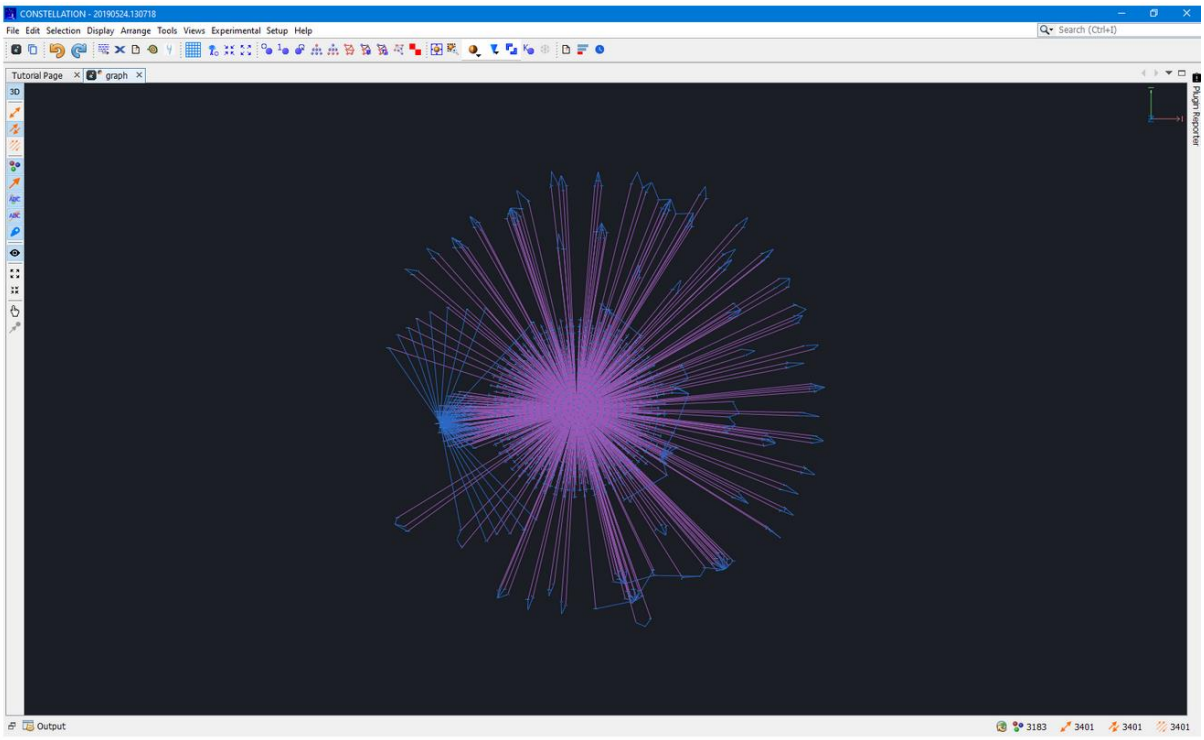
Constellation: What do?

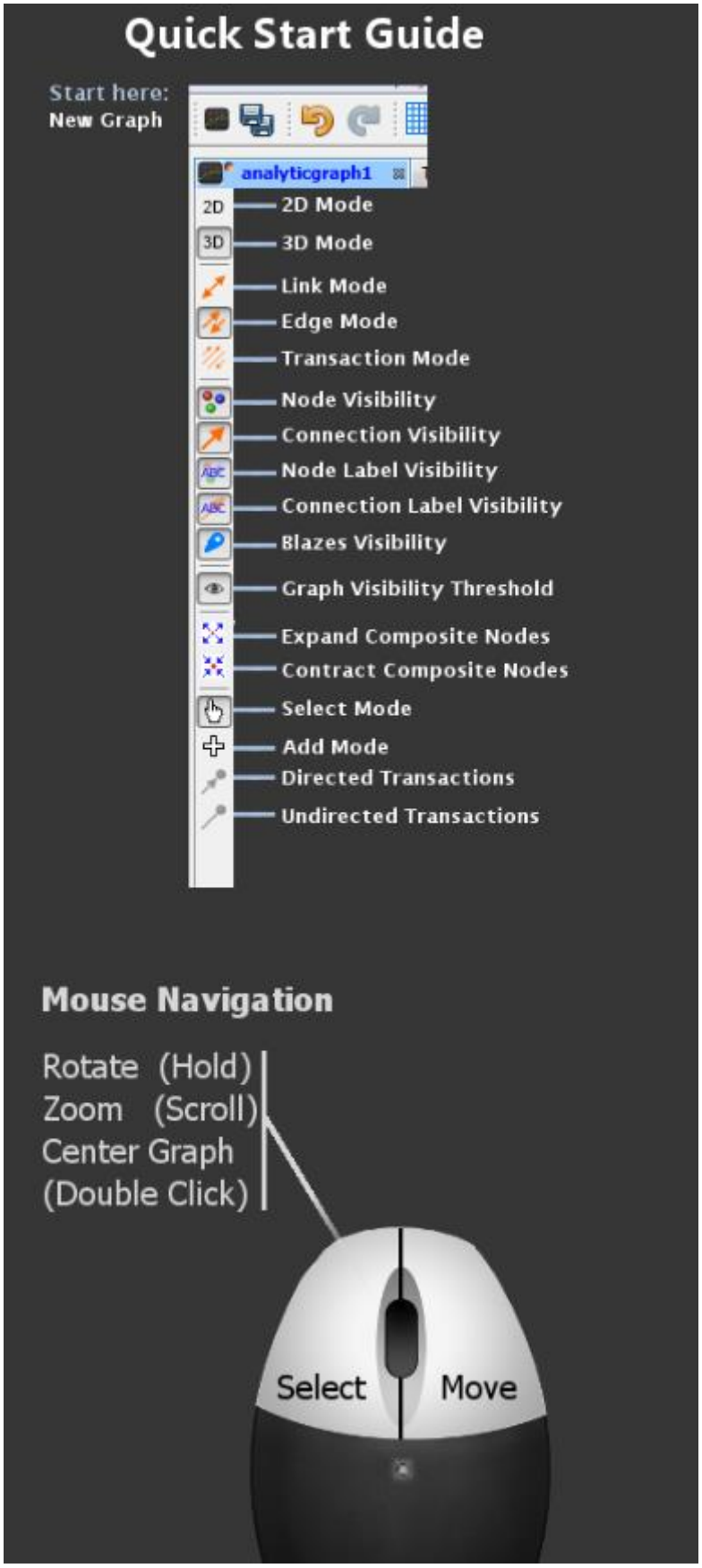
A guide to help you ***star***t.

Constellation is a graph-based data analysis application enabling data access, federation and manipulation capabilities across large and complex data sets. The application is useful for analysing networks, transactions, interactions, and relationships. It can query data stores directly or import various file types (csv, xlsx, etc.).



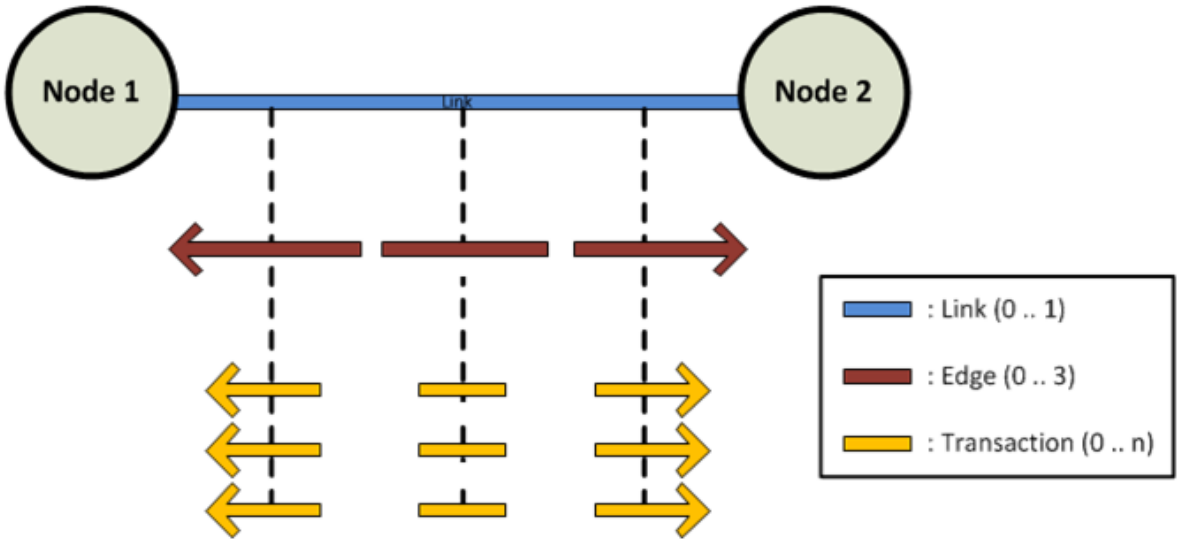
Constellation allows for approaching questions that only a graph structure can answer, as well as providing the capacity to visualise data in a variety of different ways concurrently through a multitude of views.

When Constellation opens, you will be presented with a tutorial page which includes some instructions for navigating the graph and a description of the visualisation settings available on the left-side toolbar.



