

Identifying Informal Opinion Leaders in Online Review Networks

Network structure, review outcomes, and language use in Sephora’s review ecosystem
Author: Ishika Narang; MSc Analytics & AI, ESMT Berlin



INTRODUCTION

Online consumer reviews have become a central source of information in digital marketplaces, shaping product discovery, evaluation, and adoption. However, prior research shows that influence in review systems is highly unequal: a small subset of reviewers disproportionately shapes collective perceptions (Dellarocas, 2023). Meaning, **online reviews look democratic — but influence is not**. When influence is unevenly distributed in the reviews, **organizations face the risk of misinterpreting feedback and overreacting** to highly visible opinions while overlooking broader patterns, for example by discontinuing a product or pricing it much lower to sell more. Thus, my project treats Sephora’s review environment as an informal influence network and asks: who acts as “opinion leaders,” and what signals make their influence stronger?



RESEARCH QUESTIONS

1. **Network Leadership:** Can **structurally central** reviewers be identified as informal opinion leaders?
2. **Impact Relevance:** Do these structurally powerful reviewers **differ in review outcomes** (ratings/helpfulness) compared to peripheral reviewers, **beyond pure activity?**
3. **Language as influence:** Is influence reflected in language and do opinion leaders **use distinctive vocabulary or narrative styles** that “travel” across product contexts?

