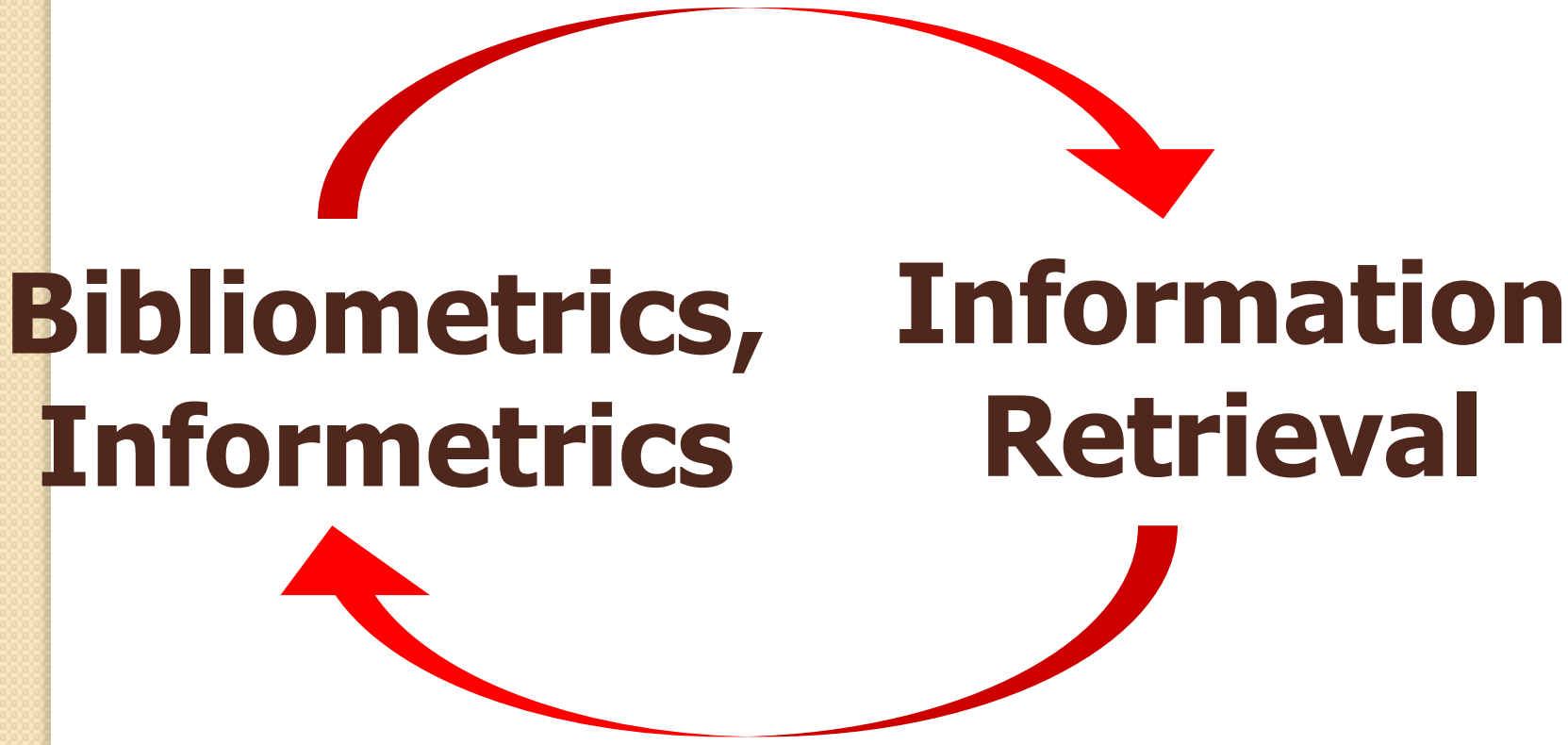
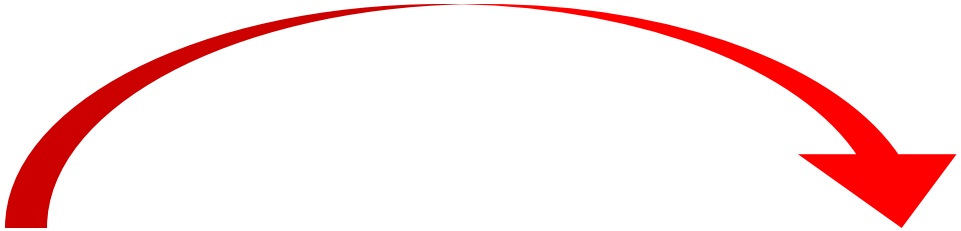
# The Symbiotic Relationship Between Bibliometrics and Information Retrieval

