

Ofício, Engenharia e Arte: Inspiração e Técnica na Análise de Dados Qualitativos

Craft, Ingenuity and Art: Inspiration and Technique in the Analysis of Qualitative Data

Artesanía, Ingenio y Arte: La Inspiración y la Técnica en el Análisis de Datos Cualitativos

Edson Armando Silva

Universidade Estadual de Ponta Grossa, Brasil
edameister@gmail.com

Joseli Maria Silva

Universidade Estadual de Ponta Grossa, Brasil
joseli.genero@gmail.com

Resumo

Resumo: O objetivo deste texto é explorar os procedimentos que podem ajudar o pesquisador de ciências sociais e humanas a analisar aspectos da documentação de sua pesquisa por meio de ferramentas computacionais. O uso de ferramentas informacionais para a pesquisa qualitativa apresenta algumas vantagens. Inicialmente, elas permitem a percepção de tendências de sentido em grandes quantidades de documentos; ademais, diminuem a necessidade de intervenção manual do pesquisador, aumentando sua produtividade e sua capacidade de realizar mudanças de rumos quando a dinâmica da pesquisa exigir. Os procedimentos analisados neste artigo permitem o pesquisador posicionar suas escolhas qualitativas com base na centralidade dos sentidos que os enunciados das fontes expressam, aumentando sensivelmente a capacidade de interpretação da realidade social.

Palavras-Chave: Análise Qualitativa – Quantitativa; Discurso; Ferramentas.

Abstract

The aim of this paper is to explore procedures that can help the researchers on social sciences and humanities to examine aspects of the documentation of their researches by means of computational tools. The use of informational tools for qualitative research has some advantages. Initially, they allow the perception of meaning tendencies in large amounts of documents; moreover, they reduce the need for manual interventions by the researchers, increasing their productivity and their ability to make changes of direction whenever the dynamics of the search demands it. The procedures discussed in this article allow the researchers to determine their qualitative choices based on the centrality of the meanings expressed by their sources, significantly increasing the capacity of interpretation of the social reality.

Keywords: Qualitative Analysis; Quantitative Analysis; Speech; Tools.



Resumen

El objetivo de este trabajo es explorar los procedimientos que pueden ayudar a los investigadores de ciencias sociales y humanidades a analizar aspectos de la documentación de su investigación a través de herramientas computacionales. El uso de herramientas de la información para la investigación cualitativa tiene algunas ventajas. Inicialmente, permiten la percepción de tendencias de sentido en grandes cantidades de documentos. Además, reducen la necesidad de intervención manual por parte del investigador, incrementando su productividad y su capacidad para hacer cambios de dirección cuando la dinámica de la investigación lo exige. Los procedimientos analizados en este artículo permiten al investigador posicionar sus elecciones cualitativas basándose en la centralidad de los sentidos que expresan los enunciados de las fuentes, aumentando significativamente la capacidad de interpretación de la realidad social.

Palabras-Clave: Análisis Cualitativo; Cuantitativo; Discurso; Herramientas.

Introdução

Em artigo intitulado 'Toward a more fully reflexive feminist geography', Karen Fanconer Al-Hindi e Hope Kawabata¹ argumentam, ao proporem a clivagem de gênero como fundamento para desnaturalizar o ponto de vista dominante nas ciências sociais, que a geografia feminista se situa ao lado das abordagens pós-colonialistas que criticam as tendências universalistas embutidas em perspectivas teóricas e metodológicas, a saber, trata-se de revelar a posicionalidade de propostas que se apresentam como universais. Sob um aspecto metodológico, esta posição teórica conduz as autoras a uma proposta duplamente reflexiva. Em primeiro lugar, porque busca superar uma posição baseada na dualidade sujeito-objeto, isto é, uma posição onde o pesquisador (sujeito) apreende e revela a 'verdadeira' natureza do objeto pesquisado. Esta superação destaca a natureza relacional do conhecimento produzido na interação entre pesquisador e pesquisado, entrevistador e entrevistado. A consequência lógica desta constatação é perceber o caráter dinâmico do processo de pesquisa. A relação de pesquisa deve, portanto, provocar modificações nas perspectivas do pesquisador, alterar suas

questões de partida e revelar dimensões da realidade que não poderiam ser pressupostas.

É possível entender a diferença entre essas perspectivas, comparando-as em duas abordagens metodológicas diferentes. Se um pesquisador vai a campo com um instrumento de pesquisa baseado em formulário estruturado, por mais que tenha tomado todos os cuidados na sua elaboração, só poderá recolher do campo as variáveis previstas no instrumento. Entrevistas abertas, histórias de vida, observações participantes podem revelar aspectos impensados da realidade. Desta maneira, a pesquisa e o pesquisador se abrem ao radicalmente novo em termos de visões de mundo ou de categorias interpretativas da realidade. Entretanto, esta perspectiva aumenta a complexidade da pesquisa e exige procedimentos operacionais mais refinados na análise e interpretação dos dados.

O objetivo deste artigo é justamente apresentar esses procedimentos, mesmo cientes de que os problemas a serem enfrentados pelos pesquisadores podem levar a outras abordagens e perspectivas.

A Importância do Registro dos Contextos

Os procedimentos de análise em uma

pesquisa implicam recortar e reorganizar os elementos levantados em campo. Todavia, em todos os casos, o pesquisador deve ter o cuidado de registrar e manter o contexto de produção dos documentos. A plena interpretação da palavra somente é possível quando levamos em consideração o contexto de sua produção, o contexto de seu uso e o contexto cultural onde os enunciados acontecem. A posicionalidade do autor do enunciado, a quem ele se dirige, o contexto da fala e a cultura na qual se insere a relação entre emissor e receptor do discurso são elementos que constantemente devem ser revisitados pelo pesquisador no ato de confrontar os resultados produzidos com os procedimentos técnicos de análise.

É o contexto que deve dar as referências para a interpretação dos dados e isto não é uma ideia nova. Ainda no século XIX, a proposta de método histórico baseado na heurística, na crítica e na hermenêutica já valorizava o contexto como elemento central no discernimento da leitura das fontes. A postura pós-colonialista acrescentou a ênfase na posicionalidade dos atores. O contexto já não era universal, mas os atores situavam-se em distintos contextos, a partir dos quais formulavam sua visão de mundo e, conseqüentemente, sua lógica argumentativa.

Feitas essas observações de caráter teórico, vamos nos concentrar nos procedimentos metodológicos que permitem revelar os aspectos da realidade que melhor correspondam aos objetivos das nossas pesquisas. Embora não existam fórmulas mágicas, é útil conhecer procedimentos que, comparando com ferramentas nas mãos de um lapidador, permitem revelar aspectos da realidade, que não obstante lá estivessem, necessitavam de lapidação para deixar ver sua beleza. Cada face da pedra exige seu próprio procedimento – às vezes o disco, eventualmente a espátula –, mas a forma final é determinada pela relação entre o material

bruto e o artesão.

De certa maneira, esses procedimentos se encaixam no que Laurence Bardin chamou, no prefácio de seu livro, de análise de conteúdo:

Um conjunto de instrumentos metodológicos cada vez mais subtis em constante aperfeiçoamento, que se aplicam a “discursos” (conteúdos e continentes) extremamente diversificados. O factor comum destas técnicas múltiplas e multiplicadas – desde o cálculo de frequências que fornece dados cifrados, até à extracção de estruturas traduzíveis em modelos – é uma hermenêutica controlada, baseada na dedução: a inferência².

O objetivo deste texto é explorar os procedimentos que podem ajudar o pesquisador a analisar aspectos da documentação de sua pesquisa. Neste sentido, exploram-se as ferramentas que podem ser usadas em conjunto mas cujos resultados devem sempre ser confrontados com o contexto.

Para cada procedimento metodológico serão utilizados alguns softwares que facilitarão sua execução. O uso de ferramentas informacionais para a pesquisa qualitativa apresenta algumas vantagens. Inicialmente, elas permitem a percepção de tendências de sentido em grandes quantidades de documentos; ademais, diminuem a necessidade de intervenção manual do pesquisador, aumentando sua produtividade e sua capacidade de realizar 'mudanças de rumos' quando a dinâmica da pesquisa exigir.

Nossa proposta de articulação entre procedimentos quantitativos e qualitativos, como veremos a seguir, traz um elemento que merece atenção especial: a

verificabilidade das escolhas qualitativas. Em outras palavras, ao citar trechos lapidares dos nossos documentos, podemos também demonstrar a centralidade dos sentidos que os enunciados expressam, ou, ao contrário, deixar clara a sua singularidade.

A ênfase deste artigo, entretanto, não está nos *softwares*, mas nos procedimentos. Embora já exista uma vasta oferta de ferramentas proprietárias para cobrir grande parte das funcionalidades necessárias às pesquisas qualitativas – facilitando o processo de codificação e recuperação de informações –, elas permitem executar apenas o que foi anteriormente previsto pelos programadores, limitando a liberdade do pesquisador em explorar novos procedimentos.

Para este artigo escolhemos executar os procedimentos com ferramentas livres, cuja arquitetura interna fosse suficientemente aberta e flexível a fim de permitir a comunicação com outros aplicativos, importando e exportando os dados de modo a produzir relações imprevistas pelos seus criadores. Além disso, por serem ferramentas gratuitas e multiplataformas (rodam nos principais sistemas operacionais disponíveis), os leitores deste trabalho podem simplesmente instalar os programas e testar os procedimentos em suas pesquisas, experimentando criar outros procedimentos que melhor respondam às suas perguntas.

