



Pennant diagrams, what is it, what are the possibilities and are they useful?

Jesper W. Schneider, Birger Larsen & Peter Ingwersen,
Royal School of Library & Information Science, Denmark




Outline

- Introduction and motivation
 - Bibliometric distributions
 - Relevance theory
 - White's synthesis
- Pennant diagrams
 - Notion
 - Examples
- Summary points

Introduction
Pennant diagrams
Summary points

Nordic Bibliometric Workshop 2007, Copenhagen




Introduction

- Pennant diagrams is an idea of Howard White
 - A combination of bibliometrics, information retrieval and relevance theory
 - For a thorough theoretical and methodical introduction please consult
 - White, H. (2007a) Combining bibliometrics, information retrieval, and relevance theory, part 1: First examples of a synthesis. JASIST 58(4), p. 536-559
 - White, H. (2007a) Combining bibliometrics, information retrieval, and relevance theory, part 2: Some implications for information science. JASIST 58(4), p. 583-605

Nordic Bibliometric Workshop 2007, Copenhagen

Introduction
Pennant diagrams
Summary points

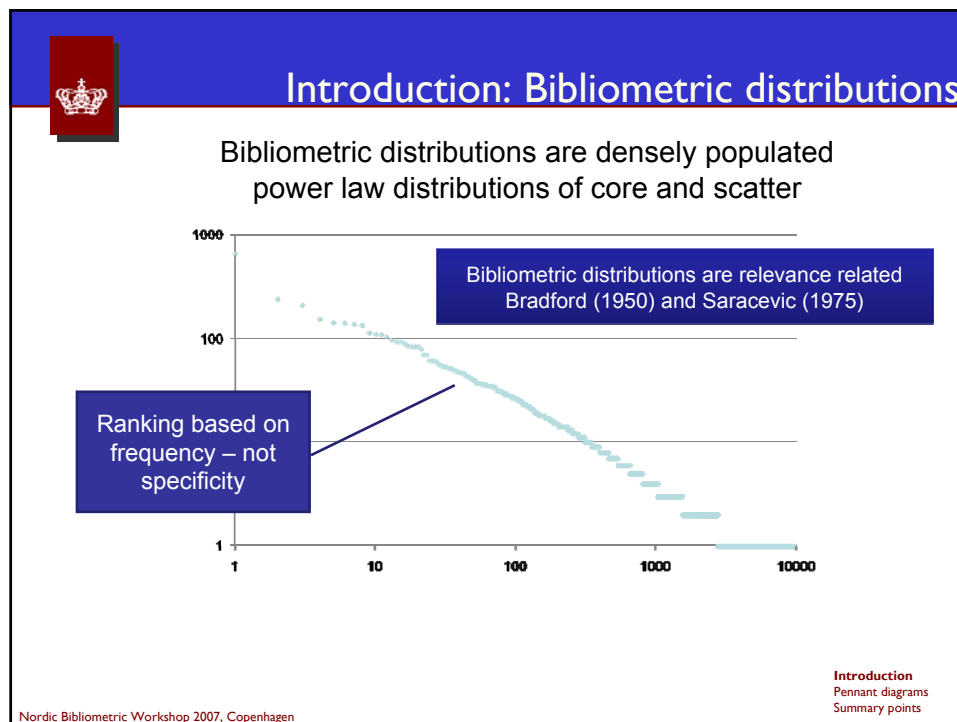


Introduction

- Our research aim is two-fold
 - Further development of pennant diagrams
 - Validation of the diagrams and their implicit variables
- Our research is at a preliminary stage where we are exploring a number of different diagrams in order to identify common traits and interesting variables
- This presentation outlines the main ideas and mechanisms of pennant diagrams and briefly introduce a couple of research questions

Nordic Bibliometric Workshop 2007, Copenhagen

Introduction
Pennant diagrams
Summary points




Introduction: Relevance theory

- Relevance theory (dialogues between persons)
 - ‘Assumptions relevant in context’ (Sperber & Wilson, 1995, p. 125)
 - An ‘assumption’ is relevant in a context to the extent that its cognitive effects in this context are large
 - An assumption is relevant in a context to the extent that the processing effort required to process it in this context is small
 - Cognitive effect = “is a worthwhile difference to the individual's representation of the world” (p. 251)
 - The ratio Cognitive Effects/Processing Effort defines the relevance of a communication

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition*. 2nd Ed. Oxford: Blackwell

Introduction
Pennant diagrams
Summary points

Nordic Bibliometric Workshop 2007, Copenhagen




Introduction: Relevance theory interpreted by White

- White's interpretation (a dialogue between system and user)
 - Assumptions \approx terms (noun phrases)
 - A measure of relevance based on term counts is a system measure, and the assumptions are the system's as instructed by its human designers
 - “Suitably marshalled the counts permit a responsive answer within a limited bibliographic domain – one that is qualitatively similar to what a well-informed person could supply” (p. 537)
 - ... all the user need to do is set a context with a seed term designating an interest

White, H.D. (2007) . Combining bibliometrics, information retrieval, and relevance theory, part 1: First examples of a synthesis. JASIST 58(4), p. 536-559.

