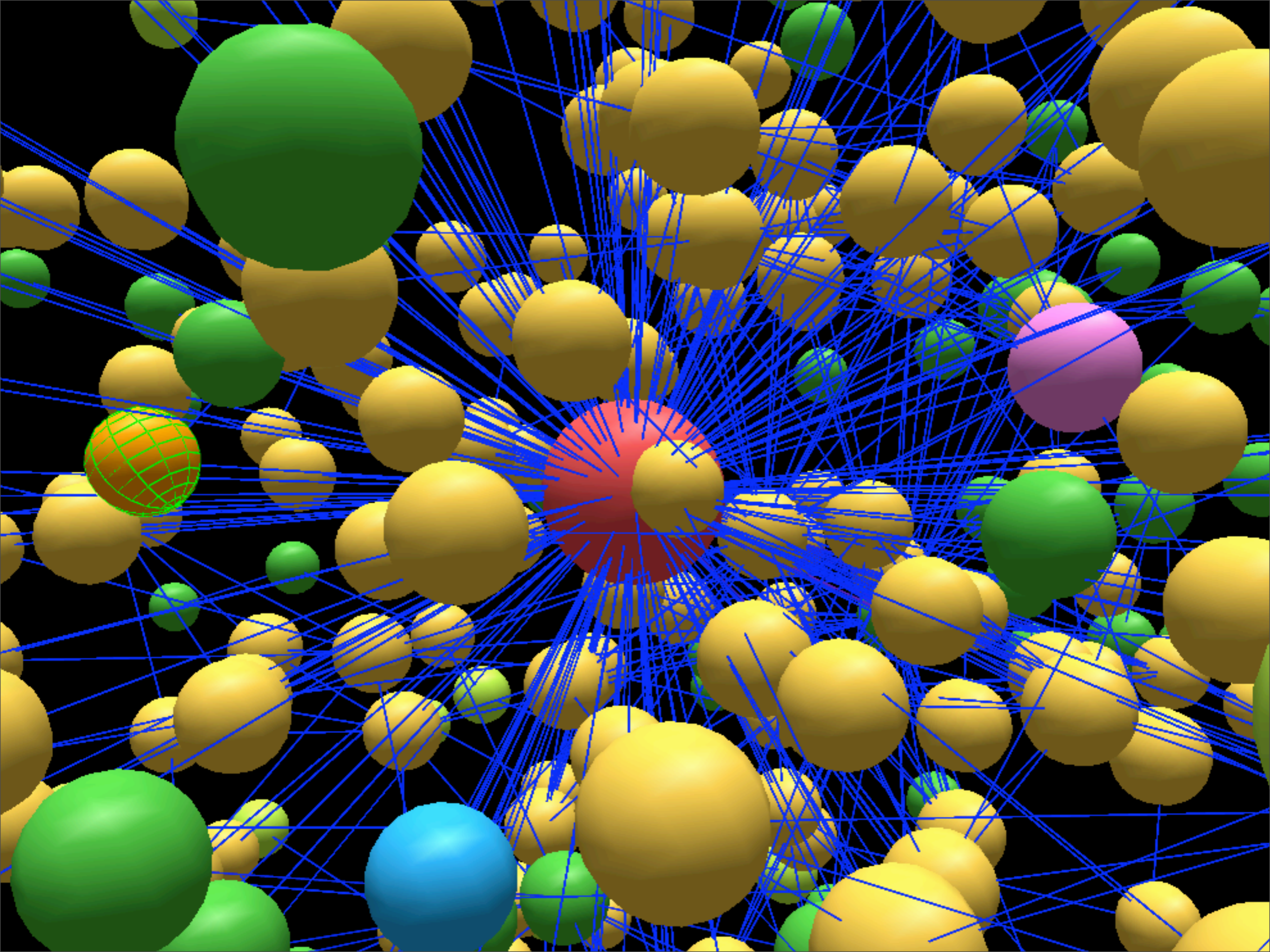


# Navigating gene networks by text mining and statistical analysis

Simon Vanden Bussche ··· Leon-Charles Tranchevent ··· Shi Yu  
<http://homes.esat.kuleuven.be/~vandenbs/>



