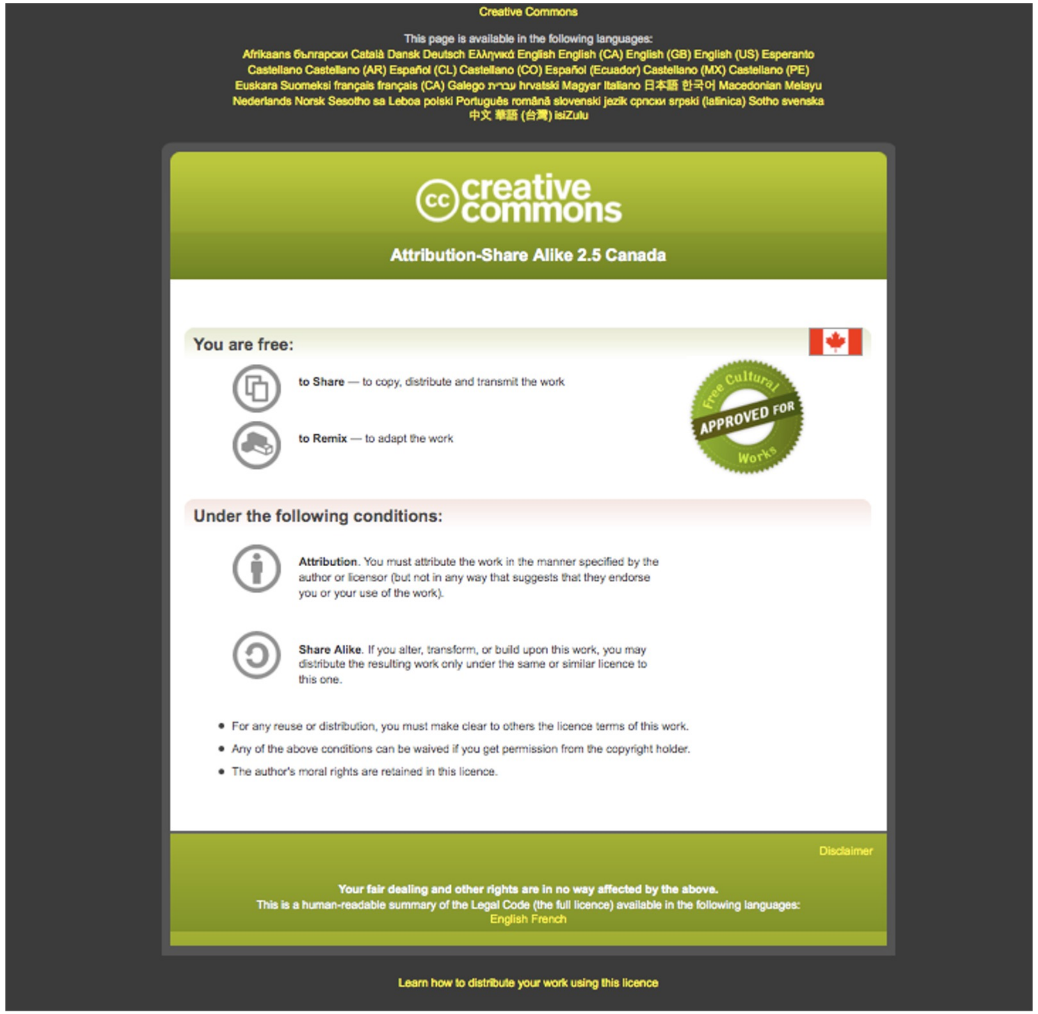
Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io



Creative Commons

This page is available in the following languages:
Afrikaans български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE) Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska 中文 華語 (台灣) isiZulu

creative commons

Attribution-Share Alike 2.5 Canada

You are free:

- **to Share** — to copy, distribute and transmit the work
- **to Remix** — to adapt the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French

Learn how to distribute your work using this licence

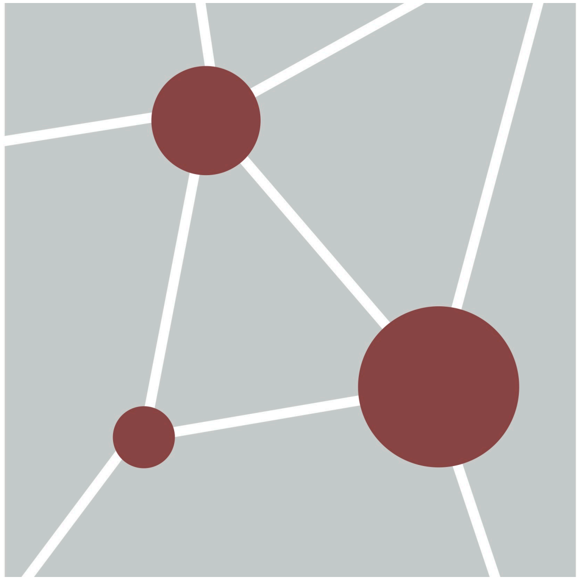
Module 3: Network Visualization and analysis with Cytoscape

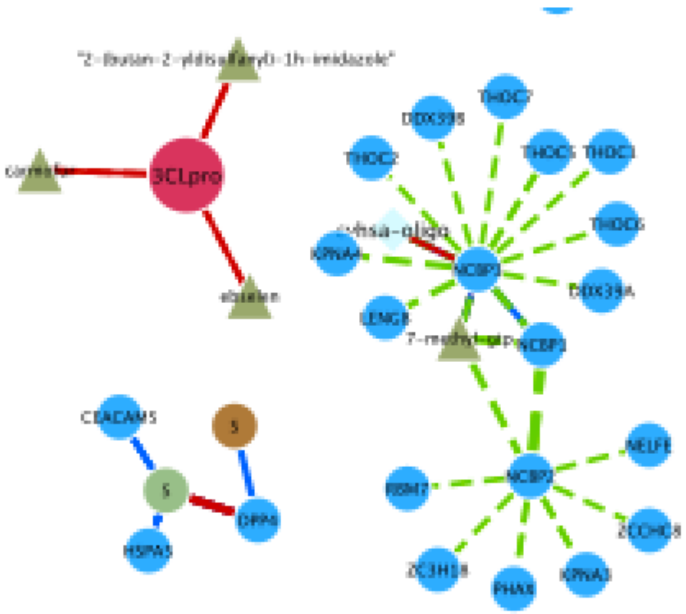
Ruth Isserlin

Pathway and Network Analysis

June 5-7, 2023

bioinformatics.ca





BADER LAB

UNIVERSITY OF TORONTO

Donnelly Centre
Cellular & Biomolecular Research
UNIVERSITY OF TORONTO

Learning Objectives

- By the end of this lecture, you will:
    - Understand the advantages of network visualization.
    - Understand how to choose and install Cytoscape Apps
    - Be able to use basic Cytoscape features
    - Be able to create and optimize a network in Cytoscape

Network Visualization and Analysis Outline

- Introduction
- Network Basics
- Network Visualization
- Network Analysis

- Demo: Cytoscape software tool for Network visualization and analysis

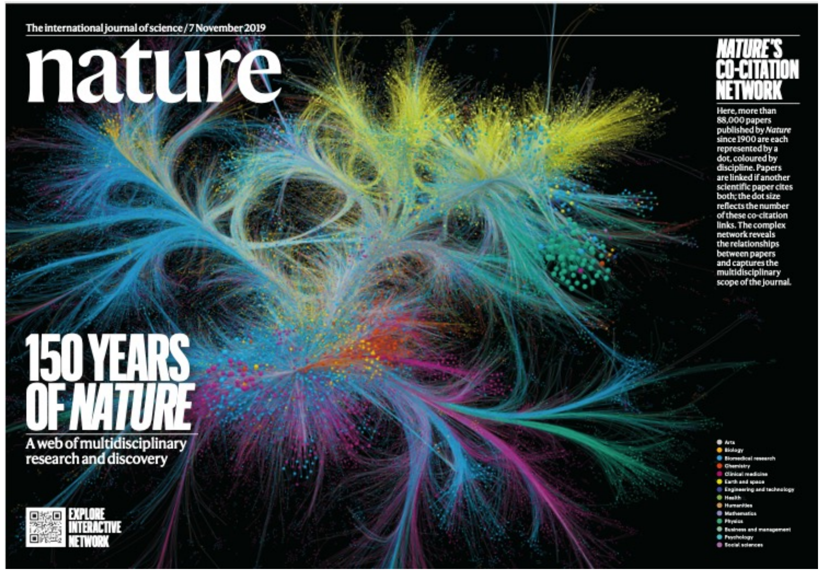
Six Degrees of Separation

- Everyone in the world is connected by at most six links.
- Originally demonstrated by Stanely Milgram at Harvard in 1960s in an experiment to find how many random acquaintances it would take to connect two people…. sometimes defined as the the **social distance** between two people.
- Popularized in the 1990s with six degrees of Kevin Bacon.
- Demonstrated the interconnectivity of our social structures.
- Networks exists everywhere – power grids, cell tower networks, Fed ex delivery networks, supply chain networks ….
- Is this Biologically relevant?

Why Networks?

Networks are everywhere….

- Molecular networks
- Cell-Cell Communication
- Nervous systems
- Social Networks



The international journal of science / 7 November 2019

nature

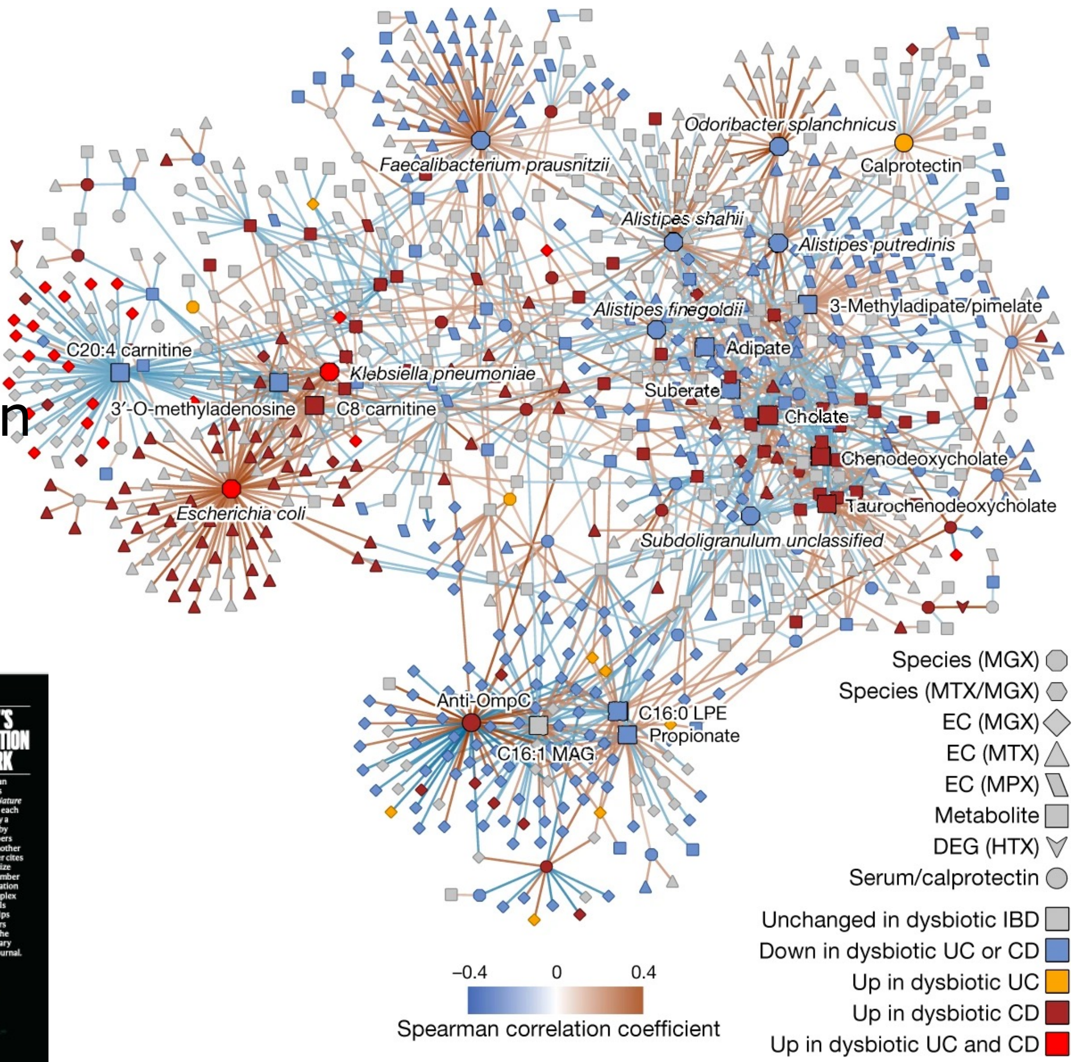
NATURE'S CO-CITATION NETWORK

Here, more than 88,000 papers published by *Nature* since 1900 are each represented by a dot, coloured by discipline. Papers are linked if another scientific paper cites both; the dot size reflects the number of these co-citation links. The complex network reveals the relationships between papers and captures the multidisciplinary scope of the journal.