Introduction
Pennant diagrams
Summary points

Nordic Bibliometric Workshop 2007, Copenhagen




Introduction: The new synthesis proposed by White (1/2)

- ‘Assumptions relevant in context’ then becomes
 - ... terms relevant to a seed term
- Notion
 - To use logarithmic versions of the *tf* and *idf* weighting scheme to rank bibliometric distributions by their degrees of relevance to a seed term used as a query
 - The purpose of *tf* \times *idf* is to break out comparatively small fractions of the densely populated bibliometric distributions
 - those most relevant in Sperber and Wilson's sense

Introduction
Pennant diagrams
Summary points

Nordic Bibliometric Workshop 2007, Copenhagen

Introduction: The new synthesis proposed by White (2/2)



- $\text{Log}(tf)$ of terms co-occurring with a seed term \approx
 - Measures the predicted cognitive effects within the context of that seed term (system side)
- $\text{Log}(idf)$ for the same distribution \approx
 - Measures the predicted processing effort of the terms co-occurring in that context (system side)


$$Tf \times idf \approx \text{Cognitive Effects/Processing Effort}$$

- When bibliometric distributions predicts degrees of cognitive effect and processing effort \approx
- User-oriented and instrumental

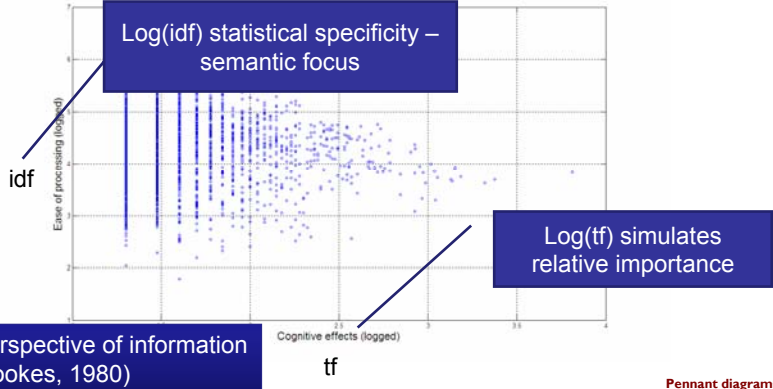
Introduction
Pennant diagrams
Summary points

Nordic Bibliometric Workshop 2007, Copenhagen

Pennant diagrams: Plotting Cognitive Effects and Ease of Processing



- Pennant diagrams use bibliometric data and information retrieval weighting schemes on the system side to mimic a relevance theoretic model of cognition on the user side



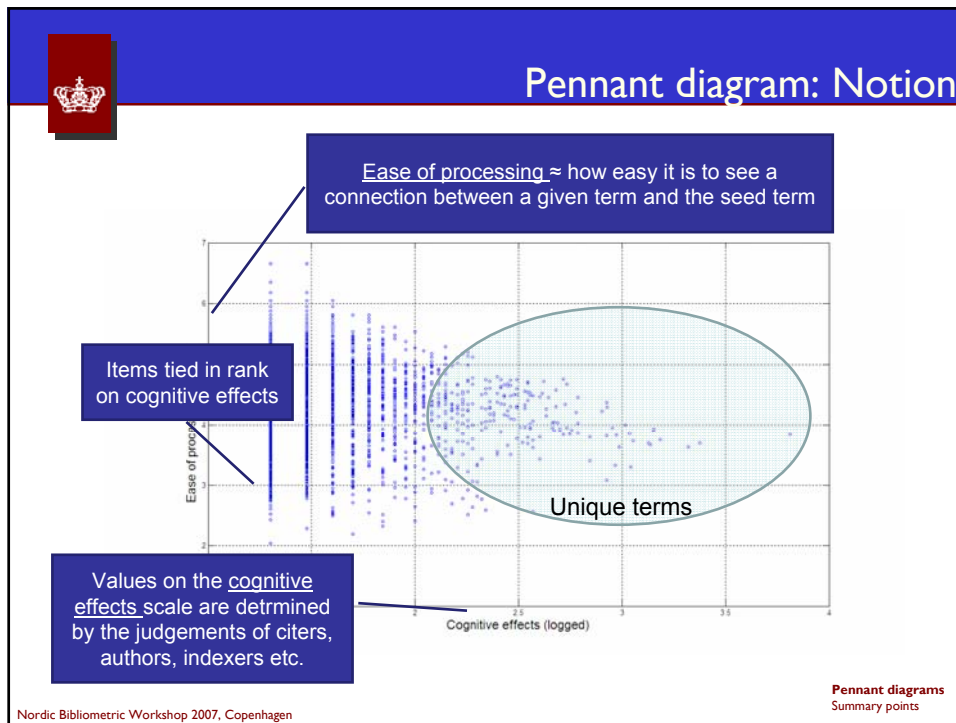
Log(idf) statistical specificity – semantic focus

Log(tf) simulates relative importance

Ordinal perspective of information space (Brookes, 1980)

Pennant diagrams
Summary points

Nordic Bibliometric Workshop 2007, Copenhagen



Pennant diagram: An example (1/2)