# Text mining

# Text mining

searching for genes in the PubMed database

# Text mining

searching for genes in the PubMed database

- Java

# Text mining

searching for genes in the PubMed database

- Java
- Apache Lucene



# Text mining

searching for genes in the PubMed database

- Java
- Apache Lucene
- MySQL



# Text mining

searching for genes in the PubMed database

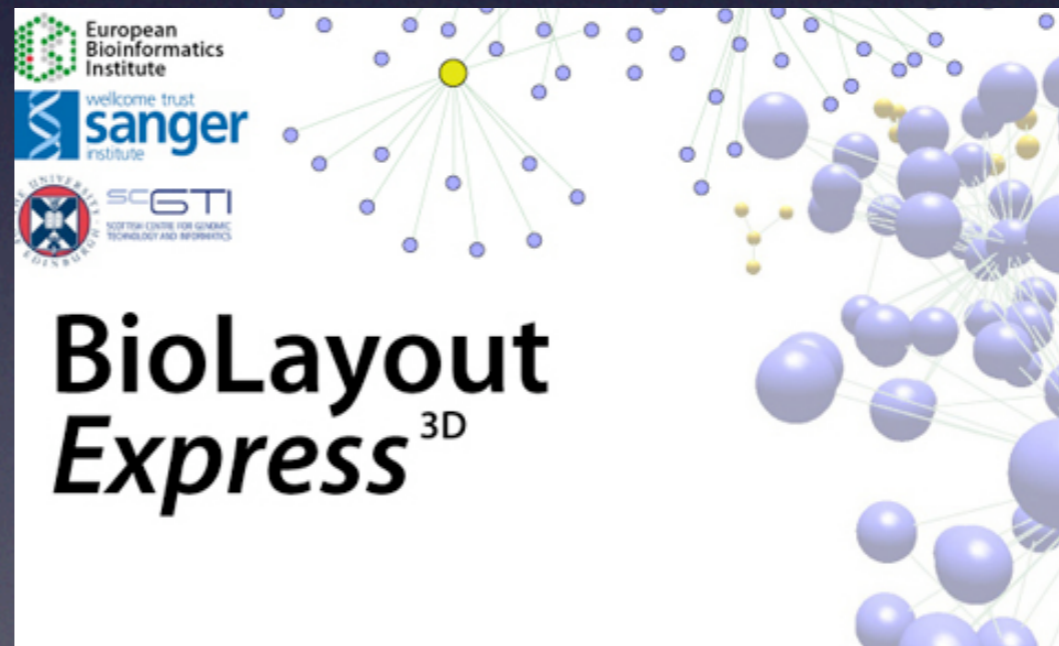
- Java
- Apache Lucene
- MySQL
- Matlab



# Text mining

searching for genes in the PubMed database

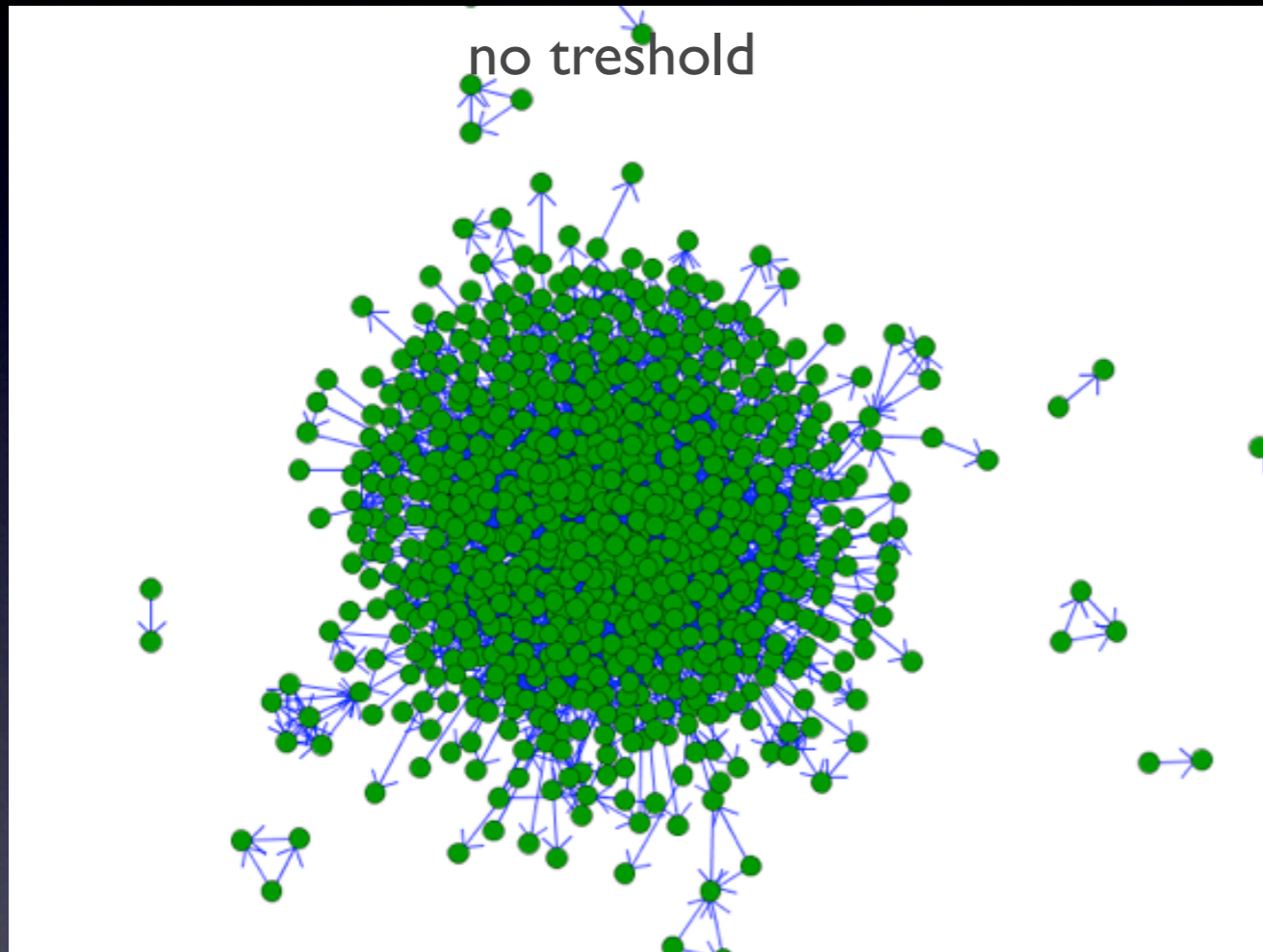
- Java
- Apache Lucene
- MySQL
- Matlab
- BioLayout



