Preference Fields on Semantic Manifolds

Directional Constraints, Spectral Admissibility, and Random Geometric Graph Structure

Flyxion

June 2026

Abstract

We develop a geometric theory of preference-guided navigation over corpora represented as high-dimensional random geometric graphs. A user's evolving preference is modeled as a time-dependent unit vector Φ_t in a semantic embedding space $\mathcal{E} = \mathbb{R}^d$, inducing a scalar preference field $f_t(v) = \langle \Phi_t, v \rangle$ over the corpus manifold $\mathcal{M} \subset S^{d-1}$. The admissible neighborhood $A_t(\theta) = \{i : f_t(v_i) \geq \theta\}$ forms the primary object of study: the set of corpus items accessible to the current preference state. We establish that this neighborhood has the structure of a subgraph of a spherical random geometric graph, with clique probabilities governed by the triangle-penalty corrections of Hunter, Milojević, and Sudakov. We identify the anisotropic centroid bias of pretrained embeddings as an inadmissible subspace whose removal via spectral filtering sharpens the discriminative geometry of the preference field. The central result is a dimensionless order parameter $\Lambda_t(\theta) = n_\theta^3 p_\theta^3 / d$ governing a phase transition: below the critical threshold the admissible neighborhood is statistically indistinguishable from an Erdős–Rényi random graph and geometric structure is undetectable; above it, the spherical geometry is present and the preference field resolves genuine semantic clusters. This order parameter unifies the field dynamics, spectral admissibility, and graph-theoretic clique structure into a single geometric framework applicable to literature exploration, memory retrieval, recommendation, and semantic search.

Contents

1. INTRODUCTION

Many information systems can be understood as the problem of navigating a high-dimensional semantic manifold under evolving constraints. Classical approaches emphasize retrieval, ranking, and optimization against a fixed query. The present work examines a structurally different setting: navigation driven by a *preference field* that evolves in response to the navigator's encounters with the manifold itself. The query is not given in advance; it is constructed through interaction, and the geometry of the space shapes both what is accessible and what the evolving field can discriminate.

This setting arises whenever users explore a corpus whose internal structure is represented by embeddings in a high-dimensional metric space. When preference is encoded as a direction Φ_t in that space, evaluating items by $\langle \Phi_t, v_i \rangle$ is not a retrieval operation but a *field evaluation*: a measurement of how well each point in the corpus aligns with the current constraint. The natural questions then become geometric. How does the field evolve? What is the structure of the region it selects? How does the ambient geometry of the embedding space govern the possible shapes of coherent preference neighborhoods? When does the geometry become detectable and meaningful?

We address these questions by developing three interlocking theoretical structures. The first is a formal framework for preference fields as directional constraints over manifolds, including a precise account of how ratings update the field and how the field induces a family of admissible neighborhoods (Section 2). The second is a random geometric graph characterization of the high-alignment sub-corpus, drawing on recent results in Ramsey theory to bound the coherent cluster size and identify the triangle density as a diagnostic of geometric regime (Sections 3 and 5). The third is a spectral admissibility analysis of the embedding geometry itself, connecting the anisotropy of pretrained embeddings to a natural notion of inadmissible subspace and characterizing spectral filtering as admissibility enforcement (Section 4).

The paper is organized to be progressively more abstract. Early sections establish concrete objects; later sections develop the connections between them. We assume throughout that $\mathcal{E} = \mathbb{R}^d$ with d large (specifically $d = 512$ in the CLIP ViT-B/32 setting, but most results hold for general d) and that items are represented as unit vectors $v_i \in S^{d-1}$. Section 6 places the framework in the context of related field theories; no familiarity with those frameworks is required for the main results.

2. PREFERENCE FIELDS AS DIRECTIONAL CONSTRAINTS

2.1. *Semantic Embedding and the Corpus Manifold*

Let $\mathcal{E} = \mathbb{R}^d$ be a semantic embedding space equipped with the standard inner product. A *corpus* $\mathcal{C} = \{c_1, \ldots, c_N\}$ is a finite collection of items, each associated with a unit vector $v_i \in S^{d-1}$ via an embedding map $\phi : \mathcal{C} \to S^{d-1}$.

