COMPUTATIONS ON COLLABORATION SPOTTING DATASETS

RICHARD FORSTER

AGENDA

- Graph generation
- Community Detection
- ForceAtlas
- Performance results
- Future work

GRAPH GENERATION

GRAPH GENERATION

- This is a process required for every single graph by any user
- First, database returned data needs to be transformed
- Have to generate:
  - Collaborations
  - Nodes
  - Edges

GRAPH GENERATION

	Silicon	Database	3D	CT
Collaborations	33,45	13,85	18,64	12,65
Nodes	0,69	0,65	2,48	0,38
Edges	1,48	1,87	0,71	0,69

Computation time for specific parts of the graph generation in seconds

COMMUNITY DETECTION

COMMUNITY DETECTION

- Used to reveal groups in real world data
- Louvain method
- Parallel heuristics

LOUVAIN METHOD

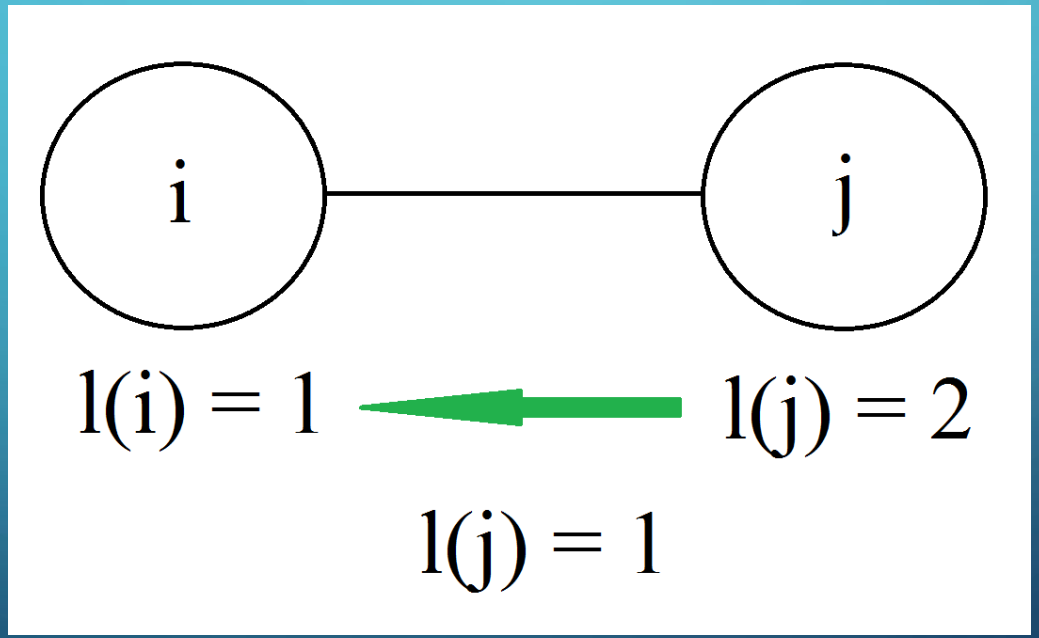
- Multi-phase, iterative, greedy algorithm
- Monotonically increasing modularity
- Inherently sequential

