PubChem Bioassays as a Source of Polypharmacology

Bin Chen, David Wild, Rajarshi Guha

School of Informatics
Indiana University

236th ACS National Meeting
17th August, 2008

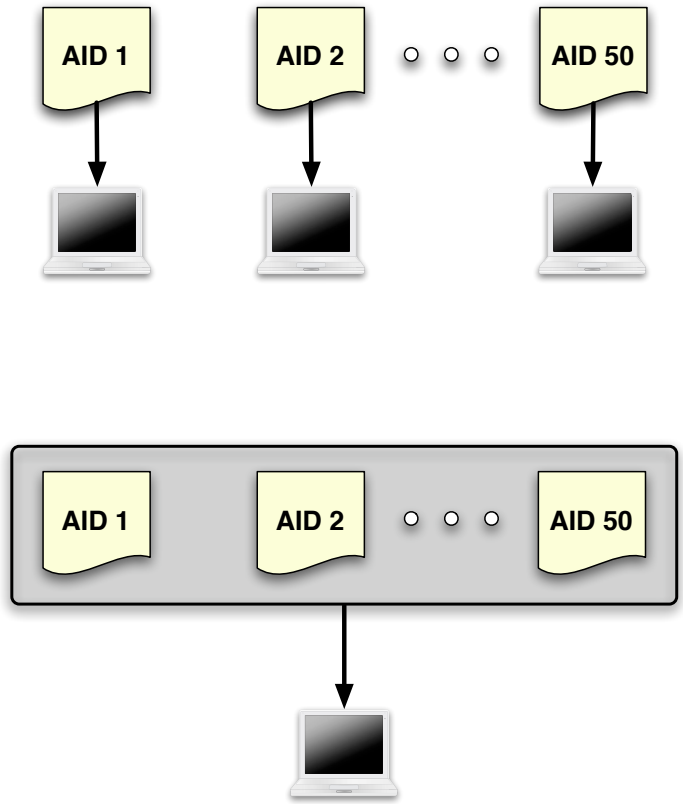
PubChem Bioassays

- Currently contains 1157 assays
  - A number are follow ups of primary screens
- Assay size ranges from 2 to 224,000 molecules
- Many compounds tested in multiple assays
- PubChem web interface support queries that focus on individual assays
- Cross-assay queries can be tough

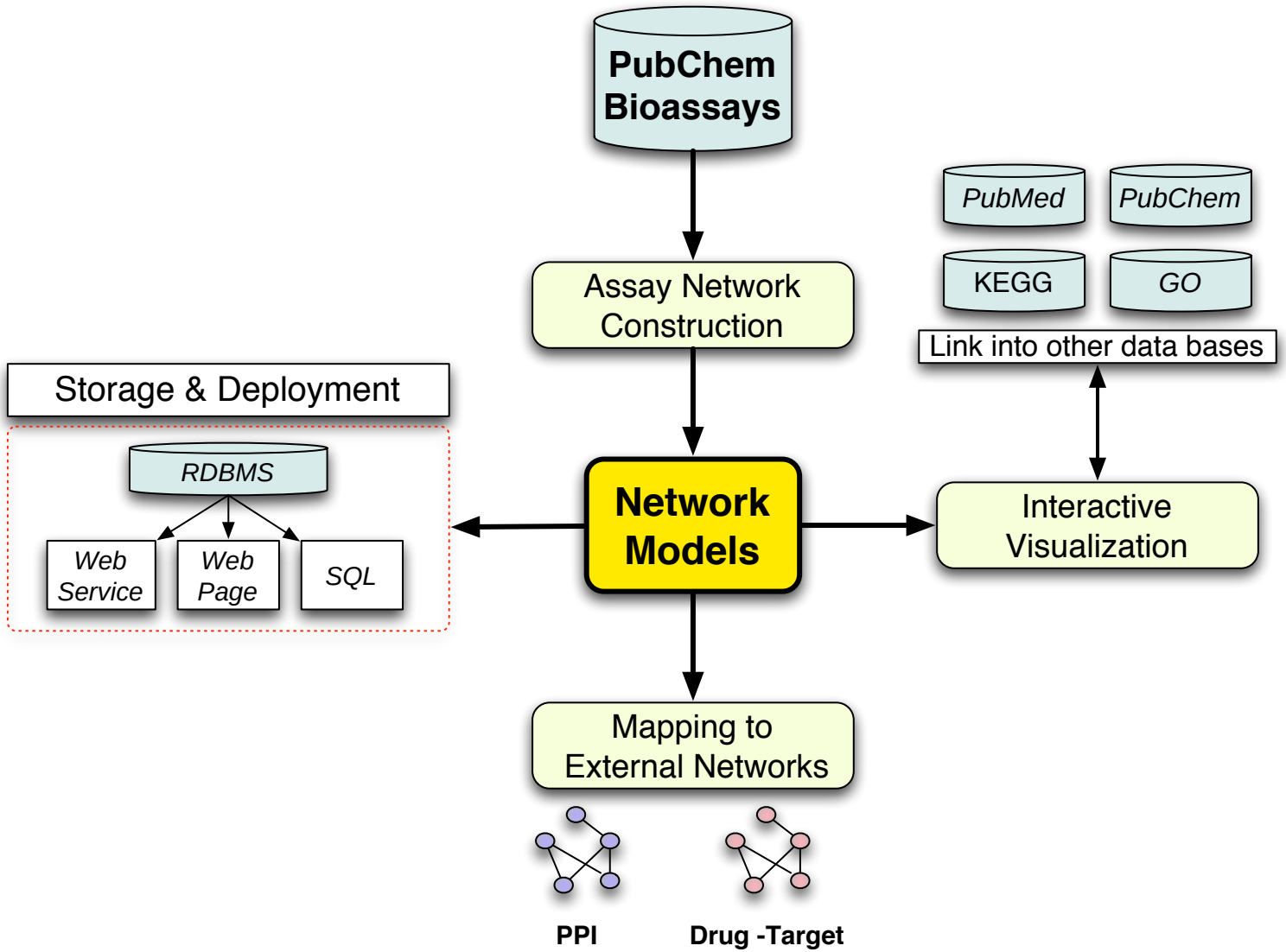
Assay Content

- The data is obviously primary
- But the assay description and target are also useful pieces of information
- Can we combine
  - data
  - target
  - description

  across multiple assays to draw conclusions, gain insight?



A Network Model of Bioassays - Goals



Mapping Assay Networks to Real Networks

- An assay network is an artificial network - does not necessarily have physical meaning
- We need to map the assay network onto a *real* biological network
  - PPI networks
  - metabolic networks
  - drug target networks
- Using the mapping, we'd like to identify MLSCN compounds that might be active against one or more nodes in the *real* network

The stepping stones . . .

- How do we construct the assay network?
- How do we map the network?

Mapping Assay Networks to Real Networks

- An assay network is an artificial network - does not necessarily have physical meaning
- We need to map the assay network onto a *real* biological network
  - PPI networks
  - metabolic networks
  - drug target networks
- Using the mapping, we'd like to identify MLSCN compounds that might be active against one or more nodes in the *real* network

The stepping stones ...

