Welcome to MSMExplorer
Tutorial and Reference
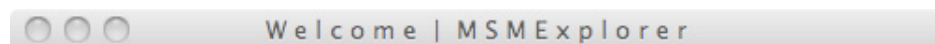
contact: Bryce Cronkite-Ratcliff <brycecr@stanford.edu>

MSMExplorer is a tool for visualizing Markov State Models for protein folding applications as produced by MSMBuilder. The software aims to provide tools for analysis and visualization production.
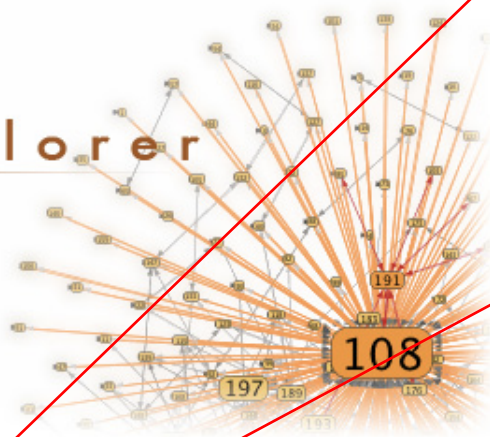
This guide is primarily a reference for the graphical interface of the program. Some additional documentation, including a tutorial video, is available on the Documents page of MSMExplorer's SimTK site (https://simtk.org/docman/?group_id=529). Please direct all questions and issues not answered by documentation to msmexplorer@gmail.com .

Welcome Screen

When you first start MSMExplorer, you are greeted with this screen:

Visualize an MSM with the primary network visualization window (called Graph View). This is the default option. Press ALT-ENTER to select.

Skips loading the graph visually and goes straight to a dialog from which to select nodes for TPT. The utility of this is that you may be able to run TPT using MSMExplorer's setup while bypassing loading the whole graph for visualization. Make sure you know the state number (0-indexed for dats, 1-indexed for mtx) of your start and end states.
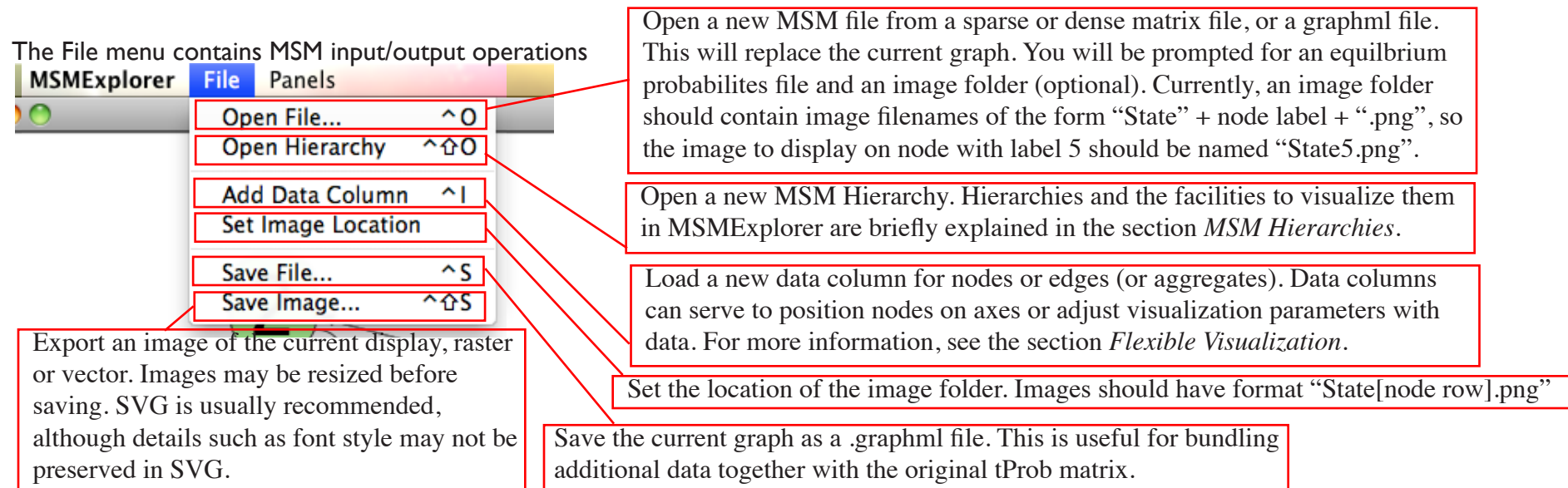
Opening and Saving MSMs

MSMExplorer can open data directly from MSMBuilder output. The main file MSMExplorer uses is a Transition Probability (or Transition Counts) matrix. These matricies may be in dense (space delimited, .dat or .txt) format, or sparse (.mtx) format. MSMExplorer can also open GraphML files (.graphml or .xml) files, which are convenient because they can package information of arbitrary dimension along with the basic graph structure. That is, sparse and dense matrix files only contain one piece of auxiliary information per edge (transition probability), while graphml files can store an arbitrary number of additional data fields for both edges and nodes.

