

Fidelity at Scale

HOW TIME AFFECTS DATA

Scott B. Weingart
@scott_bot / scottbot.net
Indiana University

#HDHD

“We encourage the participants to focus on practical issues such as how to (and not to) approach historical change methodologically and what type of data to use (and not to use).”

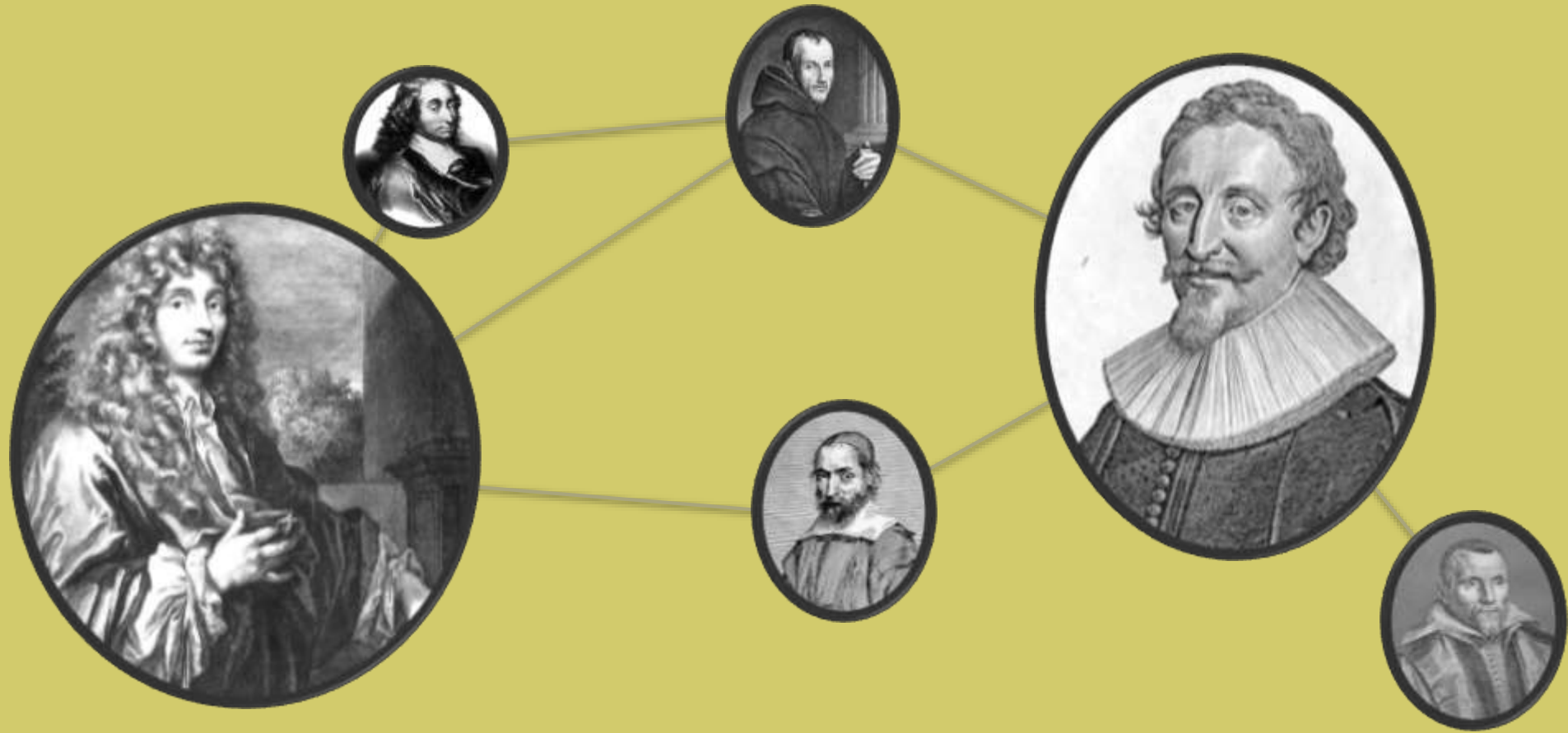
LACK OF EVIDENCE

“I am content to define history as the past events of which we have knowledge and refrain from worrying about those of which we have none - until, that is, some archaeologist digs them up.”