idf = items in file

tf = items ranked

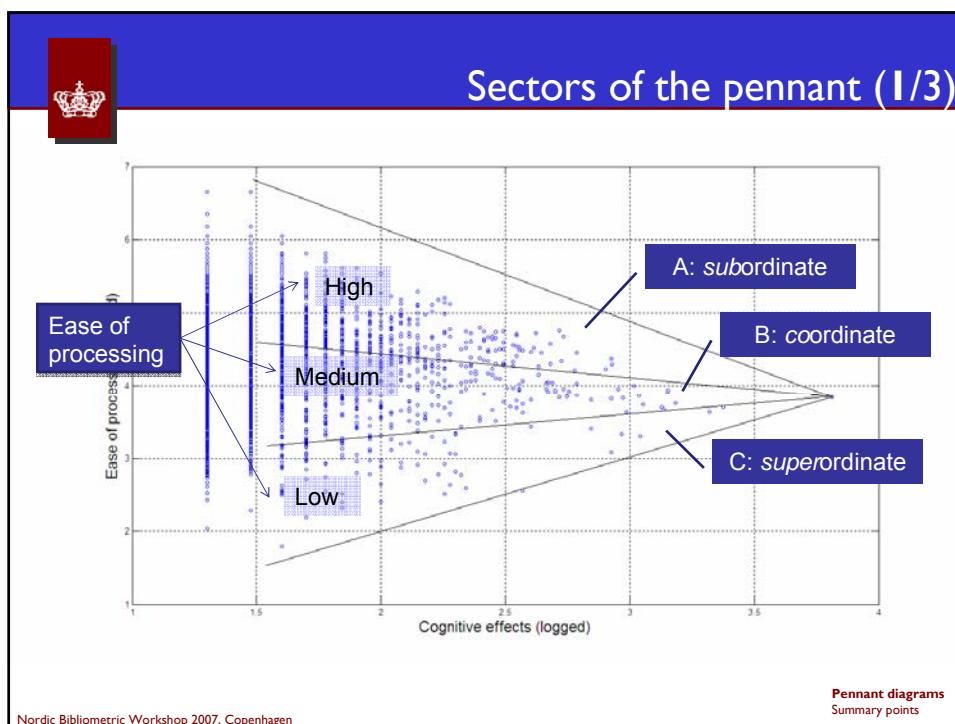
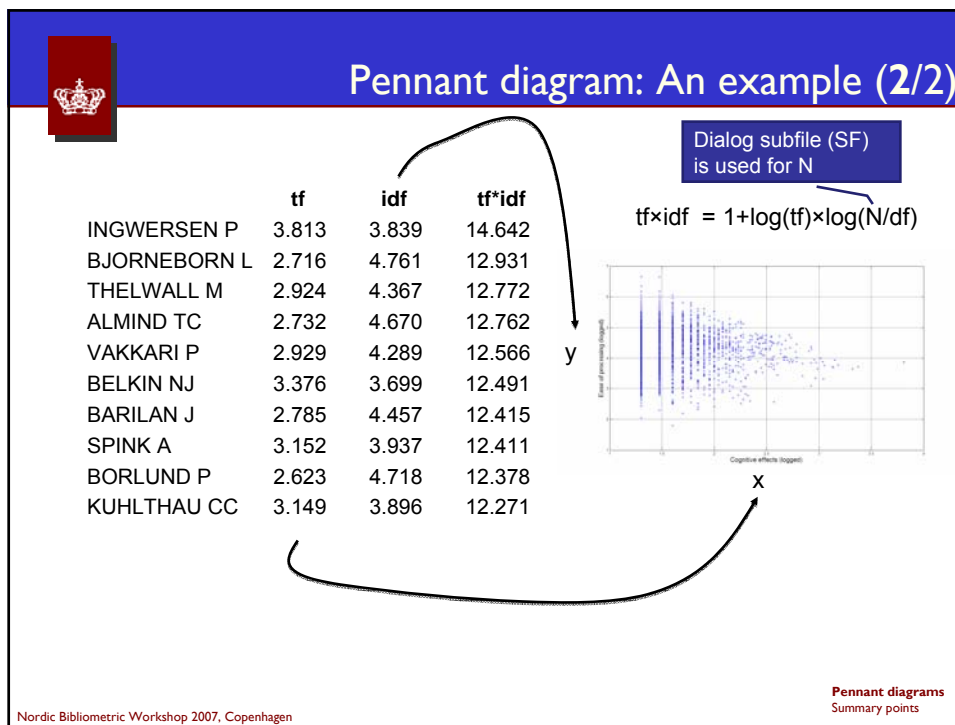
DIALOG RANK Results (Detailed Display)