Frequência e Relação

Ao analisar fundos documentais de grande porte, um dos procedimentos é avaliar a frequência no emprego de termos. Isto corresponde a um dito do Evangelho: “a boca fala do que o coração está cheio” (Mt, 12:34). De fato, a ênfase no emprego de palavras e expressões revela valores e sentimentos comuns expressos em um conjunto de discursos. Esta observação é consequência do

papel da linguagem na formação do mundo cultural. A linguagem não apenas expressa o mundo, mas ajuda a produzi-lo. O uso da linguagem é, portanto, um dos principais indicadores que os pesquisadores dispõem para realizar mapeamentos de fenômenos culturais.

Quando nos referimos à análise de frequência, não estamos nos referindo apenas à análise estatística da repetição com variações morfológicas ou semânticas, como foi feito brilhantemente por Jenni Klemettinen ao analisar discursos de Charles de Gaulle³, mas incorporar a relação existente entre os termos como fez Dmitry Paranyushkin⁴.

A proposta de Dmitry Paranyushkin é evitar qualquer influência subjetiva ou cultural no processo de construção do gráfico, na medida em que ele trata somente da frequência e da relação das palavras no discurso a ser analisado, revelando as estruturas de sentido do texto em si. Em outras palavras, ele busca revelar a 'verdade' a partir da multiplicidade e pretende – ao usar como critério único de organização do 'grafo' as relações de intensidade na conexão entre as palavras empregadas no discurso – deixar que a forma resultante do procedimento 'fale por si mesma'.

Uma rede, entretanto, é uma relação binária, isto é, devemos criar uma conexão entre dois nós, e esta conexão é a estrutura básica sobre a qual vão sendo construídos os demais componentes. Como critério para procedimento inicial, Dmitry Paranyushkin escolheu a proximidade entre as palavras partindo de um intervalo de '2'. Vale dizer que cada palavra geraria um nó na rede e ao intervalo de até duas palavras produziríamos as relações entre essas palavras chamadas de arestas.

Neste artigo propomos um caminho diferente. Vamos criar uma relação binária entre cada uma das palavras de uma frase.

Tomar a frase como unidade onde as conexões serão criadas nos parece mais coerente com a estrutura natural de produção do sentido em um discurso. E para criar esta rede vamos nos utilizar de três *softwares*, todos livres, descrevendo os passos de maneira que não haja a necessidade do uso de técnicas de programação. Usaremos apenas

número de identificação correspondente à linha onde a frase estiver colocada. Feito isso, poderemos exportar a planilha para continuar os procedimentos em outro aplicativo. Os passos seguintes serão realizados em um aplicativo chamado '*OpenRefine*'⁶. Abaixo há um exemplo de planilha que poderia ser importada e tratada no *OpenRefine*:

Figura 1.

Identificação da Frase	Frase
1	Sim, eu divido o meu dia entre cuidar da família e o trabalho.
2	Então quando eu tô em casa, tento privilegiar a família e não levar trabalho pra casa, não precisar fazer trabalho fora dos horários em que eu teria que trabalhar.
3	E quando minhas filhas tão em casa, eu privilegio as duas em vez de atender à vida profissional.

as funções básicas de cada um dos programas indicados.

A representação gráfica das relações entre as palavras exige um trabalho de preparação antes que os dados sejam importados para um programa de análise de redes. Começaremos nossa preparação em um processador de textos e em uma planilha eletrônica – neste caso escolhemos o pacote '*LibreOffice*'⁵.

A tarefa inicial é simples: basta usar os comandos de 'localizar e substituir' para localizar os 'pontos' e substituí-los por 'ponto e parágrafo'. Isto pode ser feito rapidamente no texto todo usando a seguinte expressão regular no campo de localização: '\ ' e a expressão '\n' no campo de substituição. O objetivo deste procedimento é simplesmente colocar cada frase em um parágrafo diferente. O passo seguinte é transformar texto em tabela, de modo que cada frase seja colocada numa linha da tabela. Em seguida podemos copiar e colar a tabela em uma planilha. Na planilha criamos uma nova coluna com o recurso de autonumeração. Assim teremos uma planilha na qual toda frase terá um

A padronização dos termos é um procedimento necessário antes de organizarmos a rede. O aplicativo '*Open Refine*' foi construído para organizar e padronizar dados em uma tabela ou banco de dados. Do ponto de vista prático, a *string* (sequência de caracteres) '*Bondade*' e '*bondade*' é compreendida pelo computador como termos diferentes, assim como '*profissional.*' e '*profissional*' (observe o ponto no final da *string*). Para que possamos analisar as redes de sentido existentes em um texto é necessário padronizar os termos, excluir a pontuação e retirar os termos não significativos.

A partir das operações feitas no *LibreOffice*, o *OpenRefine* recebe uma planilha contendo duas colunas. Na primeira há um número identificador da frase, e na segunda, a frase em si. O procedimento inicial é a exclusão dos símbolos de pontuação que, ligados às palavras, poderiam gerar erros de interpretação na coluna contendo as frases. Isto pode ser feito rapidamente no menu '*Edit Cells*' com o

comando `'value.replace(?,')'`. O símbolo de interrogação significa a pontuação a ser excluída e as aspas vazias indicam sua substituição por nada. A operação deve ser repetida até que toda a pontuação seja excluída do arquivo.

Embora pareça um trabalho repetitivo, ele deve ser realizado apenas uma vez, pois na

aba “*Undo/Redo*” podemos usar o botão 'Extract' para obter todo o histórico de operação na forma de um *script* no formato JSON (*JavaScript Object Notation*), que poderá ser aplicado a todos os demais arquivos a ser tratados. O *Script* pode ter uma forma parecida com o exemplo a seguir:

```
[
  {
    "op": "core/row-removal",
    "description": "Remove rows",
    "engineConfig": {
      "facets": [
        {
          "query": "Edson",
          "name": "Nome",
          "caseSensitive": false,
          "columnName": "Nome",
          "type": "text",
          "mode": "text"
        }
      ],
      "mode": "row-based"
    }
  },
  {
    "op": "core/text-transform",
    "description": "Text transform on cells in column Nome using expression
grel:value.replace(\"i\",\"\")",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "Frase",
    "expression": "grel:value.replace(\"?\",\"\")",
    "onError": "keep-original",
    "repeat": false,
    "repeatCount": 10
  }
]
```

A operação seguinte busca transformar todas as palavras para a forma de minúsculas. Podemos fazer isso com o comando '*Edit Cells>Common Transforms>To lowercase*'. Na sequência é necessário dividir as frases em palavras alocadas em diferentes linhas da tabela, procedimento que podemos realizar com o comando '*Edit Cells>Split multi-valued cells*', colocando um espaço no box que se apresenta para ser completado. Na verdade, estaremos dizendo ao programa que ele separe todas as strings que tenham um espaço de separação entre si. Deste modo teríamos uma tabela mais ou menos como a abaixo:

Identificação da Frase	Frase
1	Sim
	eu
	divido
	o
	meu
	dia
	entre
	cuidar
	da
	família
	e
	o
	trabalho

Devemos observar que ainda não temos um formato que possa ser transformado em rede, por duas razões. Em primeiro lugar, ainda não temos a relação binária para

construir uma rede. Além disso, verificamos a presença de termos muito frequentes que, entretanto, se descontextualizados da frase, não terão sentido próprio, por exemplo, os termos 'o','e', 'da', 'eu', etc. Então vamos executar mais alguns procedimentos para poder criar uma rede semântica.

O primeiro procedimento é completar as lacunas criadas na coluna 'Identificação da Frase' com o comando '*Edit Cells>Fill down*'. Isto cria uma relação entre cada palavra e a frase a que ela pertence através do número de identificação da frase. O procedimento 'copia' o número de identificação para as linhas abaixo até encontrar um novo identificador.

O próximo passo é a exclusão das '*stopwords*', palavras que não guardam significado fora do contexto da frase. Normalmente são artigos, preposições, conjunções, pronomes e advérbios.

Existem diversas listas de *stopwords* para a língua portuguesa na internet⁷, porém muitas vezes podem ser necessárias algumas adaptações, conforme o contexto da pesquisa. O comando no *OpenRefine* para fazer esta exclusão é '*Facet>Text facet*'. Isto cria uma lista das palavras da tabela na aba à direita da interface gráfica, onde as palavras da lista podem ser encontradas e deletadas as respectivas linhas no mesmo procedimento. Vale aqui o mesmo comentário feito acima.

Trata-se de uma tarefa mecânica e repetitiva que, contudo, será realizada uma única vez, pois o *OpenRefine* pode gerar um script que poderá ser executado posteriormente em outros arquivos e adaptado aos novos problemas de pesquisa.

O resultado desses procedimentos é uma tabela com duas colunas preenchidas: uma contendo o identificador da frase e outra contendo as palavras que fazem parte da frase, como podemos observar em seguida:

Identificador	Frase
1	sim
1	divido
1	dia
1	cuidar
1	família
1	trabalho
2	então
2	tô
2	casa
2	tento
2	privilegiar
2	família
2	não
2	levar
2	trabalho
2	pra
2	casa
2	não
2	precisa
2	fazer
2	trabalho
2	horários
2	trabalhar

Quando trabalhamos com pesquisas concretas, as diferentes questões formuladas pelos pesquisadores podem exigir diferentes procedimentos, e a simples aplicação mecânica destas técnicas pode não ser suficiente. Por exemplo, neste caso, como se trata de transcrição de entrevistas, temos a presença de expressões como 'pra' e 'tô' comuns no uso informal da língua e que não poderiam ser detectadas quando aplicamos o script de remoção das 'stopwords'. Portanto, além do processo automatizado, o arquivo deve ser revisto para uma “limpeza fina” manual, que pode ser facilitada pela interface do *OpenRefine*, pois nela as palavras estão agrupadas e organizadas em ordem alfabética, permitindo a intervenção em várias linhas simultaneamente.

