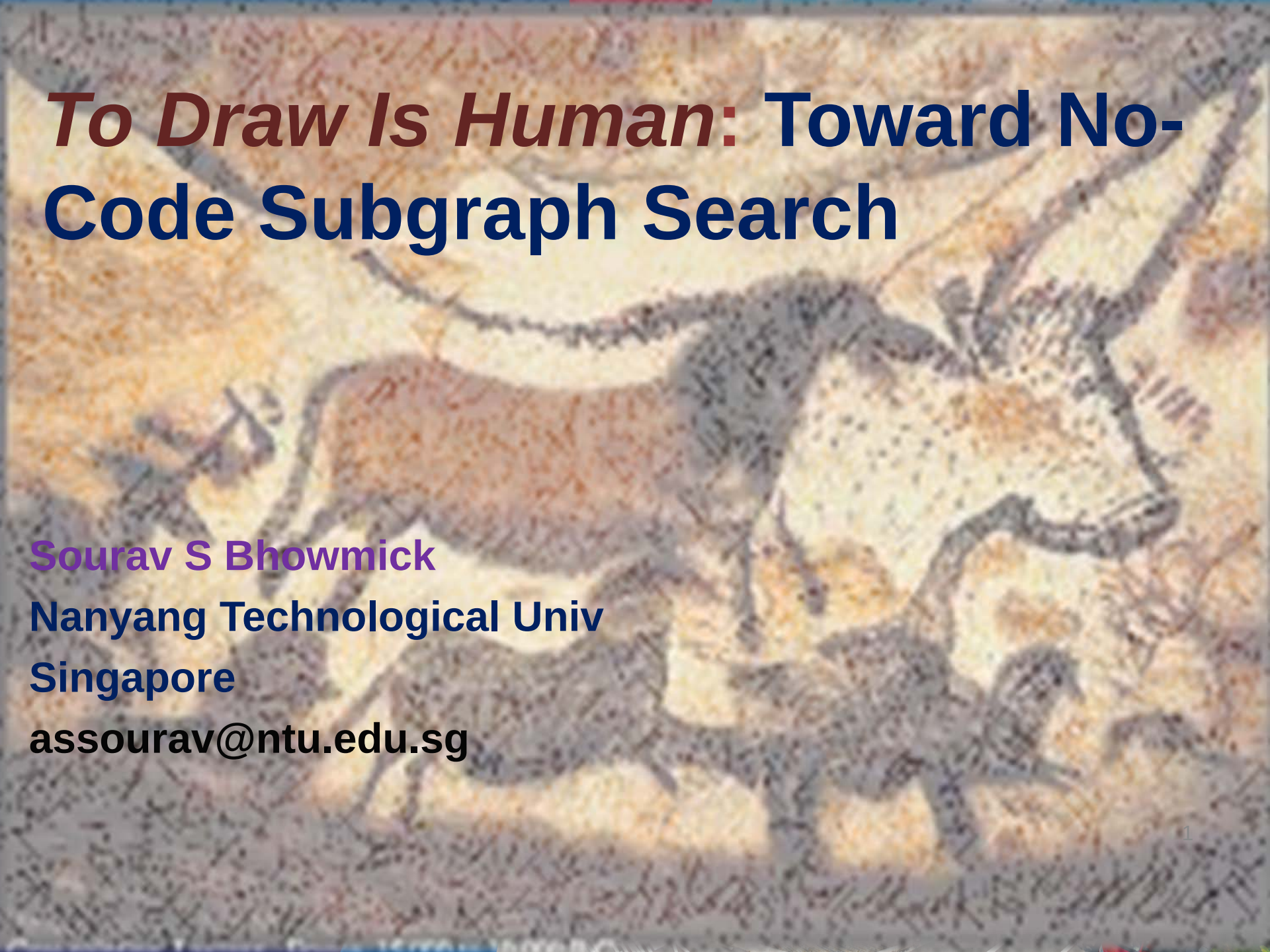


To Draw Is Human: Toward No-Code Subgraph Search

Sourav S Bhowmick
Nanyang Technological Univ
Singapore
assourav@ntu.edu.sg

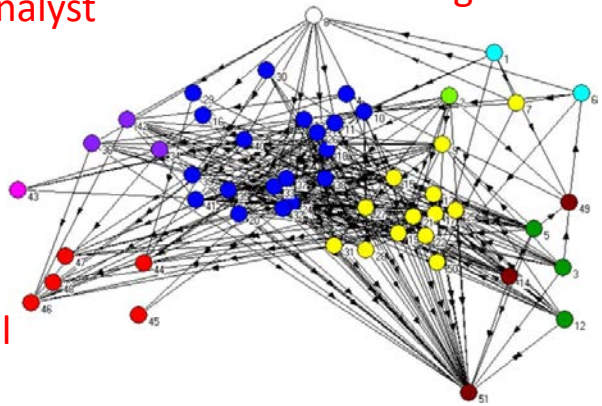
World of Networks

Security analyst

biologist

Pharmacist/chemist

Techies



Financial analyst

ecologist

journalist

social scientist

DB Approach of Searching Graphs

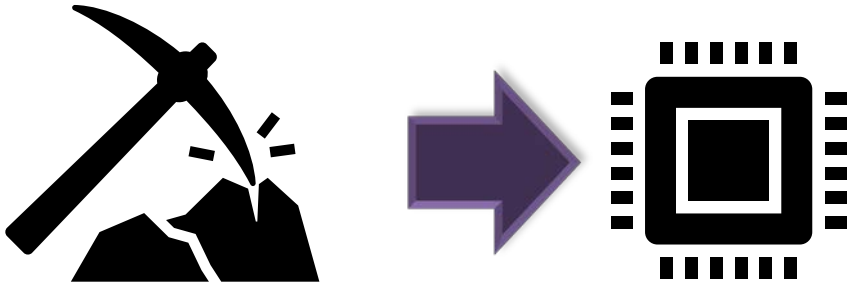
Query Formulation

- Formal query language

Query Processing

Efficient algorithms and optimization techniques to process queries “quickly”