-Barbara W. Tuchman, *Practicing History* (1981)

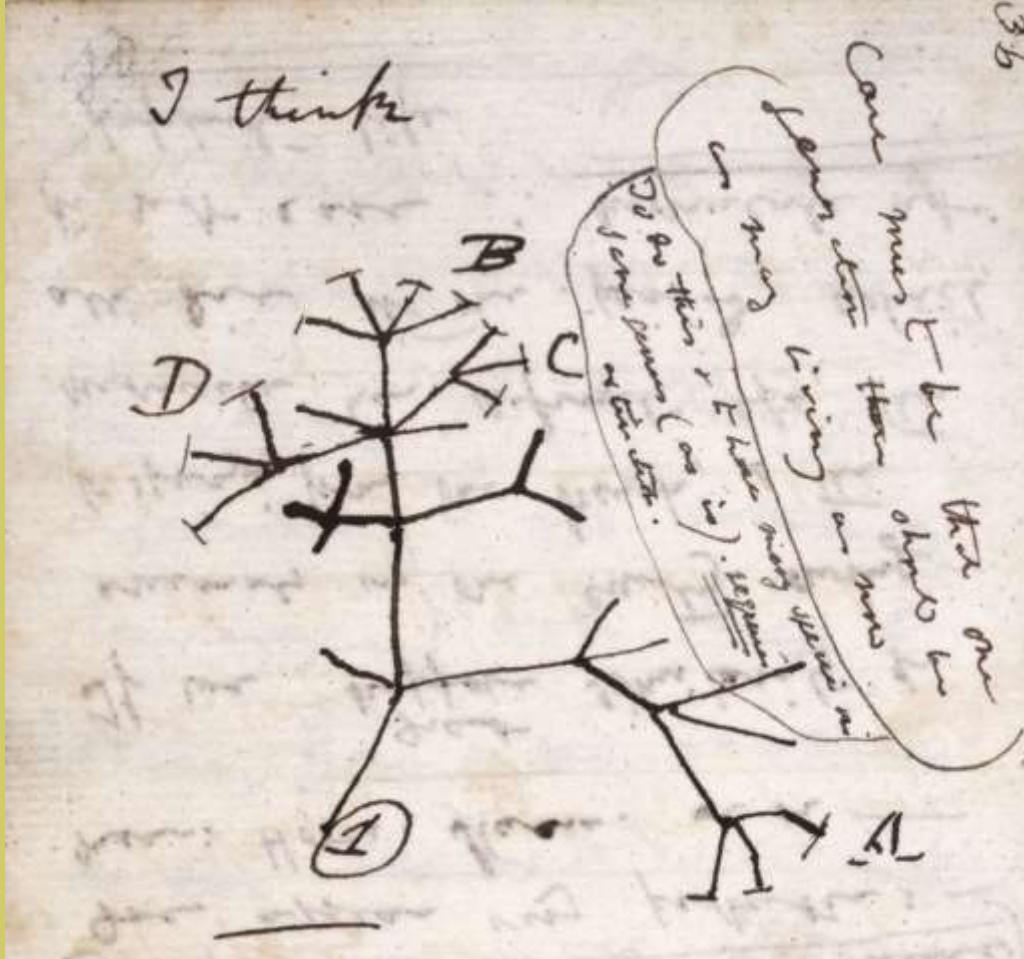
FIDELITY AT SCALE

- 13.30 Introduction
- 13.35 The Republic of Letters
- 13.40 Fidelity and Transmission
- 14.05 Accounting for Bias
- 14.10 Q&A



