An Introduction of Gephi

CITI Visualization Group

February 21, 2017

Contents

1 Introduction

Gephi(https://gephi.org/) is an open-source visualization platform for graphs and networks. This documents focuses on giving readers a brief overview of this software by introducing its configurations, concepts, and functions. To access more detailed tutorials and demonstrations we refer readers to Gephi's official website: https://gephi.org/users/.

2 File Formats and Basic Data Structures

While Gephi can import data from various file formats as listed below, GEXF, CVS, and Spreadsheet are the widely adopted because of their simple structures.

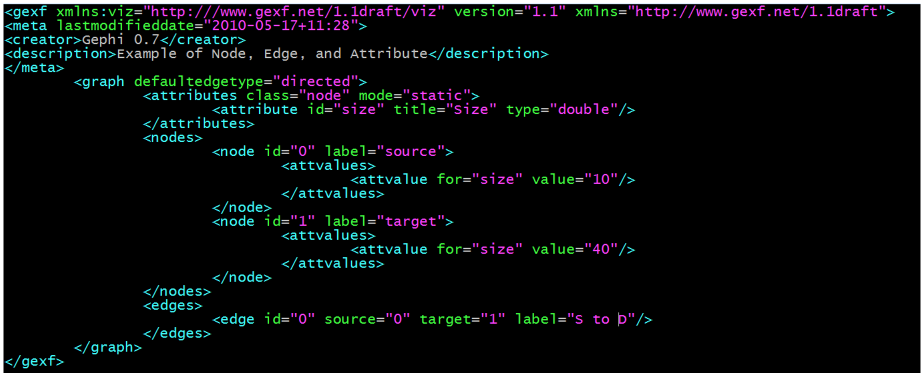
- GEXF
- GDF
- GML
- GraphML
- Pajek NET
- GraphVis Dot
- CVS
- UCINET DL
- TLP
- Netdraw VNA
- Spreadsheet(Excel)

GEXF is a XML based structure which is able to encode almost all of the data structures supported by Gephi, while CVS and Spreadsheet can only include a few of them. However, CVS and Spreadsheet can be read and edited by a wider range of applications.

In addition to these formats, Gephi can import and export its own state files having an extension .gephi, that not only include original data but also encode operations and filters have been applied to the data. In other words, a .gephi file stores an actual visualization.

The essential data structure supported by Gephi are node, edge, and attribute, where nodes and edges are referred to as the network topology while attributes are referred to as network data. A node represents a basic entity in the data set and two nodes can be connected by an edge. Attributes are data associated to nodes or edges, e.g. string, integer, boolean, etc..

Figure 1 illustrates how the aforementioned data structures are encoded in a GEXF file. In this example, we declared a directed graph with a "source" node



```
<gexf xmlns:viz="http:///www.gexf.net/1.1draft/viz" version="1.1" xmlns="http://www.gexf.net/1.1draft">
<meta lastmodifieddate="2010-05-17+11:28">
<creator>Gephi 0.7</creator>
<description>Example of Node, Edge, and Attribute</description>
</meta>
        <graph defaultedgetype="directed">
                <attributes class="node" mode="static">
                        <attribute id="size" title="Size" type="double"/>
                </attributes>
                <nodes>
                        <node id="0" label="source">
                                <attvalues>
                                        <attvalue for="size" value="10"/>
                                </attvalues>
                        </node>
                        <node id="1" label="target">
                                <attvalues>
                                        <attvalue for="size" value="40"/>
                                </attvalues>
                        </node>
                </nodes>
                <edges>
                        <edge id="0" source="0" target="1" label="S to D"/>
                </edges>
        </graph>
</gexf>
```

Figure 1: A GEXF file encoded two nodes attached with one attribute, and a directed edge connects them.