Query Formulation Precedes Query Processing



Gremlin

Cypher

GQL

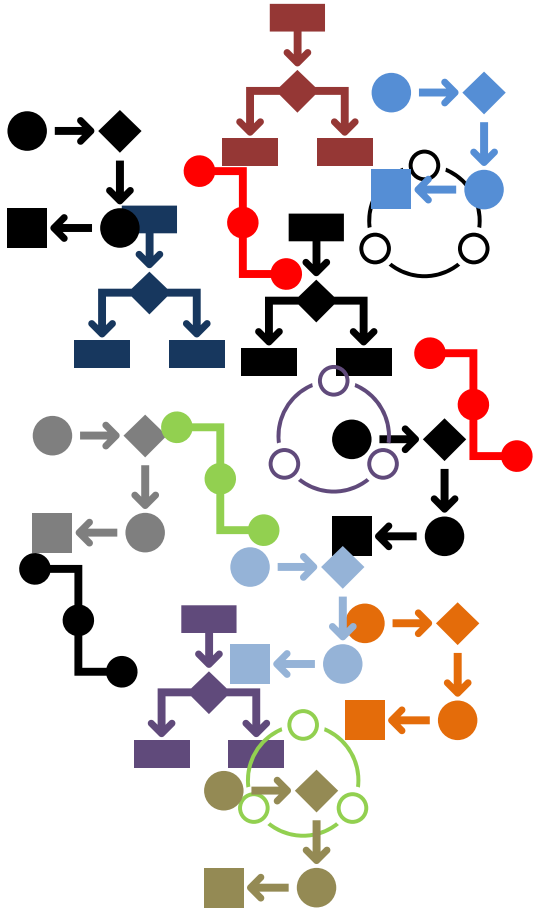
GCORE

GraphQL

PGQL

SparQL

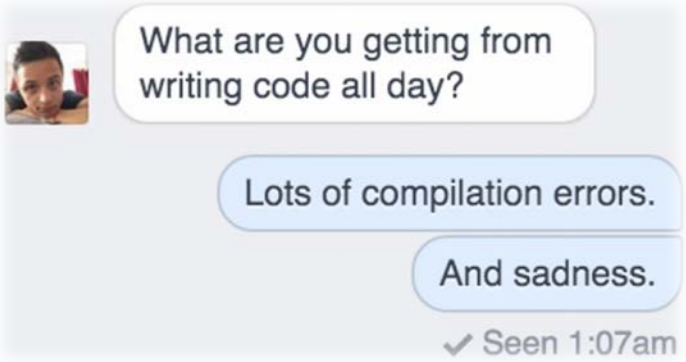
AQL

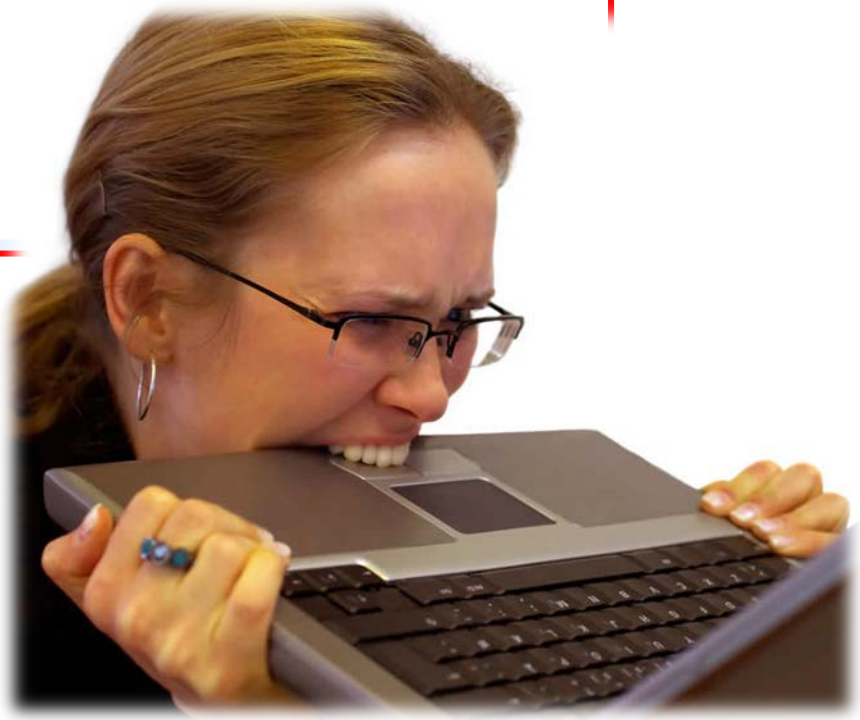


Graph Query Formulation using QL

```
prefix wp:      <http://vocabularies.wikipathways.org/wp#>
prefix dcterms: <http://purl.org/dc/terms/>
prefix foaf:    <http://xmlns.com/foaf/0.1/>

select (str(?organismName) as ?organism) ?page ?gene1 ?gene2 ?interaction where {
  ?gene1 a wp:GeneProduct .
  ?gene2 a wp:GeneProduct .
  ?interaction wp:source ?gene1 ;
    wp:target ?gene2 ;
    a wp:Conversion ;
    dcterms:isPartOf ?pathway .
  ?pathway foaf:page ?page ;
    wp:organismName ?organismName .
  FILTER (?gene1 != ?gene2)
} ORDER BY ASC(?organism)
```





A Significant Barrier for Democratizing Graph Databases

If users cannot formulate queries, a powerful query engine is of no use to them!

What Do Researchers Think?

> Today's data consumers may not know how to formulate a query at all – e.g., a journalist who wants to "find the average temperature of all cities with population exceeding 100,000 in Florida" over a structured data set. Our community's challenge is to make it possible for such people to get their answers themselves, directly. This requires new query interfaces, e.g., interfaces based on multitouch, not just console-based SQL interfaces. We need interfaces that combine visualization, querying, and navigation.

