We Don’t Need No Education: From Building for Coders to Building for Users

Sourav S Bhowmick

Nanyang Technological Univ

Singapore

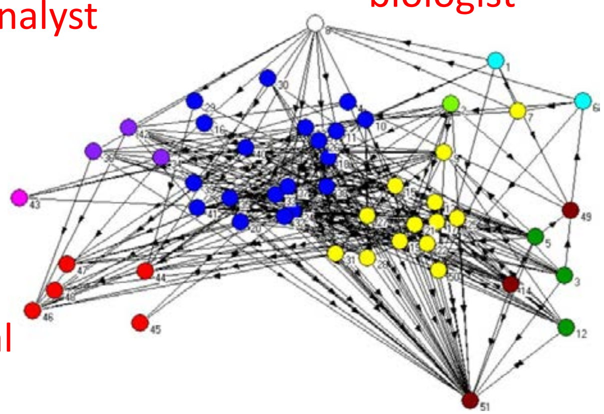
World of Graphs

Security analyst

biologist

Pharmacist/chemist

Techies



Financial analyst

social scientist

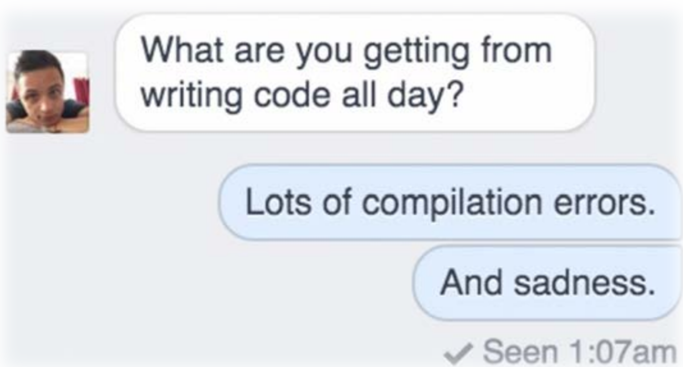
ecologist

journalist

DB Approach to Graph Interaction

```
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)
```



QL – Command Language Style of Interaction

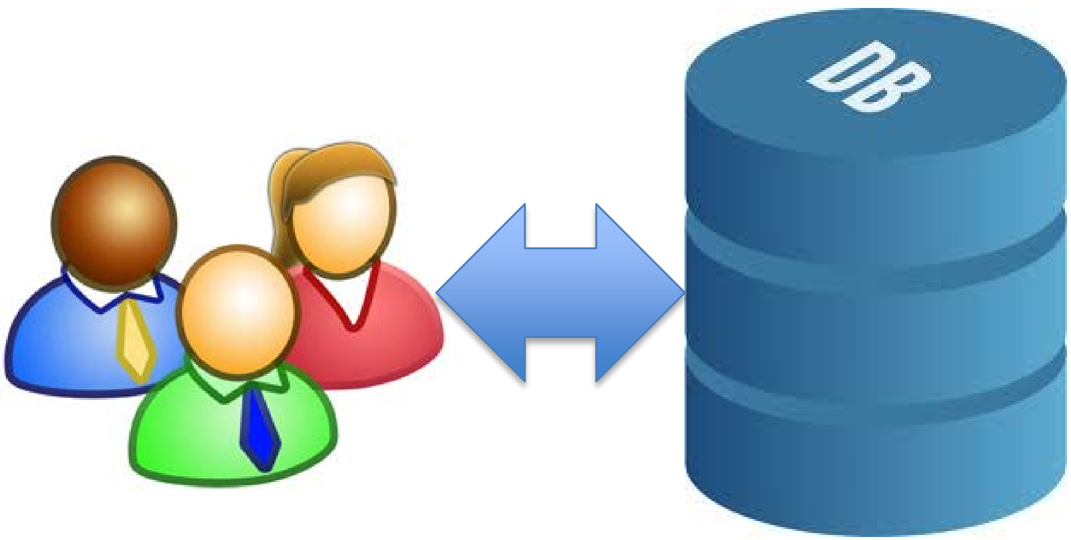
Pros

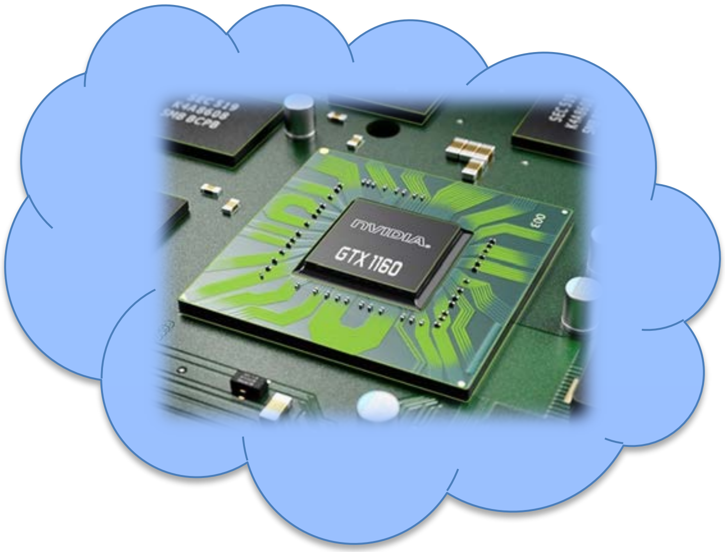
- Flexible
- Appeals to “power” users
- Allows convenient creation of user-defined macros
- Domain of expert frequent users

Cons

- Error rates are high
- Substantial training is necessary
- Retention may be poor

Are We Nice to End Users?





Shneiderman’s View

Evolution of Users

“In the first decades of computer-software development, technically oriented programmers designed text editors, programming languages, and applications for themselves and their peers. The substantial experience and motivation of these users meant that complex interfaces were accepted and even appreciated. Now, the user population for mobile devices, instant messaging, e-business, and digital libraries is so vastly different from the original that programmers’ intuitions may be inappropriate. Current users are not dedicated to the technology; their background is more tied to their work needs and the work tasks they perform.

Ben Shneiderman & Catherine Plaisant
Designing the User Interface (5th Ed, 2010)

Proof By Authority (2005)

> Thirty years of research on query languages can be summarized by: we have moved from SQL to XQuery. At best we have moved from one declarative language to a second declarative language with roughly the same level of expressiveness. It has been well documented that end users will not learn SQL; rather SQL is notation for professional programmers.

The Lowell Database Research Self-Assessment,
Communication of the ACM (May 2005)

Proof By Authority (2016)

> 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?