Several sample graphs are included in the "demos" directory inside the main directory of the application ("MSMExplorer"). Try opening and playing with these to experiment with MSMExplorer.

A simple tProb matrix defines nodes (which are assigned integer labels that are 0-indexed for dense matrix files, 1-indexed for mtx files, and given by the "label" field for GraphML files), edges, and the transition probability for each node. Usually, equilibrium probabilities are also desired. Upon opening a matrix file, MSMExplorer will attempt to automatically find an equilibrium probabilities file, which should be a newline deliminted file of probabilities where row index indicates corresponding node index. MSMExplorer assumes this filename starts something like "Population" but you may also specify the eqProb file manually when prompted.

You may also be prompted for an image folder, which allows the nodes to contain the image you specify. Currently, an image folder should contain image filenames of the form "State" + node label + ".png", so the image to display on node with label 5 should be named "State5.png".

The File menu contains MSM input/output operations

Open a new MSM file from a sparse or dense matrix file, or a graphml file. This will replace the current graph. You will be prompted for an equilbrium probabilites file and an image folder (optional). Currently, an image folder should contain image filenames of the form "State" + node label + ".png", so the image to display on node with label 5 should be named "State5.png".

Open a new MSM Hierarchy. Hierarchies and the facilities to visualize them in MSMExplorer are briefly explained in the section *MSM Hierarchies*.

Load a new data column for nodes or edges (or aggregates). Data columns can serve to position nodes on axes or adjust visualization parameters with data. For more information, see the section *Flexible Visualization*.

Set the location of the image folder. Images should have format "State[node row].png"

Save the current graph as a .graphml file. This is useful for bundling additional data together with the original tProb matrix.

Export an image of the current display, raster or vector. Images may be resized before saving. SVG is usually recommended, although details such as font style may not be preserved in SVG.
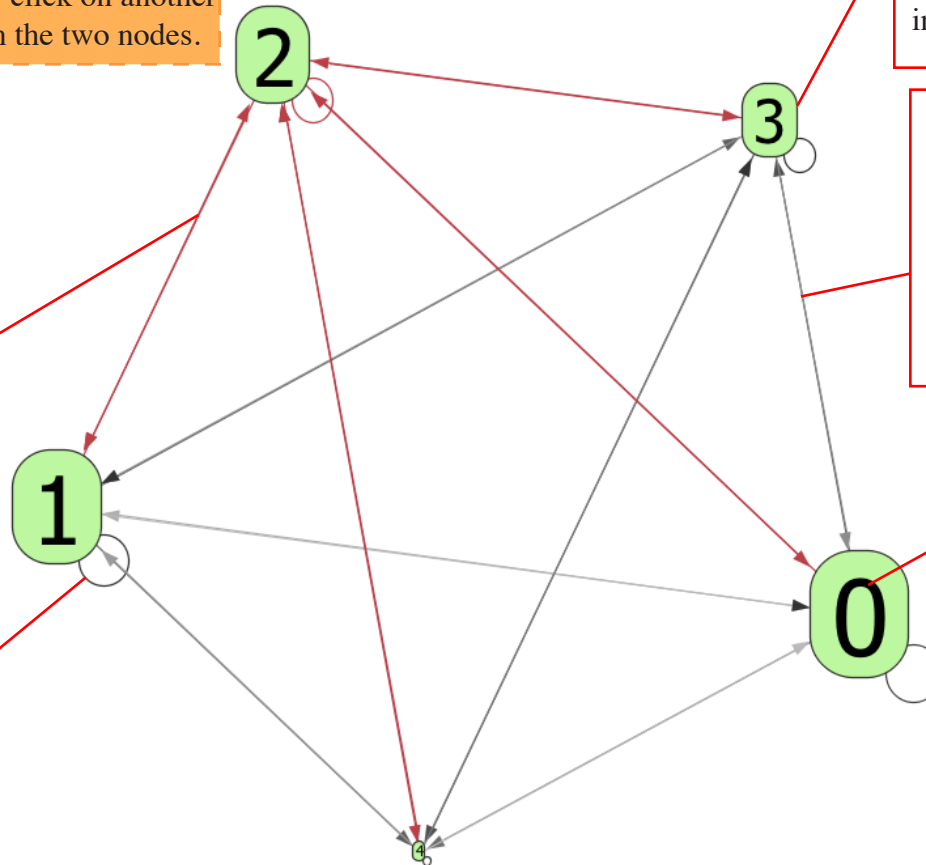
Interacting with the Default Graph

By default, node size varies directly with equilibrium probability. The node size range and the data that determine node size can be changed in the Vis Settings Panel.

