

# Path-based approach for Important Node Exploration

In collaboration with



## Introduction

Our main goal is to provide people with good insights on technological areas, they have little knowledge about. We do that by obtaining graph of patents citations with respect to provided request and finding core nodes in it.

**Approach:** get clear data and utilize GNN architectures to highlight core nodes.

**Challenge:** there are no labels, validation can be done only with the help of specialists from the specific field.

## Objectives

- Create pipeline with
  - Scrapping relevant patents from google patents
  - Enrichment of citation graph with more relevant nodes and edges
  - Pruning of non-relevant nodes with preservation of graph topology
- Research potential ways to highlight core nodes
- Validate the results

## Process & Methods

1. Designed data load pipeline
2. Tested and incorporated relevance score estimation with modification of BERT model
3. Designed and implemented algorithm for graph enrichment.
4. Experimented with various versions of algorithm for finding core node
  - GNN for link prediction
  - Attention on paths
  - Attention on random walks
5. Validation on 5 technological domains
  - photolithography simulation
  - etch simulation
  - plasma discharge simulation
  - physical photoresist model
  - microscope image stitching

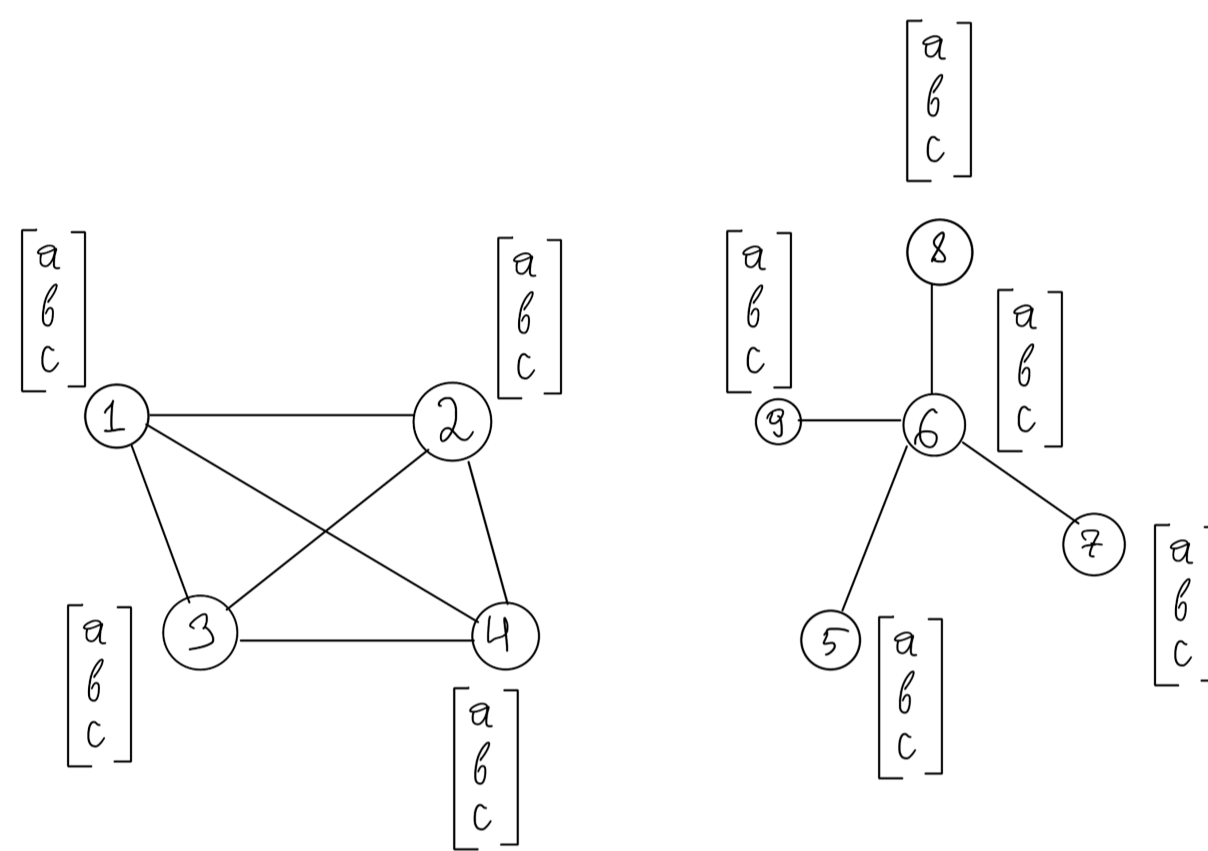


Fig.1 Graph with vector features

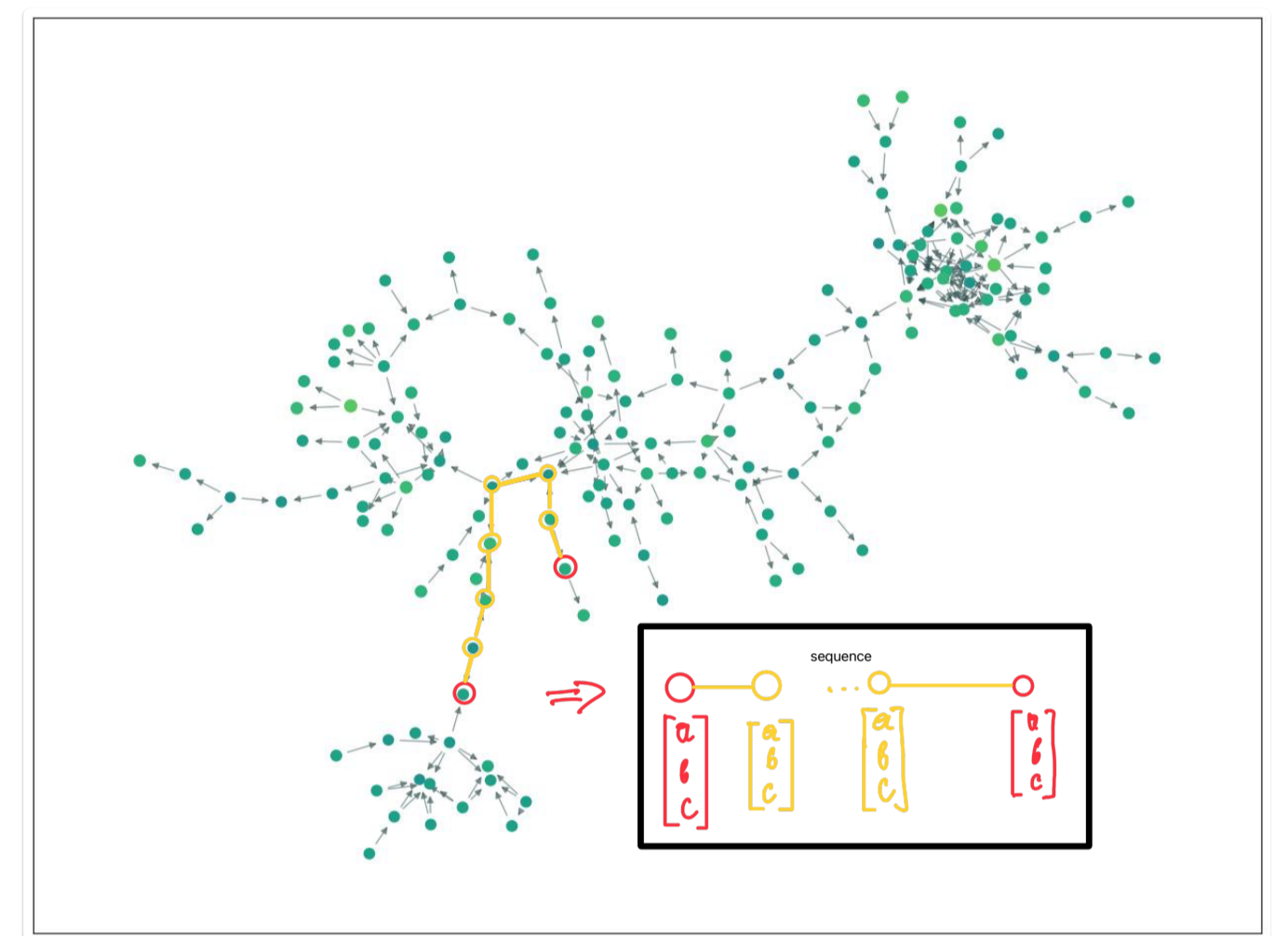


Fig.2 Attention on paths

## Results

Repository for forming request to google patents and:

- Data preparation
- Enrichment of graph with
  - patent citations, obtained from merging duplicates
  - new patents, highly cited by patents from existing set
- PINE Algorithm for finding core nodes
- Validation from experts