Before we get started, first let's cover some terminology:

Terminology



By definition, a graph is simply a collection of entities, or (Nodes) and the relationships between them (Links). Sometimes, data is complex; Constellation expands on this concept by providing Edges and Transactions.

Node/Vertex

- An entity on the graph.
- Displayed on the bottom right-hand summary as:
    - Number of Selected Nodes / Total Number of Nodes

Transaction

- A connection between two nodes representing a single relationship.
- A transaction can be directed or undirected.
- The maximum number of rendered transactions is configurable in the Attribute Editor
- Displayed on the bottom right-hand summary as:
    - Number of Selected Transactions / Total Number of Transactions

Edge

- A collection of Transactions of a single direction between two nodes.
- There can be a maximum of three edges between two nodes (left, right, and undirected).
- Edges are the default connection type in Constellation.
- The red arrow on the count label points to the edge it applies to.
- Displayed on the bottom right-hand summary as:
    - Number of Selected Edges / Total Number of Edges

Link

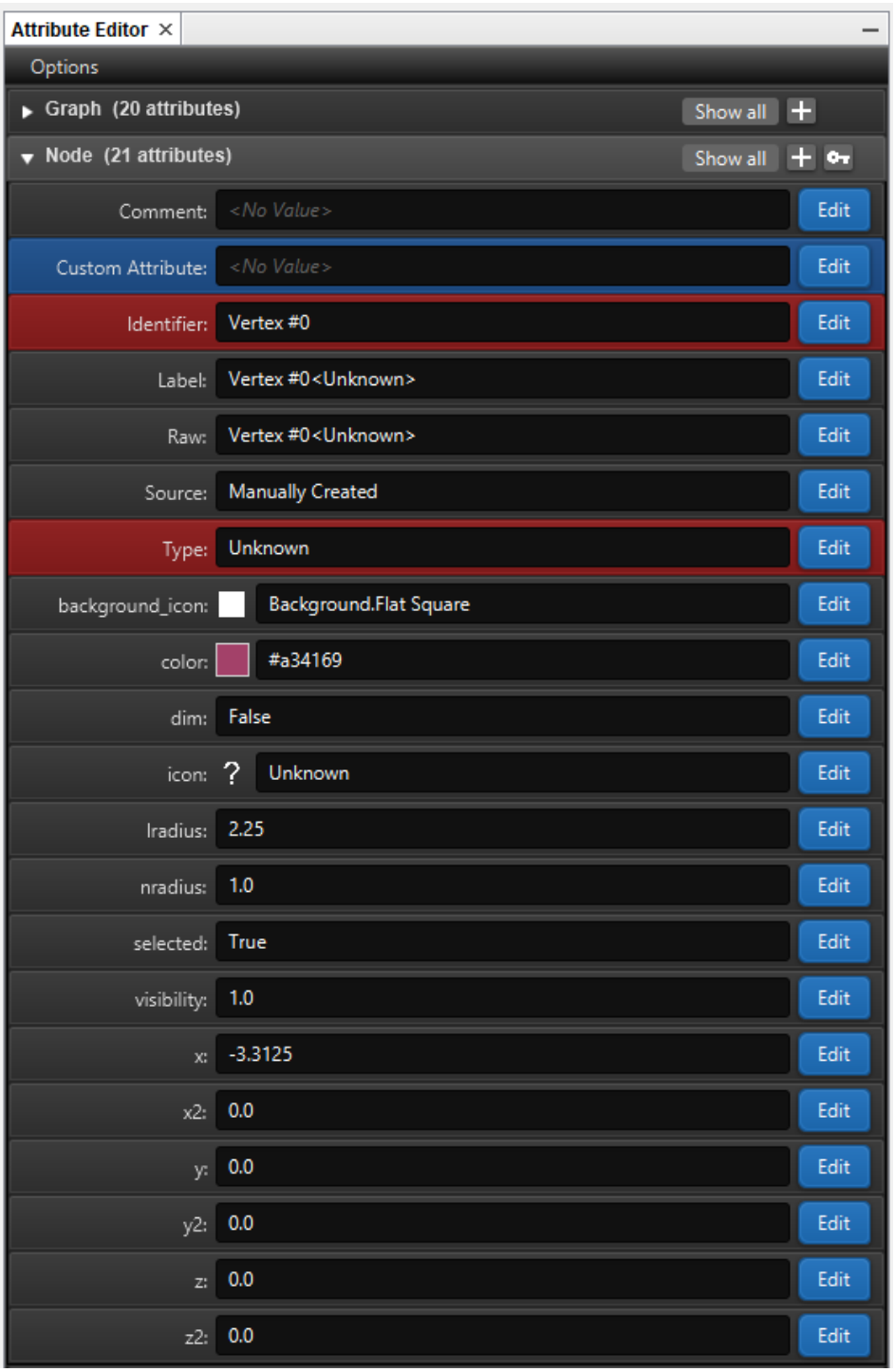
- A collection of all Transactions (direction agnostic) between two nodes.
- There can only ever be one link between two nodes.
- Displays on the bottom right-hand summary as:
    - Number of Selected Links / Total Number of Links

Attributes

Attributes are used for holding information about the elements of a graph. For example, the 'Type' attribute for a node might hold 'city', indicating that the node represents a city, and transactions to and from that node represent interactions or relationships with that city (for example, flights in and out to other 'city' nodes).

You can view the attributes for various objects on the graph by opening the **Attribute Editor**.

(Views > Attribute Editor, or Ctrl+Shift+E)



The Attribute Editor provides an interface for viewing and editing the attributes of any selected elements on the current graph. These attributes are colour-coded based on their type:

- Key Attributes:
    - Coloured red, these attributes are used to determine how nodes and transactions on your graph are merged together. An object on the graph is considered unique as long as all of its key attribute values together are not shared with any other object. Changing the value of a key attribute may cause merging in your graph to occur.
- Schema Attributes:
    - Coloured grey, these attributes are understood by the schema of your graph, and might be used to enable Constellation's various views and plugins.
- Custom Attributes:
    - Coloured blue, these attributes are not used by Constellation in any special way, but can be added to aide users.

You can change the value of any attribute by clicking on its 'edit' button within the Attribute Editor.

The Attribute Editor can also be used to add new attributes to your graph. This is done using the plus button next to the desired graph element. Attributes known to the schema can be added by selecting one from a category, or custom attributes can be defined by selecting 'Custom'.

Key attributes can be changed by clicking on the key button next to the desired graph element.

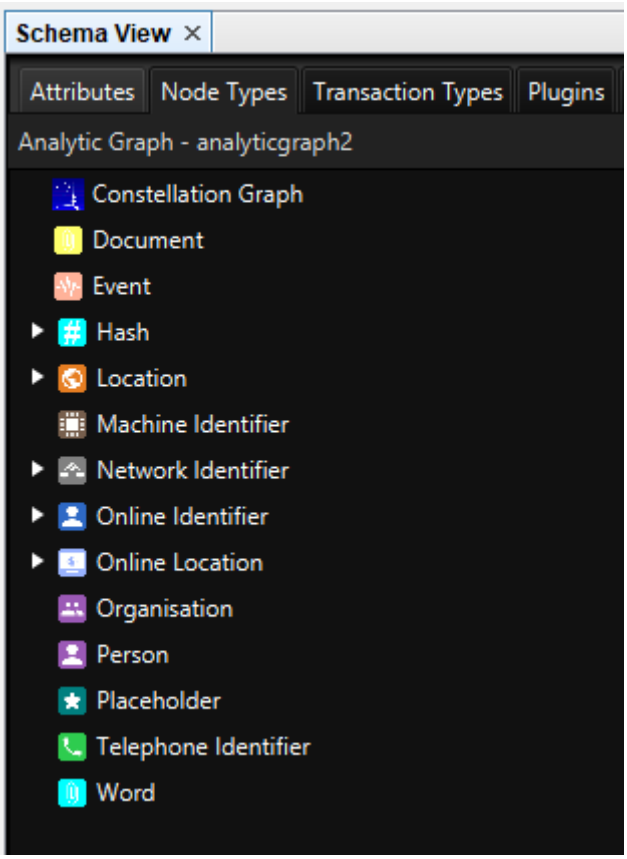
Graph Attributes allow you to configure the properties of the graph, such as the background colour, the number of transactions that are rendered, etc.

Schemas

Schemas are used to help Constellation interpret the data that is stored on the graph. Each graph type in Constellation has a 'schema' associated with it, allowing it to understand certain types of data and giving it the opportunity to improve data quality and visualisation based on this knowledge. When you open a new graph, you can specify the schema that the graph will use:

- Simple Graph:
    - The *Simple Graph* knows how to deal with visualisation, meaning colours, icons, labels and sizing will all work and can be edited, but the graph won't do anything smart with these properties on your behalf. Adding a node to a Simple Graph will create something visible in your graph with a default colour and a label.
- Analytic Graph:
    - The *Analytic Graph* is all about general graph analysis, meaning it knows about basic node and transaction types, and provides attributes which can be used for calculating properties of your graph such as clusters and centralities. Adding a node to an Analytic Graph will create a visual element where its colour and icon are determined by its type.

The **Schema View** (Views > Schema View, or Ctrl+Shift+S) provides a dictionary of what attributes, node and link types that Constellation will try to interpret, and how it will visualise them.



