Evaluating research and patenting performance using elites: A preliminary classification Scheme

Max Kuan
National Taiwan University of Science and Technology

Background of study

For a number of *entities*

Each has a set of *publications*

Background of Study

How to compare
the performance of these entities
in terms of
their respective publication sets?

Background of Study

- Many factors to consider
  - Quantity of publications
  - Quality of publications
  - Size of entity
  - Entities involve multiple disciplines
  - Publications may be affiliated with multiple entities
  - …

Background of Study

A type of evaluation methods

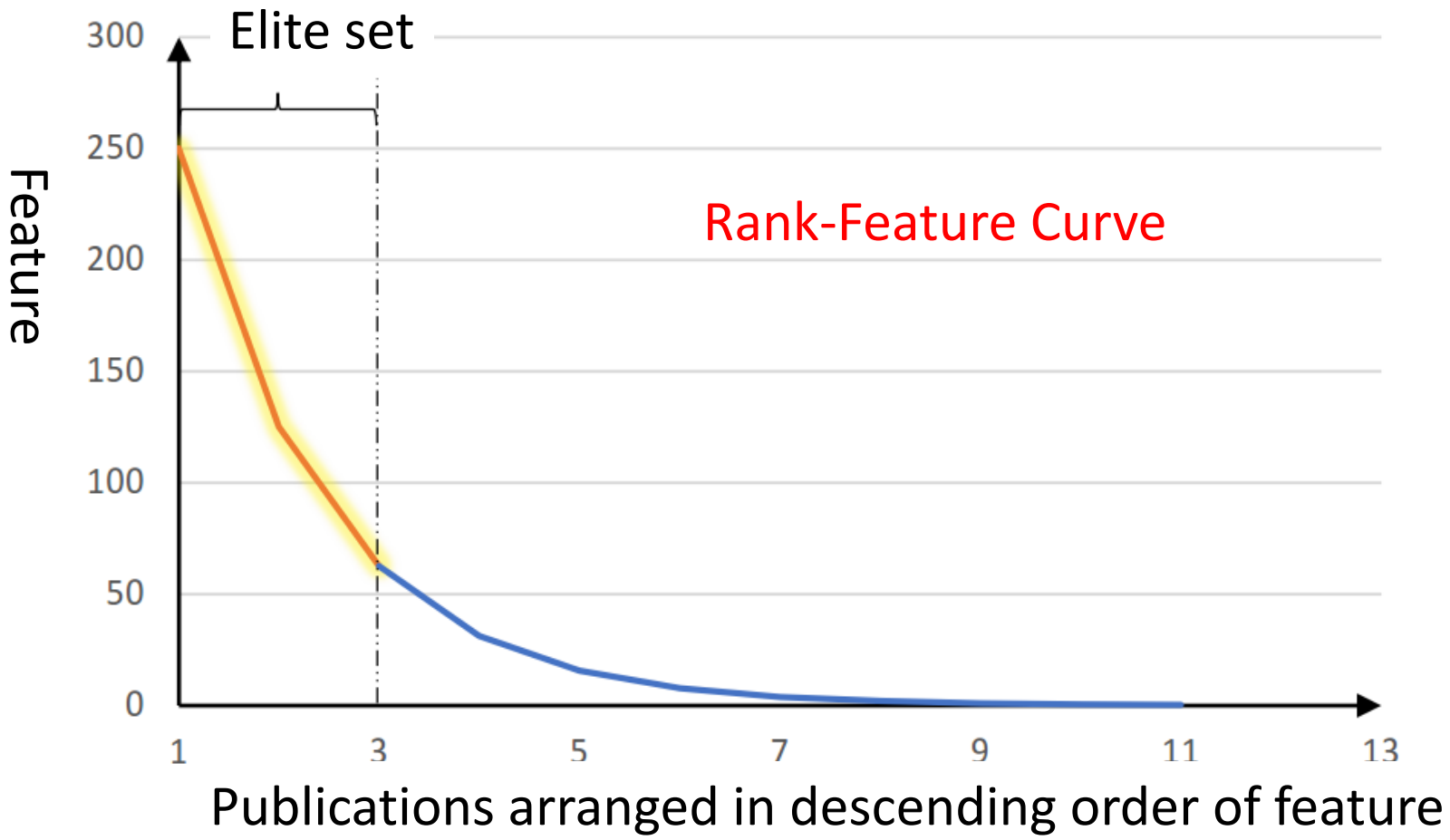
1.Using a few outstanding *elites* (Peter Vinkler) as representatives

2. Comparing the entities using their *elites*

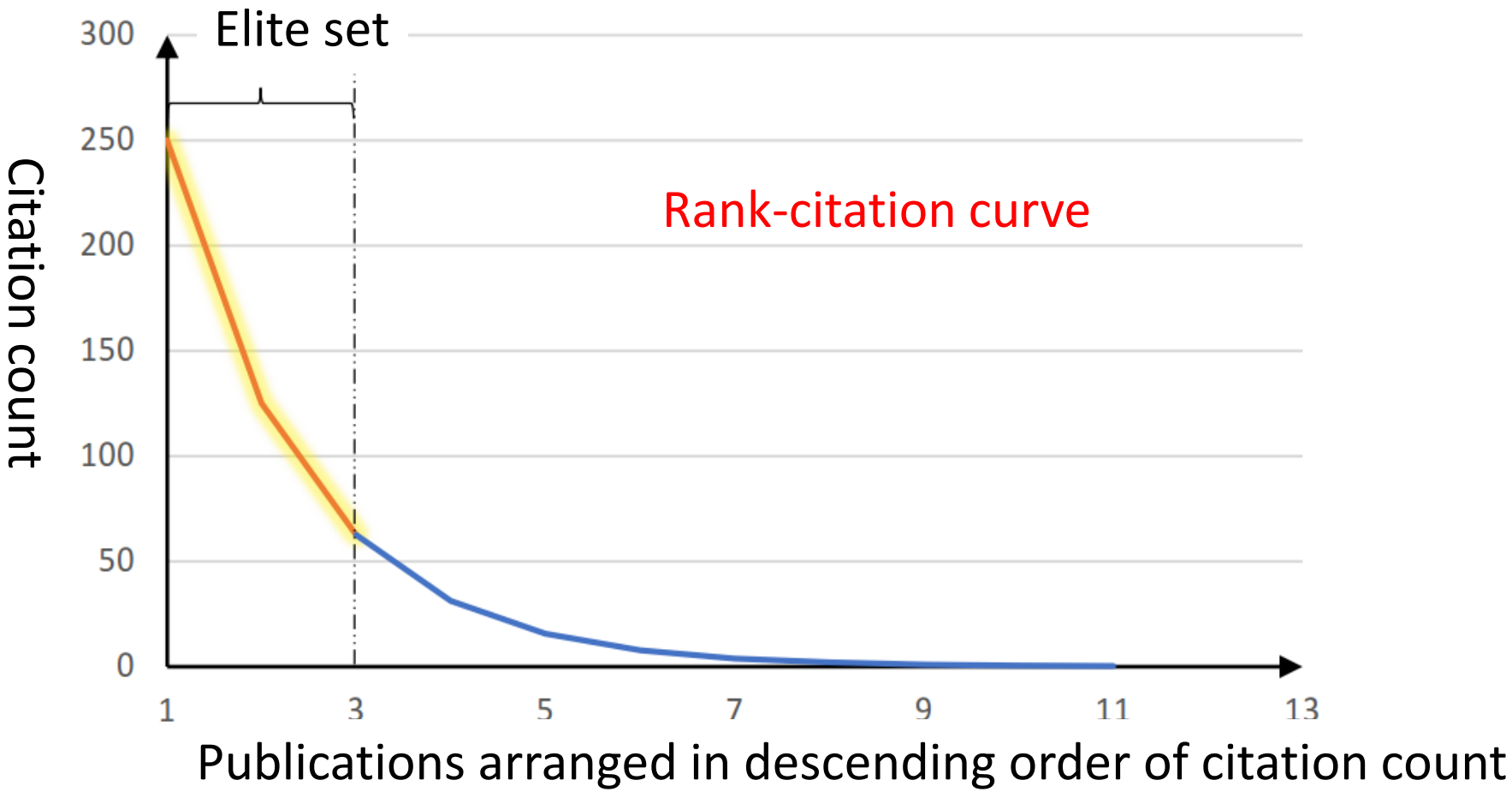
Elite-based Evaluation

- Some Examples
  - Academic Ranking of World Universities (ARWU) has scores determined by the alumni and faculty receiving Nobel prizes or Fields medal
  - the number of highly cited papers (HCPs) as a measure to a researcher
  - *h*-index
  - …

