UC San Diego

LinkedIn Unleashed

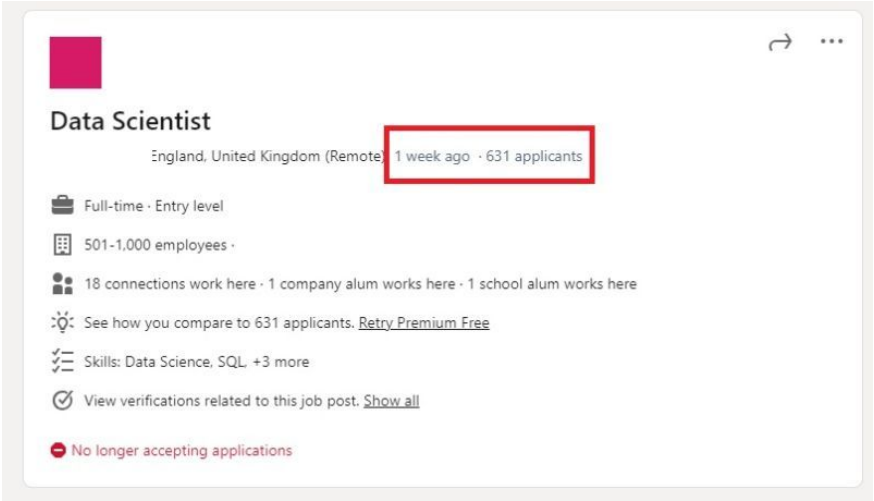
Getting Users to their Future Jobs

Abdullah Ashfaq, Aagaaz Sayed, Jacob Ryan

DSC 202 Final Project | December 12, 2023

The Background

The entry-level listing asking for 3+ years of experience



Applying for jobs is hard

UC San Diego

Optimize your search using LinkedIn Unleashed

- Built from datasets scraped directly from LinkedIn

- Find better …
  - Connections
  - Groups
  - Advice
  - Jobs

UC San Diego

Goals

I’m looking to…

Expand my network strategically

Find the types of companies that need employees

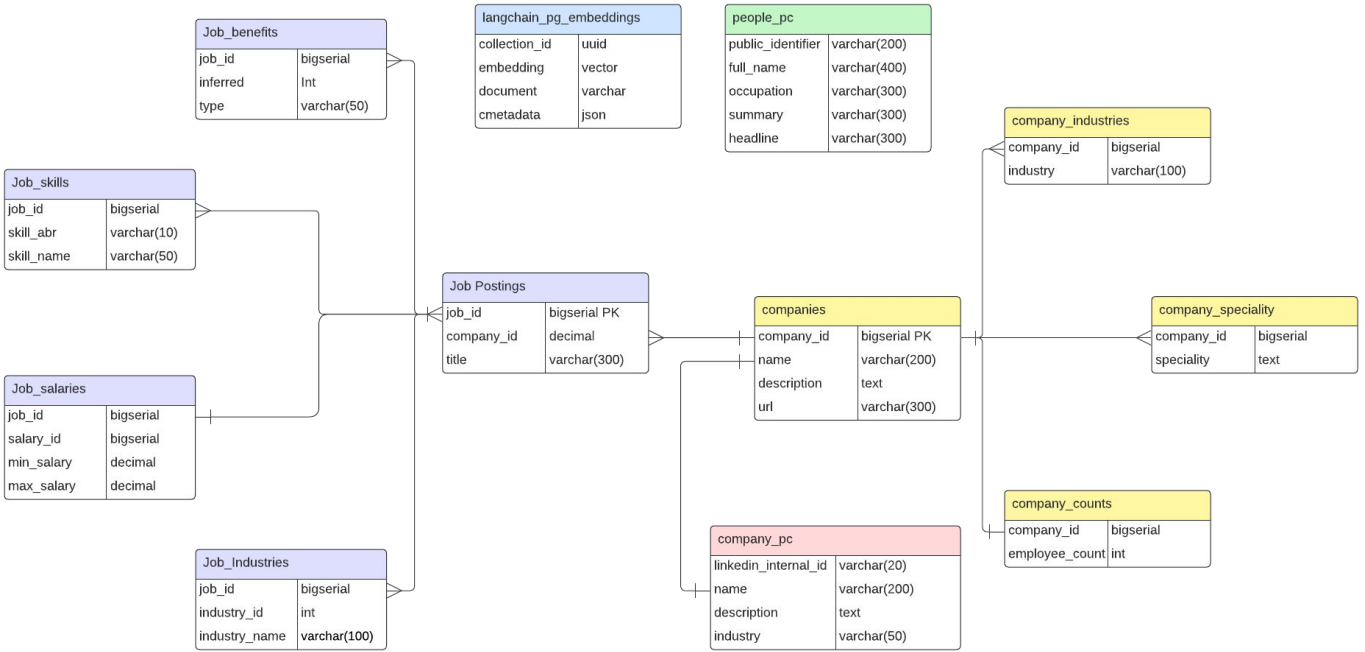
Understand the skills required to work at a top company

How to best connect with individuals at my desired company

Find jobs that are actually similar to my current role

The Data

PostGres (ERD)



PostgreSQL

UC San Diego

PostGres

Job Postings

	job_id	company_id	title	description	formatted_experience_level
0	3757940104	553718.0	Hearing Care Provider	Overview\n\nHearingLife is a national hearing ...	Entry level
1	3757940025	2192142.0	Shipping & Receiving Associate 2nd shift (Beav...	Metalcraft of Mayville\nMetalcraft of Mayville...	None
2	3757938019	474443.0	Manager, Engineering	\nThe TSUBAKI name is synonymous with excellen...	None
3	3757938018	18213359.0	Cook	descriptionTitle\n\n Looking for a great oppor...	Entry level
4	3757937095	437225.0	Principal Cloud Security Architect (Remote)	Job Summary\nAt iHerb, we are on a mission to ...	Mid-Senior level
5	3757937037	13727.0	Territory Manager - New Haven	Location: Remote, CT, United States of America...	Mid-Senior level
6	3757937004	10515052.0	Auto Body Technician	Company: Gerber Collision & Glass\n\nWELCOME T...	Entry level
7	3757936167	2915.0	ACME D8- Asst Store Director (ASD) Sussex, NJ	The First Assistant Store Director is actively...	Mid-Senior level
8	3757936097	18213359.0	Dishwasher	descriptionTitle\n\n $2,000 Sign-on Bonus Guar...	Entry level
9	3757932736	73013724.0	Sales Manager	Position Summary: Our Sales Manager has managi...	Mid-Senior level