53 practitioners (36 from IT)

Reality

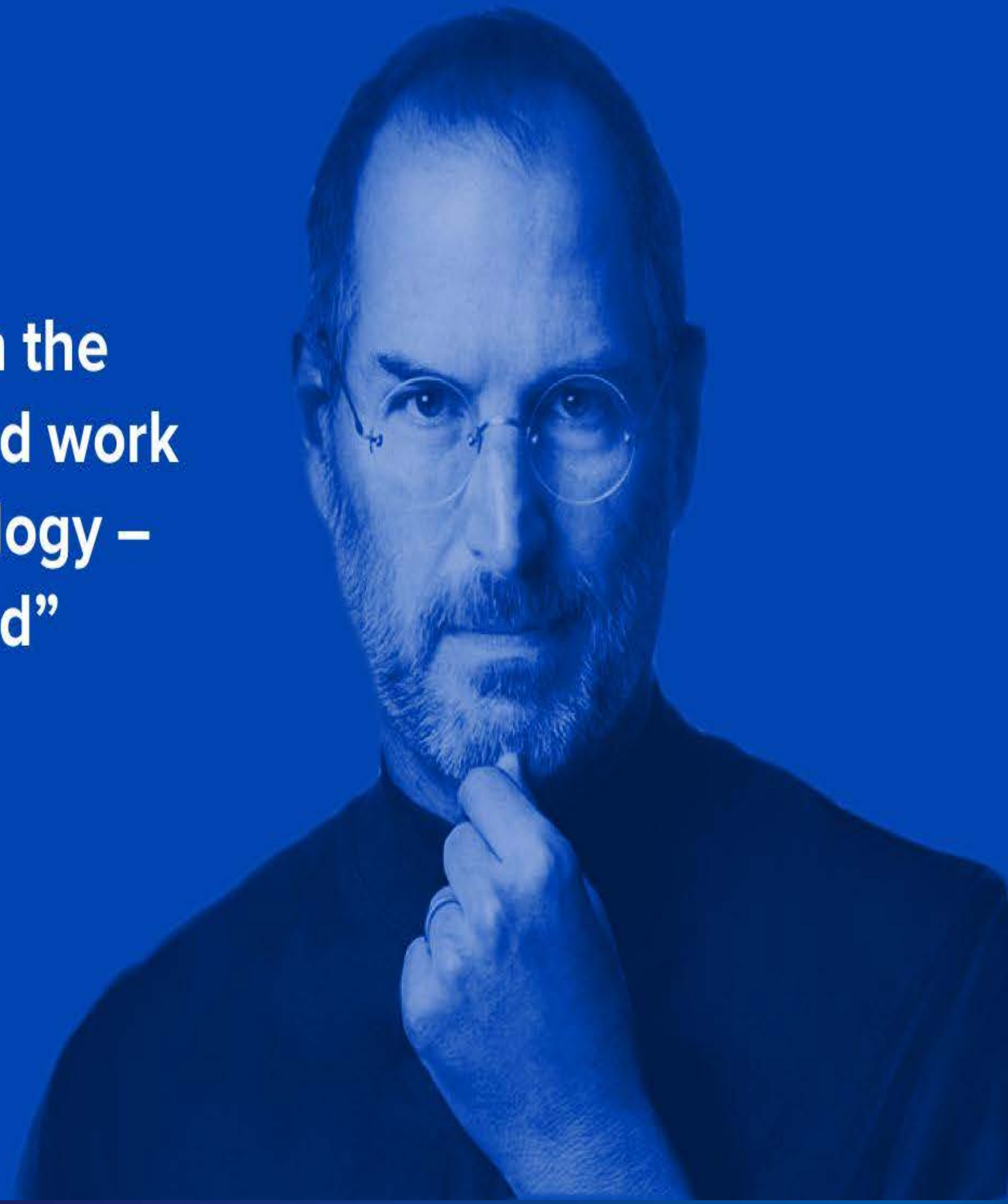
“ The overwhelming majority of the emails and issues were routine engineering tasks, such as users asking how to write a query...

Challenge	Total	R	P
Scalability (i.e., software that can process larger graphs)	45	20	25
Visualization	39	17	22
Query Languages / Programming APIs	39	18	21
Faster graph or machine learning algorithms	35	19	16
Usability (i.e., easier to deploy, configure, and use)	25	10	15
Benchmarks	22	12	10
Extract & Transform	20	6	14
More general purpose graph software (e.g., that can process offline, online, and streaming computations)	20	9	11
Graph Cleaning	17	7	10
Debugging & Testing	10	2	8

Siddhartha Sahu, et al. **The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing.** PVLDB 11(4): 420-431 (2017)

"You've got to start with the customer experience and work back toward the technology – not the other way around"

