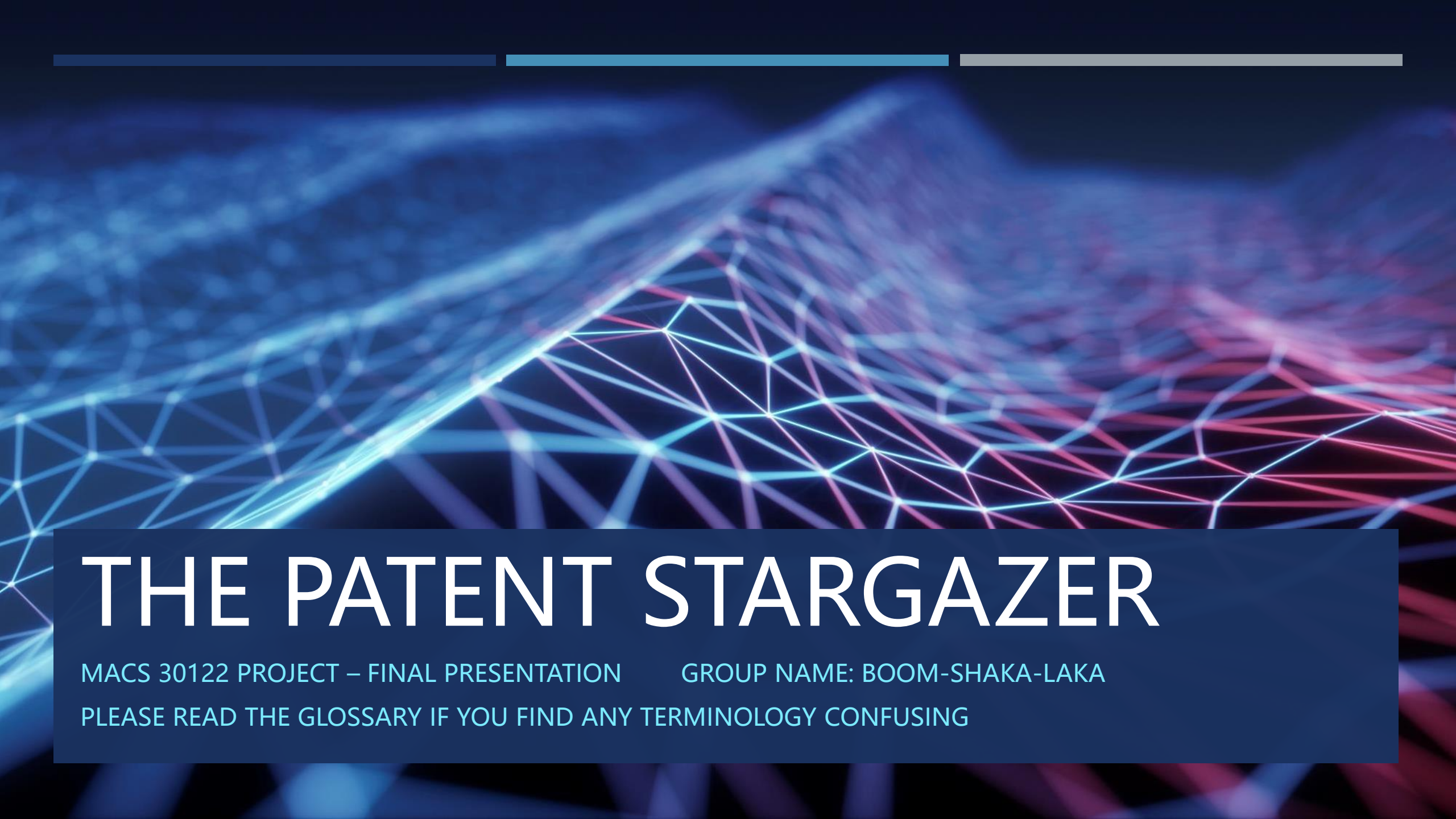


THE PATENT STARGAZER

MACS 30122 PROJECT – FINAL PRESENTATION GROUP NAME: BOOM-SHAKA-LAKA

PLEASE READ THE GLOSSARY IF YOU FIND ANY TERMINOLOGY CONFUSING

STRUCTURE OF OUR REPORT

- Background Introduction & Raw Data Collection
- Barriers We Have Encountered
- Processing Time, Legal Status, Inventor City, CPC
- Network Analysis
- Abstract Analysis (NLP)
- User Interface

BACKGROUND INTRO

- Patent analysis has been an important discipline in both managerial science and macroeconomics. Hence, we are interested in using patent data for technology analysis and planning and will explore in our project "Patent Stargazer" with a **focus on cloud computation patents**.
- We chose the top 3 companies with the largest market share in the cloud computation industry:

Amazon Technologies Inc.

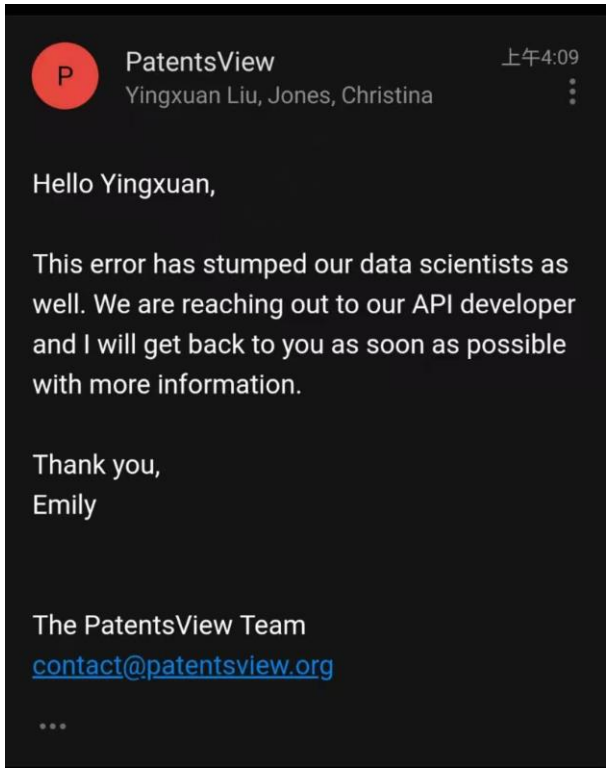
Microsoft Technology Licensing, Llc

Google Inc.

