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

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

Why? – Search for “magic bullet”

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: treatments that compounds induce

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

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

What? – The data set
- Drug dataset consists of about 1500 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

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

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