150 YEARS OF NATURE

A web of multidisciplinary research and discovery

EXPLORE INTERACTIVE NETWORK



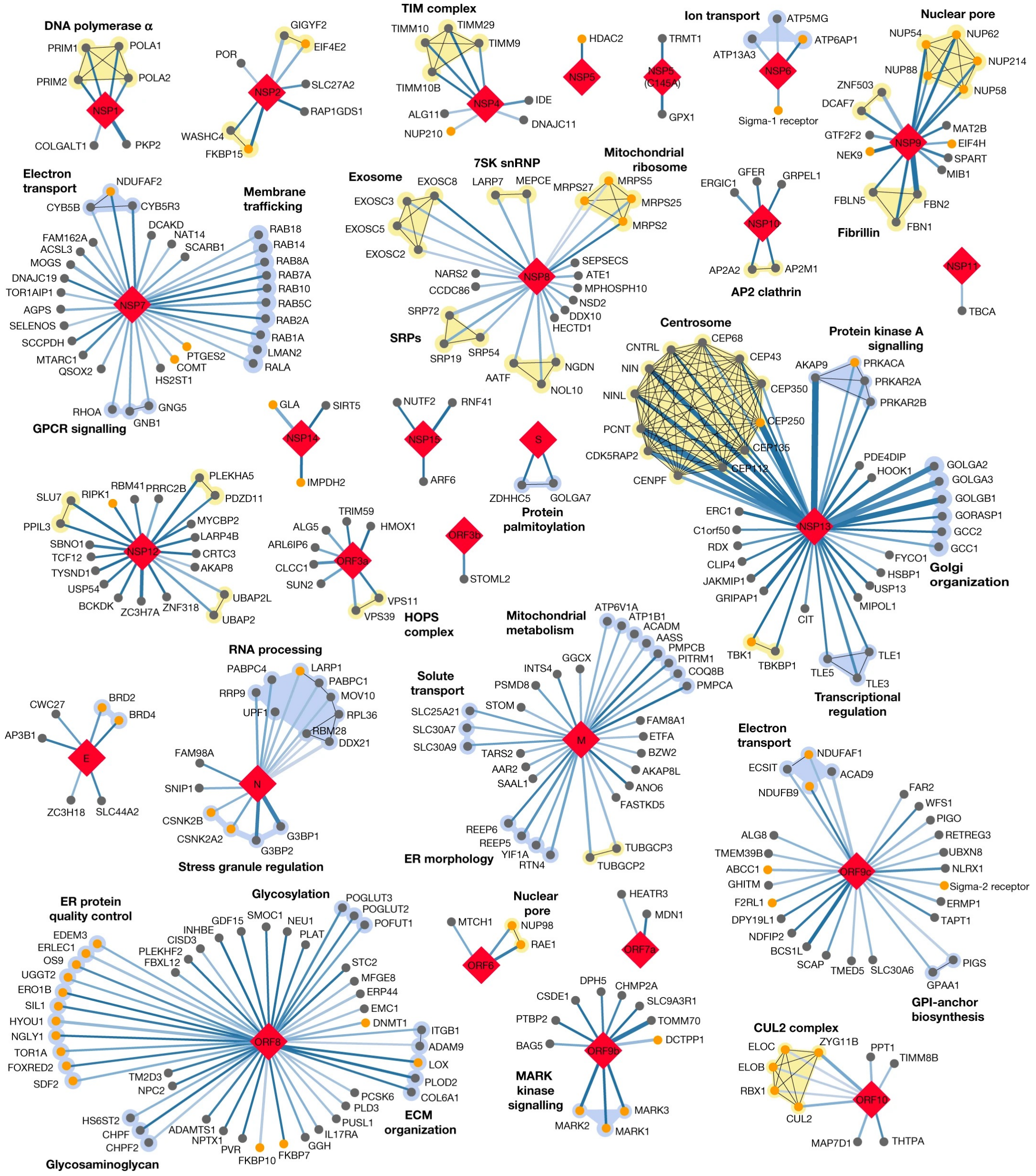
Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569,** 655–662 (2019)

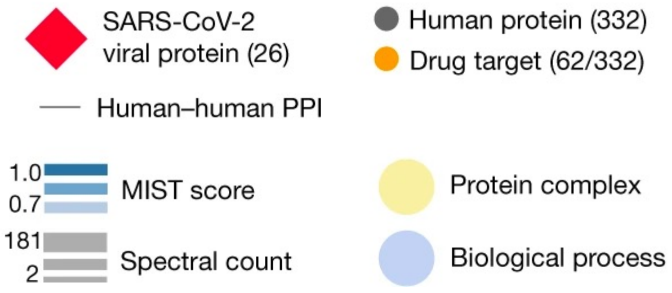
Why Networks?

Networks are powerful tools….

- Reduce complexity
- More efficient than tables
- Great for data integration
- Intuitive visualization

SARS=CoV-2 protein-protein interaction Network





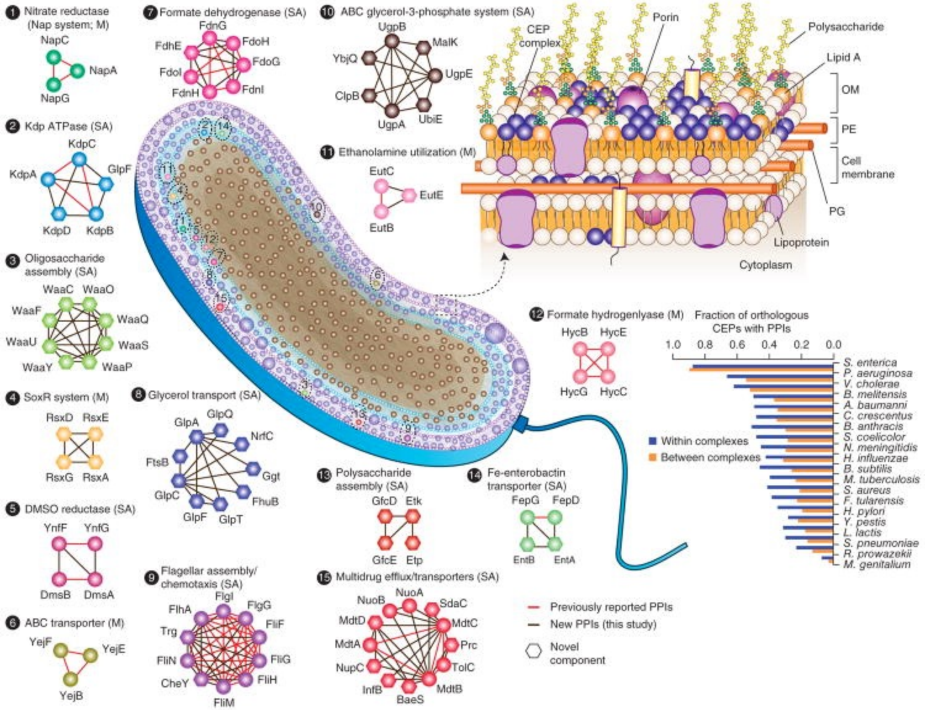
Gordon, D.E., Jang, G.M., Bouhaddou, M. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583,** 459–468 (2020).

Why Would We Use Network Visualization for Biological Data?

- Represent relationships of biological molecules
  - Physical, regulatory, genetic, functional interactions
- Useful for discovering relationships in large data sets
  - Better than tables in Excel
- Visualize multiple data types together
  - Discover interesting patterns
- Network analysis
  - Finding sub-networks with certain properties (densely connected, co-expressed, frequently mutated, clinical characteristics)
  - Finding paths between nodes (or other network “motifs”)
  - Finding central nodes in network topology (“hub” genes)

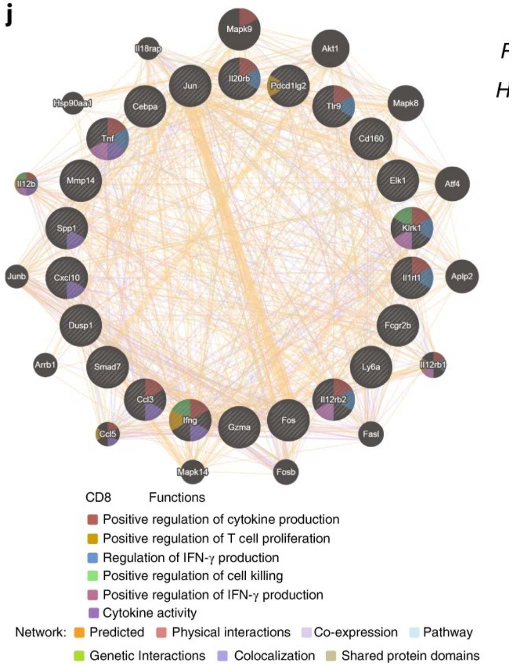
Applications of Network Biology

Detection of protein complexes/other modular structures – identification of subnetworks of interconnected nodes that might work together to perform a specific function, i.e. molecular complexes.



Babu M, Bundalovic-Torma C, Calmettes C, et al. Global landscape of cell envelope protein complexes in Escherichia coli. *Nat Biotechnol*. 2018;36(1):103-112

Gene Function Prediction – associate novel gene to existing function or pathway



Wang, Z., Aguilar, E.G., Luna, J.I. *et al.* Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. *Nat Med* **25,** 141–151 (2019)

Applications of Network Biology

Subnetwork based diagnosis – biomarker or biomarker panel. Identification of subnetworks activated in disease

Motif analysis– discover higher order organization such motifs, feedback loops (small network patterns that are over-represented when compared to a randomized version of the same network)

Network Alignment and comparison– analyze evolution of networks between different species to identify similarities and evolutionary relationships

Pathway based gene association – uncovering common pathways disturbed during disease.