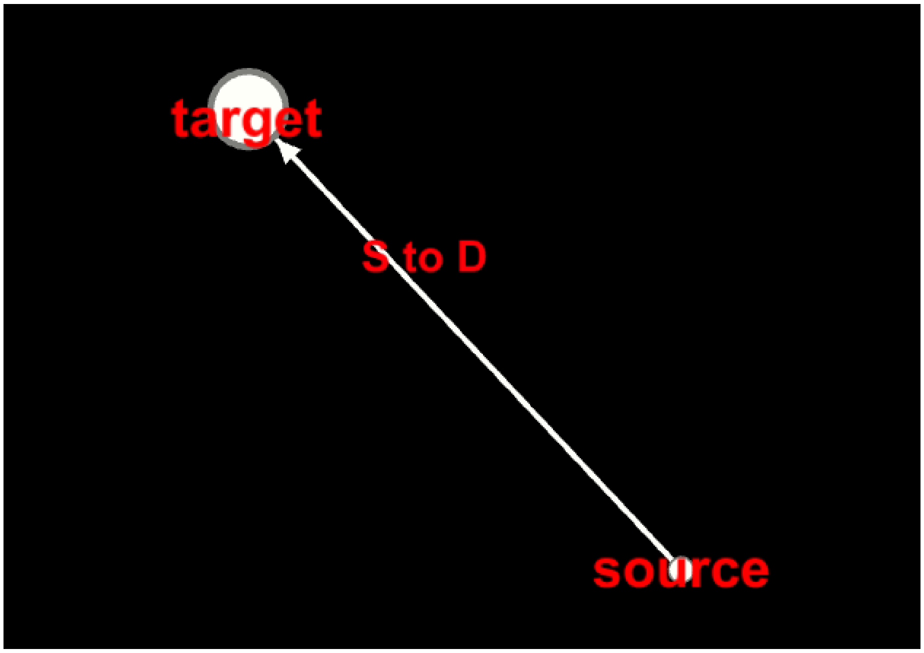


Figure 2: The corresponding visualization of the example illustrated in Figure 1, where sizes of nodes were determined by their “size” attribute.

and a “target” node that were distinguished by their unique node ids, i.e. 0 and 1. They were also superimposed with a “size” attribute which was represented by a scalar value. An edge between them was declared to indicate a directed path from source to target. The corresponding visualization of this example is shown in Figure 2, where the sizes of nodes were assigned based on the “size” attribute.

3 Creating Visualizations

This section will introduce the basic usage and functions of Gephi by demonstrating the process of exploring an Airline sample data set.

To load the data, we can simply click the **File** drop-down menu, press **Open**, and select *airlines.gexf*. Once the data has been imported successfully, Gephi

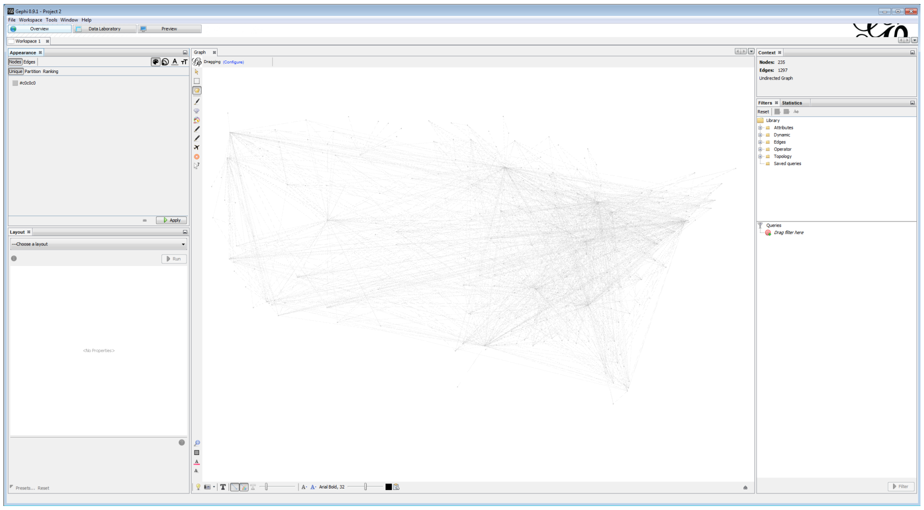
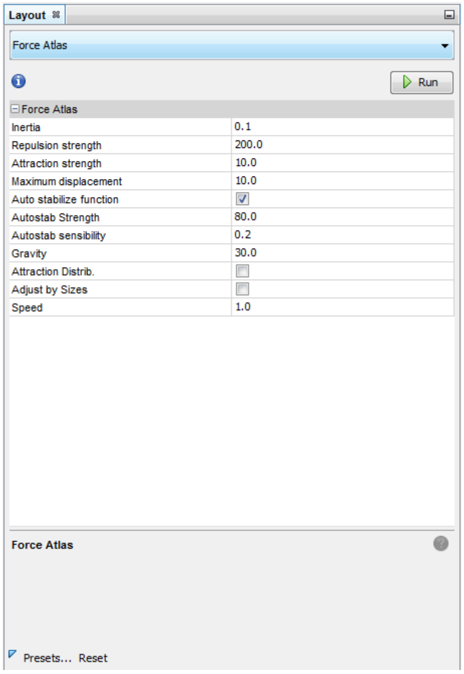

Nodes: 235
Edges: 1297
Undirected Graph

Figure 3: A default view the airline data set in Gephi.

