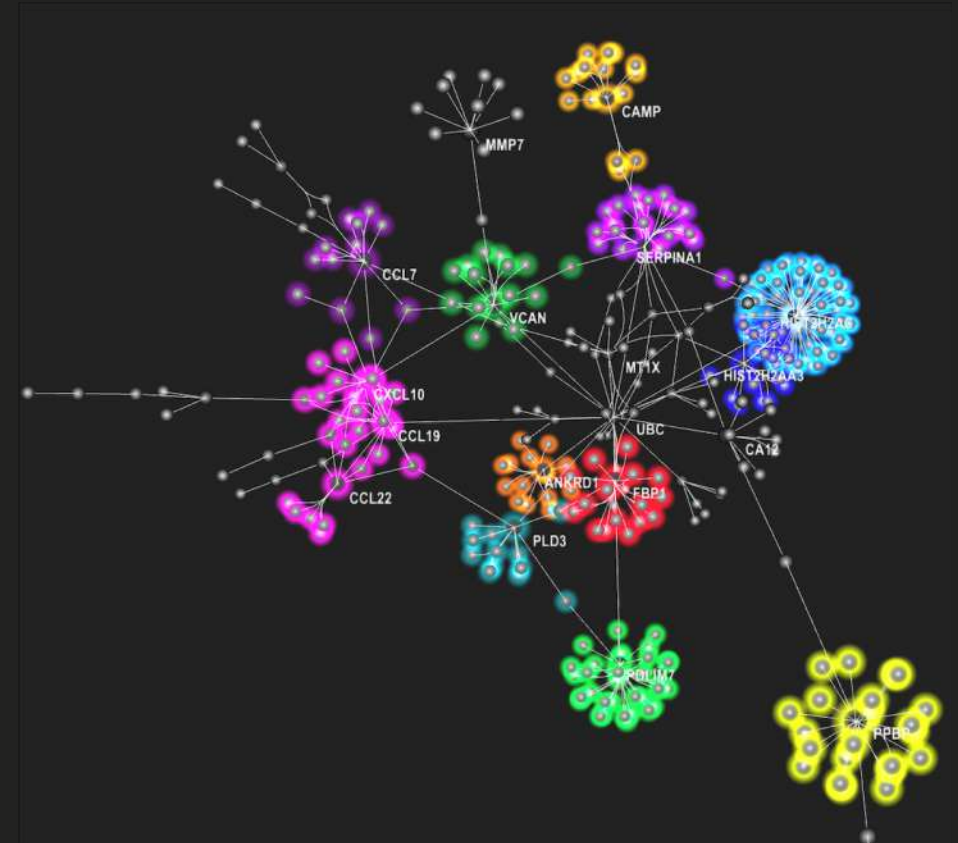
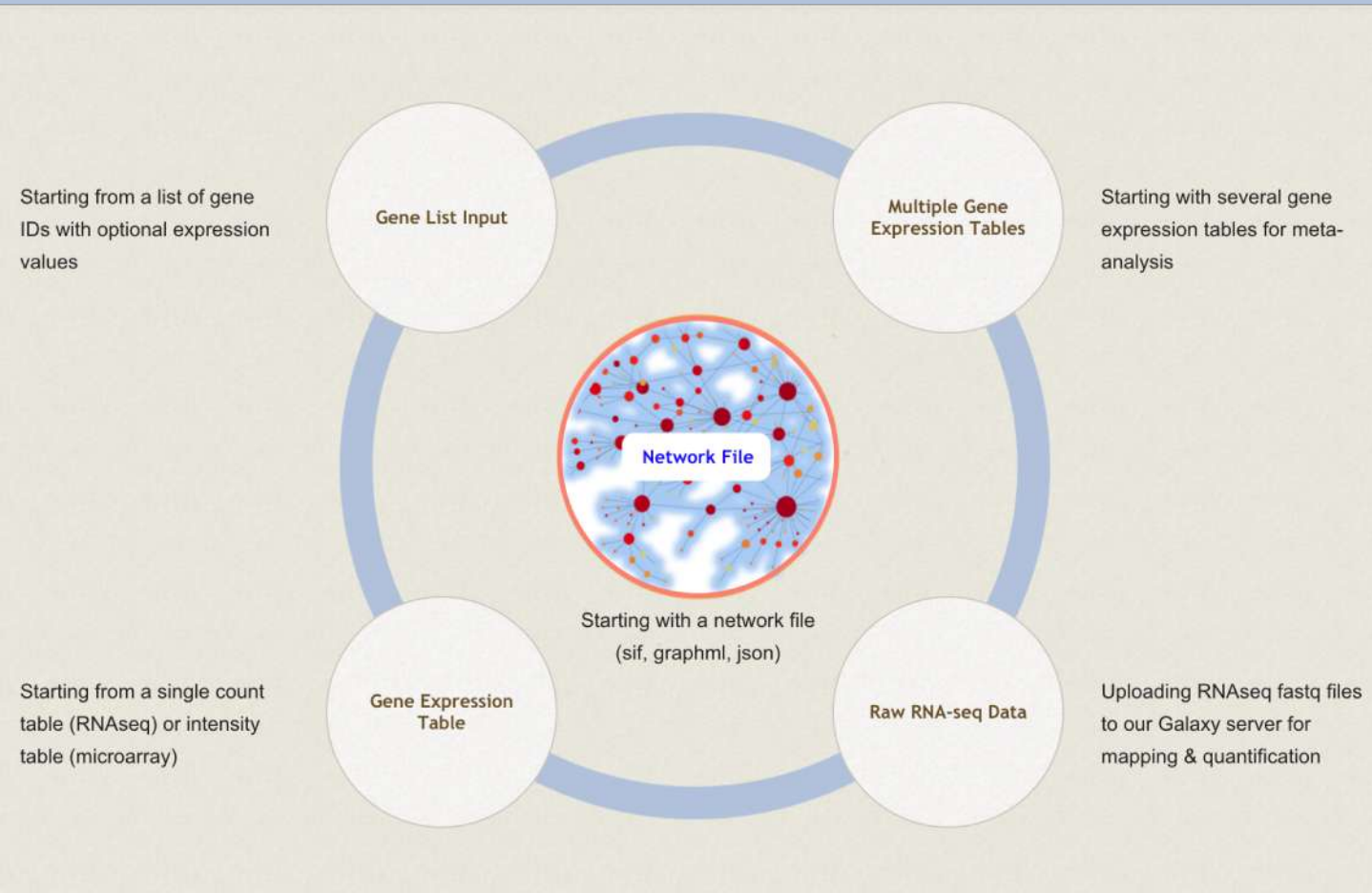