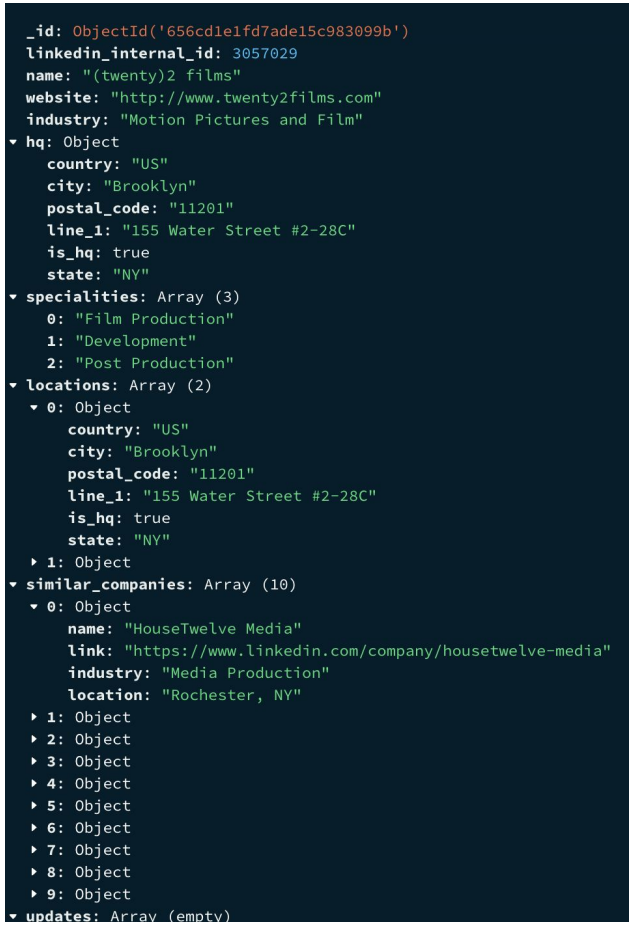
PostGres

Companies

	company_id	name	country	url
0	1009	IBM	US	https://www.linkedin.com/company/ibm
1	1016	GE HealthCare	US	https://www.linkedin.com/company/gehealthcare
2	1021	GE Power	US	https://www.linkedin.com/company/gepower
3	1025	Hewlett Packard Enterprise	US	https://www.linkedin.com/company/hewlett-packa...
4	1028	Oracle	US	https://www.linkedin.com/company/oracle
5	1033	Accenture	IE	https://www.linkedin.com/company/accenture
6	1038	Deloitte	OO	https://www.linkedin.com/company/deloitte
7	1043	Siemens	DE	https://www.linkedin.com/company/siemens
8	8296	Aerojet Rocketdyne	US	https://www.linkedin.com/company/aerojet-rocke...
9	1044	PwC	GB	https://www.linkedin.com/company/pwc

MongoDB

Companies



mongoDB®

UC San Diego

MongoDB

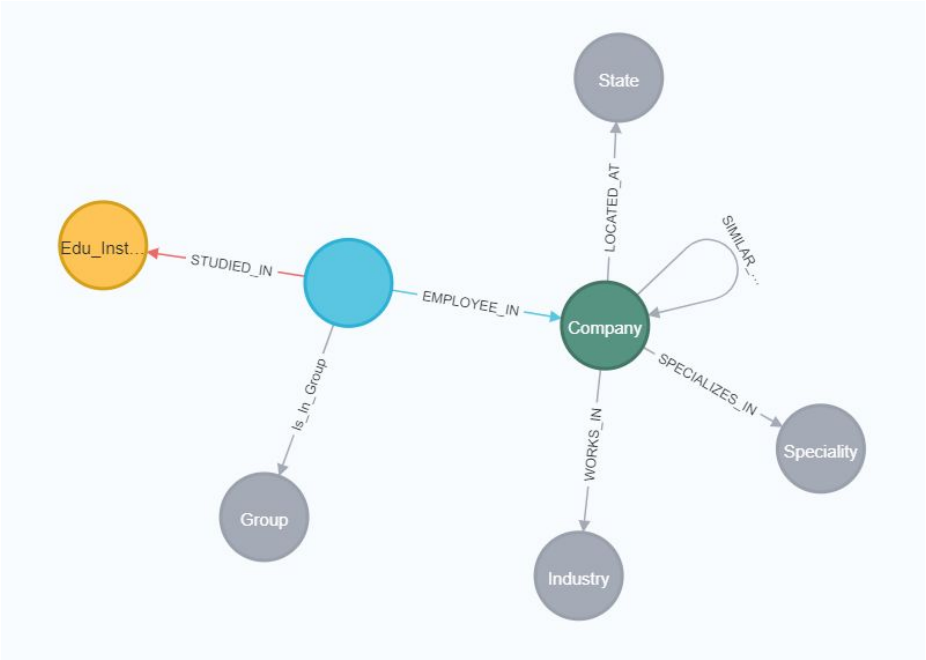
People


```
_id: ObjectId('656cd1dcfd7ade15c982e2d6')
public_identifier: "a-arone"
full_name: "Aaron Jones"
▾ experiences: Array (2)
  ▾ 0: Object
    ▸ starts_at: Object
      ends_at: null
      company: "Dish Network"
      company_linkedin_profile_url: "https://www.linkedin.com/company/dish-network"
      title: "Retention Specialist"
      description: null
      location: "Roseland NJ"
      logo_url: "https://media-exp1.licdn.com/dms/image/C560BAQE0KxtLULI5lQ/company-log…"
  ▸ 1: Object
▾ education: Array (1)
  ▾ 0: Object
    ▸ starts_at: Object
    ▸ ends_at: Object
      field_of_study: "Marketing"
      degree_name: "Bachelor of Science - BS"
      school: "Cheyney University of Pennsylvania"
      school_linkedin_profile_url: "https://www.linkedin.com/school/cheyney-university-of-pennsylvania/"
      description: null
      logo_url: "https://media-exp1.licdn.com/dms/image/C510BAQFqhBmXswrzww/company-log…"
▸ languages: Array (empty)
```


mongoDB®

Neo4j

General Schema



Overview

Node labels

* (7) | Industry (1) | Group (1) | Company (1) | Speciality (1) | State (1) | Edu_Institution (1) | People (1)

