Liber

Rata

Liberata

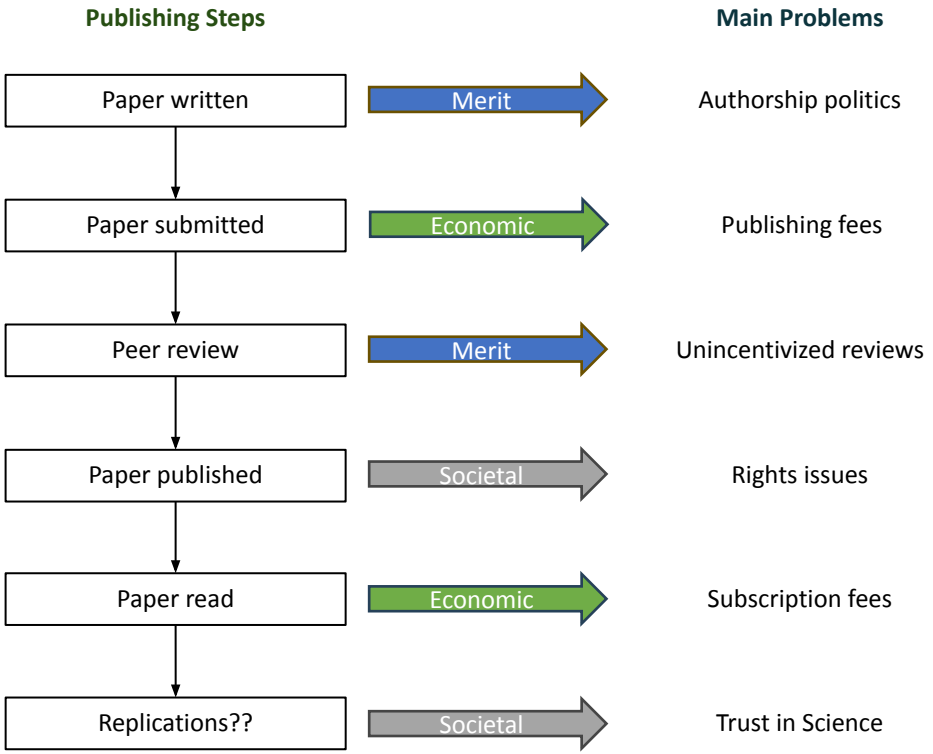
Liberating academic publishing with shares

Han Zhang & Bill Zheng

CS 521

Fall 2024

Contents



Academic Publishing Today

Academic publishing today suffers many incentive misalignment problems, broadly categorizable into three types.

- Merit and accreditation
- Economic rewards
- Societal trust

The root of much of these problems have to do with a lack of any incentive structure for critical roles in science, such as peer review or replication.

Academic community over time relies on weak signals of quality such as journal name and scientist fame.

Liberata – Shares Based Academic Publishing

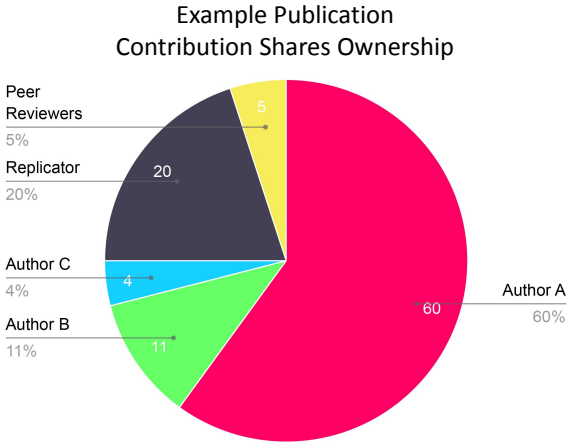
- Current attributions on academic papers: [Discrete]

Paper Title:	*Some Novel Scientific Discovery about X*
Authors:	Author A, Author B, Author C, …

- Liberata attributions on academic papers: [Continuous]

Paper Title:	*Some Novel Scientific Discovery about X*
Authors:	Author A$^{(60)}$, Author B$^{(11)}$, Author C$^{(4)}$, …
Peer Reviewers:	Peer Reviewer C$^{(2)}$, Peer Reviewer D$^{(1.5)}$, Peer Reviewer E$^{(1.5)}$
Replicators:	Replicator F$^{(12.5)}$, Replicator G$^{(7.5)}$