RAW DATA COLLECTION

- This project has used 3 sources of data:
- The PatentsView API provides web developers and researchers programmatic access to longitudinal data and metadata on patents, inventors, companies, and geographic locations.
- incoPat provides as many as 260 searching fields, making an integration of data like patent litigation, licensing, etc.
- 1) incoPat Global Patent Database – Manual download
- 2) PatentsView Patent API
- 3) PatentsView Assignee API

BARRIERS WE HAVE ENCOUNTERED

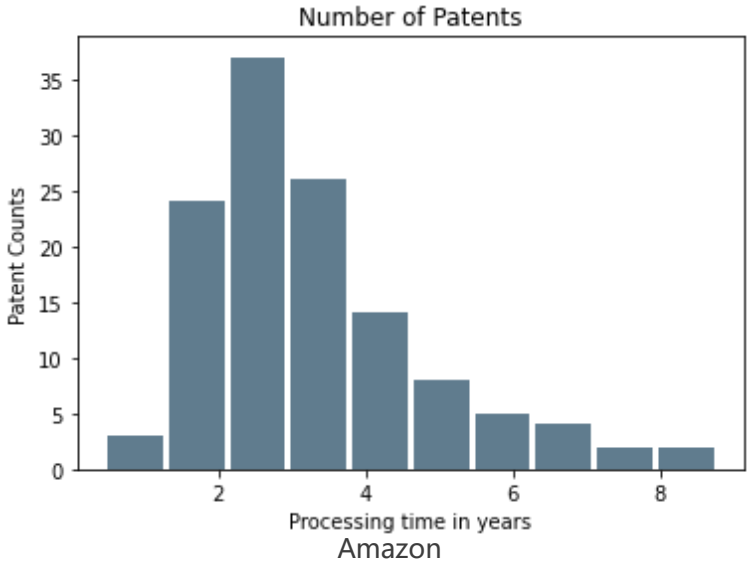


While scraping for citation, the API imposed visit limitation and the service provider experienced internal issues which affects our schedule.

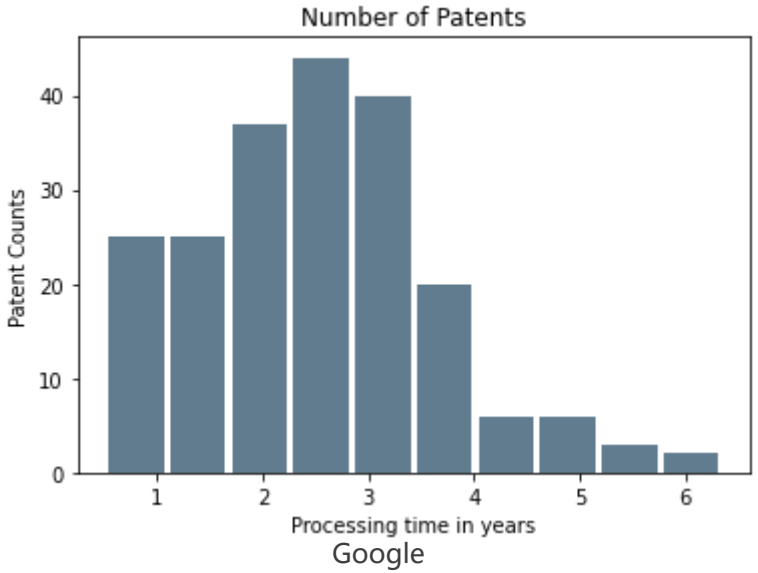
The packages for network analysis is slow while processing massive amount of data.

Nobody in the group had prior experience with GUI.