Relationship Types

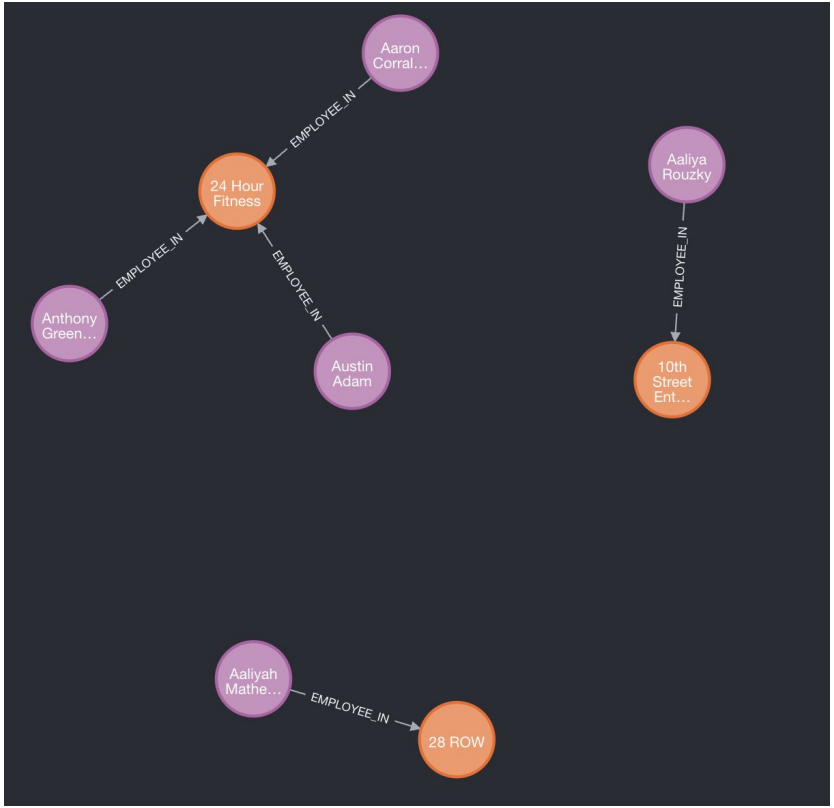
* (7) | SIMILAR_TO (1) | STUDIED_IN (1) | LOCATED_AT (1) | WORKS_IN (1) | EMPLOYEE_IN (1) | SPECIALIZES_IN (1) | Is_In_Group (1)

Displaying 7 nodes, 7 relationships.

UC San Diego

Neo4j

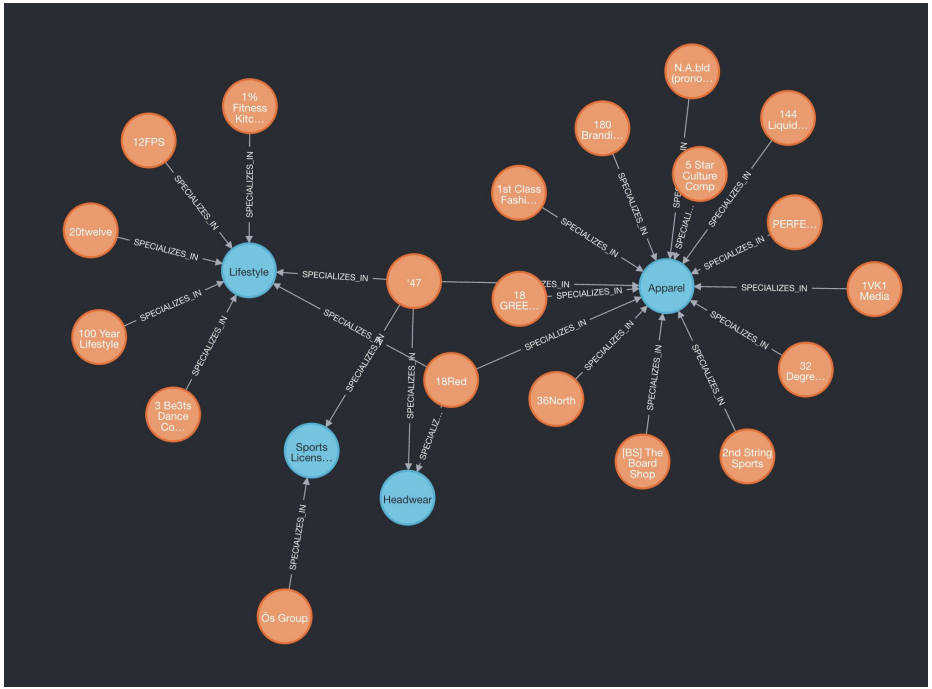
People Employed by Companies



UC San Diego

Neo4j

Companies Specializes In Industry



ChromaDB

A Vector Database

Vectorized Job Postings

```
["""Data Science grad with 2 years of experience in the entire data science life cycle.
Have multiple internships and projects in computer vision and natural language processing. Proficient with Python,
 R and querying languages like SQL"""]
```

