

Zoomgraph version 0.3 (alpha)

Eytan Adar & Joshua Tyler

1. Zoomgraph features

This tool is:

- A zoomable interface to large graphs. Zoomable means you can smoothly zoom in and out and easily move between nodes. It's not exactly what you would expect and is a little hard to explain, you'll have to try it.
- A database driven system. Nodes and edges have features that you can query and use to control what gets displayed (e.g. show all the nodes and interactions for yeast genes that have a metabolic function, show all the email communications between two departments).
- Ability to save state and to smoothly morph between states (new in .3)
- Facility to describe and manage subgraphs
- Robust language for selecting and managing nodes and edges
- Writes out SVG and jpeg
- A few layout algorithms
- Talks to R (some functions, but it's easy to add more)

This tool isn't:

- A replacement to UCINET, Pajek, whatever else you may be using. It has some overlapping features, but really it's meant to be an additional resource.
- Anywhere near perfect. If we had to do it again there are lots of things we would have done differently (written a parser, thought the language through, better object model, etc.)
- Complete. There are certainly features we would want to add or extend to make this truly useful (labeling on saved files, complete R integration, etc.).

2. Getting Started

Installation

You're going to need 3 things:

- The Java runtime (version 1.4). You can get the SDK, but all you really need is the JRE. <http://java.sun.com>
- The Zoomable class file and sample files. You can download the zip file at: <http://www.hpl.hp.com/shl/projects/graphs/> and unzip it somewhere.
- A bunch of 3rd party class files. I've packaged them all up for you at the URL above. Please read the various copyright notices.

Specifically, you're getting the Colt libraries (used for some of the visualization), HSQLDB an embedded database, Rserve (modified the included jar so that the classes sit in a package), the Batik libraries for SVG output (put all the separate jar files into one), and finally the Jazz visualization library (also modified this slightly because of some thread safety issues. The modified copies just catch exceptions a little better). Put those jar files in the directory where you put the zoomable jar.

Zoomgraph is launched from the command line, so start up a command prompt. You're going to want to set the CLASSPATH so that it includes all the JAR files. We've included a file called setvarbs.bat which you can run to set everything for you.

Running

To build a database you can run:

```
java com.hp.hpl.zoomgraph.DBServer db_name db_definition_file
```

To run the browser do:

```
java com.hp.hpl.zoomgraph.ZoomableGraph db db_name
```

There are a few other ways to run zoomgraph. You can type:

```
java com.hp.hpl.zoomgraph.ZoomableGraph db edge_def_file
```

This will take the edge definition file: edge_def_file (see Section 5 for formatting instructions) and will generate a temporary database (overwriting the previous temporary database). You can also work without a database by doing:

```
java com.hp.hpl.zoomgraph.ZoomableGraph edge_def_file
```

Finally, you can also run zoomgraph as an applet (see Section 5).

Tutorial

Let's start with a simple example. There is a sample database (sample.database) in the zip file. It includes about 400 nodes and 700 edges. Take a look at it to get a sense of what goes into a data definition file. But don't get intimidated, almost none of it is required.

After setting your classpath (see above), type the following to transform the text file to an actual SQL server:

```
java com.hp.hpl.zoomgraph.DBServer sample sample.database
```

then do:

```
java com.hp.hpl.zoomgraph.ZoomableGraph db sample
```

You'll see something that looks like Figure 1. (If you ever find yourself running into an out of memory error you can try upping the stack size by using the `-Xmx` and `-Xms` commands, e.g. `java -Xmx300m -Xms128 com.hp.hpl.zoomgraph.ZoomableGraph db sample`).

The graph that popped up represents a corporate communication network. Each node represents an employee (with a department property), and each edge represents communication between two employees (with a frequency property on the edge indicating the number of undirected communications).

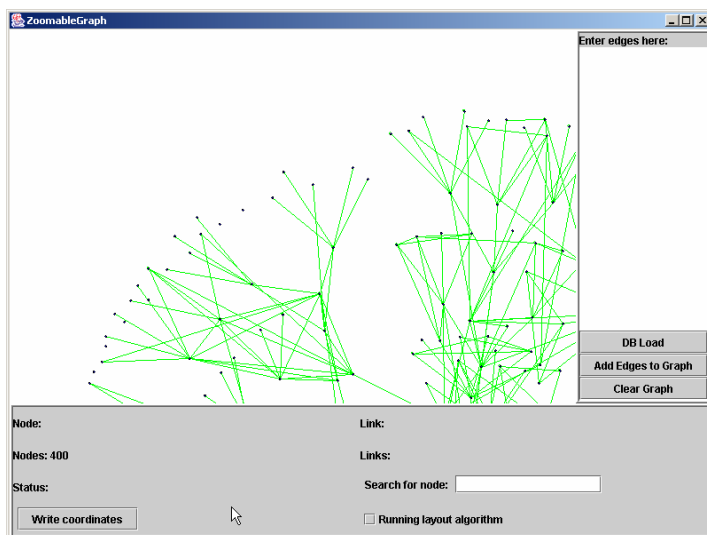


Figure 1

Try moving around in this space. If you hover over a node or edge you can see some details pop up. If you click on the node it will center in the display. Clicking on an edge will bring both end points into view dynamically. Left clicking and dragging on the background will allow you to move the display around. Right clicking and moving the mouse will zoom you in and out of the display.

