

Design of an InfoVis system

Example: Co-authorship network

Alessandra Tappini

Co-authorship network

- Each node is an author
- There is an edge between two nodes if the corresponding authors share at least one publication (aka paper)

Co-authorship network

- Each node is an author
- There is an edge between two nodes if the corresponding authors share at least one publication (aka paper)
- Example:

Nodes: {Tappini, Liotta, Piselli}

Edges: {(Tappini, Liotta), (Tappini, Piselli), (Liotta, Piselli)}

Co-authorship network

- Each node is an author
- There is an edge between two nodes if the corresponding authors share at least one publication (aka paper)
- Example:

Nodes: {Tappini, Liotta, Piselli}

Edges: {(Tappini, Liotta), (Tappini, Piselli), (Liotta, Piselli)}

The edge (Tappini, Liotta) represents papers whose titles are

- *Comparative Study and Evaluation of Hybrid Visualizations of Graphs*
- *Nonplanar Graph Drawings with k Vertices per Face*

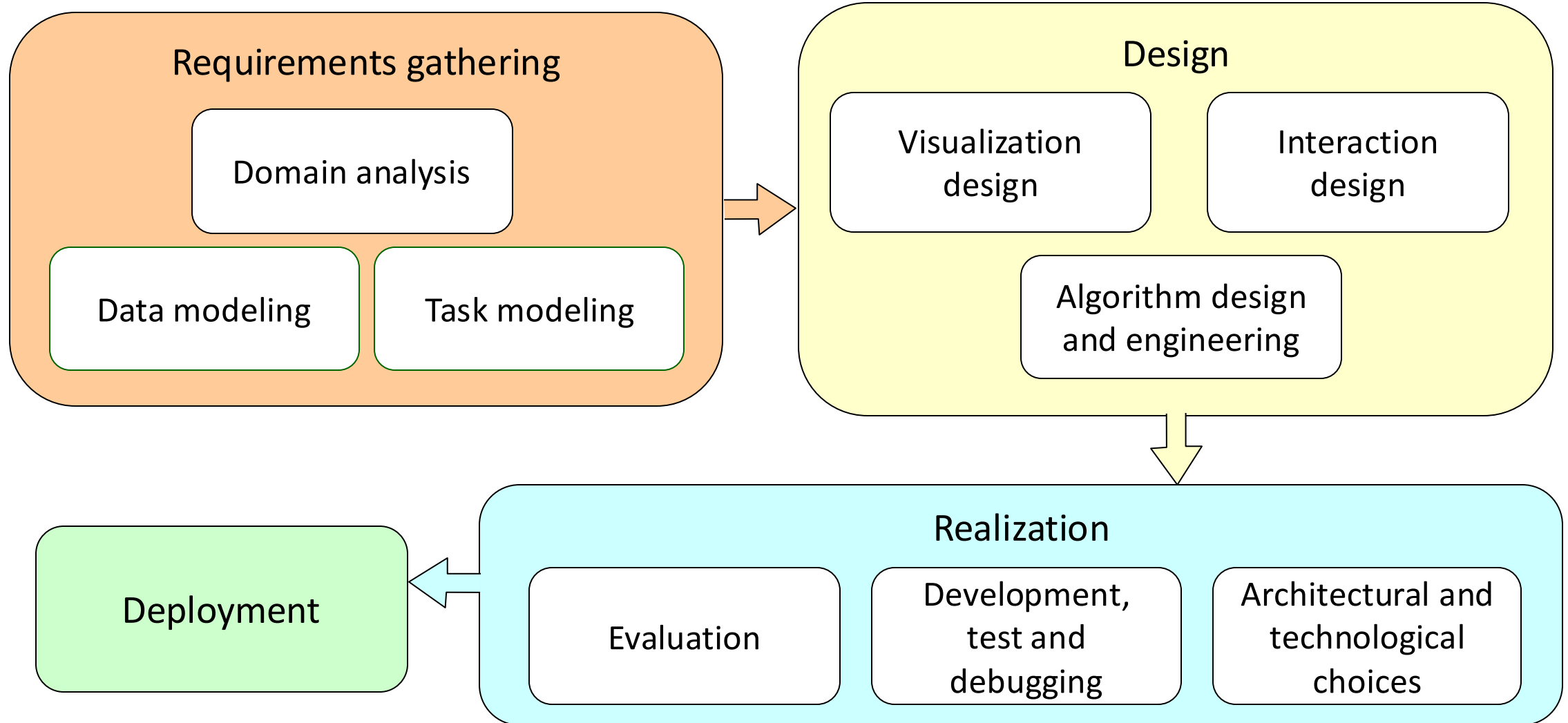
The edge (Tappini, Piselli) represents a paper whose title is