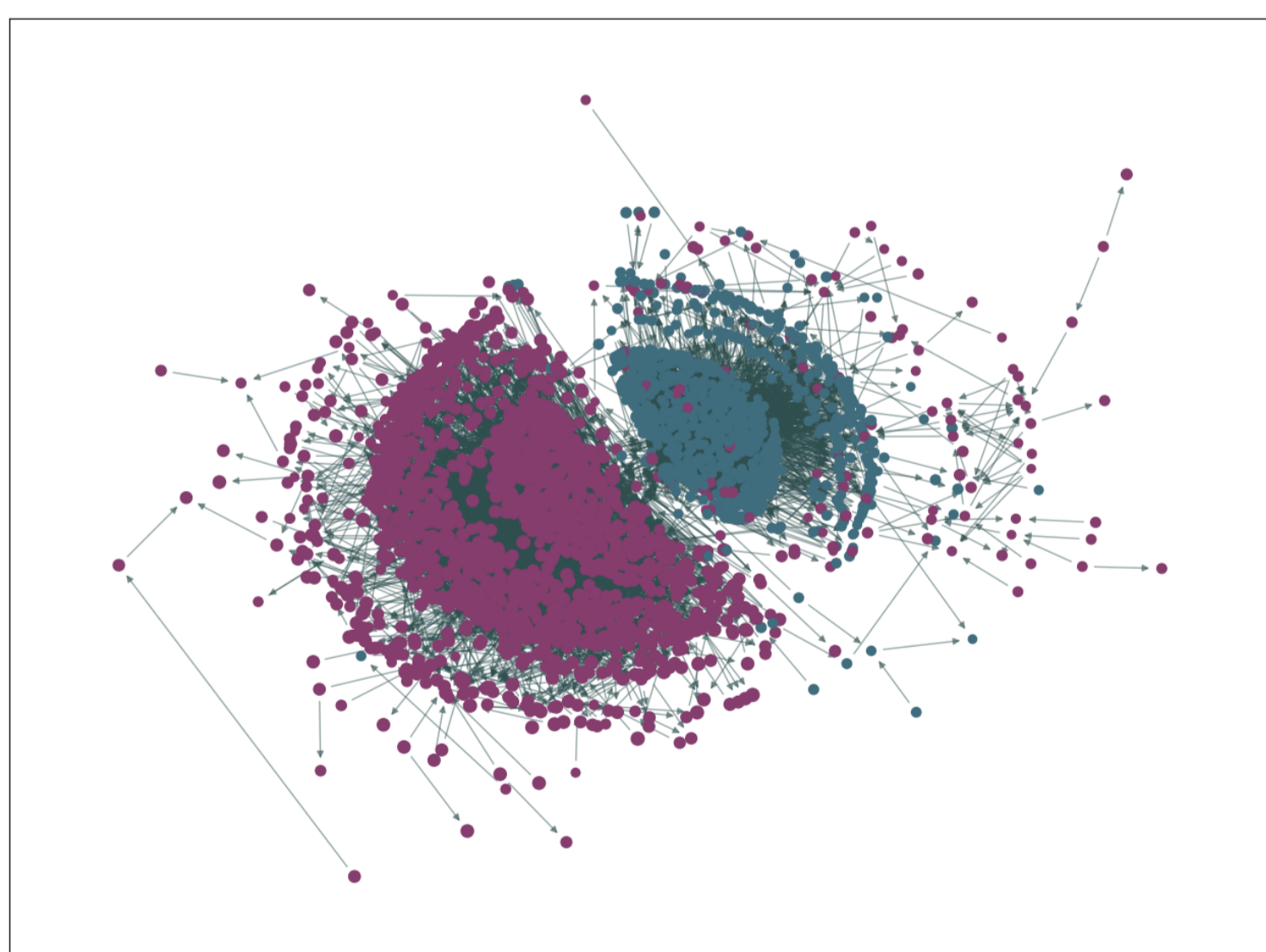


Fig.3 KMeans clusterization visualized on graph for lithography patents