Definition 2.1 (Corpus manifold)**.** The *corpus manifold* $\mathcal{M} \subset S^{d-1}$ is the image $\phi(\mathcal{C})$. We write $\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i$ for the corpus centroid (not necessarily of unit norm) and $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\|v_i - \bar{v}\|^2$ for the corpus dispersion.

The embedding space for pretrained vision-language models (CLIP) and language models approximates a semantic manifold: inner products $\langle v_i, v_j \rangle$ correlate with human-perceived similarity between items c_i and c_j [1]. This grounding is what makes the preference field semantically meaningful; without it the framework reduces to arbitrary geometry.

2.2. *The Preference Field*

Definition 2.2 (Preference field)**.** A *preference field* at time t is a unit vector $\Phi_t \in S^{d-1}$. It induces a scalar alignment function

$$f_t : S^{d-1} \to [-1, 1], \qquad f_t(v) = \langle \Phi_t, v \rangle, \tag{1}$$

and its restriction to the corpus manifold

$$f_t \restriction_{\mathcal{M}} : \mathcal{M} \to [-1, 1], \qquad f_t(v_i) = \langle \Phi_t, v_i \rangle \tag{2}$$

is the *discrete preference field* over $\mathcal{C}$.

The discrete preference field is not a ranking: it is the trace of a continuous linear functional on S^{d-1} evaluated at the corpus sample points. This distinction matters. A ranking selects a permutation of $\mathcal{C}$ with no structure beyond order. The preference field selects a *direction* in $\mathcal{E}$, and that direction generalizes continuously to any item whose embedding can be computed, whether or not it appears in $\mathcal{C}$.

2.3. *Field Dynamics Under Preference Accumulation*

Let $\mathcal{R}_t = \{(i, r_i)\}$ be the set of rated items at time t, where $r_i \in \mathbb{R}$ is a scalar rating with $\mathrm{sgn}(r_i)$ indicating positive or negative preference. Define the weighted sum

$$\tilde{\Phi}_t = \sum_{(i,r_i)\in\mathcal{R}_t} w(r_i)\, v_i, \qquad w(r) = \operatorname{sgn}(r)\cdot(1+\kappa|r|), \tag{3}$$

for a sensitivity parameter $\kappa > 0$. When $\tilde{\Phi}_t \neq 0$, the updated preference field is

$$\Phi_t = \frac{\tilde{\Phi}_t}{\|\tilde{\Phi}_t\|}. \tag{4}$$

Proposition 2.1 (Monotone convergence under consistent preference)**.** *Suppose all ratings in $\mathcal{R}_t$ are positive ($r_i > 0$ for all i) and the rated items lie within a geodesic ball $B_\rho(\Phi^*) \subset S^{d-1}$ of radius $\rho < \pi/2$ around some direction $\Phi^* \in S^{d-1}$. Then $\langle \Phi_t, \Phi^* \rangle \geq \cos\rho > 0$ and $\langle \Phi_t, \Phi^* \rangle$ is non-decreasing as new positively-rated items are added from $B_\rho(\Phi^*)$.*

Proof. The normalized sum of vectors within a geodesic ball of radius $\rho < \pi/2$ has positive inner product with any vector in the ball. Adding a new vector $v_j \in B_\rho(\Phi^*)$ to $\tilde{\Phi}_t$ moves the sum further into the half-space $\{v : \langle v, \Phi^* \rangle > 0\}$, since $\langle v_j, \Phi^* \rangle > \cos\rho > 0$. Normalization preserves the sign and the bound follows from the convexity of geodesic balls on S^{d-1}. □

2.4. Admissible Neighborhoods

Definition 2.3 (Admissible neighborhood)**.** For threshold $\theta \in [-1, 1]$, the *admissible neighborhood* of the field Φ_t at level θ is

$$A_t(\theta) = \{i \in \mathcal{C} : f_t(v_i) \geq \theta\}. \tag{5}$$

The *inadmissible set* at level θ is $\mathcal{C} \setminus A_t(\theta)$.