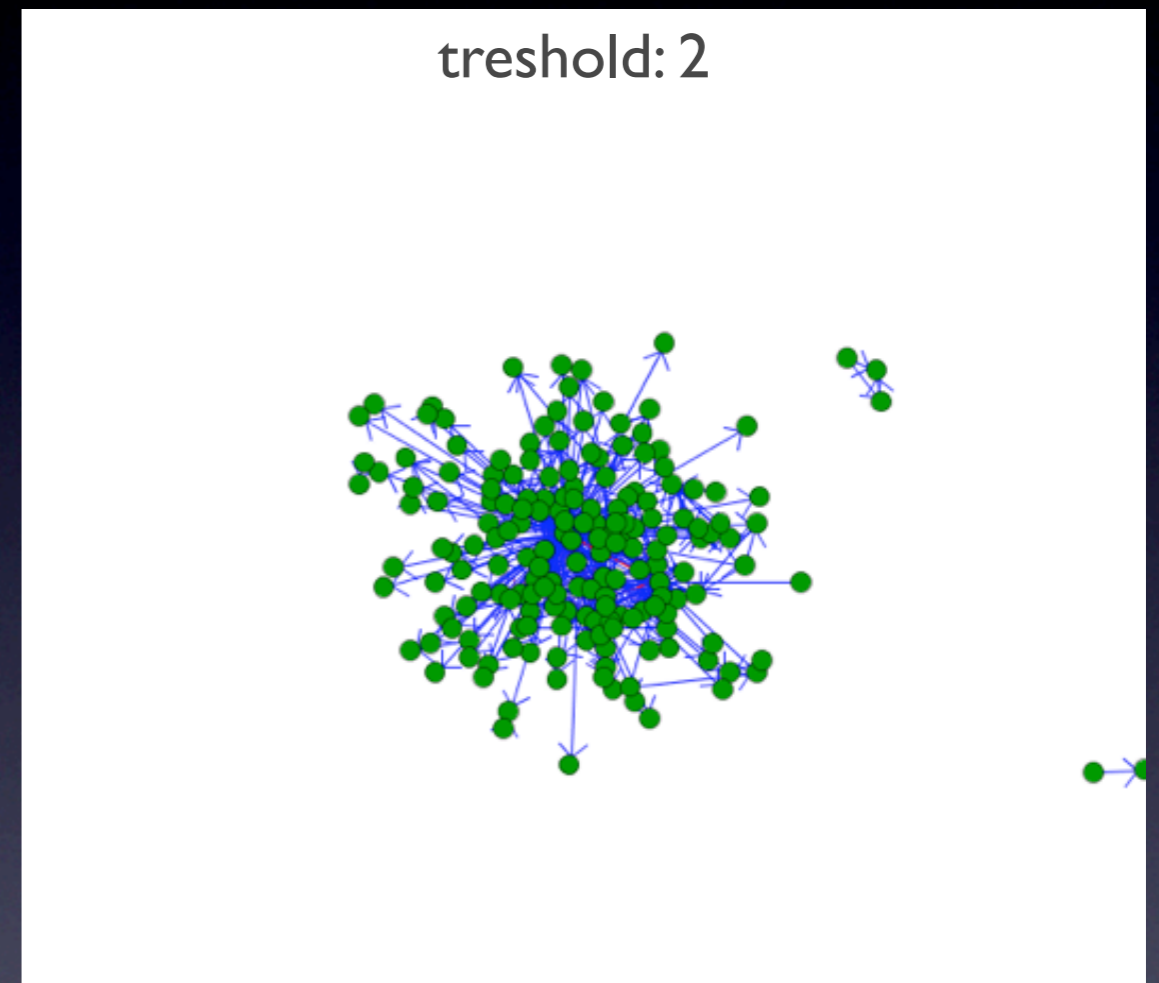
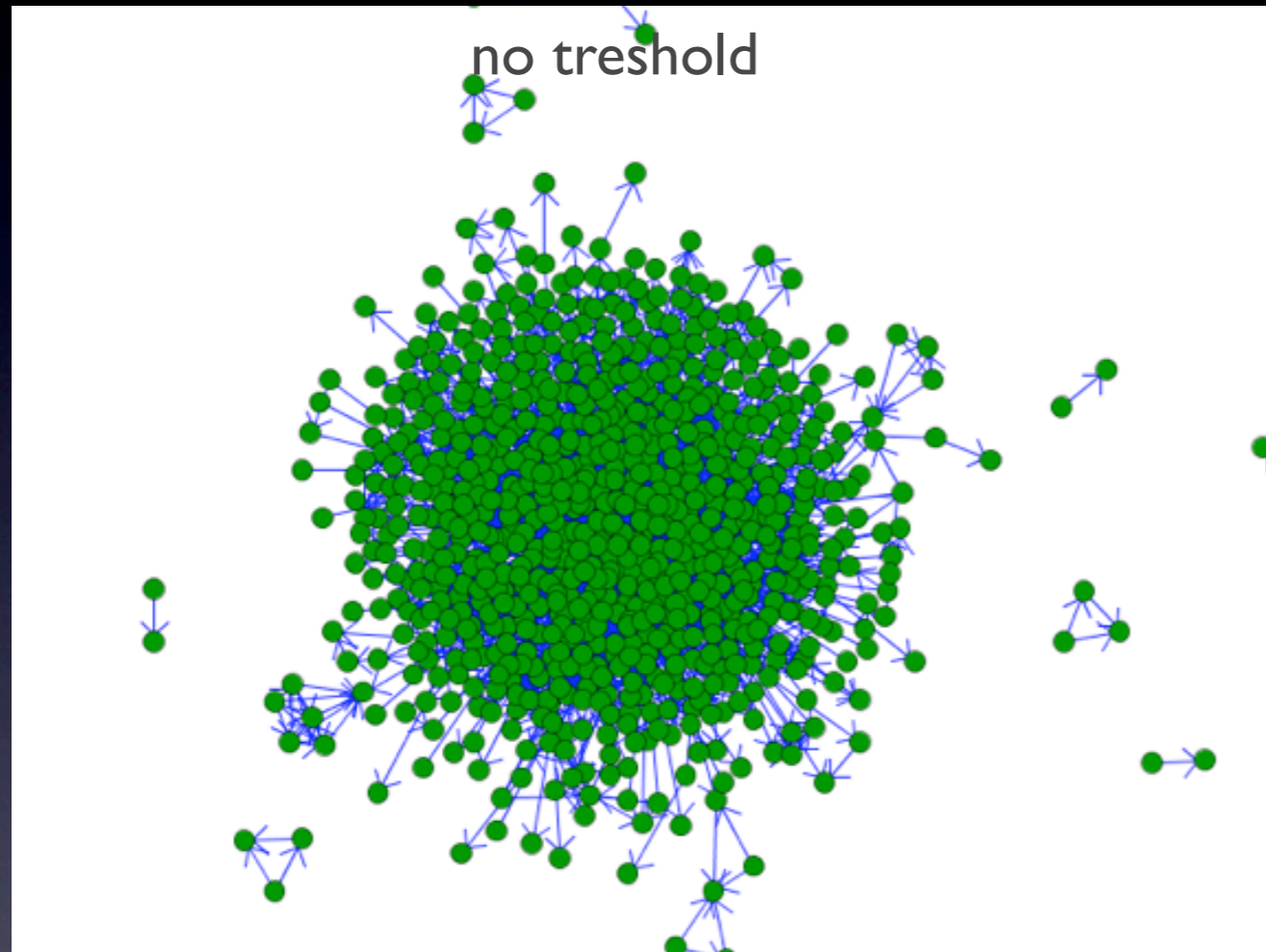
# Text mining