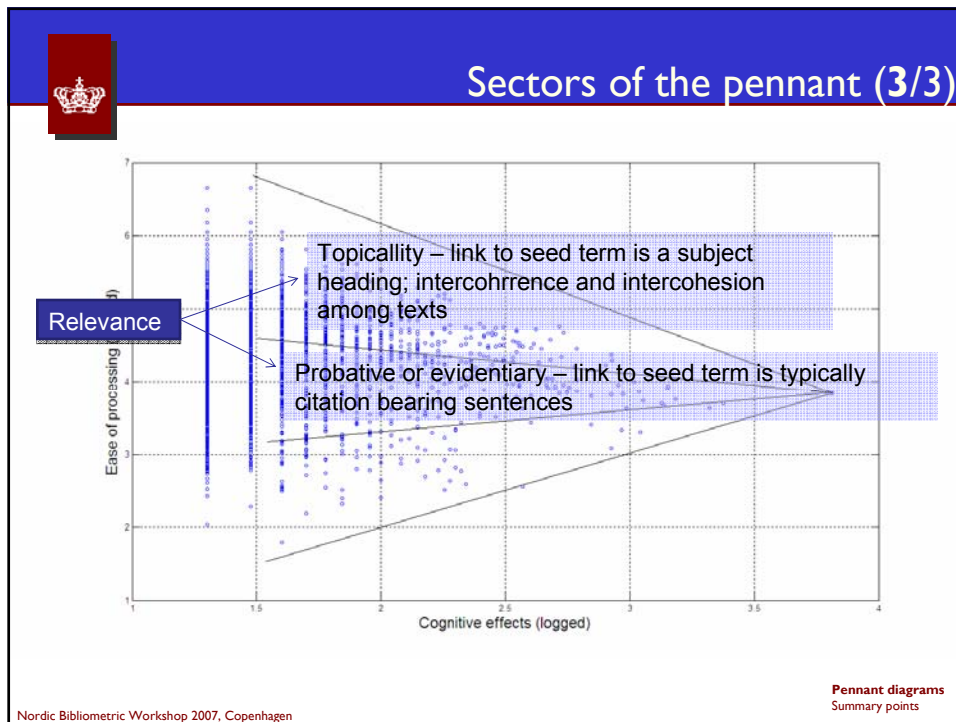
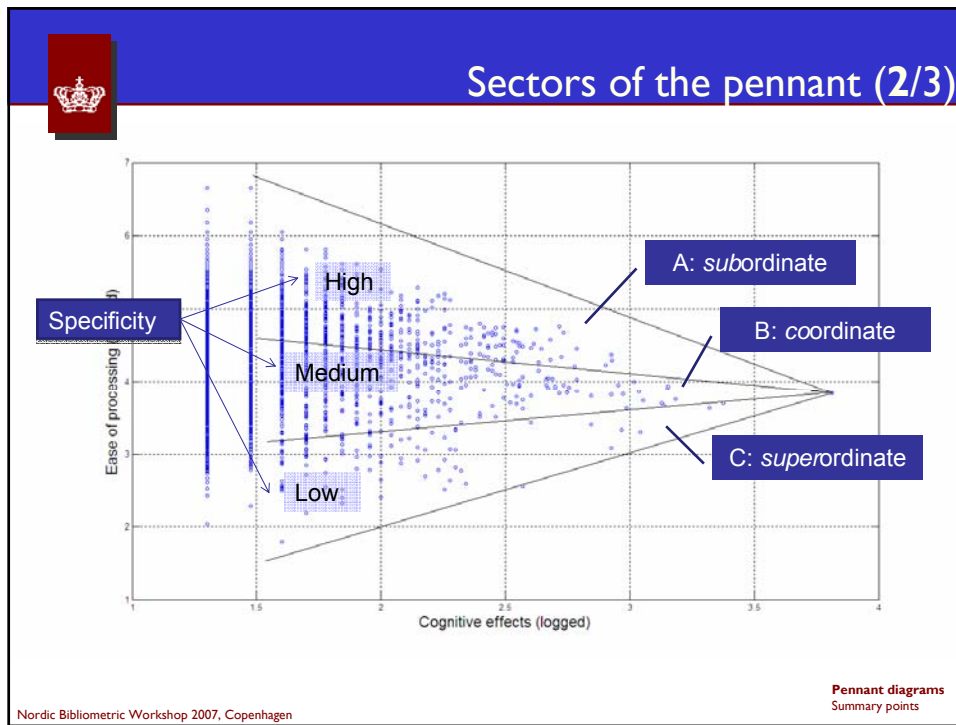
RANK: S2/I-651 Field: CA= File(s): 7
(Rank fields found in 651 records -- 9545 unique terms)

