Introduction to Network Visualizations and Analysis in Cytoscape

Igy Pang and Tara K. Bartolec

SYSTEMS BIOLOGY INITIATIVE

UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Overview

1. **Introduction**

2. **What you need to know about networks?**

3. **Cytoscape Workshop**

Sydney rail network

M Metro **T** Trains

NSW GOVERNMENT | Transport

T1 North Shore
Berowra
Mount Kuring-gai
Mount Colah
Asquith
To Central Coast & Newcastle Line
Northern **T9** **Hornsby**
Normanhurst
Waitara
Thornleigh
Wahroonga
Pennant Hills
Warrawee
Beecroft
Turramurra
Cheltenham
Pymble
Gordon
Killara
Gordon **T9**
Lindfield
Roseville
Chatswood
Artarmon
St Leonards
Wollstonecraft
Waverton

Richmond **T5** **T1** Richmond
Richmond
East Richmond
Clarendon
Windsor
Mulgrave
Vineyard
Riverstone
Schofields
Quakers Hill
Marayong

Tallawong
Rouse Hill Kellyville Bella Vista Norwest Hills Showground Castle Hill Cherrybrook
M Tallawong

Epping
Eastwood
Macquarie University Macquarie Park North Ryde
Denistone
Chatswood **M**
West Ryde
Crows Nest
Meadowbank
Victoria Cross
North Sydney
Milsons Point
Circular Quay
Barangaroo
City
Kings Cross Edgecliff
Bondi Junction
Rhodes
Wynyard
Martin Place
Concord West
Pitt St
Town Hall
St James
North Strathfield
Central
Museum
T4 Eastern Suburbs
Redfern
Sydney Metro City and Southwest under construction

To Blue Mountains Line
T1 Western
Emu Plains
Penrith
Kingswood
St Marys
Werrington Mount Druitt Rooty Hill Doonside
Blacktown
Seven Hills
Toongabbie Pendle Hill Wentworthville Westmead
Inner West **T2**
Parramatta
Harris Park
Granville Clyde Auburn
Lidcombe **T3**
Olympic Park **T7**
Olympic Park
Merrylands
Guildford
Yennora
Fairfield
Canley Vale
Cabramatta
Warwick Farm
Liverpool
T3 Liverpool
Casula
Leppington
Edmondson Park
Leppington **T5**
Leppington **T2**
Glenfield
Macquarie Fields
Ingleburn
Minto
Leumeah
Campbelltown
Macarthur
South **T8**
To Southern Highlands Line

Flemington Homebush Strathfield Croydon Ashfield Summer Hill Lewisham Petersham Stanmore Newtown Macdonaldtown Erskineville St Peters **Sydenham**
Berala
Regents Park
Carramar Villawood Leightonfield Chester Hill Sefton
Birrong
Bankstown
Punchbowl Wiley Park Lakemba Yagoona
Hurlstone Park Canterbury Campsie Belmore
Dulwich Hill Marrickville
Waterloo
Green Square
Mascot
Domestic Airport Station Access Fee applies **T8** Airport
International Airport Station Access Fee applies
Tempe
Wolli Creek
Arncliffe
Banksia
Rockdale
Kogarah
Carlton
Allawah
Hurstville
Penshurst
Mortdale
Oatley
Como
Jannali
Sutherland
Kirrawee Gymea Miranda Caringbah Woolooware
Cronulla
T4 Cronulla
Loftus
Engadine
Heathcote
Illawarra **T4** **Waterfall**
To South Coast Line

Turrella
Bardwell Park
Bexley North
Kingsgrove
Beverly Hills
Narwee
Holsworthy **East Hills** Panania **Revesby** Padstow Riverwood

Sydney metro and train lines

M *Metro North West Line* Chatswood, Tallawong	**T1** *North Shore & Western Line* North Shore, Western, Richmond	**T2** *Inner West & Leppington Line* Inner West, Leppington, City	**T3** *Bankstown Line* Liverpool, Lidcombe, City	**T4** *Eastern Suburbs & Illawarra Line* Eastern Suburbs, Illawarra, Cronulla
T5 *Cumberland Line* Leppington, Richmond	**T7** *Olympic Park Line* Olympic Park, Lidcombe	**T8** *Airport & South Line* Airport, South, City	**T9** *Northern Line* Northern, Gordon	