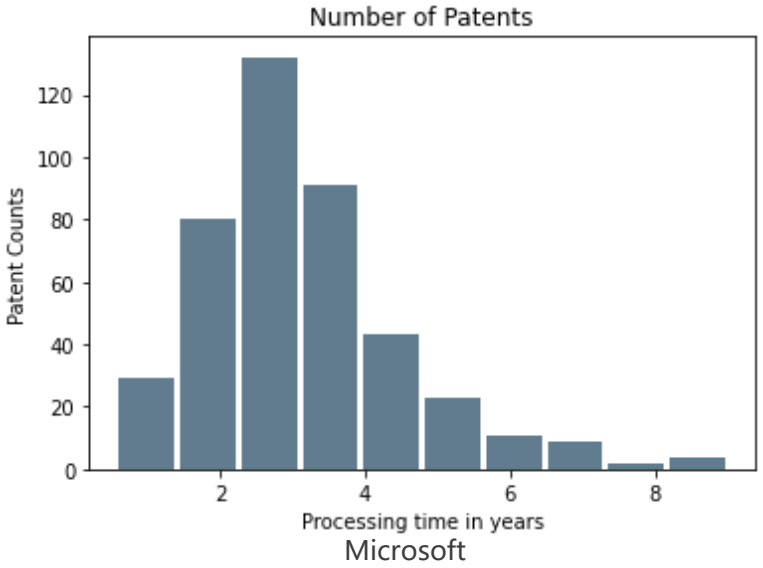
PROCESSING TIME



Amazon



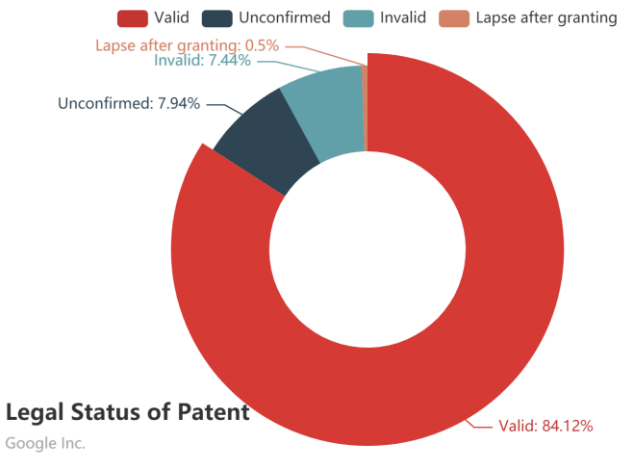
Google

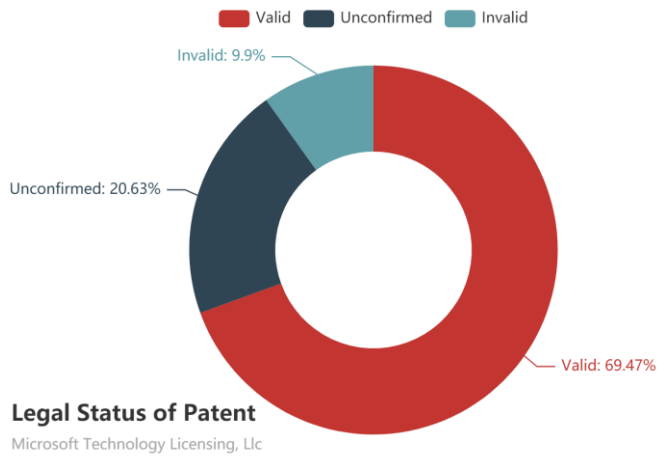


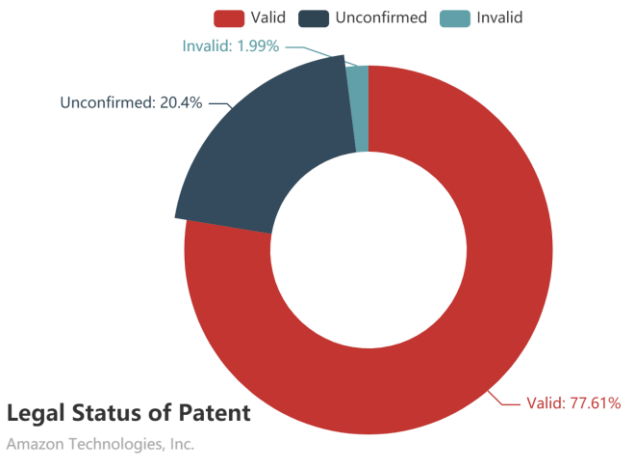
Microsoft

- How long are Amazon's patents usually processed?
- Mainly 2-3 yrs
- This time length is appropriate and shows that Amazon maintains a moderate speed with applying its patents rather than acquiring through litigation.