Visual Representation



Visual Representation

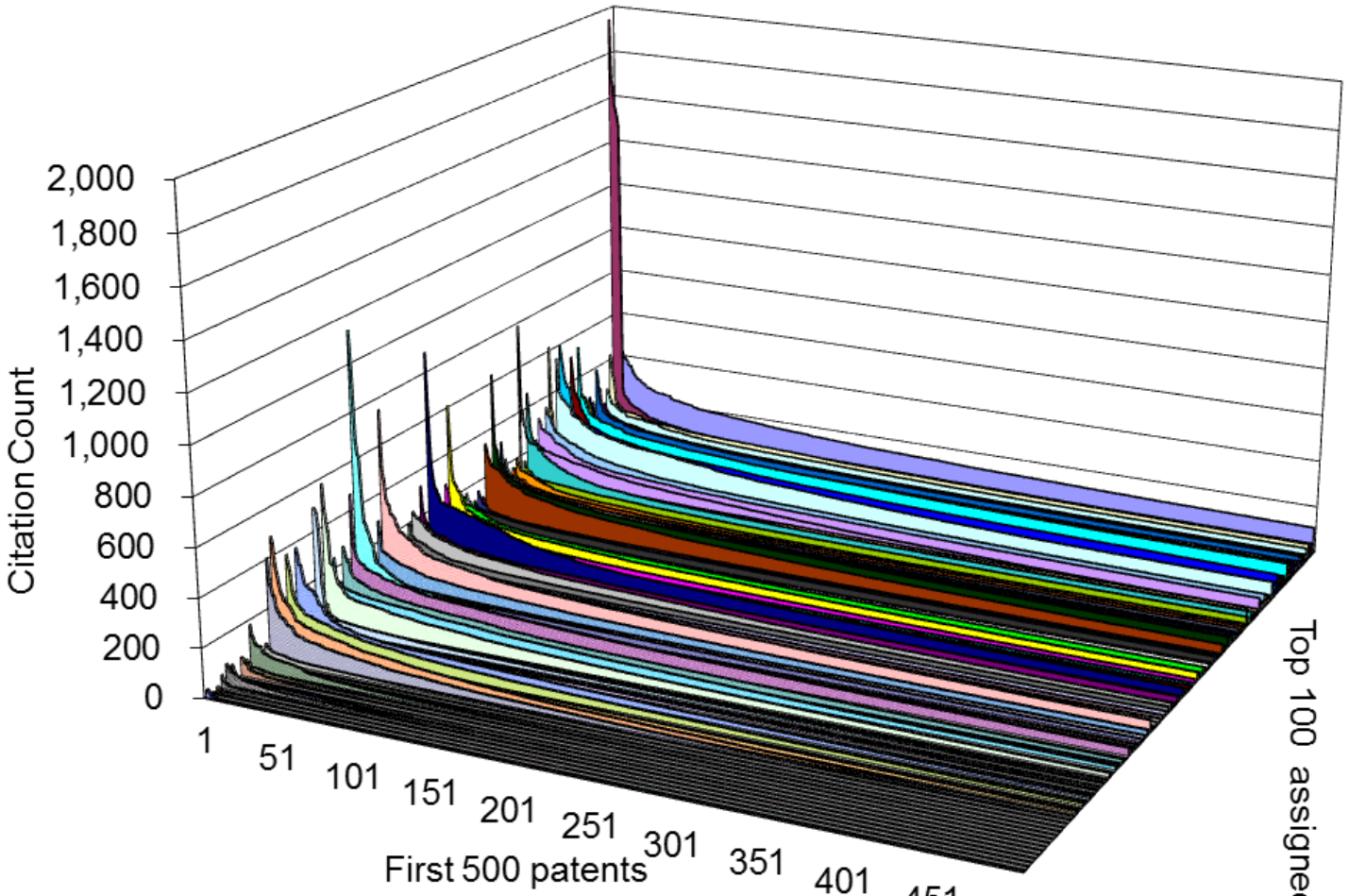


Elite-based Evaluation

- Advantages
  - Especially suitable for entities with very *skewed* distributions of feature
    - A great majority of mediocre publications
    - A very small number of outstanding publications
    - Typical example: patent assignee and its patent portfolio
  - Entity size is of little impact

Elite-based Evaluation

- 2009 Top 100 Patent Assignees of USPTO

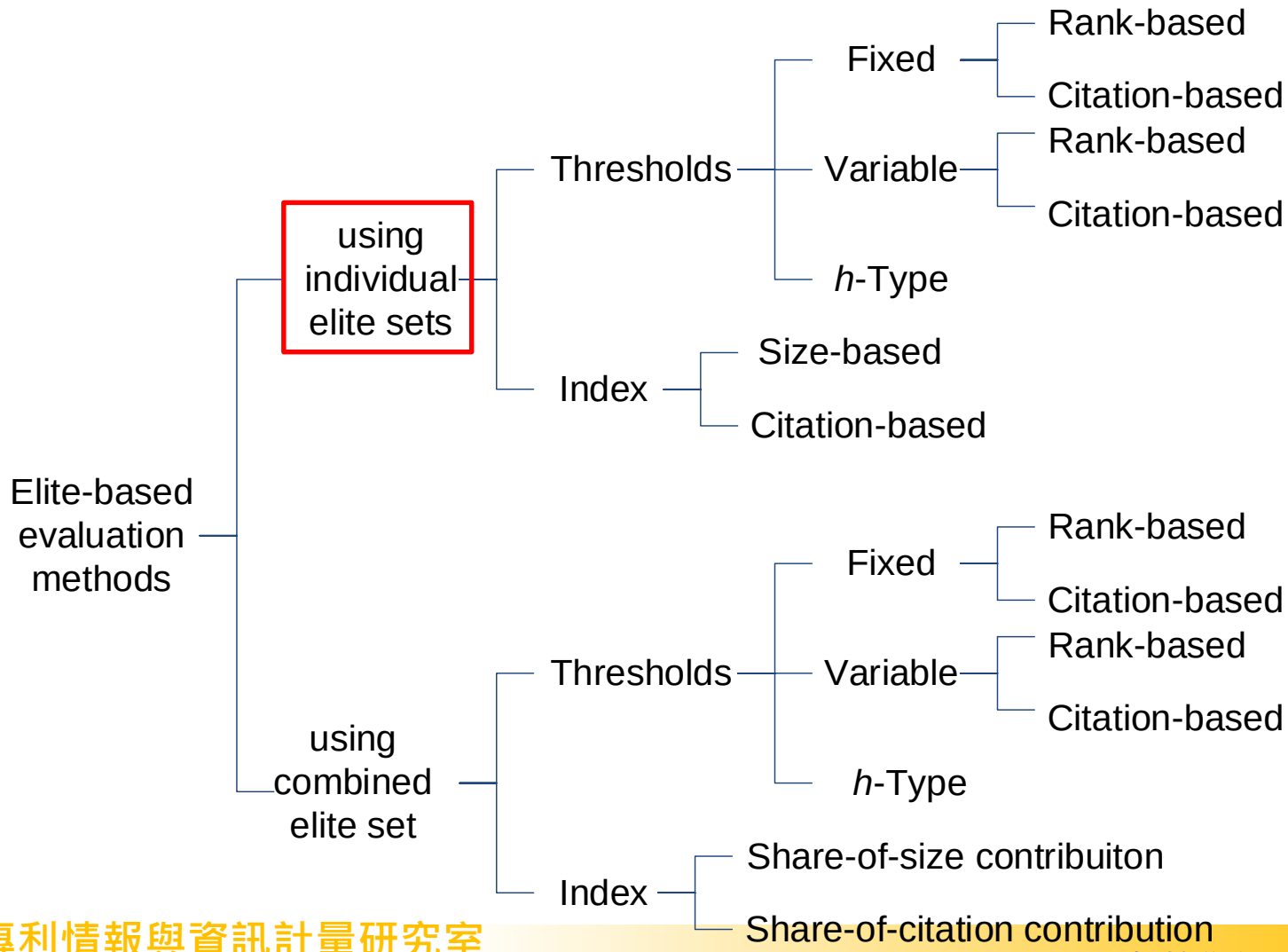


A Classification Scheme

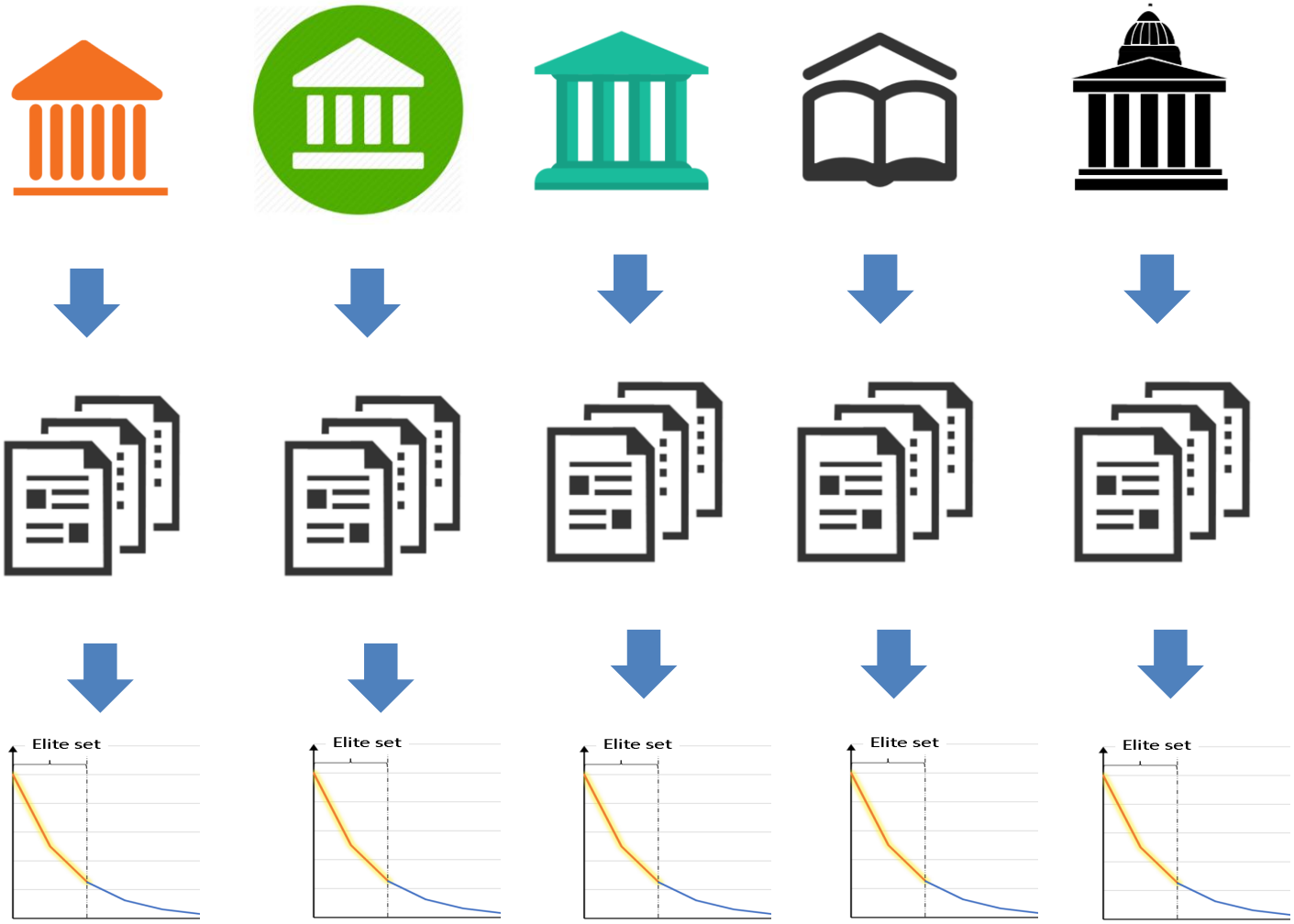
Classification Scheme

- Purpose
  - Grouping similar methods together
  - Positioning one's own method
  - So as to observe their differences