Stop Interchange Interchange End of line

NORTH

Line under construction

Check timetables and trip planners for train services and connections

Visit transportnsw.info
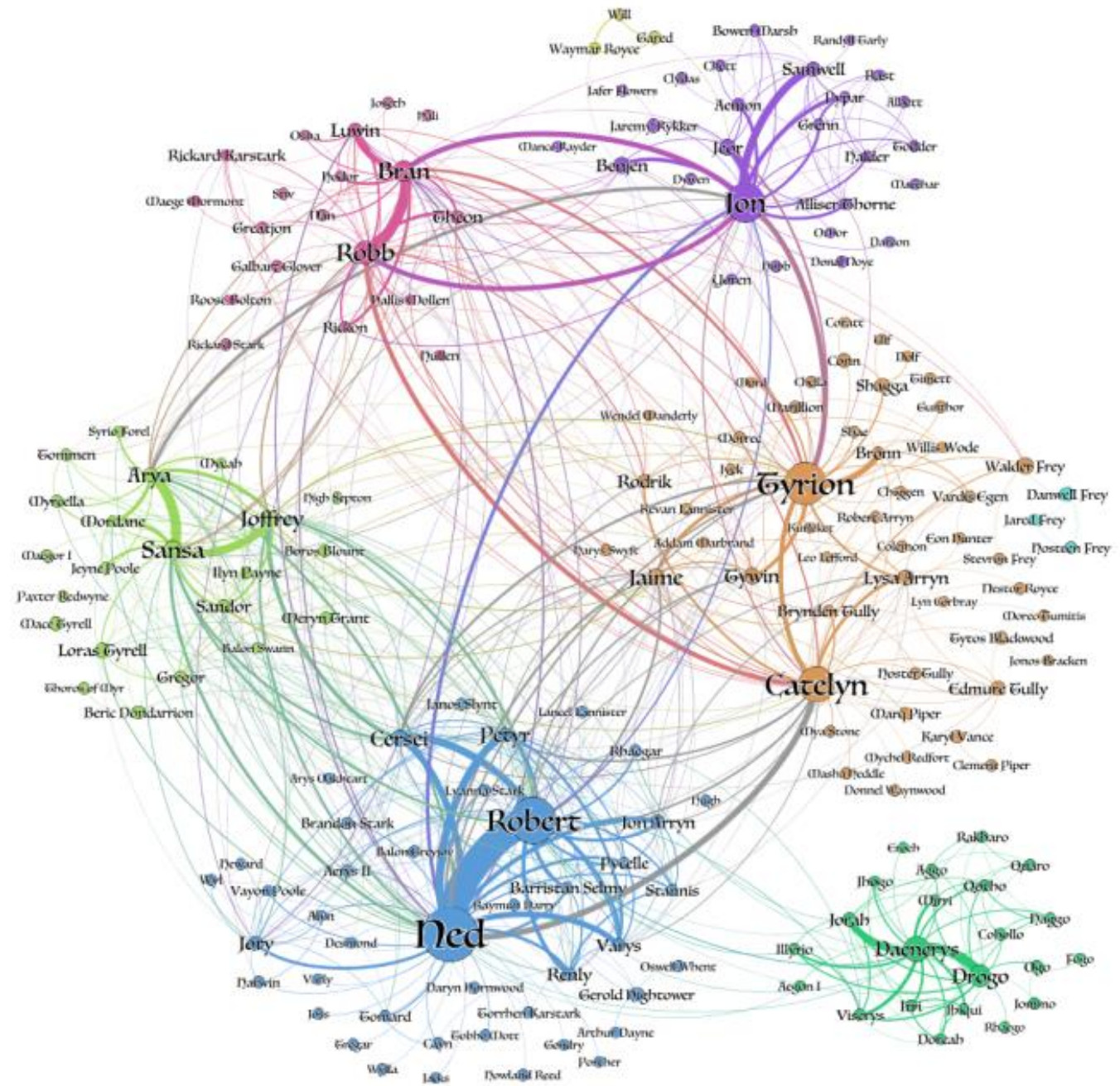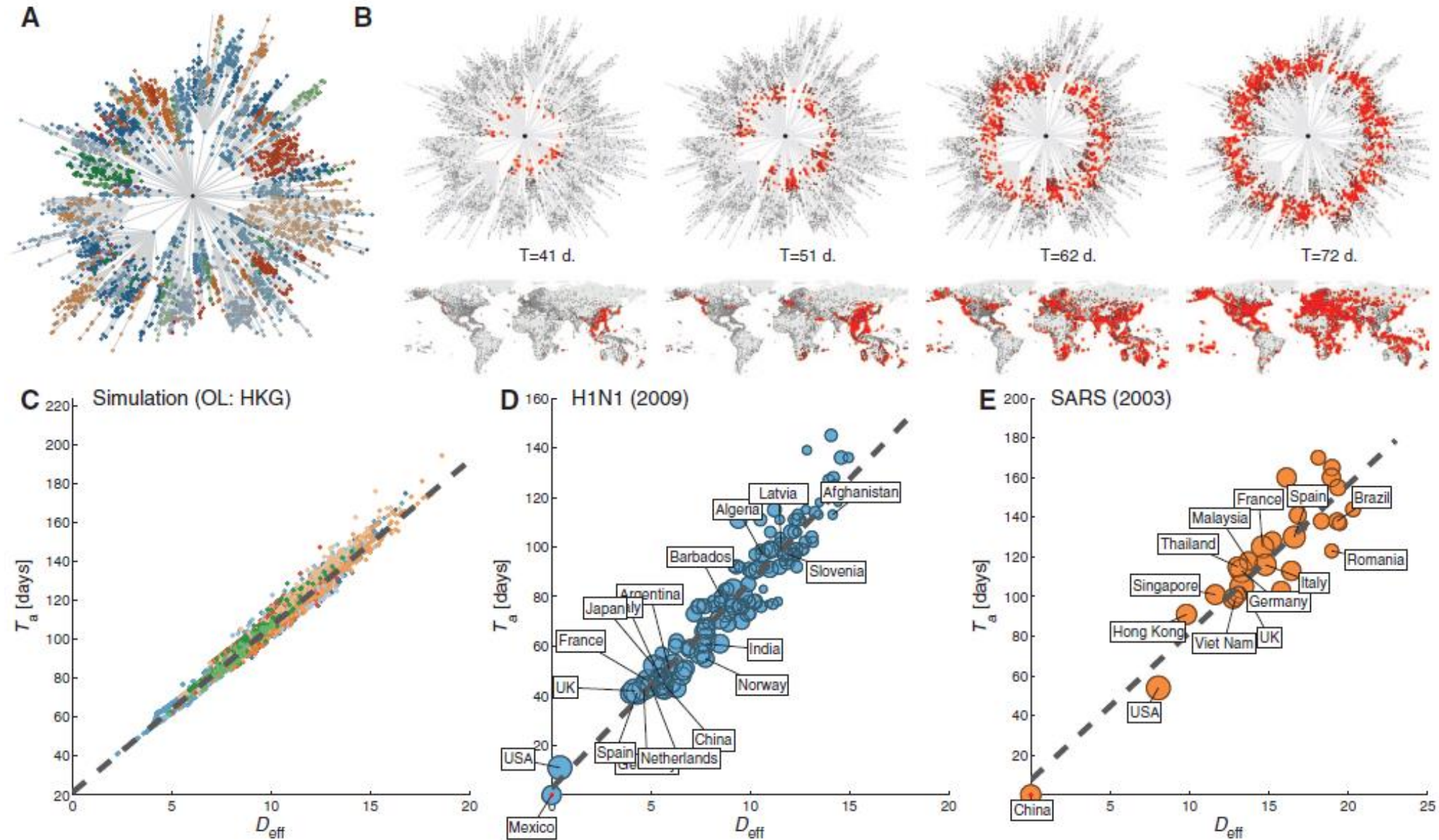
Copyright © 2019 Transport for NSW

1909TMS-P-COR-A4P

GoT

Cryptocurrency Transaction Networks

- Evidence trail
- Fraud detection
- Lots of data inherent in the network

Model of Virus Spread

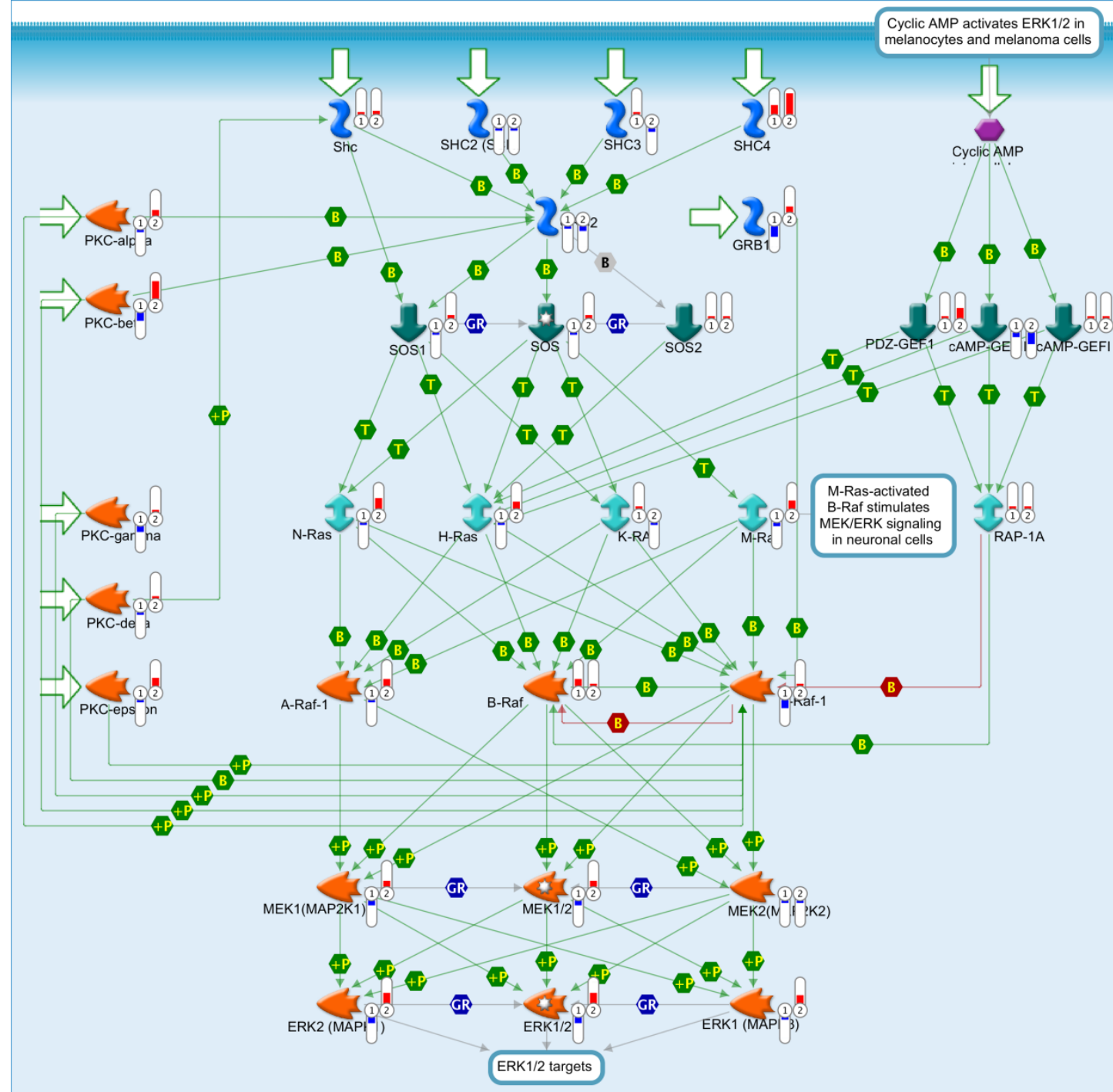A

B

T=41 d.   T=51 d.   T=62 d.   T=72 d.

