

Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms

PD Dr. Michael Hutter, Kerstin Reuter, Daria Gaidar
Wintersemester 2016/2017

Saarland University

Department of Computational Biology

Tutorial 9 12. Januar 2017

Tutoren: Christina Kiefer, Markus Dillmann

Netzwerke

Im heutigen Tutorial geht es um die Analyse von *Protein-Protein* und *Protein-DNS* Netzwerken mit Hilfe des Programms Cytoscape (<http://www.cytoscape.org>). Die Daten stammen aus den Cytoscape-Tutorials und basieren auf Experimenten zum Galaktose-Metabolismus von *S. cerevisiae* (Science 2001, 292:929-34).

Exercise 9.1: Vorbereitung

Alle für dieses Tutorial benötigten Dateien befinden sich auf unsere Webpage.

```
galFiltered.sif  
galExpData.mrna
```

Um Cytoscape zu starten, öffnen Sie ein Terminal und führen Sie diese Zeile aus:

```
/home/stud/dabiku/Tutorial_Cytoscape/v2.8.3/cytoscape.sh
```


Exercise 9.2: Einlesen und Darstellen des Netzwerks

Wir beginnen mit dem Laden der Dateien und passen die grafische Darstellung an.

- Über **File > Import > Network** wählen Sie die Datei “galFiltered.sif” aus dem Verzeichnis `/home/stud/dabiku/Tutorial_Cytoscape/data` aus und laden sie. Schliessen Sie den Bestätigungsdialog. Das Netzwerk wird mit einem (für uns hier unbrauchbaren) Gitter-Layout dargestellt.
- Maximieren Sie das Fenster mit dem Netzwerk (betitelt mit galFiltered.sif) und probieren Sie ein paar der im Layout-Menü zur Verfügung stehenden Layout-Algorithmen aus.



Wählen Sie für die weiteren Schritte das **yFiles > Organic-Layout**.

- Einzelne Knoten und Kanten können mittels Mausklick ausgewählt werden, eine Gruppe wird mit gedrückter Maustaste ausgewählt. Mit den Zoom-Icons  können Sie die Darstellung verkleinern und vergrößern, auf die ausgewählten Knoten fokussieren, bzw. auf maximal darstellbare Grösse bringen. Zu den ausgewählten Knoten, bzw. Kanten werden unterhalb des Netzwerks im Data Panel die vorhandenen Informationen angezeigt. Dazu muss dieses Panel zwischen Node-, Edge- oder Network-Modus umgeschaltet werden.
- Um weitere Informationen zu den Knoten zu erhalten, importieren Sie noch die Datei `galExpData.na` mit den “Node Attributes”.

```
File > Import > Node Attributes
```

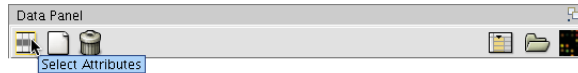


Abbildung 1: Attributauswahl

Klicken Sie im Data Panel auf das **Select Attributes**-Icon und wählen Sie **Common Name** aus (1).

Exercise 9.3: Einzelne Knoten

Suchen Sie nun im Suchfeld oberhalb des Netzwerks nach dem Knoten **YMR043W**. Der Knoten wird ausgewählt und gezoomt dargestellt. Das Data Panel sollte nun auch den *Common Name* **MCM1** anzeigen. Zoomen Sie ein Stück raus. Markieren Sie die Nachbarn dieses Knotens über **Select > Nodes > First Neighbors**. Im *Network Panel* links sehen Sie, dass 19 der insgesamt 331 Knoten ausgewählt sind.

- Wie oft können Sie weitere Nachbarn auswälen, bis alle Knoten dieses Clusters markiert sind? (Tipp: **Ctrl-6** auf der Tastatur geht schneller). Wenn Sie herauszoomen, können sie den Fortschritt am besten sehen. Wie lange ist der Pfad, wieviele Knoten befinden sich im Cluster? Merken Sie sich die zuletzt markierten Knoten.
- Markieren Sie nun einen der letzten Knoten und wiederholen Sie die Suche nach den jeweils nächsten Nachbarn. Wie lang ist nun der Pfad zwischen dem Startknoten und dem im Netz davon am weitesten entfernten Knoten?
- Das Icon rechts neben dem Suchfeld führt zu einem Einstellungsdialog. Stellen Sie das Kriterium auf **Common Name** und suchen Sie den Knoten **SIN4**. Die Knoten sind noch mit ihrer ID beschriftet. Um das zu ändern, öffnen Sie das *VizMapper*-Kontrollfeld und wählen *Common Name* für das *Node Label* aus.

Wieviele Knoten gibt es, deren *Unique Identifier* (ID) mit **YMR** beginnt?

