ACTIVIDAD 4. GRAFOS

Nombre: David Alfonso Velasco Sedano

Expediente: ms671002

Fecha: 05/10/17

# Finding the average degree and average weighted degree of a graph

The degree of a node in a graph is defined as the number of edges that are incident on that node. The loops—that is, the edges that have the same node as their starting and end point— are counted twice. In this recipe, we will learn how to find the average degree and average weighted degree for a graph.

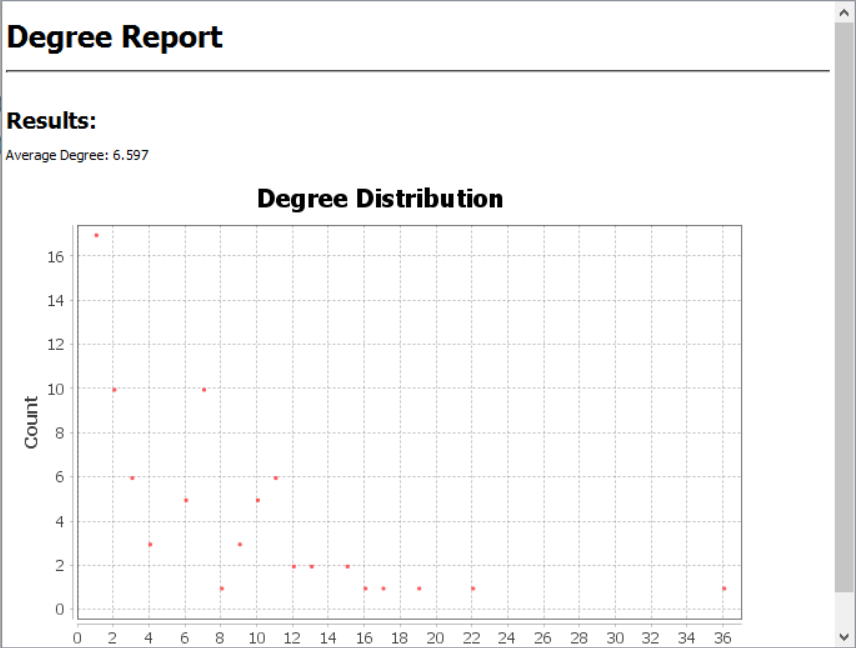
How to do it…

The following steps illustrate the process to find the average degree and weighted degree of a graph:

1. Load or create a graph in Gephi. For this recipe, we will consider the Les Misérables network that is already available in Gephi and can be loaded at the Welcome screen.

2. In the Statistics panel located on the right-hand side of the Gephi application window, under the Network Overview tab, click on the Run button located beside Average Degree.

3. This opens up a window containing the degree report for the Les Misérables network. In the case of directed graphs, the report contains the in-degree and out-degree distributions as well.

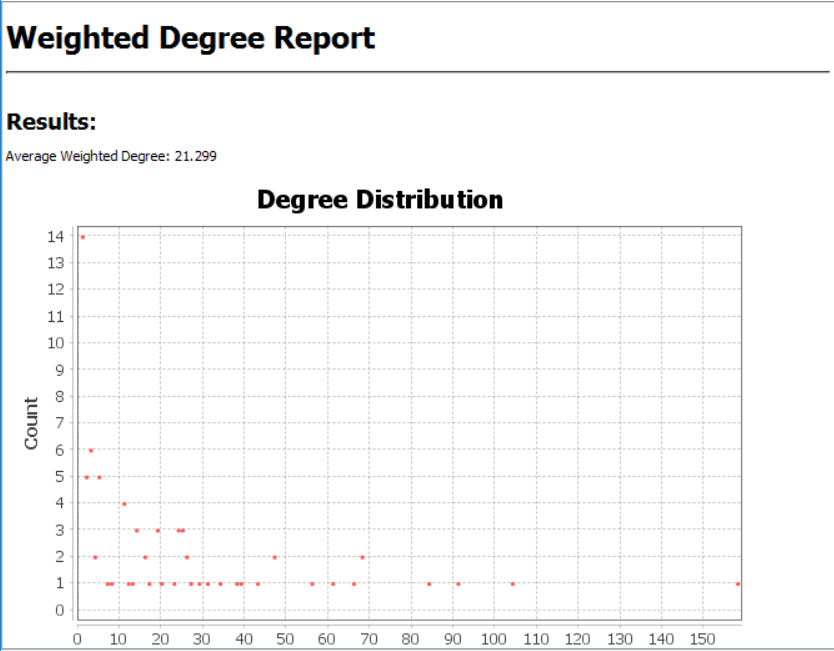


The average degree of the Les Misérables network is now displayed in the Statistics panel beside the Run button for Average Degree.