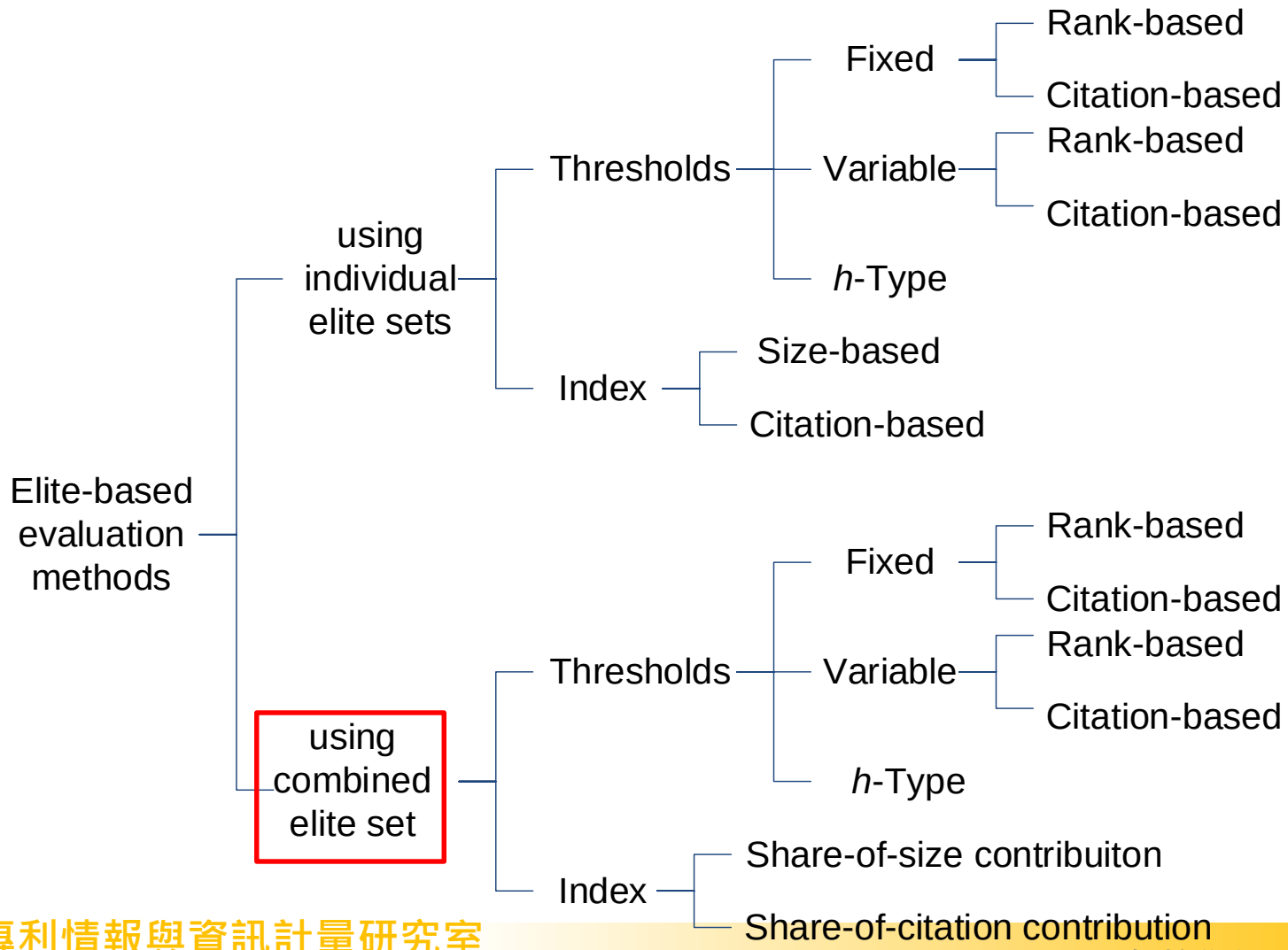
Classification Scheme



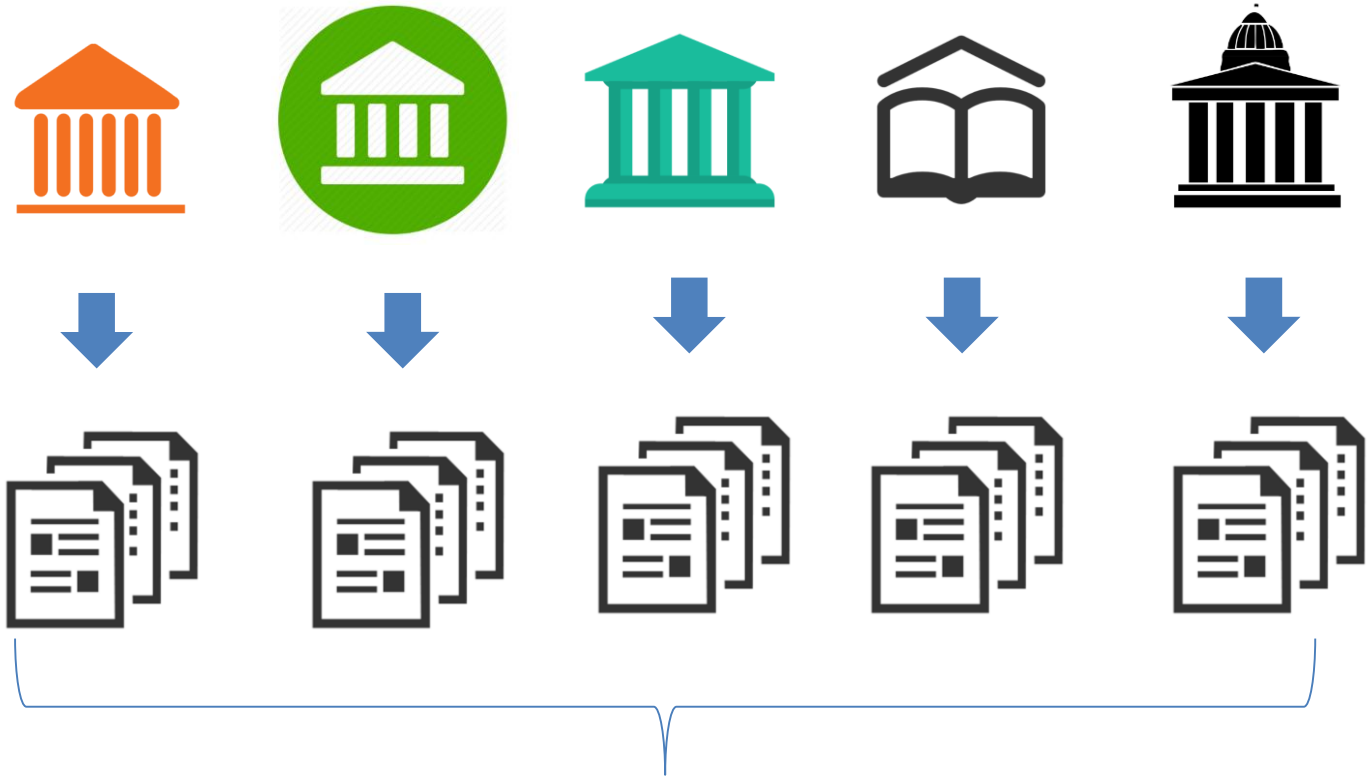
Using individual elite sets

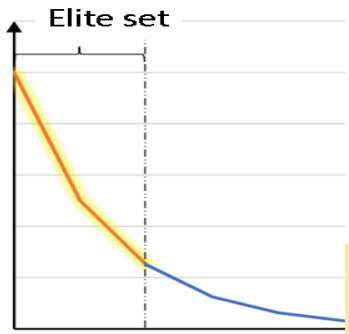


Classification Scheme

