

Gephi

Um software open source de manipulação e visualização de grafos

Oficina Gephi

Mapeando e analisando a vida das redes sociais



www.labic.net

Pesquisadores:

Allan Cancian Marquez
Bianca Bortolon Gonçalves
Jean Maicon Rickes Medeiros
Nelson Aloysio Reis

Índice

1. Introdução

1.1 O que é Gephi	3
1.2 Formatos de entrada e saída	3
1.3 Formatos para exportação	3
1.4 Glossário de termos	4
1.5 Links úteis	4
1.6 Exemplos de grafos	5

2. Preparação

2.1 Prerequisites do programa	9
2.2 interface gráfica	9
2.3 Plugins	12

3. Teoria

3.1 Estatísticas comumente utilizadas	12
3.2 Filtros comumente utilizados	13

4. Prática

Criando um <i>dataset</i> e exemplos de crawlers	14
--	----

1. Introdução

1.1 O que é Gephi?

É uma plataforma *open source* para a visualização e manipulação de grafos dinâmicos e hierárquicos, incluindo todos os tipos de redes e sistemas complexos. É uma ferramenta similar ao Adobe Photoshop: em vez de imagens, o usuário é livre para modificar o visual e interagir com a estrutura das redes que importa, posteriormente adicionando filtros para ressaltar os aspectos que deseja e enfim exportando o resultado final em SVG, PNG ou PDF.

Sua principal função é servir como método de análise de dados, elaboração de hipóteses, descoberta de padrões sociais e de comportamento e isolamento de estruturas importantes dentro de redes hierarquizadas. Também é largamente utilizado na visualização de redes de relações entre indivíduos e dos conteúdos que (re)produzem, tornando-se uma ferramenta poderosa para pesquisas de opinião e no campo da cibercultura.

O download pode ser feito no site oficial, onde também encontram-se mais informações e novidades sobre o programa, assim como plugins com funções adicionais interessantes.

Nesta apostila, estaremos trabalhando com a versão 0.8.2 beta do programa, lançado em 3 de janeiro de 2013.

1.2 Formatos de entrada e saída

- **.GEXF, .GDF, .GraphML:**

Grafos com nós e arestas definidos; caso o único programa a ser usado seja o Gephi, recomenda-se utilizar o .GEXF por possuir mais recursos.

- **.Gephi:**

Grafos com nós, arestas, estrutura e visual definidos; é o projeto da rede conforme desenvolvido pelo usuário (equivalente ao .psd do Photoshop);

- **.CSV (*Comma Separated Values*):**

Tabela de nós e/ou arestas separada por por vírgulas, ponto e vírgulas ou pipes (caractere "|");

1.3 Formatos para exportação

- **.PNG (*Portable Networks Graphics*):**

Formato de imagem de pequeno tamanho e consequente versatilidade;

- **.SVG (*Scalable Vector Graphics*), PDF (*Portable Document Format*):**

Ambos formatos de imagens vetoriais, ou seja, apropriados para ilustrações, mapas, grafos, logos, etc., em especial para redimensionamento e zoom.

1.4 Glossário de termos

- **Nó:** chamado também de *node*, vértice;
- **Aresta:** chamado também de *link*, ligação, *edge*. Ela pode ser:
 - *Simples:* quando a ligação não possui peso;
 - *Ponderada:* quando diferentes arestas possuem pesos distintos, ou seja, ligações mais ou menos fortes, a depender do que representam na rede;
- **Cluster:** conjunto de nós fortemente conectado, grupo de nós;
- **Grafo:** representação gráfica de uma rede interativa. Ele pode ser:
 - *Direcional ou dirigido:* quando um nó possui ligação com outro não necessariamente recíproca (ex: Instagram, Twitter);
 - *Não direcional ou não dirigido:* quando a ligação entre dois nós é necessariamente recíproca (ex: Facebook e Orkut).
- **Crawl:** importado do inglês, o verbo “crawlear” (“*to crawl*”) neste contexto significa minerar/coletar dados da web, de mídias, de redes sociais etc.