LEGAL STATUS

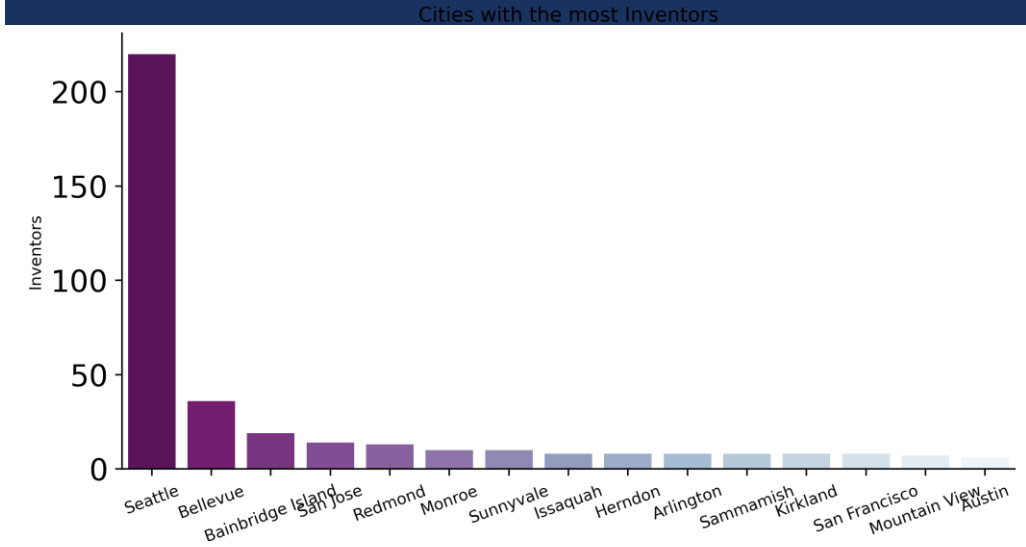


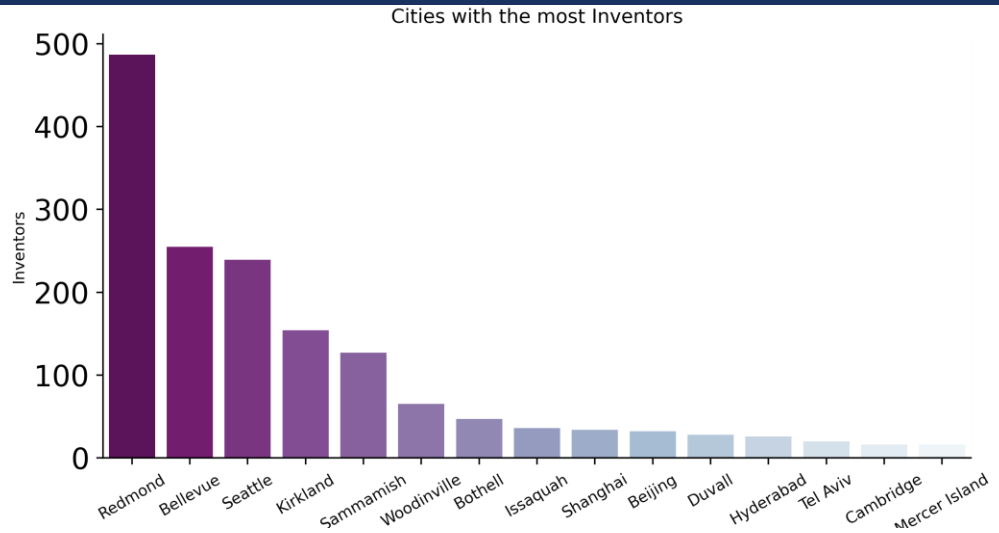


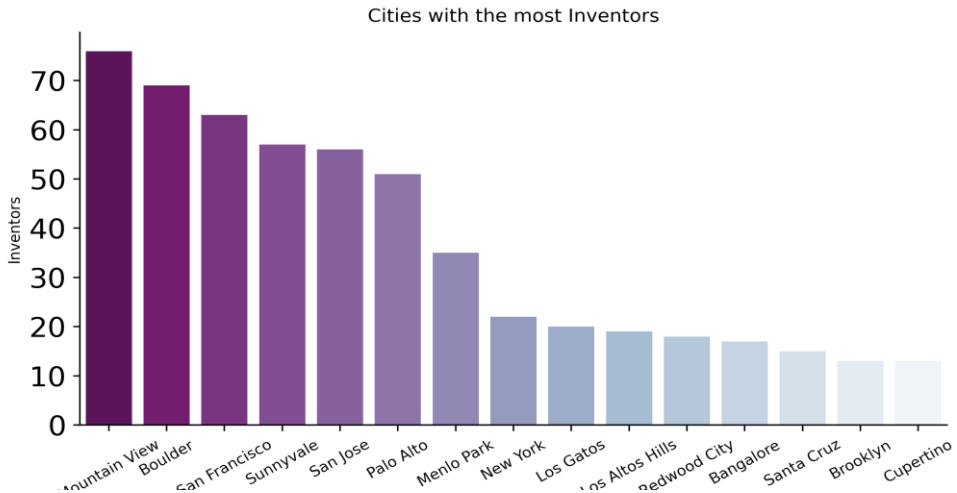


- Knowing the legal status actually explains some glitches happened during scraping.
- Having highest share of valid patents is good? (Like Google) Wait a sec…
- Drawn via Pyecharts

INVENTOR CITY

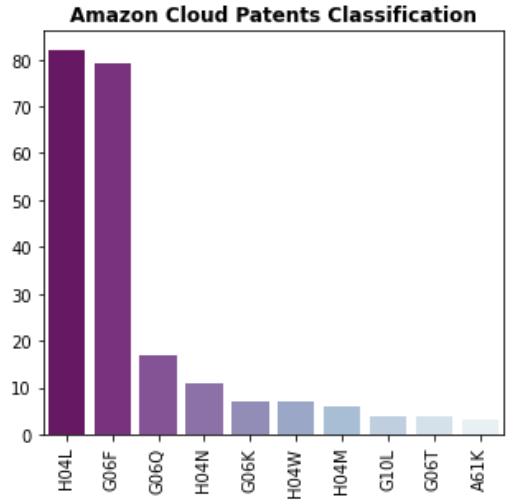


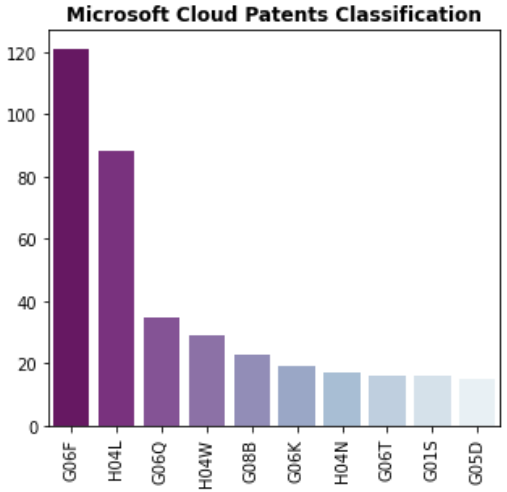


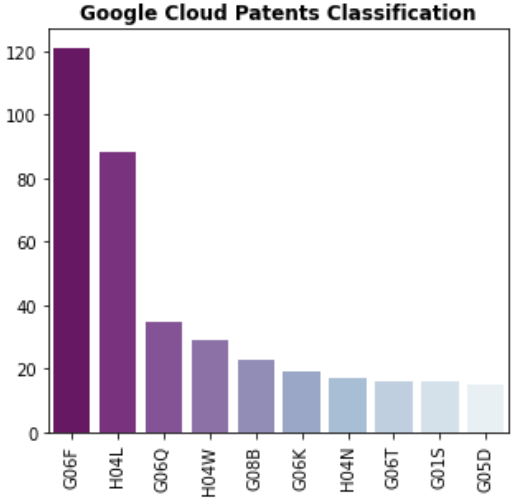


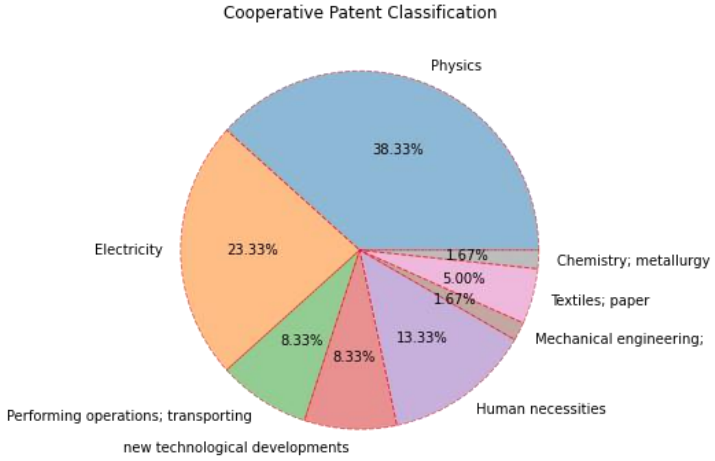
- Where are Amazon's cloud computation patents' inventors are at?
- Mainly the coastal cities
- Cities which are tech-heavy and have good tech and patent policies.
- *Useful for industrial economics

