Learning Module – Social Network Analysis and the Analytic View

In this module we will go through the basics of Social Network Analysis. We will cover:

- What is Social Network Analysis
- Using the Analytic View
- Example Metric: Eccentricity
- Example Metric: Connectivity Degree
- Centrality
- Using the Documentation

What is Social Network Analysis

Social network analysis answers the question: "What can the structure of a graph tell us about a network?"

It is the process of investigation social structures using network and graph theories.

- Entity resolution
- Behaviour analysis
- Link prediction

It helps us determine who an entity is and where they fit in the network.

Constellation provides selection shortcuts for a number of interesting graph structures.

- Singletons - nodes with no neighbours, completely disconnected. This might be useful in situations where you are importing results, or maybe filtering via the histogram view.
- Pendants - nodes with only one neighbour. Typically found on the outskirts of the network 0 if you have done a multi-hop query, it might indicate nodes we are less interested in.
- Loops - transactions where the source = the destination. It might be worth looking at the attributes - but think about the context. It could be erroneous.
- Sources - nodes with only outgoing transactions - a trait of spam/broadcasting nodes - look for sources with a high degree.
- Sinks - only have incoming transactions - identifies non-participants or network exits, sometimes can be an indicator of hierarchy.
- Backbone - the inverse of pendants - identifies more highly connected nodes. The backbone is all nodes with more than one neighbour and the transactions between them.
- Shortest paths - Paths are a crucial structural element of a graph. Why? Because there isn't always going to be a direct connection between each entity - maybe show a regular graph to demonstrate what that looks like - so how do we track how information flows from one node to

another? It is most likely to follow the shortest path - minimum number of hops - let's face it, laziness rules.

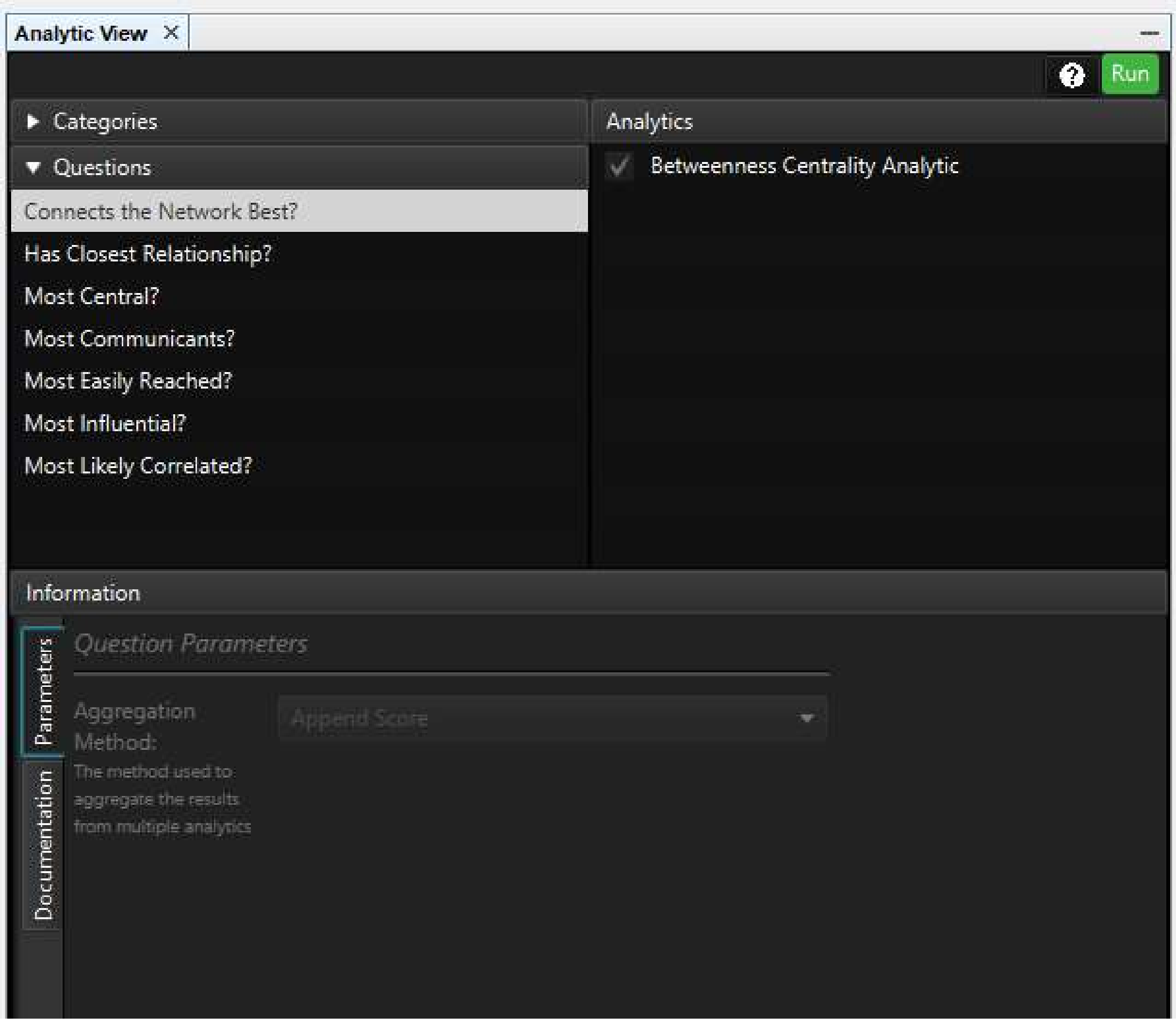
There are two types of SNA - egocentric and sociocentric.

- Egocentric - Personal networks; focus is on the individual and their connections
- Sociocentric - Whole networks; looking at patterns across the entire network

Examples of ego-centric social network analysis include looking at graph properties in the Histogram View. Neighbour Count can indicate interconnectivity, popularity, a whole range of things. Transaction Count tells us how active someone is, or how many observed events are in the graph for an individual. Between neighbours, transaction count can also the indicate importance of the relationship – this is sometimes referred to as the Weight of the relationship. There are other factors which also help determine this. For example: Multiplexity and Reciprocity.

- Multiplexity is the number of types of relationships between entities - the number of observed means of interaction. In SNA, the more ways two people keep in touch, the stronger the bond. Facebook, twitter, phone, email, gaming, in person - all these interactive means denotes a stronger bond than just one gaming or social media platform.
- Reciprocity - the extent to which a relationship is one sided or two-sided. One sided relationships are weak. Two sided relationships are strong. Calculate reciprocity by measuring the ratio of outgoing to incoming transactions.

To calculate metrics such as Multiplexity or Reciprocity, you could either do it by hand for each link on the graph, or you can just use the Analytic View in Constellation.

