



# DSC 202 FINAL PROJECT

---

## Technology Forecasting

Sheetal Srivastava , Shreya Reddy Pakala , Yashi Shukla

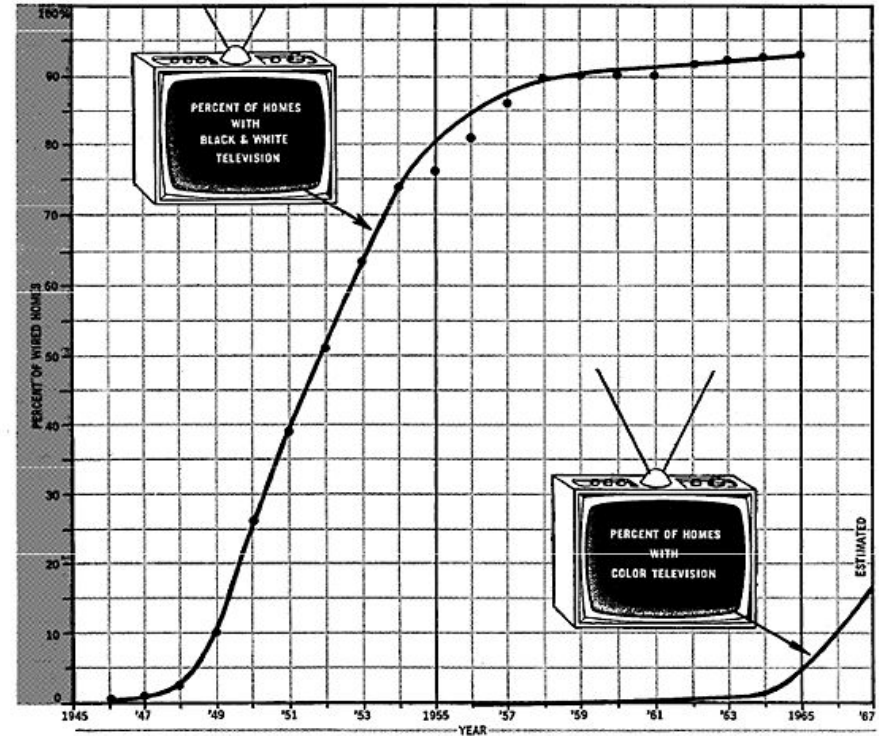


# Introduction

# Motivation

- Technology monitoring
- Trend Analysis
- Market Predictors

EXHIBIT II. GROWTH CURVES OF BLACK-AND-WHITE AND COLOR TELEVISION



SOURCES: *The Economic Almanac 1964* (New York, National Industrial Conference Board, 1964), p. 414; *Electronic Industries Yearbook 1966* (Washington, Electronic Industries Association, 1966), p. 6; *Financial World*, August 24, 1966, p. 6. Figures derived by R. C. Goldstein.



# Data Sources



## Datasets and Databases used

### DATASETS

- Patent data
- SBIR Award data
- Wikipedia (scraped) data

### DATABASES

- PostgreSQL – (RDBMS)
- MongoDB – (NoSQL)
- Neo4j – (Graph DBMS)

patentid	application_number	inventor_name
EP2329291B1	EP09781768.8A	{Jan Laarhuis}
EP2049915B1	EP20070836612	{Raymond S. Di Esposti}
EP2589108B1	EP11729873.7A	{Mihai Banu,Yiping Feng,Vladimir Ivanov ...
EP0428565A4	EP19890908752	{Inna G. Talmy,Deborah A. Haught}
EP2169420B1	EP20090171231	{Siegfried Rouzes,Lilian Lacoste}
EP2492706A1	EP11155483A	{Jan Laarhuis}
EP1818687A1	EP20070101871	{John B. Mckitterick}
EP1445453A2	EP03020505A	{Eyvind Frederiksen,Lars Frederiksen,Sve...
EP1623949A2	EP05016327A	{Frank Neerhut}
EP2100842A1	EP20090154929	{Michael Dr.-Ing. Tilkorn,Ronny Ischgane...
EP3280541A1	EP16777498.3A	{Sungho Jin,Calvin GARDNER,Stewart MATTH...
EP8009895A		{Cameron Dale Hinman,David J. Danitz}
EP20040791702		{Davide Filizola,Dario Parata,Emanuele G...