- *On the Parameterized Complexity of s -club Cluster Deletion Problems*

The edge (Liotta, Piselli) represents a paper whose title is

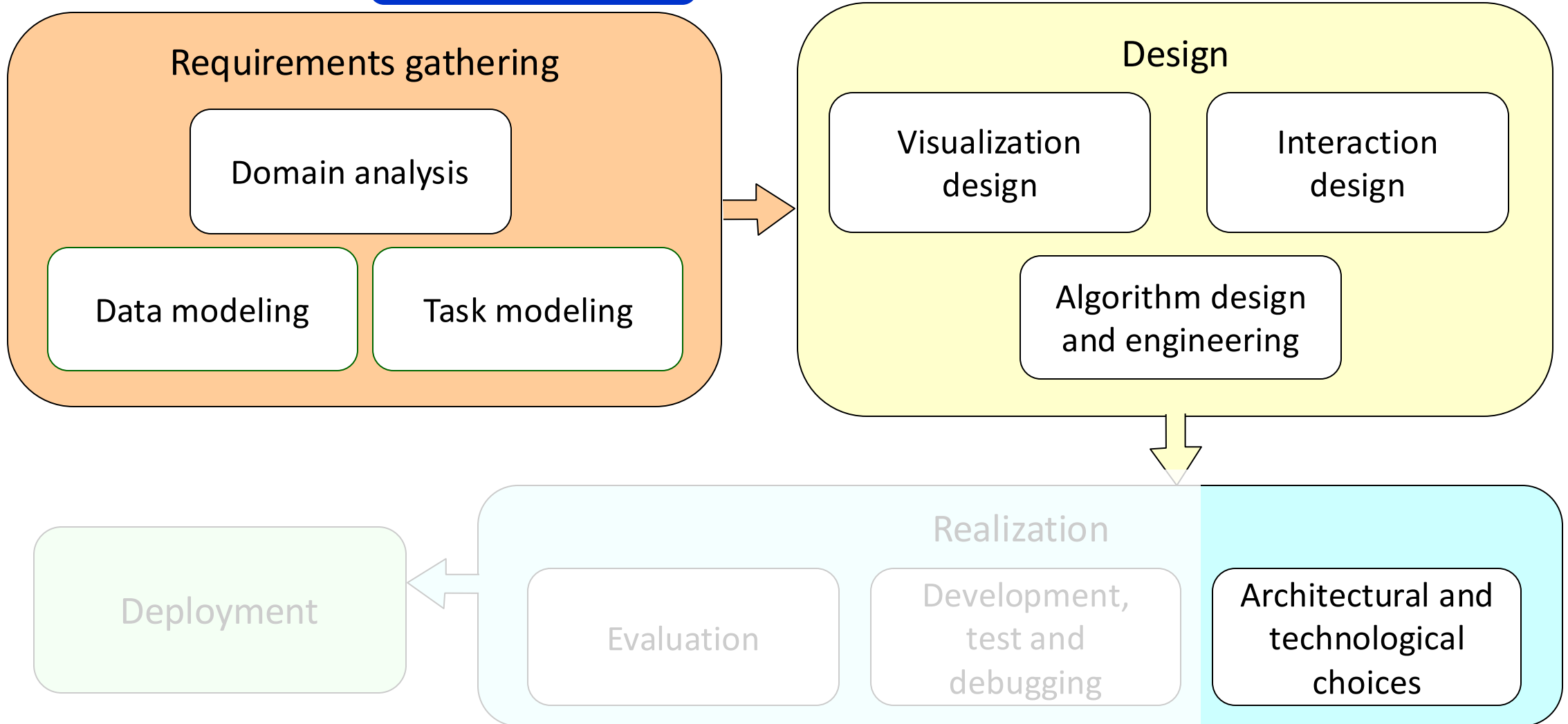
- *On the Parameterized Complexity of Computing st -Orientations with Few Transitive Edges*