Type "node1" into the search box and hit enter. The display will automatically shift to center on that node.

Ok, now back in the command prompt where you started zoomgraph you should see a prompt that looks like this ">." Here you can type whatever commands you want to manipulate the graph. Type "center" and hit enter. The display will automatically center to include all the nodes. Your display should look like Figure 2. (note: type quit at any time to exit or just close the display window... don't ctrl-c as you may corrupt your database).

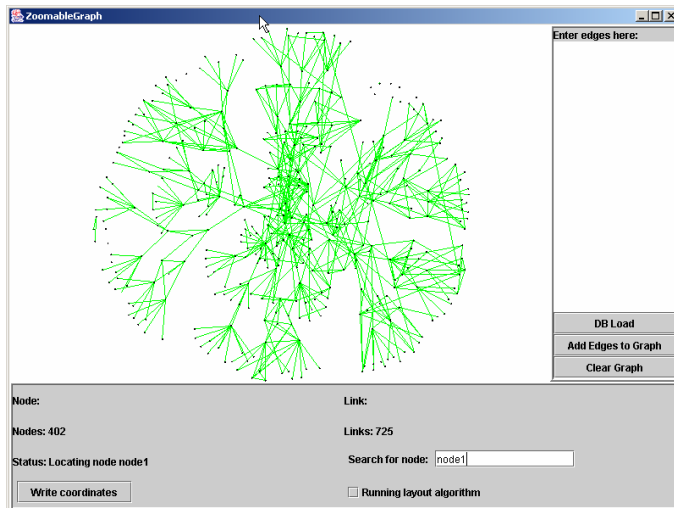


Figure 2

Let's make the nodes a little bigger so we can see them a little better. Type "nodesize 10" to make all nodes 10 pixels. If things don't immediately change on the display for any of these commands just type "redraw."

Nodes can either be selected by name or through a SQL query. For example, try typing: "nodecolor red node5,node6" This will make nodes 5 and 6 red. Our sample database has other properties on nodes. Specifically, nodes here have a department. To select nodes by a SQL query you can just type what you would after the WHERE clause. For example, "nodecolor black dept = 'dept5'" will set all the people in department 5 to a black color. Some commands just assume you want all the nodes if you don't enter a list, otherwise the character '*' means all nodes (or edges if it's an edge command).

Edges are accessed in a slightly different way. Edges have names that are the start and end nodes. For example, "edgecolor red node67-node76" changes the edge between person 67 and 76 to red. You can also access edges by SQL queries. As mentioned earlier, edge in this case have an attribute called freq (frequency). So if we wanted to hide edges where the communication frequency was under 100 we would type: "hideedges freq < 100" The '-' also implies directionality. If the database indicated directions

(which this one doesn't) you could talk about node67->node76, node67<-node76, or node67<->node76.

The last mechanism for accessing edges is by defining node sets. Let's say we only care about communications between dept 4 and 9. Let's hide everything: `hideall`. Then show only the nodes in departments 4 or 9: `shownodes dept = 'dept4' OR dept = 'dept9'`. Finally, we can change the color for inter-departmental edges by typing: `"edgecolor red {dept = 'dept9'}-{dept like 'dept4'}"` This command tells the Zoomgraph to find all nodes in dept 4 and all nodes in department 9, and through some SQL magic that goes on in the background it will find all edges between them (in this case only one). We can also do `"edgecolor blue {dept = 'dept9'}-{dept = 'dept9'}"` to just get intra-departmental links blue. You should see something like Figure 3.

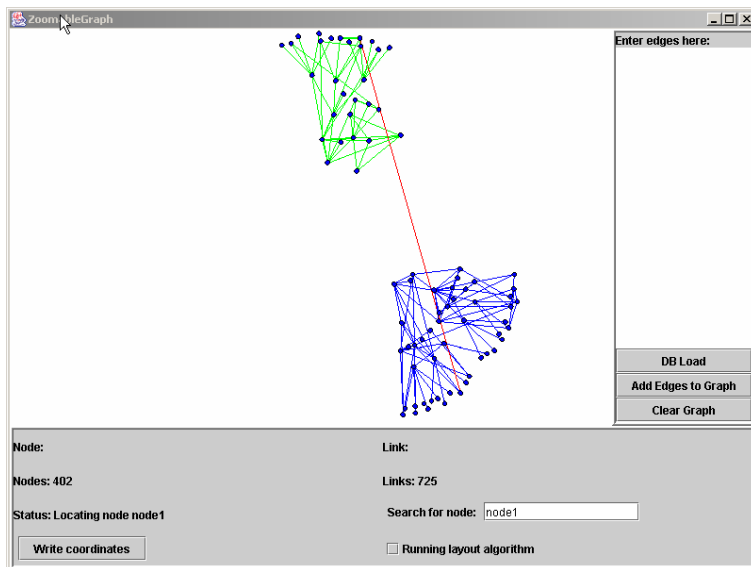


Figure 3

The Zoomable system also contains a number of analysis modules to simply basic tasks (calculating graph metrics, etc.) These are described in much more detail elsewhere, but just to give you a flavor try this... First, reset the graph to it's starting state. Type: `showall`. Then type: `edgecolor green` and finally `nodecolor blue` (you should see the same thing as before). We do this because when nodes and edges invisible it is the same as if they weren't there and are not counted in various calculations. Type `"analysis density *"` This should calculate the density of the graph (.00827...).

