Reproducibility & the Performativity of Methods

Theresa Velden

Workshop ‚Reproducible Scientometrics Research‘
October 17, 2017
ISSI 2017, Wuhan, China

Introduction

- Concern: Reliability of computational methods to map scientific fields

- Example: **Topic extraction**, foundational for e.g. specialty studies, field delineation, measurements of interdisciplinary

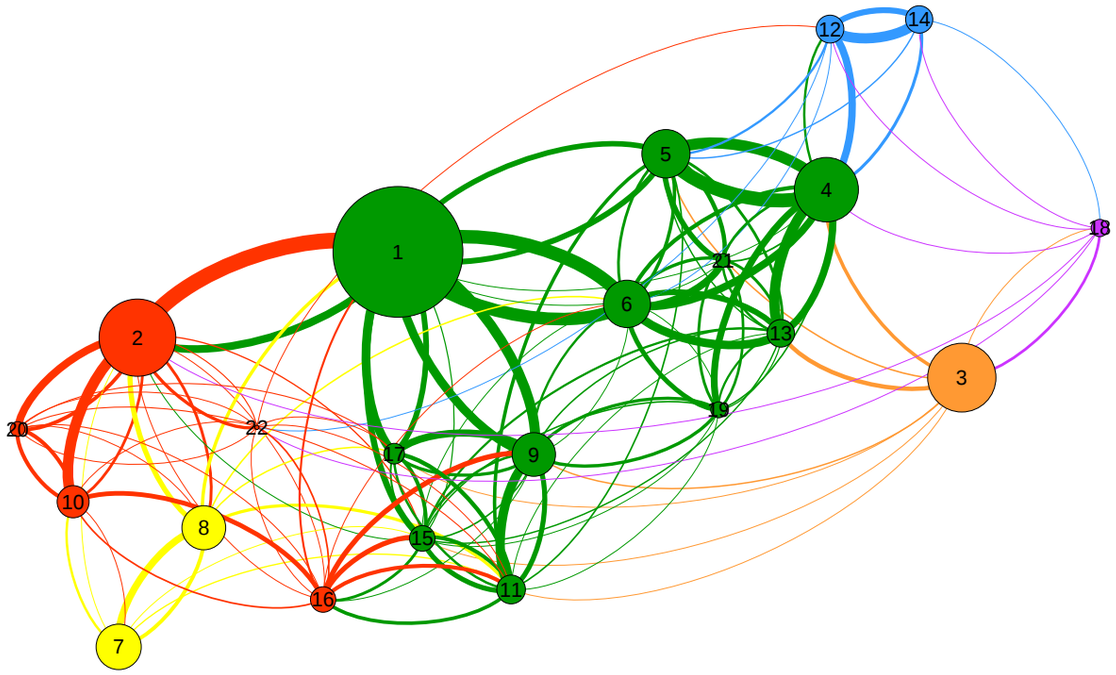
- Numerous approaches to topic extraction (data models, algorithms)

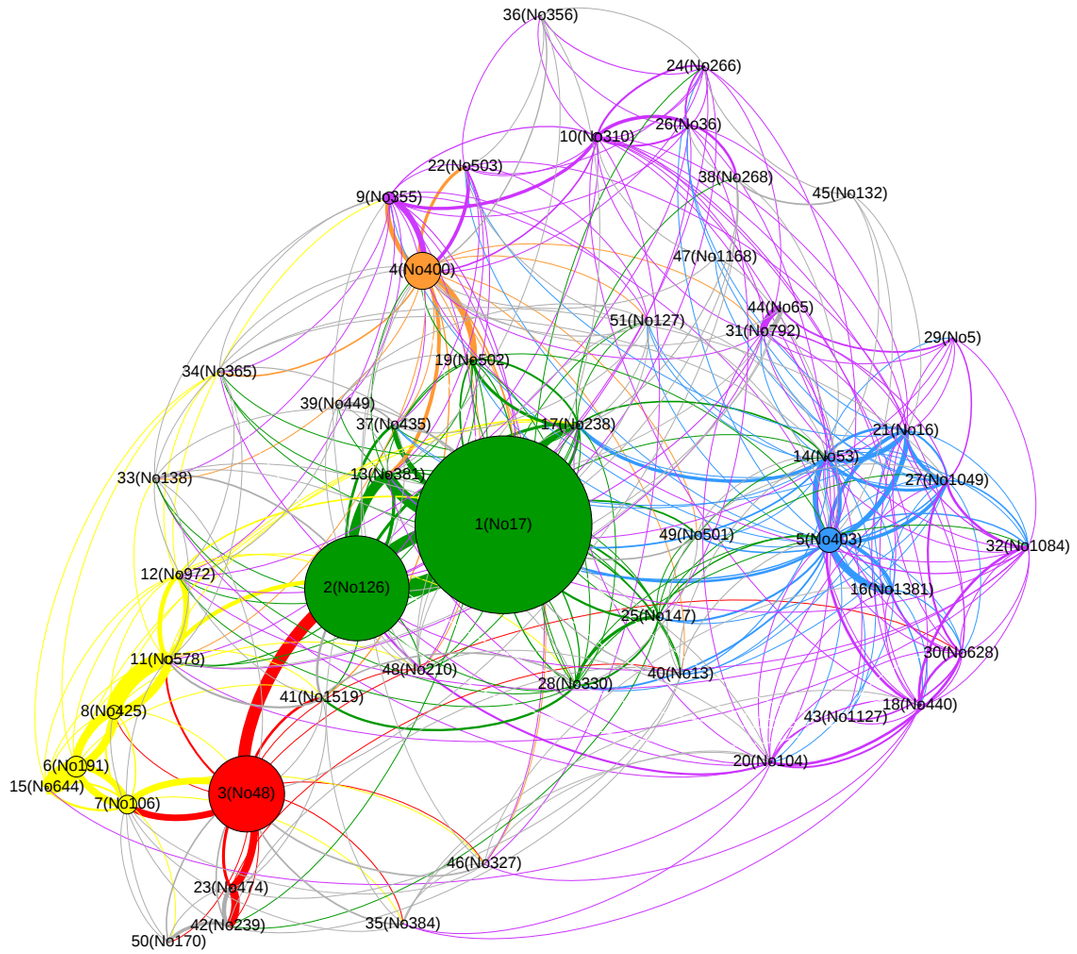
- The topical structures we ‚detect' depend on the methods we use (***performativity of methods***).

