

Conversation Cartographer: Mapping Relational Trajectories in Human–AI Dialogue

Abstract

Human–AI conversations are commonly analyzed as transcripts or aggregate metrics, obscuring relational dynamics that unfold over time. We present Conversation Cartographer, a visual system that renders dialogue as trajectories through a three-dimensional relational space defined by functional–social orientation, linguistic alignment, and affective intensity. Each conversational turn is encoded as a coordinate and connected into continuous paths, revealing drift, volatility, and breakdowns that are invisible in conversation-level labels. We analyze 562 validated conversations drawn from Chatbot Arena, WildChat, OASST, and a small set of human–human dialogues.

Across hierarchical and K-means clustering, trajectory features (spatial movement + affective dynamics) account for 70–83% of cluster separation (83.3% in hierarchical).

Conversations with identical role classifications differ by 41×–82× in affective variance, demonstrating “same destination, different journeys.”

The pictorial contribution is a set of comparative terrain visualizations that make these temporal dynamics legible for evaluation and design.



Motivation

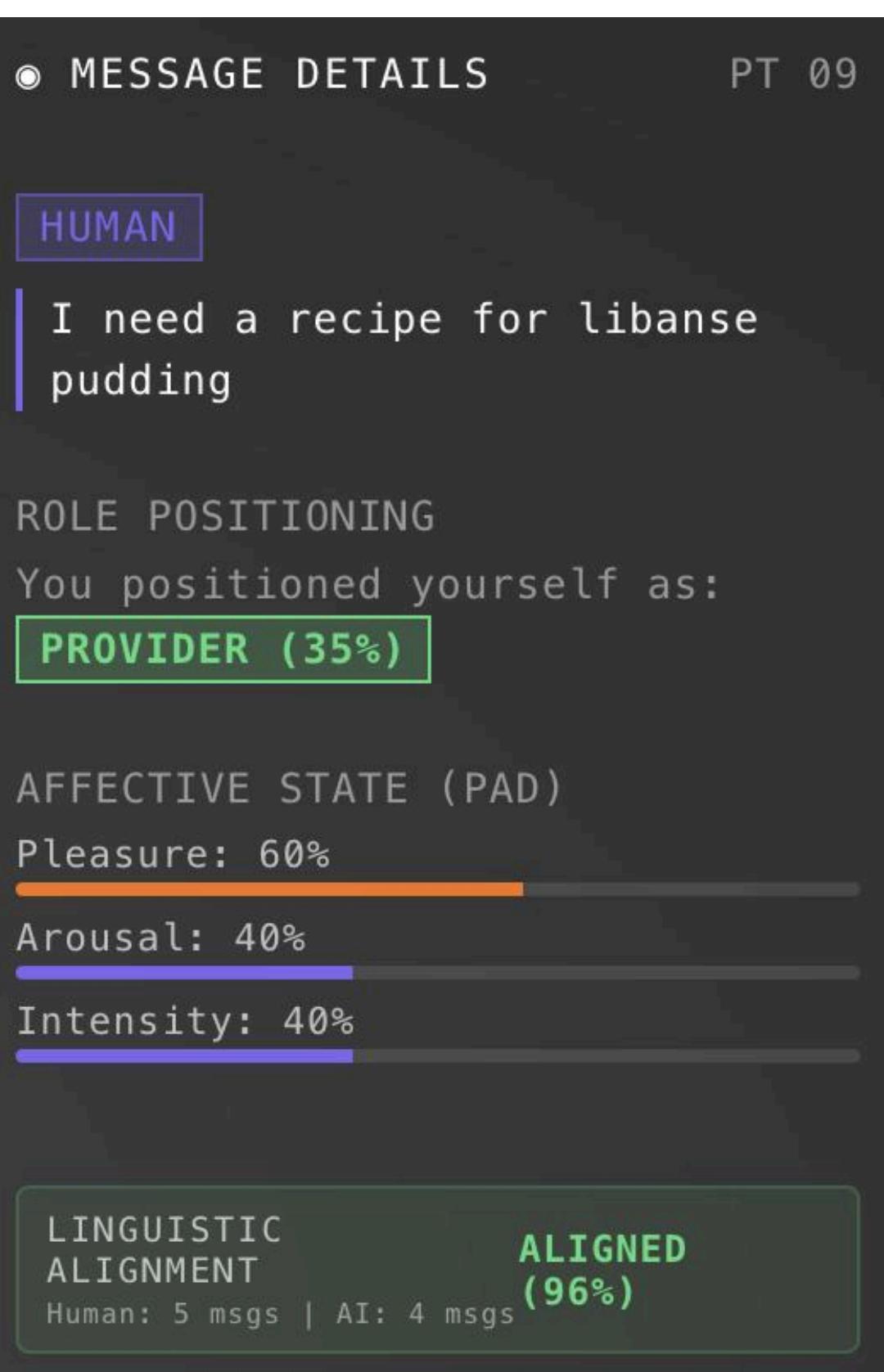
Traditional dialogue analysis treats conversation as text or point summaries. These representations are effective for content inspection, but they obscure how interaction unfolds: whether the exchange stabilizes, escalates, meanders, or ruptures.

Conversation Cartographer reframes dialogue as movement through relational-affective space, allowing us to read interaction quality as trajectory shape rather than only topic or task outcome.



[FIGURE 1: Terrain Overview vs. Transcript]

Visual suggestion: Left: raw transcript / JSON log. Right: the same exchange rendered as a spatial trajectory.
Caption: From text to terrain: transforming linear conversation logs into relational journeys.

