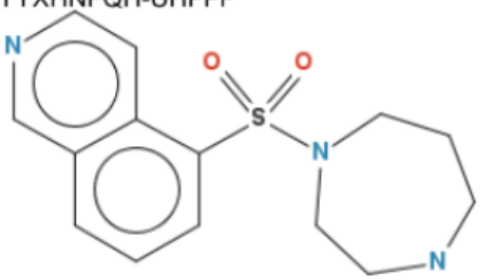
ConTour:
Data-Driven Exploration of Multi-Relational Datasets for Drug Discovery

Christian Partl, Alexander Lex, Marc Streit, Hendrik Strobelt,
Anne-Mai Wassermann, Hanspeter Pfister and Dieter Schmalstieg

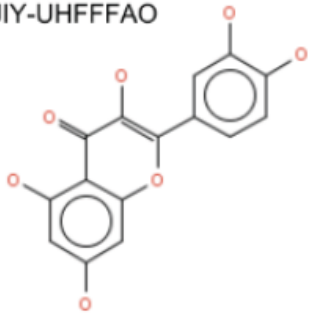
VIS 2014

Why? – Search for “magic bullet”

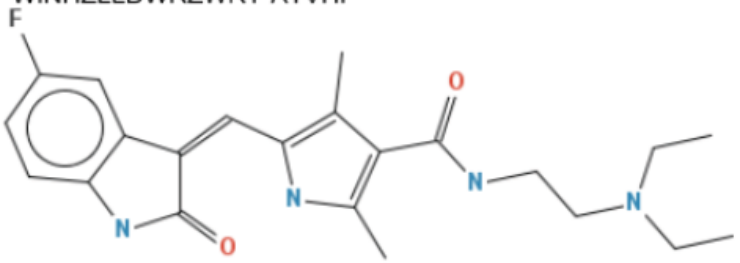
NGOGFTYYXHNFQH-UHFFF



REFJWTPEDVJJIY-UHFFFAO

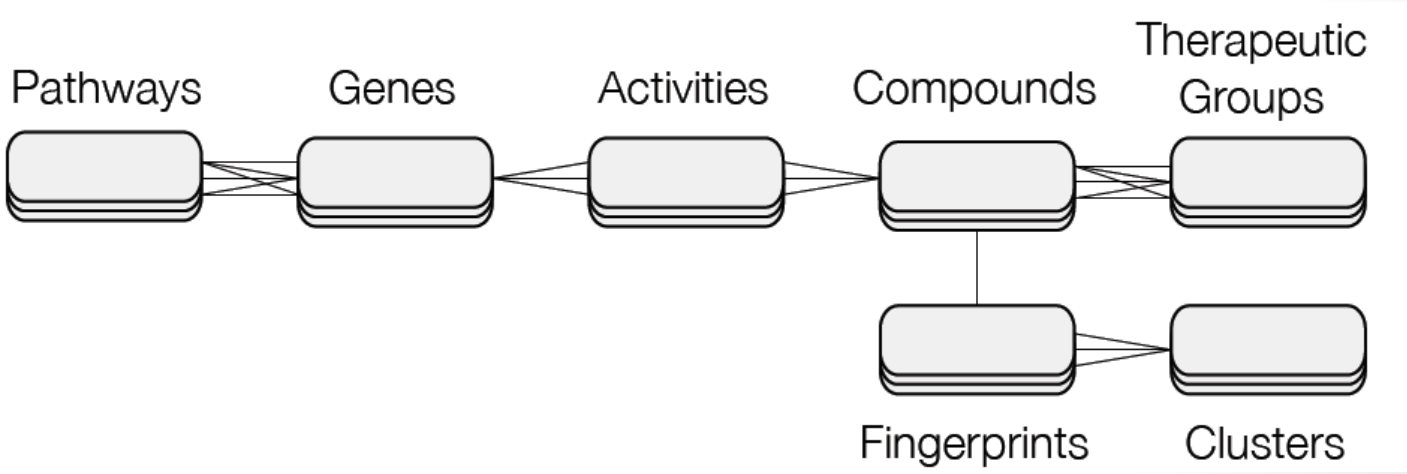


WINHZLLDWRZWRT-ATVHP



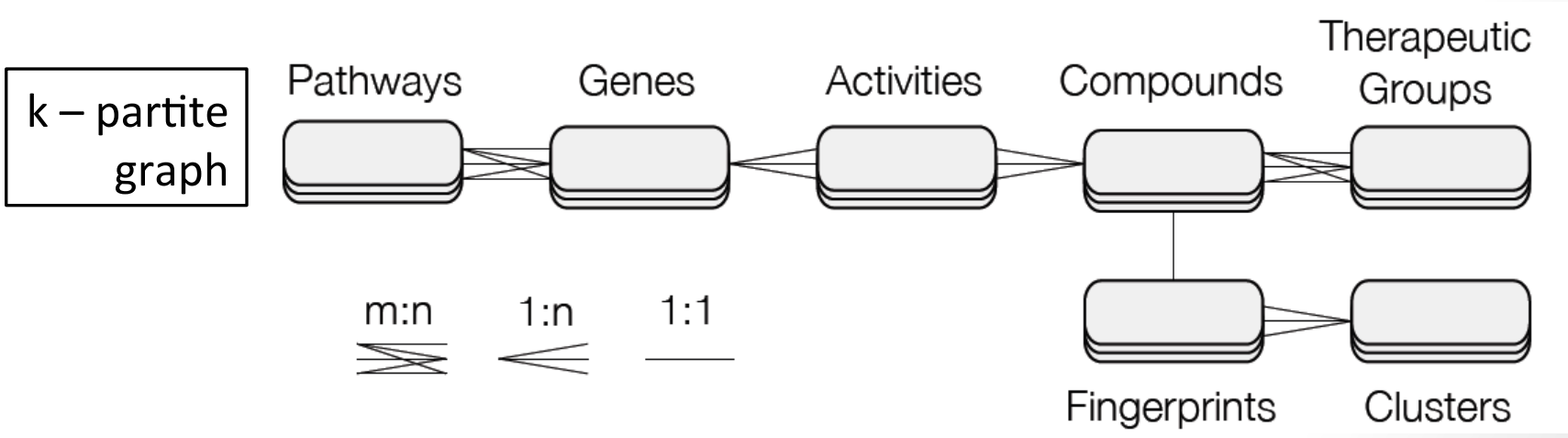
Key word definitions

- Compounds: drugs or drug candidates
- Biological fingerprints: activity of a compound across several experiments of cellular processes
- Pathways: series of action that leads to a change in the cell; regulation of genes and transmission of signals
- Therapeutic groups: treatment that compounds induces