Fig. 1: nó

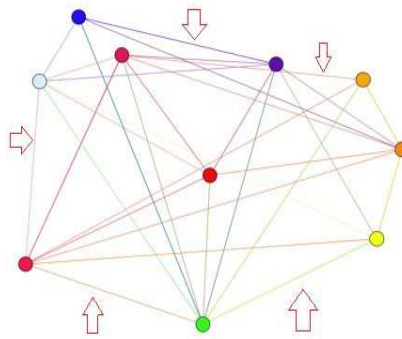


Fig. 2: arestas (indicadas por setas) de um grafo

1.5 Links úteis

- Site oficial: www.gephi.org;
- Blog: www.gephi.org/blog;
- Fóruns: www.forum.gephi.org;
- Marketplace (plugins): www.marketplace.gephi.org/plugins
- Tutoriais oficiais e feitos por usuários: www.gephi.org/users;
- Twitter: www.twitter.com/Gephi
- Wikipédia: www.wiki.gephi.org;
- Grupo no Facebook: www.facebook.com/groups/gephi/

1.6 Exemplos de grafos

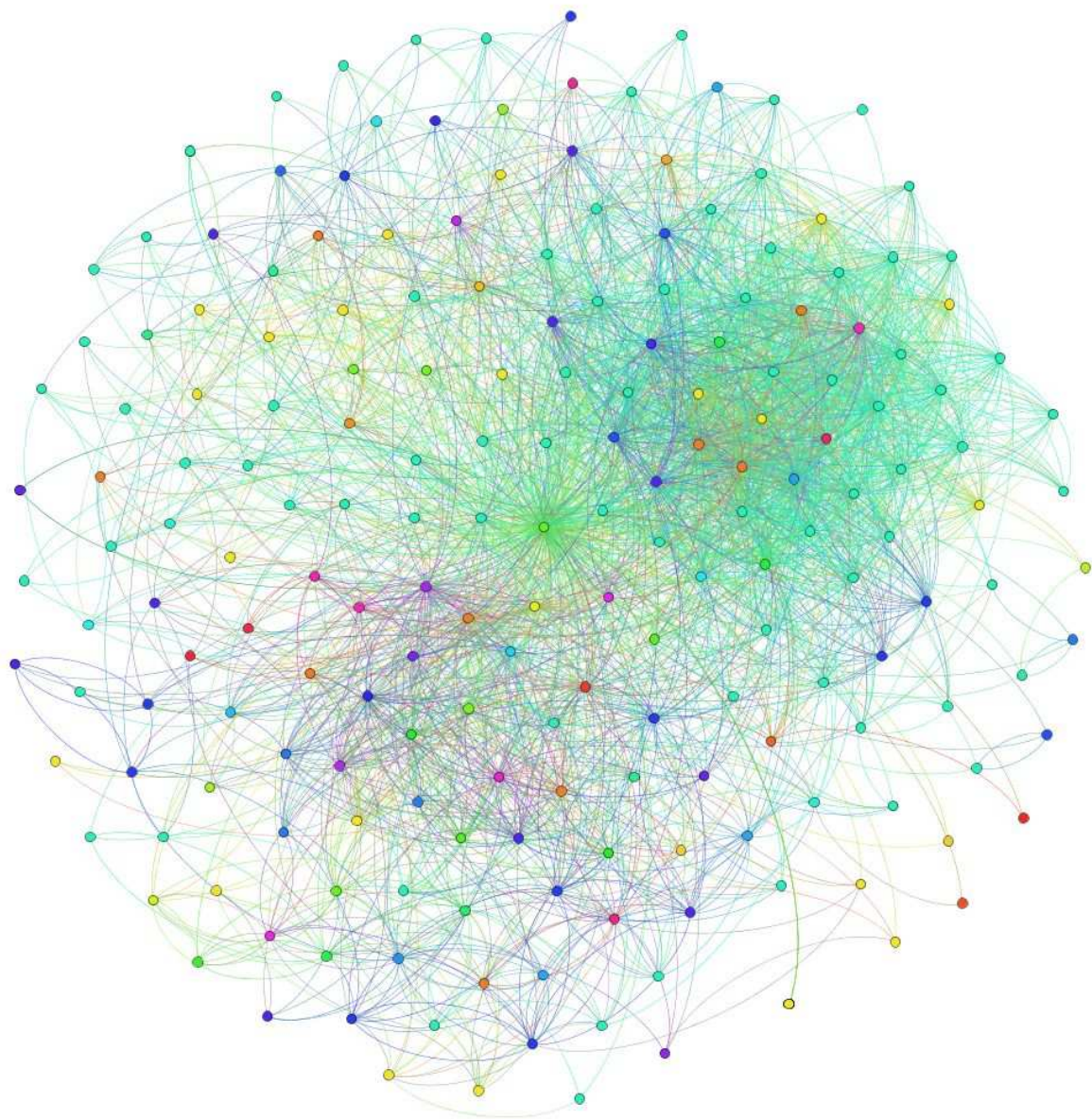


Fig. 3: Grafo não-direcionado de uma rede

Observam-se nós representados por cores distintas e arestas cujas cores são o resultado da interação destes nós - opção padrão no Gephi.

