The Evolution of Knowledge in the Refactoring Research Field
Matteo Orrù, Simone Porru, Michele Marchesi and Roberto Tonelli

International Workshop on Refactoring and Testing (REFTEST)

Department of Electrical and Electronic Engineering (DIEE) University of Cagliari
Contents

• Science Mapping
• Conceptual Structure and Evolution
• Science Mapping workflow
• Results and discussion
• Conclusions and Future works
Science Mapping

• Provides a spatial representation of a discipline, highlighting the reciprocal relationships among its elements.

• **Intellectual, conceptual and social structure** and **evolution** of a discipline.

• Network of bibliometric unit
  • Nodes ↔ Bibliometric unit.
    – E.g. paper, author, affiliation, etc
  • Edges: relationships among bibliometric units.
    – E.g. co-authorship
Conceptual Structure

- Network of reciprocal relationships between topics and sub-topics in a research field

- **Keywords Co-Occurrence Network**
  - **Nodes**: are the keywords found in publications corpus.
  - **Edges**: co-occurrence of two keywords within the same publication.
Science Mapping Workflow

- **Data Retrieval** → Query bibliographic databases
- **Preprocessing** → Remove errors, duplicated items, etc.
- **Network Extraction** → Create a network with the selected bibliographic units and relations
- **Normalization** → Apply a measure of similarity
- **Clustering** → compute the way the network is split in subnetworks
- **Mapping** → Map the document in the corresponding cluster

And finally ...

- **Analysis, Visualization and Interpretation**
Strategic Diagram (*)

- **Density (internal cohesion)**
  - How strong are the links among keywords inside the same theme

- **Centrality (external cohesion)**
  - How strong are the ties among theme

- **Node Size**
  - Depends on the performance measure (e.g. n. of docs).

(*) Callon, Courtial and Laville, Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer Chemistry, Scientometrics September 1991, Volume 22, Issue 1, pp 155-205
Experimental Settings 1/2

- **Data Source:** ISI Web of Science (ISIWoS)
  - 432 (we discarded 16 of them).

- **Periods:**
  - 3 periods, each of them of around 5 years, starting from 2001 to 2014.

<table>
<thead>
<tr>
<th>Years</th>
<th>N. Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: from 2001 to 2004</td>
<td>66</td>
</tr>
<tr>
<td>P2: from 2005 to 2009</td>
<td>151</td>
</tr>
<tr>
<td>P3: from 2010 to 2014</td>
<td>225</td>
</tr>
</tbody>
</table>
• **Network:** co-occurrence network, focusing on author’s keywords.

• **Clustering Algorithm:** Simple Centre Algorithm.

• **Mapper:** Mapping to nodes for documents that are present in at least k nodes (k=2, Core Mapper).
Strategic Diagram - 2001-2004
2001-2004 – OO Design
2005-2009 – Languages
2005-2009 – Agile Methodologies
2005-2009 – OO Systems
Strategic Diagram – 2010-2014
2010-2014 - Experimentation
Conclusions

- Insights on the structure and dynamics of the research on Refactoring
- Threats to validity
  - Co-word analysis limitations (internal)
  - Use of Scimat (external)
  - Experimental setting, etc (construction)
- Future work
  - Extend the base of data
  - Extend the analysis to other aspects (social network analysis, intellectual base analysis)
Thank You for Paying Attention
Any Questions?