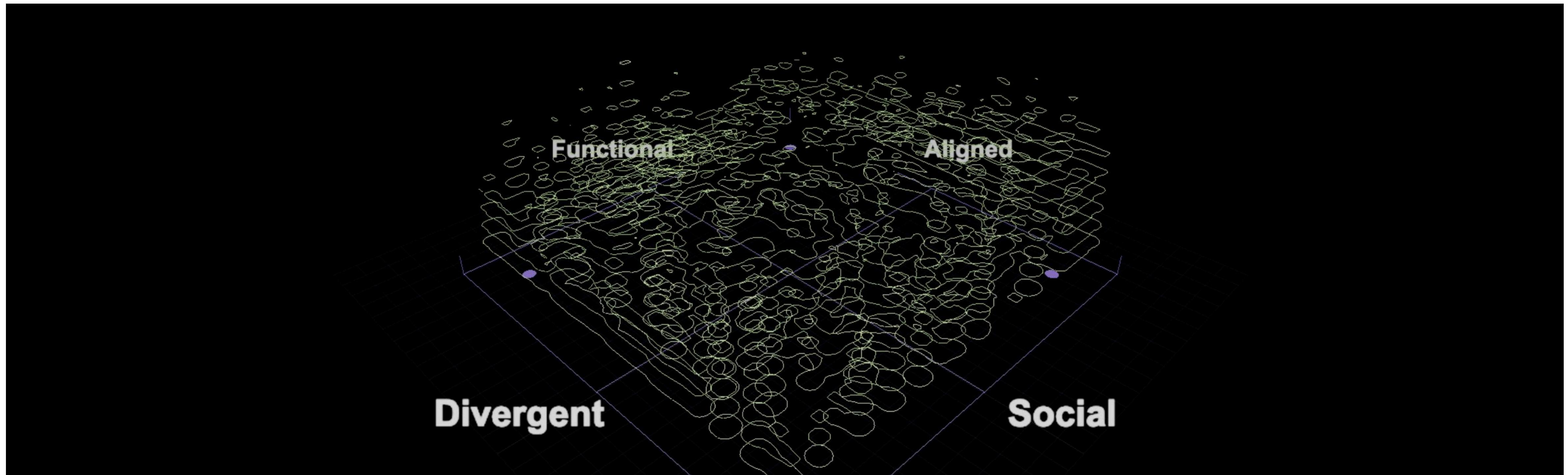


Spatial Encoding Model

We construct a three-dimensional relational space:

- Functional \leftrightarrow Social Orientation (X)
Degree to which interaction is task-driven versus socially oriented, derived from role distributions.
- Linguistic Alignment (Y)
Structural and stylistic convergence between participants.
- Affective Intensity (Z)
Emotional prominence encoded as height using a PAD-derived metric.

Together, these dimensions form a continuous coordinate system in which conversations can be situated and compared.



Patterns & Landscapes

We analyzed 562 validated conversations across multiple sources. Clustering reveals 7–9 recurring positioning patterns (depending on granularity), but the boundaries are continuous rather than discrete.

Across hierarchical clustering, trajectory features dominate separation:

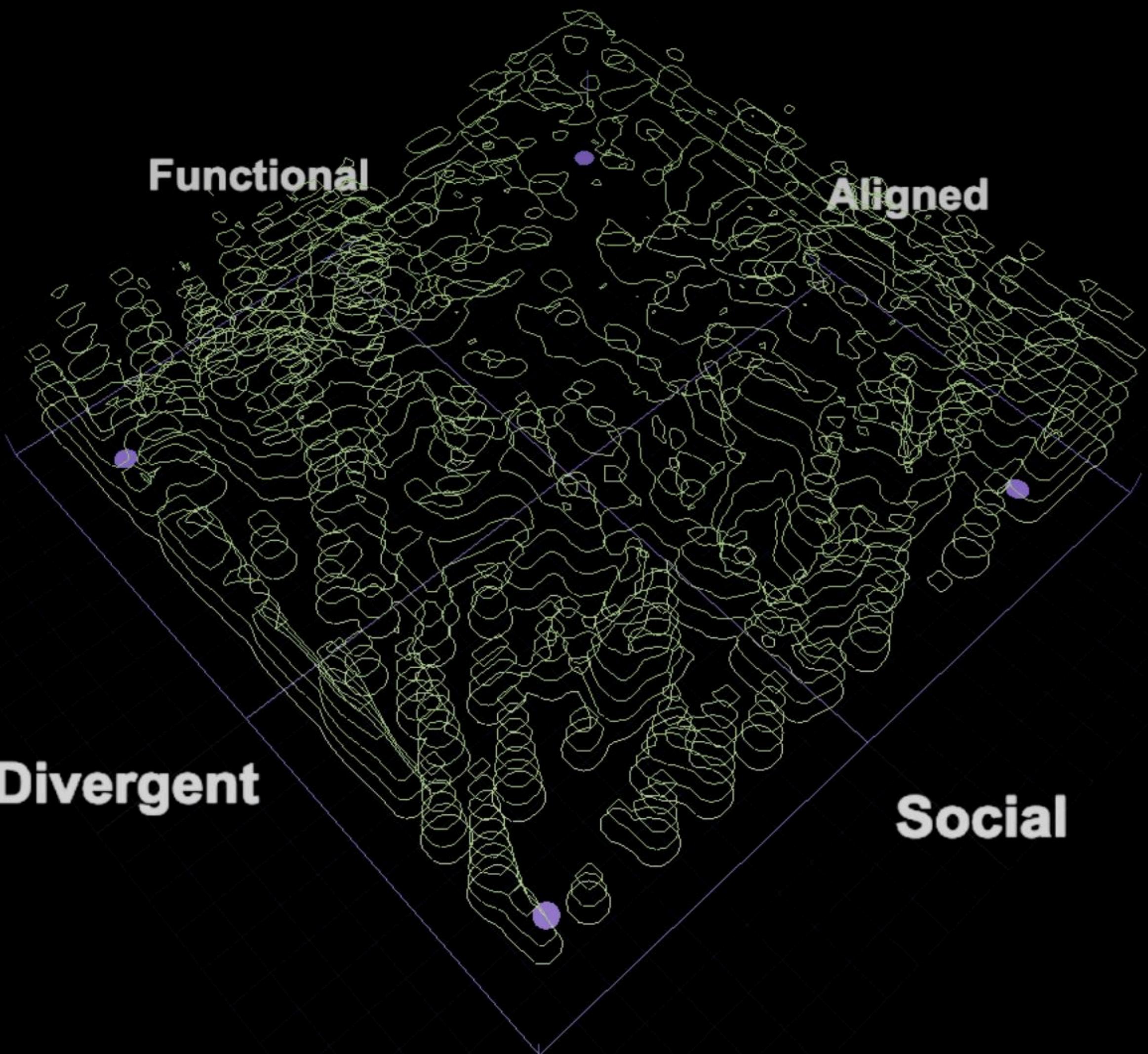
Spatial trajectory: 50.4%

Affective dynamics: 32.9%

Total trajectory contribution: 83.3%

All categorical labels combined: 16.7%

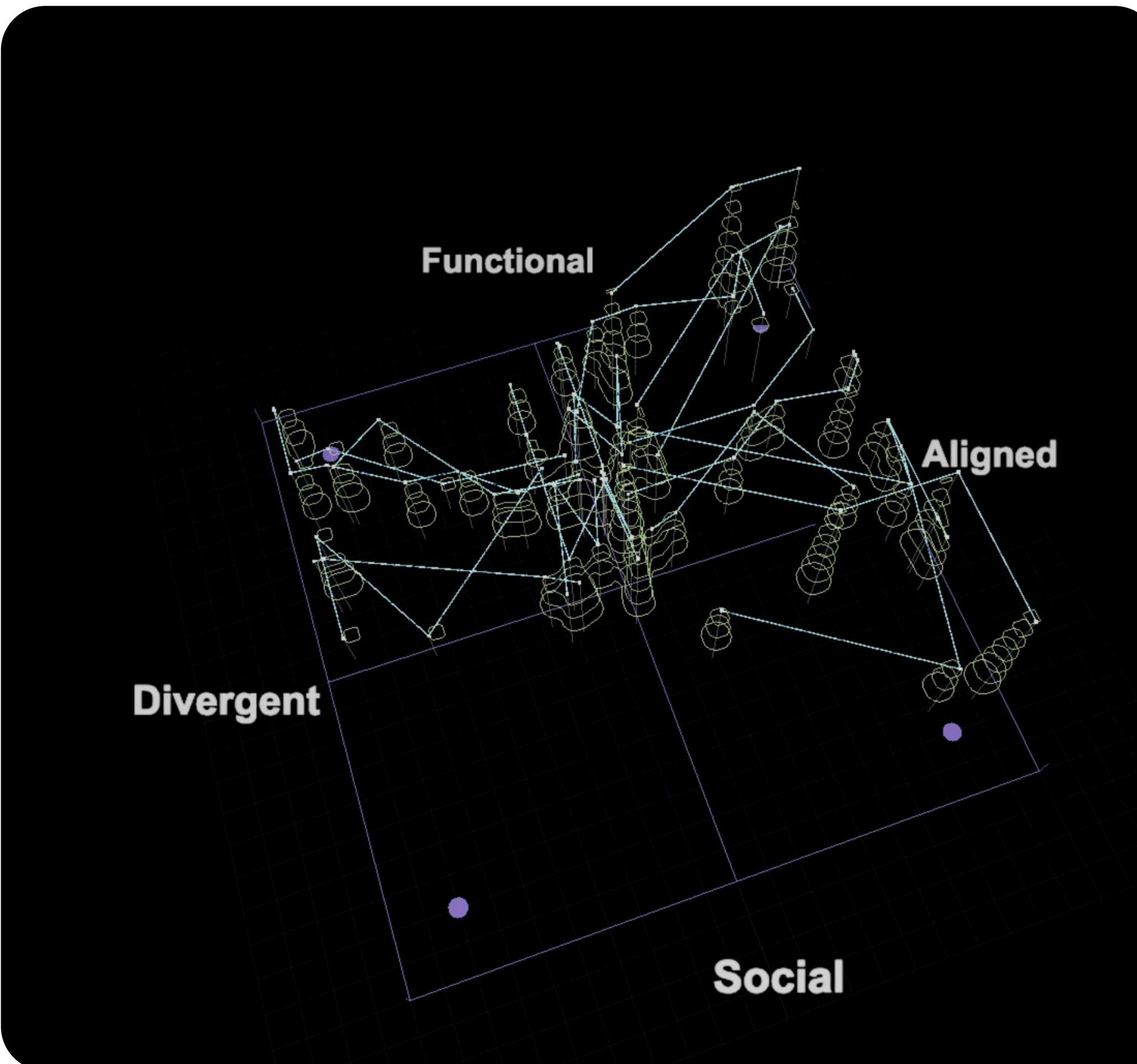
Cluster validity metrics remain modest (silhouette 0.198), indicating genuine structure but fuzzy transitions between patterns—consistent with a continuous terrain rather than stable “types.”