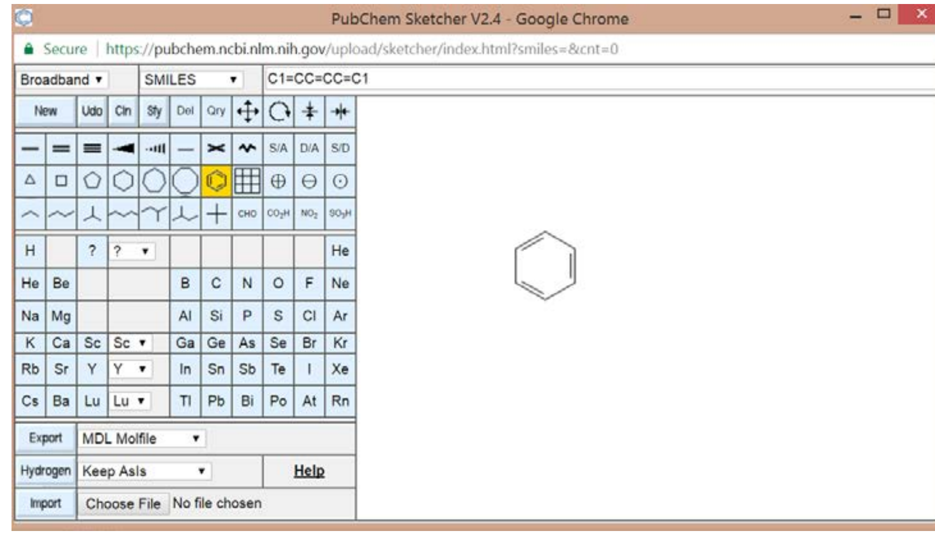
Steve Jobs

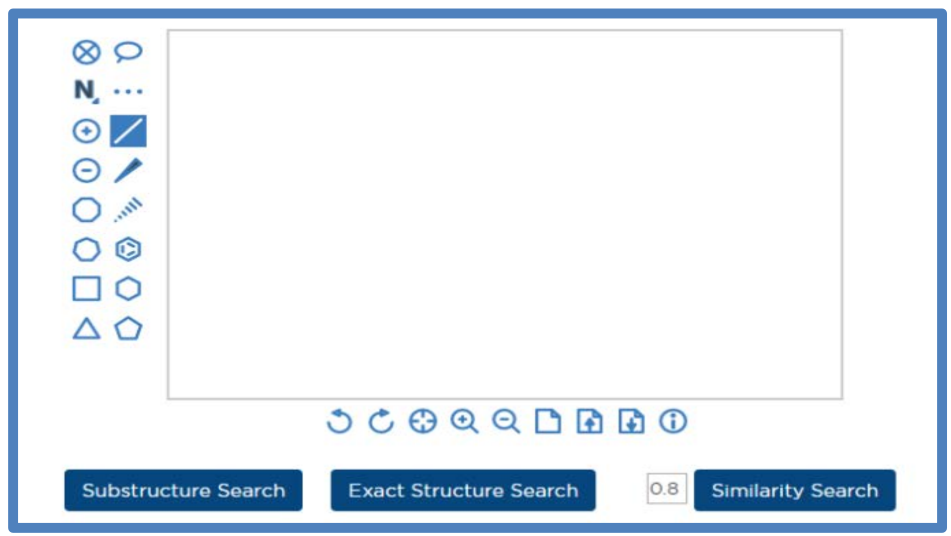


Direct-Manipulation Interfaces

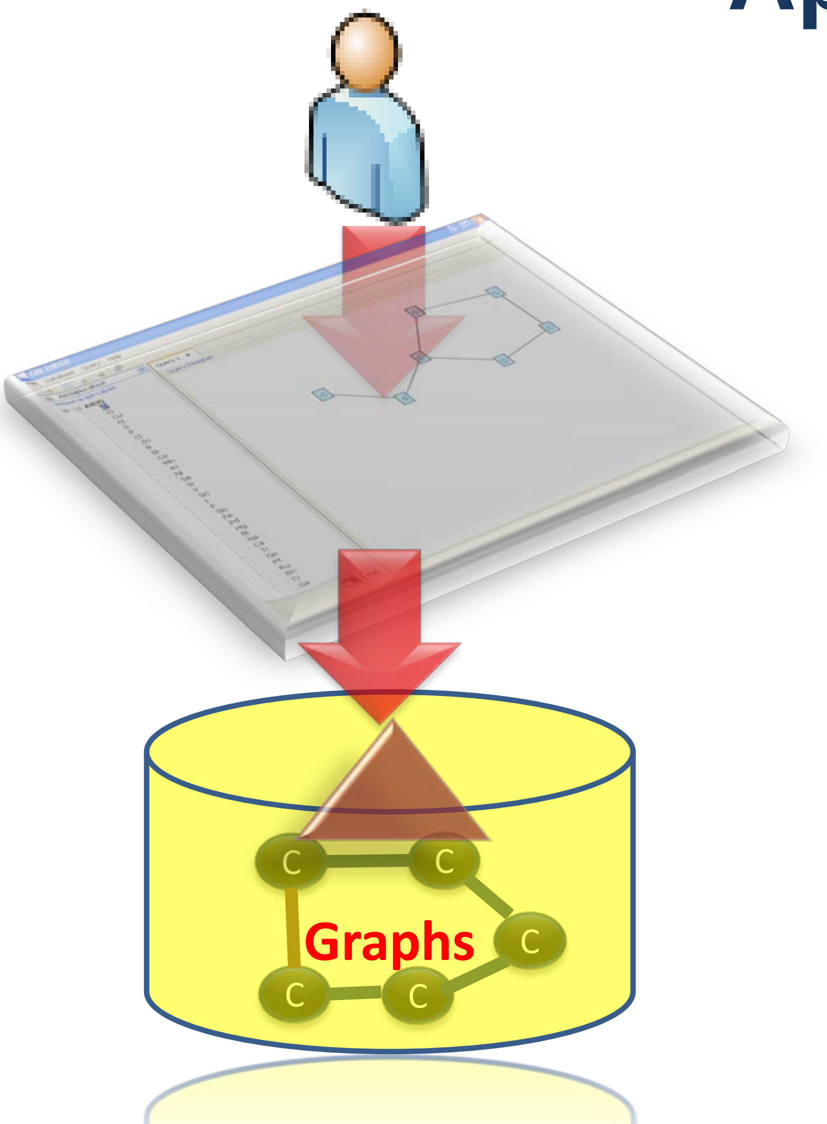
Pros

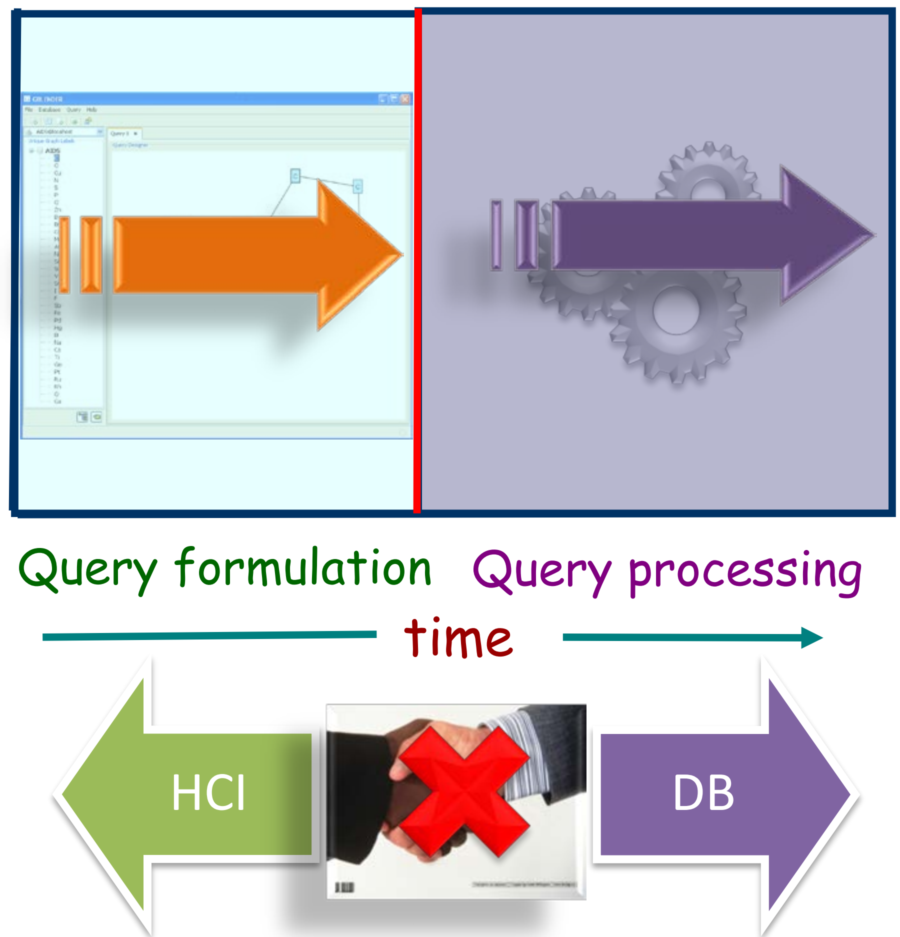
- Visually present task concepts
- Allows easy learning
- Allows easy retention
- Allows errors to be avoided
- Encourages exploration
- Appealing to novice, is easy to remember for intermittent users, and can be rapid for frequent users