Por todos possuírem o mesmo peso, o grafo objetiva, portanto, uma representação da rede de relações de nós e dos grupos a que pertencem. Para o layout, utilizou-se do algoritmo de distribuição *Fruchterman-Reingold*, cujo objetivo é uma estrutura o mais limpa, visualmente, possível.

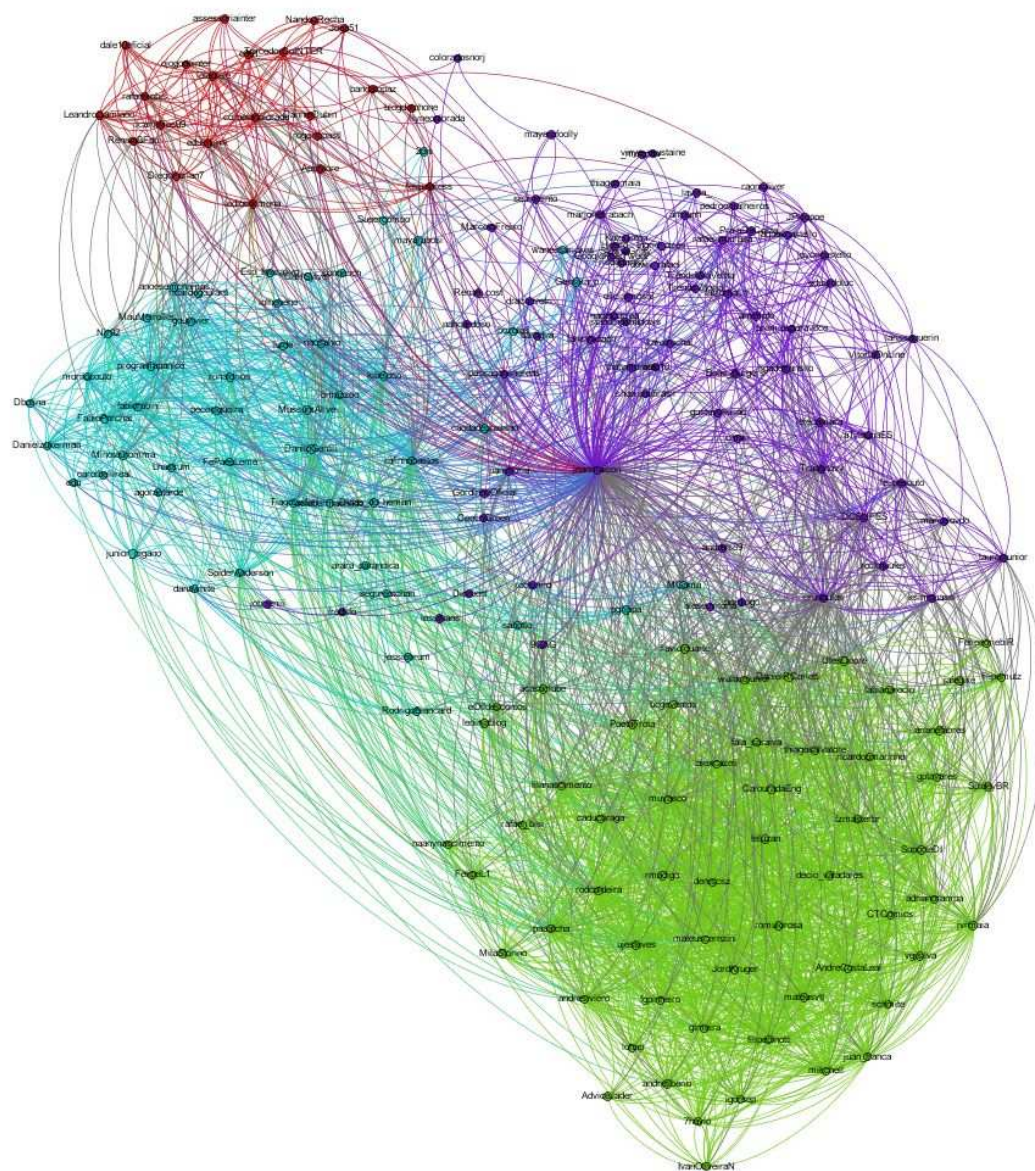


Fig. 4: Grafo de uma rede extraída do Twitter

Observam-se quatro diferentes clusters de nós conectados a um componente central, responsável por ser o hub neste caso.