Click on a node to select it. This will highlight the edges going in and out of this node in red. The Graph Distance control on the Control Dock will now count distances starting from the slelected node.

By default, edge darkness is directly proportional to transition probability along that edge. The field that controls edge color and the range of edge colors can be changed in the Vis Settings Panel.

Labels are assigned automatically based on the tProb matrix. They can be turned off by switching to "Shape" rendering, modified by opening the Node Table, or changed in color from the Vis Settings panel.

These somewhat funny looking circles are representations of self-edges. They can be toggled off in the Edge panel of the Vis Settings dialog.

See Motion? Layout Misbehaving?

MSMExplorer uses a Force-based layout algorithm to automatically space nodes apart from each other. Nodes act as masses and edges act as springs. The parameters of this layout can be changed in the Force Panel, accessible from the Panels menu and from the main Control Dock.
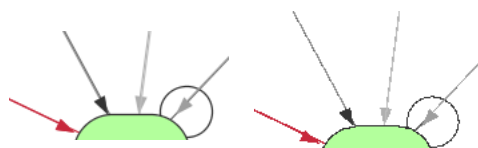
Customized Visualization?

Node size, node color, node shape, edge color, edge weight, and aggregate color can all be set to vary with any data of your choosing, and you are in control of the color and size ranges that are assigned as well. See *Flexible Visualization* for more information.

The Control Dock

Only those nodes with equilibrium probabilities above that indicated on this slider or text box (whichever was last modified) will be shown.

Only nodes within the number of hops indicated on this slider from the currently selected node will be shown.
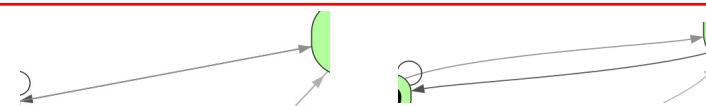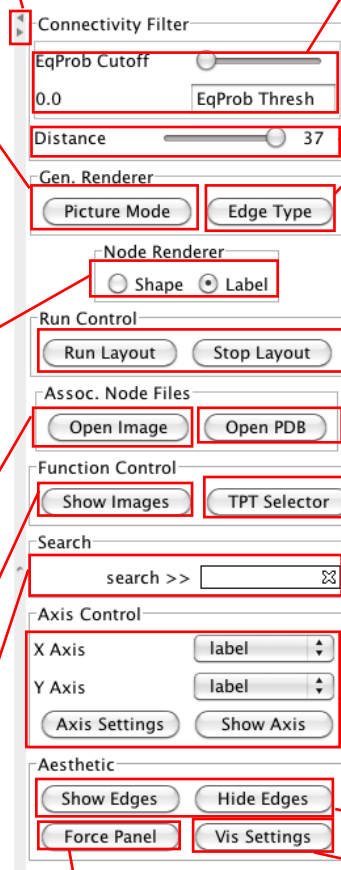
Show or Hide the Control Dock.

AA ON   **AA OFF**

Toggle anti-aliasing. Anti-aliasing makes the image look much nicer, but causes a significant slowdown in rendering speed. Turn aliasing off to sacrifice rendering quality for crisper interactivity.

Straight   **Curved**

Toggle between curved and straight edges. Curved edges make it easier to distinguish between the color/weight of the edges going either direction between the same two nodes. Straight edges are more space effecient.

Alternate between nodes with labels and shape-based nodes. Only Label mode supports images (you can turn off labels when images are visible in Vis Settings). In Shape mode, nodes can take on different shapes based on a node data field.

Run or stop the layout algorithm. Running allows the layout to dynamically adapt to movement, while stopping puts the user in complete control. With large graphs, run will just re-run the one-shot layout algorithm instead of animating.

Open the full-size image for the selected node in another window.

Open the corresponding PDB file for the selected node in PyMol. Prompts for folder with PDBs if not previously indicated. PyMol must be command line program, and PDB's should have format: "State[row number]-0.pdb"

Show images on nodes. Must have Label node renderer selected. Images are in the folder indicated upon opening this MSM. Change node size in Vis Settings to rescale images.

Open a dialog to select the start and end nodes for TPT. Supports selecting multiple source and target nodes.

Axis Controls. See *MSMs on Axes* for more information.

Search for a node by label. If found, the node will be highlighted and centered in the display.