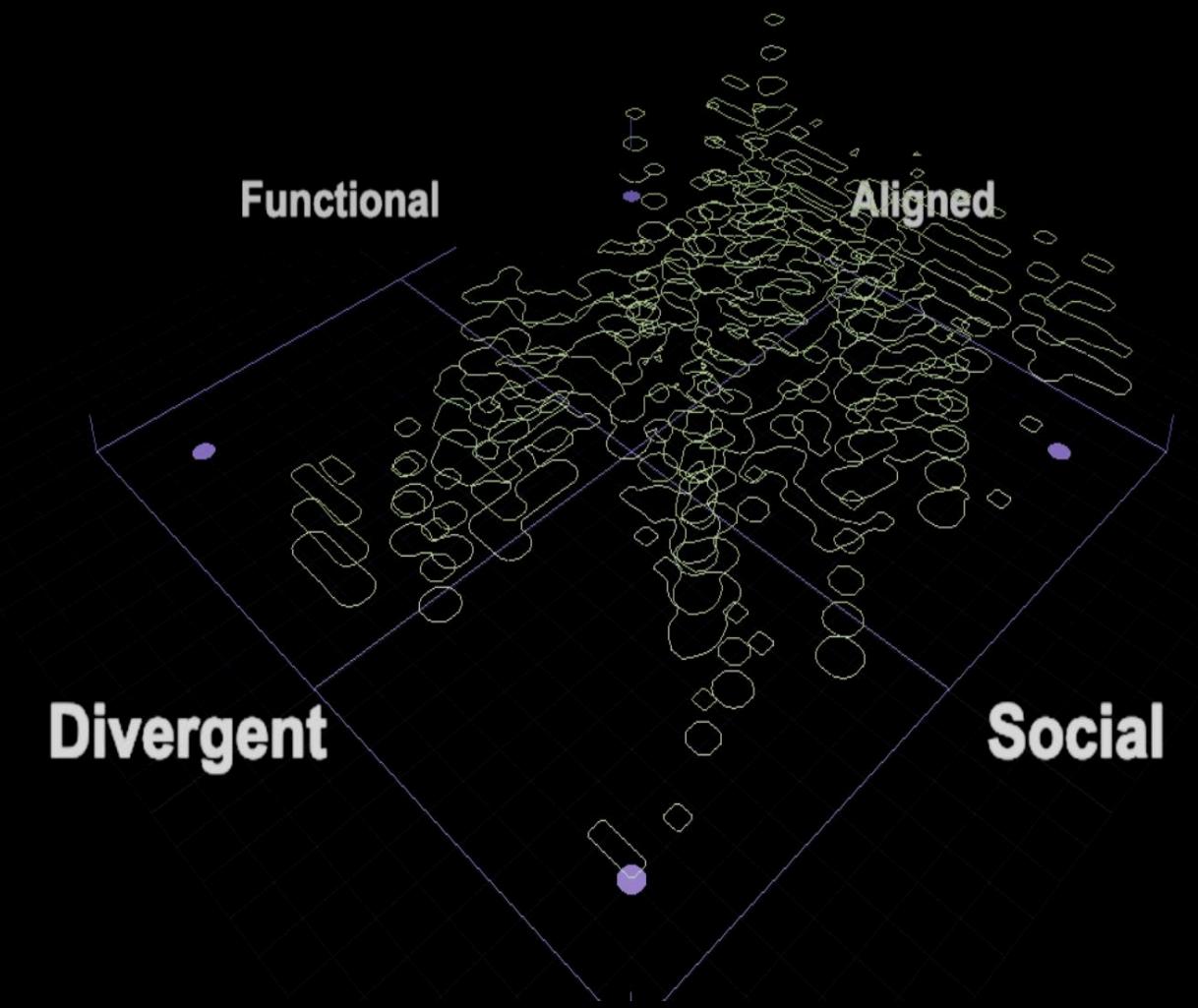


Conversation Trajectories

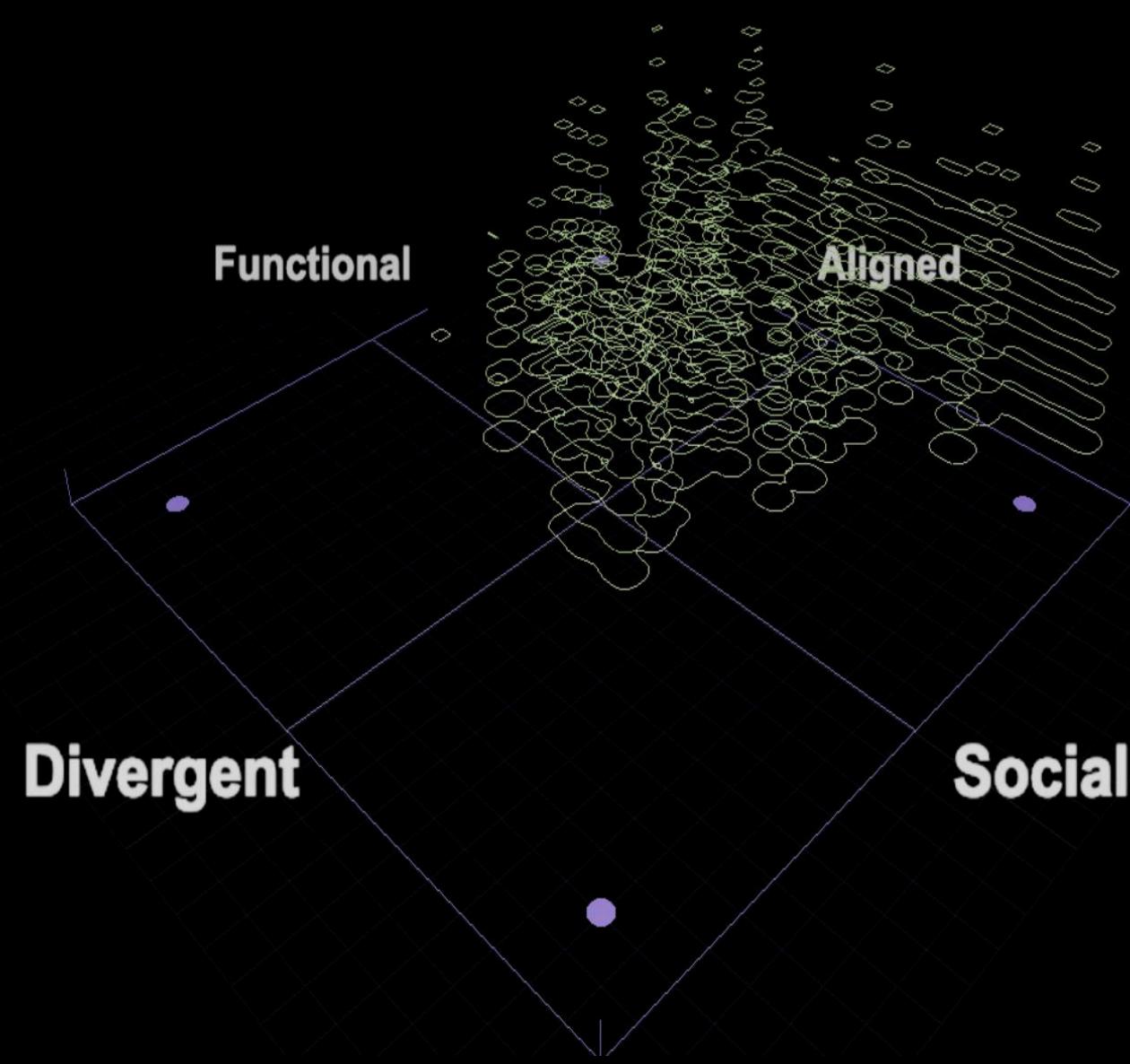
Each message is mapped to a point in relational space. Sequences of messages form continuous polylines—conversational trajectories.

These trajectories expose properties that transcripts cannot:
directional drift, volatility, emotional escalation, stabilization, and rupture.