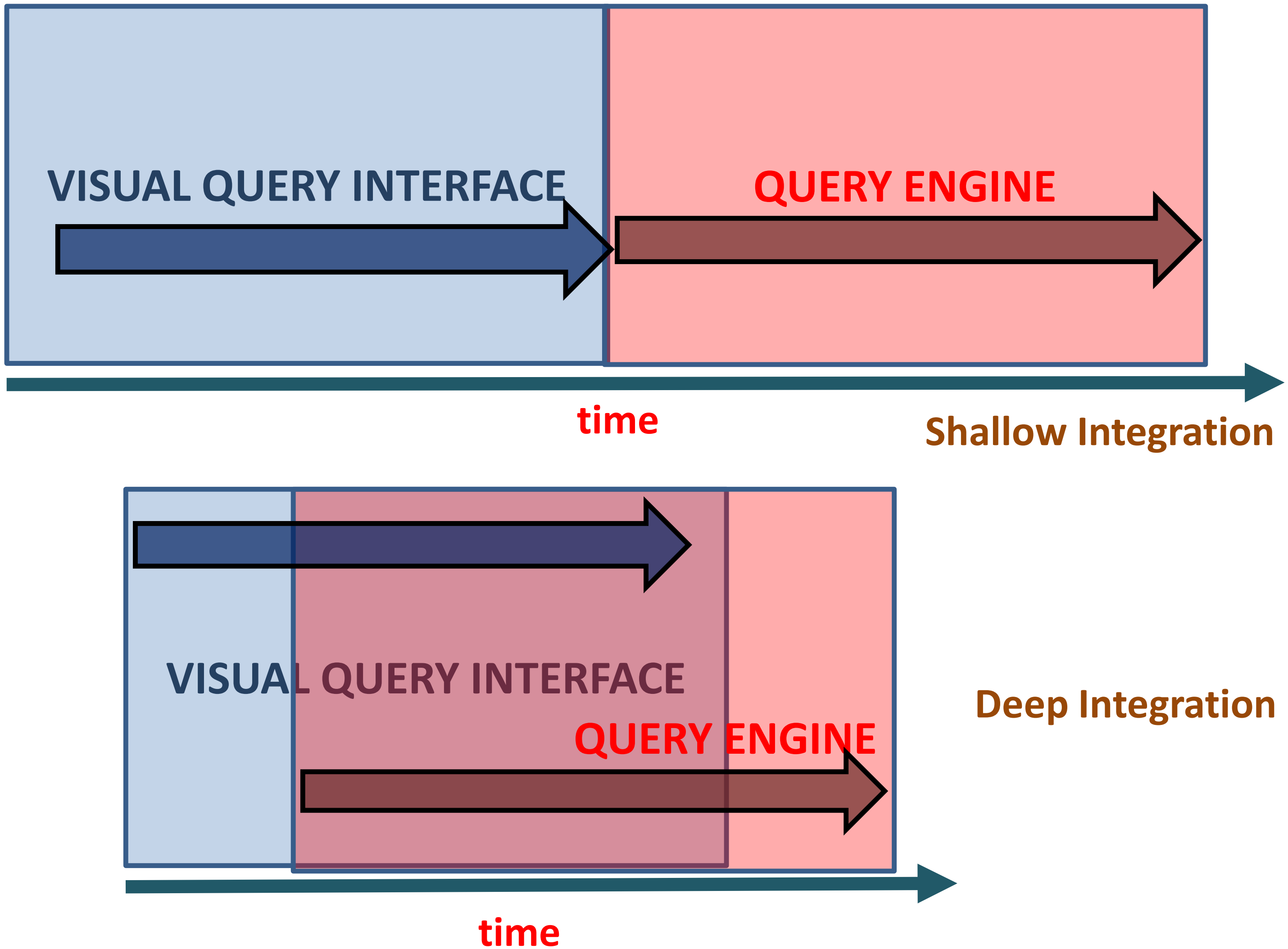


Classical Direct Manipulation-based Approach





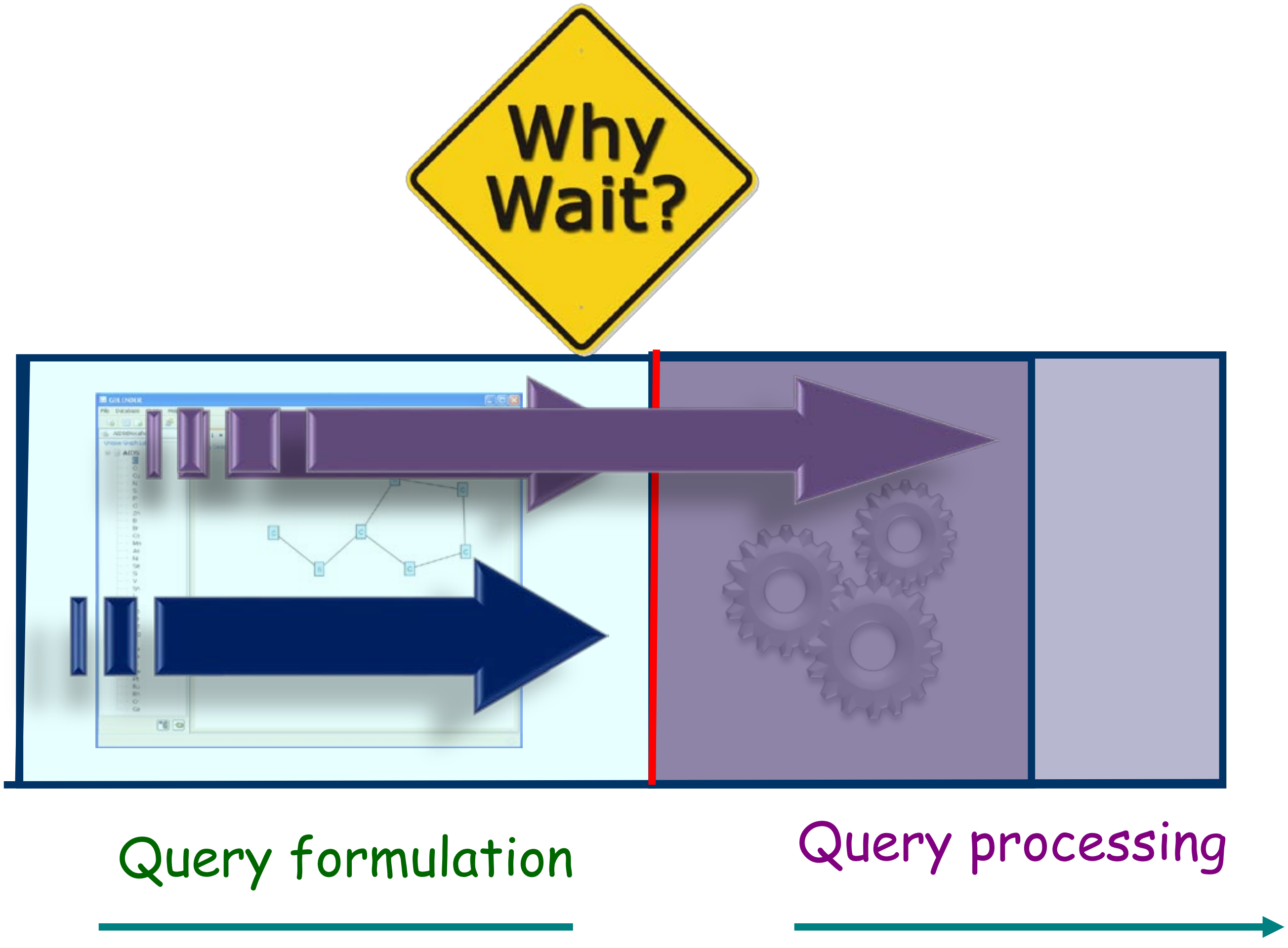
Taking Advantage of Direct-Manipulation Interaction Style



Benefits of Deep Integration

- 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]
- Interactive search and exploration [ICDE 19, VLDB 17]
- Interactive visualization of results [VLDB 17, VLDB J 14]

