Evaluating scientific products by means of citation-based models

Gianna M. Del Corso
joint work with Dario A. Bini and Francesco Romani

Dipartimento di Informatica, Università di Pisa, Italy

The Mathematics of Ranking,
Brussels, October 15 2008

The Problem

- Number of scientific journals and papers is increasing at an almost exponential rate
- What to read? What to cite? Which journals subscribe? How to evaluate research?
- This burden affects researchers, funding agencies, university administrators, reviewers

Difficult to give an in-depth evaluation of the research
Use indirect indicators of quality

The Problem

- Number of scientific journals and papers is increasing at an almost exponential rate
- What to read? What to cite? Which journals subscribe? How to evaluate research?
- This burden affects researchers, funding agencies, university administrators, reviewers

Difficult to give an in-depth evaluation of the research
Use indirect indicators of quality

The Problem

- Number of scientific journals and papers is increasing at an almost exponential rate
- What to read? What to cite? Which journals subscribe? How to evaluate research?
- This burden affects researchers, funding agencies, university administrators, reviewers

Difficult to give an in-depth evaluation of the research
Use indirect indicators of quality

The Problem

- Number of scientific journals and papers is increasing at an almost exponential rate
- What to read? What to cite? Which journals subscribe? How to evaluate research?
- This burden affects researchers, funding agencies, university administrators, reviewers

Difficult to give an in-depth evaluation of the research

Use indirect indicators of quality

The Problem

- Number of scientific journals and papers is increasing at an almost exponential rate
- What to read? What to cite? Which journals subscribe? How to evaluate research?
- This burden affects researchers, funding agencies, university administrators, reviewers

Difficult to give an in-depth evaluation of the research
Use indirect indicators of quality

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

The Problem

Most of "automatic" methods rely on citation analysis

The evaluation of research using citation analysis has weaknesses...

- Is a citation always a trusting vote?
- Data source and coverage
- How do authors choose the papers to cite?

... but also some pros

- Peer review is not always practicable
- There are plausible assumptions underlying the use of citation analysis as a heuristic
- Simple and objective

Notation

We can represent the citation process as a graph and hence as a binary matrix

$$C_{ij} = 1 \text{ iff } p_i \text{ cites } p_j.$$

Assume that receiving a citation is always good!

Common metrics:

Different metrics for different purposes

- Ranking journals - Libraries, scholars for deciding where to publish, ...
- Ranking papers - What to read, what to cite, ...
- Ranking authors - distribution of grants, hiring people, ...

Common metrics:

Different metrics for different purposes

- Ranking journals - Libraries, scholars for deciding where to publish, ...
- Ranking papers - What to read, what to cite, ...
- Ranking authors - distribution of grants, hiring people, ...

Common metrics:

Different metrics for different purposes

- Ranking journals - Libraries, scholars for deciding where to publish, ...
- Ranking papers - What to read, what to cite, ...
- Ranking authors - distribution of grants, hiring people, ...

Common metrics:

Different metrics for different purposes

- Ranking journals - Libraries, scholars for deciding where to publish, ...
- Ranking papers - What to read, what to cite, ...
- Ranking authors - distribution of grants, hiring people, ...

Common metrics: Journals

Citation Statistics: Impact Factor, AMS MR, Citeseer,...

Pros: Easy to calculate, time aware, objective, etc.

Cons: Depend on the area.
In the same journal articles with different citation rates.
The ranking provided doesn't always agree with the widely accepted journal's reputation.

Common metrics: Journals

Citation Statistics: Impact Factor, AMS MR, Citeseer,...

Pros: Easy to calculate, time aware, objective, etc.

Cons: Depend on the area.
In the same journal articles with different citation rates.
The ranking provided doesn't always agree with the widely accepted journal's reputation.

Common metrics: Journals

Citation Statistics: Impact Factor, AMS MR, Citeseer,...

Pros: Easy to calculate, time aware, objective, etc.

Cons: Depend on the area.
In the same journal articles with different citation rates.
The ranking provided doesn't always agree with the widely accepted journal's reputation.

The ISI Impact factor

The ISI IF defines the status of a journal for a specific year.

It is defined as the mean number of citations that occurred in the considered year y to articles published in a given journal j during the previous two years.

$$IF(j,y) = \frac{\sum_k \left(c_{kj}(y, y-1) + c_{kj}(y, y-2)\right)}{n(j,y)},$$

where

$n(j,y)$ = number of papers published in journal j in years $y-2$,$y-1$

The ISI Impact factor

The ISI IF defines the status of a journal for a specific year.

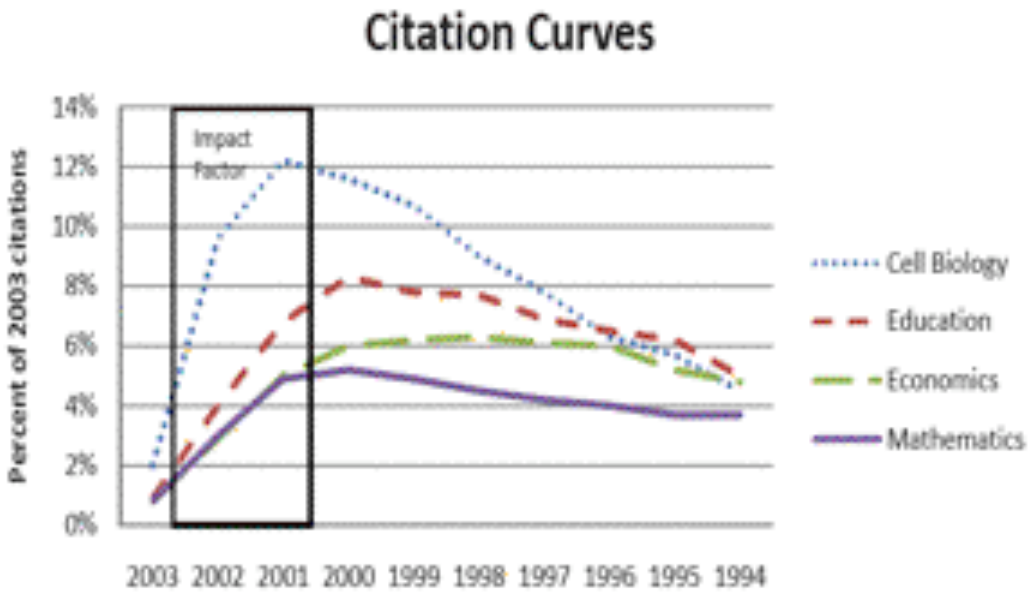
It is defined as the mean number of citations that occurred in the considered year y to articles published in a given journal j during the previous two years.

$$IF(j,y) = \frac{\sum_k \left(c_{kj}(y, y-1) + c_{kj}(y, y-2)\right)}{n(j,y)},$$

where

$n(j,y)$ = number of papers published in journal j in years $y-2$,$y-1$

The ISI Impact factor



[IMU, 2008]

The time window of two years is too small for many disciplines.

The ISI Impact factor

Journals Ranked by Impact: Mathematics

Rank	2006 Impact Factor	Impact 2002-06	Impact 1981-2006
1	J. Amer. Math. Soc. (2.55)	J. Amer. Math. Soc. (5.08)	Annals of Mathematics (24.82)
2	Annals of Mathematics (2.43)	Acta Mathematica (4.79)	Comm. Pure Appl. Math. (24.12)
3	Bull. Amer. Math. Soc. (2.39)	Bull. Amer. Math. Soc. (4.46)	Acta Mathematica (22.93)
4	Comm. Pure Appl. Math. (2.03)	Annals of Mathematics (4.28)	Inventiones Mathemat. (18.64)
5	Inventiones Mathemat. (1.66)	Fdns. Computat. Math. (3.93)	J. Different. Geometry (17.31)
6	J. Eur. Math. Soc. (1.49)	Comm. Pure Appl. Math. (3.74)	Bull. Amer. Math. Soc. (16.31)
7	Duke Mathematical J. (1.41)	Inventiones Mathemat. (3.50)	Ann. Sci. Ecole Norm. (12.99)
8	Publ. Mathematiques (1.35)	Ann. Sci. Ecole Norm. (2.59)	J. Amer. Math. Soc. (12.40)
9	Acta Mathematica (1.33)	Duke Mathematical J. (2.57)	SIAM J. Algebr. Discr. (12.35)
10	Geometry & Topology (1.27)	J. Different. Geometry (2.53)	Duke Mathematical J. (10.01)

[ScienceWatch.com, 2008]

Common metrics: Journals

PageRank-like techniques: Eigenfactor, SCImago, RedJasper, ...

Based on the idea that not all the citations are equal

Common metrics: Journals

PageRank-like techniques: Eigenfactor, SCImago, RedJasper, ...