Design phases



Design phases

THESE SLIDES



Design phases

1. Identify the target users and their needs
2. Translate from domain-specific language to visualization language
 - A. Data modeling
 - B. Task modeling
3. Design visual encoding idioms and interaction idioms
4. Device efficient algorithms
5. Make architectural and technological choices

Design phases

- 1. Identify the target users and their needs**
2. Translate from domain-specific language to visualization language
 - A. Data modeling
 - B. Task modeling
3. Design visual encoding idioms and interaction idioms
4. Device efficient algorithms
5. Make architectural and technological choices

1. Identify the target users and their needs

Users are researchers who want to analyze data from some editions of a conference.

- Identify the most prolific authors
- Identify the authors who have the most collaborations
- Analyze the collaborations between authors with different affiliations (i.e., university or research institute)
- etc.

Design phases

1. Identify the target users and their needs
2. Translate from domain-specific language to visualization language
 - A. **Data modeling**
 - B. Task modeling
3. Design visual encoding idioms and interaction idioms
4. Device efficient algorithms
5. Make architectural and technological choices

2.A. Data modeling

Phases:

- Data extraction
- Data cleaning and transformation
- Definition of dataset types and attribute types

Data extraction

- “The **dblp** computer science bibliography provides open bibliographic information on major computer science journals and proceedings”
- dblp provides a search engine, which allows the user to execute queries and retrieve the information of interest
- dblp arranges the retrieved information on an xml file, which contains a list of publications

Example:

- URL for query:
https://dblp.org/search/publ/api?q=streamid%3Aconf%2Fgd%3A%20type%3AConference_and_Workshop_Papers%3A%20year%3A2021%3A|%20year%3A2022%3A&h=1000

Data extraction

- Extract of xml file:

```
▼<hit score="3" id="593923">
  ▼<info>
    ▼<authors>
      <author pid="b/GiuseppeDiBattista">Giuseppe Di Battista</author>
      <author pid="38/5614">Walter Didimo</author>
      <author pid="40/2522">Luca Grilli 0001</author>
      <author pid="327/9301">Fabrizio Grosso</author>
      <author pid="225/7828">Giacomo Ortali</author>
      <author pid="83/321">Maurizio Patrignani</author>
      <author pid="192/0799">Alessandra Tappini</author>
    </authors>
    <title>Small Point-Sets Supporting Graph Stories.</title>
    <venue>GD</venue>
    <pages>289–303</pages>
    <year>2022</year>
    <type>Conference and Workshop Papers</type>
    <access>closed</access>
    <key>conf/gd/BattistaDGGOPT22</key>
    <doi>10.1007/978-3-031-22203-0_21</doi>
    <ee>https://doi.org/10.1007/978-3-031-22203-0_21</ee>
    <url>https://dblp.org/rec/conf/gd/BattistaDGGOPT22</url>
  </info>
  <url>URL#593923</url>
</hit>
▼<hit score="3" id="593924">
  ▼<info>
    ▼<authors>
      <author pid="06/1457">Michael A. Bekos</author>
      <author pid="09/2980">Stefan Felsner</author>
      <author pid="130/4006">Philipp Kindermann</author>
      <author pid="98/2841">Stephen G. Kobourov</author>
      <author pid="31/6569">Jan Kratochvíl</author>
      <author pid="99/44">Ignaz Rutter</author>
    </authors>
    <title>The Rique-Number of Graphs.</title>
    <venue>GD</venue>
    <pages>371–386</pages>
    <year>2022</year>
    <type>Conference and Workshop Papers</type>
    <access>closed</access>
    <key>conf/gd/BekosFKKKR22</key>
    <doi>10.1007/978-3-031-22203-0_27</doi>
    <ee>https://doi.org/10.1007/978-3-031-22203-0_27</ee>
    <url>https://dblp.org/rec/conf/gd/BekosFKKKR22</url>
  </info>
  <url>URL#593924</url>
</hit>
```

Data cleaning and transformation

- From the xml file, keep only the information that is of interest
 - For each publication: the title, the authors' names, and their affiliations
- Make authors' names and affiliations homogeneous
- Parse the xml file to model the data as a network
 - Each node is equipped with the name of the author it represents, the author's affiliation, and the number of the author's publications
 - Each edge is equipped with a list of the titles of the papers shared by its end-nodes and the number of shared publications
- Create a json file, which contains a list of nodes and a list of edges