Além da retirada das palavras não significativas, é também necessário realizar o processo de 'radicalização' (*stemming*), isto é, de agrupar palavras com o mesmo radical que ficariam separadas no arquivo porque possuem sufixos diferentes, como 'o' ou 'a' para indicar gênero, 's' para indicar plural, 'inho' para indicar diminutivo, etc. Sem esses procedimentos, expressões referentes à mesma realidade ficariam em nós diferentes na rede como, por exemplo: 'trabalho' (eu) 'trabalha' (ela). Esses procedimentos também podem ser feitos no *OpenRefine* com o comando '*Facet>Text facet*'. Dentro da aba resultante deste comando é útil usar os algoritmos oferecidos na opção 'cluster'. Este processo pode adiantar bastante o trabalho de padronização do arquivo.

Feito isso, podemos exportar o arquivo no formato .csv (*Comma-separated Values*), que pode ser lido pelo programa *Gephi*⁸, no qual vamos executar os demais procedimentos. O *Gephi* é um programa de análise de rede com aplicações em diversas áreas, como na sociologia, para estudar redes sociais; na biologia, para mapear propagação de epidemias; na comunicação social, para

entender a propagação de ideias ou conceitos, entre outras. A análise de rede toma a realidade a partir de suas relações, e a partir delas podemos calcular centralidades, densidade, estruturação em comunidades, etc. Ao tomar o discurso considerando a relação entre as palavras, a análise de redes poderá revelar as tendências de sentido produzidas em um conjunto discursivo.

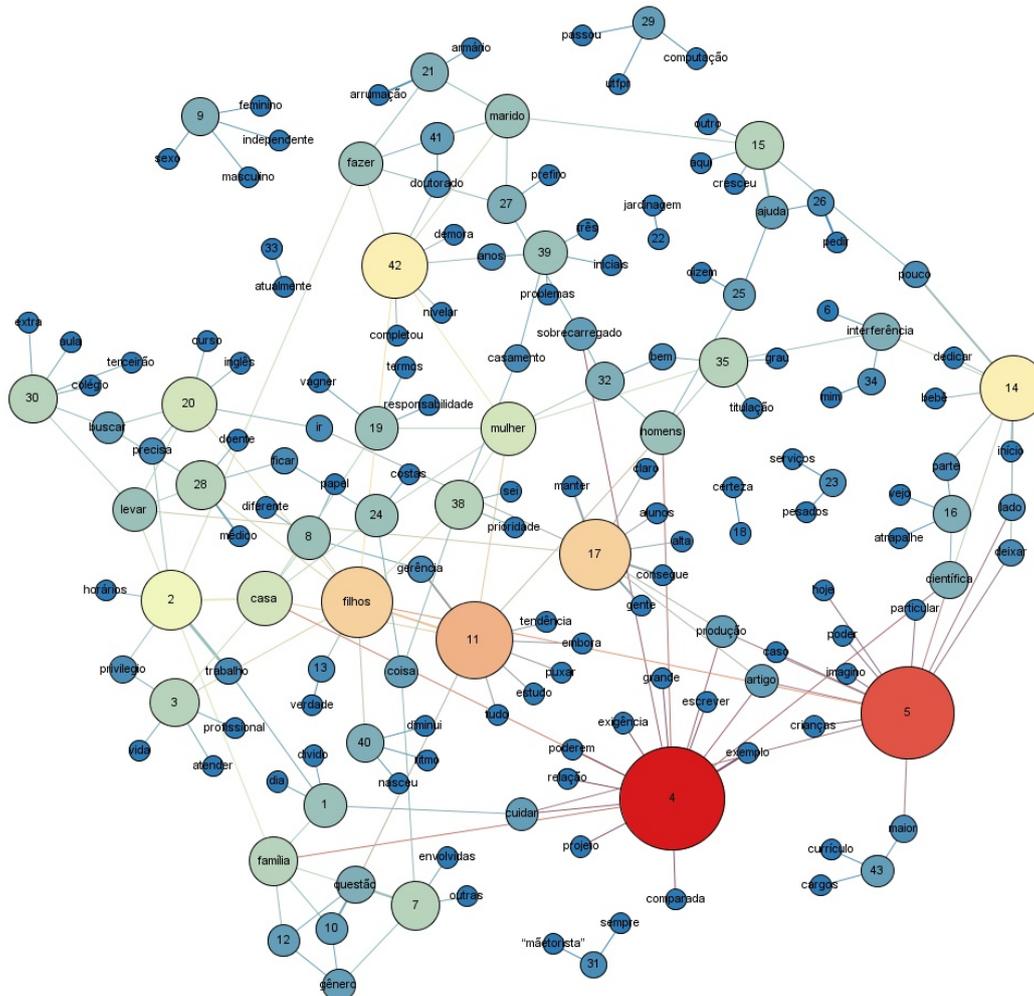
O primeiro passo para criar a rede no *Gephi* é a importação da planilha para a tabela de arestas, sendo imprescindível nomear as colunas de acordo com o padrão do *Gephi*. Assim, a primeira coluna de identificação das frases será renomeada para '*Source*', a coluna contendo as palavras será renomeada para '*Target*', e acrescentaremos uma coluna chamada '*Type*', que iremos preencher com o valor '*Undirected*', o qual determina do tipo de relação entre os nós. Outro passo é deixar ligada a opção “criar os nós automaticamente”, possibilitando ao programa extrair os nós com base na relação entre as arestas.

A tabela de nós será criada automaticamente com este procedimento, mas precisamos ainda acrescentar uma nova informação que iremos utilizar posteriormente para caracterizar o tipo de nós em relação ao seu papel na rede. Então, vamos criar uma coluna que chamaremos de '*Natureza*', que iremos preencher com os valores 'F' para o identificador das frases e 'P' para palavra. No intuito de realizar este procedimento de maneira rápida, vamos exportar o conteúdo da tabela de nós e preencher os registros em uma planilha eletrônica para, em seguida, importar os dados para a tabela de nós no *Gephi*. Um detalhe importante a ser observado é deixar desativada a opção 'forçar os nós a serem criados como novos'. Isto permitirá que o programa apenas preencha os campos faltantes sem criar novos nós, mantendo a consistência entre a tabela de nós e a tabela

de arestas.

Esses procedimentos irão gerar uma rede a partir da relação entre frases e palavras. Podemos visualizar esta rede a seguir:

Rede de Frases e Palavras



Organização Edson Armando Silva usando o programa Gephi.

Esta rede ainda não nos traz muitas informações, apenas marca quais palavras significativas foram usadas numa determinada frase e em quais outras elas foram também utilizadas. Isto acontece porque os nós que representam as frases 'poluem' o grafo e ocultam a topologia que uma rede de palavras poderia revelar.

Com a finalidade de representar graficamente os sentidos, precisamos ainda executar dois procedimentos. O primeiro será projetar a rede bipartida (*dual mode*) em uma rede monopartida (*one mode*). No nosso caso,

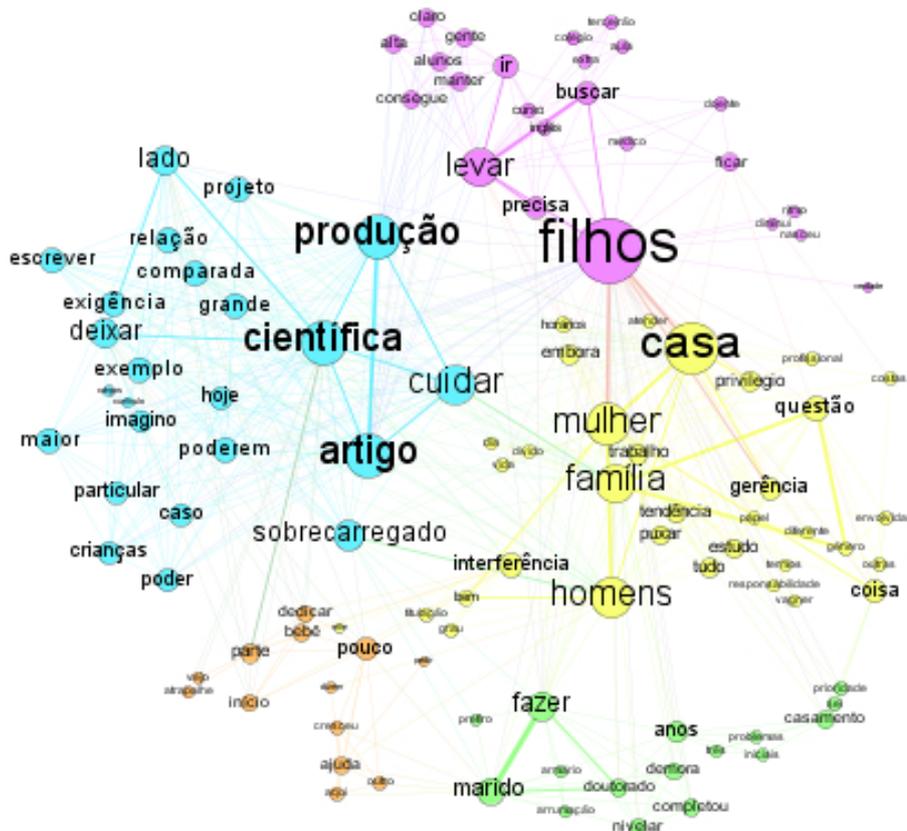
iremos transformar uma rede baseada na relação FRASE-PALAVRA em uma rede baseada na relação PALAVRA-PALAVRA. A projeção irá assim criar uma aresta entre as palavras que estão na mesma frase. Podemos exemplificar com a frase 23, que contém dois termos significativos: serviços-23-pesados. A projeção elimina a codificação da frase '23' e

cria uma ligação direta: serviços-pesados. Para executar este procedimento, iremos utilizar um plugin criado por Jaroslav Kuchar chamado 'Multimode Networks Transformations'⁹.

