Publication performance evaluation for academic institutes by their contributions to the most visible publications across multiple fields

Max Kuan
Assistant Professor
National Taiwan University of Science and Technology

We report a method for evaluating

- the **publication performance** of
  - Academic papers, or
  - Patents, etc.
- an **entity**
  - academic/research institutes, or
  - companies, etc.
- **across a number of fields**
  - engineering, life science, etc.
- using **no. of publications (quantity)** and **no. of citations (quality)**

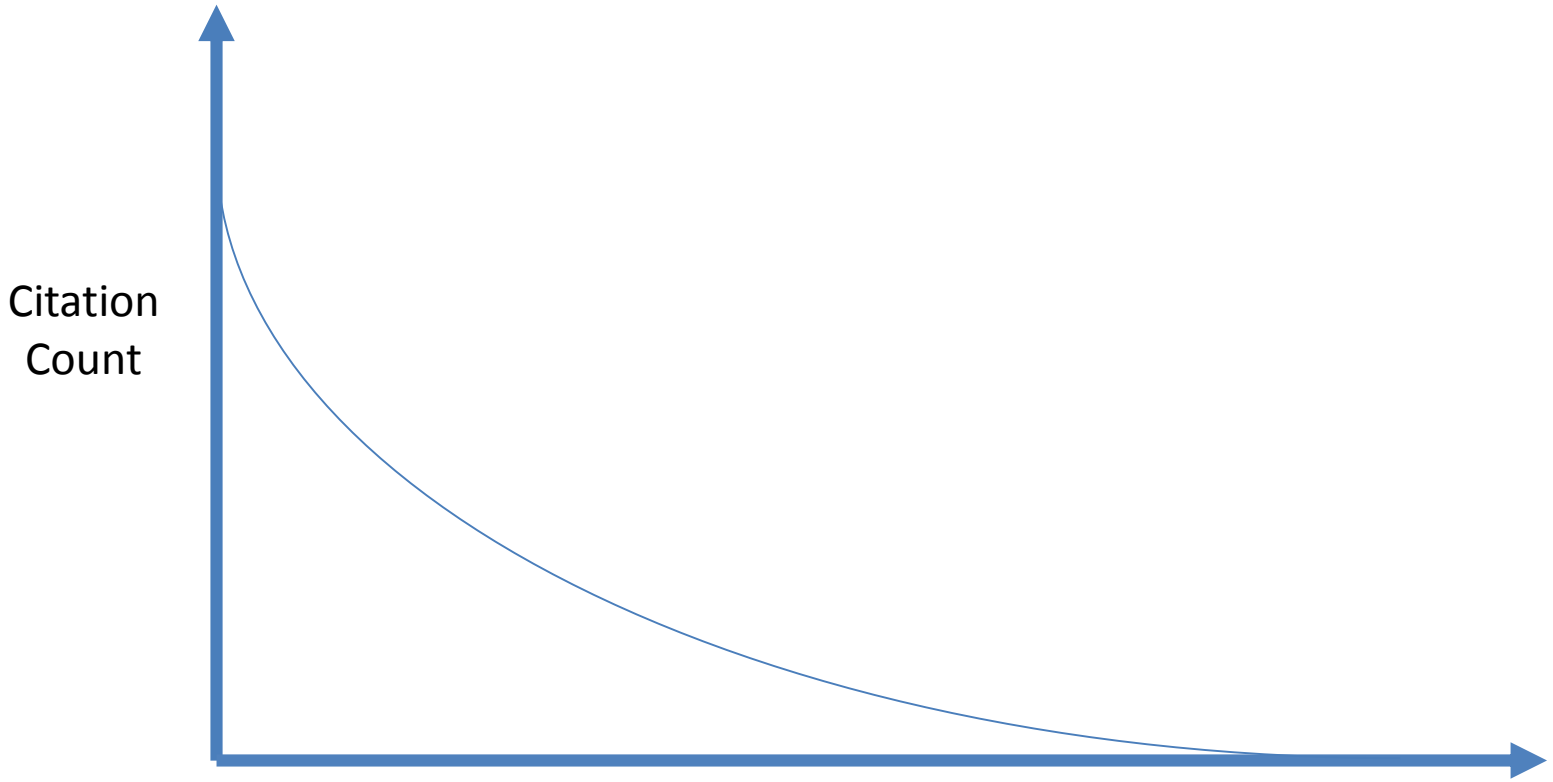
Concepts behind the method

- Based on a concept called ***elite set*** (proposed by Vinkler, P.)
- Using ***h*-index** to determine the elite set
  - Elite set is also similar to the so-called ***h*-core** (proposed by Rousseau, R.) in *h*-related research
- It is called ***Most Visible Publications* (MVP)** in this study.
- Publication performance is measured by an entity's ***contribution*** to a field's MVPs