4. To find the average weighted degree of the Les Misérables graph, hit the Run button adjacent to Avg. Weighted Degree in the Network Overview tab of the Statistics panel in the Gephi window.

This will open up a window containing the weighted degree report of the Les Misérables network.





How it works…

The average degree for a graph is the measure of how many edges there are in the graph compared to its number of vertices. To find out the average degree for a graph, Gephi computes the sum of the degrees of individual nodes in the graph and divides that by the number of nodes present in it.

To find the average weighted degree for a graph with weighted edges, Gephi computes the average mean of the sum of the weights of the incident edges on all the nodes in the graph.

[Qué es el grado medio y el grado medio con pesos para un grafo? Cómo los calcula Gephi?]

|  |  |  |
| --- | --- | --- |
|  | ¿Qué es? | ¿Cómo los calcula? |
| Grado Medio | Es la relación entre la cantidad de aristas con respecto a los vértices presentes en el grafo | Obtiene todos los vértices del grafo y los divide por el total de aristas |
| Grado Medio con Pesos | Es el promedio del grado de todos los vértices | Calcula los grados individuales de los nodos, de ahí los suma todos y divide entre el número total de vértices presentes |

There's more…

If you have closed the report window and wish to see it once again, click on the small button with a question mark adjacent to the Run button. This will reopen the degree report.

# Finding the network diameter

The diameter of a network refers to the length of the longest of all the computed shortest paths between all pair of nodes in the network.

How to do it…

The following steps describe how to find the diameter of a network using the capabilities offered by Gephi:

1. Click on Window in the menu bar located at the top of the Gephi window. From the drop-down, select Welcome. Click on Les Miserables.gexf.

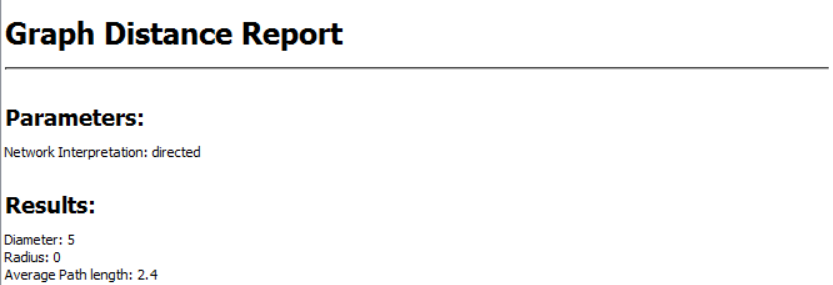
2. In the pop-up window, select Graph Type as Directed. This opens up the directed version of the Les Misérables network into Gephi.

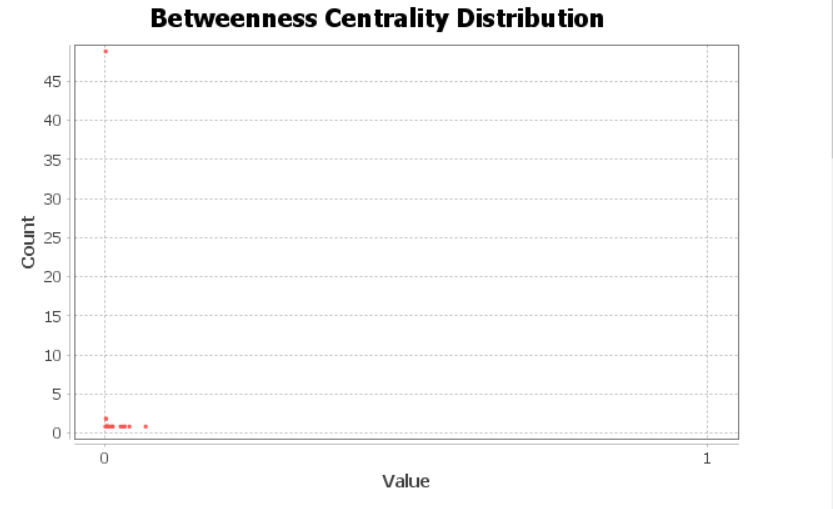
3. In the Statistics panel, under the Network Overview tab, click on the Run button, which is next to Network Diameter, to open the Graph Distance settings window:

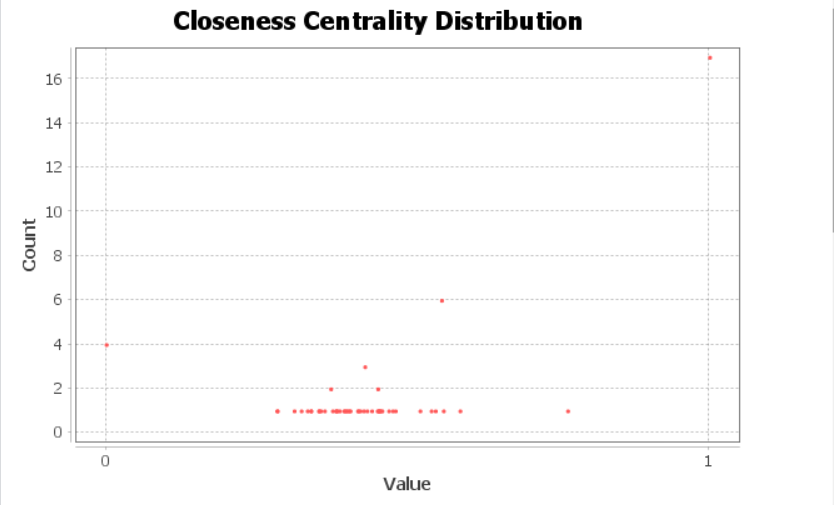
4. In the Graph Distance settings window, you can decide on which type of graph, Directed or UnDirected, the diameter algorithm has to be run. If you have loaded an undirected graph, the Directed radio button will remain deactivated. If a directed graph is chosen, you can choose between the directed and undirected versions of it to find the diameter.

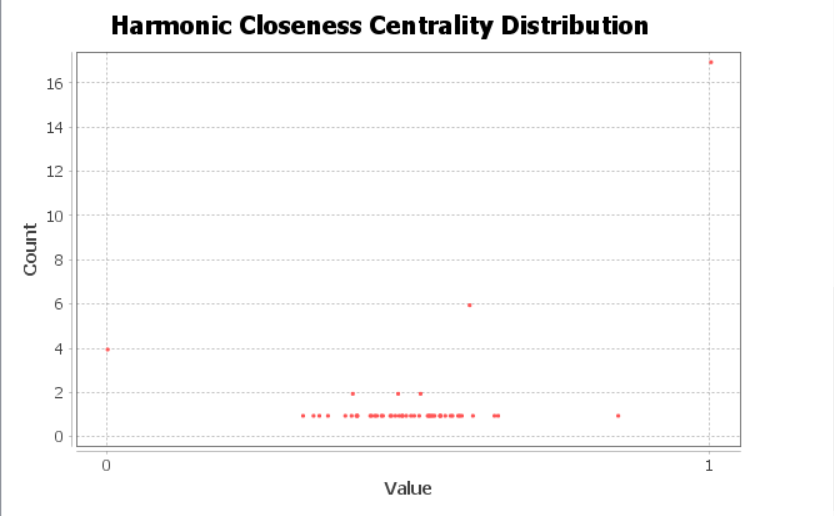
5. Check the box next to Normalize Centralities in [0, 1] to allow Gephi to normalize the three centralities' values between zero and one. The three centralities being referred to here are Betweenness Centrality, Closeness Centrality, and Eccentricity.

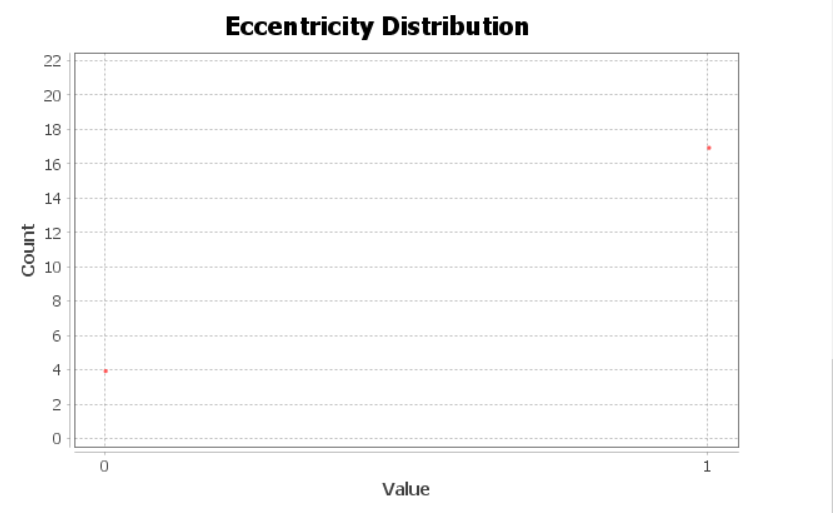
6. Click on OK. This opens up the Graph Distance Report window, that shows the value of the network diameter, network radius, average path length, number of shortest paths, and three separate graphs depicting betweenness centrality distribution, closeness centrality distribution, and eccentricity distribution.











