



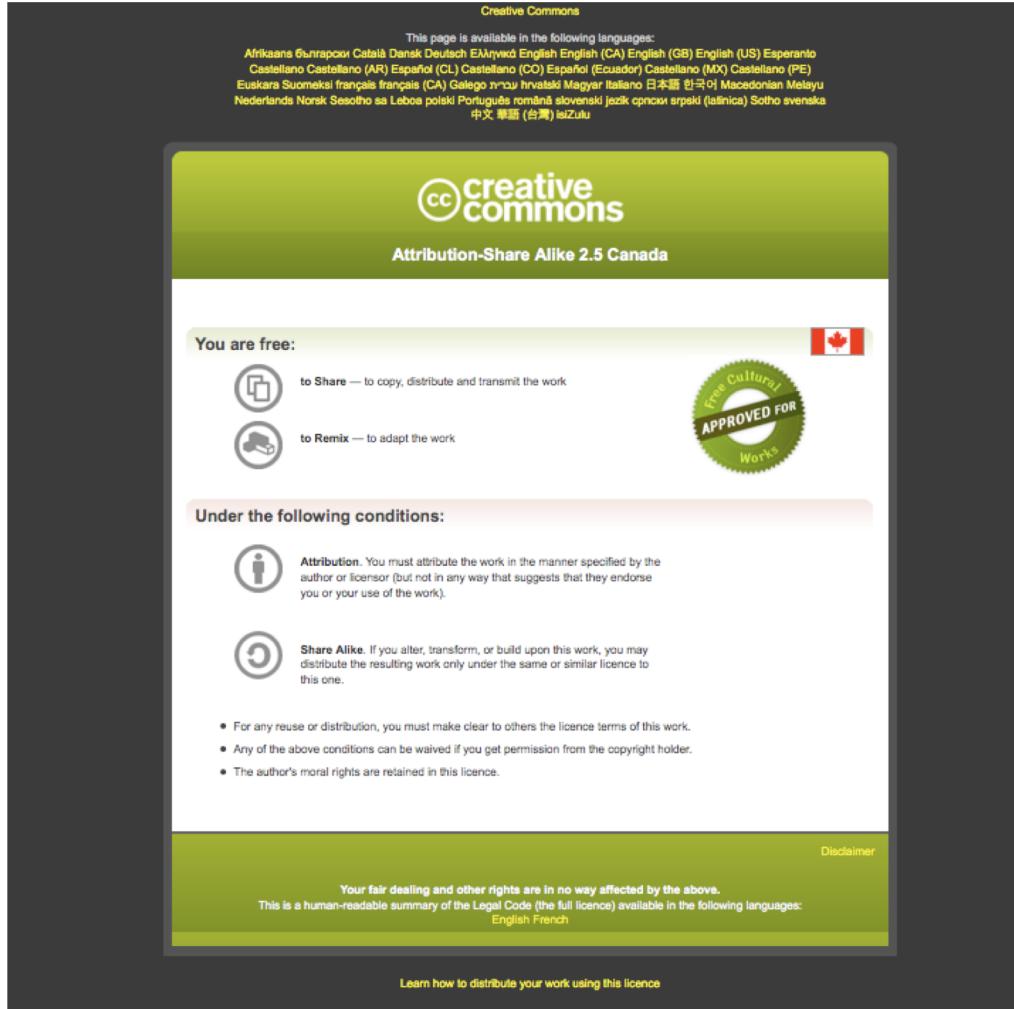
Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by





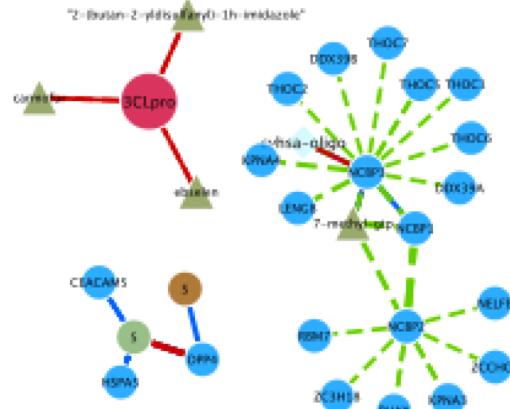
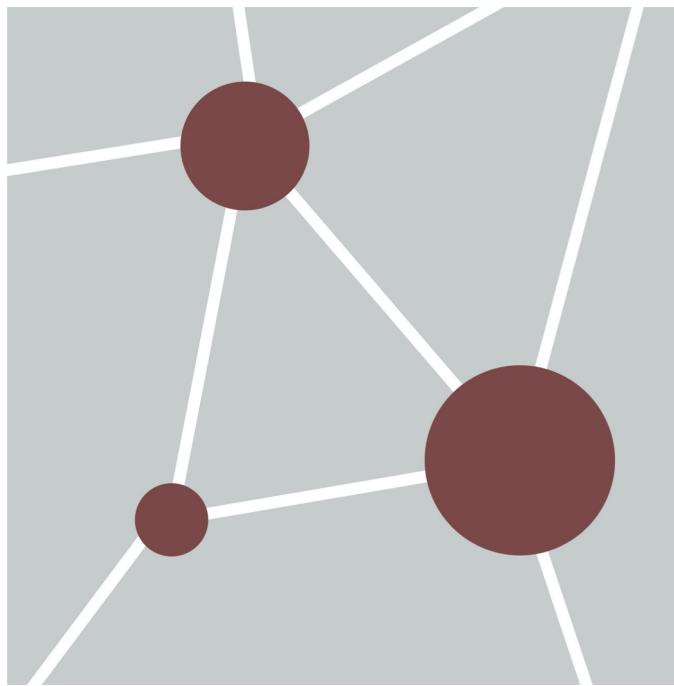
Module 3