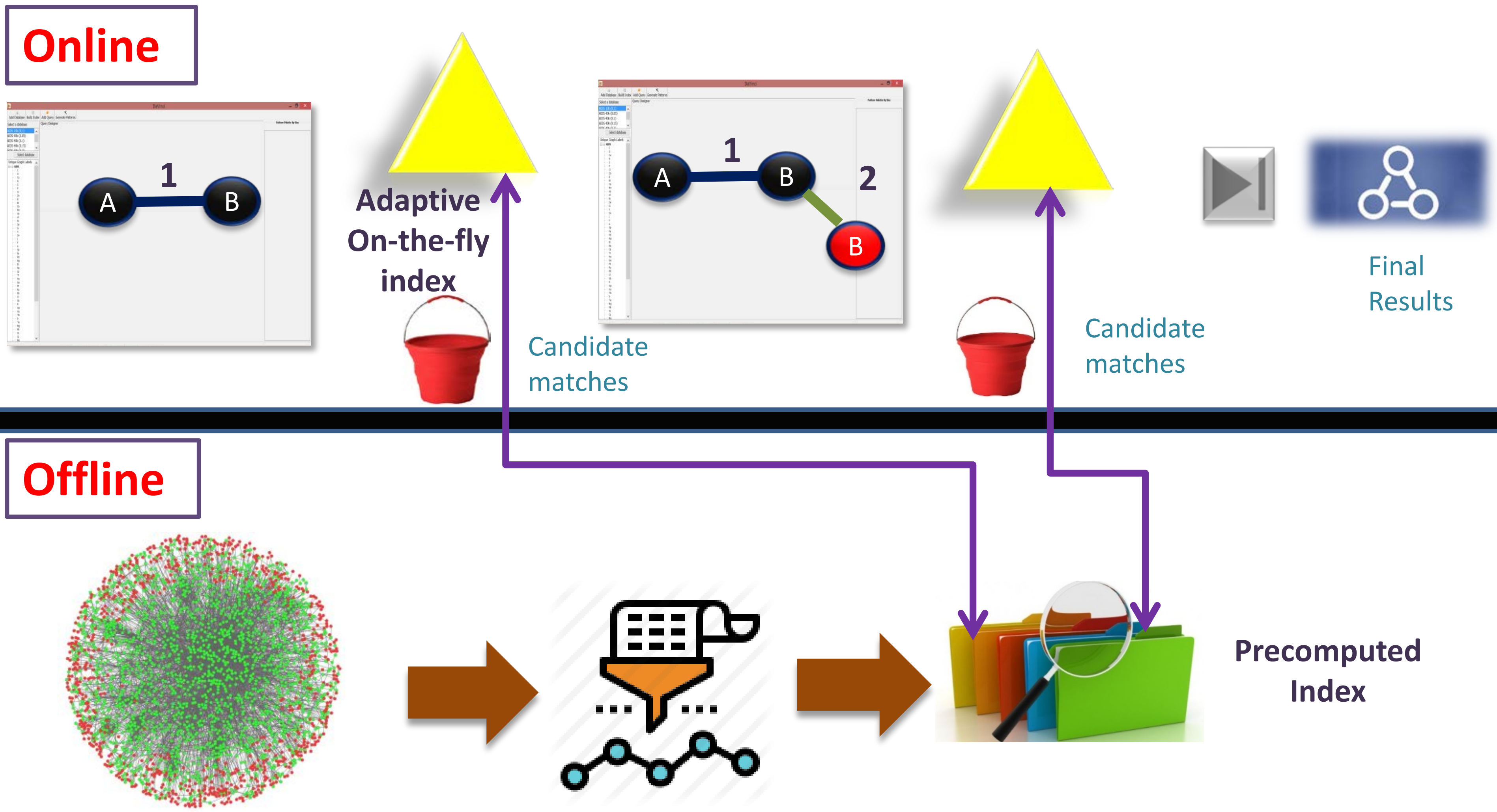
Faster Query Response Time



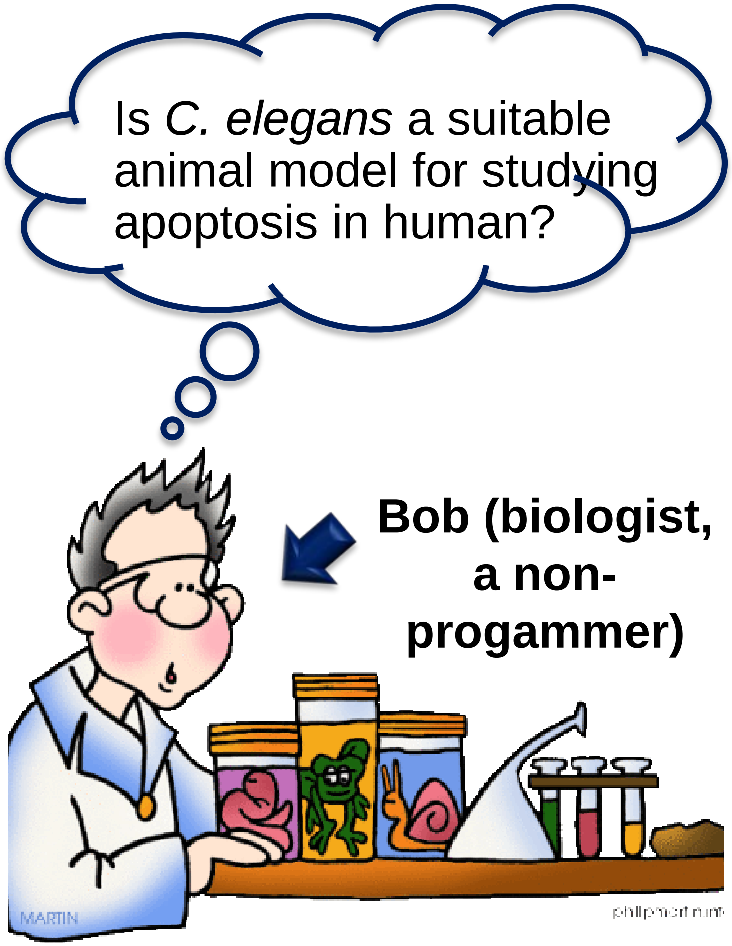
Non-traditional Challenges

- Partial query-aware indexing schemes
- Materialization of intermediate results
- Selectivity-free query processing
