

【NBIS 2014】

Frequent-Pattern based Facet Extraction from Graph Data

Takahiro Komamizu, Toshiyuki Amagasa, Hiroyuki Kitagawa
University of Tsukuba, Japan

Graph data

- General model to be able to represent real world information
 - e.g., social networks, chemical compounds
- Consist of vertex set and edge set
 - e.g., a user is a vertex and an edge is relationship between users in social networks
- Two classes of graph data: **a single large graph** and a set of multiple small graphs

Graph Data Search

- Find subgraphs matching with given query
 - e.g., user search on social networks, search for a bunch of co-authors in co-author network

- Querying

- Pattern query (e.g., SPARQL →)

- input: a desired pattern with variables
 - output: values of the variables

```
select ?x ?y
where{
  ?a foaf:knows ?b.
  ?a foaf:name ?x.
  ?b foaf:name ?y.
}
```

- Keyword query

- input: a set of keywords
 - results: subgraphs containing all keywords

Motivation

- Problems
 - Pattern query: users are expected to know
 - query languages of pattern query, and
 - structure of graph data.
 - Keyword query: getting appropriate result subgraphs is still difficult
- Supports to graph data search

Basic Idea

- Applying faceted search for graph data search over a single graph
 - need to extract objects (target subgraphs) and facets (attributes of objects)
- Extracting meaningful subgraphs as objects
 - e.g., frequent subgraphs

Related work

- [1] applies faceted search to construct SPARQL queries by selecting predicate and subject.
- [2] gives graphical interface to construct chemical compound pattern queries.
 - the dataset consists of a set of graphs

[1] E. Oren, R. Delbru, and S. Decker, “Extending Faceted Navigation for RDF Data,” in Proc. International Semantic Web Conference, 2006, pp.559–572.

[2] C. Jin, S. S. Bhowmick, X. Xiao, J. Cheng, and B. Choi, “GBLENDER: Towards Blending Visual Query Formulation and Query Processing in Graph Databases,” in Proc. SIGMOD Conference, 2010, pp. 111–122.

Faceted Search

- One of the exploratory searches
- Search process
 1. (system) shows current results, associated facets and values of the facets.
 2. (user) selects one of the values of the facets.
 3. continue.
- Real applications
 - DBLP, eBay, Amazon, etc.

Faceted Search: an example

Facet

		Make	Count
		Honda	3
		Toyota	2
		Suzuki	2

Year	Count	Color	Count
2009	3	Red	3
2010	1	Blue	3
2011	2	Black	1

Car database

Make	Year	Color
Honda	2011	Red
Honda	2009	Blue
Honda	2009	Black
Toyota	2010	Blue
Toyota	2009	Red
Suzuki	2011	Red
Suzuki	2010	Blue

Frequent Subgraph Mining

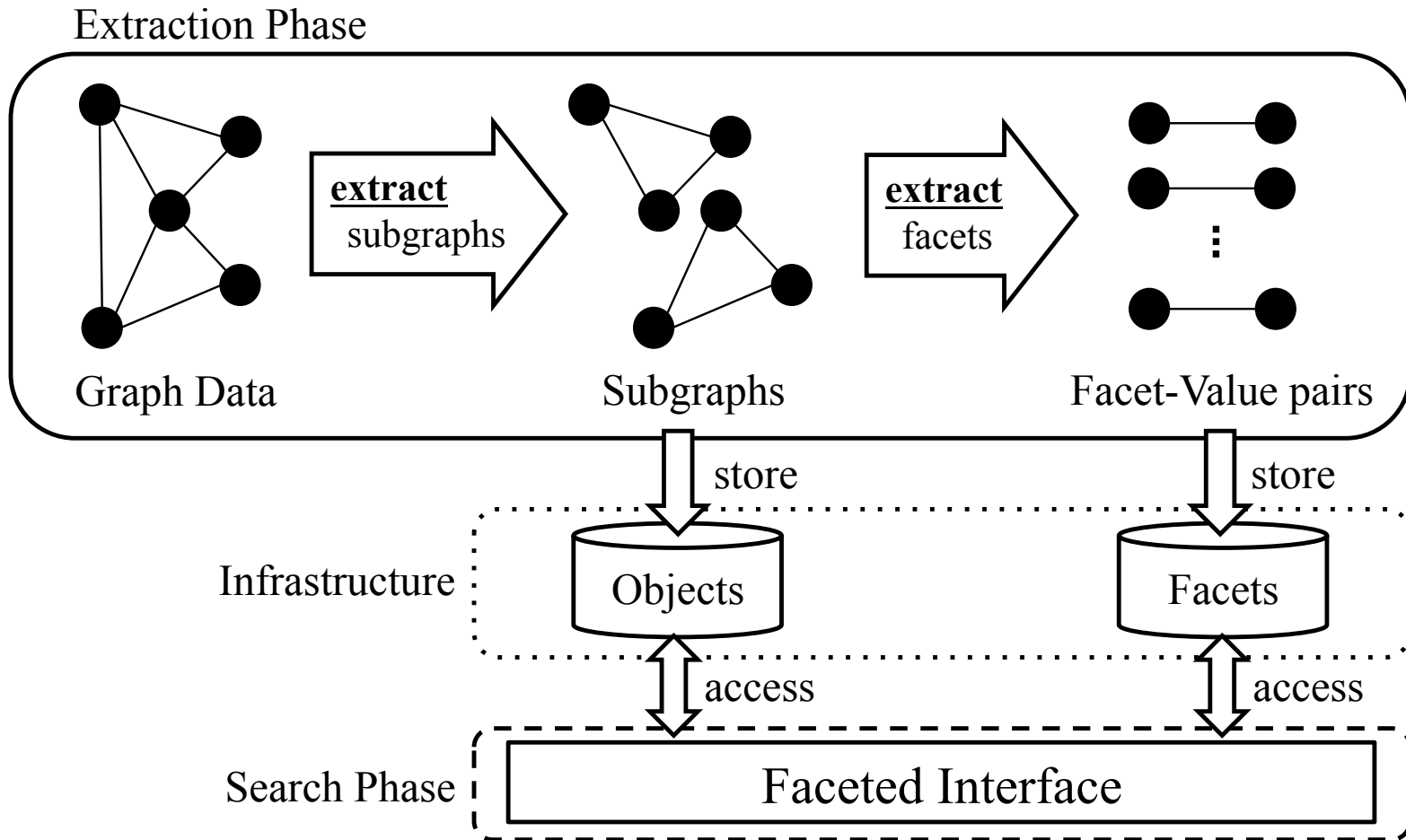
- A technique extracting frequently occurring subgraphs in graph data
- A subgraph in a frequent subgraph is also frequent subgraphs
 - ➔ extract maximal frequent subgraphs
- Existing work [3, 4, 5]