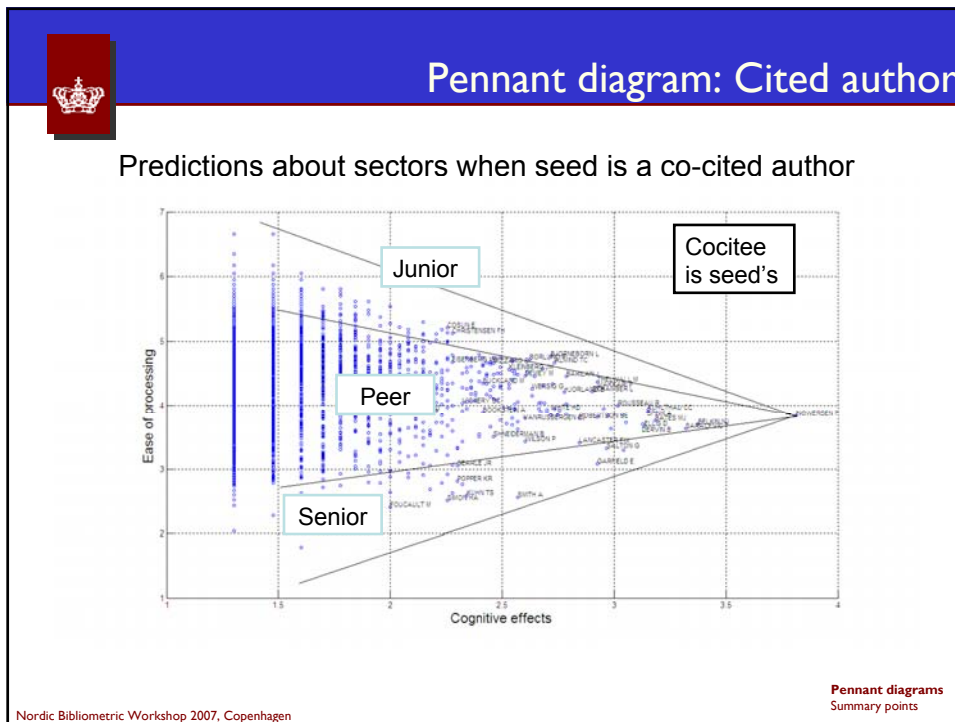
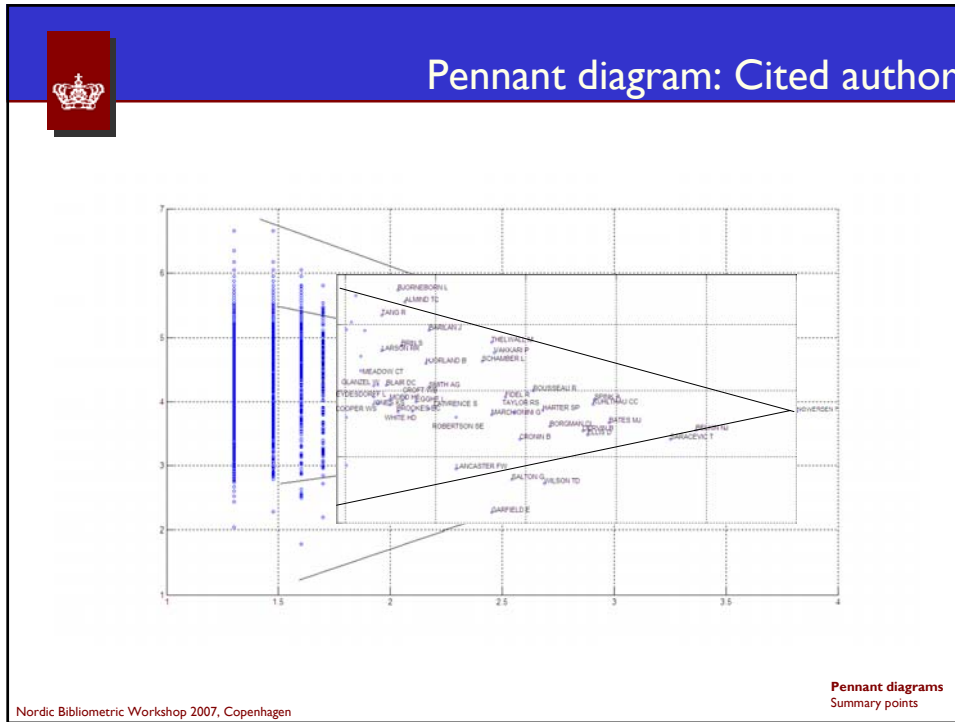
RANK No.	Items in File	Items Ranked	%Items Ranked	Term
1	651	651	100.0%	INGWERSEN P
2	899	238	36.6%	BELKIN NJ
3	1048	209	32.1%	SARACEVIC T
4	789	153	23.5%	BATES MJ
5	520	142	21.8%	SPINK A
6	571	141	21.7%	KUHLTHAU CC
7	967	137	21.0%	ELLIS D
8	901	134	20.6%	DERVIN B
9	833	113	17.4%	BORGMAN CL
10	2269	110	16.9%	WILSON TD

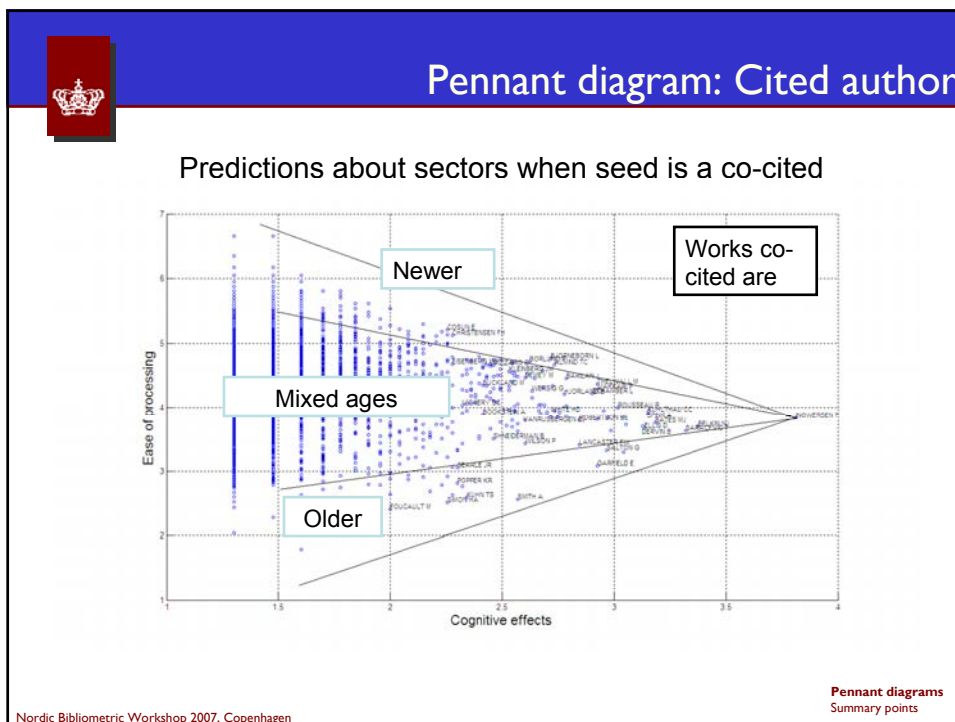
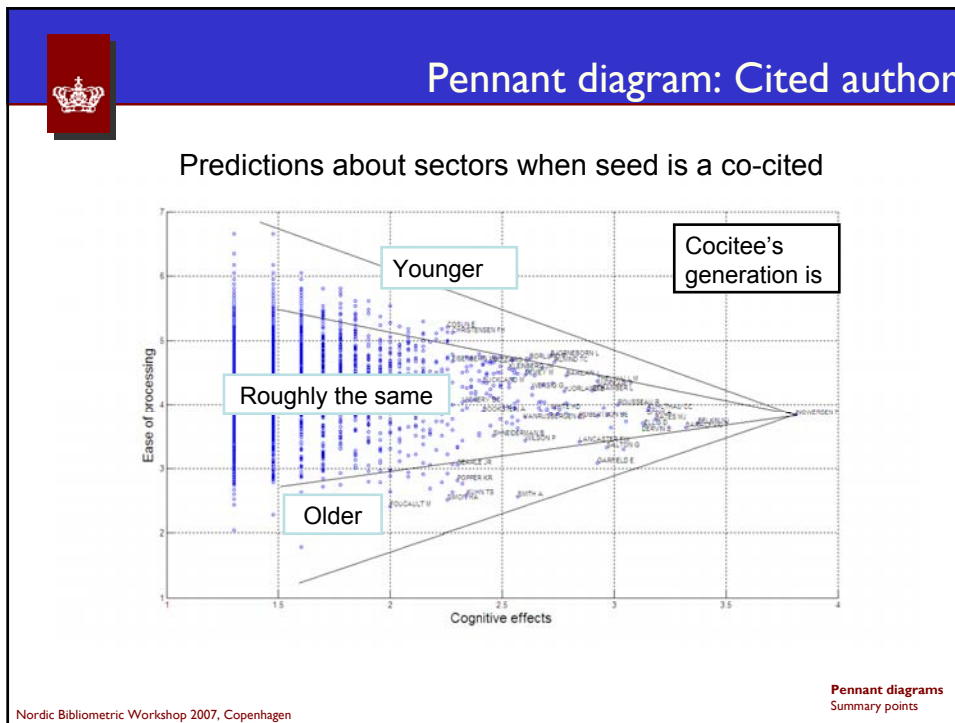
Pennant diagrams
Summary points

