5 Topic 4: Clustering-Based Recommender System

Standard recommender systems assume a continuous, convex latent space where any linear interpolation between User A and User B is a valid theoretical state.

Hypothesis: The state space is discontinuous; "Customer Identity" (Static) and "Behavior" (Temporal) are governed by strict sociological and market rules.

5.1 Ghost Archetypes, Structural Voids & directional asymmetry

We model the system as a collection of observed local manifolds (dense clusters). Transitions are valid only if they traverse regions of high local density. Low-density regions represent **Structural Voids** (impossible states). Traditional linear averaging creates "Ghost Archetypes":

Table 1: Linear Interpolation vs. Reality

Metric	Profile A	Linear Avg	Profile B
Type	Solo Consultant	*Ghost*	Summer Family
Comp.	0 (Alone)	1.5 People	3 (Spouse+Kids)
Timing	Tue/Wed	Oct Thu	July/Aug
Stay	1 Night	7.5 Nights	14 Nights

Furthermore, transitions on the manifold are subject to **Asymmetry Constraints**. While a Solo Consultant may transition to a Family profile (cross-sell leisure), a Summer Family traveler cannot be forced into a Consultant profile via promotion.

5.2 Geodesics represent causality and natural lifecycles

Euclidean distance ignores covariance. In airline data, High Spend is often inextricably coupled with High Frequency (tier bonuses). We utilize **Geodesics**—locally length-minimizing curves constrained to the manifold surface. Geodesics respect causality: to move from Start to Aurora status, the path may force an increase in Frequency (Cluster A $\rightarrow$ B) before Yield can increase (Cluster B $\rightarrow$ C).

5.3 Customer Physics: Energy & Inertia

We treat the customer as a particle with mass (inertia) moving through a vector field (trends).

- **Natural Trajectory:** If we do nothing, the customer follows the flow of their current cluster.
- **Marketing Energy:** To change state, we apply an Exogenous Force.

The total energy E required to traverse a path is defined by velocity change, void resistance, and directional change:

$$E_{total} = \sum_{t=0}^{T} \left(\|\Delta \mathbf{v}\|^2 + \frac{\lambda}{\rho(z_t)} + \theta(1 - \cos\phi) \right) \tag{1}$$

Where:

- $\|\Delta \mathbf{v}\|^2$: Kinetic energy required to accelerate (e.g., fly more).
- $\rho(z_t)$: Local density. Low density (voids) creates infinite resistance ($1/\rho$), making "unnatural" behavior prohibitively expensive to incentivize.
- $\theta(1 - \cos\phi)$: Directional work. Changing customer type is harder than accelerating current behavior.

5.4 Pathing solution: Dijkstra on the k-NN Graph bound by clusters

We solve for the Energy Minimization Path using Dijkstra's Algorithm on the k-NN graph. Unlike a Geodesic (shortest vector path), this method identifies the "Golden Path" of intermediate clusters where marketing resistance is lowest, guiding customers through natural lifecycles rather than forcing impossible jumps.

6 Case Study: converting to Aurora

6.1 Hyperpersonalized marketing

Figure 1: Visualization of the local knn manifold in 3D.

The white cone represents the customer's position and their Momentum Vector (Velocity). The cone points in the direction of their natural inertia. This path (Target 2/Cyan) is the Geodesic. It is short, direct, and crucially, it follows the dense "wireframe" of the data.

The path traverses a region of relatively low density. We are pushing the customer into a "less traveled" behavioral segment. This represents Uncertainty Risk (in red).

The spider graph shows the difference between current user trajectory and the trajectory towards each Aurora path. It's represents the **exogenous marketing vector**.

Recommendation: The system has identified that the "Path of Least Resistance" to high value is Cross-Selling. The customer is likely a frequent solo flyer, and the algorithm suggests converting them into a Leisure/Companion Traveler.