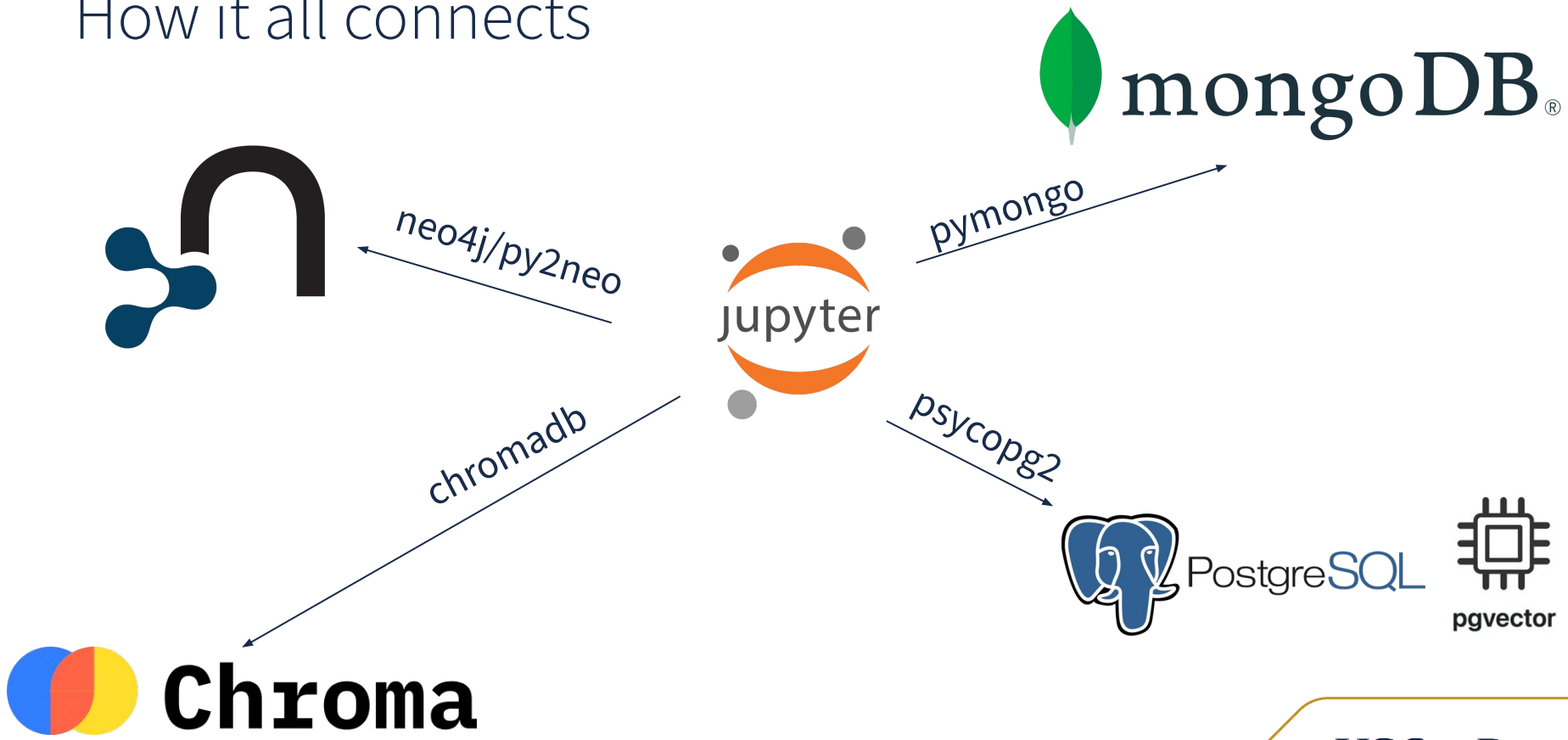
Vectorized User Descriptions

```
[[ 6.73589930e-02  7.83936083e-01  2.70018369e-01  9.58026499e-02
   3.89930516e-01  1.33015066e-01  8.51949275e-01 -2.25674719e-01
   2.49909591e-02  1.85417011e-01  2.15236664e-01 -1.49790362e-01
   1.57957390e-01  1.01382211e-01  8.90305102e-01  2.33439162e-01
   7.86206722e-01  3.00011307e-01 -7.10195661e-01 -3.11727017e-01
   8.51575077e-01 -6.04339428e-02 -1.21499293e-01 -5.63790619e-01
  -2.75132746e-01  1.39450043e-01  1.07049279e-01 -2.37949759e-01
   2.08920732e-01 -2.28108257e-01 -2.22002849e-01 -7.31366128e-02
   1.53638765e-01 -5.19387163e-02 -5.07328771e-02  3.23610961e-01
   2.23328486e-01  7.97795594e-01  9.15637463e-02  2.07635630e-02
  -2.19855204e-01 -2.75381386e-01  2.45380420e-02  2.46632341e-02
  -2.31250256e-01 -5.35810888e-01 -3.87015603e-02 -4.36115486e-04
  -7.61911692e-03 -4.71225560e-01 -4.18657601e-01 -7.82947540e-02
```

Chroma

UC San Diego

How it all connects



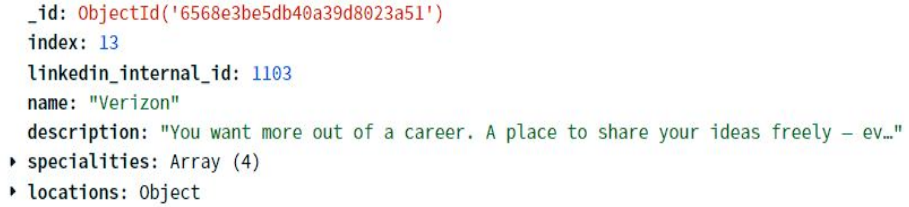
The Process to Employment

Example Case: New to the Field

Task 0

Find a domain

- Maybe you're not sold on any specific type of job
- Maybe any job that is hiring is the right type of job
- Find companies hiring the most and the industries they specialize in!