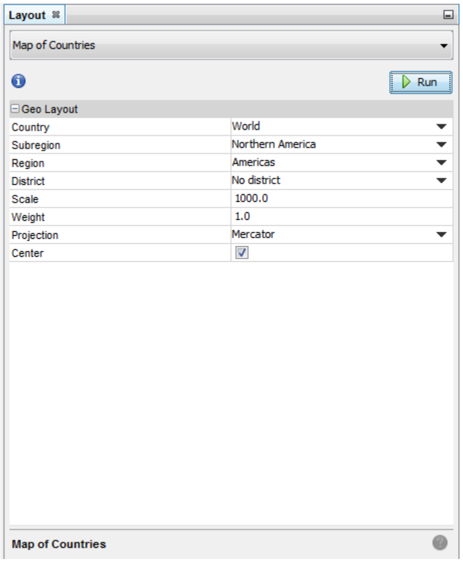
will create a default display which is similar to Figure 3. As seen in this figure, Gephi contains four primary panels: **Appearance**, **Layout**, **Graph**, and **Filters and Statistics**. Appearance panel provides the user interface to manipulate rendering options of nodes and edges, i.e. color, size, label color, and label size. Layout panel lists a set of built-in algorithms to adjust the graph layout. As shown in Figure 4, we can configure the properties and parameters of a particular algorithm by selecting it in the drop down menu, and apply the algorithm by clicking the **Run** button. Graph panel shows the actual visualization, and also provides a suite of user interfaces, e.g. selection of entities, creation of nodes and edges, configuration of displays, etc.. Filters tab provides a list of filters and operations that can be applied for supporting specific queries of the dataset, while the Statistics tab implements statistical calculations that can be used for analysis of the dataset.



Force Atlas	
Inertia	0.1
Repulsion strength	200.0
Attraction strength	10.0
Maximum displacement	10.0
Auto stabilize function	☑
Autostab Strength	80.0
Autostab sensibility	0.2
Gravity	30.0
Attraction Distrib.	☐
Adjust by Sizes	☐
Speed	1.0

Figure 4: Selecting Force Atlas algorithm in Layout panel.

Even though the default visualization shown in Figure 3 may lead viewers to some superficial interpretations of the data set, for instance there are many airports and they established flights between one another, it can hardly support more detailed explorations accurately, such as where are the airports, which airport maintains the most airlines, etc.. The first step to enhance the visualization is to superimpose a geographical map by applying a **Map of Countries** layout algorithm. Figure 5 shows our specified configurations of the layout, meaning that the display region of the map is restricted to U.S. After running the algorithm, we obtained an incorrect result, as shown in Figure 6, because of the inconsistence between geological information encoded by the dataset and that of the map. This can be fixed by applying a **Geo Layout** algorithm, with Latitude and Longitude parameters setup correctly, as shown in Figure 7.



Country	World
Subregion	Northern America
Region	Americas
District	No district
Scale	1000.0
Weight	1.0
Projection	Mercator
Center	☑

Figure 5: Configurations of Map of Countries layout algorithm.

In Figure 7, each circle represents an airport on the map, and a straight forward idea to improve this visualization is to encode the number of airlines maintained by individual airports by sizes of the circles, i.e. an airport has more airlines will be rendered as a bigger circle, and vice versa. We can simply make the modification in Appearance panel by changing the size of nodes. However, since the map shown in the visualization is also constructed by nodes and edges, a direct manipulation of all nodes will also change appearance of the map, resulting in undesired display. This can be avoided by creating filters to extract nodes only associated with airlines.s To achieve the proper filters, we first examine properties of all nodes and edges contained in the visualization by clicking the **Data Laboratory** button. Figure 8 exemplifies the data table, and we notice that a *backgroundmap* attribute is associated to each node

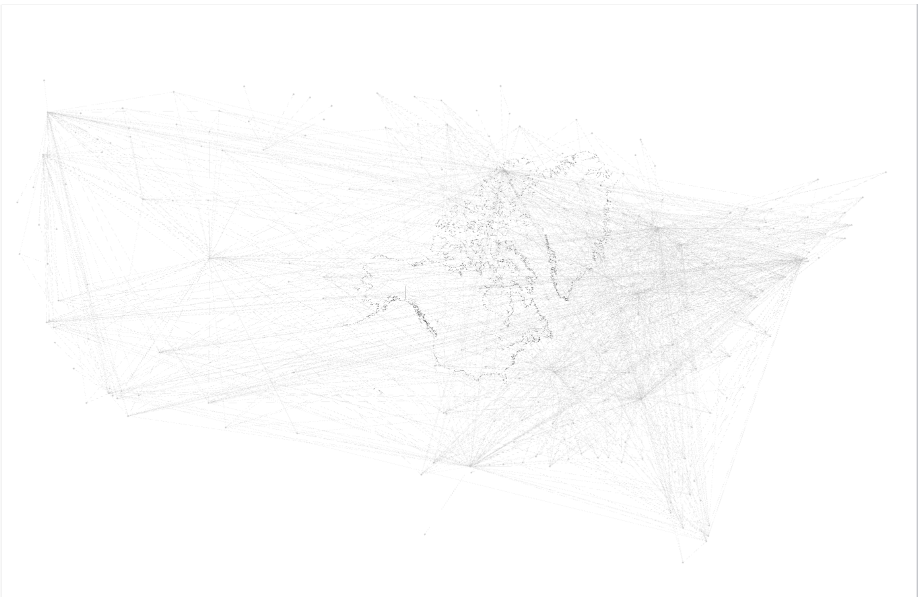


Figure 6: An incorrect map layout.