The Beckman Report on Database Research, Communication of the ACM (Feb 2016)

What do Practitioners Think?

> Another potential drawback is that developers have to write their queries using Java as there is no Standard Query Language (SQL) to retrieve data from graph databases, which means employing expensive programmers or developers have to use SparcQL or one of the other query languages that have been developed to support graph databases, however, it would mean learning a new skill. This results in the lack of standardization and programming ease for graph database systems. There are visualization tools available for graph databases, but they are still in the developing stage.

Marketsandmarkets.com
2021

https://www.marketsandmarkets.com/Market-Reports/graph-database-market-126230231.html?gclid=Cj0KCQiAxc6PBhCEARIsAH8Hff1pUb5PI2peZmHQa-AvoPd2MRWXyPwGfEKYFu6I86Z-SgGyQ2a8G88aAmgmEALw_wcB

What do Practitioners Think?

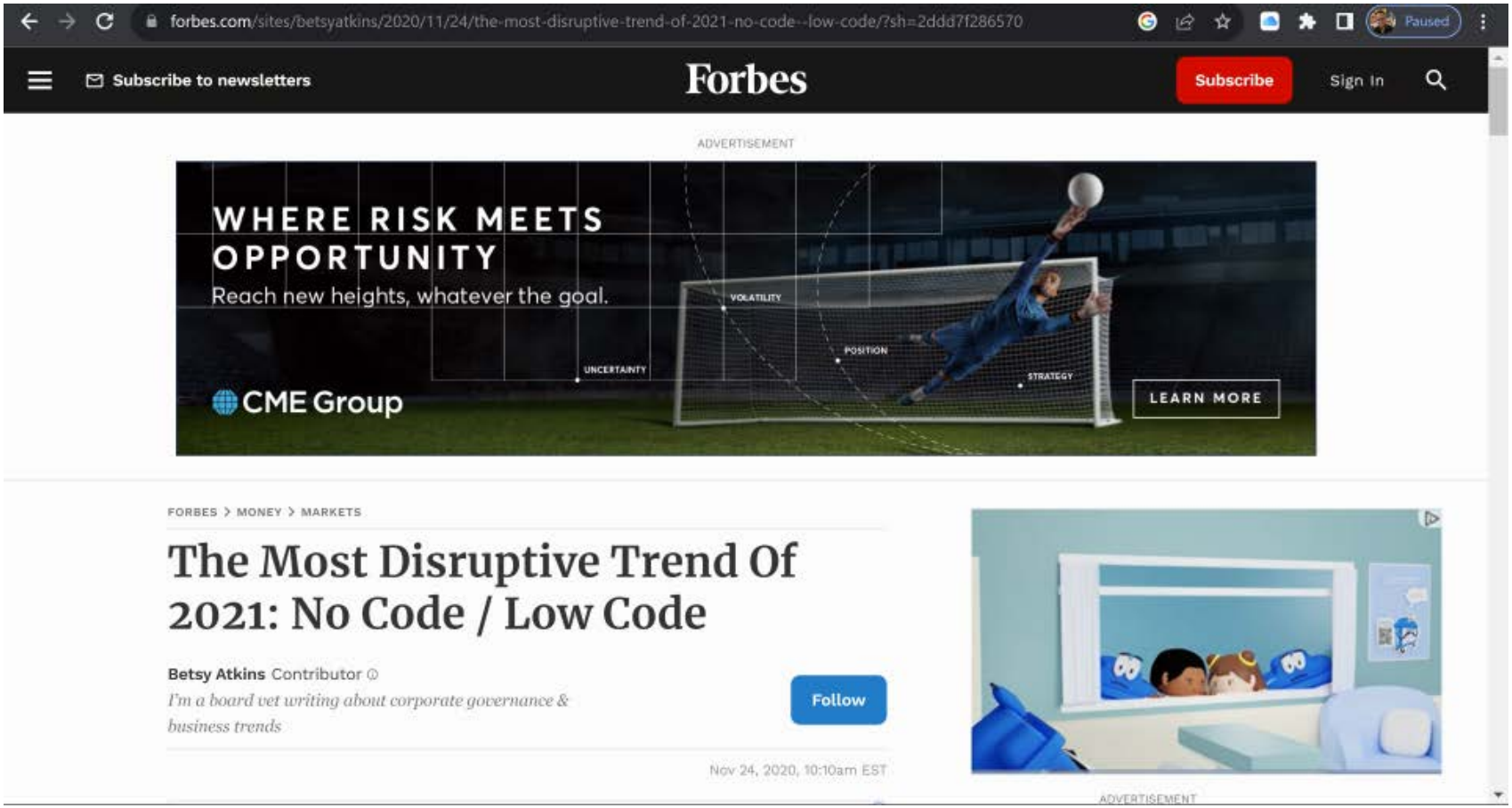
> *If the user can't use it, it doesn't work.*

Susan Dray, President
Dray and Associates, Inc

Why This Is The Reality?

	Traditional Subgraph Search	Search in the Data-driven World
Who?	IT professionals	Analysts in a specific line of business
Why?	Search data, design & implement search strategies	Search only
What?	Use graph query languages	Use programming-agnostic tools
Where?	Typically, in the IT department	Anywhere in an organization