Based on the idea that not all the citations are equal

Common metrics: Journals

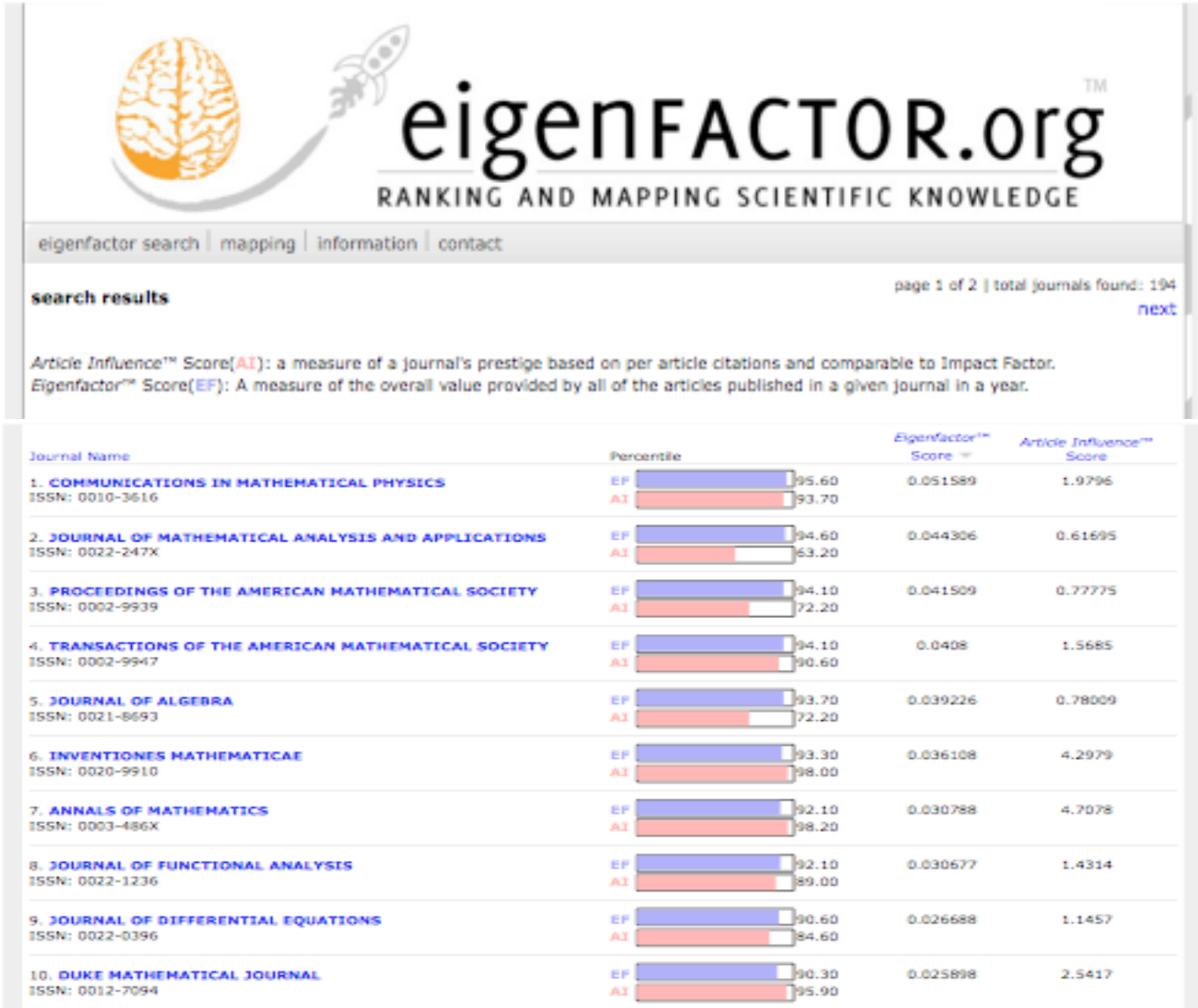
PageRank-like techniques: Eigenfactor, SCImago, RedJasper, ...

Based on the idea that not all the citations are equal

Common metrics: Journals

Pros: Quality is more important than quantity, metric of prestige, nice mathematical properties [see the assiomatic approach Palacios-Huerta,Volij, 04]

Common metrics: Eigenfactor



eigenFACTOR.org™

RANKING AND MAPPING SCIENTIFIC KNOWLEDGE

eigenfactor search | mapping | information | contact

search results

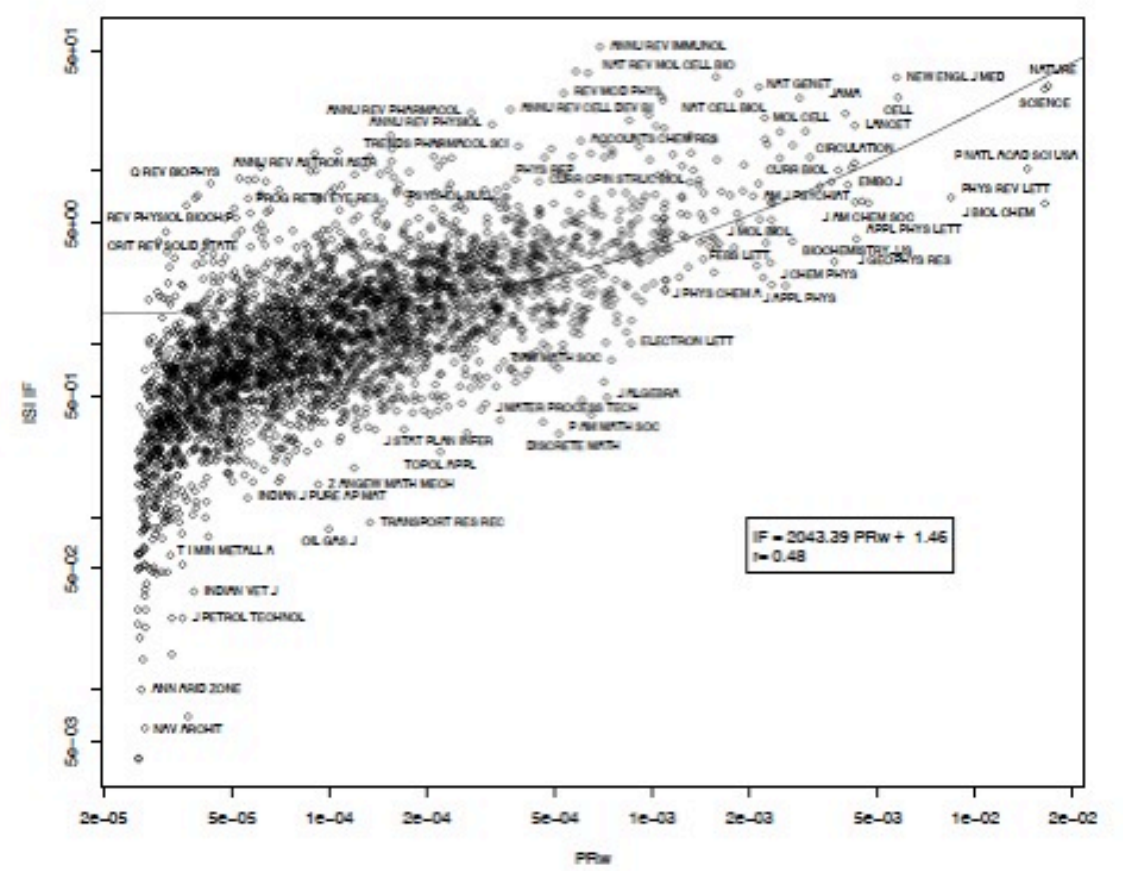
page 1 of 2 | total journals found: 194
next

Article Influence™ Score(AI): a measure of a journal's prestige based on per article citations and comparable to Impact Factor.
Eigenfactor™ Score(EF): A measure of the overall value provided by all of the articles published in a given journal in a year.

Journal Name	Percentile	*Eigenfactor*™ Score	*Article Influence*™ Score
1. COMMUNICATIONS IN MATHEMATICAL PHYSICS ISSN: 0010-3616	EF 95.60 AI 93.70	0.051589	1.9796
2. JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS ISSN: 0022-247X	EF 94.60 AI 63.20	0.044306	0.61695
3. PROCEEDINGS OF THE AMERICAN MATHEMATICAL SOCIETY ISSN: 0002-9939	EF 94.10 AI 72.20	0.041509	0.77775
4. TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY ISSN: 0002-9947	EF 94.10 AI 90.60	0.0408	1.5685
5. JOURNAL OF ALGEBRA ISSN: 0021-8693	EF 93.70 AI 72.20	0.039226	0.78009
6. INVENTIONES MATHEMATICAE ISSN: 0020-9910	EF 93.30 AI 98.00	0.036108	4.2979
7. ANNALS OF MATHEMATICS ISSN: 0003-486X	EF 92.10 AI 98.20	0.030788	4.7078
8. JOURNAL OF FUNCTIONAL ANALYSIS ISSN: 0022-1236	EF 92.10 AI 89.00	0.030677	1.4314
9. JOURNAL OF DIFFERENTIAL EQUATIONS ISSN: 0022-0396	EF 90.60 AI 84.60	0.026688	1.1457
10. DUKE MATHEMATICAL JOURNAL ISSN: 0012-7094	EF 90.30 AI 95.90	0.025898	2.5417