O procedimento seguinte será a aplicação do algoritmo de detecção de comunidades. Comunidades, também chamadas de módulos ou clusters, são conjuntos de nós cuja relação intra é mais densa do que a relação extracomunidade. Se estivéssemos analisando redes sociais a partir da comunicação por e-mail, a detecção das comunidades seria a identificação das “panelinhas” ou grupos que se comunicam mais intensamente entre si. Neste caso, o algoritmo irá identificar o conjunto de palavras que costumam aparecer juntas, revelando os sentidos produzidos nos discursos.

A rede abaixo foi produzida por um conjunto de discursos produzido por mulheres ao responderem à seguinte questão: “Achas que o teu papel social de gênero interfere na tua produtividade científica?” Estas entrevistas, que estamos usando como exemplos, foram realizadas pelo pesquisador Vagner André Morais Pinto, integrante do GETE (Grupo de Estudos Territoriais) em um projeto intitulado “O Gênero como Componente da Produtividade Científica no Espaço Acadêmico da UEPG”¹⁰. Ele entrevistou seis casais de professores da Universidade Estadual de Ponta Grossa, todos submetidos a idênticas regras funcionais e pertencendo à mesma carreira acadêmica. O exercício seguinte busca revelar os caminhos de sentido produzidos nos discursos femininos e masculinos sobre a mesma realidade.

Discursos Femininos: frequência e relação



Organização Edson Armando Silva usando o programa Gephi.

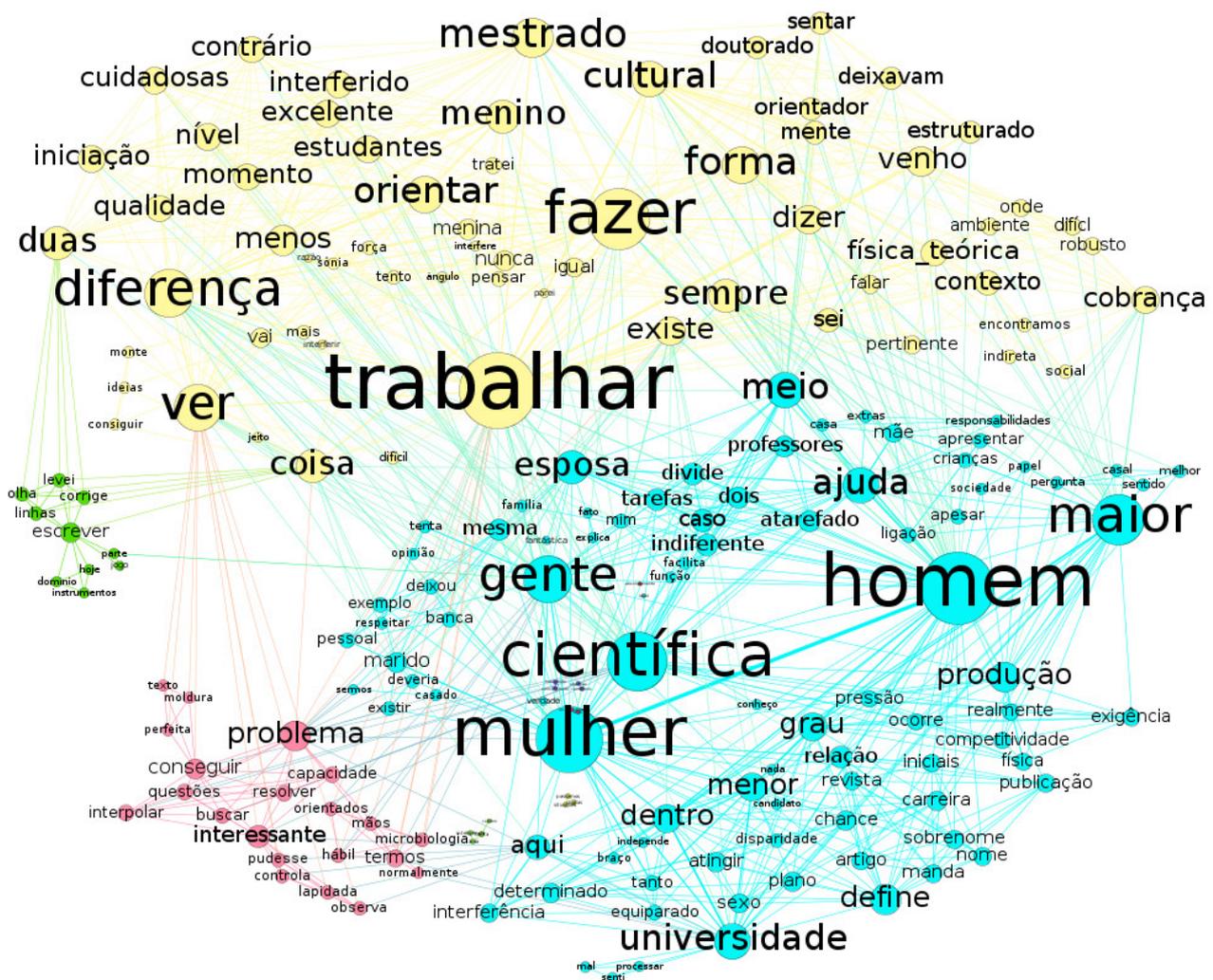
Ofício, Engenho e Arte: inspiração e técnica na análise de dados qualitativos

A topologia da rede revela uma tensão entre três polos: um dos polos tem como palavra de maior centralidade o termo 'filhos'; o outro, o termo 'casa', seguido de 'mulher' e 'família'. Juntas, essas duas comunidades significam 46,4 % da rede. O terceiro polo, que tensiona os demais, tem como termos centrais as palavras 'produção', 'científica' e 'artigo' e representa 19,2 % da rede. A quarta comunidade, em termos de representatividade na rede, é a que tem as palavras 'marido' e 'fazer' e representa 12,8% da rede, mas se

coloca em uma posicionalidade marginal, se comparada com as demais.

No discurso das mulheres, a frequência e as relações no uso das palavras, três das quatro primeiras comunidades se referem ao ambiente doméstico, e uma delas reúne os termos comuns no ambiente de produção acadêmica. O significado dessa distribuição fica mais claro quando a comparamos com o grafo produzido pelo discurso dos homens:

Discursos Masculinos: frequência e relação



Organização Edson Armando Silva usando o programa Gephi.

Mesmo vivendo realidades muito próximas e compartilhando o mesmo cotidiano, o discurso produzido pelos homens produz uma estrutura bastante diferente. A palavra 'trabalhar' é o termo com maior centralidade de grau na rede, entretanto o grafo não apresenta tensões entre os termos mais frequentes, uma vez que, mesmo colocados em comunidades distintas, encontram-se posicionados no centro do grafo. As palavras 'homem', 'mulher' estão colocadas na mesma comunidade dos termos que se referem à produção acadêmica como 'científica', 'universidade', assim como os termos que expressam colaboração, como 'ajuda' ou 'divide'.

O que mais chama a atenção, ao compararmos os grafos, é a quase ausência dos termos que se referem ao cuidado dos filhos ou da casa. Apenas um pequeno núcleo em torno da palavra 'ajuda', que se liga a 'divide', 'tarefas', 'mãe' e 'crianças', é quase insignificante se comparado à sua importância no grafo produzido pelo discurso das mulheres.

A Análise através de Categorias

Como vimos acima, a análise de frequência e de relação revela estruturas presentes em um conjunto de textos, entretanto, grosso modo, uma parte dos sentidos produzidos se perde. A grande vantagem da aplicação dessa técnica é que esses sentidos são produzidos apenas a partir das falas dos depoentes ou entrevistados, sem intervenção do pesquisador e, portanto, sem pressupostos teóricos pré-formulados. Feita esta observação, temos de admitir que a capacidade humana de interpretar a frase ainda não foi igualada por qualquer procedimento automatizado.

A análise categorial implica a criação de uma nova camada sobre o texto, na qual a intervenção do pesquisador é fundamental. O

pesquisador 'traduz' o texto em unidades de sentido que possam ser medidas e comparadas. A categorização é o processo através do qual o pesquisador atribui sentidos a passagens do documento. Esta 'marcação' pode ser feita a um caractere ou ao documento como um todo, e a uma passagem podem ser atribuídas diversas categorias. Vamos a um exemplo: na expressão "Não. Acho que não, né?" que na análise frequencial fica praticamente destituída de sentido porque as 'StopWords' são retiradas da análise, podem ser atribuídas categorias que expressem que se trata do discurso de um homem e que o sentido dela é a negação da pergunta formulada.