Show or hide edges. Hiding edges can particularly useful for producing visualizations on axes.

Open a panel with force layout parameters to adjust automatic layout output. The Default Spring Length paramater is useful for directly changing how long edges are to space out or contract graphs.

Open a dialog with lots of settings to customize visualizations. See *Flexible Visualization* for more informaiton.

Control Dock Panel

Connectivity Filter
EqProb Cutoff
0.0 | EqProb Thresh
Distance | 37

Gen. Renderer
Picture Mode | Edge Type

Node Renderer
○ Shape ⦿ Label

Run Control
Run Layout | Stop Layout

Assoc. Node Files
Open Image | Open PDB

Function Control
Show Images | TPT Selector

Search
search >>

Axis Control
X Axis | label
Y Axis | label
Axis Settings | Show Axis

Aesthetic
Show Edges | Hide Edges
Force Panel | Vis Settings

MSMs on Axes

MSMExplorer can lay graphs out on one or two axes of any data type that has been loaded into the nodes. New data can be loaded from File->Import Data Column or from "Load new" in the axis field drop down menus. MSMExplorer supports custom axis labels and grid labels; the size and contents of these can be set in "Axis Settings" and they can be repositioned with clicking and dragging. Select "Hide Edges" to produce a graph with just edges. Note that upon selecting "Show Axis" the layout algorithm will stop so as to not interfere with the axis layout; it can be re-enabled by selecting "Run Layout." In addition to label management, "Axis Settings" supports changing the axis range for numerical data fields and spacing of grid lines.

Change the node data field that is used to layout the nodes along the specified axis. "No Axis" will not layout graphs in that direction. New data may be added by selecting "Load new" in either drop-down menu, which will pop up a dialog to specify the name and type of the new data. It's important to specify type correctly in this dialog, as some program functions depend on this.

Axis Control

| X Axis | eqProb |
| Y Axis | label |

Axis Settings   Show Axis

Show or hide the axes. NOTE: In order to affect changes made to the axis data fields or axis settings, toggle this button of and on.

Opens this dialog to change axis settings, labels, spacings, and font sizes.

Automatically determine layout bounds based on the range of values observed. Uncheck to enable bounds boxes.

Auto Set Bounds   Note: Limits for non-numerical datatypes not shown.

| X Min: | 0 | X Max: | 1 |
| Y Min: | 0 | Y Max: | 1 |

Set the upper and lower limit for the axis range. These boxes are only visible if the data type selected for that axis is numerical.

Set the label and font size for each axis. The label can be repositioned by dragging if desired.

| X Label: | Size: | Y Label: | Size: |
| | 4 | | 4 |

X Grid Spacing: 50   Y Grid Spacing: 50

Change the minimum spacing for the grid lines and labels. Units are pixels, so larger values mean fewer gridlines.