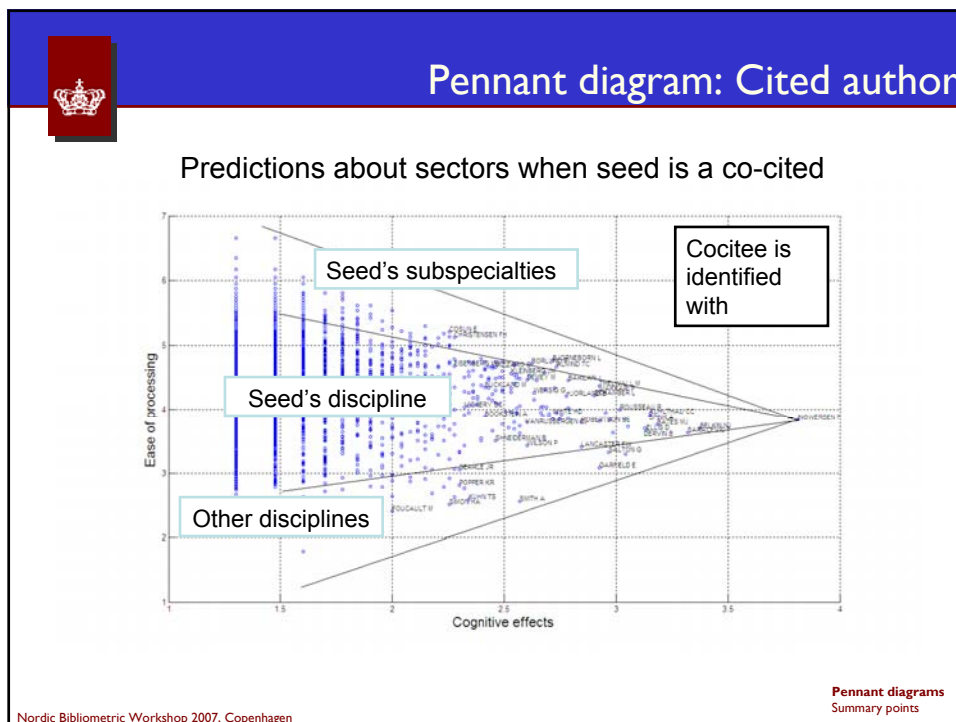
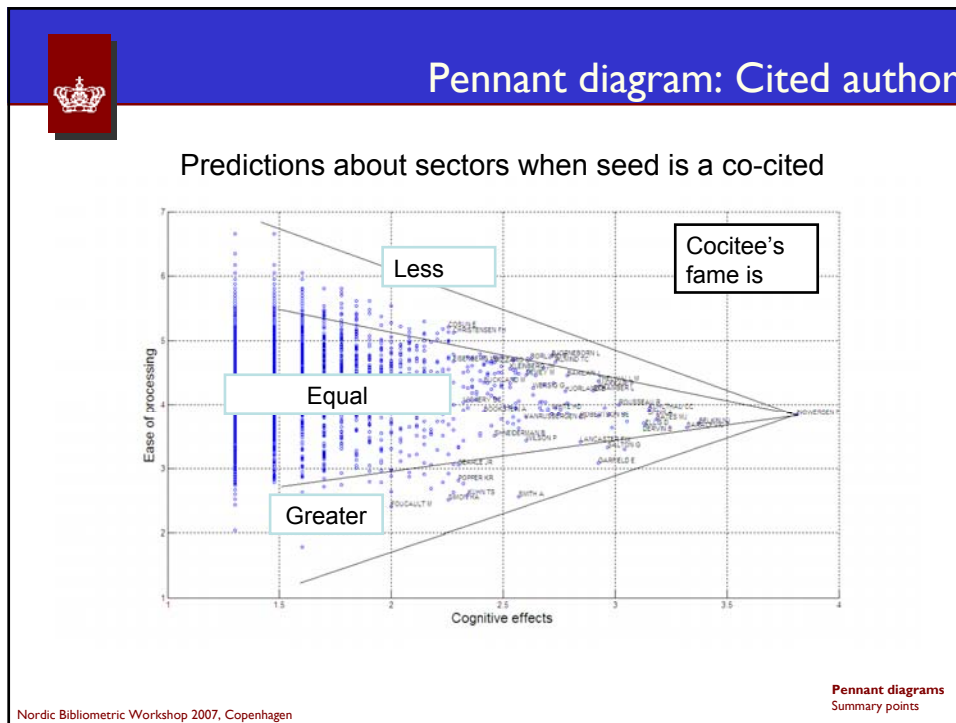
Nordic Bibliometric Workshop 2007, Copenhagen