Common metrics: Journals

Hybrid techniques: Y-factor [Bollen,et al. 2006]



Common metrics: Journals

Hybrid technique: Y-factor

$$Y(j) = IF(j) \times PR_w(j)$$

Common metrics: Journals

Hybrid technique: Y-factor

$$Y(j) = IF(j) \times PR_w(j)$$

Common metrics: Papers

- Relevance of the journal where the paper is published
  - Not all the papers in a journal have the same quality
- Number of citations received
  - Citations gathered can be a way also to measure

Common metrics: Papers

- Relevance of the journal where the paper is published
  - Not all the papers in a journal have the same quality
- Number of citations received
  - Citation gathering can be a very slow process

Common metrics: Papers

- Relevance of the journal where the paper is published
  - Not all the papers in a journal have the same quality
- Number of citations received
  - Citation gathering can be a very slow process

Common metrics: Papers

- Relevance of the journal where the paper is published
  - Not all the papers in a journal have the same quality
- Number of citations received
  - Citation gathering can be a very slow process

Common metrics: Authors

- Top author if she publishes in "top" journals
- More accurate measures h-index, m-index, g-index, g_1-index

Common metrics: Authors

- Top author if she publishes in "top" journals
- More accurate measures h-index, m-index, g-index, g_1-index

Our Proposal [BDR ETNA 08]

- In the classical approach the ranking of journals is based on citations
- The ranking of papers and authors follows from the rank of the journals where the research is published

We proposed an integrated ranking of authors, journals, papers, areas, and institutions

Mutual reinforcement between papers, journals, authors

Our Proposal [BDR ETNA 08]

- In the classical approach the ranking of journals is based on citations
- The ranking of papers and authors follows from the rank of the journals where the research is published

We proposed an integrated ranking of authors, journals, papers, areas, and institutions

Mutual reinforcement between papers, journals, authors

Our Proposal [BDR ETNA 08]

- In the classical approach the ranking of journals is based on citations
- The ranking of papers and authors follows from the rank of the journals where the research is published

We proposed an integrated ranking of authors, journals, papers, areas, and institutions

Mutual reinforcement between papers, journals, authors

Our Proposal [BDR ETNA 08]

- In the classical approach the ranking of journals is based on citations
- The ranking of papers and authors follows from the rank of the journals where the research is published

We proposed an integrated ranking of authors, journals, papers, areas, and institutions

Mutual reinforcement between papers, journals, authors

Our Proposal

We have seen there are different uses of ranking

- Research evaluation by funding agencies
- Hiring in University or in a Industrial context
- Choosing individuals for a research team
- Many others ...

We try to design a tunable method to capture the different needs

Our Proposal

We have seen there are different uses of ranking

- Research evaluation by funding agencies
- Hiring in University or in a Industrial context
- Choosing individuals for a research team
- Many others ...

We try to design a tunable method to capture the different needs

Our Proposal

We have seen there are different uses of ranking

- Research evaluation by funding agencies
- Hiring in University or in a Industrial context
- Choosing individuals for a research team
- Many others ...

We try to design a tunable method to capture the different needs

Our Proposal

We have seen there are different uses of ranking

- Research evaluation by funding agencies
- Hiring in University or in a Industrial context
- Choosing individuals for a research team
- Many others ...

We try to design a tunable method to capture the different needs

Our Proposal

We have seen there are different uses of ranking

- Research evaluation by funding agencies
- Hiring in University or in a Industrial context
- Choosing individuals for a research team
- Many others ...

We try to design a tunable method to capture the different needs

Our Proposal

We have seen there are different uses of ranking

- Research evaluation by funding agencies
- Hiring in University or in a Industrial context
- Choosing individuals for a research team
- Many others ...

We try to design a tunable method to capture the different needs

General principles

- A paper is important if published in an important journal but also if cited by important papers and authored by important authors
- An author is important if she has important co-authors and has written important papers published in important journals
- A journal is important if collects citations from important journals, publishes important papers by important authors

	J	A	P
J	citation	publication	publication
A	publication	co-authorship	authorship
P	publication	authorship	citation

General principles

- A paper is important if published in an important journal but also if cited by important papers and authored by important authors
- An author is important if she has important co-authors and has written important papers published in important journals
- A journal is important if collects citations from important journals, publishes important papers by important authors

	J	A	P
J	citation	publication	publication
A	publication	co-authorship	authorship
P	publication	authorship	citation

General principles

- A paper is important if published in an important journal but also if cited by important papers and authored by important authors
- An author is important if she has important co-authors and has written important papers published in important journals
- A journal is important if collects citations from important journals, publishes important papers by important authors

	J	A	P
J	citation	publication	publication
A	publication	co-authorship	authorship
P	publication	authorship	citation

General principles

- A paper is important if published in an important journal but also if cited by important papers and authored by important authors
- An author is important if she has important co-authors and has written important papers published in important journals
- A journal is important if collects citations from important journals, publishes important papers by important authors

	J	A	P
J	citation	publication	publication
A	publication	co-authorship	authorship
P	publication	authorship	citation

General principles

- A paper is important if published in an important journal but also if cited by important papers and authored by important authors
- An author is important if she has important co-authors and has written important papers published in important journals
- A journal is important if collects citations from important journals, publishes important papers by important authors

	J	A	P
J	citation	publication	publication
A	publication	co-authorship	authorship
P	publication	authorship	citation

We described and analyze three models

- One-class model, made up by Papers only
- Two-class model, made up by Papers and Authors
- Three-class model, made up by Papers, Authors and Journals

One-class model

$C = (c_{i,j})$ citation matrix $n \times n$

$\mathbf{e} = (1, 1, \ldots, 1)^T$, $\mathbf{d} = C\mathbf{e}$, assume that $d_i \neq 0$, $i = 1 : n$

$$P = (p_{i,j}) := \mathrm{Diag}(\mathbf{d})^{-1} C$$

- P is row stochastic, i.e., $P\mathbf{e} = \mathbf{e}$
- If P is irreducible

by the Perron-Frobenius theorem it exists unique a vector $\pi > 0$ such that

$$\pi^T = \pi^T P, \quad \mathbf{e}^T \pi = 1, \quad \pi_j = \sum_{i=1}^{n} \pi_i \frac{c_{ij}}{d_i}$$

π_j is the rank of paper j

- P is row stochastic, i.e., $P\mathbf{e} = \mathbf{e}$
- If P is irreducible

by the Perron-Frobenius theorem it exists unique a vector $\pi > 0$ such that

$$\pi^T = \pi^T P, \qquad \mathbf{e}^T \pi = 1, \quad \pi_j = \sum_{i=1}^{n} \pi_i \frac{c_{i,j}}{d_i}$$

π_i is the rank of paper i

- P is row stochastic, i.e., $P\mathbf{e} = \mathbf{e}$
- If P is irreducible

  by the Perron-Frobenius theorem it exists unique a vector $\boldsymbol{\pi} > 0$ such that

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T P, \qquad \mathbf{e}^T \boldsymbol{\pi} = 1, \quad \pi_j = \sum_{i=1}^{n} \pi_i \frac{c_{i,j}}{d_i}$$

π_i **is the rank of paper** i

- P is row stochastic, i.e., $P\mathbf{e} = \mathbf{e}$
- If P is irreducible

  by the Perron-Frobenius theorem it exists unique a vector $\boldsymbol{\pi} > 0$ such that

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T P, \qquad \mathbf{e}^T \boldsymbol{\pi} = 1, \quad \pi_j = \sum_{i=1}^{n} \pi_i \frac{c_{i,j}}{d_i}$$

π_i **is the rank of paper** i

The principle contained in the equation

$$\pi_j = \sum_{i=1}^{n} \pi_i \frac{c_{i,j}}{d_i}$$

is that each paper equally distributes its importance among all the cited papers.

Problems

- Reduciblility of P
- Dangling nodes, i.e., papers which cite no papers,
- Even when P is irreducible it may be periodic

Problems

- Reduciblility of P
- Dangling nodes, i.e., papers which cite no papers,
- Even when P is irreducible it may be periodic

Problems

- Reduciblility of P
- Dangling nodes, i.e., papers which cite no papers,
- Even when P is irreducible it may be periodic

Example

If P is stochastic and reducible

$$P \approx \begin{bmatrix} P_{1,1} & P_{1,2} \\ O & P_{2,2} \end{bmatrix}$$

then $[\mathbf{0}^T, \boldsymbol{\pi_2}^T]$, where $\boldsymbol{\pi}_2^T$ is such that

$$\boldsymbol{\pi_2}^T P_{2,2} = \boldsymbol{\pi_2}^T.$$

Example

If P is stochastic and periodic we have more than an eigenvalue with modulo $\rho(P)$.

