

VI Argentinian conference on bioinformatics and Computational Biology

# Introduction to Cytoscape for network analysis



## Attribution-ShareAlike 4.0 International (cc BY-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#)



### You are free to:

**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material  
for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

### Under the following terms:



**Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



**ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

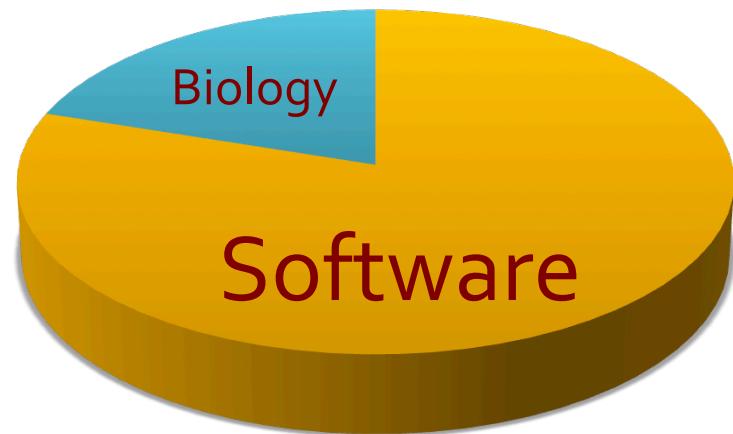
### Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

# Who am I?

- Juan José Díaz Montaña (aka Juanjo Diaz)
- Software Engineer
- Ph.D. candidate at  
Pablo de Olavide University
- Author of GFD-Net
  - Cytoscape app for visualizing and analyzing the functional dissimilarity of gene networks based on Gene Ontology (GO)
- Contact: [jjdiamon@alumno.upo.es](mailto:jjdiamon@alumno.upo.es)



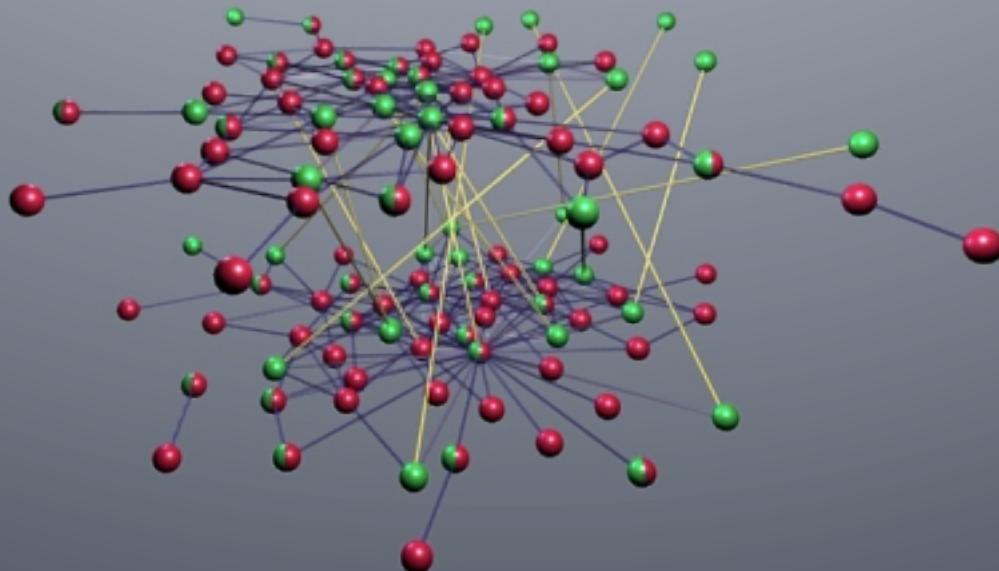
# Today's outline

- Intro
  - Biological Networks
  - Cytoscape
- Network Analysis Workflow
  - Load network
  - Load attributes
  - Analyze
  - Visualize
- Advance topics
  - App development
  - Cytoscape as a service
- Wrap up

# Biological networks

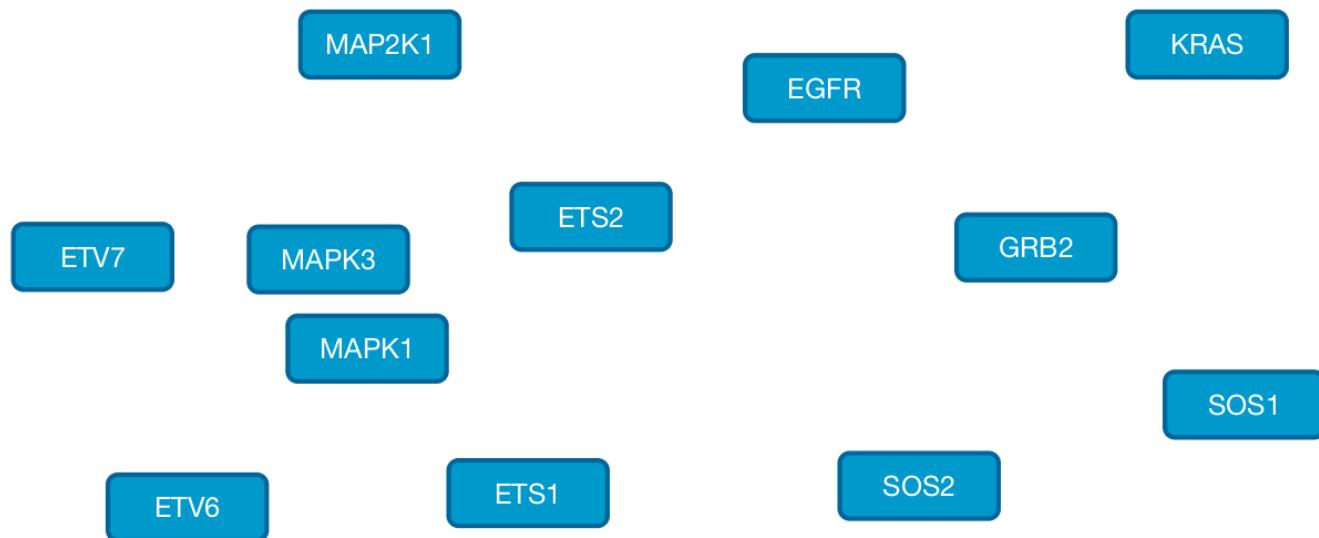
Genes/Proteins are not independent  
They regulate each other and act collectively

$$G = (V, E)$$



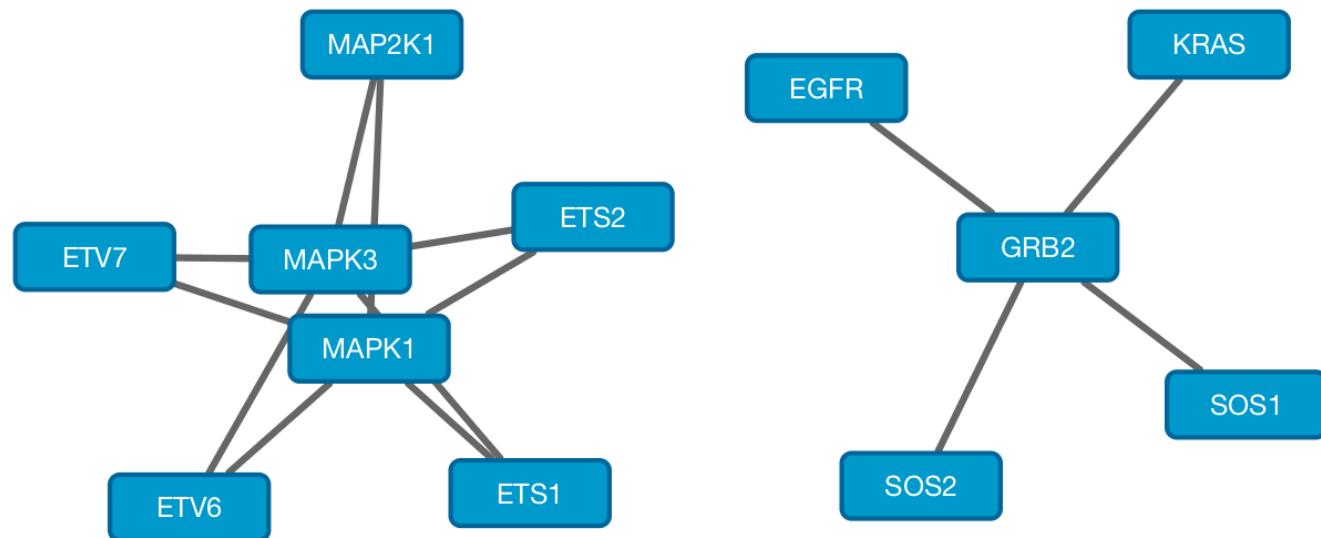
# Biological networks

- Node -> represents a gene or protein



# Biological networks

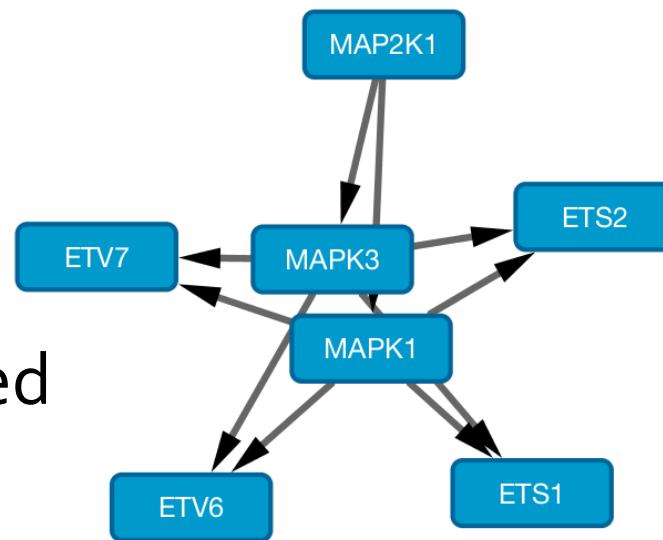
- Edge-> represents the relationship between two nodes