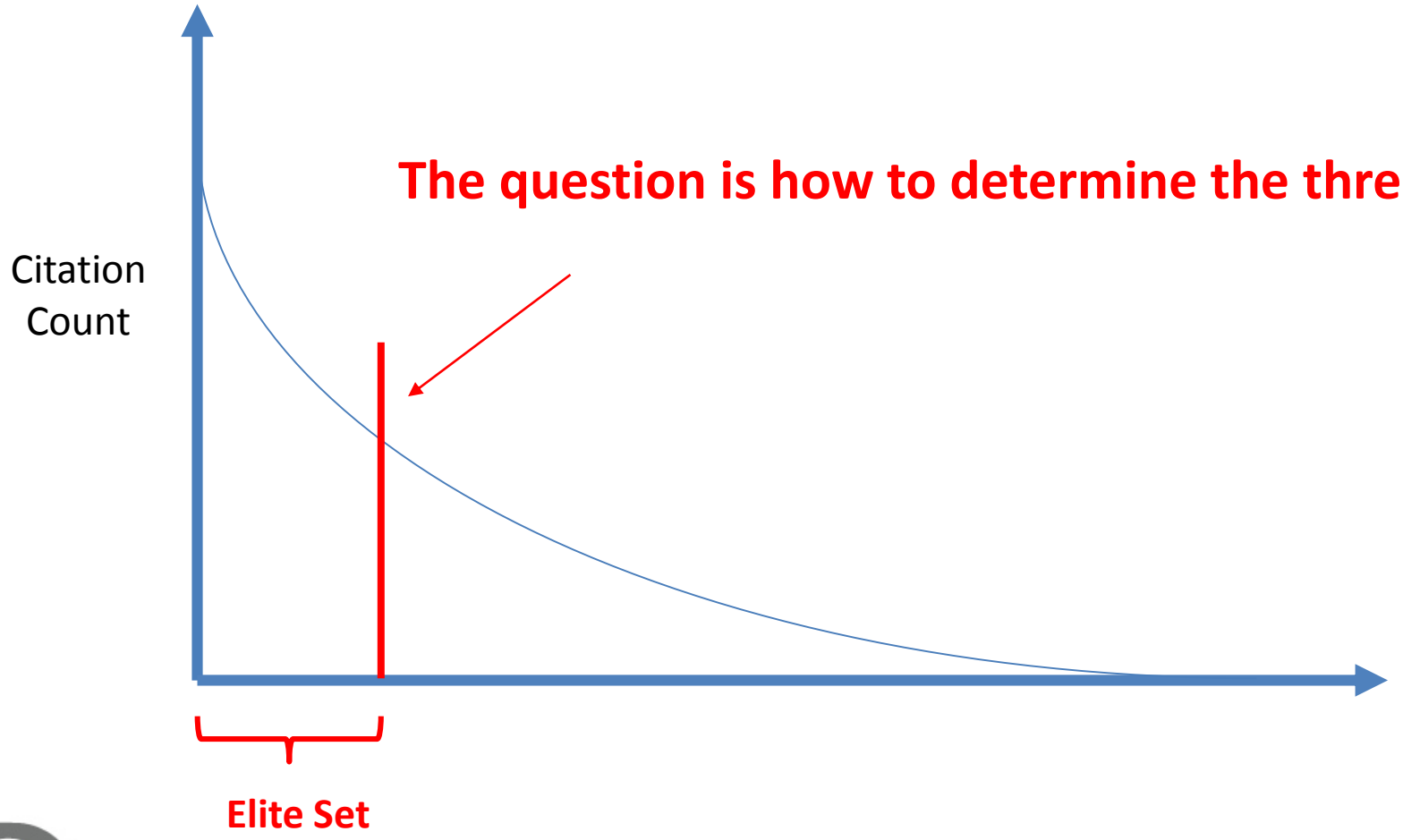
Elite set

- The **relatively most important publications** of a journal are jointly referred to as the journal's elite set (Vinkler, P.)
  - Use elite sets to determine the eminent journals
- This concept is not new
  - Highly cited paper: Papers in the upper first percentile with respect of their year of publication and subject area
  - Academic Ranking of World Universities (ARWU): no. of Nobel prize winners as a factor