UC San Diego

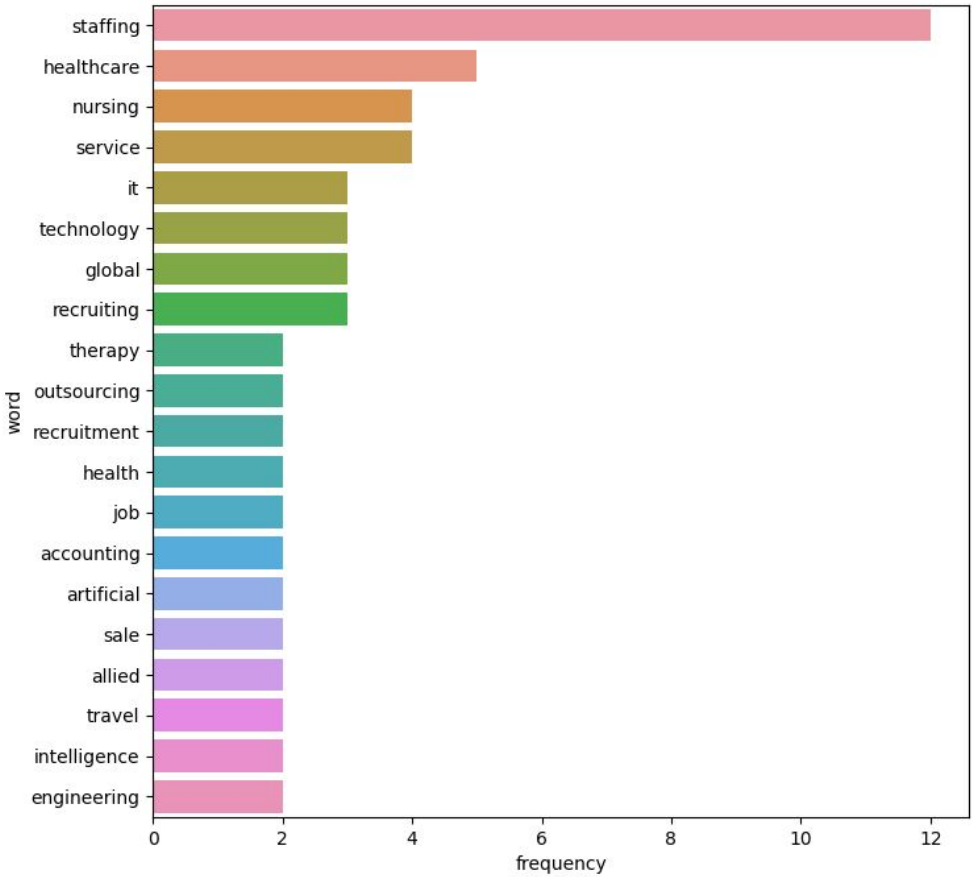
Task 0

How it works

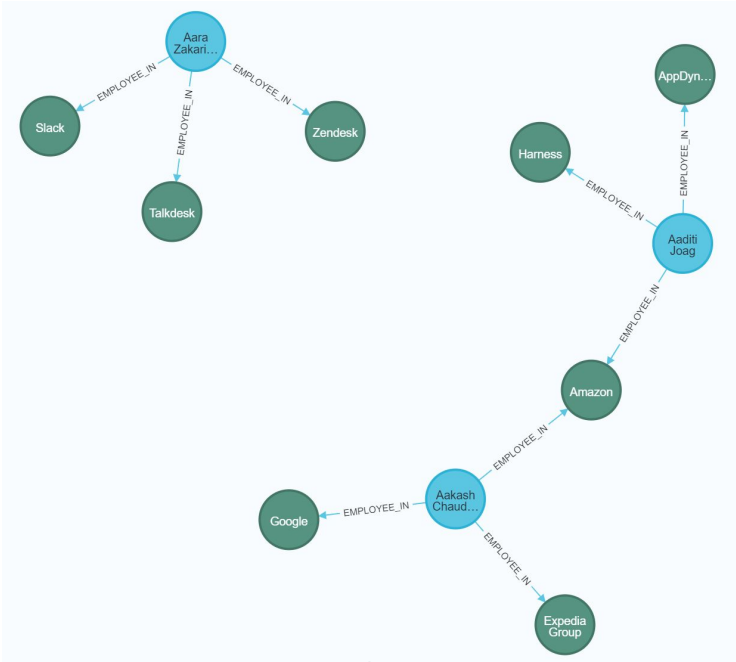
Step 1: Find number of unique job postings per company divided by number of employees at the company using PostGres

Step 2: Get top 10 companies with best ratio

Step 3: Find companies using MongoDB, collect their specializations, and visualize important domains



UC San Diego



Fetch data from PostGres

	full_name	occupation	headline	summary	city	country
0	Aaditi Joag	Software Development Engineer II at Amazon	Software Development Engineer at Amazon	None	San Francisco Bay Area	US
1	Aakash Chaudhary	Software Development Engineer 3 at Expedia Group	SDE 3 at Expedia Group | Former SDE at Amazon ...	Expert Software Development professional with ...	Seattle	US
2	Aara Zakariaei	Healthcare Account Executive at Slack	Account Executive at Slack | NASM Health Coach	None	Denver	US

Task 1

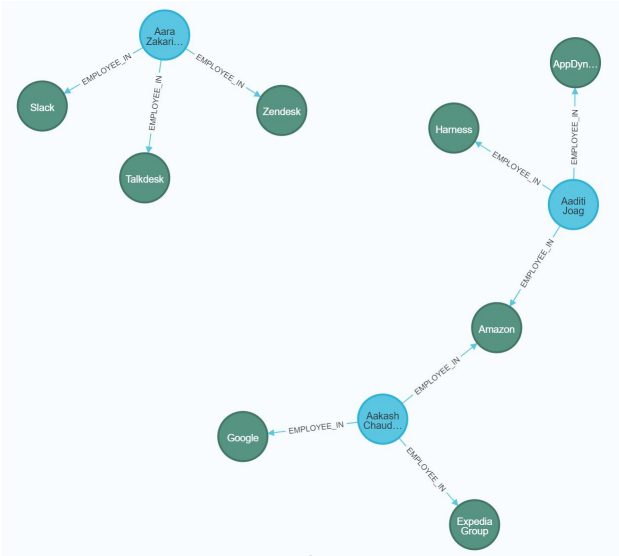
Find a mentor

- You're close to graduation
- You love your field
- You don't know where to get started

So …

- You find a mentor!
- i.e. someone who has significant work experience in your desired domain

UC San Diego



Fetch data from PostGres

How it works

Step 1: Input is target domain e.g Software here

Step 2: Find top 3 people who have worked in most companies from the targeted domain

```
match (c:Company)-[:WORKS_IN]->(i:Industry)
where i.name =~ '(?i).*Software.*'
match (p:People)-[e:EMPLOYEE_IN]->(c)
return p, collect(e) as exp, collect(c) as
comps
order by size(exp) desc limit 3;
```

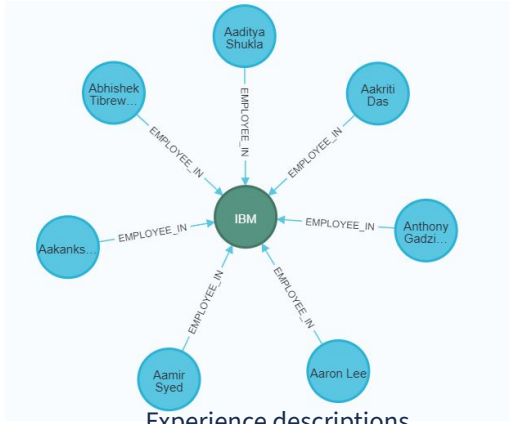
Step 3: Get information about these people from PostgreSQL

	full_name	occupation	headline	summary	city	country
0	Aaditi Joag	Software Development Engineer II at Amazon	Software Development Engineer at Amazon	None	San Francisco Bay Area	US
1	Aakash Chaudhary	Software Development Engineer 3 at Expedia Group	SDE 3 at Expedia Group | Former SDE at Amazon ...	Expert Software Development professional with ...	Seattle	US
2	Aara Zakariaei	Healthcare Account Executive at Slack	Account Executive at Slack | NASM Health Coach	None	Denver	US