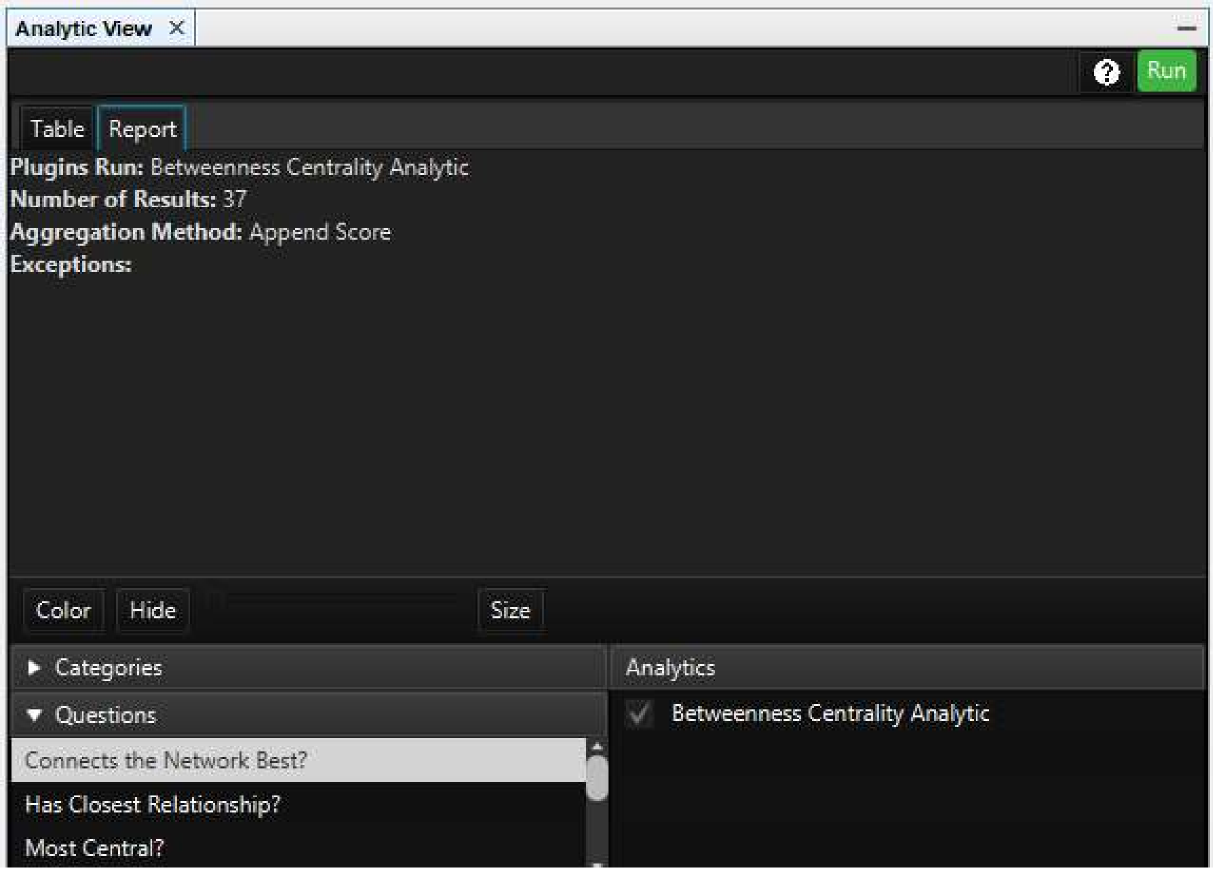


Using the Analytic View

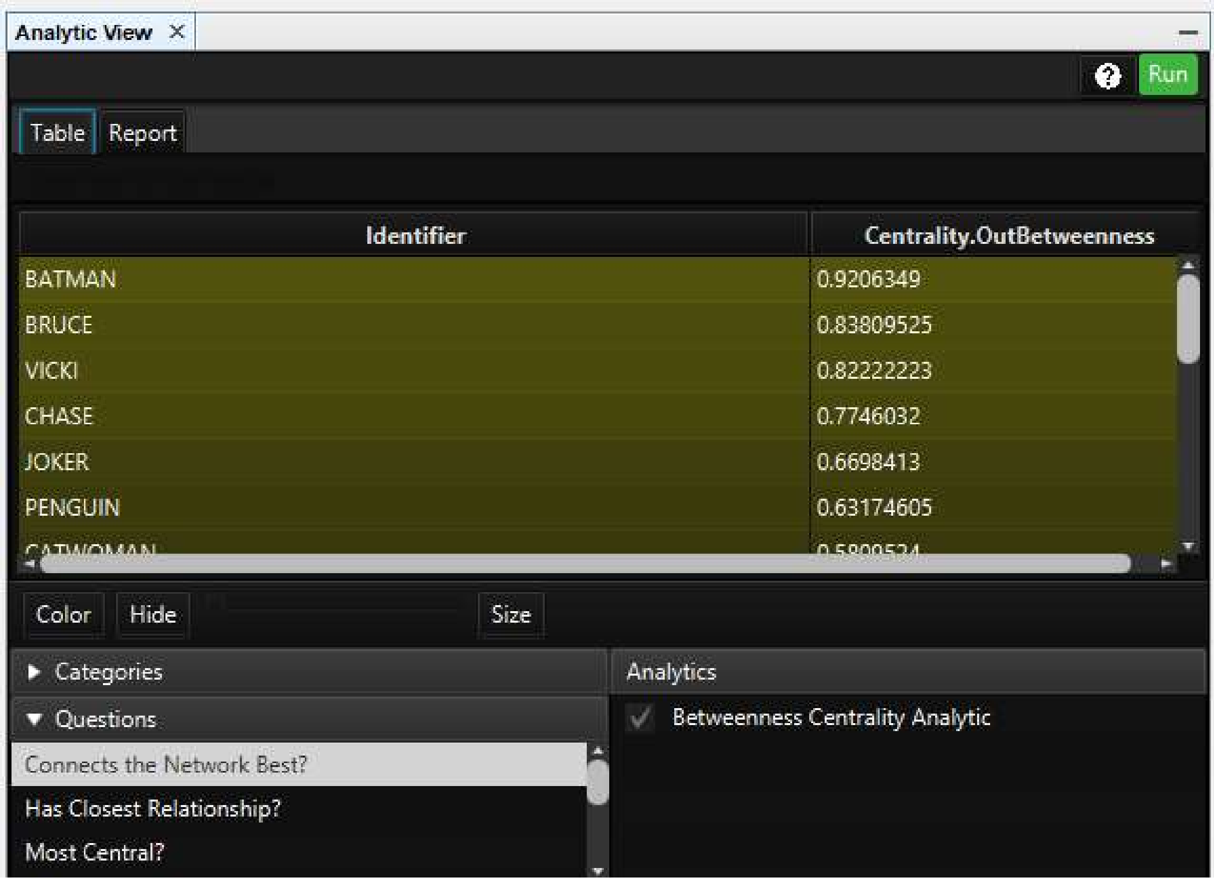
The Analytic View is an interface for running different analytics over the graph to answer a variety of questions. Nodes and transactions are scored according to how highly the measure against the analytic. Attributes are added to the graph and can be viewed using the Attribute Editor/Histogram View.