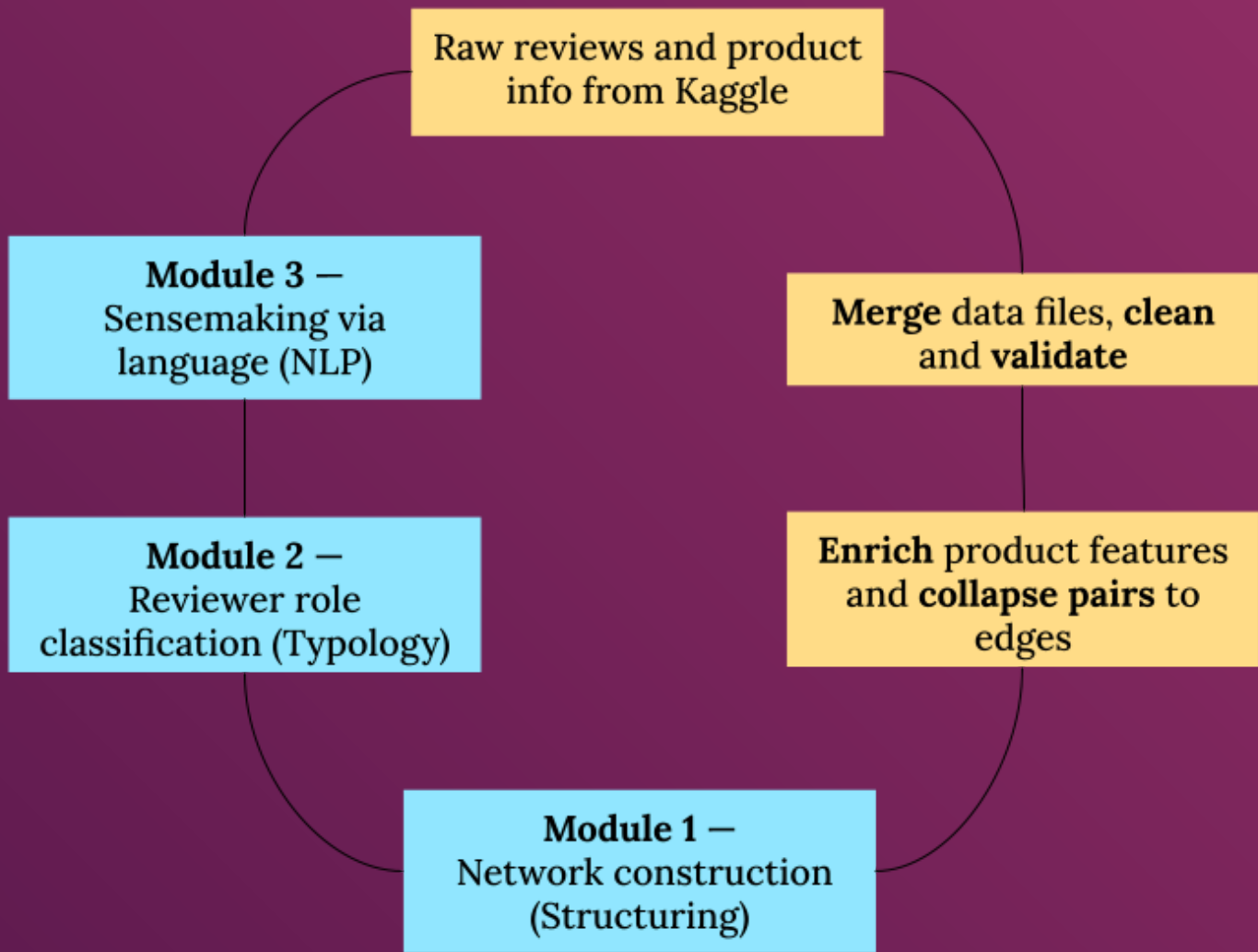
BIBLIOGRAPHY

Dellarocas, C. (2023). The digitization of word of mouth: Promise and challenges of online feedback mechanisms | request PDF. ResearchGate.
https://www.researchgate.net/publication/5175425_The_Digitization_of_Word_of_Mouth_Promise_and_Challenges_of_Online_Feedback_Mechanisms

DATA SOURCE & SCOPE

- Kaggle: Sephora Products & Skincare Reviews
- **≈ 1.09M reviews, 503k reviewers, 8k+ products**
- Unit of analysis:
 - Review events enriched with product metadata + timing + choice complexity signals. indiv-project
- Core design choice:
 - repeated reviews for the same reviewer-product pair are collapsed into one edge (influence is about exposure, not repetition), while keeping edge weights for intensity.
- Outputs created:
 - edges_rp: reviewer-product edge table (the “network backbone”)
 - reviewer-level network metrics

METHODOLOGY



Module 1 — Network construction (Structuring)

- Bipartite reviewer-product network:
 - nodes = reviewers/products; edges = review events.
- **Metrics:**
 - activity (degree), brokerage (betweenness), portfolio breadth (entropy across brands/categories).



Module 2 — Reviewer role classification (Typology)

- “Opinion leadership” defined as
 - **high activity + high breadth, not activity alone.**
- Four reviewer roles:
 - **Boundary Spanners:** high activity + high portfolio diversity
 - **High-Activity Specialists:** active but narrow
 - **Low-Activity Generalists:** broad but low visibility
 - **Peripheral Specialists:** majority, situational contributors



Module 3 — Sensemaking via language (NLP)

- Compare language signatures by reviewer role
 - (TF-IDF terms, vocabulary diversity/TTR, category-spanning language).



Identifying Informal Opinion Leaders in Online Review Networks

Network structure, review outcomes, and language use in Sephora's review ecosystem