THE REPUBLIC OF LETTERS

A cautionary tale



FIDELITY AND TRANSMISSION

How time affects data

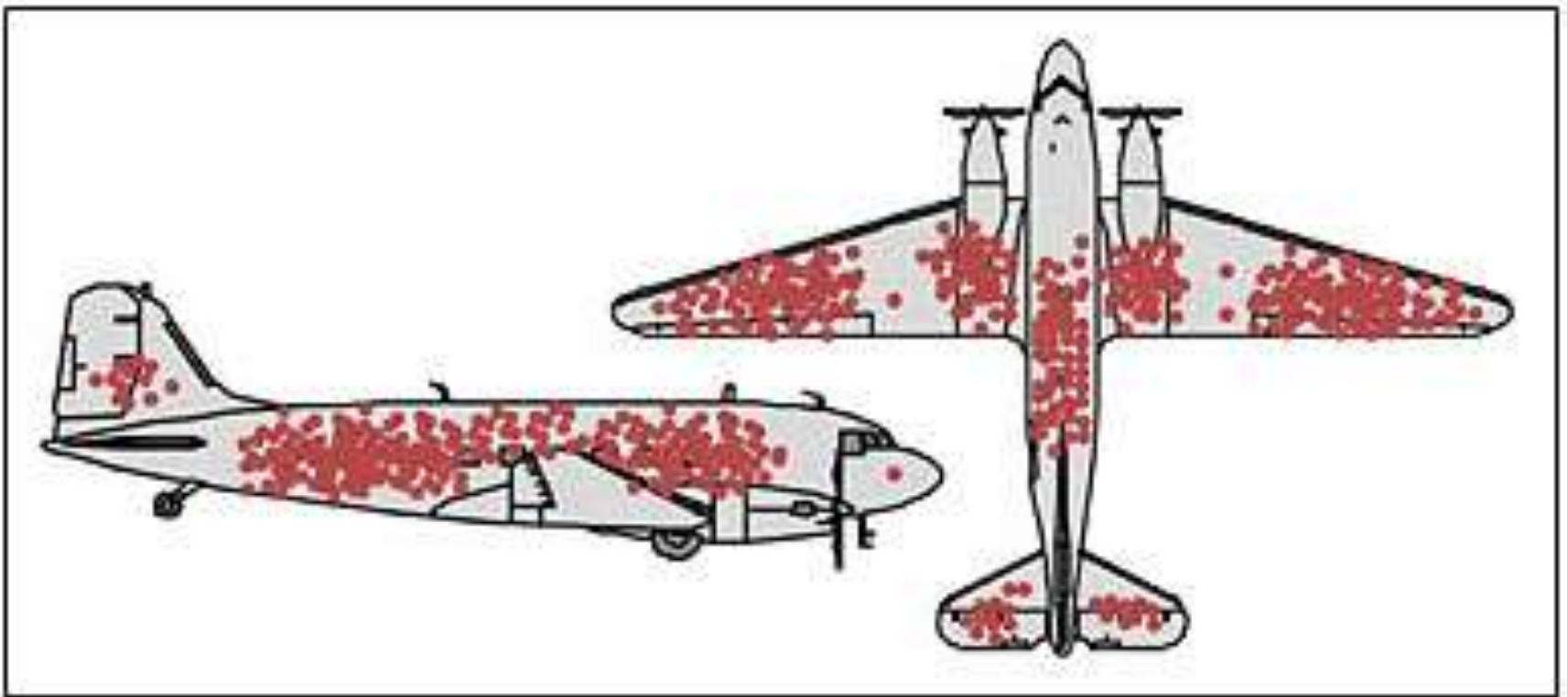
HISTORICAL FIDELITY

“Some processes tend to preserve in their end states information from their initial state more than others. **Fidelity** measures the degree to which a unit of evidence tends to preserve information about its given cause.”

[...]

“Historiography leaped forward following the recognition that oral transmission has a lower fidelity than written transmission and that memorized prose has a lower fidelity than memorized verse.”