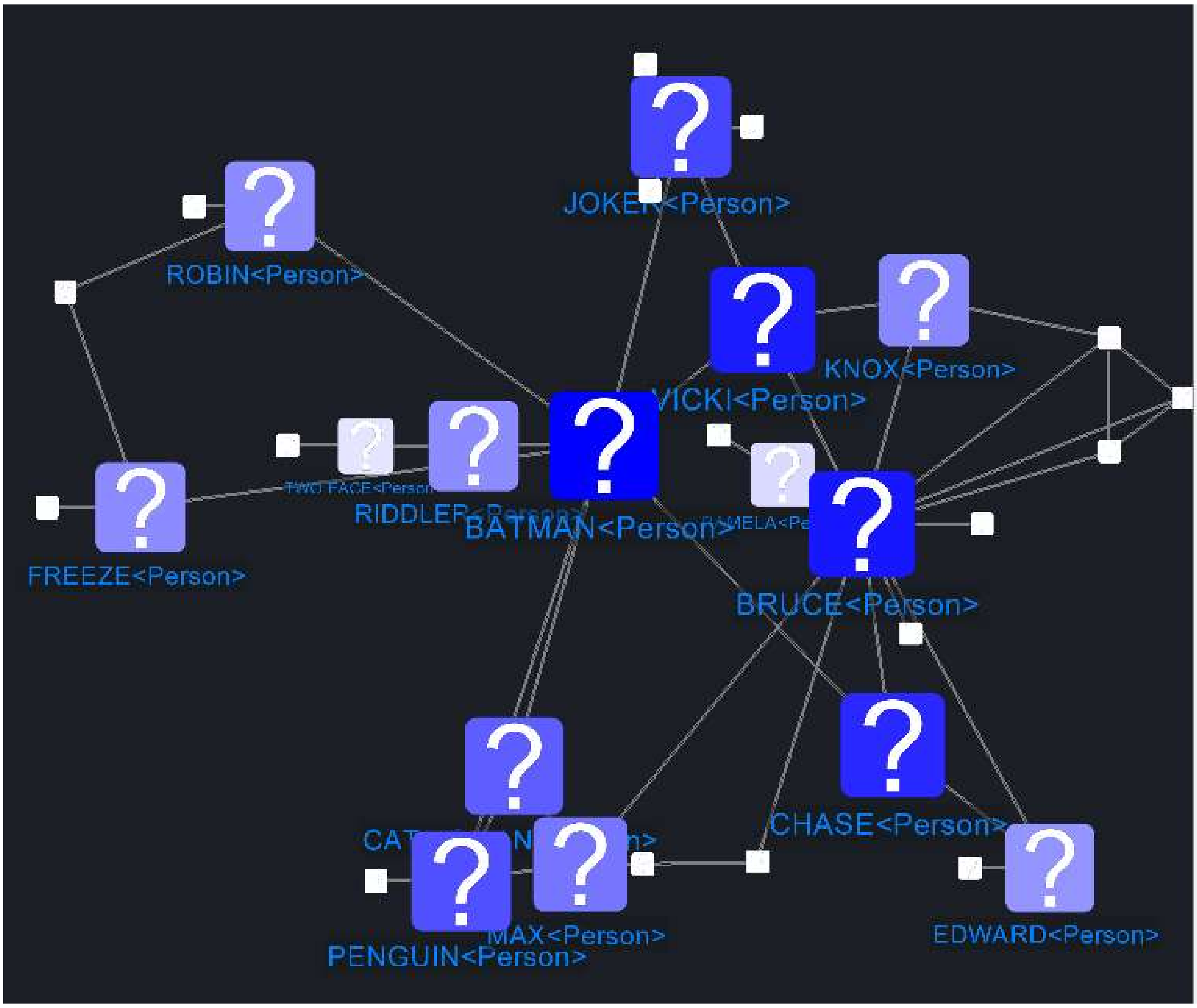
The Questions Tab holds a list of prepared questions analysts are likely to ask. Parameters are fixed. It is more accessible this way to those people who do not have a background in graph theory, and it lists the analytics used to answer the questions on the right.

Once you run an analytic, it will produce a Report visualisation, indicating the plugins that ran, the number of results, and any exceptions or errors that occurred.

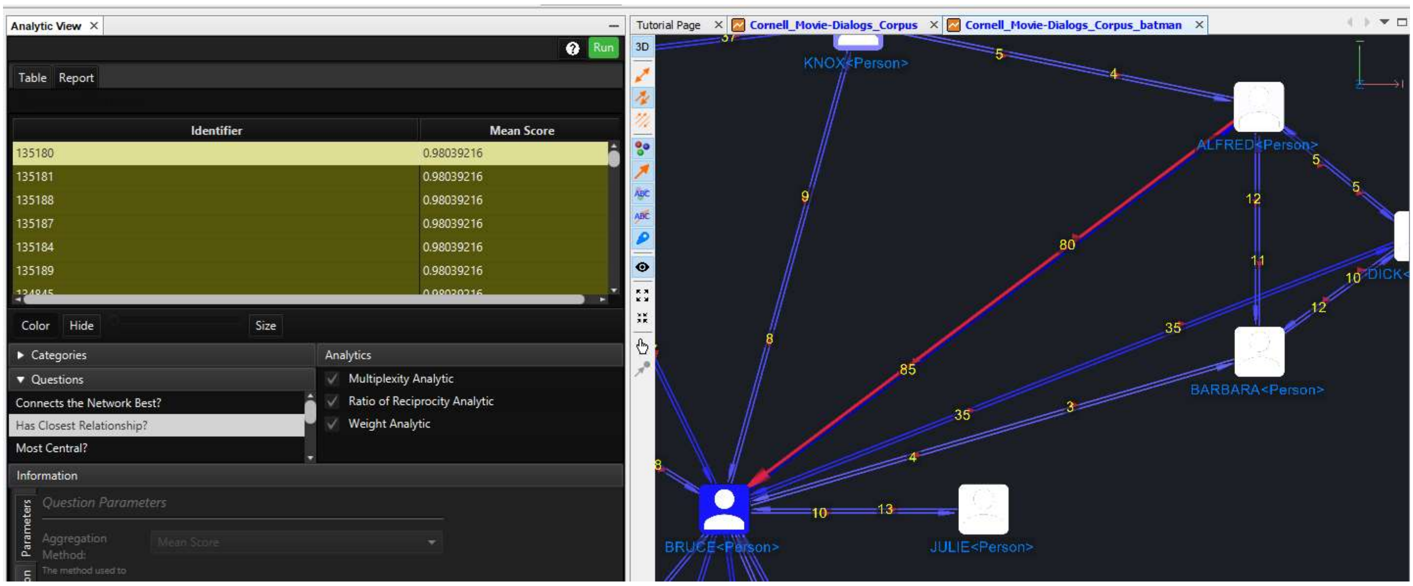


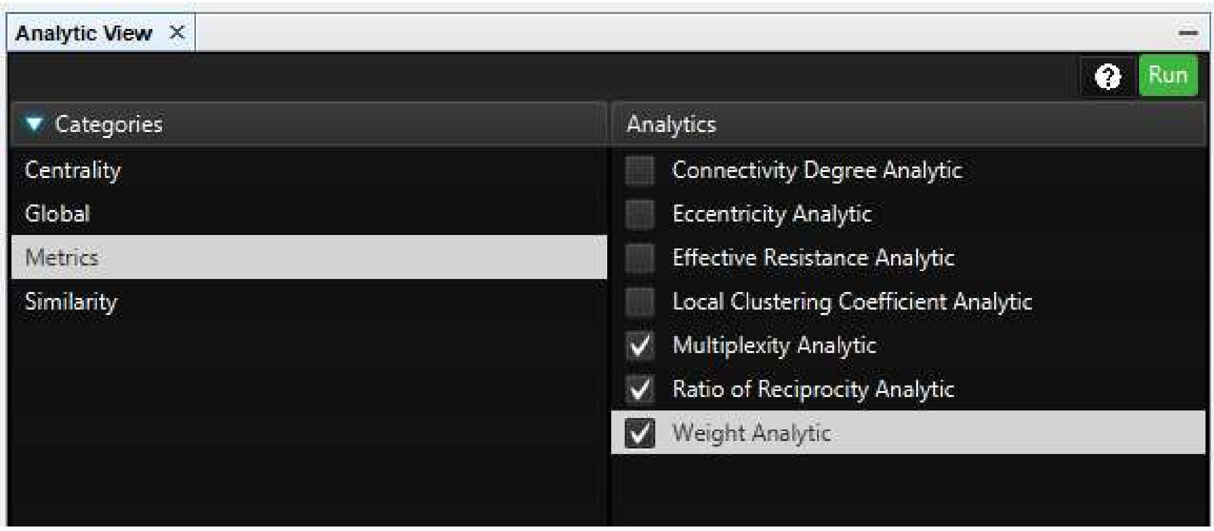
The Table visualisation lists nodes in descending order of Score - rows selected here will select nodes on the graph - the same level of interactivity you've come to know and love. You can also color and size nodes based on their score.