Grid Label Size: 4   Cancel   Ok

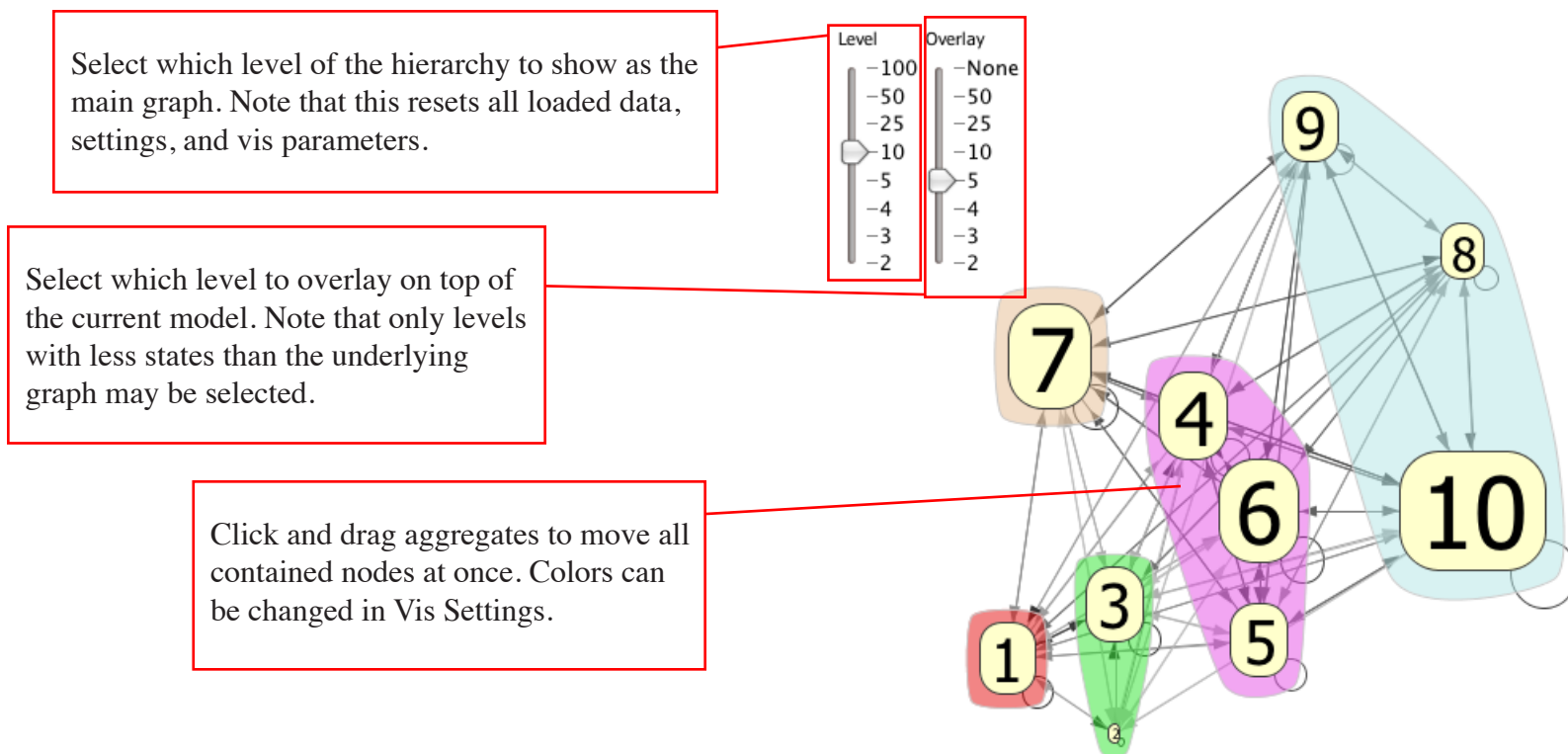Set grid label font size.

Discard changes.

Accept the settings in this dialog. Toggle "Show Axis" off and onn to affect changes.

MSM Hierarchies

      MSMExplorer can load and display MSM Hierarchies, compound models that contain different granularity macrostate models for the same system. A minimal example of the required MSM Hierarchy format can be browsed in the "simple_hierarchy" folder in "demos." Note that the filenames (tProb, Populations, and MacroMapping) are important for loading correctly, although MSMExplorer attempts to provide some flexibility. In particular, each subfolder participating in the hierarchy must contain a transition probability matrix (called tProb with an extension specifying the format) and a file specifying the mapping from a hierarchy-common base (the most populated MSM in the hierarchy) to the macrostates in the transition probability matrix (mapping file called MacroMapping). These should be MSMBuilder defaults. Equilibrium probabilities may be optionally provided (recommended) in a file called Populations.

      Load Hierarchies from the File->Open MSM Hierarchy. Once loaded, the Level and Overlay sliders will be visible, allowing you to switch between levels of the hierarchy and overlay higher-level models on lower level models to observe membership. The colored blobs used to show this mapping are called aggregates; their color can be changed in Vis Settings.

Select which level of the hierarchy to show as the main graph. Note that this resets all loaded data, settings, and vis parameters.

Select which level to overlay on top of the current model. Note that only levels with less states than the underlying graph may be selected.

Click and drag aggregates to move all contained nodes at once. Colors can be changed in Vis Settings.
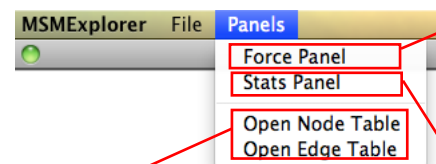
Flexible Visualization: Vis Settings and Backing Data

MSMExplorer stores MSMs as a pair of tables, a Node Table and an Edge Table. The backing data for both of these are accessible (and, to a degree, editable) from the Panels menu. Additional data columns can be added to either table via File->Import Data Column.

Visualization parameters, including node shape, node size, node color, edge weight, and edge color can all be set to vary according to data in any column. Visualization parameters based on numerical data columns (of type Integer, Double, Float, and Long) will assign values from the size or color palette set in Vis Settings by interpolating across the represented numerical data space and matching that to an interpolated position in the palette. Visualization parameters based on non-numerical data types will simply assign different colors or shapes to different data values. Non-numerical columns cannot be used for edge weight or node size because MSMExplorer cannot understand notions of relative size from non-numerical types. In order to save loaded data columns along with graphs, use File->Save Image to save as a GraphML file.

