

# Smarter RAG Starts with Cleaner Graphs: The Case for Entity Resolution

Clair J. Sullivan, PhD

[clair@clairsullivan.com](mailto:clair@clairsullivan.com)



[github.com/cj2001/  
odsc\\_east\\_2025](https://github.com/cj2001/odsc_east_2025)

# Outline

---

- Terminology: graphs and knowledge and entities, oh my!
- Why is entity resolution (ER) hard
- How to NOT do ER
- Why does ER matter for RAG?
- A worked example
- Conclusion

# Some definitions

---

- Graph
- Knowledge graph (KG)
- Entity resolution (ER)
- Entity-resolved knowledge graph (ERKG)

# What is an entity?

---

- Distinct thing
  - Person
  - Place
  - Organization
  - Object
- Think of them like rows in a table
- Defined by a combination of distinguishing characteristics
  - Some characteristics are more distinguishing than others!

# Considering the following scenarios...

---

Five different people, each with a single transaction

A single person with five different transactions

# Why does it matter?

---

- Counting records is the most basic task of a data scientist
- If you can't count, you can't predict
- If you can't count, you can't quantify
- Enables things like:
  - Good news / bad news
  - Red light / green light
  - To market to them or not
  - Benign / threat

Hot take:

Regex is not the answer



# Names are complicated

---

Johnathan Edward Smith	Jonathan Edward Smith	Jonathon Edward Smith	Jonothan Edward Smith	Jonatan Edward Smith	John Edward Smith	Jon Edward Smith	Johnny Edward Smith
Jonny Edward Smith	Jack Edward Smith	Jono Edward Smith	Nate Edward Smith	Johnathan Ed Smith	Jonathan Ed Smith	Jonathon Ed Smith	Johnathan Eddie Smith
Jonathan Eddie Smith	Jon Eddie Smith	Johnathan Ted Smith	Johnny Ted Smith	Johnathan Teddy Smith	Johnathan Ned Smith	Johnathan E. Smith	Jonathan E. Smith
Jonathon E. Smith	Jon E. Smith	John E. Smith	Jack E. Smith	Jono E. Smith	Johnny E. Smith	J. Edward Smith	J. E. Smith
J.E. Smith	J. Smith	Jon Smith	John Smith	Johnny Smith	Jack Smith	Jono Smith	Eddie Smith
Ted Smith	Teddy Smith	Ned Smith	Nate Smith	Johnathan Edward Smithe	Jonathan Edward Smithe	Jonathon Edward Smithe	Jon E. Smithe
Johnny E. Smithe	Johnathan E. Smyth	Jonathan E. Smyth	Jon Smyth	Johnny Smyth	John Smyth	Jack Smyth	Jono Smyth
J. Edward Smyth	J. E. Smyth	J.E. Smyth	Johnathan Edward Smythe	Jonathan Edward Smythe	Jonathon E. Smythe	Jon Smythe	John Smythe
Johnny Smythe	J. E. Smythe	Ted Smythe	Eddie Smythe	Nate Smythe			