C Simulation (OL: HKG)

D H1N1 (2009)

E SARS (2003)

Brockmann and Helbing 2013 Science 342:1337-1342
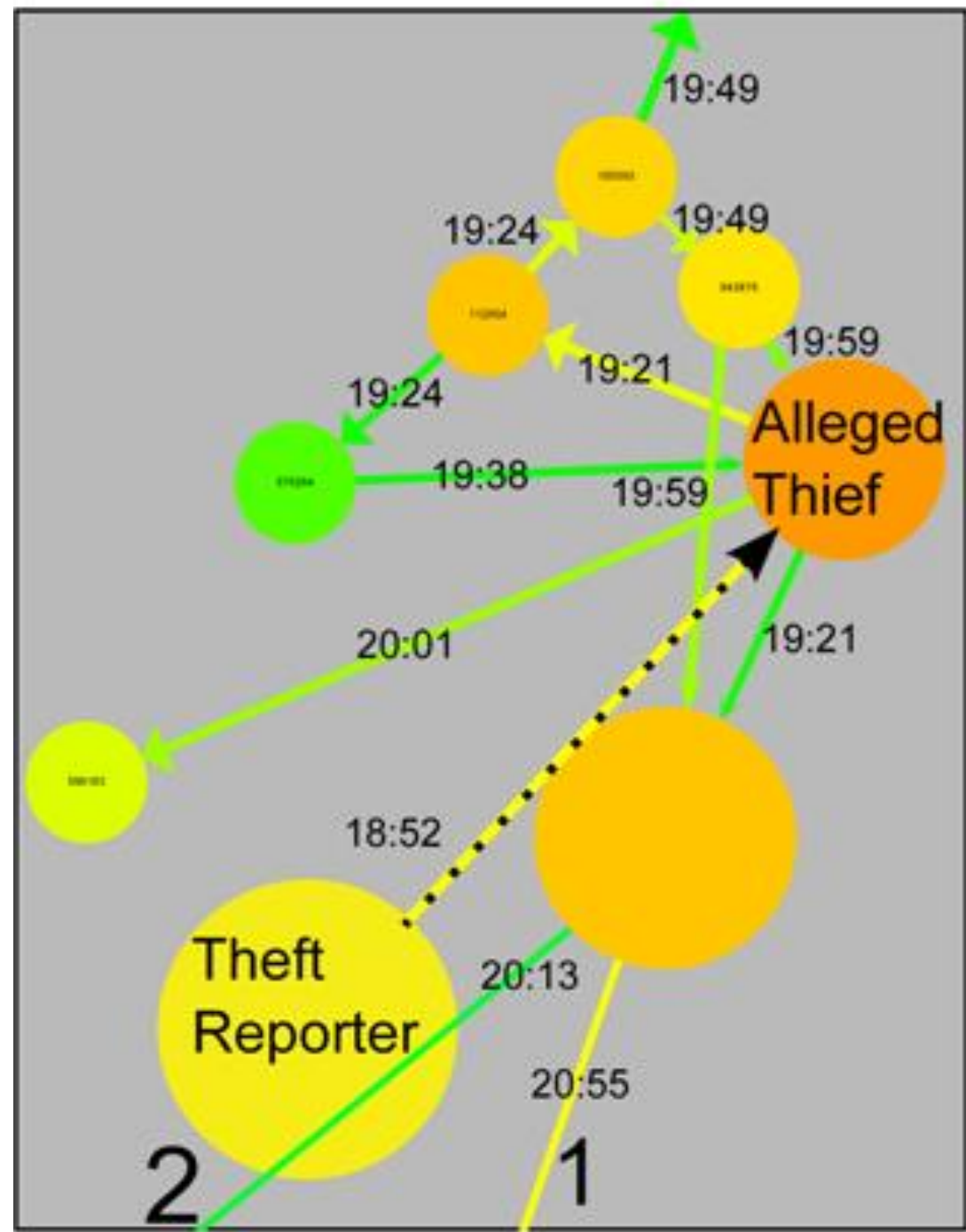
Canonical Signaling Pathways

- Idealized pathways
- Only well-known paths
- Cross-talks?

Metacore Pathway Map

Network Terminologies

- **Node** – an entity in the graph
  - person
  - gene
  - protein
  - metabolite
  - Also called 'vertex'
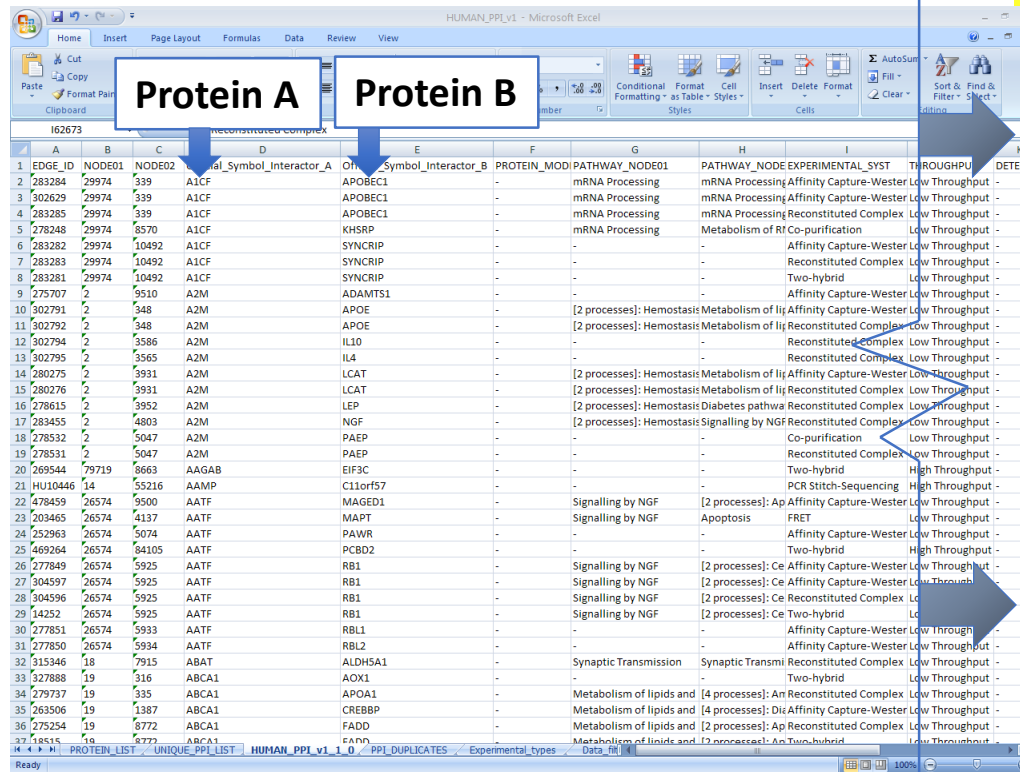
- **Edge** – represents relationships
  - Bob and Alice are friends
  - also called an 'interaction'
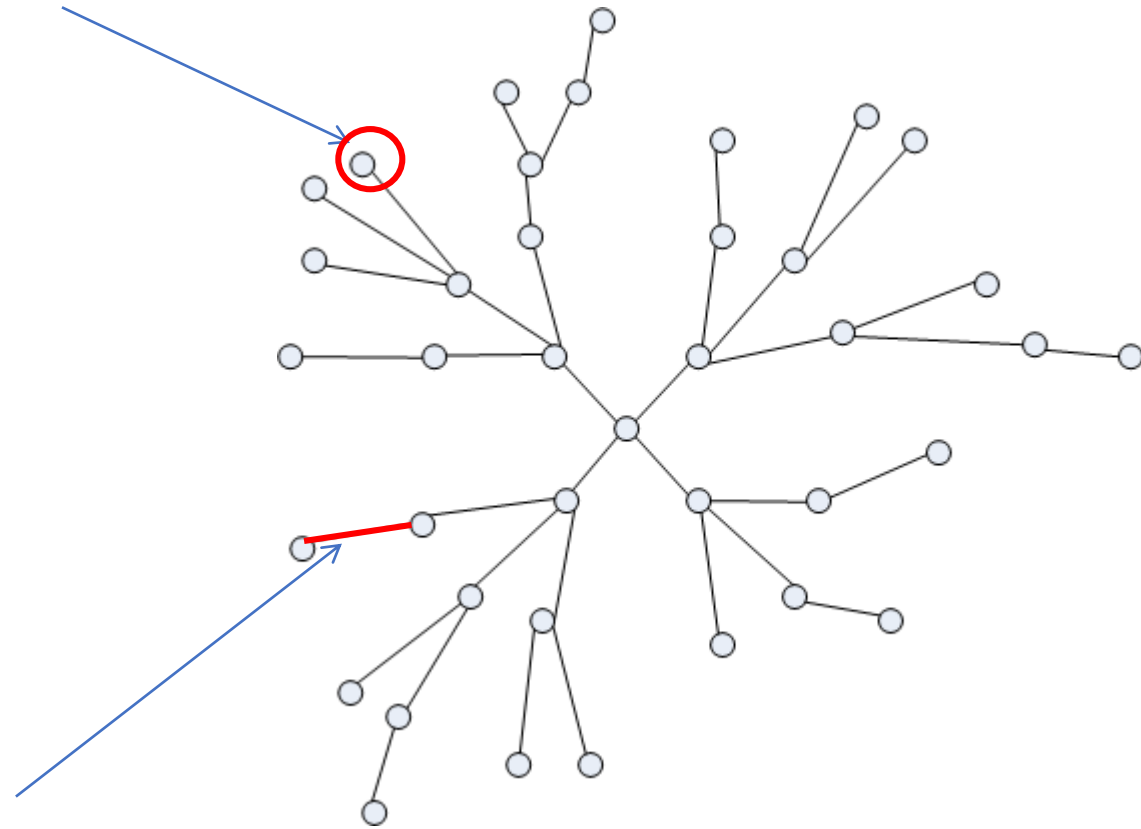
Data to Network

