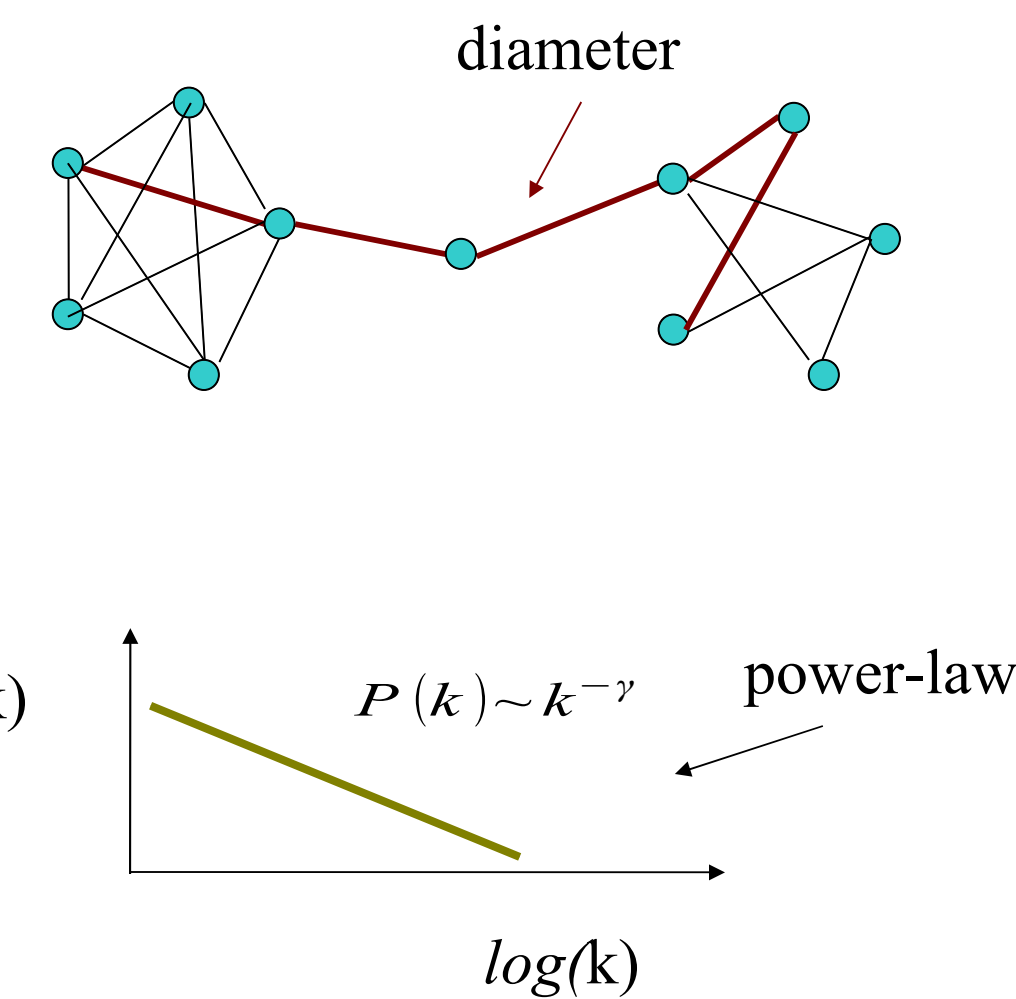


System Description

- A system for analysing molecule interaction networks
 - Interactions retrieved from the MiMI database
 - Uses the Network module of Clairlib
 - Molecule-Specific Networks:**
 - Network of interactions in the neighborhood of a molecule
 - Disease-Specific Networks:**
 - Network of interactions in the neighborhood of known disease genes
 - Prostate Cancer, Type 1 Diabetes, Type 2 Diabetes, Bipolar Disorder
 - General network statistics
 - Ranking molecules based on centrality and second neighbor statistics

