

富嶽三十六景 神奈川沖
浪裏

江戶 葛飾 柴田 画

Big_Ocean

Presentation_report_8

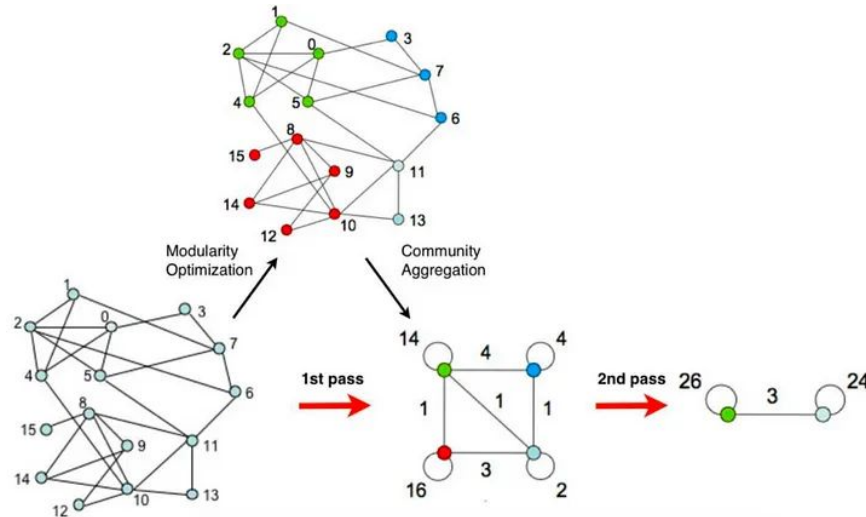
Zombie Apocalypse graph



Source: <https://github.com/neo4j-examples/graphgists/blob/master/medical/zombie.adoc>

Louvain - description

- Hierarchical clustering algorithm designed to detect communities
- Maximizing modularity score (maximize edges within community and minimize edges between community)



Louvain- Results



Leiden - Description

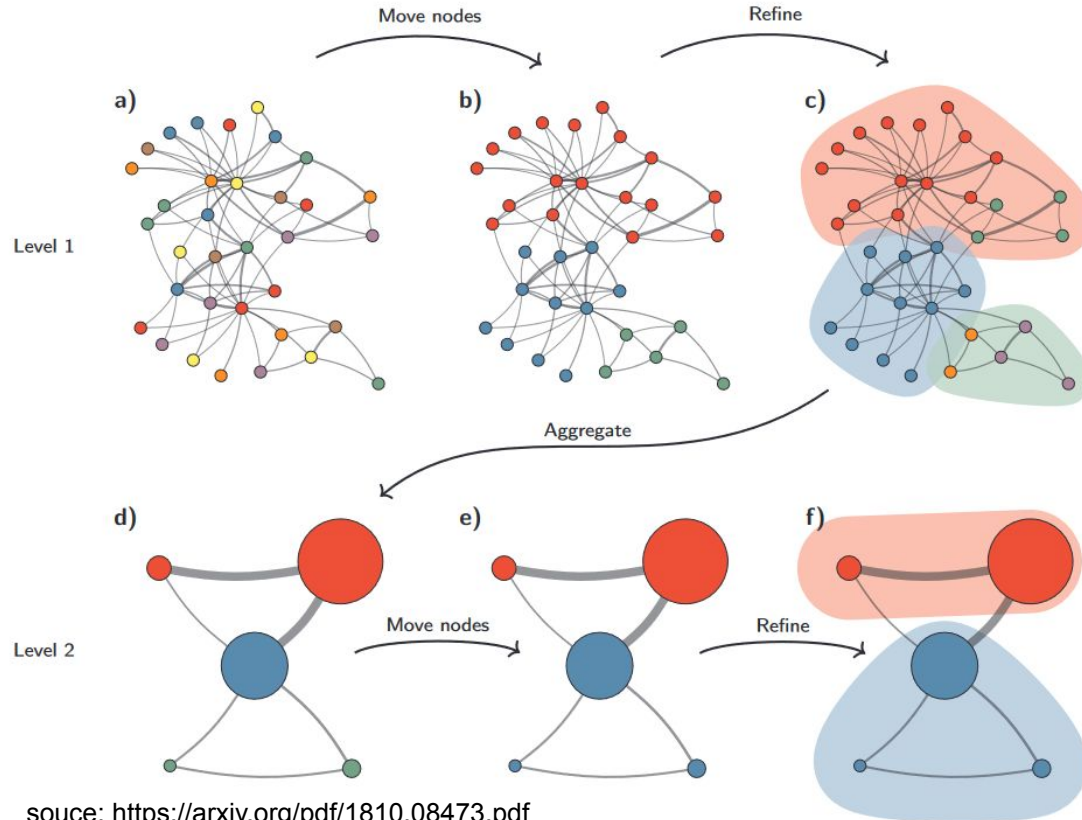
Problem with Louvain:

- Louvain algorithm has major defect -> unnoticed until 2019
- experimental analysis: 25% of the communities are badly connected & 6% are disconnected

Leiden:

- hierarchical clustering algorithm
- randomly breaks down communities into smaller well-connected ones.
- faster than the Louvain algorithm.
- communities are more locally optimally assigned.

Leiden - Description



source: <https://arxiv.org/pdf/1810.08473.pdf>

Leiden-Results



Results- Louvain - Leiden:

Louvain:

- 5 communities detected.
- some big communities

Leiden:

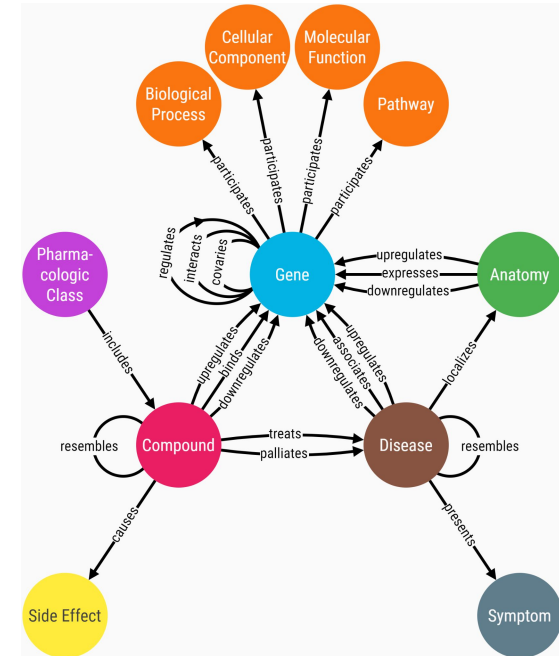
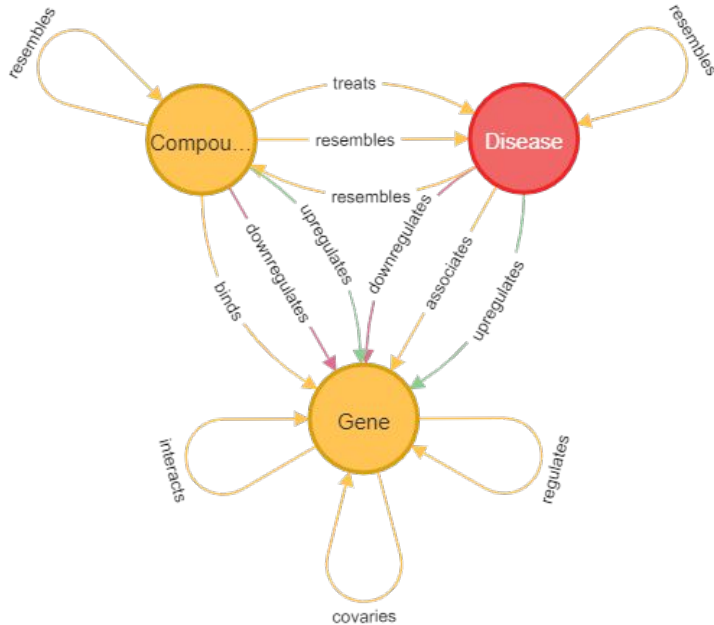
- 6 communities detected
- split in smaller communities

> Of all detected communities by both algorithm , only one same community(blue).