- How do we construct the assay network?
- How do we map the network?

Why Perform a Mapping?

- Identify compounds that interacts with two targets in different pathways
- Alternatively, identify compounds that interact with a target in a pathway but not in another pathway
- Identify compounds capable of disrupting protein-protein interactions
- Our ability to do these will depend on the quality of assay data and the way we map the assay network to the *real* network

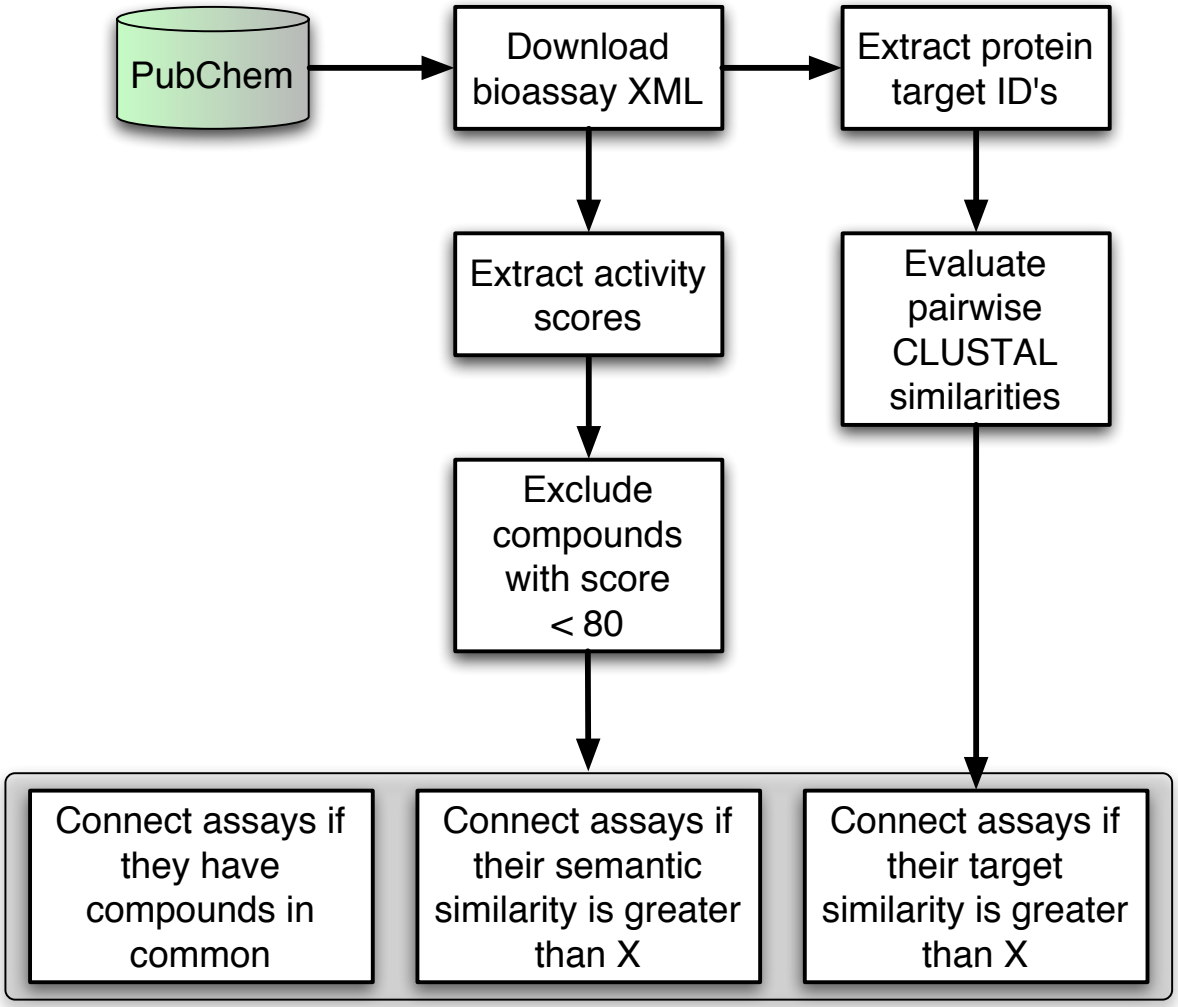
Hopkins, A.L. et al, *Curr. Opin. Chem. Biol*, **2006**, *16*, 127–136

Why Perform a Mapping?

- Identify compounds that interacts with two targets in different pathways
- Alternatively, identify compounds that interact with a target in a pathway but not in another pathway
- Identify compounds capable of disrupting protein-protein interactions
- Our ability to do these will depend on the quality of assay data and the way we map the assay network to the *real* network