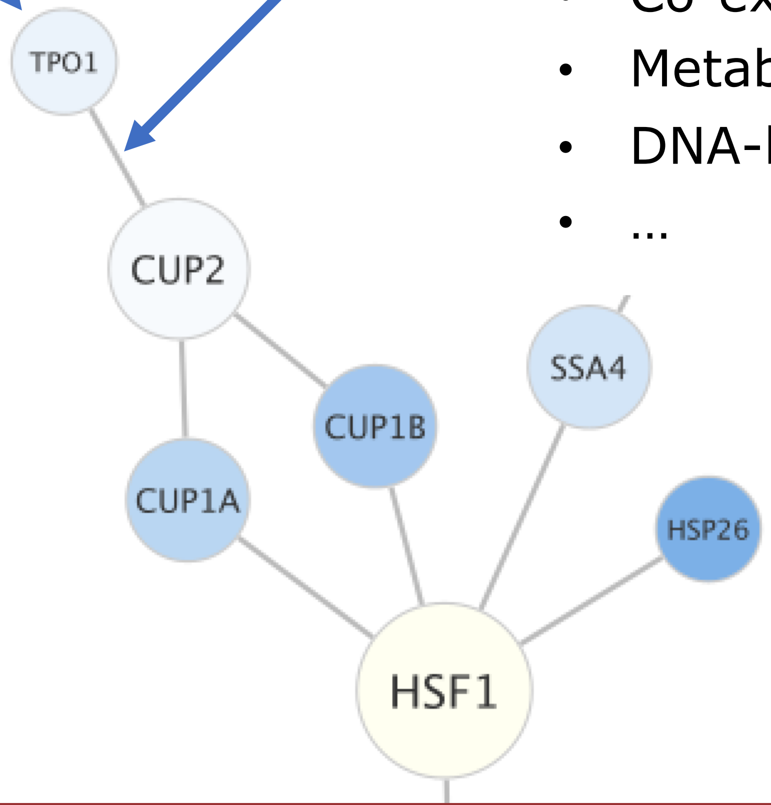
Network Basics

Node (molecule/entity)

- Gene
- Protein
- Transcript
- Drug
- MicroRNA
- …

Edge (interaction/relationship)

- Genetic interaction
- Physical protein interaction
- Co-expression
- Metabolic reaction
- DNA-binding
- …

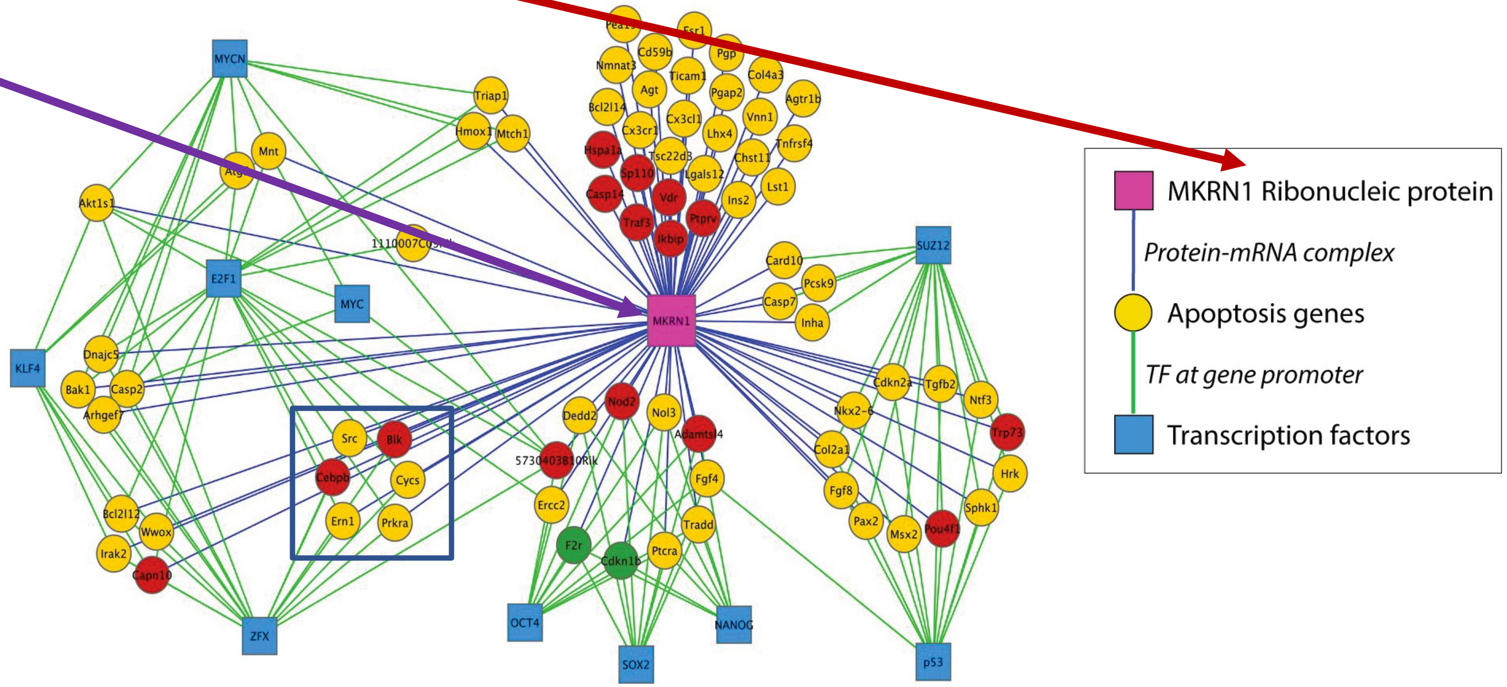


Networks as Tools

Analysis

- Topological properties
- Hubs and subnetworks
- Classify, cluster, diffuse and over-representation
- Data integration

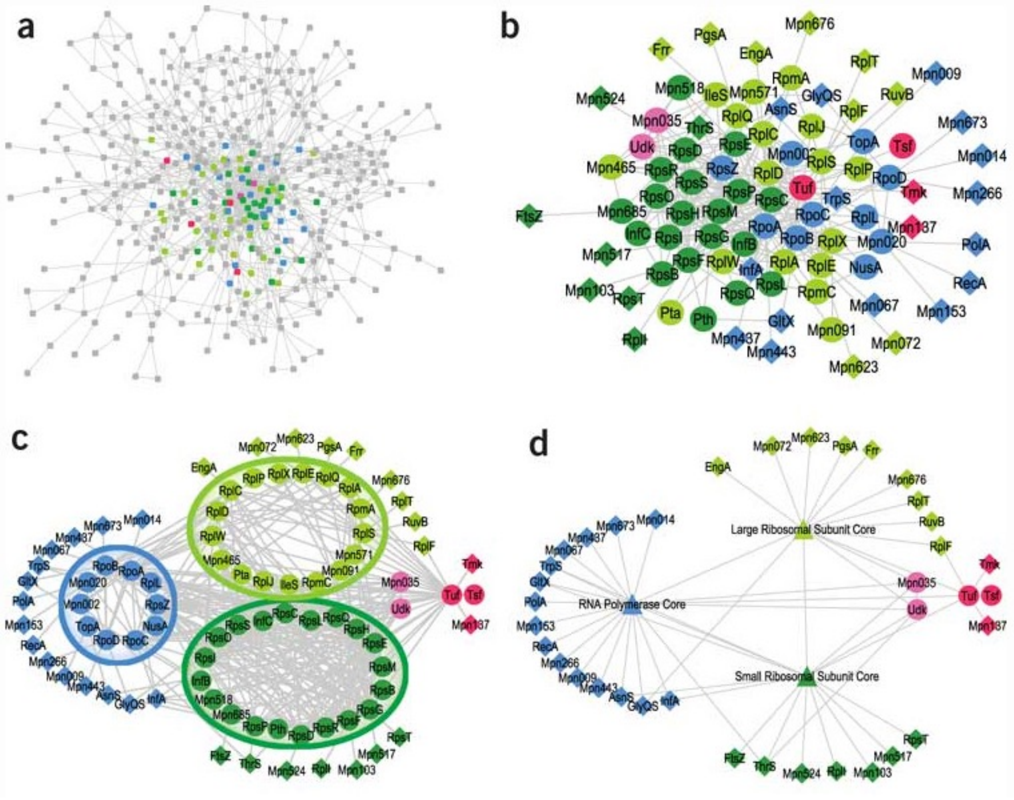
Topological properties:
Number of nodes and edges
Node degree
Degree distribution
Clustering coefficients



Networks as Tools

Visualization

- Data overlays – integrate multiple data types
- Layouts
- Exploratory analysis
- Context and interpretation



(a) Mass spectrometry analysis: 400 PPI in pneumonia microbe.
(b) Subnetwork with functional annotations
(c) Same network as (b) with added knowledge of complexes
(d) Same network as (c) with complexes collapsed to group nodes to reduce complexity

Gehlenborg N, O'Donoghue SI, Baliga NS, et al. Visualization of omics data for systems biology. *Nat Methods*. 2010;7(3 Suppl):S56-S68

What’s Missing?

- Dynamics
  - Pathways/networks represented as static processes
    - Difficult to represent a calcium wave or a feedback loop
  - More detailed mathematical representations exist that handle these e.g. Stoichiometric modeling, Kinetic modeling (VirtualCell, E-cell, …)
    - Need to accumulate or estimate comprehensive kinetic information
- Detail – atomic structures
- Context – cell type, developmental stage

What Have We Learned?

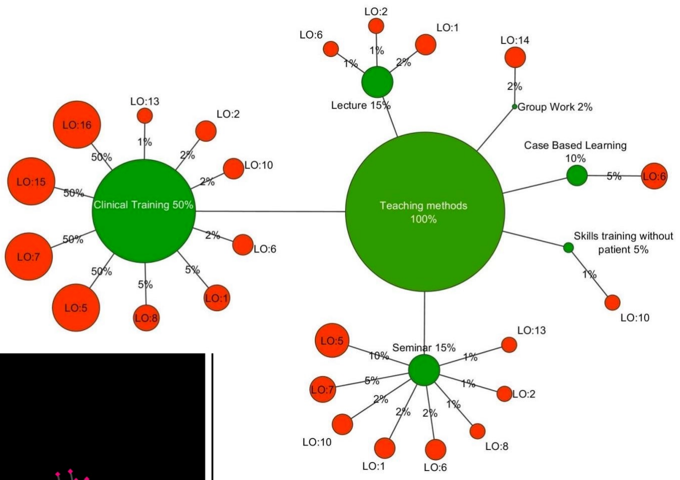
- Networks are useful for seeing relationships in large data sets
- Important to understand what the nodes and edges mean
- Important to define the biological question - know what you want to do with your gene list or network
- Many methods available for gene list and network analysis
  - Good to determine your question and search for a solution
  - Or get to know many methods and see how they can be applied to your data

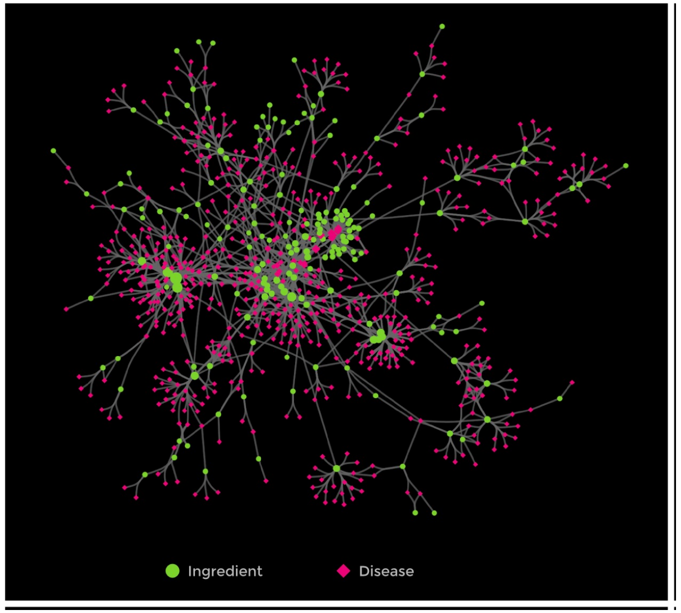
Network Visualization and Analysis using Cytoscape

