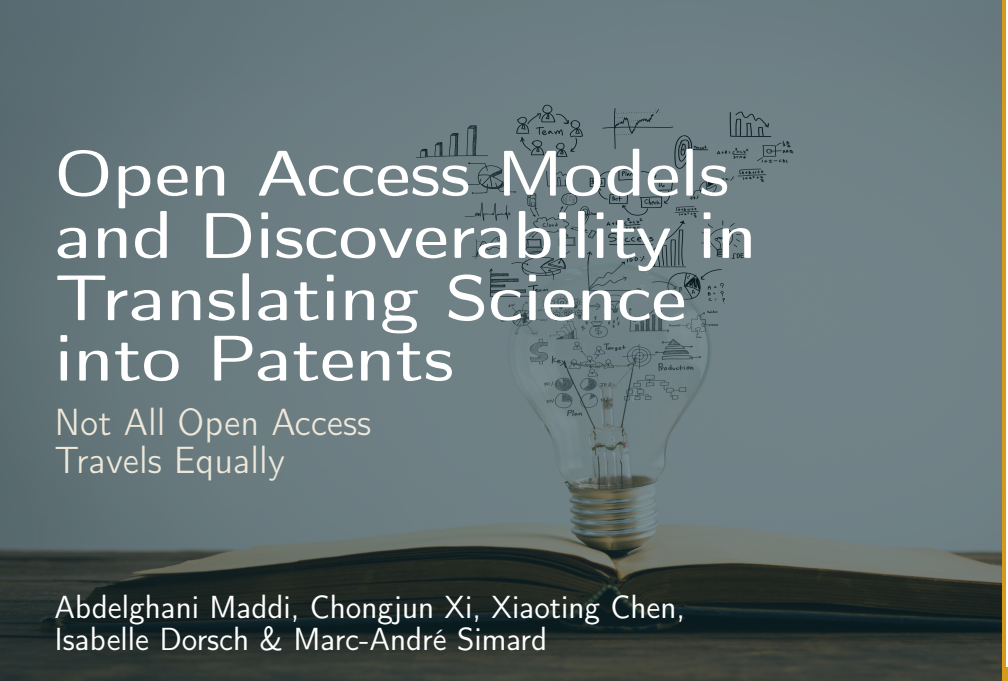


Open Access Models and Discoverability in Translating Science into Patents

Not All Open Access Travels Equally

Abdelghani Maddi, Chongjun Xi, Xiaoting Chen,
Isabelle Dorsch & Marc-André Simard

PUBMET 2026

10 September



“In the final years of the 20th century we entered a knowledge-based society.”

European Commission

European Commission (2000)



“Economic and social development will depend essentially on knowledge in its different forms, on the production, acquisition and use of knowledge.”

European Commission

European Commission (2000)

KNOWLEDGE DOES NOT CIRCULATE IN ONE SPACE

Science circulates in different worlds

01
Academic

Citation, recognition, and field-building.

02
Societal

Policy, public debate, and collective problem-solving.

03
Economic

Innovation, patents, and industrial applications.

Our observable trace of this economic circulation: scientific publications cited in patents.

Patent citations as a trace

NON-PATENT REFERENCES

NPRs

Scientific papers, books, reports and other non-patent documents cited in patent records.

WHAT THEY ARE USED TO STUDY

01 Science–technology linkages

02 Technological relevance of research



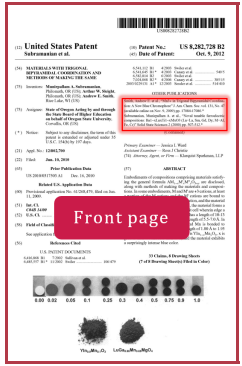
A citation makes a possible connection between science and technology visible in the patent record.

Reminder an NPR is not direct evidence that a publication was read, used or absorbed by an inventor.

THE SAME SCIENTIFIC REFERENCE CAN APPEAR IN TWO PLACES

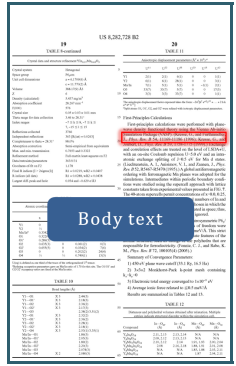
Front page or body text?

FRONT PAGE



Formal citation list

BODY TEXT



Technical description or argument

The location of the citation changes what we can reasonably infer from it.

Citation position changes the interpretation

Patent citations are not equivalent: their location provides information about the kind of connection we observe.

01
Front page

A formal trace of visibility in the patent record.

02
Body text

A closer trace of technical or semantic proximity.

03
Both

Formal visibility and technical proximity combined.

Citation position helps separate visibility from technical proximity.

Open access promises wider circulation.

If knowledge is open, it should be easier to move beyond academia; into society, policy and innovation.

Read

Remove access barriers

Find

Make research visible

Reuse

Enable application

But access is only the first condition for circulation.