Other analysis modules do more interesting things. For example, `colorize` will color nodes and edges by different features. Try typing `"analysis colorize dept."` Each node will now be a different (random) color. The

colorize function will also generate a bunch of subgraphs. Then try "analysis colorize freq [table=edges,linear=true]" which will assign a color over a linear range (from blue to red) based on the frequency of communications.

Subgraphs let us bundle a related edges and nodes together and give them a name. You can type "sg list" to see a list of named subgraphs. Running the colorize command gave us a different subgraph for each department and one main one "AUTO_nodes_dept_all" that holds all the subgraphs (yes, subgraphs can be nested). Operations that work on nodes and edges will also work on named subgraphs in the same way. You can type "hidenodes AUTO_nodes_dept_dept1" to hide all the nodes in that subgraph. "sg details subgraphname" will tell you which nodes and edges and nested subgraphs are part of the named subgraph. Nested subgraphs are referenceable by using a .* at the end of the subgraph name. For example "hideall AUTO_nodes_dept_all.*" will hide each individual dept subgraph contained in AUTO_nodes_dept_all (this command isn't all that interesting since you could do without the .*, but try "sg details AUTO_nodes_dept_all.*" and "sg details AUTO_nodes_dept_all" to see the difference).

Our last example that combines everything and adds some new twist. Let's visualize all the pairwise connections between departments, one at a time. We can do this by using the foreach loop. The foreach loop takes two arguments a variable and a set. For each element in the set, the variable will be set to that element and the loop will be done. So try this:

```
> foreach sg1 AUTO_nodes_dept_all.*
  > foreach sg2 AUTO_nodes_dept_all.*
    > hideall
    > shownodes sg1
    > shownodes sg2
    > hideedges *
    > showedges {sg1}-{sg2}
    > redraw
    > pause 2000
    > .
  > .
```

What we're doing is looping over all the colorized subsets. The variables "sg1" and "sg2" are set by the foreach loop. The system knows that you're talking about subgraphs and not nodes or edges because of the naming convention. If you wanted all the edges in the subgraph you would do "foreach edge1 AUTO_nodes_dept_all" similarly all nodes are "foreach node1 ..."

States & Movies

As of release .3 of Zoomgraph, we are supporting "states" and the smooth transitions between them (caution: this is still a little rough). Here's a simple example:

1. Let's reset all the nodes/edges... do "showall" followed by "nodecolor blue" and "edgecolor green" You should hopefully have something that looks like the view you started from.
2. We're going to call this state 0 and save it... so type "savestate 0"
3. Let's make some changes... type "analysis concom [color=true]" which will color each connected component differently and then "layout spring" which will re-layout the display.
4. Let's call this state 1... save it by typing "savestate 1"
5. Now we can switch back and forth between states... simply type "loadstate 0" to switch back to the first view and "loadstate 1" to go back.
6. Next trick... the morph command let's us smoothly transition between various states. Go to state 1 by typing "loadstate 1" and then try "morph 0 1 0" which will morph the display from the current display to state 0 and then to 1 and back to 0 (you can write down any set of states you want here).
7. If your computer is slow you may find that this process doesn't look great with a lot of nodes (a little jerky). You can get a better display by freezing the on screen rendering process and saving it out to a movie. Go back to state 1 ("loadstate 1"). Start a movie by typing "startmovie test.mov" which will create a QuickTime movie called "test.mov" in your current directory. Freeze the display by typing "freeze" and then try to morph again ("morph 0 1 0"). Type "stopmovie" to save the movie and "unfreeze" to get the display back. You can then watch the movie in QuickTime or the Java Media Framework viewer. If the edges look a little jagged for you try typing "rq high" before starting the movie next time so it will render in slightly higher quality.

Once you start a movie anything you do to the display will be saved out (moving nodes by hand, changing the display, whatever...). The morph and movie commands support a great many features (tracking nodes with the camera, how fast things go, controlling which frames get saved, etc.). You should look at the documentation for these commands below.

Listing Objects

Most of the commands you'll see below allow you to enter a list of objects that you are applying the command to. The tutorial gave you some

examples of how to call graphs and edges, but a quick summary may be helpful.

Nodes can be referred to by their name.

- You can submit a list of comma delimited nodes when you have the option to list. For example "node5,node6,node7" is a valid list.
- You may also refer to nodes by means of a SQL like query acting on node properties. For example you may pass "size < 50" as the list to something acting on nodes (this will select nodes that are under 50 pixels in size). SQL queries can be complex. For example: "size < 50 OR (xloc > 500 AND yloc < 400)"
- You cannot mix SQL style selects and standard node names in a list. If you want to be clever you could add the nodes to a subgraph, generate a node view (see commands), and then apply the SQL query to the view.
- * refers to all nodes
- If you have a subgraph defined with nodes in it, any node command applied to that subgraph will act on the nodes held within the subgraph.