COOPERATIVE PATENT CLASSIFICATION (CPC)









- Method:
  - According to the CPC, we visualize **Amazon, Google and Microsoft** 's CPC group id using PatentsView Patent API
  - This is useful for determining the **main application field** of cloud computing
  - Identify the **diversified competition strategies** for each targeted company
  - CPC subclassification and section

NETWORK ANALYSIS

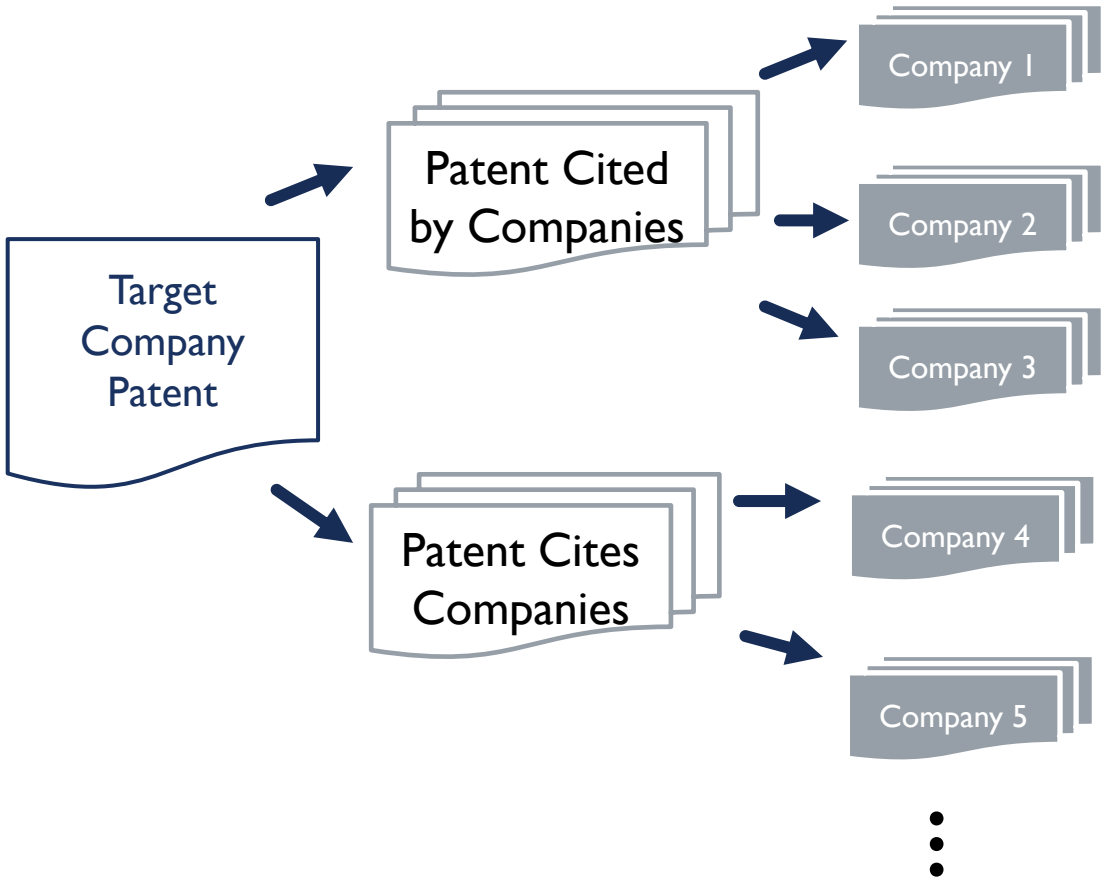


Figure 1. Snowball sampling

- **Data Collection:**
  - Snowball sampling + UPSTO API
  - Around 8,000 pairs of citation pairs
  - 200 companies
- **Data Processing:**
  - Networkx and Gephi
- **Data analysis:**
  - degree distribution
  - centrality metrics
  - smallworld coefficient
  - community detection – Modularity algorithm

$$Q = \frac{1}{4m}\sum_{vw} B_{vw} s_v s_w = \frac{1}{4m}\mathbf{s}^{\mathrm{T}}\mathbf{B}\mathbf{s},$$

