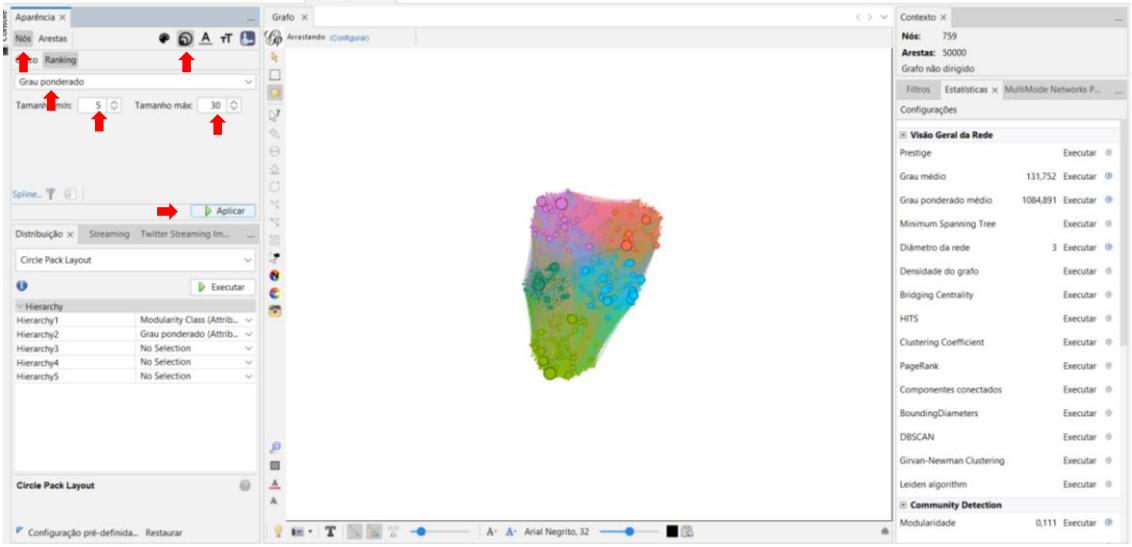
ii) Na sequência, deverão ser aplicados as seguintes configurações:



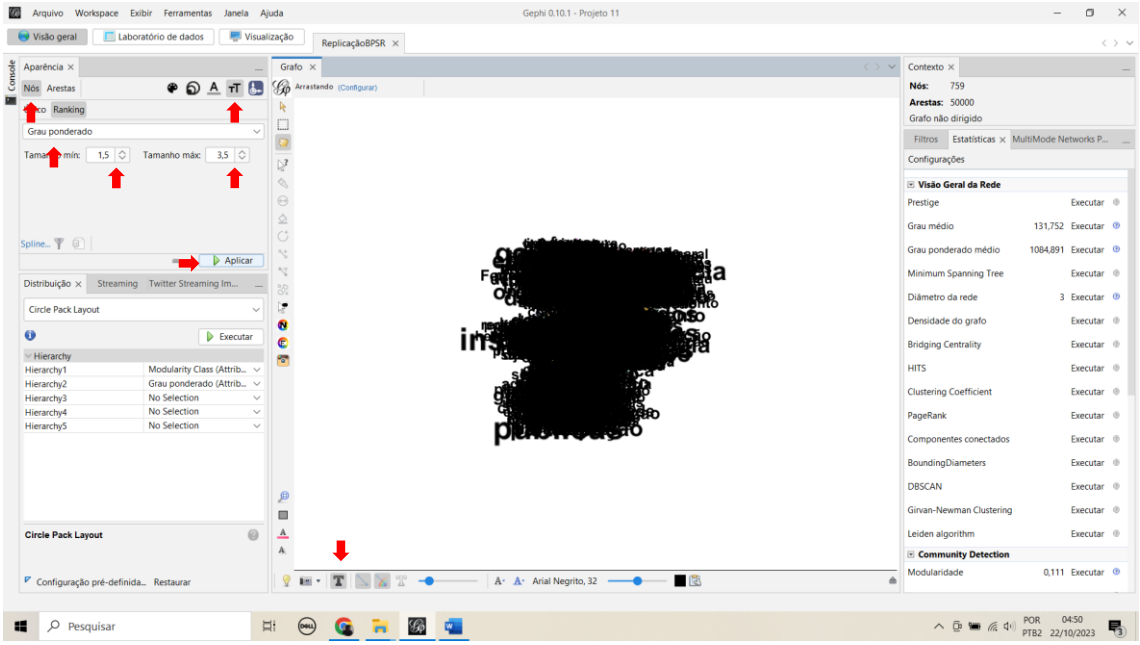
Em distribuição, selecione o algoritmo Circle Pack Layout > Em “Hierarchy 1” escolha modularidade > Em “Hierarchy 2” escolha grau ponderado > Execute