LOUVAIN PARALLEL HEURISTICS

- **Singlet minimum label heuristic**
- **Generalized minimum label heuristic**

SINGLET MINIMUM LABEL HEURISTIC

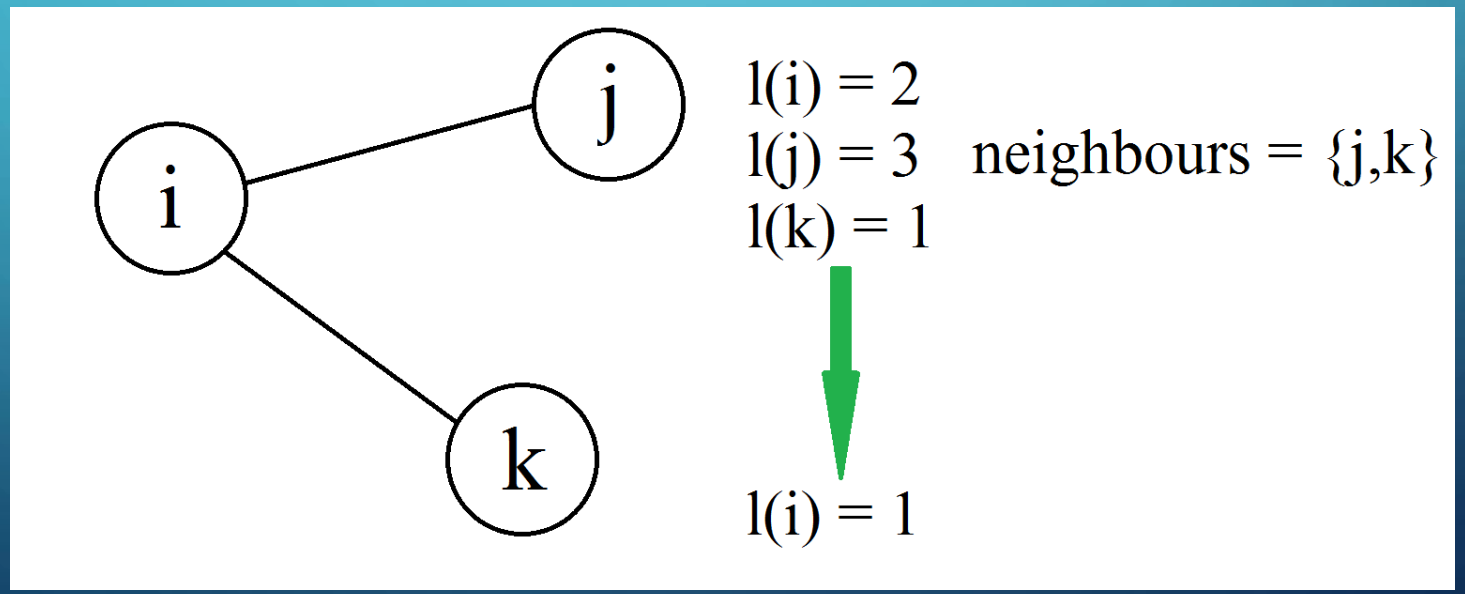
- Move node (j) only if: $l(C(i)) < l(C(j))$

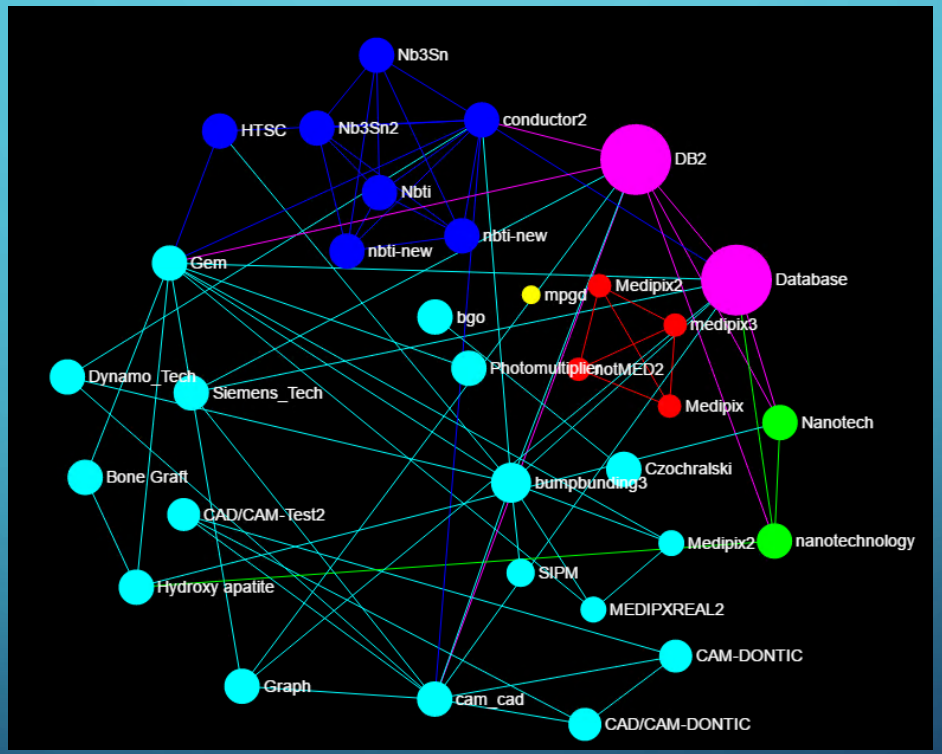


GENERALIZED MINIMUM LABEL HEURISTIC

- Move node (i) to neighbour community only if:

$$\min_{n \in neighbours} l\big(C(n)\big) < l(C(i))$$




Nb3Sn
HTSC
Nb3Sn2
conductor2
DB2
Nbti
nbti-new
nbti-new
Gem
Database
mpgd
Medipix2
bgo
medipix3
Photomultiplier
notMED2
Dynamo_Tech
Siemens_Tech
Medipix
Nanotech
Bone Graft
Czochralski
bumpbunding3
CAD/CAM-Test2
Medipix2
nanotechnology
SIPM
Hydroxy apatite
MEDIPXREAL2
CAM-DONTIC
Graph
cam_cad
CAD/CAM-DONTIC

FORCEATLAS

FORCEATLAS

- Force-directed layout based on n-body simulation
- Repulsion-attraction
- Makes visual interpretation easier
- Result depends on starting state

PARALLEL FORCEATLAS

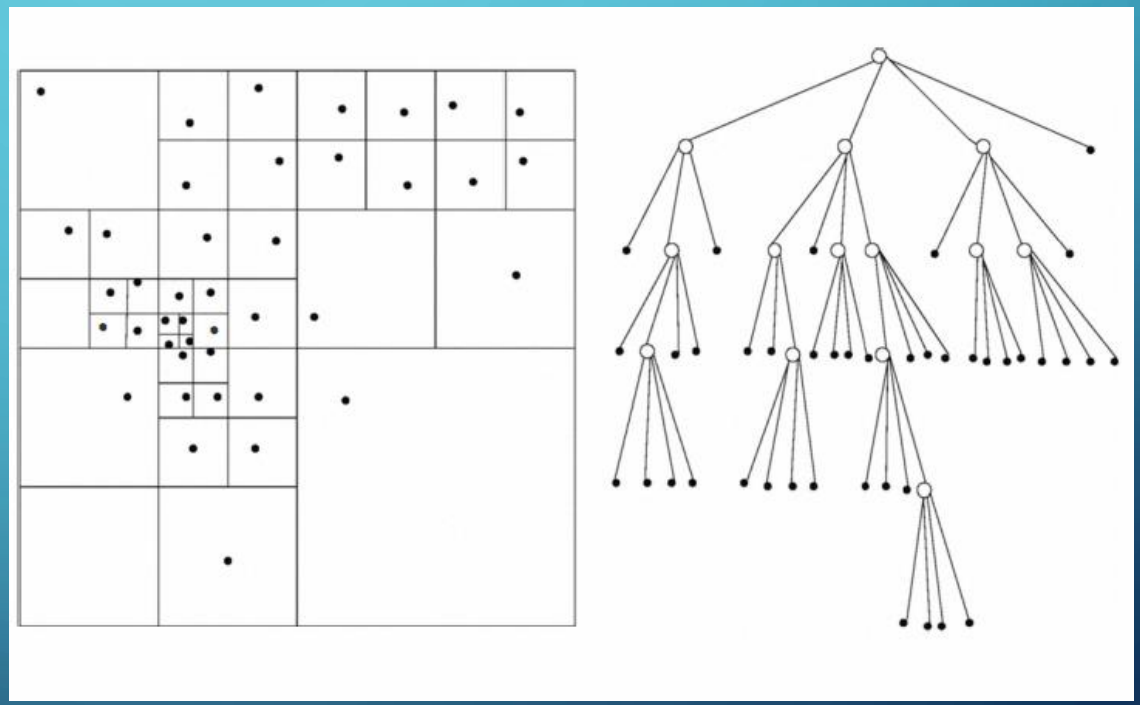
- Repulsion in parallel
- Attraction in parallel

PARALLEL REPULSION

- Barnes-Hut algorithm

- Repulsion on regions
  - Complexity:

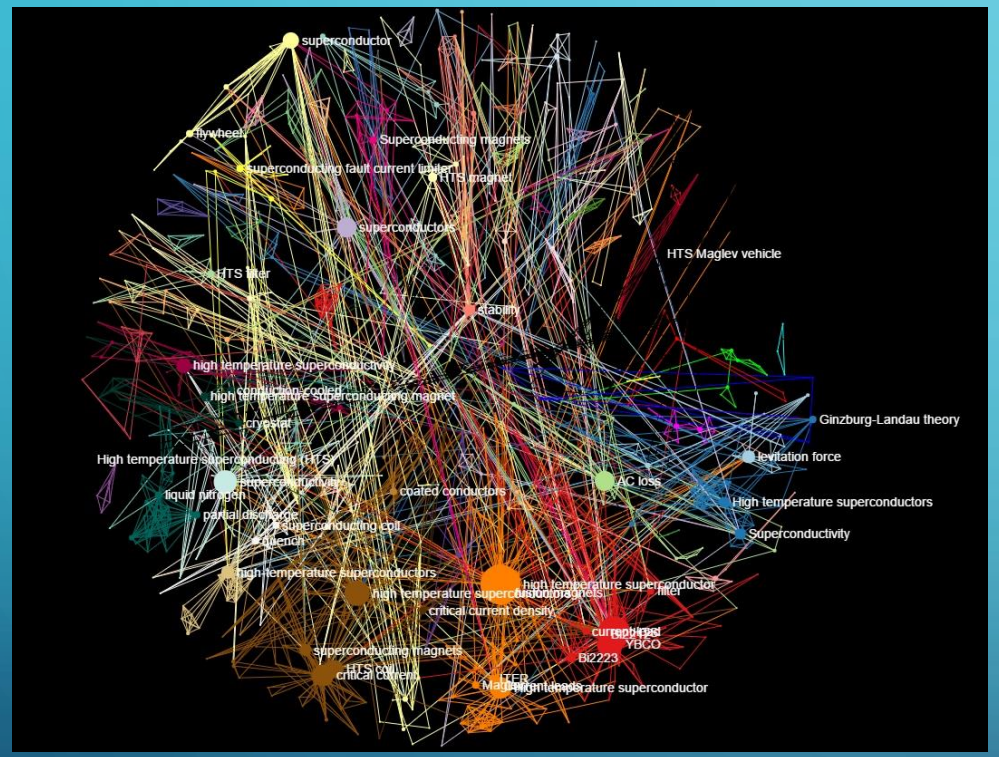
    $O(N*LOG(N))$ instead of $O(N^2)$



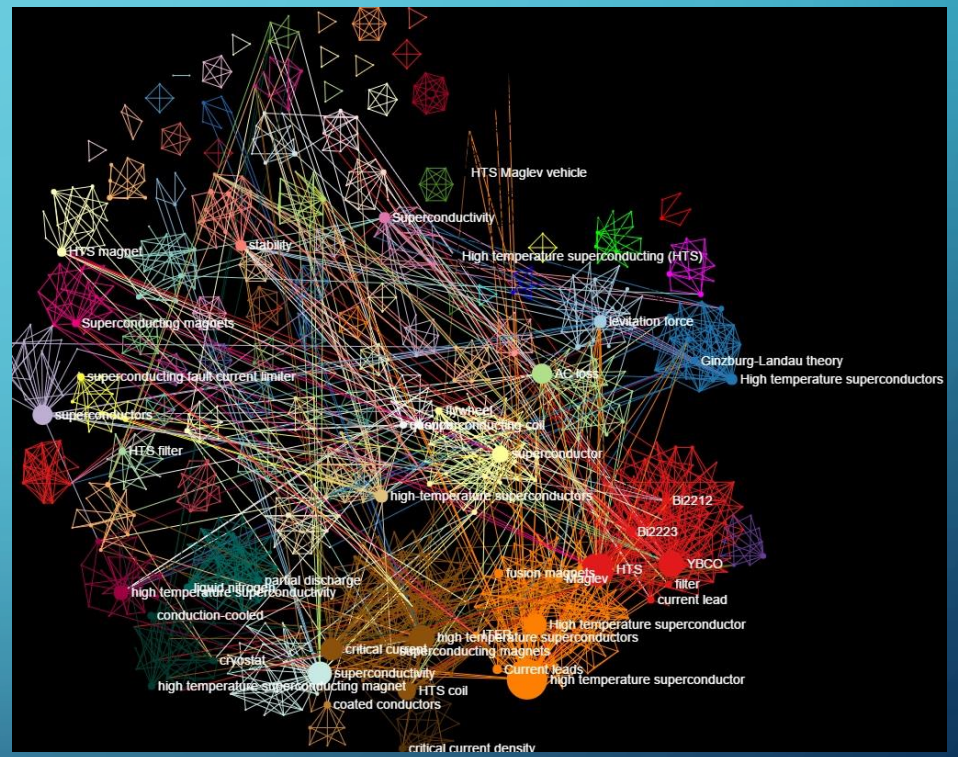
PARALLEL ATTRACTION

- Connected nodes are attracted to each other
- Attraction intensity controlled by the user