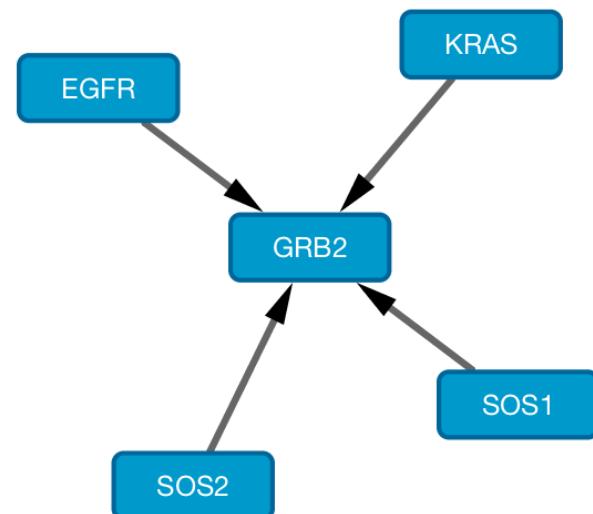
# Biological networks

- Edge-> represents the relationship between two nodes

- Directed



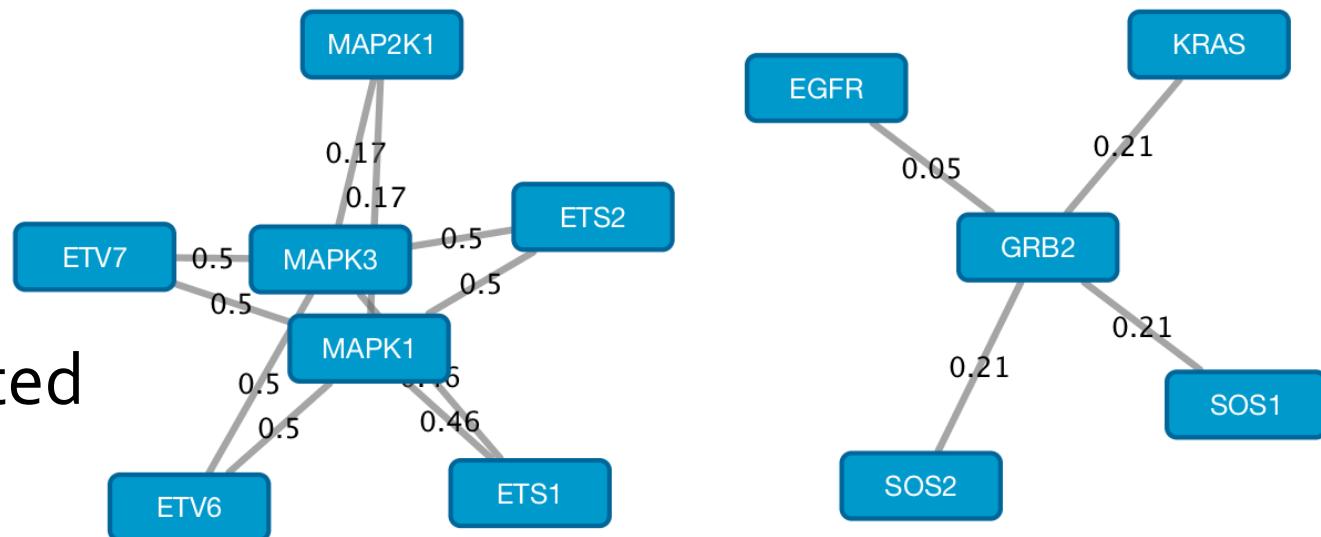
- Undirected



# Biological networks

- Edge-> represents the relationship between two nodes

- Weighted
- Unweighted



# Biological networks – Types

- **Gene Regulatory Network**
  - How genes regulate other genes
- **Protein-Protein Interaction (PPI) Network**
  - How proteins physically interact between them
- **Similarity networks**
  - Edges represent the similarity between two nodes
- **Others**
  - Gene co-expression network, metabolic network, signaling network, etc.

# Biological networks – Why?

- **Gene Function Prediction**
  - Connections might show how a set of genes/proteins are involved in the same biological process
- **Detection of protein complexes/other modular structures**
  - Discover modularity & higher order organization
- **Prediction of new interactions and functional associations**
  - Orthology-based methods
  - Bayesian networks inference methods
  - Domain-domain correlation methods

# Biological networks – Why?

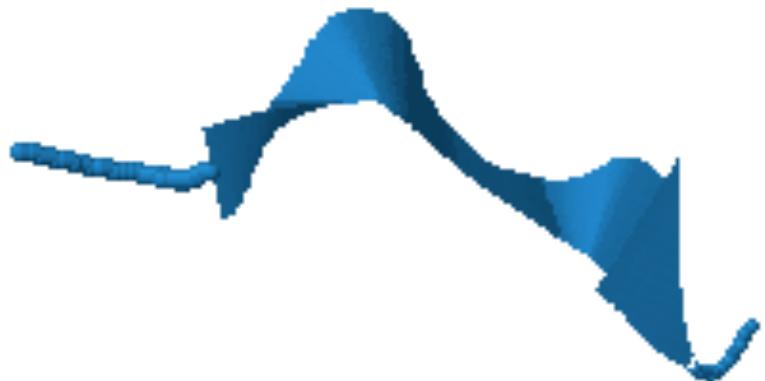
- **Identification of disease subnetworks**
  - Identification of disease network subnetworks that are transcriptionally active in disease
- **Subnetwork-based diagnosis**
  - Source of biomarkers for disease classification
  - Identify interconnected genes whose aggregate expression levels are predictive of disease state
- **Subnetwork-based gene association**
  - Map common pathway mechanisms affected by collection of genotypes

# Biological networks – Pros

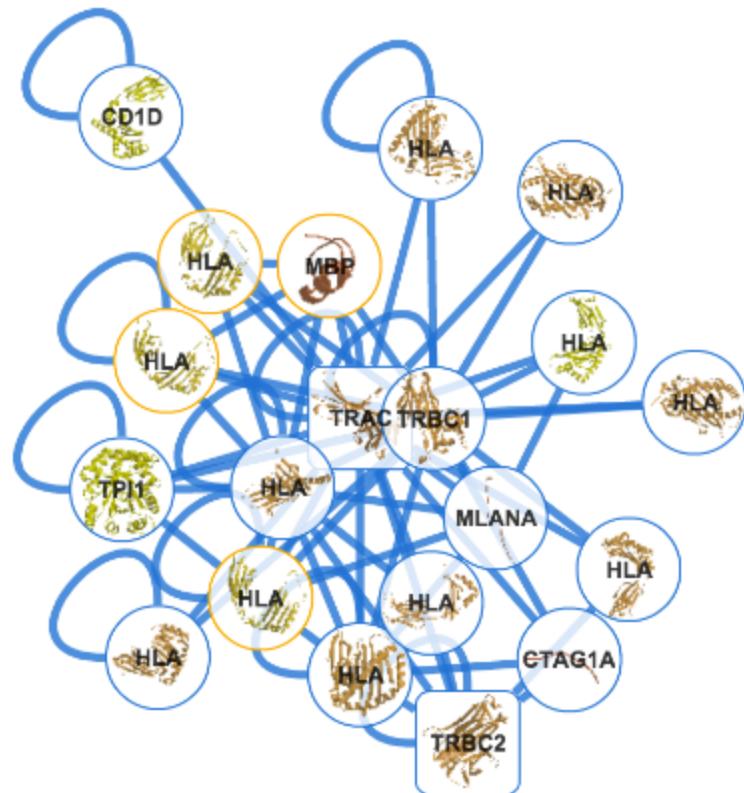
- Widely used and understood
- Diagrammatic representation
  - Concepts vs relationships
  - Reduce underlying complexity
- Graph theory
  - Allows mathematical analyses
- Different networks for different cases

# Biological networks – Pros

From genes, proteins, etc.



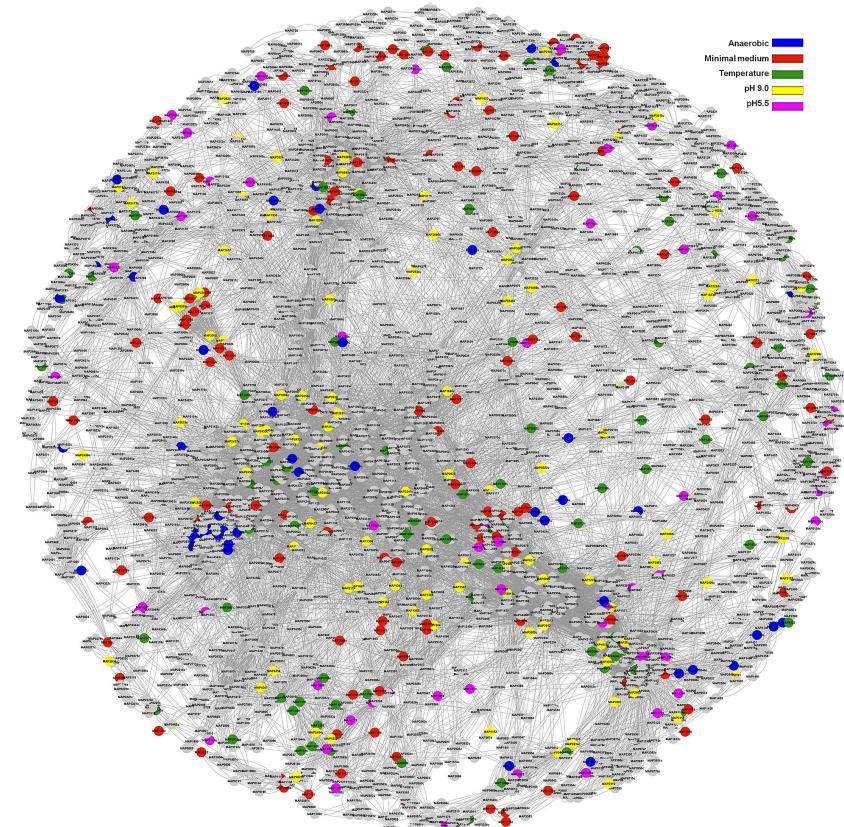
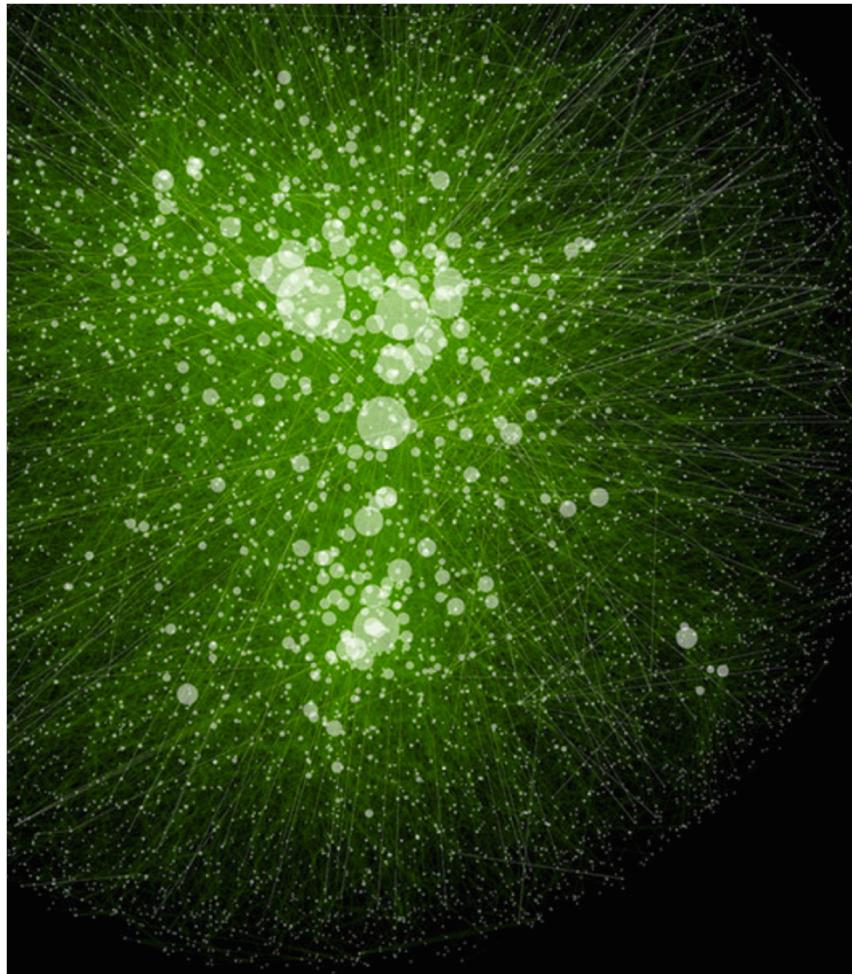
To complex networks