THE FRICTION BETWEEN OPENNESS AND USE

Open does not automatically mean used.

01 Access

Can someone read the publication?

→ **02 Discoverability**

Can someone find it in the systems they use?

→ **03 Proximity**

Is it relevant to the problem being solved?

Open ≠ Findable ≠ Used

Patent citations can be shaped by all three conditions, not by access alone.

OPEN ACCESS IS NOT A SINGLE CHANNEL

OA models travel through different infrastructures.

The publication model is not neutral: it affects the routes through which research becomes visible.

Journals

Publisher brands, editorial platforms and established venues.

Metadata

Indexing, identifiers, links and machine-readable records.

Routines

Search platforms and professional knowledge practices.

Some OA is highly visible. Some OA is open, but less institutionally visible.

THE RESEARCH QUESTION

Do OA models shape the journey of science into patents?

We distinguish two questions that are often conflated.

01 Being seen

Which publications become visible in patent citations?

02 Being used

Once cited, how close are they to patent content?

THREE EXPECTATIONS

OA models may structure this journey differently.

H1 Selection

OA models differ in their likelihood of appearing in patent citations.

H2 Decoupling

Greater citation visibility does not necessarily imply greater semantic proximity.

H3 Position

Front-page, body-text and both-position references capture different relations.

The same openness does not guarantee the same route into technology.

Data: from patent citations to semantic proximity

32.7M
Patent–publication links
distinct USPTO NPR links

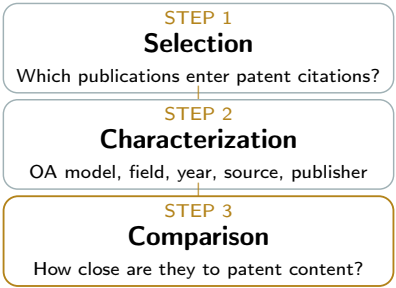
2.27M
semantic comparisons
patent + publication abstracts

6
access categories
closed, bronze, green, hybrid, gold, diamond

Four data layers

- **Patent–science links** from non-patent references
- **Patent content** from PatentsView abstracts
- **Publication metadata and OA status** from OpenAlex
- **Semantic proximity** using SPECTER2 embeddings

Unit of analysis: one patent–publication link.



Method: two empirical lenses

1. Selection lens

Compare OA shares in patent-cited publications with the OpenAlex baseline.

$$\log_2\left(\frac{\text{patent-cited share}}{\text{baseline share}}\right)$$

2. Proximity lens

Compare patent abstracts and publication abstracts using SPECTER2 embeddings.

$$\cos(V_{patent}, V_{paper})$$

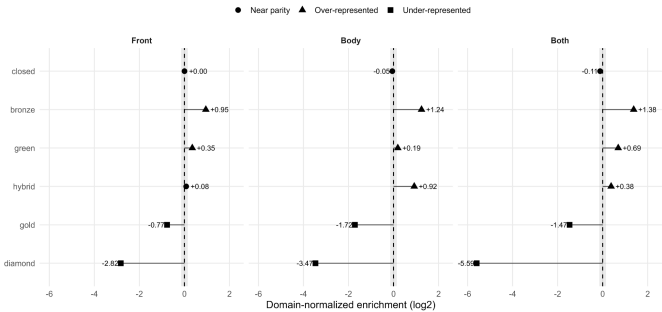
Validation: cited pairs are distinguishable from controls. AUC = **0.894** for SPECTER2, **0.906** for SBERT MiniLM, and **0.807** for TF–IDF.

FINDING 01: SELECTION

Patent citations favour visible OA channels.

Figure 2. Domain-normalized representation of OA models in USPTO patent-publication links

Values are log2 ratios relative to the OpenAlex baseline; the shaded band marks near-parity deviations.



Bronze and hybrid

are overrepresented among patent-cited publications.

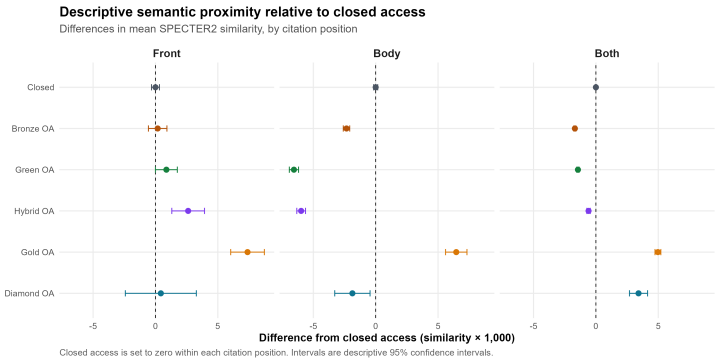
Gold, and especially diamond, OA are underrepresented.

Patent citations do not sample open science evenly.

FINDING 02: SEMANTIC PROXIMITY

Visibility and proximity diverge.

The models most visible in patents are not necessarily the closest to patent content.