PostgreSQL

```

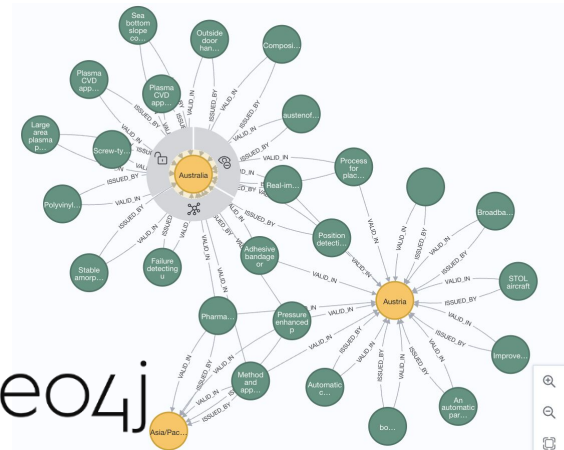
_id: ObjectId('638d92d9e2aee8fe48168124')
name: "Biocon Limited"
logo: "Biocon Logo.svg"
type: "[[Public company|Public]]"
traded_as: "{{BSE|532523}} <br /> {{NSE|BIOCON}}"
founder: "[[Kiran Mazumdar-Shaw]]"
key_people: "[[Kiran Mazumdar-Shaw]]<small><br />(Executive Chairperson)</small><br...\"
ISIN: "INE376G01013"
industry: "[[Biopharmaceutical]]"
products: "[[Biologics]] <br /> Small molecules <br /> Branded formulations <br /> ..."
revenue: "[[Increase]] {{INRConvert|8396.70|c}}"
revenue_per_share: "FY2022"
operating_income: "{{{up}} {{INRConvert|1119|c}}"
income_per_share: "FY2022"
net_income: "{{Decrease}} {{INRConvert|648.4|c}}"