- Attributes Tab:
    - The attributes the current schema knows about. These attributes are often used by Constellation itself, for example, the attribute ‘nradius’ will be used to size the nodes in a graph, and the attribute ‘Datetime’ will be used to populate the Timeline.
- Node Types Tab:
    - The node types the current schema knows about. Nodes of a type will be provided certain benefits such as improved visualisation (colours, icons, etc.), and better interaction with Constellation analysis features.
- Transaction Types Tab:
    - The transaction types the current schema knows about. As with node types, transactions of this type will be provided certain benefits.

Importing Data

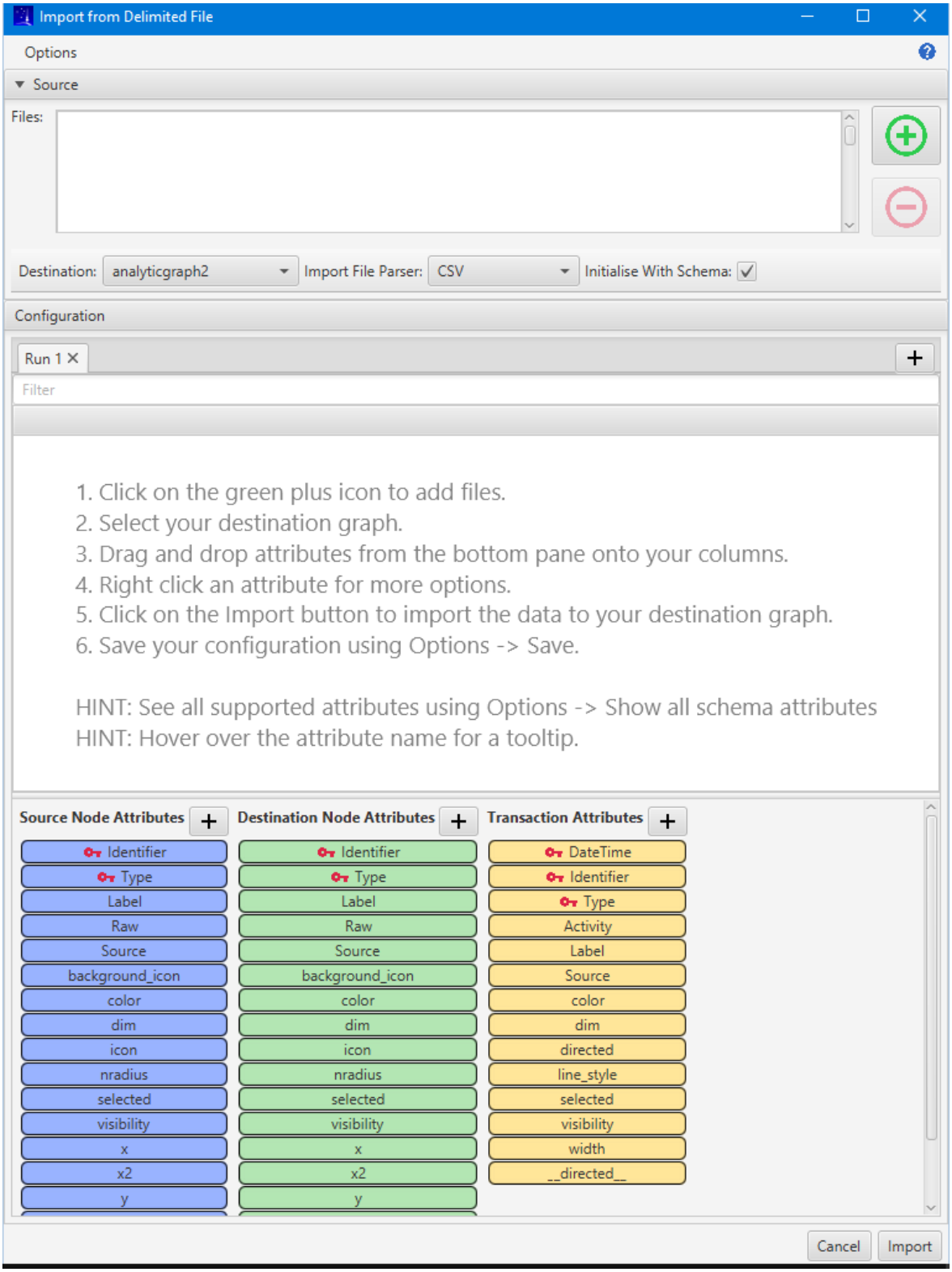
Now that we've covered the terminology, open up a new graph and we'll look at importing data.

If you have a delimited file that describes some relationships, we can load it into Constellation using the Delimited File Importer.

Delimited File Importer

File > Import > From Delimited File.

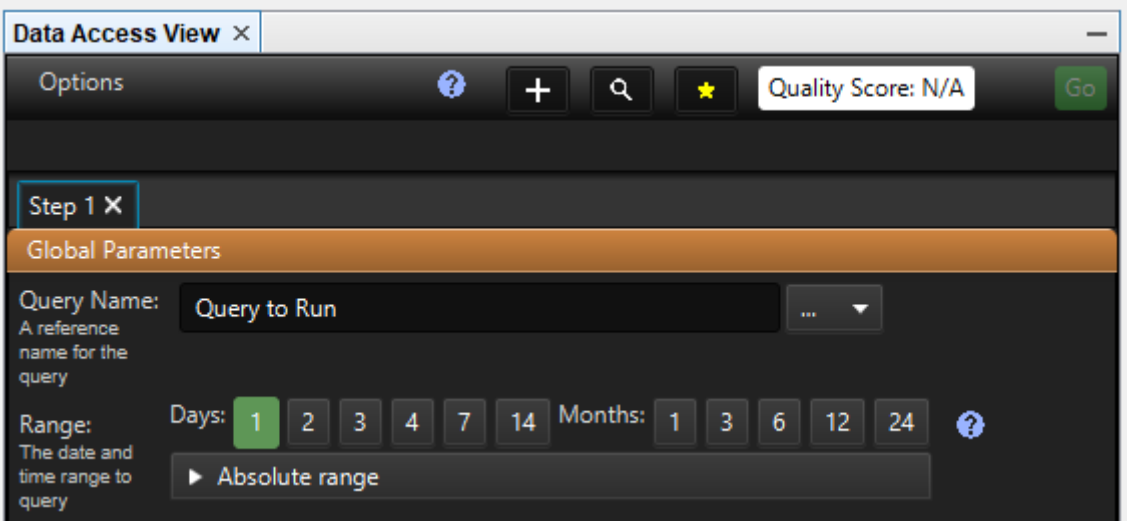
1. Click on the green + button and browse to the file.
2. Set the 'Destination' to either an existing graph or a new graph.
3. Set the 'Import File Parser' to CSV.
4. Leave 'Initialise with Schema' ticked, in order to allow the schema to repair data.
5. Drag and drop the source and destination vertex 'Identifier' attributes to the appropriate columns in the table. This will indicate which entities represent the "from" and "to" nodes.
6. Drag other source/destination attribute labels to appropriate columns. You can also right-click to specify default values.
7. If you have timestamps, drag the transaction DateTime label to the appropriate column. Right click to Set Custom Format to match the date time style.
8. Drag other transaction descriptions to the appropriate columns.
9. You can also create custom attributes where needed by clicking the + button next to the Source, Destination and Transaction Attribute column headers.
10. Once all attributes have been assigned to columns, click the 'Import' button to add the table to your graph.



Alternately, if a specialised data importer has been created for importing, chaining, or enrichment data, you will find it in the Data Access View.

Data Access View (Ctrl + Shift + D)

The Data Access View is a versatile view used to import data, depending on how database administrators have enabled access to data stores.



At the very top, there is a list of tabs. Pressing the + button will add tabs, and the X in the tab name will remove it. Tabs can be customised to perform different queries, and when ready, pressing the Go button will cause each tab to run their queries sequentially. The Go button will only highlight once a plugin check box has been selected and a graph has been opened.

From the list of plugins, you can select which queries they wish to run. The arrows allow options to be visible or hidden. Selecting the checkbox next to the name of the query will cause that query to run when Go is pressed.

Note that the queries in the Data Access View only run using graph elements that have been selected.

Note that when Go is pressed, the fields in the Global Parameters section like **Query Name** are saved as part of the graph's state so that when you save and load the graph, the last used values are restored.

You can filter plugins by clicking on the magnifying glass button and typing into the text field beneath. Alternatively, you can right click anywhere on the Data Access View and choose the **Search plugins** option.

You can add plugins to your favourites category by clicking on the yellow star button which will allow you to add or remove plugins you have selected from the favourites category.

The top of the Data Access View contains the set of parameters, which apply to every plugin within the Data Access View. This includes a name for your query, the date range for which the data is based on (if available).

The datetime range control allows you to enter a datetime range in either relative or absolute form. The relative section provides predetermined relative periods. The absolute section provides an unchanging from and to range.

A relative range indicates a period ending "now" and starting at the indicated time in the past. For example, if 2 days is selected, and a data access query is run at 2015-01-30 09:30, the range used for the query will be 2015-01-28 09:30:00 to 2015-01-30 09:30. If another query is run ten minutes later, the range will be 2015-01-28 09:40:00 to 2015-01-30 09:40.