A criação do conjunto de categorias que servirá de parâmetro para a análise depende, obviamente, da questão central que a pesquisa se propõe a responder, contudo a pesquisa também deve estar aberta ao radicalmente novo, que pode vir dos documentos encontrados ou produzidos. Nesse sentido, é útil ter como referência os marcos estruturais produzidos pelos procedimentos anteriores (análises frequencial e relacional), além de experimentar a aplicação das categorias que podem posteriormente ser aprofundadas, divididas ou ampliadas para retratar com mais fidelidade as nuances presentes na documentação.

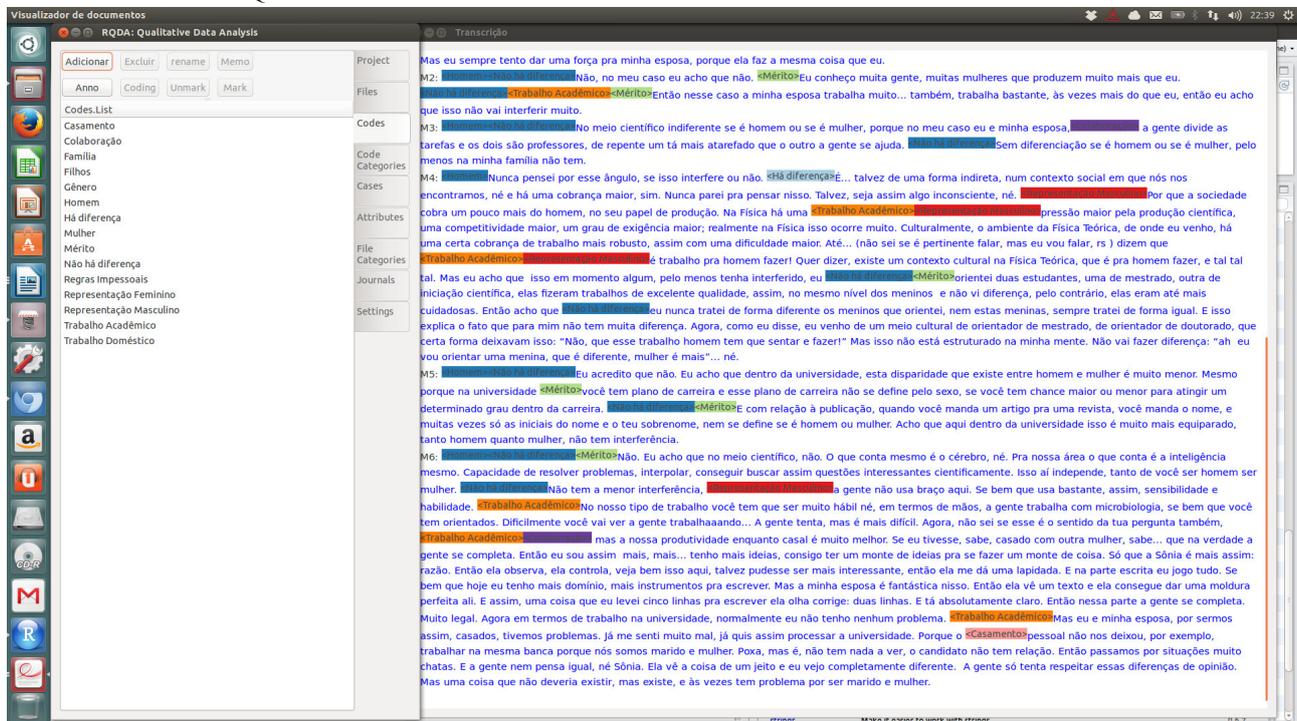
A ferramenta que escolhemos para esta fase da pesquisa é um *software* pertencente ao pacote estatístico 'R' chamado 'RQDA - a R package for Qualitative Data Analysis'¹¹. Trata-se de uma ferramenta livre com as principais funcionalidades exigidas pela análise de dados qualitativos, que funciona nos principais sistemas operacionais (*Windows*, *Linux* e *MacOSX*) e possui uma interface gráfica simples e intuitiva. O mais interessante, porém, é que esta plataforma possui uma arquitetura aberta e armazena os dados num banco de dados chamado

SQLITE¹². Assim, o pesquisador pode trabalhar com os comandos oferecidos pela interface gráfica ou fazer cruzamentos mais sofisticados usando a linguagem de programação em banco de dados chamada SQL. Neste trabalho vamos usar um misto das duas coisas, porque vamos fazer a categorização no RQDA, extrair os dados com um comando SQL simples e exportar os dados para análise e visualização no Gephi.

Depois de conceber o projeto, importar os

arquivos e criar os códigos (ou categorias), o processo de categorização é simples. Basta selecionar o trecho desejado, clicar na categoria que se deseja atribuir ao texto e clicar no botão 'Mark'. Trabalhar com um *software* facilita o trabalho de reanálise que se torna necessário na medida em que nos aprofundamos no conhecimento da documentação. Abaixo segue uma imagem da interface gráfica do RQDA e de uma parte do texto codificado:

Interface Gráfica do RQDA



Screenshot feita por Edson Armando Silva.

Nesta imagem temos duas janelas: uma com os principais comandos da interface gráfica, organizada em abas, e outra com a transcrição das entrevistas com a marcação dos códigos. O interessante é que ao criarmos um código e marcarmos o texto com ele, tanto o código como o texto são armazenados em um banco de dados, de onde eles podem ser extraídos com um simples comando como este: `'SELECT name, seltext FROM freecode,`

`coding WHERE freecode.id=coding.cid AND coding.fid=3'`. Traduzindo para o português, estamos pedido que o banco nos forneça o campo 'name' da tabela 'freecode', onde estão armazenados os códigos, e o campo 'seltext' da tabela 'coding', onde está colocado o texto marcado com o respectivo código.

Isso resulta em uma tabela com códigos e frases como a abaixo:

Os procedimentos a seguir são os mesmos

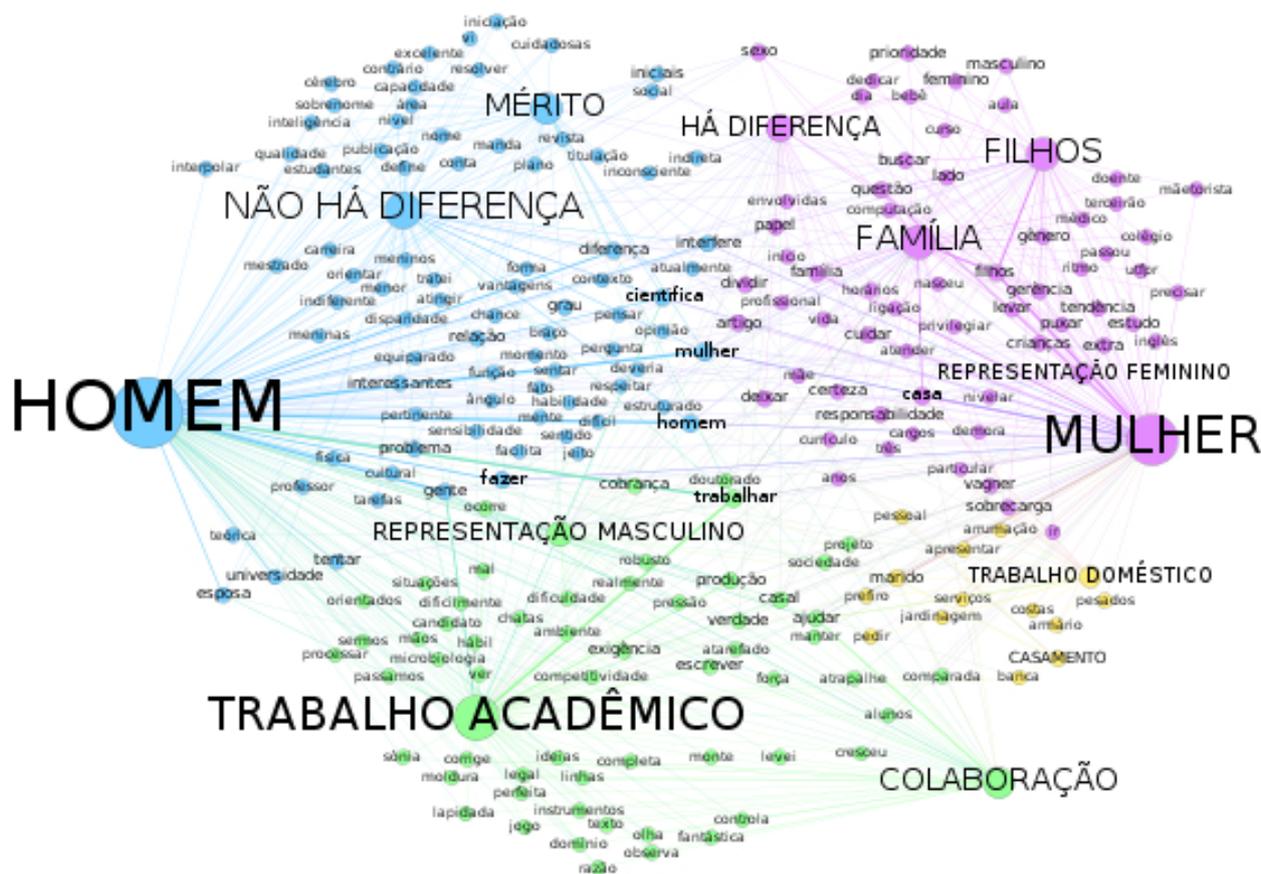
Categoria	Palavras
Homem	Não. Acho que não né. Tem algumas vantagens... Eu ajudo bastante em casa, mas sempre a mãe tem responsabilidades extras né. As crianças têm uma ligação um pouco maior com a mãe, apesar de eu sempre me apresentar pra ajudar. Então eu acho que ajuda, facilita um pouco pra mim né ser homem em função disso né. Mas eu sempre tento dar uma força pra minha esposa, porque ela faz a mesma coisa que eu.
Mulher	Sim, eu divido o meu dia entre cuidar da família e o trabalho. Então quando eu tô em casa tento privilegiar a família e não levar trabalho pra casa, não precisar fazer trabalho fora dos horários em que eu teria que trabalhar. E quando minhas filhas tão em casa eu privilegio as duas do que atender a vida profissional [...] Sobrecarregam porque a nossa exigência com relação à produção científica é grande, eu sou comparada com homens, por exemplo, quando estão em casa às vezes não estão cuidando da família pra poderem escrever um artigo, um projeto, por exemplo. No meu caso em particular, eu deixo o artigo de lado pra poder cuidar das crianças, então se eu não tivesse filhos, eu imagino que a minha produtividade científica seria muito maior do que hoje.
Família	Acho que é diferente você ter filho, ter esse papel, ter uma casa pra gerenciar e você não ter... independente do sexo ser masculino ou feminino.