```

mongoDB



neo4j



Node properties

Country

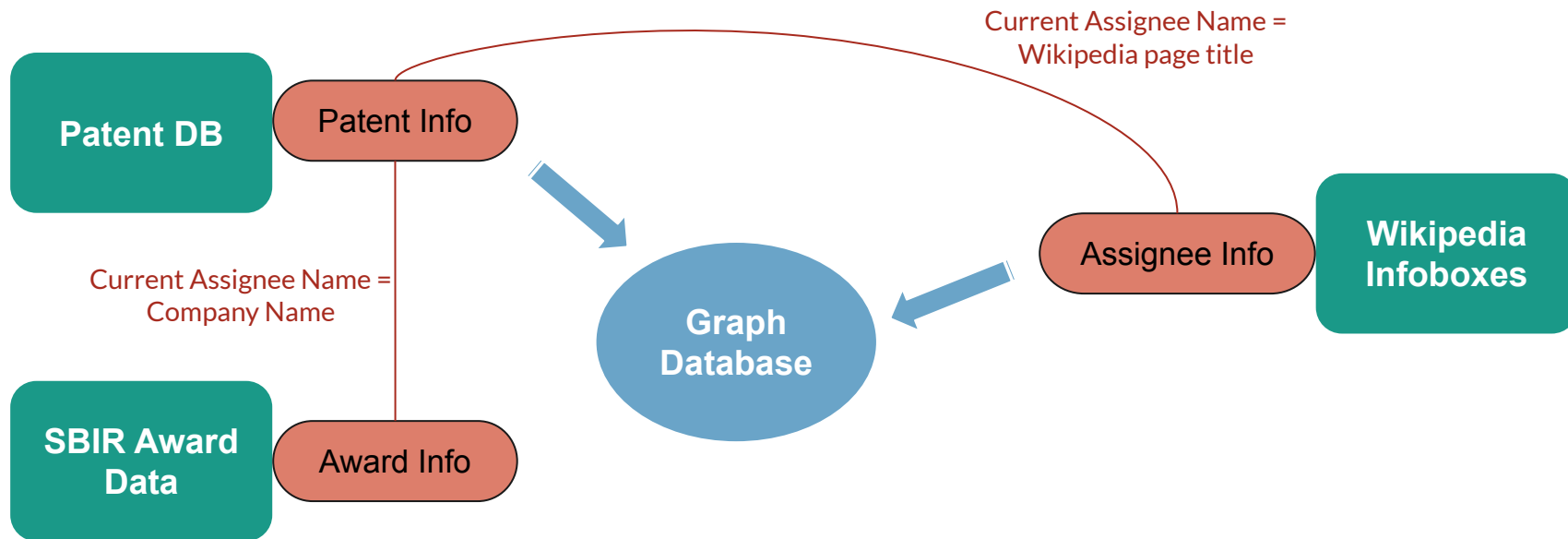
<id> 89871

countryid AU

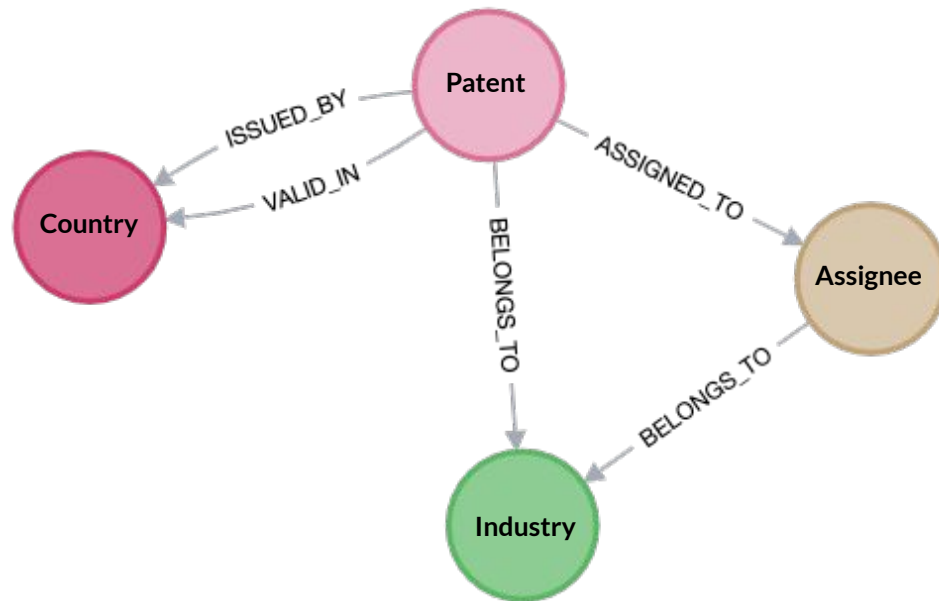
countryN Australia

ame

## Collating Data Sources



# Graph Data Model







# Methodology

# Technical Steps

## Installing Libraries/Softwares

1. Neo4j Desktop
2. Mongo Atlas
3. Psycopg2
4. Neo4j
5. PyMongo

## Pre-processing

### Patent DB:

1. Industries
2. Valid Countries
3. File date

### Wikipedia:

Extracting core industries of the assignees.

### PyCountry:

Extracting country names from country code.

