
MCODE Documentation

Release 2.0.0

Gary Bader, Vuk Pavlovic, Christian Lopes

Nov 09, 2020

USER GUIDE:

1	Feature Requests and Reporting Bugs	3
2	Cite MCODE	5
2.1	Installation	5
2.2	Running MCODE	6
2.3	The User Interface	8
2.4	Getting and Interpreting Results	10
2.5	Fine-Tuning Your Analysis	13
2.6	Exploring Results Interactively	16
2.7	Outputting Results	19

MCODE is a [Cytoscape](#) app that finds clusters (highly interconnected regions) in a network. Clusters mean different things in different types of networks. For instance, clusters in a protein-protein interaction network are often protein complexes and parts of pathways, while clusters in a protein similarity network represent protein families. Cytoscape is freely available network visualization and analysis software.

MCODE is a relatively fast method of clustering. With an intuitive interface, it is suited for both computationally and biologically oriented researchers. Current features include:

- Fast network clustering
- Fine-tuning of results with numerous node-scoring and cluster-finding parameters
- Interactive cluster boundary and content exploration
- Multiple result set management
- Cluster sub-network creation and plain text export

Currently, MCODE does not provide any statistical score on the resulting clusters but can be used as a discovery tool in network analysis. Please see the [MCODE paper](#) for more information.

FEATURE REQUESTS AND REPORTING BUGS

The MCODE GitHub issue tracker can be used to report a bug or request a feature.

To Report a bug:

- Go to <https://github.com/BaderLab/MCODE/issues>
- Click on *New Issue*
- Write a short description of the issue. It is very helpful to provide a series of steps that can be taken to reproduce the issue.
- If possible attach a session file (.cys) or example input files.
- Enter App version, Cytoscape version and operating system.
- Click on *Submit new issue*

CITE MCODE

- **An automated method for finding molecular complexes in large protein interaction networks**

Bader GD, Hogue CW

BMC Bioinformatics. 2003 Jan 13;4:2.

PubMed Abstract.

2.1 Installation

1. Install Cytoscape

- Download and install the latest version of Cytoscape at cytoscape.org.

Note: Cytoscape minimum version **3.8** is required for this version of MCODE.

If you want to run MCODE on older versions of Cytoscape, the **App Manager** (see below) should show the latest compatible version of MCODE.

2. Install MCODE

2.1. Automated Installation (via the **App Manager**)

- Open Cytoscape.
- In the main menu select **Apps > App Manager**.
- In the App Manager select *MCODE* in the list of **all apps** and click the **Install** button.

Note: MCODE can also be installed from the [Cytoscape App Store](#).

2.2. Manual Installation

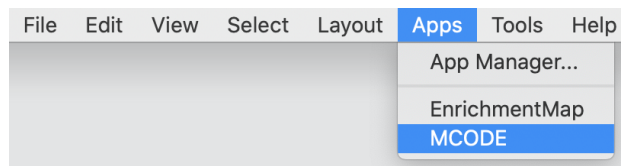
- Download the MCODE *jar* file from the [Cytoscape App Store](#), where you have two options:
 - a) Click the **Download** button on the MCODE app page to get the latest version.
 - b) Go to the **Release History** section and click on the specific version you want to download.
- Copy the downloaded *jar* file (e.g. `mcode-app-2.0.0.jar`) to your `[...]/CytoscapeConfiguration/3/apps/installed` directory.
- Check that **MCODE** appears in the **Apps** menu of Cytoscape. It may take a few seconds for the app to load up and appear in the menu. If it still does not appear, the most common causes are:
 - a) You placed the MCODE *jar* file in the wrong directory.

- b) You may have one or more newer versions of MCODE already installed, in which case you just need to delete all the other `mcode-[...] .jar` files you find in the `installed` directory.
- c) The downloaded version is not compatible with the running version of Cytoscape.

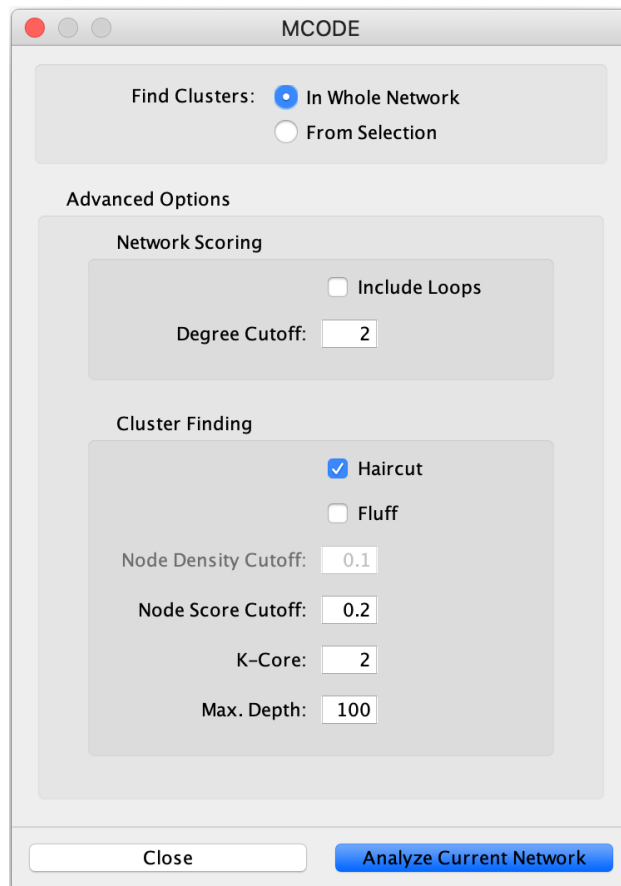
2.2 Running MCODE

MCODE is an extension for Cytoscape and can only be accessed through Cytoscape:

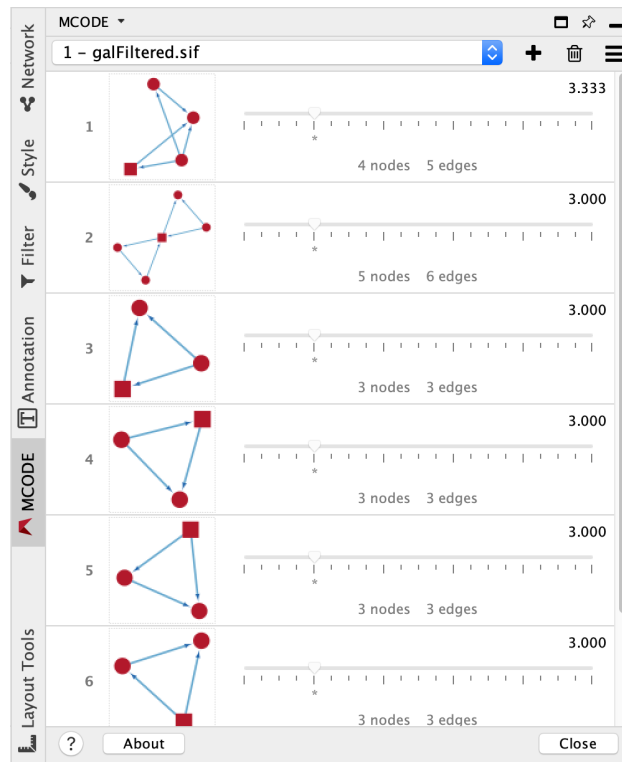
1. Start Cytoscape.
2. Import a network or open a Cytoscape session that has one or more networks.
3. Make sure a network is selected in Cytoscape's.
4. Select the menu option **Apps > MCODE**, which should open MCODE's *New Analysis Dialog*.



5. Click the **Analyze Current Network** button.

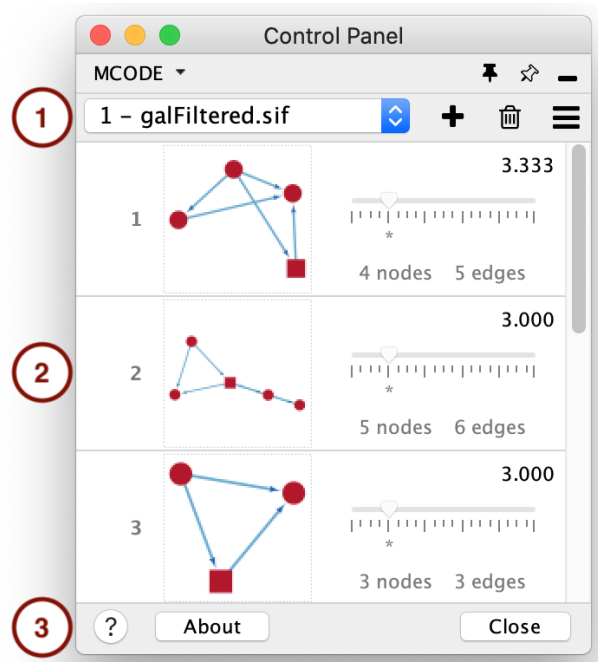


6. The main MCODE interface will appear with the results as a tab in the left-hand panel of Cytoscape.



2.3 The User Interface

2.3.1 MCODE Panel



The MCODE Panel is added to Cytoscape's **Control Panel** whenever the MCODE app is opened (menu option **Apps > MCODE**), and it contains three main sections:

1. Toolbar
2. Cluster Browser
3. Footer

Note: The first time MCODE is opened after installing the app or starting Cytoscape, the *MCODE Panel* may not appear right away, because the *New Analysis Dialog* will pop up automatically instead. This happens only if the current Cytoscape session has at least one network and it is selected. Either way, the *MCODE Panel* will show up right after the *New Analysis Dialog* is closed.

Toolbar

Use the toolbar components to add and manage your analysis results:



1. Results:

Selector that lists all your MCODE results and allows you to switch between them.

2. New Analysis:

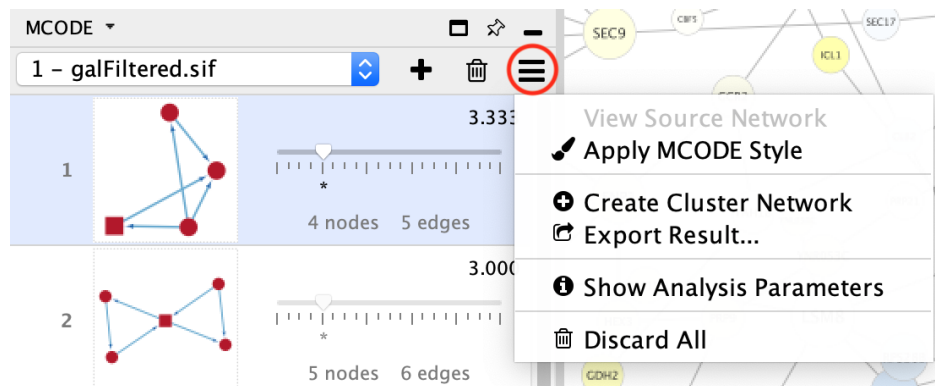
Click this button whenever you want to perform a new analysis. It is disabled when no network is selected.