Mapping Astrophysics

Comparison of Topic Extraction Approaches and Results

Velden et al., *Scientometrics*, 2017



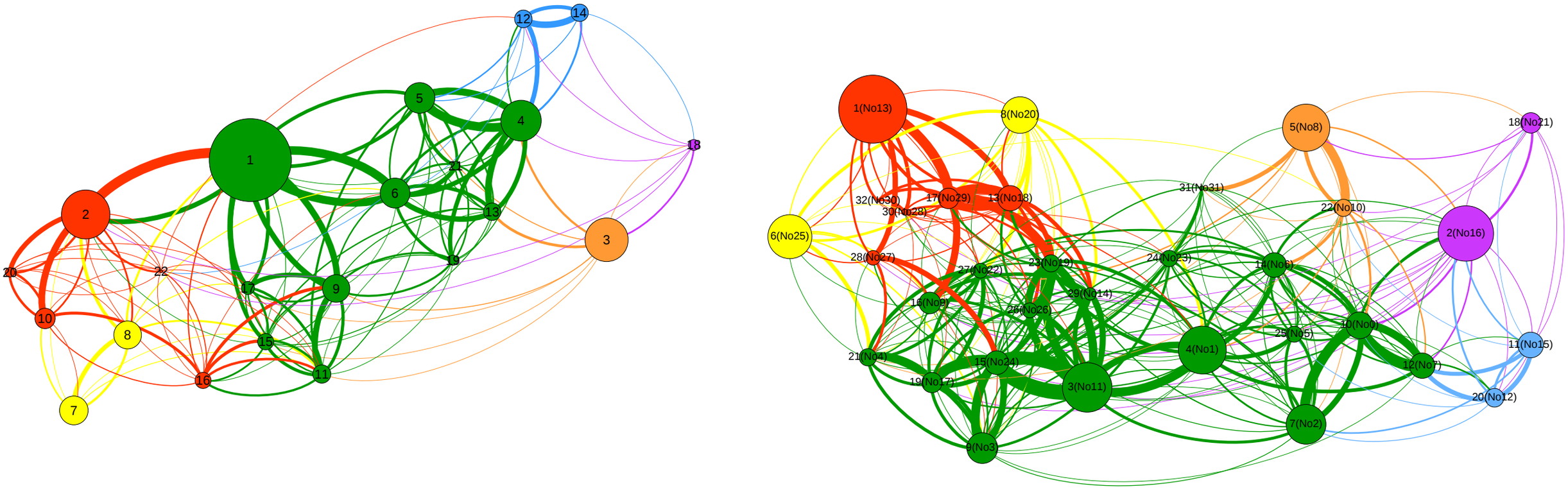


Astro Data: 111,616 bibliographic records from WoS (59 journals in Astronomy & Astrophysics, 2003-2010)

Mapping Astrophysics

Comparison of Topic Extraction Approaches and Results

Velden et al., *Scientometrics*, 2017



Astro Data: 111,616 bibliographic records from WoS (59 journals in Astronomy & Astrophysics, 2003-2010)

The Problem

- Lack of methodology for topic extraction in scientometrics
    - Theoretical properties of topics
    - Validity of operationalizations (for given purposes)
    - Influence of choices on results (data model, algorithm, thresholds and parameters)

- Conceptual reproducibility elusive as we cannot compare across studies

- To establish a methodology of topic extraction
    - Improve theoretical foundation
    - Conduct systematic comparisons of approaches and results

Answers to workshop questions

1a) Threats to reproducibility?

- Lack of a methodology of topic extraction
- Pre-dominance of case studies, applied work

1b) Why bother?

- Topic delineation fundamental to many types of studies

2) Concern about exact or conceptual reproducibility?

- Conceptual (& exact as a means)

3) Through what measures can these threats be addressed?

- Fundamental research (theory, systematic comparisons)
- Shared benchmark data sets
- Improved documentation (incl. open code)

www.topic-challenge.info

—> **How to access the data** (provided by Clarivate Analytics) **& participate**

Related Papers

- A principled methodology for comparing relatedness measures for clustering publications
  Ludo Waltman, Kevin W. Boyack, Giovanni Colavizza, Nees Jan van Eck **(Wednesday 8:30 AM)**
- Overlapping Thematic Structures Extraction with Mixed-Membership Stochastic Blockmodel
  Shuo Xu, Junwan Liu, Zheng Wang **(Thursday 8:30 AM)** - *- uses Astro Data set -*
- Methodological Challenges for the Comparison of Results of Topic Extraction from Scientific Literature
  Theresa Velden **(Thursday 4:30 PM)**