Network Visualization Tool

List of Edge
- **protein A interacts with protein B**

Network

Different Representations of Networks Data
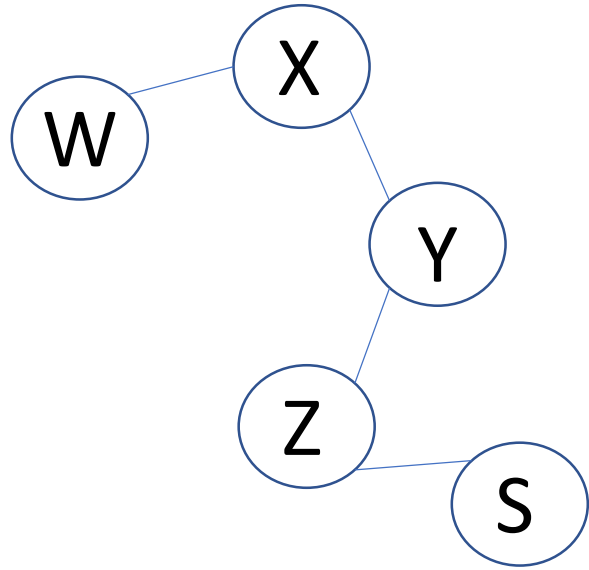
Data file format