3. Discard Selected Result:

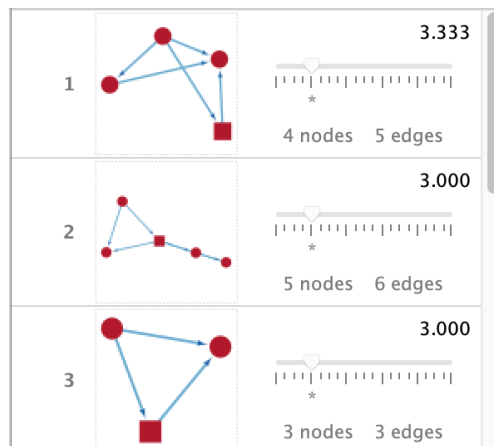
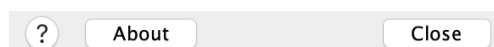
Click this button to permanently remove the selected result.

4. Options:

Menu that provides additional options, such as exporting the results and creating a network from the selected cluster.

**Cluster Browser**

This section lists the clusters which were found for the selected result.

**Footer**

1. **Online Help:** Opens this documentation.
2. **About:** Opens MCODE's *About Dialog*.
3. **Close:** Disposes all the MCODE results and closes the *MCODE Panel*.

2.4 Getting and Interpreting Results

This section covers some of the basic steps of running MCODE on a network.

2.4.1 1. Load Your Network

To begin, load the network to be analyzed into Cytoscape. You can load as many networks as your computer system can handle, large or small.

MCODE will analyze the *current* network, which is usually the network that is selected in Cytoscape's **Network** tab (left-hand panel) or the network *View* that is currently rendered.

2.4.2 2. Choose The Analysis Options

2.1. Go to the **MCODE Panel**.

2.2. Click the **New Analysis** button to open the *New Analysis Dialog*.

2.3. Choose The Scope:

Clusters can be found in the entire network or from a selection of nodes. Chose one of these options in the **Find Clusters** section of the *New Analysis Dialog*:

- **Find Clusters in Whole Network** - MCODE will find and report all clusters in the entire network.
- **Find Clusters from Selection** - Only those clusters which include the selected node(s) within them will be reported. - Selections can be made either directly in the view, the *Node Table* or Cytoscape's handy search tools.

Note: The choice of scope is dependent on your familiarity with the network in question and the desired outcome. If you have a particular protein of interest within a network, for example, it may be appropriate to search for only those clusters involving such a protein. On the other hand, exploratory work will benefit most from analyzing the whole network.

2.4.3 3. Analyze Your Network

3.1. Click the **Analyze Current Network** button:

If everything is OK, this will display a progress meter of the scoring, finding, and drawing steps, provided that the task is not too quick.

However, the analysis cannot be performed in some situations, where you may see these *Analysis Interrupted* messages:

a) *No Selection* Message

- This message appears when the *From Selection* scope is used without an actual node selection.
- You must select the desired node(s) and try again before MCODE can attempt to find clusters.

b) *Parameters Unchanged* Message

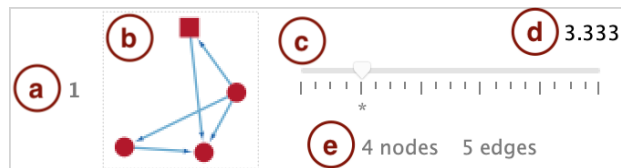
- Parameters are discussed in detail in the following section. For now, you should know that if you attempt to analyze a network twice without changing any of the settings, such as the scope of the analysis, MCODE will let you know that this analysis was previously conducted.

2.4.4 4. Browse Your Results

If the analysis finds clusters, the result will be added to MCODE's main panel as `result_number - network_name` (e.g. "1 - galFiltered.sif").

4.1 Cluster Browser

- The *Cluster Browser* entries are ordered from highest to lowest cluster's computed *score*:



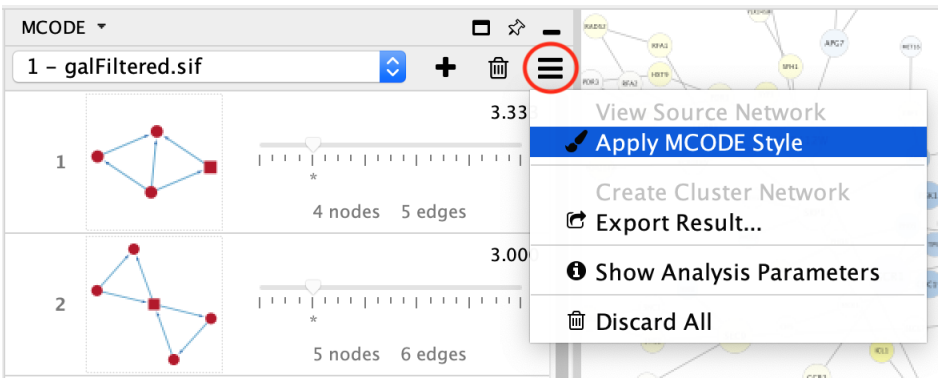
- Rank:** Just a number based on the cluster's computed score, used to identify the clusters within each result.
 - For example, *Cluster 1* is the highest ranked cluster in a given result, and thus, at the top of the list.
- Cluster:** The graphical representation of the cluster.
 - Cluster members are coloured red.
 - The highest scoring node in the cluster is called the **Seed**. It is the node from which the cluster was derived and is represented as a square.
 - Other cluster members are circles.
 - Edges, representing interactions for example, are blue.
 - Edge directionality is represented by arrows.
- Size Threshold:** Slider that allows you to expand or reduce the cluster based on the node score.
- Score:** The cluster's computed score.
- Nodes and Edges:** A simple enumeration of the cluster's members and their interconnections.