As θ increases, $A_t(\theta)$ shrinks: only items tightly aligned with the current field direction remain admissible. The trajectory of $A_t(\theta)$ as t increases characterizes how navigation concentrates the preference field. In an embedding space with good semantic structure, $A_t(\theta)$ for large θ should correspond to a semantically coherent neighborhood of $\phi^{-1}(\Phi_t)$—the "concept nearest to the current field direction."

Remark 2.1. In practice, θ is not set explicitly but is induced by a softmax sampling distribution with temperature $T > 0$:

$$\mathcal{P}_t(i) \propto \exp\left(\frac{s_t(i)}{T}\right), \tag{6}$$

where $s_t(i) = \alpha f_t(v_i) +$ (auxiliary terms) and $\alpha > 0$ bounds the alignment contribution. Items in $A_t(\theta)$ receive disproportionate sampling probability, inducing a soft version of the hard thresholding above.

3. Random Geometric Graph Structure of the Corpus

3.1. *The Spherical RGG Model*

We now establish the formal relationship between the preference field threshold structure and spherical random geometric graphs. The *Gaussian random geometric graph* $G(n, d, p)$ of Hunter, Milojević, and Sudakov [2] samples n vectors $x_1, \ldots, x_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \frac{1}{d} I_d)$ and connects x_i, x_j if $\langle x_i, x_j \rangle \geq c_p/\sqrt{d}$, where c_p is chosen so that each edge has marginal probability p. This is the Gaussian analogue of the spherical RGG; in high dimension, Gaussian vectors concentrate on a sphere of radius ≈ 1, making the two models asymptotically close.

Definition 3.1 (Threshold graph of the preference field)**.** For threshold θ, the *threshold graph* $G_t(\theta)$ has vertex set $A_t(\theta)$ and connects $i, j \in A_t(\theta)$ if $\langle v_i, v_j \rangle \geq \theta'$ for some proximity threshold θ' induced by θ and the corpus geometry.

Proposition 3.1 (RGG identification)**.** *If the corpus embeddings $\{v_i\}$ are drawn from a distribution on S^{d-1} whose conditional distribution given $\langle v_i, \Phi_t \rangle \geq \theta$ is approximately uniform on the spherical cap $\{v \in S^{d-1} : \langle v, \Phi_t \rangle \geq \theta\}$, then $G_t(\theta)$ is approximately distributed as a spherical RGG on $|A_t(\theta)|$ vertices with edge probability determined by the angular diameter of the cap.*

The "approximately uniform" condition holds to a first approximation for corpora generated by diffusion models conditioned on a fixed prompt, since the prompt induces a Gaussian-like perturbation around the prompt direction in CLIP space rather than a hard angular cutoff. The approximation improves as d grows relative to the angular radius of the cap.

3.2. *Triangle Penalty and Clique Suppression*

The key geometric fact distinguishing an RGG from $G_{n,p}$ is the *triangle penalty*: edges in the RGG are not independent because shared proximity to the preference direction correlates pairs of items. Hunter, Milojević, and Sudakov quantify this precisely for the Gaussian model.

Let $a = e^{-c_p^2/2}/\sqrt{2\pi}$ where c_p is the inner-product threshold parameter. Define the per-triangle correction factors

$$\delta_r = 1 - \frac{a^3 p^3}{\sqrt{d}}, \qquad \delta_b = 1 + \frac{a^3(1-p)^3}{\sqrt{d}}. \tag{7}$$