## Data Ingestion

### Nodes:

1. Patent
2. Industry
3. Assignee
4. Country

### Relationships:

1. Belongs\_to
2. Assigned\_to
3. Valid\_in
4. Issued\_by

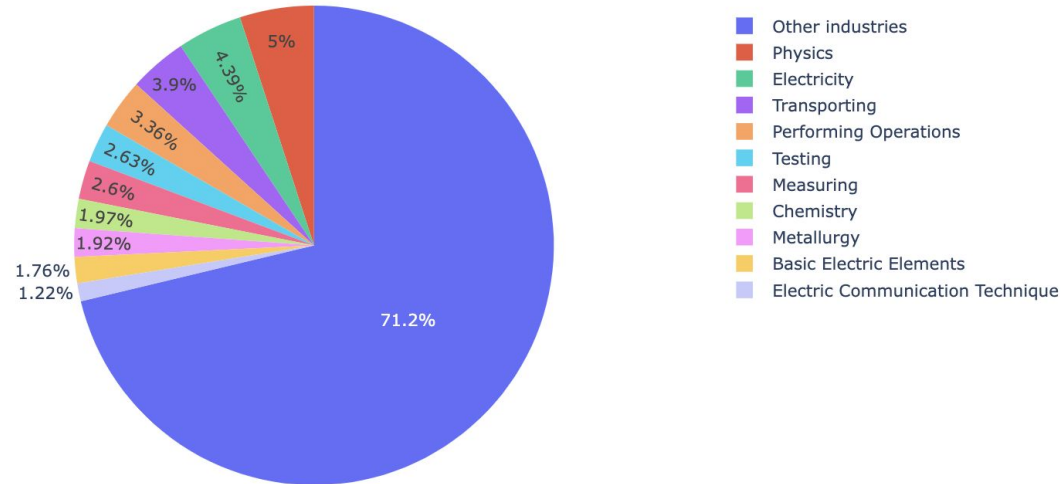


# Analysis

# Popular research industries (degree centrality)



Top Research Industries

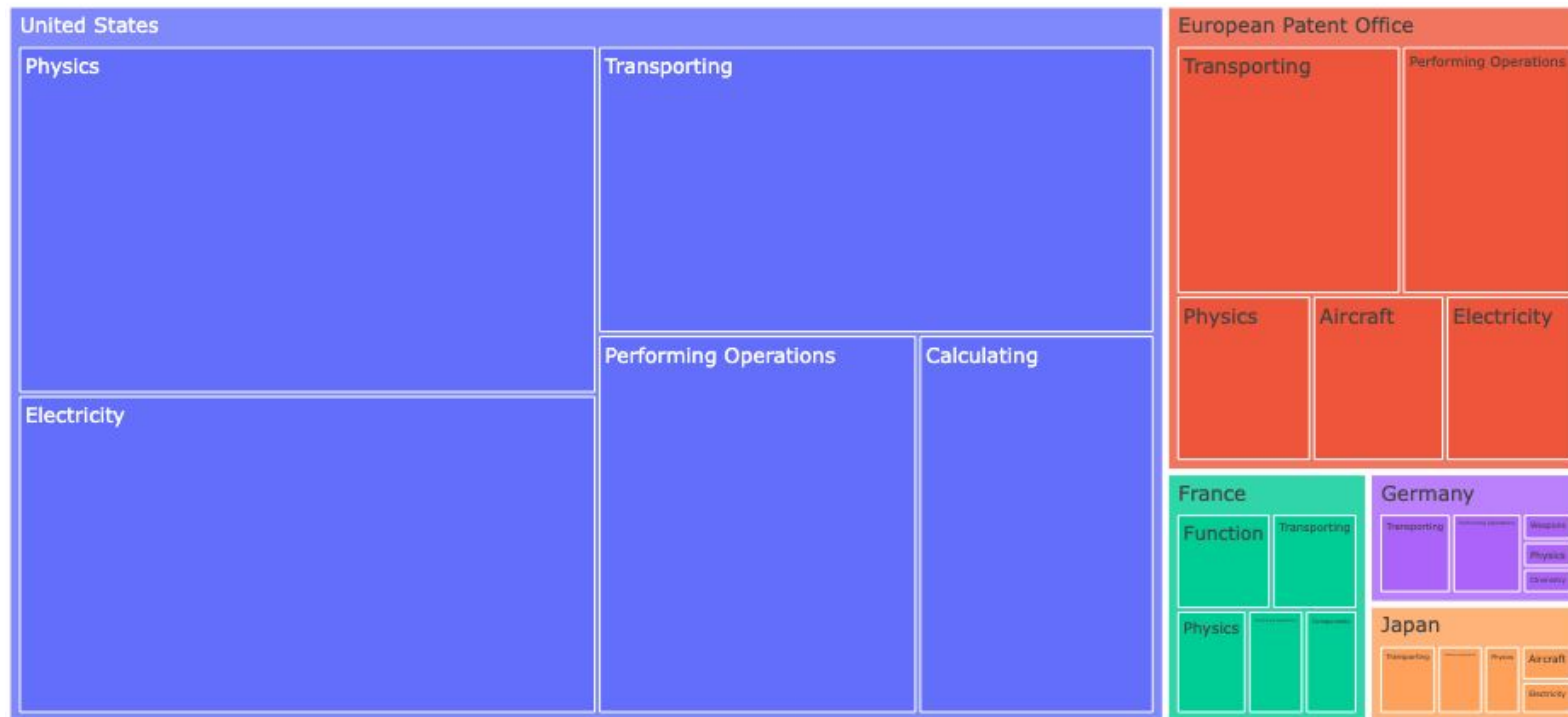