4.2. Network View

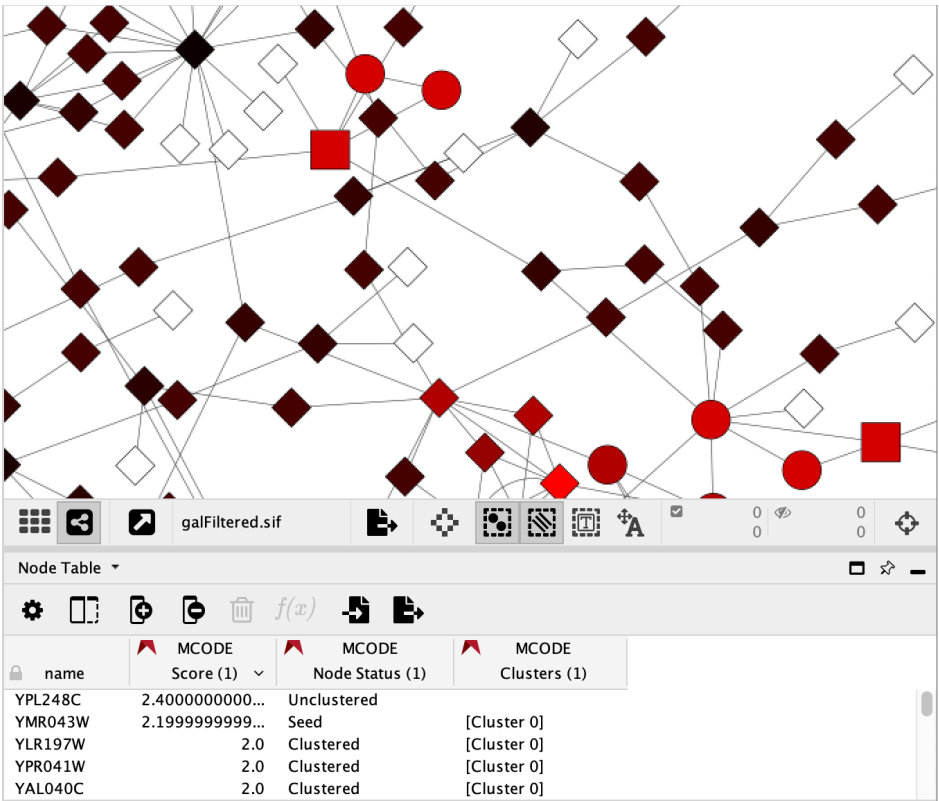
If the network being analyzed has a view, you can apply the *MCODE Style* by clicking MCODE's **Options** button and selecting the corresponding menu item.

This style utilizes two MCODE generated *Node* columns:

- Node Status:** Node *shapes* indicate the cluster status of the nodes.
 - Square*: seed (highest scoring node in the cluster)
 - Circle*: clustered
 - Diamond*: unclustered
- Score:** Node *colors* represent the node score.



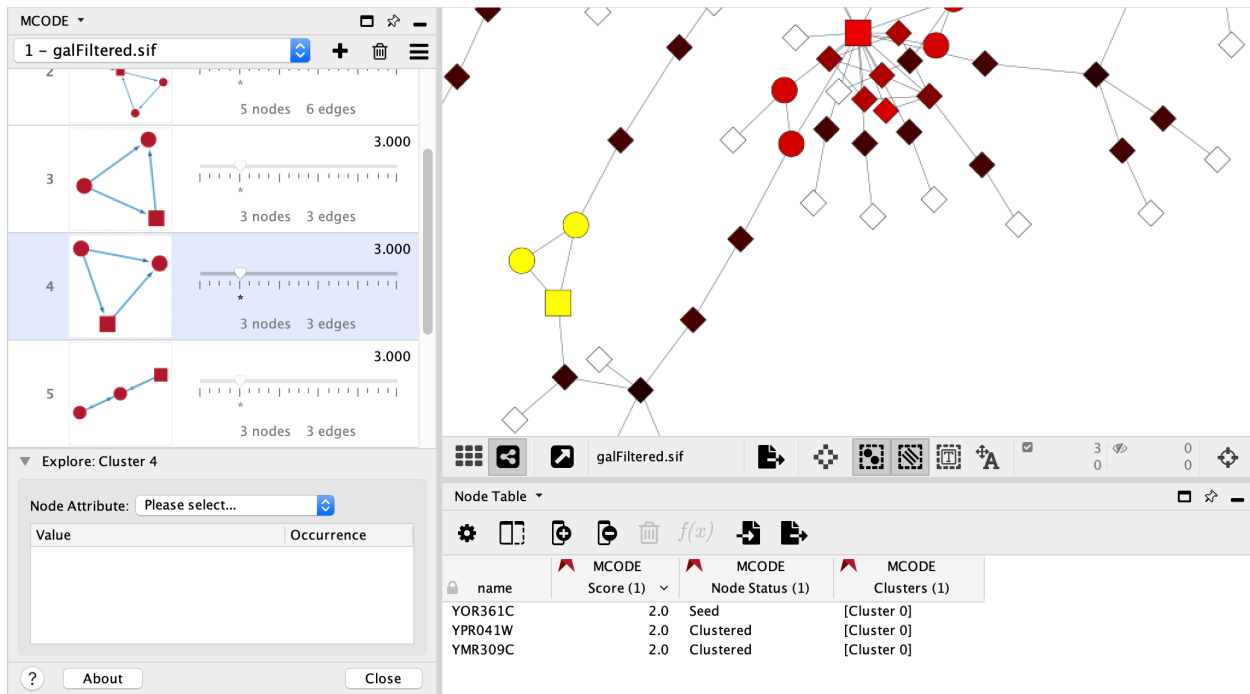
- A range from *black* to *red* indicates the MCODE computed node score (lowest to highest, respectively).
- *White* indicates a score of zero.