How it works…

The diameter of a network gives us the maximum number of hops that must be made to travel from one node in the graph to the other. To find the diameter, all the shortest paths between every pair of nodes in the graph are computed and then the length of the longest of them gives us the diameter of the network. If the network is disconnected—that is, if the network has multiple components that do not share an edge between them—then the diameter of such a network is infinite.

Note that, in the case of weighted graphs, the longest path that determines the diameter of the graph is not the actual length of the path but the number of hops that would be required to traverse from the starting vertex to the end vertex. The computation of the diameter of a graphical network makes use of a property called the eccentricity of nodes. The eccentricity of a node is a measure of the number of hops required to reach the farthest node in the graph from this node. The diameter is then the maximum eccentricity among all the nodes in the graph.

¿Qué es y cómo (con y sin peso) se calcula el diámetro de la red?

El diámetro de la red es el camino más largo de entre los cortos entre dos puntos del grafo. Se calcula primero obteniendo todos los caminos cortos entre todos los vértices. De ahí, se busca cuál es el camino más largo y ese es considerado el diámetro del grafo. Esto simplemente aplica en casos de grafos dirigidos, de ser no-dirigidos, se considera que el diámetro es infinito.

¿Qué es la intermediación, cercanía y excentricidad de un grafo?

* Cercanía: Nos muestra la distancia (más corta) entre un nodo dado con respecto a todos los demás
* Intermediación: Muestra cuántas veces un nodo actúa como un punto intermedio para llegar al nodo deseado
* Excentricidad: Es la distancia máxima (a partir de los más cortos encontrados) entre un nodo y otro

¿Qué diferencia hay entre grafos sin pesos y con peso?

En el caso de encontrar el diámetro, no hay diferencia. Esto simplemente ocurre a la hora de calcular el camino más corto.

There's more…

There are three concepts—betweenness centrality, closeness centrality, and eccentricity— that have been introduced in this recipe. Eccentricity has already been covered in the *How it works…* section of this recipe. Betweenness centrality and closeness centrality are yet more important statistical properties of a network and are applied in a lot of real-world problems such as finding influential people in a social network, finding crucial hubs in a computer network, finding congestion nodes in wireless networks, and so on.

The betweenness centrality of a node is an indicator of its centrality or importance in the network. It is described as the number of shortest paths from all the vertices to all the other vertices in the network that pass through the node in consideration.

The closeness centrality of a node measures how accessible every other node in the graph is from the considered node. It is defined as the inverse of the sum of shortest distances of every other node in the network from the current node. Closeness centrality is an indicator of the speed at which information will transfuse into the network, starting from the current node.

Yet another concept that has been mentioned in this recipe is the radius of the graph. The radius of a graph is the opposite of its diameter. It is defined as the minimum eccentricity among the vertices of the graph. In other words, it refers to the minimum number of hops that are required to reach from one node of the graph to its farthest node.

# Finding graph density

One another important statistical metric for graphs is density. In this recipe, you will learn what graph density is and how to compute it in Gephi.

How to do it…

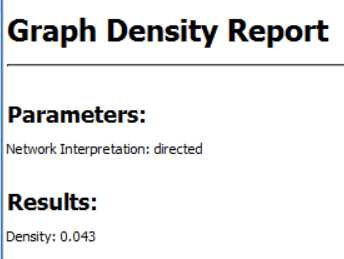
The following steps illustrate how to use Gephi to figure out the graph density for a chosen graph:

1. Load the directed version of the Les Misérables network in Gephi, as described in the *How to do it…* section of the previous recipe.

2. In the Statistics panel located on the right-hand side of the Gephi application window, click on the Run button that is placed against Graph Density.