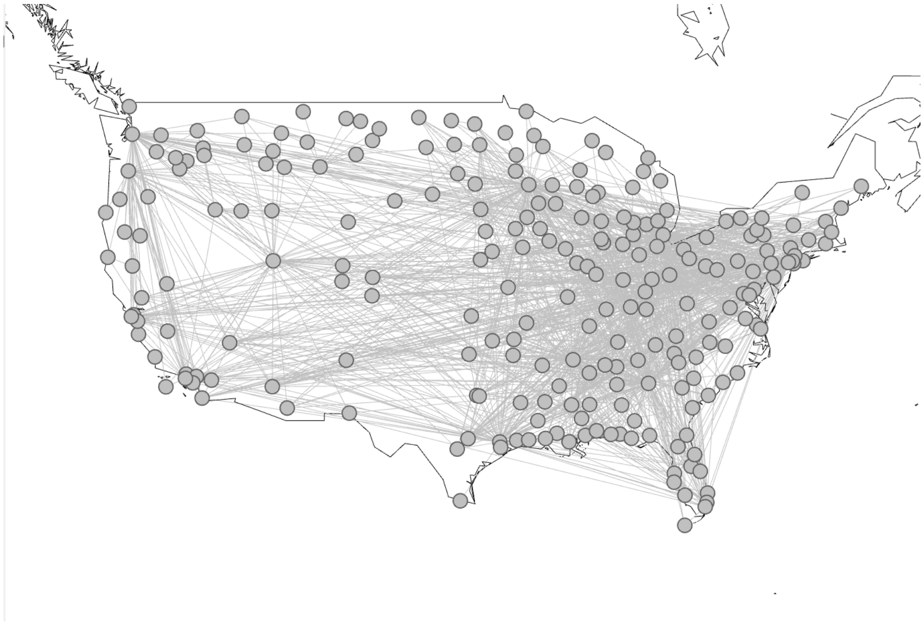
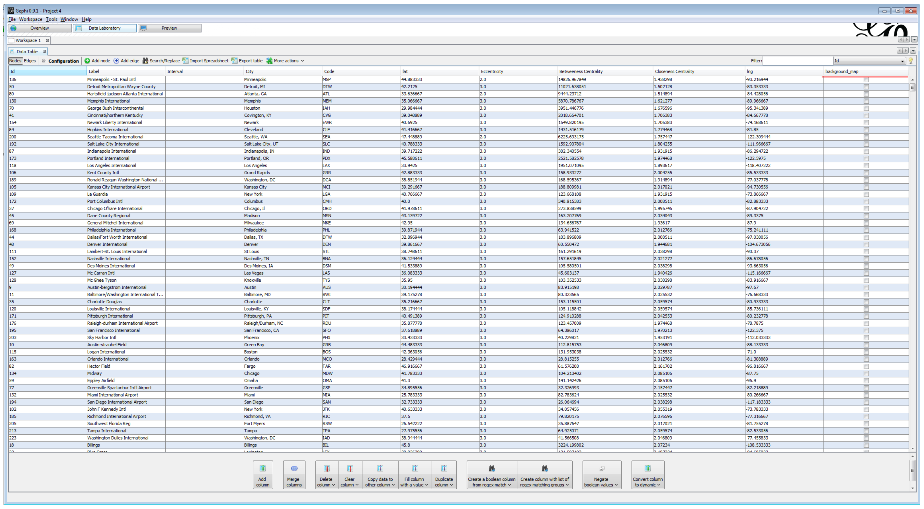


Figure 7: A correct map layout adjusted by Geo Layout algorithm.