Note: The **Clusters** column is an additional list type attribute that indicates which cluster the node belongs to. The *MCODE Style* does not use this column, but it exists should you need it.

4.3. Cluster Selection

- The cluster entries in the *Cluster Browser* are selectable and will automatically select the corresponding nodes in the network view (if it exists). If no network view is available, the selected nodes can be reviewed in Cytoscape’s **Node Table**.
- Also, the cluster exploration panel will appear below the cluster entries titled “Explore: Cluster [Rank]”. This panel can be collapsed for now – its use will be discussed in the *Fine-Tuning Your Analysis* section of this Manual.

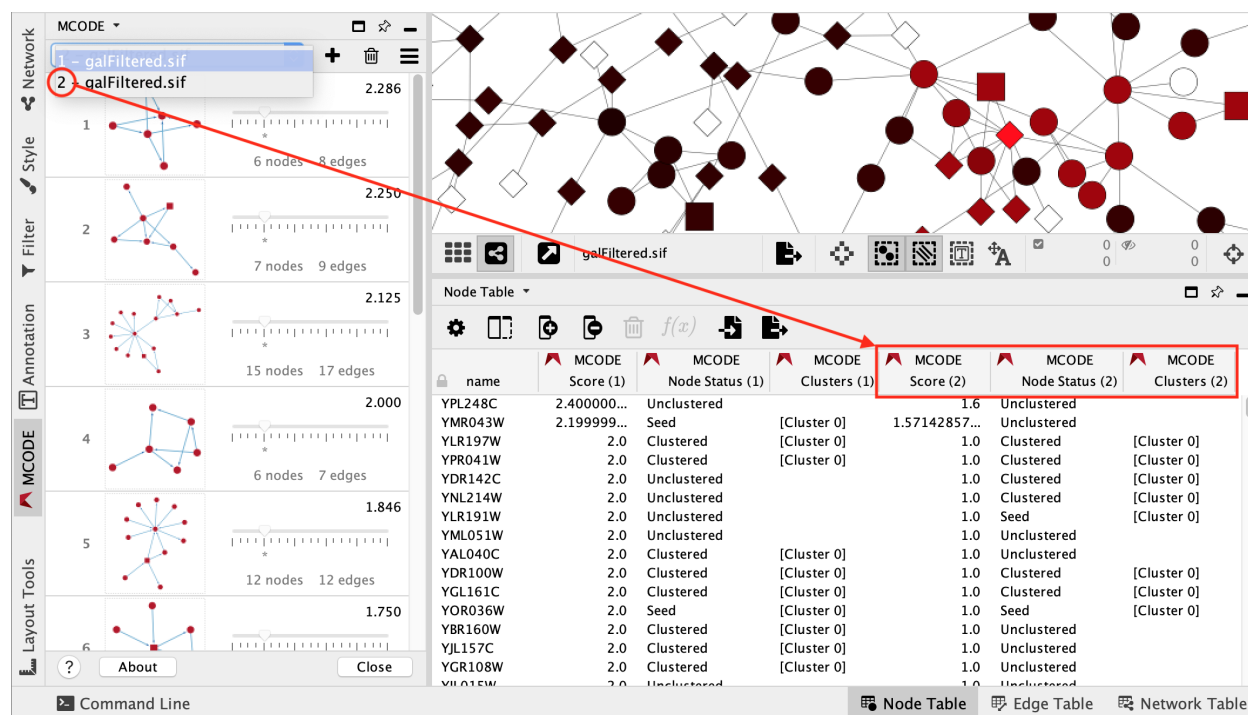


2.5 Fine-Tuning Your Analysis

By default, MCODE analyzes networks using scoring and finding parameters that have been optimized to produce the best results for the average user and network. However, you may achieve better results for your network by familiarizing yourself with these parameters and changing them appropriately. Sometimes even slight customizations can produce considerable differences, reduce unwanted or false results, and increase relevance of results. This is only an overview – for detailed parameter information, consult the [MCODE paper](#).