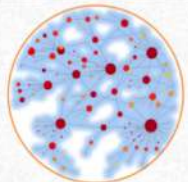
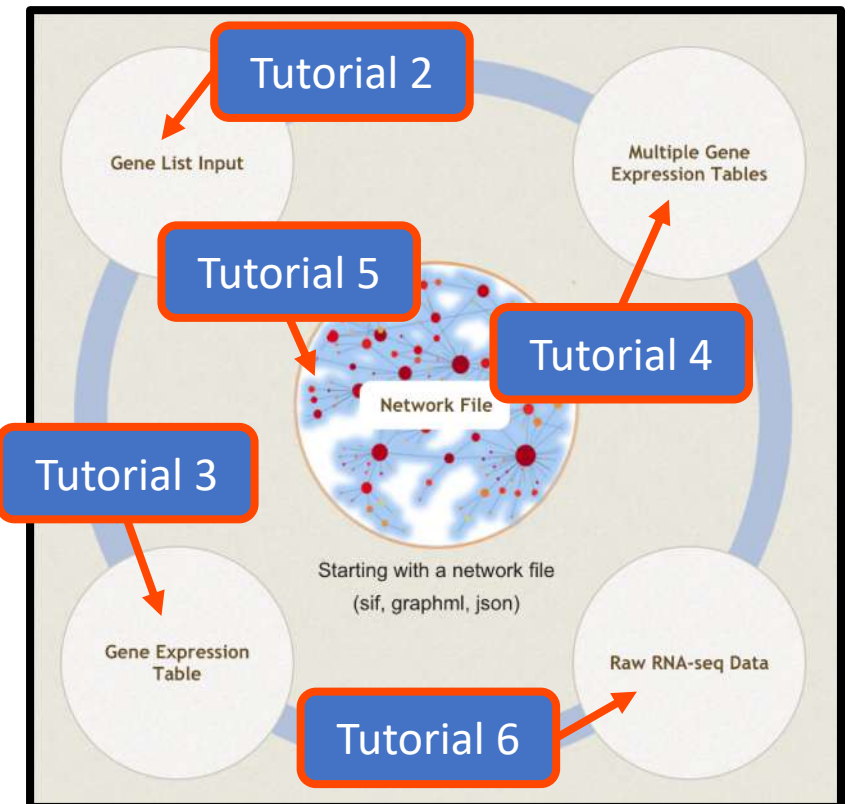
Tutorial 5: network file



Intro to NetworkAnalyst

- Web application that enables complex meta-analysis and visualization
- Designed to be accessible to biologists rather than specialized bioinformaticians
- Integrates advanced statistical methods and innovative data visualization to support:
 - Efficient data comparisons
 - Biological interpretation
 - Hypothesis generation

Tutorial 1: Overview



NetworkAnalyst -- a web-based platform for gene expression profiling & biological network analysis

Computer and browser requirements

- A modern web browser with Java Script enabled
 - Supported browsers include Chrome, Safari, Firefox, and Internet Explorer 9+
- For best performance and visualization, use:
 - Latest version of Google Chrome
 - A computer with at least 4GB of physical RAM
 - A 15-inch screen or bigger (larger is better)
- Browser must be WebGL enabled for 3D network visualization
- 50MB limit for data upload
 - ~300 samples for gene expression data with 20 000 genes

Goals for this tutorial

- A challenge for 'omics analysis is to find ways to benefit from big data without being overwhelmed by it
- Visualizing key functions and relationships within large datasets is an integral part of this
- Networks capture the topology of statistical and functional relationships in graphical form – thus they are a powerful tool for 'omics visual analytics
- The goal of this tutorial is to introduce NetworkAnalyst's tools for visualizing 2D and 3D networks from user-defined network files

Network file formats

- NetworkAnalyst currently accepts four different file formats:
 - [JSON graph format](#) – JSON-schema based format
 - [Edge list](#) – a .txt file
 - [SIF format](#) – commonly used in Cytoscape
 - [GraphML](#) – XML based format
- The file formats will be explored in more detail on the next slides

JSON graph format

After a header section, a .json network file contains information on each “node” and “edge”.

Each node is assigned a numerical id, a name, and a Cartesian position.

