Project Proposal
Organisational Analytics
Ishika Narang, MAAI

Who Shapes Collective Opinion? Identifying Informal Leaders in Online Review Organizations

Context (organisational problem):
Retail organizations increasingly rely on customer review systems as a decentralized "voice of the customer," but these systems often function like informal organizations where a small subset of reviewers can shape shared beliefs about product quality. 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. Despite the importance of this issue, most review analytics treat all reviewers as equal contributors.

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? The project will make use of structuring logic (patterns of relations, brokers/influencers) combined with sensemaking (how text expresses strengths/weaknesses).

Substantive Question:
Who are the informal opinion leaders in Sephora's review ecosystem, and how do they shape collective product evaluations?

Analytics Questions:

- Can informal opinion leaders be identified using network position in a reviewer–product network (for example, using centrality, brokerage)?
- Do influential reviewers systematically use different language (for example, higher confidence, stronger evaluative tone, broader topical scope)?
- Are products reviewed by influential reviewers more likely to experience subsequent changes in ratings or rating volatility?

Data accessibility plan:

- Kaggle Sephora Products and Skincare Reviews (≈8k products and ~1M skincare reviews, multiple CSV files)

Technical approach:

1. **Network construction (structuring)**
    1.1. Construct a bipartite reviewer–product network, where:
        1.1.1. Nodes represent reviewers and products

1.1.2. Edges represent review events

1.2. Derive reviewer-level network metrics:

1.2.1. Degree centrality (number of products reviewed)

1.2.2. Betweenness centrality (reviewers bridging otherwise weakly connected product clusters)

1.2.3. Product diversity index (ex: entropy or category spread of products reviewed)

This allows identification of informal opinion leaders based not only on activity level, but on breadth of exposure and brokerage across product spaces.

2. Reviewer portfolio analysis (boundary spanning)

2.1. Classify influential reviewers as:

2.1.1. Specialists (focused on a narrow product/category set)

2.1.2. Generalists / boundary spanners (reviewing across multiple categories and brands)

2.2. Examine whether influence is associated with specialization in high-visibility products or with broader exploratory behavior.

3. Linguistic sensemaking analysis (NLP)

3.1. Preprocess review text (cleaning, lemmatization, stopword removal).

3.2. Extract linguistic features including:

3.2.1. Sentiment intensity (strength rather than polarity alone)

3.2.2. Linguistic certainty vs hedging language

3.2.3. Topical diversity using topic modeling

3.3. Compare language patterns between influential and non-influential reviewers.

4. Linking influence to collective outcomes (modeling)

4.1. Model downstream product outcomes (ex: changes in average rating or rating volatility) as a function of:

4.1.1. Exposure to influential reviewers

4.1.2. Reviewer portfolio diversity

4.1.3. Linguistic characteristics

4.2. Use regression models with controls for product category, brand, price, and review volume.

Value:

- This approach reveals how influence emerges, not just who is influential.
- Results enable organizations to:
    - Distinguish between narrow but loud voices and broad, credible evaluators
    - Identify which feedback signals should be prioritized for decision-making
    - Detect products that are particularly sensitive to influential reviewers.