This is also true if a relative range is saved and loaded at a later time: the datetime range will always end at "now".

Note that selecting a relative period will update the absolute range fields. However, this is done as an indication of the period when it was selected. Since the relative period ends "now", and is therefore continually changing, the values displayed in the absolute section are immediately obsolete.

An absolute range allows you to specify the start and end of the range as fixed instances in time. For example, if 2015-01-23 09:30 and 2015-01-30 09:30 are used, the same range will be used whenever a query is run.

The absolute range also allows to you select a timezone to be used to display the absolute time range. A drop-down list contains all available timezones, with convenience buttons for UTC and local timezones.

The **Use Maximum Time Range** button sets the time range to start at 00:00:00 and end at 23:59:59.

When the timezone changes, the time instant of each absolute datetime remains the same: only the timezone in which the datetimes are displayed is changed.

The date range is highlighted in green to indicate whether a relative or absolute range is being used.

The Quality Score next to the Go button in the Data Access View brings us to the Quality Control View.

Quality Control View (Ctrl + Shift + Q)

The Quality Control View is a quick and easy way to look at nodes on the graph and determine whether or not they have any obvious quality issues based on a series of rules built into Constellation. Selecting nodes in the graph will cause them to be processed against any registered quality control rules, and ranked in the Quality Control View according to their quality. This quality rating will be a score out of 100 (with higher scores indicating lesser quality) along with reasoning as to why they may pose a quality control issue. You can then highlight rows in the quality control view and apply options via the buttons at the bottom of the view.

There are four levels of quality control for nodes that are selected.

- Any node which is specifically disallowed (whether it be disallowed for queries, or simply should not be in your graph) will have a black background, and may be blocked from further analysis.
- Any node which is considered of particularly bad quality will have a red background, but will never be blocked.
- Any node which is considered of questionable quality will have a blue background.

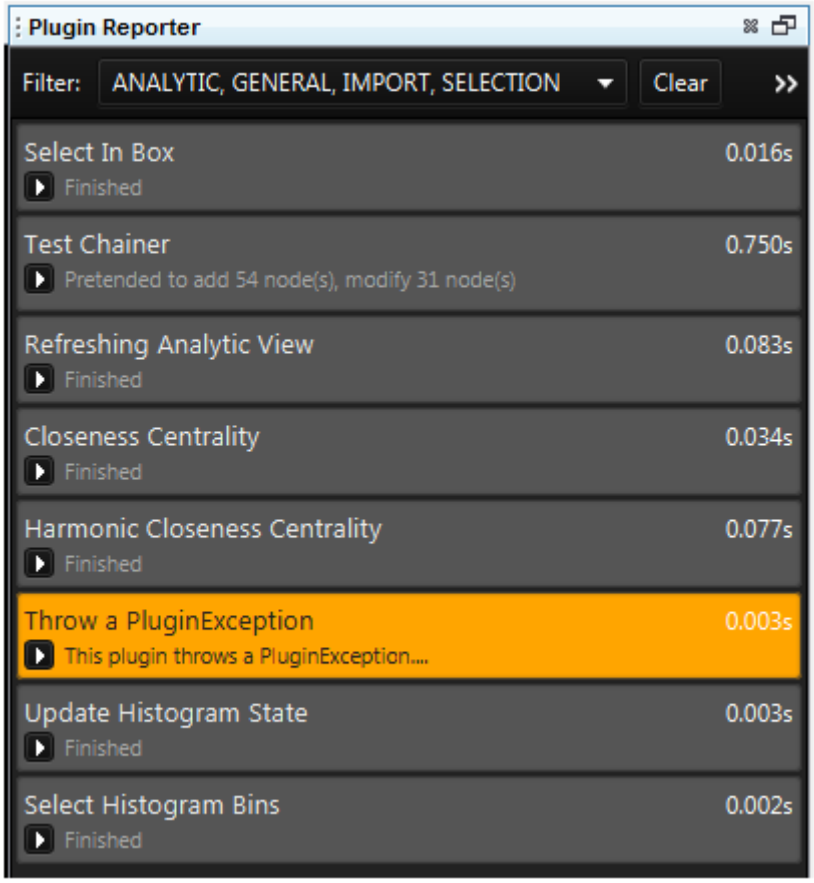
- Any node which is considered of good quality will have a white background. This does not mean that these nodes are necessarily of high quality, only that no quality control rules matched - such nodes require manual checking by the user.

Double-clicking a row in the Quality Control View will open a dialog box that shows the scores given to the node in that row by the rules.

- Zoom on Graph: Resets the camera on the graph to make every selected node visible.
- Select on graph: Selects the nodes highlighted in the Quality Control View and deselects everything else in the graph.
- Deselect on Graph: Deselect the nodes highlighted in the Quality Control View from the selection in the graph.
- Delete from Graph: Deletes the nodes highlighted in the Quality Control View from the graph.

If you've run some plugins with the Data Access View and want to check how they've gone, you can use the Plugin Reporter

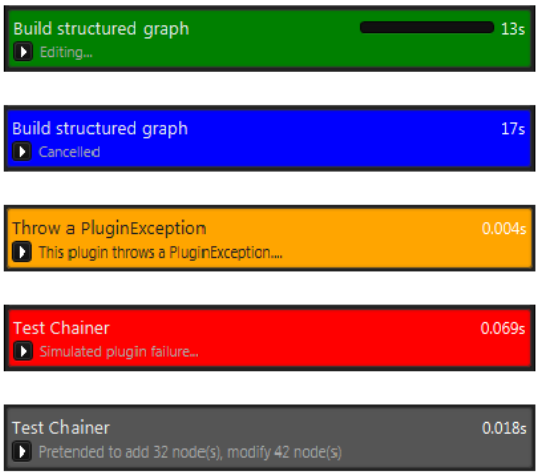
Plugin Reporter (Ctrl + Shift + P)



The Plugin Reporter contains a history of all plugins that have been run on a graph. In Constellation, almost every function is a plugin, so this essentially describes the analytic process that an analyst has taken for a graph.

A plugin report will be created whenever a plugin is run, and added to the Plugin Reporter. These reports will change their colour and message based on the current status of the plugin they represent.

- **Green indicates the plugin is currently running.**
- **Blue indicates that the plugin was canceled.**
- **Orange indicates that an anticipated error occurred.**
- **Red indicates that an unexpected error occurred.**
- **Grey indicates that the plugin finished successfully.**



You are able to filter the plugins which are displayed by category, if you choose only the ‘IMPORT’ and ‘SEARCH’ categories, you will see all Data Access View plugins, including whether they failed or succeeded, and in the latter case how many records they added to the graph.

The Plugin Reporter allows filtering of the plugin history using a tag-based system. You can filter on plugin reports using by their tags using the 'Filter' drop down menu.

- **Analytic** Plugins which perform some analytic calculation based on information on the graph.
- **Import** Plugins which add information to the graph, generally these will be found in the Data Access View.
- **General** Plugins which don't fit into any specific category.
- **Low Level** Plugins which are generally hidden from the user.
- **Selection** Plugins which change what is selected on the graph.

The 'Clear' button will clear all currently displayed plugin reports, and only display new plugin reports for the active graph.

The 'Show All' button will make the Plugin Reporter display all plugin reports for the active graph.

The 'Help' button will open this help documentation.

Analysing Data

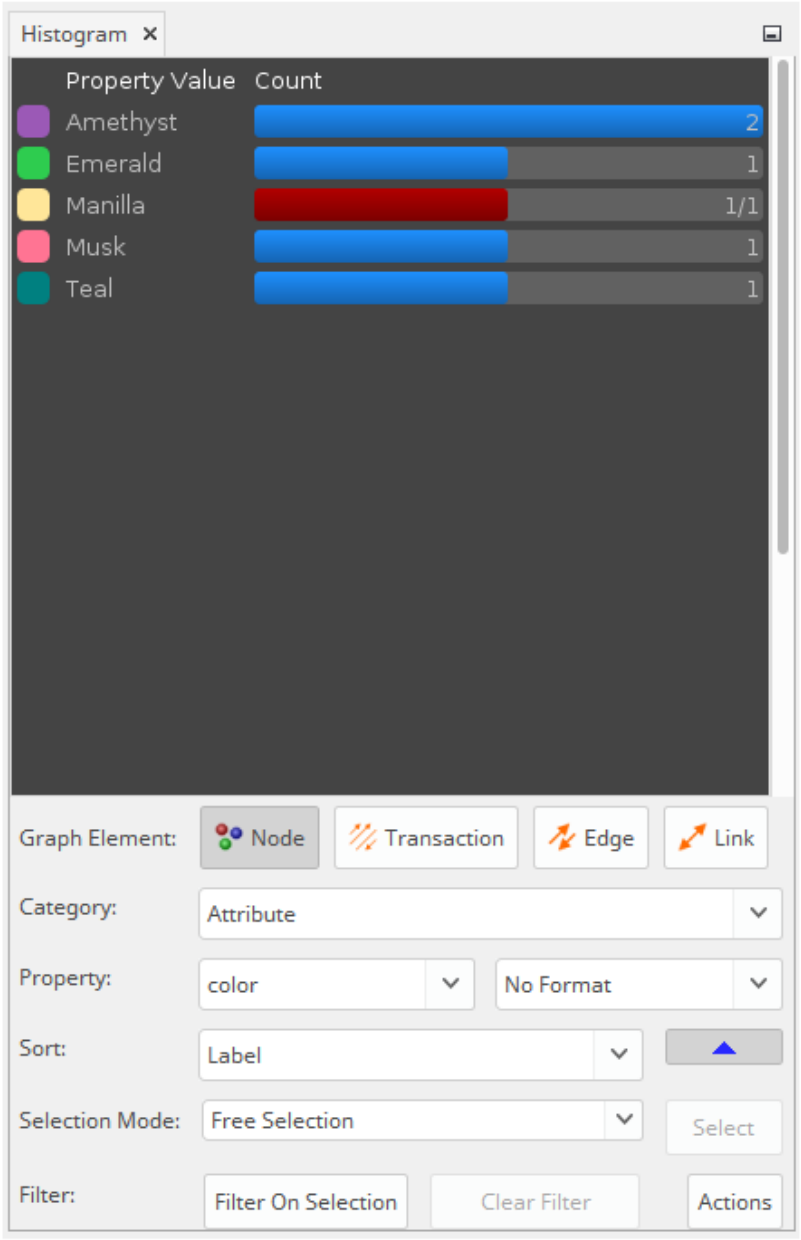
Now that we have some data in Constellation, let's cover some of the ways we can view/study the data.