Problems of convergence of numerical algorithms for computing eigenvectors

Remedy:

We introduce a dummy paper which cites and is cited by all the existing papers except by itself

The dummy paper collects the importances of all the papers and redistributes them uniformly to all the subjects by creating no privileges

Remedy:

We introduce a dummy paper which cites and is cited by all the existing papers except by itself

The dummy paper collects the importances of all the papers and redistributes them uniformly to all the subjects by creating no privileges

Example

In matrix terms we add one row and one column to the matrix C made of all ones

Example:

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 1 & 1 & 1 & 0 \end{array}\right]$$

The new matrix is irreducible and aperiodic if $C \neq 0$ (there exist cycles of length 2 and 3)

Similar to Google's PageRank?

Same problems, but solved differently!

Theoretical issues: Model validation

- What happens if we add a new citation from a paper to another one?
- What happens if we add a new paper with citations?
- Does the rank of the newly cited paper increase more than that of the other papers?
- Can we formalize and quantify this property?

Theoretical issues: Model validation

- What happens if we add a new citation from a paper to another one?
- What happens if we add a new paper with citations?
- Does the rank of the newly cited paper increase more than that of the other papers?
- Can we formalize and quantify this property?

Theoretical issues: Model validation

- What happens if we add a new citation from a paper to another one?
- What happens if we add a new paper with citations?
- Does the rank of the newly cited paper increase more than that of the other papers?
- Can we formalize and quantify this property?

Theoretical issues: Model validation

- What happens if we add a new citation from a paper to another one?
- What happens if we add a new paper with citations?
- Does the rank of the newly cited paper increase more than that of the other papers?
- Can we formalize and quantify this property?

Theoretical results

Let C be an irreducible adjacency matrix

(r, s) such that $c_{r,s} = 0$

define $\widehat{C} = (\widehat{c}_{i,j})$ such that $\widehat{c}_{r,s} = 1$, $\widehat{c}_{i,j} = hc_{i,j}$ otherwise;

$$C = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & & & s & \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow r \\ \\ \\ \end{array} \qquad \widehat{C} = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & & & s & \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow r \\ \\ \\ \end{array}$$

Define $P = \mathrm{Diag}(C\mathbf{e})^{-1}C, \quad \widehat{P} = \mathrm{Diag}(\widehat{C}\mathbf{e})^{-1}\widehat{C}$

Theoretical results

Let C be an irreducible adjacency matrix

(r, s) such that $c_{r,s} = 0$

define $\widehat{C} = (\widehat{c}_{i,j})$ such that $\widehat{c}_{r,s} = 1$, $\widehat{c}_{i,j} = hc_{i,j}$ otherwise;

$$C = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & & & s & \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow r \\ \\ \\ \end{array} \qquad \widehat{C} = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} & & & s & \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow r \\ \\ \\ \end{array}$$

Define $P = \mathrm{Diag}(C\mathbf{e})^{-1}C, \quad \widehat{P} = \mathrm{Diag}(\widehat{C}\mathbf{e})^{-1}\widehat{C}$

Note:

The flipping of c_{rs}, will produce a change in all the nonzero entries of row r of $\widehat{P}$.

Note:

The flipping of c_{rs}, will produce a change in all the nonzero entries of row r of $\widehat{P}$.

$$P = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \overset{s}{\downarrow} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1/q & 0 & \cdot & 0 & 1/q & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow r \\ \\ \\ \\ \end{array}$$

Note:

The flipping of c_{rs}, will produce a change in all the nonzero entries of row r of $\widehat{P}$.

$$\widehat{P} = \begin{bmatrix} & & & s & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1/(q+1) & 0 & \cdot & 1/(q+1) & 1/(q+1) & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \leftarrow r$$

Theorem: [BDR, ETNA 08]

For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \frac{\widehat{\pi}_s}{\pi_s} \quad j = 1, \ldots, n, \qquad \sigma = q/(q+1) < 1$$

where q is the number of ones in the r-th row of C;

$$1 < \frac{\widehat{\pi}_s}{\pi_s} \quad \text{and} \quad \frac{\widehat{\pi}_j}{\pi_j} < \frac{\widehat{\pi}_s}{\pi_s}, \quad \text{if} \quad c_{r,j} \neq 0$$

Theorem: [BDR, ETNA 08]

For the left Perron vectors $\boldsymbol{\pi}$ and $\widehat{\boldsymbol{\pi}}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \frac{\widehat{\pi}_s}{\pi_s} \quad j = 1, \ldots, n, \qquad \sigma = q/(q+1) < 1$$

where q is the number of ones in the r-th row of C;

$$1 < \frac{\widehat{\pi}_s}{\pi_s} \quad \text{and} \quad \frac{\widehat{\pi}_j}{\pi_j} < \frac{\widehat{\pi}_s}{\pi_s}, \quad \text{if} \quad c_{r,j} \neq 0$$

Theorem: [BDR, ETNA 08]

For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \frac{\widehat{\pi}_s}{\pi_s} \quad j = 1, \ldots, n, \qquad \sigma = q/(q+1) < 1$$

where q is the number of ones in the r-th row of C;

$$1 < \frac{\widehat{\pi}_s}{\pi_s} \quad \text{and} \quad \frac{\widehat{\pi}_j}{\pi_j} < \frac{\widehat{\pi}_s}{\pi_s}, \quad \text{if} \quad c_{r,j} \neq 0$$

The paper which receives a new citation has an increase of rank larger than the increase of any other paper

Theorem: [BDR, ETNA 08]

For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \frac{\widehat{\pi}_s}{\pi_s} \quad j = 1, \ldots, n, \qquad \sigma = q/(q+1) < 1$$

where q is the number of ones in the r-th row of C;

$$1 < \frac{\widehat{\pi}_s}{\pi_s} \quad \text{and} \quad \frac{\widehat{\pi}_j}{\pi_j} < \frac{\widehat{\pi}_s}{\pi_s}, \quad \text{if} \quad c_{r,j} \neq 0$$

There are other perturbation results in the literature, but this is stronger!

Can we prove a stronger result?

If there are more than one citation we can prove that at least one of the newly cited papers has an increase of rank

More precisely...

Can we prove a stronger result?

If there are more than one citation we can prove that at least one of the newly cited papers has an increase of rank

More precisely...

Can we prove a stronger result?

If there are more than one citation we can prove that at least one of the newly cited papers has an increase of rank

More precisely...

Add references from paper r to papers $s_1, s_2, \ldots, s_k$. Let $\widehat{P}$ the stochastic matrix obtained.

Theorem: [BDR, 08]

For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \left(\prod_{i=1}^{k} \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \right)^{1/k} \leq \max_i \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \quad j = 1, \ldots, n,$$

for $\sigma = q/(q+k)$.

However...

There are even cases where one of newly cited paper has a decrease of rank

Add references from paper r to papers $s_1, s_2, \ldots, s_k$. Let $\widehat{P}$ the stochastic matrix obtained.

Theorem: [BDR, 08]

For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \left(\prod_{i=1}^{k} \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \right)^{1/k} \leq \max_i \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \quad j = 1, \ldots, n,$$

for $\sigma = q/(q+k)$.

However...

There are even cases where one of newly cited paper has a decrease of rank

Add references from paper r to papers $s_1, s_2, \ldots, s_k$. Let $\widehat{P}$ the stochastic matrix obtained.

Theorem: [BDR, 08]

For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \left(\prod_{i=1}^{k} \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \right)^{1/k} \leq \max_i \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \quad j = 1, \ldots, n,$$

for $\sigma = q/(q+k)$.

However...

There are even cases where one of newly cited paper has a decrease of rank

Add references from paper r to papers $s_1, s_2, \ldots, s_k$. Let $\widehat{P}$ the stochastic matrix obtained.

Theorem: [BDR, 08]

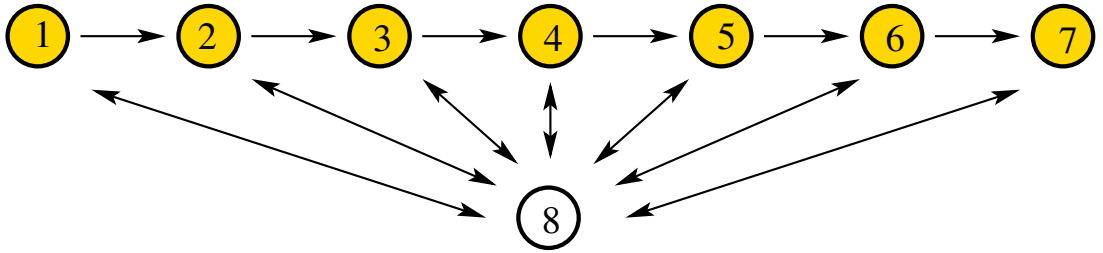
For the left Perron vectors π and $\widehat{\pi}$ of P and $\widehat{P}$ it holds