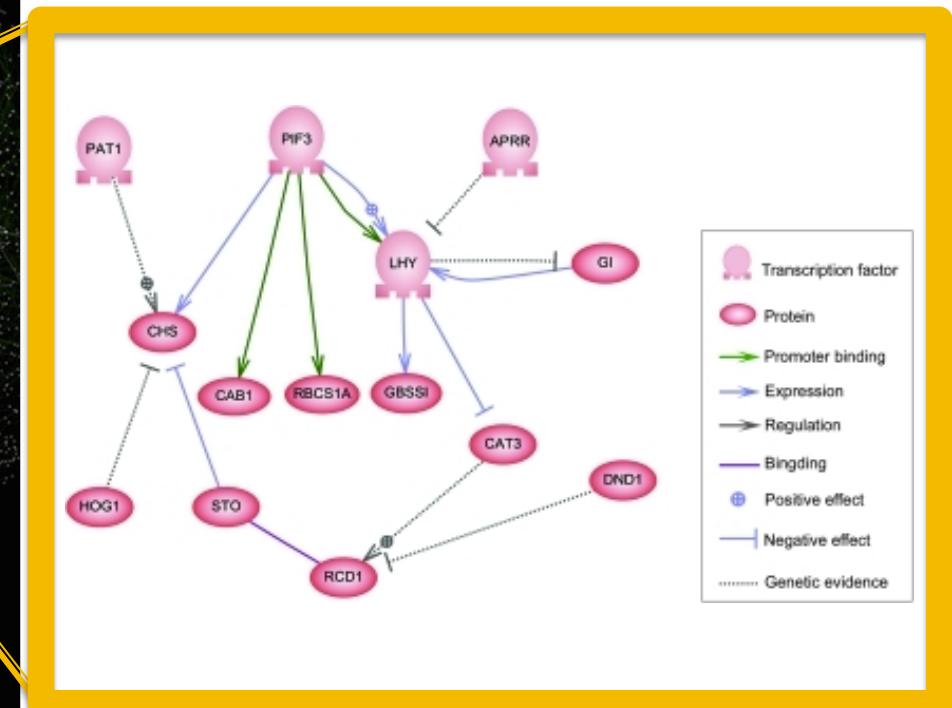
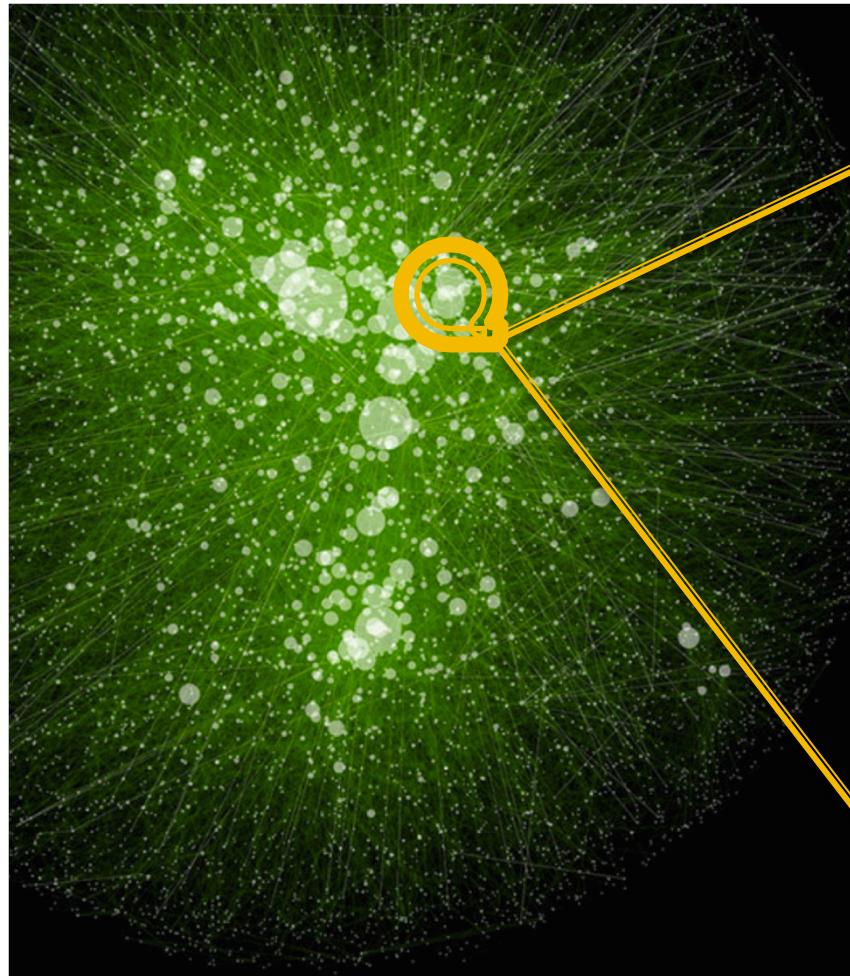
# Biological networks – Cons

- Static representation
  - Difficult to capture temporal and concentration information
- Pairwise edges
  - Not all relationships fit that model

# Biological networks – Cons



# Biological networks – Cons



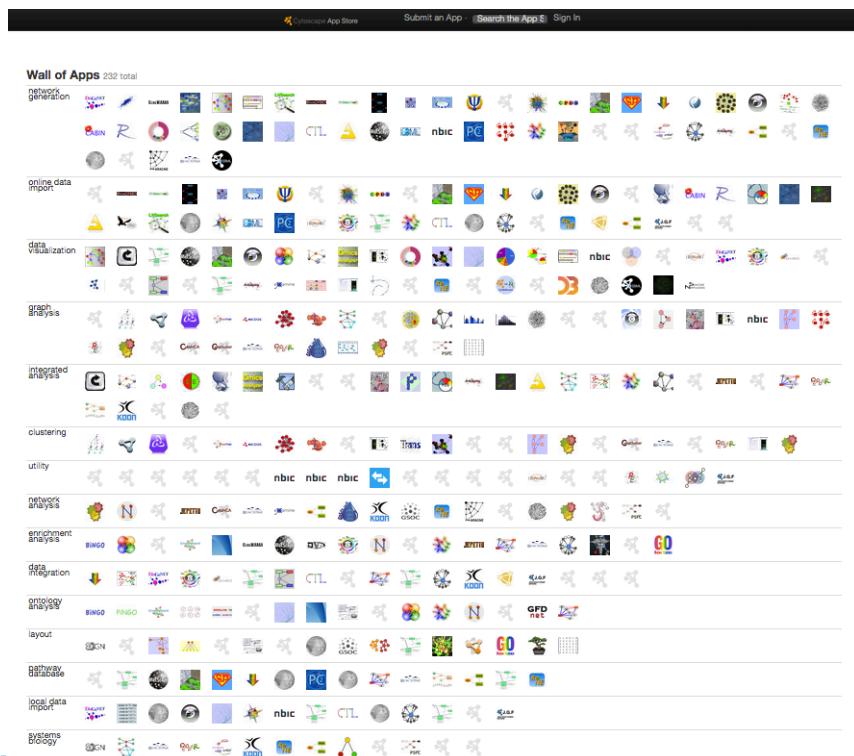
# Cytoscape

Open source software platform for visualizing complex networks and integrating these with any type of data



# Cytoscape

- Cytoscape is a platform that includes
  - Data integration
  - Analysis
  - Visualization
- Expandable by Apps  
(before called Plugins)



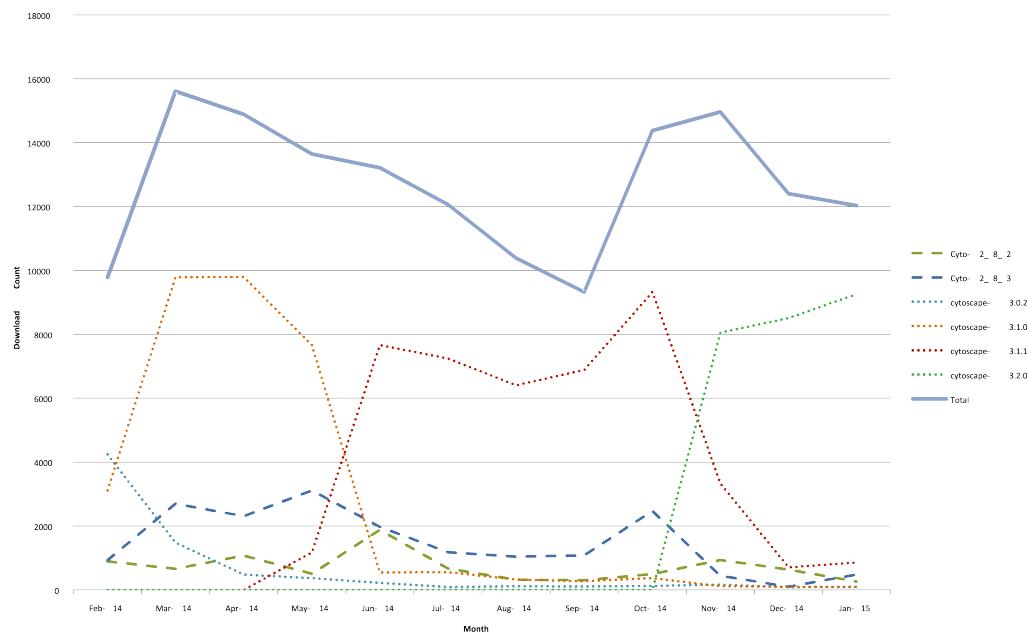
<http://apps.cytoscape.org>

# Cytoscape

- Open Source (LGPL)
  - <https://github.com/cytoscape>
- Backed by universities, companies and research institutions
- In active development since 2002  
(latest version 3.2.1)

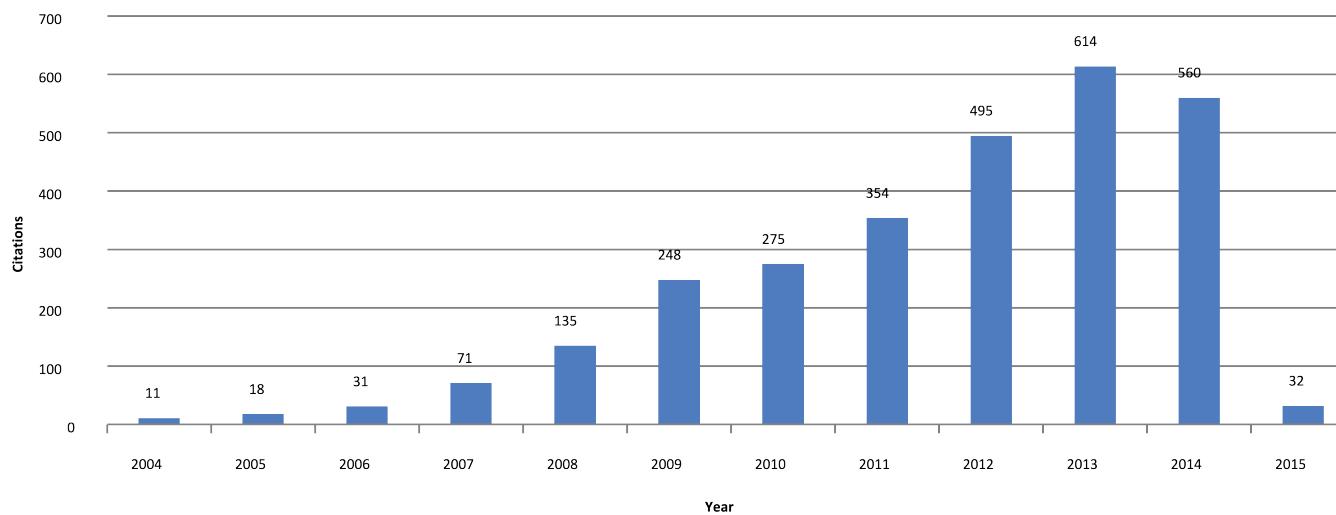
# Cytoscape

- De-facto standard software in biological network research
  - ~ 13000 downloads/month