# Addresses are complicated

123 Main, Apt 1A, Las Vegas, NV, 89111	123 Main, #1A, Las Vegas, Nevada, 89111	123 Main St, Apt 1A, Las Vegas, NV, 89111	123 Main St, #1A, Las Vegas, Nevada, 89111	123 Main St., Apt 1A, Las Vegas, NV, 89111	123 Main St., #1A, Las Vegas, Nevada, 89111	123 Main Street, Apt 1A, Las Vegas, NV, 89111	123 Main Street, #1A, Las Vegas, Nevada, 89111
123 Main, Apt 1A, Las Vegas, Nevada, 89111	123 Main, Apartment 1A, Las Vegas, NV, 89111	123 Main St, Apt 1A, Las Vegas, Nevada, 89111	123 Main St, Apartment 1A, Las Vegas, NV, 89111	123 Main St., Apt 1A, Las Vegas, Nevada, 89111	123 Main St., Apartment 1A, Las Vegas, NV, 89111	123 Main Street, Apt 1A, Las Vegas, Nevada, 89111	123 Main Street, Apartment 1A, Las Vegas, NV, 89111
123 Main, Apt. 1A, Las Vegas, NV, 89111	123 Main, Apartment 1A, Las Vegas, Nevada, 89111	123 Main St, Apt. 1A, Las Vegas, NV, 89111	123 Main St, Apartment 1A, Las Vegas, Nevada, 89111	123 Main St., Apt. 1A, Las Vegas, NV, 89111	123 Main St., Apartment 1A, Las Vegas, Nevada, 89111	123 Main Street, Apt. 1A, Las Vegas, NV, 89111	123 Main Street, Apartment 1A, Las Vegas, Nevada, 89111
123 Main, Apt. 1A, Las Vegas, Nevada, 89111	123 Main, Unit 1A, Las Vegas, NV, 89111	123 Main St, Apt. 1A, Las Vegas, Nevada, 89111	123 Main St, Unit 1A, Las Vegas, NV, 89111	123 Main St., Apt. 1A, Las Vegas, Nevada, 89111	123 Main St., Unit 1A, Las Vegas, NV, 89111	123 Main Street, Apt. 1A, Las Vegas, Nevada, 89111	123 Main Street, Unit 1A, Las Vegas, NV, 89111
123 Main, #1A, Las Vegas, NV, 89111	123 Main, Unit 1A, Las Vegas, Nevada, 89111	123 Main St, #1A, Las Vegas, NV, 89111	123 Main St, Unit 1A, Las Vegas, Nevada, 89111	123 Main St., #1A, Las Vegas, NV, 89111	123 Main St., Unit 1A, Las Vegas, Nevada, 89111	123 Main Street, #1A, Las Vegas, NV, 89111	123 Main Street, Unit 1A, Las Vegas, Nevada, 89111
123 Main, #1A, Las Vegas, Nevada, 89111	123 Main St, Apt 1A, Las Vegas, NV, 89111	123 Main St, #1A, Las Vegas, Nevada, 89111	123 Main St., Apt 1A, Las Vegas, NV, 89111	123 Main St., #1A, Las Vegas, Nevada, 89111	123 Main Street, Apt 1A, Las Vegas, NV, 89111	123 Main Street, #1A, Las Vegas, Nevada, 89111	123 Main Street, Apartment 1A, Las Vegas, Nevada, 89111
123 Main, Apartment 1A, Las Vegas, NV, 89111	123 Main St, Apt 1A, Las Vegas, Nevada, 89111	123 Main St, Apartment 1A, Las Vegas, NV, 89111	123 Main St., Apt 1A, Las Vegas, Nevada, 89111	123 Main St., Apartment 1A, Las Vegas, NV, 89111	123 Main Street, Apt 1A, Las Vegas, Nevada, 89111	123 Main Street, Apartment 1A, Las Vegas, NV, 89111	
123 Main, Apartment 1A, Las Vegas, Nevada, 89111	123 Main St, Apt. 1A, Las Vegas, NV, 89111	123 Main St, Apartment 1A, Las Vegas, Nevada, 89111	123 Main St., Apt. 1A, Las Vegas, NV, 89111	123 Main St., Apartment 1A, Las Vegas, Nevada, 89111	123 Main Street, Apt. 1A, Las Vegas, NV, 89111		

## Different character sets

---

جوناثان إدوارد سميث

A شارع مين، شقة 1 123

لاس فيغاس، نيفادا 89111

# Cultural variations

---

**Argentina:** Piedras 623, Piso 2, depto 4, C1070AAM, Capital Federal

**Belgium:** Rue du Vivier 7C bte 5, 1000 Bruxelles, BELGIQUE

**India:** e-506 street number 78, uttam viharblock d, uttam nagar, bindapur, 110059, new delhi

**New Zealand:** 43 Vogel Street, Roslyn, Palmerston North 4414

**Ukraine:** vul. Shevchenka, bud. 17, m. Bila Tserkva, Kyivs'ka obl., 09117, UKRAINE

Your RAG is only as  
good as your ER!