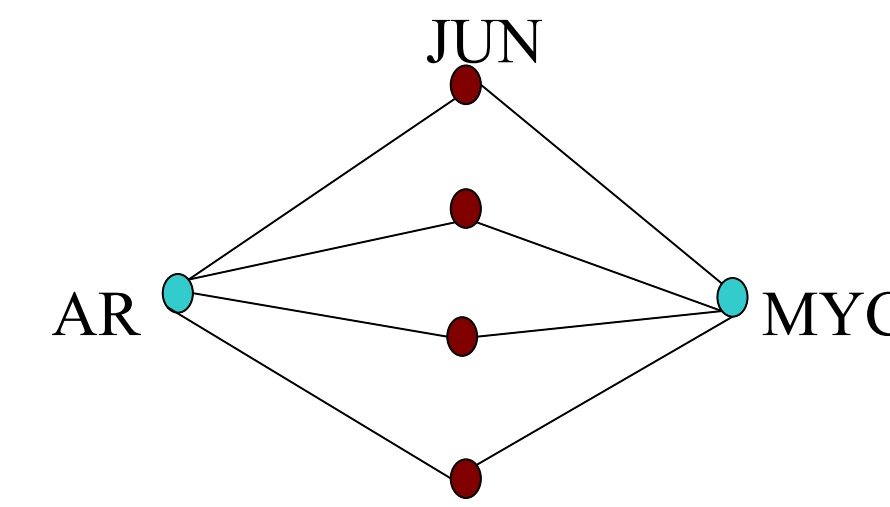
General Network Statistics

- Diameter
- Degree Statistics:
 - Average degree
 - Power-law degree distribution (exponent and r-squared statistics)
- Clustering:
 - Watts Strogatz clustering coefficient:**
 - How well connected a molecule's neighbors are.
 - (# of interactions among a molecule's neighbors) / (# of possible interactions among them)
 - Newman clustering coefficient:**
 - Measures the density of triangles (local clusters) in the network.
 - $3 \times (\text{\# of triangles in the network}) / (\text{\# of connected triples in the network})$
 - Shortest Path Statistics:**
 - Clairlib average undirected shortest path:
 - Average of the shortest paths between all molecules
 - Ferrer average undirected shortest path:
 - Average of all the average path lengths between all molecules
 - Harmonic mean geodesic distance:
 - Harmonic mean of the shortest paths between all pairs of molecules



Ranking Molecules: Second Neighbor Statistics

- Rank second neighbors based on the number of common immediate neighbors
 - AR – JUN – MYC
 - AR and MYC not directly connected
 - They have 45 common neighbors such as JUN.