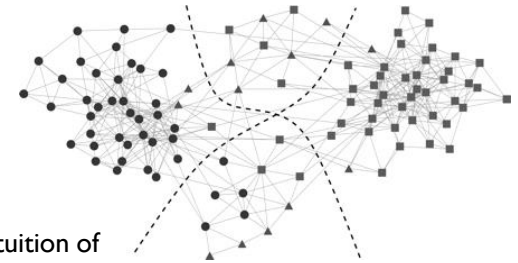
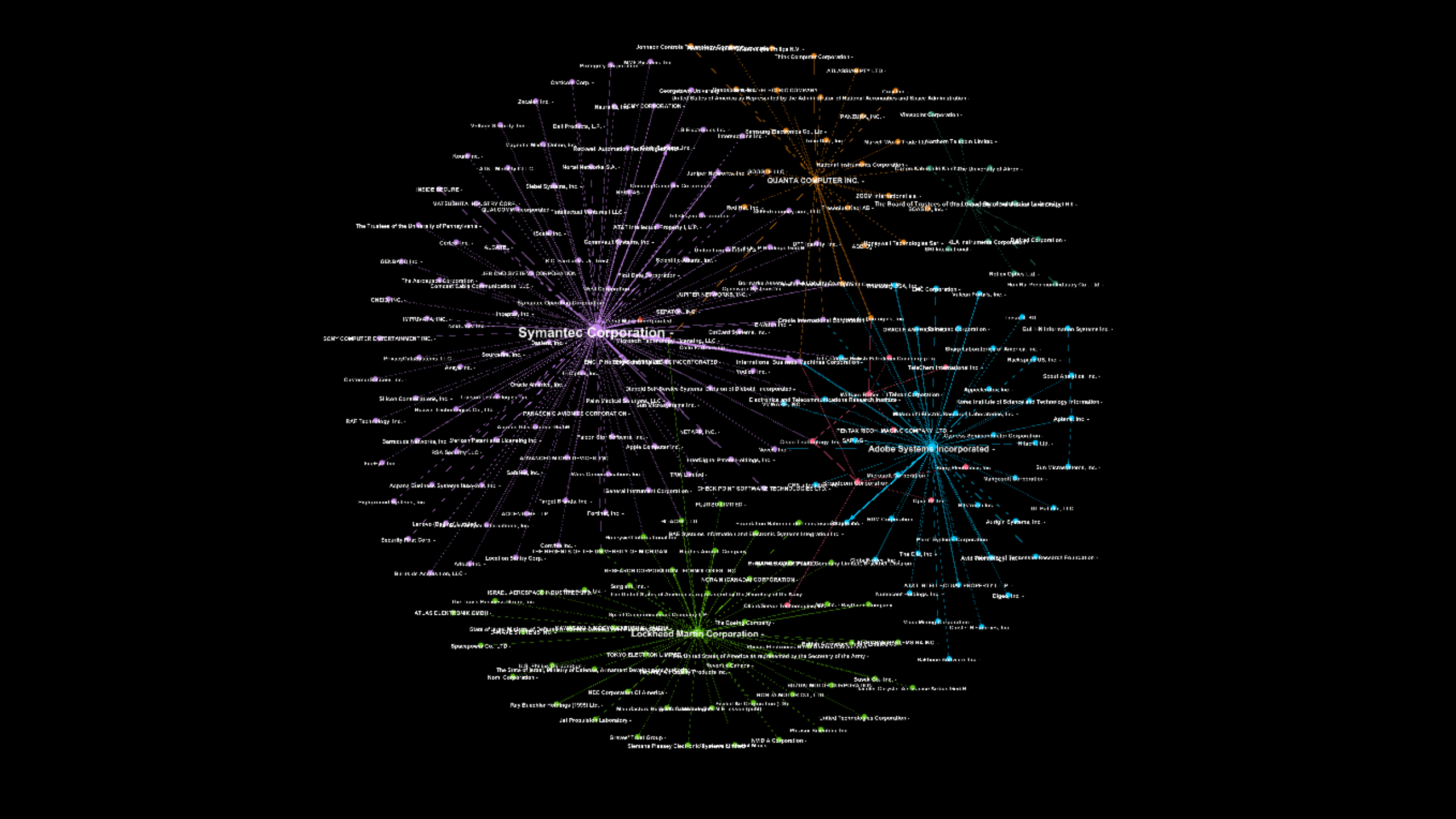


Figure 2. Intuition of modularity algorithm


Symantec Corporation
Adobe Systems Incorporated
Lockheed Martin Corporation
QUANTA COMPUTER INC.

NETWORK ANALYSIS

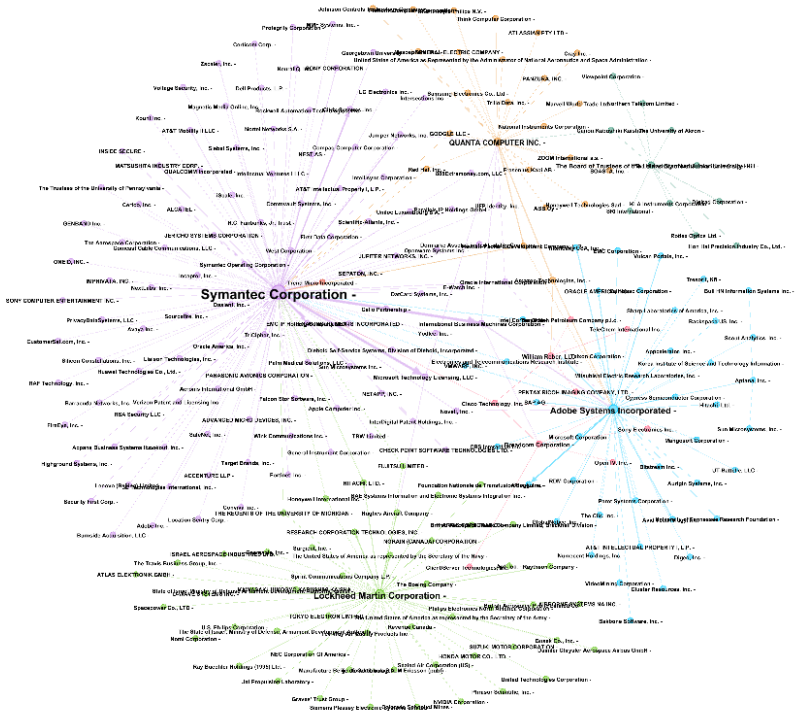


Figure 1. Company network

Network Properties:

- **Power-law distribution of degrees**: few companies with significantly high degree than average companies, forming hubs in networks
- **Small-world effect**:
  - Average length path is about one
  - check the small world robustness: $\sigma = \frac{\frac{C}{C_r}}{\frac{L}{L_r}}$ small world holds true if $\sigma > 1$ ($C \gg C_r$ and $L \approx L_r$)