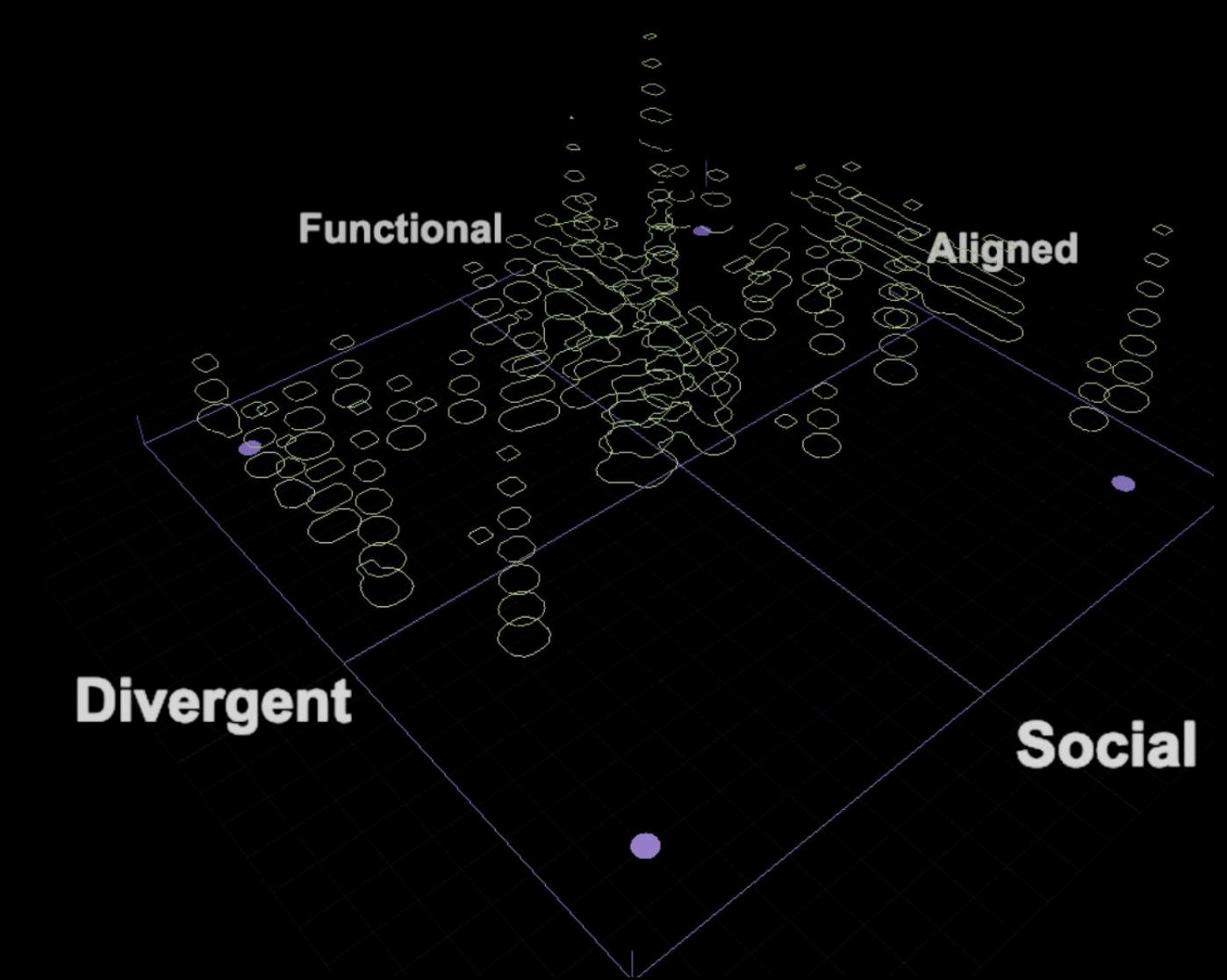




OASST

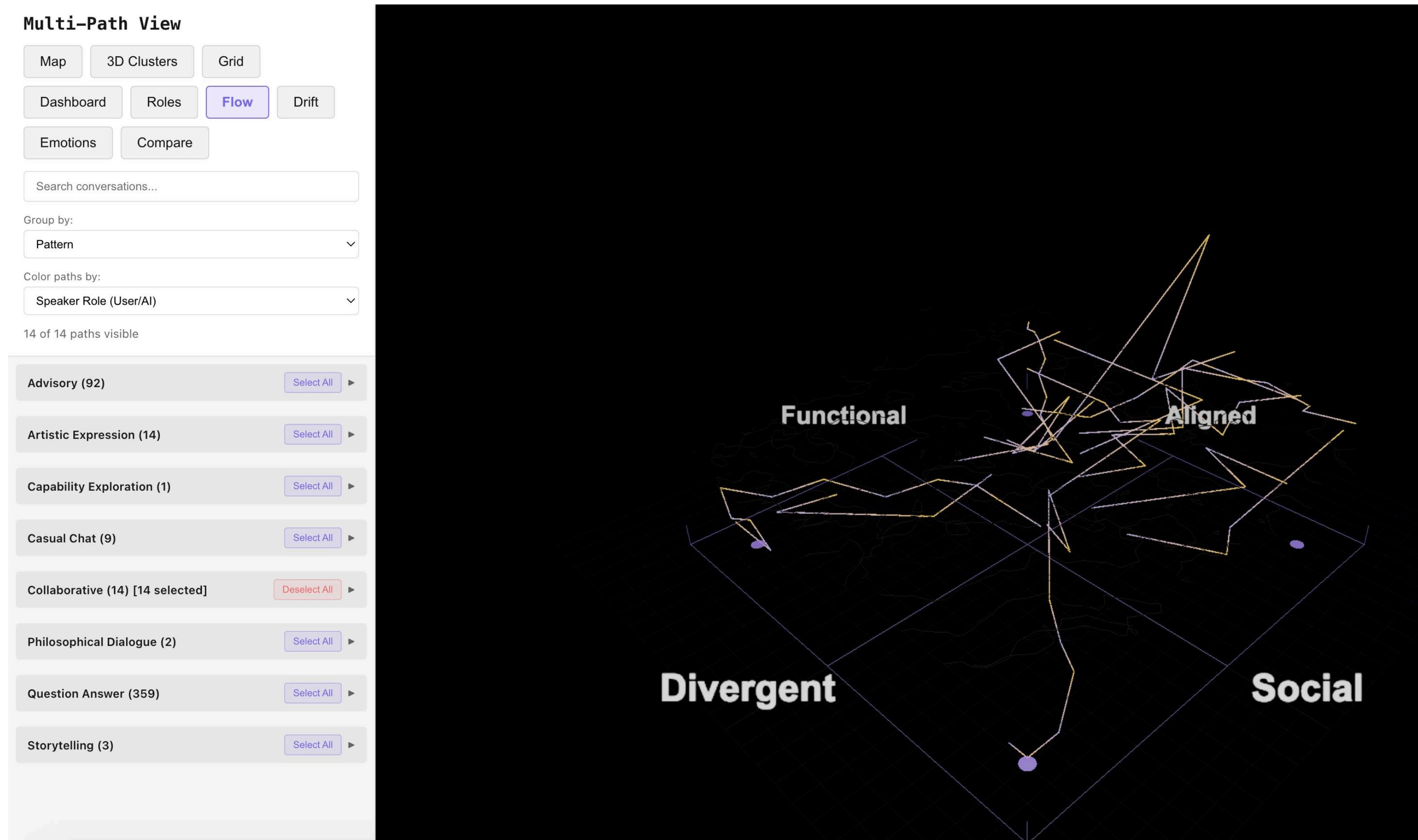


WILDCAT

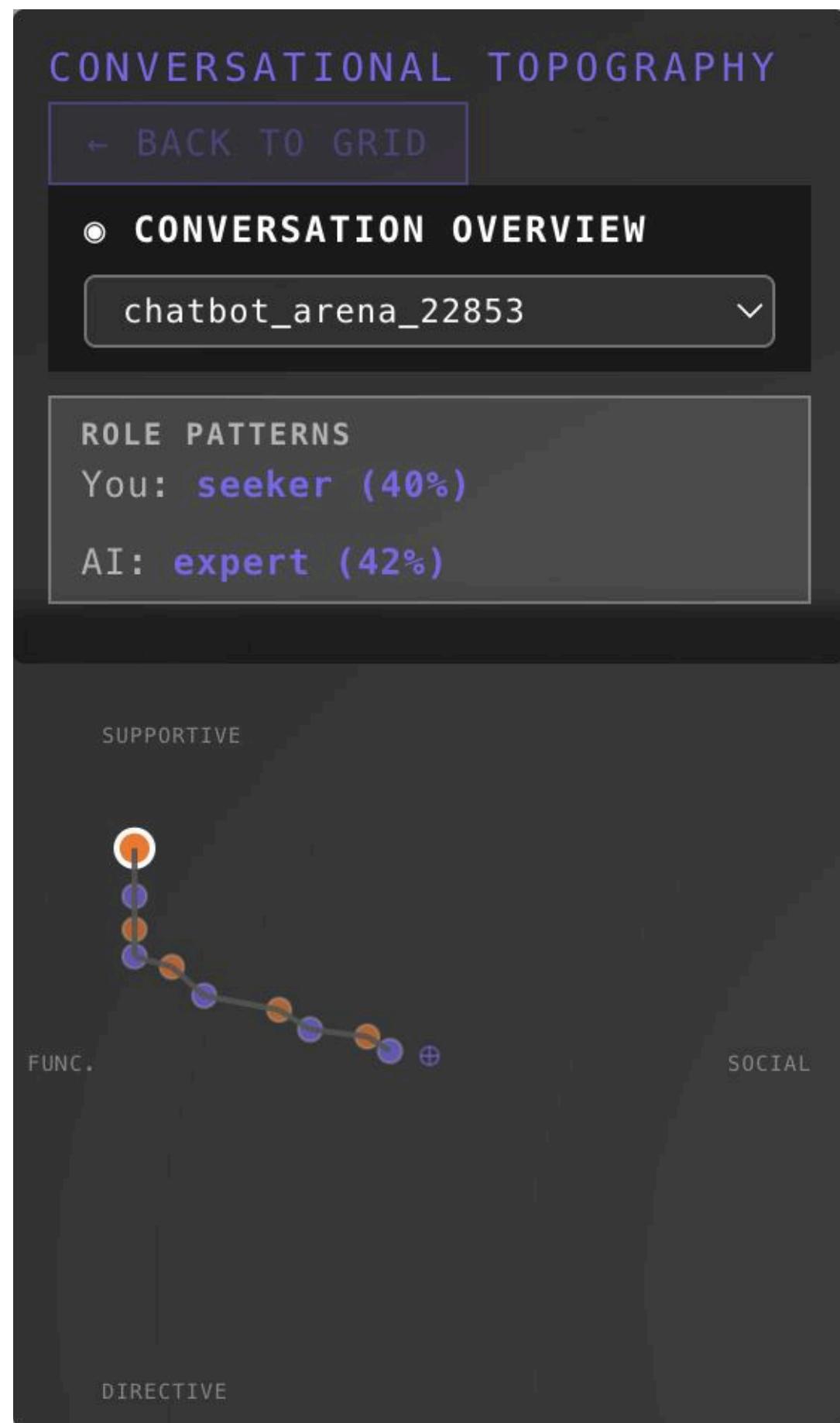


Chatbot Arena

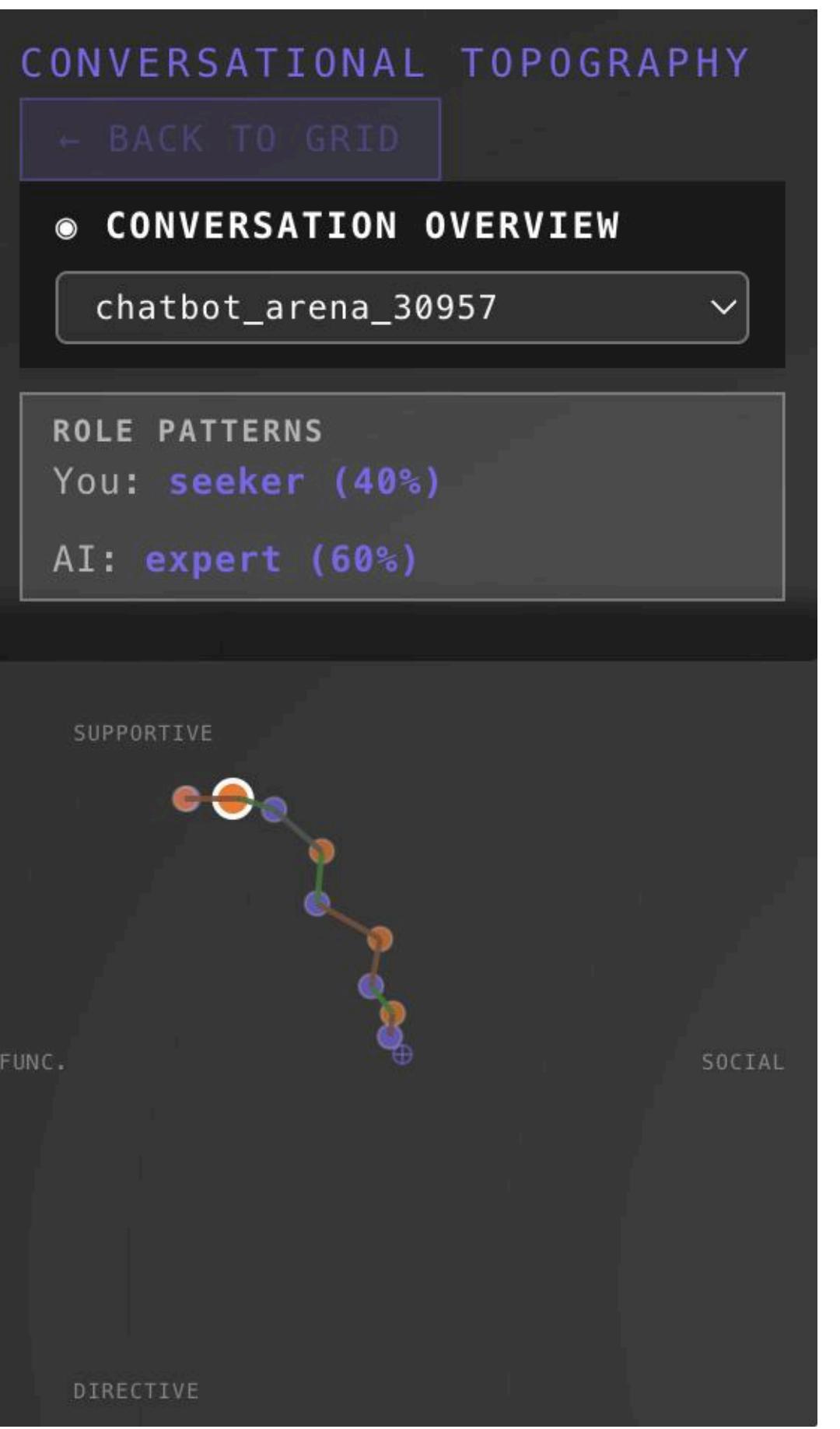
Same Roles, Different Journeys



Conversations with identical role classifications frequently diverge in affective shape. In