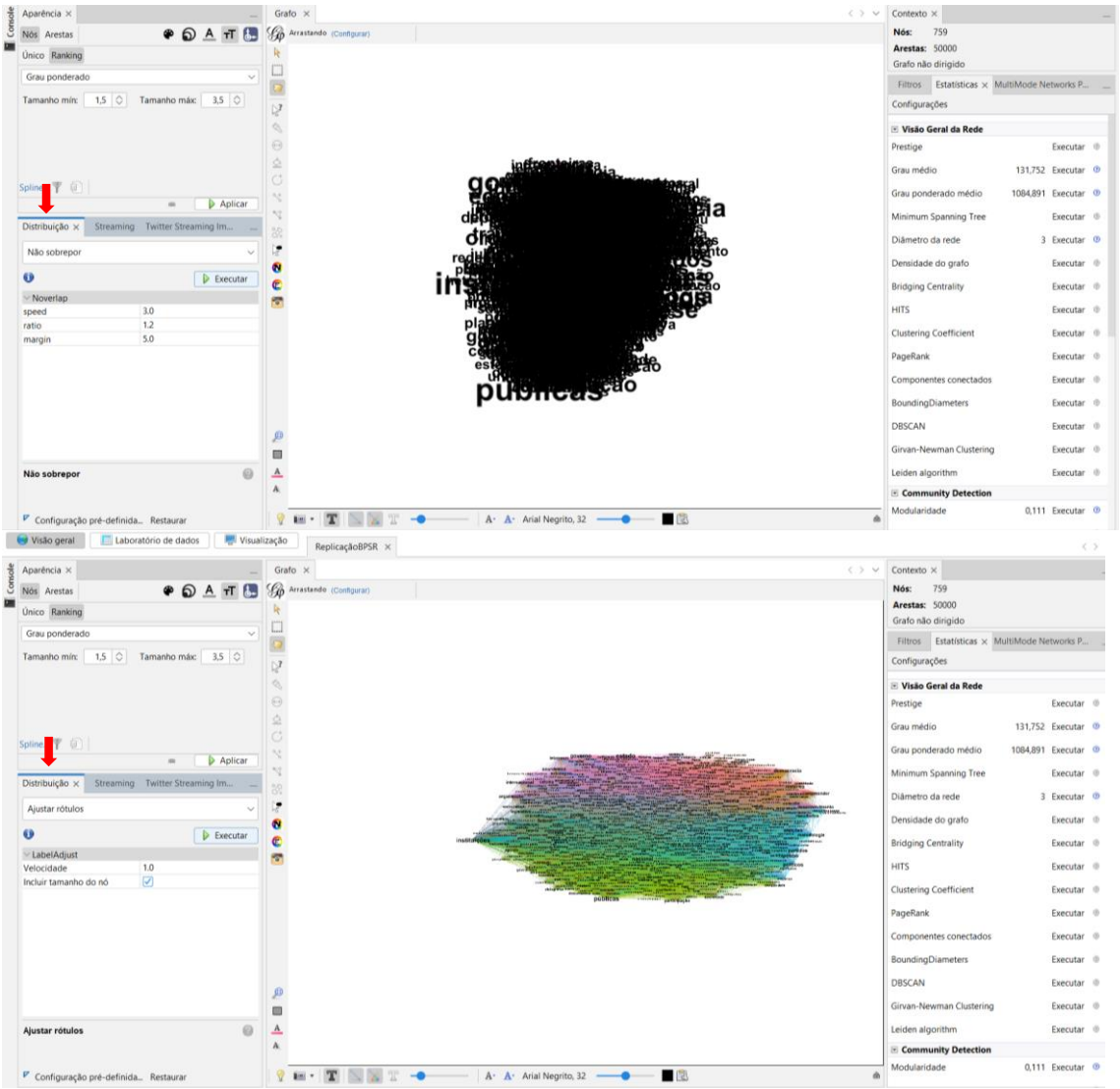
Em aparência > Nós > Cor > “Partition” > Escolha por modularidade > Aplique.