[3] L. B. Holder, D. J. Cook, and S. Djoko, “Substructure Discovery in the SUBDUE System,” in Proc. KDD Workshop, 1994, pp. 169–180.

[4] S. Ghazizadeh and S. S. Chawathe, “SEuS: Structure Extraction Using Summaries,” in Proc. Discovery Science, 2002, pp. 71–85.

[5] F. Zhu, Q. Qu, D. Lo, X. Yan, J. Han, and P. S. Yu, “Mining Top-K Large Structural Patterns in a Massive Network,” PVLDB, vol. 4, no. 11, pp. 807–818, 2011.

Proposed Framework



Infrastructure

- Relational database to store object and facet information

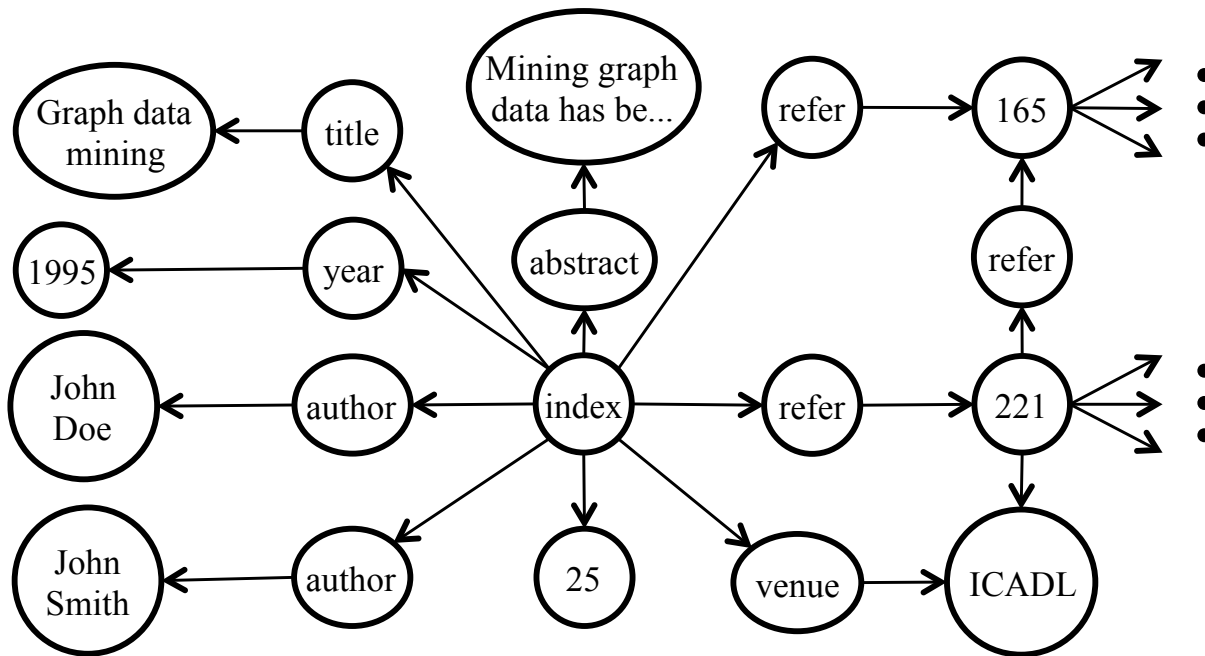
`object(object_id, object_instance)`

`facet_name(value, object_id)`

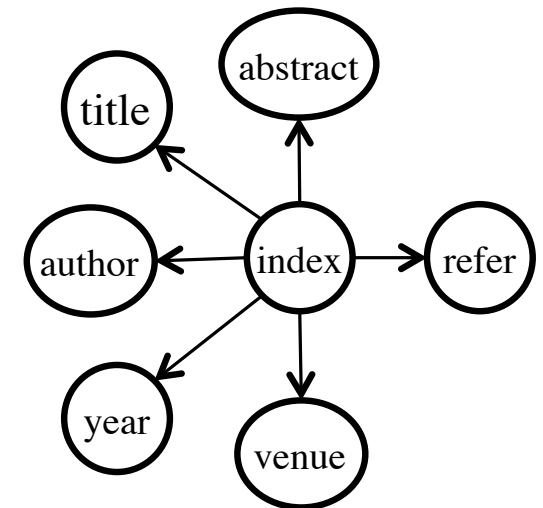
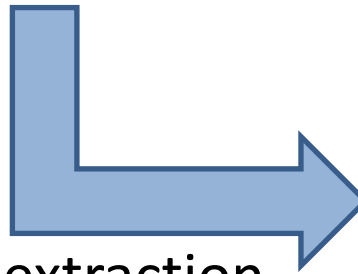
- With this info., subjects are searchable by SQL language.

```
SELECT    object.object_instance
FROM      object_instance, year
WHERE     year.value = '2014'
          and year.object_id = object.object_id
```

Case study: citation network



Frequent subgraph extraction



Case study: citation network (cont.)

Faceted Search Interface over Graph Data

1572279 results	
<p>author</p> <ul style="list-style-type: none"> Wei Wang (864) Philip S. Yu (609) Chin-Chen Chang (579) Elisa Bertino (572) Thomas S. Huang (564) Wei Zhang (539) Lei Zhang (536) <p>more</p>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "173881") (index, title, "Piccola - A Small Composition Language") (index, year, "1999") (index, venue, "ECOOP Workshops") (index, author, "Oscar Nierstrasz")</p> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "23835") (index, title, "Cellular Automata with Majority Rule on Evolving Network.") (index, year, "2004") (index, venue, "ACRI"), (index, author, "Danuta Makowiec")</p> </div>