- Focus on waiting time of users

Graph Query Processing



An Example



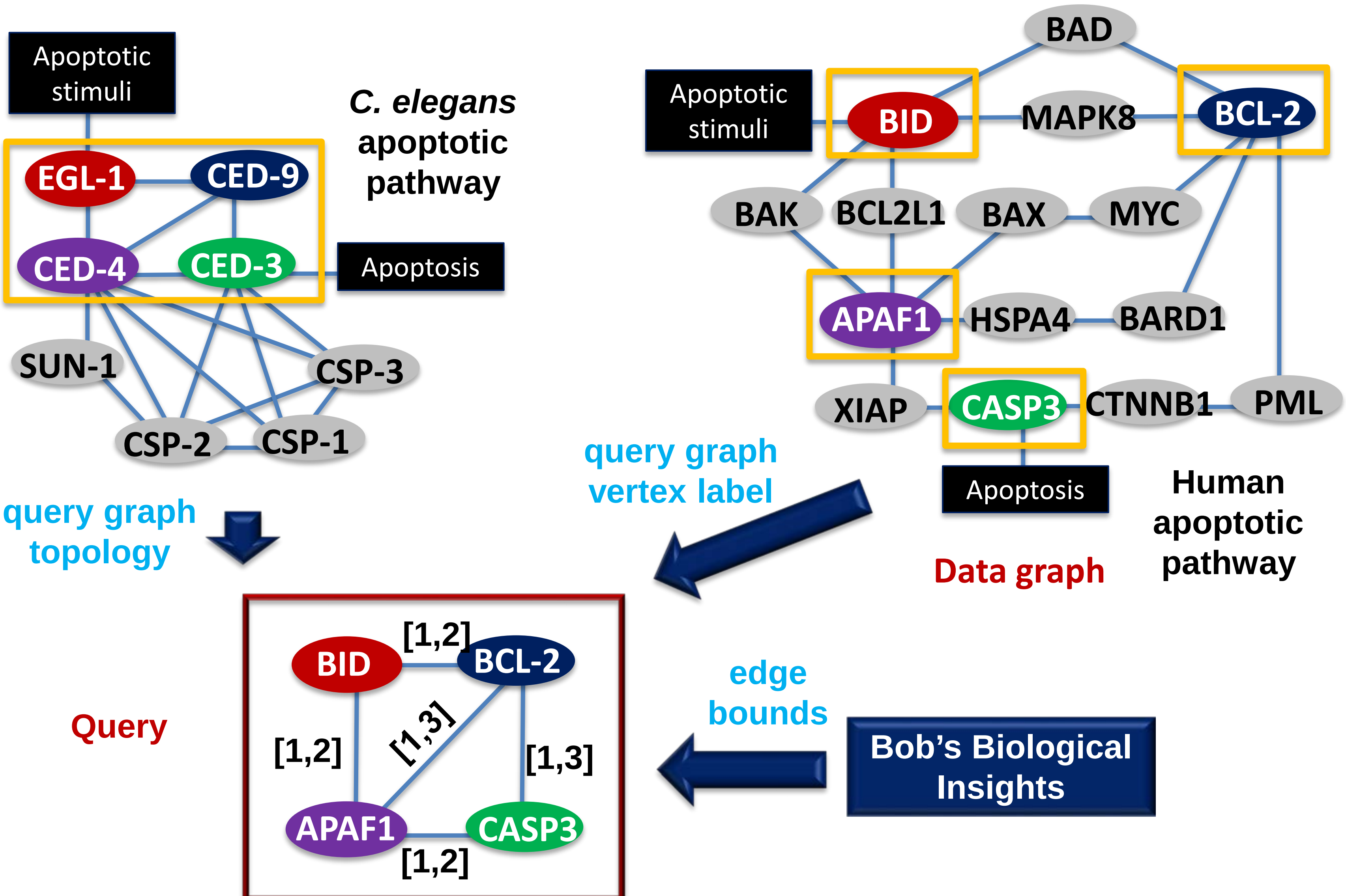
Bob's Biological Insights

C. elegans is likely to be a suitable model if **genes** related to apoptosis **behave similarly** to those that are in human