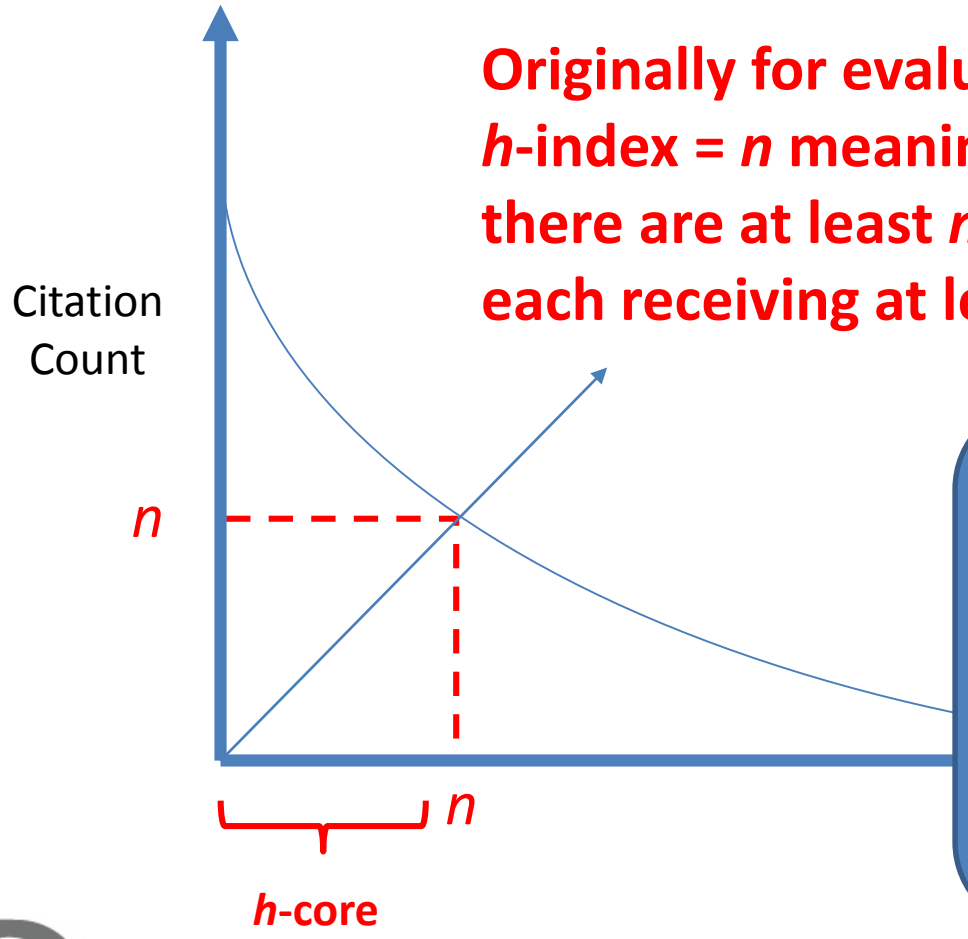
Geometric Interpretation of elite set



Geometric Interpretation of elite set



h-index



Originally for evaluating researcher performance
***h*-index = *n* meaning**
there are at least *n* publications
each receiving at least *n* citations

We choose *h*-index because its ready availability from online databases, its simplicity, and that it integrates both quantity and quality

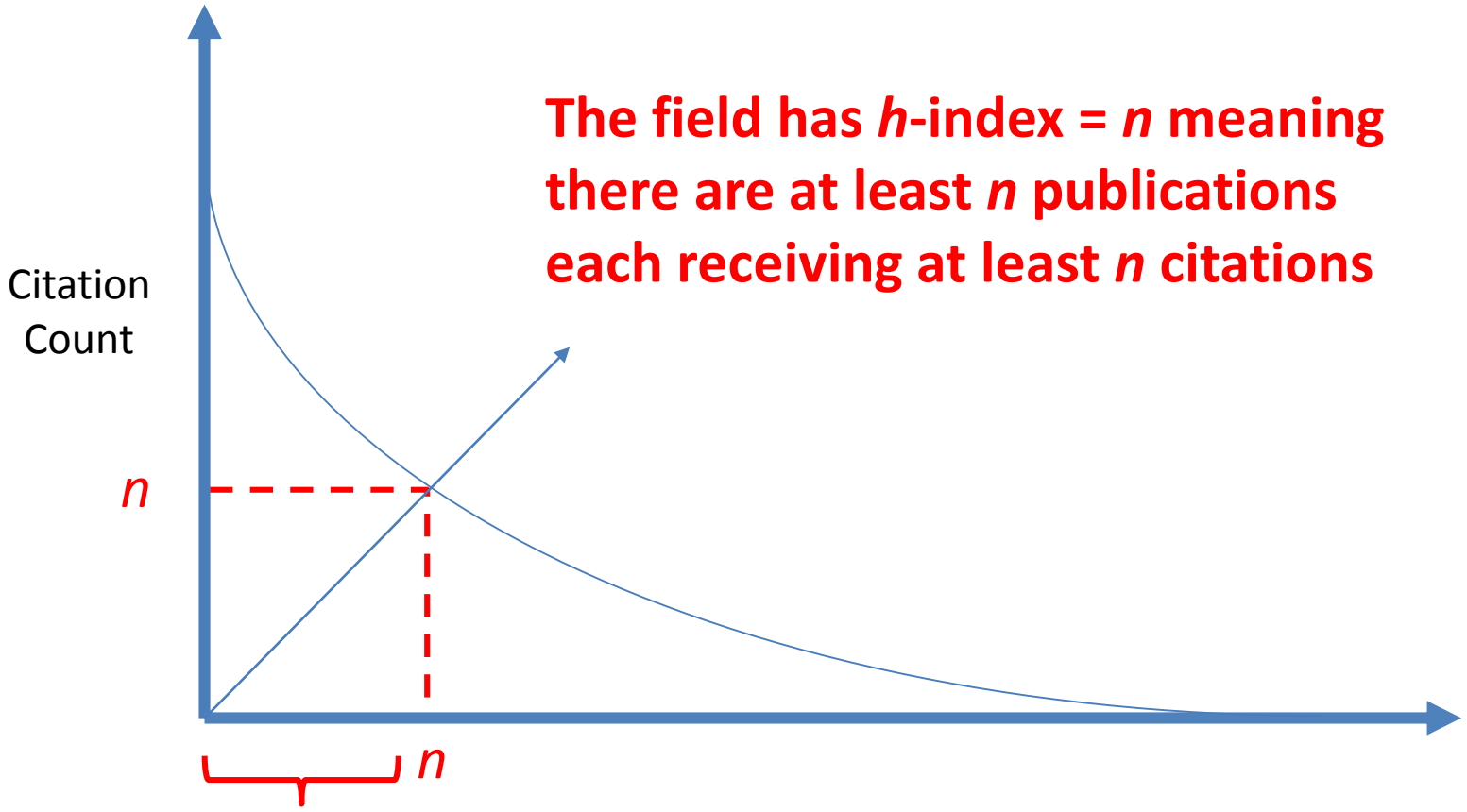
Most visible publications (MVPs)