Task 2

Research your dream position

- You know where you want to work
- You want to know what people in this position actually do
- e.g. you are interested in a data position at IBM



Experience descriptions

pgvector

NLP
(spacy, nltk)

UC San Diego

Step 1: Find people working in a particular role and in a particular company

```
match (c:Company)-[r:EMPLOYEE_IN]-(p)
where r.title =~ '(?i).*data.*' and c.name='IBM'
return *;
```

Step 2: Extract description of the work that was done at these positions by these people

Step 3a: Use these descriptions as context to ask LLM some questions about these positions (Retrieval Augmented Generation)

embedding	document
[-0.02651085,-0.0014169131,0.014719803,-0.0076880706,0.0894172…	Designed and implemented
[-0.009534597,0.0013303668,0.008652128,0.007419174,0.049020953…	- Responsible for creati
[-0.009761419,-0.018089341,-0.008261888,-0.019824786,0.0479405…	- Collaborates with key
[0.017789382,0.007855268,-0.013564722,-0.01738247,0.0500754,0.…	- Responsible for creati
[-0.0025171842,0.00546929,-0.01428999,-0.011579876,0.054069903…	10/2003 - 04/2004: Globa
[-0.01566332,0.015311546,0.0020318169,0.011340049,0.051750876,…	09/1995 - 01/1996: Call
[-0.052344844,-0.013785417,0.009193125,0.032965314,0.023065548…	* Automate the end-to-en
[-0.021815913,0.0066488856,0.008292637,-0.010481652,0.0494375…	* Design data models for

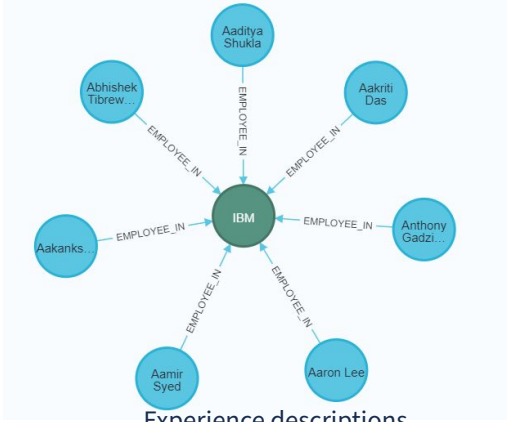
```
- **Technologies:**
  - Python
  - Node.js
  - SQL
  - JS (JavaScript)
  - DataStage
```

Step 3b: Use spacy to find named entities in the descriptions

```
nlp = spacy.load('en_core_web_sm')
text = ';'.join(description_lst)
doc = nlp(text)
doc.ents[:10]
```

✓ 1.3s

```
(Watson Cloud,
 Node.js.;07/2019,
 Incentives,
 SQL,
 JS,
```

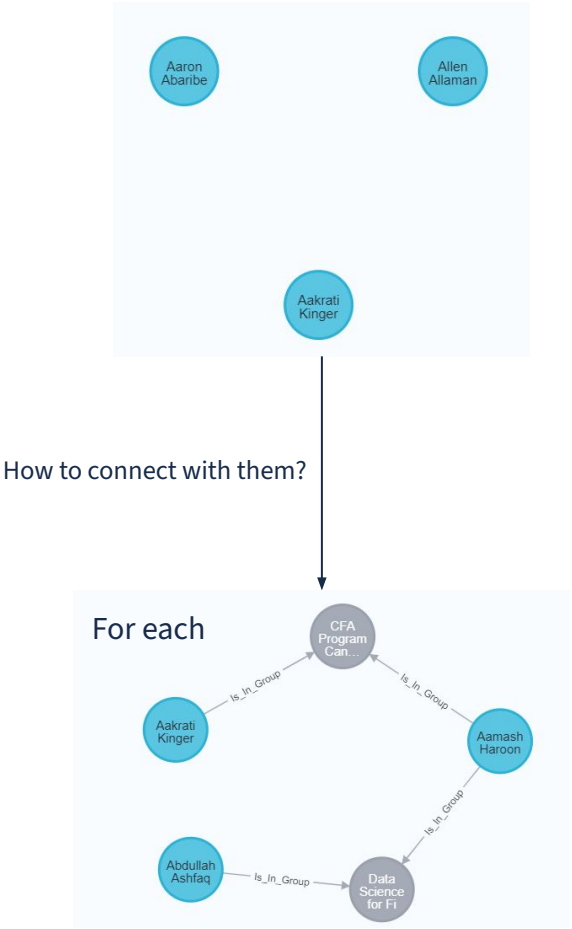


Experience descriptions

pgvector

NLP
(spacy, nltk)

UC San Diego



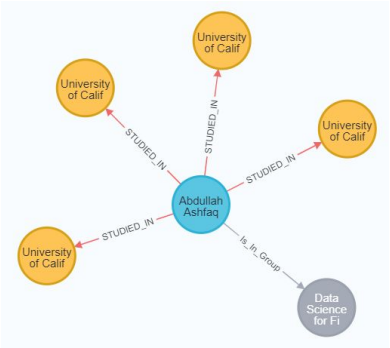
Task 3

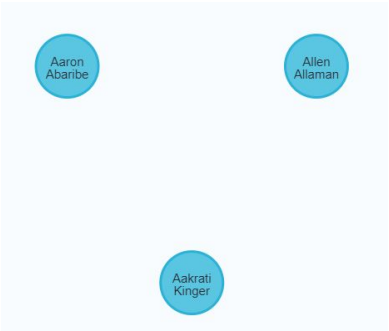
Find a connection path

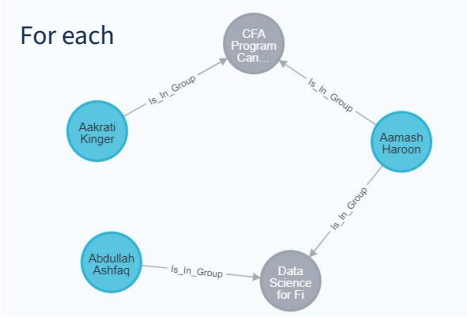
- You’re qualified for the position
- You want your best chance of success: a referral
- How do you find the right person?

- Find the most sensible connection!