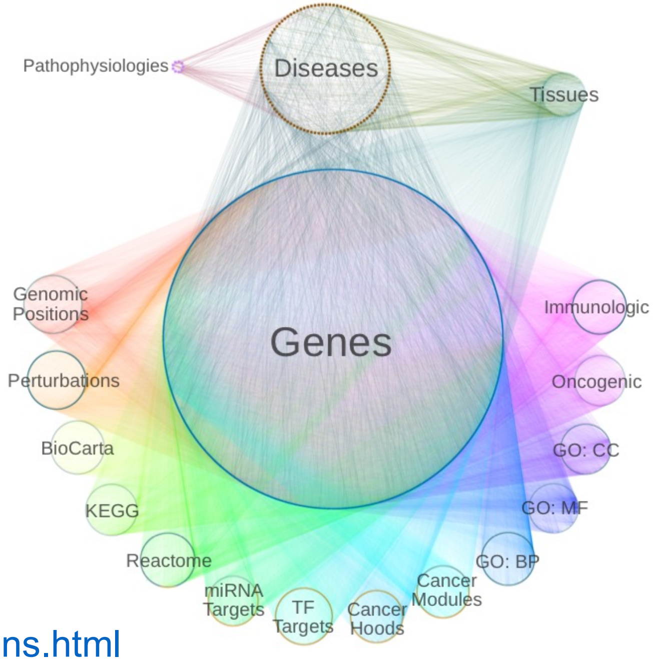
- Network visualization and analysis using Cytoscape software
- Cytoscape basics
- Cytoscape network analysis examples

Cytoscape (cytoscape.org)

- An open source software platform for visualizing and analyzing complex networks.







https://nrnb.org/competitions.html

Cytoscape (cytoscape.org)

Partners and Collaborators

GLADSTONE INSTITUTES

UCSF University of California San Francisco

UCSD

VE RI TAS HARVARD MEDICAL SCHOOL

DANA-FARBER CANCER INSTITUTE

UNIVERSITY OF TORONTO

Google

INDIANA UNIVERSITY

UNIVERSITY OF COPENHAGEN

Cancer Genome COLLABORATORY Cloud computing for collaborative research

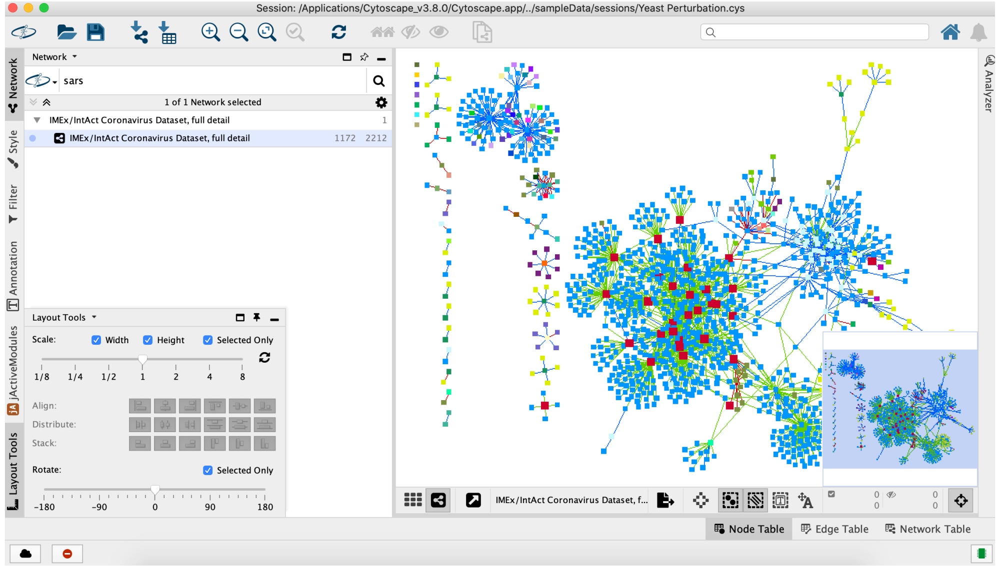
BROAD INSTITUTE

GENEONTOLOGY Unifying Biology

Pfizer

BOSTON UNIVERSITY

Manipulate Networks



Session: /Applications/Cytoscape_v3.8.0/Cytoscape.app/../sampleData/sessions/Yeast Perturbation.cys

Network search: sars

1 of 1 Network selected

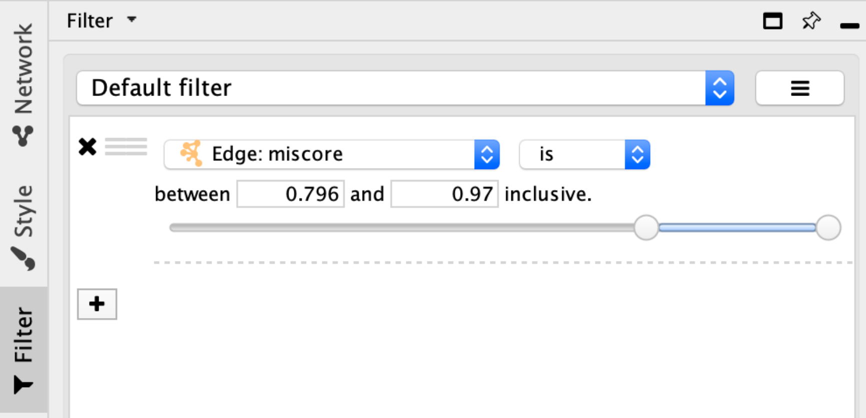
IMEx/IntAct Coronavirus Dataset, full detail	1	
IMEx/IntAct Coronavirus Dataset, full detail	1172	2212

Layout Tools

Scale: ☑ Width ☑ Height ☑ Selected Only

Rotate: ☑ Selected Only

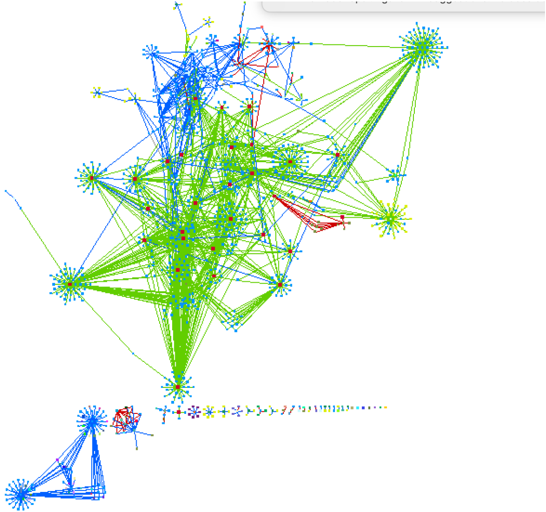
Filter/Query



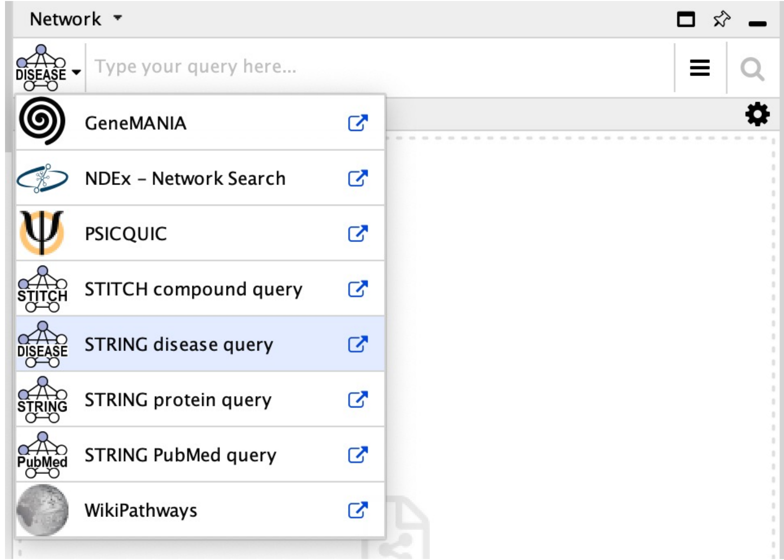
Default filter

Edge: miscore is between 0.796 and 0.97 inclusive.

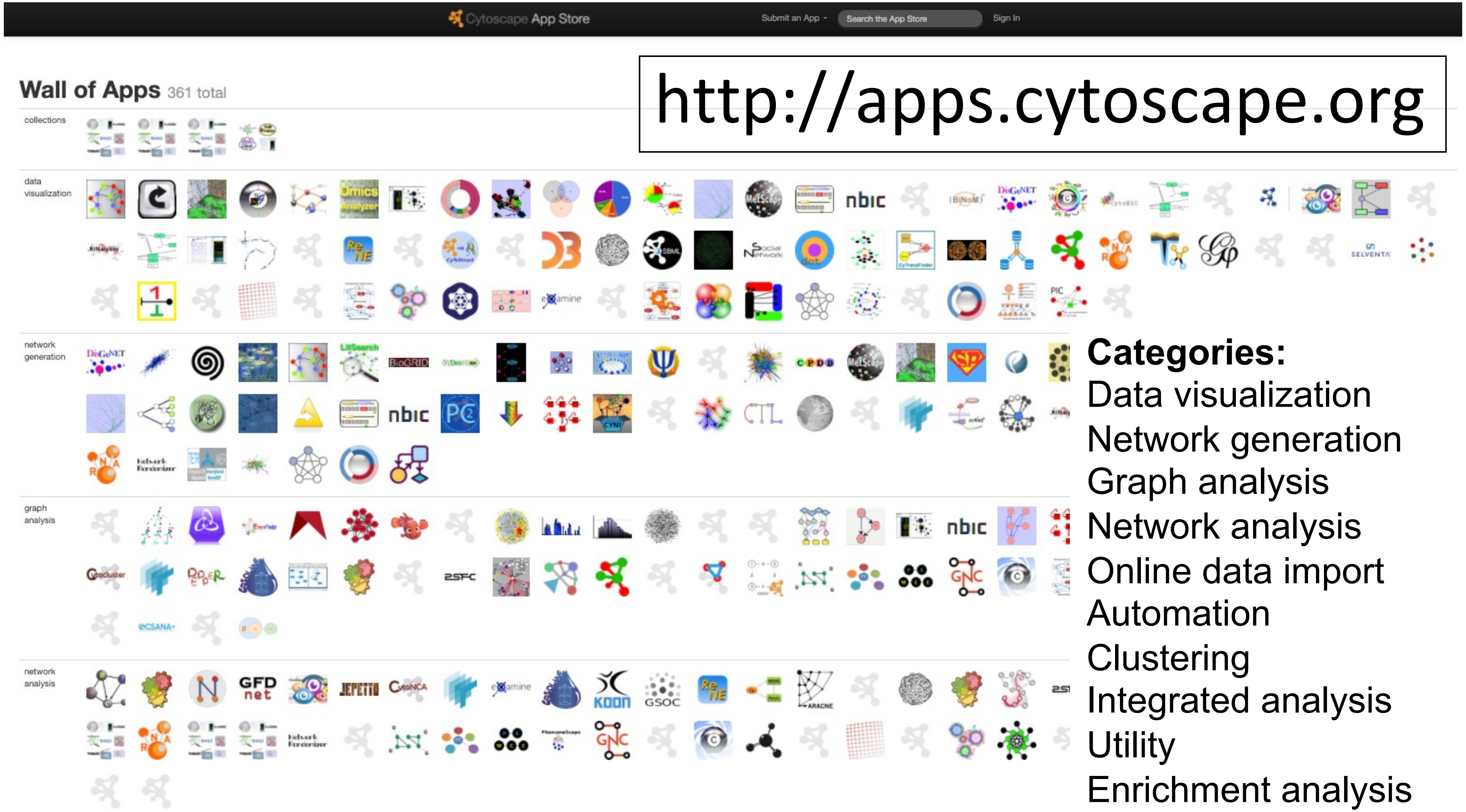
Automatic Layout



Network Import



The Cytoscape App Store



http://apps.cytoscape.org

Categories:
Data visualization
Network generation
Graph analysis
Network analysis
Online data import
Automation
Clustering
Integrated analysis
Utility
Enrichment analysis
Layout
Network Comparison