Hinweis: Verwenden Sie zur Suche ein Wildcard ("*****").

Exercise 9.4: Expressionsdaten

Wir betrachten nun die zu diesem Netzwerk zugehörigen Expressionsdaten.

- Importieren Sie die Datei "**galExpData.mrna**" (**Import > Expression Matrix**). Bei den drei Szenarien **gal1**, **gal14** und **gal80** wurde jeweils die Expression von **GAL1**, **GAL4** oder **GAL80** gestört und die entsprechenden Expressionsänderungen gemessen.
- So, wie Sie vorher im Data Panel den *Common Name* für die Knoten zur Darstellung hinzugefügt haben, markieren Sie nun die Label **gal1RGEExp**, **gal4RGEExp** und **gal80RGEExp**. Welcher der Knoten, deren *Common Name* mit **GAL** beginnt, ist am stärksten exprimiert bzw unterdrückt, wenn **GAL80** unterdrückt wird?

Hinweis:

- Suchen wie oben, dann Liste nach **gal80RGEExp** sortieren
 - Die Tabellenspalten sind durch Klick in die Kopfzeile sortierbar.
- Um die Expression sichtbar zu machen, färben wir nun die Knoten nach ihrer Expression ein. Im *VizMapper* kopieren Sie (unter **Options**) den aktuellen Stil und nennen ihn **Gal80**. Suchen Sie **Node Color**, Doppelklick. Wählen Sie als Wert "**gal80RGEExp**" aus und als Mapping **Continuous Mapper** (evtl. auf das Plus vor **Node Color** klicken). Editieren Sie den Gradienten. Mit Doppelklick auf die vier Pfeile können Sie den Bereichen Farben zuweisen. Wählen Sie ein Farbschema wie in 2 gezeigt und schliessen Sie das Fenster.

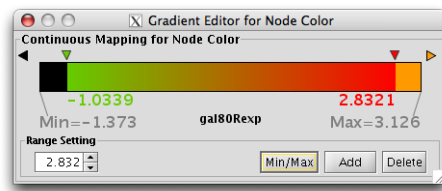


Abbildung 2: Farbschema zur Darstellung der Genexpression

- (d) Im ganzen Netz sollten drei Knoten orange gefärbt sein. Zoomen Sie zu diesen hinein. Wie heißen Sie?

Exercise 9.5: Daten filtern

Das Netzwerk besteht aus *Protein-Protein* (pp) und *Protein-DNS* Kanten (pd). Im folgenden werden nur noch die pd-Links benötigt.

- (a) Legen Sie zunächst eine Kopie des aktuellen Netzwerks an:

`File > New > Network > Clone current network`

Verifizieren Sie im *Network-Control Panel*, dass das Netzwerk aus 331 Nodes und 362 Edges besteht.

- (b) Wechseln Sie zum *Filters*-Tab und erstellen Sie über das *Options*-Popup einen neuen Filter. Nennen Sie ihn "Select pp". Wählen Sie unter *Filter Definition* `edge.interaction` aus und klicken Sie `Add`. Aus dem *Interactions*-Popup wählen sie dann "pp". Nun sollten 251 Kanten ausgewählt sein. Löschen Sie die markierten Kanten:

`Edit > Delete Selected Nodes and Edges`

Vergleichen Sie das reduzierte Netzwerk mit dem Originalen, indem Sie im *Network*-Tab abwechselnd die beiden Netze auswählen.

- (c) Um die vielen nun unverbundenen Knoten zu entfernen, erstellen Sie einen neuen *Topology Filter*, der alle Knoten mit mindestens einem Nachbarn auswählt. Nach `Apply` sind alle verbundenen Knoten ausgewählt. Erstellen Sie ein neues Netz nur mit diesen Knoten:

`File > New > Network > From Selected Nodes, all Edges`

Im *Network*-Tab erscheint dieses Netz eingerückt. Rufen Sie `Layout > yFiles > Organic` auf, um die Darstellung zu verschönern.

Exercise 9.6: Analyse des Experiments

Zoomen Sie in das Netzwerk hinein und identifizieren Sie die beiden Knoten, die mit den drei orange markierten (stark Überexprimierten) Knoten `GAL1`, `GAL7` und `GAL10` verbunden sind. Suchen Sie in der Gendatenbank von <http://www.ncbi.nlm.nih.gov/> nach `YPL248C` (`GAL4`). Unter *General Info* finden Sie, dass `GAL4` die für den Galactose-Metabolismus benötigten Gene aktiviert und von `GAL80` unterdrückt wird. Die Expressionsdaten kann man daher so verstehen, dass wenn `GAL80` künstlich unterdrückt wird, sich die Expression von `GAL4` selbst zwar kaum verändert, aber die *downstream*-Produkte nun verstärkt exprimiert werden.