

US Innovation: The SBIR-NIH-USPTO Axis

Tamar Schaap, Guruprasad Parasnis, Caden Stewart

March 18, 2025

1. Challenge
2. Data Sources
3. Individual Analysis
 - a. SBIR
 - b. NIH
 - c. USPTO
4. Combined Analysis
5. Conclusions

Index & Content

- Economists estimate that 50% of Gross Domestic Product (GDP) can be attributed to innovation, which makes it essential for nations to support and cultivate it. [1] A good proxy variable for innovation is patents, as each one represents a novel idea or invention.
- One way the US seeks to promote innovation is through grants for groups to develop and test ideas.
- Two key vehicles for this are Small Business Innovation Research (SBIR) grants, and National Institute of Health (NIH) grants.

Background

How can we improve the disbursement of SBIR and NIH grants to maximize
patent approvals and national innovation levels?

Challenge

SBIR

- **Size:** Thousands of awards annually
- **Date range:** Since 1982
- **Key fields:** Award amount, project title, agency, company details
- **Focus:** Funding for small business R&D with commercialization potential
- **Source:** [SBIR.gov](https://www.sbir.gov) [2]

NIH

- **Size:** Millions of funded projects
- **Date range:** Since 1985 (digitally available)
- **Key fields:** Grant ID, funding amount, project abstract, PI details
- **Focus:** Biomedical and health-related research funding
- **Source:** [NIH RePORTER](https://reporter.nih.gov) [4]

USPTO

- **Size:** Over 13M patent applications, 1M+ PCT applications
- **Date range:** Since 1790 (modern databases from 1976)
- **Key fields:** Patent number, assignee, claims, filing/grant date
- **Focus:** Patent and trademark datasets for IP research
- **Source:** [USPTO.gov](https://www.uspto.gov) [3]

Data Sources

SBIR

- **Size:** ~200,000 awards
- **Date range:** 2023
- **Uploaded to Neo4j:** ~16,000 awards

NIH

- **Size:** ~85,000 funding records
- **Date range:** 2023

USPTO

- **Size:** ~800,000 applications
- **Date range:** 2023-2024
- **Uploaded to Neo4j:** ~200,000 applications

Data Sources - Project Specifics

Individual Analysis

SBIR

- Image shows the schema for the SBIR awards table

```
CREATE table if not exists award_data
(company_name TEXT,
  Award_Title TEXT,
  Agency TEXT,
  Branch varchar(255),
  Phase varchar(20),
  Program varchar(20),
  Agency_Tracking_Number Varchar(50),
  Contract Varchar(50),
  Proposal_Award_Date varchar(20),
  Contract_End_Date varchar(20),
  Solicitation_Number varchar(20),
  Solicitation_Year varchar(20),
  Solicitation_Close_Date varchar(20),
  Proposal_Receipt_Date varchar,
  Date_of_Notification varchar,
  Topic_Code varchar(20),
  Award_Year varchar(20),
  Award_Amount varchar(20),
  Duns varchar(20),
  HUBZone_Owned char(1),
  Socially_and_Economically_Disadvantaged char(1),
  Women_Owned char(1),
  Number_Employees int,
  Company_Website varchar(100),
  Address1 TEXT,
  Address2 TEXT,
  City varchar(50),
  State varchar(2),
  Zip varchar(20),
```


SBIR Analysis - Postgres

Queries Written:

- Extracted principal investigators from SBIR awards.
- Linked SBIR records to companies for organizational insights.
- Merged with NIH and USPTO to find cross-schema common entities.

Insights Drawn:

- Some researchers receive funding from both SBIR and NIH.
- Identified companies awarded multiple SBIR grants, signaling strong R&D potential.
- Name inconsistencies (middle initials, abbreviations) impacted result accuracy.

Key Inferences:

- Investigators receiving both SBIR & NIH grants indicate high research impact.
- Companies with multiple SBIR grants likely have sustained innovation funding.
- Needed `STRING_TO_ARRAY ()` to separate multiple names in a single field.

