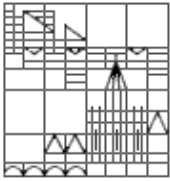




**Nycomed Chair for
Bioinformatics & Information Mining**

Universität Konstanz



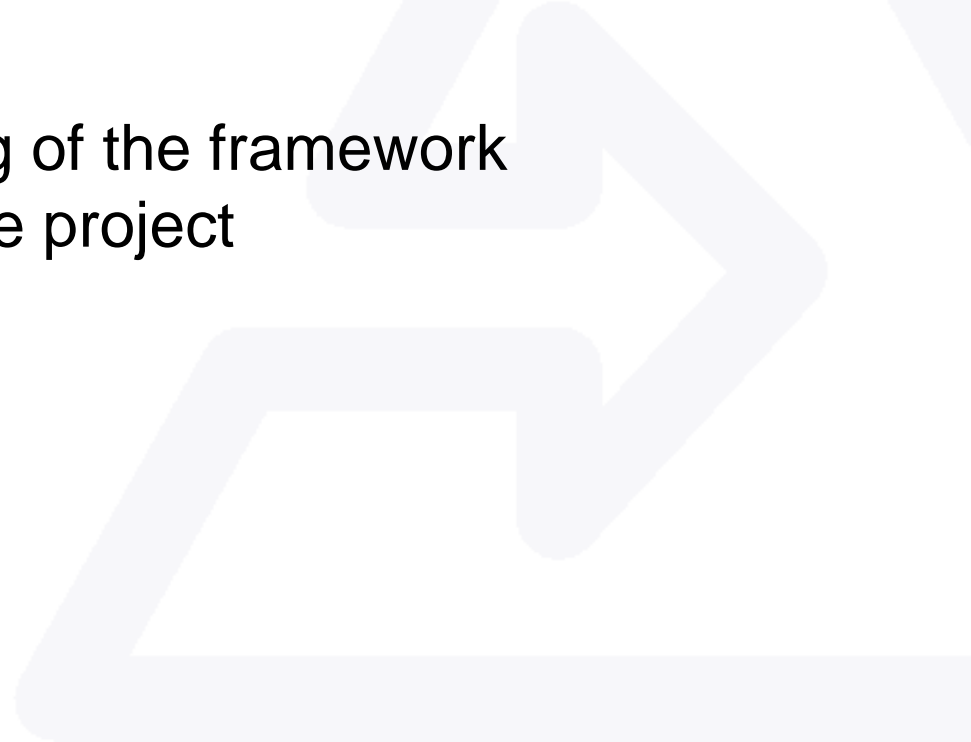
Network integration

Tobias Kötter
University of Konstanz



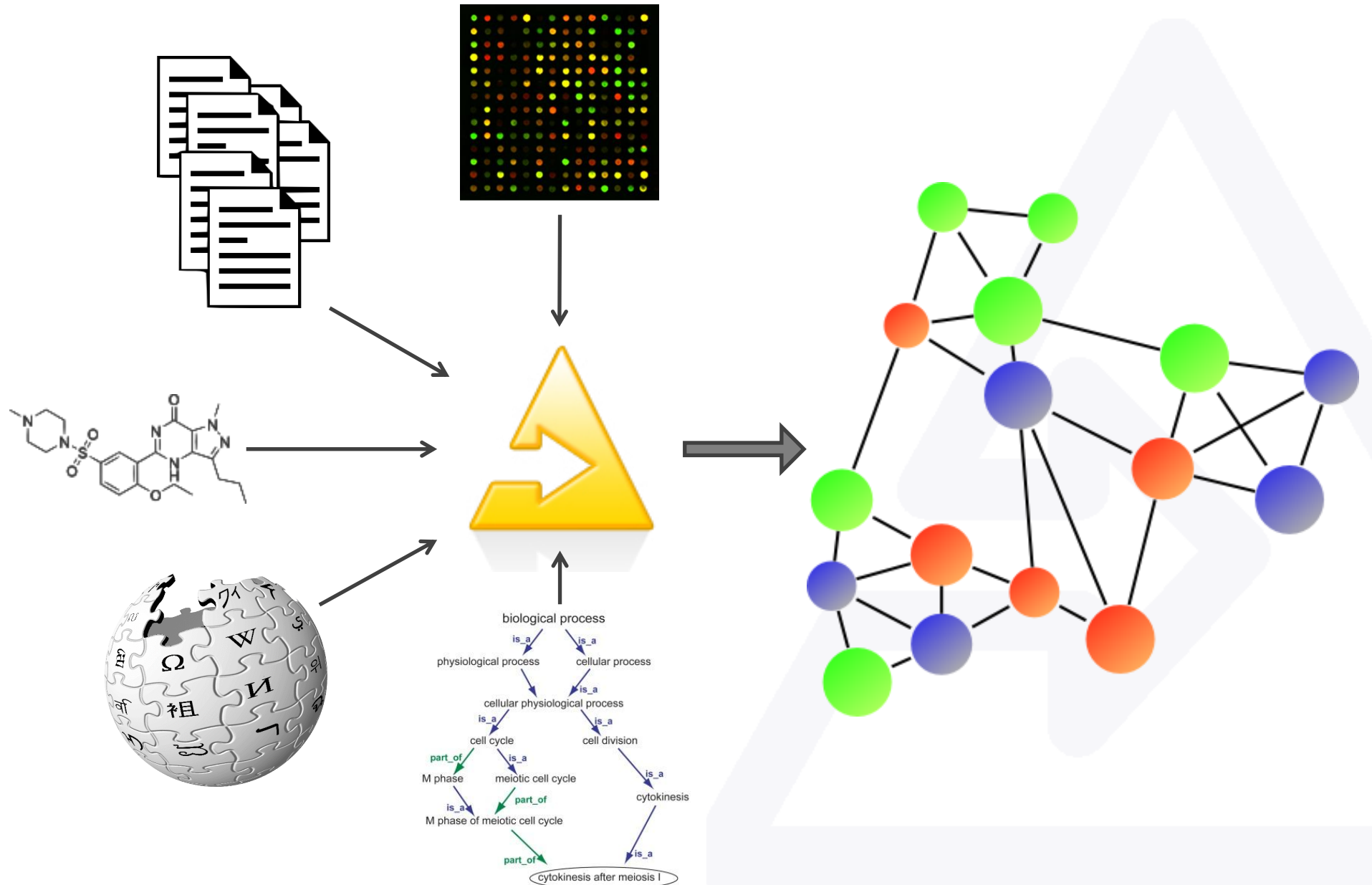
Background

- EU Project: Bison www.bisonet.eu
- June 2008 – June 2011
- Discovery of domain crossing pattern in heterogeneous data sources
- Network feature is an offspring of the framework developed and used during the project