Network Visualization and analysis with Cytoscape

Ruth Isserlin

Pathway and Network Analysis of -omics Data

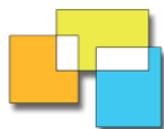
July 27-29, 2020



**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

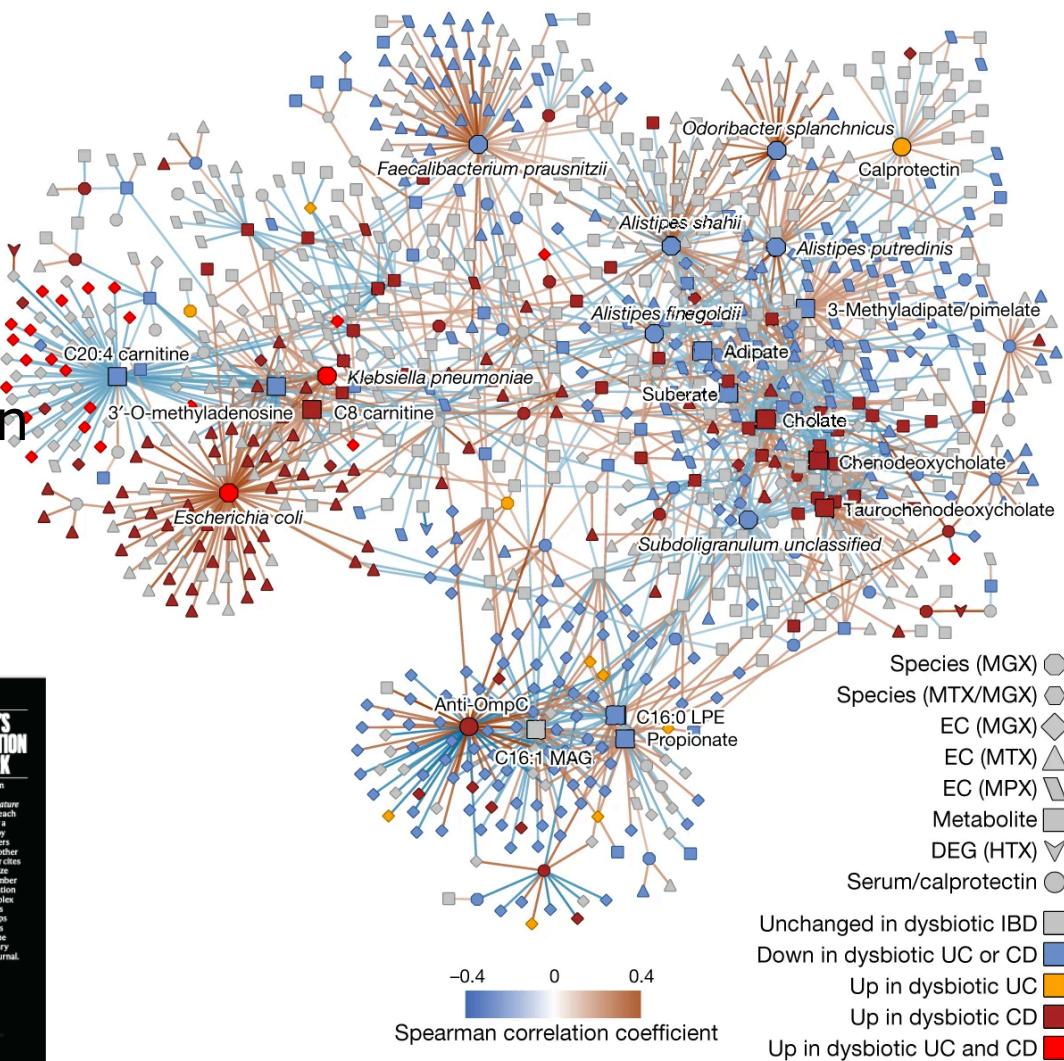
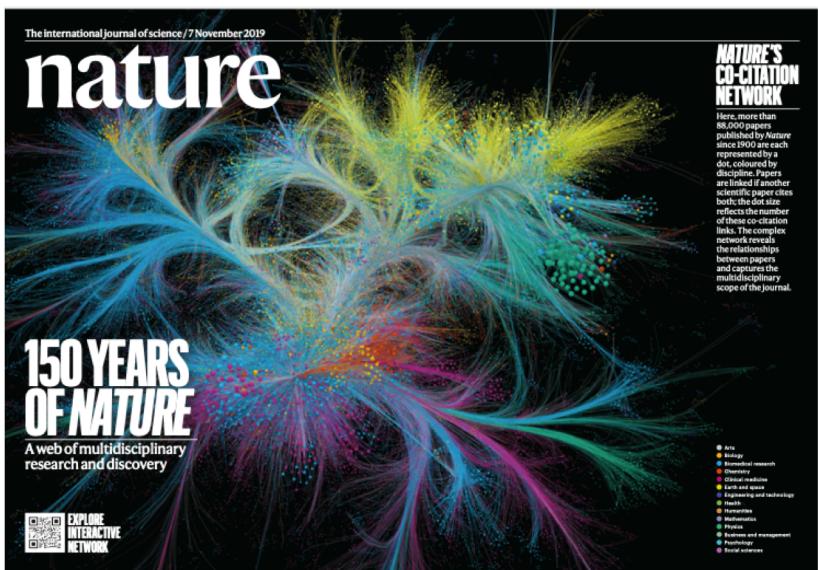
Six Degrees of Separation

- 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

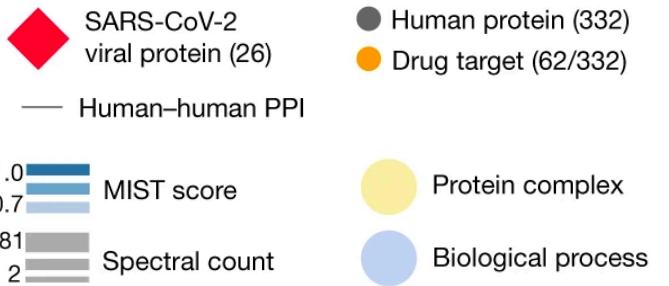


Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019)

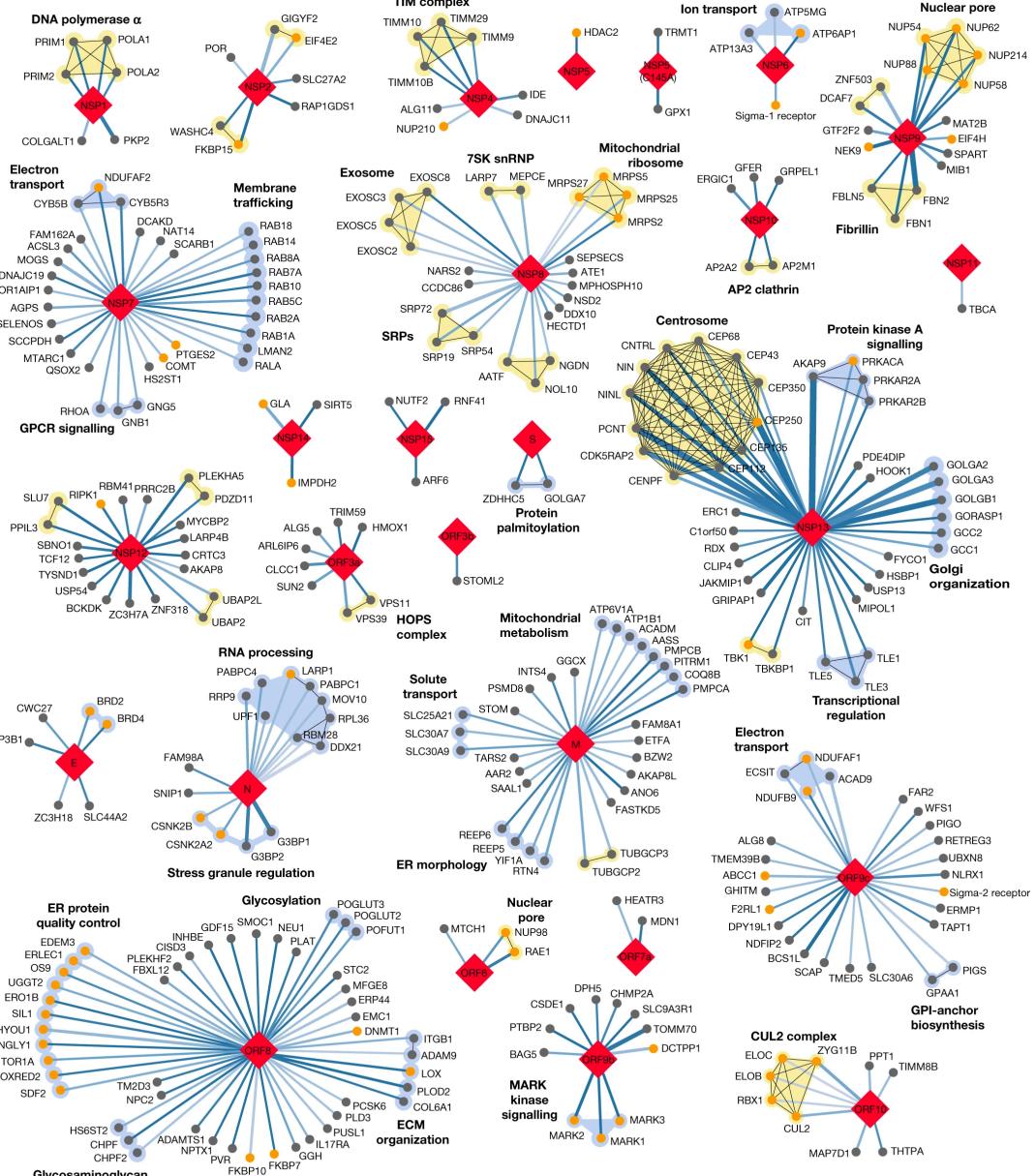
Why Networks?