Analytic View
Run
Table
Report
Identifier
Centrality.OutBetweenness
BATMAN
0.9206349
BRUCE
0.83809525
VICKI
0.82222223
CHASE
0.7746032
JOKER
0.6698413
PENGUIN
0.63174605
Color
Hide
Size
Categories
Questions
Connects the Network Best?
Has Closest Relationship?
Most Central?
Analytics
Betweenness Centrality Analytic



The Has Closest Relationship Question attempts to identify pairs of nodes that have a close bond, taking in 3 metrics: Multiplexity, Ratio of Reciprocity and Weight (each found individually under the metrics category).

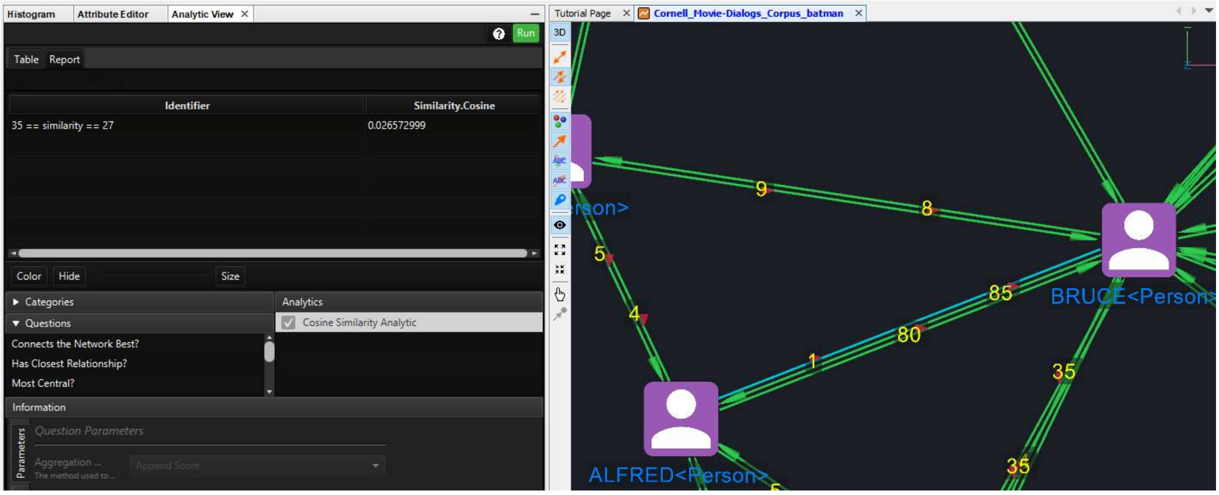




Speaking of categories - there are 4 - each one attempting to group analytics of a particular theme.

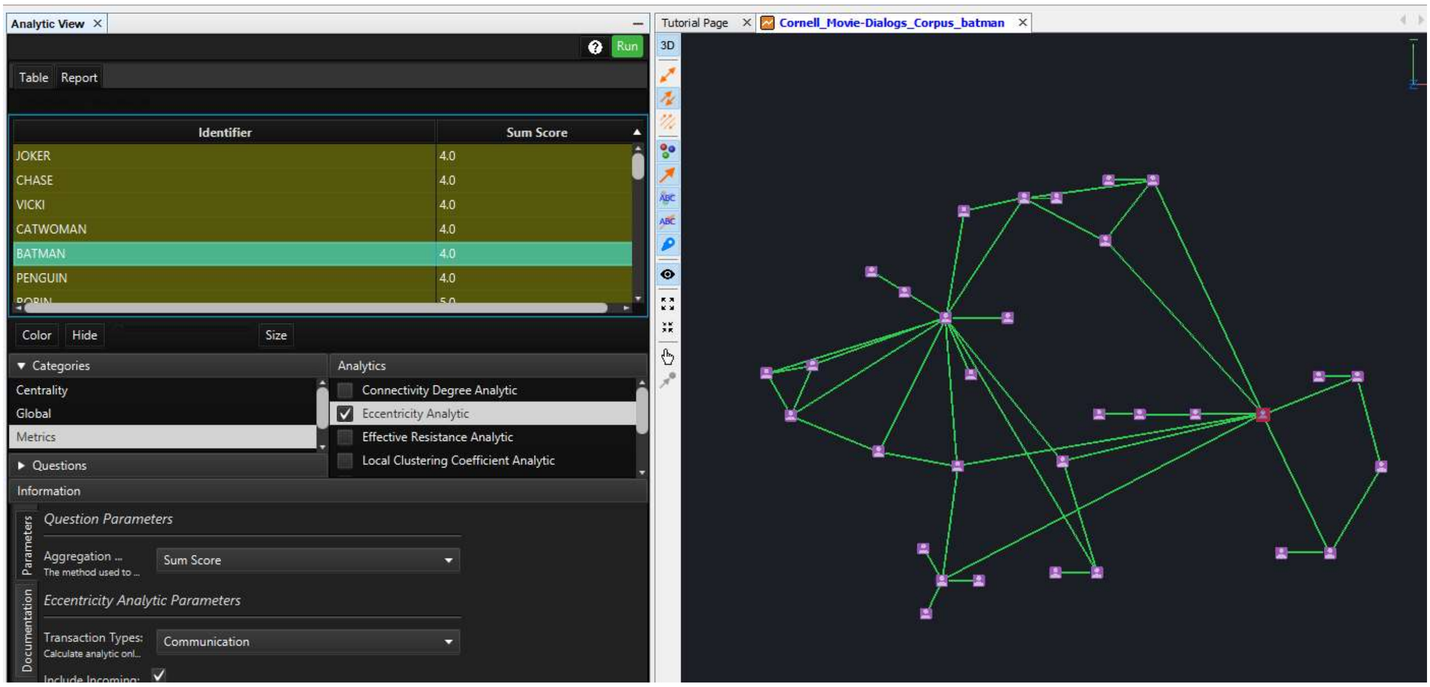
- Centrality - identifying entities that are "important" on the graph (entity resolution)
- Global - Analytics that describe the entire network (sociocentric social network analysis)
- Metrics - Describe nodes and links based on their structure
- Similarity - Connects nodes that exhibit similar features (link prediction)

Note that Similarity analytics behave slightly differently in that they do not just add attributes. As these are attempting to make predictions on where links might exist, Similarity transactions will be added to the graph with a score. Use the Histogram View to clean these up to find the most interesting, highest scoring links.