Theorem 3.1 (Clique probability bounds, Hunter–Milojević–Sudakov [2]). *Let $G \sim G(n, d, p)$ with $d \geq D^2\ell^2$ where D is sufficiently large. Then:*

$$P_{\mathrm{red},\ell} \leq p^{\binom{\ell}{2}} \exp\left(-\frac{a^3p^3}{\sqrt{d}}\binom{\ell}{3}\right)\left(1 + O\left(\tfrac{1}{D}\right)\right), \tag{8}$$

$$P_{\mathrm{blue},\ell} \leq (1-p)^{\binom{\ell}{2}} \exp\left(\frac{a^3(1-p)^3}{\sqrt{d}}\binom{\ell}{3}\right)\left(1 + O\left(\tfrac{1}{D}\right)\right). \tag{9}$$

The interpretation in our setting is the following. Color an item $i \in \mathcal{C}$ "blue" if $f_t(v_i) \geq \theta$ (it lies in the admissible neighborhood) and "red" if $f_t(v_i) < \theta$. Blue items are those close to Φ_t; red items are those far from it. Then:

- *Blue cliques* (fully admissible clusters) form with slightly higher probability than $G_{n,p}$ predicts ($\delta_b > 1$): proximity to the field direction is self-reinforcing, and items aligned with Φ_t tend to be mutually aligned.
- *Red cliques* (fully inadmissible clusters) form with exponentially lower probability than $G_{n,p}$ predicts: items forced far from Φ_t are correspondingly close to each other, making large mutually-distant (inadmissible) sets geometrically unlikely.

The exponential suppression in (8) grows as $\binom{\ell}{3}$—one triangle per triple in the clique—reflecting the cumulative geometric cost of each triangular constraint. This is the mathematical expression of the intuition that "inadmissible regions do not support large coherent clusters."

3.3. *Clique Number of the Admissible Neighborhood*

Definition 3.2 (Field clique number). The *field clique number* $\omega_t(\theta)$ is the size of the largest clique in the threshold graph $G_t(\theta)$: the maximum k such that k items in $A_t(\theta)$ are mutually within inner-product distance θ' of each other.

Corollary 3.1 (Clique bound under geometric constraint). *Under the RGG identification of Proposition 3.1, with $n_\theta = |A_t(\theta)|$ and edge probability p_θ:*

$$\mathbb{E}[\omega_t(\theta)] \lesssim 2\log_{1/(1-p_\theta)} n_\theta \cdot \left(1 + O\left(\frac{1}{\sqrt{d}}\right)\right). \tag{10}$$

The correction factor is negative ($\delta_b^{\binom{k}{3}} > 1$ amplifies rather than suppresses blue cliques), but grows slowly compared to the independent $G_{n,p}$ upper bound.

The corollary establishes that the field clique number is bounded independently of session length or rating history—it depends only on the corpus size

n_θ and the edge probability p_θ at the given alignment threshold. No amount of additional interaction can cause the admissible neighborhood to contain a coherent cluster larger than what the ambient spherical geometry permits. This is a structural, not empirical, constraint on preference field convergence.

4. Spectral Admissibility of Embedding Geometry

4.1. *Anisotropy as Inadmissibility*

Pretrained embedding models produce representations that are *anisotropic*: the image of the embedding map $\phi : \mathcal{C} \to S^{d-1}$ is not uniformly distributed over the sphere but concentrates in a narrow cone [4]. The centroid of this cone,

$$\bar{v} = \frac{1}{N} \sum_{i=1}^{N} v_i, \tag{11}$$

is non-zero and can be large in norm relative to the dispersion σ. When $\|\bar{v}\|$ is large, every item has non-negligible inner product with $\bar{v}$, regardless of semantic content.

This creates a systematic distortion of the preference field. In the weak-field regime (early in navigation, when Φ_t has not yet separated from $\bar{v}$), the alignment $f_t(v_i) = \langle \Phi_t, v_i \rangle \approx \langle \bar{v}, v_i \rangle$ for most items i, making $f_t \restriction_{\mathcal{M}}$ approximately constant and semantically uninformative. The field cannot discriminate until Φ_t has moved sufficiently far from $\bar{v}$ to produce meaningful variance.