> It could be a scenario that if zombie outbreak really happens in the blue community, then the new york can be sealed in order to isolate other parts of US.

Hetionet dataset - explanation

- Network with multiple node and edge (relationship) types — which encodes biology.
- Version 1.0 contains 47,031 nodes of 11 types and 2,250,197 relationships of 24 types.



Query on Hetionet Subset

A

```
MATCH ()-[rel]→()  
RETURN  
| type(rel) AS rel_type,  
| count(*) AS count  
ORDER BY count DESC
```

B

```
MATCH (node:Disease)  
| RETURN node  
| LIMIT 10
```

C

```
MATCH path =  
(:Disease {name: 'pancreatic cancer'})-[:upregulates]→()  
RETURN path LIMIT 10
```

D

```
MATCH (gene:Gene {name: "PAK2"})-[:covaries]→(covariate) RETURN covariate.name
```

Query on Hetionet Subset - explanation

A > analyze the number of relationships by counting the occurrences of each relationship type

B > retrieve the first 10 disease nodes from the graph

C> retrieve the first 10 upregulated genes for the "pancreatic cancer" disease.

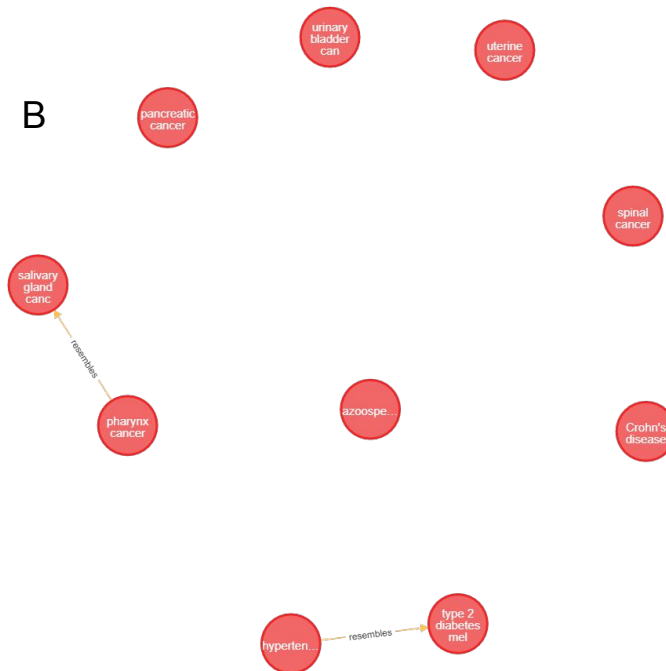
D>retrieve the names of the covariates that are associated with the "PAK2" gene that is also an upregulated gene for "pancreatic cancer" .

Query on Hetionet Subset - Some result

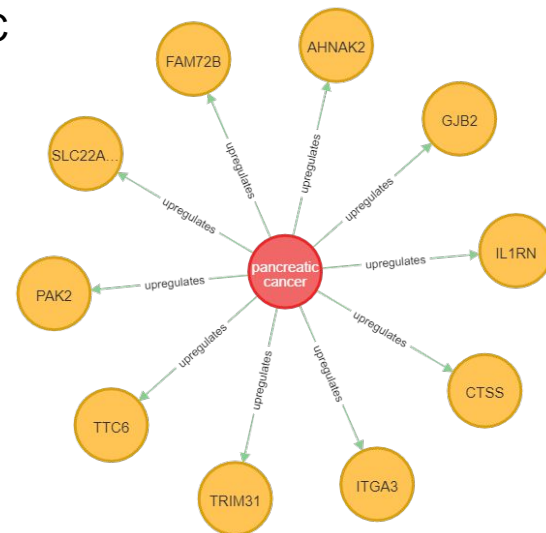
A

rel_type	count
"regulates"	265672
"interacts"	147164
"covaries"	61690
"downregulates"	28725
"upregulates"	26487
"associates"	12623
"binds"	11571
"resembles"	7029
"treats"	755

B



C



Query on Hetionet Subset - Some result

D

covariate.name
"PTBP1"
"ARF6"
"EFNA3"
"CBLN3"
"DRG1"
"SERP1"
"AVP"
"YWHAQ"