Dietmar Wolfram
University of Wisconsin-Milwaukee

# Introduction

- The intersection of two key areas of information science offers many areas of study
- A better understanding of each helps to inform and advance the other
- Historically, bibliometrics has helped to inform bibliographic IR, the reverse is now becoming more evident

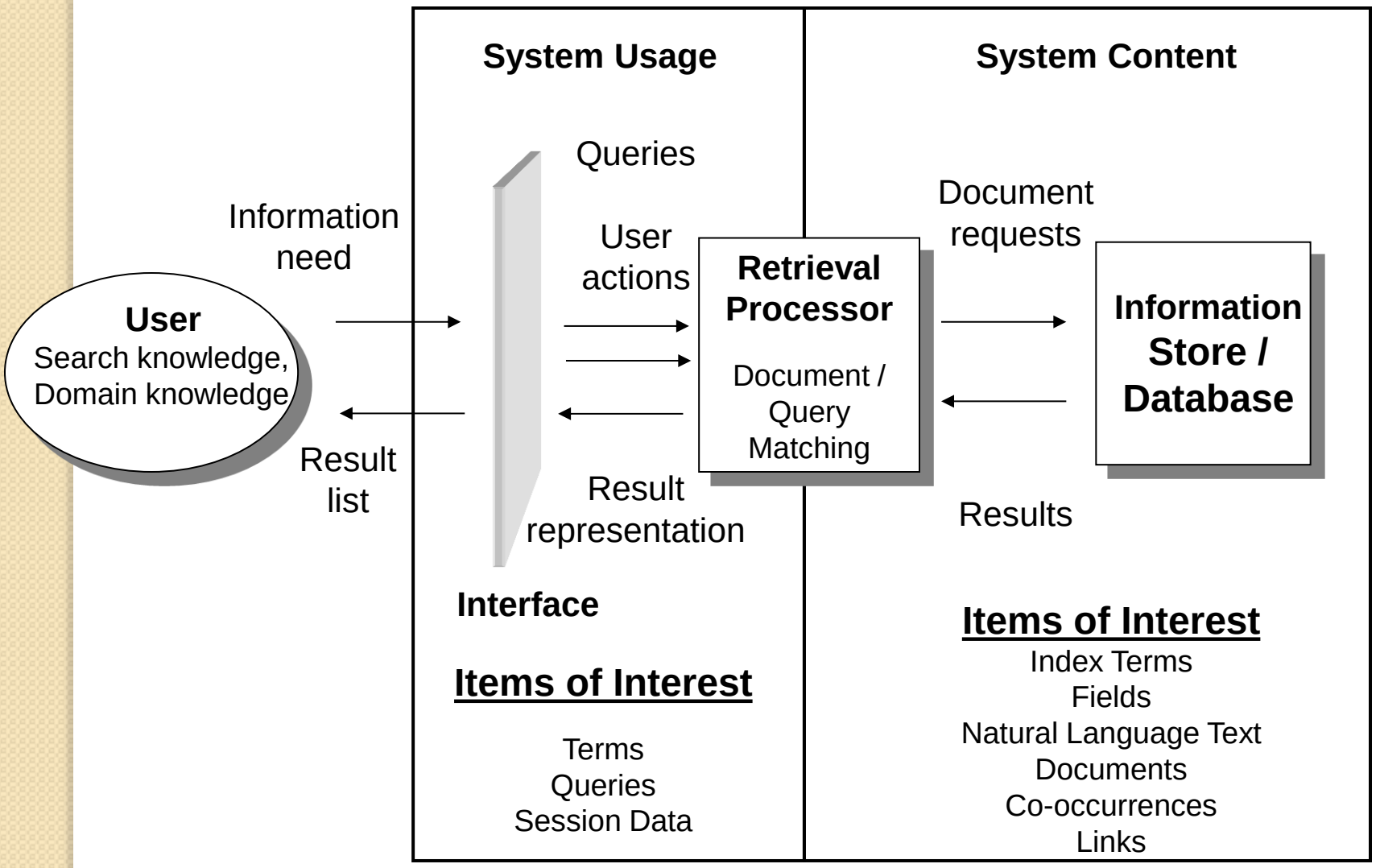

Bibliometrics,
Informetrics
Information
Retrieval



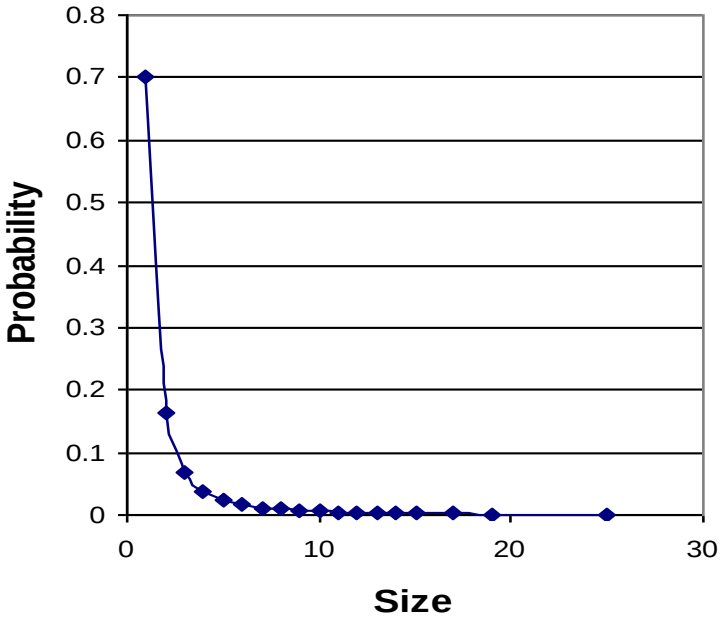
**Bibliometrics, Informetrics**

**Information Retrieval**

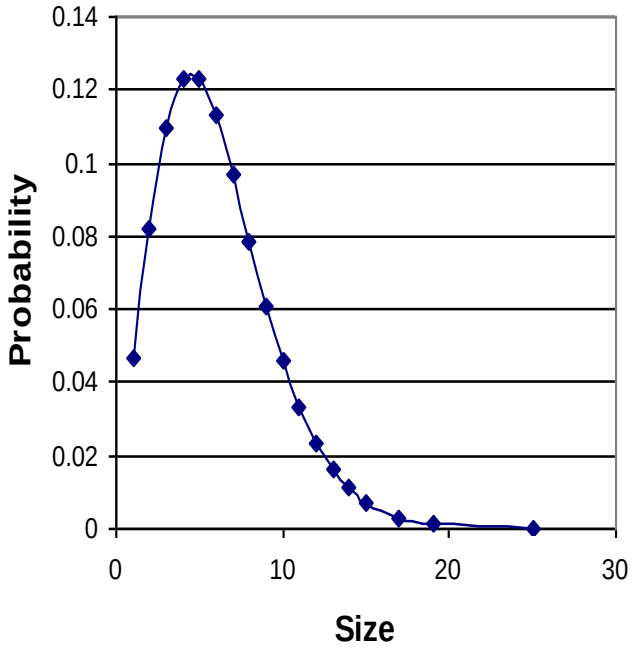
# IR Processes & Associated Data



# Observed Patterns in Content & Use Frequency

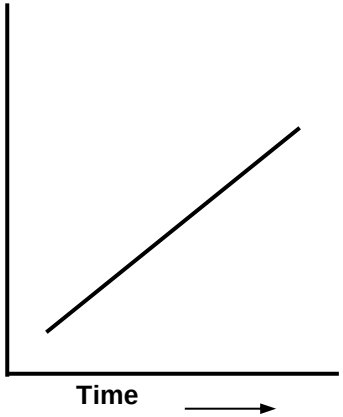



**"Zipfian" or "Lotkaian" (Power Law)**
**Mode = 1, sometimes 0**

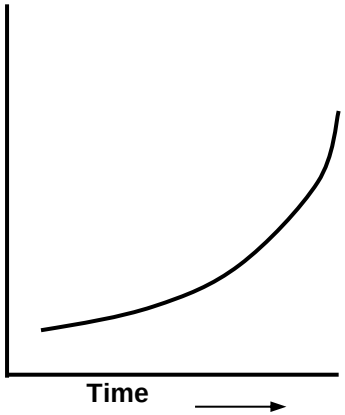