- The user can analyze a network as many times as desired with varying parameters.
- Each result is reported sequentially for reference and comparison.
- Each result creates a set of 3 “MCODE” *Node* columns, where *n* is the *result number*:
 - Score (*n*)
 - Node Status (*n*)
 - Clusters (*n*)

Note: For efficiency purposes, MCODE automatically determines which portion of the algorithm needs to be run based on the user’s parameter modifications.



- For instance, if the scoring parameters are altered, the network will be re-scored, but if only the cluster finding parameters are altered, only the cluster finding portion will be run.

These analyzes parameters are found in the *New Analysis Dialog*, as mentioned in the section [Running MCODE](#).

2.5.1 Network Scoring Parameters

Network Scoring

☒ Include Loops

Degree Cutoff:

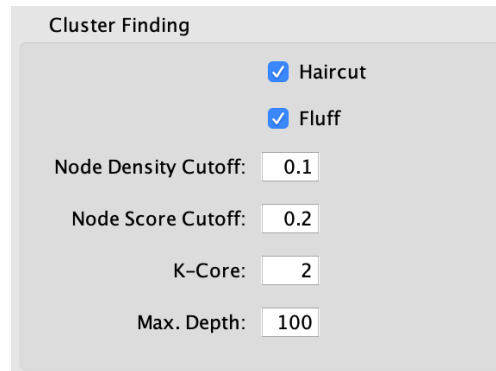
1. Include Loops

- When turned on, MCODE will include loops (self-edges) in the neighbourhood density calculation. This is expected to make a small difference in the results.

2. Degree Cutoff

- This value controls the minimum degree (number of connections) necessary in order for a node to be scored.
 - For example, nodes that share only one connection with one other node have a degree of 1.
 - Valid values are 2 or higher to prevent singly connected nodes from getting an artificially high node score.

2.5.2 Cluster Finding Parameters



The screenshot shows a 'Cluster Finding' dialog box with the following settings:

- ☒ Haircut
- ☒ Fluff
- Node Density Cutoff: 0.1
- Node Score Cutoff: 0.2
- K-Core: 2
- Max. Depth: 100

1. Haircut

- Once a cluster has been found, some nodes which may have satisfied the **Degree Cutoff** parameter during scoring may only be connected to the cluster by one edge. When haircut is turned on, MCODE removes all such singly-connected nodes from clusters. In some cases, though rare, the cluster's seed node may be only singly connected to the cluster and removed by the Haircut. For example, Cluster 2 of the *galFiltered.gml* network with default settings is one such case.

2. Fluff

- When turned on, MCODE expands cluster cores by one neighbour shell outwards, according to the fluff **Node Density Cutoff** parameter and after the optional haircut step.
- If this parameter is turned on, some nodes may belong to more than one cluster.

3. Node Density Cutoff

- Node density is calculated by dividing the node's connections by the maximum number of connections possible for that node. If Fluff is turned on, this parameter controls the neighbour inclusion criteria during 'fluffing'. Fluff expansion occurs after the cluster has already been defined by the algorithm and thus allows clusters to overlap at their edges. A higher value will expand clusters more.

4. Node Score Cutoff

- This is the most influential parameter for cluster size and is the basis for the **Size Threshold** slider in the *Getting and Interpreting Results* section.
- During cluster expansion, new members are added only if their node score deviates from the cluster's seed node's score by less than the set cutoff. This is a percentage, where a value of 0.2 allows for new members' node scores to be no more than 20% less than that of the seed node. Thus, smaller values create smaller clusters and vice versa.

5. K-Core

- This parameter filters out clusters that do not contain a maximally inter-connected sub-cluster of at least k degrees. For example, a triangle (3 nodes, 3 edges) is a 2-core (2 connections per node). Two nodes with 2 edges between them satisfy the 2-core rule as well. Since the default value is 2, this ensures that clusters must in the very least contain one of these two sub-clusters. Increasing this value will exclude smaller clusters.

6. Max. Depth

- Maximum depth limits the distance from the seed node within which MCODE can search for cluster members. By default this is set to an arbitrarily large number so that clusters are virtually unlimited. To limit cluster size, set this parameter to a small number.

2.6 Exploring Results Interactively

In addition to fine-tuning a multitude of parameters to enhance the analysis process, MCODE provides a real-time cluster exploration feature. This can be divided into two components: exploring cluster boundaries and exploring cluster content.

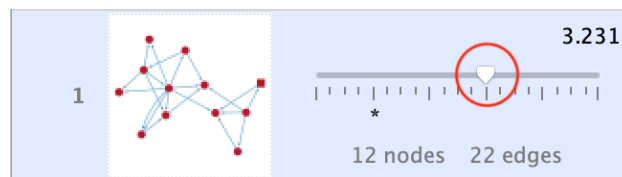
- The first exploration allows you to expand or reduce the cluster based on the node score using the **Size Threshold** slider.
- The second is the **Node Attribute Enumerator**, which provides a summary of the cluster's node attributes and their frequency in the cluster.

Together they can inform the user about the cluster's "natural" boundaries in the context of the network and ensure functional consistency.

These are both explained in greater detail below.

2.6.1 Size Threshold Slider

The slider scale ranges from **Min** to **Max** and has an 'origin' marker (^) for its starting position. **Node Score Cutoff**, which is the most influential cluster size determinant is controlled by the slider. As such, the initial position marker indicates the **Node Score Cutoff** value originally set in the *Cluster Finding Parameters* section.