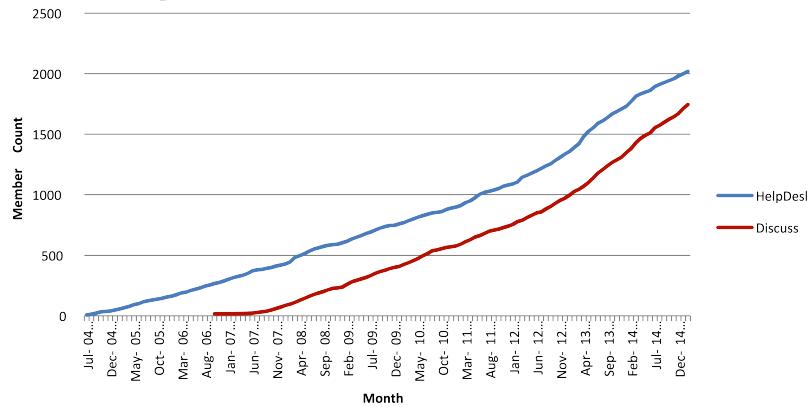
# Cytoscape

- De-facto standard software in biological network research
  - ~ 13000 downloads/month
  - + 50 citation/month



# Cytoscape

- De-facto standard software in biological network research
  - ~ 13000 downloads/month
  - + 50 citation/month
  - +1700 developers and +2000 users



# Cytoscape

- Tips
  - Size does matter
    - Cytoscape it's memory greedy
    - 4+ GB for big networks
  - Age matters too
    - Memory leaks specially with big network
    - Save often and restart if things get ugly
  - Configuration folder
    - Configuration data and downloaded apps
    - Located in home directory
    - If things get really messed up, delete (or rename) it

# Cytoscape

- Got stuck??

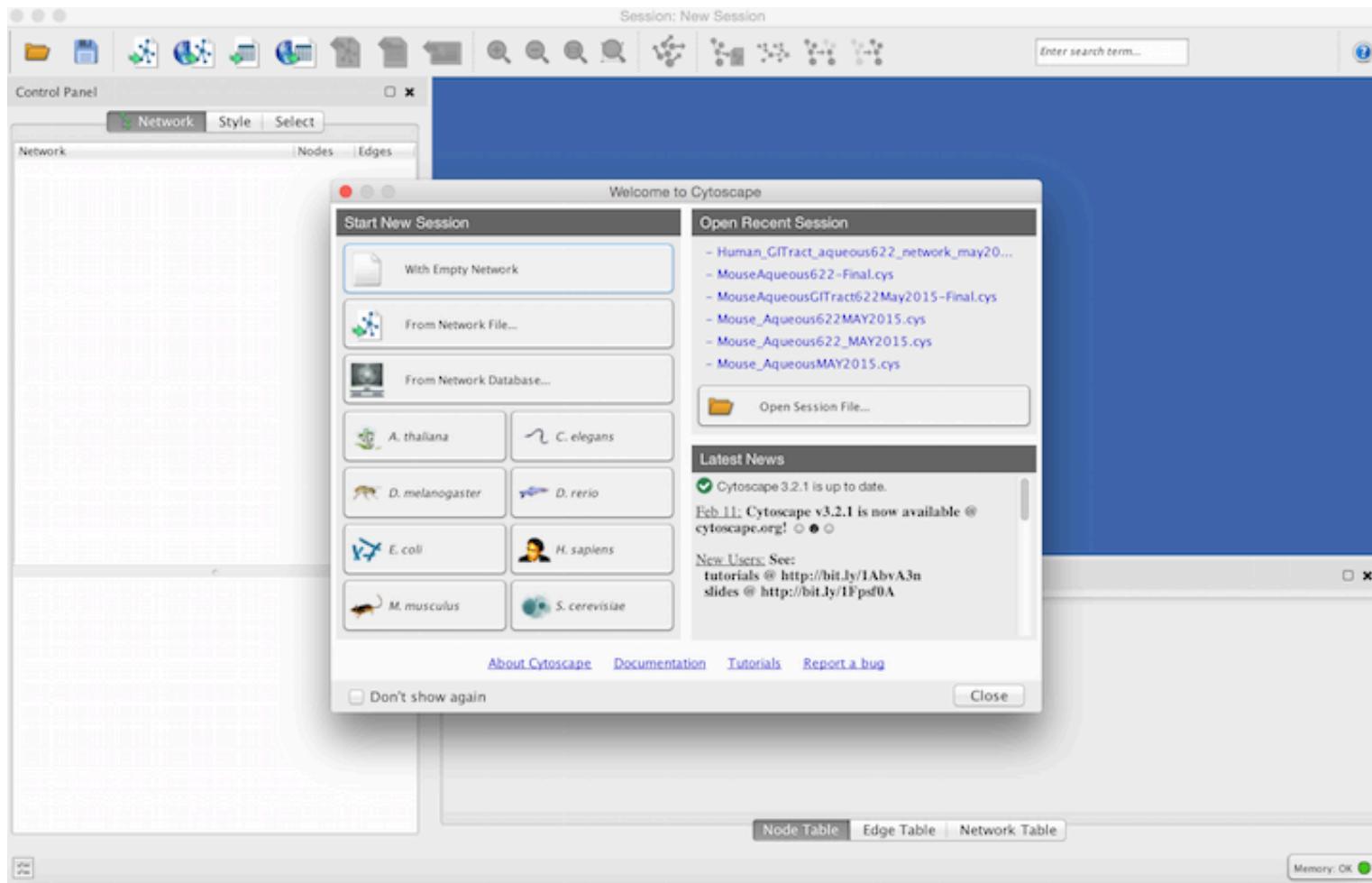
- General questions

<https://groups.google.com/d/forum/cytoscape-helpdesk>

- Development questions

<https://groups.google.com/d/forum/cytoscape-discuss>

# Cytoscape – Walkthrough



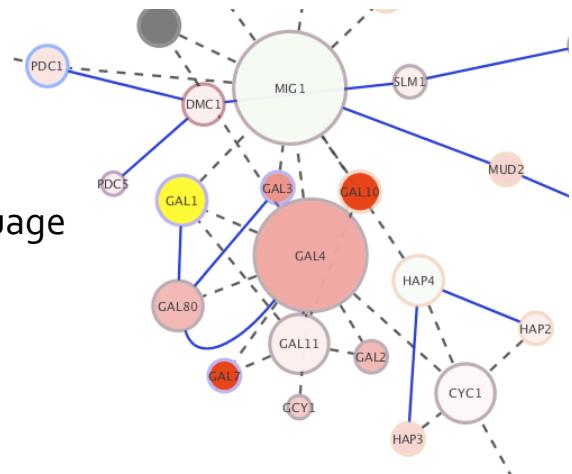
# Network analysis workflow

Cline, et al. "Integration of biological networks and gene expression data using Cytoscape", *Nature Protocols*, 2, 2366-2382 (2007)

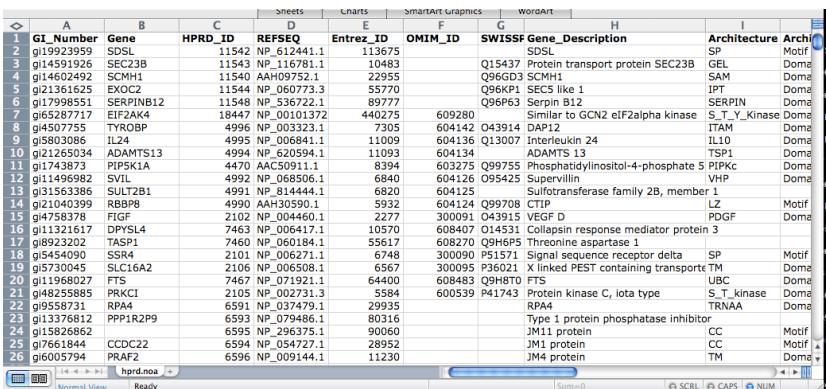


# Load network

- Manually
  - Creating an empty network and manually adding nodes and edges
- Import from public databases
  - GeneMANIA, BioGrid, IntAct, EBI-GOA-nonIntAct, UniProt, MINT, BIND, BAR, MolCon, APID, BindingDB, ChEMBL, DIP, DrugBank, HPIDb, I2D, I2D-IMEx, InnateDB, InnateDB-IMEx, Interoporc, MBInfo, MPIDB, MatrixDB, Reactome, Reactome-Fls, STRING, Spike, TopFind, VirHostNet, bhf-ucl, iRefIndex, mentha, ...
- Import from file
  - SIF Simple Interaction File
  - NNF Nested Network Format
  - GML Graph Markup Language
  - XGMML Extensible Graph Markup and Modeling Language
  - KGML Kegg Markup Language
  - SBML Systems Biology Markup Language
  - BioPAX
  - PSI-MI Level 1 and 2.5
  - Delimited text (.csv, .tsv, .txt, etc.)
  - MS Excel™ Workbook (.xls)