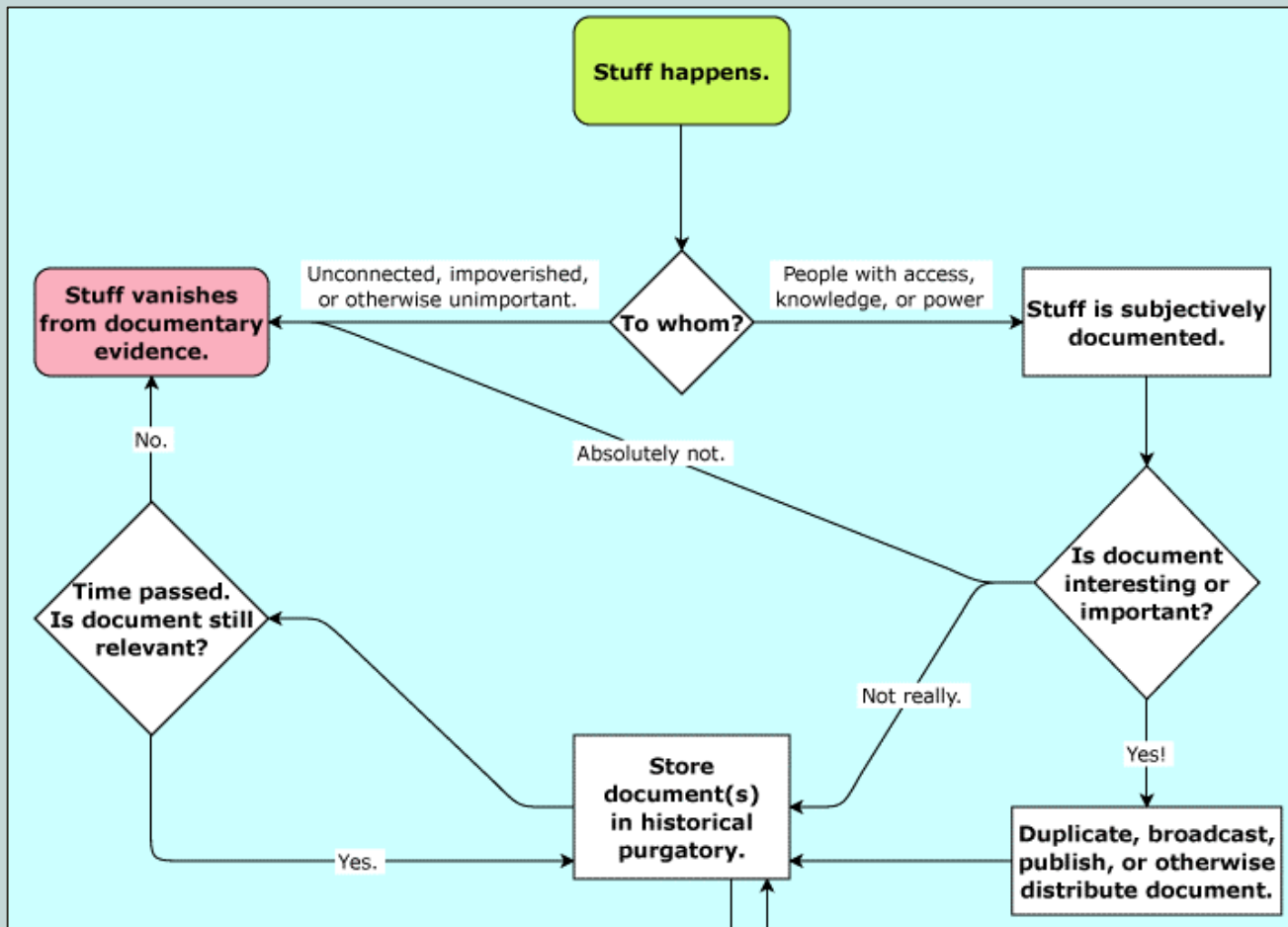
-Aviezer Tucker, “Our Knowledge of the Past”