**Networks are
powerful tools....**

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



SARS-CoV-2 protein-protein interaction Network



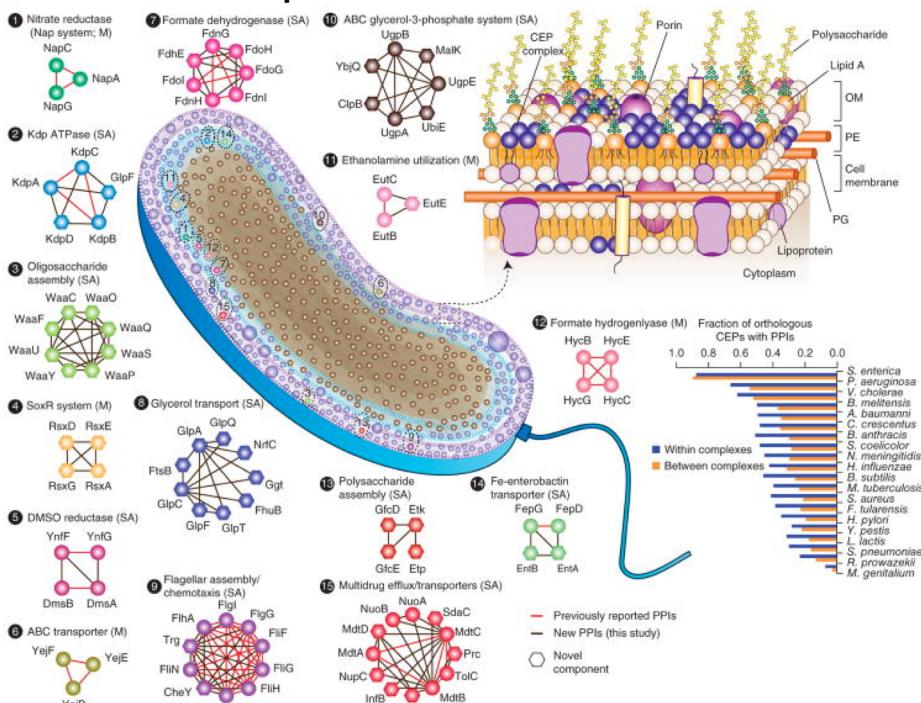
Gordon, D.E., Jang, G.M., Bouhaddou, M. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).

Why Would We Use Network Visualization for Biological Data?

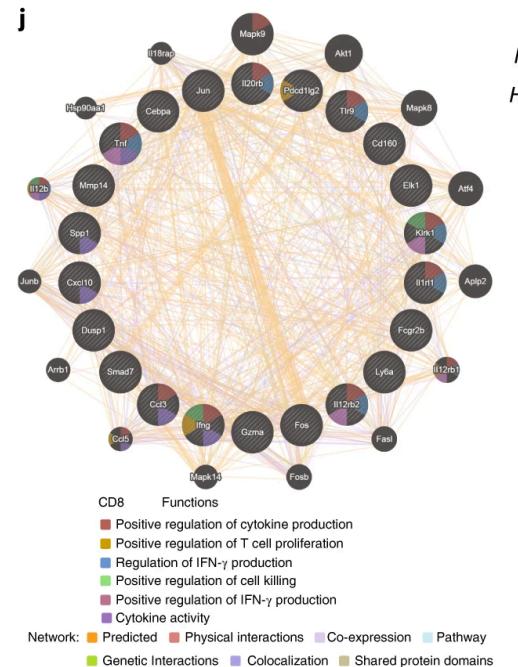
- Represent relationships of biological molecules
 - Physical, regulatory, genetic, functional interactions
- Useful for discovering relationships in large data sets
 - Better than tables in Excel
- Visualize multiple data types together
 - Discover interesting patterns
- Network analysis
 - Finding 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.



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)

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

Module 3: Network Visualization with Cytoscape

bioinformatics.ca

Applications of Network Biology

Subnetwork based diagnosis –

biomarker or biomarker panel.

Identification of subnetworks activated in disease

Network Alignment and comparison –

analyze evolution of networks between different species to identify similarities and evolutionary relationships

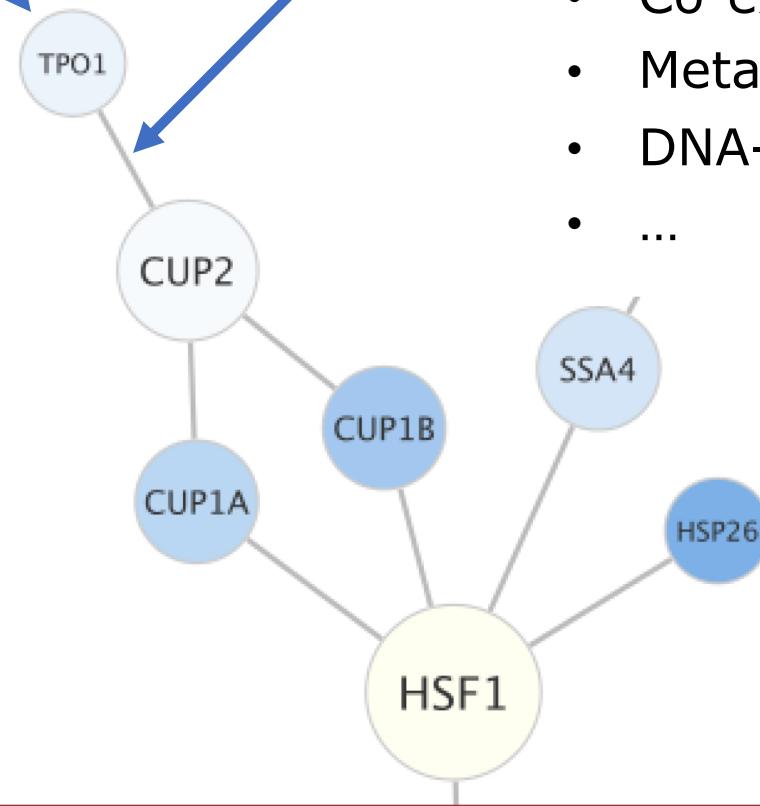
Motif analysis – discover higher order organization such motifs, feedback loops (small network patterns that are over-represented when compared to a randomized version of the same network)

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
- ...