# Load attributes

- Network
    - Description
    - Experiment details
    - Publication Id
  - Edges
    - Interaction type
    - Interaction detection method
    - Confidence
  - Node
    - Gene expression data
    - Human readable gene names
    - Gene Ontology terms
- 
- The screenshot shows a Microsoft Excel spreadsheet titled "hprd.noa". The table has 16 columns labeled A through P and includes headers for Sheets, Charts, SmartArt/graphics, and WordArt. The data consists of 26 rows of gene information, such as GL\_Number, Gene, HPRD\_ID, REFSEQ, Entrez\_ID, OMIM\_ID, SWISSFL Gene\_Description, Architecture, and Archi. The data is as follows:
- |    | A           | B         | C       | D           | E         | F       | G                                | H  | I          | Architect | Archi |
|----|-------------|-----------|---------|-------------|-----------|---------|----------------------------------|--|------------|-----------|-------|
| 1  | GL_Number   | Gene      | HPRD_ID | REFSEQ      | Entrez_ID | OMIM_ID | SWISSFL Gene_Description         |  |            |           |       |
| 2  | gi 19923959 | SDSL      | 11542   | NP_612441.1 | 113675    |         | SDSL                             |  | SP         | Motif     |       |
| 3  | gi 1452962  | SFC23B    | 11543   | NP_116781.1 | 10483     | Q15437  | Protein transport protein SEC23B |  | GEL        | Doma      |       |
| 4  | gi 14602492 | SCMH1     | 11540   | AAH09752.1  | 22955     | Q96GD1  | SCMH1                            |  | SAM        | Doma      |       |
| 5  | gi 21361622 | EXOC2     | 11544   | NP_060773.3 | 55770     | Q96KP1  | SECS like 1                      |  | IPT        | Doma      |       |
| 6  | gi 17998551 | SPINBIN12 | 11548   | NP_536722.1 | 89777     | Q96P63  | Serpin B12                       |  | SERPIN     | Doma      |       |
| 7  | gi 65287717 | EIF2AK4   | 18447   | NP_00101372 | 440275    | 609280  | Similar to GCN2 eIF2alpha kinase | S_T_Y_Kinase                             |            |           |       |
| 8  | gi 4507755  | TYROB     | 4996    | NP_003323.1 | 7305      | 604142  | DAP12                            | ITAM                                     |            |           |       |
| 9  | gi 50861024 | LOC24     | 4997    | NP_068611.1 | 11009     | 604139  | Q13007                           | CD45 isoform 24                          | ITAM       | Doma      |       |
| 10 | gi 21269034 | ADMNTS13  | 4994    | NP_068594.1 | 11003     | 604134  | ADAMTS 13                        |  | TSP1       | Doma      |       |
| 11 | gi 7438073  | PIP5K1A   | 4470    | AAC50911.1  | 8394      | 603275  | Q99755                           | Phosphatidylinositol-4-phosphate 5 PIPKc |            | Phos      |       |
| 12 | gi 1496982  | SVIL      | 4992    | NP_068506.1 | 6840      | 604126  | Q95425                           | Supervillin                              | VHP        | Doma      |       |
| 13 | gi 31563386 | SUPT2B1   | 4991    | NP_814444.1 | 6820      | 604125  |                                  | Sulfotransferase family 2B, member 1     |            |           |       |
| 14 | gi 2104399  | RBBP8     | 4993    | AAH30590.1  | 5932      | 604124  | Q99708                           | CT10                                     | LZ         | Motif     |       |
| 15 | gi 4758378  | FIGF      | 2102    | NP_004460.1 | 2277      | 300091  | Q43915                           | VEGF D                                   | PDGF       | Doma      |       |
| 16 | gi 11321617 | DPY5L4    | 7463    | NP_006417.1 | 10570     | 608407  | Q14531                           | Colapsin response mediator protein 3     |            |           |       |
| 17 | gi 8923202  | TASP1     | 7460    | NP_601841.1 | 55617     | 608270  | Q9HG65                           | Threonine aspartate 1                    |            |           |       |
| 18 | gi 8454090  | SSR4      | 2101    | NP_006271.1 | 6748      | 300090  | P15171                           | Signal sequence receptor delta           | SP         | Motif     |       |
| 19 | gi 7300845  | SLC16A2   | 2106    | NP_006383.1 | 6567      | 300090  | P36021                           | X linked PEST containing transporter TM  |            | Doma      |       |
| 20 | gi 1994027  | PTS       | 7465    | NP_601921.1 | 64000     | 608483  | Q9H625                           | Transporter C                            | VBC        | Doma      |       |
| 21 | gi 1555895  | PMLCI     | 2005    | NP_002731.3 | 3594      | 600539  | P41743                           | Protein kinase C, iota type              | S_T_Kinase | Doma      |       |
| 22 | gi 9558731  | RP44      | 6591    | NP_037479.1 | 29935     |         | RP44                             |  | TRNA       | Doma      |       |
| 23 | gi 3376812  | PPP1R2P9  | 6593    | NP_079486.1 | 80316     |         |                                  |  |            |           |       |
| 24 | gi 15826862 |           | 6595    | NP_054735.1 | 90060     |         | JM11 protein                     | CC                                       | Motif      |           |       |
| 25 | gi 7661844  | CCDC22    | 6594    | NP_054727.1 | 28952     |         | JM1 protein                      | CC                                       |            |           |       |
| 26 | gi 6005794  | PRAF2     | 6596    | NP_009144.1 | 11230     |         | JM4 protein                      | TM                                       |            | Doma      |       |

# Load attributes

- Manually
- From file
  - File format
    - Delimited text (.csv, .tsv, .txt, etc.)
    - Excel Workbook (.xls)
  - Key should map a valid id
  - Each column maps to
    - String
    - Integer
    - Floating Point
    - Boolean
    - List of the above
- Into tables

Gotcha:  
*shared name column is always a string*

# Cytocape data organization

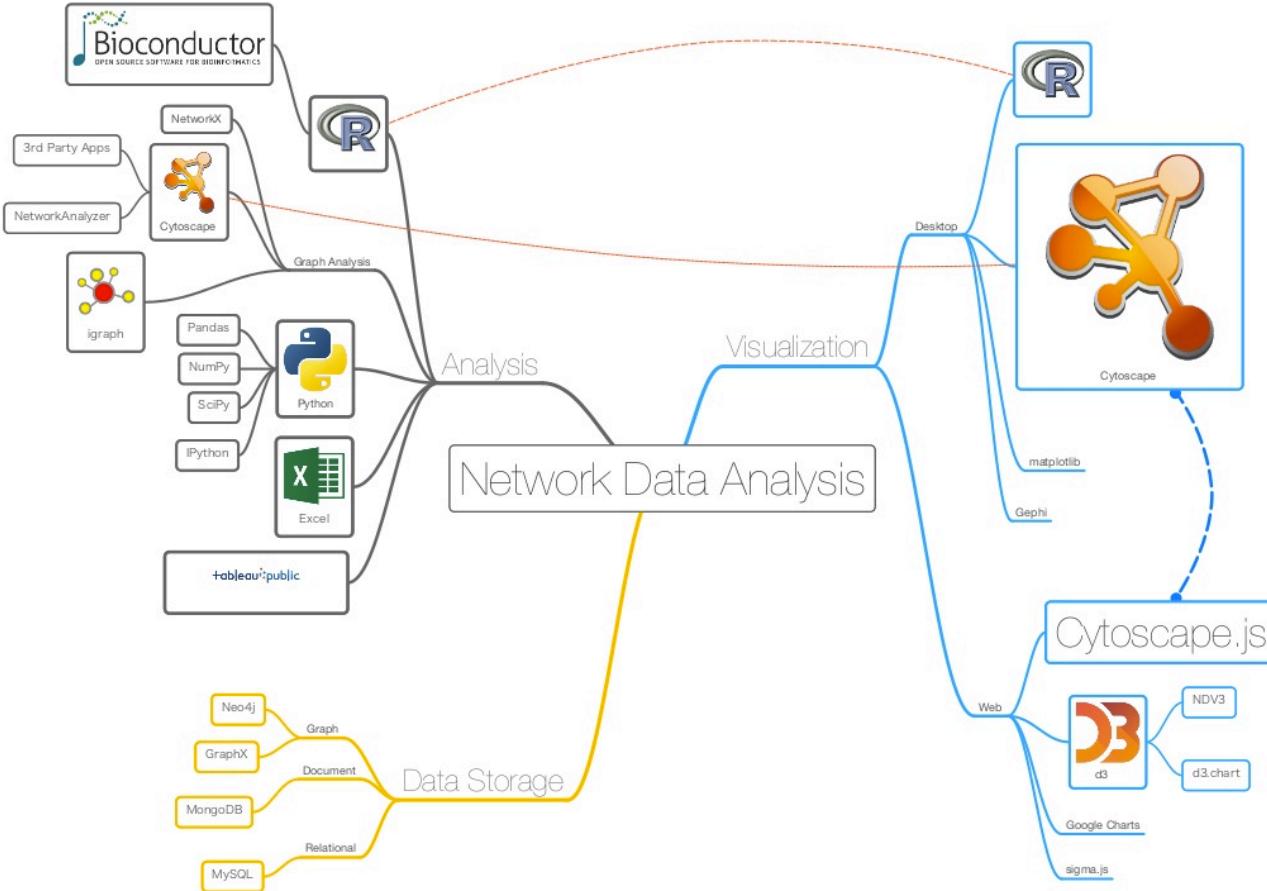
- Network collection
  - One or more Networks
    - Optional Network View
      - Levels of detail
      - Nodes can be grouped
  - Attribute tables
    - Nodes table
    - Edges table
    - Network table
    - ...

# Exercise 1

Import/export networks and attributes (~15 min)



# Network analysis



# Network analysis

- Finding and filtering nodes and edges
  - Look for nodes or edges of interest
  - Remove irrelevant nodes or edges
  - Extract a subnetwork for further investigation

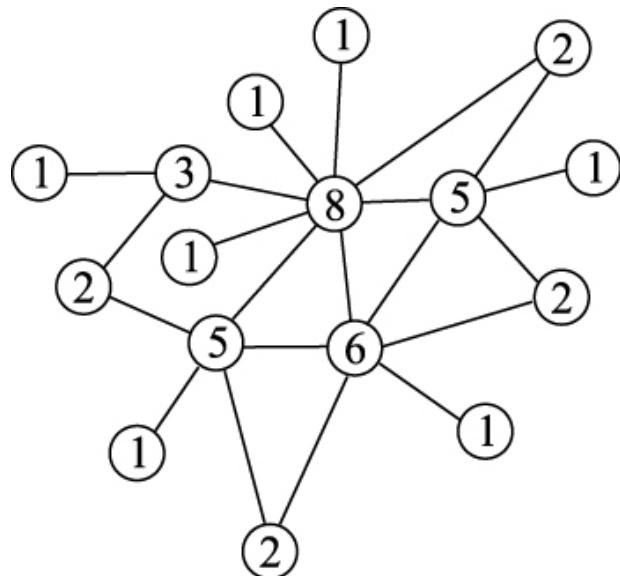
# Exercise 2

Filtering and selecting nodes/edges (~15 min)