Date of Birth: 9-22-61

Address: 123 Main St, Las Vegas, NV 89111

Phone: 555-654-1234

Email: j.smith@domain.com

Johnathan Smith

Date of Birth: 16.3.85

Address: 456 1st St, Las Vegas, NV 89111

Phone: 555-123-4567

Email: john.smith123@outlook.com

Jack Smith

Date of Birth: 16-Mar-1985

Address: 123 Main St #2B, Las Vegas, NV 89111

Phone: 555-654-1234

Email: j.smith@domain.com

John Smith

Date of Birth: Sept. 22, 1961

Address:

Phone:

Email: smith.john@gmail.com

ID#: 987654321

Dr. J. Smith

Date of Birth: 9-22-61

Address: 123 Main St, Las Vegas, NV 89111

Phone: 555-654-1234

Email: j.smith@domain.com

Johnathan Smith

Date of Birth: 16.3.85

Address: 456 1st St, Las Vegas, NV 89111

Phone: 555-123-4567

Email: john.smith123@outlook.com

Jack Smith

“Tell me about Johnathan Smith”

Date of Birth: 16-Mar-1985

Address: 123 Main St #2B, Las Vegas, NV 89111

Phone: 555-654-1234

Email: j.smith@domain.com

John Smith

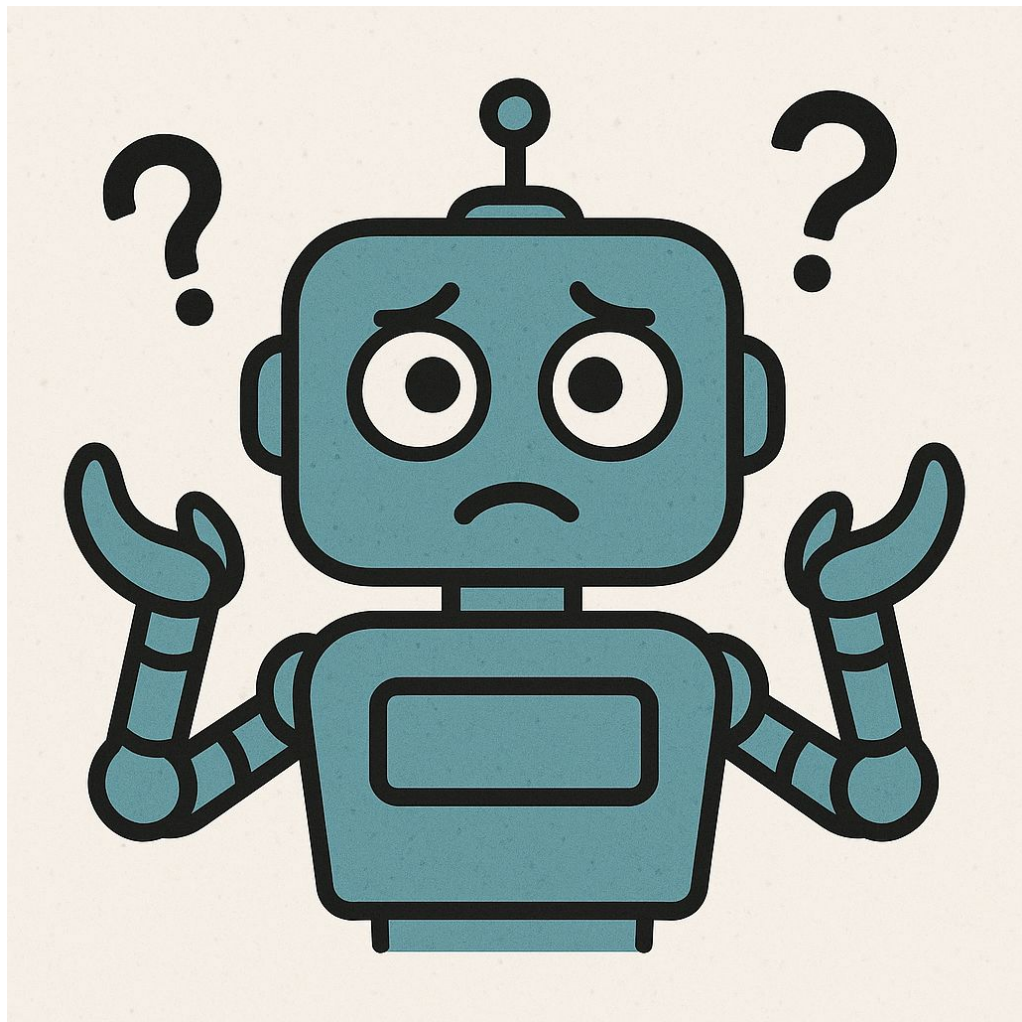
Address:

Phone:

Email: smith.john@gmail.com

ID#: 987654321

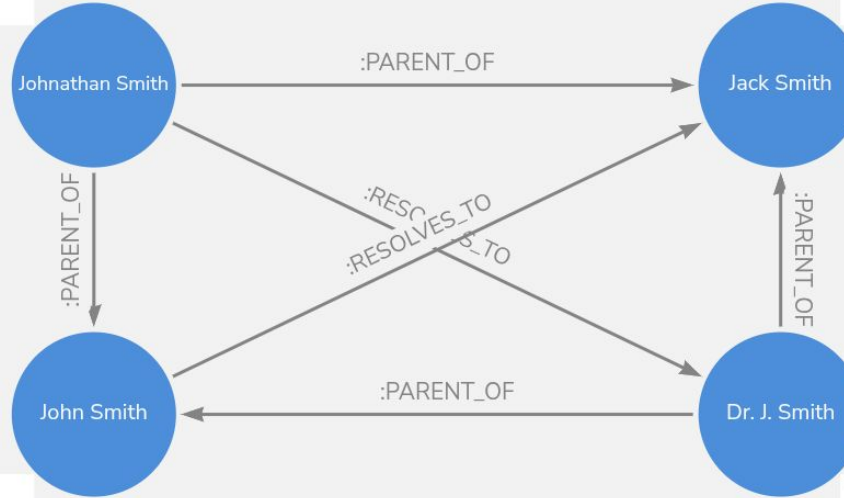
Dr. J. Smith





Date of Birth: 9-22-61  
Address: 123 Main St, Las Vegas, NV 89111  
Phone: 555-654-1234  
Email: j.smith@domain.com

Date of Birth: 16.3.85  
Address: 456 1st St, Las Vegas, NV 89111  
Phone: 555-123-4567  
Email: john.smith123@outlook.com



Date of Birth: 16-Mar-1985  
Address: 123 Main St #2B, Las Vegas, NV 89111  
Phone: 555-654-1234  
Email: j.smith@domain.com

Date of Birth: Sept. 22, 1961  
Address:  
Phone:  
Email: smith.john@gmail.com  
ID#: 987654321

***City Market***<sup>®</sup>



**Smith's**<sup>®</sup>

# Getting started

---

- Graph database
  - Neo4j
- Entity resolution
  - Senzing: real-time, purpose-built SDK for entity resolution
  - Works on your own infrastructure
  - No data shared
  - No training required!

# Example

---

- Three publicly-available data sets
  - SafeGraph (SG)
  - US Department of Labor Wage and Hour Compliance Action Data (DOL)
  - US Small Business Administration PPP Loans over \$150K (PPP)
- Senzing (<https://github.com/senzing>) ER Python package
  - Commercial-grade ER on multiple languages without the need for training/tuning
- LangChain
- Approach: RAG one of two ways
  - SQL tables
  - Run data through Senzing, ingest as an ERKG in Neo4j

# SafeGraph

---

```
{'BRANDS': '[ { "safegraph_brand_id": '
    'SG BRAND f372f9969f6e1ae119a13ff4fc78fd2d", '
    '"safegraph brand name": "Pieology Pizzeria" } ]',
'BUSINESS_ADDR_COUNTRY': 'US',
'BUSINESS_ADDR_FULL': '10965 Lavender Hill Dr Ste 130 Las Vegas NV
89135-2951',
'BUSINESS GEO LATITUDE': '36.144906',
'BUSINESS GEO LONGITUDE': '-115.332644',
'CATEGORY_TAGS': '[ "Counter Service", "Dinner", "Fast Food", "Lunch",
"Pizza" ]',
'CLOSED_ON': 'Not Provided',
'DATA_SOURCE': 'SAFEGRAPH',
'IS_INTERSECTION': 'Not Provided',
'LOCATION_NAME_ORG': 'Pieology Pizzeria',
'MAILING_VERIFIED_STATUS': 'VERIFIED_DELIVERY_POINT',
'NAICS_CODE': '722513',
'OPENED_ON': 'Not Provided',
'PHONE_NUMBER': '+17023314454',
'PLACEKEY': '226-222@5yv-hmm-whq',
'RECORD_ID': '226-222@5yv-hmm-whq',
'RECORD_TYPE': 'ORGANIZATION',
'REL_ANCHOR_DOMAIN': 'PLACEKEY',
'REL_ANCHOR_KEY': '226-222@5yv-hmm-whq',
'REL_POINTER_DOMAIN': 'PLACEKEY',
'REL_POINTER_KEY': 'zzw-223@5yv-hkm-rc5',
'REL_POINTER_ROLE': 'PARENT',
'SUB_CATEGORY': 'Limited-Service Restaurants',
'TOP_CATEGORY': 'Restaurants and Other Eating Places',
'TRACKING_CLOSED_SINCE': '2019-07-01',
'name': 'Pieology Pizzeria',
'uid': 'SAFEGRAPH.226-222@5yv-hmm-whq'}
```