Para o layout, utilizou-se do algoritmo de distribuição *Force Atlas*, cujo objetivo é representar espacialmente as ligações entre nós pertencentes a distintos grupos, de forma a criar um “mapa” de suas interações.

Seu sucessor o excede em capacidades de visualização.

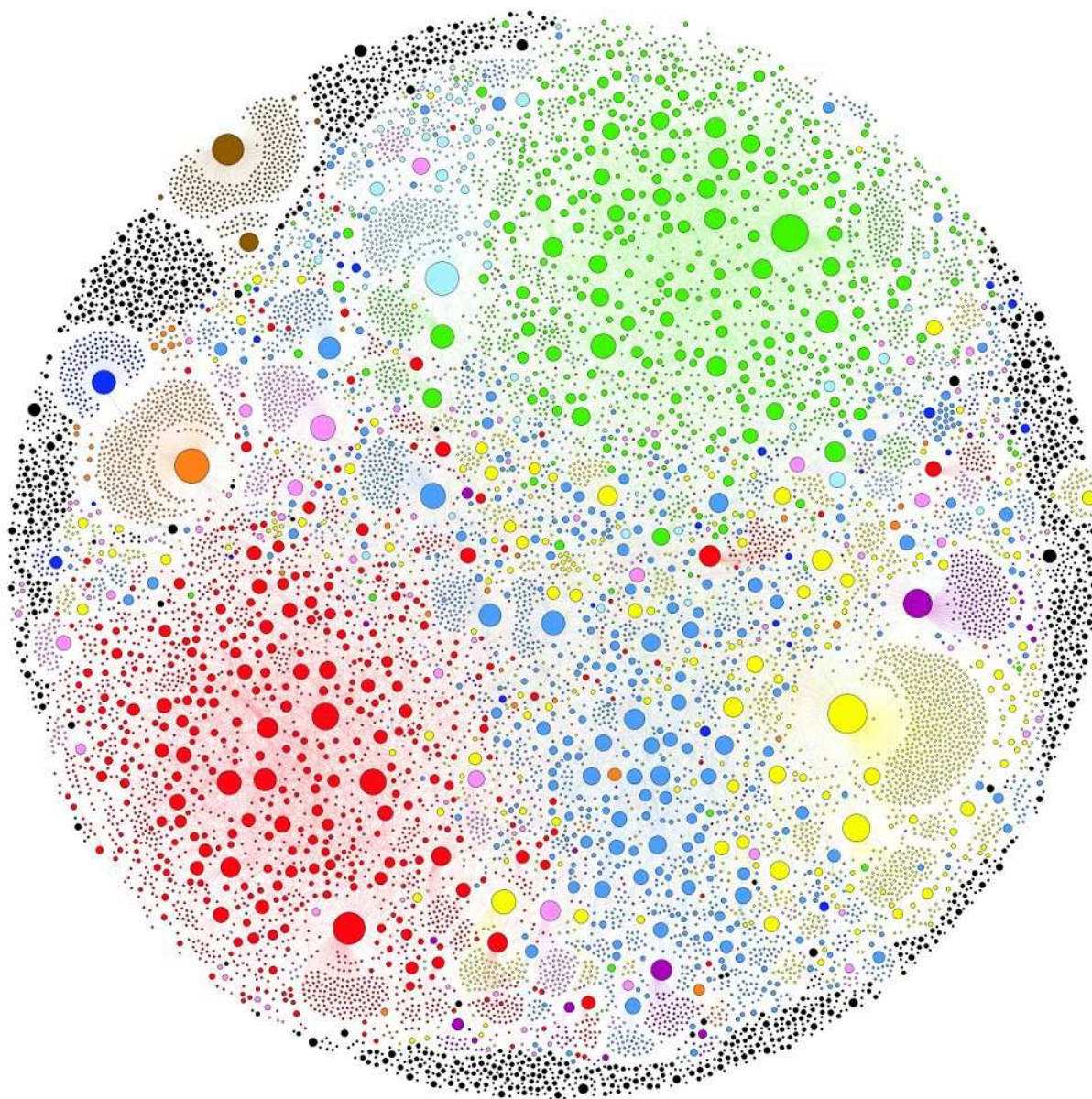


Fig. 5: Grafo de uma rede extraída do Twitter

Observam-se diferentes grupos de nós espalhados pelo grafo.