Hopkins, A.L. et al, *Curr. Opin. Chem. Biol*, **2006**, *16*, 127–136

Assay Network Construction



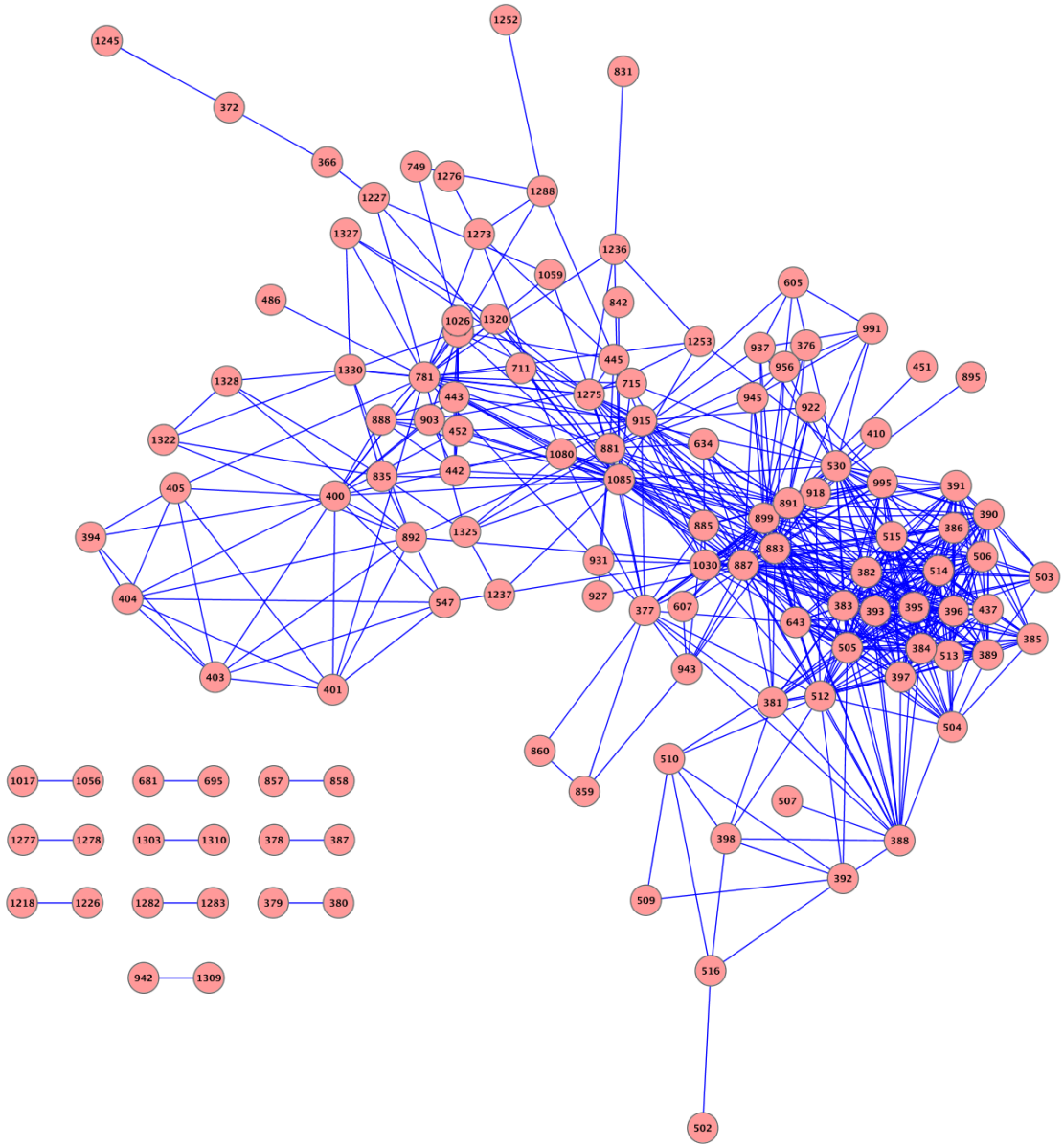
Assay Network Construction

- We will focus on a compound-centric network
- A *semantic* network requires some form of annotation on the assays
- Initial attempts at annotation assays based on GO terms (via descriptions)
- Alternatively, could consider deriving annotations based on the targets
- Using protein target similarity restricts one to enzymatic assays which leads to a relatively small assay network

Assay Network Construction – Caveats

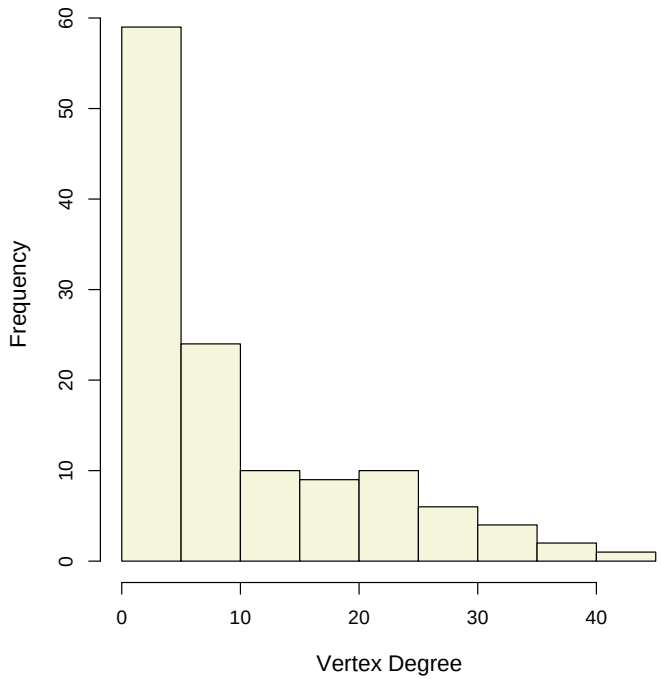
- A compound-centric network is not very rigorous
- The PubChem activity score is known to be noisy
  - Currently the only way to look at assay readouts over the whole collection
- Using an activity score cutoff of 80 is arbitrary
- We haven't considered promiscuity directly, though a filter would be useful

Assay Network - Common Compounds



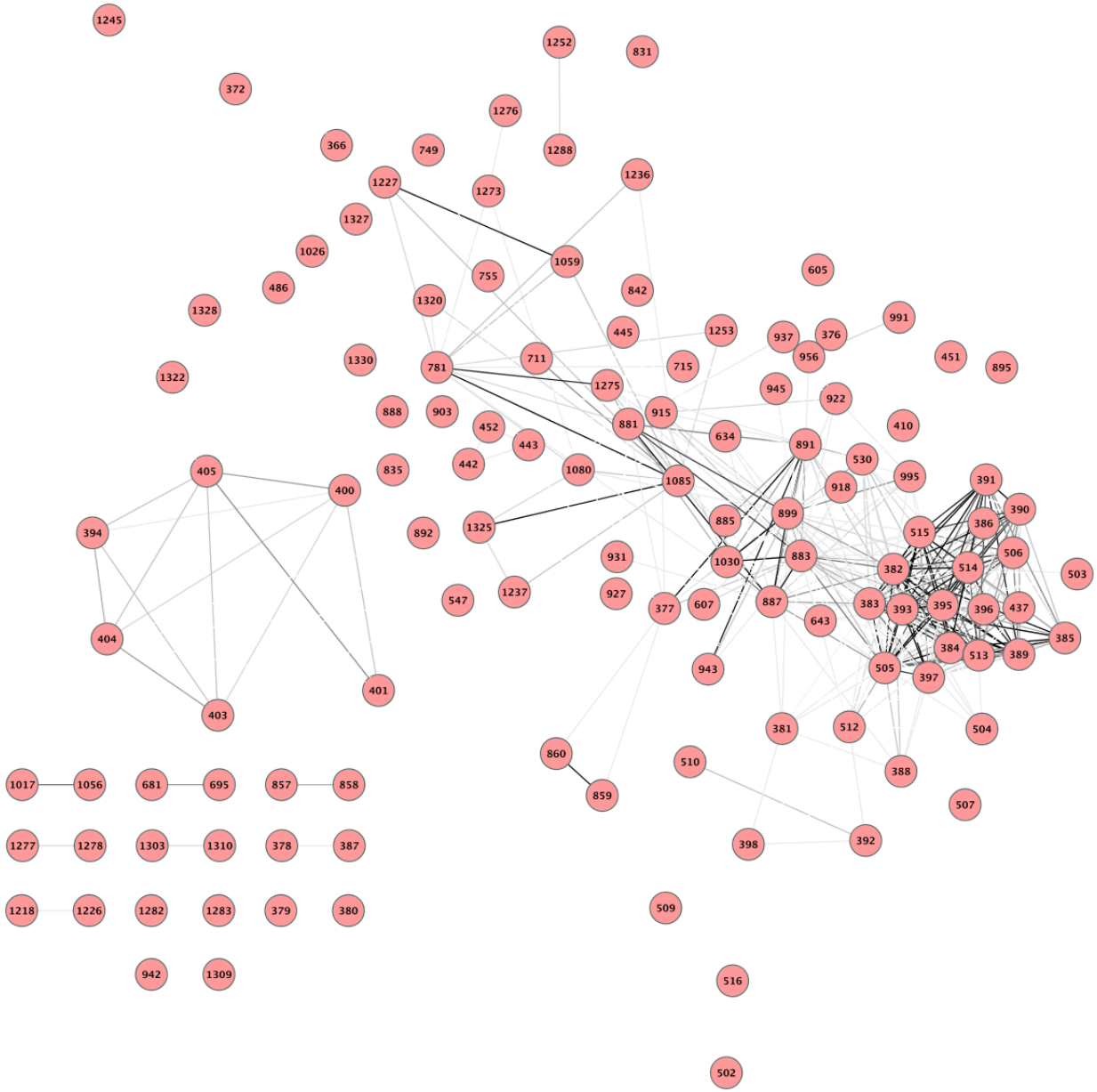
Some Network Statistics

- 222 assays with a single target
- Selected the smallest assay if more than assay had the same target
- $N = 125, E = 598$
- $V_{max} = 40, V_{avg} = 9.6$
- $\bar{C} = 0.67$