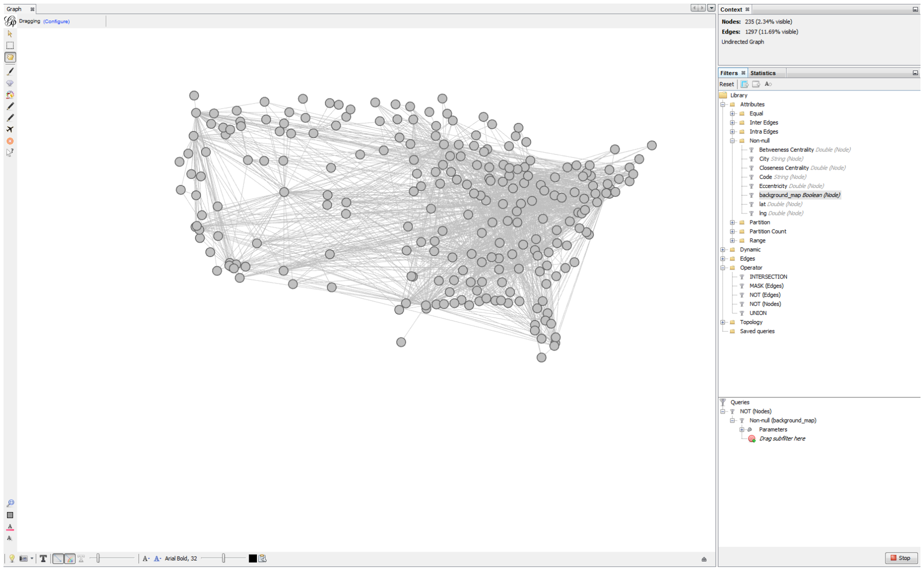
Id	Label	Interval	City	Code	lat	Eccentricity	Betweeness Centrality	Closeness Centrality	lng	background_map
136	Minneapolis - St. Paul Intl		Minneapolis	MSP	44.883333	2.0	14826.967849	1.438298	-93.216944	☐
50	Detroit Metropolitan Wayne County		Detroit, MI	DTW	42.2125	3.0	11021.638051	1.502128	-83.353333	☐
80	Hartsfield-jackson Atlanta International		Atlanta, GA	ATL	33.636667	2.0	9444.23712	1.514894	-84.428056	☐
130	Memphis International		Memphis	MEM	35.066667	3.0	5870.786767	1.621277	-89.966667	☐
70	George Bush Intercontinental		Houston	IAH	29.984444	3.0	3951.446776	1.676596	-95.341389	☐
41	Cincinnati/northern Kentucky		Covington, KY	CVG	39.048889	3.0	2018.664701	1.706383	-84.667778	☐
154	Newark Liberty International		Newark	EWR	40.6925	3.0	1549.820195	1.706383	-74.168611	☐
84	Hopkins International		Cleveland	CLE	41.416667	3.0	1431.516179	1.774468	-81.85	☐
200	Seattle-Tacoma International		Seattle, WA	SEA	47.448889	2.0	6225.693175	1.757447	-122.309444	☐
192	Salt Lake City International		Salt Lake City, UT	SLC	40.788333	3.0	1592.907804	1.804255	-111.966667	☐
87	Indianapolis International		Indianapolis, IN	IND	39.717222	3.0	382.340554	1.931915	-86.294722	☐
173	Portland International		Portland, OR	PDX	45.588611	3.0	2521.582578	1.974468	-122.5975	☐
118	Los Angeles International		Los Angeles	LAX	33.9425	3.0	1951.071095	1.893617	-118.407222	☐
106	Kent County Intl		Grand Rapids	GRR	42.883333	3.0	158.933272	2.004255	-85.533333	☐
189	Ronald Reagan Washington National ...		Washington, DC	DCA	38.851944	3.0	168.595367	1.914894	-77.037778	☐
105	Kansas City International Airport		Kansas City	MCI	39.291667	3.0	188.809981	2.017021	-94.730556	☐
109	La Guardia		New York	LGA	40.766667	3.0	123.668108	1.931915	-73.866667	☐
172	Port Columbus Intl		Columbus	CMH	40.0	3.0	340.815383	2.008511	-82.883333	☐
37	Chicago O'hare International		Chicago, Il	ORD	41.978611	3.0	273.838599	1.995745	-87.904722	☐
45	Dane County Regional		Madison	MSN	43.139722	3.0	163.207769	2.034043	-89.3375	☐
69	General Mitchell International		Milwaukee	MKE	42.95	3.0	134.656767	1.93617	-87.9	☐
168	Philadelphia International		Philadelphia	PHL	39.871944	3.0	63.941522	2.012766	-75.241111	☐
44	Dallas/Fort Worth International		Dallas, TX	DFW	32.896944	3.0	183.896809	2.008511	-97.038056	☐
48	Denver International		Denver	DEN	39.861667	3.0	60.550472	1.944681	-104.673056	☐
111	Lambert-St. Louis International		St Louis	STL	38.748611	3.0	161.291619	2.038298	-90.37	☐
152	Nashville International		Nashville, TN	BNA	36.124444	3.0	157.651845	2.021277	-86.678056	☐
49	Des Moines International		Des Moines, IA	DSM	41.533889	3.0	105.580501	2.038298	-93.663056	☐
127	Mc Carran Intl		Las Vegas	LAS	36.083333	3.0	45.603137	1.940426	-115.166667	☐
128	Mc Ghee Tyson		Knoxville	TYS	35.95	3.0	103.352533	2.038298	-83.916667	☐
9	Austin-bergstrom International		Austin	AUS	30.194444	3.0	83.915198	2.029787	-97.67	☐
11	Baltimore/Washington International T...		Baltimore, MD	BWI	39.175278	3.0	80.323565	2.025532	-76.668333	☐
35	Charlotte Douglas		Charlotte	CLT	35.216667	3.0	153.115501	2.059574	-80.933333	☐
120	Louisville International		Louisville, KY	SDF	38.174444	3.0	105.118842	2.059574	-85.736111	☐
171	Pittsburgh International		Pittsburgh, PA	PIT	40.491389	3.0	124.910288	2.042553	-80.232778	☐
176	Raleigh-durham International Airport		Raleigh/Durham, NC	RDU	35.877778	3.0	123.457009	1.974468	-78.7875	☐
195	San Francisco International		San Francisco, CA	SFO	37.618889	3.0	64.386017	1.970213	-122.375	☐
203	Sky Harbor Intl		Phoenix	PHX	33.433333	3.0	40.229821	1.953191	-112.033333	☐
10	Austin-straubel Field		Green Bay	GRB	44.483333	3.0	112.815753	2.046809	-88.133333	☐
115	Logan International		Boston	BOS	42.363056	3.0	131.953038	2.025532	-71.0	☐
163	Orlando International		Orlando	MCO	28.429444	3.0	28.815255	2.012766	-81.308889	☐
82	Hector Field		Fargo	FAR	46.916667	3.0	61.576208	2.161702	-96.816667	☐
134	Midway		Chicago	MDW	41.783333	3.0	104.213402	2.085106	-87.75	☐
59	Eppley Airfield		Omaha	OMA	41.3	3.0	141.142426	2.085106	-95.9	☐
77	Greenville Spartanbur Int'l Airport		Greenville	GSP	34.895556	3.0	32.326993	2.157447	-82.218889	☐
132	Miami International Airport		Miami	MIA	25.783333	3.0	82.783624	2.025532	-80.266667	☐
194	San Diego International Airport		San Diego	SAN	32.733333	3.0	26.064694	2.038298	-117.183333	☐
102	John F Kennedy Intl		New York	JFK	40.633333	3.0	34.057456	2.055319	-73.783333	☐
185	Richmond International Airport		Richmond, VA	RIC	37.5	3.0	79.820175	2.076596	-77.316667	☐
205	Southwest Florida Reg		Fort Myers	RSW	26.542222	3.0	35.887647	2.017021	-81.755278	☐
213	Tampa International		Tampa	TPA	27.975556	3.0	64.925071	2.059574	-82.533056	☐
223	Washington Dulles International		Washington, DC	IAD	38.944444	3.0	41.566508	2.046809	-77.455833	☐
18	Billings		Billings	BIL	45.8	3.0	3224.199802	2.07234	-108.533333	☐