# DoL

---

```
{'BUSINESS_ADDR_CITY': 'Las Vegas',  
  'BUSINESS_ADDR_LINE1': '9611 Trail Wood Drive',  
  'BUSINESS_ADDR_POSTAL_CODE': 89134,  
  'BUSINESS_ADDR_STATE': 'NV',  
  'BUSINESS_NAME_ORG': 'Fabulous Freddy's (Trailwood)',  
  'DATA_SOURCE': 'DoL WHISARD',  
  'LEGAL_NAME_ORG': 'Fabulous LLC',  
  'RECORD_ID': 53,  
  'RECORD_TYPE': 'ORGANIZATION',  
  'address': '9611 Trail Wood Drive Las Vegas NV 89134',  
  'case_id': 1658108,  
  'case_violtn_cnt': 0,  
  'naic_cd': 811192,  
  'naics_code_description': 'Car Washes',  
  'name': 'Fabulous Freddy's (Trailwood)',  
  'uid': 'DOL_WHISARD.53'}
```

# PPP

---

```
{'BUSINESS_ADDR_CITY': 'PHOENIX',  
  'BUSINESS_ADDR_LINE1': '5110 N 40TH ST STE 107',  
  'BUSINESS_ADDR_POSTAL_CODE': 85018,  
  'BUSINESS_ADDR_STATE': 'AZ',  
  'BUSINESS_NAME_ORG': 'INFINITY HOSPICE CARE OF LAS  
VEGAS, LLC',  
  'Business Type': 'Limited Liability Company(LLC)',  
  'CD': 'AZ-09',  
  'DATA_SOURCE': 'PPP LOANS',  
  'DateApproved': '05/01/2020',  
  'JobsReported': 137,  
  'Lender': 'JPMorgan Chase Bank, National Association',  
  'Loan Range': 'c $1-2 million',  
  'NAICS Code': 623110,  
  'NonProfit': '',  
  'OwnedBy': 'Male Owned',  
  'OwnedByRaceEthnicity': 'Unanswered',  
  'OwnedByVeteran': 'Unanswered',  
  'RECORD_ID': 7017,  
  'RECORD_TYPE': 'ORGANIZATION',  
  'address': '5110 N 40TH ST STE 107 PHOENIX AZ 85018',  
  'name': 'INFINITY HOSPICE CARE OF LAS VEGAS, LLC',  
  'uid': 'PPP_LOANS.7017'}
```

# The SQL approach without ER



```
sql_agent_executor.invoke('find all references  
to Union Cabs in all of the tables')
```

```
>>
```

```
> Finished chain.
```

```
{'input': 'find all references to Union Cabs  
in all of the tables',
```

```
  'output': 'The only reference to "Union Cabs"  
found in the database is in the `sg` table  
under the `LOCATION_NAME_ORG` column.\n\nHere  
is the result:\n- `sg` table: `Union  
Cabs`\n\nNo references were found in the `dol`  
or `ppp` tables.'}
```

```
sql_agent_executor.invoke('does Union Cabs have  
any violations?')
```

```
>>
```

```
> Finished chain.
```

```
{'input': 'does Union Cabs have any  
violations?',
```

```
  'output': 'It appears that Union Cabs does not  
have any violations recorded in the database.'}
```