6.2 Conversion cost analysis

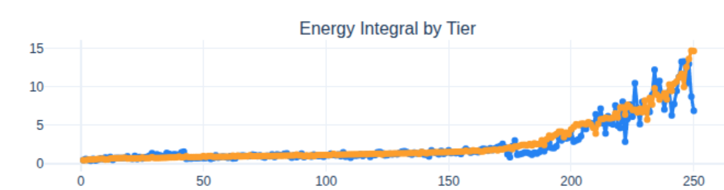

Figure 2: Cumulative energy to Aurora conversion. Yellow Star, Blue Nova. (sample size 250pts)

It visualizes the **Marginal Cost of Conversion**. For the top 70%, the effort required to upgrade them is negligible. These customers are effectively **Aurora tier in disguise**.

The Blue line (Nova) and Orange line (Star) track almost perfectly together. This means Nova is likely a default tier with multi modal distributions. We need to leverage this and convert these low hanging Novas into Star or Aurora or even build an intermediate transitional tier between Nova and Star.

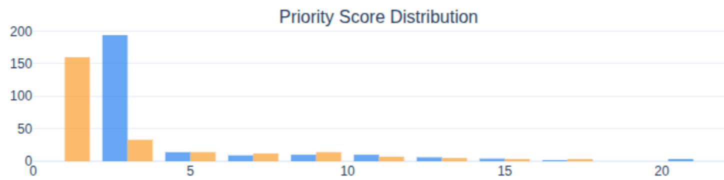

Figure 3: Prioritization of customers for marketing purposes

It's clear Star is the **natural path to Aurora**. These are the customers that we can convert with minimum cost. Then we have a selection of Nova customers, which according to the data, they are wrongly classified.

6.3 Conversion risk analysis

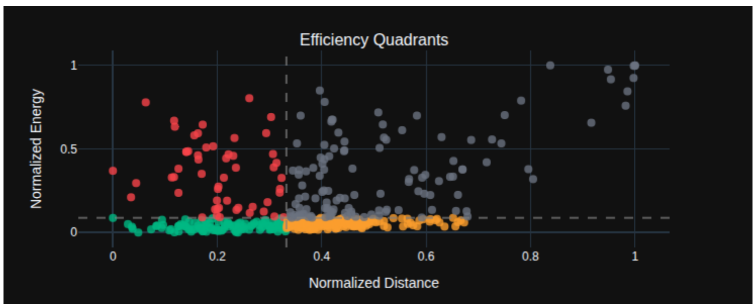

Figure 4: Geometric distance vs Energy conversion distance (sample size = 600pts)

This is one of the most interesting analysis we should make. What about those customers who are geometrically close to Aurora but face high energy barriers for behavioral change?

We defined 4 categories of customers using a risk-effort based approach:

- **Easy Wins** (35th percentile distance, 65th percentile energy): Priority targets for campaigns (green) - 23%
- **Close but Hard to convert** (25th percentile distance, 75th percentile energy): Unfeasible due to behavior barriers. These are **traditional clustering traps**, they look Aurora but their momentum is wrong. (red) - 12%
- **Far but Easy**: Good long-term targets with lower resistance (orange) - 25%

7 Conclusion

1. **Segment-Specific Engines:** By utilizing local manifold geometry rather than global averages, we created bespoke recommendation pathways that respect the distinct inertia of specific profiles (e.g., Business vs. Leisure).
2. **Cluster-Based Collaborative Filtering:** The deployment of Dijkstra's algorithm on density-weighted k-NN graphs ensures recommendations are grounded in the aggregate historical behavior of peer groups (collaborative density) rather than theoretical interpolation.
3. **Cross-Segment Analysis:** We modeled inter-segment dynamics using an Energy function, effectively filtering out "Ghost Archetypes" and identifying valid transition corridors between tiers (e.g., Nova to Aurora).
4. **Business Impact Evaluation:** The methodology translates abstract geometric distances into actionable "Marketing Energy" metrics. By categorizing customers into "Easy Wins" versus "High Energy Traps," we provide a direct mechanism to optimize marketing ROI and prioritize high-conversion targets.