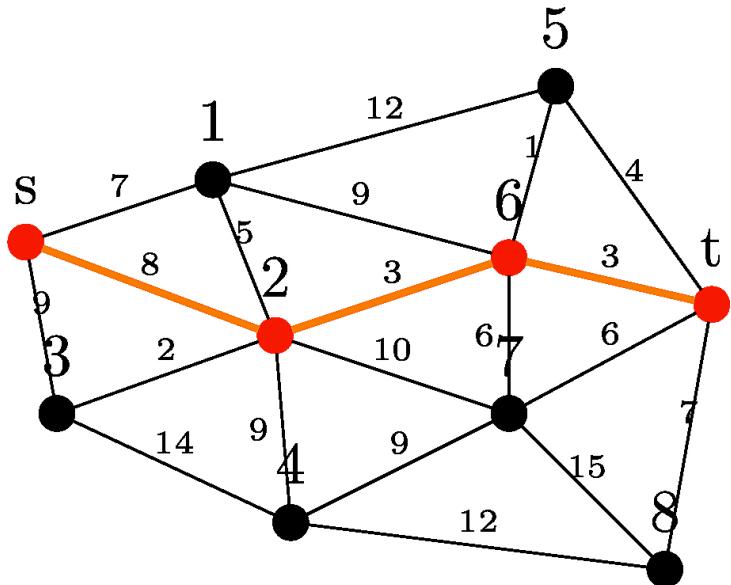


# Network analysis

- Topological Analysis
  - Degree



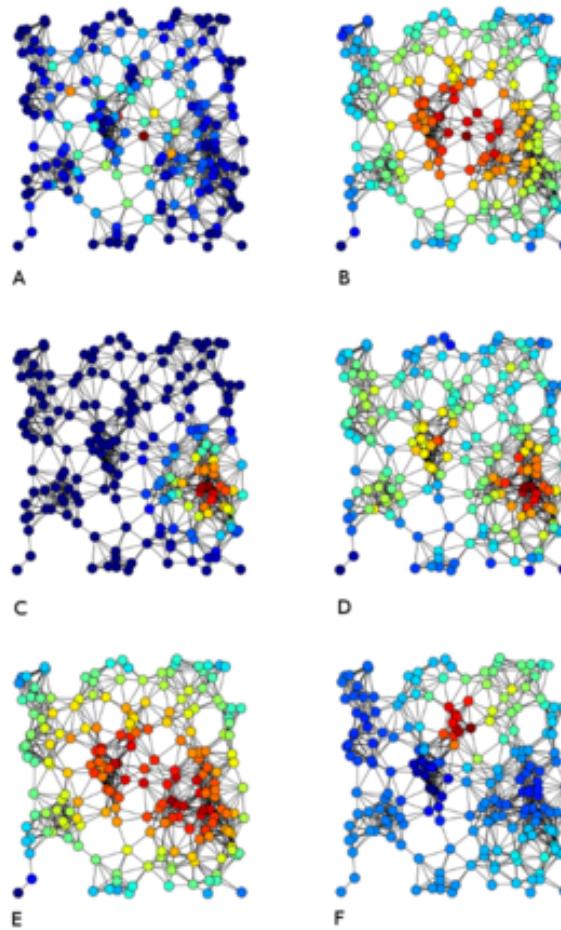
- Shortest path length



- Apps
  - NetworkAnalyzer

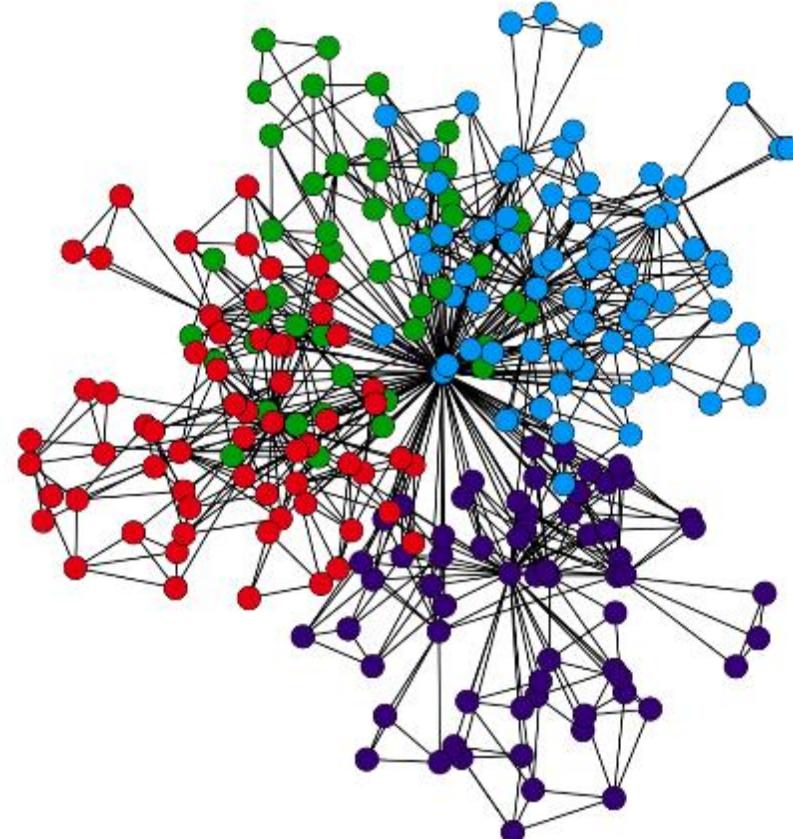
# Network analysis

- Centrality measures
  - Betweenness (A)
  - Closeness (B)
  - Eigenvector (C)
  - Degree (D)
  - ...
- Apps
  - CytoNCA
  - RINalyzer



# Network analysis

- Clustering measures
  - Connectivity-based (hierarchical)
  - Centroid-based
  - Distribution-based
  - Density-based
  
- Apps
  - MCODE
  - clusterMaker2
  - ClusterONE
  - ...



# Network analysis

- Network properties
  - Size
  - Density
  - Average degree
  - Average path length
  - Diameter
  - Average clustering coefficient
- Apps
  - NetworkAnalyzer

# Exercise 3

Topological analysis (~15 min)



# Network analysis

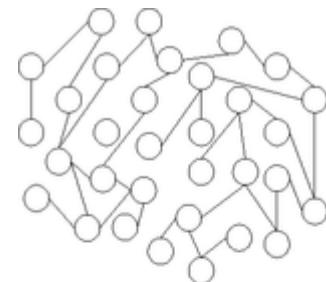
- Biological networks are scale-free (kind of...)

- Power law

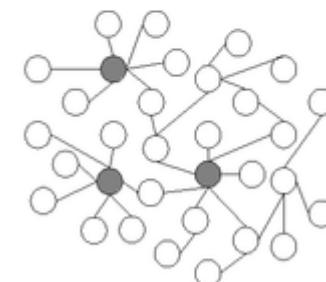
- Small-world

- Centrality and lethality principal

- Like social networks or the internet



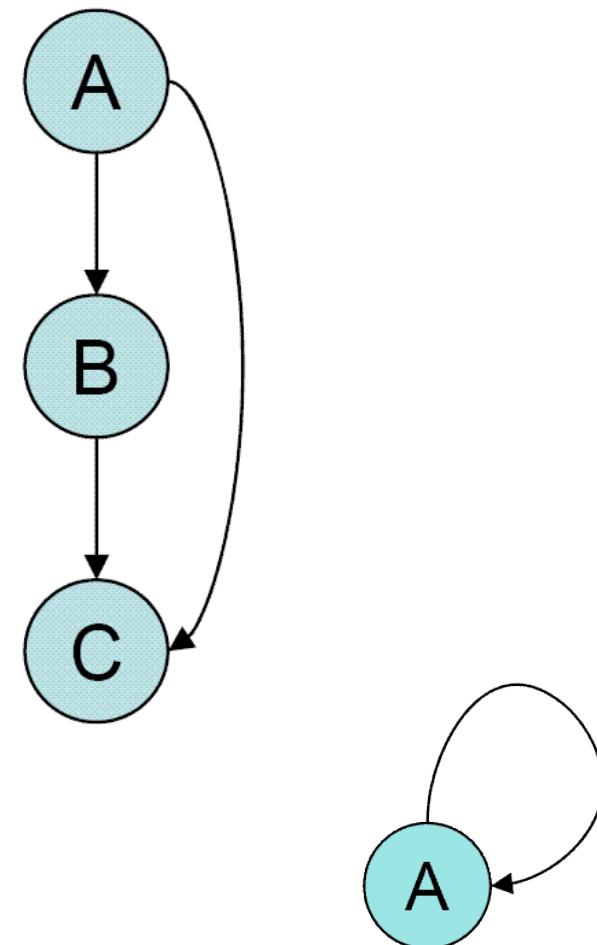
(a) Random network



(b) Scale-free network

# Network analysis

- Motif finding
  - Negative auto-regulation (NAR)
  - Positive auto-regulation (PAR)
  - Feed-forward loops (FFL)
  - Coherent type 1 FFL (C<sub>1</sub>-FFL)
  - Incoherent type 1 FFL (I<sub>1</sub>-FFL)
  - Multi-output FFLs
  - ...
- Apps
  - KeyPathwayMiner
  - CytoKavosh
  - SliMScape
  - NetMatch

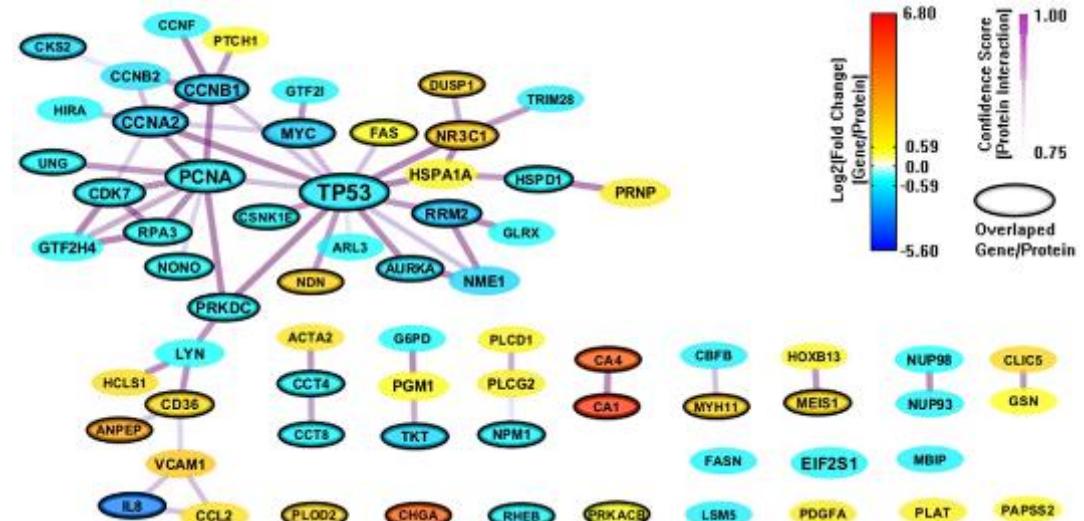


# Network analysis

- Inferred network vs Random network
  - Same idea behind
    - BLAST
    - p-value
  - Algorithms to create random networks
    - 100% random
    - Small-world
      - Watts-Strogatz
      - Erdős-Rényi
    - Scale-free
      - Barabási-Albert
  - Apps
    - Randomnetworks
    - NetDS

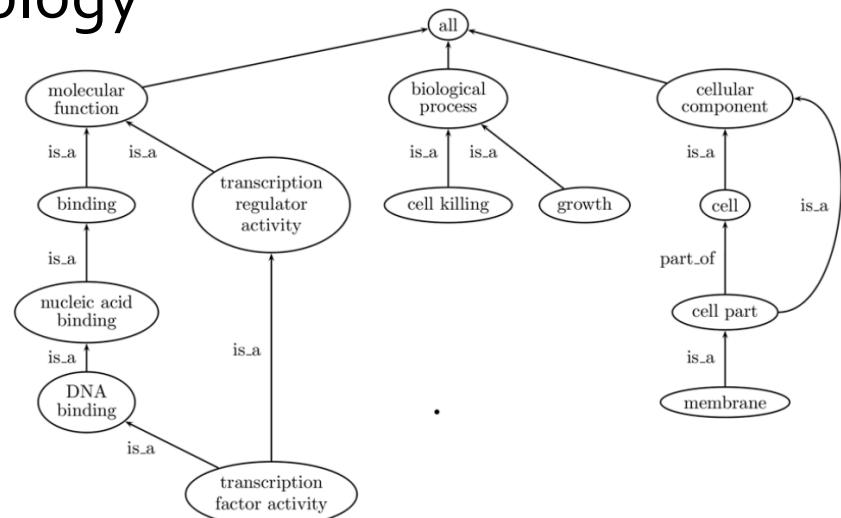
# Network analysis

- Enrichment analysis (Over-representation)
  - Uses statistical approaches to identify significantly enriched or depleted nodes
  - Gene ontology
- Apps
  - ClueGO
  - BINGO
  - EnrichmentMap
  - JEPETTO
  - ...