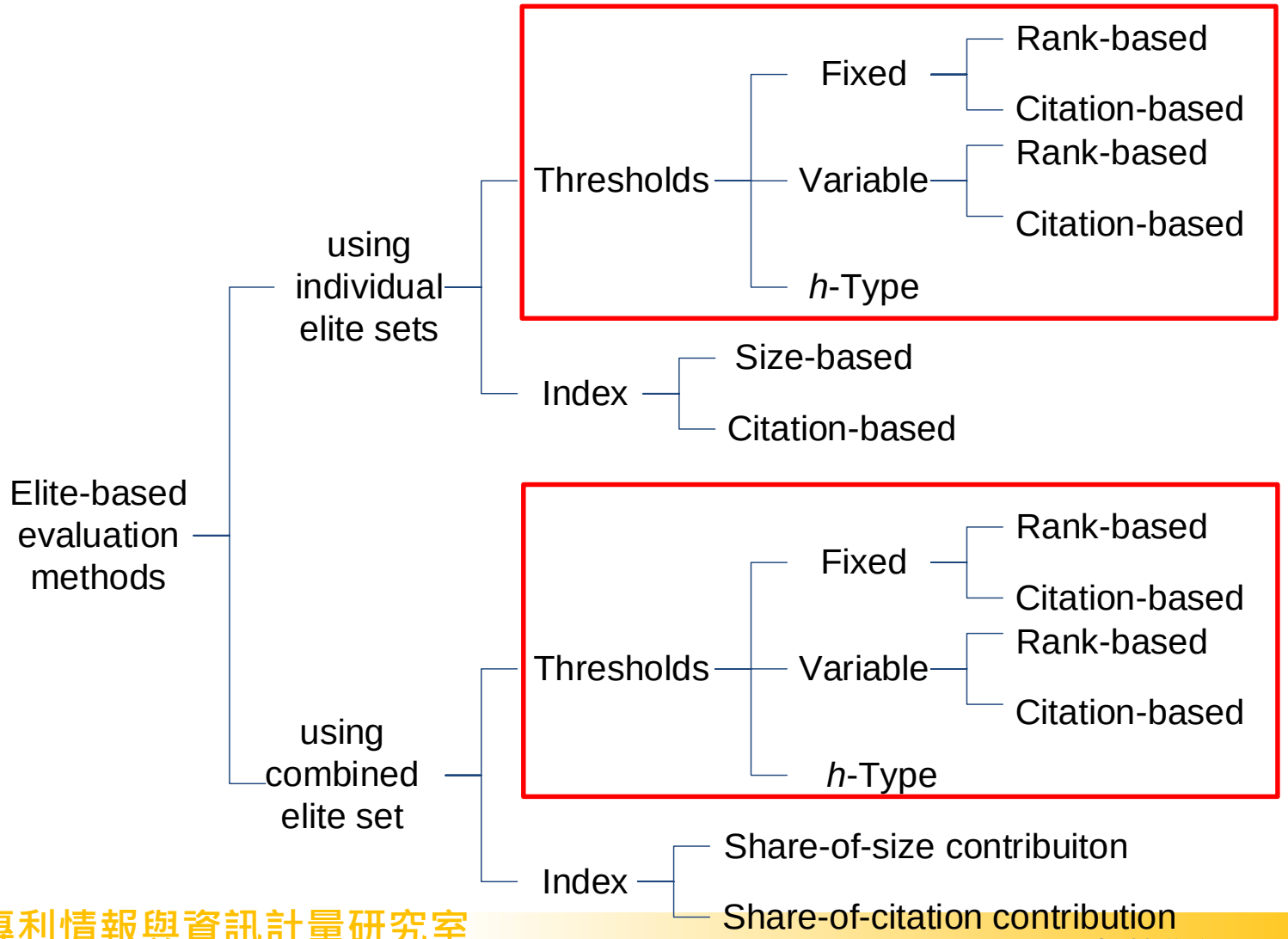


Using combined elite set

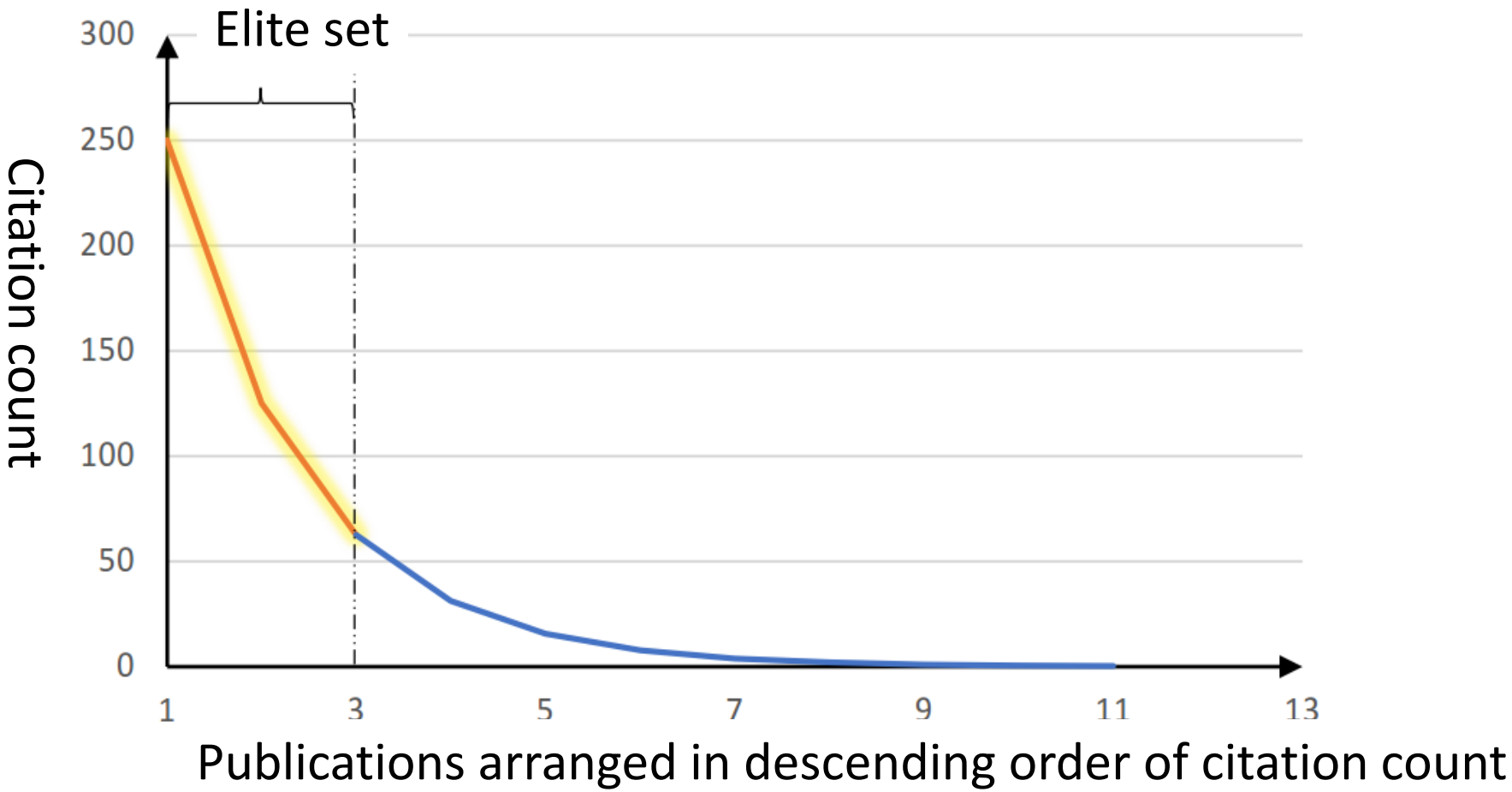




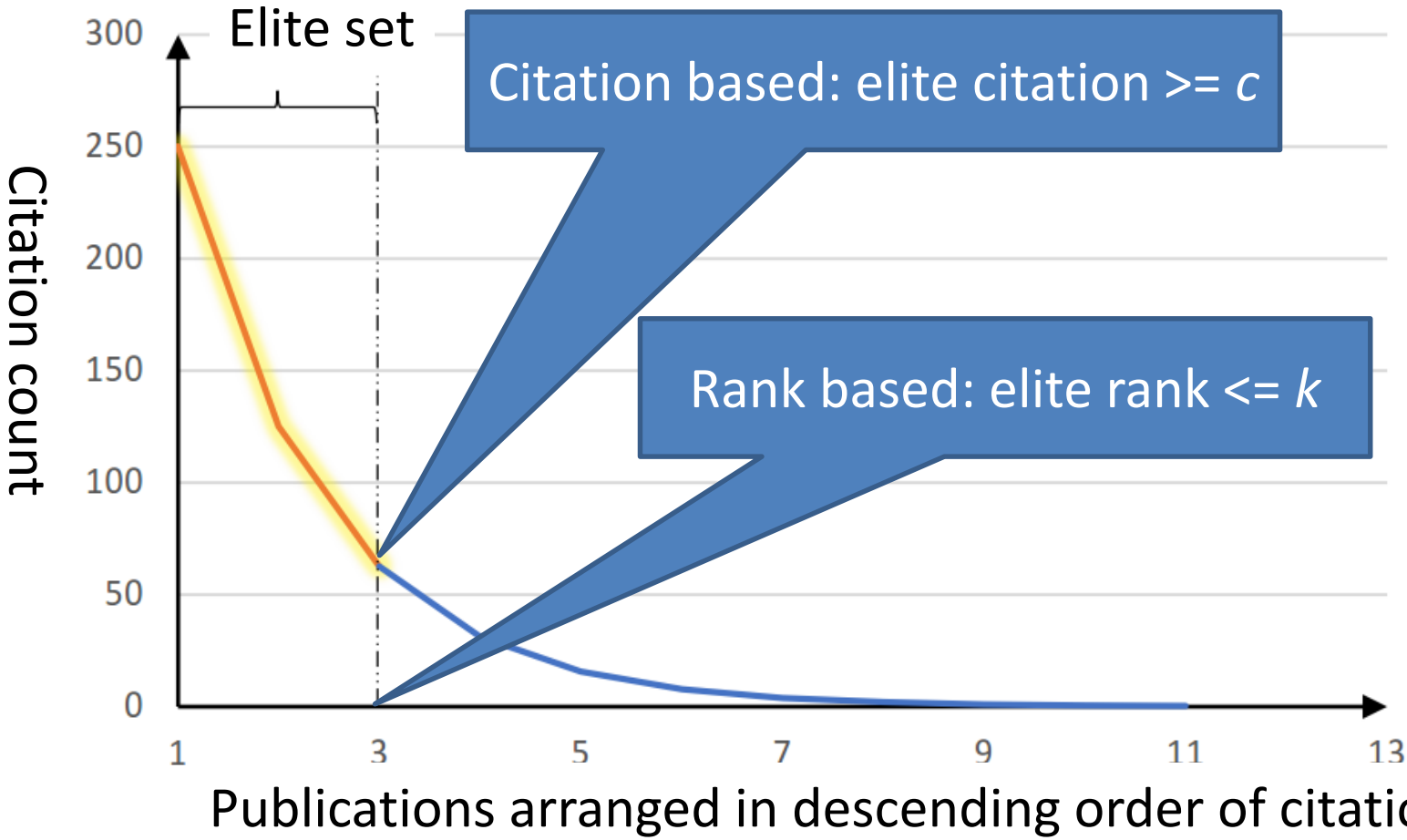
Classification Scheme



Threshold: who qualifies as elite?