Solutions? Toward No-Code Search



Towards No-Code Subgraph Search

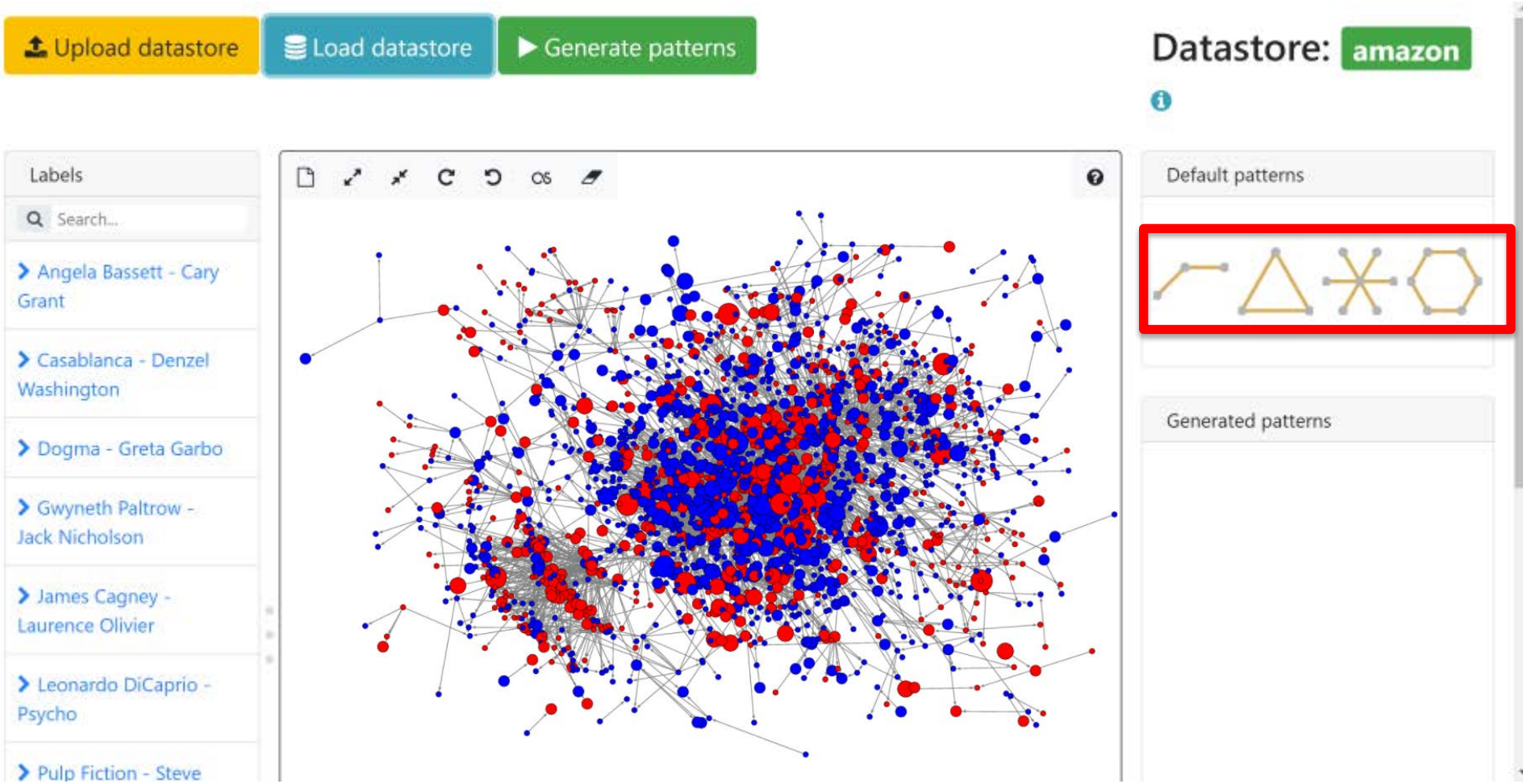
Why No-Code?

- Worldwide shortage of developers.
- Growing talent gap.
- Budgetary challenges in small- and medium-sized companies for hiring software team.

Visual Query Interface (VQI)

- Graphs are more intuitive to draw than to compose in textual format.
- Growing existence of VQIs for industrial-strength systems.

Example


Upload datastore
Load datastore
Generate patterns
Datastore: amazon
Labels
Search...
Angela Bassett - Cary Grant
Casablanca - Denzel Washington
Dogma - Greta Garbo
Gwyneth Paltrow - Jack Nicholson
James Cagney - Laurence Olivier
Leonardo DiCaprio - Psycho
Pulp Fiction - Steve
Default patterns
Generated patterns