Para o layout, utilizou-se do algoritmo de distribuição *Force Atlas 2*, com a opção *Stronger Gravity* ligada e *Prevent Overlap* para ajuste final.

Utilizou-se também a estatística de modularidade para identificar comunidades.

O usuário pode posteriormente analisar o comportamento de grupos individualmente por opções de filtragem e tratamento em outra área de trabalho (vide seções 2.2.11 e 3.2).

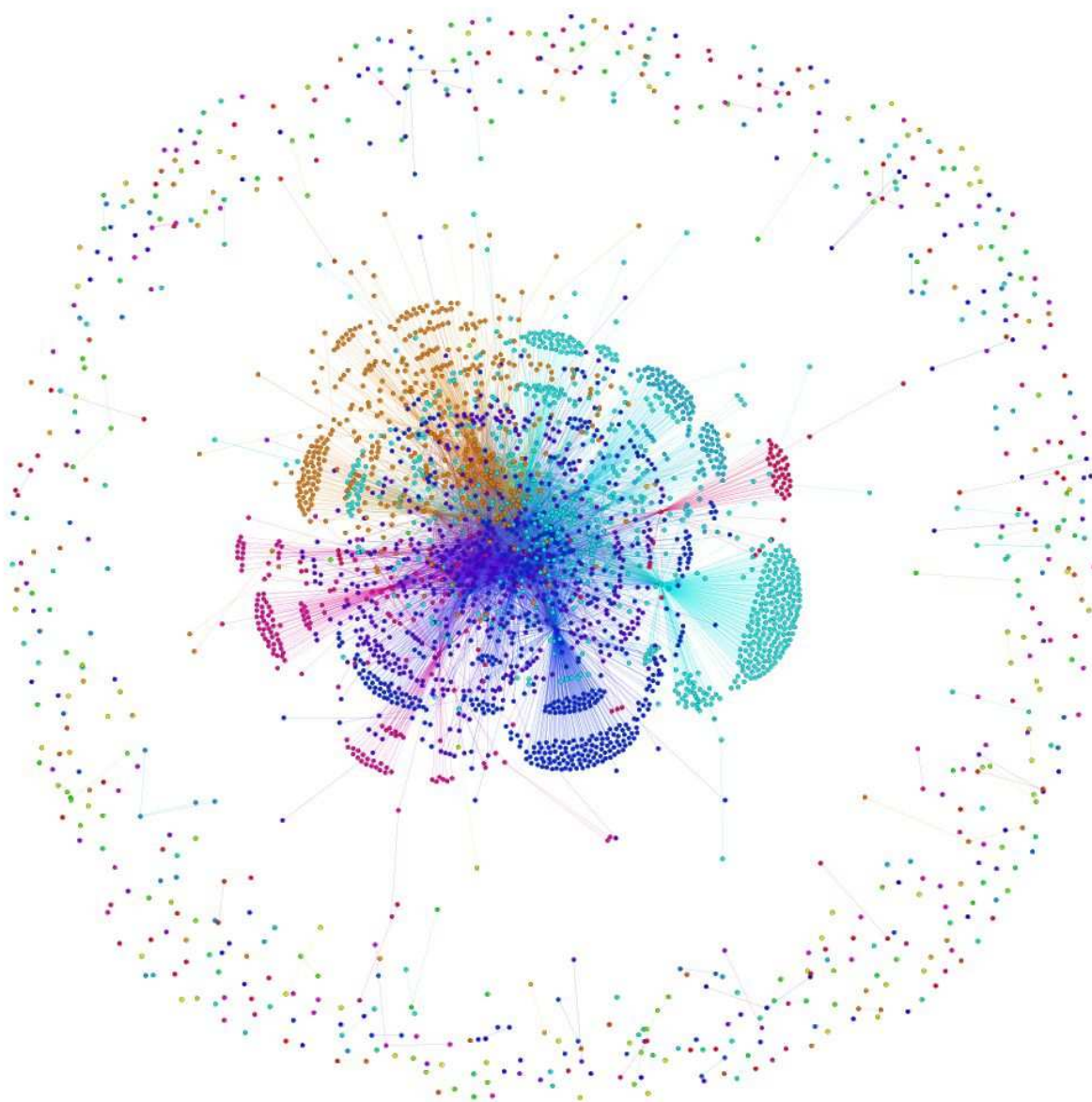


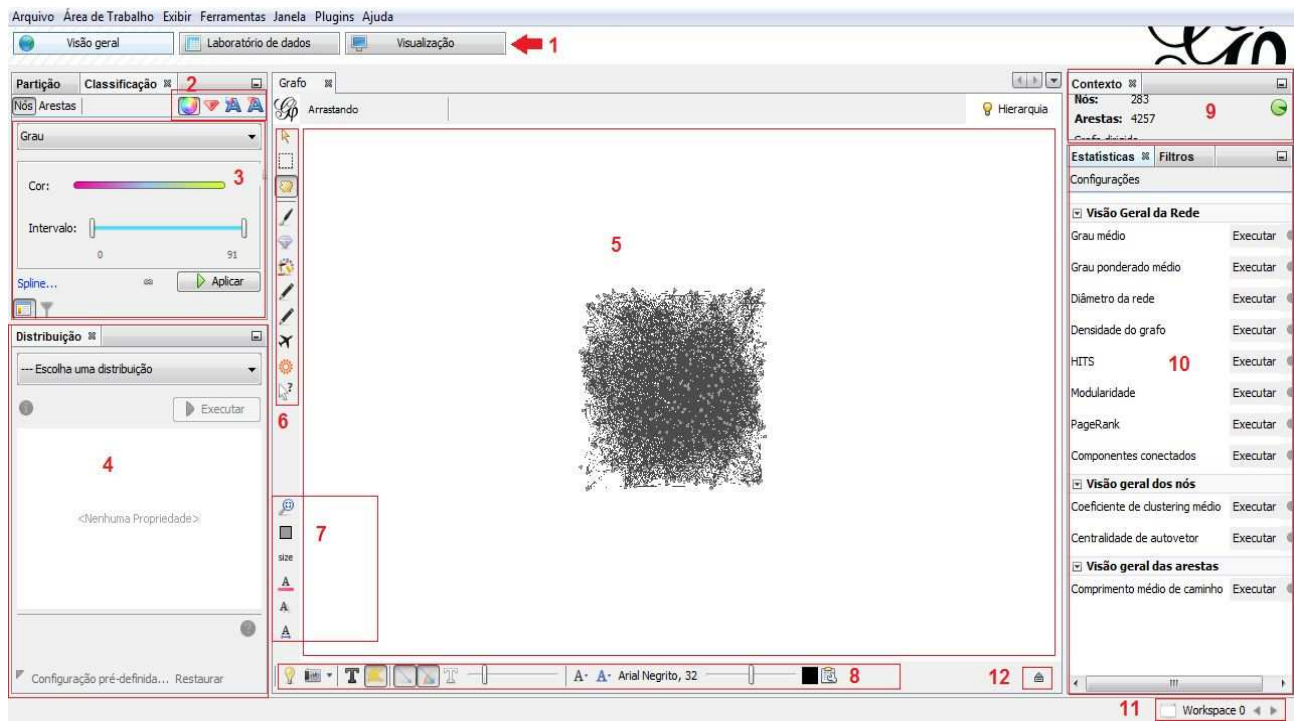
Fig. 6: Grafo de uma rede extraída da Wikipédia
Observam-se quatro diferentes *clusters* representados por diferentes cores; são subgrupos de usuários da comunidade da Wikipédia turca.
Para o layout, utilizou-se do algoritmo de distribuição *Yifan Hu*, cujo objetivo é separar os *clusters* que podem ser evidenciados com filtros de modularidade. Nota-se a rede periférica mais afastada do centro, ao contrário do grafo anterior.

2. Preparação

2.1 Prerequisitos do programa

O único prequesito do programa é a instalação do Java Runtime Environment (JRE), disponível em www.java.com. Atenção para a versão do sistema operacional, 32 ou 64 bits. Caso este seja Linux, recomenda-se instalar a versão proprietária tradicional, em vez de a escrita em código aberto (OpenJDK).

2.2 Interface gráfica



1. Barra de ambientes

- **Visão geral:** área de trabalho (local de desenvolvimento do grafo).
- **Laboratório de dados:** área de edição em tabela dos dados que compõem o grafo (ou seja, de seu *dataset*), onde é possível adicionar, editar e excluir arestas e nós individuais, assim como importar e exportar *datasets* em CSV (vide seção 1.2);
- **Visualização:** área de exportação do projeto quando concluído (vide seção 1.3).

2. Funções de visualização

- **Partição:** manipular nós e arestas de acordo com seus *grupos*.
- **Classificação:** manipular nós e arestas de acordo com seus *valores*.