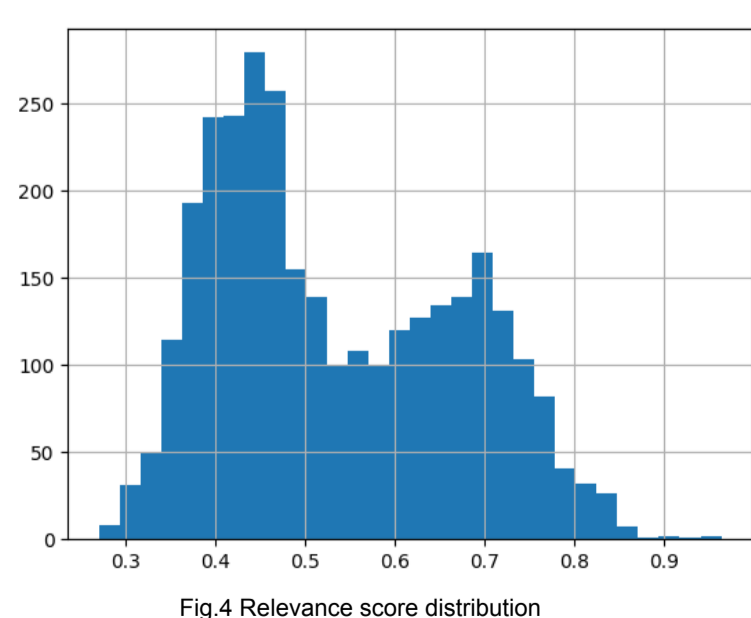


Fig.4 Relevance score distribution

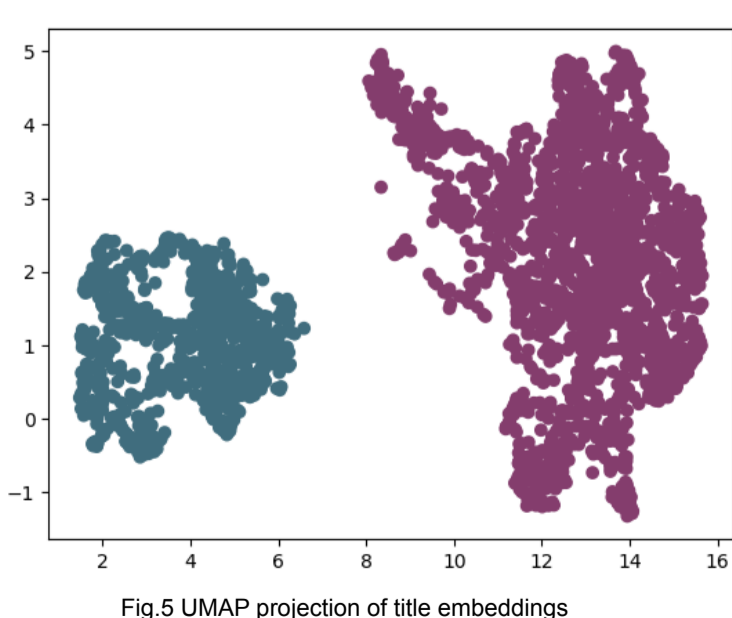