descritos acima: retirada da pontuação, transformação em minúsculas, retirada dos termos não significativos (*Stop Words*), unificação dos termos com o mesmo radical, separação por palavras, etc. Neste caso, para diferenciar na rede o que representa 'categorias' do que são os termos do discurso, transformamos a coluna de categorias em maiúsculo e mantemos a coluna de palavras

em minúsculo.

Feitos os procedimentos de padronização e importadas as planilhas no *Gephi*, o grafo resultante dos procedimentos é o seguinte:

Rede de Categorias e Palavras



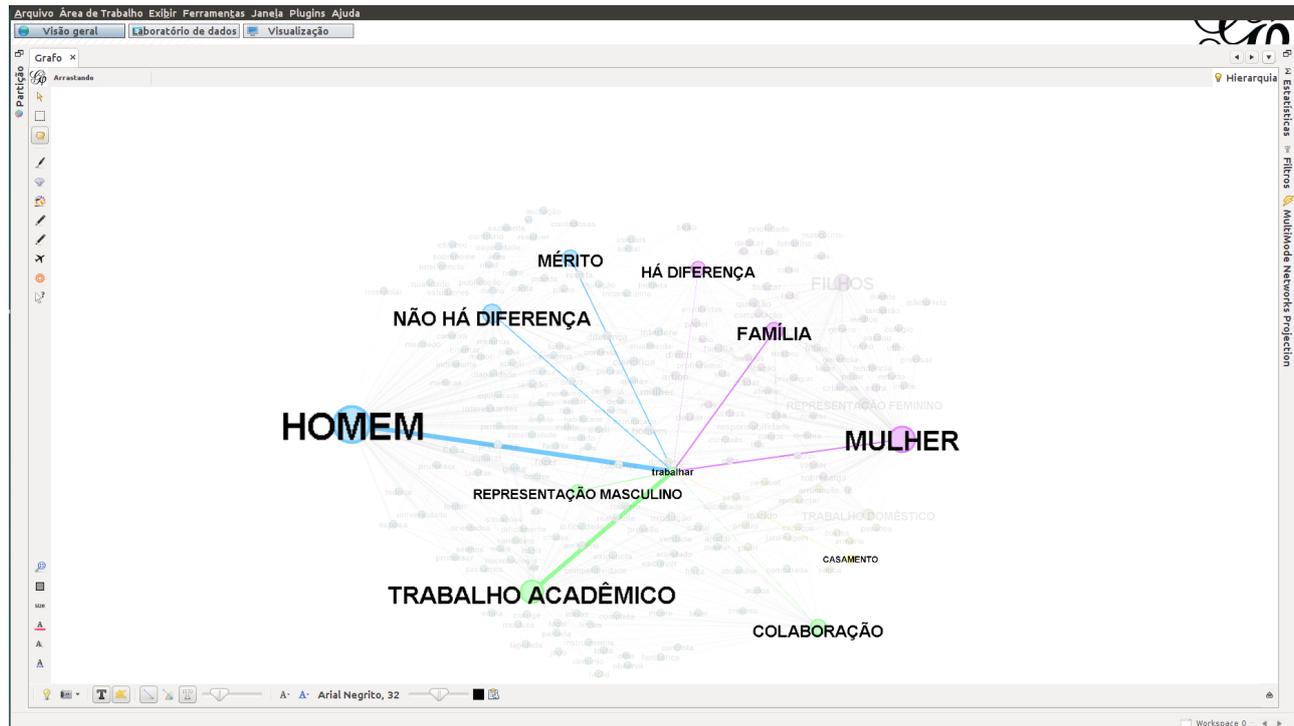
Organizado por Edson Armando Silva no programa Gephi.

As categorias servem de orientação no uso das palavras. Ao mesmo tempo em que trazem as palavras mais usadas para o centro do grafo, como 'homem', 'mulher', 'científica', 'trabalhar', aproximam por gravidade os termos mais próximos de determinadas categorias. Podemos então responder a

perguntas, como: 'Quais as palavras mais usadas quando falamos de determinado assunto?' Para exemplificar a utilidade desse recorte, podemos destacar uma categoria e as palavras a ela relacionadas como, por exemplo:

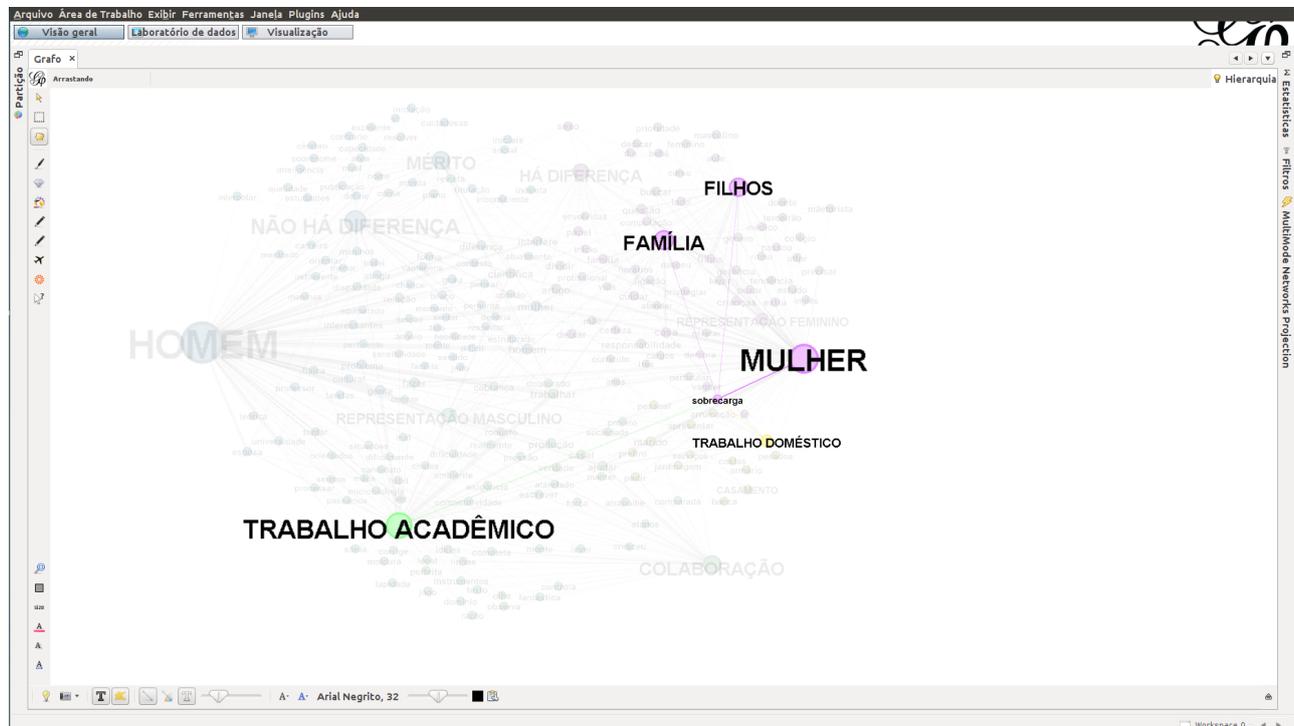
Ofício, Engenho e Arte: inspiração e técnica na análise de dados qualitativos

Interface do *Gephi*: seleção de 'trabalhar'



Organizado por Edson Armando Silva no programa *Gephi*.