3. Opções das funções de visualização

3.1 Partição

Aqui, as estatísticas são aplicadas a grupos de nós ou arestas, a depender da informação selecionada e de como ela foi inserida no laboratório de dados.

- Aplicação de parâmetros de partição;
- Porcentagens dos grupos, agrupamento e criação de gráfico em porcentagem ou em pizza das informações do grafo.

3.2 Classificação

Nesta seção, é possível aplicar estatísticas referentes a cor e tamanho de nós/arestas e seus respectivos rótulos. Abaixo da seleção de intervalo, temos:

- *Spline*: configurar a interpolação de classificação, ou seja, a proporção em que a estatística selecionada é aplicada;
- Aplicar ou auto-aplicar filtros;
- Listar dados conforme estatística aplicada;
- Escalonamento local.

4. Distribuição/Layout

Seção referente à espacialização do grafo, com diferentes algoritmos que variam em sua utilidade de acordo com a rede a ser trabalhada. Os algoritmos de espacialização mais comuns incluídos no pacote padrão do programa são *Fruchterman-Reingold* (vide pg. 4) e *Force Atlas 2*, sucessor natural do *Force Atlas* (vide pg. 5 e 6); outros podem ser conseguidos por meio de *plugins* (vide seção 2.3).

5. Área de visualização do grafo

- Mover grafo: enquanto pressionado o botão direito, arraste o mouse;
- Aumentar área de seleção: ctrl + scroll do mouse para cima ou para baixo;
- Zoom: scroll do mouse para cima ou para baixo;

6. Ferramentas de edição de grafos

- Tela cheia;
- Seleção de nós direta;
- Seleção de nós retangular;
- Mover nós;
- Colorir nós e vizinhos próximos;
- Alterar o tamanho de um nó;
- Colorir nós individuais;
- Adicionar novo nó conforme clique do mouse;
- Adicionar nova aresta conforme clique do mouse;
- Selecionar o caminho mais curto entre dois nós (grafos direcionados);

- Mapa de calor de um nó e seus vizinhos, pela distância (grafos direcionados);
- Editar atributos de nó.

7. Ferramentas de visualização

- Centralizar grafo inteiro;
- Centralizar grafo conforme coordenadas;
- Restaurar cor dos nós para preto;
- Restaurar tamanho dos nós para 1;
- Restaurar cores dos rótulos de texto do nós;
- Restaurar rótulos de nós visíveis;
- Restaurar tamanho da fonte dos rótulos.

8. Ferramentas de visualização de grafo

- Inverter cor de fundo;
- Captura de tela;
- Exibir rótulos de nós;
- Exibir bordas;
- Exibir arestas;
- Arestas com a cor do nó de origem;
- Exibir rótulos de arestas;
- Escala de peso das arestas;
- Modo do tamanho dos rótulos;
- Modo da cor;
- Fonte dos rótulos;
- Tamanho da fonte dos rótulos;
- Cor padrão da fonte dos rótulos;
- Atributos dos rótulos.

9. Contexto

Informações básicas referentes à rede, como quantidade de nós e de arestas.

10. Estatísticas e filtros

- **Estatísticas:** diferentes valores a serem calculados com diferentes algoritmos sobre atributos e propriedades da rede (vide seção 3.1).
- **Filtros:** seleciona no grafo os componentes correspondentes a uma classificação conforme selecionada pelo usuário.

11. Workspaces

É possível criar diferentes áreas de trabalho para manipular mais de um grafo simultaneamente num mesmo projeto do Gephi. Também permite-se copiar ou mover

nós existentes ou grafos inteiros para áreas distintas, selecionados manualmente ou filtrados com a ajuda do programa (vide seção 3.2).

Quando uma atualização para o programa ou algum *plugin* esteja disponível, um ícone ao lado direito desta área aparecerá, possibilitando o seu download.

12. Configurações complementares

Opções de visualização do grafo ainda no ambiente de trabalho. É possível configurar o modo que o grafo, os nós, as arestas e os rótulos de texto se comportam antes da exportação do grafo, de forma a melhorar a visualização e manipulação da rede.

2.3 Plugins

Responsáveis por adicionar funções não-existentes ao programa, a seção de *plugins* do Gephi está disposta discretamente na barra de ferramentas, em *Tools* → *Plugins*. Nela, é possível fazer o download de diferentes extensões desenvolvidas para o programa, que serão posteriormente, com a autorização do usuário, atualizadas conforme são desenvolvidas.