Fig.5 UMAP projection of title embeddings

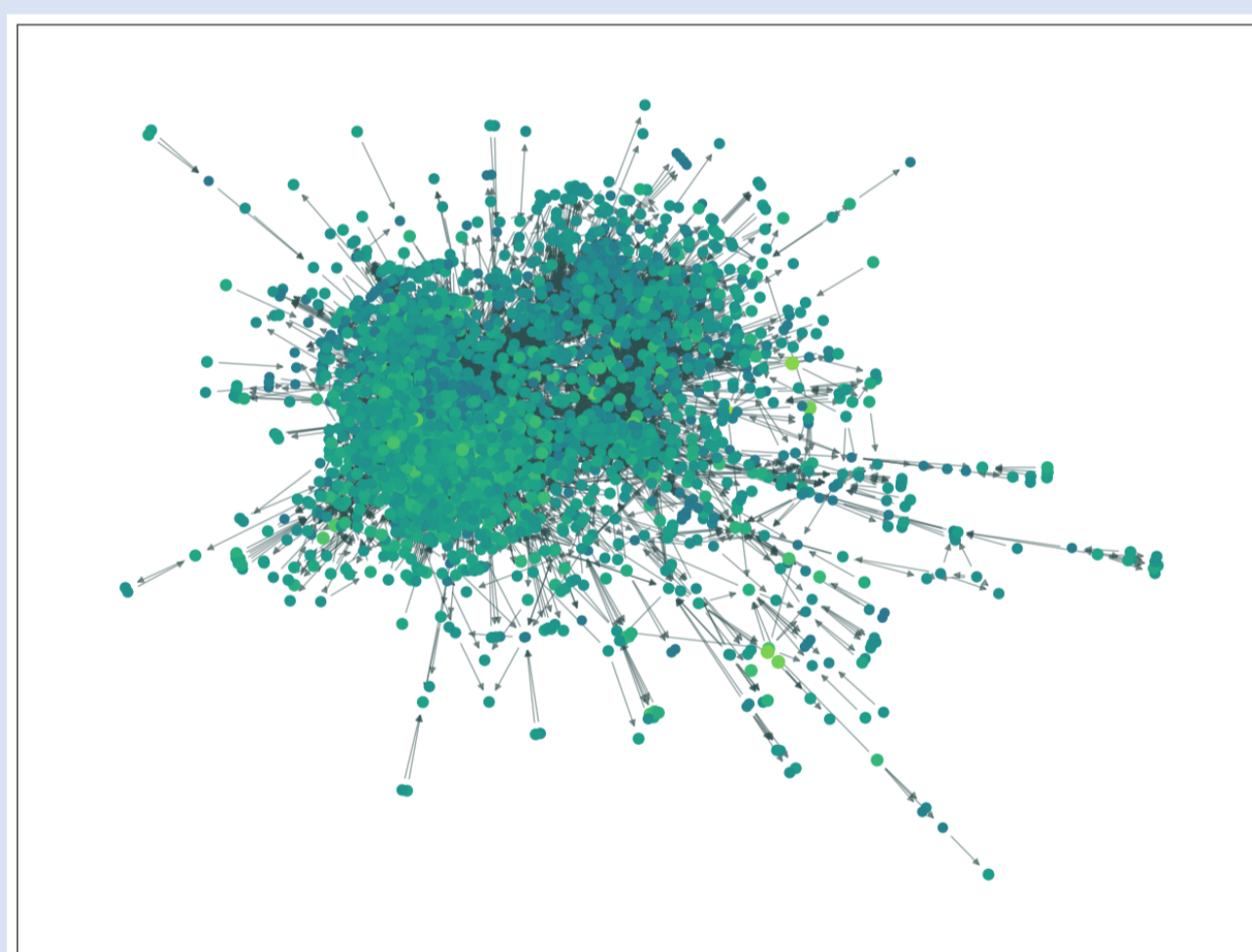