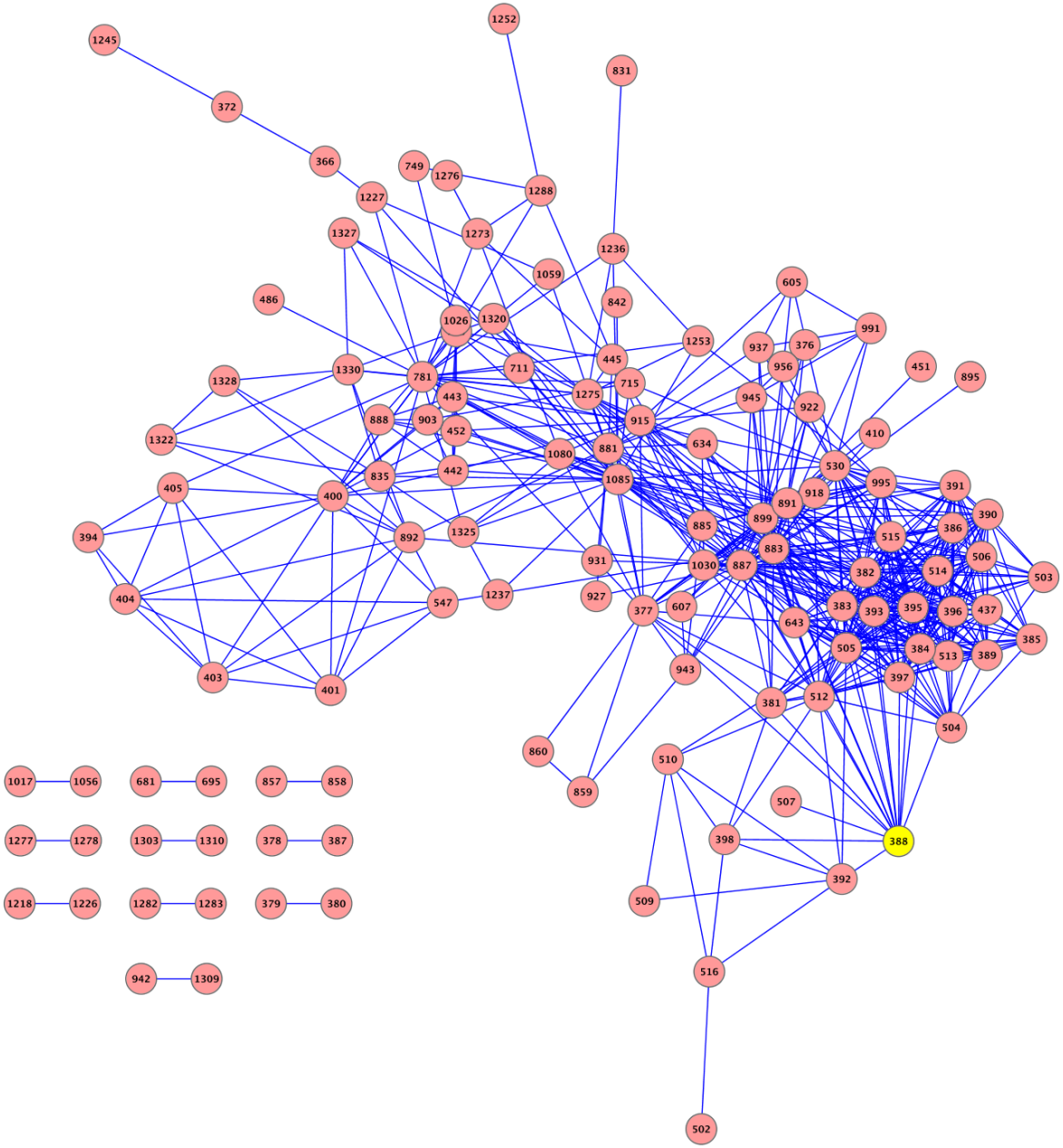


Histogram of vertex degree

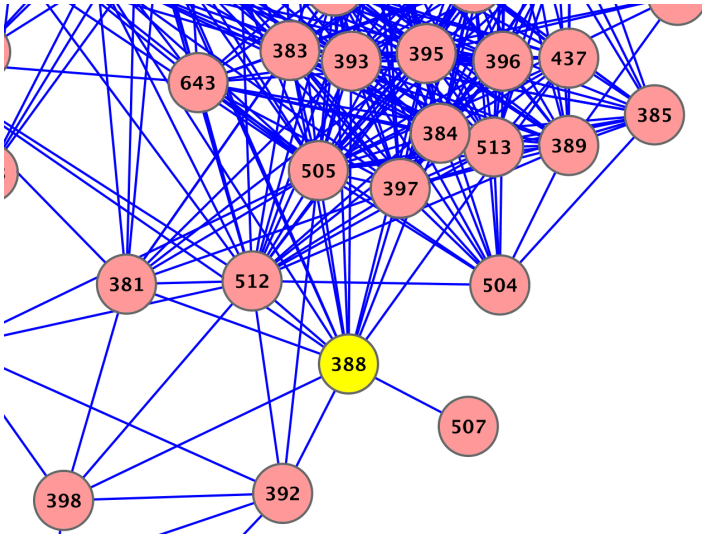
Clustering in the Assay Network



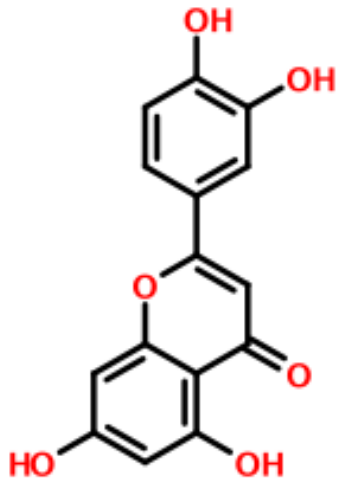
Assay Network - Common Compounds



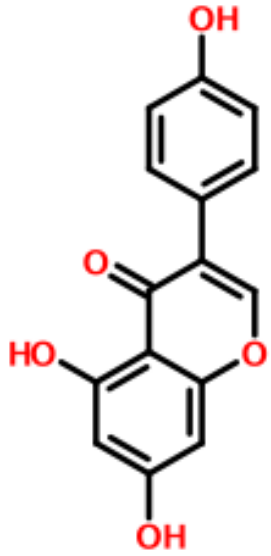
Assay Network - Common Compounds



- 388 targets NAD^+-dependent 15-hydroxyprostaglandin dehydrogenase
- Has active compounds common with
  - pim-2-oncogene (505)
  - 15-lipoxygenase (887)
  - aldo-keto reductase (381)