Interface do *Gephi*: seleção de 'sobrecarga'



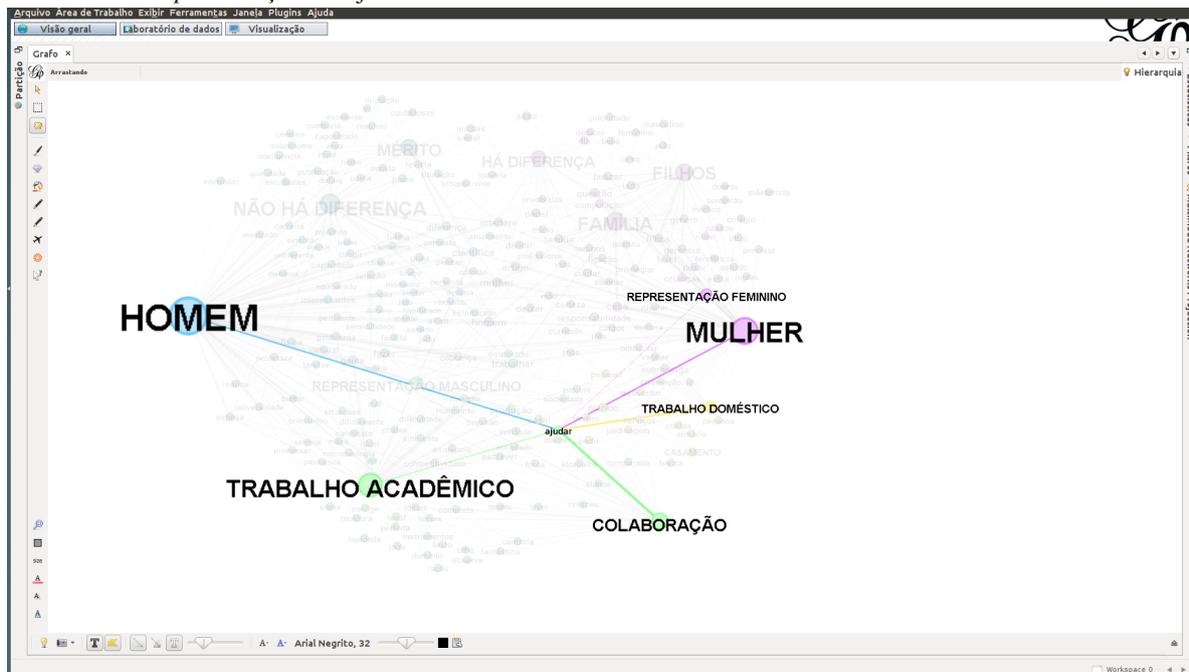
Organizado por Edson Armando Silva no programa *Gephi*.

Ofício, Engenho e Arte: inspiração e técnica na análise de dados qualitativos

Observamos que o emprego da palavra 'trabalhar' envolve os significados produzidos nos discursos masculino e feminino, ao mesmo tempo em que se relaciona com o trabalho acadêmico e com a categoria 'família', embora apresente uma ligação mais

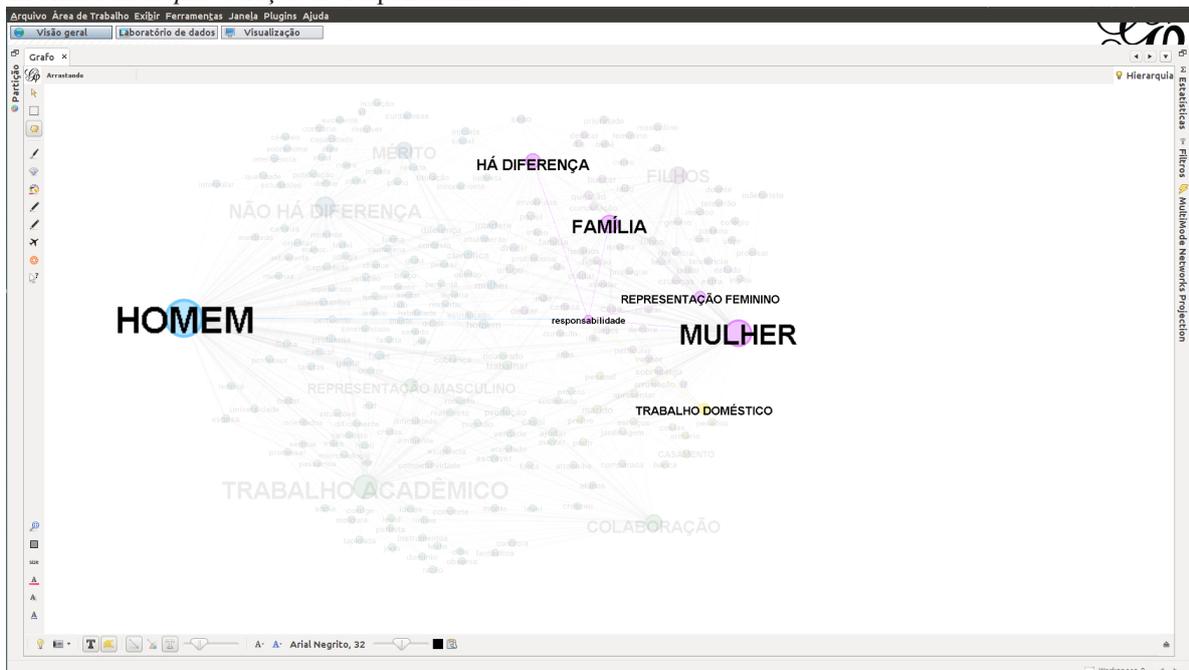
forte com os discursos produzidos no campo masculino. A palavra 'sobrecarga', todavia, está posicionada no campo feminino, tensionando o trabalho acadêmico, a família, o cuidado dos filhos e o trabalho doméstico.

Interface do *Gephi*: seleção de 'ajudar'



Organizado por Edson Armando Silva no programa *Gephi*.

Interface do *Gephi*: seleção de 'responsabilidade'



Organizado por Edson Armando Silva no programa *Gephi*.

Ofício, Engenho e Arte: inspiração e técnica na análise de dados qualitativos

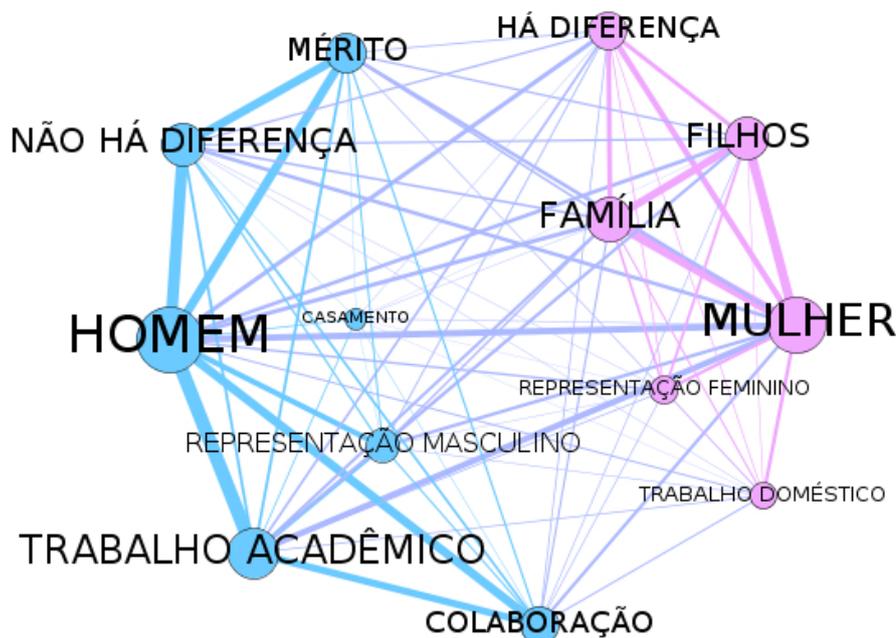
A comparação dos contextos de uso dos termos 'ajudar' e 'responsabilidade' também é reveladora porquanto 'ajudar' articula os campos do trabalho acadêmico e do trabalho doméstico e familiar, mas o uso do termo 'responsabilidade' está restrito ao campo feminino, marcando um comprometimento do discurso feminino com a esfera familiar.

Finalmente, um último procedimento deve ser executado. Embora a rede de categorias e palavras permita diversos recortes e análises, também é possível realizar a transformação da rede 'dual mode' em uma rede 'one mode': ligando 'categorias' a 'categorias' através das palavras. Este último procedimento revela a estrutura discursiva de um corpus documental, como podemos notar no grafo abaixo:

Este grafo permite a visualização da estrutura dos discursos masculino e feminino sobre sua situação de gênero e a produção acadêmica. A aplicação do algoritmo de reconhecimento de comunidades colore automaticamente o campo discursivo masculino e o campo discursivo feminino. No discurso masculino não há diferença entre homens e mulheres no trabalho acadêmico. O elemento central na organização do campo se situa nos critérios de mérito sendo, portanto, impessoal. Para os homens, o que marca o relacionamento entre homens e mulheres neste campo é o da colaboração e ajuda mútua, tanto no campo do trabalho acadêmico como no trabalho doméstico.

No discurso feminino é evidenciada a tensão entre o trabalho acadêmico e o

Rede de Categorias



Organizado por Edson Armando Silva no programa *Gephi*.

comprometimento com a família, os filhos e a organização doméstica. Se, por um lado, é verdade que não existem barreiras institucionais à produção acadêmica da mulher, a responsabilidade cultural com o cuidado dos filhos demanda tempo e atenção que concorrem com o trabalho acadêmico.

Retornando ao Texto

Reveladas as tendências estruturais na produção discursiva de um grupo, podemos então retornar ao fundo documental para selecionar as passagens lapidares que confirmam, ou mostram os limites das interpretações apontadas nas estruturas representadas nos grafos, mas com a vantagem de podermos revelar a posição que as expressões do depoente ocupam na estrutura discursiva do conjunto. Vejamos então o seguinte: nossa depoente F5¹³, mulher, nos diz que: “Atualmente não. Pra mim não tem interferência nenhuma. Eu acho que talvez o grau de titulação interfira bem mais do que ser homem ou mulher.” Este depoimento parece estar em contradição com a tendência do conjunto. Quando buscamos informações adicionais no banco de dados, verificamos que se trata de uma mulher recentemente casada e ainda sem filhos. Ora, isto tem correspondência com outros enunciados, por exemplo, F3, mulher, nos informa que: “quando eu tive o bebê interferiu um pouco porque eu tive que me dedicar mais a ele e aí deixar um pouco a parte científica de lado...”. Podemos ainda confirmar a mesma tendência com a declaração de F2, também mulher, que, segundo ela, as mulheres “tenham a tendência de puxar mais a questão de gerência de casa, de gerência de filhos, estudo de filho, tudo...”. O conjunto das citações, então, tende a confirmar a estrutura discursiva apresentada no grafo. O núcleo da tensão se encontra entre o tempo e a atenção

destinados à produção acadêmica e à atenção aos filhos e à casa.