Figure 8: Data laboratory shows attributes of all nodes and edges of the visualization.

to indicate whether it belongs to the background map or not. Now we can go back to **Filter** tab and extract all background nodes by clicking *Attributes*, *Non-null*, *backgroundmap*, and the filter will be listed in the **Queries** text box. Then we attached this filter to a *NOT(nodes)* operation, meaning that we select all nodes except those we obtained from the filter. Figure 9 shows an updated visualization where the *Queries* box at the lower right corner listed the filters and operations been conducted.

With the filters running we are able to manipulate appearance of airport nodes by clicking **Nodes** button in **Appearance** panel. We select size option and click **Ranking** button to vary node size based on one of its attached scalar attributes. In this example, we associate the *degree* attribute, which indicates the number of airlines connected to individual airports, to node size, meaning that an airport shown by a larger circle is considered to be busier. Once we apply modifications of appearance of nodes, we stop the filter created before to restore the background map to the display, as shown in Figure 10.

Similarly, we can also adjust color codings to represent additional information of the graph in **Appearance** panel. For instance, we can associate node color to **betweenness centrality** attribute, which measures the number of shortest paths between any two nodes that passes through a particular node. In other words, an airport node has higher betweenness centrality is connecting various different parts of the airline network, being considered to be more important globally. Figure 11 exemplifies this modification, where darker green represents higher betweenness centrality, and vice versa. Not surprisingly, we can notice that bigger nodes tend to have darker colors, meaning that busier airports are also playing more important roles in this network.



Nodes: 235 (2.34% visible)
Edges: 1297 (11.69% visible)
Undirected Graph

Queries
NOT (Nodes)
Non-null (background_map)
Parameters
Drag subfilter here

Figure 9: An updated visualization only shows nodes that are not background nodes.