# Network analysis

- Semantic similarity analysis
  - Assesses the degree of relatedness between two entities by the similarity in meaning of their annotations
  - Annotations are stored in a structured vocabulary (ontology) like Gene Ontology
- Apps
  - GFD-Net
  - Sim-Trek (2.8)
  - CytoSeVis (2.7)

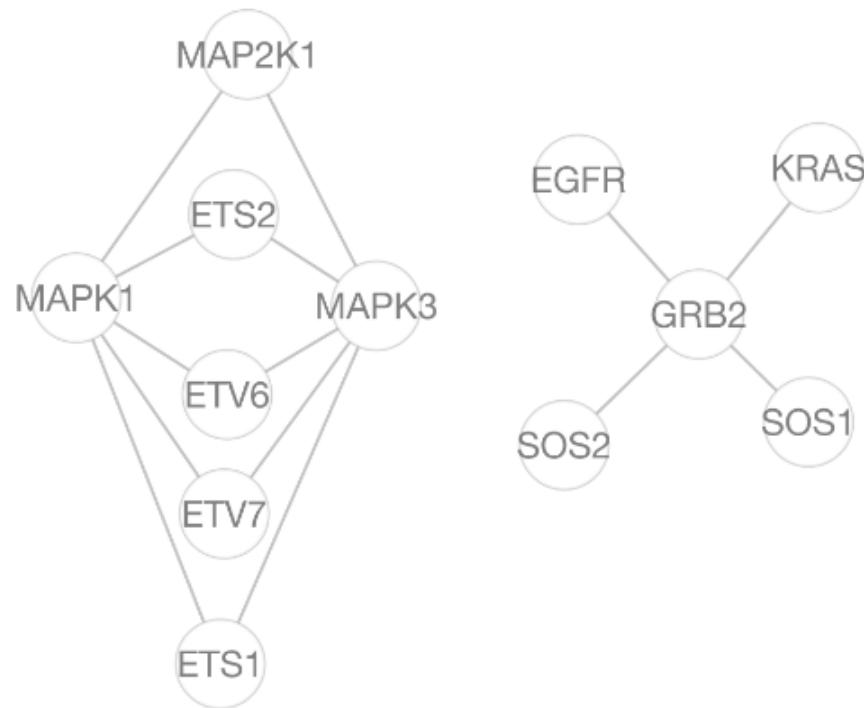


# Visualization

## ■ Facilitate

- Interpretation, new insights, new hypotheses
- Sharing results

	A	B	C
1	EGFR	pp	GRB2
2	KRAS	pp	GRB2
3	MAPK1	pp	ETS1
4	MAPK1	pp	ETS2
5	MAPK1	pp	ETV6
6	MAPK1	pp	ETV7
7	MAPK3	pp	ETS1
8	MAPK3	pp	ETS2
9	MAPK3	pp	ETV6
10	MAPK3	pp	ETV7
11	MAP2K1	pp	MAPK1
12	MAP2K1	pp	MAPK3
13	SOS1	pp	GRB2
14	SOS2	pp	GRB2



# Visualization

What is Great  
Visualization...?



# Visualization

## ■ Layout

- How to distribute nodes and edges on the canvas
- Reduce overlap and cluttering
  
- Automatic layouts
  - Grid
  - Hierarchical
  - Circular
  - Force-directed
  - etc.

# Visualization

- **Data mapping (VizMapper)**
  - Maps network properties to visual properties
  - Emphasize certain properties visually
  
- Node
  - Label, fill color, border color, shape, width, height, opacity and label
- Edge
  - Label, type, color, thickness, arrow type, size and color.
  
- Mapping types
  - Passthrough
    - Label all nodes with their gene symbols
  - Discrete
    - Color nodes differently depending on its gene function
  - Continuous
    - Map gene expression levels to brightness of color of corresponding node

# Visualization

## ■ Animation

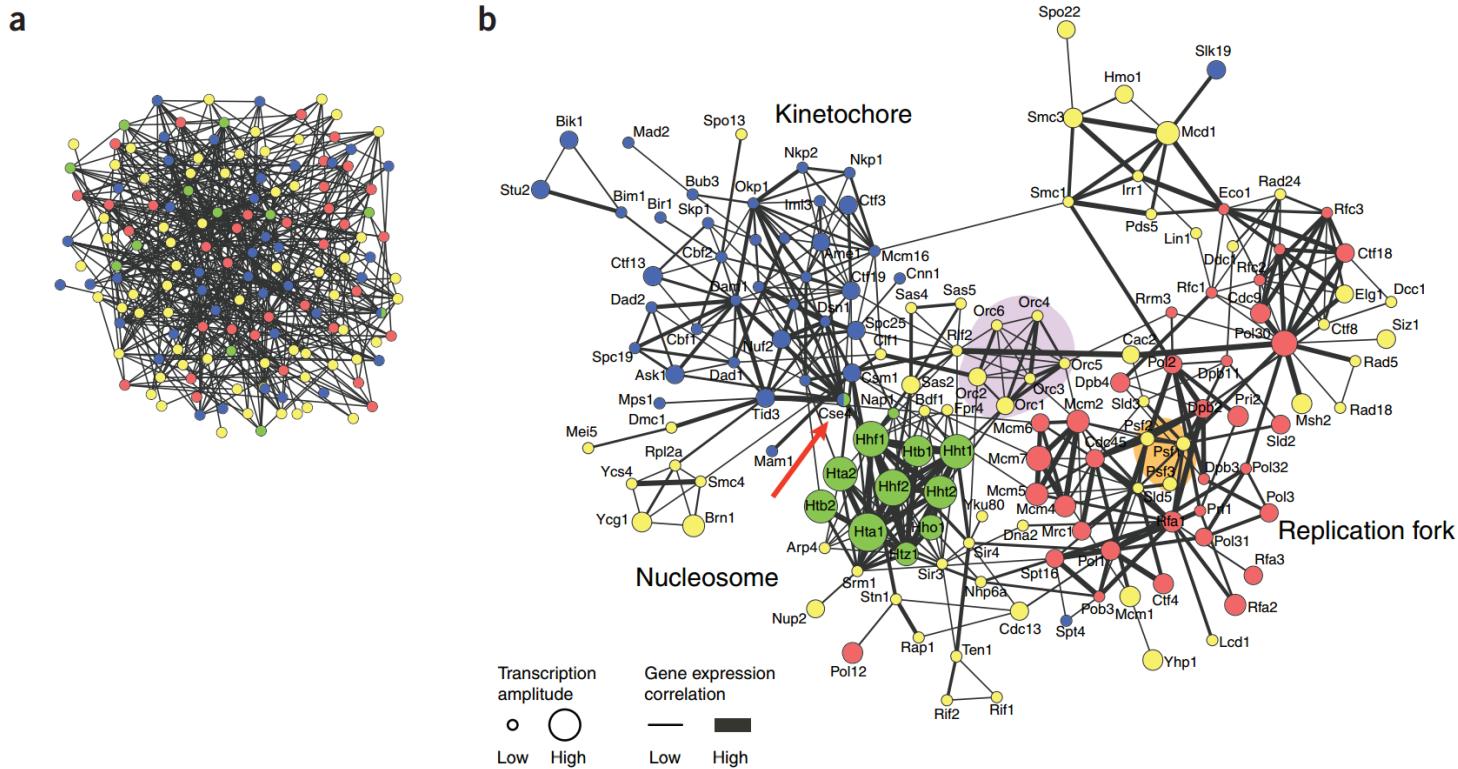
- To add non-static data
- Apps
  - DynNetwork
  - CyAnimator

# Visualization

## ■ Tips

- You need a story
- Too much info = no info
- Too cool = not cool

# Visualization



Merico, Daniele, David Gfeller, and Gary D. Bader. "How to visually interpret biological data using networks." *Nature biotechnology* 27.10 (2009): 921-924.

# Exercise 4

Enrichment analysis (~25 min)



# Prepare for Publication

- Export network data
  - Cyjs            Cytoscape.js JSON
  - SIF            Simple Interaction File
  - NNF            Nested Network Format
  - XGMML        Ext. Graph Markup and Modeling Lang.
  - PSI-MI Level 1 and 2.5
- Export attributes data
  - CSV

# Prepare for Publication

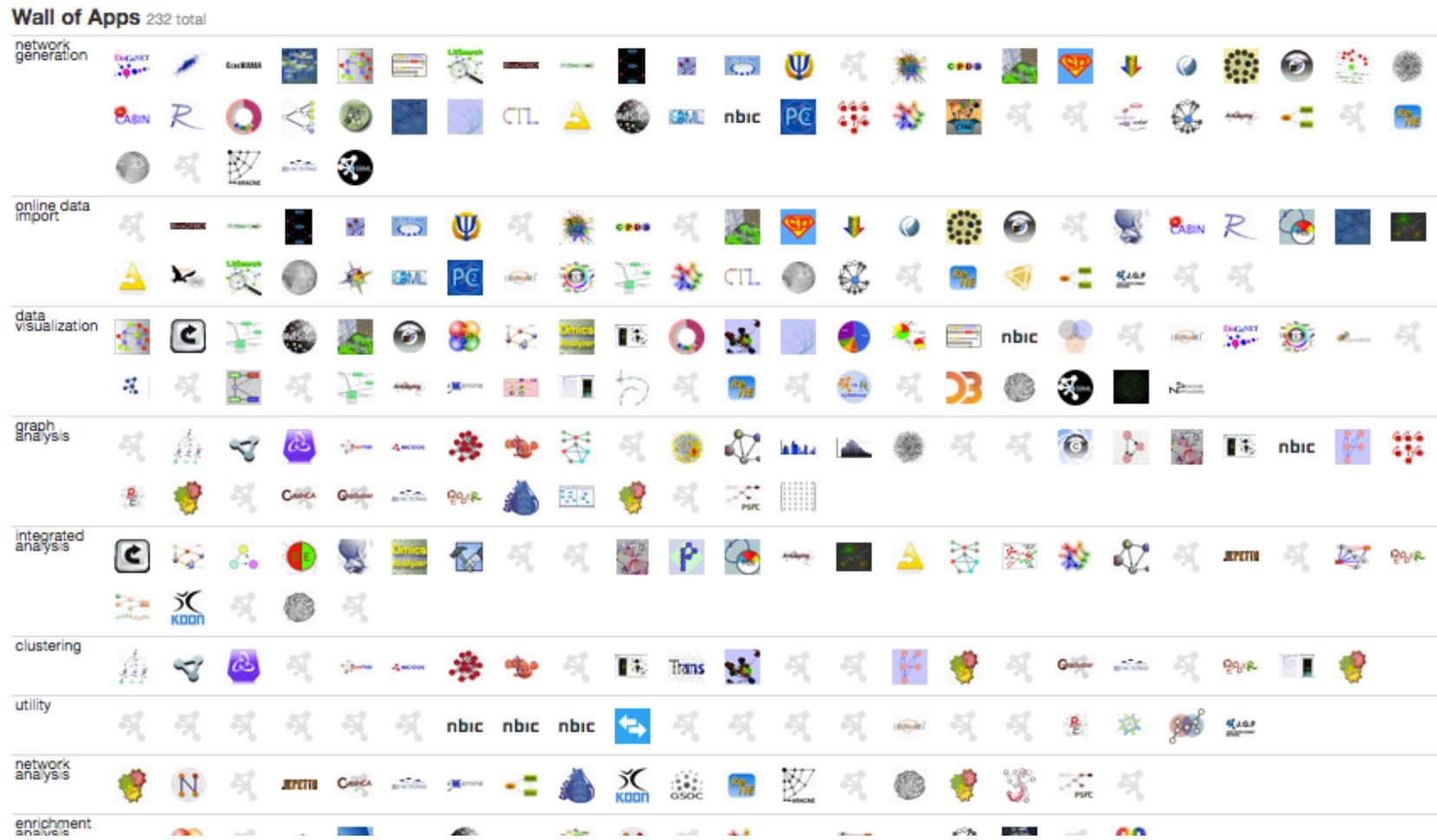
- Export as graphic
  - JPEG Image File (\*.jpg, \*.jpeg)
  - PDF File (\*.pdf)
  - Portable Network Graphics (\*.png)
  - Post Script (PS) File (\*.ps)
  - Scalable vector Graphics (SVG) File (\*.svg)
- Import for web
  - Cytoscape.js JSON (Network, style or both)
  - Full website