Fig.6 Citation graph for Etch simulation

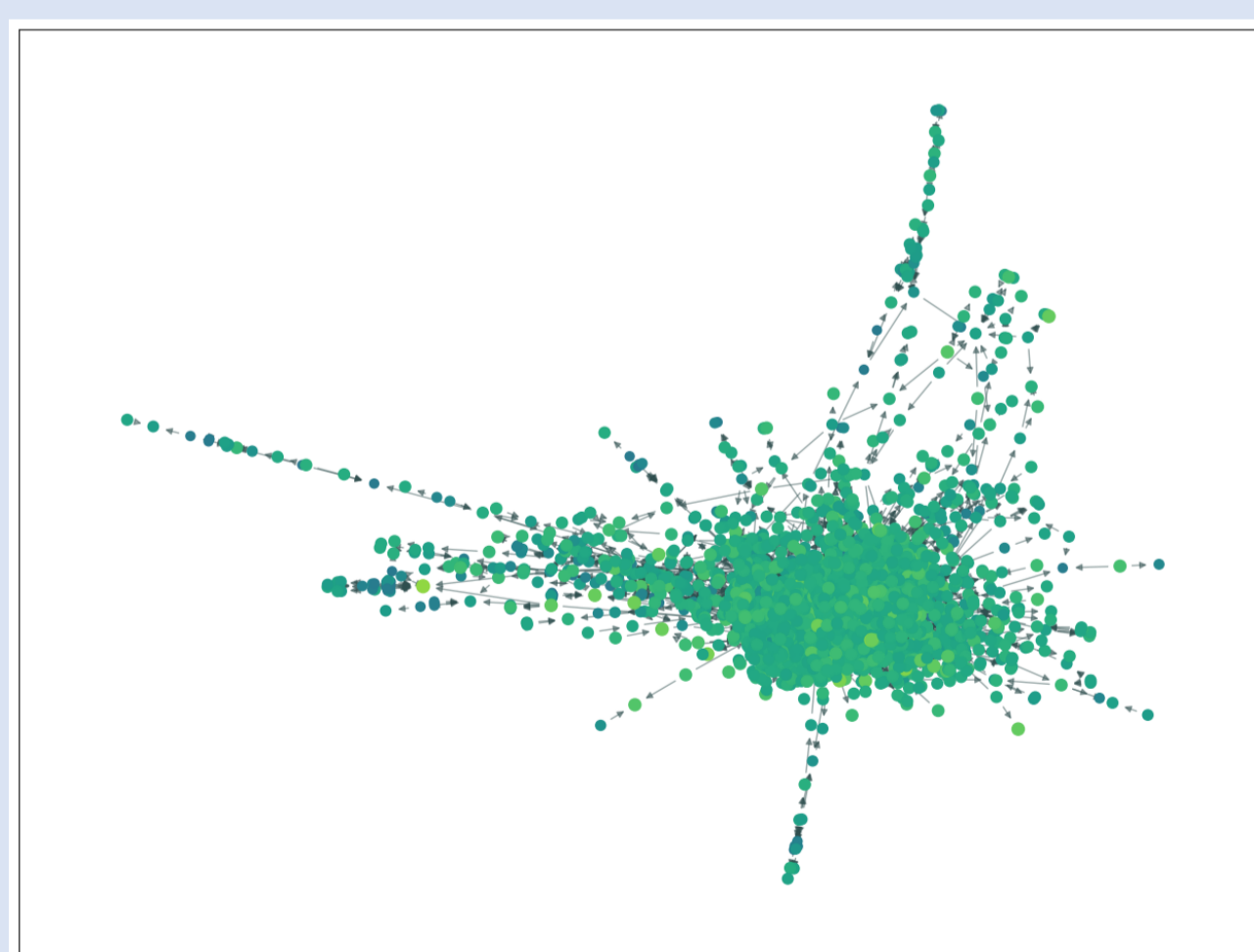


Fig.7 Citation graph for plasma