Credit: Cameron Moll

SELECTION BIAS

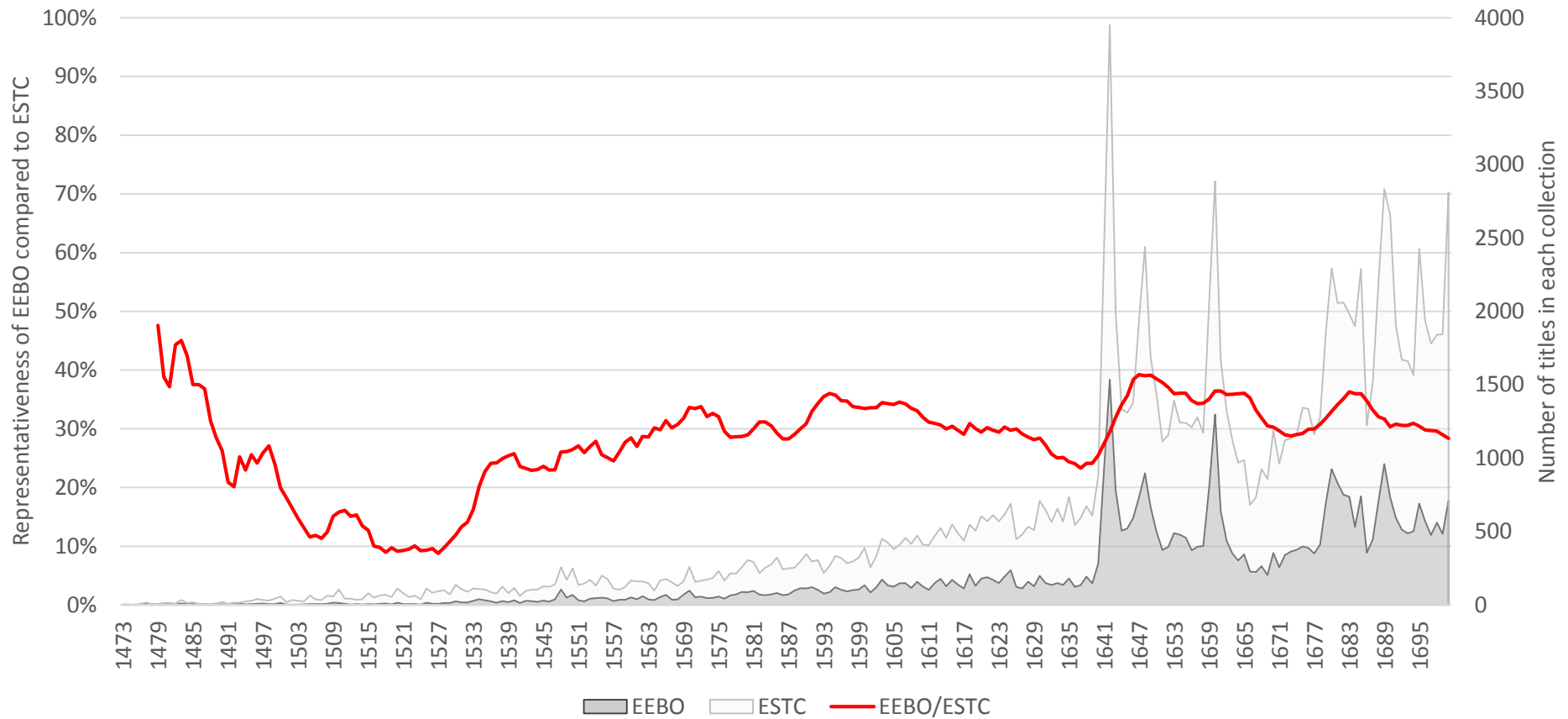
Bombers and earthquakes



EVIDENTIARY SURVIVAL

What lasts?