Edges are more complex.

- If you had an edge between node5 and node6 you could refer to it as node5-node6 or node6-node5. If you have multiple edges between node5 and node6, edges also have a numerical id property that you can use as the name. You can list edges the same was as nodes: "node5-node6,node8-node9"
- If edges are directed you replace the "-" character by "->" "<->" or "<-" For example, "node5->node6,node8<->node9" is a valid way to refer to the directed link between node 5 and 6, and any link between node8 and node9.
- Finally, you can reference edges through a set-to-set operation. A node set is defined within the curly braces..."{}" Inside you can place any node list as described above. This can either be a list of nodes or a SQL selection (or subgraphs). For example, you could say "{xloc<500 AND yloc < 500}-{xloc > 500 AND YLOC > 500}" to get all the edges going from one nodes in one quadrant to nodes in the other.
- * refers to all nodes
- As before, if a subgraph contains edges, any edge command applied to that subgraph will act on edges within the subgraph.

3. Getting Your Data in: DBServer

The DBServer program will initialize your database for you. It takes as input a database name and a database description file (see above on the command line). The description file has two components a list of nodes and a list of edges.

The node definition section starts with the line: "nodedef> name ..."

The only required column for nodes is a name, which needs to be a string. So a valid node definition line would be: "nodedef> name VARCHAR(256)" After that you can simply put down a list of nodes, one per line that match the specification in the nodedef line (in this case all you need is a name). Similarly, edges are defined through an "edgedef>" line. An edgedef line must define two columns n1 and n2 (the start and end nodes of the edge). A valid database description file is then:

```
nodedef> name VARCHAR(256)
A
B
C
edgedef> n1 VARCHAR(256),n2 VARCHAR(256)
A,B
B,C
A,C
```

Which basically represents an undirected graph with three nodes and three edges.

In addition to the required columns there are a number of optional ones for both nodes and edges. These are created for you by the DBServer if you don't do it yourself. Note that the def describes what comes next. You have to have the same number of columns in each of your node and edge lines as you did in your definition lines.

You may choose different defaults, but try to use the same types – DOUBLE, INT, etc. that are described here. The Zoomgraph makes certain assumptions about what's in a database. Node definition lines may include:

- XLOC DOUBLE DEFAULT 500 – which is the x location of the node
- YLOC DOUBLE DEFAULT 500 – the y location of the node
- VIS BIT DEFAULT TRUE – is the node visible (true = yes)
- COLOR VARCHAR(32) – the color of the node... be careful about using r,g,b values since the comma delimiter may break the database load
- FIXED BIT DEFAULT FALSE – is the node fixed in place (true = yes)

- SHAPE TINYINT DEFAULT 0 – the shape of the node, currently 0 = circle, 1 = square. May add more later. You won't be able to change this dynamically right due to constraints of the visualization.
- SIZE DOUBLE DEFAULT 2 – the size of the node

Edges can have the following properties:

- VIS BIT DEFAULT TRUE – is the edge visible (true = yes)
- COLOR VARCHAR(32) – the color of the node... be careful about using r,g,b values since the comma delimiter may break the database load
- WIDTH DOUBLE DEFAULT .3 – the width of the edge
- WEIGHT DOUBLE DEFAULT 1 – the weight of the edge
- DIRECTED TINYINT DEFAULT 0 – the direction of the edge. This doesn't work all that well yet. 0 means undirected, 2 means "n1->n2", 3 means "n1<-n2", and 4 means "n1<->n2"

Beyond these basics everything is fair game. You can add whatever columns and properties you want and then use them to control your visualization. For example, let's extend our basic definition above to indicate node size and a new column called city, and edges will have a number representing the number of planes (totally fake):

```

nodedef> name VARCHAR(256), SIZE DOUBLE DEFAULT 2, CITY VARCHAR(256)
A,10,new york
B,6,boston
C,4,san jose
edgedef> n1 VARCHAR(256),n2 VARCHAR(256),ROUTES INT DEFAULT 0
A,B,40
B,C,30
A,C,20

```

4. Commands

Display Commands

`center <list>`

Centers the display to include all nodes (if no argument is given), or only those that are in the list. This will only center on visible nodes.

Note: You may notice an exception sometimes when you do this even though the display does the right thing. There seems to be some race condition in Jazz.