**"Unimodal" Mode > 1**

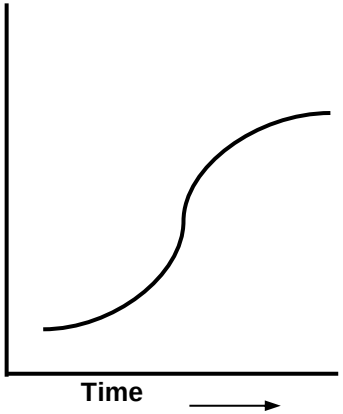
# Observed Patterns in Content Growth



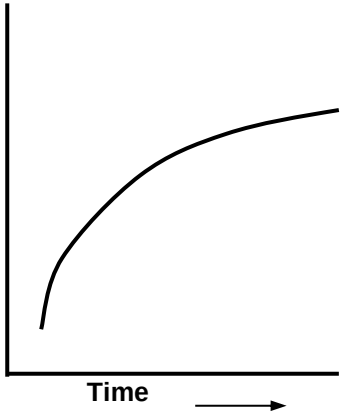

LINEAR




"EXPONENTIAL"
(NONLINEAR)




LOGISTIC




LOGARITHMIC

# Applications for IR

- IR system simulation
- File design & space planning
- System indexing & retrieval - PageRank, SEO
- IR system design & evaluation – Link analysis
- Targeting of IR services

**Bibliometrics, Informetrics**

**Information Retrieval**

# Reciprocal Contributions

- With growing datasets, new ways to store, process and display data are needed
- IR frameworks provide tools & approaches for -metrics researchers
  - Database design for bibliographic datasets
    - Relational DBMSs, IR software
  - Application of vector space & probabilistic IR models to compare data

# Examples

- White (2007) – applied IR measures of term weighting (tf*idf) to bibliometric data
- Applications of Web link analysis
  - Research by Thelwall, Vaughan
  - Bollen et al. (2006) – use of weighted PageRank for journal ranking

# Examples

- Applications of language & topic modeling
  - Tang et al. (2008) – applied LDA to academic search
  - Ding (2010) – applied to co-authorship and citation networks
  - Lu & Wolfram (2012) – compared author similarity using text from author oeuvres

# Future Directions

- Complexities of bibliometric datasets lend themselves to IR techniques
  - Resulting "big data" require data and text processing/mining techniques to identify overt & hidden patterns
- Topic modeling show great promise in providing complementary approaches to co-citation & co-authorship data