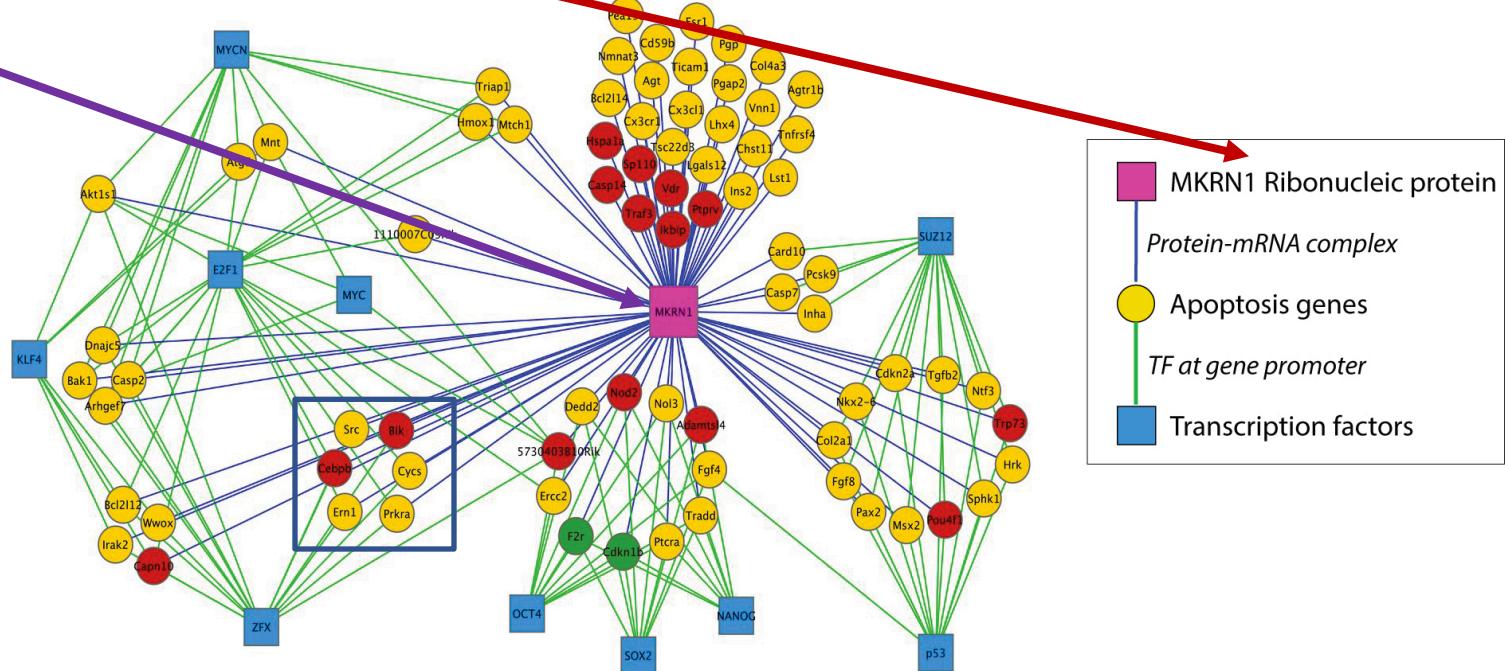
Networks as Tools

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

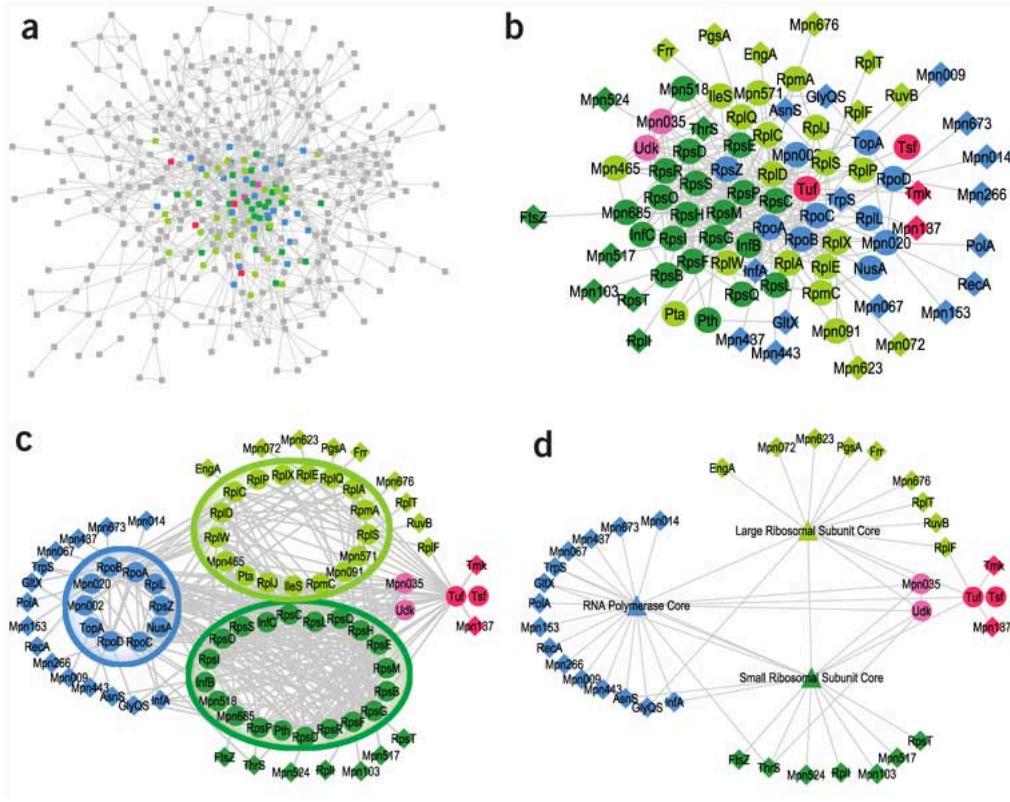


Networks as Tools

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

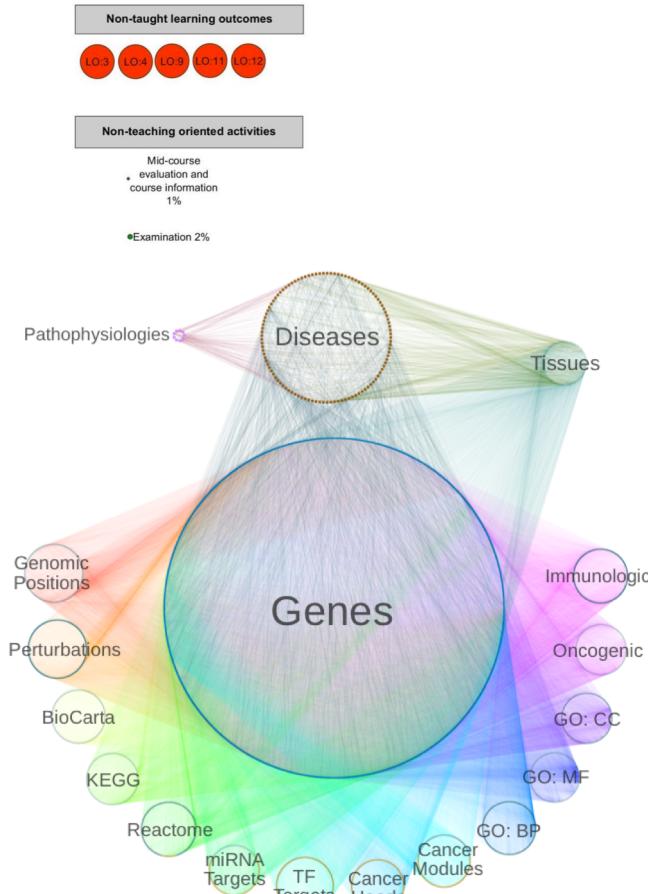
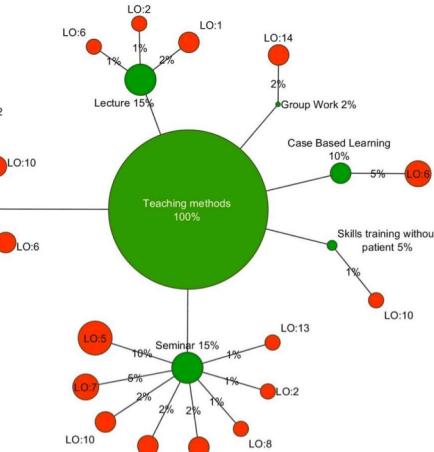
Network Visualization and Analysis using Cytoscape