# Advance topics

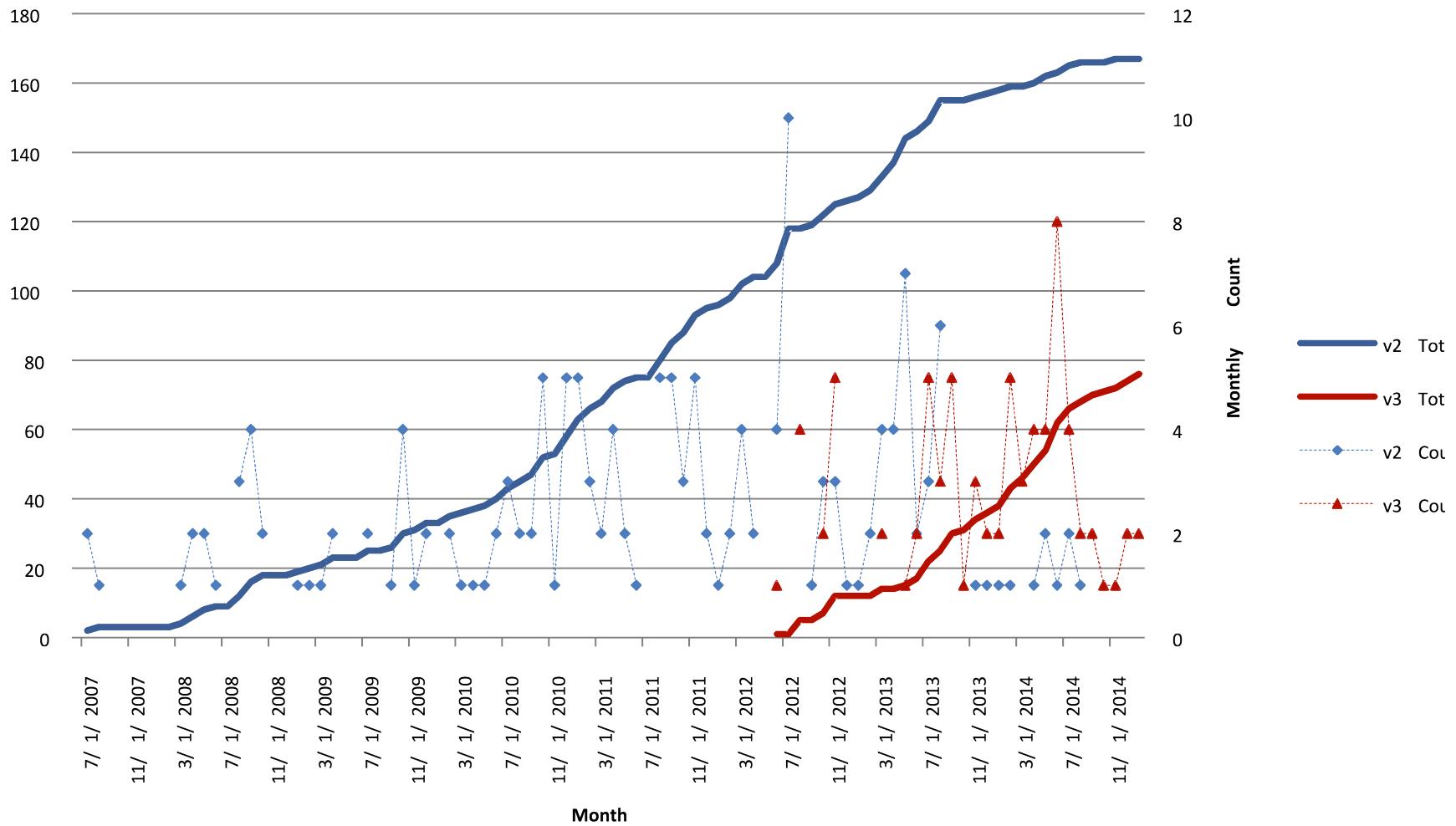
Software knowledge required!

```
function() { var i<a.length&&(x=a[i])&&x.offsetTop>0&&x.offsetHeight>0){if(!d.MM_p) d.MM_p=new Array();  
d.MM_p.length=a.length;preloadImages.arguments.arguments; for(i=0; i<a.length; i++)  
d.MM_p[i]=new Image; d.MM_p[i].src=a[i].src}  
@ @ document.all(p=a.indexOf("?"))>0&&parent.frames.length :  
@ @ parent.frames[p+1].document; n=n.substring(0,p);  
@ @ +@.all(n); for (i=0;i<d.forms.length;i++) x=d.forms[i].name;  
@ @ if (i<d.layers.length;i++) x=MM_findObj(n,d.layers[i].name);  
@ @ return x; }  
image() { var i<L.length; arguments; document.MM_sr=new Array; for(i=0; i<L.length; i++)  
L[i].imageArguments.arguments; document.MM_sr[i]=x; if(!x.offsetTop>0&&x.offsetHeight>0){  
x.offsetTop=0; x.offsetHeight=0; x.offsetWidth=0; x.innerHTML=""; x.style="";  
x.setAttribute("style",""); x.setAttribute("offsetTop","0"); x.setAttribute("offsetWidth","0");  
x.setAttribute("offsetHeight","0"); x.setAttribute("innerHTML","");
  }
}
```

# App development



# App development



# Apps development

## ■ Java + Maven + OSGi

```
public class MenuAction extends AbstractCyAction {  
  
    private final CyApplicationManager applicationManager;  
  
    public MenuAction(final CyApplicationManager applicationManager, final String menuTitle) {  
        super(menuTitle, applicationManager, null, null);  
        this.applicationManager = applicationManager;  
        setPreferredMenu("Apps");  
    }  
  
    public void actionPerformed(ActionEvent e) {  
        final CyNetworkView currentNetworkView = applicationManager.getCurrentNetworkView();  
        if (currentNetworkView == null)  
            return;  
  
        // View is always associated with its model.  
        final CyNetwork network = currentNetworkView.getModel();  
        for (CyNode node : network.getNodeList()) {  
  
            if (network.getNeighborList(node,CyEdge.Type.ANY).isEmpty()) {  
                currentNetworkView.getNodeView(node).setVisualProperty(BasicVisualLexicon.NODE_VISIBLE, false);  
            }  
        }  
        currentNetworkView.updateView();  
    }  
}
```

# Apps development

- Links:
  - Review existing apps
    - <http://apps.cytoscape.org/>
  - Check out articles about open sourced apps
    - <http://f100oresearch.com/channels/cytoscapeapps>
  - Source of wisdom
    - <https://groups.google.com/d/forum/cytoscape-discuss>

# Cytoscape as a Service

- What if you have hundreds of data files?
- What if someone need to execute your data analysis workflows?

**GUI isn't always good**

# Cytoscape as a Service

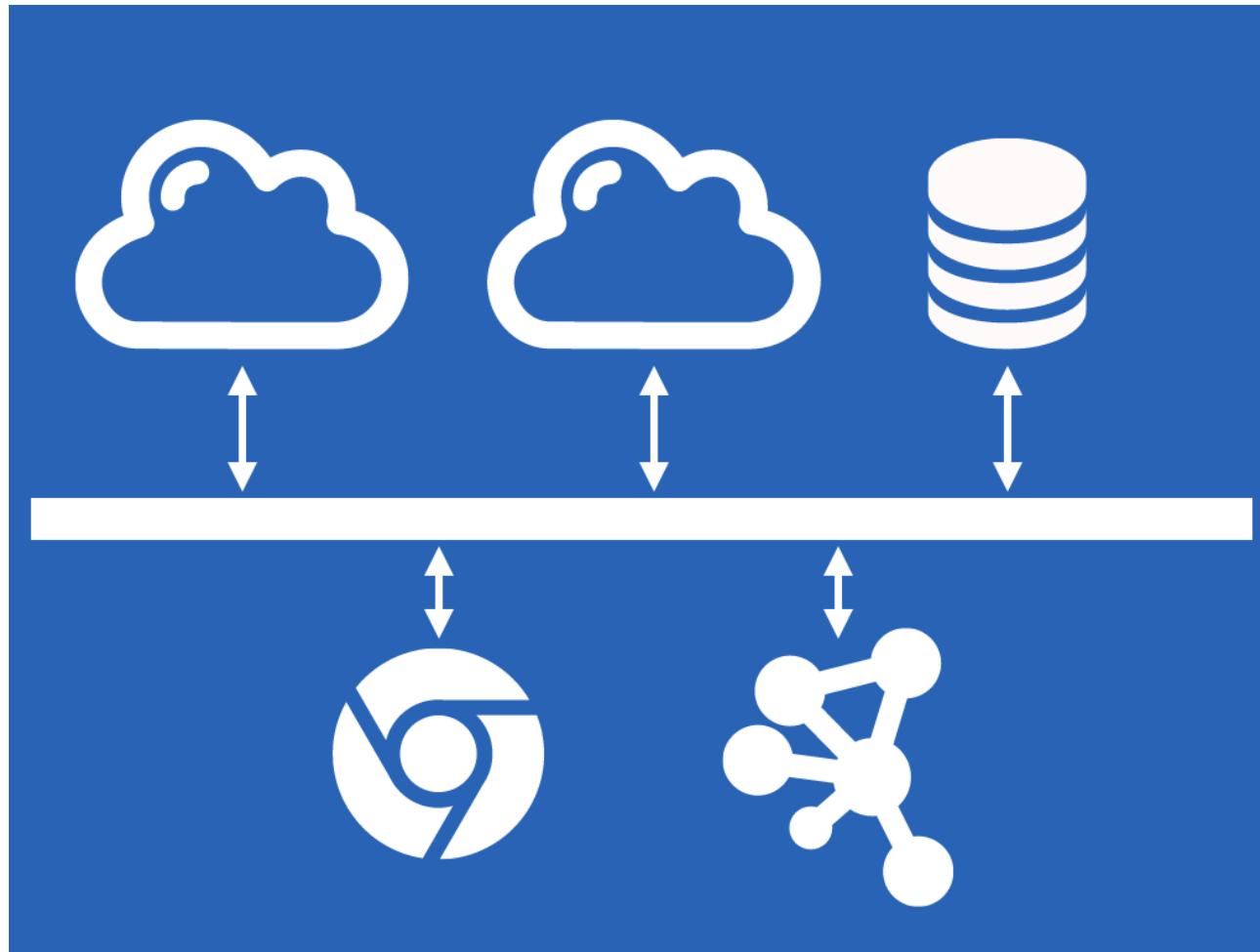
- Cytoscape is scriptable
  - Commands
    - Tools → Command Line Dialog
    - Tools → Execute Command File
    - `./cytoscape.sh -S <MY_SCRIPT>`
  - REST API
    - Built-in
      - `./cytoscape.sh -R 8888`
      - <http://localhost:8888/cytoscape/commands>
    - cyREST
      - Exposes a low-level RESTful API
      - <http://apps.cytoscape.org/apps/cyrest>

# Exercise 5

Cytoscape as a Service (~15 min)



# Cytoscape as a Service



# Wrap up

Today we learned...



# Wrap up

- Biological networks
  - Useful for seeing relationships in large data sets
  - Many different types
    - Important to understand what the nodes and edges mean
  - Widely used and understood
  - Reduce complexity
  - Seldom tell much by themselves
    - Data integration + Analysis + Visualization
  - Great to share results

# Wrap up

- Cytoscape
  - Platform for network analysis and visualization
  - Extensible → Core + Apps
  - Interoperable
    - Scriptable → commands, REST API, cyRest, ...
  - Easy to share → cys session, export
  - Cloud engine?? → Cytoscape.js, headless cytoscape

# Wrap up

- Network analysis using Cytoscape
  - There is no silver bullet
    - Use Cytoscape combined with other technologies
      - Excel, Python, R, Web Tools, etc.
  - Understand Cytoscape core features
  - Research existing apps
  - Cytoscape workflow
    - Load network + Load attribute + Analyze & visualize + Publish

# Any question?

Juanjo Díaz ([jjdiamon@alumno.upo.es](mailto:jjdiamon@alumno.upo.es))