Edge List

Protein A	Protein B
X	Y
Y	Z
W	X
S	Z

Matrix

Protein B

Protein A

	S	W	X	Y	Z
S					1
W			1		
X					1
Y					1
Z					

Edge weights

Directed versus Undirected Edges

Edge Types

Undirected Network

Directed Network
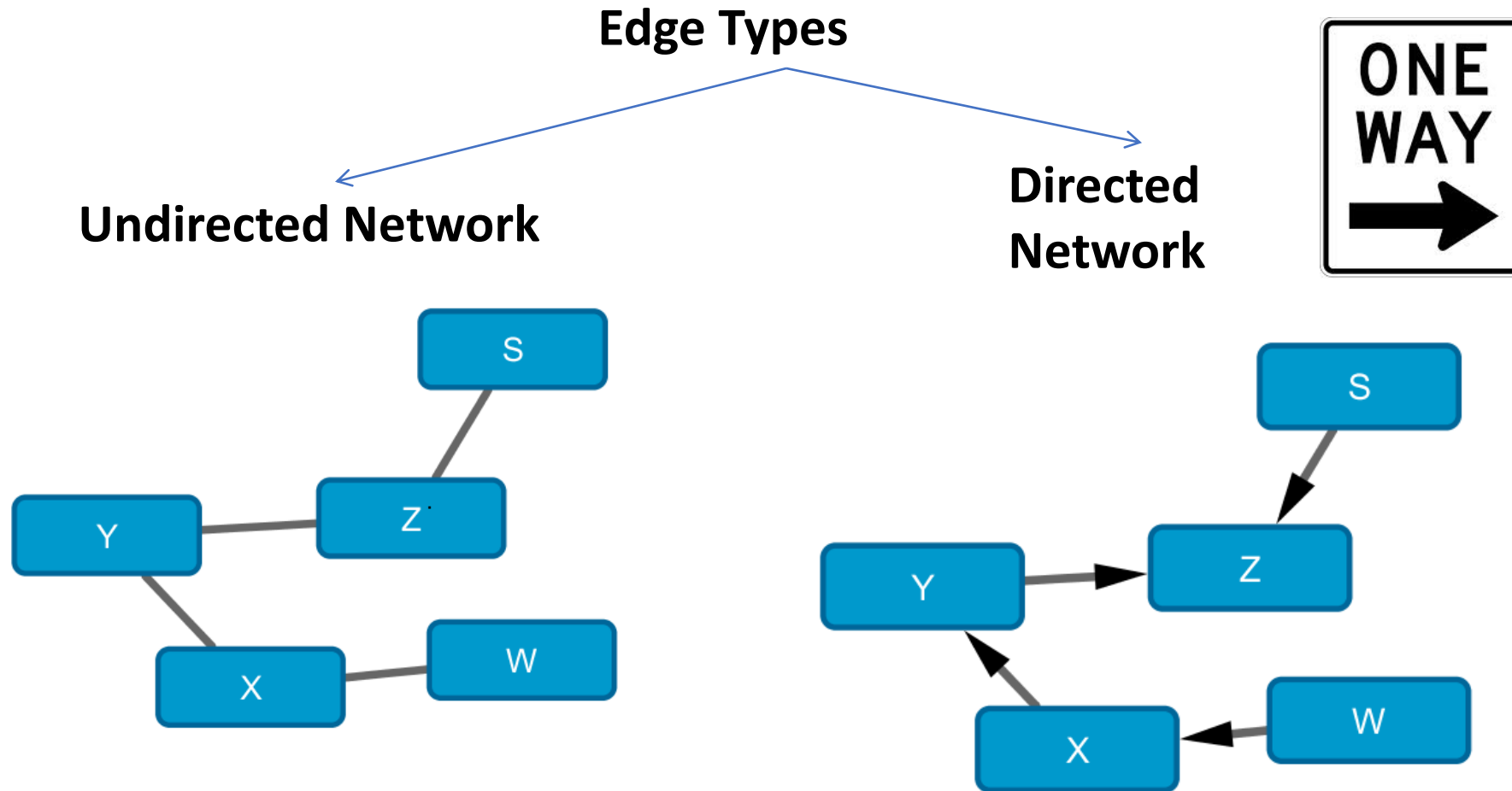
Functional Modules: Cellular Process
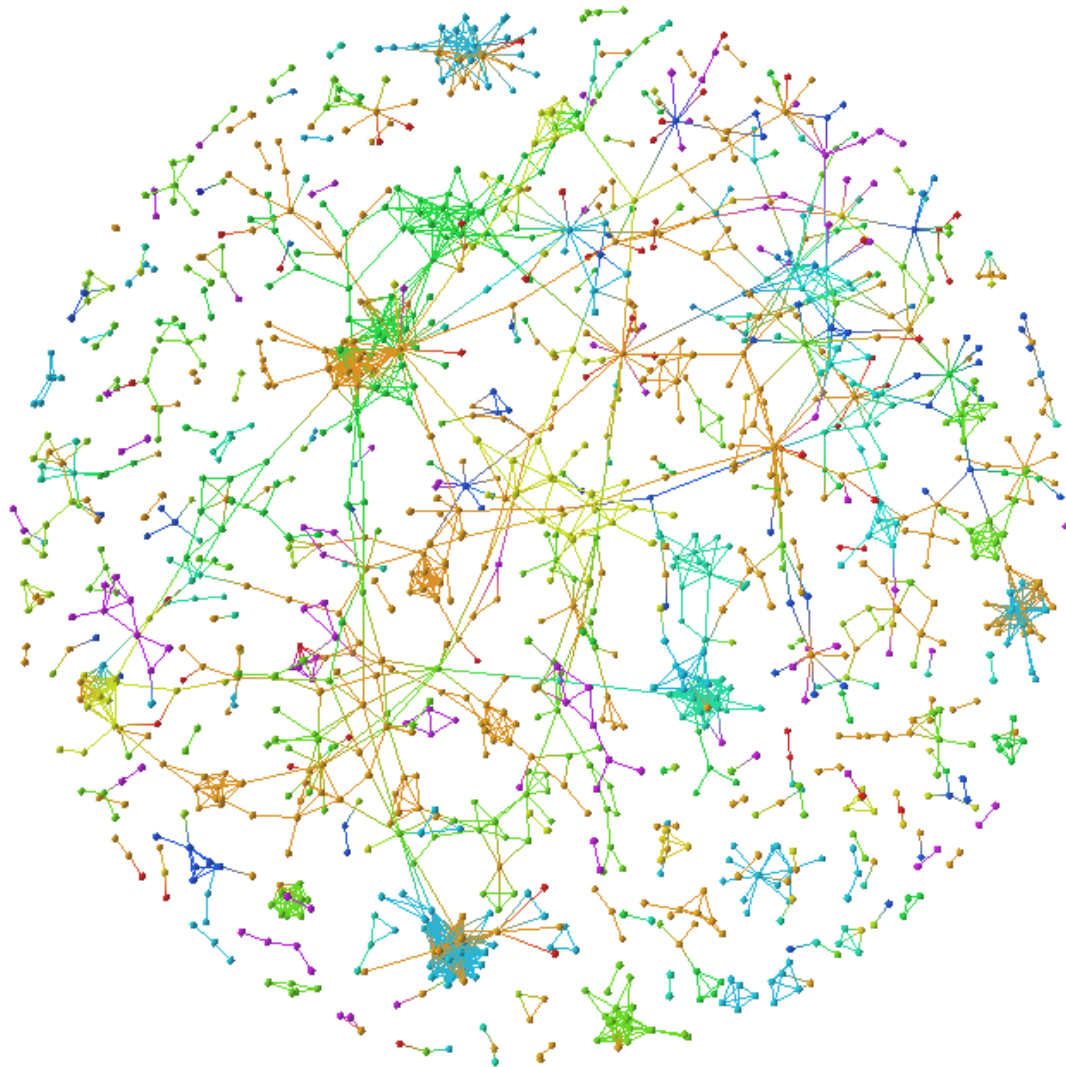
Interacting proteins are involved in the same processes

Colour:
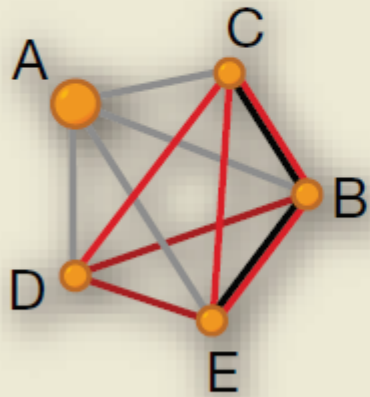Gene Ontology process

Network Measures – Node degree

Degree/
connectivity (k)

k_A = Nb of edges through A = 5
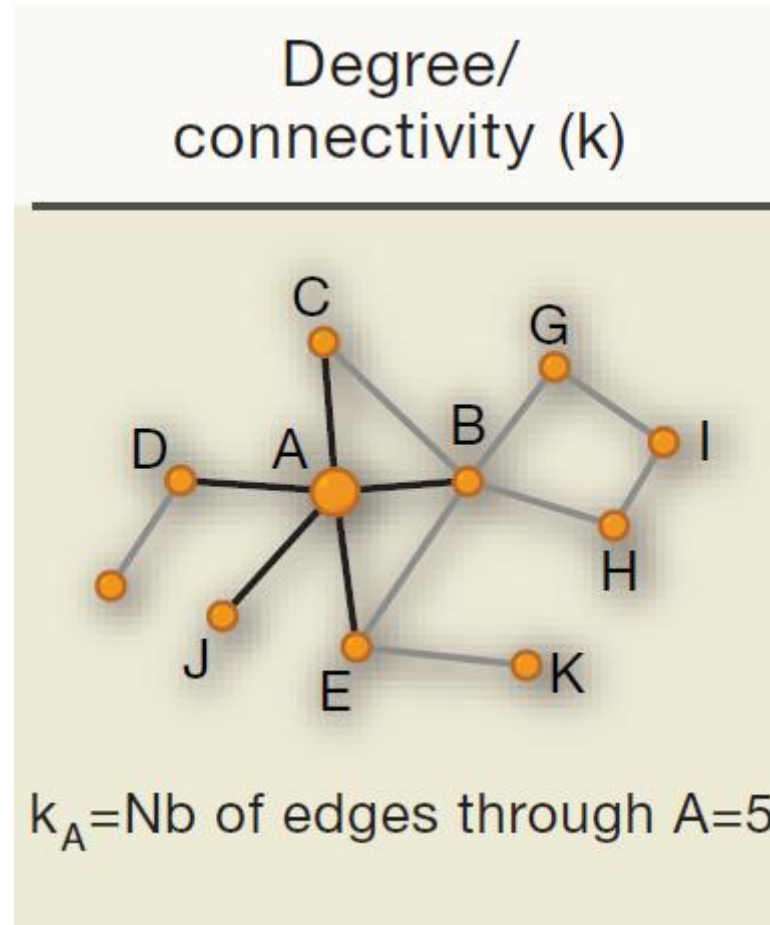
Local Network Structure: Hub Proteins
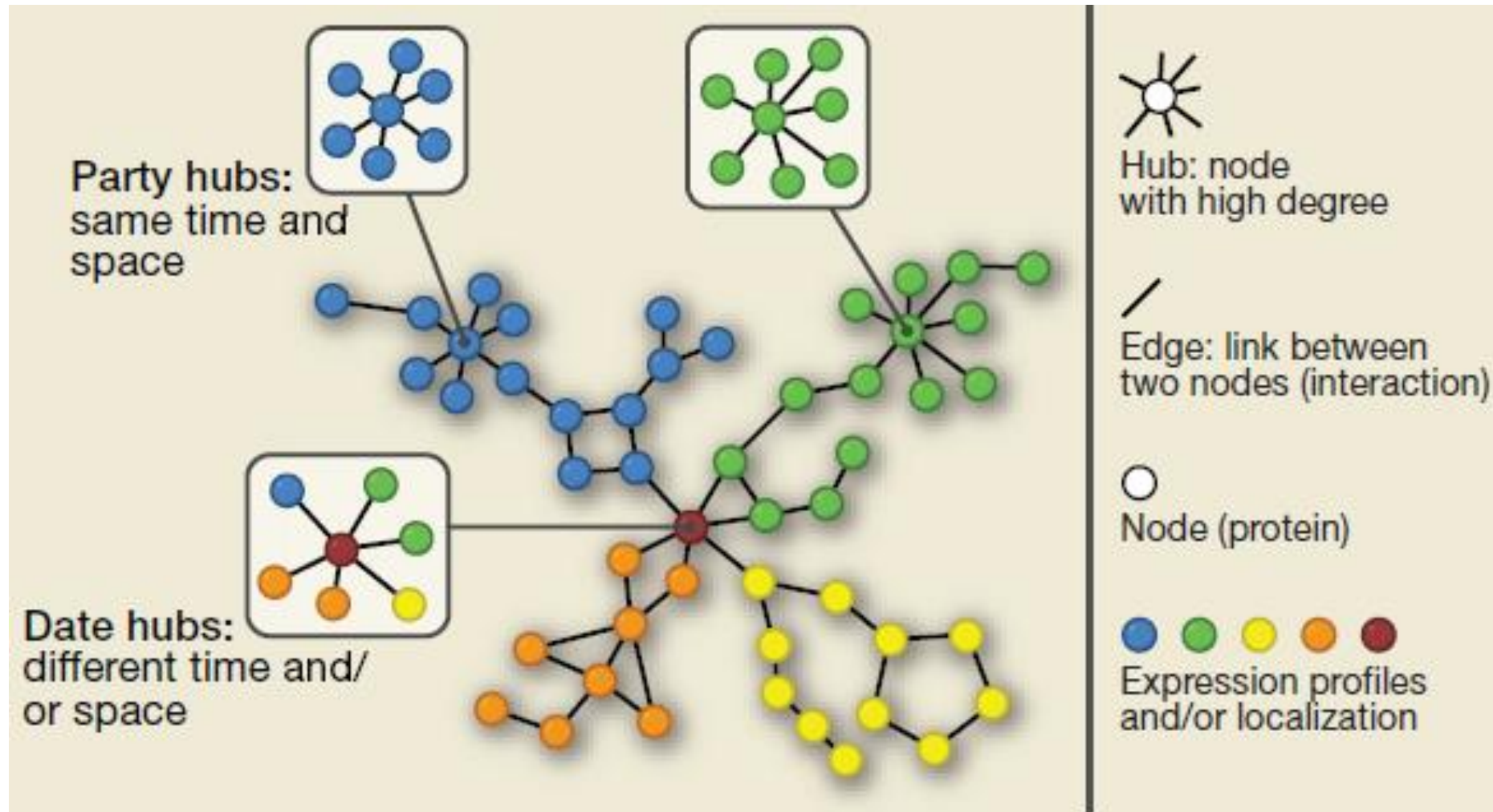
Network Measures – Clustering coefficient