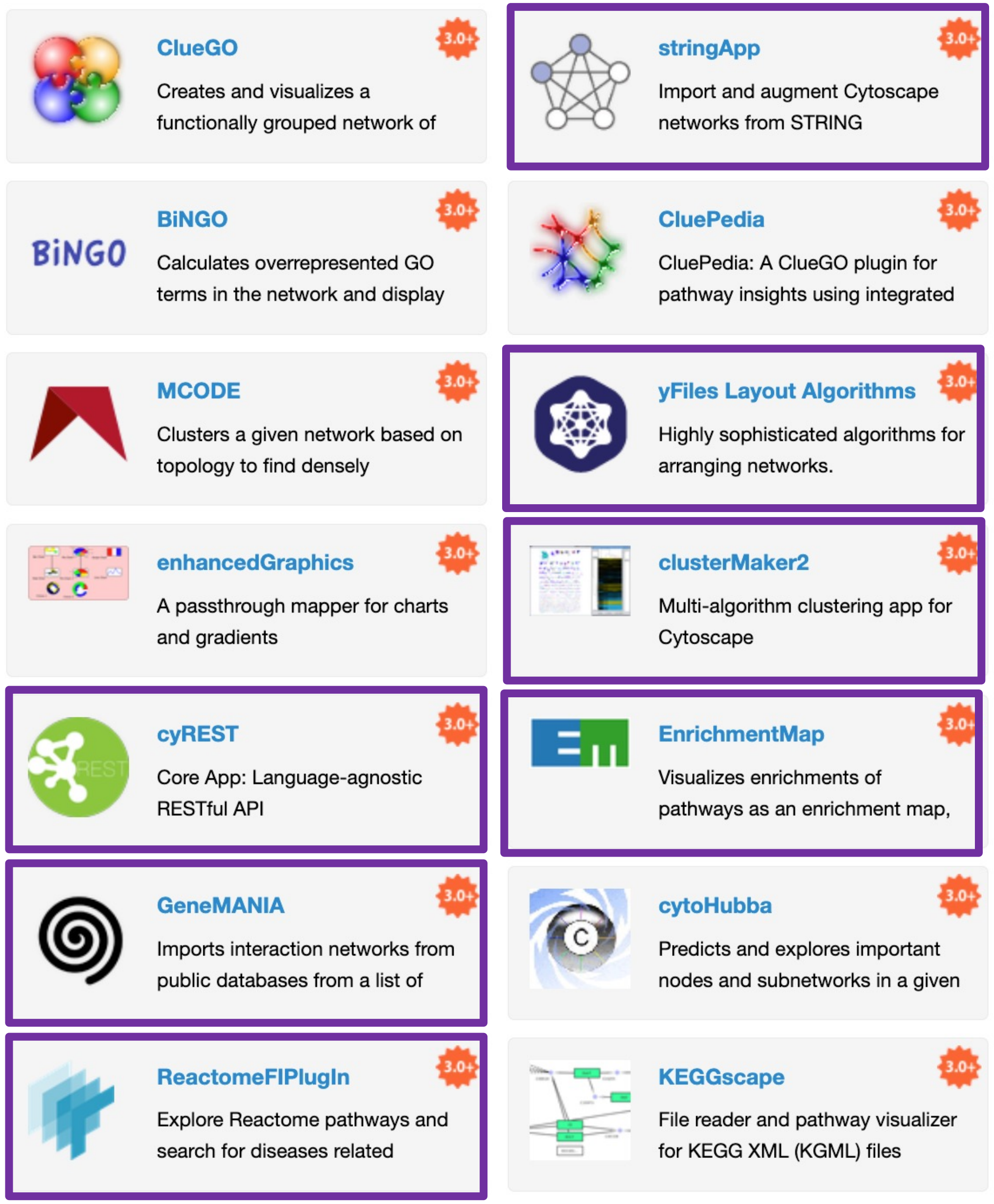
Active Community

http://www.cytoscape.org

- 10,000s users, >17,600 downloads/month (May 2018)
- Help
  - Documentation, data sets
  - Mailing lists
  - **http://tutorials.cytoscape.org**
- Automatable through R or python using cyrest
- Cited 2608 (google scholar), 1,756 (ISI web of Science) (May 2023)
- >367 active Apps Extend Functionality (April 2022)
  - Build your own, requires programming

Cline MS et al. Integration of biological networks and gene expression data using Cytoscape Nat Protoc. 2007;2(10):2366-82

Top Cytoscape Apps



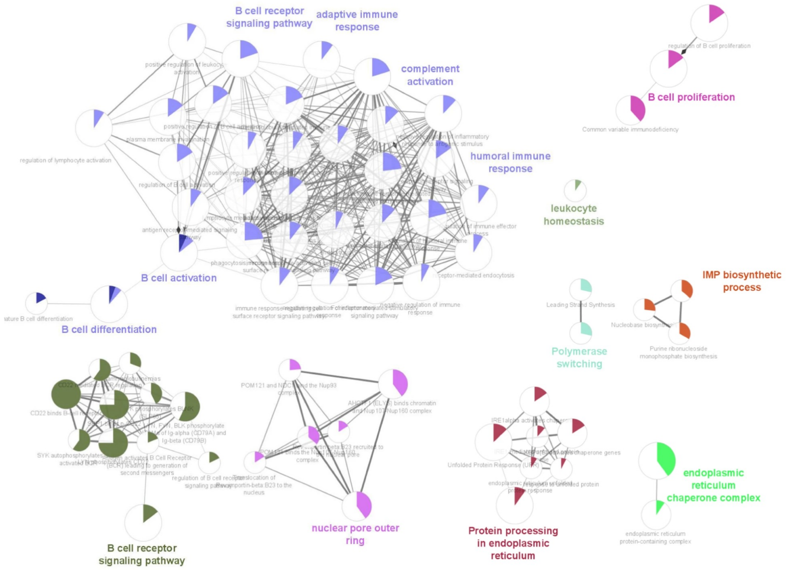
Published 2009 Citations: 2,711	**ClueGO** Creates and visualizes a functionally grouped network of	☐ **stringApp** Import and augment Cytoscape networks from STRING	Published 2019 Citations: 612
Published 2005 Citations: 2,034	**BiNGO** Calculates overrepresented GO terms in the network and display	**CluePedia** CluePedia: A ClueGO plugin for pathway insights using integrated	Published 2013 Citations: 524
Published 2003 Citations: 2,171	**MCODE** Clusters a given network based on topology to find densely	☐ **yFiles Layout Algorithms** Highly sophisticated algorithms for arranging networks.	
Core app Published 2014 Citations: 22	**enhancedGraphics** A passthrough mapper for charts and gradients	☐ **clusterMaker2** Multi-algorithm clustering app for Cytoscape	Published 2011 Citations: 259
Core app Published 2015 Citations: 33	☐ **cyREST** Core App: Language-agnostic RESTful API	☐ **EnrichmentMap** Visualizes enrichments of pathways as an enrichment map,	Published 2010 Citations: 1,080*
Published 2010 Citations: 297*	☐ **GeneMANIA** Imports interaction networks from public databases from a list of	**cytoHubba** Predicts and explores important nodes and subnetworks in a given	Published 2014 Citations: 1,699
Published 2010 Citations: 335	☐ **ReactomeFIPlugIn** Explore Reactome pathways and search for diseases related	**KEGGscape** File reader and pathway visualizer for KEGG XML (KGML) files	Published 2014 Citations: 38

* Citations based on single publication with multiple primary publications available

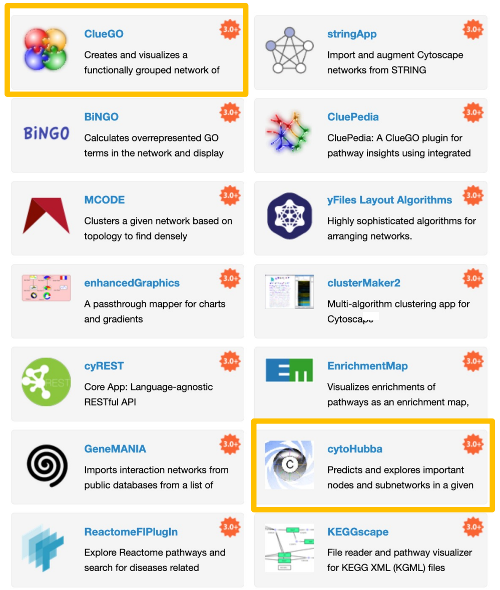
☐ Apps we covered in workshop

(Updated May 2023)

Example of Top Cytoscape Apps not covered in workshop



Riese, P., Trittel, S., Akmatov, M.K. *et al.* Distinct immunological and molecular signatures underpinning influenza vaccine responsiveness in the elderly. *Nat Commun* **13**, 6894 (2022).



ClueGO
Creates and visualizes a functionally grouped network of

stringApp
Import and augment Cytoscape networks from STRING

BiNGO
Calculates overrepresented GO terms in the network and display

CluePedia
CluePedia: A ClueGO plugin for pathway insights using integrated

MCODE
Clusters a given network based on topology to find densely

yFiles Layout Algorithms
Highly sophisticated algorithms for arranging networks.

enhancedGraphics
A passthrough mapper for charts and gradients

clusterMaker2
Multi-algorithm clustering app for Cytoscape

cyREST
Core App: Language-agnostic RESTful API

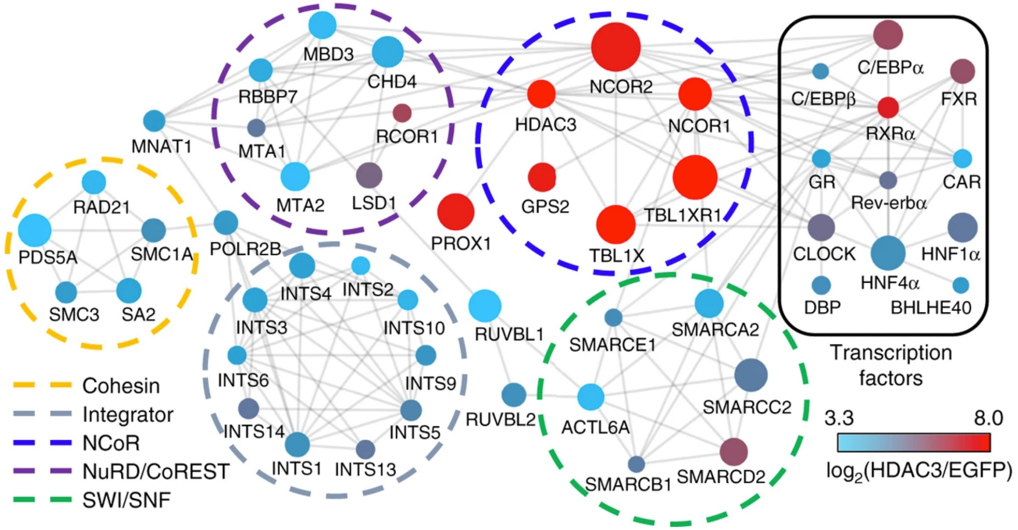
EnrichmentMap
Visualizes enrichments of pathways as an enrichment map,

GeneMANIA
Imports interaction networks from public databases from a list of

cytoHubba
Predicts and explores important nodes and subnetworks in a given

ReactomeFIPlugIn
Explore Reactome pathways and search for diseases related

KEGGscape
File reader and pathway visualizer for KEGG XML (KGML) files



Armour, S.M., Remsberg, J.R., Damle, M. *et al.* An HDAC3-PROX1 corepressor module acts on HNF4α to control hepatic triglycerides. *Nat Commun* **8**, 549 (2017). https://doi.org/10.1038/s41467-017-00772-5

What Have We Learned?