Useful Attributes and Pattern Selection

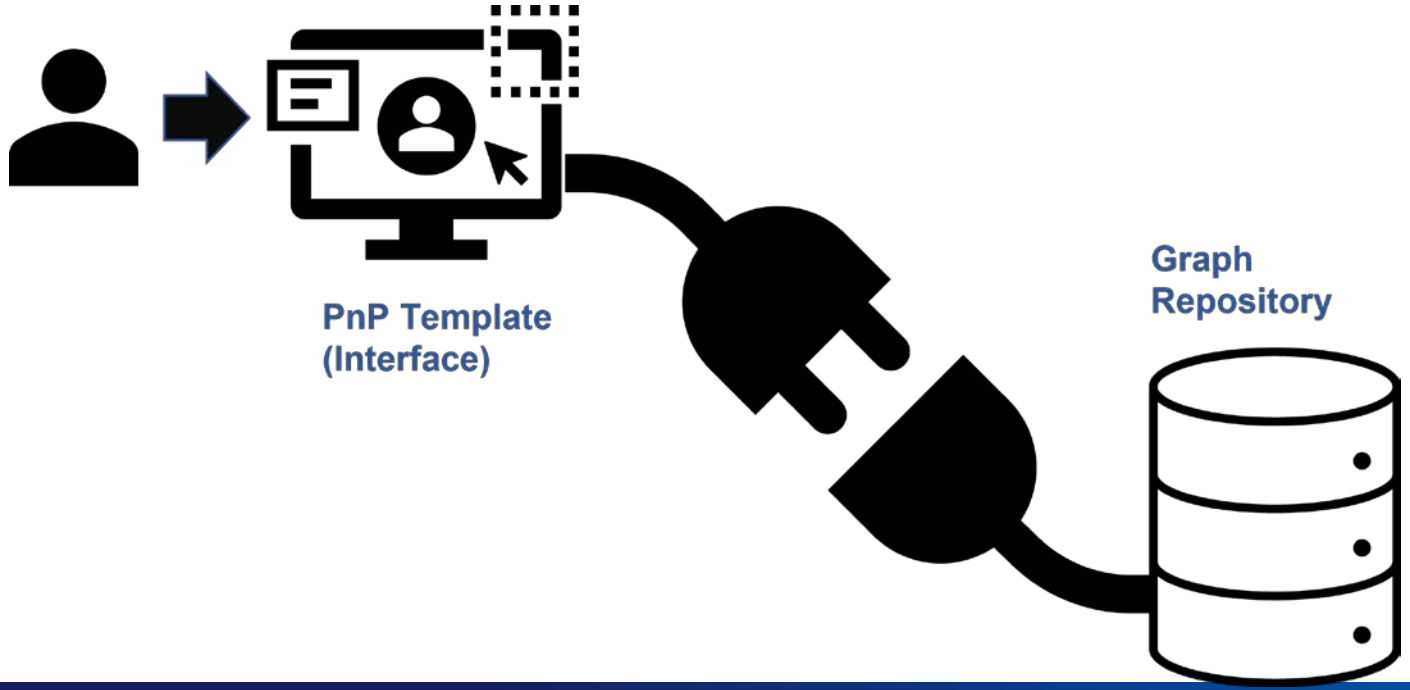
Advantages of Canned Patterns

Efficient Visual Query Formulation

- Pattern-at-a-time mode of querying incurs fewer steps and lesser time in formulating queries compared to edge-at-a-time mode.
- Less repetitive actions from users thereby reducing frustration.
- Supports both top-down and bottom-up search.

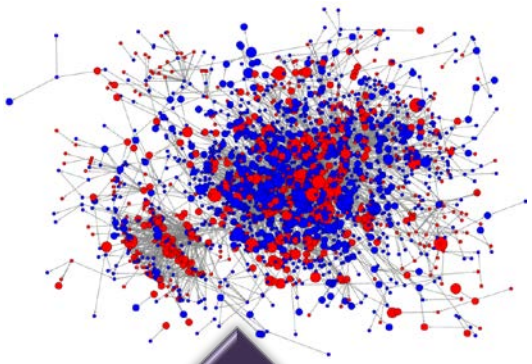
Data-driven/Plug-and-Play (PnP) VQI

Designed to give end users the freedom to easily and quickly construct and maintain a VQI in a *data-driven* manner for any graph data source without resorting to coding by simply "plugging" it on the data.



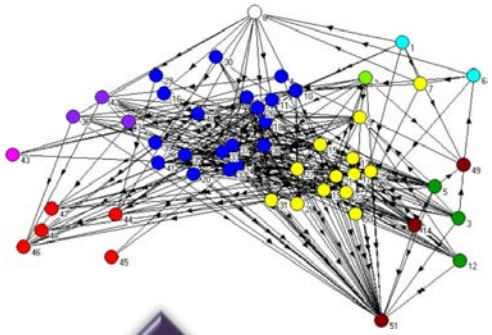
The Idea

Socket 1



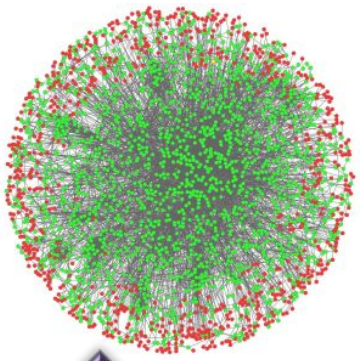
Auto

Socket 2



Auto

Socket 3



Auto

Plug

What Makes a Good Canned Pattern Set? [Huang et al., 2019]

High coverage

Canned pattern set ideally **cover as many graphs in dataset as possible**.

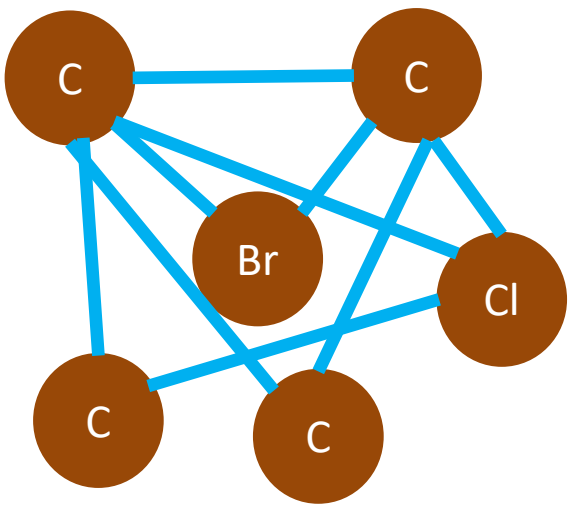
High diversity

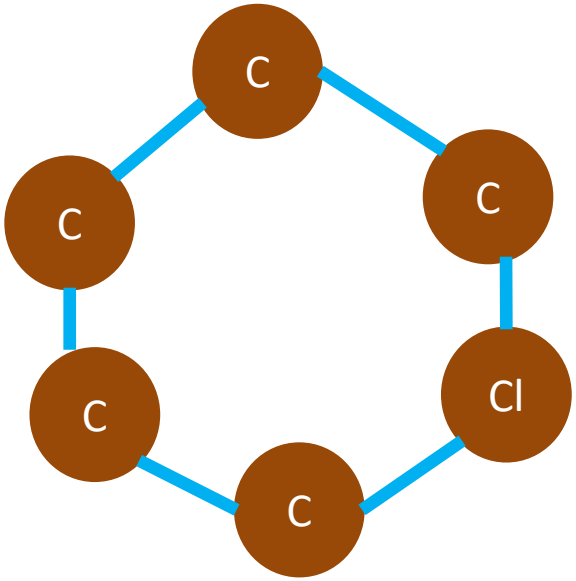
Diverse patterns makes **better use** of **limited** GUI display **space**.