UC San Diego







How it works

Step 1: Create a Profile and Insert in Graph

```
Match (a:People {full_name: "Abdullah Ashfaq",
public_identifier: "aashfaq11"}), (u:Edu_Institution), (g:Group)
where u.name =~ '(?i).*University of California.*San Diego.*'
and g.name = 'Data Science for Finance and Economics'
merge (a)-[:STUDIED_IN]->(u)
merge (g)<-[:Is_In_Group]-(a)
```

Step 2. Search shortest distance to Google employees and filter paths on the ones which don't contain "Company" node

```
MATCH (end:People)-[:EMPLOYEE_IN]->(c:Company)
WHERE c.name =~ 'Google'
WITH collect(end) as endPersons
Match (start:People {full_name: "Abdullah Ashfaq",
public_identifier: "aashfaq11"})
CALL apoc.path.subgraphNodes(start, {relationshipFilter:'',
endNodes:endPersons, limit:3, labelFilter:'-Company'}) YIELD
node
RETURN node
```

Step 3: For each of the people identified, find the shortest path to connect to them

```
Match p=shortestPath((a:People {full_name: "Abdullah Ashfaq",
public_identifier: "aashfaq11"})-[*]-(p1:People
{public_identifier:'aakratikinger'}))
    return p;
```

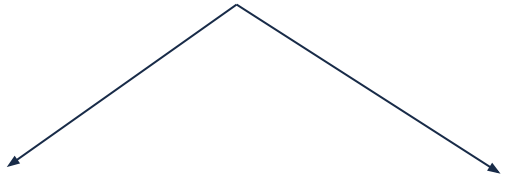

UC San Diego

Task 4

Know companies that hire you!

- You know what kind of work you want to do and have a short summary about yourself prepared.
- You want to know about and analyse jobs you could apply to right now.
- You also want to know which similar companies to keep an eye on.

```
query_string = """Data Science grad with 2 years of
experience in the entire data science life cycle.
Have multiple internships and projects in computer
vision and natural language processing.
Proficient with Python, R and querying
languages like SQL."""
```



Job Analysis

Graph of Companies

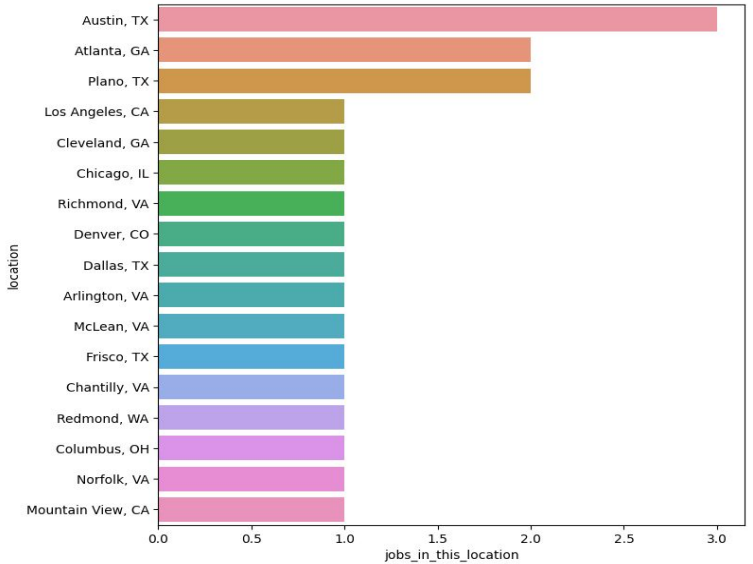
Step 1: Find top 25 job postings similar to the given profile from vectorized job postings.

```
results = collection.query(
    query_texts=[query_string],
    n_results=25
)

results
```

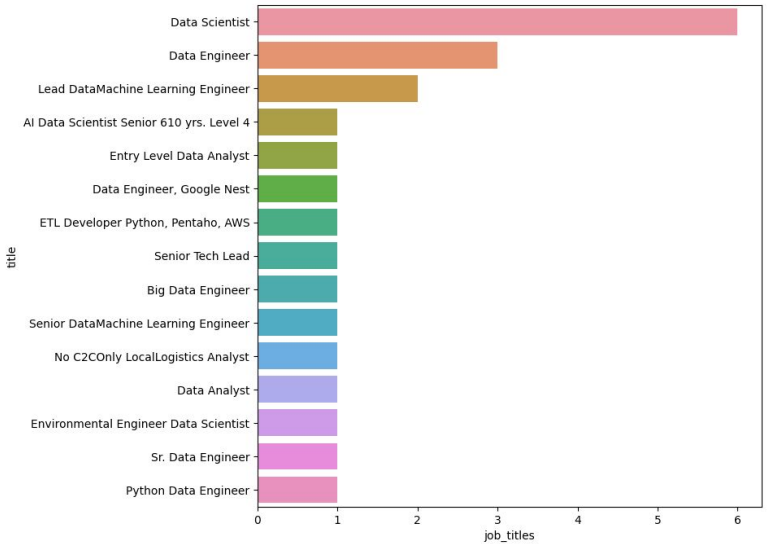
Chroma

Step 2: Location Analysis using PostgreSQL

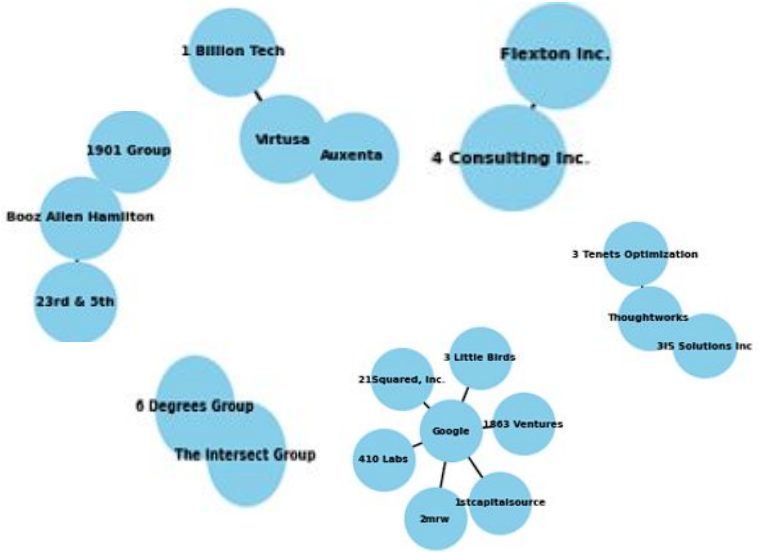


UC San Diego

Step 3: Job Title Analysis.