The Edit, Selection and Arrange menus offer a variety of ways of organising/filtering the data on the graph, and are a great place to start. There are also visualisation options for adding blazes to highlight nodes and dimming options for nodes and transactions. Each have a page in the help topics which describe what they do, so we won't dwell too much on them here

Instead, we'll explore some of the Views that allow us to answer more complex questions, starting with the Histogram View.

Histogram View (Ctrl + Shift + H)

The Histogram View provides a way to summarise both the structural properties of a graph and the values of attributes, as well as the capability to filter and make selections based on these properties.



Histogram Panel

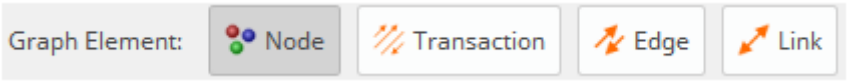
The Histogram Panel displays an interactive summary of information on the graph in a visual bar chart. The left column, **Property Value**, is a list of unique values for a particular property, and the **Count** column displays the number of graph elements which share that property value. The bar length indicates the proportion of elements that share that property when compared to the value that is most prevalent on the graph. It is blue for unselected elements, and red for selected elements.

When only a portion of the elements sharing a property value are selected on the graph, the count value will also change to the ratio of selected elements / total elements sharing that property value. It will also be coloured both red and blue, with the proportion of red matching the ratio of selected / unselected.

Left clicking on a row in the Histogram Panel will select all elements on the graph which share that Property Value.

Holding left click and dragging across multiple rows will select multiple rows.
Holding the shift key while left clicking will add the clicked row elements to the selection.
Holding the ctrl key while left clicking will toggle the selection of the clicked row elements.

Graph Element Buttons



The Graph Element Buttons will toggle which graph elements the Histogram View will list. Options include Node, Transaction, Edge, and Link.

Category Drop Down List

The Category Drop Down List will display categories of properties to display in the Histogram View. Depending on the Graph Element type, this list will change. Most commonly used categories are as follows:

Node Categories

1. Attribute - Properties listed will be the attributes of the node.
2. Graph Property - Properties will include graph properties such as neighbour count and transaction count.

Transaction Categories

1. Attribute - Properties listed will be the attributes of the transaction.
2. Graph Property - Properties will include transaction properties such as transaction direction.

Property Drop Down List

The Property Drop Down List will display properties to histogram. The properties listed will change depending on the Graph Element and Category selected - see above.

The Format Drop Down List will provide different ways of grouping values for properties, depending on the type.

1. String - Allows for grouping by substring, case (upper or lower) and find (which accepts regular expressions).
2. Numeric - Allows for binning into buckets of values.
3. DateTime - Allows for grouping by various DateTime formats - Hour of Day, Day of Week, Month of Year, Date, Year, and so on.
4. Schema Type - Allows for grouping by Top-Level Schema Types

Sort Drop Down List

The Sort Drop Down List will display options for sorting the rows of the Histogram.

The Arrow Button will toggle Ascending/Descending values.

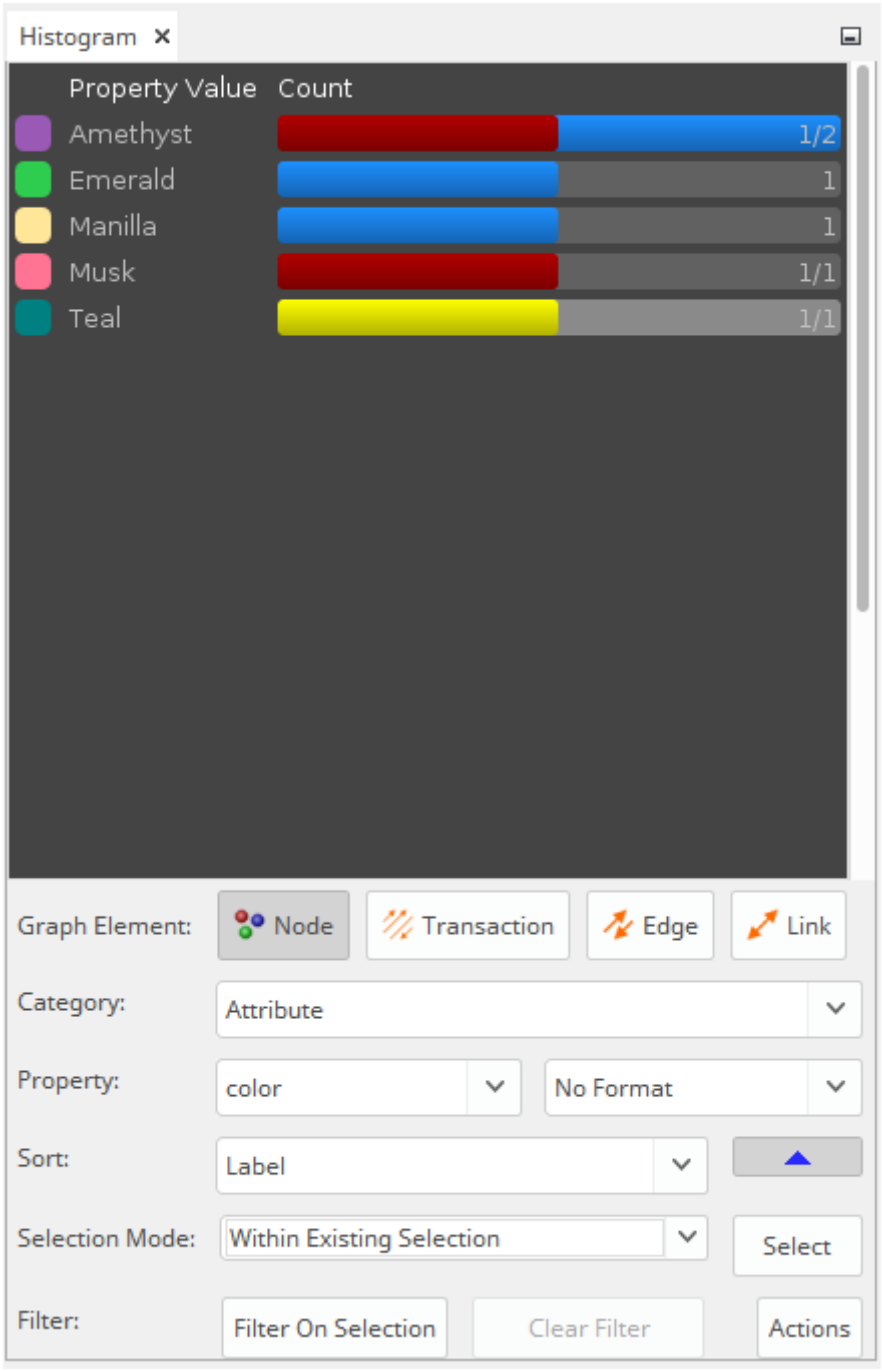
1. Label - Will sort the rows of the Histogram alphabetically by Property Value label.
2. Total Count - Will sort the rows of the Histogram in order of Count values.
3. Selected Count - Will sort the rows of the Histogram in order of the number of elements sharing a Property Value that are selected.
4. Selected Proportion - Will sort the rows of the Histogram in order of the proportion of elements sharing a Property Value that are selected.