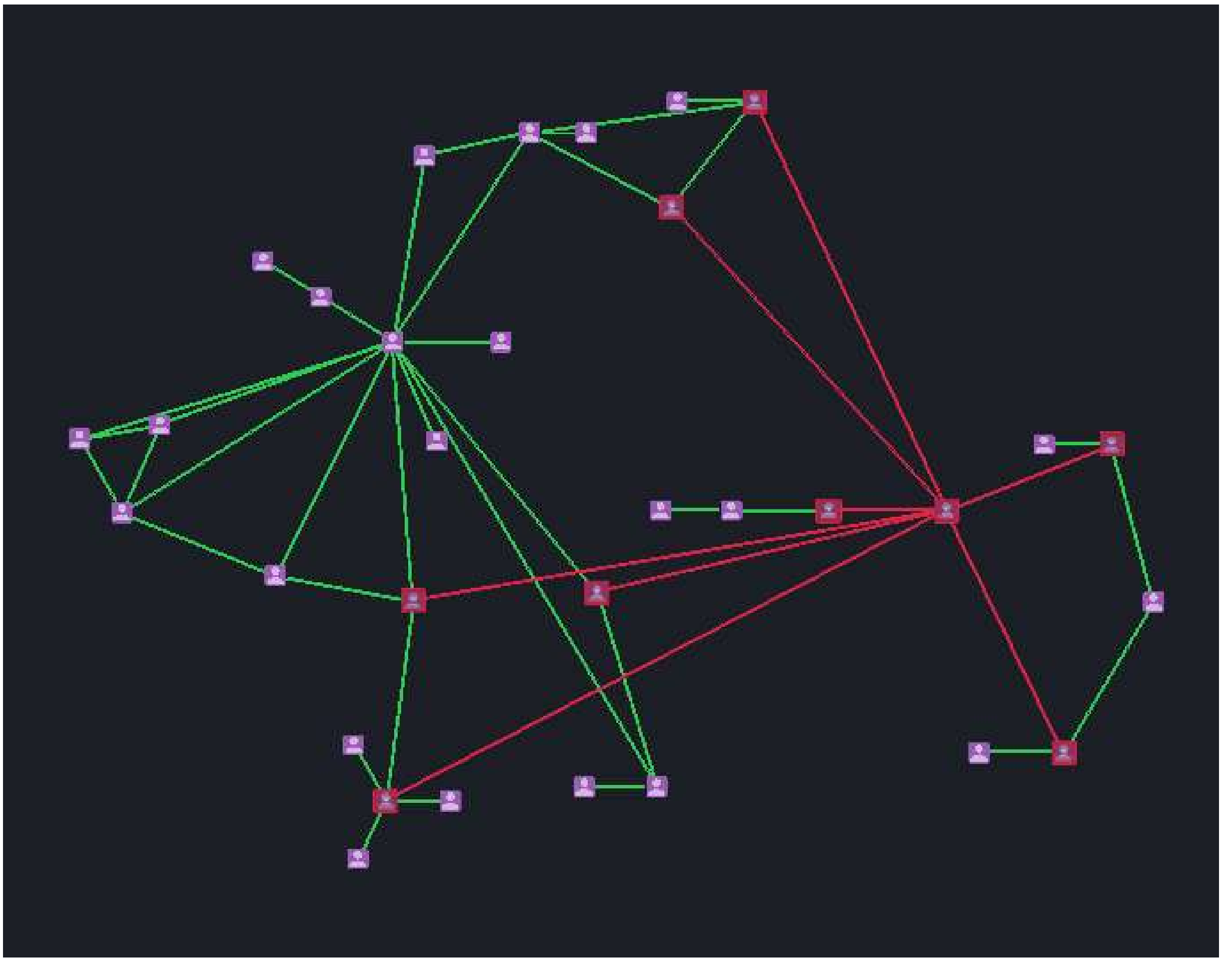
Example Metric: Eccentricity

Here is an example of a metric: Eccentricity. In graph theory, distance is measured by the number of hops taken to get from one point to another (shortest paths). The eccentricity of a node is the shortest distance to the furthest node. (The number of hops required to reach every node in the graph). Nodes with low eccentricity are positioned centrally in the network, nodes with high eccentricity are positioned on the outskirts of a network. Typically, if you are investigating a social network, the low eccentricity nodes are more interesting.

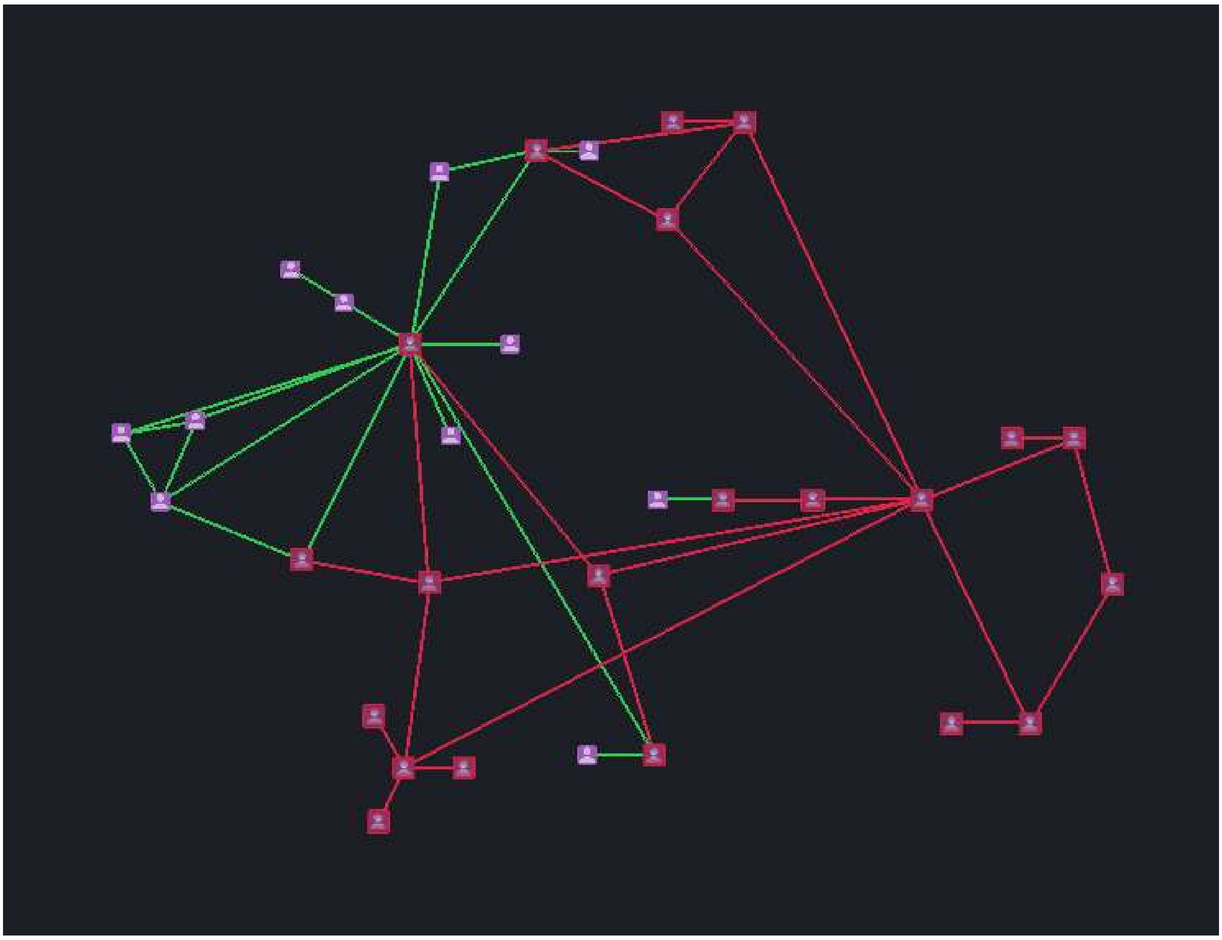


Here, Batman would reach everyone in the network in 4 hops.