Due to **evolution**, models **behave differently**

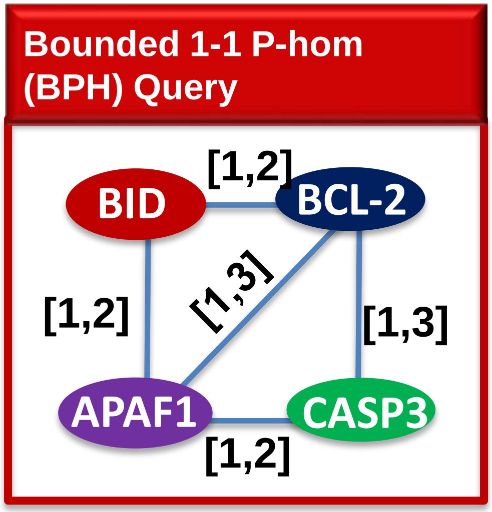
In **conserved** biological processes, **interacting gene pairs** are in **close proximity**

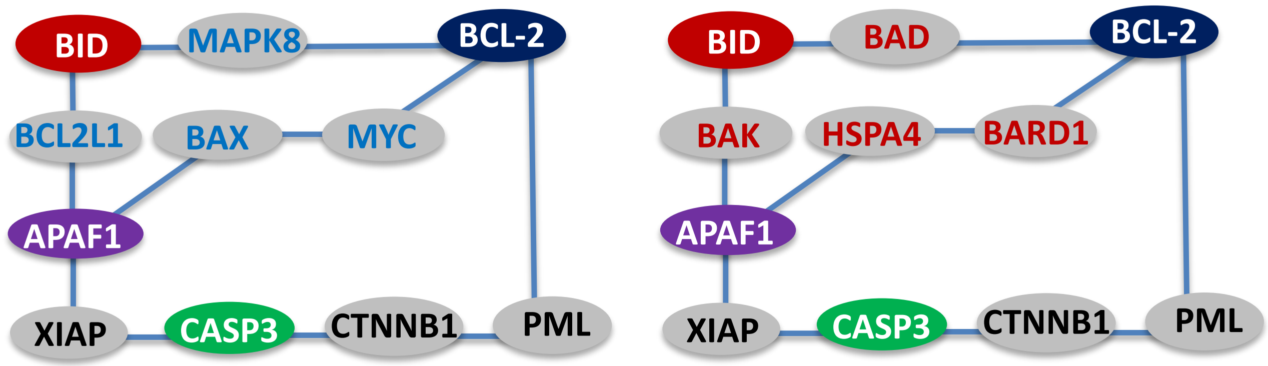
An Example



Apoptotic stimuli
EGL-1
CED-9
CED-4
CED-3
Apoptosis
SUN-1
CSP-3
CSP-2
CSP-1
C. elegans apoptotic pathway
query graph topology
Apoptotic stimuli
BAD
BID
MAPK8
BCL-2
BAK
BCL2L1
BAX
MYC
APAF1
HSPA4
BARD1
XIAP
CASP3
CTNNB1
PML
Apoptosis
Human apoptotic pathway
Data graph
query graph vertex label
Query
BID
BCL-2
APAF1
CASP3
[1,2]
[1,2]
[1,3]
[1,3]
[1,2]
edge bounds
Bob's Biological Insights

BPH Query





Result 1

Result 2

1-1 P-homomorphism (Fan et al., 2010)

G is **1-1 P-homomorphic** to G ... **1 injective** P-homomorphism from G to G'

Bounded length

- Distinct nodes have **distinct** matches
- **Edges** in G mapped to **paths (arbitrary length)** in G'
- Match vertices based on **vertex similarity**.

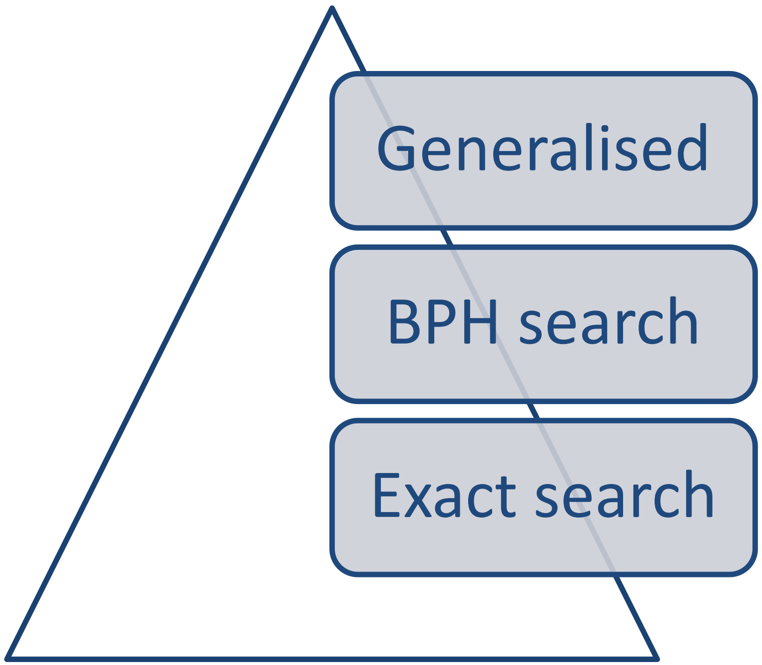
Label equality

W. Fan et al., Gra... ...sited for graph matching. VLDB, 2010

Visual BPH Search Problem

[SIGMOD 18]

Given a BPH query Q visually constructed on a visual query interface and a data graph G=(V,E,L), the goal of **visual bounded 1-1 p-hom search problem** is to retrieve all bounded 1-1 p-hom matches of Q in G by **interleaving (i.e., blending)** formulation and processing of Q.