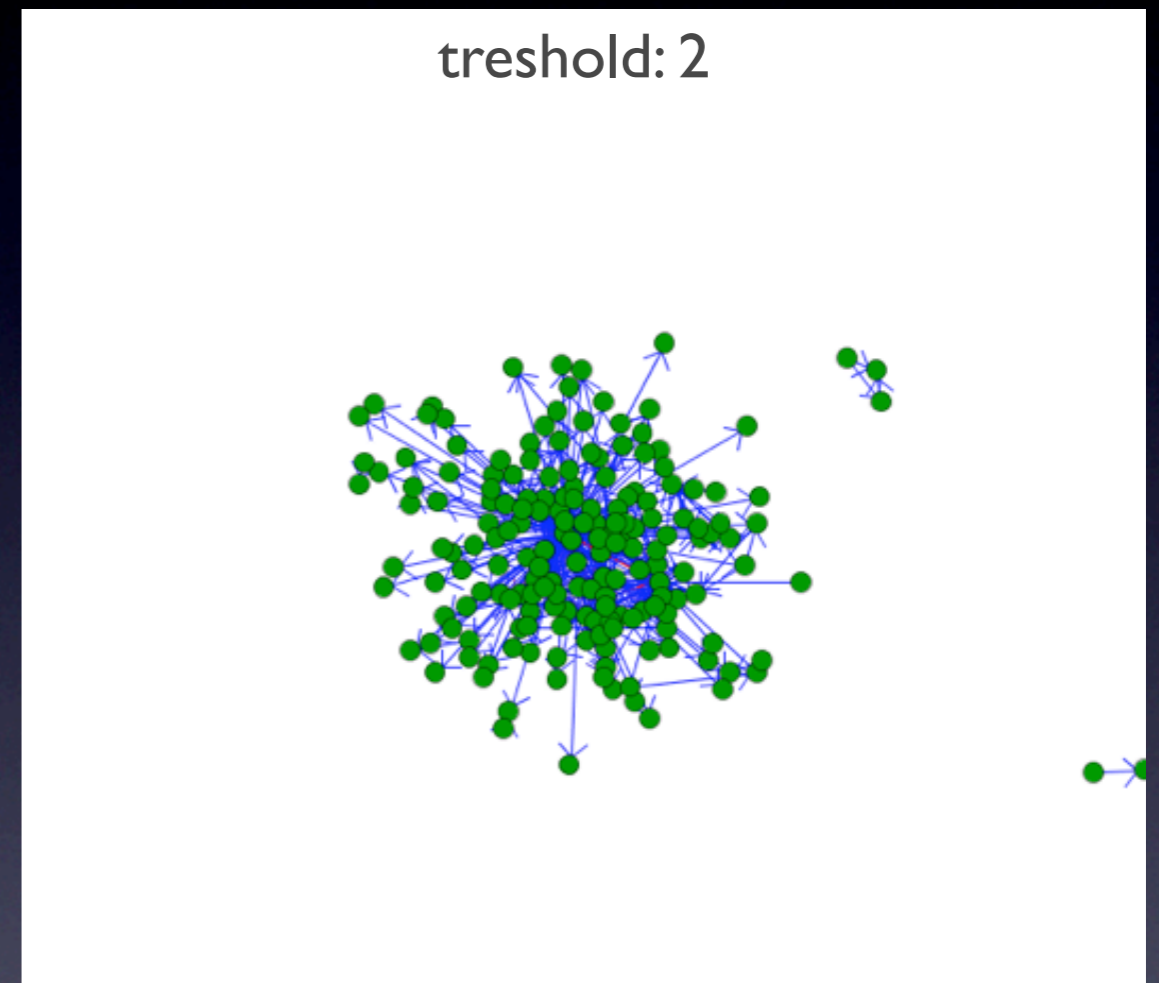
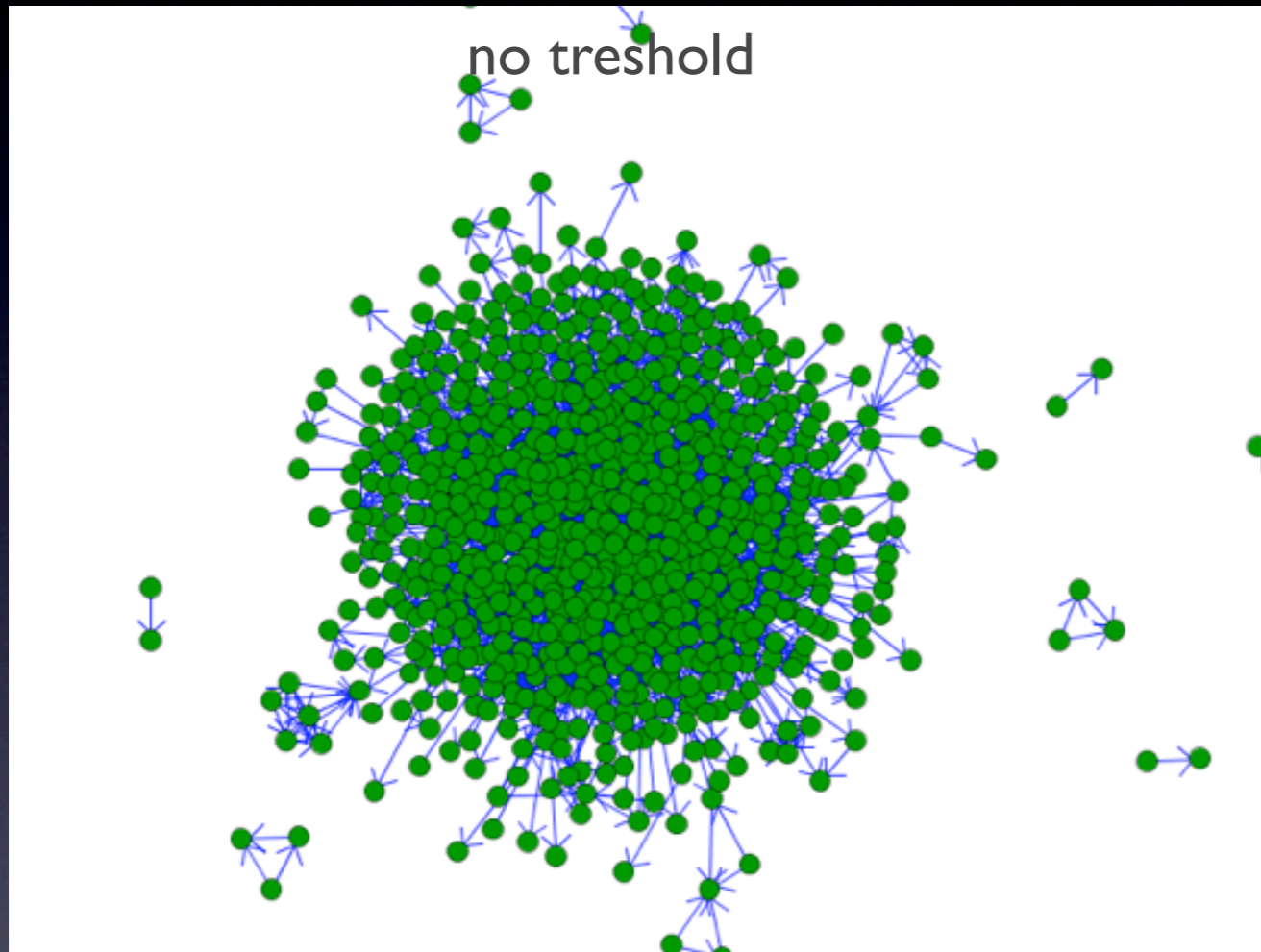
# Text mining