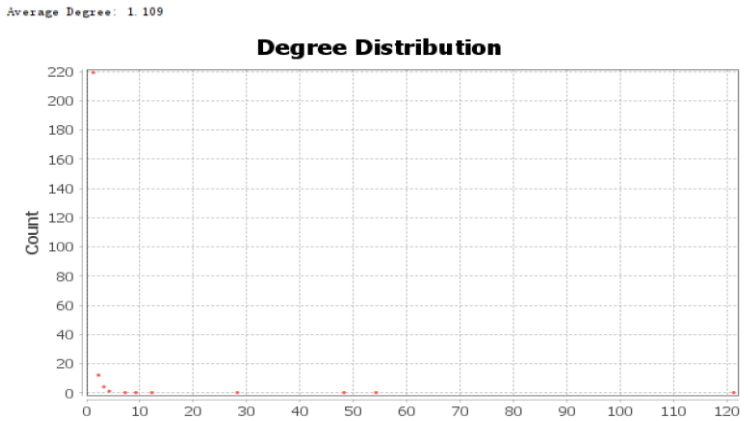


Table 1. Network Properties

	Cloud Computing Network	Random Network
Nodes / Edges	247 / 274	247 / 342
Average degree	1.109	1.385
Clustering Coefficient	0.005	0.001
Average path length	1.018	2.271
Network small-worldness	11.154	

NETWORK ANALYSIS

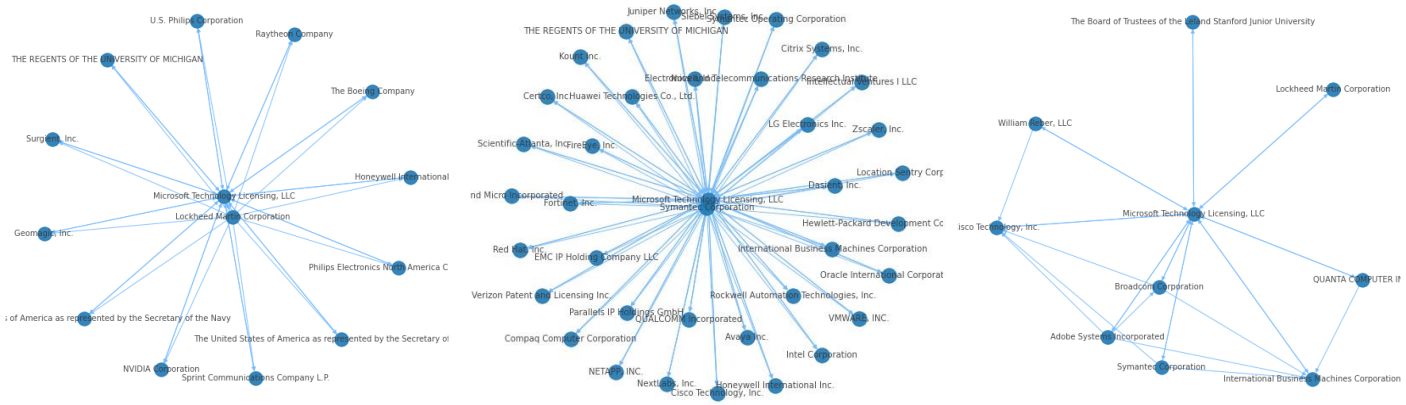


Figure 1. Subgraph network(Left: Lockheed Martin; Center: Symantec; Right: Microsoft)

	company	pagerank score	company	betweeness centrality	company	eigenvector centrality	company	degree centrality
0	Microsoft Technology Licensing, LLC	0.442151	Microsoft Technology Licensing, LLC	0.827405	Microsoft Technology Licensing, LLC	0.827692	Microsoft Technology Licensing, LLC	1.657436
1	Cisco Technology, Inc.	0.000832	Symantec Corporation	0.068618	Cisco Technology, Inc.	0.415132	Symantec Corporation	0.126154
2	Sun Microsystems, Inc.	0.000788	Lockheed Martin Corporation	0.032503	International Business Machines Corporation	0.415132	Lockheed Martin Corporation	0.057436
3	Amazon Technologies, Inc.	0.000769	Adobe Systems Incorporated	0.016358	Oracle International Corporation	0.414874	Adobe Systems Incorporated	0.052308
4	International Business Machines Corporation	0.000762	QUANTA COMPUTER INC.	0.014871	Amazon Technologies, Inc.	0.414874	QUANTA COMPUTER INC.	0.029744
5	Broadcom Corporation	0.000738	The Board of Trustees of the Leland Stanford J...	0.005099	Hewlett-Packard Development Company, L.P.	0.414874	The Board of Trustees of the Leland Stanford J...	0.014359
6	Microsoft Corporation	0.000738	William Reber, LLC	0.002549	Microsoft Corporation	0.414617	Broadcom Corporation	0.011282
7	William Reber, LLC	0.000738	Broadcom Corporation	0.001700	Novell, Inc.	0.414617	William Reber, LLC	0.009231
8	OpenTV, Inc.	0.000726	Aptana, Inc.	0.001062	Intel Corporation	0.414617	Aptana, Inc.	0.006154
9	Canon Kabushiki Kaisha	0.000717	Alcatel USA Sourcing, L.P.	0.000000	Honeywell International Inc.	0.414617	Cisco Technology, Inc.	0.006154

Subgraph analysis and centrality measures:

- **Top market shares =/= cutting-edged techniques.**
  - Companies with highest market shares are not companies who have more citation and more innovation of cloud computing techniques.
  - It implies that middle and small size companies like Symantec are more expertized in cloud computing but its business is not competitive as the big-name companies like Amazon and Google.