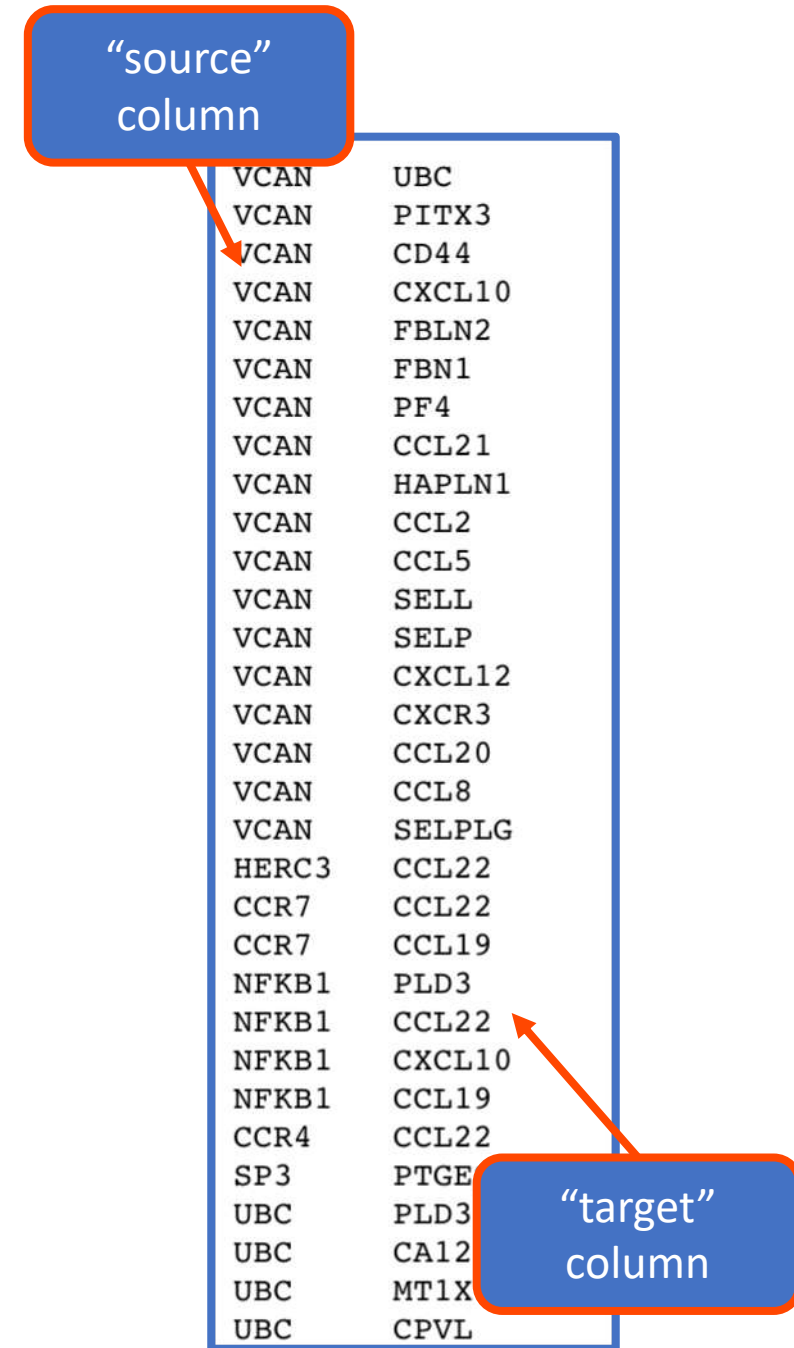
Edges contain more information because they describe relationships between nodes. In addition to their own id, they contain the ids of the nodes that they connect (“source” and “target”).

```
"elements": {
  "nodes": [{
    "data": {
      "selected": false,
      "id": "1",
      "SUID": 1,
      "name": "CRMP1",
      "shared_name": "CRMP1"
    },
    "selected": false,
    "position": {
      "x": -24.317,
      "y": -1.6134
    }
  }, {
```

```
"edges": [{
  "data": {
    "selected": false,
    "id": "2",
    "SUID": 2,
    "directed": false,
    "value": 1,
    "symbol": "NA2",
    "interaction": "pp",
    "shared_interaction": "pp",
    "name": "RBM48 (pp) CDK5RAP2",
    "shared_name": "RBM48 (pp) CDK5RAP2",
    "source": "2",
    "target": "35"
  },
  "selected": false
```

Simple interaction format (SIF)

This is the simplest file format supported by NetworkAnalyst – two columns describing the source and target of each edge. SIF files can be saved as .txt and are easy to generate.



The diagram shows a list of source-target pairs in a two-column format. A blue box with an orange border labeled "source" column has an arrow pointing to the first column. Another blue box with an orange border labeled "target" column has an arrow pointing to the second column. The list contains 30 rows of data.

VCAN	UBC
VCAN	PITX3
VCAN	CD44
VCAN	CXCL10
VCAN	FBLN2
VCAN	FBN1
VCAN	PF4
VCAN	CCL21
VCAN	HAPLN1
VCAN	CCL2
VCAN	CCL5
VCAN	SELL
VCAN	SELP
VCAN	CXCL12
VCAN	CXCR3
VCAN	CCL20
VCAN	CCL8
VCAN	SELPLG
HERC3	CCL22
CCR7	CCL22
CCR7	CCL19
NFKB1	PLD3
NFKB1	CCL22
NFKB1	CXCL10
NFKB1	CCL19
CCR4	CCL22
SP3	PTGE
UBC	PLD3
UBC	CA12
UBC	MT1X
UBC	CPVL

GraphML format

GraphML files also have a large header (not shown here), followed by node and edge definitions.

```
▼<node id="Isg15">  
  <data key="size">21.265706276554262</data>  
  <data key="type">interactor</data>  
  <data key="color">0x0080ff</data>  
  <data key="name">Isg15</data>  
  <data key="Expression">0</data>  
  <data key="positionX">172.678056951515</data>  
  <data key="positionY">181.778611453336</data>  
  <data key="positionZ">153.426310117964</data>  
  <data key="halo">0</data>  
  <data key="haloColor">0x000000</data>  
</node>
```

The node definition is more complex and allows them to be associated with additional details such as size, color, and gene expression.

```
<edge source="Skil" target="Fzr1"/>
```

Edges have a very simple definition – just the source and target node ids

Upload example data

NetworkAnalyst

https://www.networkanalyst.ca/NetworkAnalyst/faces/uploads/NetUploadView.xhtml

NetworkAnalyst -- network-based visual analytics for gene expression prof

Home ? FAQs

Upload Data

- Network Creation
- Network Visualization
- Download
- Exit

Upload a network file

Please use the panel below to upload a network file for visual exploration. The detailed descriptions on how to prepare different network files are available in the help section. For first time users, you are advised to explore the features using one of our example data sets.

Submit your data

Data Type: Edge list (.txt)

Data file: Choose File No file chosen

Submit

Try our example data

- A small network (.json)
- A medium network (.txt)
- A large network (.graphml)

Submit

Previous

Proceed

Xia Lab @ McGill (last updated 2019-01-18)

NetworkAnalyst has three different formats for example data files. We will use the medium-sized network to explore the network visualization tools.

1

2

Select "A medium network" and click "Submit"

Network creation

The network creation page allows you to manage the size of the uploaded network. The mapping overview summarizes the current network attributes.

The screenshot shows the 'Network Creation' interface. On the left, a sidebar contains buttons for 'Network Creation', 'Network Visualization', 'Download', and 'Exit'. The main area displays a table with network statistics and a 'Download' button. A 'Network Tools' panel on the right includes 'Reset Network', 'Degree Filter', 'Betweenness Filter', and 'Batch Exclusion'. Callouts point to various elements: 'Genes in network' points to the 'Nodes' column, 'Interactions between pairs of nodes' points to the 'Edges' column, 'Download to import network into other tools (i.e. Cytoscape)' points to the 'Download' button, and 'Sets of connected nodes and edges' points to the table header.