Selection Mode Drop Down List

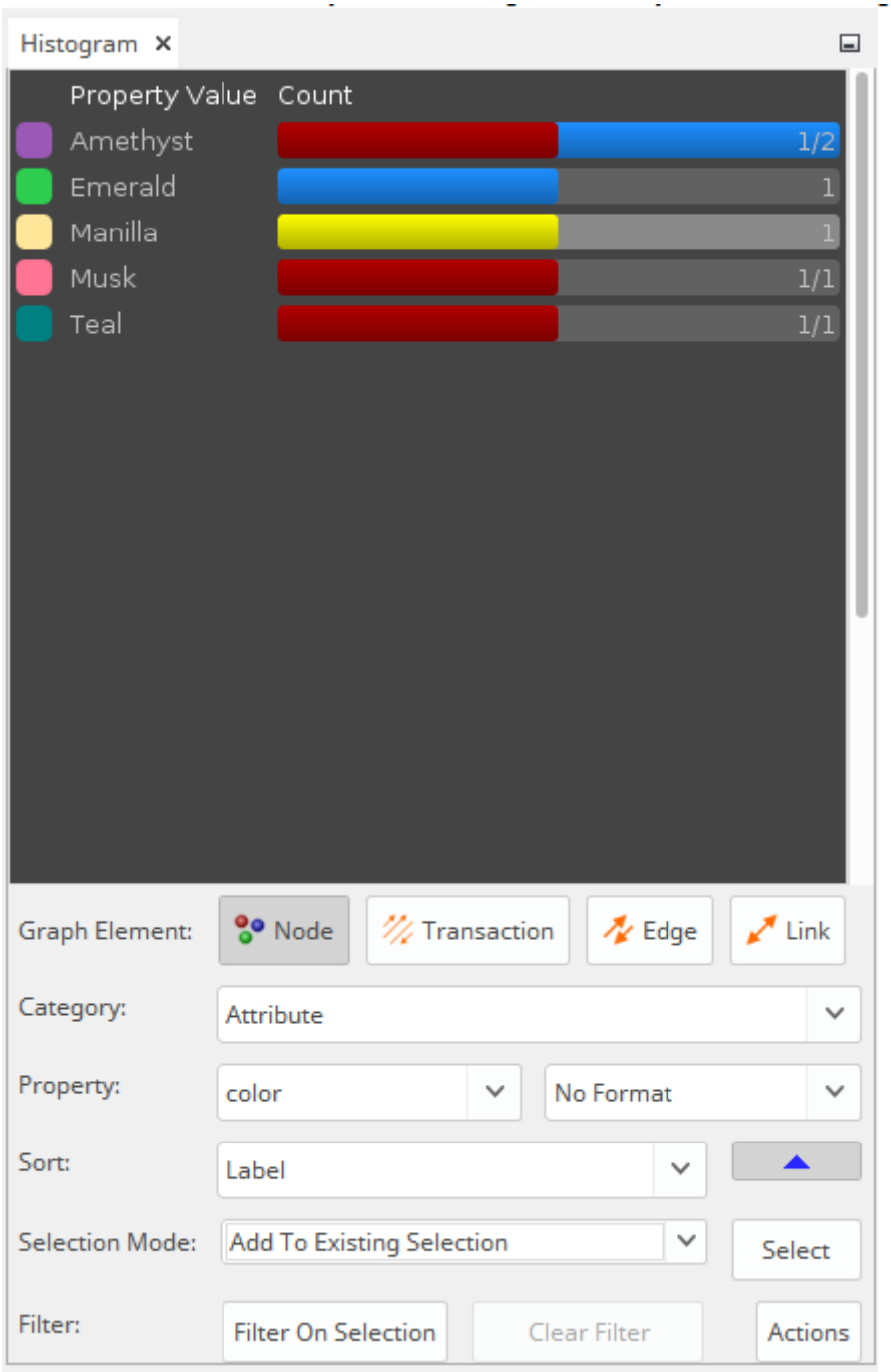
The Selection Mode will toggle the interactivity of the Histogram View.

1. Free Selection - The default setting. Clicking rows will change the selection on the graph.
    - Left clicking on a row in the Histogram Panel will select all elements on the graph which share that Property Value.
    - Holding left click and dragging across multiple rows will select multiple rows.
    - Holding the shift key while left clicking will add the clicked row elements to the selection.
    - Holding the ctrl key while left clicking will toggle the selection of the clicked row elements.

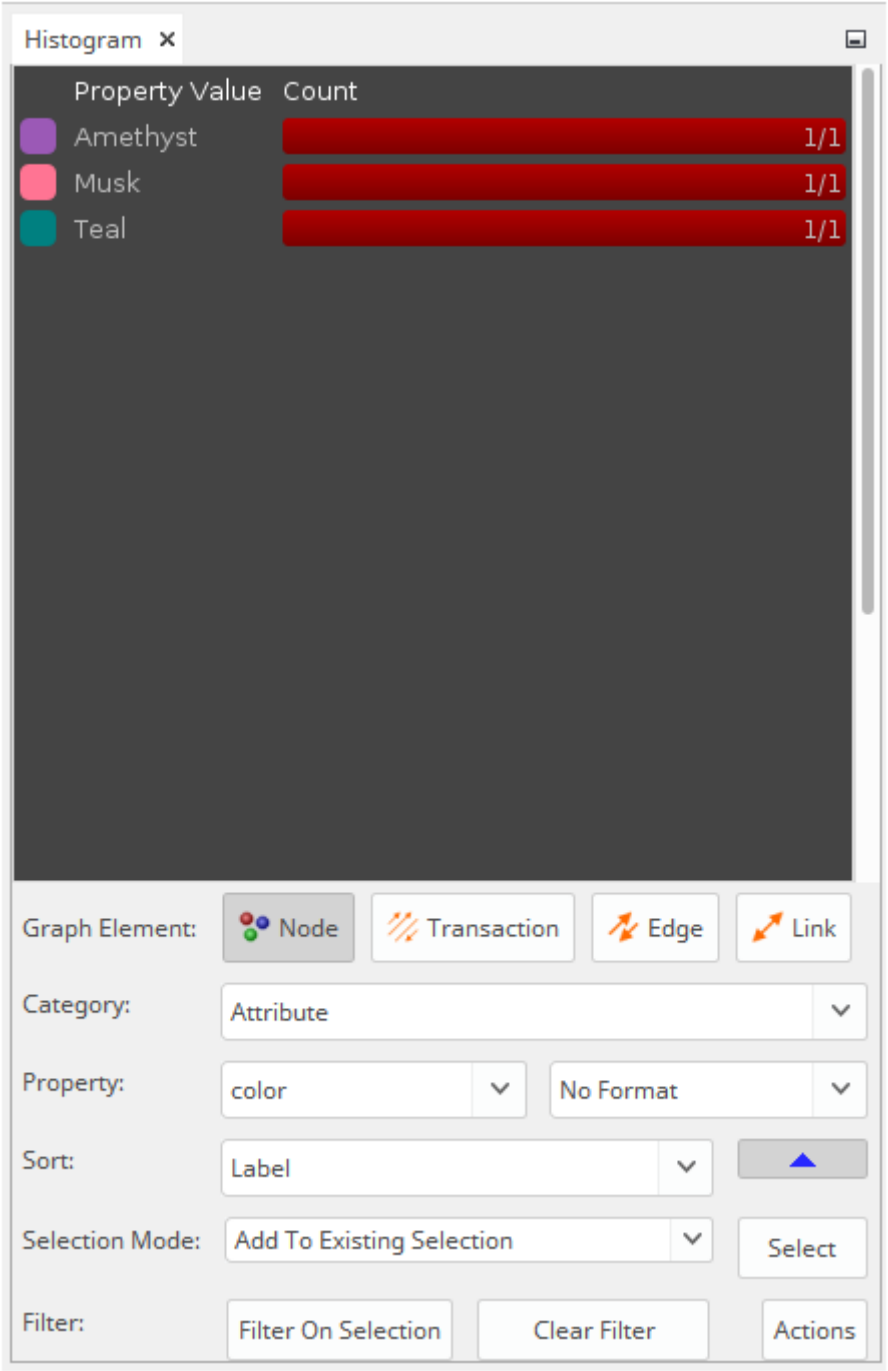
2. Within Existing Selection - Clicking rows will not change the selection on the graph until the adjacent **Select** button is pressed. This mode will allow for filtering on the existing graph selection.
    - Clicking on a row with selected (red) elements will change their colour to yellow. Pressing the adjacent **Select** button will change the selection on the graph to all yellow elements.
    - Holding the shift key while left clicking will change red elements of rows to yellow.
    - Holding the ctrl key while left clicking will toggle elements of rows between red and yellow.



3. Add To Existing Selection - Clicking rows will not change the selection on the graph until the adjacent **Select** button is pressed. This mode will allow for adding rows to the existing graph selection.
    - Clicking on a row with unselected (blue) elements will change their colour to yellow. Pressing the adjacent **Select** button will change the selection on the graph to all yellow elements.
    - Holding the shift key while left clicking will change blue elements of rows to yellow.
    - Holding the ctrl key while left clicking will toggle elements of rows between blue and yellow.


Histogram
Property Value
Count
Amethyst
1/2
Emerald
1
Manilla
1
Musk
1/1
Teal
1/1
Graph Element:
Node
Transaction
Edge
Link
Category:
Attribute
Property:
color
No Format
Sort:
Label
Selection Mode:
Add To Existing Selection
Select
Filter:
Filter On Selection
Clear Filter
Actions

Filter



The Filter on Selection button will hide unselected elements from the Histogram View, until such time as the Clear Filter Button is pressed. The Clear Filter Button will activate only when a Histogram filter is in place, and as such is useful to check when elements are missing from the Histogram View.