# The ERKG approach



(g2) get 12093

Entity summary for entity 12093: DESERT SPRINGS LANDSCAPING LLC

Record ID	Entity Data	Additional Data
PPP_LOANS 8321	BUSINESS: DESERT SPRINGS LANDSCAPING LLC RECORD_TYPE: ORGANIZATION BUSINESS: 8130 S VALLEY VIEW BLVD LAS VEGAS NV 89139	BUSINESS_TYPE: Limited Liability Company(LLC) CD: NV-03 DATEAPPROVED: 04/08/2020 JOBSREPORTED: 40 LENDER: Meadows Bank LOAN_RANGE: d \$350,000-1 million NAICS_CODE: 561730 OWNEDBY: Unanswered OWNEDBYRACEETHNICITY: Unanswered OWNEDBYVETERAN: Unanswered
SAFEGRAPH 224-226@5yv-jbb-jvz	LOCATION: Desert Springs Landscaping RECORD_TYPE: ORGANIZATION PLACEKEY: 224-226@5yv-jbb-jvz BUSINESS: 36.042112 -115.189595 BUSINESS: 8130 S Valley View Blvd Las Vegas NV 89139-7892 US PHONE: +17024368080	BRANDS: [] CATEGORY_TAGS: [ "Landscaping", "Pool Cleaner", "Pool Service" ] MAILING_VERIFIED_STATUS: VERIFIED_DELIVERY_POINT NAICS_CODE: 5617 TOP_CATEGORY: Services to Buildings and Dwellings TRACKING_CLOSED_SINCE: 2019-07-01

- └ Possible relationships (6)
  - └ ADDRESS (4)
    - └ 49974 SAFEGRAPH (1) | PPP\_LOANS (1) DESERT SPRINGS POOLS AND SPAS INC
    - └ 71900 SAFEGRAPH (1) Build Your Own Pool
    - └ 200936 PPP\_LOANS (1) HARSHMAN EXCAVATING CORP
    - └ 202069 PPP\_LOANS (1) DESERT SPRINGS RENOVATIONS LLC
  - └ ADDRESS+GEO\_LOC (1)
    - └ 49150 SAFEGRAPH (1) Oasis Springs Builders
  - └ ADDRESS+GEO\_LOC+PHONE (1)
    - └ 67084 SAFEGRAPH (1) Desert Springs Group