`centerall <list>`

Centers the display to include nodes (all or in list), visible or not.

freeze

Freezes the display. Changes will not appear on the screen until you unfreeze it.

unfreeze

Unfreezes the display

rq <low|medium|high>

Sets the render quality of the visualization to one of three states. If you just type "rq" it will tell you what the current state is. The quality only applies to a display that has stopped changing. Moving around may cause the display to shift to a different rendering quality (see rqi). Default: low

rqi <low|medium|high>

Same as rq, but sets the interactive state. This sets the rendering quality when the display is changing. Default is also low. Changing this may degrade performance in display related features since rendering will take longer.

background color

Sets the background color to color (see section 5 for more information about colors).

redraw

Re-renders the display. Sometimes this is necessary to get the display to sync up with certain commands.

iw+

opens the information window. As you mouse over nodes and edges the information window reflects details about the objects. The information is the same as the node/edge details command.

iw-

hides the information window

hullson

Turns on the convex hull rendering for subgraphs on a global scale. Default: off

hullsoff

Turns off the convex hull rendering for subgraphs on a global scale. Default: off

refresh
Redraws the screen

General Graph Commands

hideall <list>
Hides all graph objects if no argument is specified. Otherwise only hides the objects in the list. This can be a mix of edges and nodes.

hideall <list>
Shows all graph objects if no argument is specified. Otherwise only shows the objects in the list. This can be a mix of edges and nodes.

muteall <list>
Mutes all graph objects if no argument is specified. Muted nodes are shown in the muted color (default gray). If a list is supplied only mutes the objects in the list. This can be a mix of edges and nodes.

unmuteall <list>
Unmutes all graph objects if no argument is specified. If a list is supplied only unmutes the objects in the list. This can be a mix of edges and nodes.

mutecolor color
Sets the muted color for muted nodes/edges

directed
Sets the graph mode to directed. If your edges are undirected you won't see anything different. If there is a direction, you should see some arrows (this only sort of works)

undirected
Sets the graph mode to undirected.

commitoff
Turns off database commits. Changes made to nodes/edges will not be committed to the database. Default: on

commiton
Turns on database commits. Changes made to nodes/edges will be committed to the database. Default: on

Output/Scripting Commands

`savejpg file_name`

Outputs the current visualization to the specified jpeg file

`savesvg file_name`

Outputs the current visualization in svg format to the specified file

`savelog log_name`

All commands that are typed will be saved to the log specified by `log_name`. Using this you can save what you do and re-run or modify it later to make macros.

`loadlog log_name`

All commands in the log file (`log_name`) will be executed against the current environment. This can be used to reply macros that you have saved earlier.

`stoplog`

Stops logging commands

`savecsv file_name SQL_query`

Will generate a file (specified by file name) in CSV format for the database columns you are interested in

Example: `"savecsv foo.csv select name,degree from nodes"`

Will save the name and degree columns to the file `foo.csv`

`exppajek filename [nodes=list,edges=list]`

A basic export of a graph to a Pajek file. The default behavior is to save out all nodes and edges that are visible. To change this you can specify a query (this is a little broken).

`pause milliseconds`

Pauses the system for some number of milliseconds. This is useful when running a script and you don't want the display to refresh too quickly.

`foreach variable list`

The foreach loop lets you loop over a list of objects. The variable naming convention is to start the name with "node" if you want the nodes, "edge" if you want the edges, and "sg" if you want the subgraphs. So you could do: `"foreach node1 somesubgraph"` to talk about the nodes in the subgraph or `"foreach edgefoo somesubgraph"` to references the edges. Any instance of the variable name gets replaced by an element in the list each time the foreach loops. To close a foreach statement just type a period (".") on it's own line.

Node Commands

node details <list>

Gives you some details about the node (same as you would see in the information window).

hidenode(s) <list>

Hides all nodes if no argument is specified. Otherwise only hides the nodes in the list.

shownode(s) <list>

Shows all nodes if no argument is specified. Otherwise only shows the nodes in the list.

mutenode(s) <list>

Mutes all nodes if no argument is specified. Otherwise only mutes the nodes in the list.

unmutenode(s) <list>

Unmutes all nodes if no argument is specified. Otherwise only unmutes the nodes in the list.

fixnode(s) <list>

Fixes all nodes if no argument is specified. Otherwise only fixes the nodes in the list. Fixed nodes will not be moved by the layout algorithm.

unfixnode(s) <list>

Unfixes all nodes if no argument is specified. Otherwise only unfixes the nodes in the list.

nodecolor color <list>

Sets the node color of all nodes (if no argument is given) or just those listed.

hidedis <list>

Hides nodes (all if no argument, or from the subset of list) that do not have any visible edge going to them.

nodesize size <list>

Sets the node size of nodes (or all if no argument) to size

labelnodes <list>

Shows a node label next to the nodes in the list

unlabelnodes <list>
Hides node labels.

Edge Commands

edge details <list>
Gives you some details about the edge (same as you would see in the information window).

edgecolor color <list>
Sets the edge color of edges (or all if no argument) to color. Sometimes this requires an explicit redraw to be called.

edgewidth width <list>
Sets the edge width of edges (or all if no argument) to width. Sometimes this requires an explicit redraw to be called.

hideedge(s) <list>
Hides all edges if no argument is specified. Otherwise only hides the edges in the list.

showedges(s) <list>
Shows all edges if no argument is specified. Otherwise only shows the edges in the list.

muteedges(s) <list>
Mutes all edges if no argument is specified. Otherwise only mutes the edges in the list.

unmuteedge(s) <list>
Unmutes all edges if no argument is specified. Otherwise only unmutes the edges in the list.