Note: the xml file does not contain author's affiliations, which need to be retrieved by a different data source

Data cleaning and transformation

- Extract of json file:

```
"directed":0,
"weighted":1,
"nodes":[
  {"id":90, "label":"W. Didimo",
   "affiliation":"University of Perugia","weight":2},
  {"id":73, "label":"F. Grosso",
   "affiliation":"Roma Tre University", "weight":1},
  ...
],
"edges":[
  {"source":90, "target":73, "label":"Small Point-
Sets Supporting Graph Stories.", "value":1},
  {"source":90, "target":88, "label":"1: Small
Point-Sets Supporting Graph Stories.\n2: A User
Study on Hybrid Graph Visualizations.", "value":2},
  ...
]
```

Definition of dataset types and attribute types

Dataset types:

- Network → Items (nodes), links [cardinality: 90 nodes, 300 links]
- Clusters → Items (nodes) representing groups of authors with the same affiliation [cardinality: 10 = 9 main affiliations + “other”]

Attribute types for items:

- Affiliation: Categorical [cardinality: 10 levels]
- Number of publications: Ordered (quantitative) [cardinality: range from 1 to 6]

Attribute types for links:

- Number of shared publications: Ordered (quantitative) [cardinality: range from 1 to 3]

Design phases

1. Identify the target users and their needs
2. Translate from domain-specific language to visualization language
 - A. Data modeling
 - B. Task modeling**
3. Design visual encoding idioms and interaction idioms
4. Device efficient algorithms
5. Make architectural and technological choices

2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Input tasks:

1. Find the author whose name is *<Name>*
2. Find the authors having the same affiliation as *<Name>*
3. Identify the most prolific authors
4. Identify the authors who have the most collaborations
5. Identify the authors who collaborate with many authors with different affiliations
6. Determine whether *<NameA>* shares a publication with *<NameB>*
7. Determine whether *<NameA>* and *<NameB>* have a common co-author
8. Determine whether there is a strong collaboration between authors having *<AffiliationA>* and authors having *<AffiliationB>*

2.B. Task modeling

Define each task as an {**action**, **target**} pair.





Output tasks:

1. Find the author whose name is *<Name>*



Locate the **node** *<Name>*

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

2.B. Task modeling

Define each task as an {**action**, **target**} pair.





Output tasks:

2. Find the authors having the same affiliation as *<Name>*



Explore the **nodes** having the same categorical attribute as *<Name>*

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

2.B. Task modeling

Define each task as an {**action**, **target**} pair.





Output tasks:

3. Identify the most prolific authors



Locate the **nodes** with the largest value of quantitative attribute

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

3. Identify the most prolific authors







Locate the **nodes** with the largest value of quantitative attribute



Compare the **quantitative attribute** of nodes

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

➔ Query

➔ Identify



➔ Compare



➔ Summarize



2.B. Task modeling

Define each task as an {**action**, **target**} pair.





Output tasks:

4. Identify the authors who have the most collaborations



Locate the **nodes with higher degree**

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

4. Identify the authors who have the most collaborations







Locate the **nodes with higher degree**



Compare the **number of links incident to nodes**

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

➔ Query

➔ Identify



➔ Compare



➔ Summarize



2.B. Task modeling

Define each task as an {**action**, **target**} pair.





Output tasks:

5. Identify the authors who collaborate with many authors with different affiliations



Locate the **nodes with many links to nodes with different categ. attr.**

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

5. Identify the authors who collaborate with many authors with different affiliations







Locate the **nodes with many links to nodes with different categ. attr.**



Browse the **nodes having different categorical attribute**

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

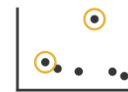
6. Determine whether $\langle \text{NameA} \rangle$ shares a publication with $\langle \text{NameB} \rangle$



Identify the **link** between node $\langle \text{NameA} \rangle$ and node $\langle \text{NameB} \rangle$

→ Query

→ Identify



→ Compare



→ Summarize



2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

6. Determine whether $\langle \text{NameA} \rangle$ shares a publication with $\langle \text{NameB} \rangle$







Identify the **link** between node $\langle \text{NameA} \rangle$ and node $\langle \text{NameB} \rangle$