Clustering coefficient/
interconnectivity (C)

$$C_A = \frac{\text{Actual links between A's neighbors (black)}}{\text{Possible links between A's neighbors (orange)}}$$

$$C_A = n_A/[k_A(k_A-1)/2]$$
$$= 2/[4 \times (4-1)/2] = 0.333$$

- Good for finding clusters / cliques
- Neighbors are highly connected
- Detect protein complexes
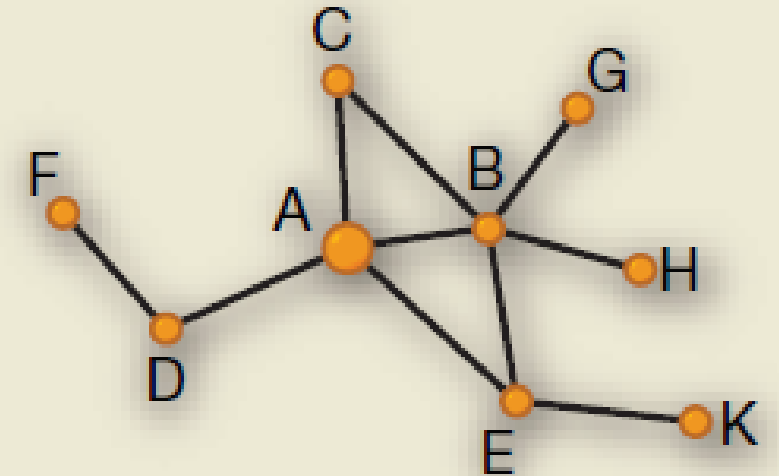
Network Measures – Shortest Path and Betweeness

- Six degrees of separation

Shortest path (SP) between two nodes
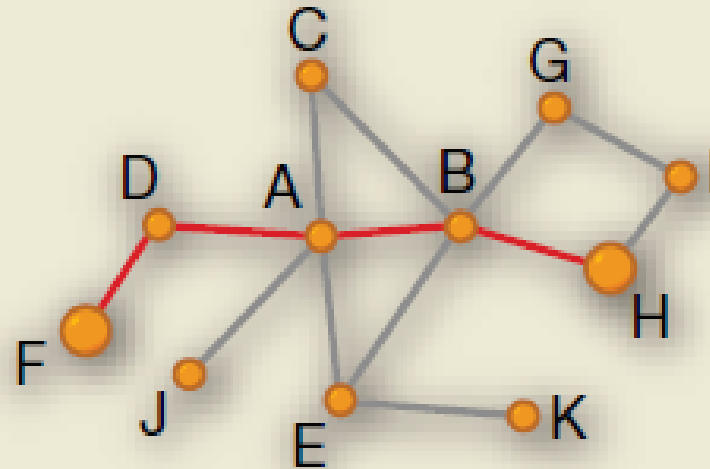
$SP_{FH} = (F,D,A,B,H) = 4$
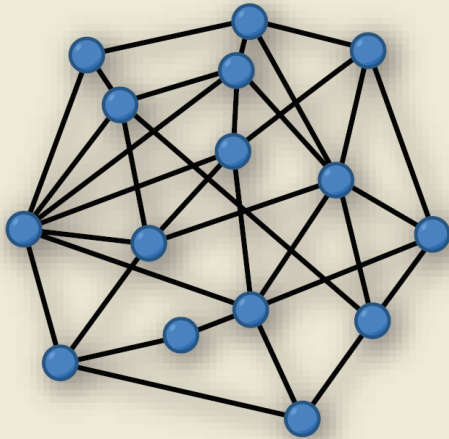
Betweenness/ centrality (B)

B_A = Fraction of SPs passing through A
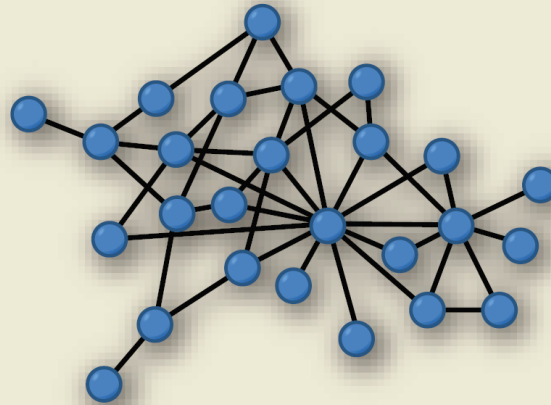
$= 0.090$
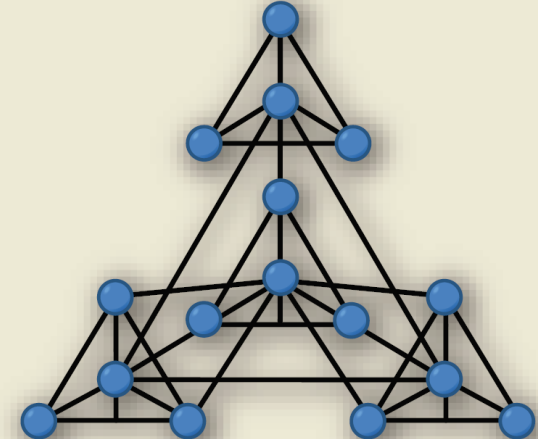
Network Topologies