Quando voltamos aos discursos feitos pelos homens, temos um leque de possibilidades no qual um dos polos destaca a impessoalidade na carreira acadêmica, por exemplo, o depoente M5, homem, afirma que: “na universidade você tem plano de carreira e esse plano de carreira não se define pelo sexo, se você tem chance maior ou menor para atingir um determinado grau dentro da carreira. E com relação à publicação, quando você manda um artigo pra uma revista, você manda o nome, e muitas vezes só as iniciais do nome e o teu sobrenome, nem se define se é homem ou mulher”. No outro polo, M1, homem, de maneira tímida reconhece “certa” desigualdade nesses termos: “Eu ajudo bastante em casa, mas sempre a mãe tem responsabilidades extras, né?” As crianças têm uma ligação um pouco maior com a mãe.”

As posições masculina e feminina em relação aos polos do trabalho acadêmico e do trabalho doméstico ficam evidentes quando observamos o uso das palavras 'ajuda' e 'responsabilidade'. Embora vários depoentes afirmem 'ajudar' nas tarefas domésticas, como no depoimento de F3, mulher: “e daí também com o meu marido aqui, um ajudando o outro então eu acho que isso ajuda um pouco. Acho que agora eu não vejo que isso atrapalhe na parte científica, não. A gente consegue manter, claro, não é uma produção muito alta, mas a gente consegue manter”, ou, menos enfaticamente, como no depoimento de F4, também mulher: “os homens dizem que ajudam... se você pedir eles ajudam, mas eu não sou de pedir, né?” Eu prefiro fazer, então eu me acho muito mais sobrecarregada do que meu marido. Se filho fica doente eu tenho que levar no médico, buscar no médico...”.

É interessante observar que a palavra

'ajuda' também aparece no contexto do trabalho acadêmico, como na afirmação de M3, homem: “No meio científico indifere se é homem ou se é mulher, porque no meu caso eu e minha esposa, a gente divide as tarefas e os dois são professores, de repente um tá mais atarefado que o outro, a gente se ajuda...”

Por outro lado, ao falar da família, dos filhos e das tarefas domésticas, as mulheres sempre empregam a expressão 'responsabilidade', como no depoimento de F4, mulher: “a responsabilidade da casa, [...] ela quase 100% das vezes é da mulher. Então se o meu filho precisa ir ao curso de inglês, eu levo, eu busco. Se é pra fazer alguma arrumação, mesmo que seja no armário do meu marido, eu arrumo, né?”. “Jardinagem, eu! Às vezes até serviços mais pesados, né?” Há uma evidente diferença de intensidade entre os termos 'ajuda' e 'responsabilidade'. Esta diferença é que explica que a tensão entre família e trabalho acadêmico apareça apenas no discurso das mulheres e esteja quase completamente ausente no discurso masculino. É a responsabilização feminina pela maternagem, noção profundamente enraizada em nossa cultura ocidental, que impede a ocultação da demanda doméstica por tempo e atenção. Tempo e atenção que também são necessários para o trabalho acadêmico... talvez tanto quanto a disponibilidade de laboratórios.

Na perspectiva masculina, apesar do compartilhamento do cotidiano, é possível ocultar essa tensão na menor intensidade da demanda por 'ajuda'. Esta douta ignorância permite reduzir as exigências da carreira universitária aos mecanismos impessoais do mérito acadêmico, retirando-as da amplitude maior da vida e circunscrevendo-as ao espaço institucional da produção e da produtividade.

Considerações Finais

O objetivo deste trabalho, na verdade, não

é discutir os resultados que possam ser obtidos na análise deste conjunto específico de fontes, mas apresentar as potencialidades de associação de um conjunto de procedimentos ao texto para revelar sua estrutura e ao mesmo tempo manter e explorar as condições das falas dos diversos atores. A associação criativa das técnicas de análise de frequência, análise de redes e procedimentos de categorização permitiu explorar os vários significados produzidos nesses discursos, e ao mesmo tempo verificar a coerência entre as representações resultantes dos procedimentos e as falas dos entrevistados.

Além dessas considerações de caráter mais geral, procuramos descrever os procedimentos passo a passo, apresentando as características dos *softwares* que possibilitam e aumentam a velocidade desse trabalho. Assim o leitor poderá, com alguma paciência, explorar os mesmos procedimentos para diferentes conjuntos de fontes.

¹ AL-HINDI, Karen Falconer; KAWABATA, Hope. Toward a More Fully Reflexive Feminist Geography. In: *Feminist Geography in Practice: Research and Methods*. Oxford UK: Blackwell Publishers Inc., 2002.

² BARDIN, Laurence. *Análise de conteúdo*. Lisboa: Edições 70, 1995, p. 9.

³ KLEMETTINEN, Jenni. Les procédés rhétoriques de répétition dans l'appel du 18 juin 1940 de Charles de Gaulle. [s.l.: s.n.], 2009.

⁴ PARANYUSHKIN, Dmitry. Identifying the Pathways for Meaning Circulation using Text Network Analysis. [s.l.]: Nodus Labs, 2011, p. 26.

⁵ THE DOCUMENT FOUNDATION. LibreOffice, [s.l.: 4.3.], 2014.

⁶ HUYNH, David. OpenRefine. [s.l.]: Open Source Community, 2014.

⁷ Neste caso, trabalhamos com a lista existente em <http://snowball.tartarus.org/algorithms/portuguese/stop.txt>. Consulta feita em 10/8/2014.

⁸ BASTIAN, Mathieu; HEYMANN, Sebastien; JACOMY, Mathieu. **Gephi: an Open Source Software for Exploring and Manipulating Networks**, [s.l.: s.n.], 2009.

⁹ KUCHAR, Jaroslav. **Multimode Networks plugin for Gephi**. [s.l.]: Open Source Community, 2014.

¹⁰ PINTO, Vagner André Morais. **O Gênero como componente da produtividade científica no espaço acadêmico da UEPG**. Ponta Grossa, 2014.

¹¹ HUANG, Ronggui. **RQDA: R-based Qualitative Data Analysis**, [s.l.: s.n.], 2012.

¹² SQLITE CONSORTIUM. **SQLite** [s.l.]: Open Source Community, 2014.

¹³ Todas as entrevistas e transcrições foram feitas por PINTO, Vagner André Morais. **O Gênero como componente da produtividade científica no espaço acadêmico da UEPG**. Ponta Grossa: UEPG, 2014.

Referências

AL-HINDI, Karen Falconer; KAWABATA, Hope. **Toward a More Fully Reflexive Feminist Geography**. In: MOSS, Pamela. **Feminist Geography in Practice: Research and Methods**. Oxford UK: Blackwell Publishers Inc., 2002, p. 103-115.

BARDIN, Laurence. **Análise de conteúdo**. Lisboa: Edições 70, 1995.

BASTIAN, Mathieu; HEYMANN, Sebastien; JACOMY, Mathieu. **Gephi: an Open Source Software for Exploring and Manipulating Networks**. [s.l.: s.n.], 2009. Disponível em: <http://www.aaii.org/ocs/index.php/ICWSM/09/paper/view/154>. Acesso em: 2 set. 2014.

HUANG, Ronggui. **RQDA: R-based Qualitative Data Analysis**. [s.l.: s.n.], 2012. (R package). Disponível em: <http://rqda.r-forge.r-project.org/>. Acesso em: 22 out. 2014.

HUYNH, David. **OpenRefine**. [s.l.]: Open Source Community, 2014. Disponível em: <http://en.wikipedia.org/wiki/OpenRefine>.

KLEMETTINEN, Jenni. **Les procédés rhétoriques de répétition dans l'appel du 18 juin 1940 de Charles de Gaulle**. [s.l.: s.n.], 2009. Disponível em: <https://jyx.jyu.fi/dspace/handle/123456789/20102>. Acesso em: 27 jul. 2014.

KUCHAR, Jaroslav. **Multimode Networks plugin for Gephi**. [s.l.]: Open Source Community, 2014. Disponível em: <https://github.com/jaroslav-kuchar/Multimode-Networks>.

PARANYUSHKIN, Dmitry. **Identifying the Pathways for Meaning Circulation using Text Network Analysis**. In: [s.l.]: **Nodus Labs**, 2011, p. 26. Disponível em: <http://noduslabs.com/publications/Pathways-Meaning-Text-Network-Analysis.pdf>. Acesso em: 26 jul. 2014.

PINTO, Vagner André Morais. **O Gênero como componente da produtividade científica no espaço acadêmico da UEPG**. Ponta Grossa, 2014.

SQLITE CONSORTIUM. **SQLite**. [s.l.]: Open Source Community, 2014. Disponível em: <http://www.sqlite.org/index.html>.

THE DOCUMENT FOUNDATION. **LibreOffice**. [s.l.: s.n.], 2014. Disponível em: <http://www.libreoffice.org>.

**Recebido em 07 de junho de 2015.
Aceito em 01 de novembro de 2015.**