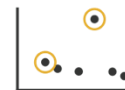
Locate the **nodes** $\langle \text{NameA} \rangle$ and $\langle \text{NameB} \rangle$

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

➔ Query

➔ Identify



➔ Compare



➔ Summarize



2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

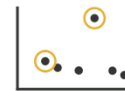
7. Determine whether $\langle \text{NameA} \rangle$ and $\langle \text{NameB} \rangle$ have a common co-author



Identify a **path of length 2** between node $\langle \text{NameA} \rangle$ and node $\langle \text{NameB} \rangle$

→ Query

→ Identify



→ Compare



→ Summarize



2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

7. Determine whether $\langle \text{NameA} \rangle$ and $\langle \text{NameB} \rangle$ have a common co-author







Identify a path of length 2 between node $\langle \text{NameA} \rangle$ and node $\langle \text{NameB} \rangle$



Locate the nodes $\langle \text{NameA} \rangle$ and $\langle \text{NameB} \rangle$

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

➔ Query

➔ Identify



➔ Compare



➔ Summarize



2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

8. Determine whether there is a strong collaboration between authors having *<AffiliationA>* and authors having *<AffiliationB>*



Identify the links between nodes with categorical attribute *<AffiliationA>* and nodes with categorical attribute *<AffiliationB>*

2.B. Task modeling

Define each task as an {**action**, **target**} pair.

Output tasks:

8. Determine whether there is a strong collaboration between authors having *<AffiliationA>* and authors having *<AffiliationB>*



Identify the links between nodes with categorical attribute *<AffiliationA>* and nodes with categorical attribute *<AffiliationB>*



Browse the nodes with categorical attribute *<AffiliationA>* and the nodes with categorical attribute *<AffiliationB>*

Design phases

1. Identify the target users and their needs
2. Translate from domain-specific language to visualization language
 - A. Data modeling
 - B. Task modeling
- 3. Design visual encoding idioms and interaction idioms**
4. Device efficient algorithms
5. Make architectural and technological choices

3. Design visual encoding idioms and interaction idioms

Visual encoding idioms:

- Identify marks and channels and associate a semantics to them
- Define one or more visual idioms

Interaction idioms:

- Identify the interaction operations
- Design an interaction paradigm that makes the interaction operations intuitive, effective, and efficient

Marks

- Items (nodes) → circle (with label reporting author's name)
- Links → connection (segment)

Channels

Channels for items (nodes):

- Categorical attribute → circle color hue (suitable to the cardinality)
Another possible option: shape
- Quantitative attribute → circle size (suitable to the cardinality)
Another possible option: color saturation (sequential color map)

Channels for links:

- Quantitative attribute → segment width (suitable to the cardinality)
Another possible option: color saturation (sequential color map)

Semantics

Items (nodes):

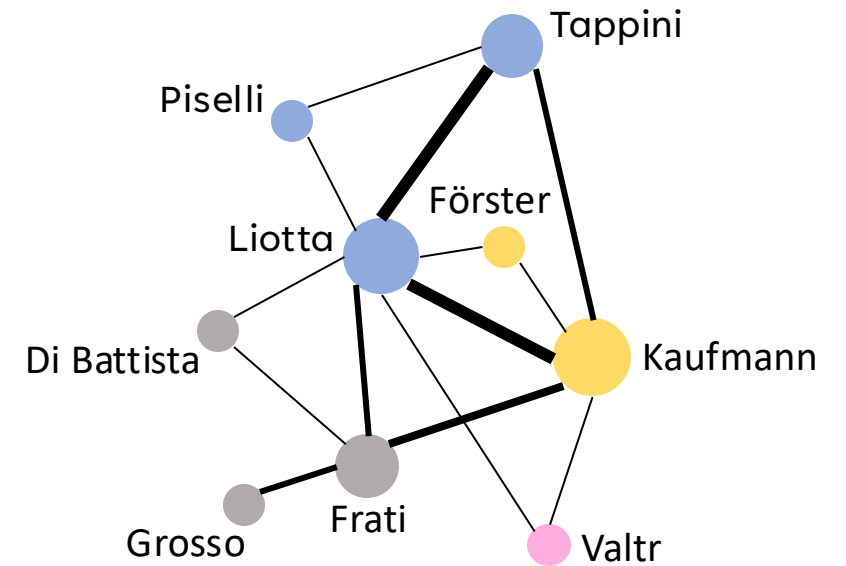
- The color hue of the circle encodes the author's affiliation
- The size of the circle is proportional to the author's number of publications