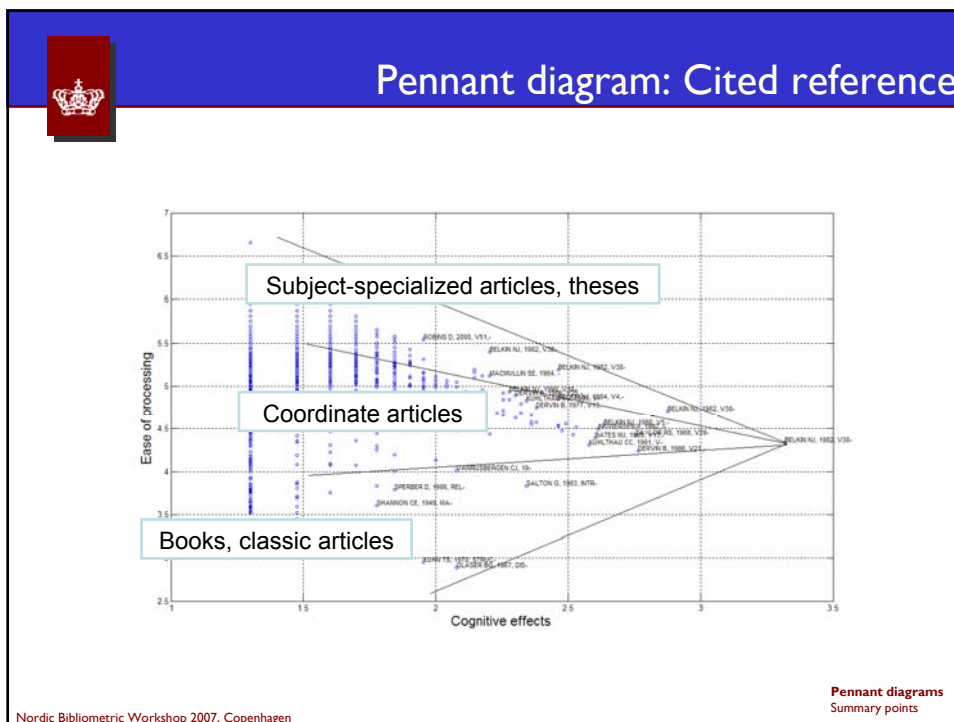
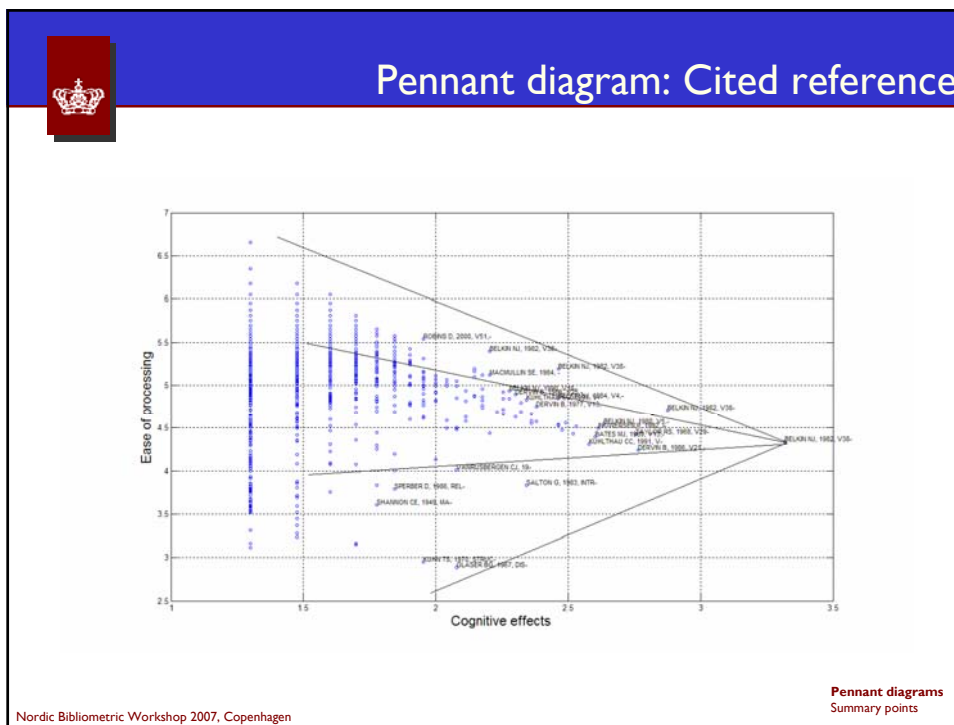