Representativeness of EEBO compared to the ESTC

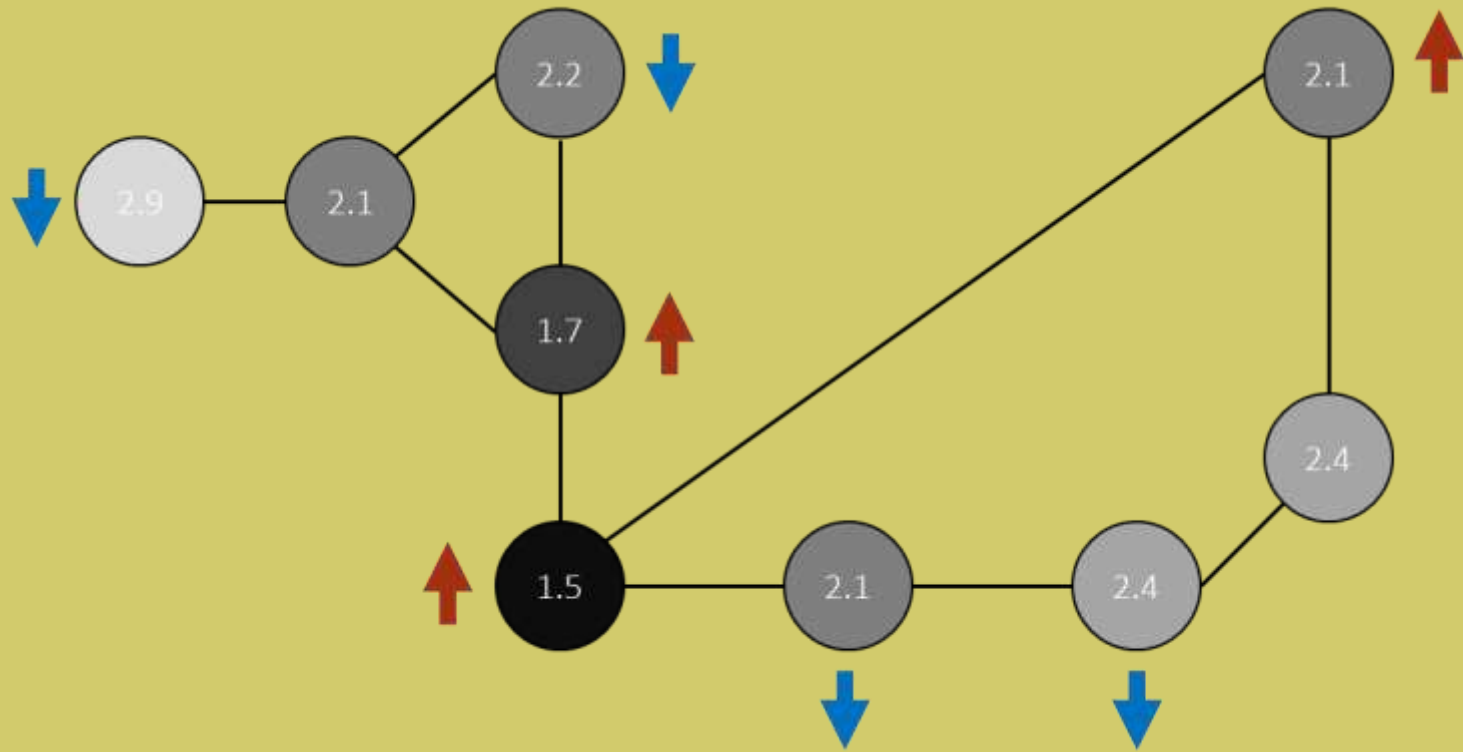


EEBO-ESTC

What's the difference?

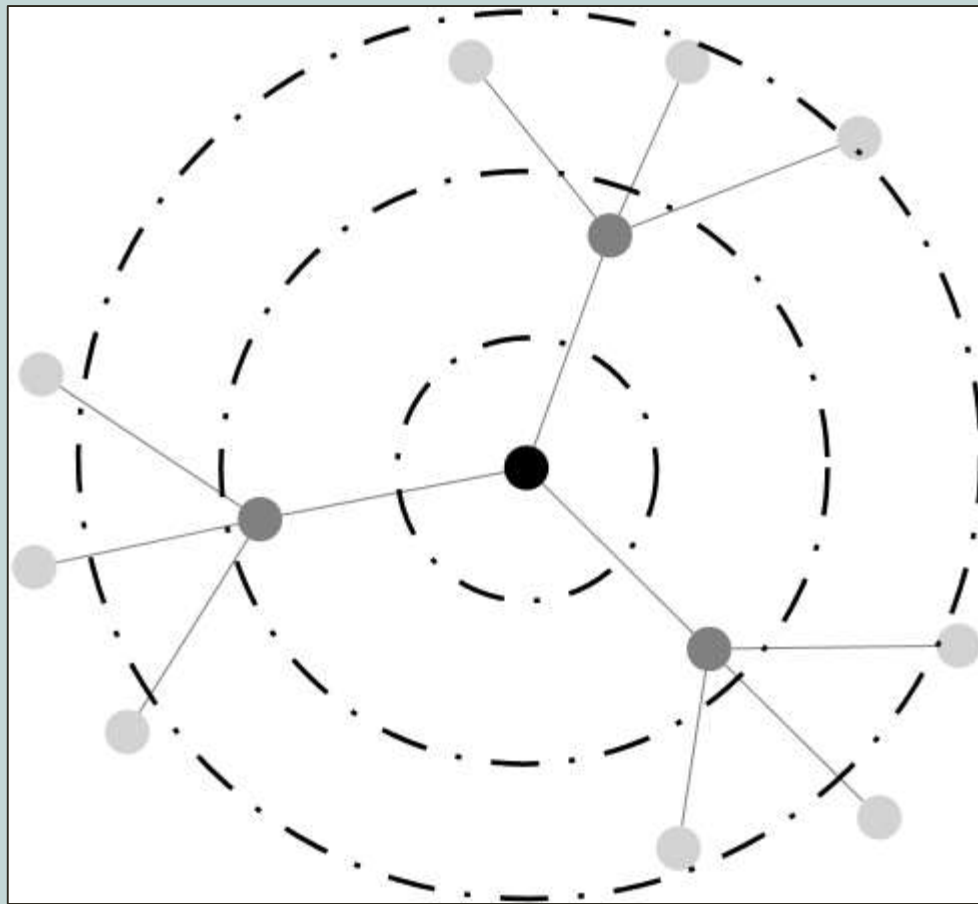
THE CASE OF GRIEVANCES

Given its historical conditions, what's the likelihood any document will survive n years?



ACCOUNTING FOR BIAS

Toward solutions



SNOWBALL SAMPLING

Everything looks the same.

Q&A

Thank you!