TALLINN UNIVERSITY

UNIVERSITY OF TARTU
1632

Digital humanities meets film history: Is there progress in films?

PEETER TINITS AND OLEG SOBCHUK

Film studies and digital humanities

Lack of film studies that:

- take a large-scale perspective;
- look for general principles in film history;
- use empirical/experimental methods (not only “interpretation”).

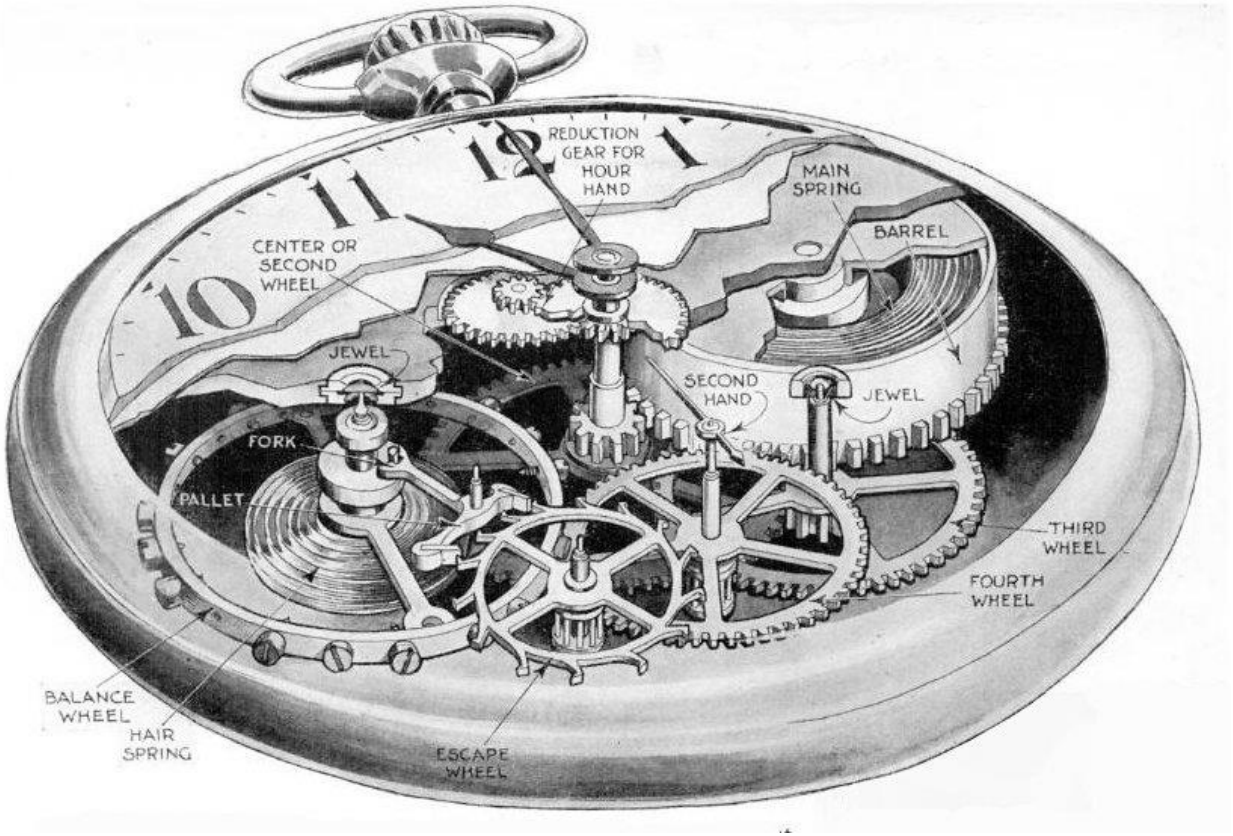
Film studies and digital humanities

Lack of film studies that:

- take a large-scale perspective;
- look for general principles in film history;
- use empirical/experimental methods (not only "interpretation").

So: Film studies badly need digital humanities.