Rank and Citation Fixed Threshold



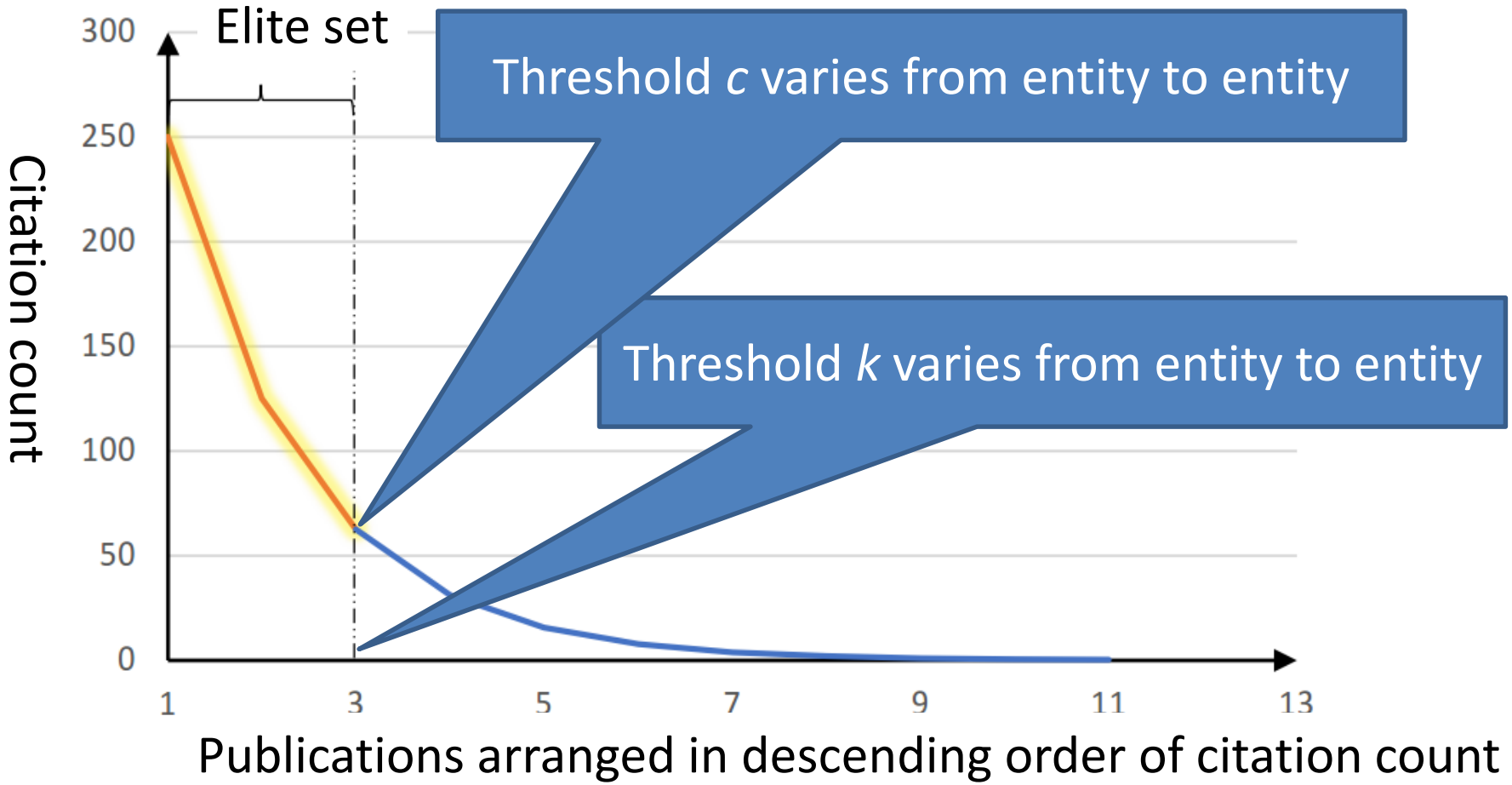
Examples of Fixed Threshold

- Rank-based
  - Garfield [20] considered the 100 most frequently cited *life science* publications published as elites.
  - Frogel [19] selected the first, the first 50, and the first 100 most frequently cited *astronomy* publications as elites.
  - Ryan and Woodall [44] applied the same concept to *statistics* publications with the rank threshold set at 25.
  - Patsopoulos, Ioannidis, and Analatos [40] chose 30 as the rank threshold for *medicine-related* publications.

Examples of Fixed Threshold

- Citation-based
  - Plomp [41] considered a researcher's elite publications are those receiving at least 25 citations.
  - The i10 and i100 indices of Google Scholar uses fixed citation thresholds 10 and 100.
  - Blessinger and Hryca [6] used 10 and 50 citations as criteria to generate two groups of elite publications.
  - Garfield [20] set the fixed citation threshold at 10.

Rank and Citation Variable Threshold



Examples of Variable Threshold

- Rank-based
  - The Highly cited papers, Hot papers, ESI most cited papers, etc. of Thomson Reuters uses variable rank thresholds with functions $0.01\%N_j$, $0.1\%N_j$, $1\%N_j$, respectively.
  - Fernandez-Alles and Ramos-Rodríguez [18] used a function $1.45\%N_j$.
  - The π-index of Vinkler [47] considered only the top $\sqrt{N_j}$ most frequently cited publications of the evaluated researchers.
  - Vinkler's another π_v index [48] is for evaluating journals, and each journal is assessed by its $\left(10 \log N_j\right) - 10$ most frequently cited publications.

Examples of Variable Threshold

- Citation-based
  - There are few *citation-based variable threshold* methods.
  - One example (cf. [21]) is that the elite publications are those receiving at least $k\bar{C}$ citations where $\bar{C}$ is the average number of citations.
  - Another example is that an entity *j*'s elite publications are most frequently cited publications jointly producing a certain percentage of the entity *j*'s all citations C_j.

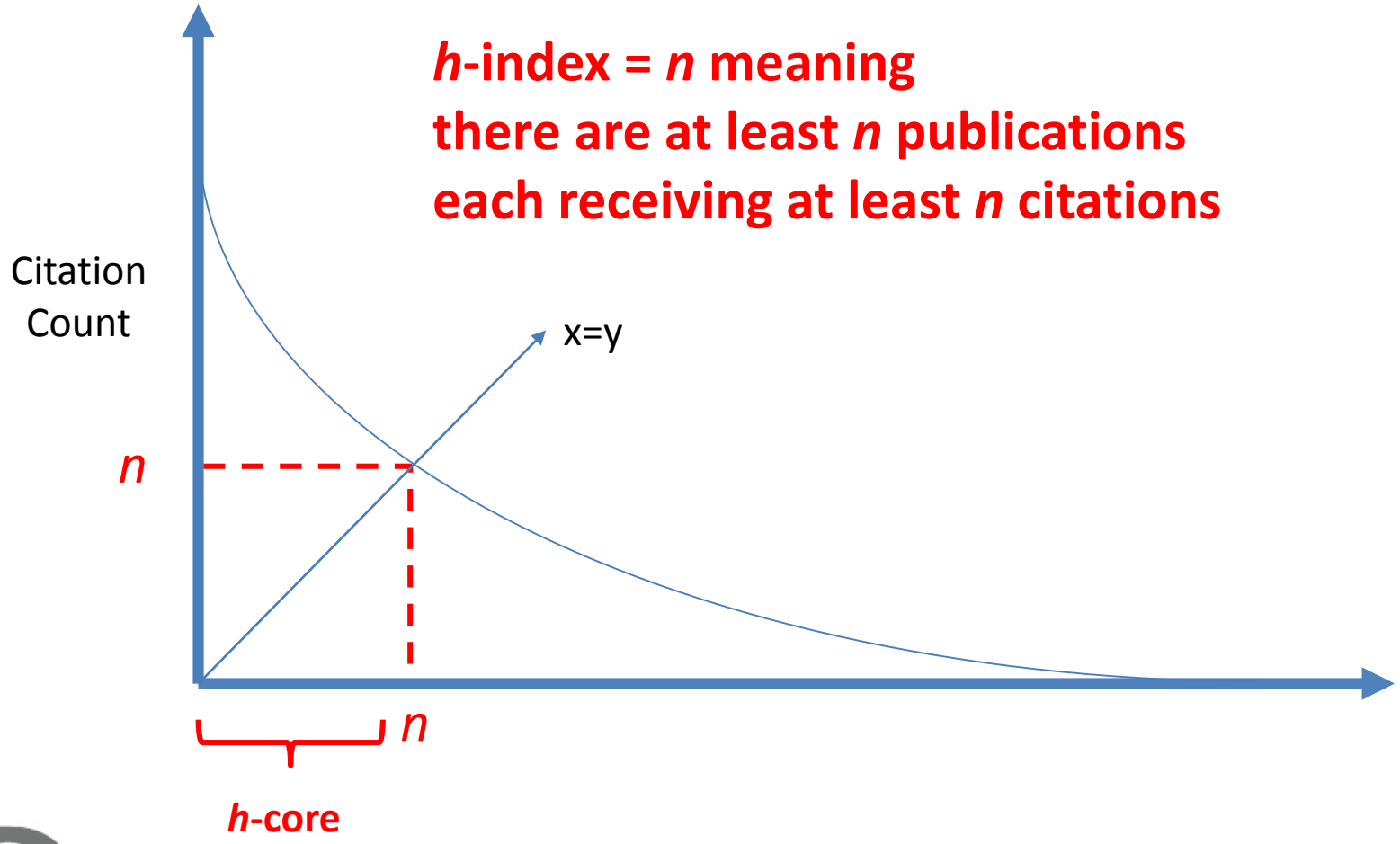
Fixed Threshold: Pro & Con

- Simple and intuitive
- Common criterion across all entities
- the choice and justification to a particular fixed threshold is probably much more complicated and difficult.
  - Usually depending on the discipline of the publications