NIH

```
CREATE TABLE nih_proj (  
  APPLICATION_ID INTEGER PRIMARY KEY,  
  ACTIVITY VARCHAR(10),  
  ADMINISTERING_IC VARCHAR(10),  
  APPLICATION_TYPE INT,  
  ARRA_FUNDED CHAR(1),  
  AWARD_NOTICE_DATE DATE,  
  BUDGET_START DATE,  
  BUDGET_END DATE,  
  CFDA_CODE INTEGER,  
  CORE_PROJECT_NUM VARCHAR(50),  
  ED_INST_TYPE VARCHAR(255),  
  OPPORTUNITY_NUMBER VARCHAR(255),  
  FULL_PROJECT_NUM VARCHAR(50),  
  FUNDING_ICs VARCHAR(255),  
  FUNDING_MECHANISM VARCHAR(50),  
  FY INTEGER,  
  IC_NAME VARCHAR(255),  
  NIH_SPENDING_CATS VARCHAR(255),  
  ORG_CITY VARCHAR(255),  
  ORG_COUNTRY VARCHAR(255),  
  ORG_DEPT VARCHAR(255),  
  ORG_DISTRICT INTEGER,  
  ORG_DUNS varchar(30),  
  ORG_FIPS VARCHAR(10),  
  ORG_IPF_CODE INTEGER,  
  ORG_NAME VARCHAR(255),  
  ORG_STATE VARCHAR(10),  
  ORG_ZIPCODE VARCHAR(20),  
  PHR TEXT,
```

NIH Analysis - Postgres

Queries Written:

- Extracted Principal Investigators (PIs) from NIH projects.
- Linked NIH projects to their funding sources and institutions.
- Cross-checked NIH investigators against SBIR & USPTO to find overlaps.

Insights Drawn:

- Some NIH PIs also received SBIR funding, hinting at commercial applications.
- Overlaps between NIH and USPTO suggest research-to-patent transitions.
- NIH projects often list multiple PIs, requiring UNNEST () for proper analysis.

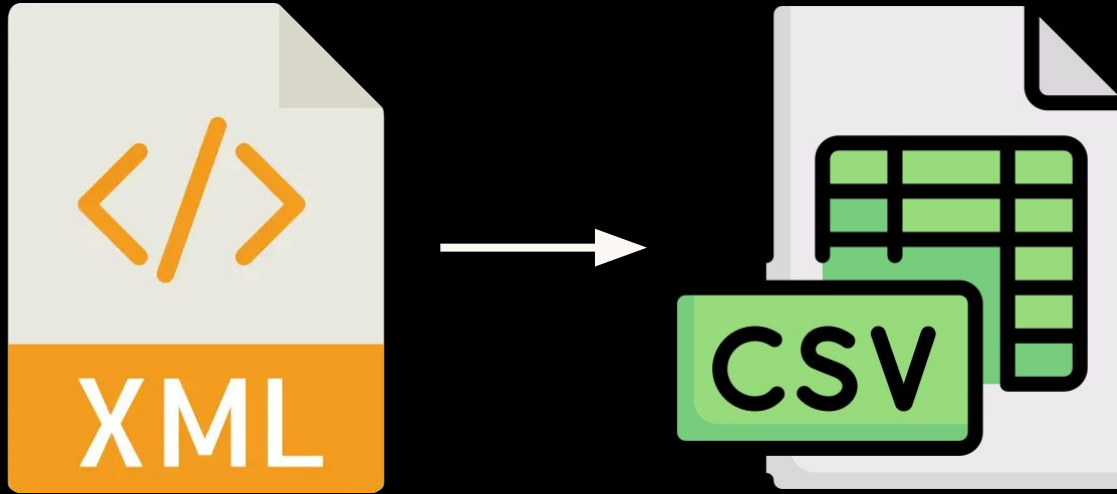
Key Inferences:

- NIH-funded researchers filing patents indicate successful tech translation.
- Finding common investigators across schemas helps track innovation impact.
- Required handling of JSON fields to extract structured insights.