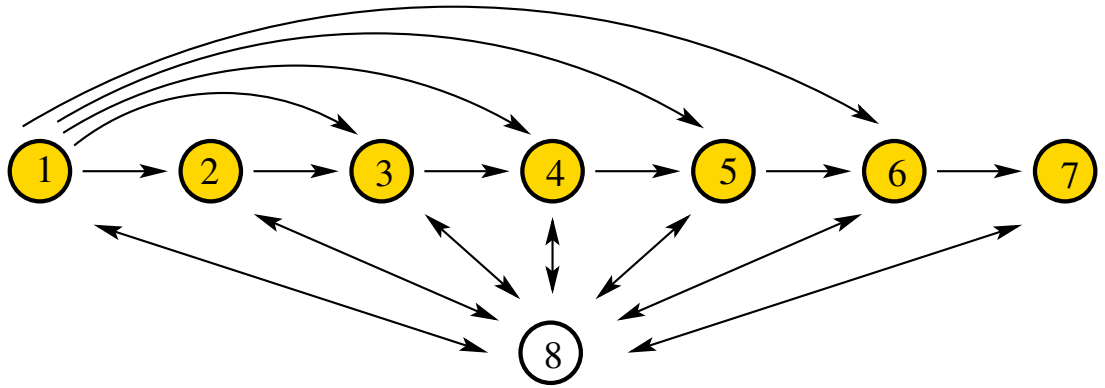
$$\sigma \frac{\widehat{\pi}_r}{\pi_r} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \left(\prod_{i=1}^{k} \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \right)^{1/k} \leq \max_i \frac{\widehat{\pi}_{s_i}}{\pi_{s_i}} \quad j = 1, \ldots, n,$$

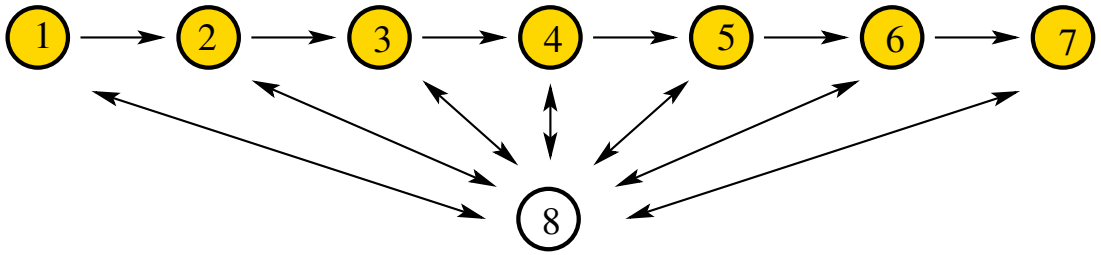
for $\sigma = q/(q+k)$.

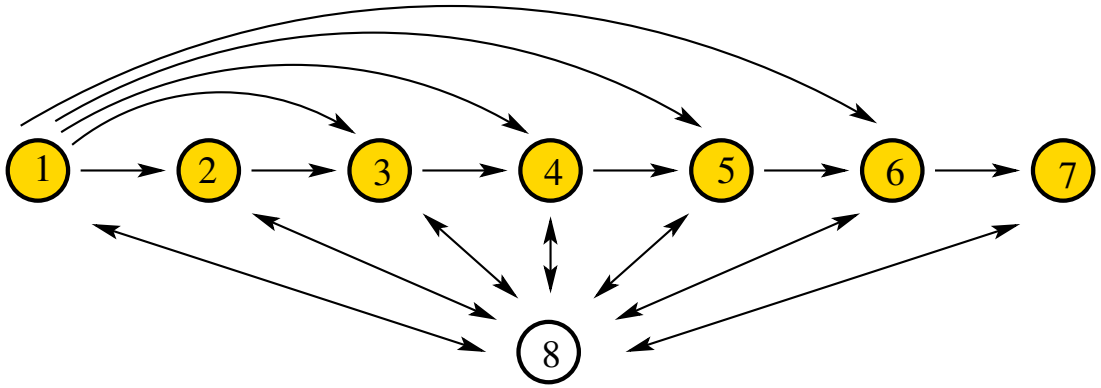
However...

There are even cases where one of newly cited paper has a decrease of rank


1
2
3
4
5
6
7
8

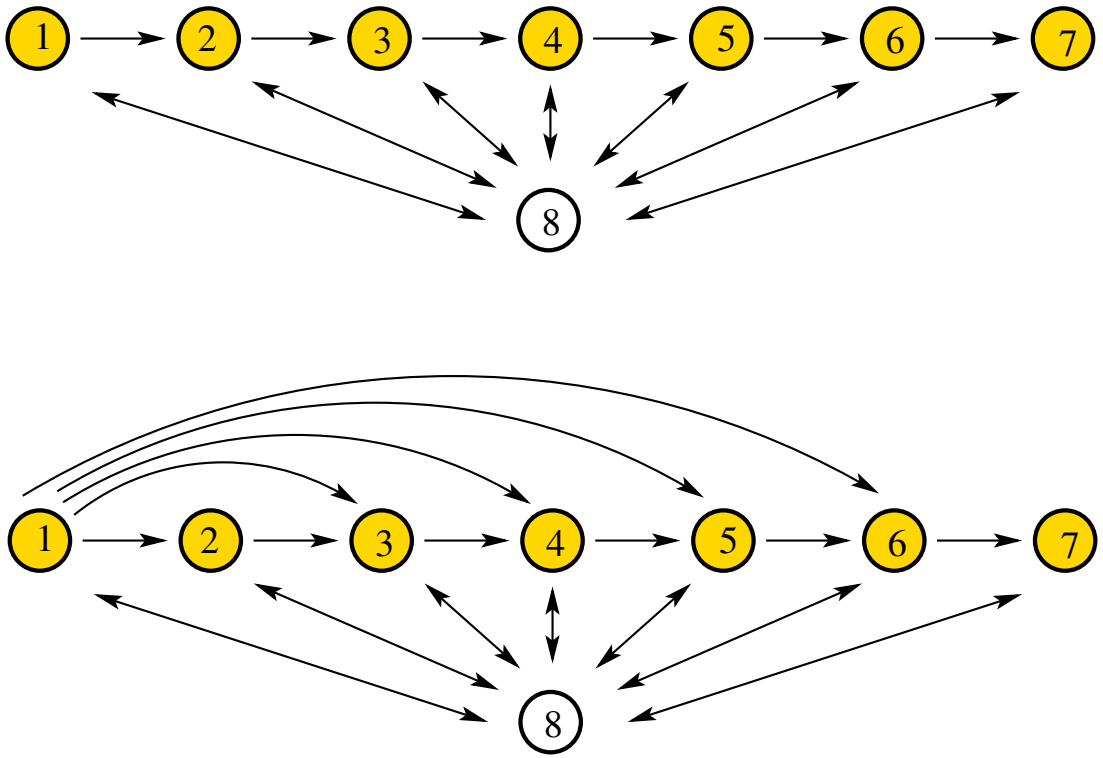

1
2
3
4
5
6
7
8





$$\pi = (0.05259, 0.07888, 0.09203, 0.09860, 0.10189, 0.10353, 0.10436)$$

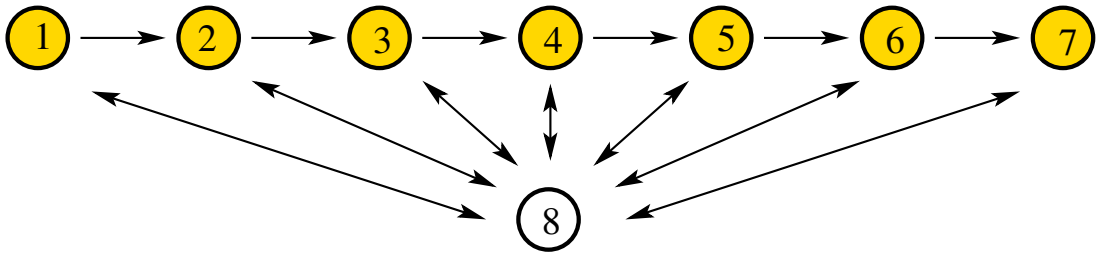
$$\widehat{\pi} = (0.05119, 0.05972, 0.08958, 0.10451, 0.11197, 0.11570, 0.10904)$$

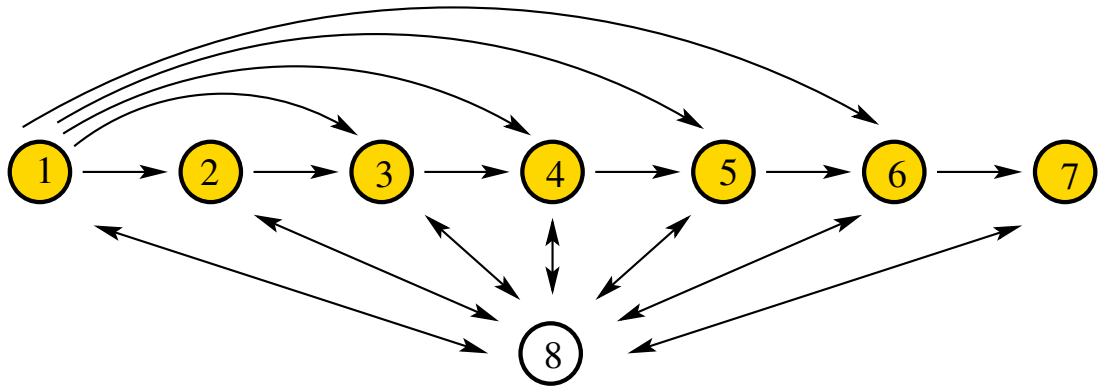


$$\boldsymbol{\pi} = (0.05259, 0.07888, 0.09203, 0.09860, 0.10189, 0.10353, 0.10436)$$

$$\widehat{\boldsymbol{\pi}} = (0.05119, 0.05972, 0.08958, 0.10451, 0.11197, 0.11570, 0.10904)$$

while their ratio is given by





$$\widehat{\pi}/\pi = (0.97334, 0.75704, 0.97334, 1.05986, 1.09893, 1.11754, 1.04487)$$

A similar result can be proved if we assume that only paper s receives a citation from papers $r_1, r_2, \ldots, r_k$.