Actions

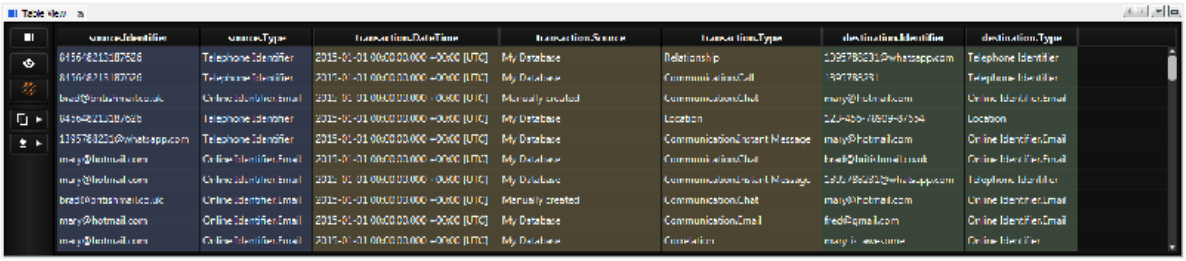
Pressing the Actions button will reveal two options: Save Bins To Graph and Save Bins To Clipboard

1. Save Bins to Graph - Will add a histogram_bin attribute to the graph, and the value of which corresponding to the Property Value of the node or transaction.

2. Save Bins to Clipboard - Will copy the Property Values and Counts displayed in the Histogram View to the clipboard, which can be pasted into a text editor in a tab delimited format.

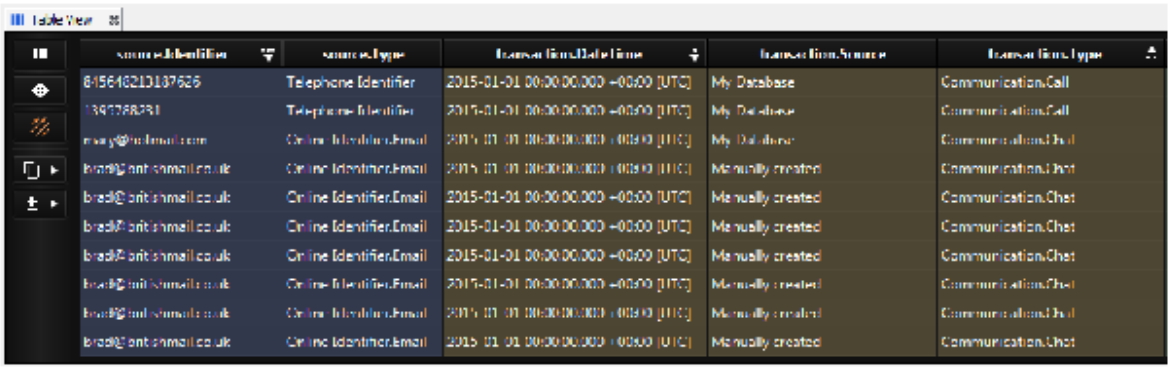
Table View (Ctrl + Shift + Y)

The table view presents attribute data from the graph in a tabular format. Highlighted rows in the table represent selected elements in the graph. Selection of the table rows (including the use of the shift and control buttons) will result in the selection of the table rows and the associated elements in the graph itself. Conversely, changes to the graph selection are reflected in the table.



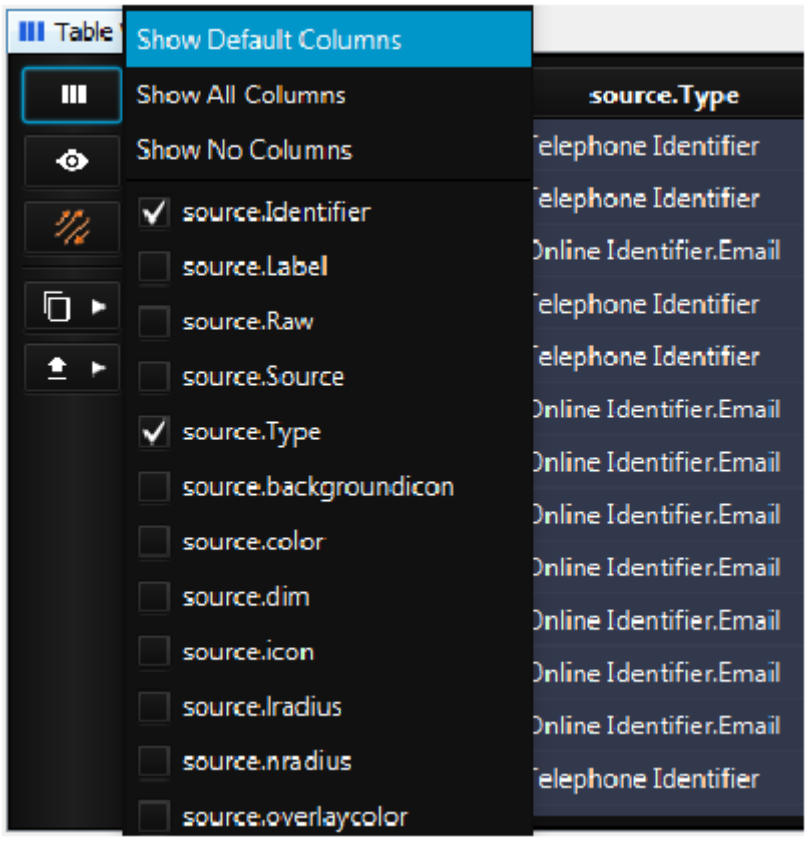
Clicking the element type toolbar button will switch between tabular views of transaction data (which includes the nodes at either end), or node data.

Left-clicking any column header will sort the table by the values in that column. A second click will reverse the sort order and a third click will remove the sort. If you hold shift while you click on column headers you can sort by multiple columns. This sorting will occur in the order that you click the columns (as indicated by the dots and numbers).



Column Visibility

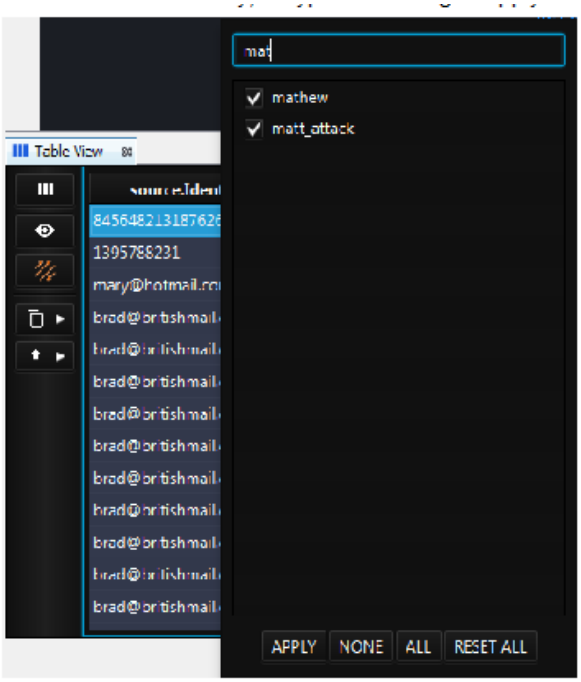
Clicking the column visibility toolbar button will open a menu allowing you to customise which attributes are displayed in the table.



Clicking the selected only toolbar button will hide any elements which are not selected on the graph. Note that while this option is enabled, selection in the table will not update selection on the graph.

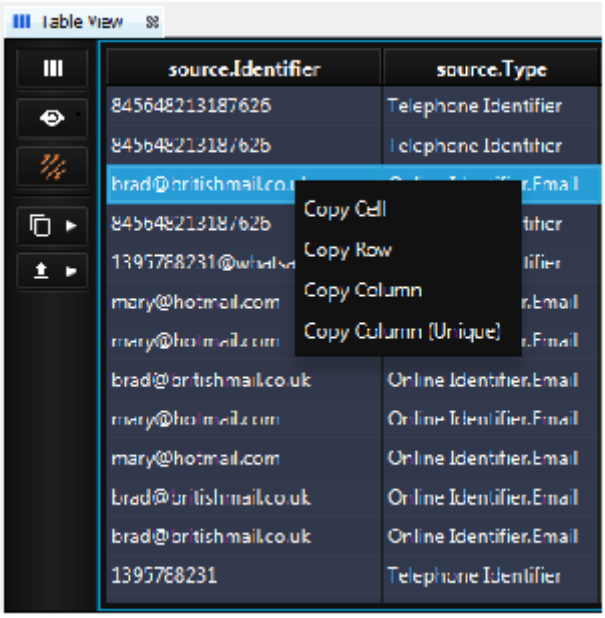
Column Filtering

Right-clicking any column header will open a filter dialog allowing you to select / deselect values manually, or type something to apply a filter to the data.



Export

Right-clicking anywhere on the table will open a context menu providing options to copy data from the clicked cell, row or column.



Copy Table

Clicking on the copy toolbar button will provide you with options to copy the table to the system clipboard. The table will be copied exactly as it appears in the Table View.