Function vs. hidden mechanism



Function vs. hidden mechanism *in arts*

Function:

- attention control (Hasson et al. 2008)
- manipulating emotions (Huron 2006)
- provide information (Bordwell 2008)

Parts, for example:

- shots
- story components
- people (involved in film production)

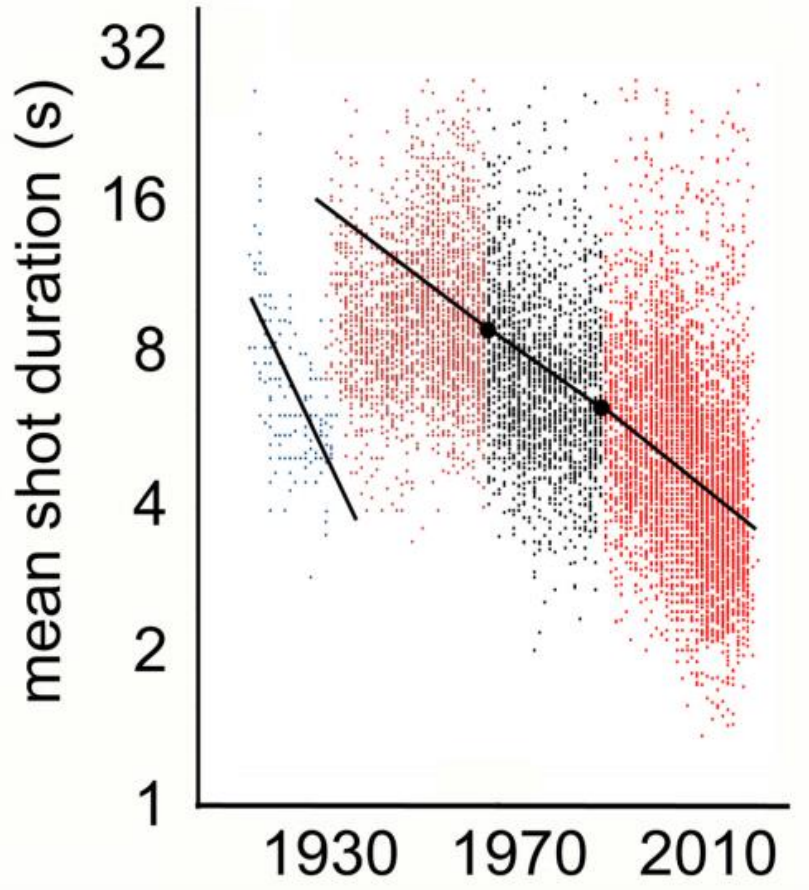
Progress in films?

Cumulative cultural evolution – is a process by which a series of social transmission events results in successive improvements in performance, arising due to an accumulation of modifications to the transmitted behaviours.

(Caldwell et al. 2016)