Theorem: [BDR, 08]

For the left Perron vectors $\boldsymbol{\pi}$ and $\widehat{\boldsymbol{\pi}}$ of P and $\widehat{P}$ it holds

$$\min_{i=1,k} \sigma_i \frac{\widehat{\pi}_{r_i}}{\pi_{r_i}} \leq \left(\prod_{i=1}^{k} \sigma_i \frac{\widehat{\pi}_{r_i}}{\pi_{r_i}} \right)^{1/k} \leq \frac{\widehat{\pi}_j}{\pi_j} \leq \frac{\widehat{\pi}_s}{\pi_s} \quad j = 1, \ldots, n,$$

for $\sigma_i = q_i/(q_i + 1)$. Moreover,

$$\frac{\widehat{\pi}_j}{\pi_j} < \frac{\widehat{\pi}_s}{\pi_s}, \quad \text{if} \quad h_{r,j} \neq 0,$$

and $1 < \frac{\widehat{\pi}_s}{\pi_s}$.

Two-class model

We consider both papers and authors

The author-paper matrix

$$K(a, p) = 1 \quad \text{if author } a \text{ has written paper } p$$

The matrix $A = KK^T$ is the matrix of co-authors

Assume that the dummy paper is written by a dummy author

Two-class model

We consider both papers and authors
The author-paper matrix

$$K(a, p) = 1 \quad \text{if author } a \text{ has written paper } p$$

The matrix $A = KK^T$ is the matrix of co-authors

Assume that the dummy paper is written by a dummy author

Two-class model

We consider both papers and authors
The author-paper matrix

$$K(a, p) = 1 \quad \text{if author } a \text{ has written paper } p$$

The matrix $A = KK^T$ is the matrix of co-authors

Assume that the dummy paper is written by a dummy author

Two-class model

We consider both papers and authors
The author-paper matrix

$$K(a, p) = 1 \quad \text{if author } a \text{ has written paper } p$$

The matrix $A = KK^T$ is the matrix of co-authors

Assume that the dummy paper is written by a dummy author

The system is represented by the matrix

$$S = \begin{bmatrix} KK^T & K \\ K^T & C \end{bmatrix}$$

which captures the relationship of authorship and citation among the different subjects of the model.

The system is represented by the matrix

$$S = \begin{bmatrix} KK^T & K \\ K^T & C \end{bmatrix}$$

which captures the relationship of authorship and citation among the different subjects of the model.

Look at the nice structure!

Example

$$P1 \to P2 \to P3 \to P1$$

$$A1 \to P1, P2, P3; \quad A2 \to P2; \quad A3 \to P3$$

Then

$$C = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ \hline 1 & 1 & 1 & 0 \end{array}\right], \quad K = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array}\right]$$

$$A = KK^T = \begin{bmatrix} 3 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$S = \left[\begin{array}{cccc|cccc} 3 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array}\right]$$

We make the four blocks row-stochastic

We combine them into a larger stochastic matrix adding weights.

$$S = \left[\begin{array}{cccc|cccc} 3 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array}\right]$$

We make the four blocks row-stochastic

We combine them into a larger stochastic matrix adding weights.

$$S = \left[\begin{array}{cccc|cccc} 3 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array}\right]$$

We make the four blocks row-stochastic

We combine them into a larger stochastic matrix adding weights.

More precisely:

$$P = \begin{bmatrix} \gamma_{1,1} A_{1,1} & \gamma_{1,2} A_{1,2} \\ \gamma_{2,1} A_{2,1} & \gamma_{2,2} A_{2,2} \end{bmatrix} \qquad A_{i,j} \text{ is row stochastic}$$

The rank vector is the Perron vector of P

$$\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$$

Let $\pi = [\pi_{\text{authors}}, \pi_{\text{papers}}]$, one has

$$P = \begin{bmatrix} \gamma_{1,1}A_{1,1} & \gamma_{1,2}A_{1,2} \\ \gamma_{2,1}A_{2,1} & \gamma_{2,2}A_{2,2} \end{bmatrix}$$

$$\begin{aligned} \pi_{\text{authors}} &= \gamma_{1,1}\pi_{\text{authors}}A_{1,1} + \gamma_{2,1}\pi_{\text{papers}}A_{2,1} \\ \pi_{\text{papers}} &= \gamma_{1,2}\pi_{\text{authors}}A_{1,2} + \gamma_{2,2}\pi_{\text{papers}}A_{2,2} \end{aligned}$$

The weights γ_{ij} can be used to tune how much of their importance each player transfers to the subject.

Let $\pi = [\pi_{\text{authors}}, \pi_{\text{papers}}]$, one has

$$P = \begin{bmatrix} \gamma_{1,1}A_{1,1} & \gamma_{1,2}A_{1,2} \\ \gamma_{2,1}A_{2,1} & \gamma_{2,2}A_{2,2} \end{bmatrix}$$

$$\begin{aligned} \pi_{\text{authors}} &= \gamma_{1,1}\pi_{\text{authors}}A_{1,1} + \gamma_{2,1}\pi_{\text{papers}}A_{2,1} \\ \pi_{\text{papers}} &= \gamma_{1,2}\pi_{\text{authors}}A_{1,2} + \gamma_{2,2}\pi_{\text{papers}}A_{2,2} \end{aligned}$$

The weights γ_{ij} can be used to tune how much of their importance each player transfers to the subject.

Let $\pi = [\pi_{\text{authors}}, \pi_{\text{papers}}]$, one has

$$P = \begin{bmatrix} \gamma_{1,1}A_{1,1} & \gamma_{1,2}A_{1,2} \\ \gamma_{2,1}A_{2,1} & \gamma_{2,2}A_{2,2} \end{bmatrix}$$

$$\begin{aligned} \pi_{\text{authors}} &= \gamma_{1,1}\pi_{\text{authors}}A_{1,1} + \gamma_{2,1}\pi_{\text{papers}}A_{2,1} \\ \pi_{\text{papers}} &= \gamma_{1,2}\pi_{\text{authors}}A_{1,2} + \gamma_{2,2}\pi_{\text{papers}}A_{2,2} \end{aligned}$$

The weights γ_{ij} can be used to tune how much of their importance each player transfers to the subject.

Let $\pi = [\pi_{\text{authors}}, \pi_{\text{papers}}]$, one has

$$P = \begin{bmatrix} \gamma_{1,1}A_{1,1} & \gamma_{1,2}A_{1,2} \\ \gamma_{2,1}A_{2,1} & \gamma_{2,2}A_{2,2} \end{bmatrix}$$

$$\begin{aligned} \pi_{\text{authors}} &= \gamma_{1,1}\pi_{\text{authors}}A_{1,1} + \gamma_{2,1}\pi_{\text{papers}}A_{2,1} \\ \pi_{\text{papers}} &= \gamma_{1,2}\pi_{\text{authors}}A_{1,2} + \gamma_{2,2}\pi_{\text{papers}}A_{2,2} \end{aligned}$$

The weights γ_{ij} can be used to tune how much of their importance each player transfers to the subject.

Let $\boldsymbol{\pi} = [\boldsymbol{\pi}_{\text{authors}}, \boldsymbol{\pi}_{\text{papers}}]$, one has

$$P = \begin{bmatrix} \gamma_{1,1} A_{1,1} & \gamma_{1,2} A_{1,2} \\ \gamma_{2,1} A_{2,1} & \gamma_{2,2} A_{2,2} \end{bmatrix}$$

$$\begin{aligned} \boldsymbol{\pi}_{\text{authors}} &= \gamma_{1,1} \boldsymbol{\pi}_{\text{authors}} A_{1,1} + \gamma_{2,1} \boldsymbol{\pi}_{\text{papers}} A_{2,1} \\ \boldsymbol{\pi}_{\text{papers}} &= \gamma_{1,2} \boldsymbol{\pi}_{\text{authors}} A_{1,2} + \gamma_{2,2} \boldsymbol{\pi}_{\text{papers}} A_{2,2} \end{aligned}$$

The weights γ_{ij} can be used to tune how much of their importance each player transfers to the subject.