Network based data integration





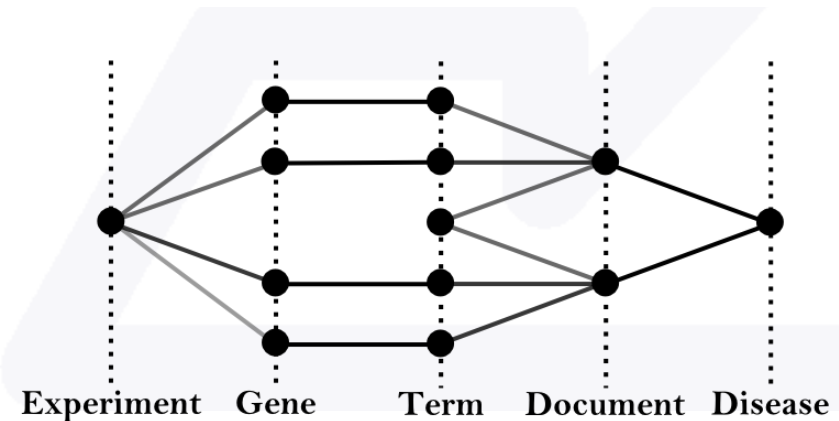
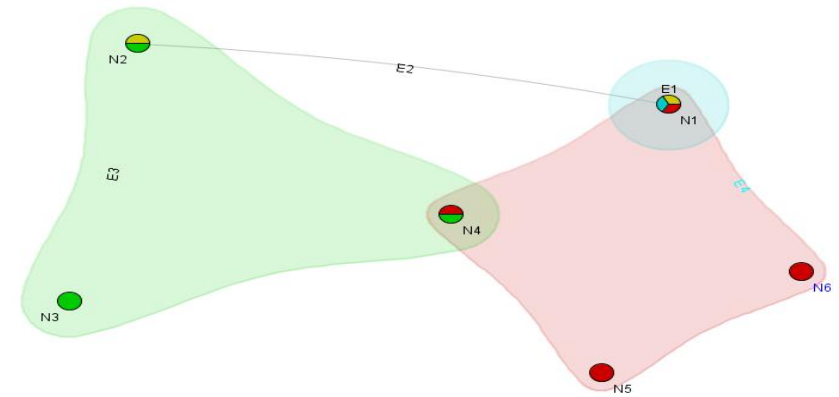
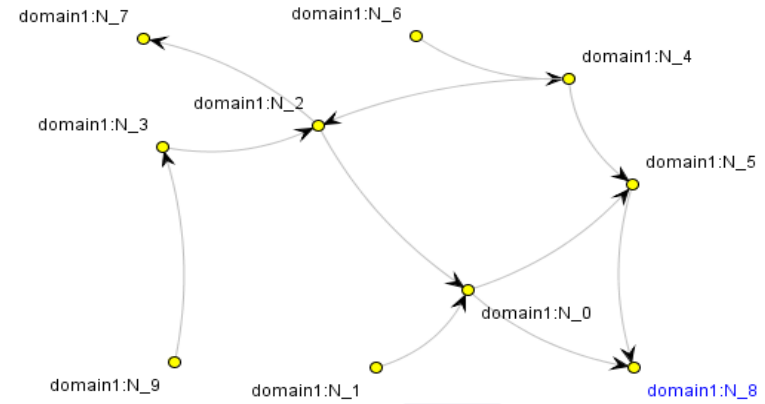
Data structure

- Supported networks:

- (un)directed
- (un)weighted
- hypergraph
- k-partite

- Features

- graph
- node
- edge
- end points
- types:
 - boolean
 - double/integer
 - string
 - uri
 - List
 - enum





Nodes

- Distance Matrix
- Network
 - IO
 - File
 - BeeF Reader
 - Network Reader
 - Network Writer
 - Assign Partition
 - Feature Inserter
 - Network Creator
 - Object Inserter
 - Convert
 - Matrix
 - Edge Adjacency Matrix
 - Incidence Matrix
 - Laplacian Matrix
 - Matrix to Network
 - Node Adjacency Matrix
 - Table
 - Edge Table
 - Feature Table
 - Node Neighbor Extractor
 - Node Table
 - Partition Table
 - Network Feature Extractor
 - Network To Row
 - Row To Network
 - SubGraph Extractor

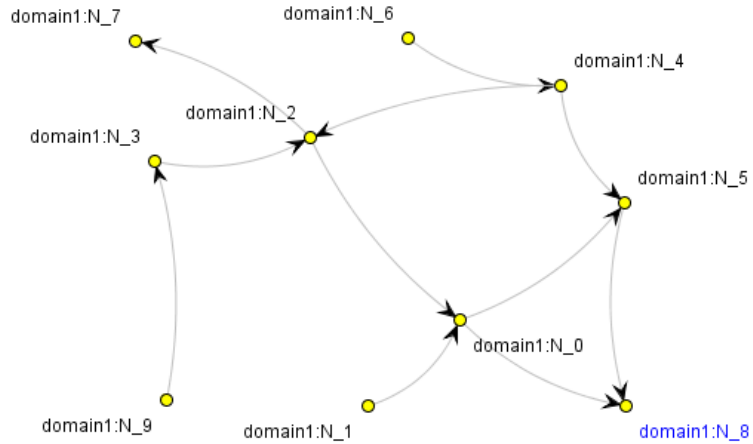
- Filter
 - Edge Degree Filter
 - Edge Weight Filter
 - Feature Filter
 - Feature List Filter
 - Feature Value Filter
 - Neighbor Filter
 - Node Degree Filter
 - Node Name Filter
 - Object ID Filter
 - Partition Filter
- Mining
 - Network Analyzer
 - Partition Graph Creator
- Visualization
 - Network Viewer
 - Viz Input Connector
 - Viz Output Connector
 - Network Generator