Evolution of film (1): **shots**

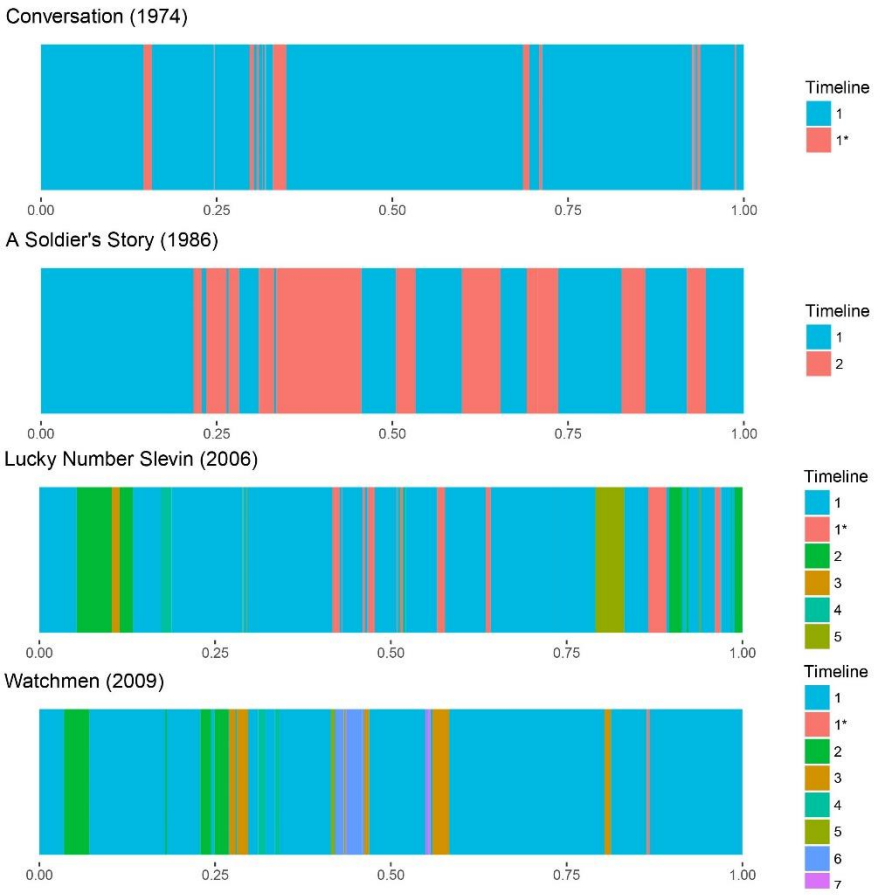
Decrease of mean shot duration



(Cutting & Candan 2015)

Evolution of film (2): **temporal structure**

Increased complexity of temporal structure



The study

COMPLEXITY OF FILM PRODUCTION CREWS

(1910-2010)

What's a film crew

A **film crew** - a group of people responsible for producing a film. (Not actors!)

Making the test

Purposive sample

Most popular films

- Good for studying trends in culture

Popularity

- is a goal for production
- is based on the product
- involves choices of many people

Measures

Complexity of film crews

- Population size
- Role distinctness and complexity
- Standardization
- Hierarchicality

Data

Collecting the data

Internet Movie Database (IMDb)

- 20+ years data collection, currently ~460k movies listed
- Data quality varies, but popular should be best

Popular films

- Just US and English
- User rankings & votes => 100 per decade (1910-2010)
- Data about film crews
- Total 1000 films, 144,756 job entries

Data

title	year	job	name
Die Hard	1988	chief engineer: [redacted]	Gene Whiteman
Die Hard	1988	special effects assistant	
Die Hard	1988	model effects: [redacted]	Tad Krzanowski
Die Hard	1988	effects technician: [redacted]	

Boss, ILM – company names

(uncredited) – not job title

Dataset

- Films, job titles, names
- Simplifying assumption: job titles ~ jobs

Pre-processing

- We removed non-essential information (e.g. model effects: ILM (uncredited))
- We removed jobs from later releases (e.g. music (1970 re-release)