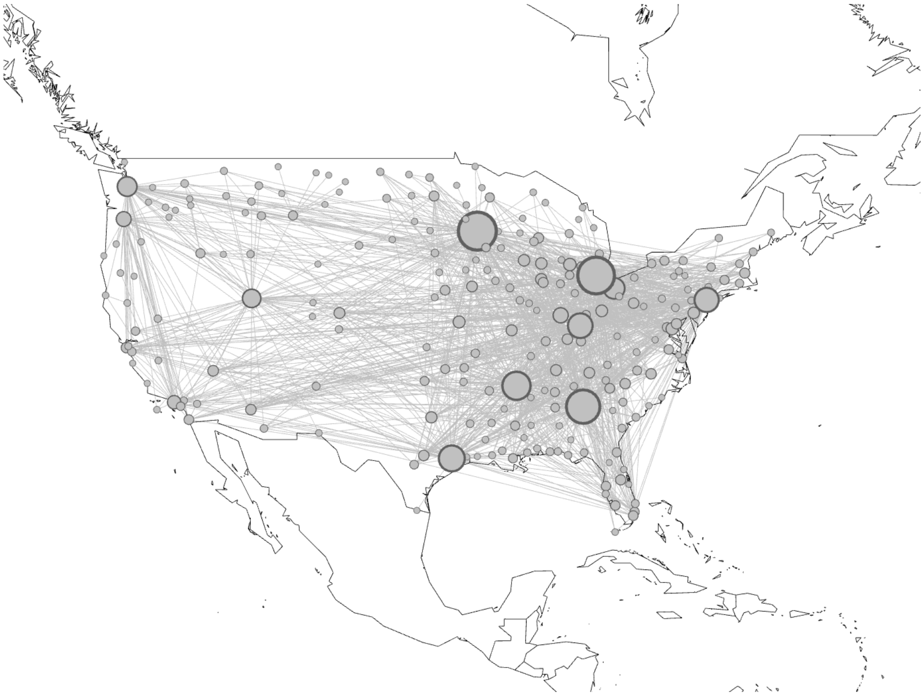


Figure 10: An updated visualization where the node size is varied based on the **degree** attribute.

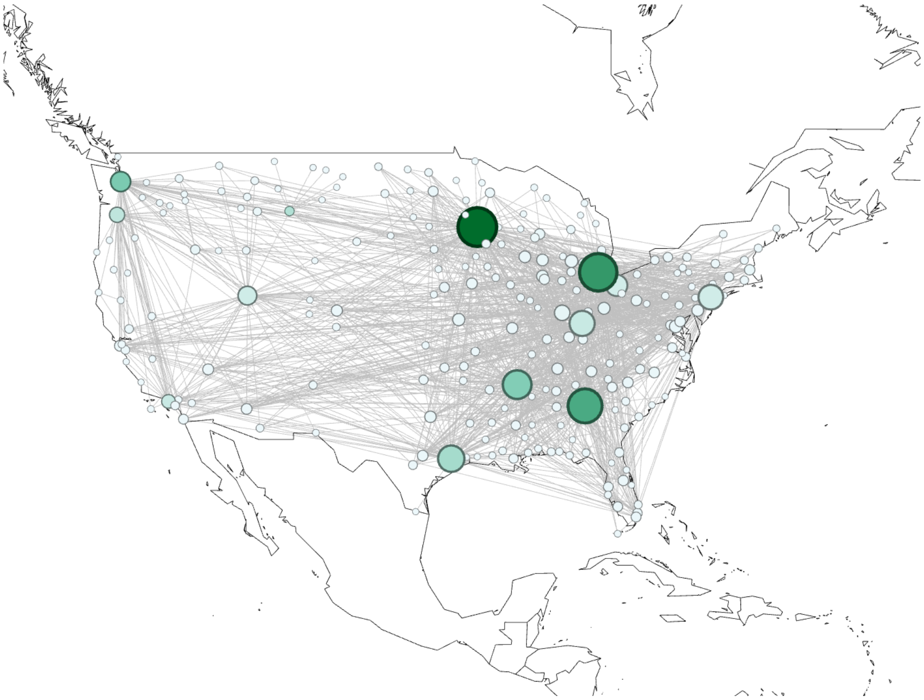
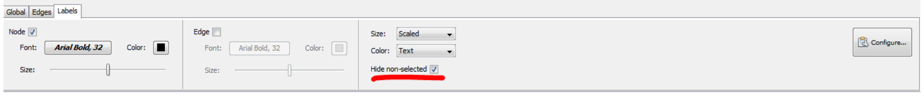


Figure 11: An updated visualization where the node color is assigned based on the **betweenness centrality** attribute.

The next straight forward question viewers would like to ask about this visualization is "what is the biggest and darkest airport?" This query can be answered by using Gephi's label and selection tools. We first open the label configurations panel, as shown in Figure 12, by clicking the small triangular icon located at the lower right corner of Graph panel, and then check the *Hide non-selected* option, as highlighted in the figure. We now can enable the *Direct selection* tool that located on the right side of Graph panel, and move the cursor onto a node we are interested in to see name of the corresponding airport, as shown in Figure 13.