The impact of co-authorship should be marginal respect to that of papers $\rightarrow$ low $\gamma_{1,1}$

Let $\pi = [\pi_{\text{authors}}, \pi_{\text{papers}}]$, one has

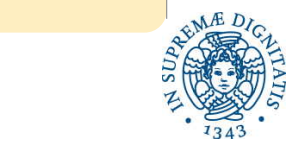
$$P = \begin{bmatrix} \gamma_{1,1} A_{1,1} & \gamma_{1,2} A_{1,2} \\ \gamma_{2,1} A_{2,1} & \gamma_{2,2} A_{2,2} \end{bmatrix}$$

$$\begin{aligned} \pi_{\text{authors}} &= \gamma_{1,1} \pi_{\text{authors}} A_{1,1} + \gamma_{2,1} \pi_{\text{papers}} A_{2,1} \\ \pi_{\text{papers}} &= \gamma_{1,2} \pi_{\text{authors}} A_{1,2} + \gamma_{2,2} \pi_{\text{papers}} A_{2,2} \end{aligned}$$

The weights γ_{ij} can be used to tune how much of their importance each player transfers to the subject.

Remark:

The previous perturbation theorems still hold in this model for perturbations concerning C only



Three-class model

We introduce in our model the class of journals

We introduce the journal-paper matrix

$$F(j, p) = 1 \quad \text{if journal } j \text{ publishes paper } p$$

- $G = (g_{i,j})$ such that $g_{i,j} = r$ if author i has published r papers in journal j
- $E = (e_{i,j})$ such that $e_{i,j} = k$ if k is the number of citations from papers published in journal i to papers published in journal j

Three-class model

We introduce in our model the class of journals

We introduce the journal-paper matrix

$$F(j, p) = 1 \quad \text{if journal } j \text{ publishes paper } p$$

- $G = (g_{i,j})$ such that $g_{i,j} = r$ if author i has published r papers in journal j
- $E = (e_{i,j})$ such that $e_{i,j} = k$ if k is the number of citations from papers published in journal i to papers published in journal j

Three-class model

We introduce in our model the class of journals

We introduce the journal-paper matrix

$$F(j, p) = 1 \quad \text{if journal } j \text{ publishes paper } p$$

- $G = (g_{i,j})$ such that $g_{i,j} = r$ if author i has published r papers in journal j
- $E = (e_{i,j})$ such that $e_{i,j} = k$ if k is the number of citations from papers published in journal i to papers published in journal j

It holds $E = FCF^T$, $G = FK^T$, the complete adjacency matrix is

$$S = \begin{bmatrix} FCF^T & FK^T & F \\ KF^T & KK^T & K \\ F^T & K^T & C \end{bmatrix}$$

It holds $E = FCF^T$, $G = FK^T$, the complete adjacency matrix is

$$S = \begin{bmatrix} FCF^T & FK^T & F \\ KF^T & KK^T & K \\ F^T & K^T & C \end{bmatrix}$$

The matrices in the first column contribute to the ranking of journals

It holds $E = FCF^T$, $G = FK^T$, the complete adjacency matrix is

$$S = \begin{bmatrix} FCF^T & FK^T & F \\ KF^T & KK^T & K \\ F^T & K^T & C \end{bmatrix}$$

The matrices in the second column contribute to the ranking of authors

It holds $E = FCF^T$, $G = FK^T$, the complete adjacency matrix is

$$S = \begin{bmatrix} FCF^T & FK^T & F \\ KF^T & KK^T & K \\ F^T & K^T & C \end{bmatrix}$$

The matrices in the third column contribute to the ranking of papers

As for the two-class model...

Normalization of blocks and the use of a 3×3 parameter matrix $\Gamma = (\gamma_{i,j})$ lead to a stochastic matrix P of which the left Perron vector π represents the ranking of the subjects

Nice mathematical properties which relates the coupling matrix Γ with the "energy" of the three classes

As for the two-class model...

Normalization of blocks and the use of a 3×3 parameter matrix $\Gamma = (\gamma_{i,j})$ lead to a stochastic matrix P of which the left Perron vector π represents the ranking of the subjects

Nice mathematical properties which relates the coupling matrix Γ with the "energy" of the three classes

As for the two-class model...

Normalization of blocks and the use of a 3×3 parameter matrix $\Gamma = (\gamma_{i,j})$ lead to a stochastic matrix P of which the left Perron vector π represents the ranking of the subjects

Nice mathematical properties which relates the coupling matrix Γ with the "energy" of the three classes

Remarks ...

- Column normalization: blocks (Authors,Papers), (Authors,Journals) and (Papers, Journals) need column normalization
- Dummy journal: Introduction of a dummy journal which publishes the dummy paper.
- Probabilistic interpretation
- The Dummy Journal is the library, the dummy author is the librarian, the dummy paper is the catalog

Remarks ...

- Column normalization: blocks (Authors,Papers), (Authors,Journals) and (Papers, Journals) need column normalization
- Dummy journal: Introduction of a dummy journal which publishes the dummy paper.
- Probabilistic interpretation
- The Dummy Journal is the library, the dummy author is the librarian, the dummy paper is the catalog

Remarks ...

- Column normalization: blocks (Authors,Papers), (Authors,Journals) and (Papers, Journals) need column normalization
- Dummy journal: Introduction of a dummy journal which publishes the dummy paper.
- Probabilistic interpretation
- The Dummy Journal is the library, the dummy author is the librarian, the dummy paper is the catalog

Remarks ...

- Column normalization: blocks (Authors,Papers), (Authors,Journals) and (Papers, Journals) need column normalization
- Dummy journal: Introduction of a dummy journal which publishes the dummy paper.
- Probabilistic interpretation
- The Dummy Journal is the library, the dummy author is the librarian, the dummy paper is the catalog

Theoretical results for the three class model?

The previous perturbation theorems are not straightforward for this model

In fact...

Introducing a citation from paper i to paper j alters also the cross citation matrix.

Theoretical results for the three class model?

The previous perturbation theorems are not straightforward for this model

In fact...

Introducing a citation from paper i to paper j alters also the cross citation matrix.

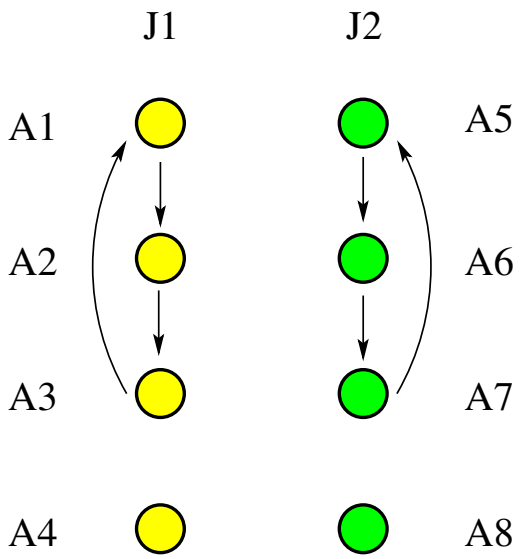
Theoretical results for the three class model?

The previous perturbation theorems are not straightforward for this model

In fact...

Introducing a citation from paper i to paper j alters also the cross citation matrix.