Networks	Nodes	Edges	Interactions (.SIF)
subnetwork1	381	426	Download

Sets of connected nodes and edges

Genes in network

Download to import network into other tools (i.e. Cytoscape)

Interactions between pairs of nodes

We recommend 200-2000 nodes for practical reasons: larger networks are difficult to interpret while smaller networks will not give systems-level understanding. NetworkAnalyst offers a range of tools for adjusting size:

- **Reset Network:** return to default network
- **Degree Filter:** the degree of a node is its number of connections to other nodes, so this filter retains hub genes
- **Betweenness Filter:** the betweenness measures the number of shortest paths through a node, so this filter retains genes that connect clusters
- **Batch exclusion:** enter a list of genes to exclude from the network

Network visualization tools

Here are some of the main network visualization functions. This tool bar is the same for all 2D networks on NetworkAnalyst.

- ✓ subnetwork1
- subnetwork2
- subnetwork3
- subnetwork4
- subnetwork5
- subnetwork6
- subnetwork7
- subnetwork8
- subnetwork9

The screenshot shows the NetworkAnalyst web application. At the top, there's a navigation bar with 'Home', 'FAQs', 'Tutorials', 'Updates', and 'Resources'. Below that is a toolbar with options for 'Network', 'Background', 'View', 'Layout', 'Node', 'Edge', 'Scope', and 'Download'. The 'View' dropdown is set to 'Topology'. On the left, there's a 'Node Explorer' table with columns for ID, Name, Degree, Betweenness, and Exp. The main area displays a network graph with nodes and edges. A 'Background Color' dialog box is open, showing a color picker with a 'choose' button and the hex code #222222. A 'Topology Expression Plain' dropdown menu is also visible, with 'Custom' selected.

Select to view a different subnetwork of interconnected nodes. Typically subnetwork1 is significantly larger (“continent”) than all others (“islands”).

Black
White
✓ Custom

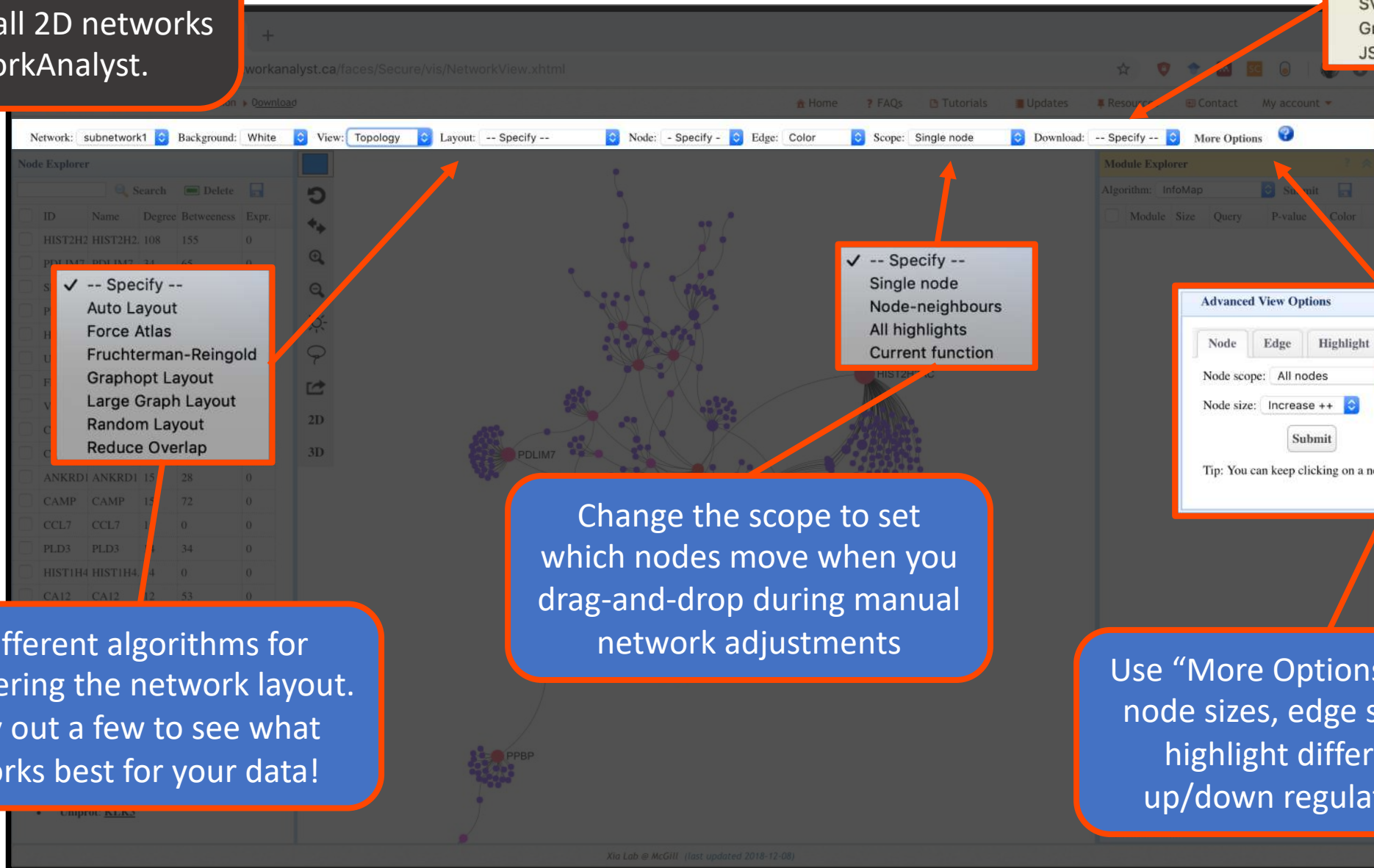
✓ Topology
Expression
Plain

Topology colors nodes according to their degree, expression colors based on uploaded fold changes, and plain makes all nodes gray (good when you want to highlight specific network regions later on)

Set the background to any color

Network visualization tools

Here are some of the main network visualization functions. This tool bar is the same for all 2D networks on NetworkAnalyst.



✓ -- Specify --
PNG Image
SVG Format
GraphML
JSON

✓ -- Specify --
Auto Layout
Force Atlas
Fruchterman-Reingold
Graphopt Layout
Large Graph Layout
Random Layout
Reduce Overlap

✓ -- Specify --
Single node
Node-neighbours
All highlights
Current function

Advanced View Options

Node Edge Highlight

Node scope: All nodes

Node size: Increase ++

Submit

Tip: You can keep clicking on a node to increase its size

Different algorithms for rendering the network layout. Try out a few to see what works best for your data!