Results

Population size

Job complexity

Standardization

Hierarchicality

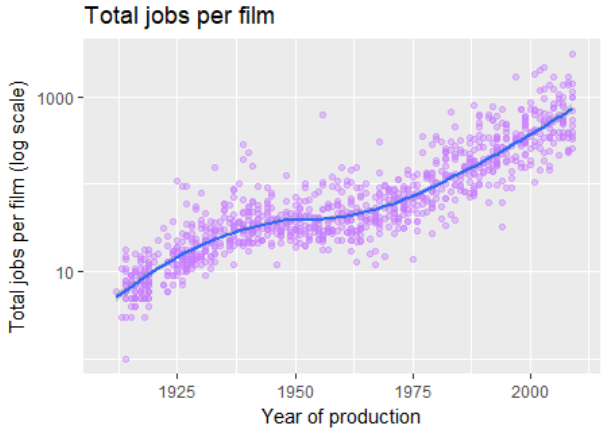
Population size

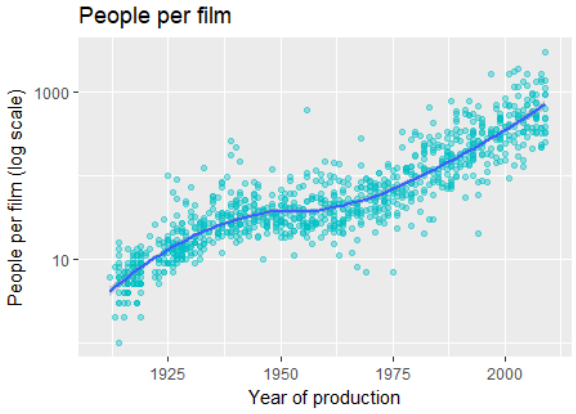
Log scale plots

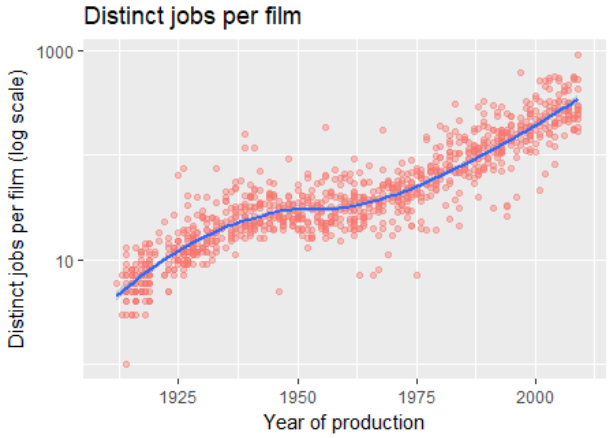
- Total jobs
- People
- Distinct jobs

Exponential growth

- Except for ~1940-1970







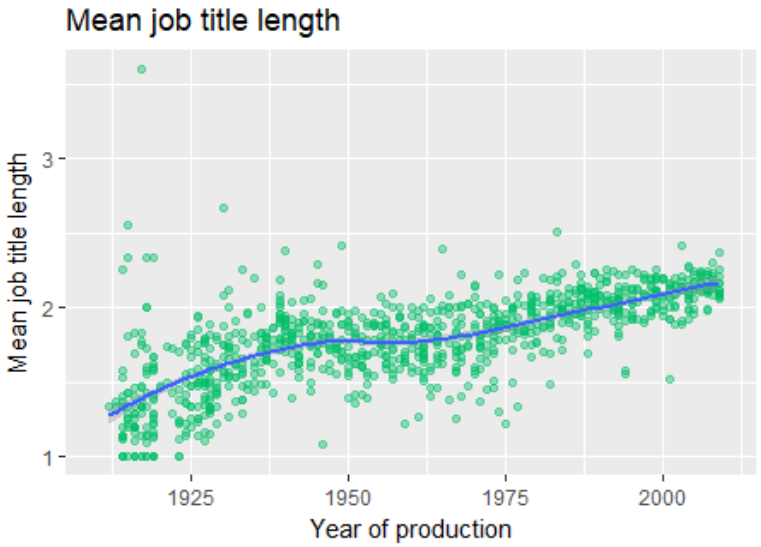
Job title complexity

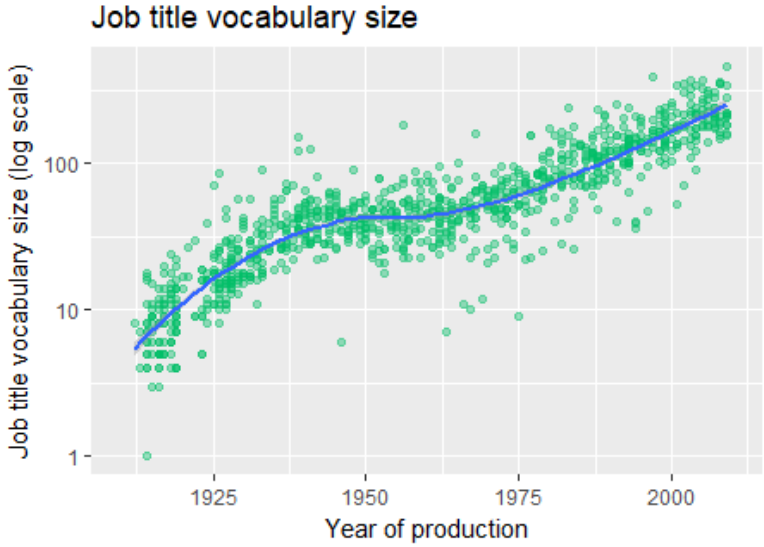
Linear
- Mean job title length

Log scale
- Vocabulary size

special effects assistant	3 words
model effects: ILM (uncredited)	2 words

Vocabulary: special + effects + assistant + model (4 words)