unkink
Some layout routines cause bends in the edges, you'll want to run this to restore the straight line mode.

edgeaverage <list>
Sets the edge color of one or more edges named in the list to the average of the two nodes connecting that the edge connects

Subgraph Commands

Subgraphs are hierarchically structured objects that contain nodes, edges, and references to other subgraphs. Figure 4 is a representation of three subgraphs, sgA, sgB, and sgC. The subgraph sgA directly contains nodes A and B, and edges A-B. It also contains a pointer to subgraphs sgB and sgC. All commands and lookups applied to sgA will be recursively applied to these embedded subgraphs. For example, setting the node color to red for sgA will cause not only nodes A and B to become red but also nodes D and E. This is usually the expected thing, but be careful when applying operations like delete as they will be applied to lower level subgraphs as well. For example deleting node D from sgA will actually cause its removal from sgB. The commands for subgraph manipulations are as follows:

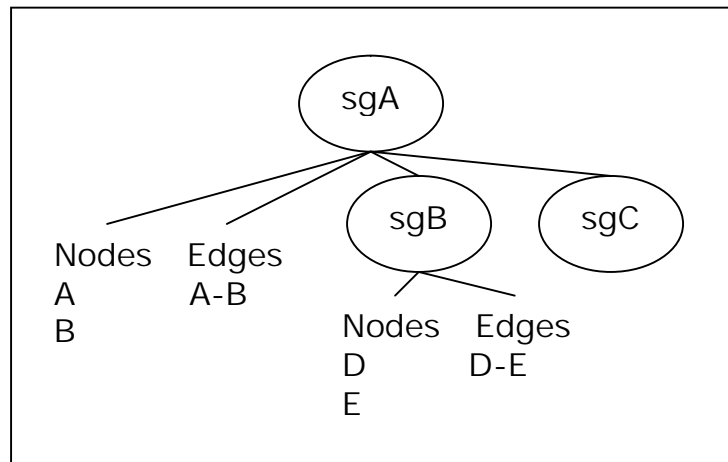


Figure 4

`sg subgraph_command <optional arguments>`

Performs a subgraph command

- `list` – lists all the named subgraphs
- `details name` – gives information about the subgraph named `name`
- `save filename` – saves the subgraph definition to `filename`. Section 5 describes the file format for the subgraph files.
- `load filename` – loads the subgraph definition from `filename`
- `create newname` – creates a subgraph with the name “`newname`”
- `delete <list>` – deletes the named subgraphs. If no argument is given it will happily delete all the subgraphs from memory (so be careful).
- `showhull <list>` - shows a convex hull around the nodes for the subgraphs named in the list. This will also turn on hull display if it is off
- `hidehull <list>` - hides the convex hull around the nodes for the named subgraphs
- `hullcolor sname color` – sets the color for the convex hull for the named subgraph

- `nview name` – creates a database view including only the nodes in the subgraph named “name.” This is useful if you want to do some heavy duty SQL manipulation.
- `eview name` – creates a database view including only the edges in the subgraph named “name” Useful for SQL manipulation.
- `dropnview name` – drops the node view for the given subgraph. You should have called `nview` before calling this.
- `dropeview name` – drops the edge view for the given subgraph. You should have called `eview` before calling this.
- `addnode(s) name list` – adds the listed nodes to the named subgraph. If the node is already contained in either the top level subgraph or any of it’s contained subgraphs the node will not be added.
- `addedge(s) name list` – adds the listed edges to the named subgraph. If the edge is already contained or is nested it will not be added. Adding an edge also causes the end points to be added.
- `addsg(s) name list` – adds the listed subgraphs to the named subgraph. Use caution in applying this command. The system does not check for cycles in subgraphs. So if subgraph A contains B, and B contains A, certain operations will freeze the system.
- `rmnode(s) name list` – remove the listed nodes to the named subgraph. It will do a recursive removal, so if the node is contained in a contained subgraph it will be removed from that lower level subgraph.
- `rmedge(s) name list` – removes the listed edges to the named subgraph. Same rules as `rmnodes`.
- `rmedge(s)+ name list` – removes the listed edges to the named subgraph and removes any nodes that are now disconnected.
- `rmsg(s) name list` – removes the listed subgraphs to the named subgraph.
- `copy name1 name2` – copies the contents of subgraph `name1` to subgraph `name2`.

Analysis Commands

`analysis analysis_command <optional arguments>`

Performs an `analysis_command`. Valid commands include:

- `concom [hullcolor=true/false, color=true/false,hullalpha=0-255]`– Figures out all the connected components of the network. It will generate a subgraph called `cc_all`. Each connected component becomes a subgraph of `cc_all`. It will be named `cc_node_size`, where

node is the name of some node in that connected component and size is the size of the connected component. If the color variable is set to true (default false) each node in connected component will be colored in the same color. If hullcolor is set to true (default false) a convex hull will be drawn around the nodes in the same color as the nodes. The hullalpha option sets the transparency of the hulls (default 127).

