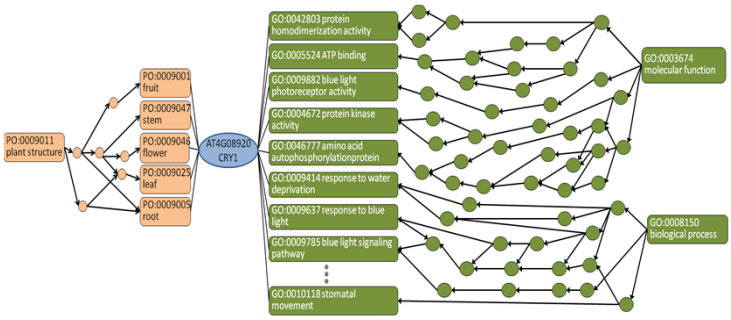
PAnG – Finding Patterns in Annotation Graphs

Philip Anderson, Andreas Thor, Joseph Benik, Louiqa Raschid, Maria Esther Vidal

Motivation

Abundance of data

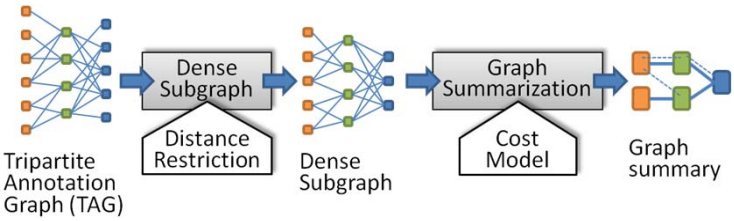
- High-throughput lab experiments in systems biology.
- Annotated datasets adorned with CV terms from ontologies.
- W3C Linking Open Data (LOD) initiative.



Goal: Explore and evaluate patterns in complex annotation graphs.

- Help scientists explore large annotation graphs.
- Generate hypothesis, e.g., interactions between groups of genes or new functional annotations.

Approach



Dense Subgraph

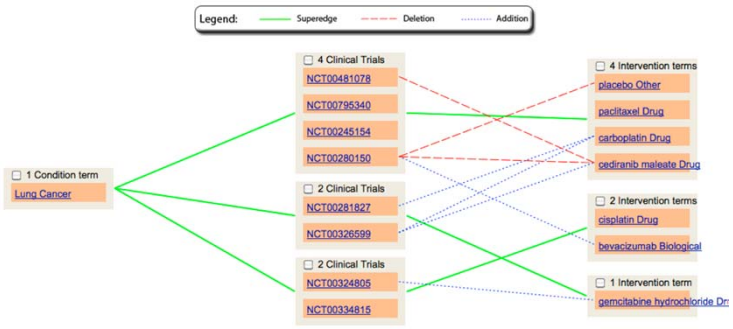
- Density as a measure of relatedness, similarity between genes.
- Identifies highly annotated candidate regions of a graph.
- Distance threshold based on path lengths between terms in the ontology.

Graph Summarization

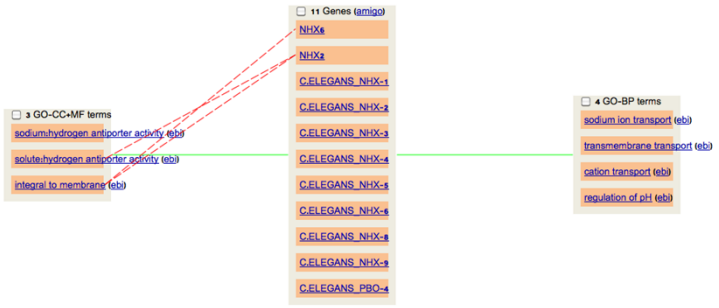
- Graph of supernodes, superedges, corrections to represent original graph.
- Intuitive way to extract and visualize graph patterns.
- Computed using cost model that gives weights to supernodes, superedges, and corrections.

Example Patterns

Clusters of clinical trials adorned with a condition "lung cancer" and corresponding treatments.

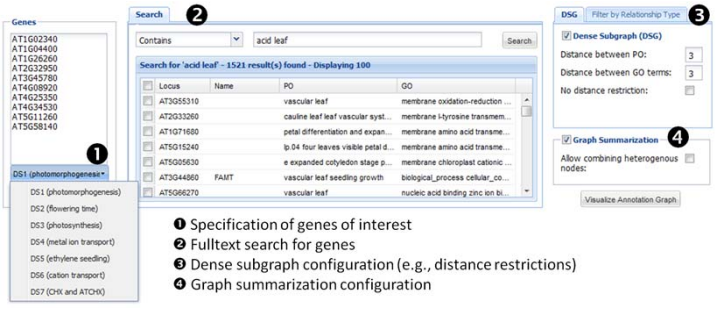


Cross genome GO annotations for cation/proton transporter genes in Arabidopsis thaliana and C. elegans.



Prototype

http://pang.umiacs.umd.edu/linkedct.html
http://pattaran.umiacs.umd.edu



❶ Specification of genes of interest
❷ Fulltext search for genes
❸ Dense subgraph configuration (e.g., distance restrictions)
❹ Graph summarization configuration

Future work

Manjal – Text Mining for MEDLINE

Annotation Visualizer – Visualize and explore annotations and patterns

Probabilistic Soft Logic

Patterns in ANnotation Graphs

PSL: Annotation computation by knowledge propagation

PAnG: Pattern identification using dense subgraphs and graph summaries

ANAPSID

Integrated access for heterogeneous data sources: adaptive query processing for SPARQL endpoints

GO the Gene Ontology — **G**ene **O**ntology

The Arabidopsis Information Resource

CT — **C**linical **T**rials