Change the scope to set which nodes move when you drag-and-drop during manual network adjustments

Use "More Options" to change node sizes, edge shapes, and highlight differently for up/down regulated nodes

Network visualization tools

Return network to default view

Autofit view to window, zoom in, and zoom out – more labels will appear as you zoom in

Manually select a batch of nodes to adjust their position (drag and drop)

Switch between 2D and 3D

Highlight all of the seed genes – genes that were in the uploaded list

Extract the highlighted nodes to visualize them separately from the rest of the network. This allows you to view regions of interest in greater detail.

Set the highlight to any color

Here are some of the main network visualization functions. This tool bar is the same for all 2D networks on NetworkAnalyst.

The screenshot displays the NetworkAnalyst web interface. At the top, a breadcrumb trail reads: Upload Data > Network Selection > Network Creation > Network Visualization > Download. Below this, a toolbar contains icons for home, autofit, zoom in, zoom out, highlight, and drag-and-drop. A central network graph shows nodes and edges, with some nodes highlighted in blue. A color selection dialog is open, showing a color spectrum and the hex code #0080ff. On the left, a table lists gene names and associated values. At the bottom, a 'Current Selections' panel shows the selected gene: HIST2H2AC (Ensembl: ENSP00000332194). The footer text reads: Xia Lab @ McGill (last updated 2018-12-12).

Gene Name	Value 1	Value 2	Value 3
ENSP00000332194	42683	2.5	
ENSP000 TGM2	50	65494	2.07
ENSP000 EGR2	46	41628	
ENSP000 PDLIM4	44	57357	3.1
ENSP000 UBC	24	281010	0
ENSP000 ITGB8	22	17543	2.09
ENSP000 CPVL	22	31956	-3.65
ENSP000 TLR7	22	21469	-2.23
ENSP000 ADAM15	21		
ENSP000 CD93	20		
ENSP000 MMP7	20		

Make network edges curved

The screenshot shows the NetworkAnalyst web interface. The browser address bar displays <https://www.networkanalyst.ca/faces/Secure/vis/NetworkView.xhtml>. The interface includes a navigation menu with links for Home, FAQs, Tutorials, Updates, Resources, Contact, and My account. Below the menu, there are several dropdown menus for configuration: Network (subnetwork1), Background (Black), View (Topology), Layout (-- Specify --), Node (- Specify -), Edge (- Specify -), Scope (Single node), and Download (-- Specify --). A 'More Options' button is also visible. On the left, there is a 'Node Explorer' panel with a search bar and a table of nodes. The main area displays a network graph with nodes and edges. An 'Advanced View Options' dialog box is open, showing settings for Node, Edge, and Highlight. The 'Edge shape' is set to 'Curve' and the 'Edge color' is set to 'Off'. A 'Submit' button is at the bottom of the dialog. On the right, there is a 'Module Explorer' panel. At the bottom, there are 'Path Explorer' and 'Batch Selection' panels. The footer text reads 'Xia Lab @ McGill (last updated 2018-12-08)'.

1 Change the background and node colors

2

3 Select "More Options" and change edge shape to "Curve"

4

Advanced View Options

Node Edge Highlight

Edge shape: Curve

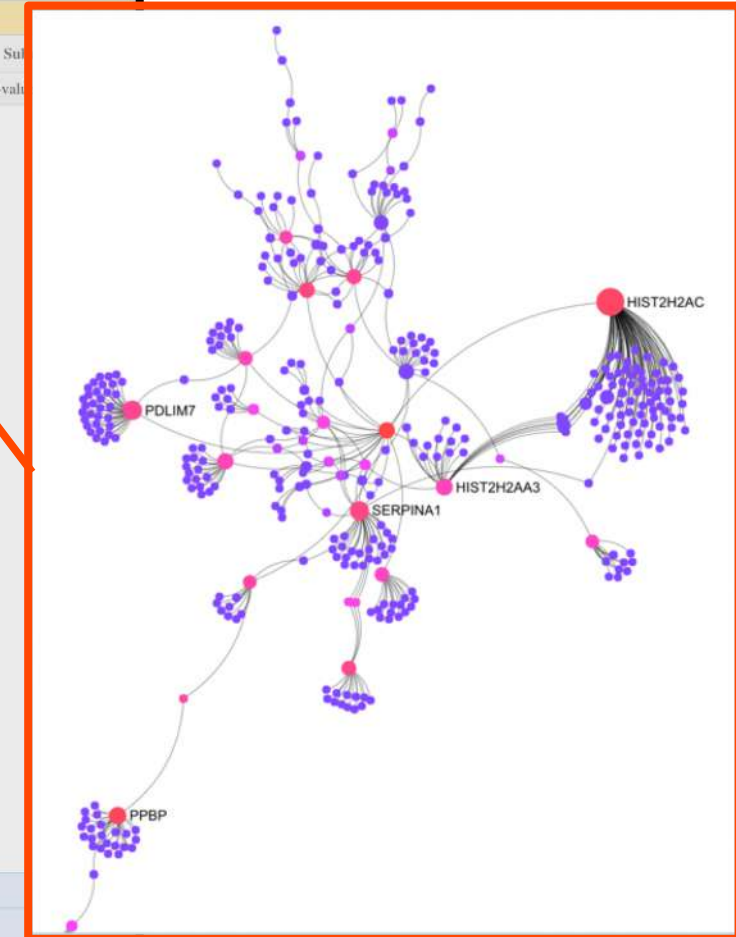
Edge color: Off

Submit

VCAN	VCAN	18	0	0
CCL19	CCL19	18	90.333	0
CXCL10	CXCL10	17	62.333	0
ANKRD1	ANKRD1	15	28	0
CAMP	CAMP	15	72	0
CCL7	CCL7	15	0	0
PLD3	PLD3	14	34	0
HIST1H4	HIST1H4	14	0	0
CA12	CA12	12	53	0
MMP7	MMP7	12	24	0

Curved edges can reduce edge clutter, which is especially useful when there are a few nodes with a very high degree (as in this network)

Xia Lab @ McGill (last updated 2018-12-08)



Generate high-resolution images

The screenshot shows the NetworkAnalyst web application. The browser address bar displays `https://www.networkanalyst.ca/faces/Secure/vis/NetworkView.xhtml`. The interface includes a navigation menu with links for Home, FAQs, Tutorials, Updates, Resources, Contact, and My account. Below the menu, there are configuration options for Network (subnetwork1), Background (White), View (Topology), Layout (-- Specify --), Node (- Specify -), Edge (Color), Scope (Single node), and Download (-- Specify --). A 'More Options' button is also present.