What? – The data set

- Dataset from public bioactivity databases *ChEMBL* [Gaulton et al. *Nucleic acids research* 2012] and *DrugBank* [Law et al. *Nucleic acids research* 2014]
- Drug dataset consists of about 1100 compounds
- Compounds have been profiled in at least 50 different cell-based screens [Petrone et al. *ACS Chemical Biology* 2012]
- Correlation-based similarity measures [Wassermann et al. *J Chem Inf Model* 2013] yielded 100 distinct clusters



Analysis Goals

1. Identify a drug's mechanism of action:

   Drugs in the same cluster are likely to have the same protein target

2. Identify the biological process a drug modulates:

   Compounds binding to different target that are clustered together are likely involved in same biological processes

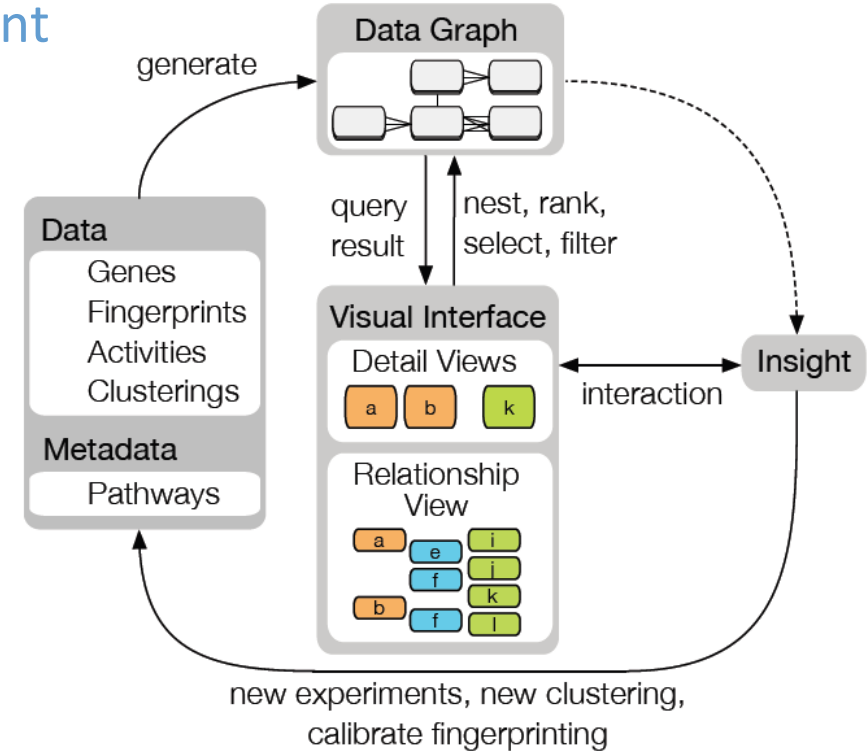
3. Identify new drugs for specific therapeutic indications:

   Compounds clustering with drugs for particular therapeutic indication could be a novel candidate for this therapy

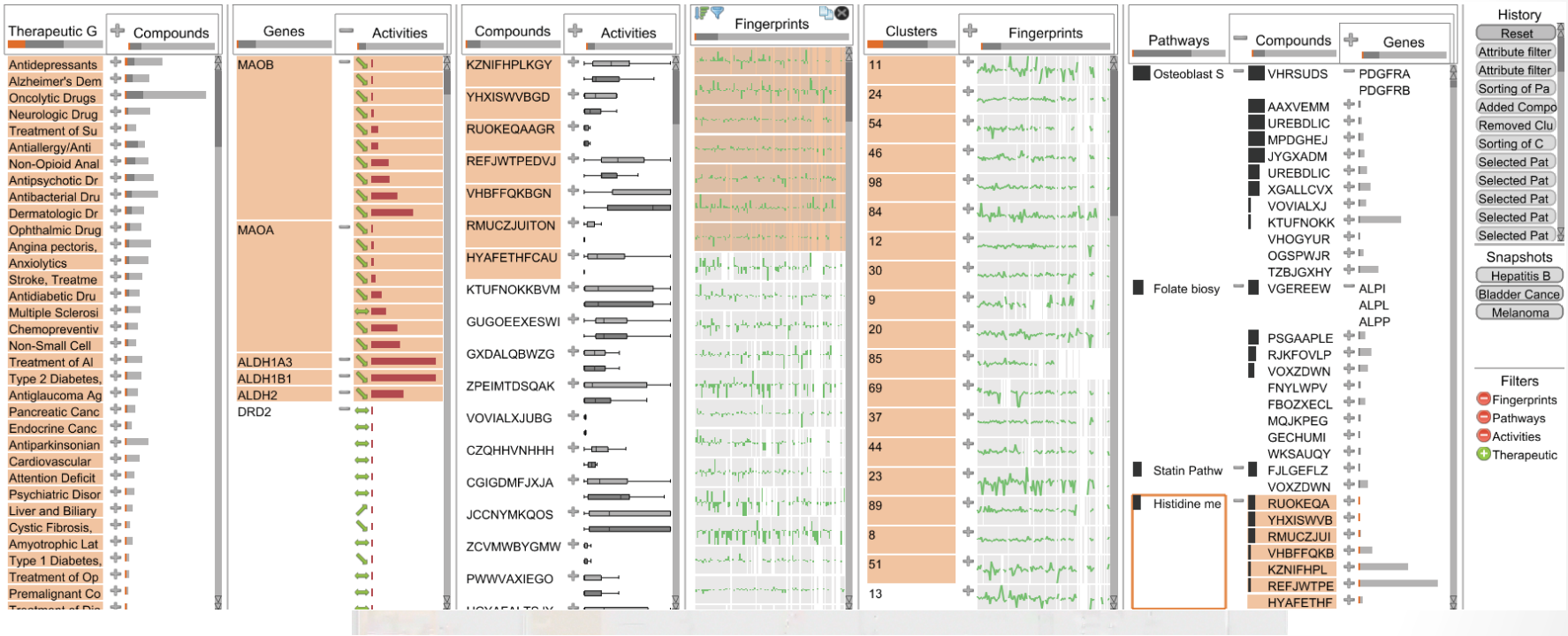
Task analysis

Tasks the analyst needs to perform to achieve previous goals:

1. Identify related items
2. Identify items that share a relationship with a set of items
3. Analyze network enrichment
4. Rank items
5. Filter items
6. View items in detail



ConTour for drug discovery



Task validation

1. Identify related items

   highlighting (hovering, clicking), selection based filtering, nesting, history view
2. Identify items that share a relationship with a set of items

   recursive nesting, history view
3. Analyze network enrichment

   enrichment scores
4. Rank items

   ranking & sorting
5. Filter items

   selection based filtering, filter view
6. View items in detail

   pathway, compound & parallel coordinates view

Conclusion

- Strength
  - Highly exploratory through ranking, sorting and filtering
  - Integrates overview, detailed view and support views
  - Simple and recursive nesting illustrates parent-child relationships
  - Case study showed that ConTour is an effective tool for interactively exploring relationships in drug discovery
  - Applicable to other biological and non-biological domains
- Weakness
  - Scaling to higher number of columns difficult due to limited space
  - Nesting approach is not very space efficient
  - Relationship between items of the sets are of arbitrary cardinality → problematic for data graphs containing cycles