No Gephi Marketplace (www.marketplace.gephi.org/plugins) é possível ver a descrição de cada *plugin* e também fazer o seu download para instalação manual.

3. Teoria

3.1 Estatísticas comumente utilizadas

Grau médio

Define o peso dos nós de acordo com a quantidade de suas conexões.

- **Grau de entrada:** número de conexões que um nó recebe de outro;
- **Grau de saída:** números de conexões que sai de um nó para outro;

Grau ponderado médio

Similar ao grau médio, mas utiliza-se os pesos das arestas em seu algoritmo para então definir o peso dos nós.

Densidade do grafo

Mostra o quão próxima a rede está de ser completa. Um grafo completo é aquele em que, para cada nó, existe uma aresta conectando esse nó aos demais; ou seja, é um grafo em que todos os seus componentes encontram-se conectados

HITS:

- **Autoridade:** estima o valor do conteúdo do nó;
- **Hub:** o valor das arestas para as outras páginas.

Diâmetro da rede:

- **Centralidade de intermediação:** mede a frequência com que um nó aparece nos caminhos mais curtos entre nós das redes;
- **Centralidade de proximidade:** distância média de um determinado nó inicial para todos os demais nós da rede;
- **Excentricidade:** distância de um determinado nó inicial até o nó mais distante dele encontrado na rede.

Modularidade

Define os *clusters* da rede de acordo com a força de suas conexões.

Centralidade de autovetor

Medida de importância do nó com base nas conexões que possui com outros.

3.2 Filtros comumente utilizados

Atributos

Filtra nós e vértices de acordo com atributos pré-estabelecidos pelo usuário no *dataset*, no laboratório de dados e nas definições de partição e classificação do grafo.

Topologia

- **Intervalo de grau:** seleciona e mantém nós que possuem grau dentro de um intervalo definido pelo usuário;
- **Rede ego:** exibe a rede de um nó em específico, filtrando todos os outros que não possuem conexão com este determinado;
- **Componente gigante:** seleciona os nós mais conectados da rede e retira da visualização os periféricos, com poucas conexões;
- **Intervalo de grau de entrada:** seleciona e mantém nós que possuem grau de entrada dentro de um intervalo definido pelo usuário;
- **K-core:** mantém no grafo apenas nós no mínimo menos K (número) de grau;
- **Intervalo de grau bidirecional/mútuo:** seleciona e mantém nós que mutuamente possuem grau dentro de um intervalo definido pelo usuário;
- **Rede de vizinhos:** mantém nós que estejam a uma profundidade x de determinado nó, identificado pelo seu ID ou rótulo;
- **Intervalo de grau de saída:** seleciona e mantém nós que possuem grau de saída dentro de um intervalo definido pelo usuário.

4. Prática

Criando um dataset

Conforme visto anteriormente, *crawlers* são os responsáveis por minerar grandes conjuntos de dados (*big data*) na web e filtrar o que é ou não pertinente, a depender do objetivo do usuário. Após a coleta, importa-se o *dataset* para o Gephi, que então tomará a função de edição de dados e esquematização visual da rede.

Alguns *crawlers* de notável utilização no campo são:

Flocker

Webapp que age como estruturador de redes de *retweets* em tempo real. Permite exportar o grafo criado para GEXF, PNG e SVG.

Licença: gratuita.

Site: www.flocker.outliers.es

Netvizz

Aplicativo do Facebook de fácil utilização que possibilita extrair as redes de amigos, páginas e grupos a que o usuário principal está conectado.

Licença: gratuita.

Site: www.apps.facebook.com/netvizz

NodeXL

Extensão para o Microsoft Excel que permite extrair dados de redes como Facebook e Twitter e posterior manipulação no programa ou exportação para Gephi.

Licença: gratuita.

Site: www.nodexl.codeplex.com

Plugin Social Network Importer: www.socialnetimporter.codeplex.com

Topsy

Permite extrair dados da rede do Twitter. Pela parceria que tem com a rede social, é o único que não possui limite de tempo de publicação do *tweet* nem limite de requisições ao servidor. É o sucessor do YourTwapperKeeper.

Licença: versão limitada gratuita e profissional paga.

Site: www.topsy.com

YourTwapperKeeper

Permite a configuração de diferentes *keywords* para monitoração, captura e armazenamento de *tweets* em tempo real. Necessita de instalação.

Licença: código aberto.

Site: www.github.com/540co/yourTwapperKeeper