# Text mining



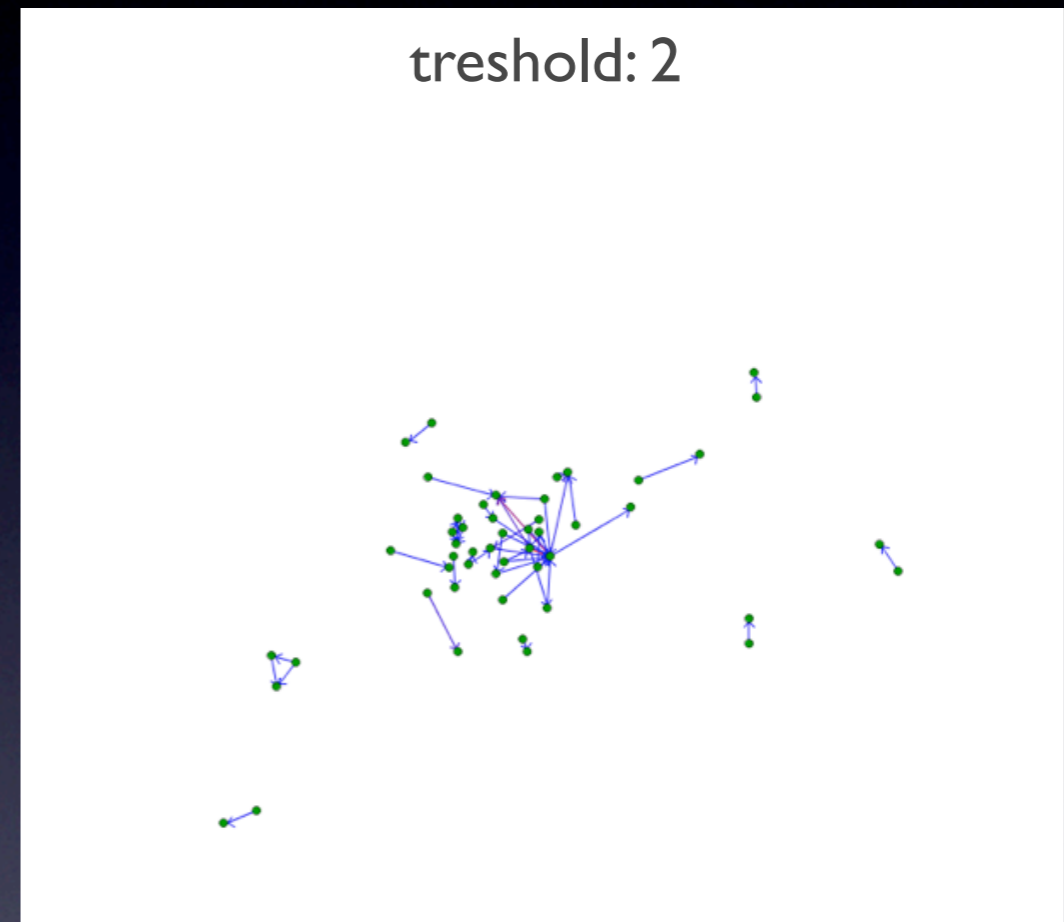
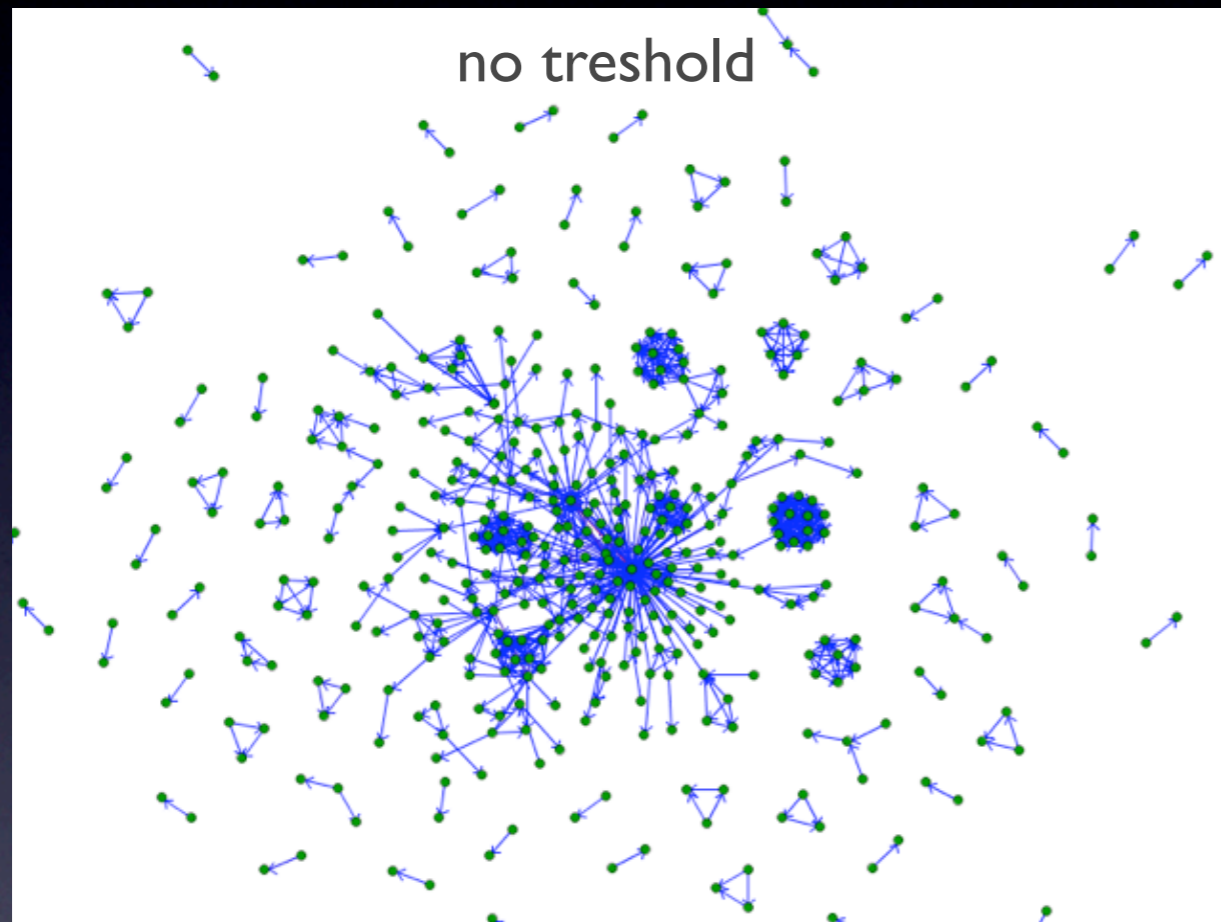
# Text mining