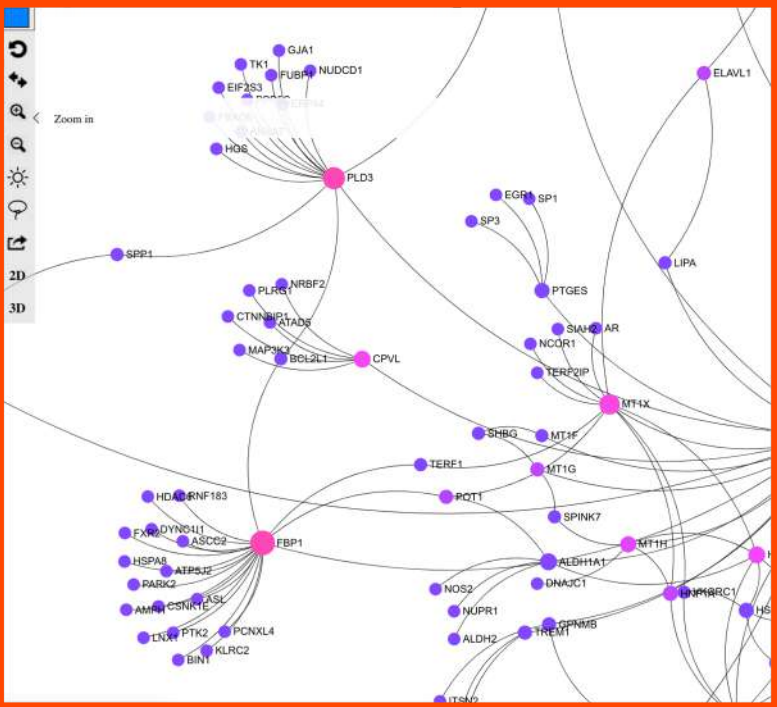
On the left, the 'Node Explorer' table lists nodes with their IDs, names, degrees, betweenness values, and expected values. The table is as follows:

ID	Name	Degree	Betweenness	Expr.
<input type="checkbox"/>	HIST2H2	108	155	0
<input type="checkbox"/>	PDLIM7	34	65	0
<input type="checkbox"/>	SERPINA	32	81.738	0
<input type="checkbox"/>	PPBP	24	176	0
<input type="checkbox"/>	HIST2H2	22	30.786	0
<input type="checkbox"/>	UBC	20	277	0
<input type="checkbox"/>	FBP1	20	35	0
<input type="checkbox"/>	VCAN	18	0	0
<input type="checkbox"/>	CCL19	18	90.333	0
<input type="checkbox"/>	CXCL10	17	62.333	0
<input type="checkbox"/>	ANKRD1	15	28	0
<input type="checkbox"/>	CAMP	15	72	0
<input type="checkbox"/>	CCL7	15	0	0
<input type="checkbox"/>	PLD3	14	34	0
<input type="checkbox"/>	HIST1H4	14	0	0
<input type="checkbox"/>	CA			
<input type="checkbox"/>	MM			
<input type="checkbox"/>	CC			
<input type="checkbox"/>	MT			
<input type="checkbox"/>	HIS			
<input type="checkbox"/>	CST			

The main visualization area shows a network graph with nodes and edges. A '2D' button is highlighted in the bottom left corner of the visualization area. A '3D' button is also visible. The 'Current Selections' panel at the bottom left shows the selected node: Name: KLK3, Uniprot: [KLK3](#).