Low cognitive load

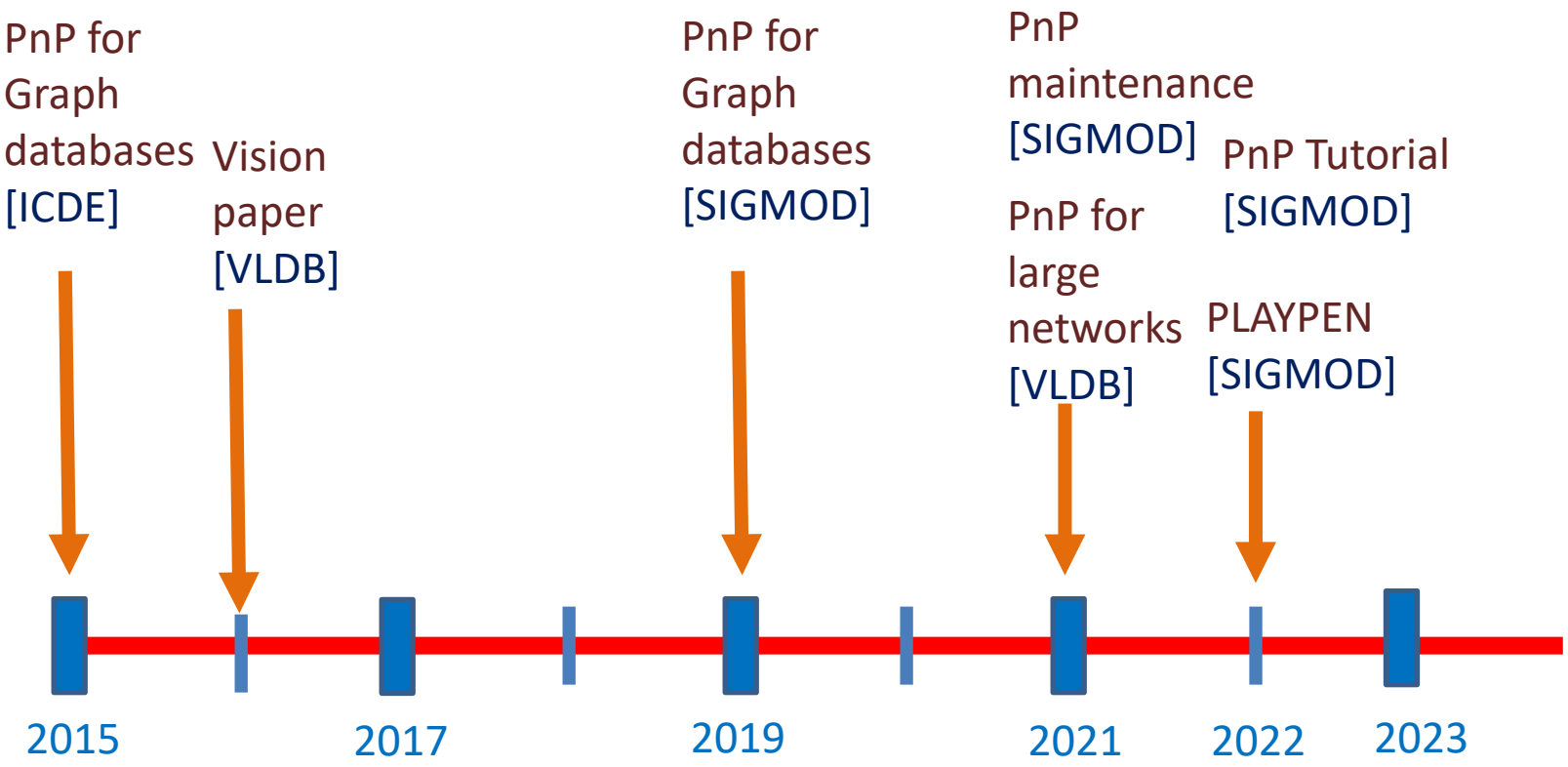
Demands **little mental effort** by users to interpret and determine if patterns can be utilize for query formulation.

Cognitive Load-aware Pattern Selection





The Odyssey of PnP Interfaces



Canned Pattern Selection (CPS) Problem For Large Networks

Given a network *G = (V,E)*, a PnP template *I* and a user-specified plug *b*, the goal of the **canned pattern selection (CPS) problem** is to select a set of *unlabeled* patterns *P* for display on *I*, which satisfies the specifications in *b* and *optimizes coverage, diversity* and *cognitive load* of *P*.