- 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>

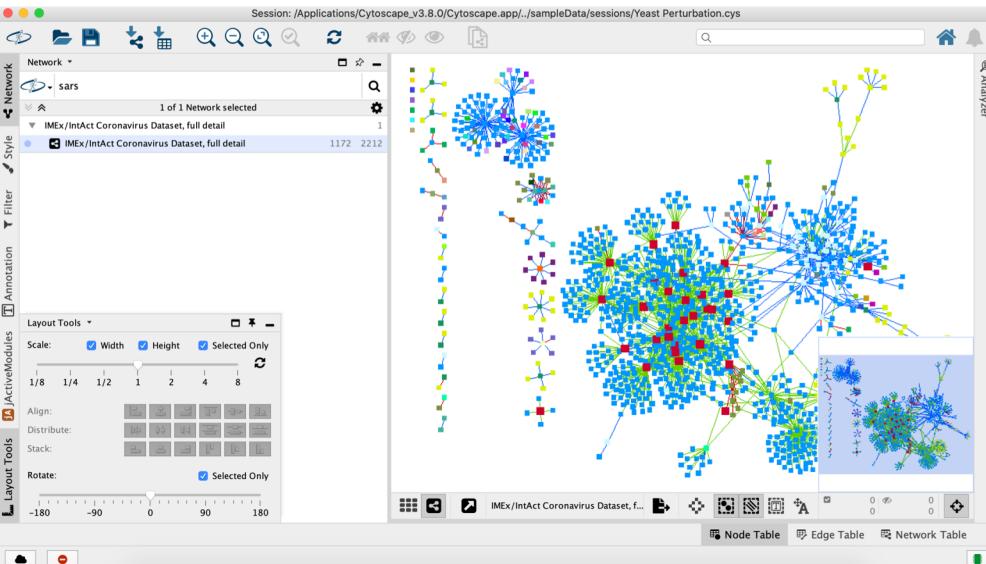
Cytoscape (cytoscape.org)



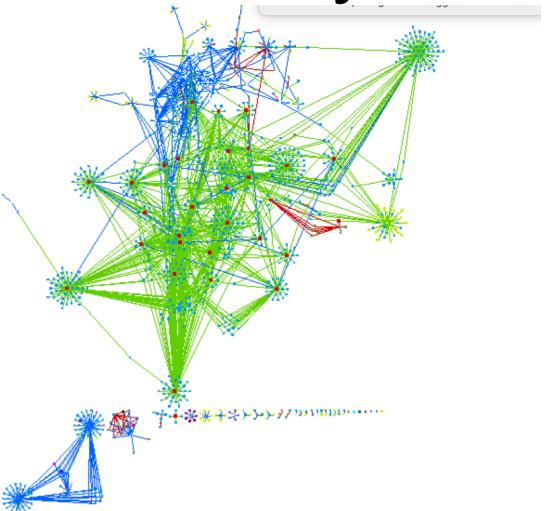
Partners and Collaborators



Manipulate Networks



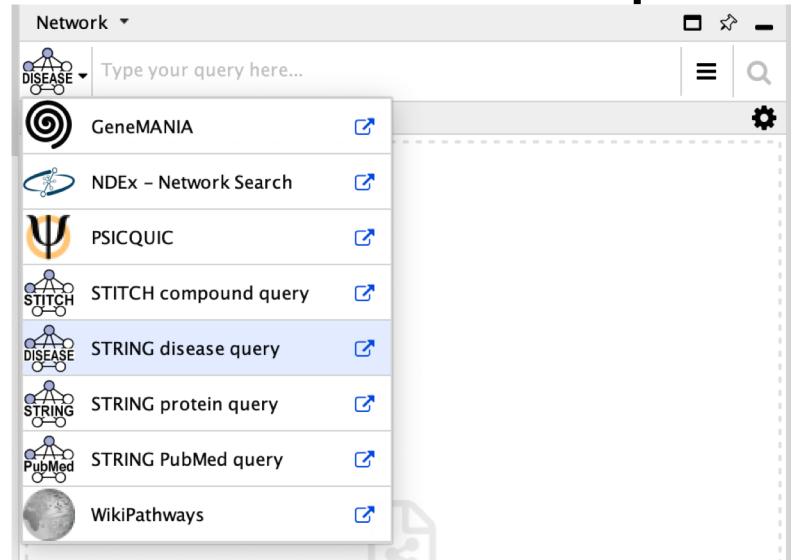
Automatic Layout



Filter/Query



Network Import



The Cytoscape App Store

Cytoscape App Store

Submit an App

Search the App Store

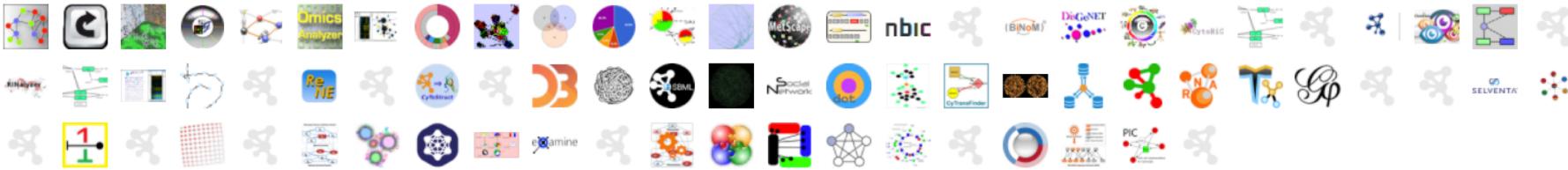
Sign In

Wall of Apps 361 total

collections



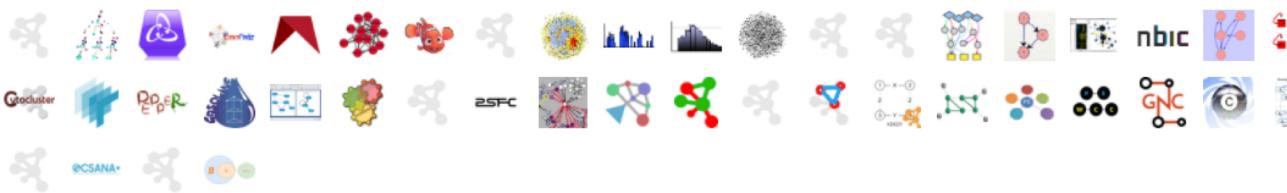
data visualization



network generation



graph analysis



network analysis



<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 2121 (google scholar), 1,437 (ISI web of Science)
- >361 Apps Extend Functionality (July 2020)
 - 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