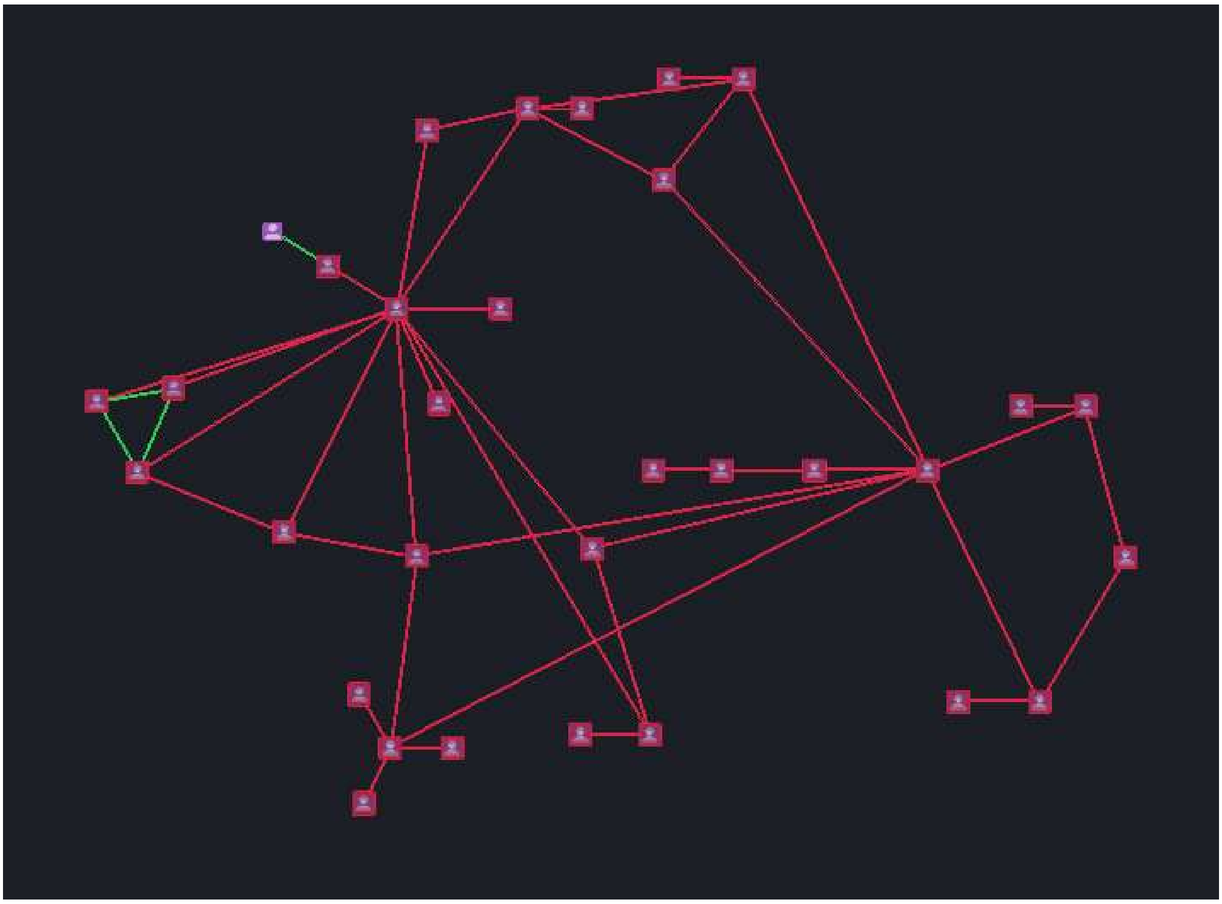
1 Hop



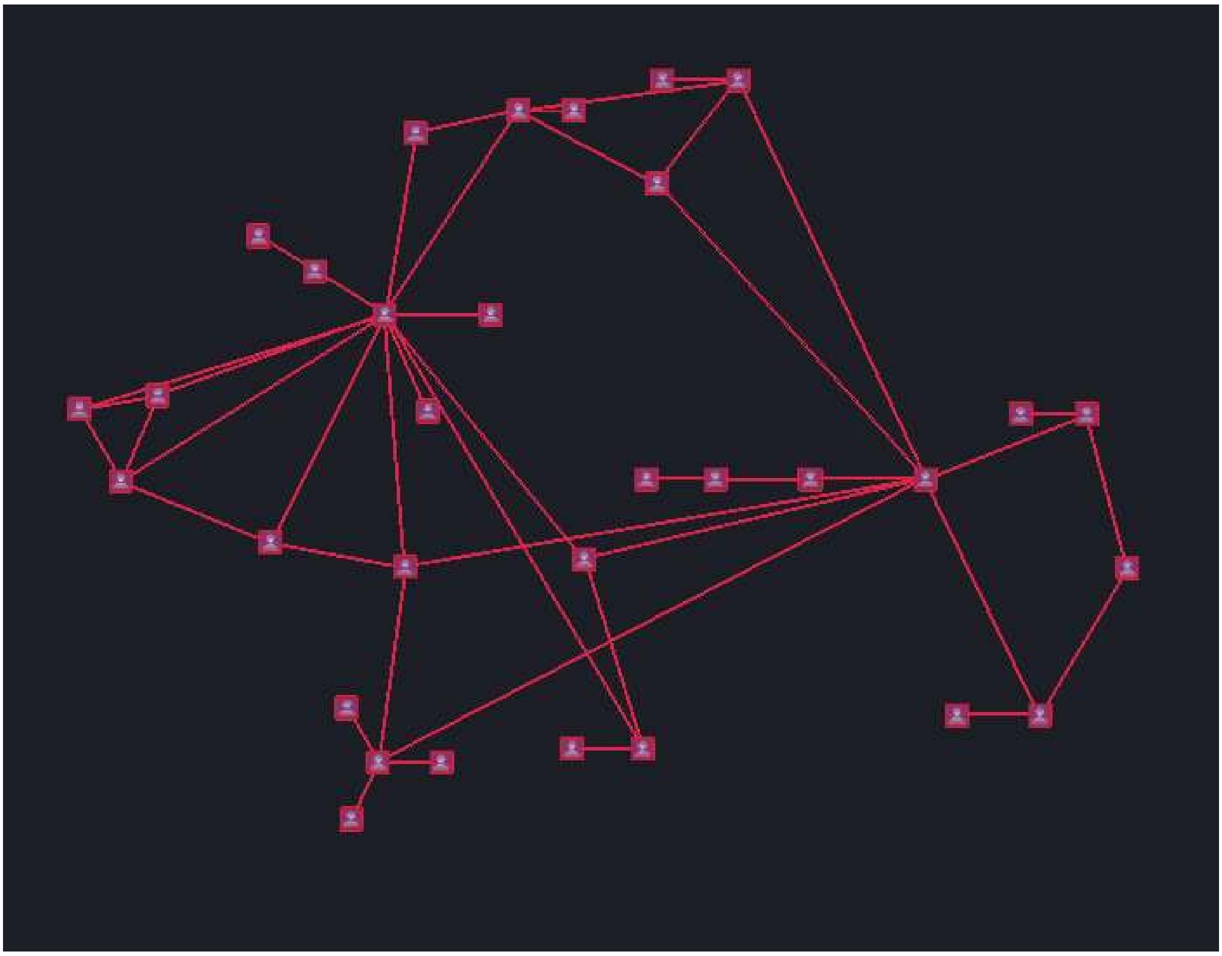
2 hops



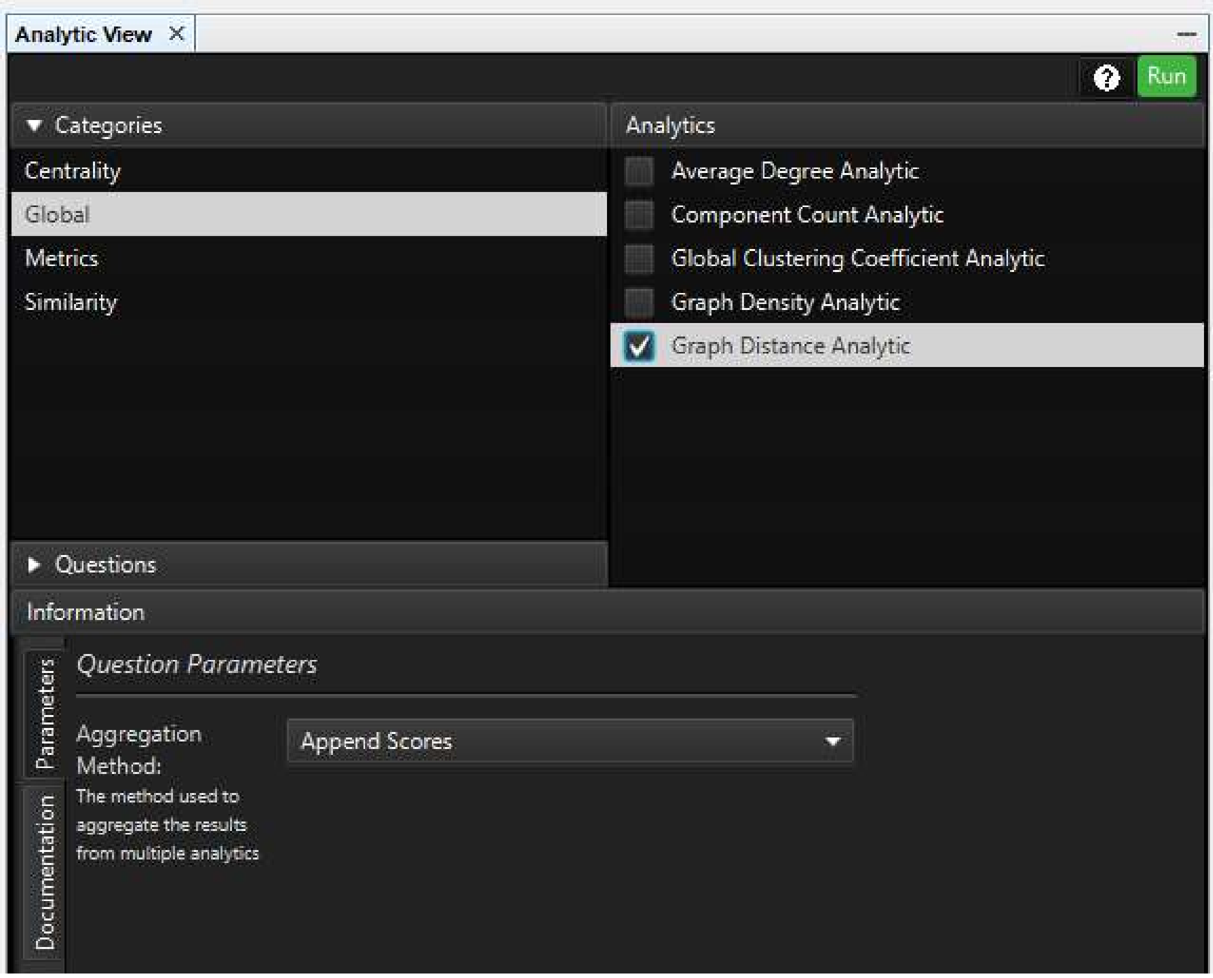
3 hops

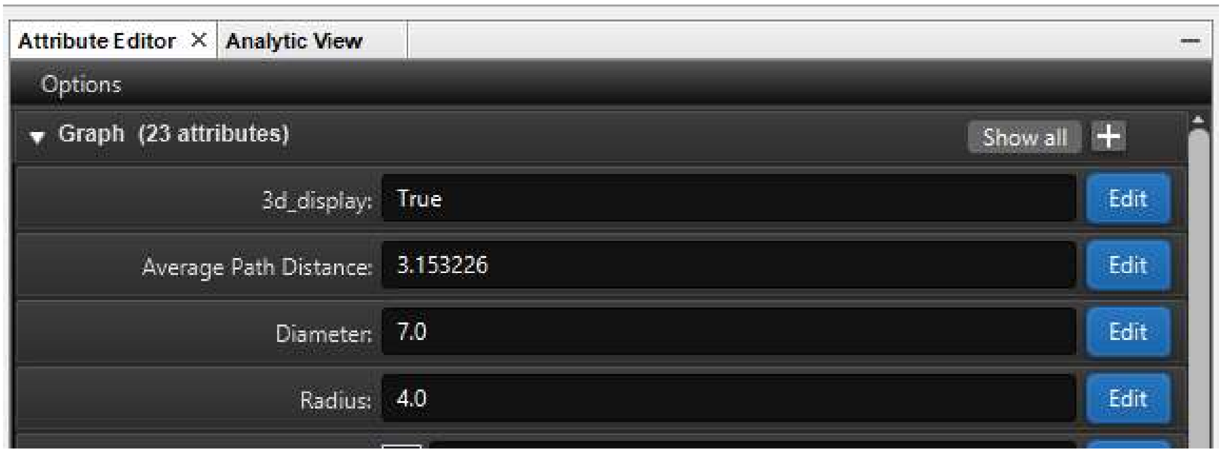


4 hops… everyone reached.



The radius of a graph is the lowest eccentricity, and the diameter of the graph is the highest eccentricity. These two attributes can be calculated using the graph distance analytic found in the global category, and the results/added attributes can be found under "Graph attributes" in the Attribute Editor. These global attributes can be useful when comparing different networks (sociocentric social network analysis) - networks with a low radius/diameter tend to be more cohesive - a high radius/diameter means that the network can probably be easily disrupted.