BOOMER In Action

https://www.youtube.com/watch?v=zRYqo-F4xcg&feature=youtu.be

Exploratory Search on Graph DB

Visual Exploratory Search

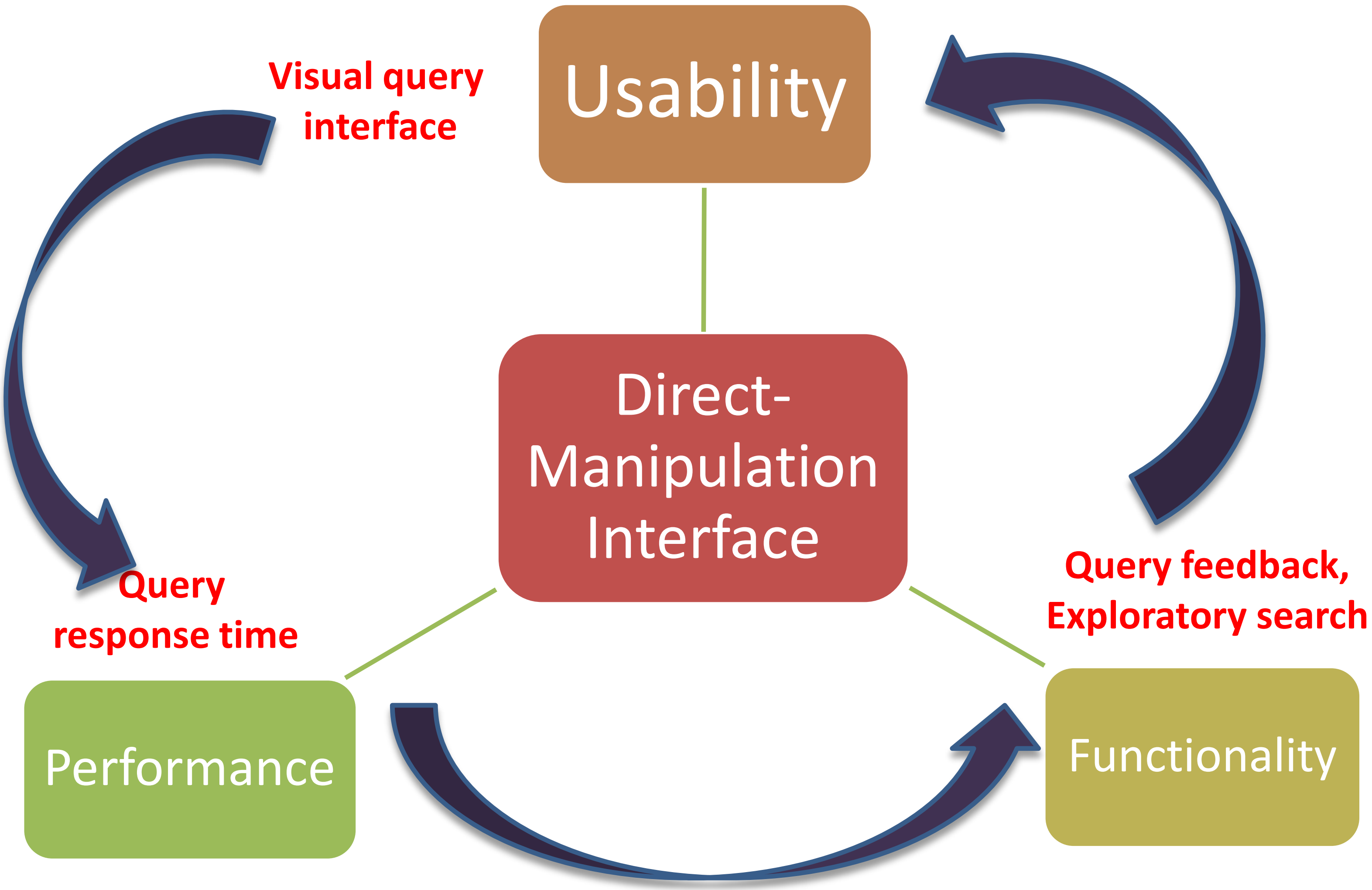
- Open ended, evolving in nature
- Search state is ambiguous in the beginning
- Multiple and iterative query formulation and execution
- Initial query may often grow in size
- Visual query interface to facilitate exploration

Let $\mathbb{A}$ be an exploratory action sequence (i.e., sequence of add(), modify(), run() actions) undertaken by a user on a visual interface for exploring a graph database $D = \{g_1, g_2, \ldots, g_n\}$. Then the goal of visual exploratory subgraph search (VESS) problem is to retrieve all the graphs $g_i \in D$ with $dist(g_i, q) \leq \delta$ for each $run(q) \in \mathbb{A}$ where δ is the subgraph similarity distance threshold.

PICASSO [ICDE 19, VLDB 2017]

https://www.youtube.com/watch?v=-_ogvGKRROY&feature=youtu.be

Summary


Visual query interface
Usability
Direct-Manipulation Interface
Query response time
Query feedback, Exploratory search
Performance
Functionality

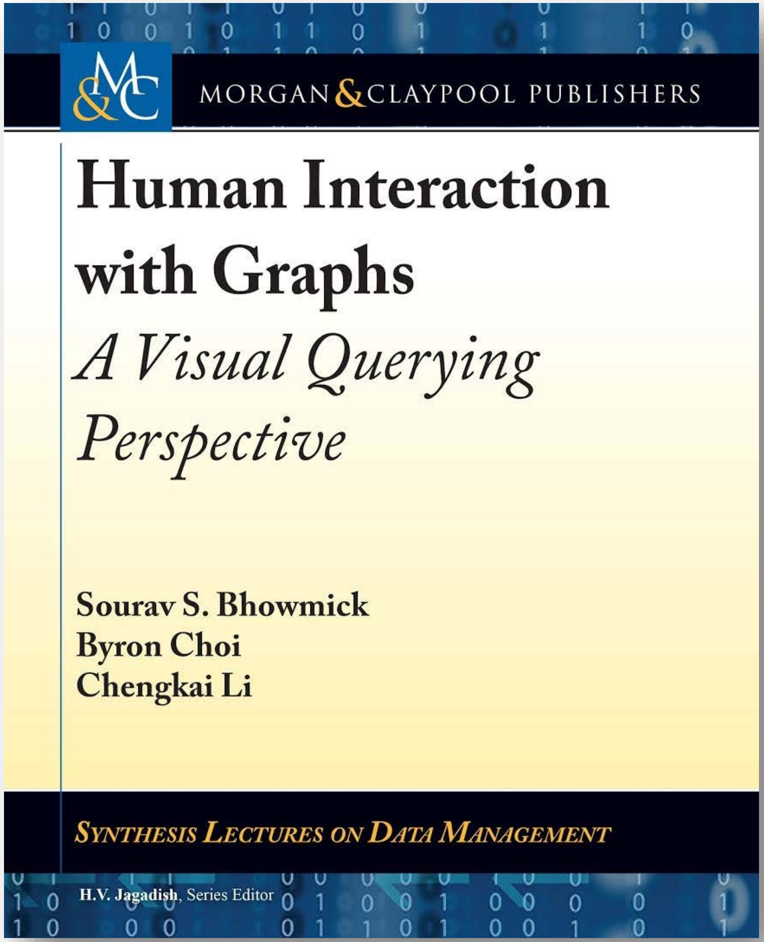
Open Problems

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

Final Remarks

Human-Graph Interaction

- Non-progammers are typically not going to invest resources in educating themselves with GQLs
- Usable graph search systems by exploiting direct manipulation-based interaction
- HCI, Cognitive Psychology & Data Management



HINT Project