Some examples

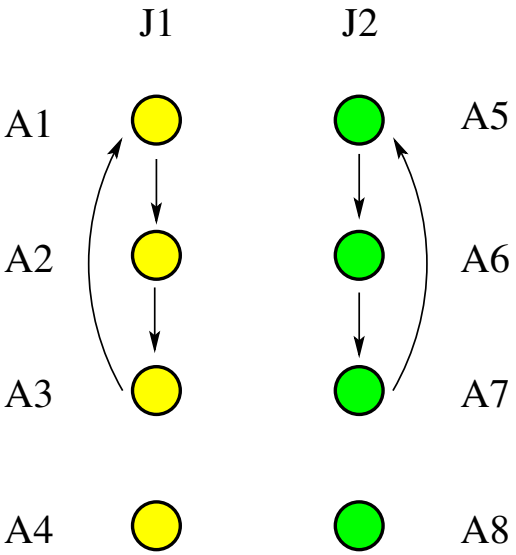


The yellow papers have the same rank of the corresponding green papers

Journal 1 has the same rank as Journal 2

Now we add a new citation

Some examples

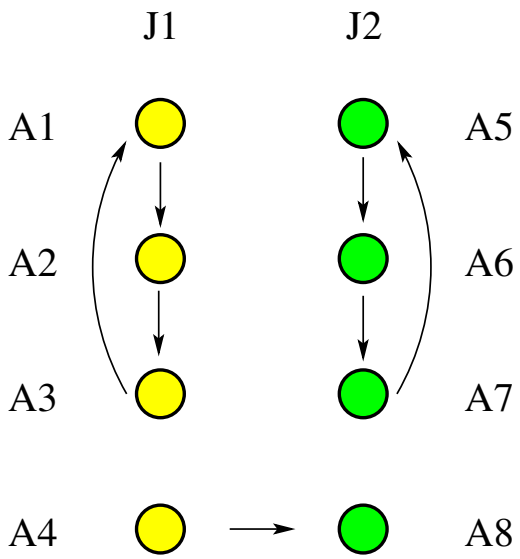


The yellow papers have the same rank of the corresponding green papers

Journal 1 has the same rank as Journal 2

Now we add a new citation

Some examples

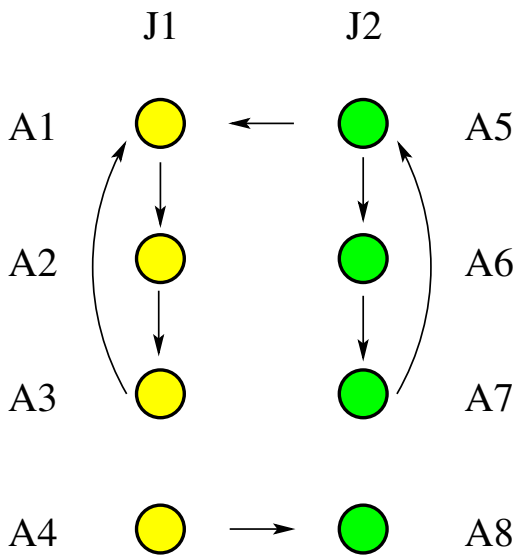


Journals: $(0.4835, 0.5165)$

Authors: $(0.1252, 0.1252, 0.1252, 0.1187$
$0.1267, 0.1267, 0.1267, 0.1254)$

Papers: $(0.1275, 0.1275, 0.1275, 0.1038$
$0.1296, 0.1296, 0.1296, 0.1248)$

Example



Journals: $(0.5015, 0.4985)$

Authors: $(0.1301, 0.1259, 0.1265, 0.1192$
$0.1255, 0.1252, 0.1234, 0.1244)$

Papers: $(0.1432, 0.1279, 0.1303, 0.1037$
$0.1269, 0.1257, 0.1193, 0.1228)$

Time

We need to add time into this mechanism

- Newly published papers do not have yet received enough citations
  Their rank is destined to be low
  The same for junior researchers whose rank remains lower than that of senior researchers
- How does it change the rank over the time?
- Was author X important in year y?

Time

We need to add time into this mechanism

- Newly published papers do not have yet received enough citations
  Their rank is destined to be low
  The same for junior researchers whose rank remains lower than that of senior researchers
- How does it change the rank over the time?
- Was author X important in year y?

Time

We need to add time into this mechanism

- Newly published papers do not have yet received enough citations
  Their rank is destined to be low
  The same for junior researchers whose rank remains lower than that of senior researchers
- How does it change the rank over the time?
- Was author X important in year y?

Time

We need to add time into this mechanism

- Newly published papers do not have yet received enough citations
  Their rank is destined to be low
  The same for junior researchers whose rank remains lower than that of senior researchers
- How does it change the rank over the time?
- Was author X important in year y?

Time

We need to add time into this mechanism

- Newly published papers do not have yet received enough citations
  Their rank is destined to be low
  The same for junior researchers whose rank remains lower than that of senior researchers
- How does it change the rank over the time?
- Was author X important in year y?

Time

We need to add time into this mechanism

- Newly published papers do not have yet received enough citations
  Their rank is destined to be low
  The same for junior researchers whose rank remains lower than that of senior researchers
- How does it change the rank over the time?
- Was author X important in year y?

Our idea:

- Scale the citation matrix C with an exponential decay function.

  A paper not cited recently looses its importance

  An old paper that is cited recently gains importance

Our idea:

- Scale the citation matrix C with an exponential decay function.

  A paper not cited recently looses its importance

  An old paper that is cited recently gains importance

Our idea:

- Scale the citation matrix C with an exponential decay function.

  A paper not cited recently looses its importance

  An old paper that is cited recently gains importance

Experiments

We used the CiteSeer dataset, focused primarily on the literature in computer and information sciences, made up by 250,000 authors, 350,000 papers.

Experiments

We used the CiteSeer dataset, focused primarily on the literature in computer and information sciences, made up by 250,000 authors, 350,000 papers. **One class model:**

paper	pos.	cit.
Diffie, Hellman- New directions in Cryptography	31	553
Rivest, Shamir, Adleman - Public Key cryptography	3	1218
Bryant -Boolean Functions Manipulation, BDD	1	1636
Kirkpatrick, Gelatt, Vecchi- Simulated Annealing	2	1337
Floyd, Jacobson - TCP/IP Protocol	4	1125
Canny - Computational approach to Edge detection	10	834

Two-class model:

Author	cit	pap.	cit./pap.
Randal Bryant	2615	83	31.5
Sally Floyd	4950	91	54.4
John K. Ousterhout	2214	23	96.3
Luca Cardelli	2112	91	23.2
Van Jacobson	4719	40	118.0
Rakesh Agrawal	4745	83	57.2
Jack J. Dongarra	2799	291	9.6
Raj Jain	1038	116	8.9
Douglas C. Schmidt	2980	329	9.1
Vern Paxson	2735	66	41.4
John Mccarthy	911	41	22.2
Thomas A. Henzinger	3694	176	21.0

Three class model: Journals (AMS-MR)

$\approx$ 3500 journals, 110,000 authors, 300,000 papers

Journal	n. cit	n. pap.	IF	EF
Trans. AMS	22796	5247	0.820	0.041
Inventiones Mathematicae	21181	2481	1.659	0.039
Annals of Mathematics	19365	2193	2.426	0.031
Proc. AMS	16722	6045	0.513	0.042
Comm. Math. Phys	21997	3748	2.077	0.052
J. Algebra	15059	4457	0.568	0.039
Duke Math. J.	11939	2161	1.409	0.026
Mathematische Annalen	11248	2670	0.902	0.023
J. Functional Analysis	13778	2437	0.866	0.031
Comm. on Pure Appl. Math.	12111	1227	2.031	0.019

Authors - time aware method (AMS-MR)

Author	num. cit	num. pap.
Lions, Pierre-Louis (FM)	2641	199
Erdös, Paul	1358	377
Bourgain, Jean (FM)	1019	156
Simon, Barry	1502	198
Shelah, Saharonh	972	333
Brezis, Haïm	1698	127
Lustzig, George	1145	87
Caffarelli, Luis	1288	131
Yau, Shing Tung (FM)	1571	136
Connes, Alain (FM)	1114	79
Arnold, Vladimir	551	90

Sorry, very few papers of Lucien Godeaux are in the database used

Authors - time aware method (AMS-MR)

Author	num. cit	num. pap.
Lions, Pierre-Louis (FM)	2641	199
Erdös, Paul	1358	377
Bourgain, Jean (FM)	1019	156
Simon, Barry	1502	198
Shelah, Saharonh	972	333
Brezis, Haïm	1698	127
Lustzig, George	1145	87
Caffarelli, Luis	1288	131
Yau, Shing Tung (FM)	1571	136
Connes, Alain (FM)	1114	79
Arnold, Vladimir	551	90

Sorry, very few papers of Lucien Godeaux are in the database used

Authors - time aware method (AMS-MR)

Author	num. cit	num. pap.
Lions, Pierre-Louis (FM)	2641	199
Erdös, Paul	1358	377
Bourgain, Jean (FM)	1019	156
Simon, Barry	1502	198
Shelah, Saharonh	972	333
Brezis, Haïm	1698	127
Lustzig, George	1145	87
Caffarelli, Luis	1288	131
Yau, Shing Tung (FM)	1571	136
Connes, Alain (FM)	1114	79
Arnold, Vladimir	551	90

Sorry, very few papers of Lucien Godeaux are in the database used

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...we believe it is a nice recipe!

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...we believe it is a nice recipe!

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...we believe it is a nice recipe!

Conclusions

- Use automatic ranking only when strictly necessary
- Cannot replace peer review
- There are parameters that have to be agreed upon at a "political" level

Pros of our approach

- Flexible to meet multiple goals
- The ranking obtained is a mixture of ingredients...we believe it is a nice recipe!

Open problems

Many open problems for mathematicians:

- More perturbation results for the three-class model: what happens if we add an author? Is it convenint to write a paper with your office-mate?
- Theoretical results for the time-aware mechanism.
- Can we apply similar ideas to "similar" problems?

Open problems

Many open problems for mathematicians:

- More perturbation results for the three-class model: what happens if we add an author? Is it convenint to write a paper with your office-mate?
- Theoretical results for the time-aware mechanism.
- Can we apply similar ideas to "similar" problems?

Open problems

Many open problems for mathematicians:

- More perturbation results for the three-class model: what happens if we add an author? Is it convenint to write a paper with your office-mate?
- Theoretical results for the time-aware mechanism.
- Can we apply similar ideas to "similar" problems?

Open problems

Many open problems for mathematicians:

- More perturbation results for the three-class model: what happens if we add an author? Is it convenint to write a paper with your office-mate?
- Theoretical results for the time-aware mechanism.
- Can we apply similar ideas to "similar" problems?

Thanks!