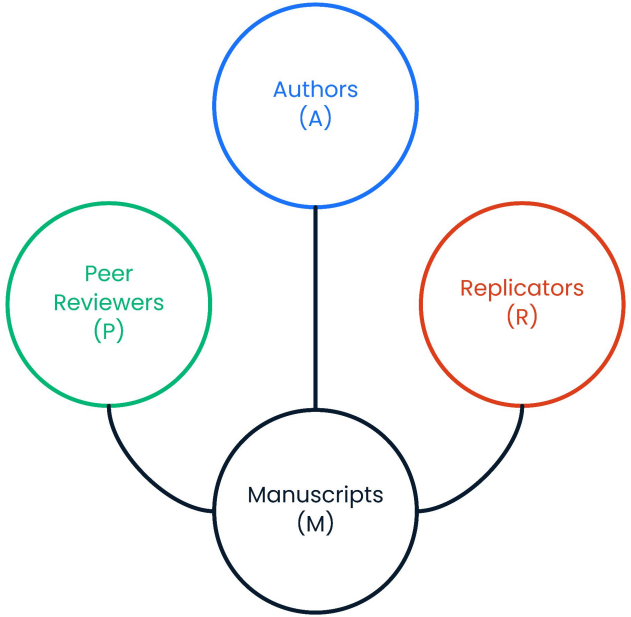
- Liberata:
  - Reduces authorship politics by improving academic metrics
  - Solves the curation problem by incentivizing peer review
  - Solves the replication crisis by incentivizing replications



Citation counts*, together with shares, gives rise to the concept of **scientific capitalization**, akin to market capitalization of a company.

* Liberata uses relative citations, which account for total no. of citations on a paper.

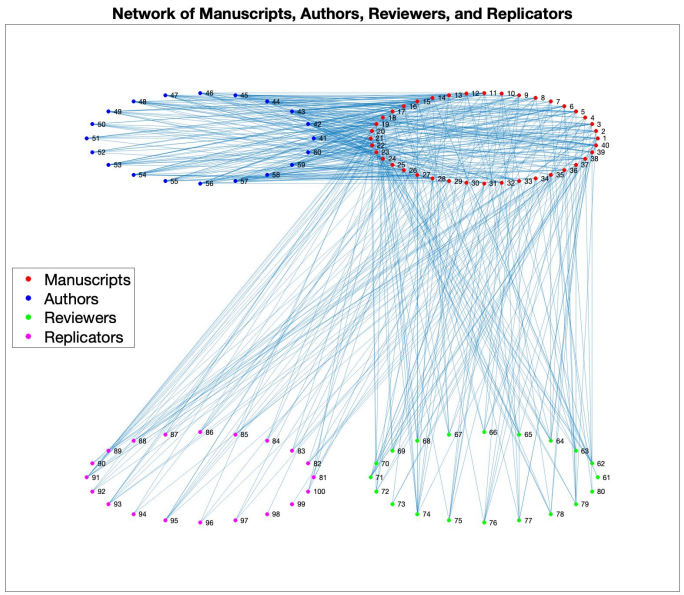
Graph and Matrix Representation

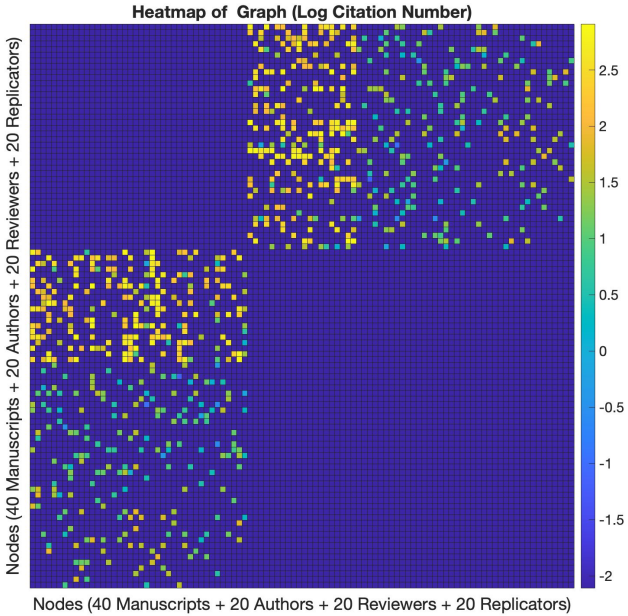


M	0	SA*CM	SP*CM	SR*CM
A	SA*CM	0	0	0
P	SP*CM	0	0	0
R	SR*CM	0	0	0

- Graph has four types of nodes: A, P, R, M
- Edge weight = person's shares * citation count of manuscript
- Edges only between people and manuscripts -> star bipartite graph
- Sparse matrix by construction

Simulated Example





Adjacency matrix is positive semi-definite

Simulation parameters:

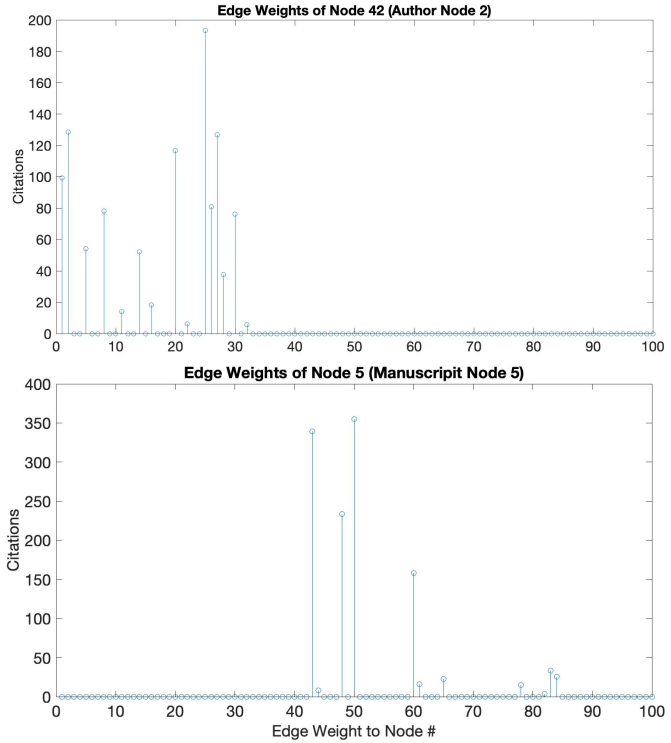
Nodes:

- 40 Manuscripts
- 20 Authors
- 20 Peer Reviewers
- 20 Replicators

Assumptions:

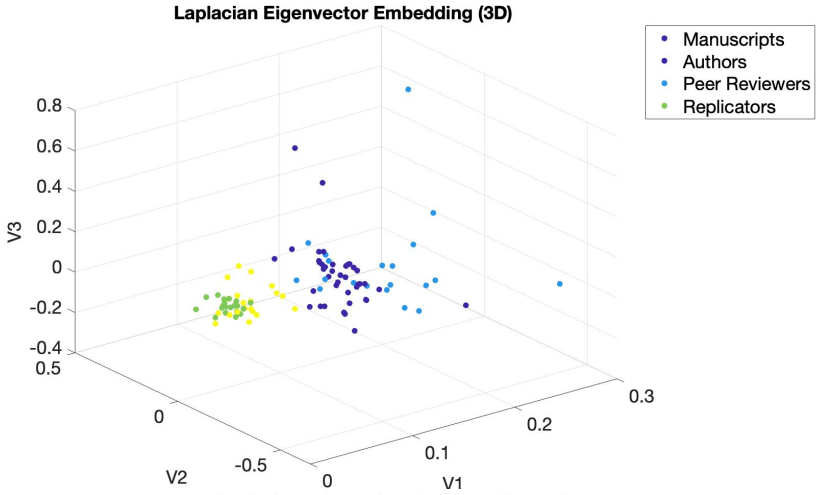
- 5% shares to P nodes
- 10% shares to R nodes
- 1 - 10 A nodes per M node
- 3 P node per M node
- 1 - 3 R node per M node
- 10 - 1000 citations per M node

Fetch Vectors & Distributions



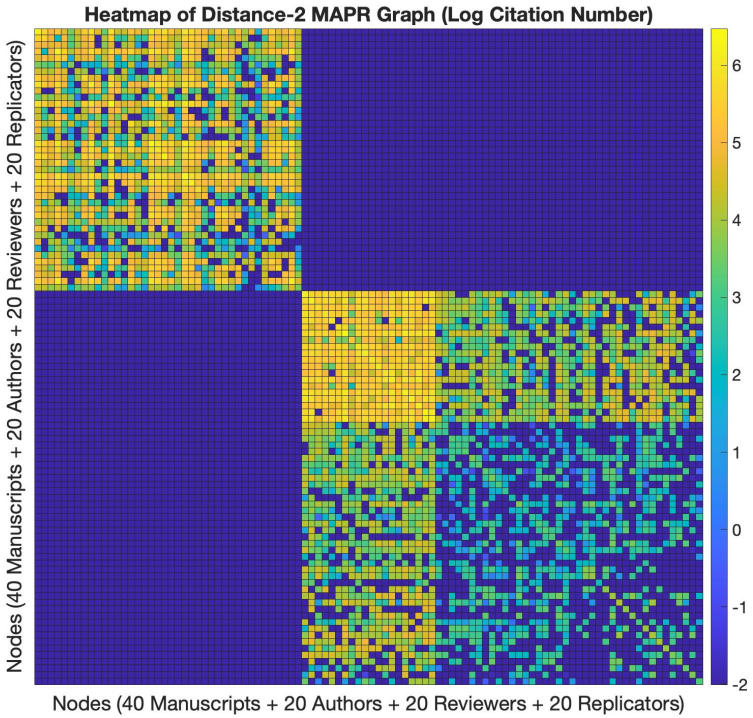
- Fetch vectors are compositions of unit vectors:
  - Extracts a distribution from the adjacency matrix
  - A, P, or R nodes -> one person's sci. cap. across all manuscripts as A, P, or R role.
  - M nodes -> total citation count divided amongst contributors.
  - Extractions -> distribution plots of sci. cap. -> check for "collaborativeness" & anomalies.
- Composition of A unit vectors -> total productivity of a lab or institution, (once corrected for domain citations and career length.)
- Composition of P or R unit vectors -> calculate average cost of review and replication tasks within a field.
- Composition of M unit vectors -> domain sci. cap.