Links:

- The width of the segment is proportional to the number of publications shared by the authors (nodes) connected by the segment

Visual encoding idiom

Unconstrained straight-line drawing

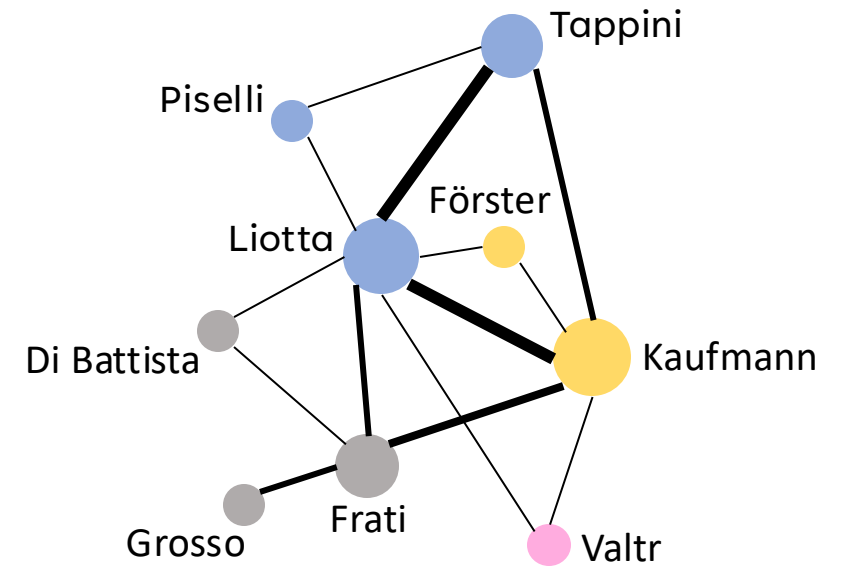


Visual encoding idiom

Unconstrained straight-line drawing

Why?

- Need to understand topology
- Suitable to the network size
- Links are represented as segments
- The position of nodes is not used to encode any information
 - even if we would like nodes in the same cluster (authors with the same affiliation) to be placed close to each other



Interaction idiom

Interaction operations:

- Exploration (panning)
- Zooming
- Reconfiguring (move nodes on the view)
- Details-on-demand (tooltips: name, affiliation, papers titles; hide/show labels)

Interaction paradigm:

- Full view (suitable to the network size)
 - “Overview first, zoom and filter, then details-on-demand”

Design phases

1. Identify the target users and their needs
2. Translate from domain-specific language to visualization language
 - A. Data modeling
 - B. Task modeling
3. Design visual encoding idioms and interaction idioms
- 4. Device efficient algorithms**
5. Make architectural and technological choices