Standardization of job titles

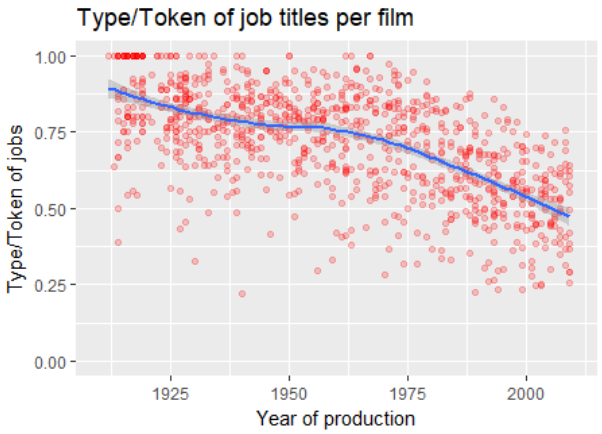
Within films

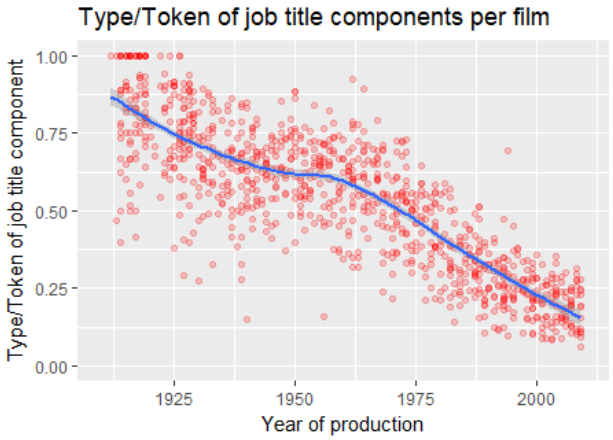
- Type/Token ratio
- 1 - proportion of repetitions

Nr of distinct words (types)

/

Nr of total words (tokens)

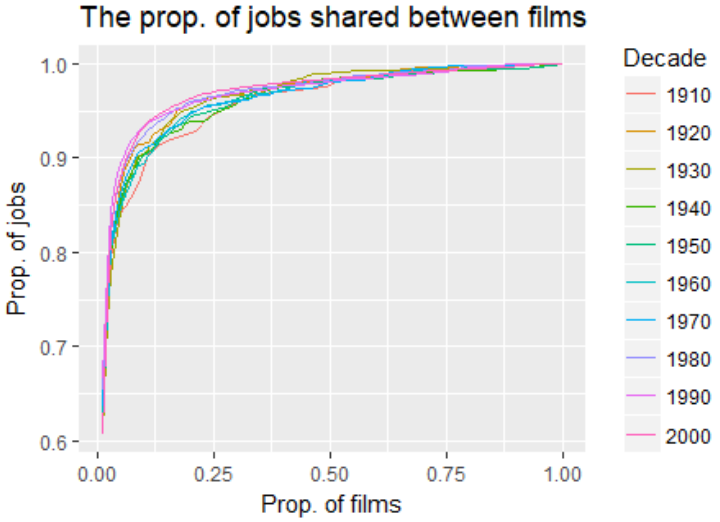


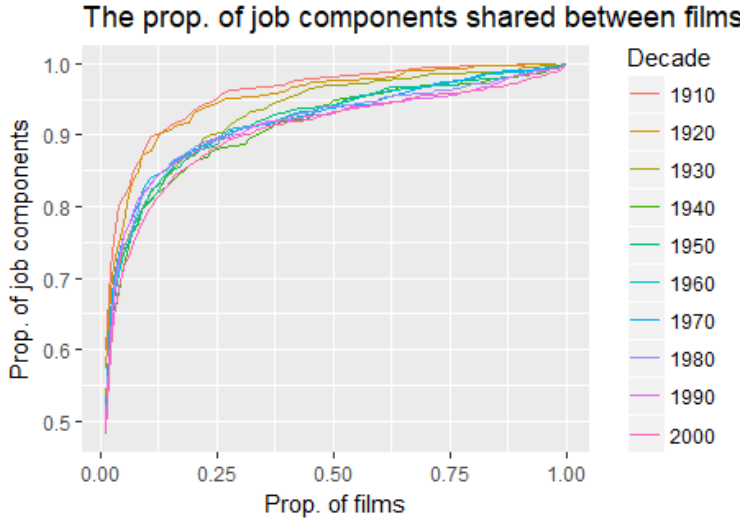


Standardization of jobs

Between films

- How many jobs were commonly shared between films
- In proportions, it is mostly a Zipf-like distribution
  - 1910-1940 had less shared components