- Our idea of elite set is different from *h*-core
  - **MVPs = The set of publications having citations ≥ *h*-index**
- For example, an entity has 5 publications with citations 5, 3, 3, 3, and 1
  - *h*-index=3, *h*-core = {5, 3, 3}
  - MVPs = {5, 3, 3, 3}

Our evaluation method

- For an entity's performance in a field
  - Determine the field's *h*-index
  - Determine the MVPs of the field using its *h*-index
  - For an entity
    - Determine its **contribution** to the field's MVPs
      - Two approaches
- For an entity's performance across a number of fields
  - Combine the entity's performance in each field

A field's *h*-index and MVPs



An entity's field performance

- An entity's performance in a field is measured by its ***contribution*** to the field's MVPs
- Two approaches
  - Contribution by no. of MVPs
  - Contribution by no. of citations of the MVPs

Contribution by no. of MVPs

- Say a field's MVPs *contains* 100 publications and they receive total 1,000 citations
- for two entities *i* and *j*
  - if 50 of the 100 MVPs are produced by entity *i*
    - Entity *i* contribution = 50%
  - if 10 are produced by the entity *j*
    - Entity *j* contribution = 10%
  - We therefore suggest that entity *i* should be considered to have better performance than entity institute *j*.

Contribution by no. of citations of the MVPs

- Say a field's MVPs *contains* 100 publications and they receive total 1,000 citations
- for two entitiy *i* and *j*
  - if 50 of the 100 MVPs are produced by entity *i*
    - The 50 publications receive 300 citations
    - Entity *i* contribution = 30%
  - if 10 are produced by the entity *j*
    - The 10 publications receive 400 citations,
    - Entity *j* contribution = 40%
  - We therefore suggest that entity *j* should be considered to have better performance than entity institute *i*.