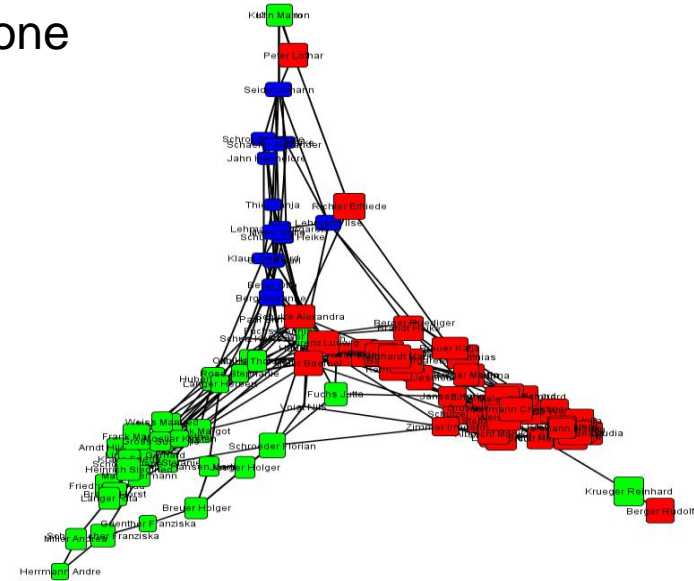
Visualization



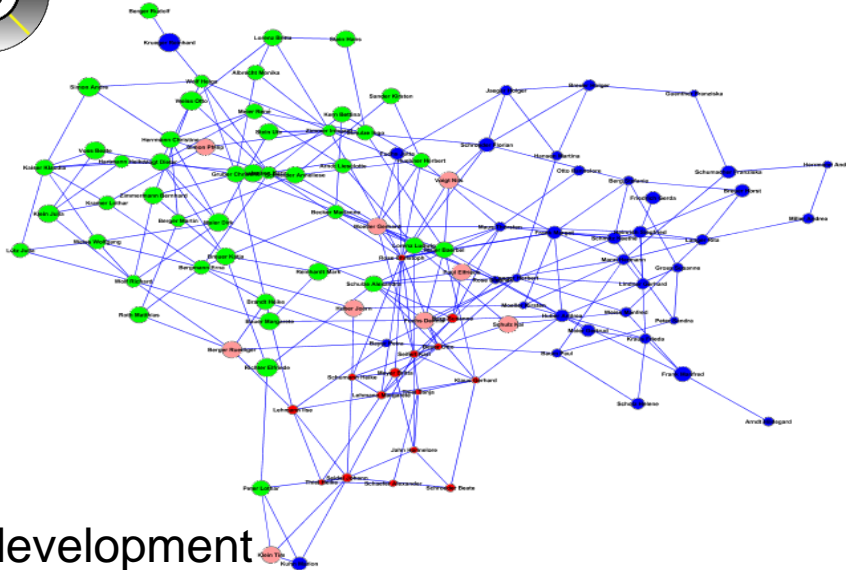
Internal based on Jung 2.0.1



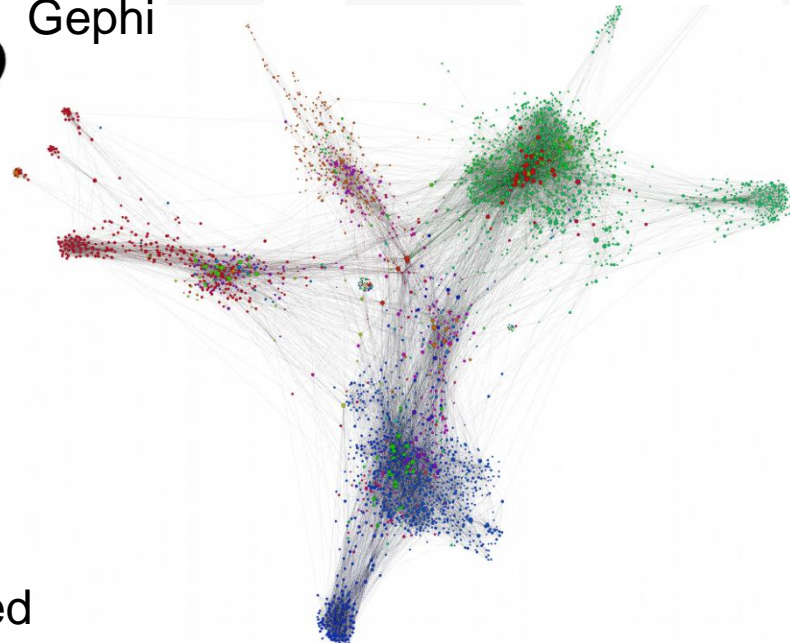
visone



Cytoscape



Gephi

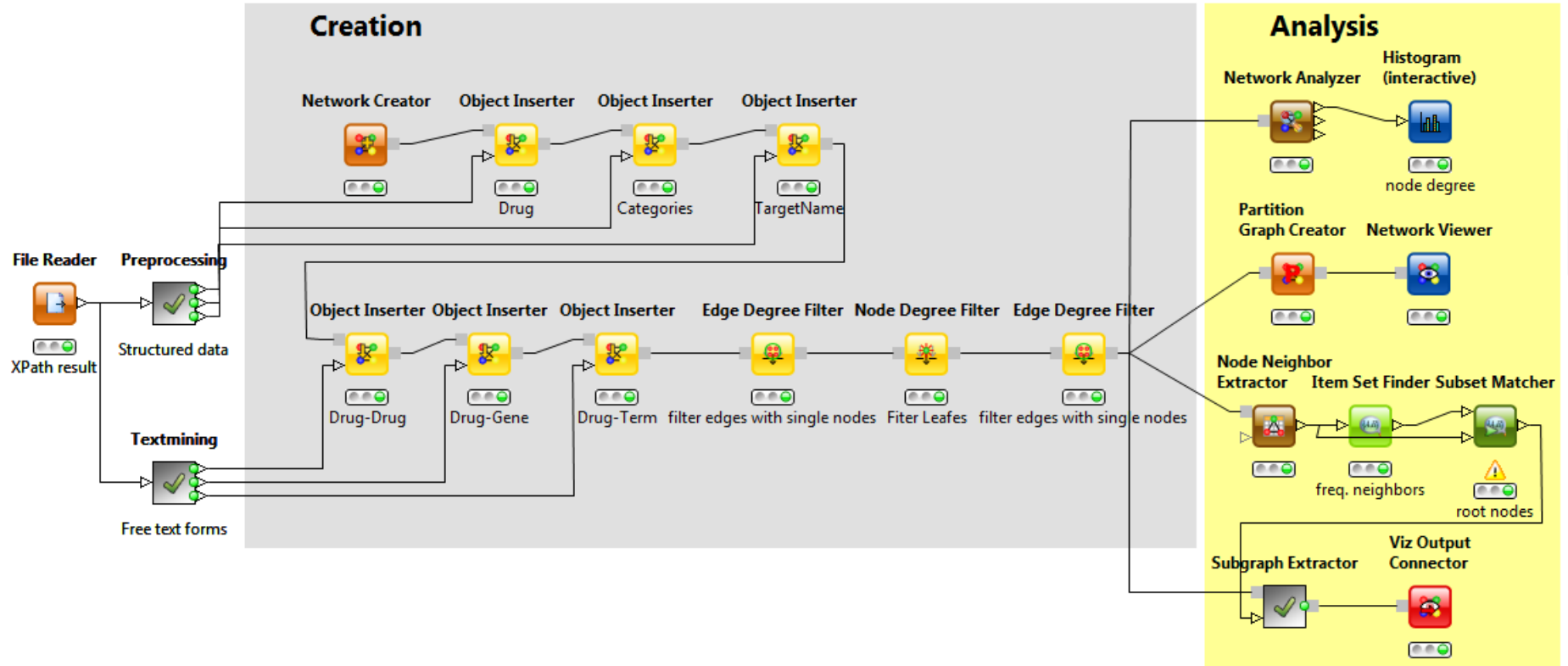


in development

planned