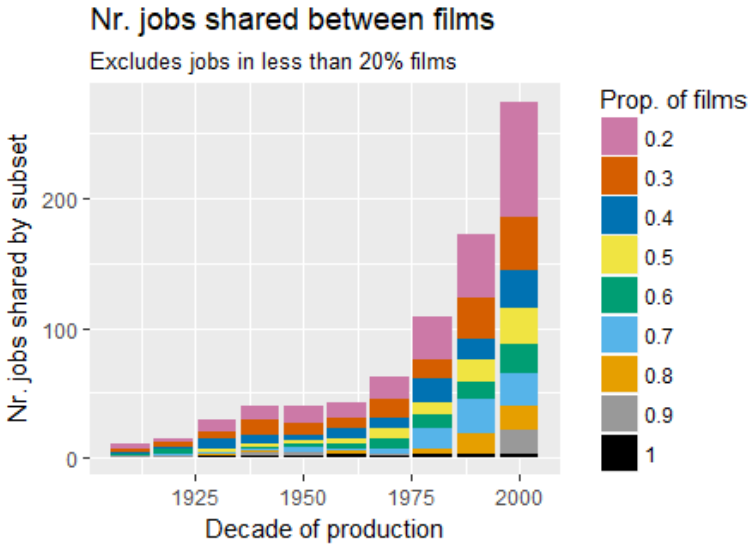


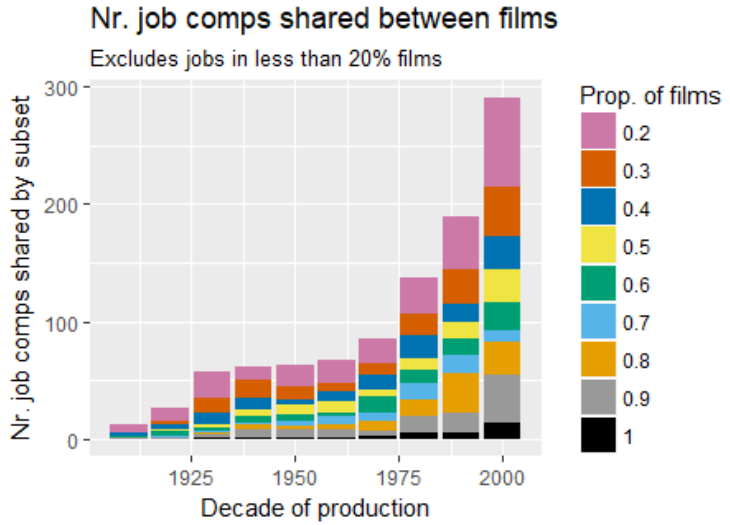


Standardization of jobs

Between films

- How many jobs were commonly shared between films
- In absolute number an increase in core jobs.
  - E.g. the jobs that were in at least 50% of films
    - Before 1930, less than 10, in 2000s, more than 100