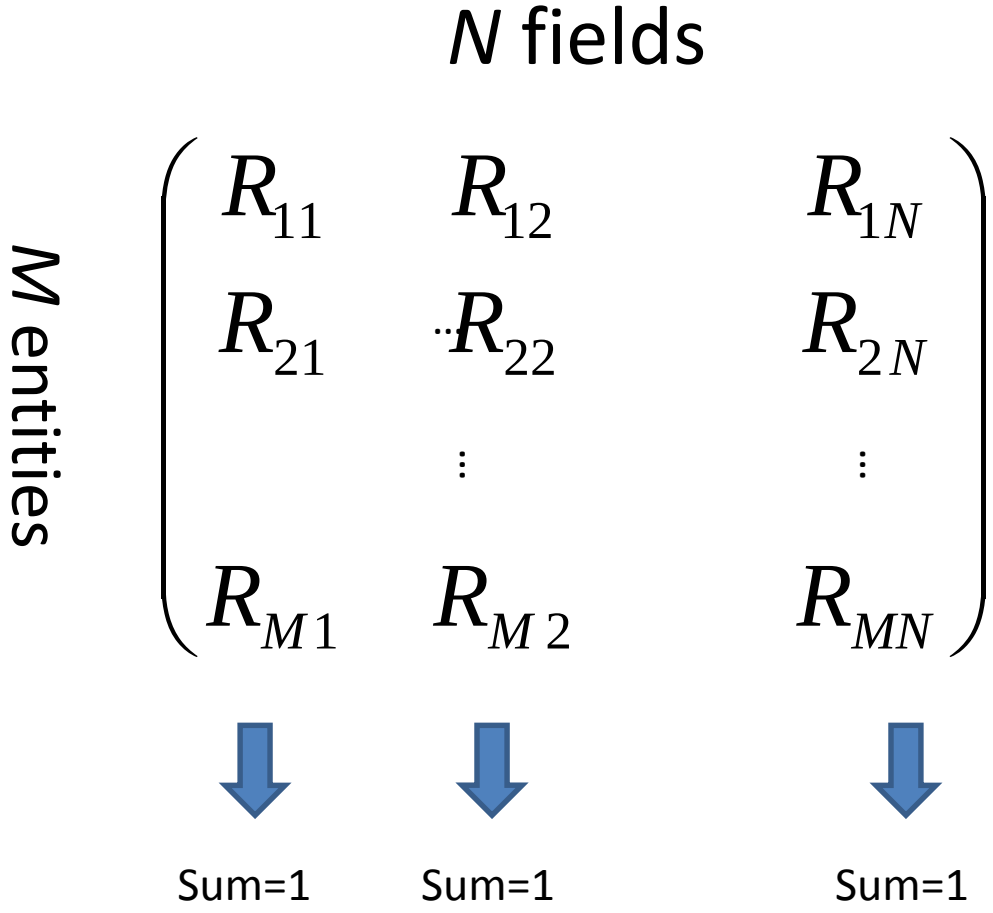
Contribution Matrix

N fields

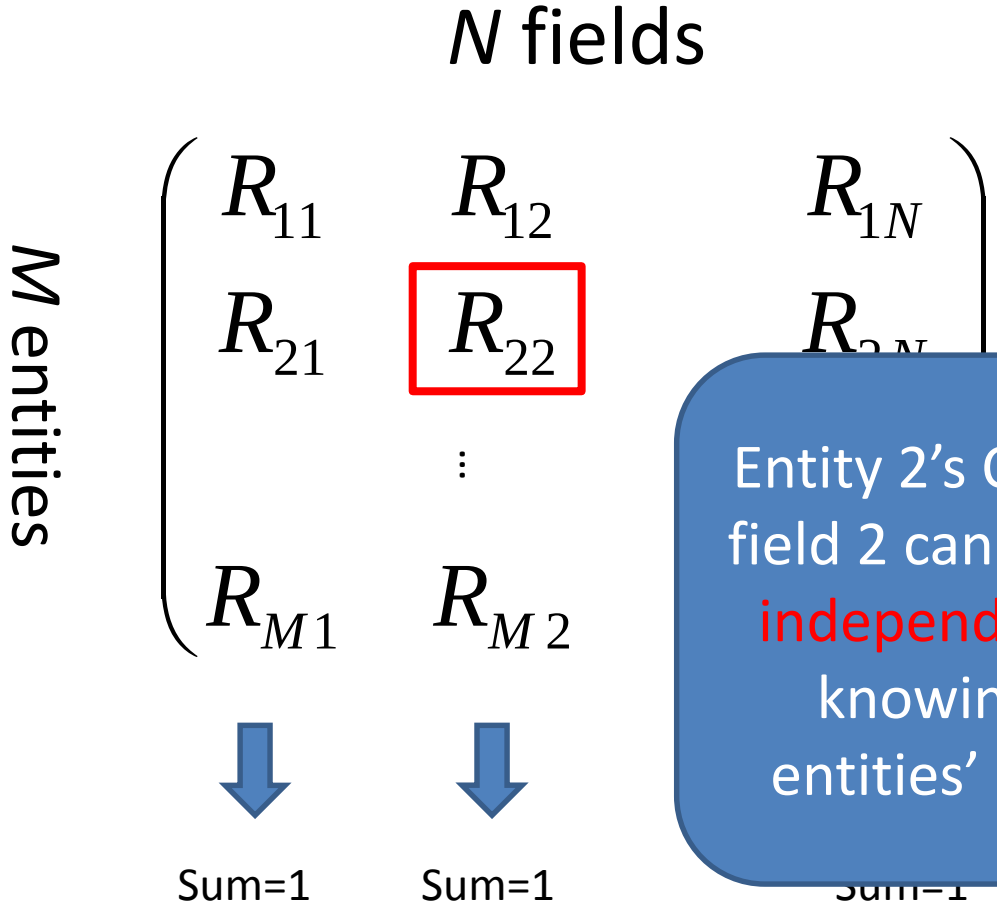
M entities

$$\begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & & \vdots \\ R_{M1} & R_{M2} & \cdots & R_{MN} \end{pmatrix}$$