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



bioinformatics.ca

Cytoscape Demo

Version 3.8.0

www.cytoscape.org

Interface Overview

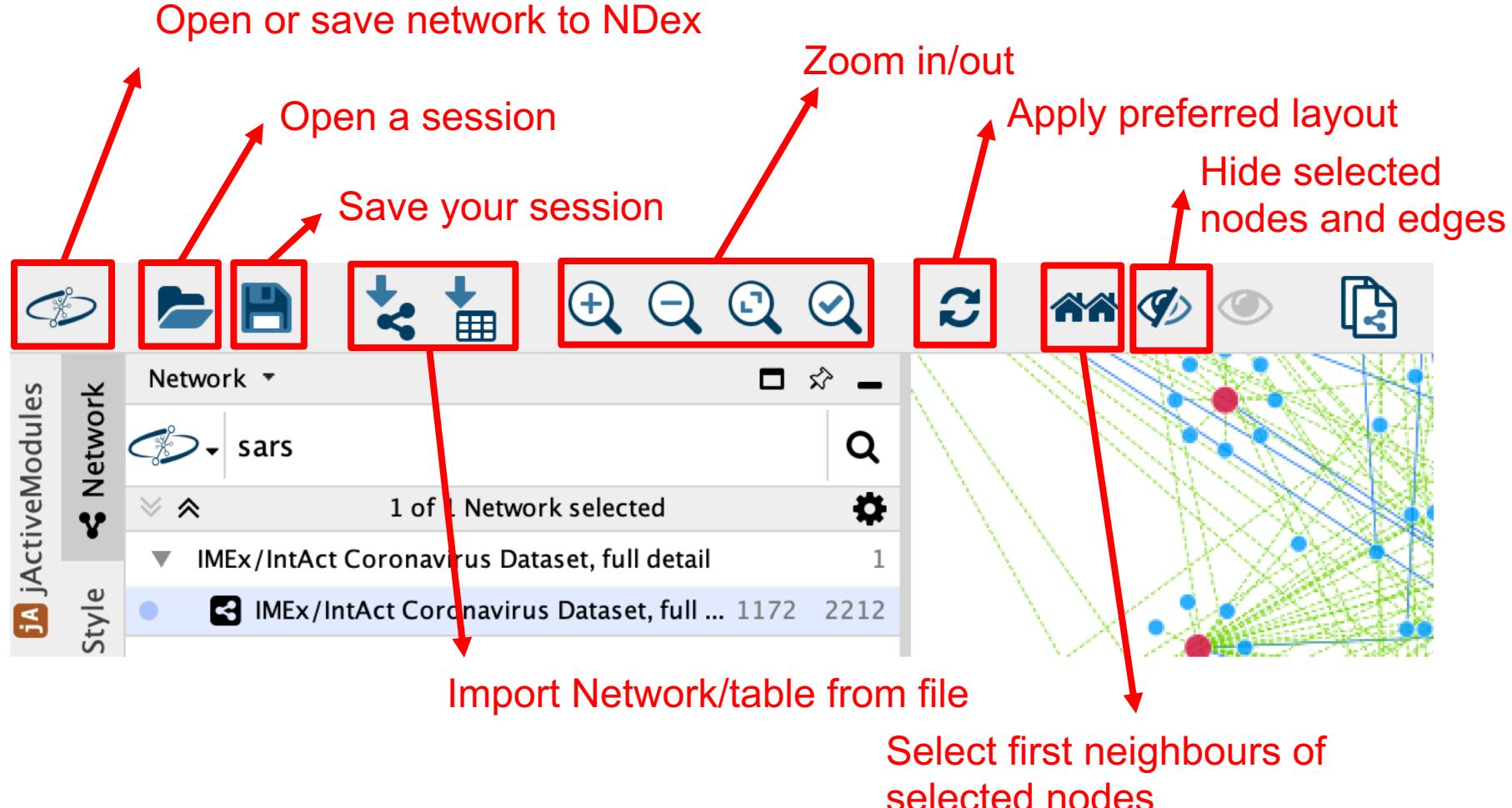
Results Panel

The screenshot displays the Cytoscape application interface. On the left, the **Control Panel** contains various icons for file operations like Open, Save, and Import. The **Network manager** panel shows a session titled "Yeast Perturbation.cys" with a network named "sars". It lists two datasets: "IMEx/IntAct Coronavirus Dataset, full detail" and "IMEx/IntAct Coronavirus Dataset, full ...". The **Canvas** is the central area where a complex network graph is visualized, featuring numerous nodes (blue and red dots) and a dense web of green and blue lines representing connections. A callout box labeled "Network overview" points to the top right corner of the canvas area. The bottom right section is the **Table Panel**, which displays a "Node Table" with the following data:

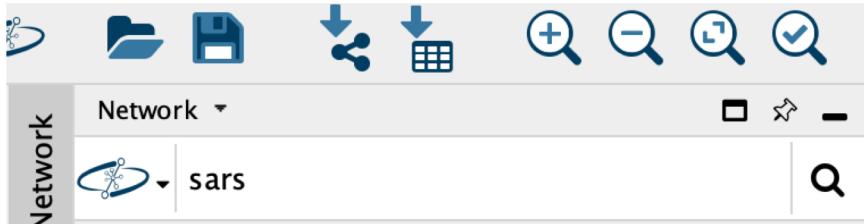
shared name	name	alias	taxid	species	m.type	m.type.mi
Q06093	Q06093	U47	10370	hhv6u	protein	MI:0326
P27228	P27228	E6	10587	Human papillo...	protein	MI:0326
P24835	P24835	E6	10588	Human papillo...	protein	MI:0326
P50804	P50804	E6	39457	Human papillo...	protein	MI:0326
Q8NHY3	Q8NHY3	GAS2L2	9606	human	protein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

Buttons at the bottom of the Table Panel allow switching between "Node Table", "Edge Table", and "Network Table". A large black arrow points from the "Results Panel" header towards the top right corner of the Cytoscape window.

Basic Navigation



Load a Network



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

Find Networks

NDEx

Sign in

sars

Click on download arrow to load network into Cytoscape

Results

	name	owner	visibility	nodes	edges	modified
	BioGRID: Protein-Protein Interactions (HIV-1)	biogrid	PUBLIC	1146	1338	2020-06-20, 12:40 p.m.
	BioGRID: Protein-Protein Interactions (X. laevis)	biogrid	PUBLIC	1153	1259	2020-06-20, 12:13 p.m.
	IMEx / IntAct Coronavirus Dataset, collapsed e...	pporras	PUBLIC	1172	1778	2020-06-15, 5:41 a.m.
	IMEx/IntAct Coronavirus Dataset, full detail	pporras	PUBLIC	1172	2212	2020-06-15, 5:40 a.m.
	IMEx/IntAct Coronavirus Dataset, binding regi...	pporras	PUBLIC	1172	2212	2020-06-15, 5:40 a.m.
	IMEx/IntAct Coronavirus Dataset, mutations de...	pporras	PUBLIC	1172	2212	2020-06-15, 5:40 a.m.
	WP3888 – VEGFA-VEGFR2 Signaling – Homo sa...	wikipathways	PUBLIC	1186	509	2019-12-16, 4:24 p.m.
	Bandyopadhyay et al., NATURE METHODS (201...	rudipillich	PUBLIC	1346	2272	2018-02-07, 3:37 p.m.
	SMPDB: Protein-Pathway Associations	ndexbutler	PUBLIC	1753	14187	2019-09-19, 9:19 p.m.
	ProteomeHD – Top 0.5% Co-Regulated Proteins...	proteomehd	PUBLIC	2718	63290	2019-09-20, 2:23 p.m.
	BioGRID: Protein-Protein Interactions (R. norve...	biogrid	PUBLIC	4148	7186	2020-06-20, 12:38 p.m.
	Drugbank Database – v4.3	drugbank	PUBLIC	4606	9652	2016-05-03, 3:40 p.m.