- `colorize column [table=nodes*/edges, random=true*/false, linear=true/false*, startcolor=color, endcolor=color, hullcolor=true/false*, hullalpha=0-255]` - * options are defaults
The `colorize` command will generate a coloring on either the nodes or edges table depending on some property (the column). The number of colors it will pick depend on the range of the column. For example, if column is a boolean column (true/false), there will be two classes and therefore two colors. Random coloring will assign a random color to each class. Linear (blue to red) usually only makes sense when the column is a number. For example if you had an edge weight column you could do "`colorize weight [table=edge,linear=true]`" and have low weight edges be blue, and high weight edges be red. This algorithm will also generate a set of subgraphs. The subgraph `AUTO_table_column_all` will hold all the subgraphs generated where each subgraph represents a different color coded class. As in the `ConCom` example, the `hullcolor` and `hullalpha` configure if and how hulls are drawn around the subgraphs.
- `resize column [table=nodes*/edges, random=true*/false, linear=true/false*, startsize=size, endsize=size, hullcolor=true/false*, hullalpha=0-255]` - * options are defaults
Same as the `colorize` command but will change the sizes of nodes or widths of edges depending on the column selected.
- `clustCoef <list>` – calculates the clustering coefficient on the list of nodes
- `density <list>` – calculates the density on the list of nodes
- `diameter <list>` – calculates the diameter on the list of nodes
- `avgPathLength <list>` – calculates the average path length on the list of nodes (really slow for long lists)
- `symmetry <list>` – calculates the symmetry on the list of nodes
- `rservehost hostname` – sets the name of the machine hosting R for doing R based analysis. (see instructions for R below)
- `jdegree <list> [column=name]`- calculates the node degree of the listed nodes. If the optional column variable is set the result will be placed in a column with that name in the node database.

The following are R commands (and so require `rserve` to be running and you connected to it). It then uses the `SNA` library for calculating various

centrality measures. The results will be placed in a new column corresponding to the name. Syntax is still "analysis command_name <list>." See the R SNA documentation for a full explanation of these commands.

- `betweenness <list>`
- `nodedegree <list>`
- `bonpow <list>`
- `evcent <list>`
- `graphcent <list>`
- `infocent <list>`
- `prestige <list>`
- `closeness <list>`

Layout Commands

`layout layout_command <extra>`

These are various layout commands. Nodes that are fixed will not be moved.

- `fruch [height=num,width=num, random=true*/false, rescale=true*/false]` – Fruchmen-Reingold layout. You can set the height and width (default 1000,1000). The default behavior is to randomize the layout first. If you would prefer not to do that, set `random` to `false` (to prevent). Rescaling causes nodes to get "blow up" to fill the area allocated for layout (some layout algorithms have a tendency to pull the nodes in)
- `random [height=num,width=num]` – random layout, height and width as above
- `circular [height=num,width=num]` – regular circular layout, height and width as above
- `kamada [height=num,width=num, random=true*/false, rescale=true*/false]` – Kamada-Kawai layout. Optional arguments as above.
- `hierarchy [height=num,width=num, column=nodecolumn name, layers=x]` – still in alpha. It will layer nodes depending on the value of `column` into `num` layers. You'll want to run `unkink` when you're done with this.
- `spring [height=num,width=num,loops=num]` – spring layout, height and width as above. The `loops` arguments sets how many times to iterate (default 50).
- `binpack subgraphs [height=num,width=num]` – to get a slightly nicer layout with multiple overlapping components you can use the `binpack` command which operates on subgraphs. To do this first get a

set of connected components ("analysis concom") and then try
"binpack cc_all.*"

- layout rescale [height=num,width=num] – blow up the nodes to fit in the area.
- move <list> [x=num,y=num] – moves the listed nodes so that the upper left most node is at the x,y coordinate specified.
- rotate <list> – not exactly a rotate but it sets the x coordinate to the y and vice versa for each node in the list.

Movie & State Commands

savestate statename

Saves the current state of the graph to the state "statename" (which can be any set of alphanumeric characters)

loadstate statename

Loads the state statename and changes the displayed nodes and edges (not subgraphs yet!) to the values of the saved state.

morph statename <more states> [localtime=time, totaltime=time, camera=true/false*, tofollow= nodelist, estart=percent, eend=percent, nstart=percent, nend=percent]

Smoothly transitions the display from the current state to statename moving edges, nodes, and changing colors and visibility. You can list any number of states. Each transition takes 3 seconds by default. Advanced options are:

- localtime – the time to spend on each state transition (default 3000)
- totaltime – the time to spend in total on all transitions (this overrides localtime and each state change will receive $1/n^{\text{th}}$ of the total time where n = the number of states)
- camera – if you want the camera to follow the certain nodes set this to true. If you don't set the tofollow variable it will keep all the nodes centered.
- tofollow – you can list a set of nodes or a subgraph you want the camera to follow around. If you're entering a list of nodes place quotes around the comma separated list (e.g. morph 0 1 [camera=true,tofollow='node1,node2']). Sometimes the animation is too fast and the nodes fly off the screen... the camera will try to catch up but it makes for some jerky transitions sometimes. Playing with the time allocated for transition helps. You can also turn on node labeling for the nodes you're trying to follow so you can see where they go.
- estart/eend and nstart/nend – when edges and nodes appear and disappear between states you sometimes have too many things happening at once. The estart/eend and nstart/nend commands let

you control when things appear and disappear. The default is 0/1 for starting/ending but you can change these to somewhere between 0 and 1 (say .8/1) which will cause the edge appearance/disappearance to be delayed until the animation reaches 80% of completion.