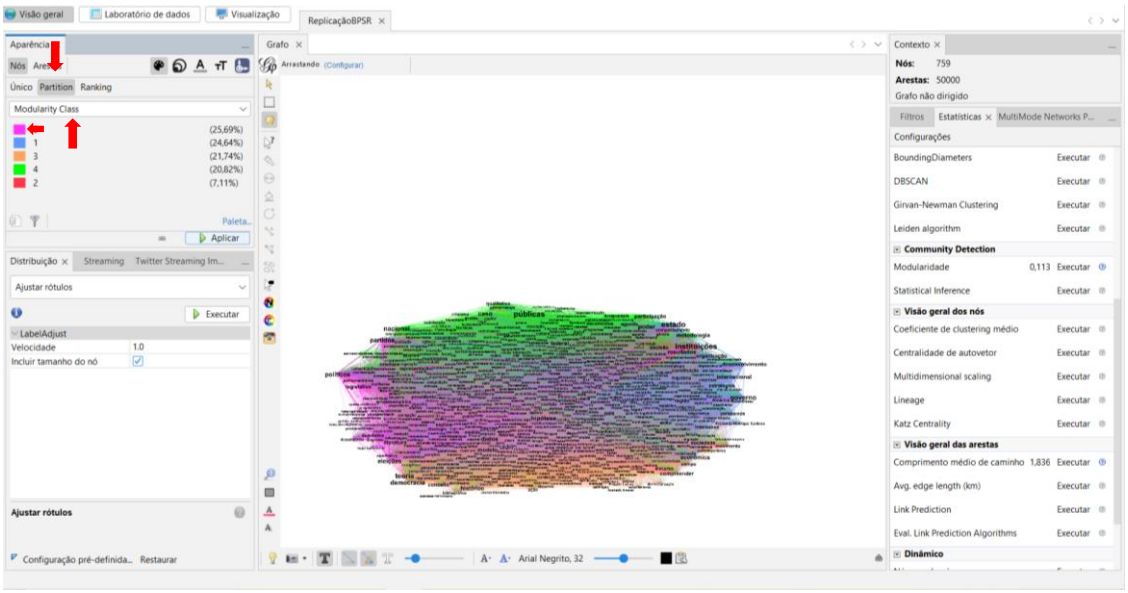
Em aparência > Nós > Tamanho (nós) > Ranking > Escolha grau ponderado configure tamanho mínimo 5 e máximo 30 > Aplique



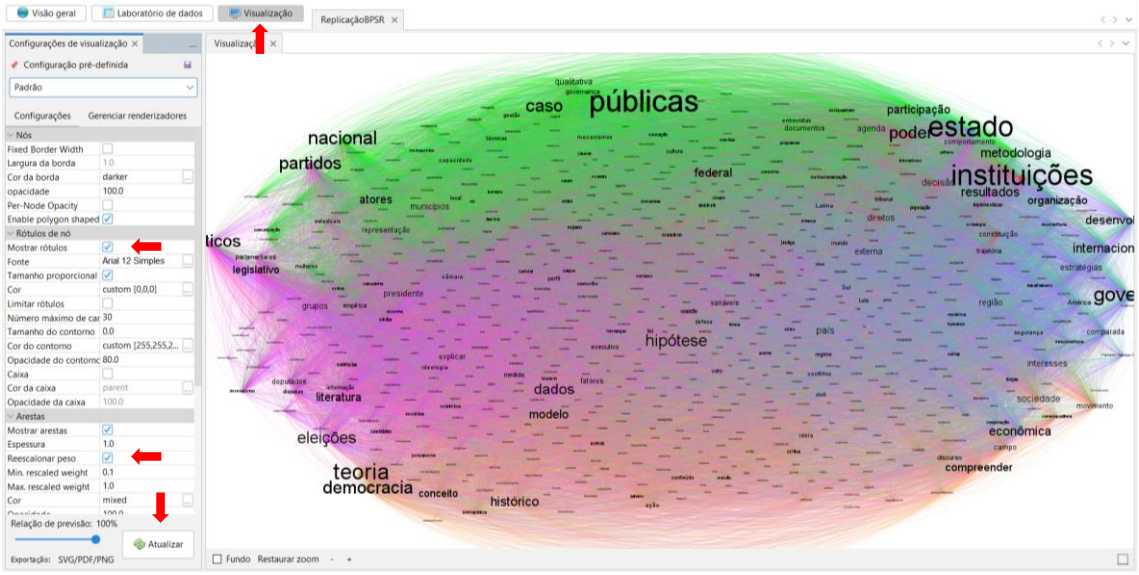
Em aparência > Nós > Tamanho do rótulo > Ranking > Escolha grau ponderado configure com tamanho mínimo 1,5 e máximo 3,5 > Aplique > Clique em “T” na parte inferior do Grafo para exibir os rótulos da rede