Molecule-Specific Networks

general network statistics

search molecule in MiMI

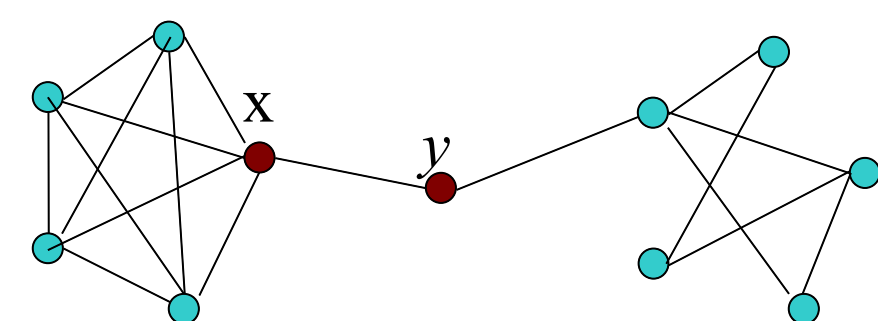
visualize network on cytoscape

most central molecules

second neighbors

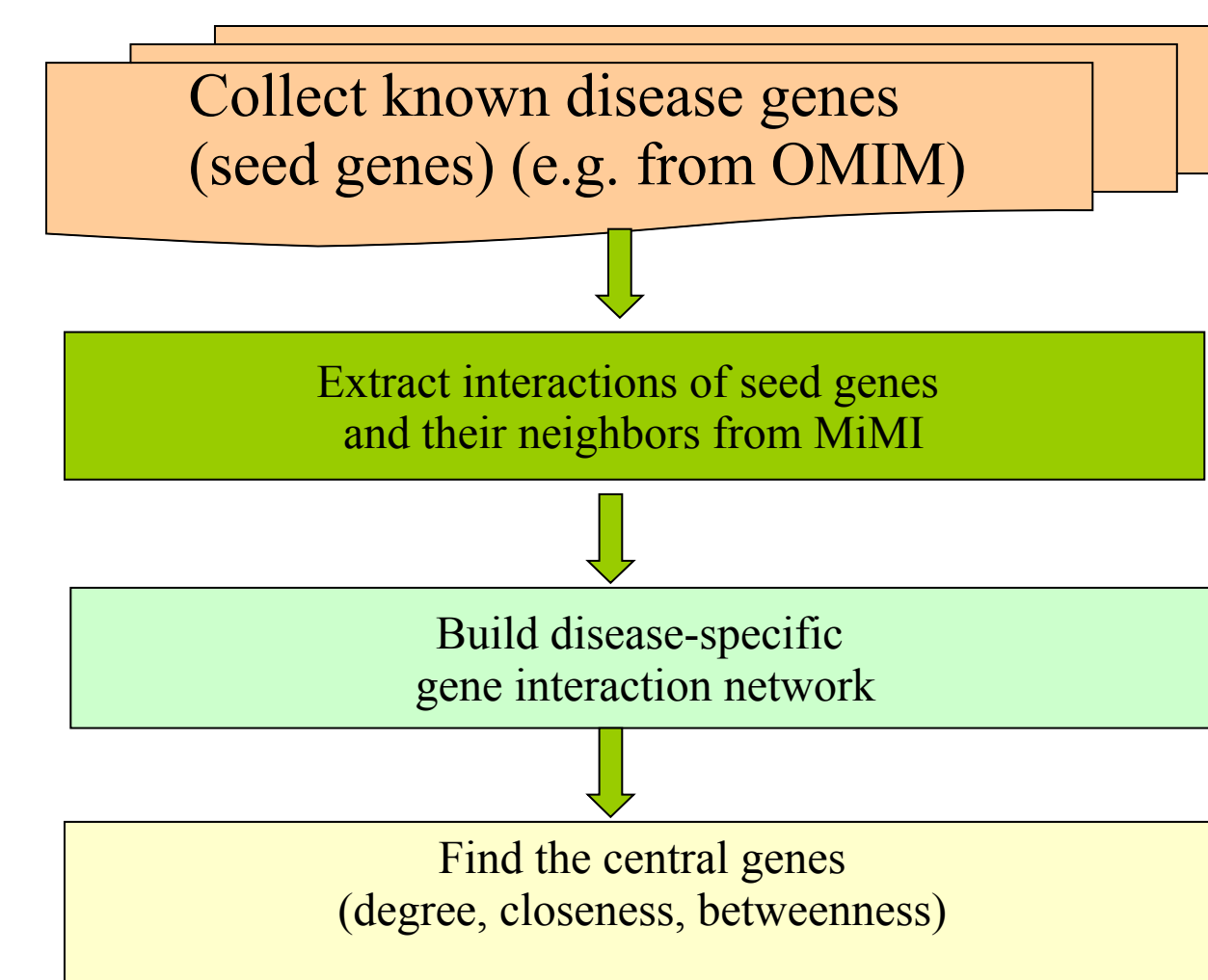
Ranking Molecules based on Network Centrality

- Centrality: Importance of a node in the graph
 - Degree:** The more neighbors a node has, the more important it is. (most central node: x)
 - Betweenness:** The more shortest paths pass through the node, the more important it is. Control of a node over the information flow of the network. (most central node: y)
 - Closeness:** The closer a node to the other nodes, the more important it is. (most central node: y)



Disease-Specific Networks

Network analysis methods to predict gene-disease associations.



Hypothesis: Genes central in the disease-specific gene interaction network are likely to be related to the disease

Disease-Specific Networks for the Driving Biological Problems

Seed genes: OMIM Morbid Map

Seed genes: (Serretti and Mandelli, 2008)

Seed genes: OMIM Morbid Map

Seed genes: (Scott et al., 2007)

References:

- A. Ozgur, T. Vu, G. Erkan, and D. R. Radev. "Identifying gene-disease associations using centrality on a literature mined gene interaction network", *Bioinformatics*, Vol. 24, Num. 13, pp. 1277-1285, 2008.
- D. R. Radev, M. Hodges, A. Fader, M. Joseph, J. Gerrish, M. Schaller, J. dePeri, and B. Gibson. Clairlib documentation v1.03. Technical Report CSE-TR-536-07, University of Michigan. Department of EECS, 2007.
- Scott et al. Genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007; 316(5829): 1341-5. 20.
- Serretti and Mandelli. The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions. *Mol Psychiatry*. 2008 Aug;13(8):742-71.

Acknowledgements

This work was supported by National Institutes of Health: Grant #US4 DA021519.

We would like to thank Jing Gao, Glenn Tarcea, Aaron Bookvich, Terry Weymouth, and H. V. Jagadish for their contributions to the project.