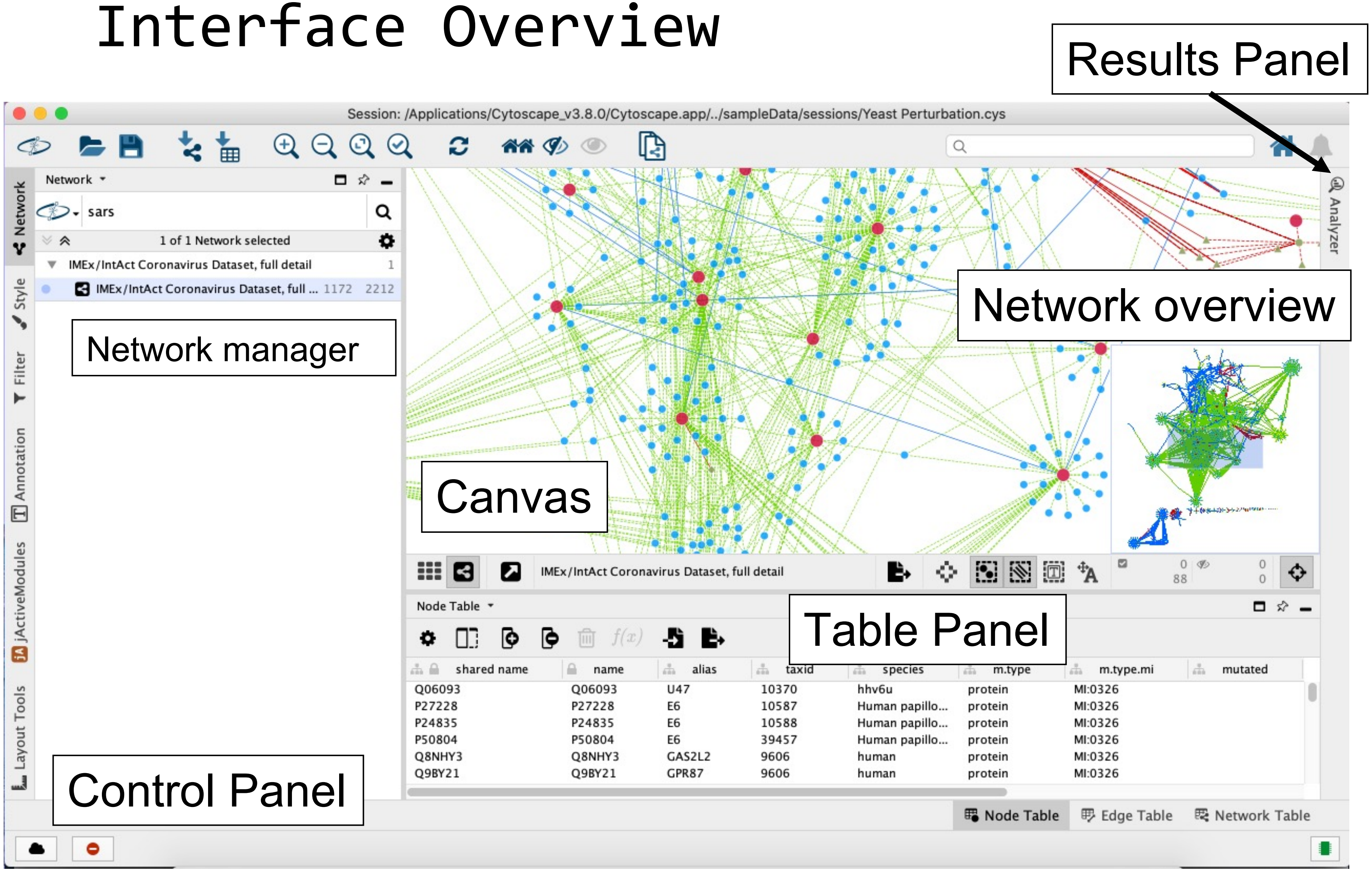
- Cytoscape is a useful, free software tool for network visualization and analysis
- Provides basic network manipulation features
- Apps are available to extend the functionality

Cytoscape Demo

Version 3.8.0

www.cytoscape.org

Interface Overview



Session: /Applications/Cytoscape_v3.8.0/Cytoscape.app/../sampleData/sessions/Yeast Perturbation.cys

sars

1 of 1 Network selected

IMEx/IntAct Coronavirus Dataset, full detail		1
IMEx/IntAct Coronavirus Dataset, full ...	1172	2212

IMEx/IntAct Coronavirus Dataset, full detail

Node Table

shared name	name	alias	taxid	species	m.type	m.type.mi	mutated
Q06093	Q06093	U47	10370	hhv6u	protein	MI:0326	
P27228	P27228	E6	10587	Human papillo...	protein	MI:0326	
P24835	P24835	E6	10588	Human papillo...	protein	MI:0326	
P50804	P50804	E6	39457	Human papillo...	protein	MI:0326	
Q8NHY3	Q8NHY3	GAS2L2	9606	human	protein	MI:0326	
Q9BY21	Q9BY21	GPR87	9606	human	protein	MI:0326	

Results Panel

Network overview

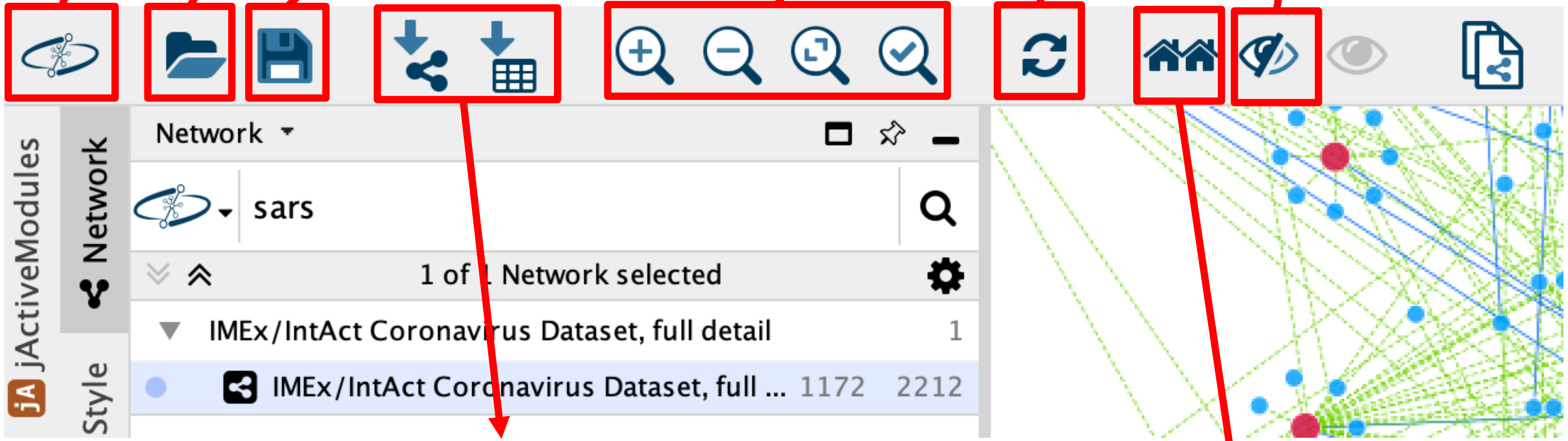
Network manager

Canvas

Table Panel

Control Panel

Basic Navigation



Open or save network to NDex

Open a session

Save your session

Zoom in/out

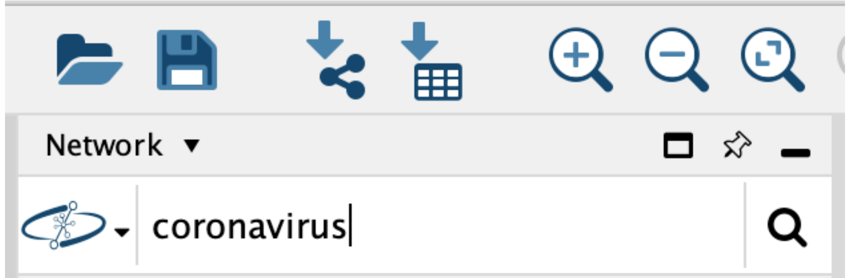
Apply preferred layout

Hide selected
nodes and edges

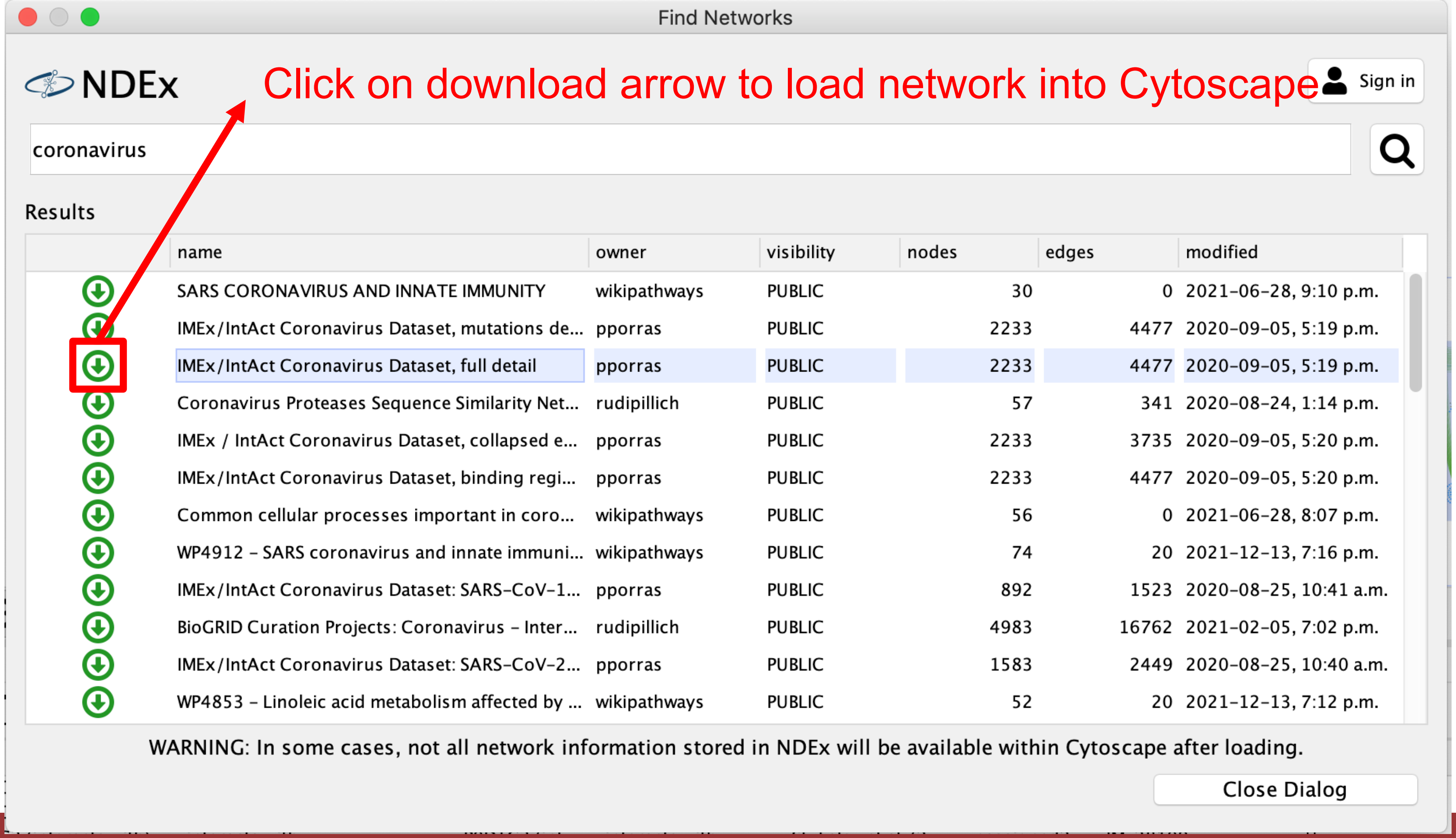
Import Network/table from file

Select first neighbours of
selected nodes

Load a Network



Enter a search term into Network search bar. I am using Ndex- (https://home.ndexbio.org/index/)