When moving the slider, the **Node Score Cutoff** is set to 0 at Min and 100 at Max, however there are several notable differences between the functions of the **Size Threshold** slider and the **Node Score Cutoff** parameter (*Cluster Finding*):

1. During exploration, the cluster is reevaluated without the requirements of satisfying the **K-Core** parameter. Thus, moving the slider leftwards of the initial position allows the cluster to be reduced to only the seed node.
2. During exploration in the Max direction, the cluster is 'unaware' of other clusters. Unlike in the analysis where every subsequent attempt at finding a cluster is only allowed to expand around previously found clusters, the slider expands the cluster despite adjacent cluster borders. Thus, moving the slider rightwards of the initial position allows the cluster to be expanded to as much as the whole network.
 - However, the 'awareness' of other clusters is intact in the range between the 'origin' marker and Min to allow the cluster to return to its original size.

Haircut and **Fluff** are applied after slider movement if they were turned on in the original production of the result that is being explored.

In response to the slider, the **Cluster Browser** is updated with the new cluster's network graphic and details (number of nodes and edges and new cluster score).

- The node selection in the main network view will also be updated.
- Since clusters can expand to large and sometimes unreasonable sizes, the layouter may need extra time to complete its task. When this occurs, a loader and progress bar will appear in the **Cluster Browser**.
- There is no need to wait for the cluster to be drawn, the details and node selections will remain responsive to the slider's movements.

- If the new cluster exceeds 300 nodes, a place holder (“Too big to show”) will be drawn instead since the graphic representation will take too long to compute and will likely be too crowded to be of any real value.

Several peculiarities may arise during size exploration:

1. Cluster Size Explosion

- When exploring a lower ranked cluster (further down the list) the cluster’s size may depend heavily on nearby higher ranked clusters. This may not always occur since the finding process starts at the highest scoring nodes while clusters are ranked afterwards based on their size and connectivity – higher scoring seed nodes may not produce higher scoring clusters. Given that, when expanding a cluster, there may be an unexpected initial discontinuity in size since the **Size Threshold** slider will ignore the presence of other clusters. If the cluster was produced around a low-scoring seed node then more nodes are likely to satisfy the **Node Score Cutoff** parameter. Such situations can indicate that the cluster in question may be part of a larger cluster.

2. Slider Dead-Zone

- Sometimes, on the other hand, moving the **Size Threshold** slider a long distance may produce no changes in cluster size. In such cases, the seed node’s score is so high compared to its immediate neighbourhood that the **Node Score Cutoff** must be increased greatly to include much lower scoring members. This indicates that the cluster is more or less well separated from the surrounding network by a local peak in node scores and as such, it is likely a well defined cluster.

3. No Change

- Lastly, if no changes occur during size exploration, the cluster in question is likely not connected to the larger network and as such cannot expand.

2.6.2 Node Attribute Enumerator

The Enumerator provides a numerical summary of node attribute values possessed by the currently explored cluster’s members. It is meant to inform the user of the cluster’s contents and aid in determining the cluster’s functional relevance. All node attributes that are available for the loaded network are listed in the select box. When an attribute selection is made in one exploration, it persists for all cluster explorations within the given result.

The table below the select box has two columns:

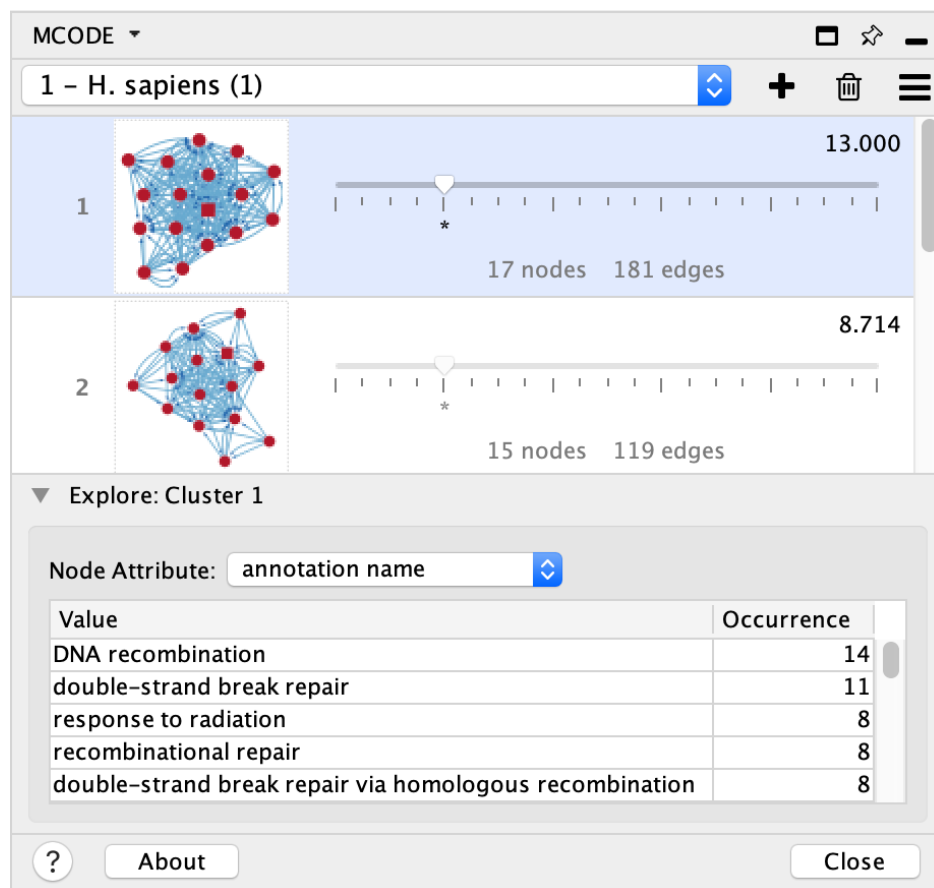
1. Value

- This column lists all node attribute values that are possessed by the cluster being explored. With a simple string type attribute, such as **Node Status**, this list will never exceed the number of cluster members since every member can have only one value and some values may be shared by several members. However, list type attributes such as *Gene Ontology (GO)* terms may outnumber the cluster members since each member can have numerous values.

2. Occurrence

- This column simply displays the number of nodes possessing the particular attribute value listed in each row. The Enumerator table orders the list by the frequency in descending order where the most commonly occurring attribute value is listed on top.
- The Occurrence numbers are best interpreted when compared with the number of nodes in the cluster. For example, when enumerating *Biological Process GO Terms*, it may be a good indicator that the given cluster is biologically relevant if 9 of the 10 cluster members share some specific value.

In combination with the **Size Threshold** slider, the Enumerator can be used to optimize clusters based on functional relevance. As the slider is being manipulated, the enumerator will automatically report changes in cluster content for the selected attribute. As such one can hone in on a size that, for example, reduces nodes with attribute values that are unrelated to some particular value of interest which is simultaneously maximized.



Note: At this stage of MCODE development, the **Node Attribute Enumerator** is a precursor to more automated methods of accomplishing similar attribute-enhanced clustering and statistical reporting.

2.7 Outputting Results

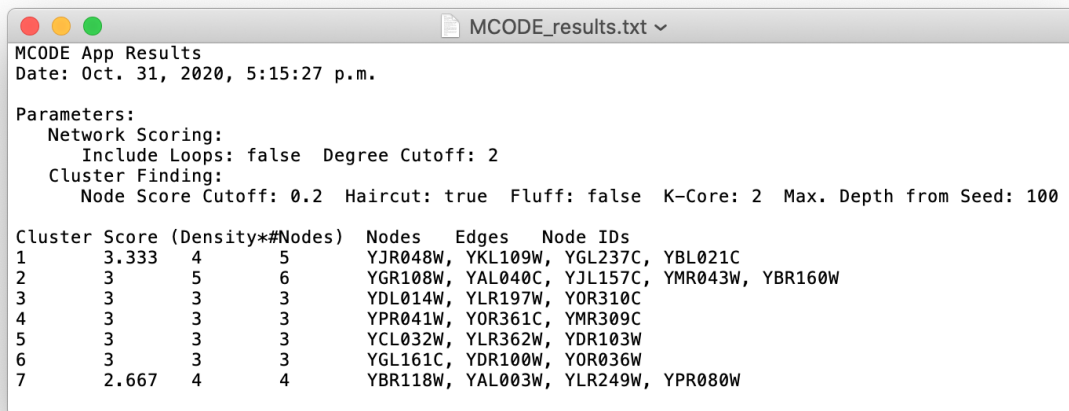
2.7.1 Create Sub-Network

Clusters can be output as sub or child-networks of the original network by selecting the **Create Cluster Network** option located on the MCODE menu which appears after clicking the *Toolbar*'s **Options** button.

- Since exploration allows for a cluster size to change, the user can create as many sub-networks of the same cluster as desired.
- New networks are named by their *result number*, *cluster rank* and *score* (e.g. 1: Cluster 2 (Score: 4.3)).

2.7.2 Export as Text

MCODE results can be exported to a time-stamped text file by selecting the **Export Result** option located on the MCODE **Options** menu (see *Toolbar* section).



```

MCODE App Results
Date: Oct. 31, 2020, 5:15:27 p.m.

Parameters:
  Network Scoring:
    Include Loops: false Degree Cutoff: 2
  Cluster Finding:
    Node Score Cutoff: 0.2 Haircut: true Fluff: false K-Core: 2 Max. Depth from Seed: 100

Cluster Score (Density*#Nodes) Nodes Edges Node IDs
1      3.333  4      5      YJR048W, YKL109W, YGL237C, YBL021C
2      3      5      6      YGR108W, YAL040C, YJL157C, YMR043W, YBR160W
3      3      3      3      YDL014W, YLR197W, YOR310C
4      3      3      3      YPR041W, YOR361C, YMR309C
5      3      3      3      YCL032W, YLR362W, YDR103W
6      3      3      3      YGL161C, YDR100W, YOR036W
7      2.667  4      4      YBR118W, YAL003W, YLR249W, YPR080W
  
```

Clusters are summarized in a tab-delimited format consisting of:

- Cluster: rank
- Score (Density*#Nodes): cluster score (density multiplied by the number of members)
- Nodes: number of nodes
- Edges: number of edges
- Node IDs: cluster member IDs (comma-delimited)

The **parameters** used in scoring and finding the exported result are included in the file as well for future reference:

- Network Scoring:
 - Include Loops
 - Degree Cutoff
 - K-Core
- Cluster Finding:
 - Node Score Cutoff
 - Haircut
 - Fluff
 - K-Core
 - Max. Depth from Seed