Author: Ishika Narang; MSc Analytics & AI, ESMT Berlin

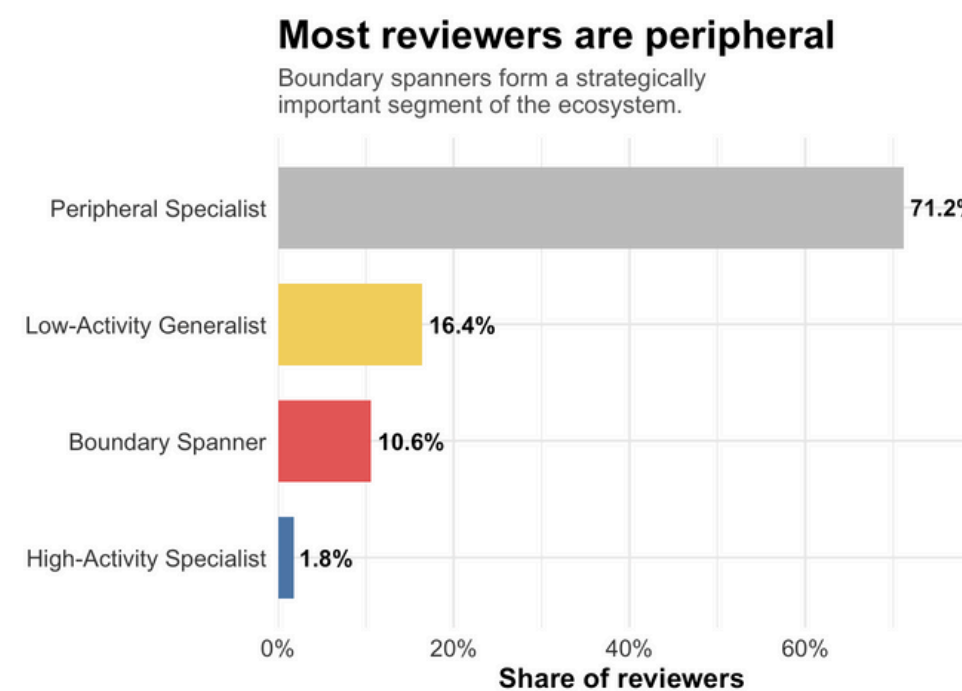


RESULTS

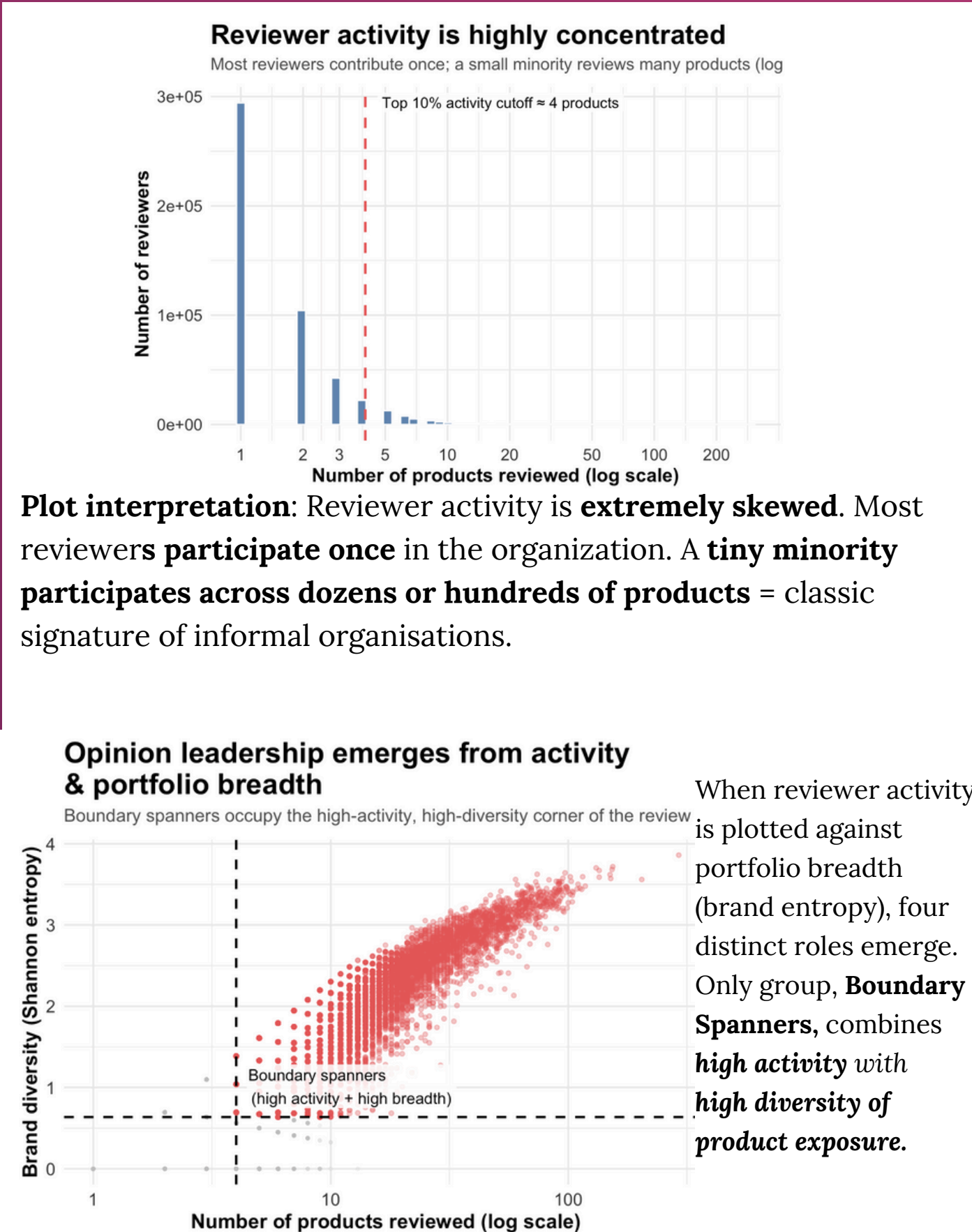
Module 1: Network Leadership

Online review systems often appear democratic, but informal organizations typically exhibit highly unequal participation. If opinion leaders exist in Sephora's review ecosystem, they should be visible as a structurally distinct minority rather than as “average” reviewers.