Data columns for nodes only can also be used to lay out the graph on axes; data for nodes can also be added to the Node Tabel from the axis field.

Open a panel to adjust force parameters of the layout.

Open a panel to show the backing data for the panel. Some data, especially strings (including the "label" field) can be modified by double-clicking on the entry.
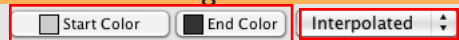
Open a panel with a few simple statistics about the current MSM.

Vis Settings

The Vis Settings panel, accessed from the button on the Control Dock, contains tools for modifying visualization parameters and other aesthetic details.

Similar controls are available for each visualization parameter, so some recurring patterns are explained here before going into the Vis Settings panel specifically.
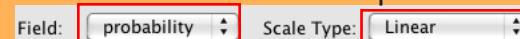
Controls for Setting Color Palettes

Open a Color Chooser dialog to select the colors for the start and end of the spectrum for an Interpolated palette.

Select the palette to use from a set of preset palettes, or select "Interpolated" to use a palette that ranges between the chosen start and end colors.

Controls for Setting Size Ranges

Set the lower and upper limits for the range to assign sizes from. Put the knobs together to set all objects the same size.

Data field controls for Visualization parameters

Select the field with which to alter the visualization parameter.

Select the interpolation scale for assigning values. Scales such as Log can help get a reasonable visualization with very different parameters. Note that this only has effect for numerical data types.

Vis Settings Panes

Open a color chooser to select a color to set all nodes, reegardless of node color data settings.
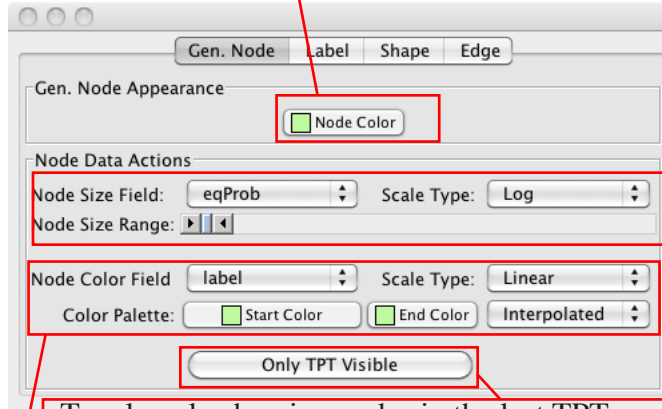
Set the position of images relative to label when images are displayed.

Set the font size for the label text on nodes. Useful for matching text size with image size.

Adjust how rounded label nodes appear

Set the node shape to vary with the selected shape field from the set of available shapes.

Set the shape of all the nodes to the selected shape.

| Gen. Node | Label | Shape | Edge |

Gen. Node Appearance

Node Color

Node Data Actions

Node Size Field: eqProb   Scale Type: Log
Node Size Range: ▶ ◀

Node Color Field: label   Scale Type: Linear
Color Palette: Start Color End Color Interpolated

Only TPT Visible

| Gen. Node | Label | Shape | Edge |

Rounding Radius: 8

Label Font Size: 10

Image Position: Left

Images Only   Text Color

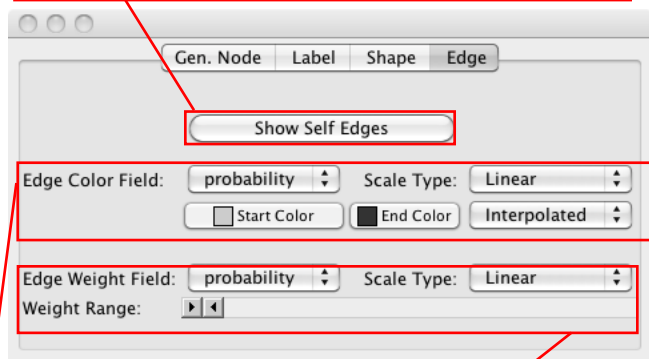| Gen. Node | Label | Shape | Edge |

Node Shape: Cross

Node Shape Field: label

Toggle only showing nodes in the last TPT run. Only visible if TPT has been run on this graph.

Node color vis parameter settings.

Node size vis parameter settings. Data field must be numerical.
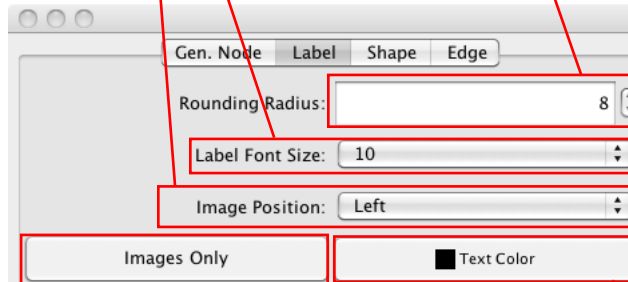
Hide labels when images are showing. Images must already be showing for this to have an effect. Recall that images cannot be shown with the Shape node renderer selected in the Control Dock.