Click on download arrow to load network into Cytoscape

coronavirus

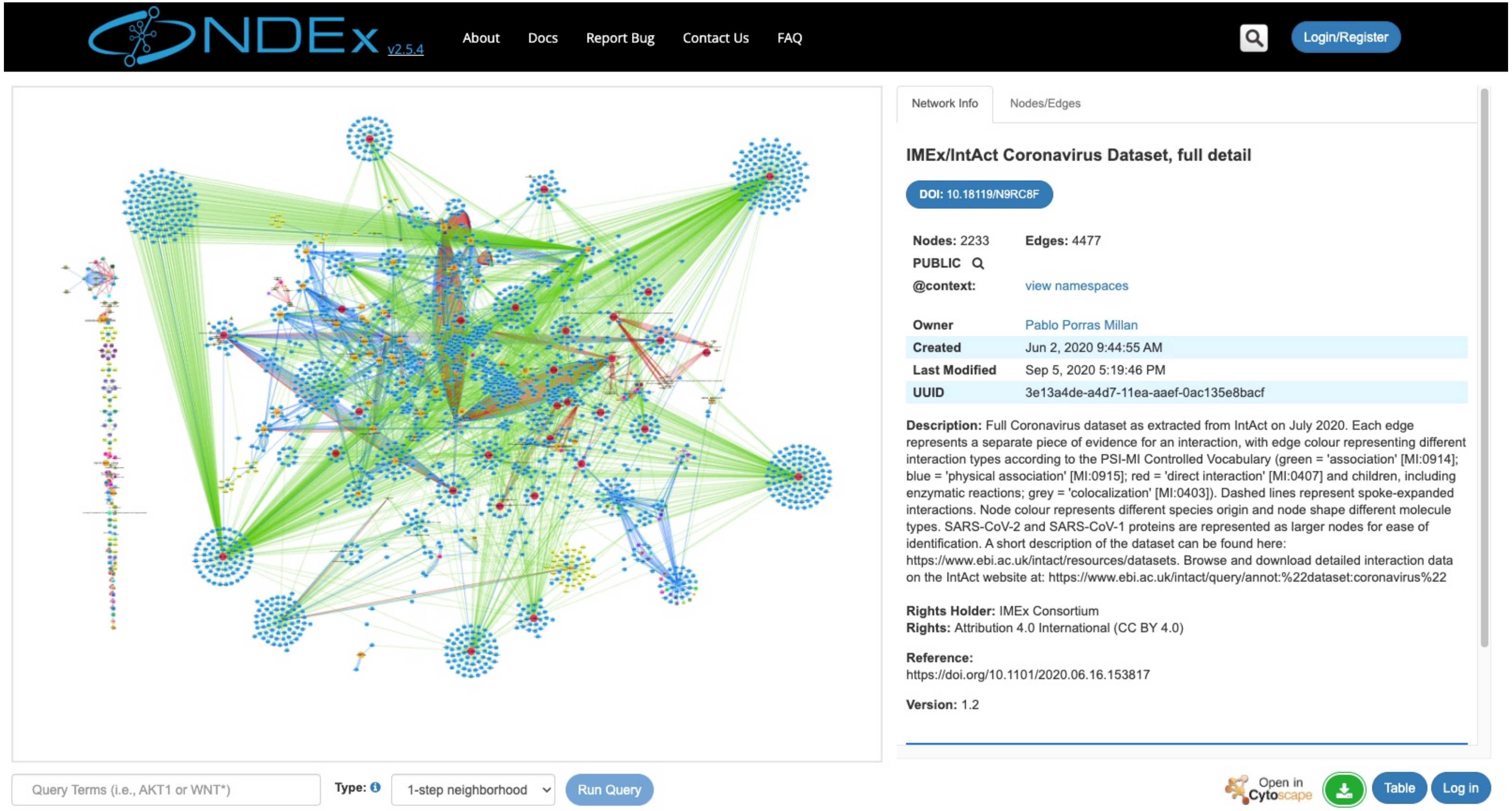
Results

	name	owner	visibility	nodes	edges	modified
⬇	SARS CORONAVIRUS AND INNATE IMMUNITY	wikipathways	PUBLIC	30	0	2021-06-28, 9:10 p.m.
⬇	IMEx/IntAct Coronavirus Dataset, mutations de...	pporras	PUBLIC	2233	4477	2020-09-05, 5:19 p.m.
⬇	IMEx/IntAct Coronavirus Dataset, full detail	pporras	PUBLIC	2233	4477	2020-09-05, 5:19 p.m.
⬇	Coronavirus Proteases Sequence Similarity Net...	rudipillich	PUBLIC	57	341	2020-08-24, 1:14 p.m.
⬇	IMEx / IntAct Coronavirus Dataset, collapsed e...	pporras	PUBLIC	2233	3735	2020-09-05, 5:20 p.m.
⬇	IMEx/IntAct Coronavirus Dataset, binding regi...	pporras	PUBLIC	2233	4477	2020-09-05, 5:20 p.m.
⬇	Common cellular processes important in coro...	wikipathways	PUBLIC	56	0	2021-06-28, 8:07 p.m.
⬇	WP4912 – SARS coronavirus and innate immuni...	wikipathways	PUBLIC	74	20	2021-12-13, 7:16 p.m.
⬇	IMEx/IntAct Coronavirus Dataset: SARS-CoV-1...	pporras	PUBLIC	892	1523	2020-08-25, 10:41 a.m.
⬇	BioGRID Curation Projects: Coronavirus – Inter...	rudipillich	PUBLIC	4983	16762	2021-02-05, 7:02 p.m.
⬇	IMEx/IntAct Coronavirus Dataset: SARS-CoV-2...	pporras	PUBLIC	1583	2449	2020-08-25, 10:40 a.m.
⬇	WP4853 – Linoleic acid metabolism affected by ...	wikipathways	PUBLIC	52	20	2021-12-13, 7:12 p.m.

WARNING: In some cases, not all network information stored in NDEx will be available within Cytoscape after loading.

About the network

http://ndexbio.org/#/network/3e13a4de-a4d7-11ea-aaef-0ac135e8bacf



IMEx/IntAct Coronavirus Dataset, full detail

DOI: 10.18119/N9RC8F

Nodes: 2233 **Edges:** 4477

PUBLIC

@context: view namespaces

Owner	Pablo Porras Millan
Created	Jun 2, 2020 9:44:55 AM
Last Modified	Sep 5, 2020 5:19:46 PM
UUID	3e13a4de-a4d7-11ea-aaef-0ac135e8bacf

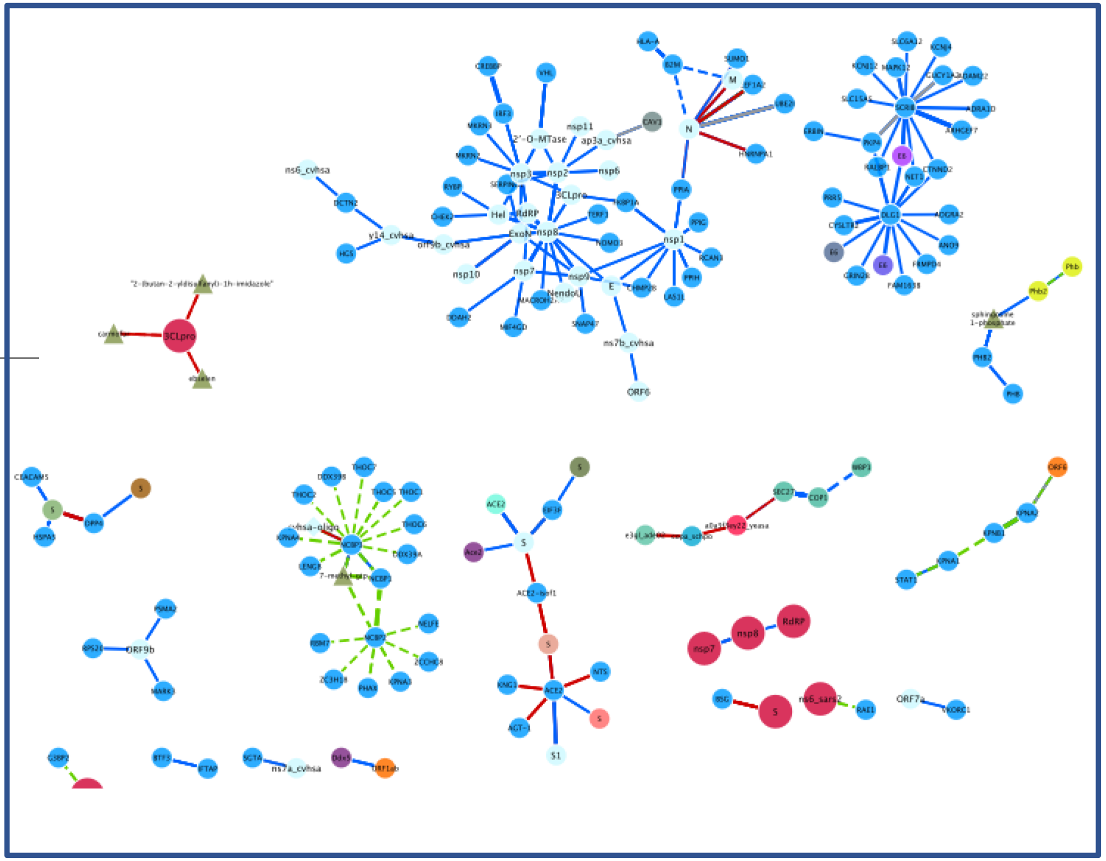
Description: Full Coronavirus dataset as extracted from IntAct on July 2020. Each edge represents a separate piece of evidence for an interaction, with edge colour representing different interaction types according to the PSI-MI Controlled Vocabulary (green = 'association' [MI:0914]; blue = 'physical association' [MI:0915]; red = 'direct interaction' [MI:0407] and children, including enzymatic reactions; grey = 'colocalization' [MI:0403]). Dashed lines represent spoke-expanded interactions. Node colour represents different species origin and node shape different molecule types. SARS-CoV-2 and SARS-CoV-1 proteins are represented as larger nodes for ease of identification. A short description of the dataset can be found here: https://www.ebi.ac.uk/intact/resources/datasets. Browse and download detailed interaction data on the IntAct website at: https://www.ebi.ac.uk/intact/query/annot:%22dataset:coronavirus%22