- Created a bi-partite network (edges_rp) that contains reviewers, products, and edge weights are defined by how strong / informative that relationship is
- Before identifying opinion leaders structurally, it is important to understand **how reviewers are distributed across roles in Sephora's review ecosystem.**



- **Most reviewers are peripheral (71.2%)** → The ecosystem is wide and shallow
- **Boundary spanners are rare (10.6%) but non-trivial** → Leadership is concentrated but meaningful
- **High activity ≠ dominance (1.8%)** → Refutes the “power users run everything” assumption
- **Sephora review ecosystem is wide and shallow, not repetitive.** Influence therefore comes from breadth of exposure (reviewing many products), not from repeatedly reviewing the same product.



Plot interpretation:

- High activity alone does not imply leadership. Reviewers who are **active but narrowly focused** (*High-Activity Specialists*) exert deep but localized influence.
- *Boundary Spanners*, in contrast, **participate widely and span brands and categories**, positioning them to compare products, transfer evaluations, and shape shared beliefs across contexts.
- This combination of **visibility and breadth is what makes them structurally powerful**.

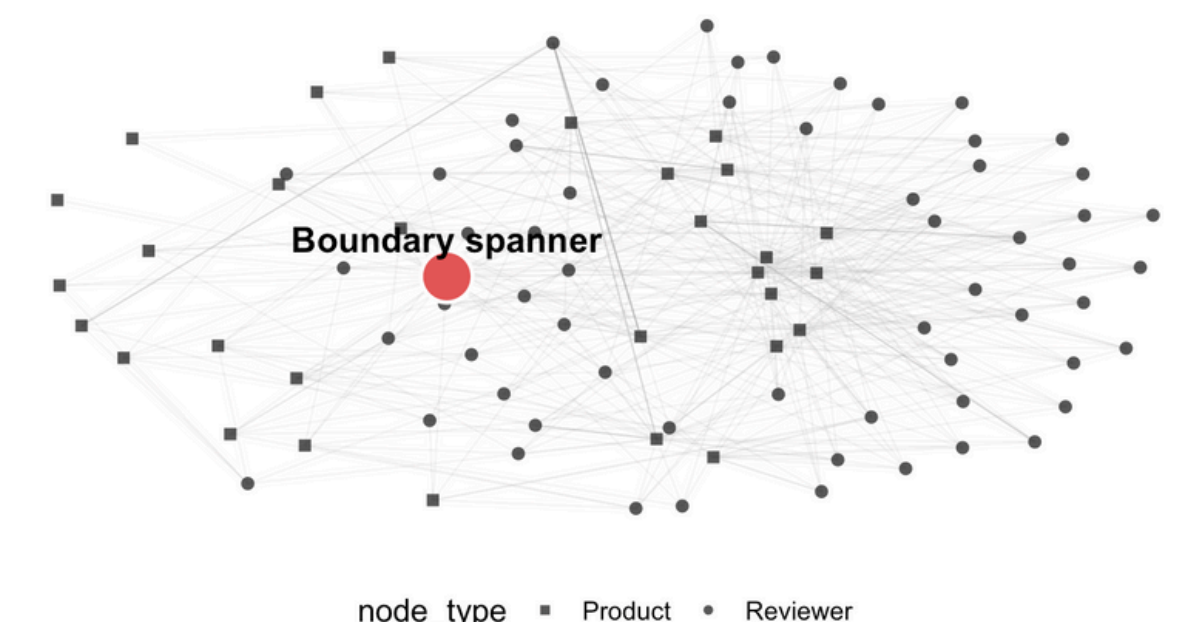
Key RQ1 takeaway - Opinion leaders are not the most active reviewers, but a small group of boundary spanners who review many different products and brands.

Module 2: Impact Relevance

If boundary spanners act as informal opinion leaders, their reviews should matter more, **not because they are louder**, but because they **appear in high-visibility**, and **statistically have higher rating and helpfulness**.

Ego-network of a boundary spanner (trimmed)

Top products reviewed by the boundary spanner + the reviewers who overlap most on those products.