Contribution Matrix

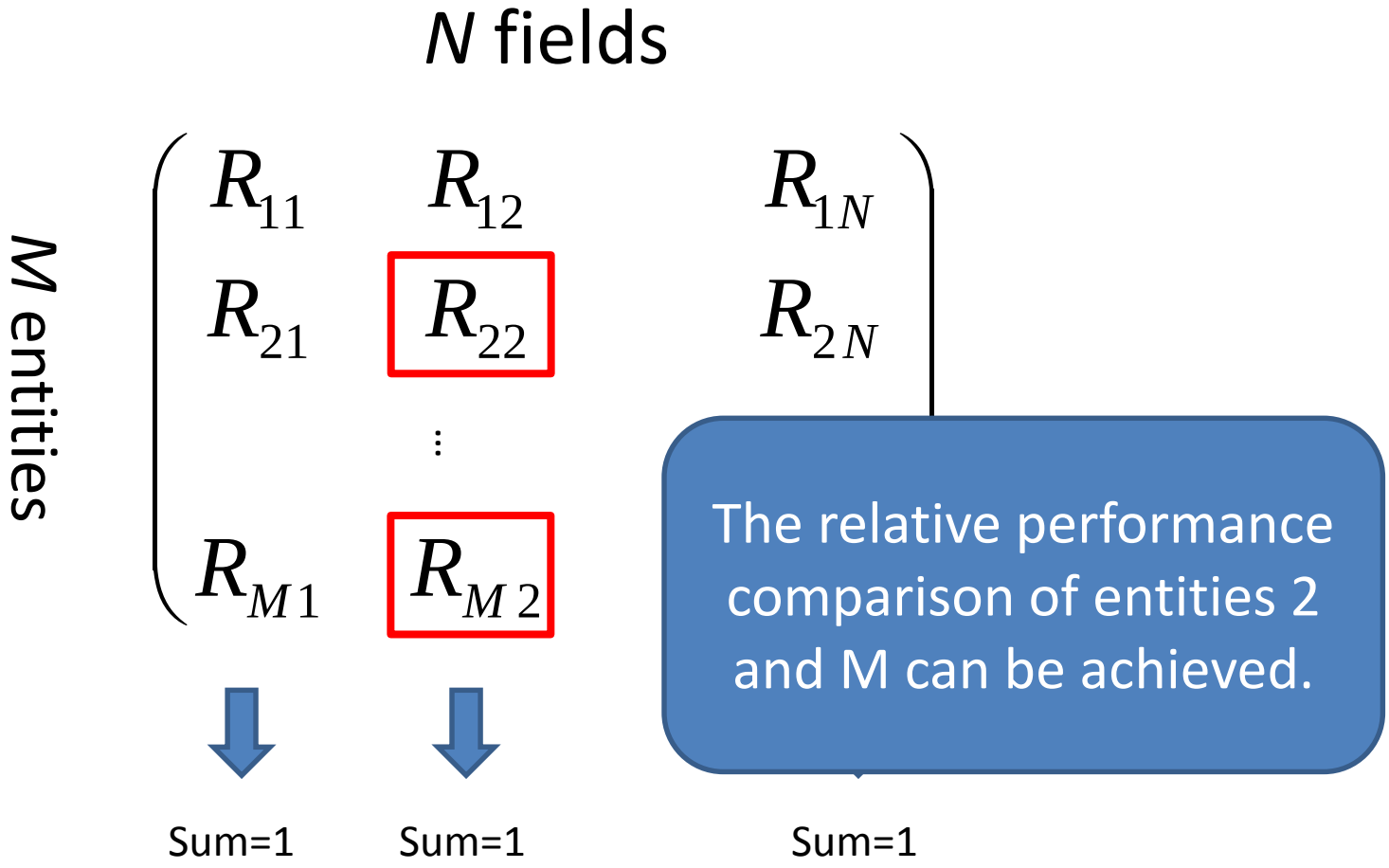


Usage of Contribution Matrix

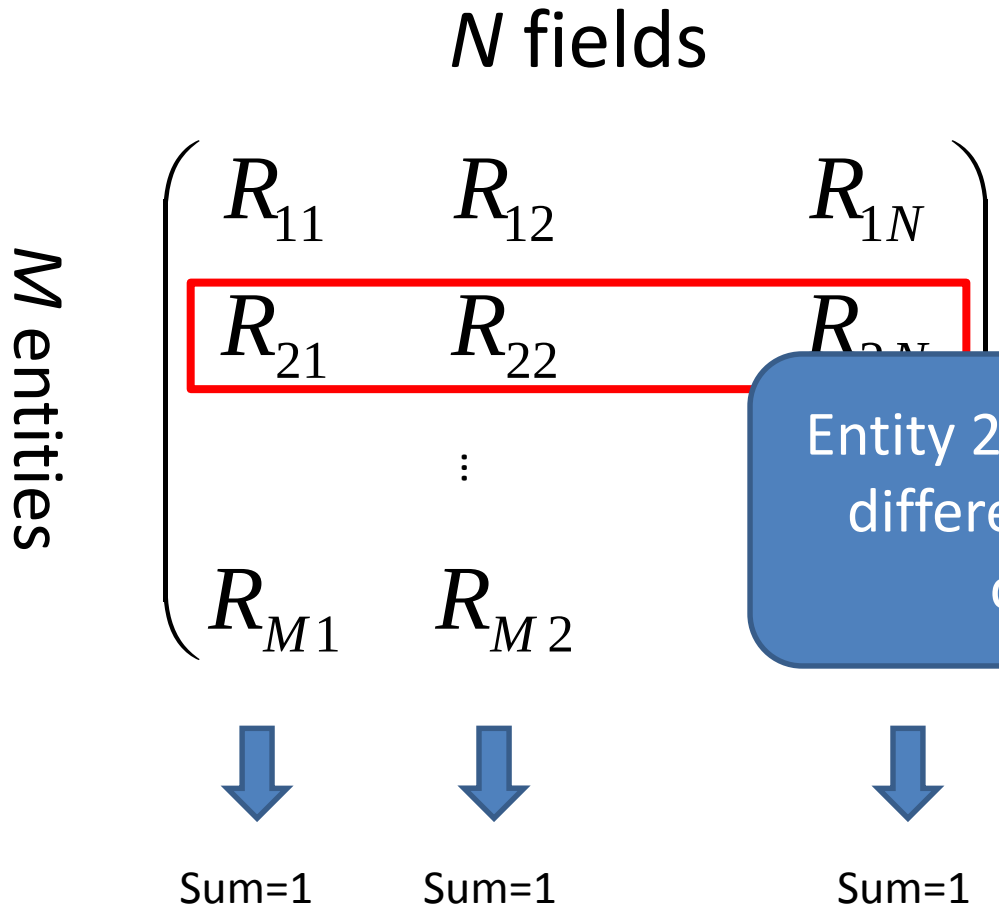


Entity 2's Contribution to field 2 can be interpreted independently without knowing the other entities' contributions

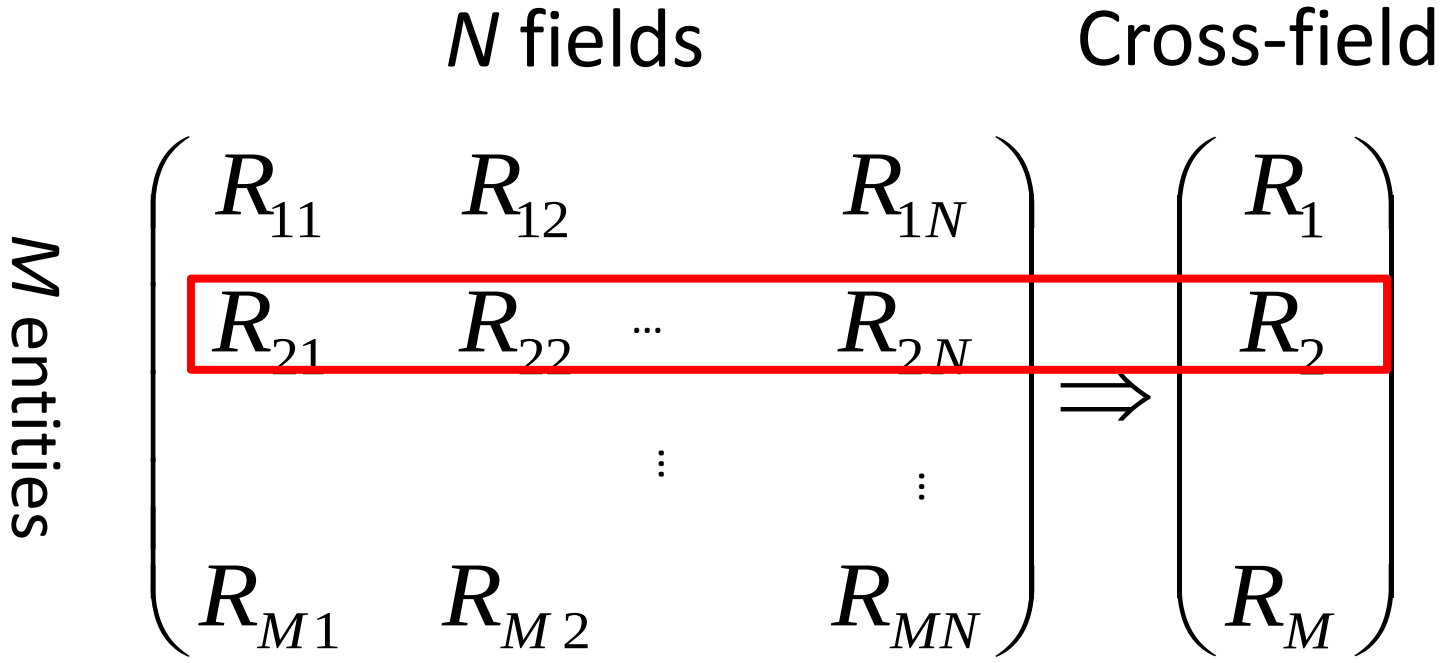
Usage of Contribution Matrix



Usage of Contribution Matrix



Cross-field performance



$$R_i = f(R_{i1}, R_{i2}, \ldots, R_{iN}) = \frac{\sum_{j=1}^{N} R_{ij}}{N}$$

Sum=1

A sample result

	Cross-field	Individual field（Contribution by publication 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

For the two types of contributions

- Contribution using publication share
  - **rather indiscriminating when it is applied to single-field evaluation**.
  - This shortcoming disappears when it is applied to cross-field evaluation.