First version



# Text mining



Case sensitive version



# Statistical analysis

# Statistical analysis

known disease genes vs. regular genes

# Statistical analysis

known disease genes vs. regular genes

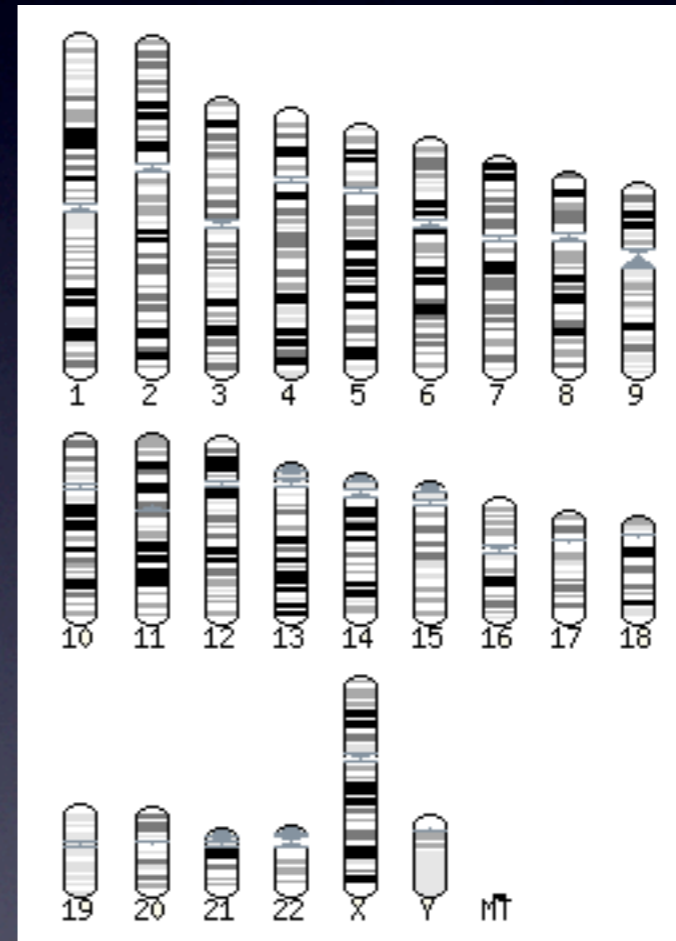
- ENSEMBL
- ASTD



# Statistical analysis

known disease genes vs. regular genes

- ENSEMBL
- ASTD
- OMIM
- GAD



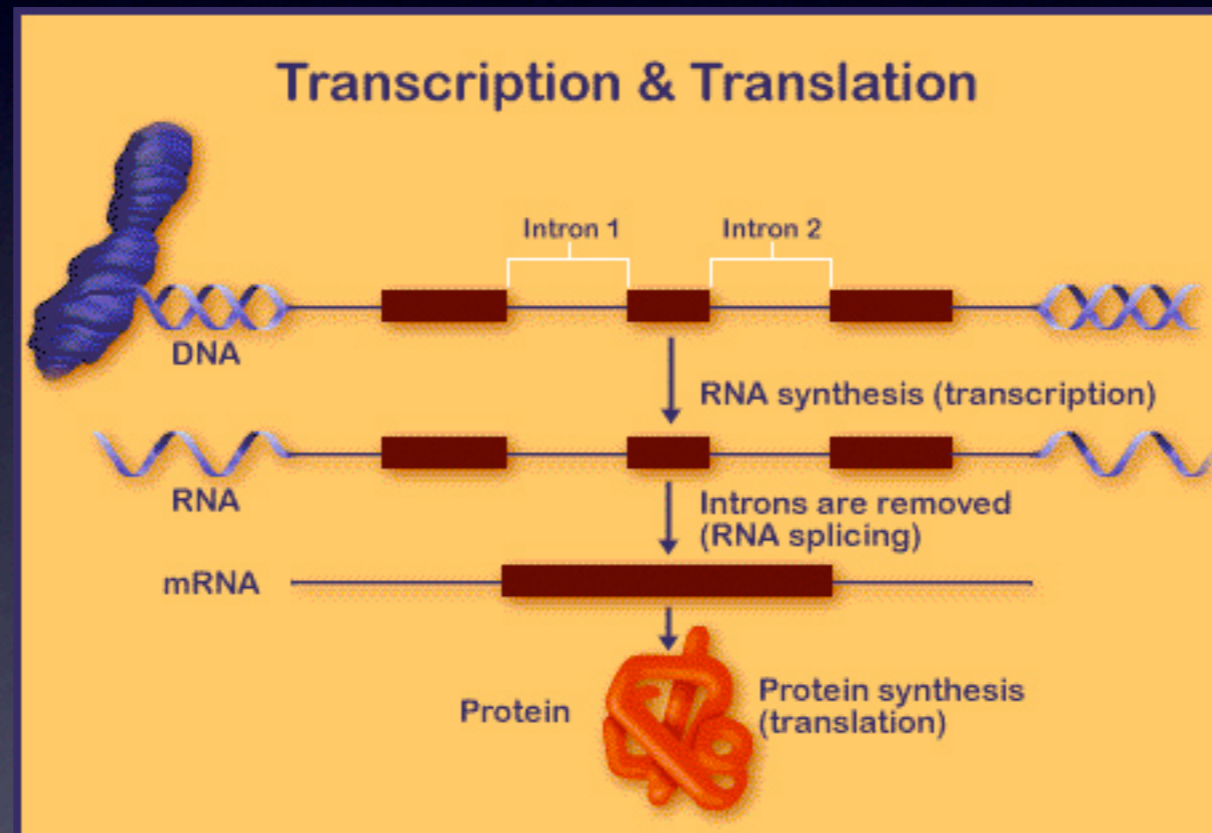
# Statistical analysis

# Statistical analysis

- What is compared?