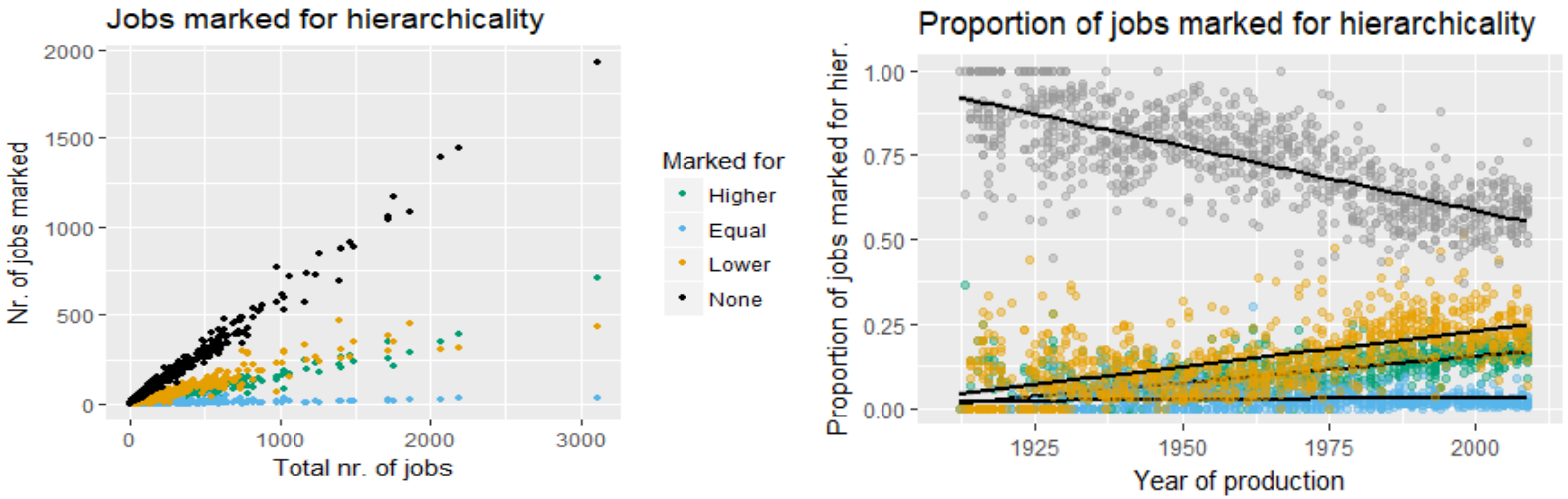


Hierarchicality

Often jobs carried explicit markers of hierarchicality

- superordinate: boss, chief, supervising ...
- subordinate: assistant, 2nd, junior ...

Checked the presence of any markers in jobs



To sum

Results

- Growth in the number of jobs, people and distinct roles
- Growth in the complexity of job titles, and the vocabulary used for it
- An increase in vocabulary standardization within films and between films
- An increase in hierarchicality markers across films

Discussion

1. The film crew “parts” are becoming more complex and more hierarchical

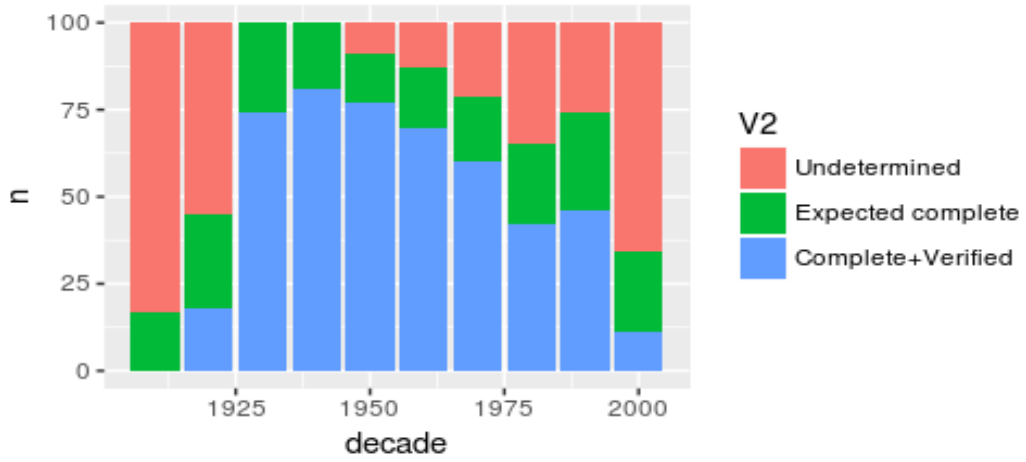
2. Digital humanities approach
   - Making generalizations
   - Novel theoretical interests

3. Complexity of art (films in particular)
   - Difficult to measure, but could prove interesting
   - More work towards collecting datasets?

Data quality

Films were marked for expected completeness of film crew data

- Confirmed
- Expected complete
- No marking



Some counterfactual history

What if there were no 1940-1970? ☺