multiple validated pairs labeled Information-Seeker → Expert-System, affective variance differs by 41× to 82×, despite similar purpose, pattern, and endpoint quadrant.



Flat journey (variance 0.0004)
Disengaged, topic-hopping
Cluster:
StraightPath_FunctionalStructured_QA_Inf
oSeeking



Adversarial evaluation:
chatbot_arena_30957
Volatile journey (variance 0.0164)
Engaged, trap-setting, sarcasm
Cluster:
Peak_Volatile_FunctionalStructured_QA_Inf
oSeeking

Defining Zones or areas that the conversations lead towards

Every conversation starts out from 0,0 and leads towards different zones

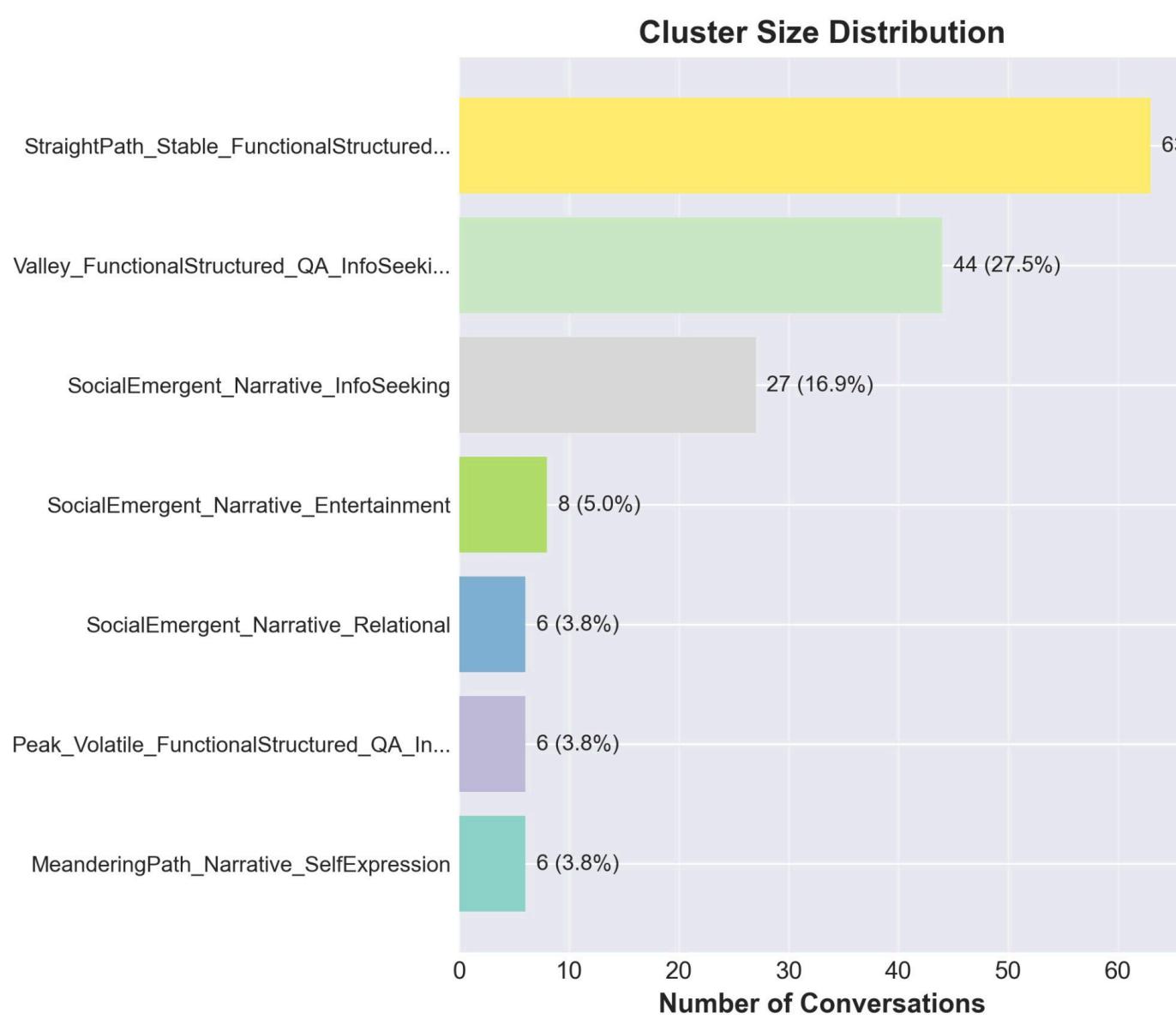
Zones (characteristic regions)