## Analysis of top Assignees (organizations to whom the patent was assigned) for each of the top 5 industries.

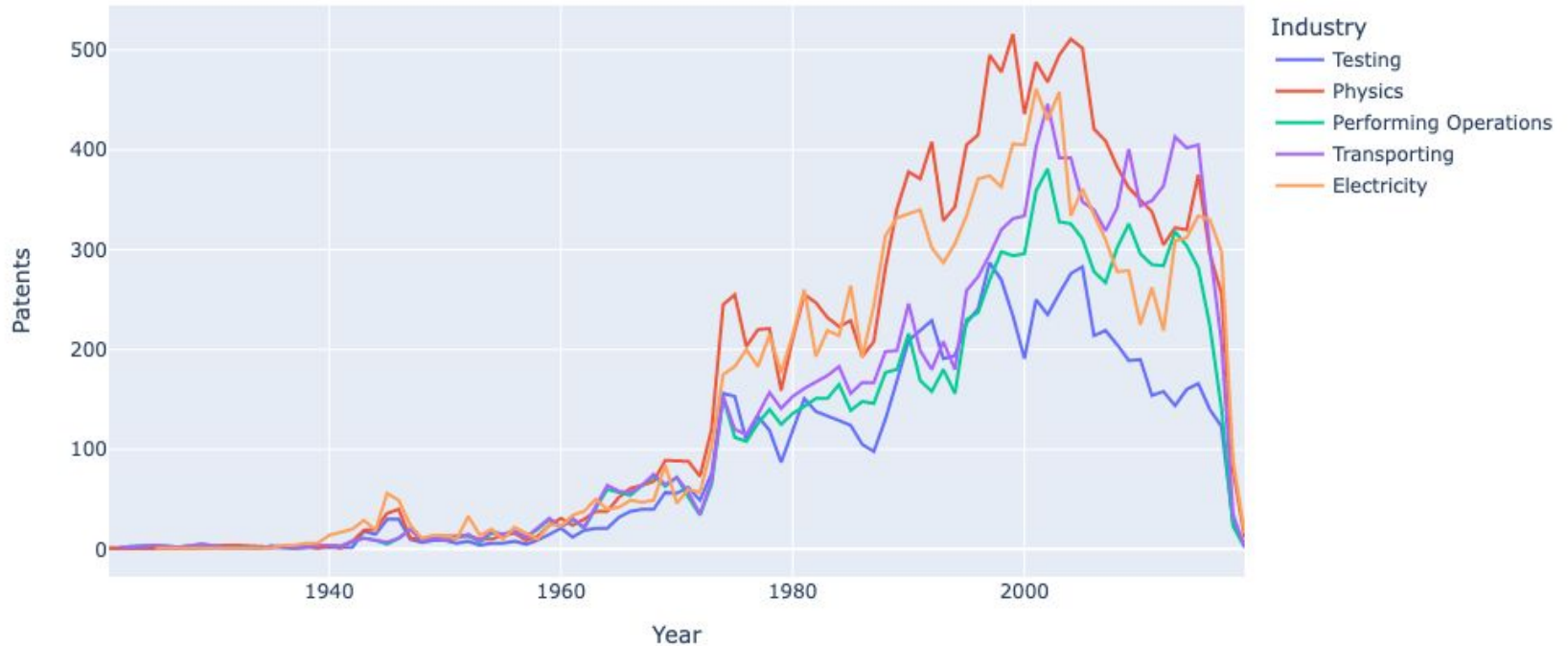
1. Physics = 17395 patents
2. Electricity = 15269 patents
3. Transporting = 13580 patents
4. Performing Operations = 11696 patents
5. Testing = 9166 patents