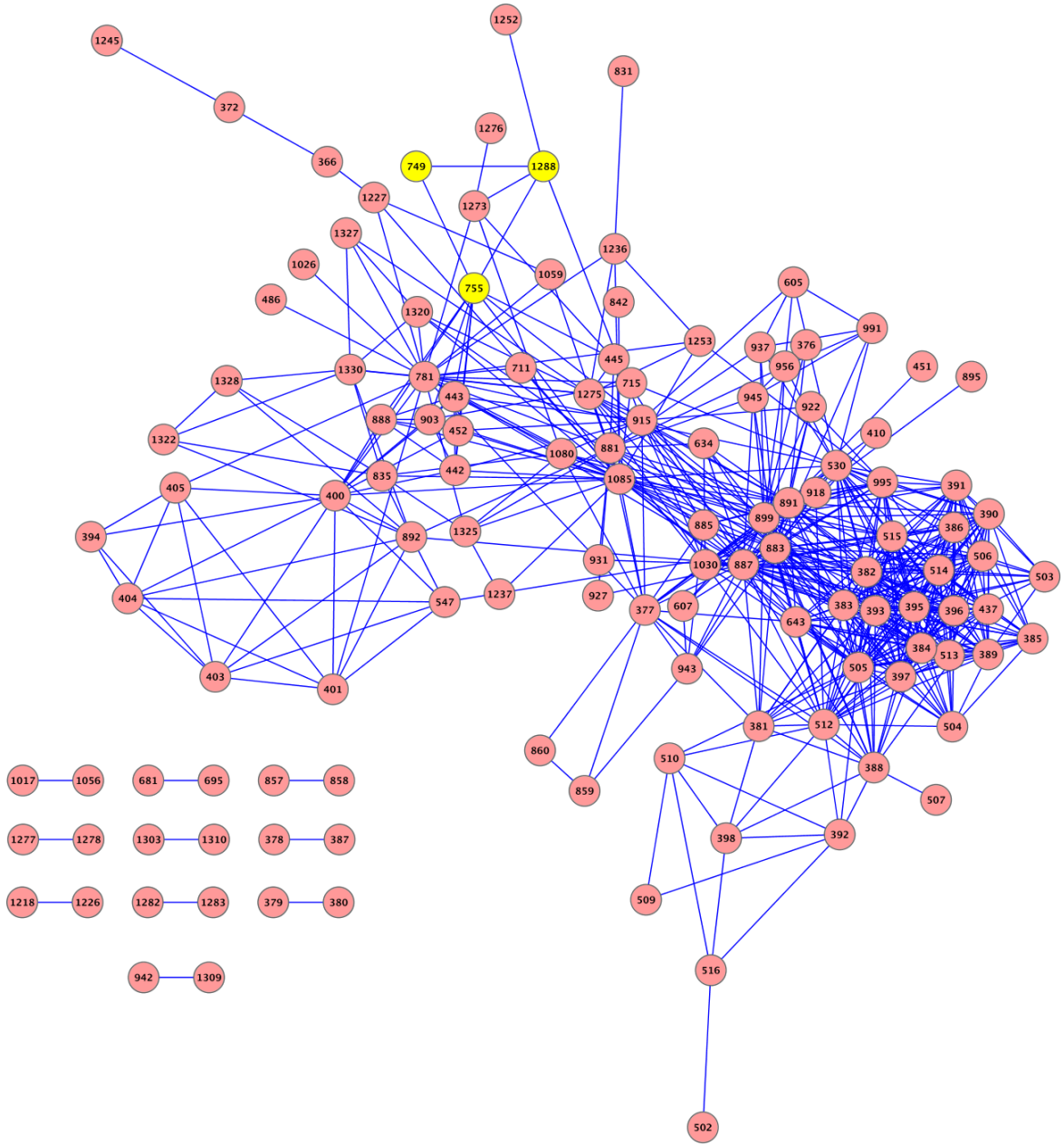


Luteonin

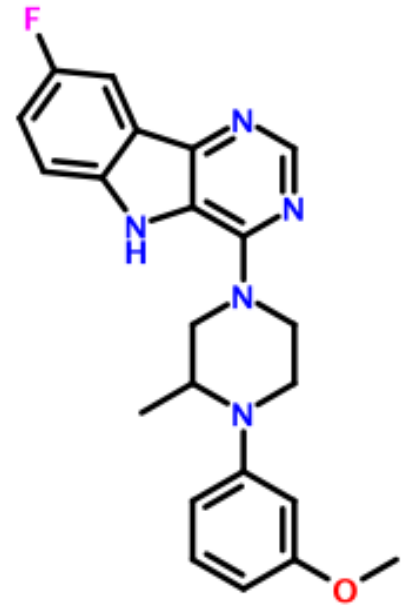


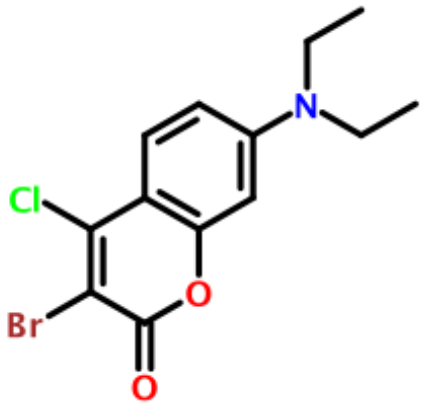
Genistein

Assay Network - Common Compounds



Assay Network - Common Compounds

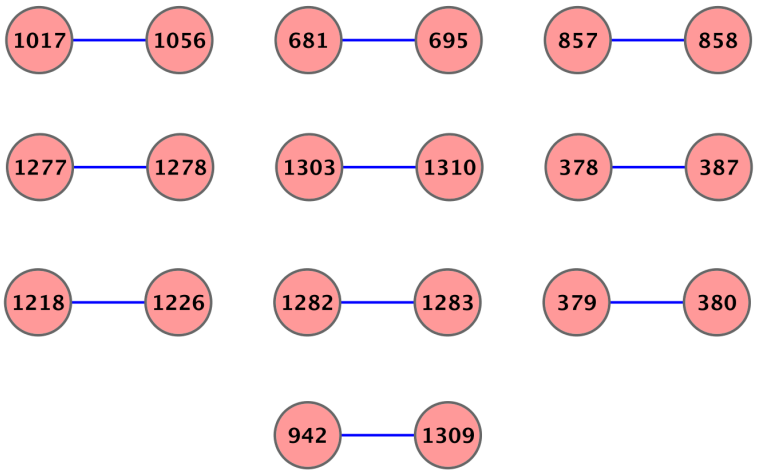




- 749 and 755 target 5-HT$_{1E}$ and 5-HT$_{1A}$ respectively
- Both have a (different) compound in common with 1288 (selectin E)
- Probably promiscuous given that they are also active in many other assays
- But a selectin inhibitor is known to reduce hyperalgesia by blocking 5-HT$_3$