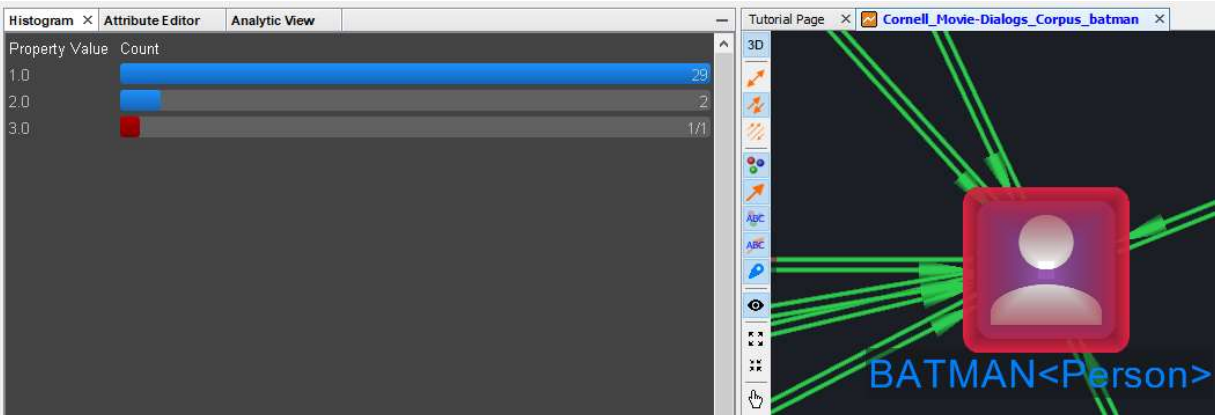




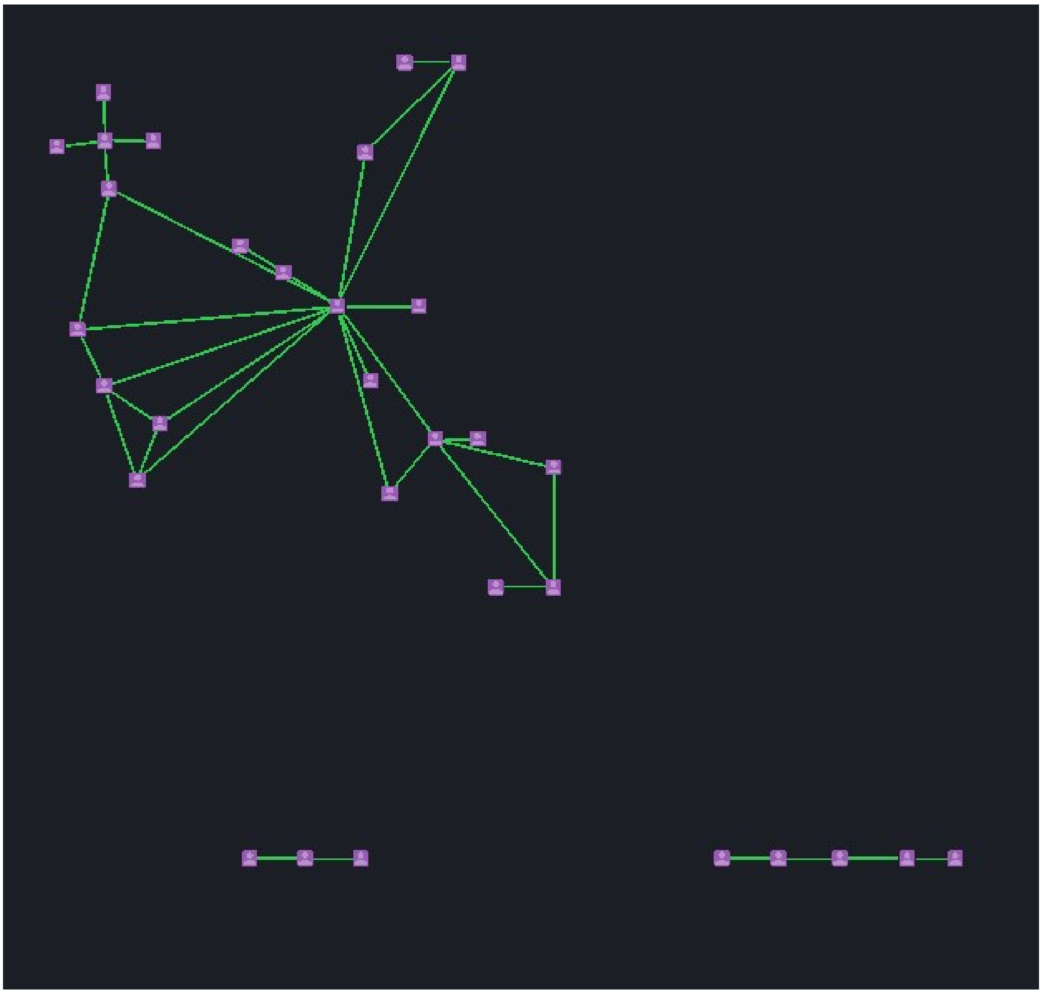
What do we mean by disrupted? Well, this brings us to our next graph metric: Connectivity Degree.

Example Metric: Connectivity Degree

In graph theory, a component is a group of nodes that are completely connected. There are some variations - for example, directed graphs can have strongly or weakly connected components, but the idea is that everything is connected. Some nodes are structurally important because they hold components together. Deleting the node results in a disconnected graph. The number of components that a node is holding together is referred to as its Connectivity Degree.



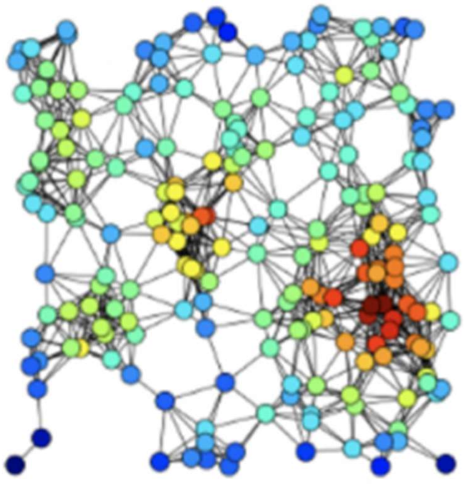
Here, Batman has a connectivity degree of 3. This node is connecting 3 components. If we delete Batman from the graph, we end up with:



Centrality

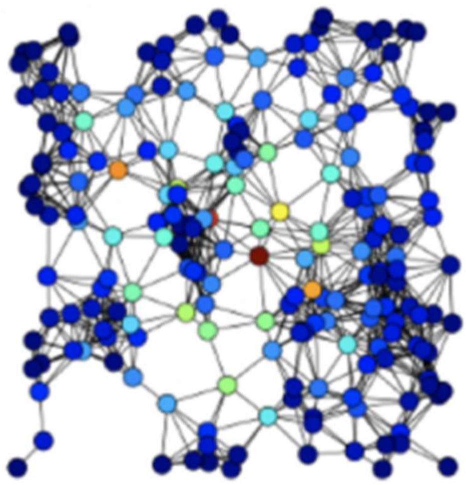
Centrality is a very important Social Network Analysis tool. It is a means of identifying which nodes are **important** in a network.

Because importance can be subjective, there are many ways to calculate centrality, depending on what you are looking for. Here are some examples of different centrality measures (the accompanying images have been borrowed from Wikipedia):



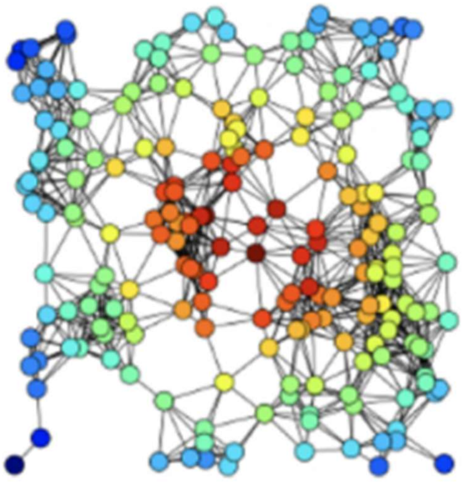
Degree Centrality

Degree Centrality — This is basically the **number of transactions** a node has (akin to looking at Transaction Count in the Histogram View). The logic behind this is that the more connections an entity has, the better. This is not always the case though - for example, nodes that are connected to lots of pendants that otherwise don't interact with the rest of the network, or nodes whose neighbours are already all connected - nodes that only connect those that are already connected. There are variations to Degree Centrality that incorporate direction – **in-degree**, which is the number of **incoming** transactions, and **out-degree**, which is the number of **outgoing** transactions.



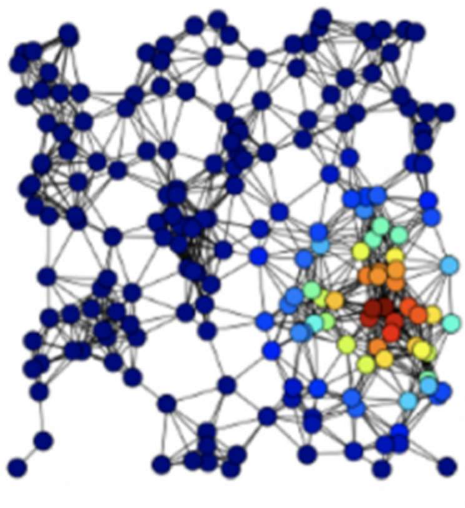
Betweenness Centrality

Betweenness Centrality — The betweenness centrality of a node is the number of times a node acts as a **bridge** along the shortest path between two other nodes. Nodes with high betweenness centrality can potentially have higher connectivity degree, as high betweenness centrality is a given when a node is holding the network together. But where a node isn't connecting network components, betweenness centrality will tell us if traffic or information is still likely to flow through that part of the network.



Closeness Centrality

Closeness Centrality — The closeness centrality of a node is the **average length of the shortest path** between that node and all other nodes in the graph. The more central a node is, the closer it is to all other nodes. This centrality measure has a close relationship with eccentricity, but the difference here is that closeness centrality will tell us which nodes fall closer to dense clusters within the network, something which eccentricity is completely insensitive to.



Eigenvector Centrality

Eigenvector Centrality – Eigenvector Centrality measures a node’s **influence**. It starts by measuring each nodes degree, and then goes a step further to account for the degree of the node’s neighbours, and so on through the network.

PageRank Centrality – This is very similar to Eigenvector Centrality, but the key difference is that it accounts for **link direction.** Each node in a network is assigned a score based on its number of incoming edges (its in-degree). These edges are also weighted depending on the relative score of its originating node. The result is that nodes with many incoming edges are influential, and nodes to which they are connected share some of that influence. This is the same algorithm that is used by the Google search engine when it comes to which websites are returned when you run a search.

Using the Documentation

If you click on the Documentation tab in the Analytic View, you will find descriptions on what each Analytic calculates.