Random network

-Degrees follow Poisson (or peaked) distribution

-Vulnerability to failure

Scale-free network
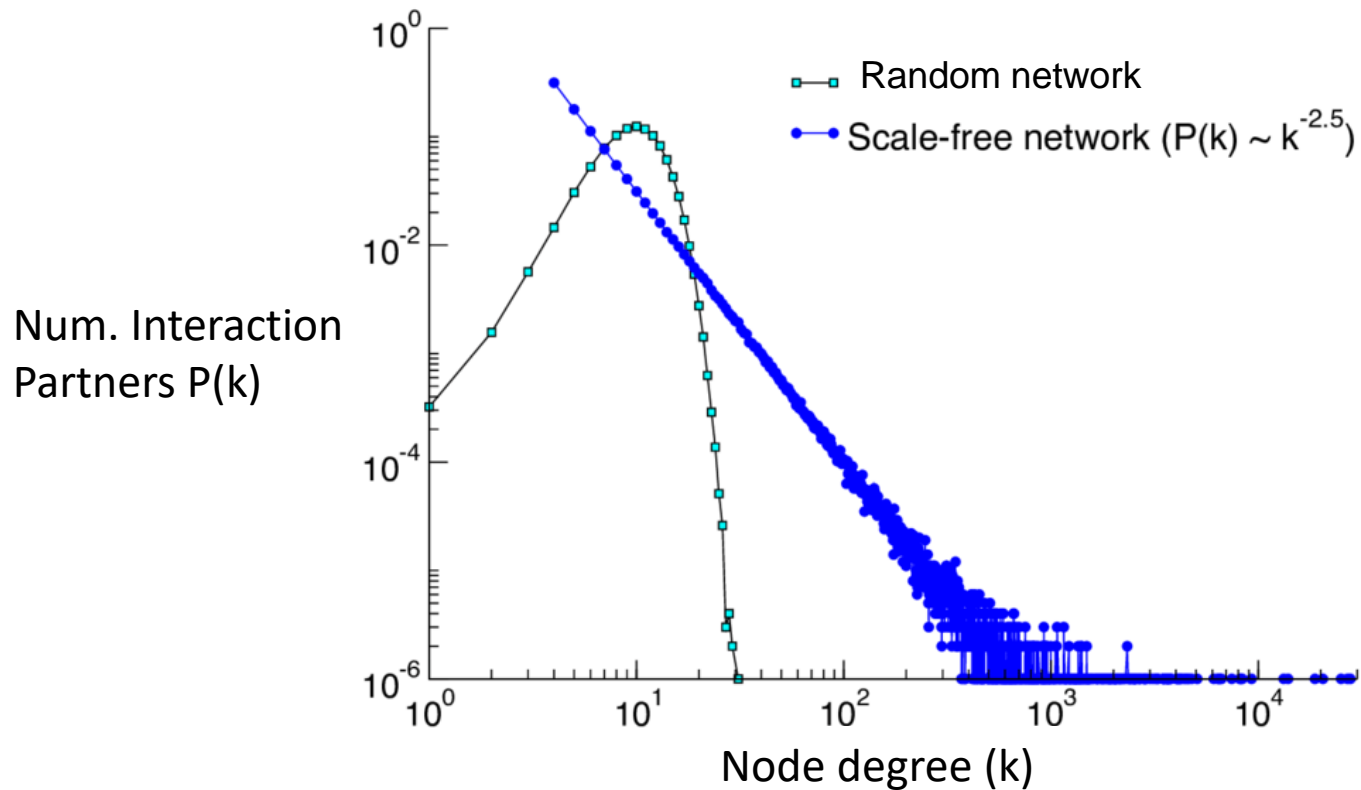(Biological/cellular networks)

-Degrees follow power-law distributions

-Robustness against random failure

-Vulnerability to targeted attacks

Hierarchical network
(Many types of real networks)

-Degrees follow power-law distributions

-Account for modularity, local clustering, and scale-free topology

-High clustering coefficient (C)

(Seebacher and Gavin, 2011, Cell, 144:1000.e1)

Degree Distributions of Random Network and Scale-free Network

- Log-log scale
- Power-law
- Scale-free network slope -2.5

$$\frac{\log P(k)}{\log k} \sim -2.5$$

Baronchelli et al. 2013 Networks in Cognitive Science. Trends Cogn. Sci 17(7):348-360

Cytoscape

- Where to download?

https://cytoscape.org/

- App store

https://apps.cytoscape.org/

Cross-linking Mass Spectrometry Analysis of the Yeast Nucleus Reveals Extensive Protein−Protein Interactions Not Detected by Systematic Two-Hybrid or Affinity Purification-Mass Spectrometry

Tara K. Bartolec,[†] Daniela-Lee Smith,[†] Chi Nam Ignatius Pang,[†] You Dan Xu,[‡] Joshua J. Hamey,[†] and Marc R. Wilkins[*,†]

[†]Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia

[‡]Centre for Advanced Macromolecular Design, School of Chemistry, University of New South Wales, Sydney, New South Wales 2052, Australia