Set the color of the labels on nodes.

Show or hide self-edges -- those funky circles attached to nodes. Self-edges can look weird, but they can give useful information as well.
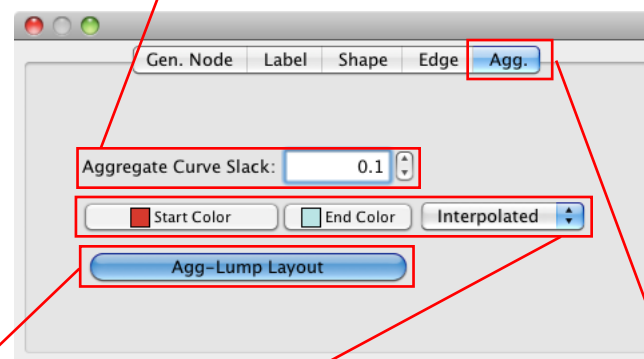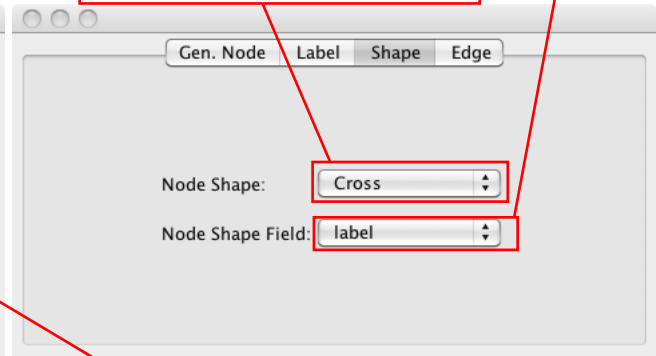
Set how curved the aggregate shapes are. Excessively high values will cause peculiar loopiness; low values produce sharp points.

A Few Reminders

:: Add new data columns via File->Import Data Column

File	Panels
Open File...	^O
Open Hierarchy	^⇧O
Add Data Column	^I

| Gen. Node | Label | Shape | Edge |

Show Self Edges

Edge Color Field: probability   Scale Type: Linear
Start Color End Color Interpolated

Edge Weight Field: probability   Scale Type: Linear
Weight Range: ▶ ◀

| Gen. Node | Label | Shape | Edge | Agg. |

Aggregate Curve Slack: 0.1

Start Color End Color Interpolated

Agg-Lump Layout

:: Change between node renders with the "Label" and "Shape" radio buttons on on the Control Dock.

Node Renderer
○ Shape ● Label

:: Inspect backing data with Panels->Show Node Table or Panels->Show Edge Table

Panels	
Force Panel	
Stats Panel	
Open Node Table	
Open Edge Table	

Edge color vis parameter settings.

Edge weight vis parameter settings.

Set the color palette for aggregate objects.

The Aggregate pane is only visible when aggregates are showing in the graph.

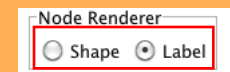Toggle whether to automatically lump nodes within an aggregate together so that aggregates are separated.

Transition Path Theory

MSMExplorer can automatically produce graphs depecting the most important paths between sets of start and end nodes according to Transition Path Theory (TPT). To run TPT between two nodes, click on (select) the start node, and then Control-Click on the destination node. A new TPTView window will appear with a visualization of the highest reactive paths between your source and target nodes. Note that the TPT Selector, accessible from the Welcome Screen or the Control Dock, allows selecting source and target node sets of arbitrary size, although the program currently does not support sets of cardinality greater than one.

TPT and GraphView

:: After TPT is run, the new data columns created by TPT are available for use in the main Graph View window.
:: The same mouse controls (dragging, zooming, panning, and fit-to-zoom) apply in TPTView as well.

The Y-axis indicates equilibrium probability.

Show or hide the axis. Showing the axis will reset node positions to axis positions and fix all nodes.

Toggle between color modes; one mode colors source, target, and other nodes differently, and the other colors fixed and unfixed nodes differently.

Export a vector or rasterimage of the current display.

Show or hide images. Image resolution will scale with current zoom level.

MSMExplorer attempts position nodes along the X-axis with some basic intelligence, but nodes can be repositioned by dragging.

Edge thickness scales with reactive flux along that edge.

Open a panel to adjust force-based layout parameters.