For the two types of contributions

- Contribution using citation share
  - **is rather discriminating in single-field evaluation**
    - it may be biased by entities having a few extremely highly cited publications.
    - Such bias is lessened in cross-field evaluation.

For the two types of contributions

- For single-field evaluation
  - We suggest using contribution by citation share
- For cross-field evaluation
  - we suggest to use
    - Contribution by publcation share because it is simpler
    - The best mode would be to use both, and an entity is indeed has a better publication performance if it is considered as such by both types of contributions.

Life is not easy: two issues

- Multiple affiliations
  - In real life, it is common that a publication has multiple affiliations.
- Cannot differentiate entities having field contribution equal to zero, and entities having cross-field contribution equal to zero

Differentiate entities with zero contribution

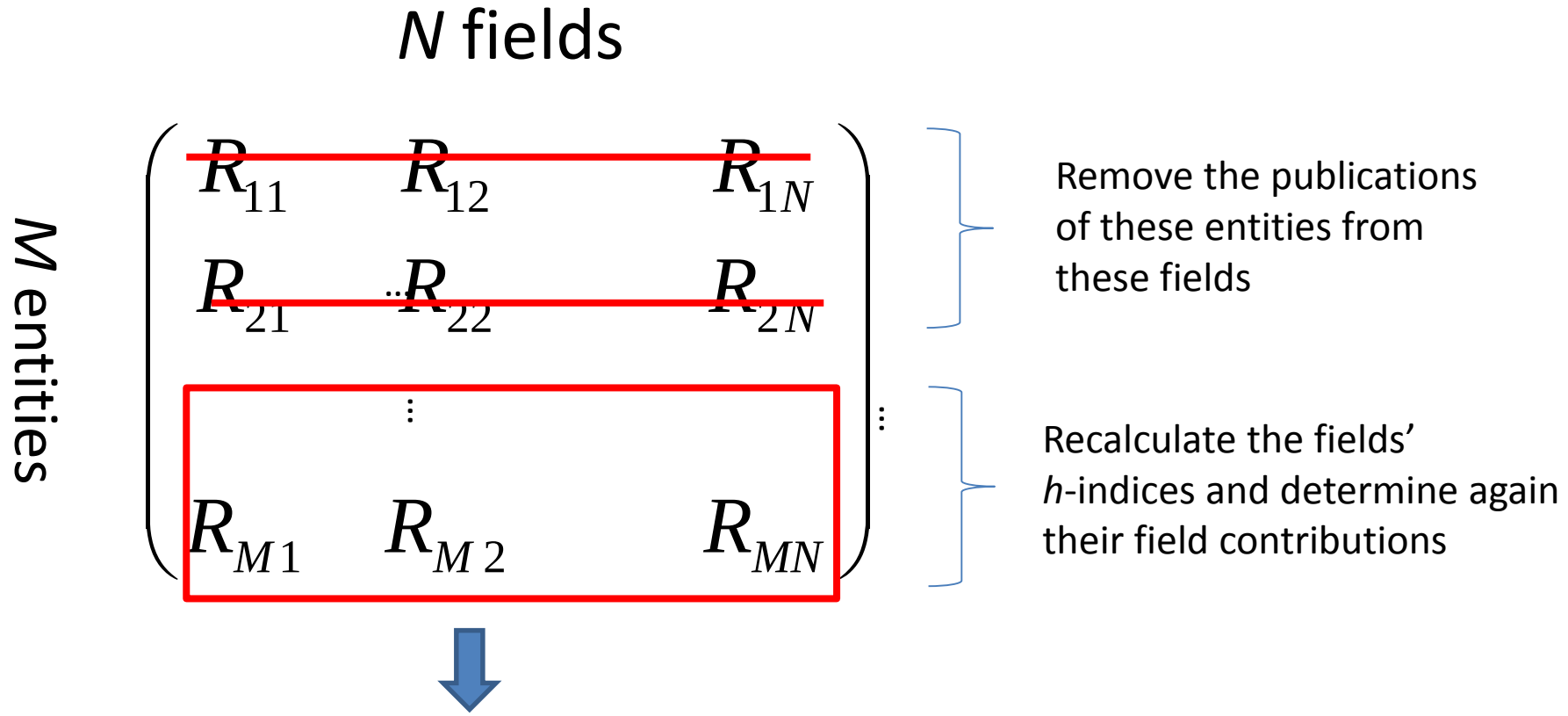
N fields

M entities

$$\begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{M1} & R_{M2} & \cdots & R_{MN} \\ \vdots & & & \end{pmatrix}$$

entities with zero contribution

Differentiate entities with zero contribution



Thank You