NP-hard problem

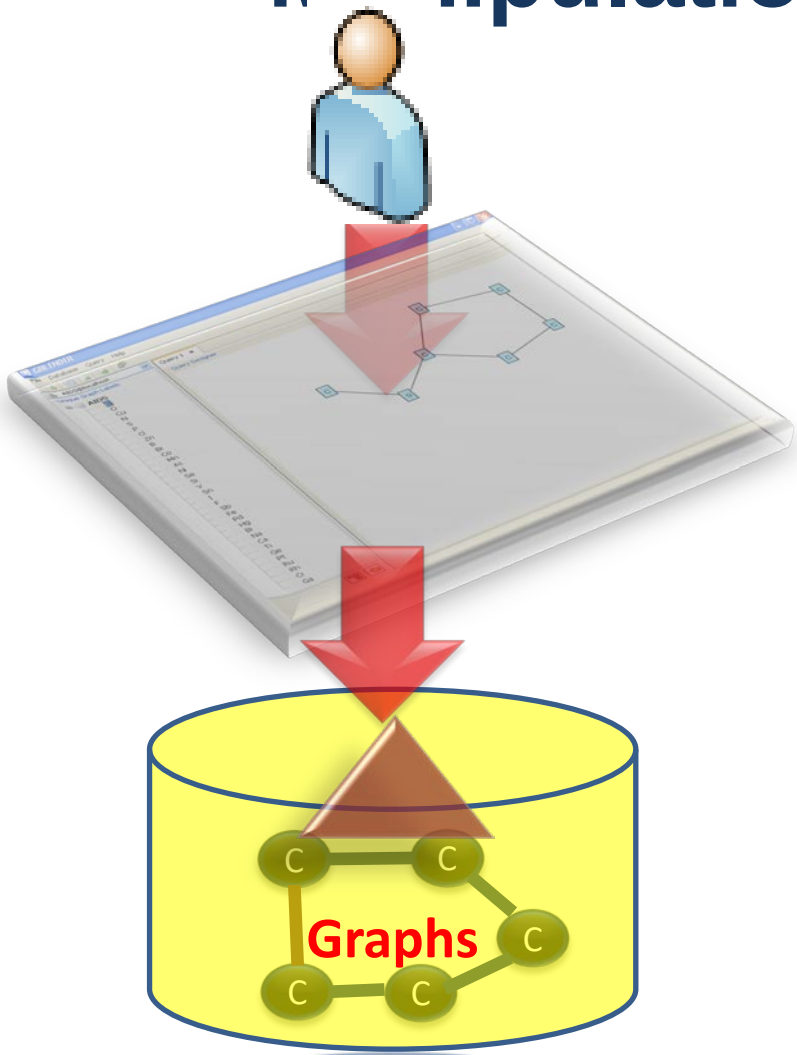
PLAYPEN In Action

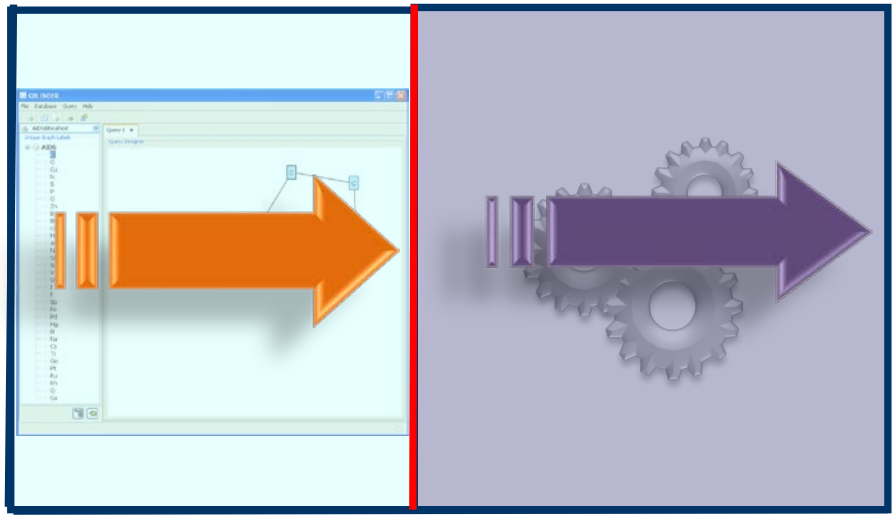
https://www.youtube.com/watch?v=nS_nFQQN_Ck

Visual Query Processing

Given a visual graph query, how can we process it?