Show or hide node labels.

| Fix Position | Show Images | Save Image | Color Mode | Show Axis | Force Panel | Node Labels | 3 |

Edge Weight ▶| | |◀   Node Size ▶|◀

Enabled when a node is selected. Toggle whether the position of the node is fixed or decided by the force-directed layout.

Sets the upper and lower bounds of the range for edge weight.
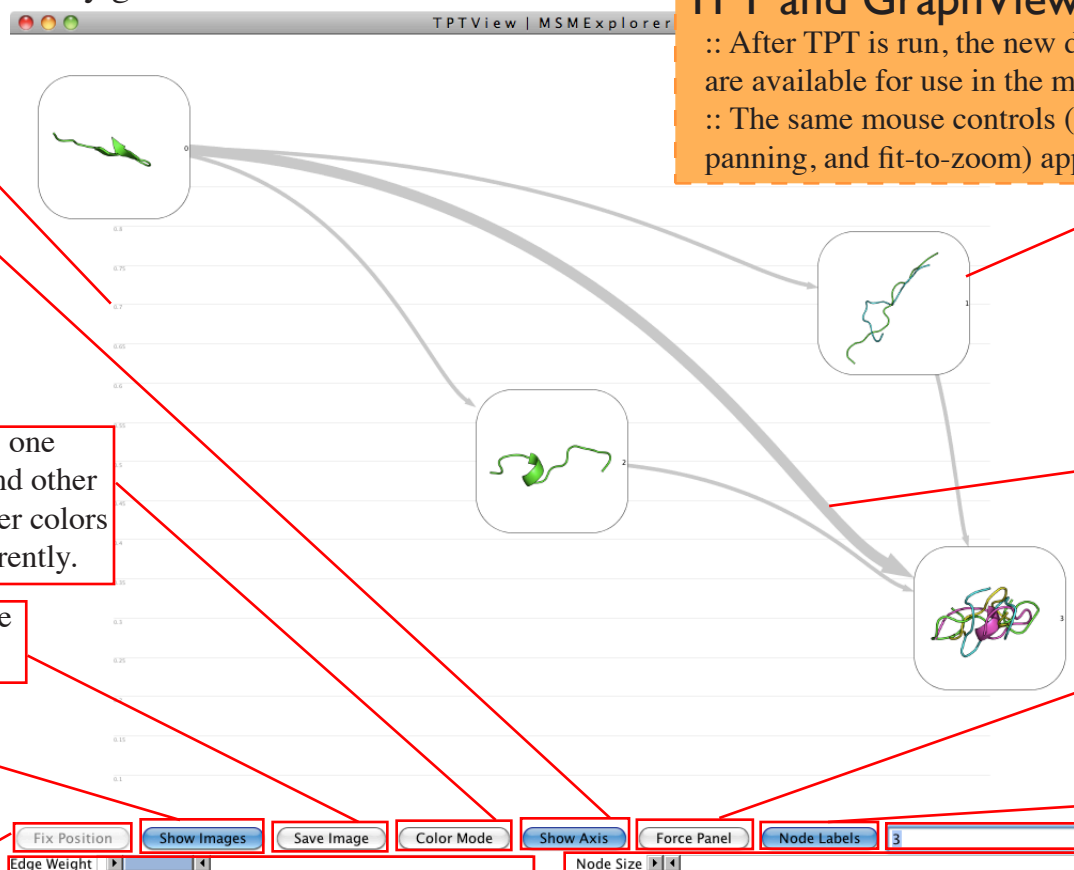
Sets the upper and lower bounds for node size.

Sets the number of top paths to show. MSMExplorer will show the minimum of the number here and the number of extant paths.

Evolving MSMExplorer

:: The information in this reference is up-to-date as of August 16th, 2012.
:: The most recent code can be acquired via github at https://github.com/SimTk/msmexplorer
:: MSMExplorer is released under LGPL. Fork it and fix it and generally have fun.
:: MSMExplorer is a work in progress. Help create it on github and we'll all love you forever!
:: This document is also released under LGPL and can be acquired at https://github.com/brycecr/msmexplorer_doc

Brief Credits

More in README.

:: MSMExplorer was written mostly by Bryce Cronkite-Ratcliff, so poke him with any complaints or questions.
:: MSMExplorer was built with funding from the Stanford VPUE in the lab of Professor Vijay Pande.
:: MSMExplorer makes extensive use of Prefuse, a framework written by Professor Jeff Heer and contributors.
:: MSMExplorer uses Java, Netbeans, Apache Ant, Apache Batik, and Apache Commons Math
:: Bryce wrote this guide without spellcheck and his spelling already is questionable, so expect errors and let him know.

Happy MSMExploring!