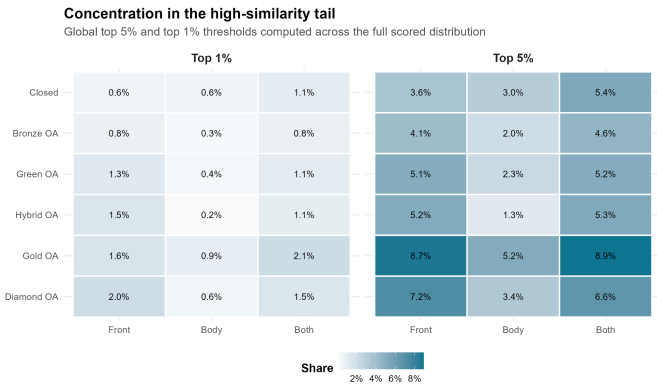


Visible does not mean used.

Citation visibility and semantic proximity capture different dimensions of the science–technology link.

FINDING 03: CITATION POSITION

The strongest proximity signal is “both”.



FRONT ONLY

Formal and bibliographic visibility.

BODY ONLY

A textual connection, but not always the closest one.

BOTH

The most stable signal of semantic proximity.

WHAT THE RESULTS CHANGE

Patent citations are a composite signal.

01 Selection

Which publications become visible in patent records?

+ 02 Discoverability

Which infrastructures make them findable?

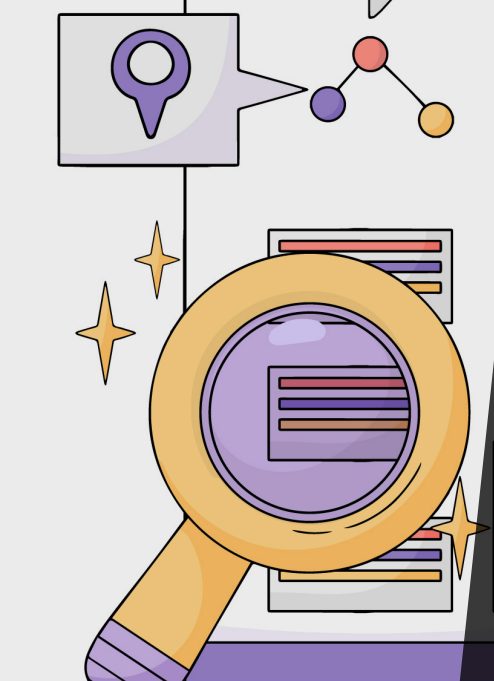
+ 03 Relatedness

How close are they to the invention's technical content?

Cited ≠ used

Open ≠ findable

Visible ≠ close



IMPLICATION

Open science needs discoverability infrastructure.

What makes research travel?

- Interoperable metadata
- Persistent IDs and stable links
- Machine-readable full text
- Search-environment integration
- Innovation and patent visibility

Open access policy is also discoverability policy.

FINAL TAKE-HOME

Open access does not travel by itself.

It travels through infrastructures that make science visible, searchable and reusable.

Not only: **Is science open?**

Also: Is it findable where innovation looks for knowledge?

Thank you

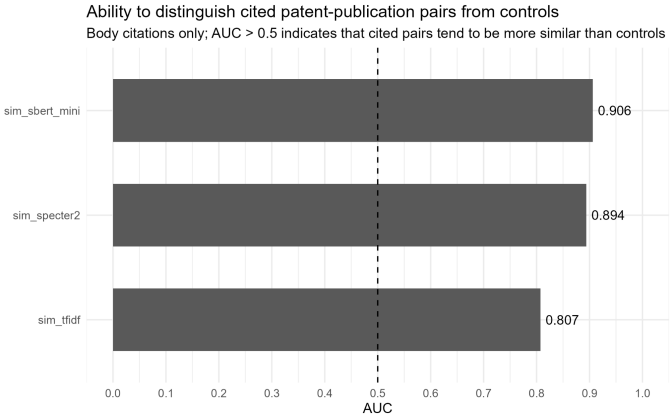
Abdelghani Maddi

GEMASS (CNRS / Sorbonne Université)

abdelghani.maddi@cnrs.fr

@amaddi.bsky.social

Backup — Semantic validation



Validation on body-citation links: cited pairs are distinguishable from controls across three similarity methods.

Backup — Data construction

Step	Links
Distinct USPTO patent–publication links	32,696,577
Front-only links	19,882,137
Body-only links	9,616,412
Both-position links	3,198,028
Scored semantic links	2,272,444

The semantic analysis is restricted to links with patent and publication abstracts.

Backup — Diamond OA caution

Diamond OA is analytically important, but it is much smaller and unevenly distributed across fields.

Citation position	Diamond links	Share
Front-only	539	1.06%
Body-only	1,552	0.75%
Both-position	5,335	0.26%

Main message: treat Diamond separately from Gold; avoid strong general claims.