FORCEATLAS LAYOUT TYPES



Original layout



Community based layout

Search: in title or abstract (high AND temp* AND supercond*)

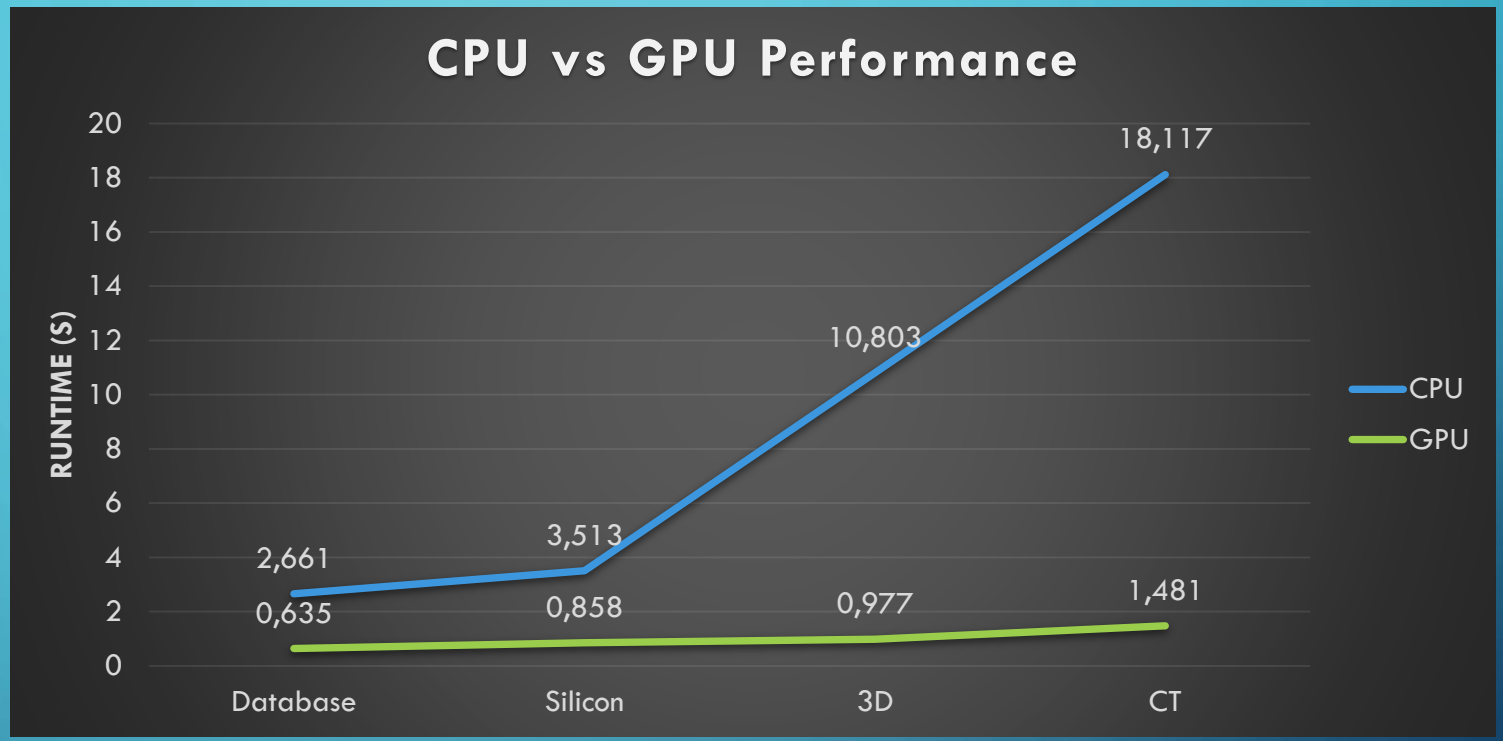
PERFORMANCE RESULTS

PERFORMANCE RESULTS

- Graphs used: *3D, CT, Database* and *Silicon* technologies
- Implementation in C++
- 8 threads in parallel execution
- System:
  - Intel Core i7 920
  - GeForce GTX 980