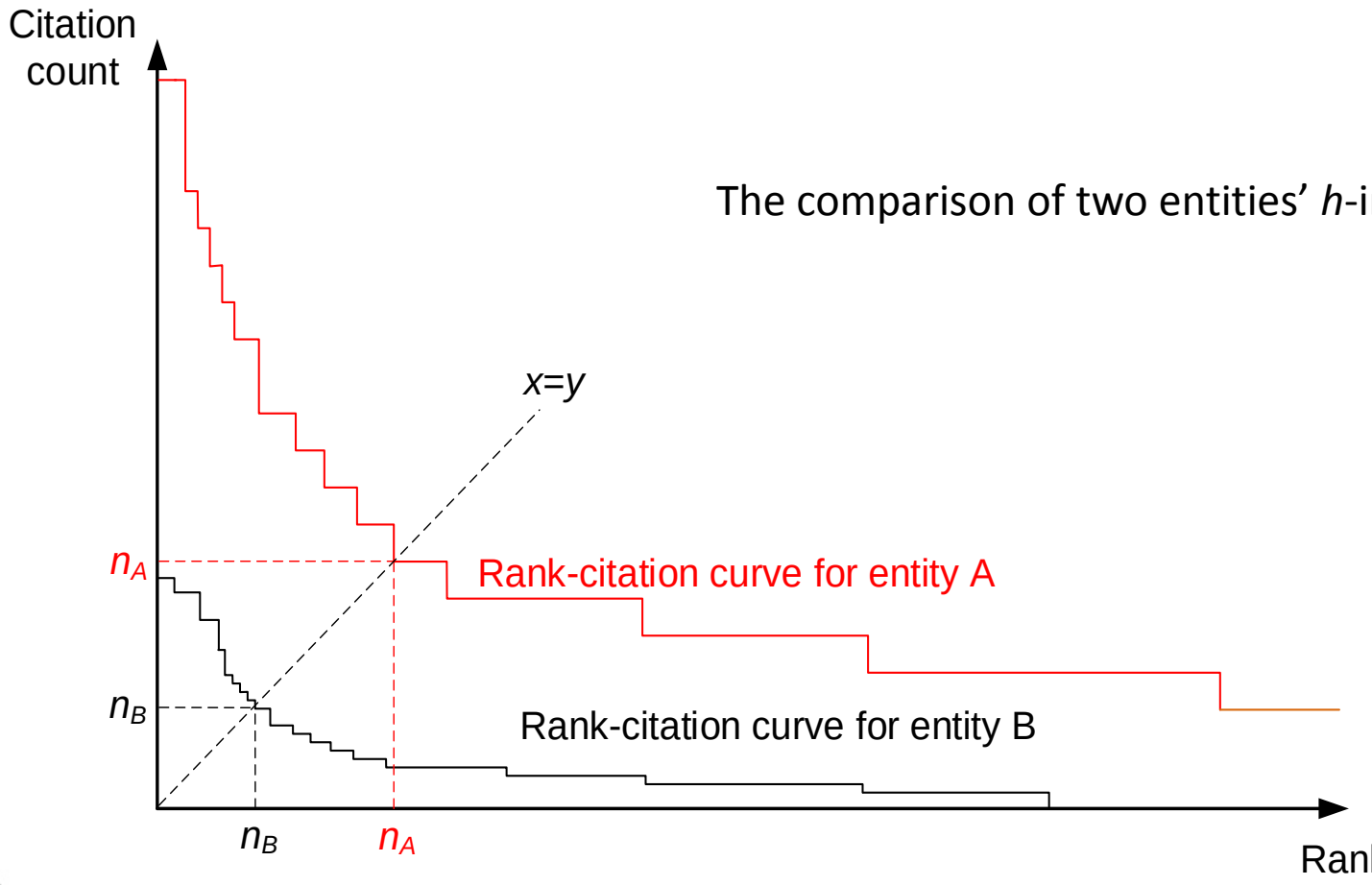
Variable Threshold: Pro & Con

- Flexible and are adaptable to different entities
- The disadvantage is that there is not a single uniform criterion