Node ☑ Font: Arial Bold, 32 Color: ■ Size:

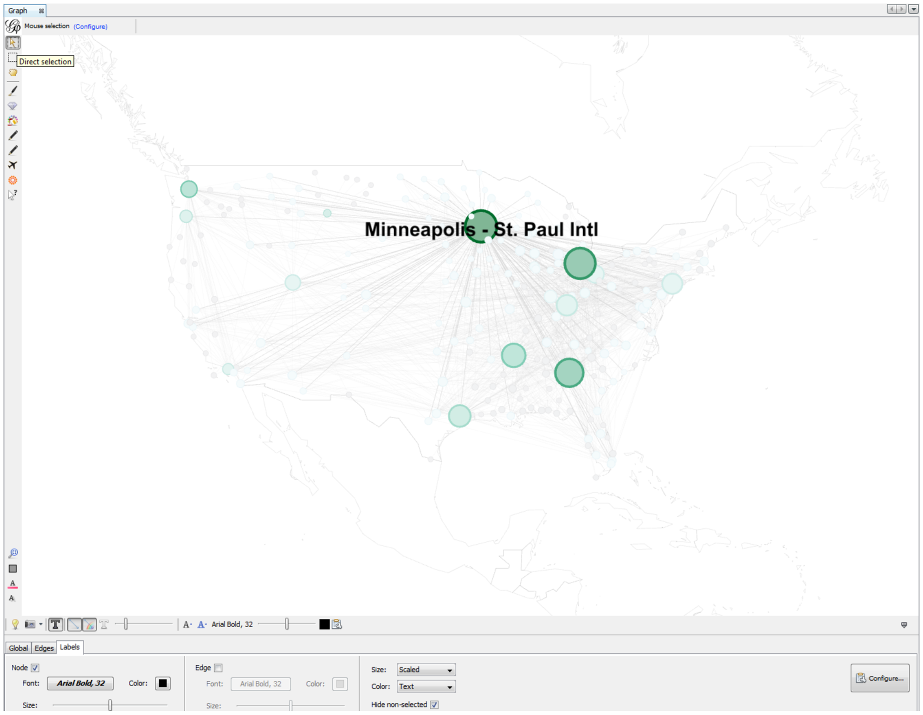
Edge □ Font: Arial Bold, 32 Color: □ Size:

Size: Scaled

Color: Text

Hide non-selected ☑

Figure 12: Configurations of label.



Node ☑ Font: Arial Bold, 32 Color: ■ Size:

Edge □ Font: Arial Bold, 32 Color: □ Size:

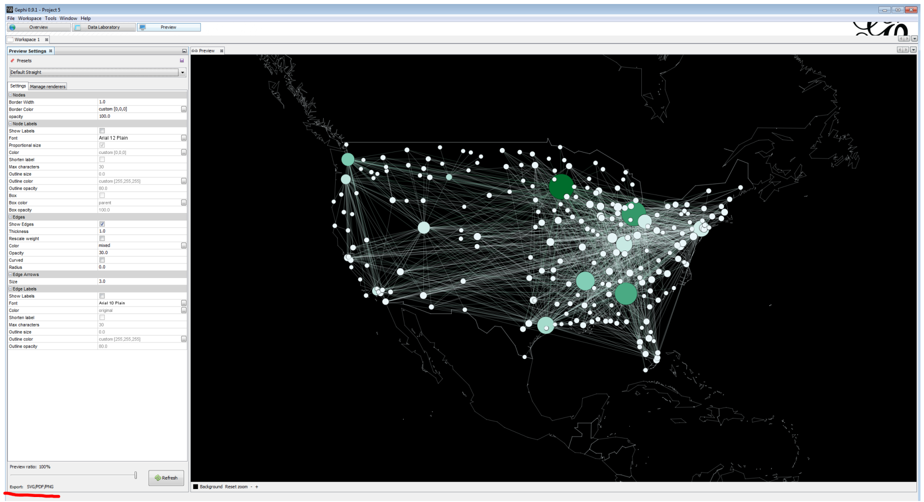
Size: Scaled

Color: Text

Hide non-selected ☑

Figure 13: An example of selecting the biggest node in the network.

Above we introduced the basic procedures of creating a visualization in Gephi, the final step is to output our visualization to an image file by using Gephi's **Preview** panel, as shown in Figure 14. It allows user to adjust rendering configurations, e.g. opacities of nodes and edges, showing labels or not,



Default Straight

⊟ Nodes	
Border Width	1.0
Border Color	custom [0,0,0]
opacity	100.0
⊟ Node Labels	
Show Labels	□
Font	Arial 12 Plain
Proportional size	☑
Color	custom [0,0,0]
Shorten label	□
Max characters	30
Outline size	0.0
Outline color	custom [255,255,255]
Outline opacity	80.0
Box	□
Box color	parent
Box opacity	100.0
⊟ Edges	
Show Edges	☑
Thickness	1.0
Rescale weight	□
Color	mixed
Opacity	30.0
Curved	□
Radius	0.0
⊟ Edge Arrows	
Size	3.0
⊟ Edge Labels	
Show Labels	□
Font	Arial 10 Plain
Color	original
Shorten label	□
Max characters	30
Outline size	0.0
Outline color	custom [255,255,255]
Outline opacity	80.0

Preview ratio: 100%

Export: SVG/PDF/PNG

Figure 14: Preview panel in Gephi.

drawing bounding boxes of texts or not, etc., and provides a preview of the configuration. After all settings have been specified, we can click **Export** button to output the display with supported file formats.