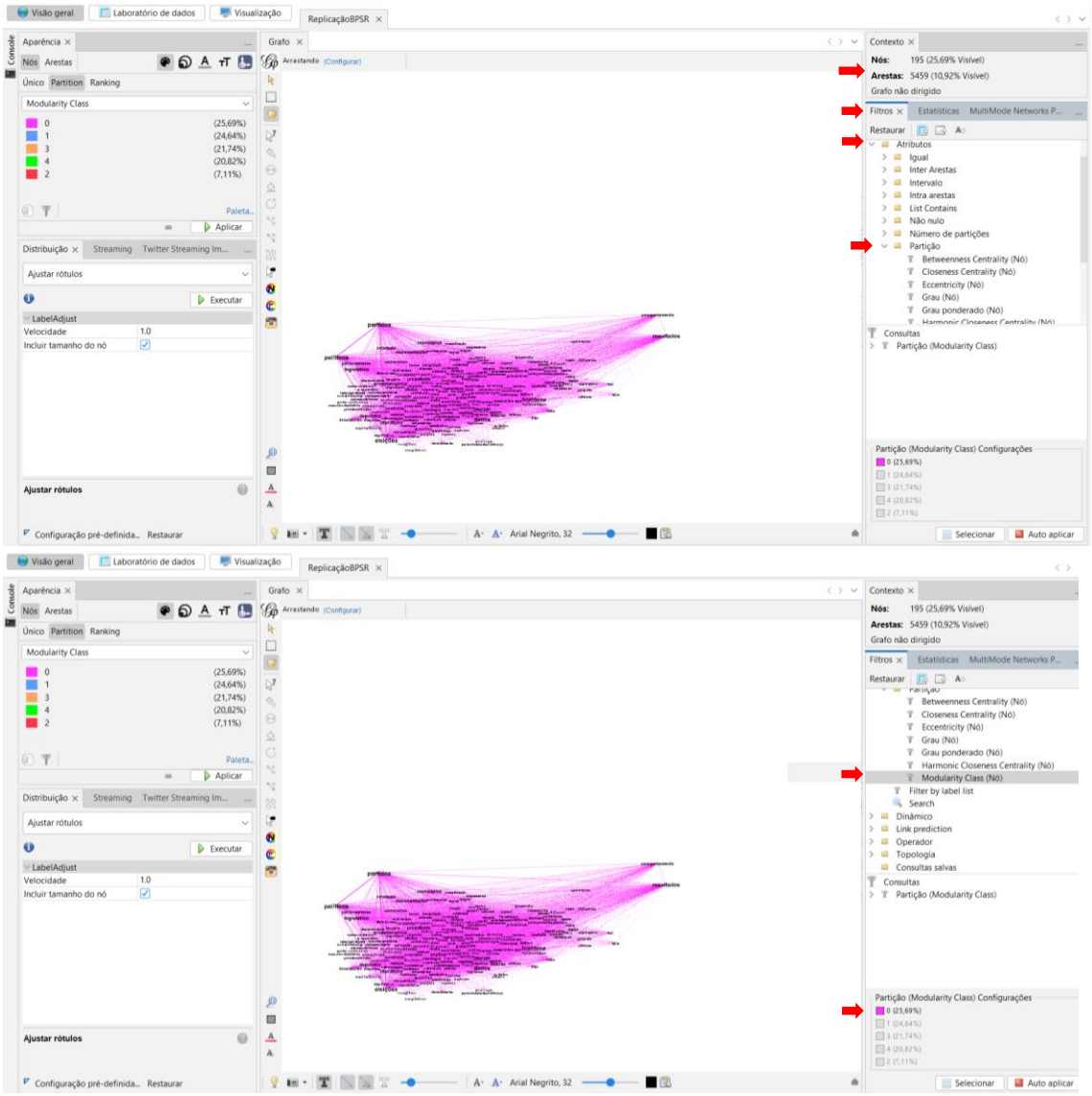
Para melhorar a visibilidade da rede, em Distribuição > execute o algoritmo “Não sobrepor” > Em seguida execute o algoritmo de “Ajustar rótulos”



Para configurar as cores dos clusters conforme o realizado no artigo publicado na BPSR > em “partition” > altere a "caixa" de cada um dos clusters > Aplique



Para melhorar a visualização da rede > em visualização > Selecione Reescalonar peso das arestas > Marque a opção mostrar rótulos > Clique em atualizar



Em Filtros > Atributos > Partição > Modularity class > Selecione um *cluster* para visualizar apenas o grupo desejado.

Com isso você poderá gerar os grafos presentes nas figuras A1 a A9 do artigo e também verificar o laboratório de dados dos *clusters* separadamente que são as fontes das tabelas presentes no artigo. Também é possível observar na aba contexto como estará configurada a distribuição de nós e arestas dos referidos clusters.