Plot interpretation: This ego network illustrates how one specific boundary spanner is embedded in the review ecosystem. The highlighted reviewer (red) is directly connected to many products (squares), but the key signal is that those products are also reviewed by multiple other reviewers (circles). That shared overlap is the mechanism through which opinions can “travel”: when the boundary spanner reviews products that many others also touch, their evaluations sit in the high-visibility part of the network, shaping the same product conversations other reviewers participate in.

Identifying Informal Opinion Leaders in Online Review Networks

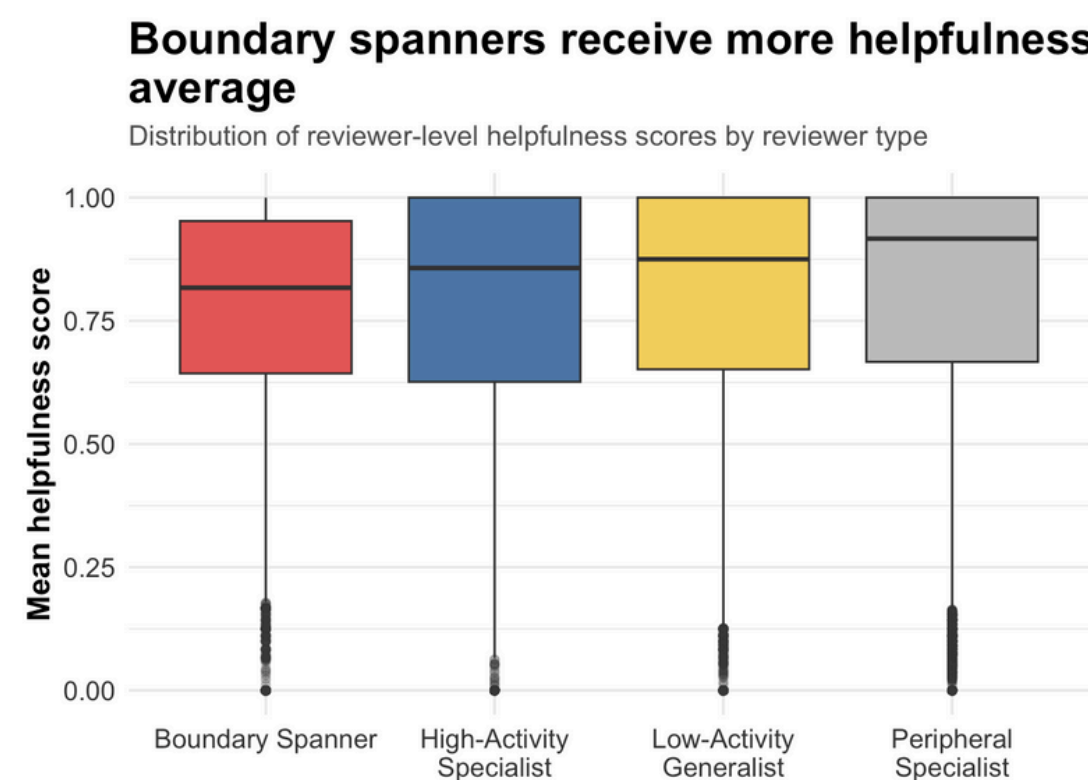
Network structure, review outcomes, and language use in Sephora's review ecosystem

Author: Ishika Narang; MSc Analytics & AI, ESMT Berlin



RESULTS

In organizational terms, the boundary spanner is not just active, they occupy a **broker-like position that links multiple product clusters and exposes their judgments across a wider audience** than a typical specialist.