(g2) get 49974

Entity summary for entity 49974: DESERT SPRINGS POOLS AND SPAS INC

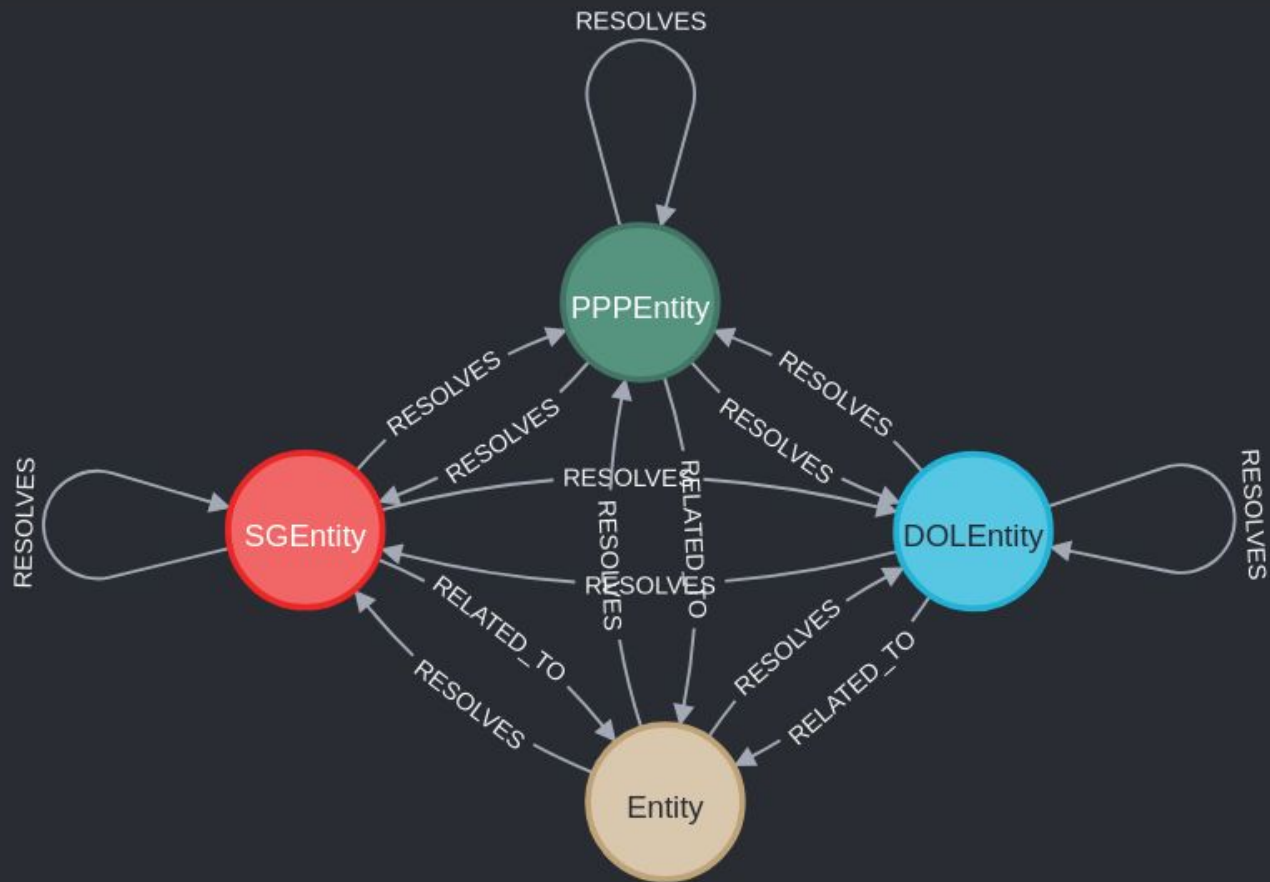
Record ID	Entity Data	Additional Data
PPP_LOANS 8323	BUSINESS: DESERT SPRINGS POOLS AND SPAS, INC RECORD_TYPE: ORGANIZATION BUSINESS: 8130 S. Valley View Blvd. LAS VEGAS NV 89139	BUSINESS_TYPE: Subchapter S Corporation CD: NV-03 DATEAPPROVED: 04/15/2020 JOBSREPORTED: 36 LENDER: Zions Bank, A Division of LOAN_RANGE: d \$350,000-1 million NAICS_CODE: 238990 OWNEDBY: Unanswered OWNEDBYRACEETHNICITY: Unanswered OWNEDBYVETERAN: Unanswered
SAFEGRAPH 224-224@5yv-jbb-jvz	LOCATION: Desert Springs Pools & Spas RECORD_TYPE: ORGANIZATION PLACEKEY: 224-224@5yv-jbb-jvz BUSINESS: 8130 S Valley View Blvd Las Vegas NV 89139-7892 US BUSINESS: 36.041785 -115.190721 PHONE: +17024525900	BRANDS: [] CATEGORY_TAGS: [ "Patio Coverings", "Pool Service" ] MAILING_VERIFIED_STATUS: VERIFIED_DELIVERY_POINT NAICS_CODE: 2389 TOP_CATEGORY: Other Specialty Trade Contractors TRACKING_CLOSED_SINCE: 2019-07-01