Tool Basin: Human-dominant, instrumental, critical

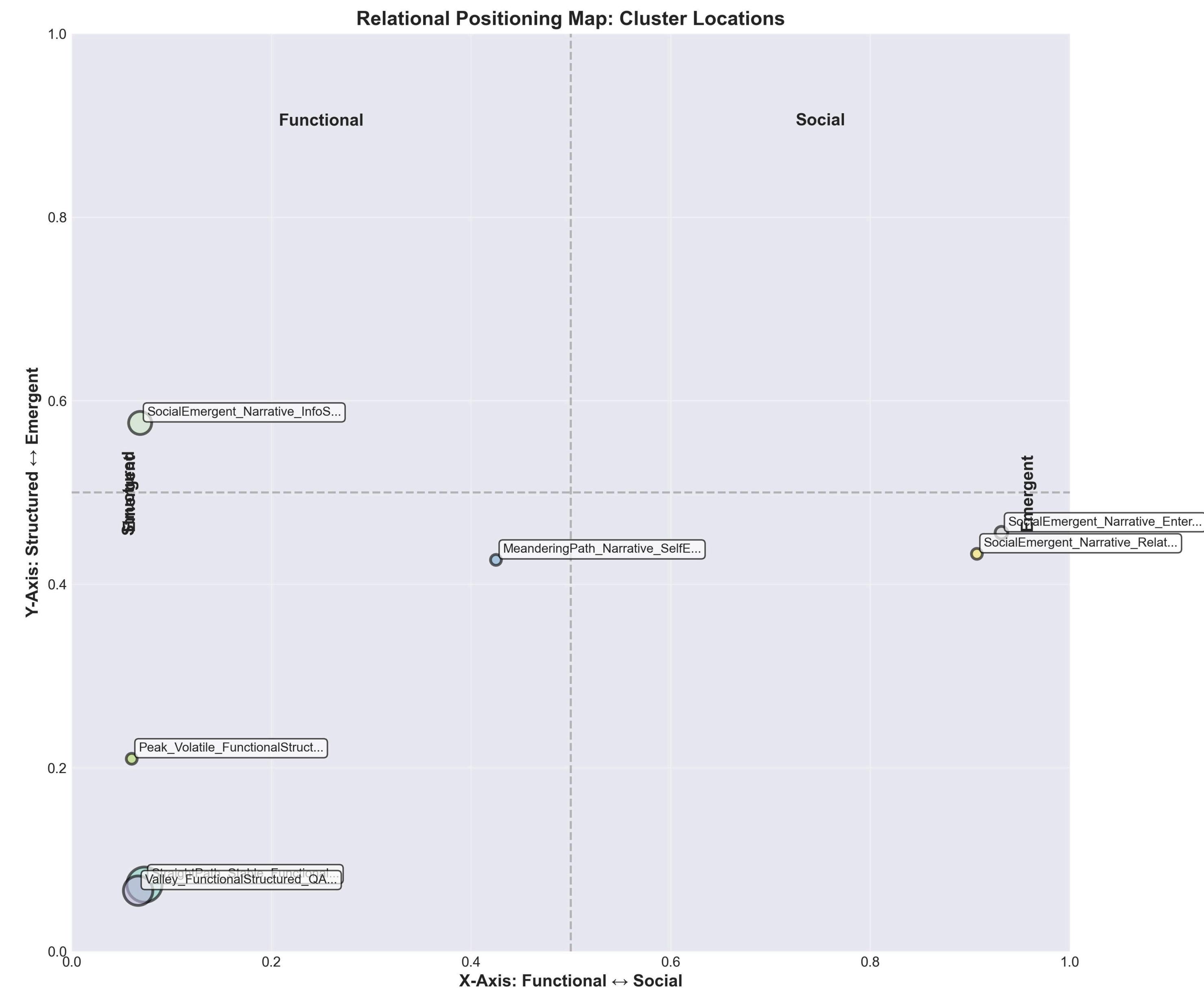
Collaborator Ridge: Shared agency, moderate intimacy

Companion Delta: Intimate, trusting, ongoing

Evaluator Heights: Testing, verification, skepticism



Soc



Dataset Bias as Visible Terrain (use your 72% + "missing territory")

Across the corpus, ~72% of conversations occupy the Functional/Aligned quadrant, reflecting the dominance of instrumental interaction norms in contemporary datasets and evaluation settings. Social-emergent territory is comparatively sparse, suggesting a narrow “relational corridor” that current systems and benchmarks repeatedly reproduce.

Rather than treating this skew as noise, the terrain makes it legible as a socio-technical condition: how datasets, evaluation practices, and interface expectations shape the kinds of relationships that are possible.