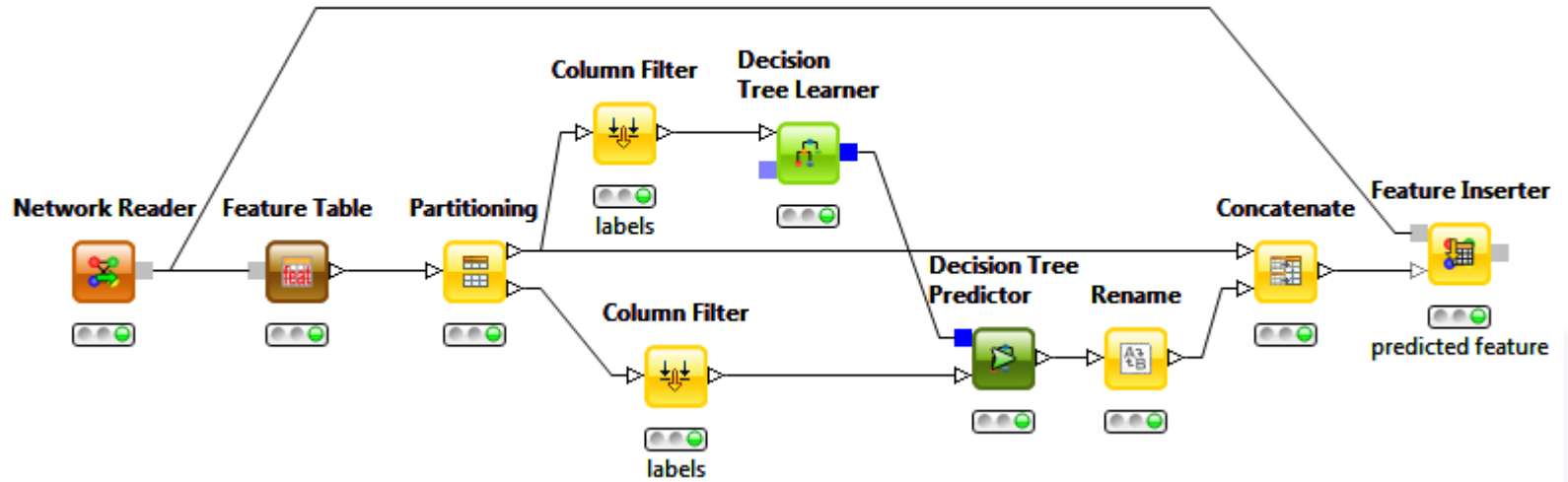
Network creation



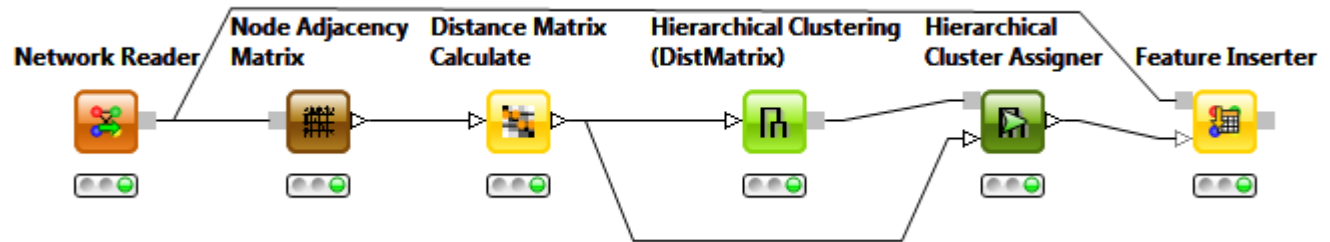


Network processing

Feature prediction



Clustering



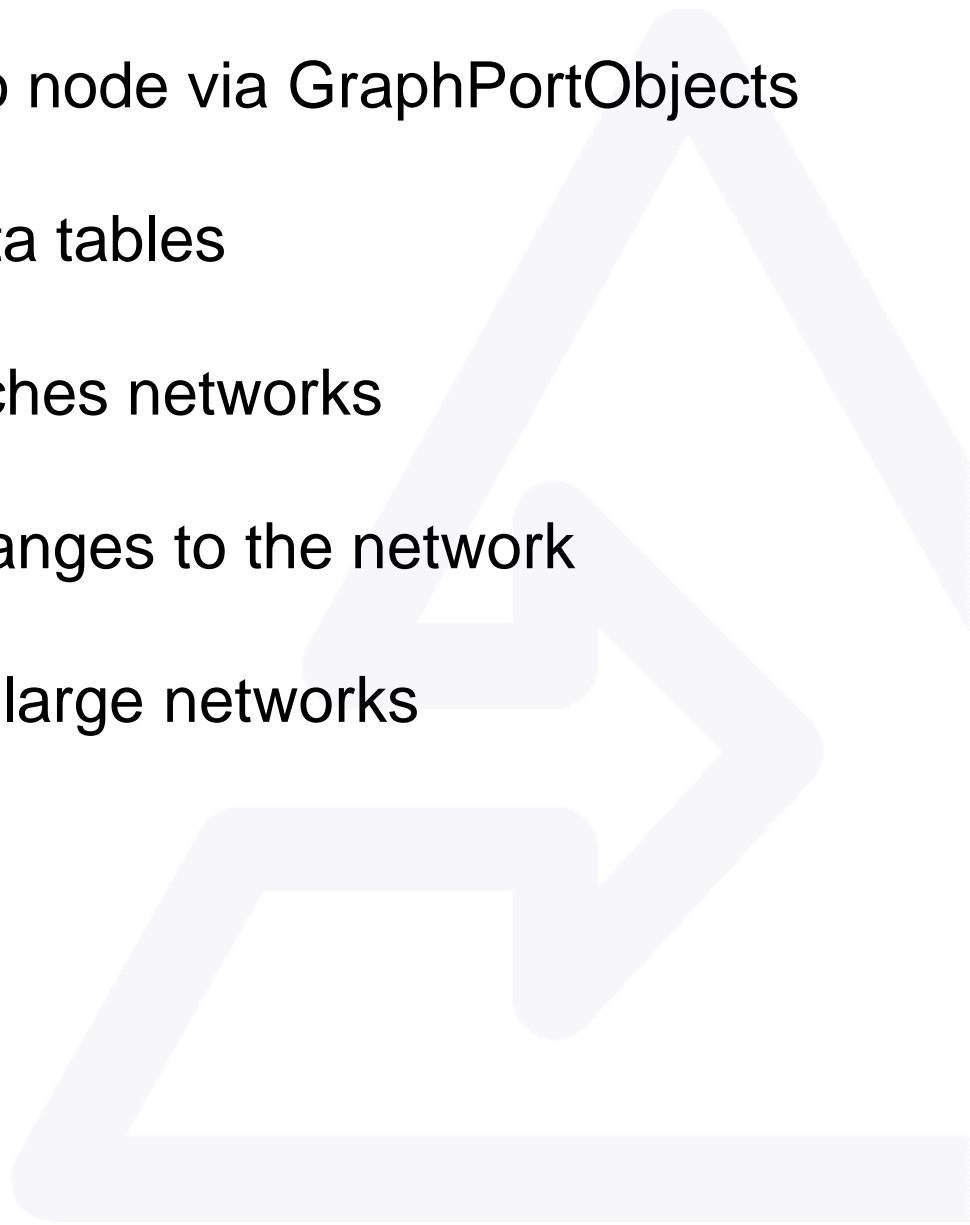
Batch processing





Technical Details

- Networks passed from node to node via GraphPortObjects
- GraphCells for networks in data tables
- Central network repository caches networks
- Modifying nodes save only changes to the network
- Optional database support for large networks





Outlook

- More nodes
 - network mining algorithm
 - network abstraction
 - Lucene integration for text search in features
 - parser GraphML, RDF, OBO, ...
 - ...
- Graph database as persistence layer (e.g. neo4j)
- Additional feature types such as byte for images
- Available via the KNIME Labs site with the next KNIME release