Export Table

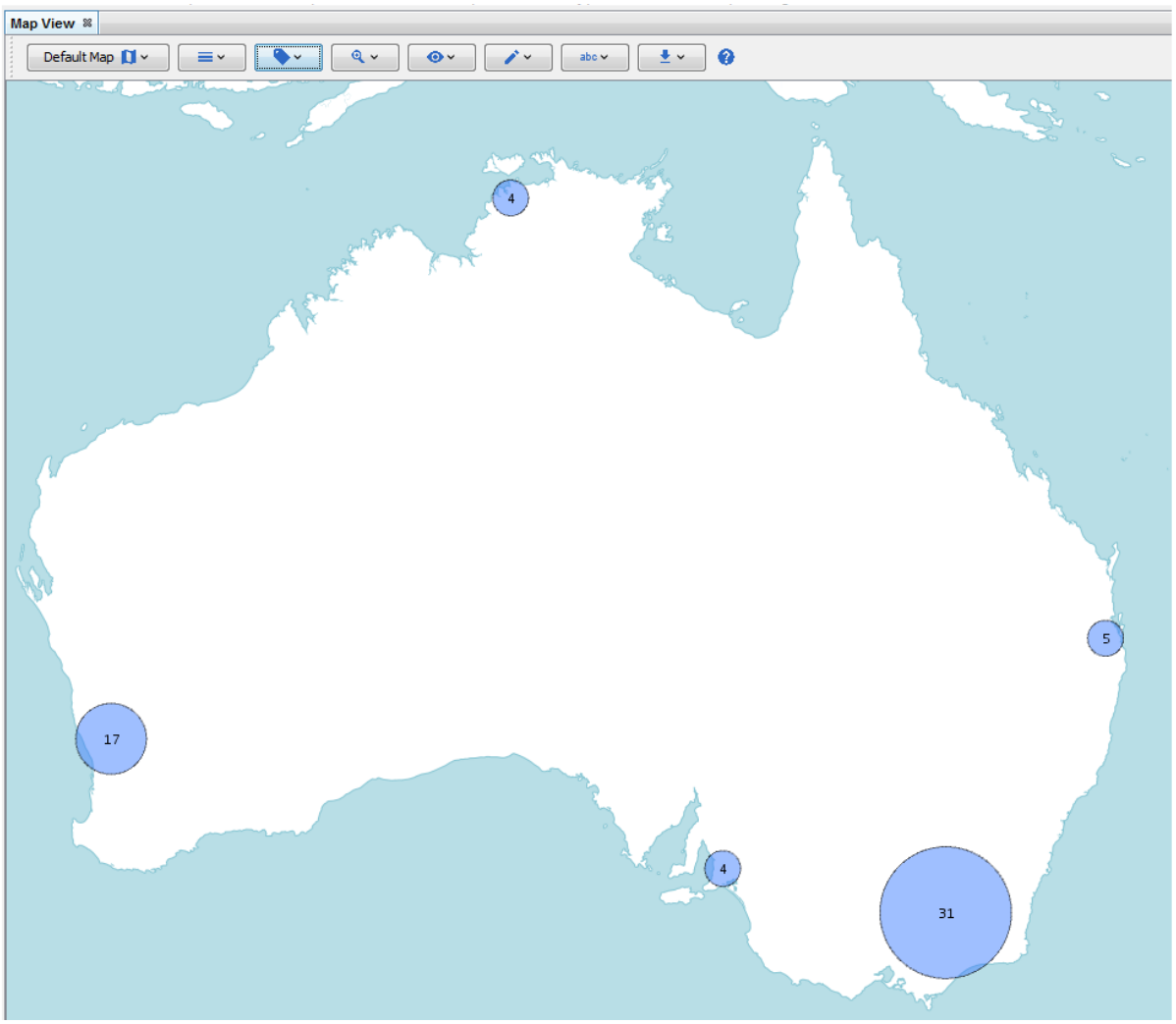
Clicking on the export toolbar button will provide you with options to export the table in various file formats. The table will be copied exactly as it appears in the Table View.

Context Specific Views

There are also views that are catered to visualising specific attributes, or specific types of attributes. Again, there are help topics which cover these in more detail.

- Timeline View (Ctrl + Shift + T)
    - This view is useful for studying temporal/datetime attributes

- Map View (Ctrl + Shift + M)
    - This view is useful for visualising geospatial attribute data (Geo.Latitude, Geo.Longitude, Geo.Shape)

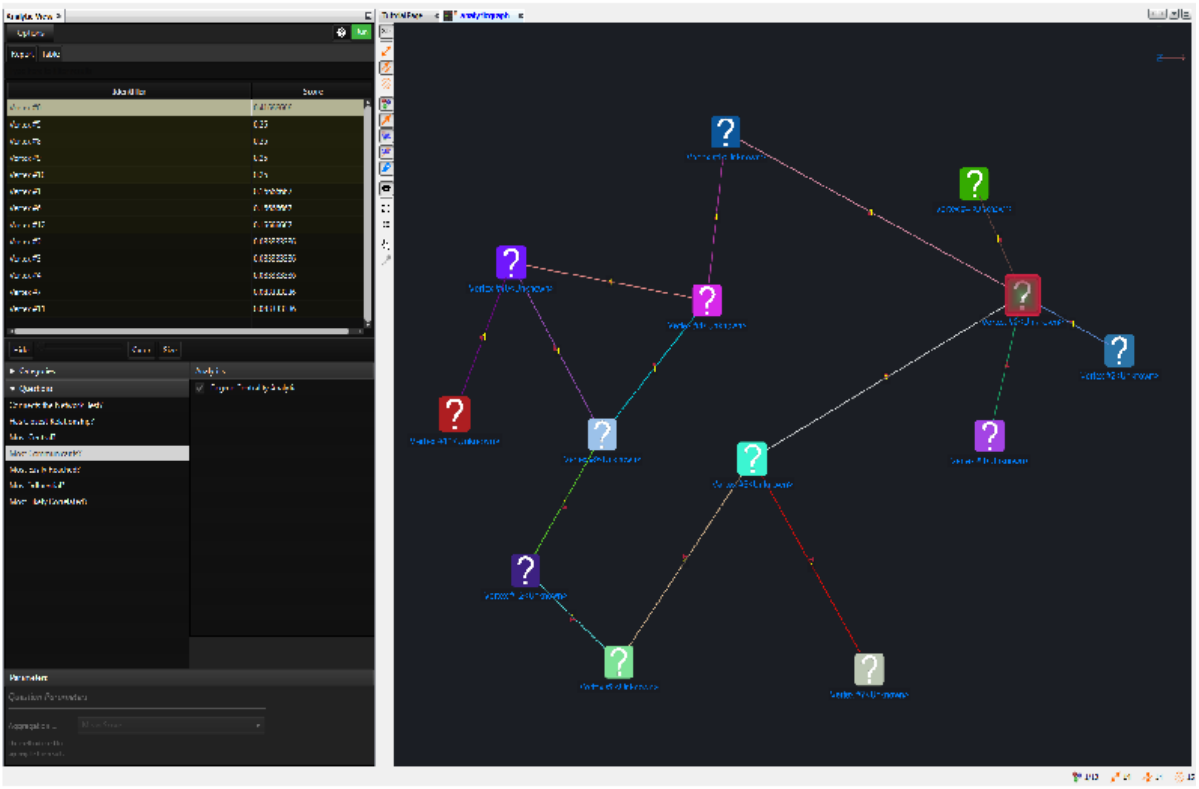


- Conversation View (Ctrl + Shift + C)
    - This view will visualise transaction Content data in a more interpretable fashion, as you might see conversations in a chat window.
- Scatter Plot View
    - This is useful for comparing attribute values.

As well as visualising, there are also views that allow you to ask complex questions of the data. For example, the Analytic View.

Analytic View (Ctrl + Shift + X)

The Analytic View is a tool for computing analytics across a graph in a user-friendly manner. This is done by posing complex algorithms as simple questions you might ask of the graph. These algorithms range in functionality from scoring based on the structure of a graph, to clustering and partitioning the graph, and even analysing content stored on a graph. Once an analytic is executed, results will be added to nodes and transactions as attributes.



- **Configuration:** This is the options toolbar, which allows you to configure the Analytic View, browse to this help document, or execute the currently selected analytic question.
- **Analytic Visualisation:** After a question is executed, the results of the chosen analytic question will appear. The visualisations available will differ based on the particular result type of the question executed.
- **Graph Visualisation:** After a question is executed, you will also be given a series of options allowing you to change the graph visualisation based on the result set. These options will also differ depending on the type of result.
- **Question Definition:** This is where you can construct an analytic question to ask of the graph.
    - You can choose to run a pre-defined question from the 'Questions' list, The 'Analytics' list will be populated with the analytics specified by that question, and each analytic will have pre-defined parameters which can be viewed by clicking each analytic.
    - Alternatively, you can choose to define your own custom question by selecting a category under the 'Categories' list, and then selecting any number of the analytics which will populate the 'Analytics' list and configuring their individual parameters. A documentations tab is available to explain what each analytic does.
- **Analytic Parameters:** This is where you can configure the parameters related to a question. If you are using a pre-defined question, the parameters cannot be changed but can still be viewed here. If you are building your own question, you have full control over these parameters. Selecting a question or category will populate the parameters pane with any parameters affecting the question as a whole. Selecting an individual analytic will additionally populate the parameters pane with any parameters specific to that analytic.

Finally

This guide is designed to help you get started finding your way around Constellation. Hopefully, you are now familiar with the graph concepts behind Constellation, and are capable of importing, exploring, visualising and analysing your data.