`startmovie filename [fps=num, auto=true*/false]`

Tells Zoomgraph to start saving the display to a (QuickTime) movie file. The fps option allows you to control the frames per second. Playing with this as well as the morph times can make for much smoother animations. The auto option (by default true) tells Zoomgraph if you want all display changes automatically to be saved.

`saveframe [loop=count,pause=true/false*]`

Saves the current view as a frame (you can only do this if you started a movie in the non-auto mode). The current view will only be one frame unless you set the loop to something else (e.g. if loop=5 the current display will be saved as 5 frames). You can also force Zoomgraph to pause between each frame renderings to allow certain repaint commands to happen.

`stopmovie`

Stops recording the movie.

Power User Commands

Backdoor

Typing "backdoor" puts you in direct contact with the SQL server. Your prompt will change to "b>" and any commands typed after that will be directly routed as SQL. You should be careful if you choose to use this. The results of select commands will be dumped to the screen. Typing backdoor again when in this mode will put you back in the regular mode. Typing quit will exit zoomgraph not just the backdoor mode.

`select/SELECT sql_stuff`

Does a SQL select and dumps the results.

`CREATE VIEW sql_stuff`

If you're into doing some heavy-duty table merges and selects you can use this. It will route this type of command to the SQL server. You should remember to drop this if you don't want it to be persistent.

`DROP VIEW`

Lets you drop the view you created. Routed directly to the SQL server.

temp sql_stuff

This will run a full query on the database and place the resulting nodes and/or edges into the subgraph __temp. This comes in handy if you're doing weird selects and joins. You can then copy the nodes/edges out of the temporary subgraph. For example: the command "temp SELECT * from nodes where xloc < 500" will load up the __temp with nodes that are on the left side of the display. The routine basically sees if the table you're querying on has a column called "name" which indicates that you want nodes, and/or columns named n1 and n2. You can get both through different select and view operations.

5. Additional Information

Colors

Colors are defined by a comma delimited numbers representing red, green, and blue. There can be no space between them. For example the color red is "255,0,0." There are also a number of predefined named colors: black, blue, cyan, darkGray, gray, green, lightGray, magenta, orange, pink, red, white, and yellow.

Subgraph File Format

Saved subgraphs are pretty straightforward. A subgraph is named on a line preceded by the ">" character. Any items found until the next subgraph line are taken to be either nodes, edges, or other subgraphs. Other subgraph names should be preprend by two "_" characters. For example, the following file:

```
>subgraphA
a
b
a-b
__subgraphB
>subgraphB
c
d
c-d
```

describes two subgraphs, subgraphA which contains two nodes (a and b) and one edge (a-b) as well as the nodes and edges defined by subgraphB (c,d, and c-d).

If the subgraph definition file includes an edge or node that is not defined in the database you will see an error.

Valid SQL

Wherever SQL can be used we more or less allow anything that HSQLDB allows. Please see the `hsqSyntax.html` file included in the distribution. It should give you a sense of what you can (most things) and can't do.

Edge Definition File

If you want to do things without a database or run just have Zoomgraph make you a temporary database, you can create an edge definition file. Each line is:

```
node1-node2 <node1x,node1y node2x,node2y>
```

See the files `sample` `sampleedgefile_nocoordinates.txt` or `sampleedgefile_withcoordinates.txt` as examples

Zoomable Graphs as Applets

You can run the Zoomable Graph as an applet without the database features. Look at `applet.html` for an example. You can simply paste in the edge definition file into the `INITEDGES` parameter. This is in the same format as the edge definition file above but with a semicolon at the end.

Using R-based Functions

Zoomgraph currently contains a couple of R-based functions... specifically, we pass the matrix representing the graph to R and can calculate betweenness and node degree. The values are deposited into a column of the database.

To use this, you will need to:

- Install R from <http://www.r-project.org> (I use 1.6.2),
- Install Rserve from <http://stats.math.uni-augsburg.de/Rserve/>.
- You will also need to install the SNA library (available at the CRAN site on the r-project.org page above). If you're using windows, you can just do it through the Package menu (download package from CRAN...).
- Finally, run Rserve as specified for your platform (Windows: type "Rserve", Unix: type "R CMD Rserve")

When running Zoomgraph you will need to tell it the machine on which you have R installed. You do this by typing:

```
analysis rservehost hostname
```

Where hostname is the machine you are Rserve on. If you're running Zoomgraph and R on the same machine you probably don't need to do anything since the default is "localhost."

You can then test the connection by typing: `analysis nodedegree *`

In version 0.3 we have also added a command "rmode" which will send any commands you enter after wards to R and then display the results. For example if you type "rmode" you will see that the prompt changes to "r>" Typing "2+2" will result in "4.0" Type rmode again to exit this mode. There are other things you can do here, but they're still broken in this release.