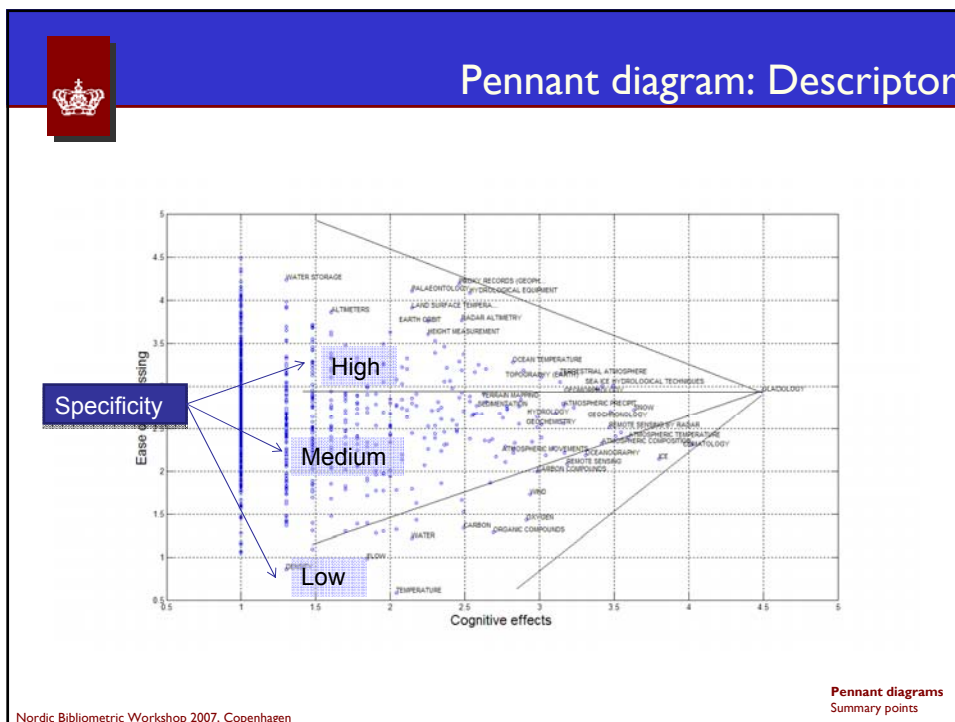
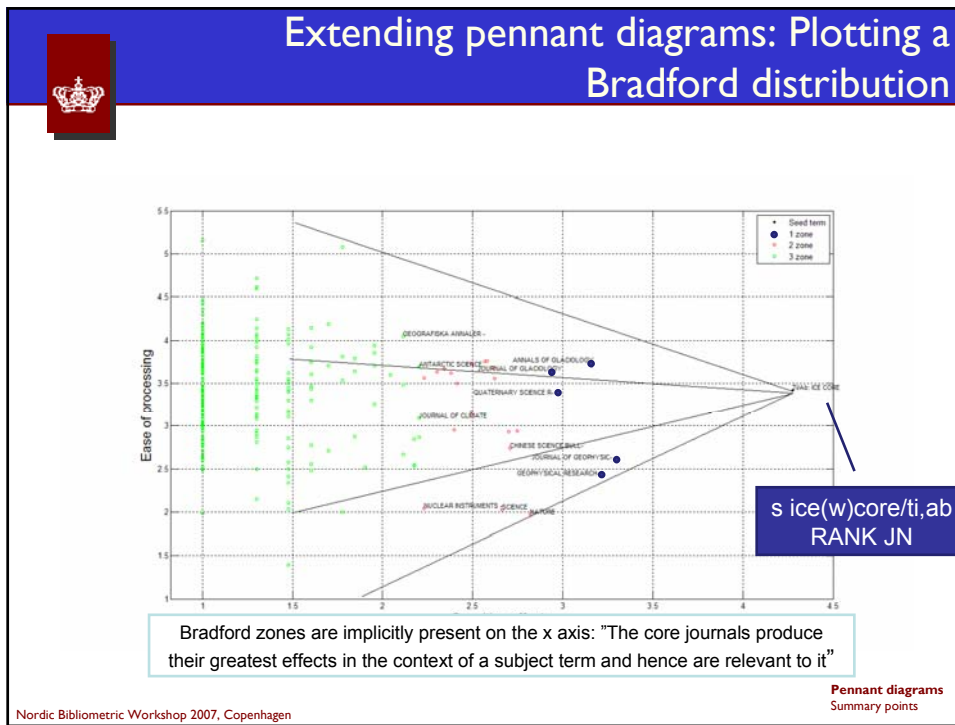
















Summary points

- Most interestingly, when bibliometric data are subjected to $tf \times idf$, plotted as pennants, and interpreted according to relevance theory, the results evoke major variables in information science
 - Topicality (intercoherence and intercohesion among texts)
 - Other types of relevance stratified by sectors
 - Cognitive effects in relation to people's questions
 - Levels of expertise as a precondition for cognitive effects
 - Processing effort (principle of least effort)
 - Specificity of terms as it affects processing effort
 - Relevance as the effects/effort ratio
 - Authority of texts and their authors

Nordic Bibliometric Workshop 2007, Copenhagen Summary points



Summary points

- What is their potential use?
 - Instrumental for information retrieval and bibliometric analyses?
 - It is evident that statistical specificity (idf) is an extremely interesting model for bibliometric distributions
- Research questions
 - Investigate whether the three sectors can be algorithmically set so that qualitative gradations in the specificity of terms are preserved at some level
 - Validation of the topical dimension of specificity through content analysis
 - Validation of ease of processing through trials with domain experts
 - Experimenting with merged diagrams based on more seed terms (polyrepresentation)

Nordic Bibliometric Workshop 2007, Copenhagen Summary points