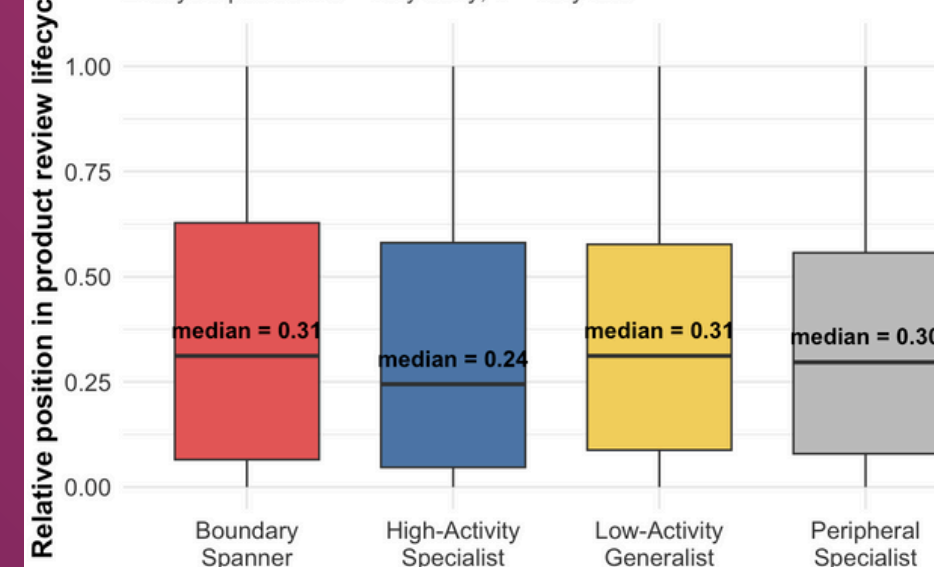
- **Ratings are consistently high across all roles ($\approx 4-5$),** indicating that influence \neq negativity.
- Boundary spanners do not drive impact by being harsher or more extreme
- Helpfulness scores are comparable or slightly higher
- Boundary spanners' influence does not come from writing more critical reviews. Instead, it comes from where and how often their reviews appear. They shape interpretation in complex choice environments, even when sentiment is broadly positive.

Reviewer roles differ not only in structure, but also in when they contribute reviews in a product's review lifecycle.

Plot interpretation (next column): Boundary spanners are not first movers. Instead, they act as interpreters and consolidators, synthesizing early signals into evaluations that later reviewers encounter. **Influence here is not about being first — it is about being strategically timed.**

When do different reviewer types show up in a product's review lifecycle?

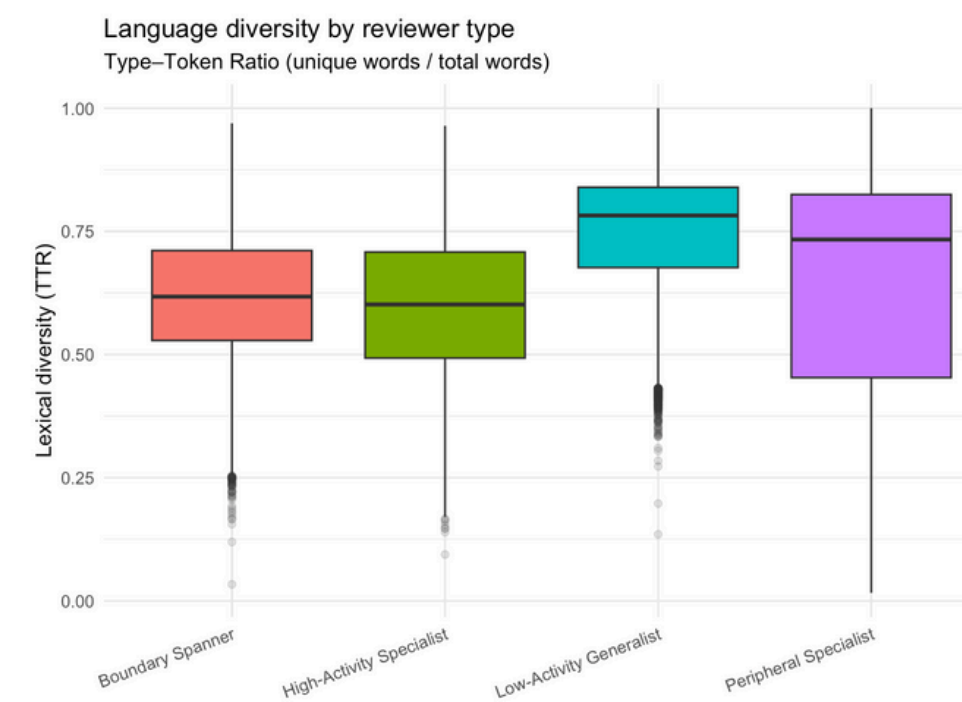
Lifecycle position: 0 = very early, 1 = very late



Key RQ2 takeaway - Boundary spanners matter because their reviews appear where many other reviewers and consumers see them, not because they are more negative or extreme.