USPTO

- Detailed analysis combining USPTO data with the previous SBIR and NIH data to gain insights
- Data derived from xml files after extracting relevant fields

USPTO Bulk Upload facility



Data Preparation

Extracted fields:

- "publication_country"
- "publication_doc_number"
- "publication_kind"
- "publication_date"
- "grant_length"
- "application_country"
- "application_doc_number"
- "application_date"
- "applicant_organization"
- "applicant_location"
- "organization_residence_country"
- "assignee_organization"
- "assignee_location"
- "inventor"
- "inventor_location"
- "agent_organization"
- "agent_location"
- "invention_title"
- "primary_examiners"
- "assistant_examiners"
- "reference_name"
- "reference_date"
- "reference_country"
- "reference_categories"
- "file_name"

USPTO Analysis - Postgres

Queries Written:

- Extracted inventors and their associated patents.
- Identified patents linked to the same company/organization.
- Found inventors appearing across multiple patents.

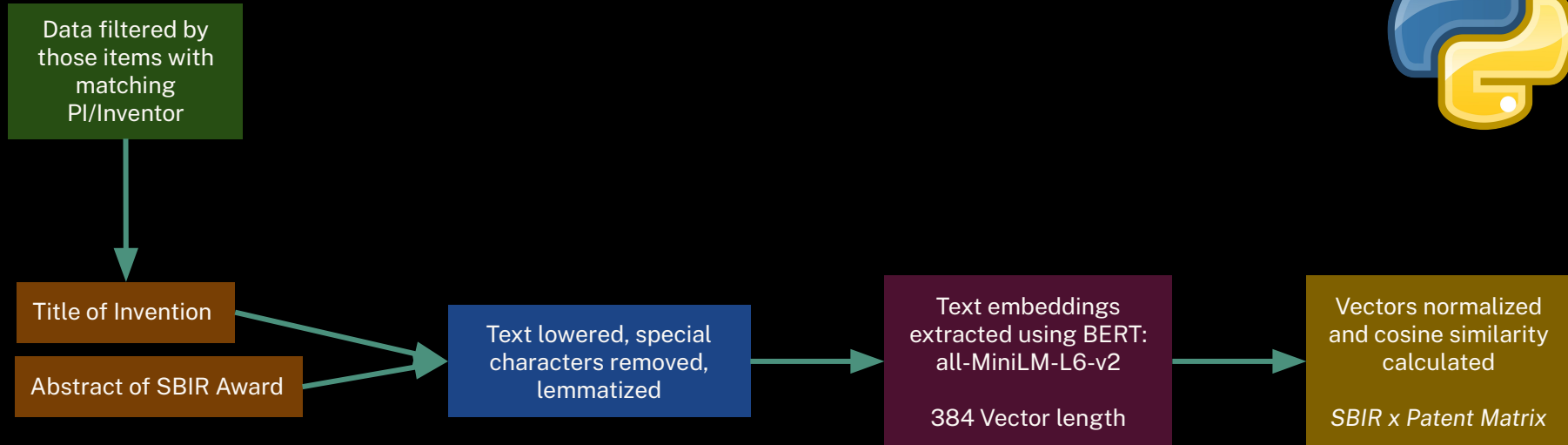
Insights Drawn:

- Some inventors appear on multiple patents, indicating active researchers.
- Companies with multiple patents suggest strong innovation pipelines.
- Issues with name formatting (e.g., multiple names in one field) required data cleaning.