4. Device efficient algorithms

Force-directed algorithms

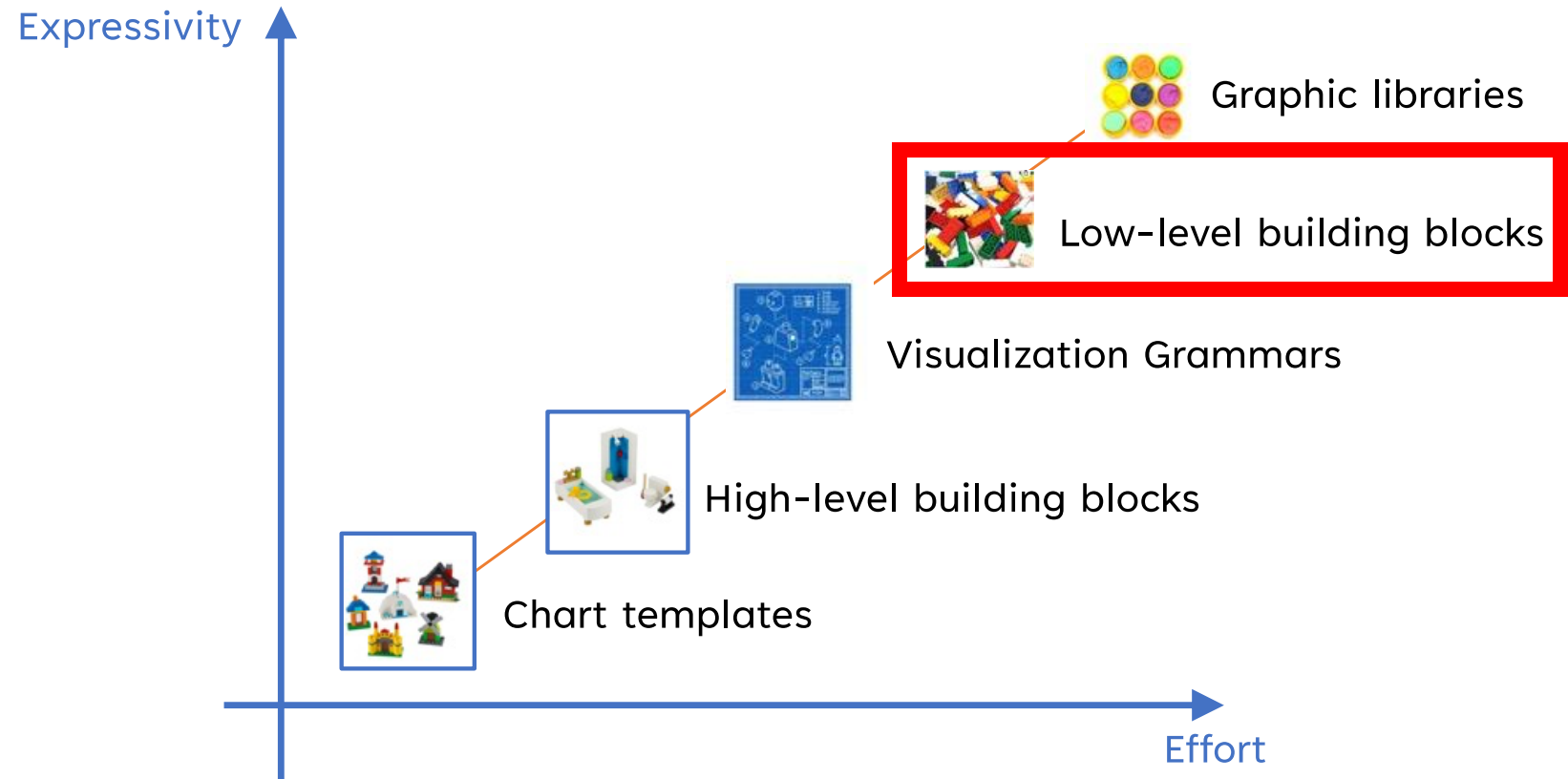
- Produce unconstrained straight-line drawings
- Suitable for small/medium networks (up to 100 nodes)
- Usually place groups of strongly connected nodes close to each other
 - Typically, there is a strong collaboration among authors having the same affiliation

Design phases

1. Identify the target users and their needs
2. Translate from domain-specific language to visualization language
 - A. Data modeling
 - B. Task modeling
3. Design visual encoding idioms and interaction idioms
4. Device efficient algorithms
- 5. Make architectural and technological choices**

5. Make architectural and technological choices

- Web application (JavaScript, html, css)
- D3.js library



(D3.js provides an implementation of a force-directed algorithm)

Interaction techniques and tools

Interaction operations and techniques:

- Exploration (panning) → grab the scene and move it with the mouse
- Zooming → +/- buttons
- Reconfiguring → grab and drag nodes with the mouse
- Details-on-demand → hover on nodes/links to see tooltips, dropdown for node labels

Interaction tools:

- Mouse

Datasets

Some useful links for open datasets.

Mostly non-relational:

- <https://www.cs.ubc.ca/group/infovis/resources.shtml#data-repos>
- https://data.world/search?scope=_all
- <https://www.kaggle.com/datasets>
- <https://www.tableau.com/learn/articles/free-public-data-sets>

Relational:

- https://visdunneright.github.io/gd_benchmark_sets/
- Graph Drawing contest (*creative topic* of the various editions):
 - <https://mozart.diei.unipg.it/gdcontest/history/>