Circa 2008: Classical Direct Manipulation-based Approach

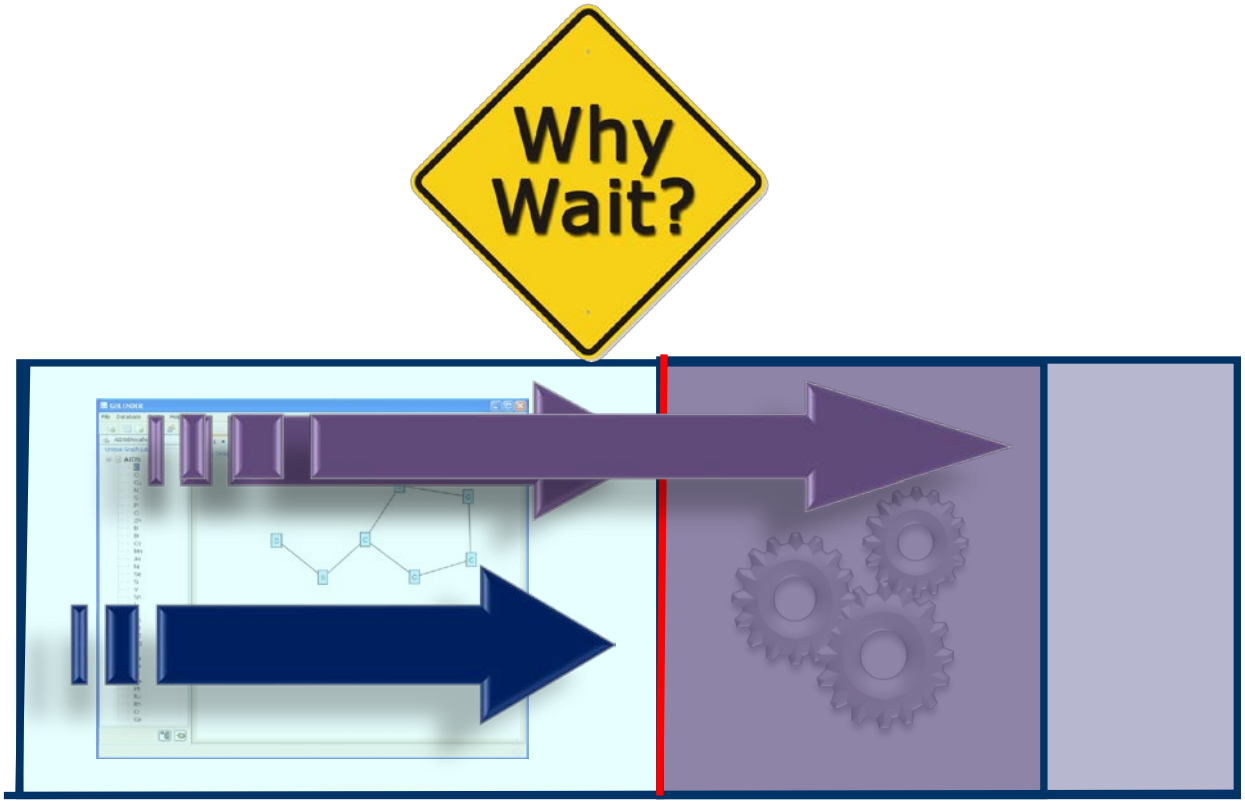




Query formulation Query processing

time

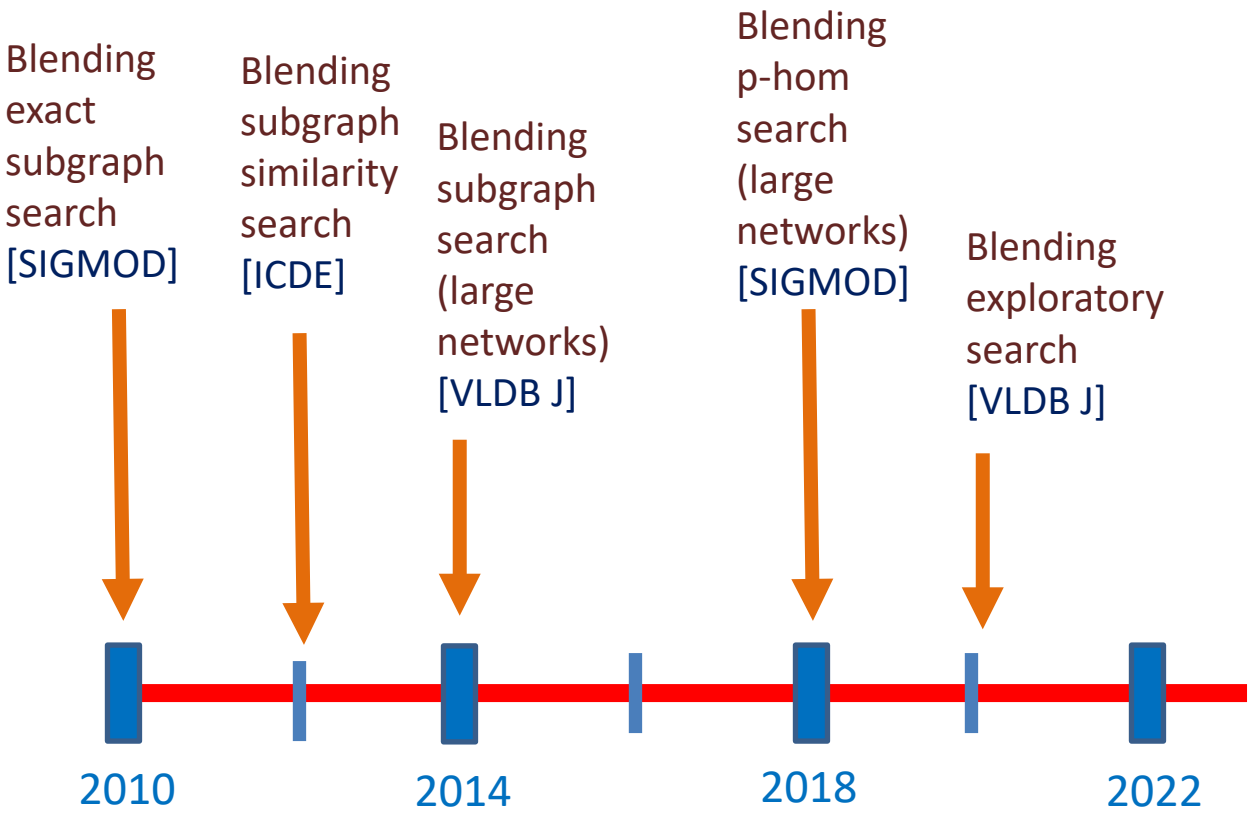
Why Wait?



Benefits of Blending

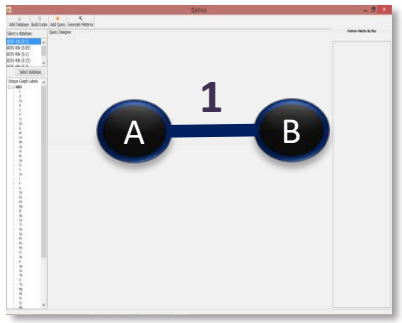
- Query suggestions and feedback [ICDE 19, VLDB J 17, VLDB 15, CIKM 15]
- Faster query response time [SIGMOD 10, ICDE 12, SIGMOD 13, CIDR 13, VLDB J 14, SIGMOD 18, SIGMOD 20]
- Interactive search and exploration [ICDE 19, VLDB 17, VLDBJ 20]
- Interactive visualization of results [VLDB 17, VLDB J 14]

The Odyssey of Blending

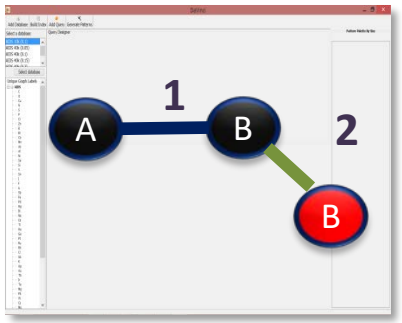


Graph Query Processing

Online



Adaptive On-the-fly index

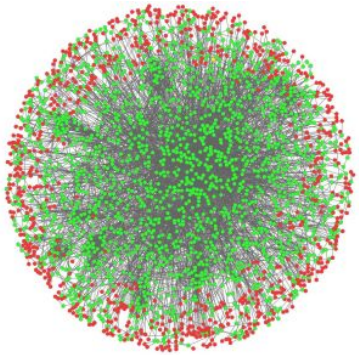


Candidate matches

Candidate matches

Final Results

Offline



Precomputed Index

Conclusion

> "The primary bottleneck is not the quality of the inner-loop algorithms, but the lack of technology enabling domain experts to perform end-to-end **Graph search** without programming experience. Hellerstein, Heer, Kandel

Toward No-code Subgraph Search

1990-2015: Visual query interfaces are constructed manually

2015: Automatic, data-driven construction of visual graph query interface

1970s-2005s: Query Formulation → Query Processing

2010s: Visual query form. ⟷ Query Processing

Multi-disciplinary effort:

Data management

HCI

Cognitive psychology

HINT Project

Open Problems

- Rethinking in a distributed environment
- On property graphs
- Multi-faceted exploration and visualization
- Expanding the paradigm to other data types
- Direct Manipulation-driven analytics

Thank You!

Serengeti National Park, Tanzania, July 2023