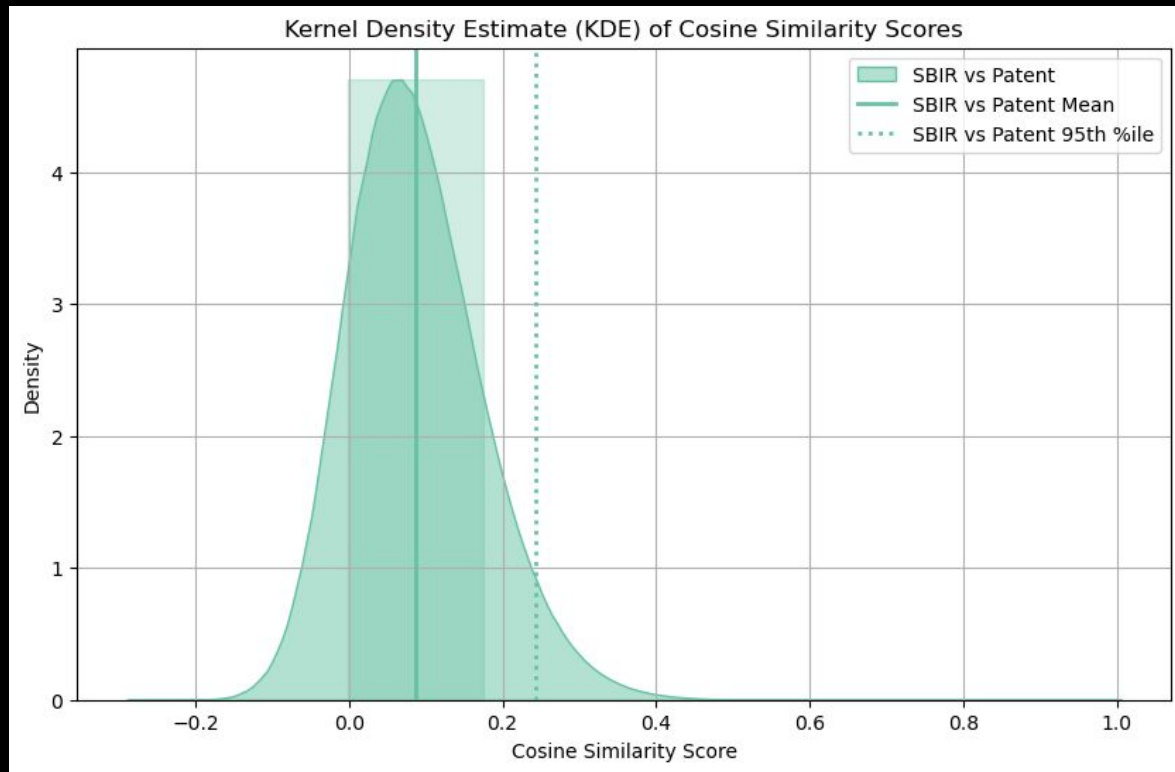
Key Inferences:

- High patent count per inventor could indicate subject matter expertise.
- Overlapping inventors in different companies may suggest industry collaboration.
- Required UNNEST () to handle multiple inventors correctly.

Combined Analysis



Patent - SBIR Award Comparison



**188,060,457 total
similarity scores**

**~1000 similarity
scores > 0.6**

Cosine similarity distribution

Top Matches: SBIR - Patent

Cosine similarity: 0.77

SBIR Abstract:

LM Group Holdings Inc. (LMGH) partnering with Fabrisonic LLC is proposing a program to investigate manufacturing of amorphous metal alloy laminate composites and **cladding of metallic surfaces** by using **ultrasonic additive manufacturing** (UAM), a solid-state 3D metal printing technology.

Patent:

Applicant Organization: LM Group Holdings, Inc., Fabrisonics LLC

Title: Ultrasonic additive manufacturing of cladless amorphous metal products

Cosine similarity: 0.77

SBIR Abstract:

nLight proposes to develop and commercialize an innovative kilo-watt class fiber laser pump combiner which is especially suited for counter-pumping architectures that is indispensable for various coherent combining techniques.... nLight will demonstrate kilowatt-class **amplifier** using this pump combiner and its **proprietary active fiber**.

Patent:

Applicant Organization: NLIGHT INC.

Title: Cladless fiber for fiber laser pump and combiner

Cosine similarity: 0.74

SBIR Abstract:

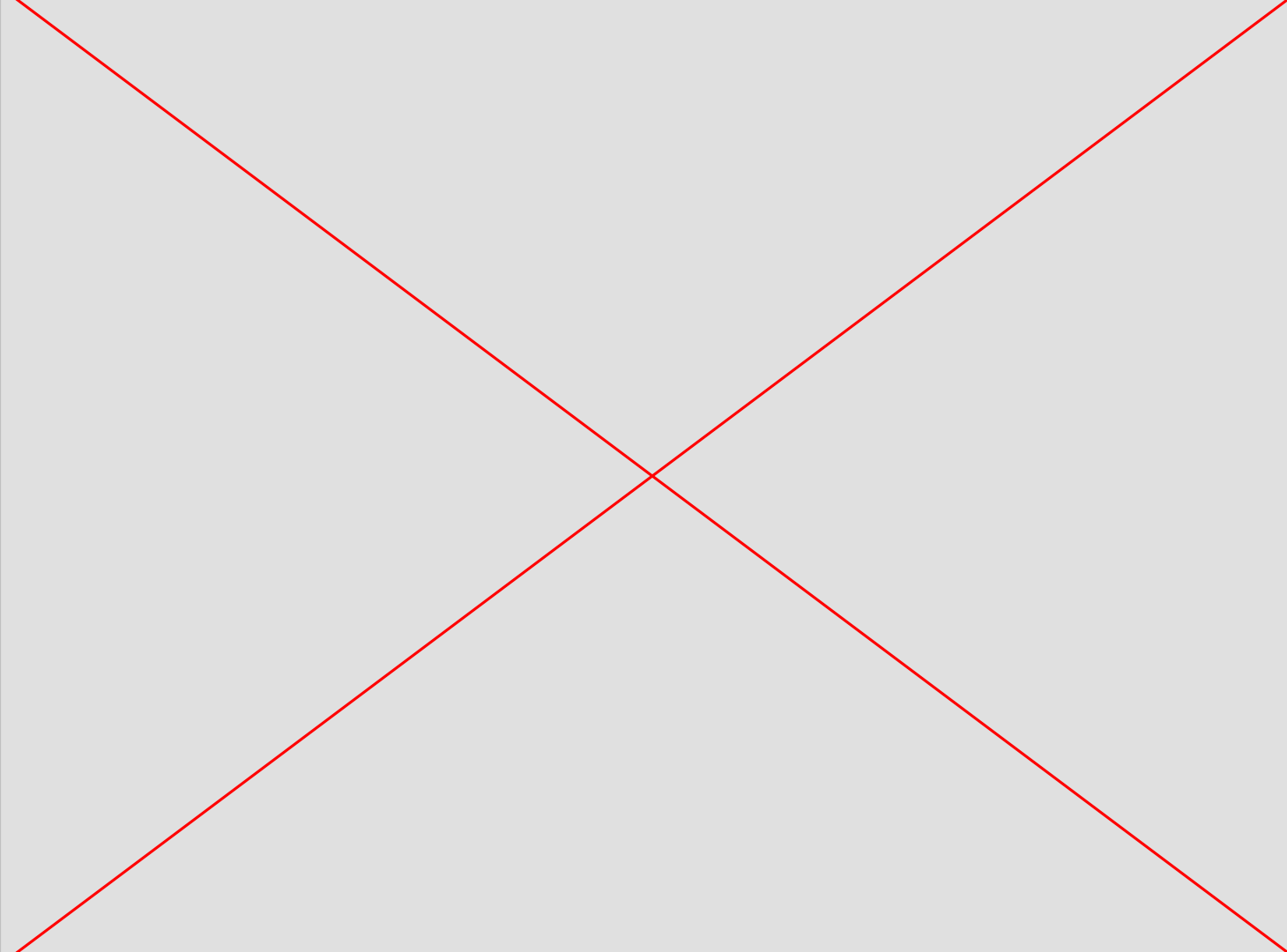
Flexible silica aerogel composites, a class of super-insulation material recently developed by **Aspen Aerogels**, has not been utilized before in **high temperature** TPS designs. Thermophysical characterization data will be collected...for **high-temperature** durable, oxidatively stable, flexible **aerogel composites** at different densities, pressures and temperatures....The aerogels will be compatible with all **high temperature** capable face-skin materials.

Patent:

Applicant Organization: Aspen Aerogels

Title: Aerogel compositions for high temperature applications

Neo4j



Organization Networks



Individual Networks



Challenge - Conclusion

Government grant disbursement and patent application approvals demonstrate strongly connected network patterns, which could be used to improve innovation to funding ratio in the United States.

[1] McKinney, P. (2023, January 1). Innovation Fuels 50% of GDP Growth per Economists. *The Innovators Network*.

<https://theinnovators.network/innovation-fuels-50-of-gdp-growth-per-economists/>

[2] <https://legacy.www.sbir.gov/data-resources>

[3] <https://developer.uspto.gov/data>

[4] <https://report.nih.gov/nihdatabook/>

References