Oliviera, M.C.G. et al, *Neuroscience*, **2007**, *145*, 708–714

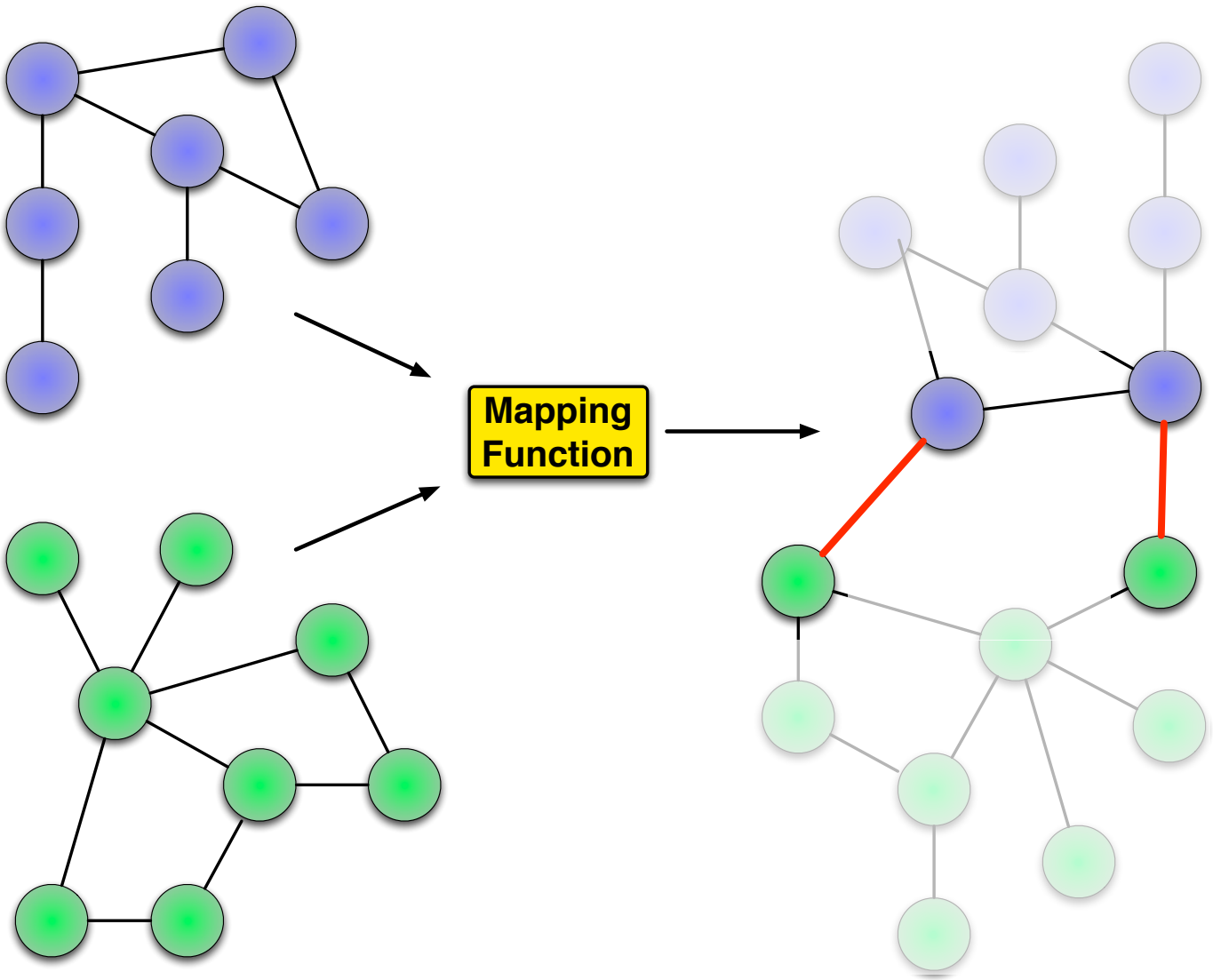
Assay Network - Common Compounds



- Most of these assay pairs have closely related targets
- Tissue non-specific alkaline phosphatase and intestinal alkaline phosphatase (1056 & 1017)
- STAT1 and STAT3 (1303 & 1310)
- ER-α and ER-β (1226 & 1228)

lethal factor (*B. anthracis*) and nF-κB (942 & 1309) have one compound in common - podophyllotoxin