discharge simulation

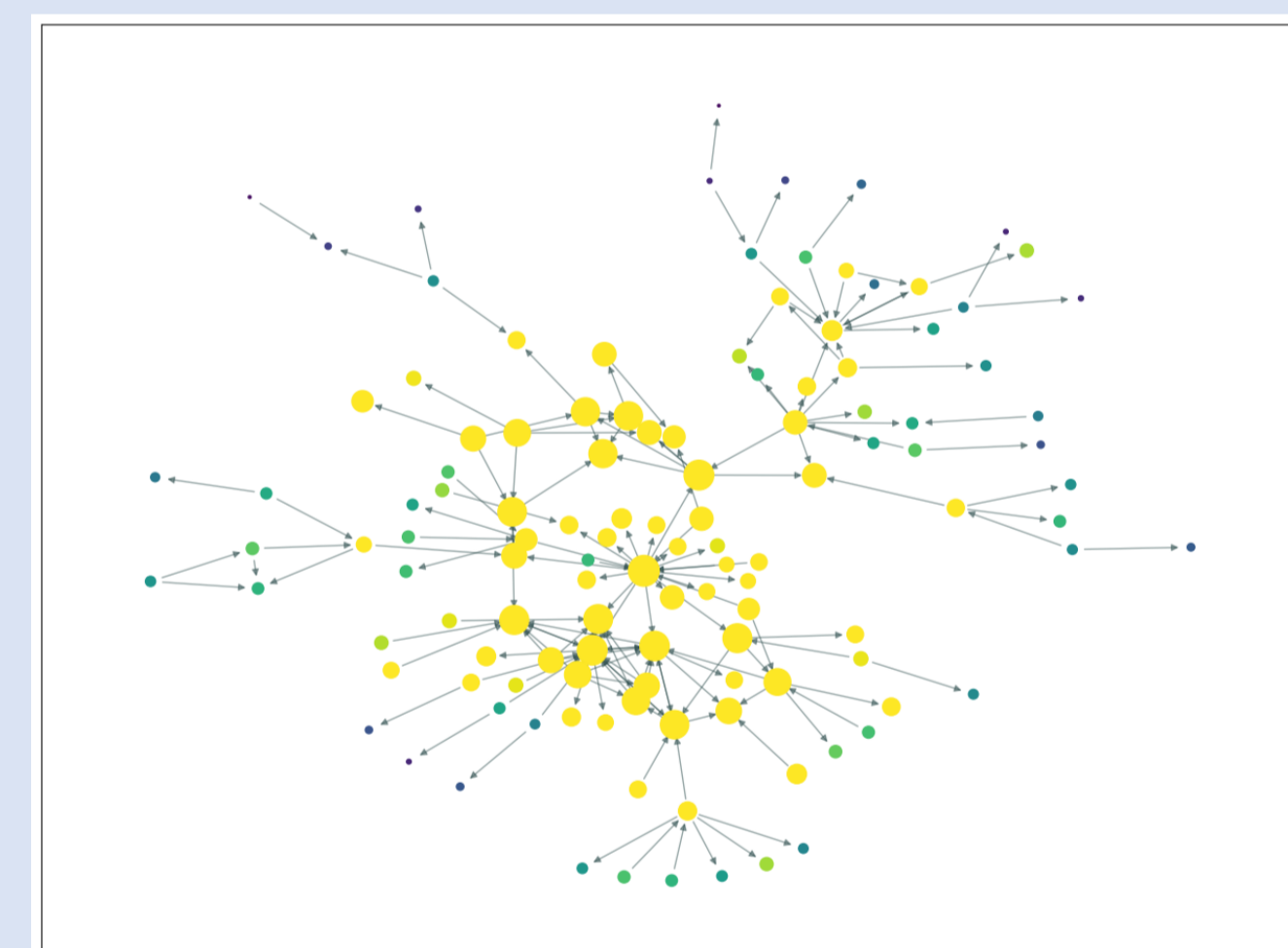


Fig.8 Plasma discharge model, resistance