Spectral Analysis with Laplacians



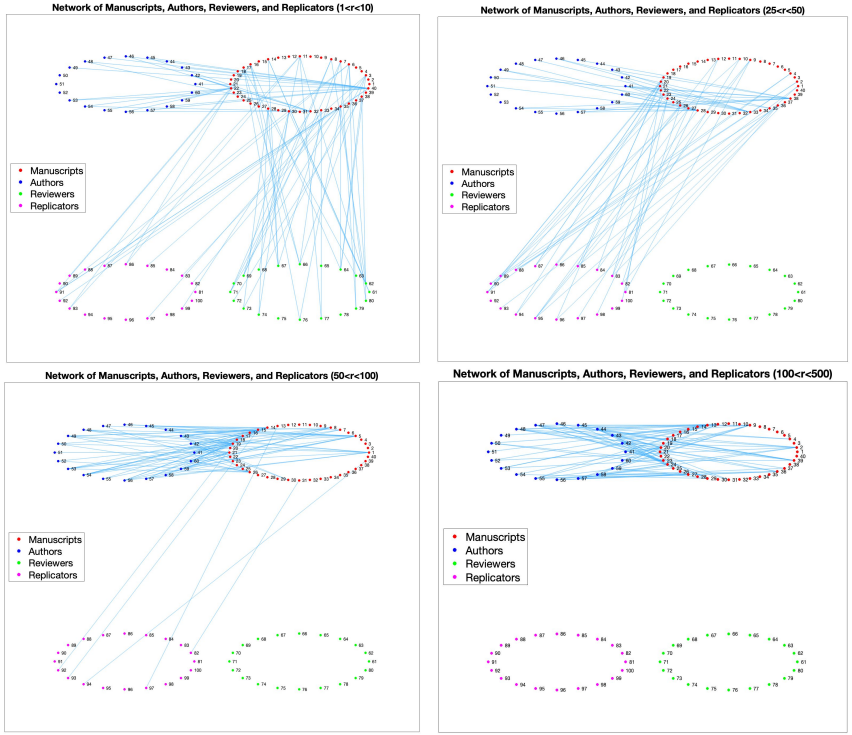
- The multiplicity of the 0 eigenvalue:
  - number of disconnected components
  - Can be used together with sci. cap. to detect budding fields of science as well as siloed fields of science.
- The eigenvectors of null space -> reveals fragmented scientific domains, identifying its publications, authors, peer reviewers, and replicators.
- Smallest non-zero eigenvalues' eigenvectors
  - embeds the graph into global structure clusters
  - reveals collaboration networks
  - visualizes insularity between subdomains
- Largest eigenvalues' eigenvectors
  - embeds graph into local structure clusters
  - edge weights matter more than edge count
  - Embedding results in "impactfulness" graph

2 Step Graphs



- A, P, R nodes <-> A, P, R nodes
- M nodes <-> M nodes
- A^2 elements -> sum of products between edges of A, P, R nodes with each other, and M nodes with M nodes.
- Applications:
  - Discover collaboration networks
  - Measure exploitiveness of PIs
  - Group self-citation tendencies
  - Equality of split

Orbital Band Neighbors



- Higher edge weights -> higher scientific capital
- OBN -> check for $t_1 < r < t_2$, filtering for edges within a band of scientific capital
  - A -> high impact authors.
  - P -> reviewers paid at premium or discount for their services.
  - R -> cost of replication for different scientific fields.
  - M -> detect “riskiness” via compensation of P and R nodes.
- Comparing filtered edges to total original edges can be used as an anti-collusion, anti-corruption metric.

Liberata Compared

Feature	Liberata	arXiv	Private Journals
Online repository of knowledge	✔	✔	✔
Anonymous peer review	✔	✘	✔
Incentivized peer review	✔	✘	✘
Incentives for replication studies	✔	✘	✘
Accurate academic metrics	✔	✘	✘
Open source access	✔	✔	✘
Non-profit academic organization	✔	✔	✘

Curious to learn more?

Visit liberata.online in Q2. 2025

Acknowledgements

- Prof. Xiaobai Sun's code was used in the generation of some of the figures in this presentation.
- ChatGPT was used to assist with coding for plotting and matrix operations in this presentation.