RESULTS
(COMMUNITY DETECTION)

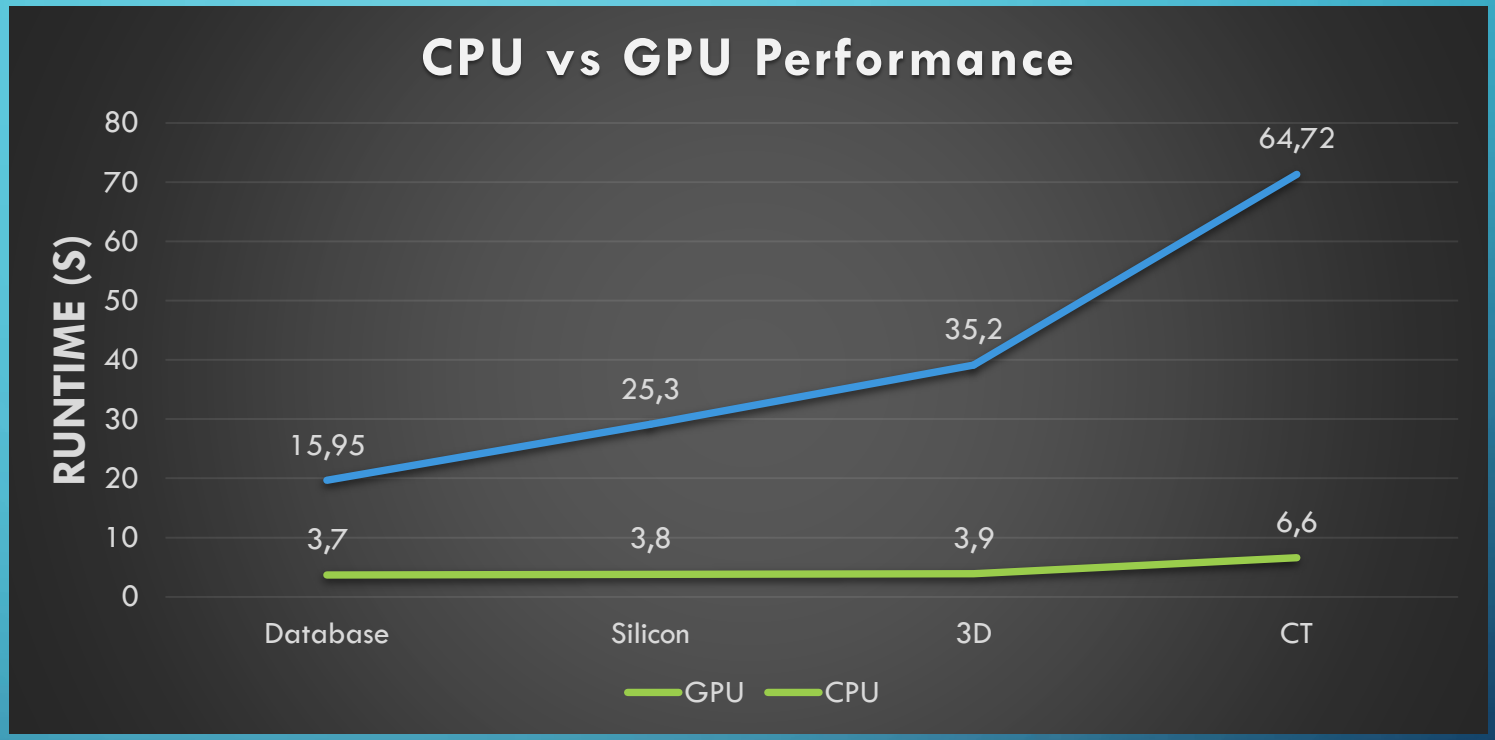
- Database: 4x
- Silicon: 4x
- 3D: 11x
- CT: 12x



	Database	Silicon	3D	CT
Nodes	17832	24923	35763	53039
Edges	113622	185594	192336	325490

RESULTS
(FORCEATLAS)

- Database: 4,31x
- Silicon: 6,65x
- 3D: 9x
- CT: 9,8x



	Database	Silicon	3D	CT
Nodes	17832	24923	35763	53039
Edges	113622	185594	192336	325490

FUTURE WORK

- Further optimization of the graph data generation
- ForceAtlas optimization

THANK YOU