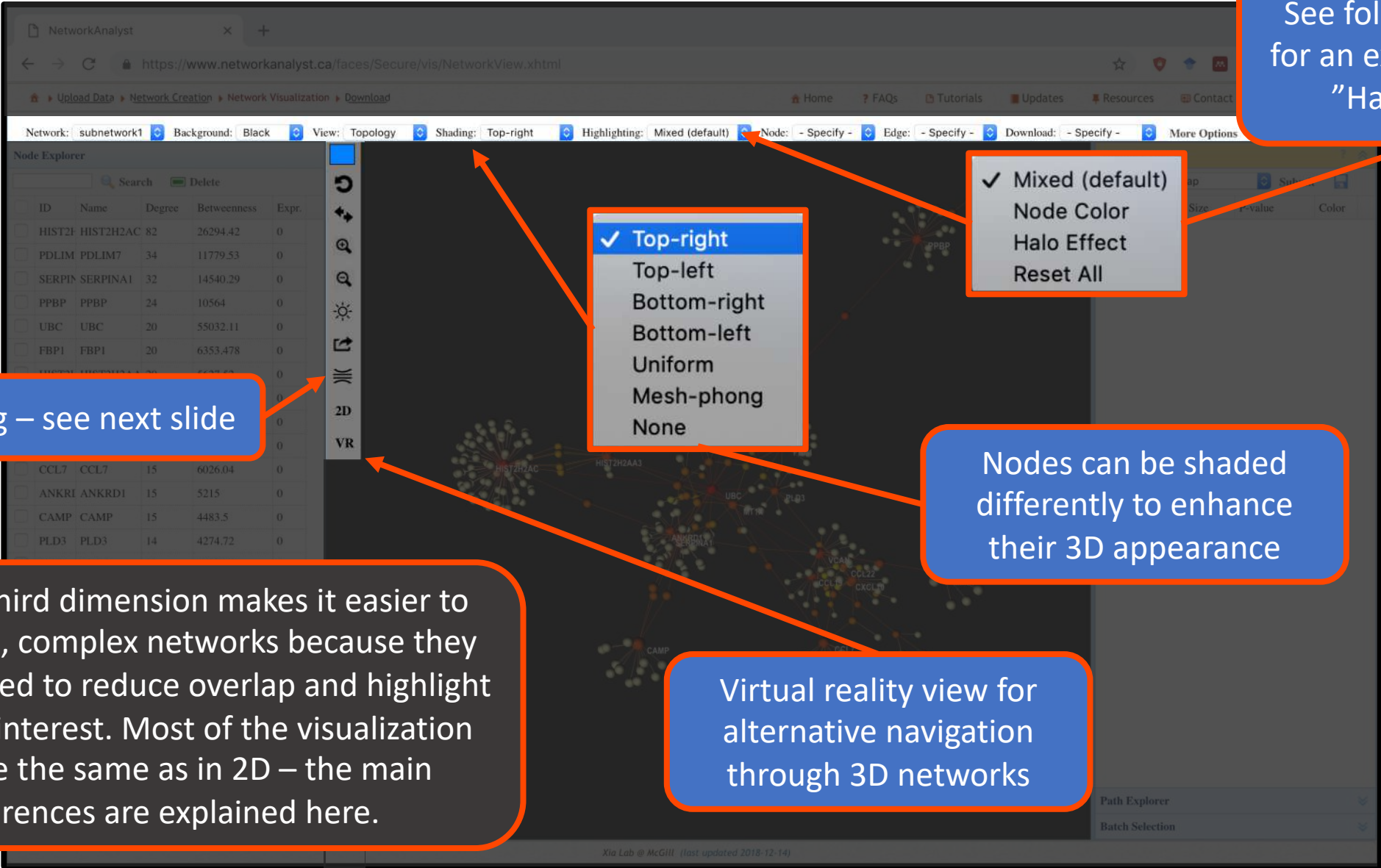
Download high resolution images

1 Zoom in to see more labels



3 When you are finished exploring the network in 2D, move on to 3D

3D network visualization tools



Edge bundling – see next slide

See following slides for an example of the "Halo Effect"

- ✓ Top-right
- Top-left
- Bottom-right
- Bottom-left
- Uniform
- Mesh-phong
- None

- ✓ Mixed (default)
- Node Color
- Halo Effect
- Reset All

Nodes can be shaded differently to enhance their 3D appearance

Adding a third dimension makes it easier to view dense, complex networks because they can be rotated to reduce overlap and highlight regions of interest. Most of the visualization tools are the same as in 2D – the main differences are explained here.

Virtual reality view for alternative navigation through 3D networks

Use edge bundling to enhance 3D visualization

Click the edge bundling icon and "Confirm" in the message box

2

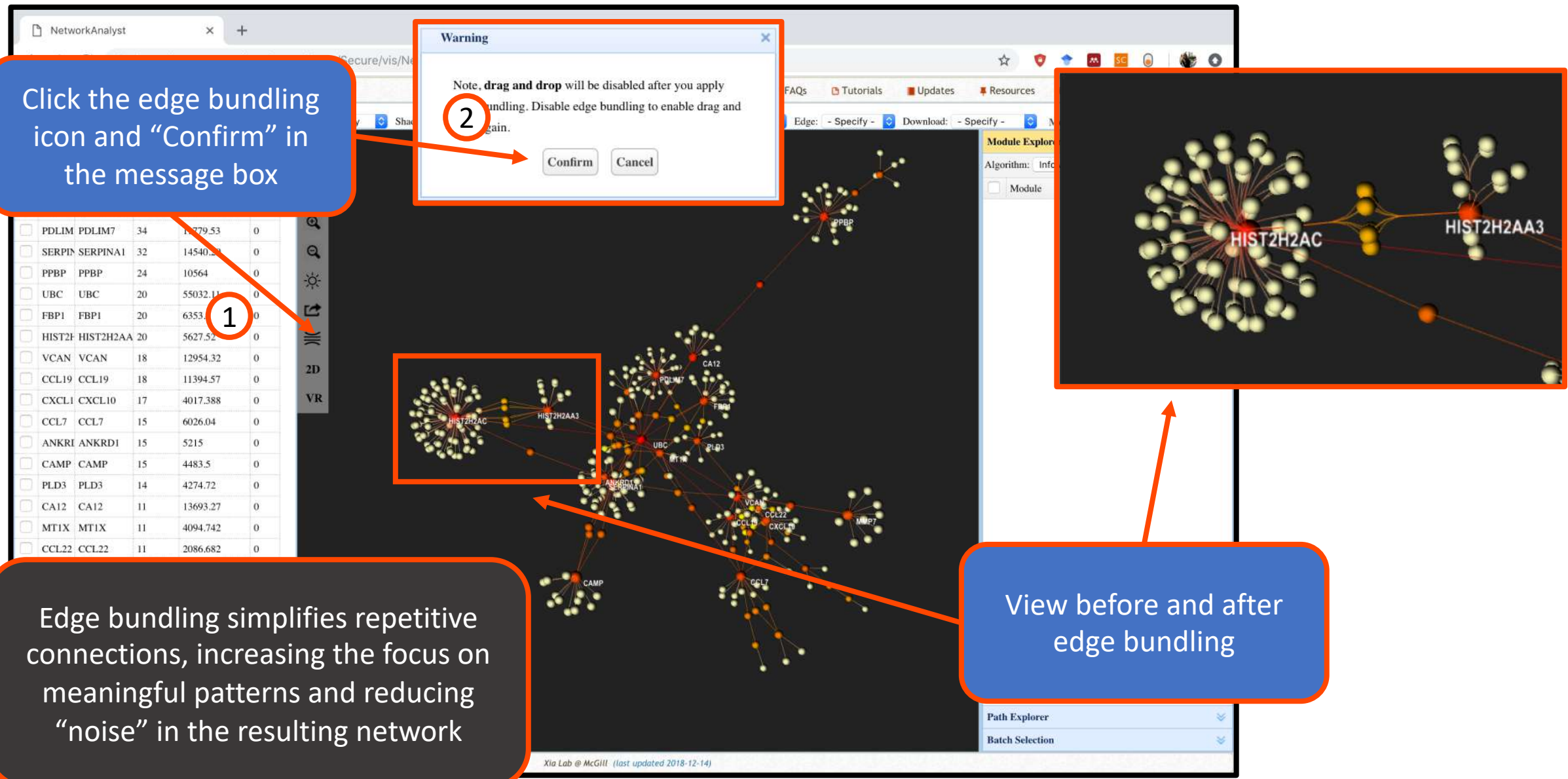
Note, drag and drop will be disabled after you apply edge bundling. Disable edge bundling to enable drag and drop again.