Rights Holder: IMEx Consortium
Rights: Attribution 4.0 International (CC BY 4.0)

Reference:
https://doi.org/10.1101/2020.06.16.153817

Version: 1.2

Visual Styles



Node Properties

Node Shape Mapping

Node Shape	m.type
◇	messenger rna
◇	micro rna
◇	nucleic acid
○	peptide
○	protein
◇	ribonucleic acid
△	small molecule
◇	small nuclear rna

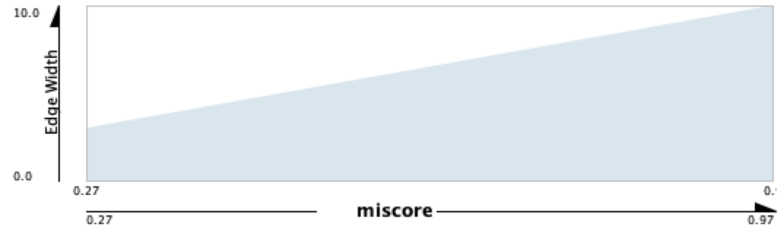
Node Fill Color Mapping

Node Fill Color	species
🟥	SARS-CoV-2
🟦	human
🟨	mouse
🟪	rat

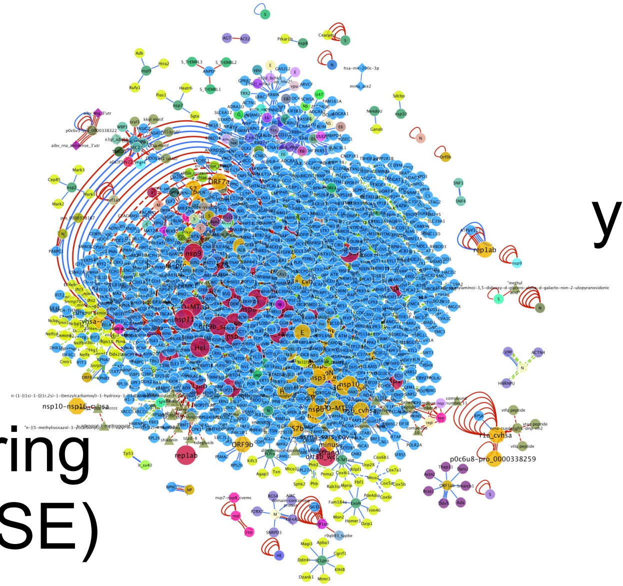
Edge Stroke Color (Unselected) Mapping

Edge Stroke Color (Unselected)	int.type
🟩	association
🟥	cleavage reaction
◼	colocalization
🟥	covalent binding
🟥	direct interaction
🟥	hydroxylation reaction
🟦	physical association
🟥	protein cleavage
🟥	rna cleavage

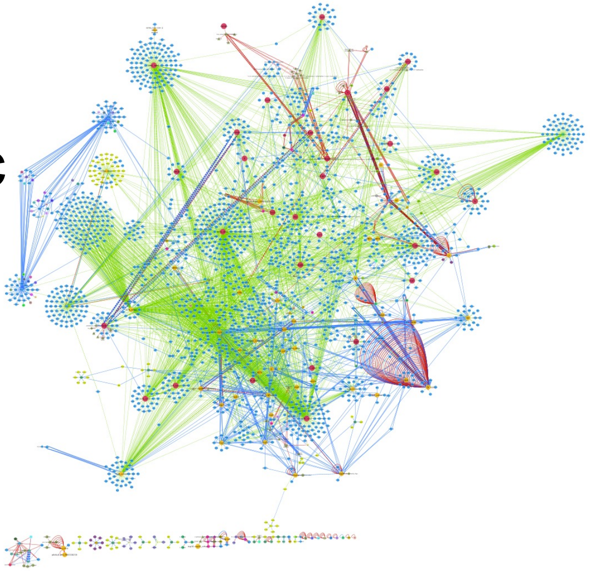
Edge Width Mapping



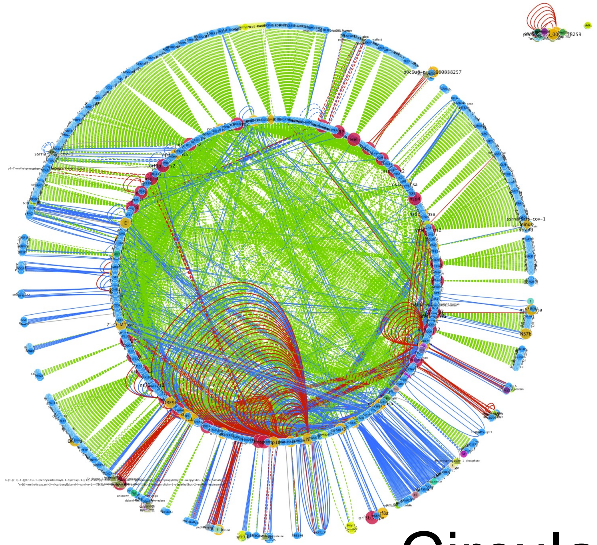
Experiment with different layouts



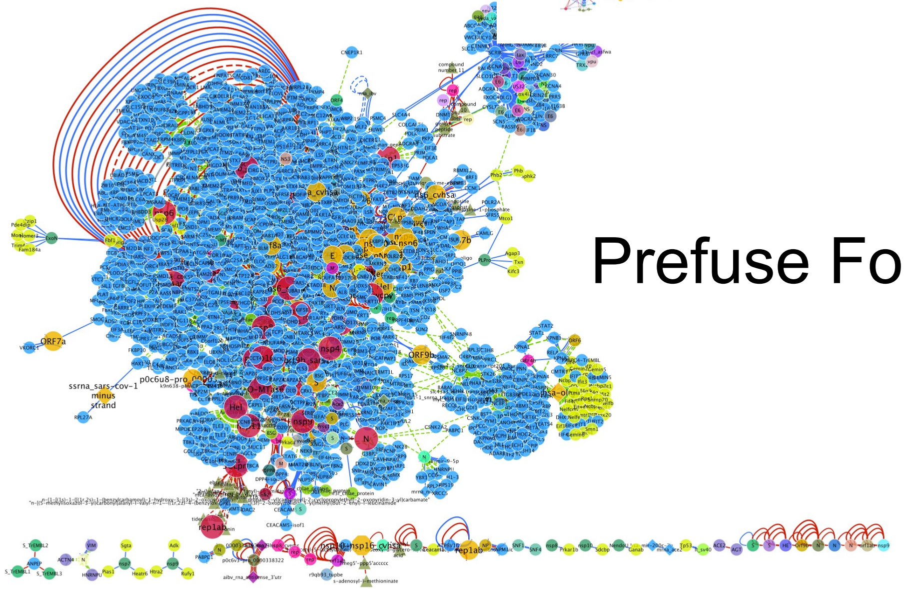
Compound Spring Embedder (CoSE)



yfiles organic

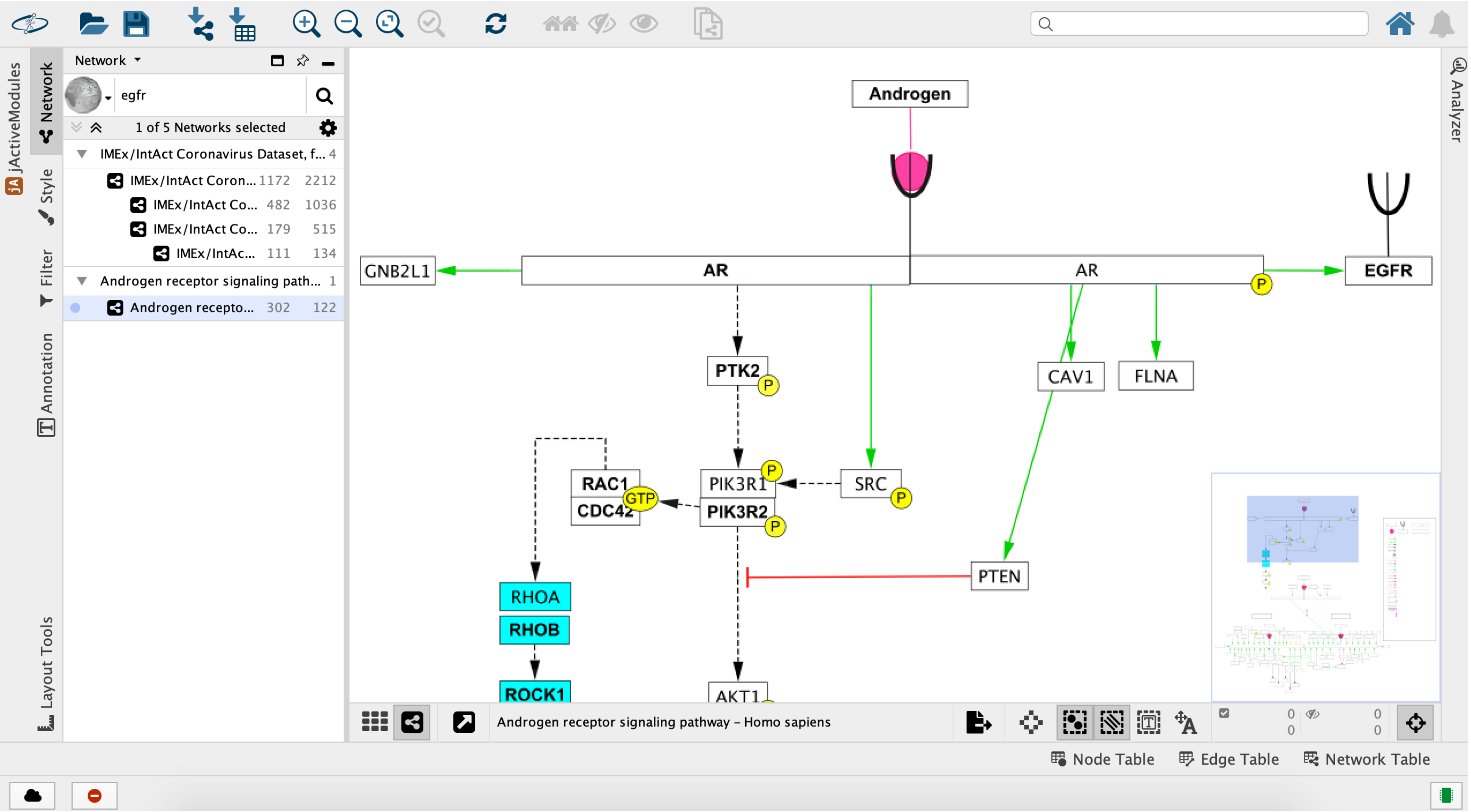


Circular Layout



Prefuse Force directed

Load different types of networks



egfr

1 of 5 Networks selected

IMEx/IntAct Coronavirus Dataset, f...		4
IMEx/IntAct Coron...	1172	2212
IMEx/IntAct Co...	482	1036
IMEx/IntAct Co...	179	515
IMEx/IntAc...	111	134
Androgen receptor signaling path...		1
Androgen recepto...	302	122

Androgen receptor signaling pathway – Homo sapiens

Androgen Receptor Pathway from WikiPathways (WP138) – https://www.wikipathways.org/index.php/Pathway:WP138

We are on a Coffee Break & Networking Session

Workshop Sponsors:

Canadian Centre for Computational Genomics

HPC4Health

OICR Ontario Institute for Cancer Research

GenomeCanada