distance on relevance scores

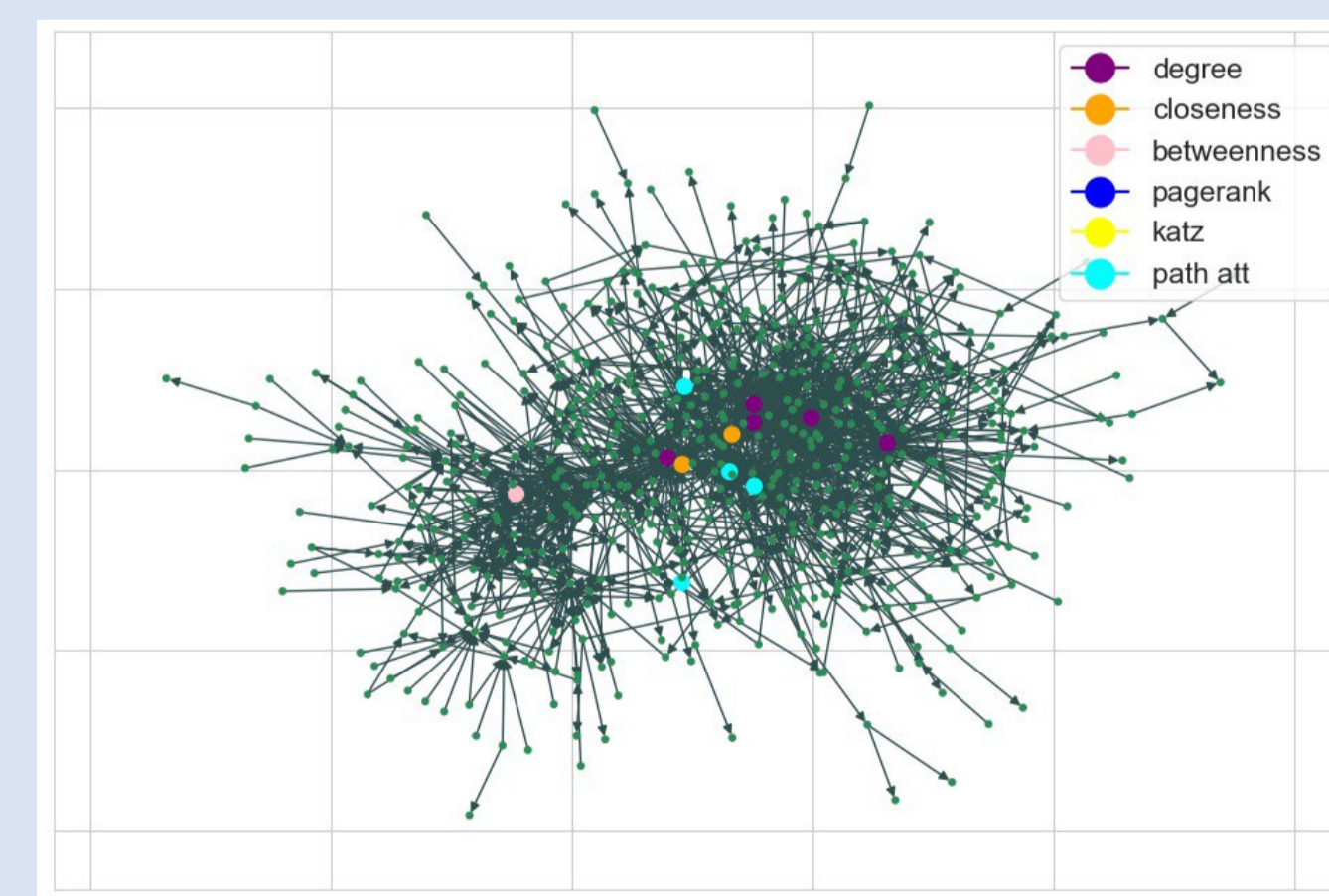


Fig.7 Etch modelling patents pruning