3. This opens up the Density settings window, as shown in the following screenshot, where you can choose between the directed or the undirected version of the graph to be considered for the computation of graph density:

4. Click on OK. This opens up the following Graph Density Report window.



How it works…

A complete graph is a graph in which every pair of nodes is connected via a direct edge. The density of a graph is a measure of how close the graph is to a complete graph with the same number of nodes. It is defined as the ratio of the total number of edges present in a graph to the total number of edges possible in the graph. The total number of edges possible in a simple undirected graph is mathematically computed as *(N(N-1))/2*, where *N* is the number of nodes in the graph. A simple graph is a graph that has no loops and not more than one edge between the same pair of nodes.

¿Qué es la densidad de un grafo?

Es una medición para decir que tan cercano es un grafo dado a ser uno completo (todos los vértices están conectados)

¿Cómo calcula Gephi la densidad?

Debajo de ‘Visión General de la Red’, seleccionamos ‘Densidad de grafo’



El cálculo per sé se consigue obteniendo todas las aristas del grafo y se divide entre las aristas que contiene un grafo completo con la misma cantidad de vértices.

¿Cuál de las fórmulas vistas en la sección de permutación y combinatoria del curso es la que se utiliza para calcular el número total de aristas posibles en un grafo no dirigido?

Combinatoria (N ( N – 1 ) ) / 2. N siendo la cantidad de vértices en nuestro grafo.

¿Qué es un grafo simple?

Un grafo que no tiene auto-aristas, aristas en paralelo, ni circuitos.

There's more…

The density of the undirected version of a graph with *n* nodes will be twice of that of the directed version of the graph. This is because, in a directed graph, there are two edges possible between every pair of nodes, each with a different direction.

# Finding a graph's modularity

The modularity of a graph is a measure of its strength and describes how easily the graph could be disintegrated into communities, modules, or clusters. In this recipe, the concept of modularity, along with its implementation in Gephi, is described.

How to do it…

To obtain the modularity score for a graph, follow these steps:

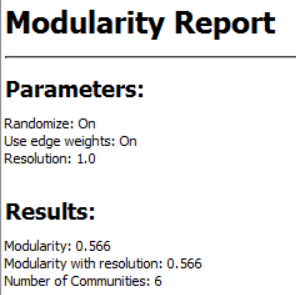
1. Load the Les Misérables graph in Gephi.

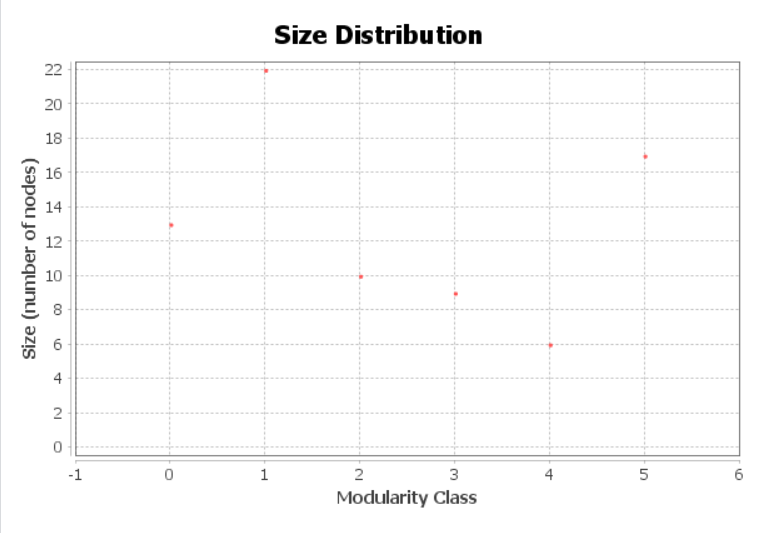
2. In the Network Overview tab under the Statistics panel, hit the Run button adjacent to Modularity.

3. In the Modularity settings window, enter a resolution in the textbox depending on whether you want a small or large number of communities. You can choose to randomize to get a better decomposition of the graph into communities, but this increases the computation time. You can also choose to include edge weight in computing modularity.

4. Hit OK once done.

5. This opens up the Modularity Report window, which the size distribution of communities into various modularity classes. The report also shows the number of communities formed, along with the overall modularity score of the graph.





How it works…

Modularity is defined as the fraction of edges that fall within the given modules to the total number of edges that could have existed among these modules. Mathematically, modularity is computed , where is the probability that an edge is in module *i* and is the probability that a random edge would fall into the module *i*.

Modularity is a measure of the structure of graphical networks. It determines the strength of the network as a whole. It describes how easily a network could be clustered into communities or modules.