Confirm Cancel

1

Edge bundling simplifies repetitive connections, increasing the focus on meaningful patterns and reducing "noise" in the resulting network

View before and after edge bundling



Highlight modules in a 3D network

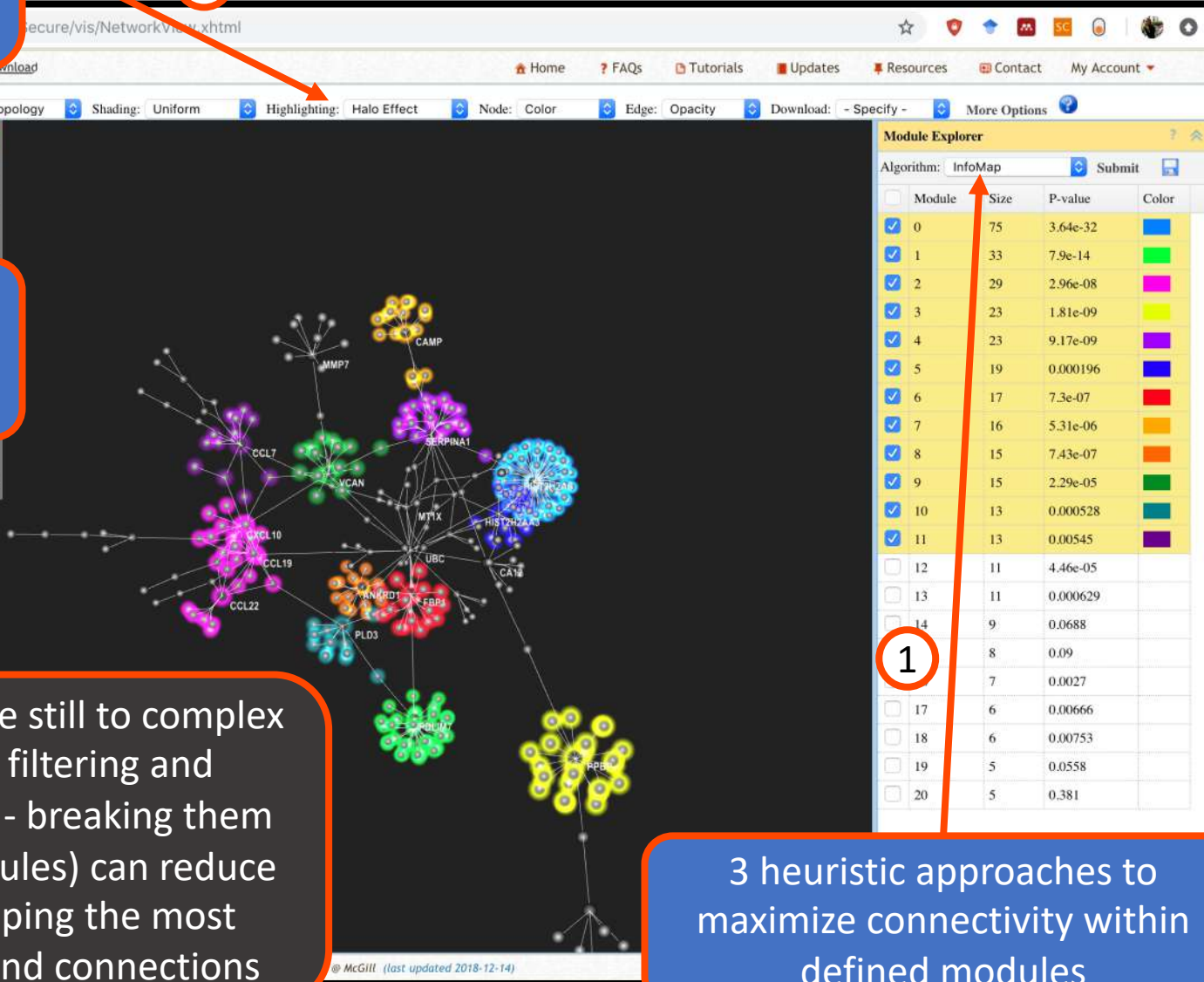
Halo effect highlighting is particularly effective at drawing attention to modules in 3D

2

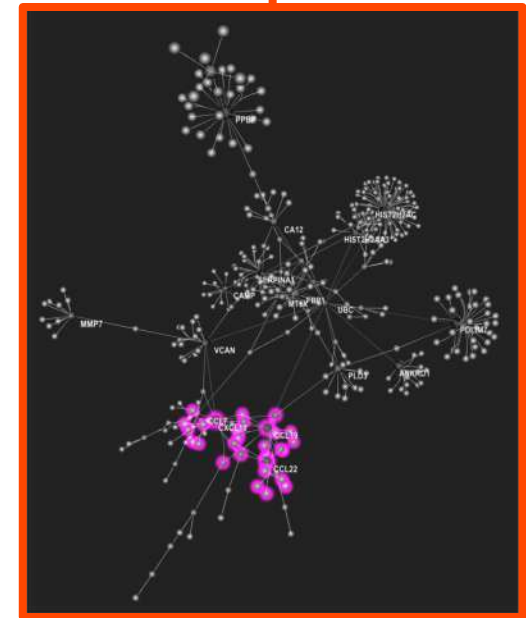
3

ID	Name	Degree	Centrality	Expr.
HIST2F	HIST2H2AC	82	26294.42	0
PDLIM	PDLIM7	34	4270.53	0

Change the highlighting color for each module

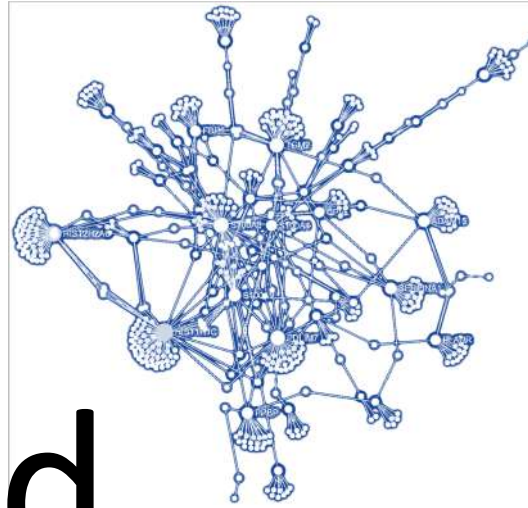


Highlight only one module and rotate to emphasize an area of interest



Sometimes networks are still too complex to understand after filtering and adjusting visual factors - breaking them into smaller units (modules) can reduce complexity while keeping the most interesting functions and connections

3 heuristic approaches to maximize connectivity within defined modules



The End

*For more information, visit the **FAQs, Tutorials, Resources,**
and **Contact** pages on www.networkanalyst.ca*