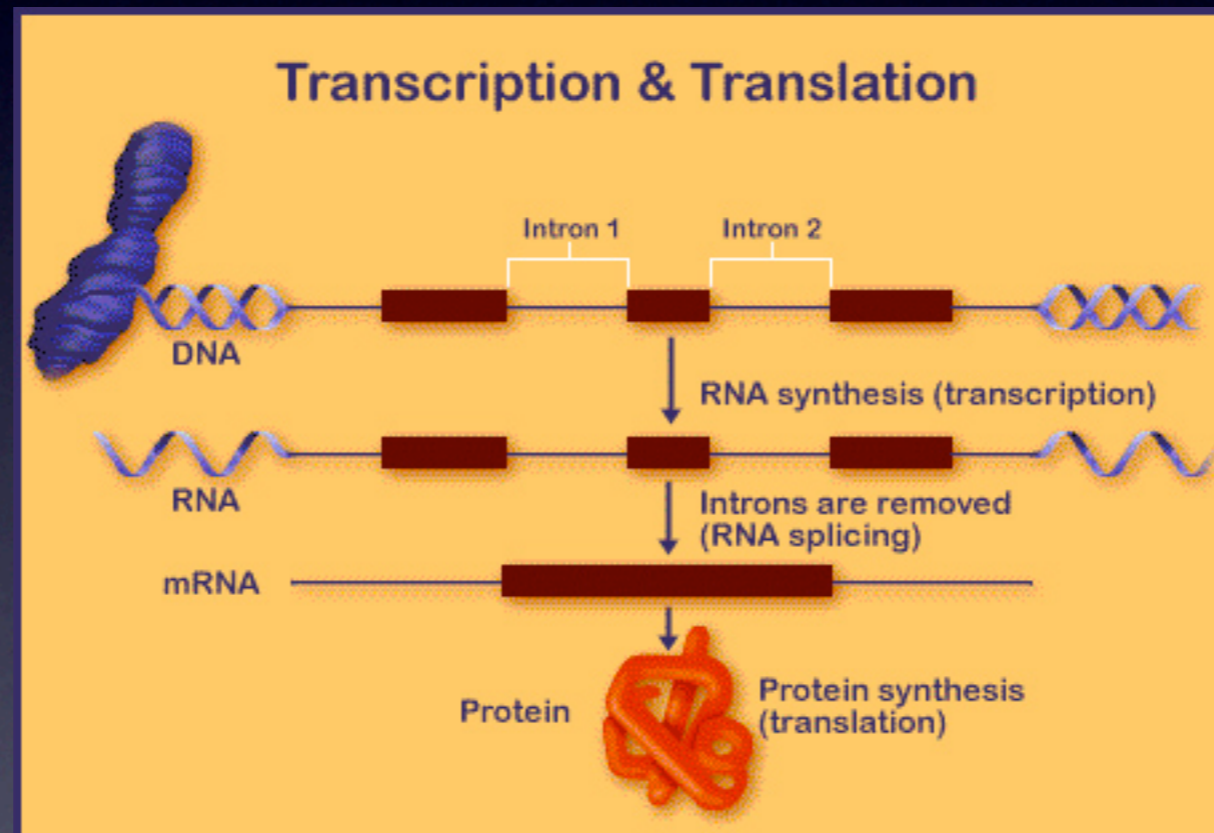
# Statistical analysis

- What is compared?



# Statistical analysis

- What is compared?