Step 4: Network Graph of similar companies for potential applications



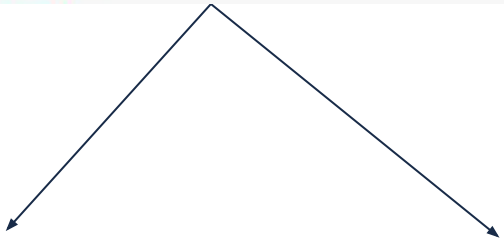
UC San Diego

Task 5

Know people who could refer you and the proven set of skills to succeed in your domain!

- You know what kind of work you want to do and have a short summary about yourself prepared.
- You want to know about people working in similar domains and connect with them for a potential referral.
- You also want to know what the proven set of skills required to succeed in this domain are.

```
query_string = """Data Science grad with 2 years of
experience in the entire data science life cycle.
Have multiple internships and projects in computer
vision and natural language processing.
Proficient with Python, R and querying
languages like SQL."""
```



Potential Referrals

Proven set of skills

Step 1: Find top 50 people who have a similar profile to you on vectorized summary on people data.

```
results = collection.query(
    query_texts=[query_string],
    n_results=50
)

results
```

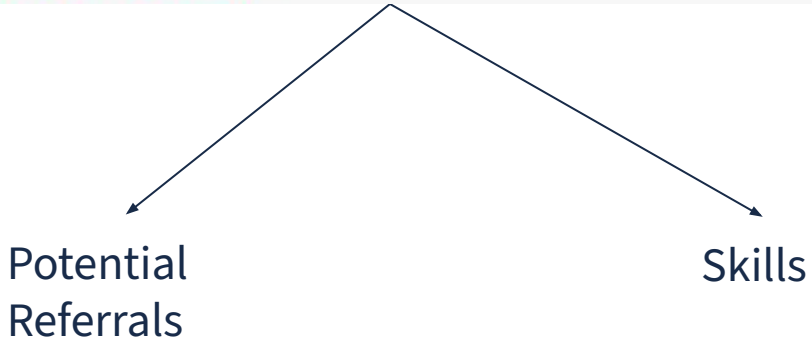
Step 2: Top skills highlighted using text data and word cloud.

UC San Diego

Step 3: Potential companies that you can get a referral from

UC San Diego

```
query_string = """Data Science grad with 2 years of
experience in the entire data science life cycle.
Have multiple internships and projects in computer
vision and natural language processing.
Proficient with Python, R and querying
languages like SQL."""
```



Task 5

Connect with similar people and keep an eye out for referrals.

- You always want to be ready for new opportunities
- How do you know a new position/company would be worth it?
- Find people in similar positions at other companies, and ask them for referrals if they're happy there!


```
{'ids': [['Aadit Vyas',
   'Aakriti Gupta',
   'Abdullah Aburomeh',
   'Adam R.',
   'Azaan Barlas',
   'Aabhashree Lamichhane',
   'Aalap M',
   'Aaditya Bhat',
   'Aarohi Mehta',
   'Aakanksha Patil',
   'Aanchal Gosain',
   'Aakanksha Bharde',
   'Aabir Abubaker Kar',
   'Aamir Goriawala',
   'Aaditya Mohapatra',
   'Aaron Carney',
   'Ahmed Jaafar',
   'Aadithya Viswanath Ramasubramaniam',
   'Aafaz Ilahi',
```

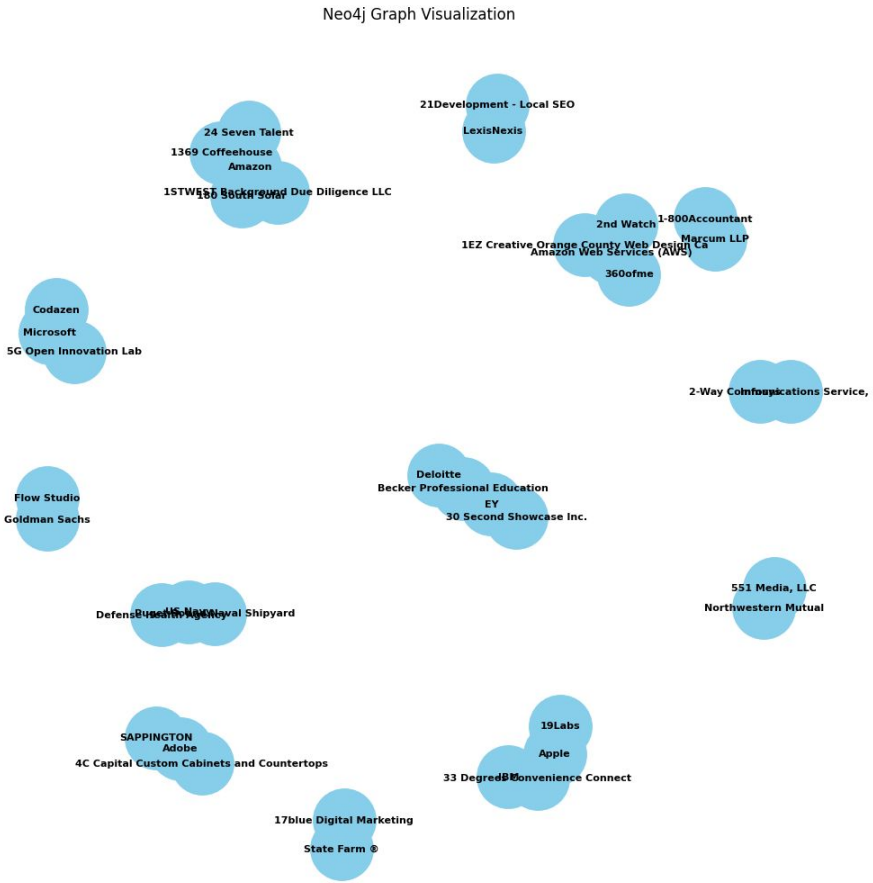

How it works

Step 1: Input your current position description

Step 2: Find others with similar profile using ChromaDB vector search

Step 3: Visualize important similarities between positions

UC San Diego



How it works

Step 4: Identify the companies of similar users using Neo4js

Step 5: Group companies based on similar specializations and visualize

UC San Diego

Putting it all together

1. Find the industries with best job postings to employee count ratio
2. Find People with most connections to companies in these industries
3. How to connect with these people at a company
4. Researching what people at a particular position in a company do
5. Finding the best current job openings and potential future openings based on my profile.
6. Finding people working in the industry in the domains you want to work in who could refer you to various companies.

Good luck on the job search.