h-type Variable Threshold



h-type Variable Threshold

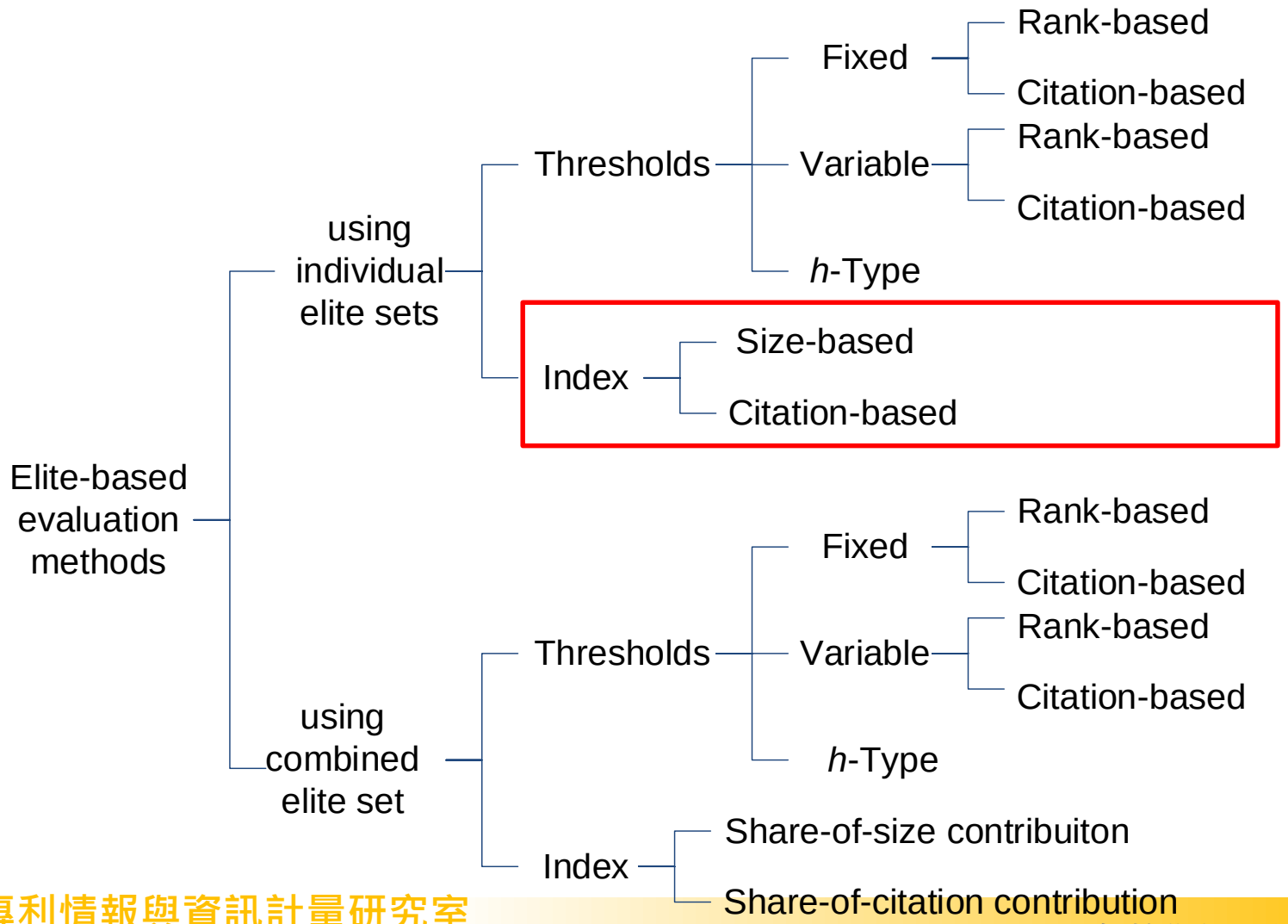


The comparison of two entities' *h*-indices

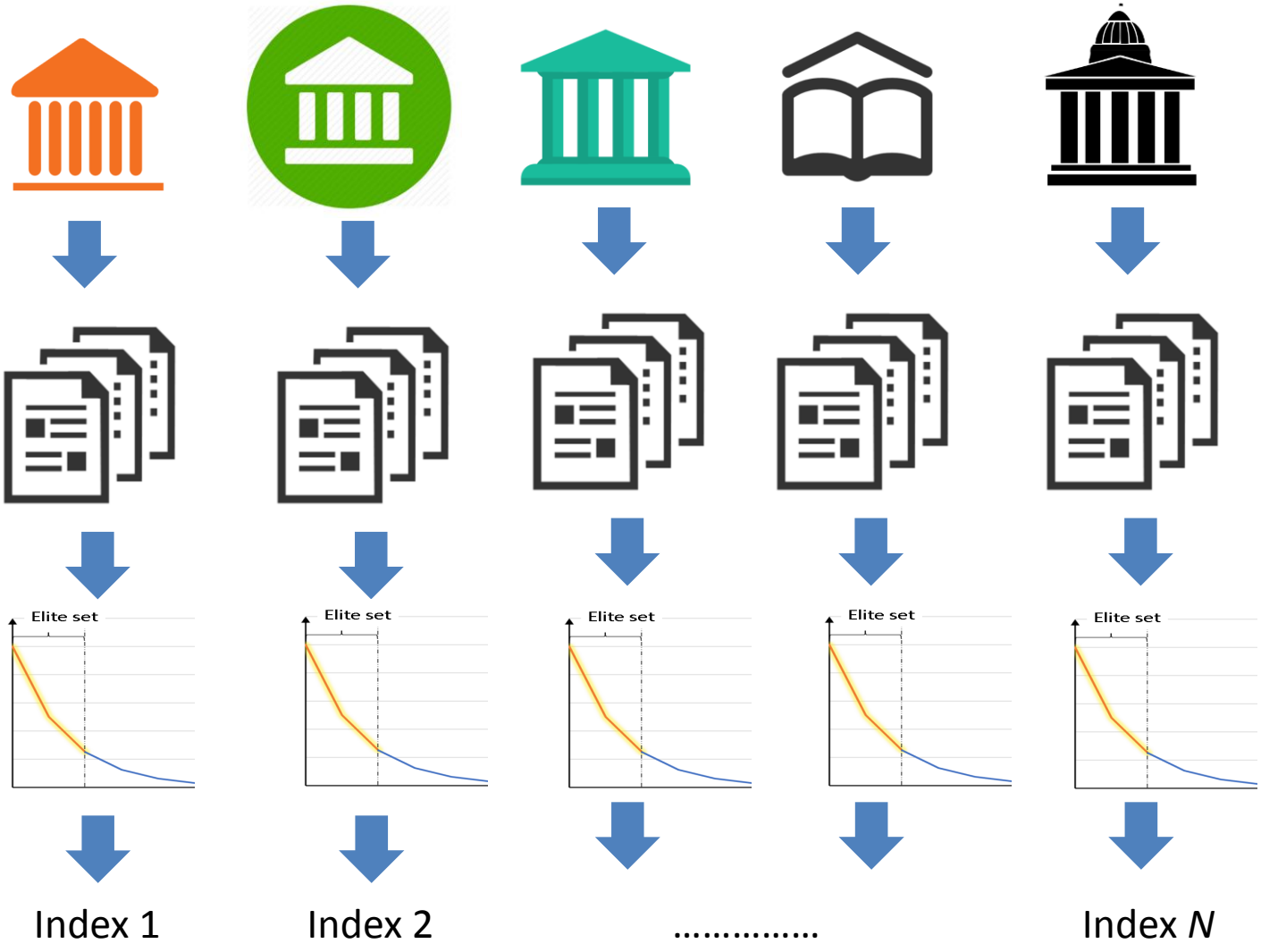
h-type Variable Threshold

- Neither rank- nor citation-based
- *h*-Type thresholds provides a uniform approach similar to the fixed thresholds
- But also adaptable to different entities, like the variable thresholds

Classification Scheme



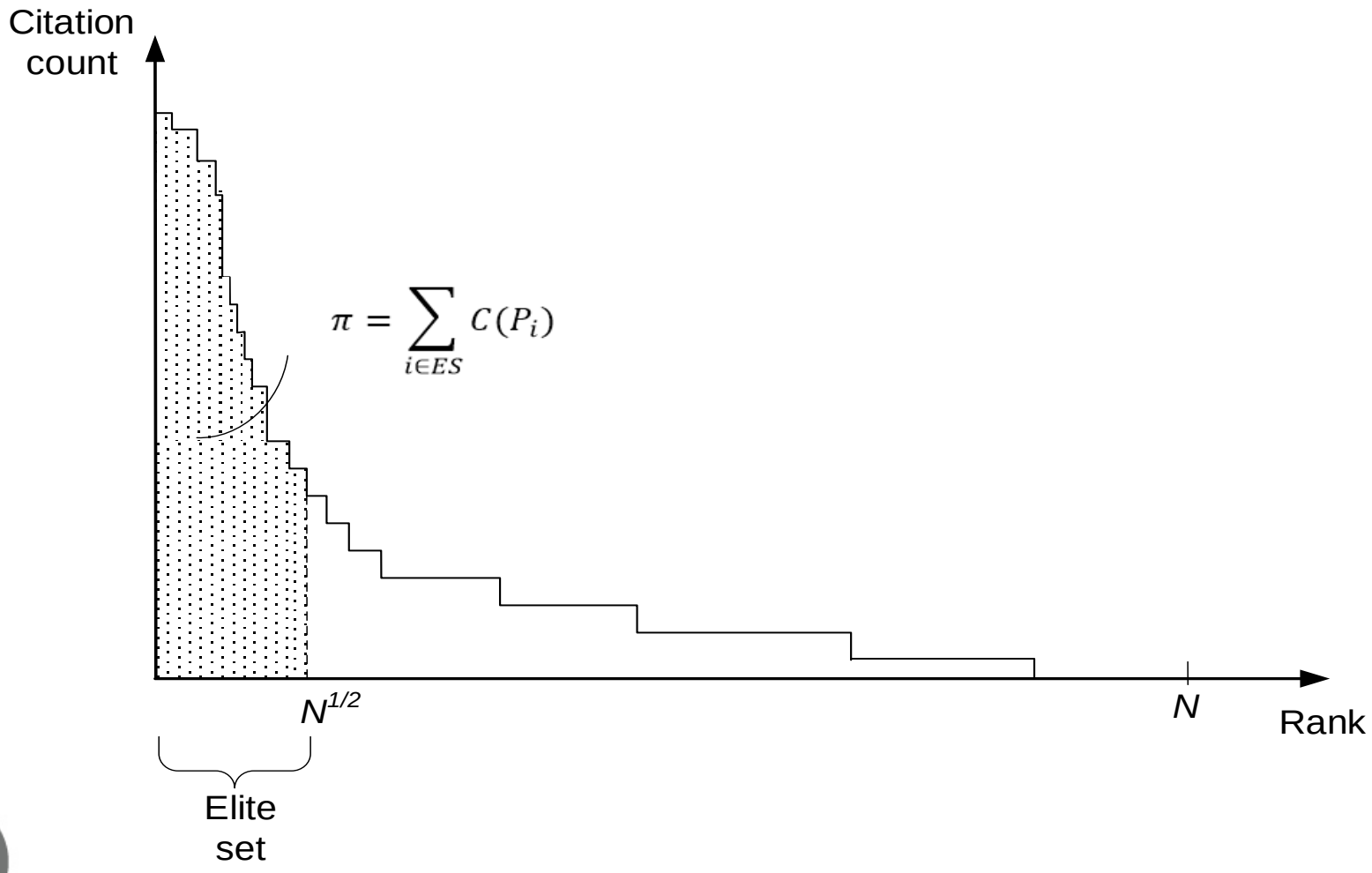
Using individual elite sets



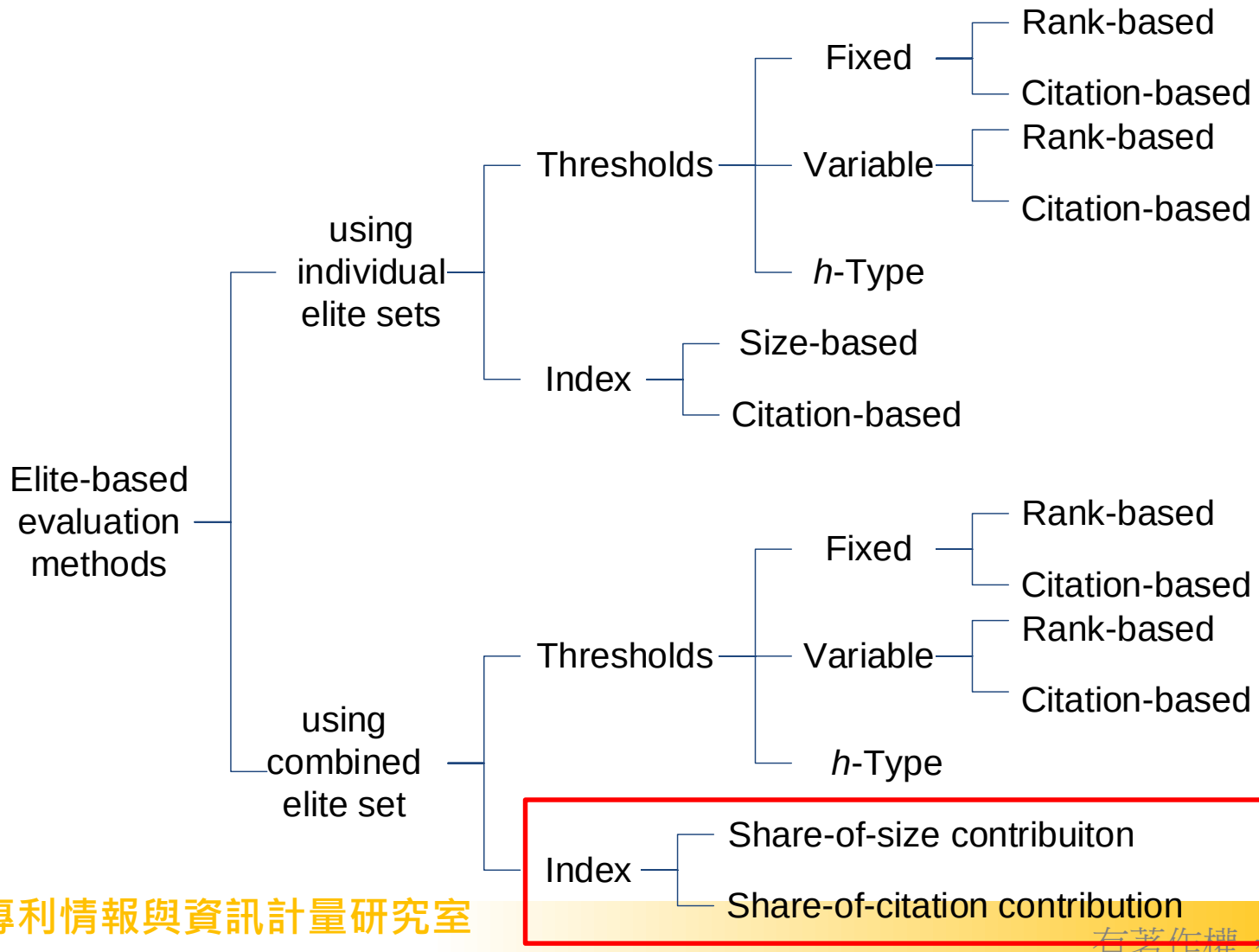
Indices of the elite sets

- Size-based index
  - No. of elites in the elite set
- Citation-based index
  - Citations received by the elites
    - therefore is an “area” under the rank-citation curve