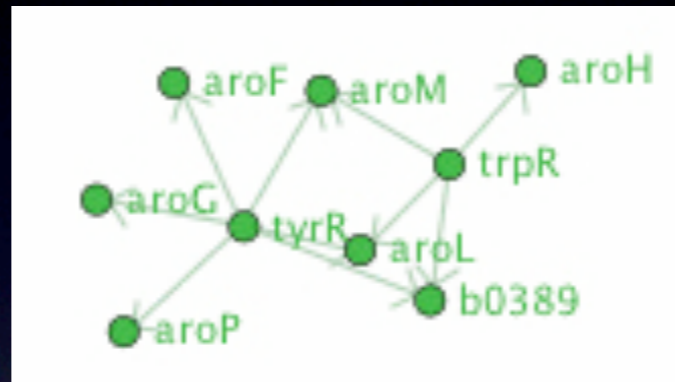


- Matlab programming

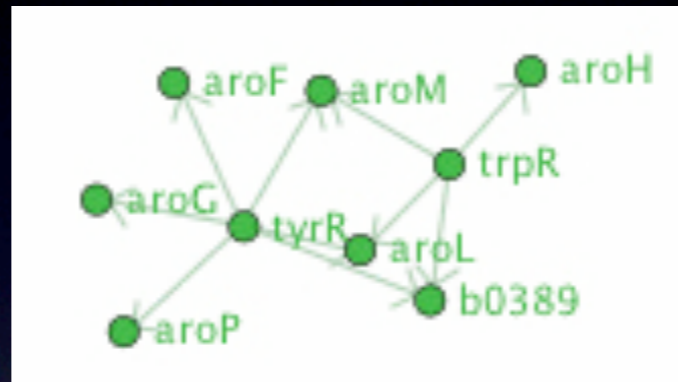


# Combined results

# Combined results

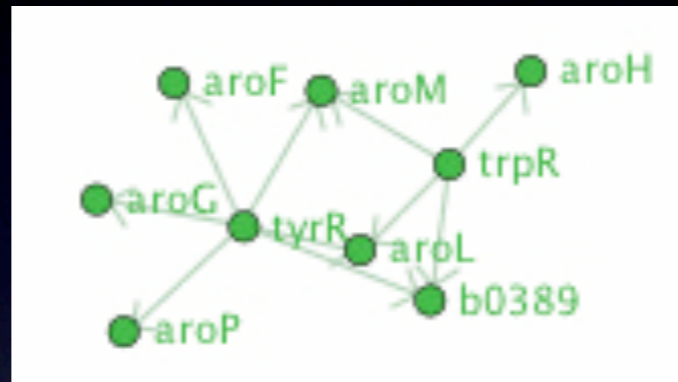


# Combined results



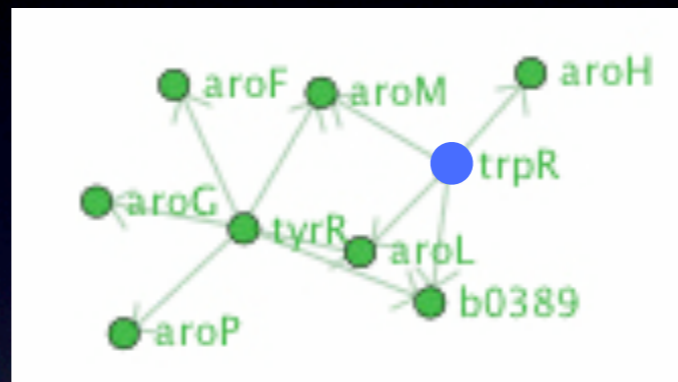
- Text mining result

# Combined results



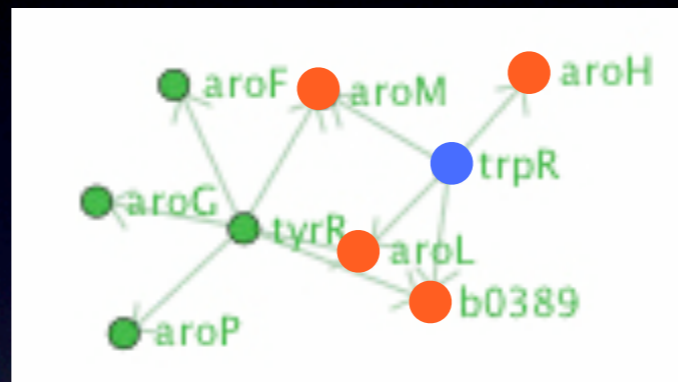
- Text mining result
- Different thresholds

# Combined results



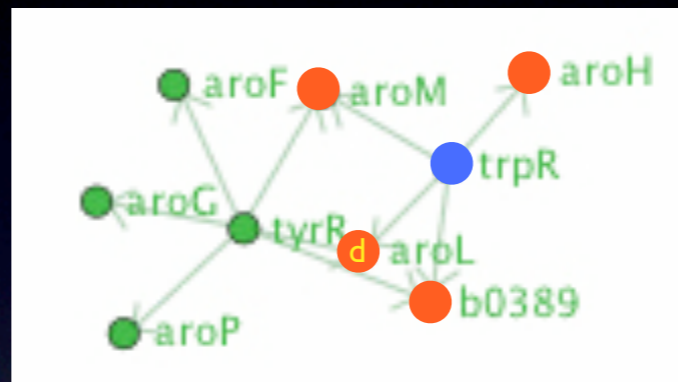
- Text mining result
- Different thresholds
- Number of first neighbours
- Number of disease neighbours

# Combined results



- Text mining result
- Different thresholds
- Number of first neighbours

# Combined results



- Text mining result
- Different thresholds
- Number of first neighbours
- Number of disease neighbours

# Combined results

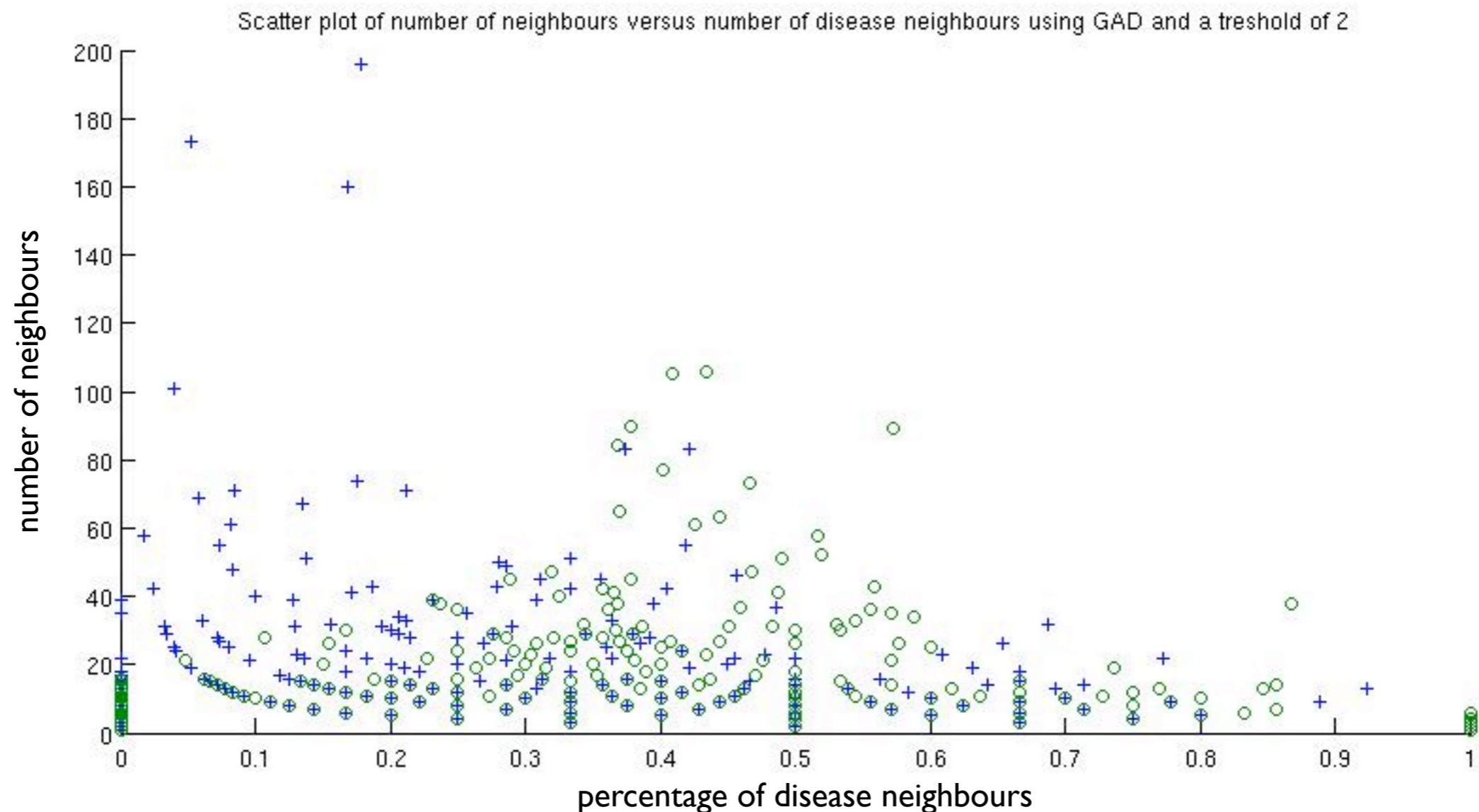
# Combined results

Scatter plots

# Combined results

## Scatter plots

number of neighbours versus percentage of disease neighbours



# Combined results

# Combined results

General comparison

Homology

# Combined results

General comparison

Homology

- extra boxplot for neighbours of disease genes

# Combined results

General comparison

Homology

- extra boxplot for neighbours of disease genes
- doing the same comparisons



<http://homes.esat.kuleuven.be/~vandenbs/>

<http://homes.esat.kuleuven.be/~vandenbs/>



Questions?