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

Close Dialog

About the network

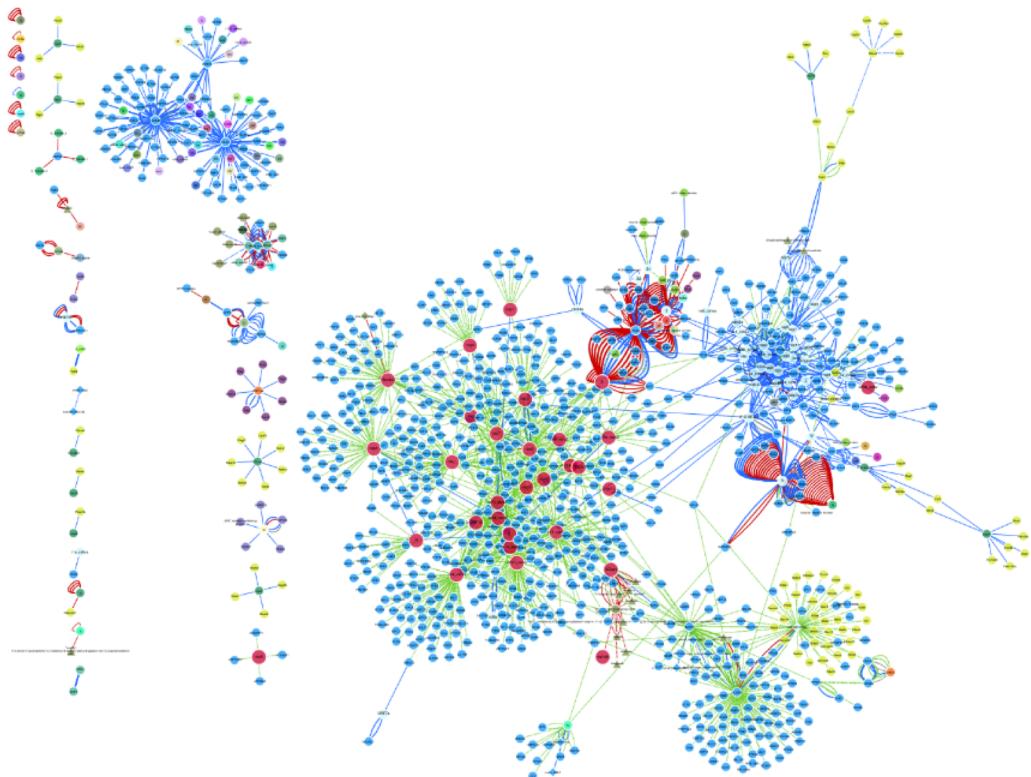
<http://ndexbio.org/#/network/3e13a4de-a4d7-11ea-aaef-0ac135e8bacf>



About Docs Report Bug Contact Us FAQ



Login/Register



Network Info Nodes/Edges

IMEx/IntAct Coronavirus Dataset, full detail

DOI: 10.18119/N9RC8F

Nodes: 1172 Edges: 2212

PUBLIC @context: [view namespaces](#)

Owner Pablo Porras Millan

Created Jun 2, 2020 9:44:55 AM

Last Modified Jun 15, 2020 5:40:49 AM

UUID 3e13a4de-a4d7-11ea-aaef-0ac135e8bacf

Description:

Full Coronavirus dataset as extracted from IntAct on May 2020. Each edge represents a separate piece of evidence for an interaction, with edge colour representing different interaction types. Dashed lines represent spoke-expanded interactions. Node colour represents different species origin and node shape different molecule types. SARS-CoV-2 and SARS 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)

Version: 1.0

Properties:

author IMEx Consortium Curators & Pablo Porras

disease COVID19, SARS

organism SARS-CoV-2, SARS, Coronaviridae, human, mouse, mammal

Visual Styles

Edge Stroke Color (Unselected) Mapping

Edge Stroke Color (Unselected)	int.type
green	association
red	cleavage reaction
grey	colocalization
dark red	covalent binding
dark red	direct interaction
dark red	hydroxylation reaction
blue	physical association
dark red	protein cleavage
dark red	rna cleavage

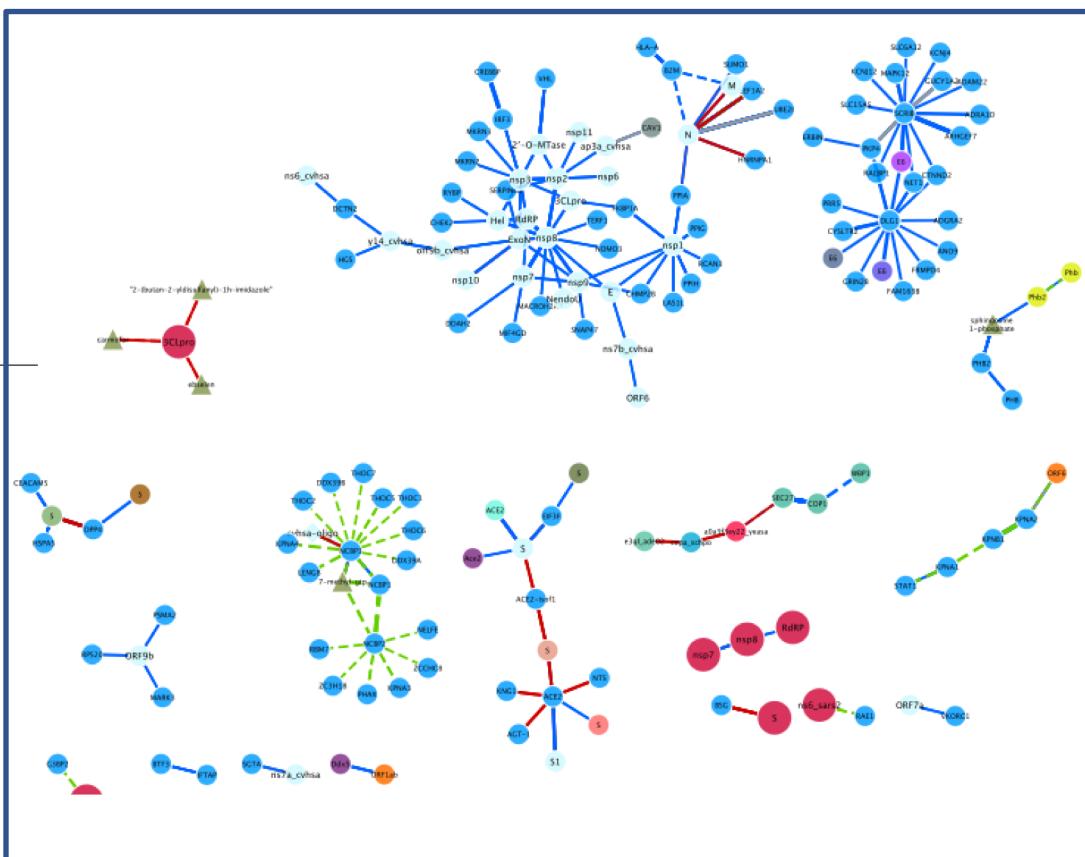
Node Properties

Node Shape Mapping

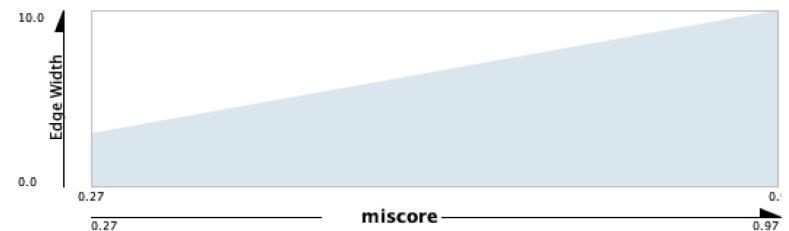
Node Shape	m.type
diamond	messenger rna
diamond	micro rna
diamond	nucleic acid
circle	peptide
circle	protein
diamond	ribonucleic acid
triangle	small molecule
diamond	small nuclear rna