Mapping an Assay Network



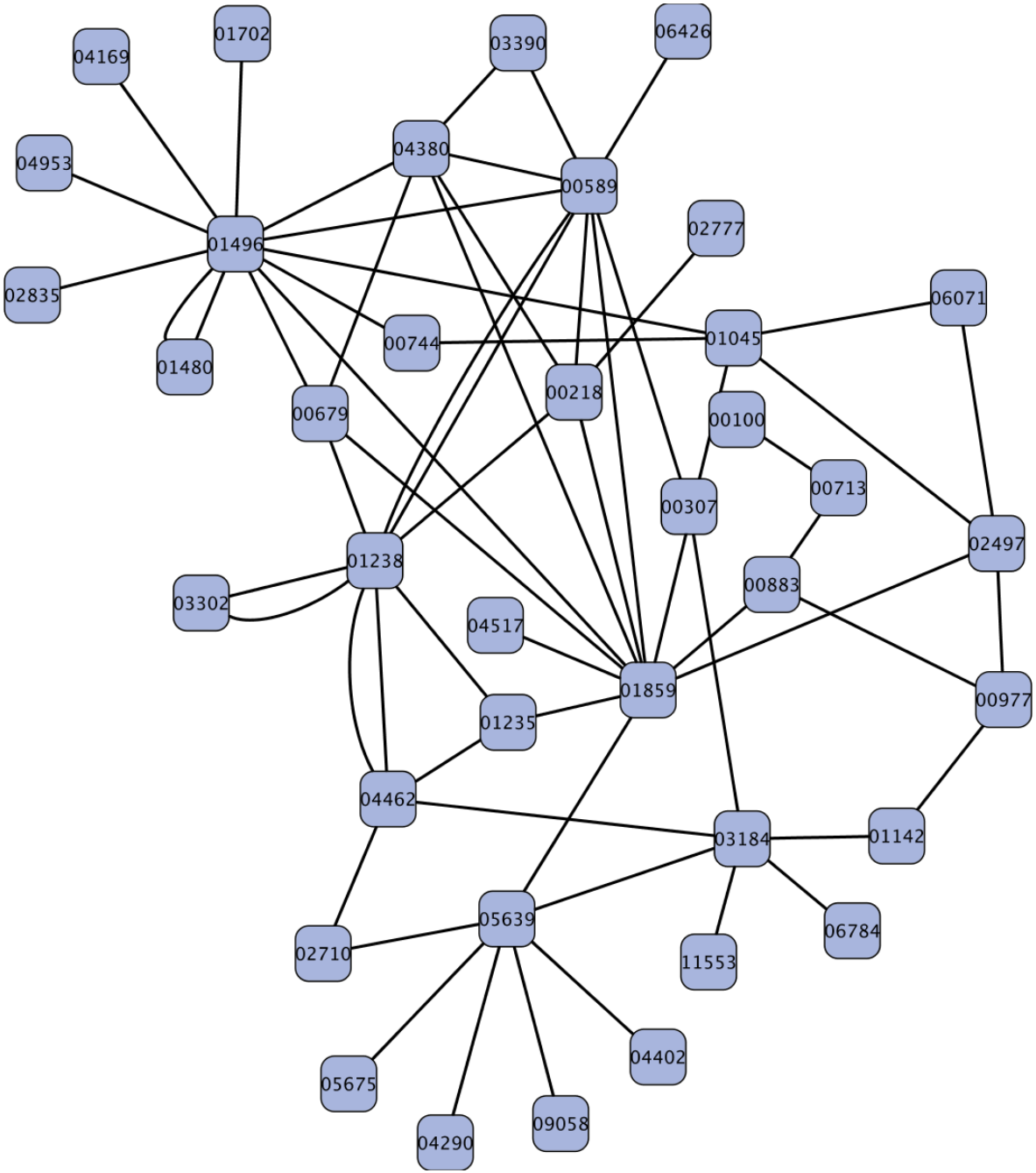
Defining a Mapping Function

- Multiple mapping functions can be defined
  - exact matches between assay target and external targets
  - similarity between target sequences
  - similarity between target binding sites
- One could also map *edges* of one network onto another
  - Dependent on the nature of the external network
- Depending on the nature of the definition, the mapping procedure can be a trivial search or may require an optimization scheme if multiple mappings are possible

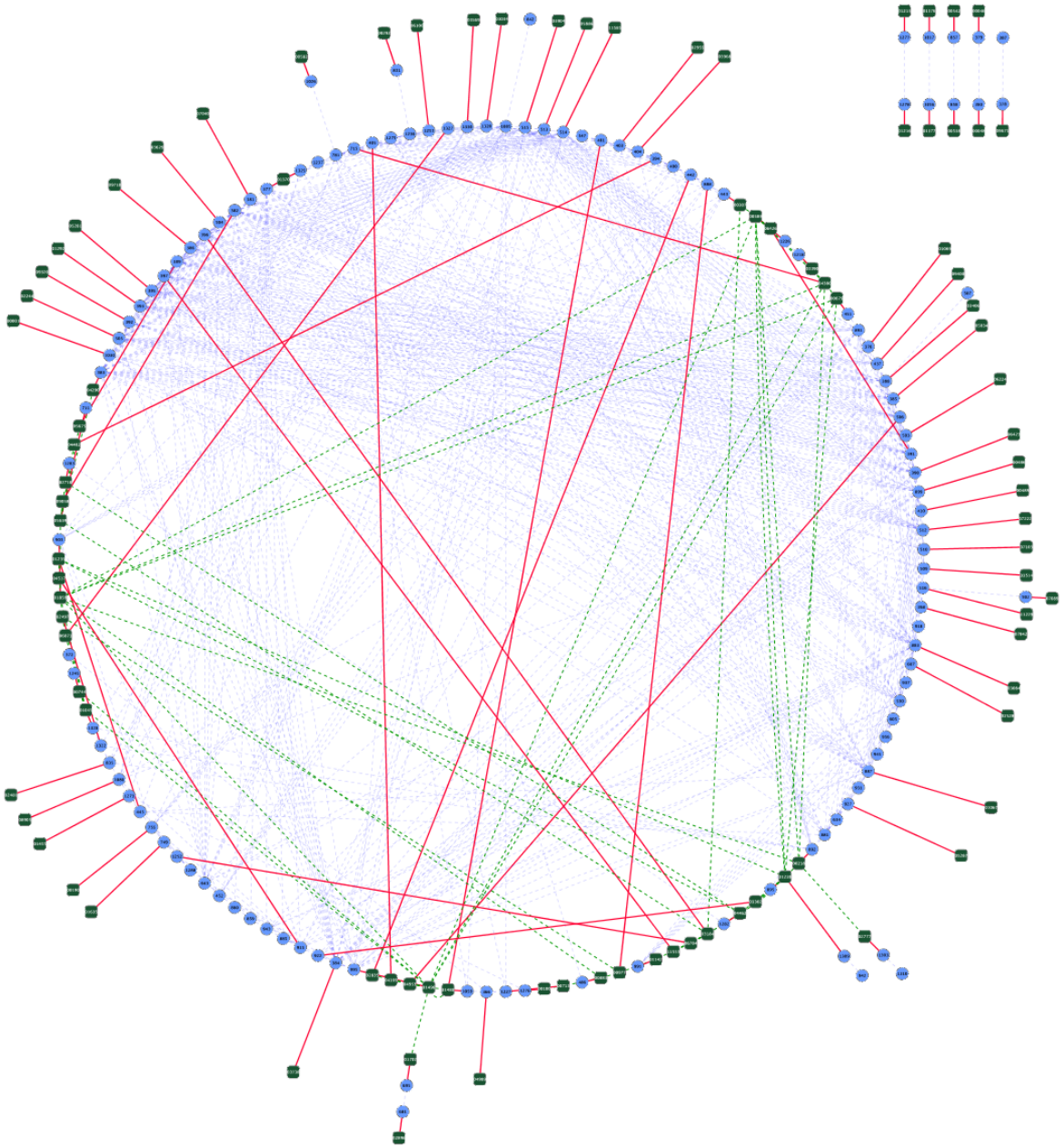
Assay Network to HPRD

- The HPRD database collects protein-protein interaction data and pathway membership
- The July 2007 release lists 31,708 PPI's
- 96 assays can be mapped to the unique proteins in HPRD
- We construct a *HPRD network* by identifying the pairs from the 96 proteins that have a listed interaction
- When mapping the HPRD network to the assay network, we include singleton HPRD nodes

HPRD Network



Assay - HPRD Network Mapping

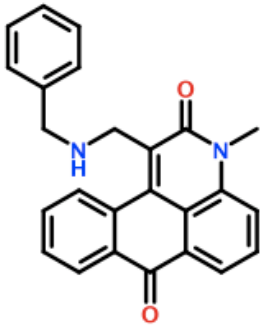


Assay - HPRD Network Mapping

Is this a useful mapping?