<p>venue</p> <ul style="list-style-type: none"> ISCAS (10780) ICRA (10688) ICIP (10468) IEICE Transactions (10462) Discrete Mathematics (10315) Commun. ACM (9400) Theor. Comput. Sci. (8524) <p>more</p>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "897238") (index, title, "Projector-Based Color Simulator for Print Industry.") (index, year, "2006") (index, venue, "IEICE Transactions") (index, author, "Kumiko Ueda") (index, author, "Norimichi Tsumura") (index, author, "Shoji Yamamoto") (index, author, "Toshiya Nakaguchi")</p> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "394958") (index, title, "Adversary Centered Design: Threat Modeling Using Anti-Scenarios, Anti-Use Cases and Anti-Personas.") (index, year, "2008") (index, venue, "IKE") (index, author, "Adam Steele") (index, author, "Xiaoping Jia")</p> </div>
<p>year</p> <ul style="list-style-type: none"> 2009 (155299) 2008 (146714) 2007 (135277) 2010 (129273) 2006 (126537) 2005 (114033) 2004 (95725) <p>more</p>	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "570860") (index, title, "Safety and Security Issues in Electric Power Industry.") (index, year, "2000") (index, venue, "SAFECOMP") (index, refer, "570803") (index, author, "Bartosz Nowicki") (index, author, "Janusz Górski")</p> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "1319988") (index, title, "Strong Simultaneous Stabilization for a Class of Generalized Linear System.") (index, year, "2009") (index, venue, "HIS") (index, author, "Jing Ma") (index, author, "Kun Ma") (index, abstract, "This paper investigate stable right...")</p> </div>
	<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> <p>(index,, "575185") (index, title, "Traffic-based Load Balance for Scalable Network Emulation.")</p> </div> <div style="border: 1px solid black; padding: 5px;"> <p>(index,, "325879") (index, title, "Object tracking based on area weighted centroids shifting with spatiality constraints.")</p> </div>

Conclusion and Future works

- Conclusion
 - Faceted search framework for graph data
 - Objects are extracted using frequent subgraph mining approach.
 - Case studies on simpler data
 - citation network and review network
- Future work
 - enabling faceted search for more complex graph data, e.g., multiple connectable objects
 - interface design

**THANK YOU FOR
YOUR ATTENTION**