Node Fill Color Mapping

Node Fill Color	species
red	SARS-CoV-2
blue	human
yellow	mouse
purple	rat

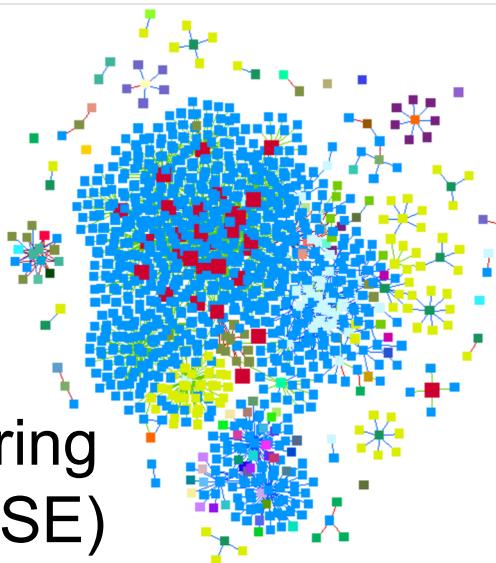


Edge Width Mapping

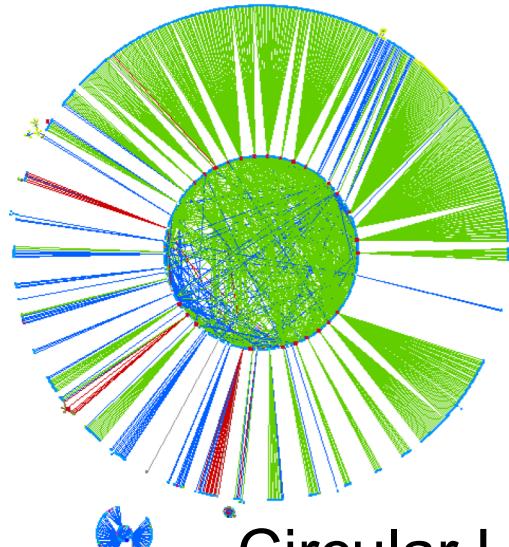
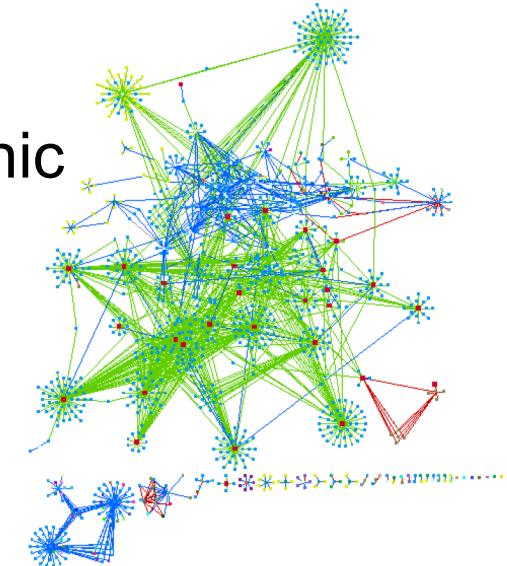


Experiment with different layouts

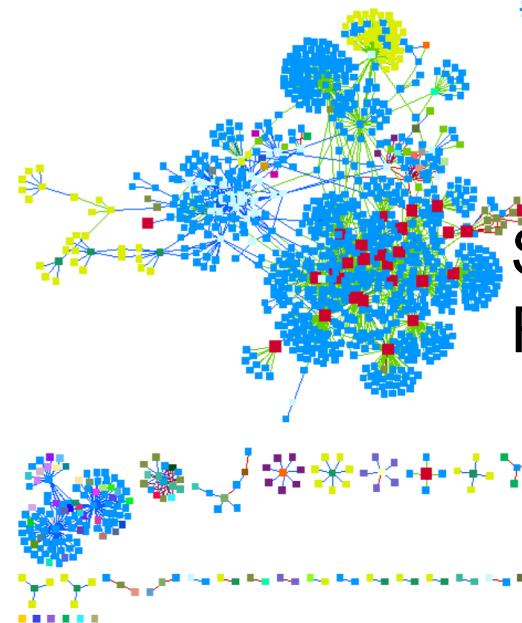
Compound Spring
Embedder (CoSE)



yfiles organic

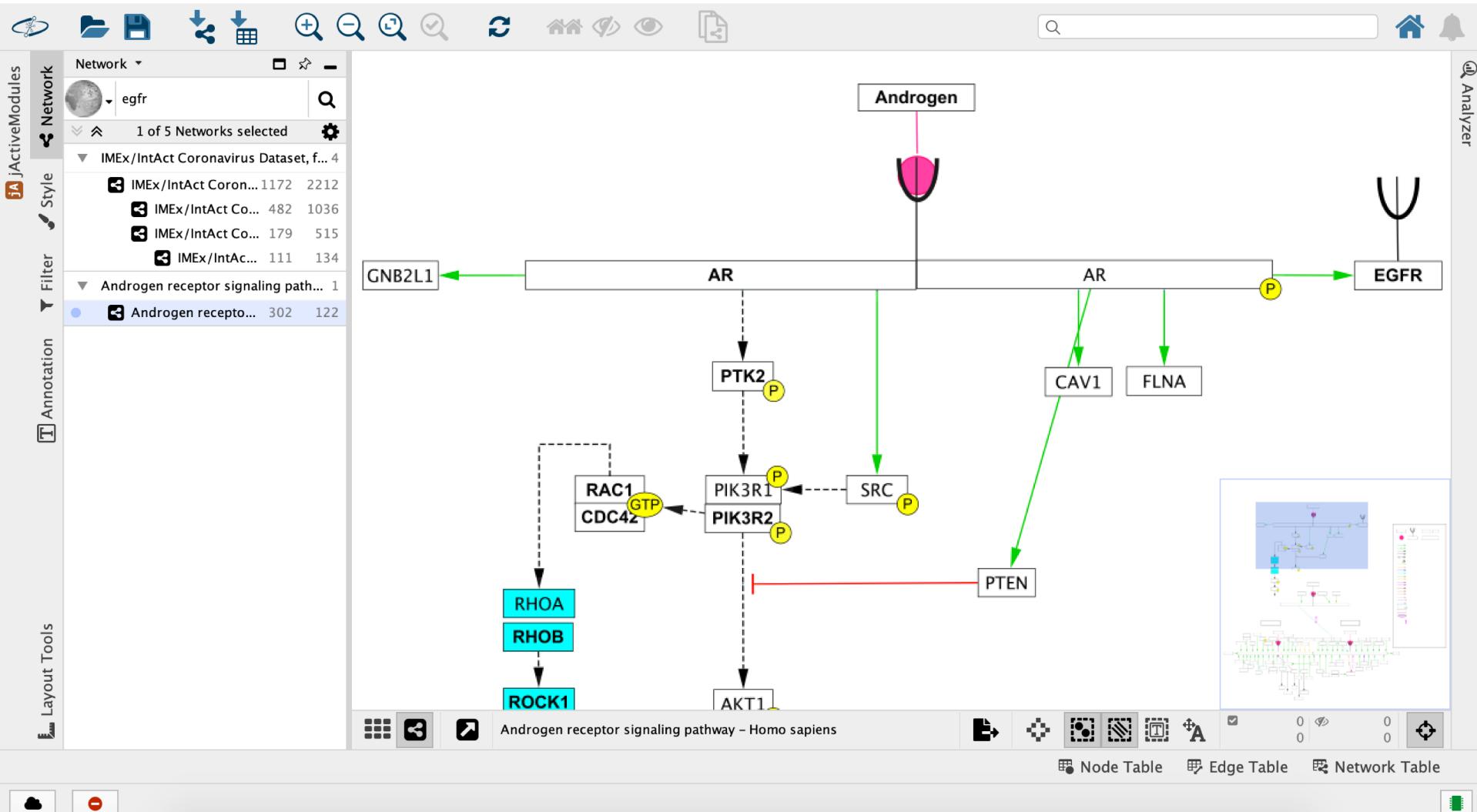


Circular Layout



Spring embedded
Force directed

Load different types of networks



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

We are on a Coffee Break & Networking Session

compute | calcul
canada | canada



Workshop Sponsors:

Canadian Centre for
Computational
Genomics

MicM McGill initiative in
Computational Medicine