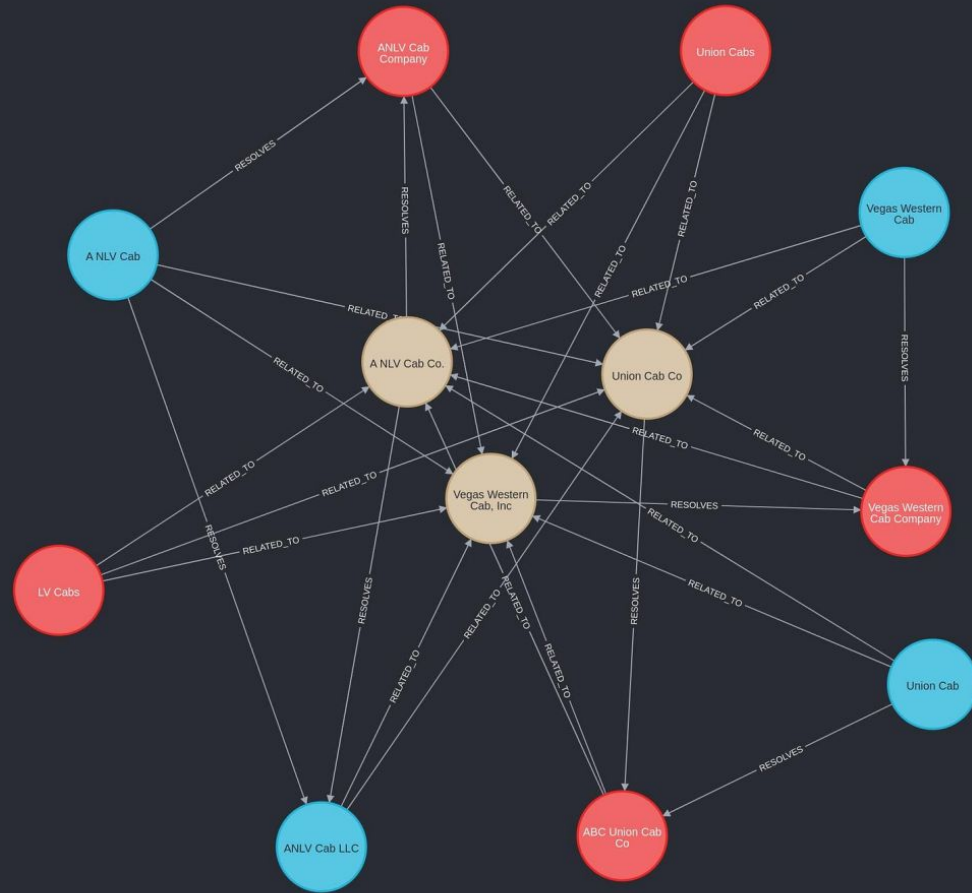
Possible relationships (7)

- ADDRESS (6)
  - 12093 SAFEGRAPH (1) | PPP\_LOANS (1) DESERT SPRINGS LANDSCAPING LLC
  - 49150 SAFEGRAPH (1) Oasis Springs Builders
  - 67084 SAFEGRAPH (1) Desert Springs Group
  - 71900 SAFEGRAPH (1) Build Your Own Pool
  - 200936 PPP\_LOANS (1) HARSHMAN EXCAVATING CORP
  - 202069 PPP\_LOANS (1) DESERT SPRINGS RENOVATIONS LLC
- GEO\_LOC (1)
  - 37912 SAFEGRAPH (1) Discount Tire

Disclosed relationships (1)

- PARENT (1)
  - 65977 SAFEGRAPH (1) Blue Diamond Crossing







```
response = chain.invoke({"query": "Tell me about Union Cab  
Co"})
```

```
...
```

```
>>
```

```
> Finished chain.
```

```
{'query': 'Tell me about Union Cab Co',  
  'result': 'Union Cab Co, also known as ABC Union Cab Co, was  
located at 5010 S Valley View Blvd, Las Vegas, NV 89118-1705.  
It operated in the Taxi Service industry, specifically under  
the NAICS code 485310. The company was categorized under "Taxi  
and Limousine Service"  
and had tags such as "Airport Shuttles" and "Taxis." Union Cab  
Co had a verified mailing status and was tracked as closed  
since July 1, 2019, with its official closure on March 1,  
2022. The company had one recorded case violation and was  
related to other entities like Vegas Western Cab, Inc and A  
NLV Cab Co.'}
```

```
response = chain.invoke({"query": "How many  
violations have entities Union Cab Co is  
related to been involved in?"})
```

...

```
> Finished chain.
```

```
{'query': 'How many violations have entities  
Union Cab Co is related to been involved in?',  
  'result': 'Entities related to Union Cab Co  
have been involved in a total of 6  
violations.'}
```

# Conclusions

---

- RAG is important

# Conclusions

---

- ~~RAG is important~~
- GraphRAG is important

# Conclusions

---

- ~~RAG is important~~
- ~~GraphRAG is important~~
- Doing ER properly is much more difficult than can be solved with regex, edit distance algorithms, clustering, etc.

# Conclusions

---

- ~~RAG is important~~
- ~~GraphRAG is important~~
- Doing ER properly is much more difficult than can be solved with regex, edit distance algorithms, clustering, etc.
- GraphRAG with ERKG is important-est!!!

# Thank you!

[clair@clairsullivan.com](mailto:clair@clairsullivan.com)

[https://github.com/cj2001/odsc\\_east\\_2025](https://github.com/cj2001/odsc_east_2025)