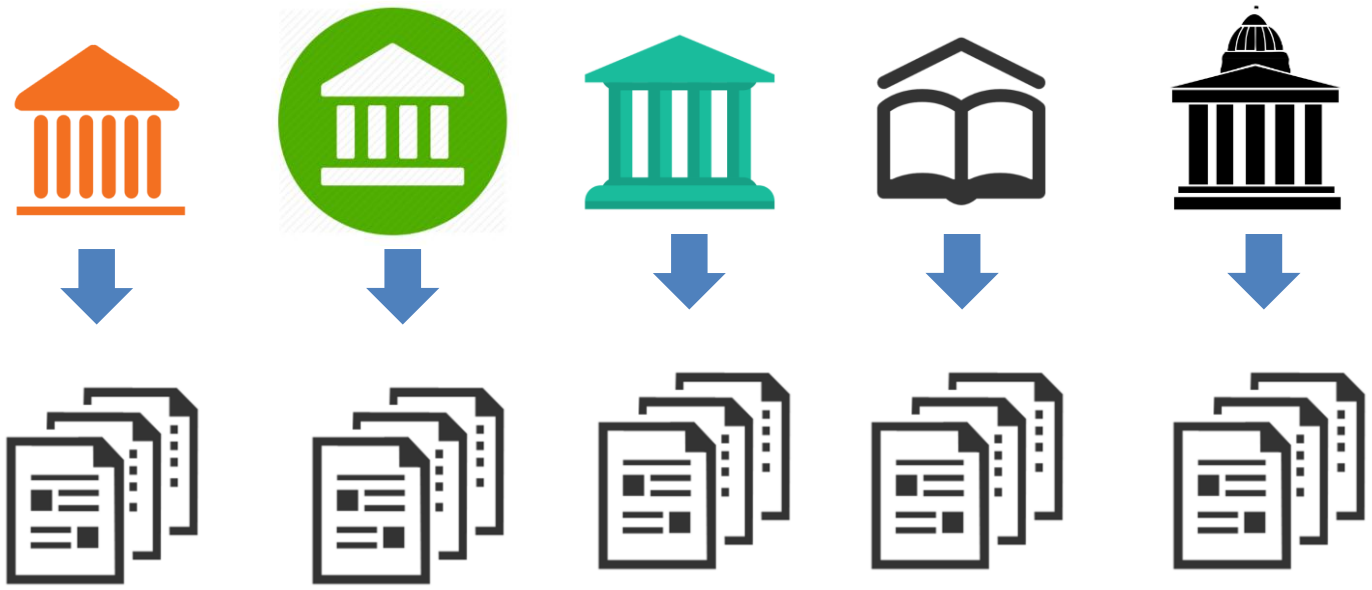
Citation-based index: Vinkler's π index



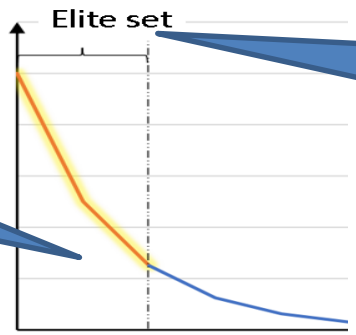
Classification Scheme



Using combined elite set



share-of-citation contribution: Who contributes the elites' citations?



share-of-size contribution: Who contributes the elites?

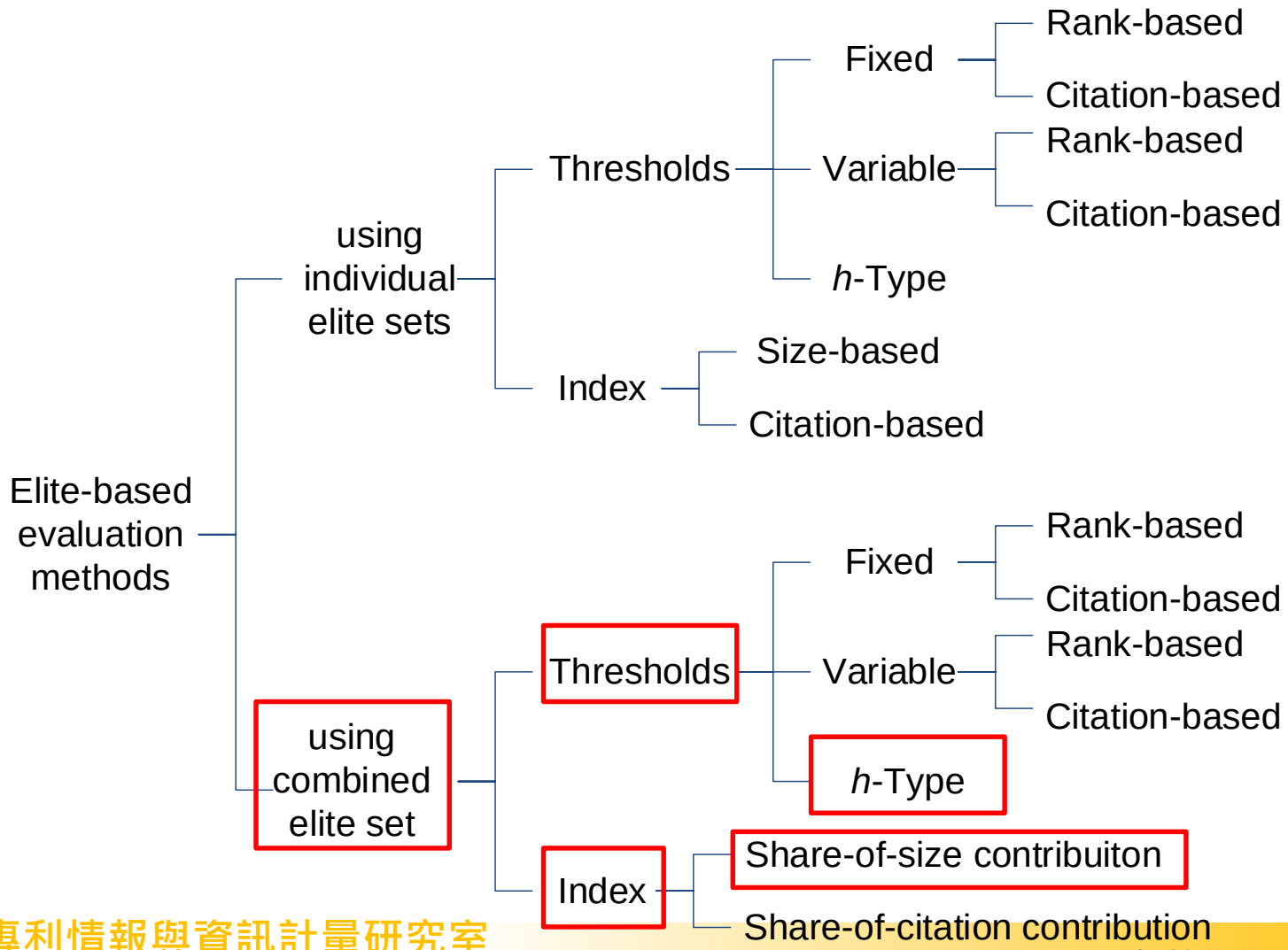
Size-based Contribution Example

- Say the combined elite set *contains* 100 publications and they receive total 1,000 citations
- for two entities *i* and *j*
  - if 50 of the 100 elites are from entity *i*
    - Entity *i* size-based contribution = 50%
  - if 10 are from the entity *j*
    - Entity *j* size-based contribution = 10%
  - It is suggested that entity *i* should be considered to have better performance than entity *j*.

Citation-based Contribution Example

- Say the combined elite set *contains* 100 publications and they receive total 1,000 citations
- for two entities *i* and *j*
  - if 50 of the elites are from by entity *i*
    - The 50 publications receive 300 citations
    - Entity *i* citation-based contribution = 30%
  - if 10 are from the entity *j*
    - The 10 publications receive 400 citations,
    - Entity *j* citation-based contribution = 40%
  - It is suggested that entity *j* should be considered to have better performance than entity *i*.

An Example



An Example

	Average	Individual field（Contribution by size share）					
		Agr	Cli	Eng	Lif	Phy	Soc
Harvard U.	9.85,1	5.49,5	13.69,1	4.03,7	16.31,1	2.97,8	16.60,1
UC - Berkeley	5.96,2	8.79,1	0.91,19	12.63,1	2.29,14	6.96,2	4.15,8
MIT	5.52,3	0.73,16	0.52,21	10.75,2	7.16,2	5.25,3	8.68,3
Stanford U.	4.70,4	3.30,9	3.52,8	8.33,3	4.73,6	3.42,7	4.91,6
UW - Seattle	3.87,5	3.30,9	5.22,4	4.84,5	3.81,7	4.91,5	1.13,14
Johns Hopkins U.	3.80,6	1.47,14	7.69,2	2.15,11	5.03,5	4.22,6	2.26,11
UC - Los Angeles	3.58,7	1.10,15	4.69,6	5.65,4	2.90,12	2.63,10	4.53,7
UC - San Diego	3.53,8	3.30,9	4.82,5	2.69,9	6.55,3	2.28,13	1.51,13
U. of Pennsylvania	3.36,9	0.00,18	3.39,9	0.54,17	3.05,11	2.97,8	10.19,2
UMich - Ann Arbor	3.25,10	1.47,14	4.56,7	2.15,11	2.59,13	3.42,7	5.28,5

Thank You