# Country-wise Research in Last 10 years



# Patents Filed over last 100 years in Top Industries

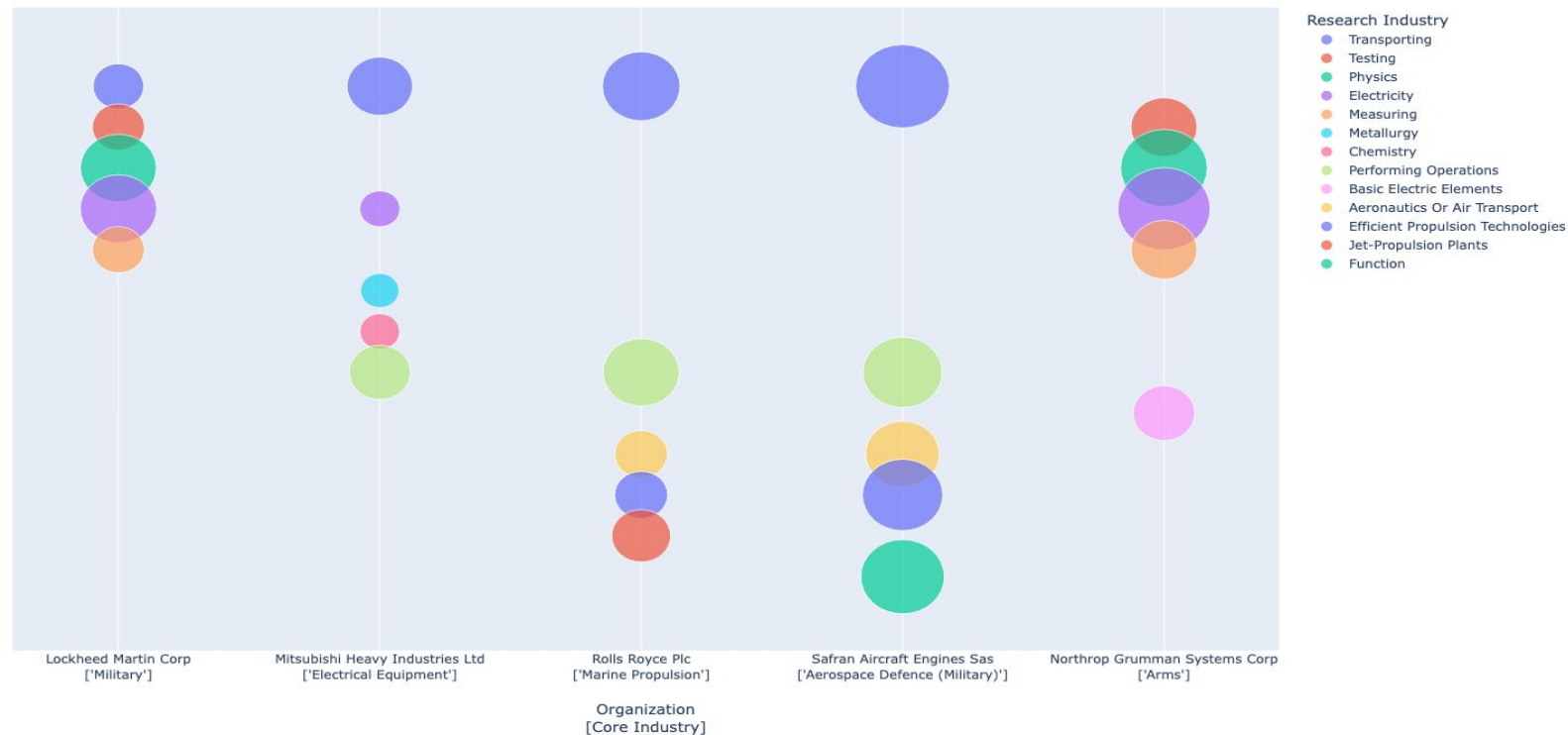


# Patents Filed over last 50 years in Top Industries





# Top 5 Companies Core Industries vs Their Research Industries







# Demonstration