Module 3: Language as influence

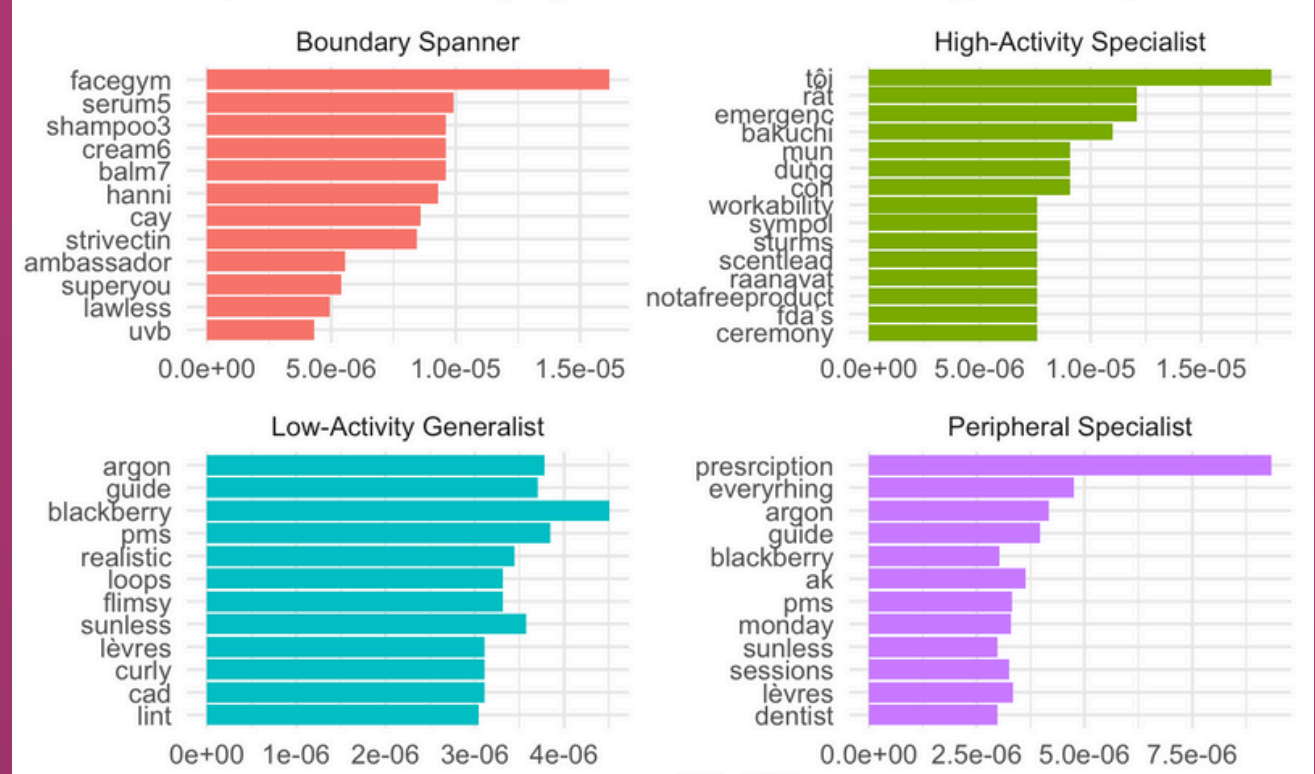
If boundary spanners act as informal opinion leaders, this should be reflected not only in how much they write, but in how they use language to frame products for others.



Plot interpretation: Boundary Spanners exhibit moderate, controlled lexical diversity. High-Activity Specialists show similarly constrained vocabularies. **Low-Activity Generalists and Peripheral Specialists display higher and more variable lexical diversity, probably due to situational reviewing.**

Distinctive language by reviewer type

Top TF-IDF terms highlight how different reviewer types frame product evaluation



- **Boundary spanners** - integrative, cross-category language centered on brands, product formats, and routines
- **High activity specialist** - technical and diagnostic terminology (ingredients, formulation, performance)
- **Low activity generalist** - high lexical diversity and narrative expression
- **Peripheral specialist** - situational and problem-specific language tied to specific needs or contexts

Key RQ3 takeaway - Boundary spanners influence others by using clear, comparable language that helps people understand and compare products across brands and categories.

CONCLUSIONS & NEXT STEPS

Informal opinion leaders are boundary spanners whose influence comes from visibility, timing, and integrative language, not extremity.

Next steps:

- Model influence over time, Link language to downstream behavior (Test whether integrative language predicts review helpfulness or sales outcomes), Extend beyond Sephora (ex: Amazon)