WORD CLOUD

Patents Raw Data

Extract abstract and title

Focus on the cloud computing related abstracts

Text Normalization Pipeline

- Tokenization
- Remove stop words / punctuation
- Lemmatization
- POS: NN + VB

Word counts, distribution

Visualization

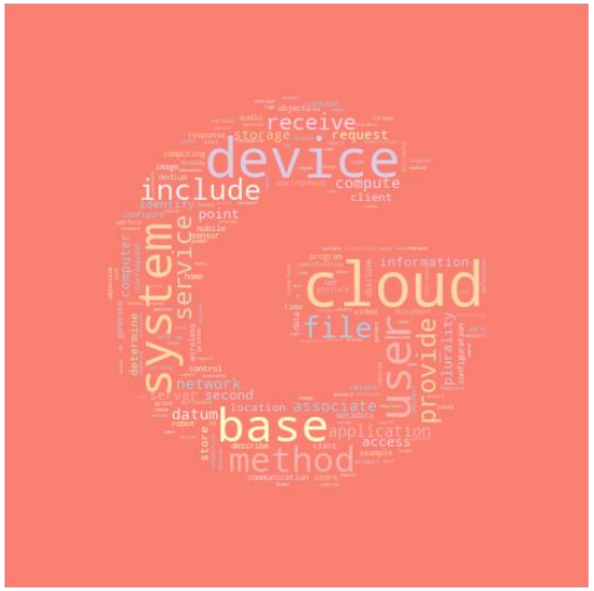


Figure 1. Google patents – word cloud

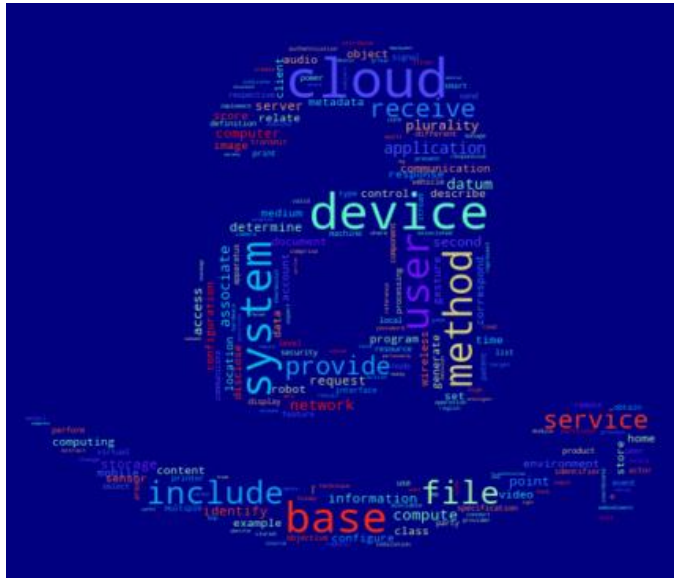


Figure 2. Amazon patents – word cloud

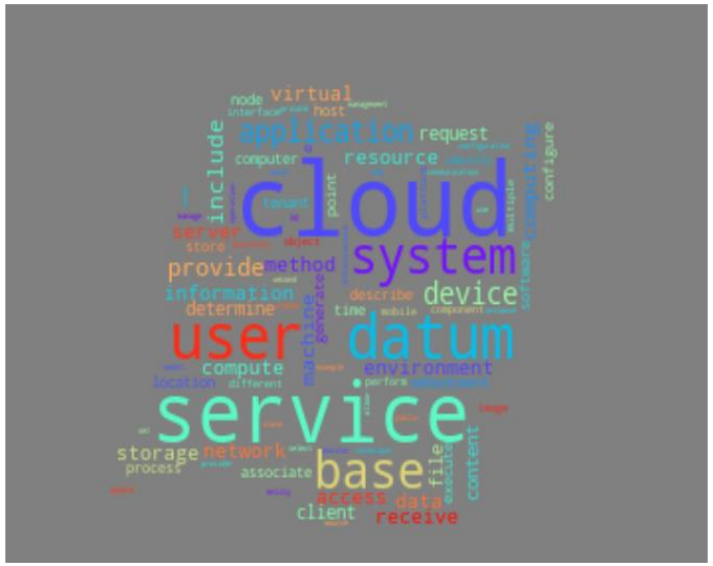
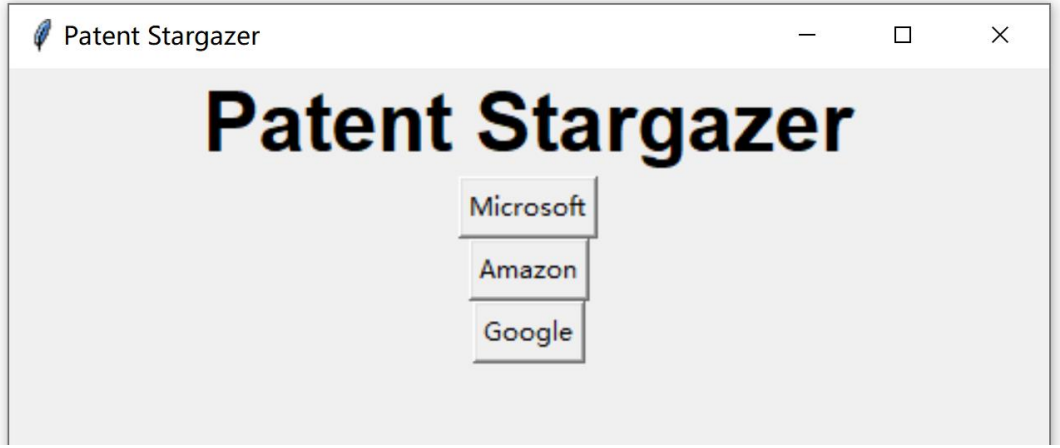


Figure 3. Microsoft patents – word cloud

USER INTERFACE & FUNCTIONAL DEMONSTRATION

- We build a visualization display interface with the GUI framework tkinter and PIL.
- Tkinter has several strengths. 1) It's **cross-platform**, so the same code works on Windows, macOS, and Linux. 2) It's light-weight and already built into the Python package.
- A main page where users can click on each button to quickly view each company and three sub-pages.
- Let's play!



REFERENCES

- Mary Ellen Mogee (1991) Using Patent Data for Technology Analysis and Planning, Research-Technology Management, 34:4, 43-49, DOI: 10.1080/08956308.1991.11670755
- Muro, M., Maxim, R., & Whiton, J. (2019). Automation and artificial intelligence: How machines are affecting people and places.

THANK YOU!

FEEL FREE TO REACH OUT TO ANY OF US IF THERE ARE QUESTIONS, THE GROUP MEMBERS ARE:

CHUQING ZHAO

YI QIAN

YINGXUAN LIU