S *Supporting Information*

ABSTRACT: *Saccharomyces cerevisiae* has the most comprehensively characterized protein−protein interaction network, or interactome, of any eukaryote. This has predominantly been generated through multiple, systematic studies of protein−protein interactions by two-hybrid techniques and of affinity-purified protein complexes. A pressing question is to understand how large-scale cross-linking mass spectrometry (XL-MS) can confirm and extend this interactome. Here, intact yeast nuclei were subject to cross-linking with disuccinimidyl sulfoxide (DSSO) and analyzed using hybrid MS2-MS3 methods. XlinkX identified a total of 2,052 unique residue pair cross-links at 1% FDR. Intraprotein cross-links were found to provide extensive structural constraint data, with almost all intralinks that mapped to known structures and slightly fewer of those mapping to homology models being within 30 Å. Intralinks provided structural information for a further 366 proteins. A method for optimizing interprotein cross-link score cut-offs was developed, through use of extensive known yeast interactions. Its application led to a high confidence, yeast nuclear interactome. Strikingly, almost half of the interactions were not previously detected by two-hybrid or AP-MS techniques. Multiple lines of evidence existed for many such interactions, whether through literature or ortholog interaction data, through multiple unique interlinks between proteins, and/or through replicates. We conclude that XL-MS is a powerful means to measure interactions, that complements two-hybrid and affinity-purification techniques.
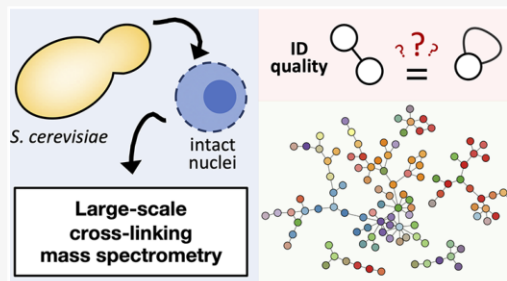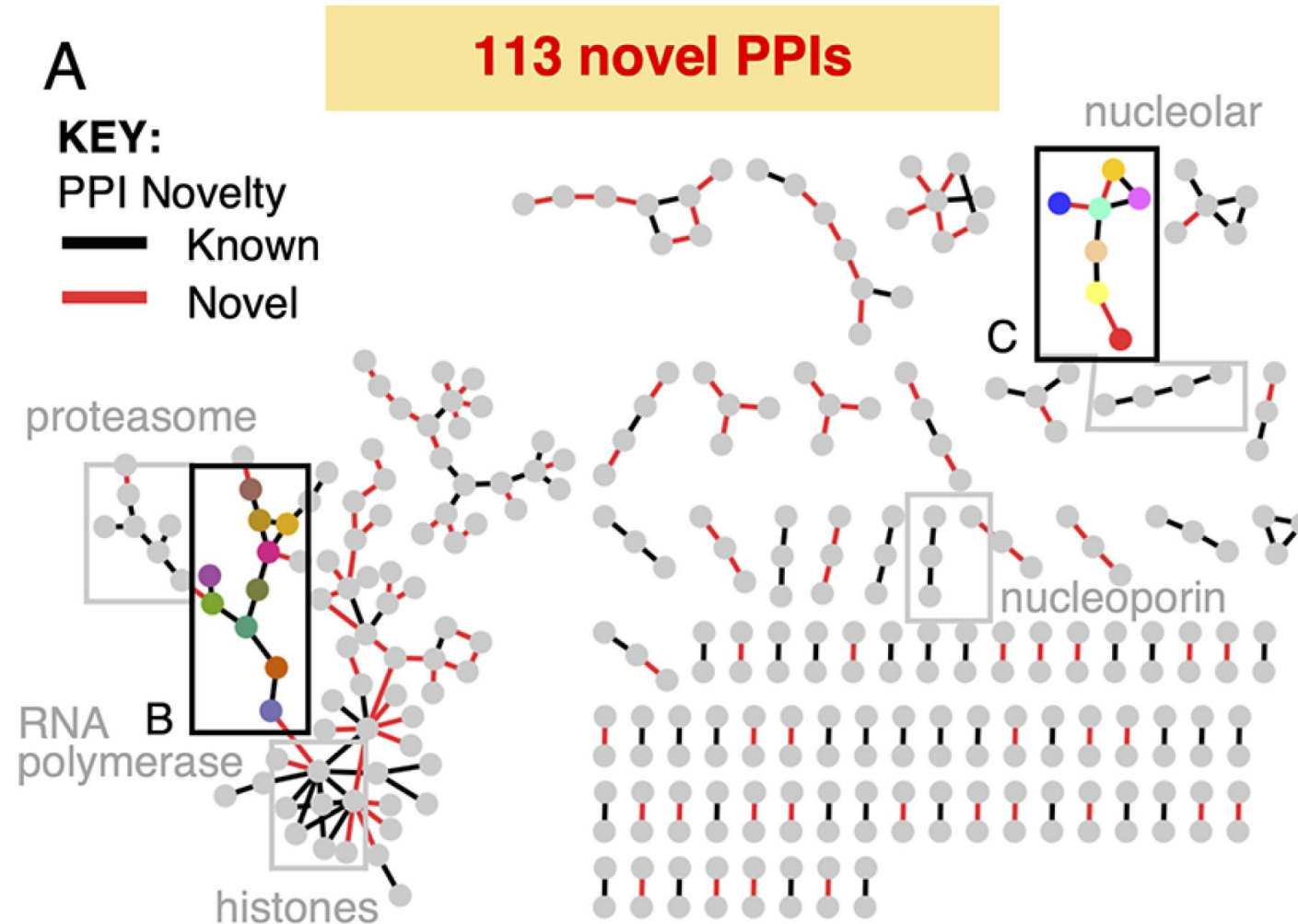
Tara Bartolec

Cross-linking Mass Spectrometry
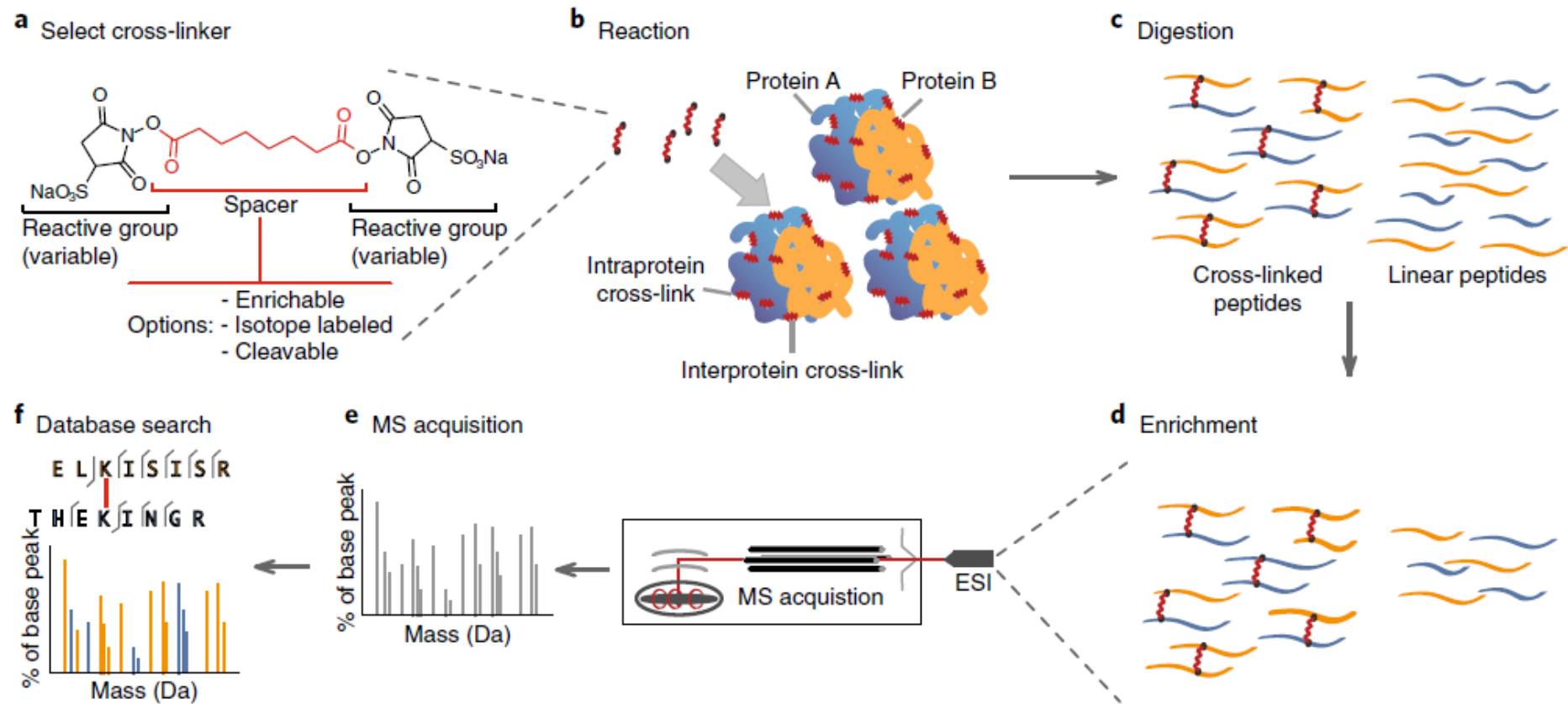
O'Reilly, F.J., Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. Nat Struct Mol Biol 25, 1000–1008 (2018).

Yeast Nuclear XL-MS Protein Interaction Network

Yeast XL-MS Interactions Edge List Table

Protein 1	Protein 2	Name Protein 1	Name Protein 2	PPI Novelty	PPI Evidence Info Group	Number Unique Lys-Lys Contacts
P02293	P04911	H2B1	H2A1	Known	Structure	12
P02293	P02309	H2B1	H4	Known	Structure	6
P02994	P32471	EF1A	EF1B	Known	Structure	5
P0CX51	P38011	RS16A	GBLP	Known	Structure	1
P02406	P0CX49	RL28	RL18A	Novel	STRING	2

Tutorial Google Drive and Document

https://tinyurl.com/te3b5c4

Input datasets and tutorial document

Please feel free make save a copy of the data and documents and make notes on the document.