A network with high modularity points to strong relationships within the same communities but weaker relationship across different communities. It is one of the fundamental methods used during community detection in graphs. Modularity finds its applications in a wide range of areas such as social networks, biological networks, and collaboration networks.

¿Qué es modularidad?

Describe qué tan fácil se puede dividir un grafo en comunidades. Se dice a esto que tan “fuerte” es la conexión en sí misma.

¿Cómo se define?

¿Qué significa una alta modularidad en un grafo?

Significa que los nodos dentro de una comunidad tienen una conexión fuerte entre ellos, pero muy débil con elementos de otras comunidades.

¿En dónde se aplica?

Redes sociales, redes biológicas y redes de colaboración.

# Applying individual filters on a graph

Very often, anyone studying a graph is interested in certain parts of it and doesn't need to look at the entire graph, and that is where we need a way of filtering out required portions of the graph. Gephi has a wide set of filters, based on which specific parts of a graph can be filtered out and then studied in detail. For example, one might filter out nodes based on modularity and then run statistics such as degree, PageRank, and so on, one after the other.

How to do it…

To filter out selected portions of a graph based on certain pre-specified criteria, follow these steps:

1. Load the Les Misérables network in Gephi.

2. In the Filters panel located on the right-side of the Gephi application window next to the Statistics panel, you will see a library of filters categorized into buckets based on which part of the graph the filter is going to be applied to.

3. Expand one of the buckets, say Attributes, and under that select a subbucket, say Partition, and then select a criterion for filtering. Here, it is Modularity Class.

4. Double-click on Modularity Class or drag-and-drop it in the Queries panel below the Filters panel to select the filter:

3. Expand one of the buckets, say Attributes, and under that select a subbucket, say Partition, and then select a criterion for filtering. Here, it is Modularity Class.

4. Double-click on Modularity Class or drag-and-drop it in the Queries panel below the Filters panel to select the filter.

6. Hit Filter to apply this filter on the graph.

[Insert screenshot from the Les Misérables network when the Modularity Class filter has been applied on partitions corresponding to modularity classes 8, and then 8 and 2]

How it works…

On applying a specific filter, Gephi extracts only the node and edge information corresponding to selections made in the filter settings. With this new information, it constructs a new graph and displays it on the screen. This feature comes in handy when one is dealing with large networks and wants to clearly visualize only parts of interest in the graph under consideration.

Apart from Modularity Class, there are a lot of other interesting filters available in Gephi, such as edge weight, filtering based on node labels, filtering based on edge labels, filtering based on ID, filtering based on degree range, and so on.

# Applying a combination of filters on a graph

In the previous recipe, we learned how to apply individual filters on graphs. In this recipe, we will learn how to apply a combination of filters.

How to do it…

The following steps illustrate how to apply a combination of filters to a graph in Gephi.

1. Load the Les Misérables network in Gephi.

2. In the Filters panel located on the right-side of the Gephi application window, next to the Statistics panel, expand the Operator option and drag-and-drop the INTERSECTION filter into the Queries panel below.

3. Now, in the Filters panel, expand the Topology option and drag-and-drop the Degree Range filter under the INTERSECTION filter in the Queries panel.

4. Again, in the Filters panel, expand the Edges option and drag-and-drop the Edge Weight filter under the INTERSECTION filter in the Queries panel.

5. Click on Degree Range in the Queries panel to open the Settings panel for Degree Range. Move the sliders to select the range of degrees to be filtered.

6. Click on Edge Weight in the Queries panel to open the Settings panel for Edge Weight. Move the sliders to select the range of edge weights to be filtered.

7. Hit Filter to apply this combination of degree range and edge weight filters to the graph. The following screenshot shows the Les Misérables network when only its edges having a weight between 3.61 and 18.1, and nodes having degrees between 6 and 36, have been retained.

[INSERT the graph with the filters applied]

How it works…

When a logical operator is selected, followed by one or more filters, Gephi selects edges and nodes after applying all the filters on all of them and retains only the ones matching the criteria.

There's more…

The results of applying a combination of filters on nodes or edges, based on what sort of filter has been selected, can be exported as columns to be used later on. To do so, click on the first button next to the Reset button, which is located at the top of the Filters panel.

When the mouse pointer is hovered on it, the pop-up text should read, Export filtered graph as a true/false data column. In the pop-up box, enter the title of the column that will contain the result of this filtering and hit OK.

***Extracted from:*** KHOKHAR, Devangana. Gephi cookbook: over 90 hands-on recipes to master the art of network analysis and visualization with Gephi. 2015.