- Since we map assays to HPRD entries by target ID, we aren't getting new information on the assays individually
- But we are able to easily identify assay targets that interact with each other (or not)

Comparing Two Assays



CID 126298

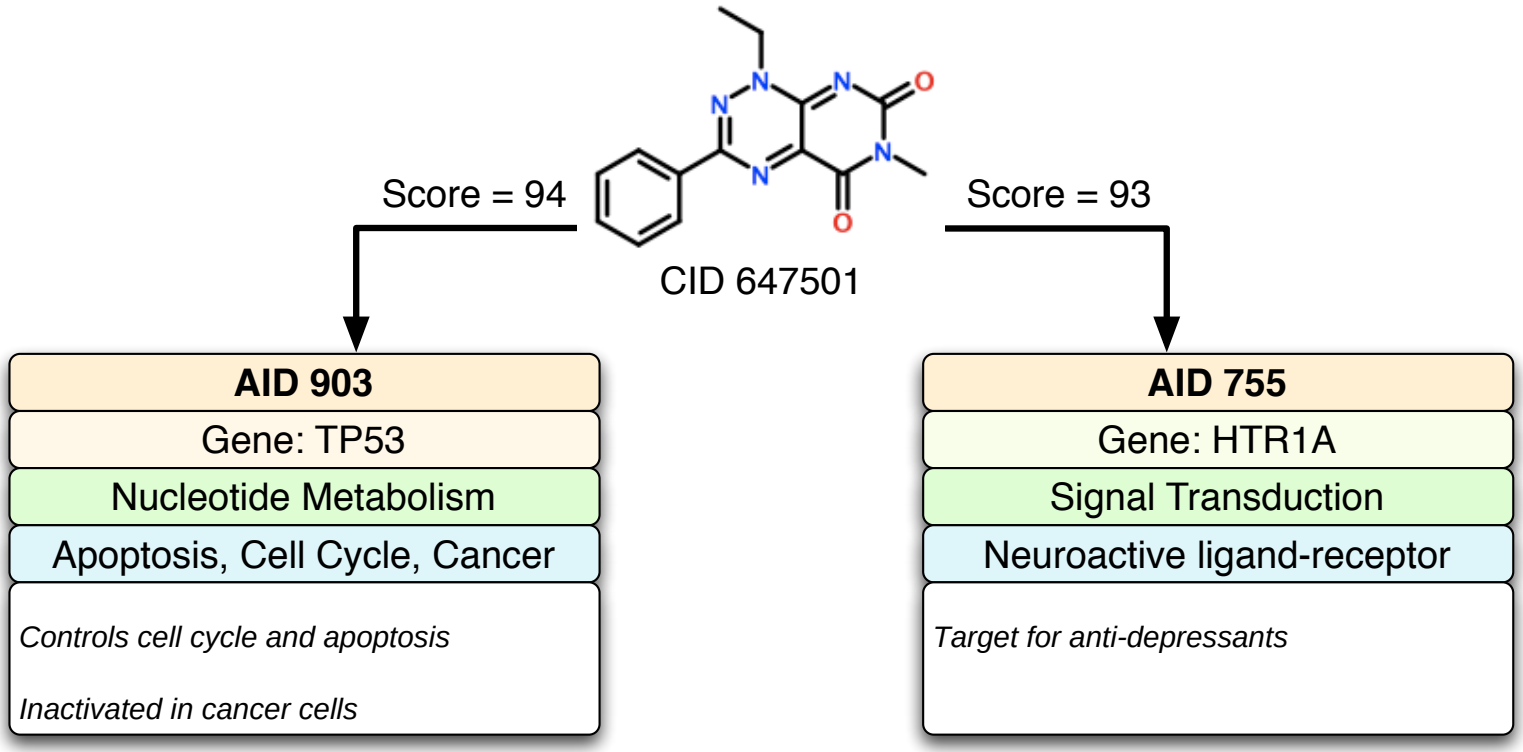
Score = 95

Score = 100

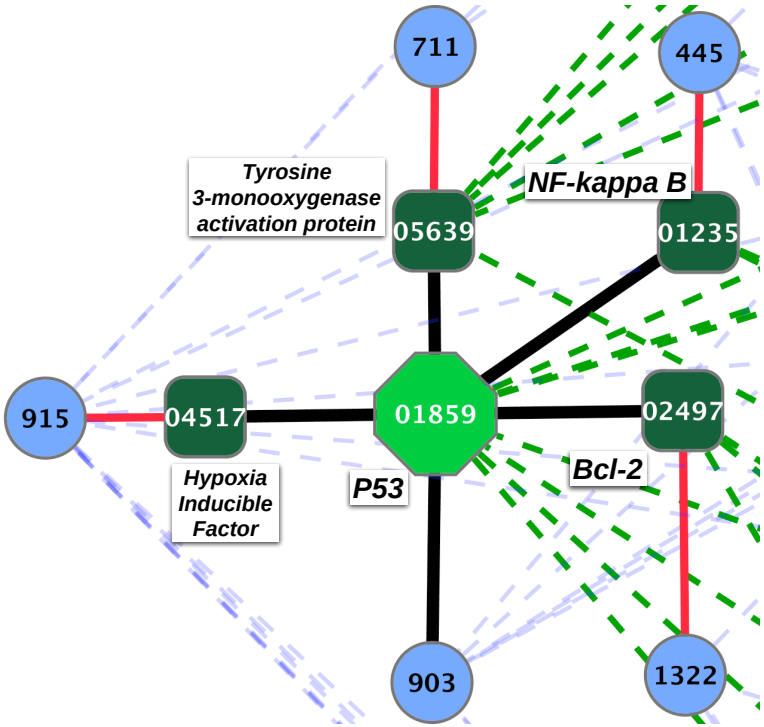
AID 835
Gene: EPHB4
Signal Transduction
Axon guidance pathway
Over expressed in breast carcinoma *Preferentially expressed in veins* *Required for angiogenesis*

AID 1325
Gene: ABCB1
Transporter
ABC transporters
Overexpression is related to multidrug resistance in chemotherapy *Involved in the BBB* *Not expressed very highly in vascular tissue*

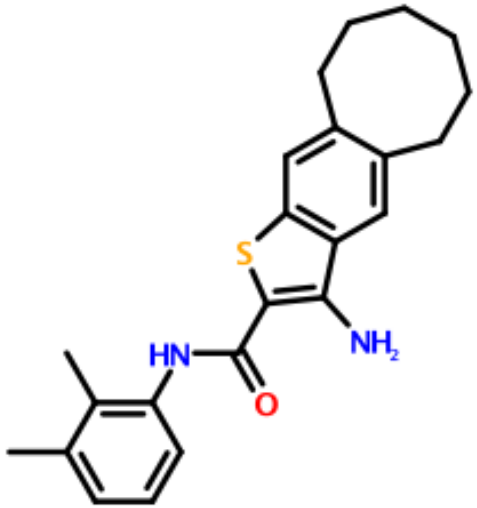
Comparing Two Assays



Disrupting PPI's



- The pairs of interacting targets have compounds tested against both of them
- Majority are inactive or inconclusive in both of them
- CID 1025314 is active in AID 445 but inactive in AID 903



Summary

- A network view of assays provides with a novel tool for visualization and summary of the assay collection
- It's utility beyond visualization is dependent on the way we construct the network
- A compound-centric network allows us to use the assay collection as a probe into external networks
- Future work will investigate different forms of the assay network focusing on protein target and GO annotation similarity