Definition 4.1 (Admissible and inadmissible subspaces)**.** Let W_U be the unembedding matrix of a language or vision-language model, with singular value decomposition $W_U = U \Sigma V^\top$. The *edge-spectrum subspace* is

$$\mathcal{V}_{\text{edge}} = \text{span}\{V_{[k]} : k \in \mathcal{K}_{\text{edge}}\}, \tag{12}$$

where $\mathcal{K}_{\text{edge}}$ indexes the singular dimensions with either the largest or smallest singular values—the boundary of the spectrum. The *bulk subspace* is its complement: $\mathcal{V}_{\text{bulk}} = \mathcal{V}_{\text{edge}}^{\perp}$.

A vector $v \in \mathbb{R}^d$ is *spectrally admissible* if $\|\Pi_{\text{edge}} v\| / \|v\| < \varepsilon$ for some threshold $\varepsilon > 0$, where Π_{edge} is the orthogonal projector onto $\mathcal{V}_{\text{edge}}$.

Wu et al. [3] establish, via Logit Spectroscopy, that $\mathcal{V}_{\text{edge}}$ is precisely the subspace responsible for writing high-frequency, semantically uninformative tokens into the embedding space. The corpus centroid $\bar{v}$ lies predominantly in

$\mathcal{V}_{\text{edge}}$; removing the edge-spectrum projection eliminates the centroid bias and redistributes embeddings more isotropically over S^{d-1}.

4.2. *EmbedFilter as Admissibility Enforcement*

Definition 4.2 (EmbedFilter transformation [3])**.** The *Bulk Spectrum Transformation* with filtering ratio τ is the linear map

$$\Phi_\tau : \mathbb{R}^d \to \mathbb{R}^k, \qquad \Phi_\tau(v) = V[l_\tau : r_\tau]^\top v, \tag{13}$$

where $k = r_\tau - l_\tau \approx d/\tau$, and l_τ, r_τ bound the retained bulk indices. The *filtered embedding* is $\tilde{v} = \Phi_\tau(v)$.

Proposition 4.1 (EmbedFilter as inadmissibility removal)**.** *The EmbedFilter transformation satisfies:*

(i) Projection identity. $\Phi_\tau = (I - \Pi_{\text{edge}}) + \Pi_{\text{bulk}}$*, up to re-indexing. In particular,* $\tilde{v} = (I - \Pi_{\text{edge}})v$ *up to rotation within the bulk.*

(ii) Distance preservation. *For any* $v, w \in \mathbb{R}^d$*,* $\|\Phi_\tau(v) - \Phi_\tau(w)\|_2 = \|v - w\|_2$ *within the bulk subspace.*

(iii) Centroid elimination. *If* $\bar{v} \in \mathcal{V}_{\text{edge}}$ *then* $\Phi_\tau(\bar{v}) = 0$.

(iv) Admissibility enforcement. *Every* $\tilde{v}_i = \Phi_\tau(v_i)$ *is spectrally admissible with respect to the original edge-spectrum projector.*

Proof. (i) follows from the orthogonality of the singular vectors of V. (ii) follows from the fact that $V[l_\tau : r_\tau]$ consists of columns of an orthogonal matrix, hence is an isometry on its domain. (iii) holds when the centroid is exactly in the edge subspace; in practice this is approximate, and the quality of centroid elimination depends on how concentrated $\bar{v}$ is in $\mathcal{V}_{\text{edge}}$. (iv) is immediate from (i): after projection, the edge-spectrum component of $\tilde{v}_i$ is zero. □

4.3. *Effect on Preference Field Discriminability*

After applying EmbedFilter, the filtered preference field

$$\tilde{f}_t(v_i) = \langle \tilde{\Phi}_t, \tilde{v}_i \rangle \tag{14}$$

operates in the bulk subspace $\mathcal{V}_{\mathrm{bulk}}$. Because the centroid bias has been removed, the variance

$$\mathrm{Var}_i[\tilde{f}_t(v_i)] = \frac{1}{N}\sum_{i=1}^{N} \tilde{f}_t(v_i)^2 - \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{f}_t(v_i)\right)^2 \tag{15}$$

is strictly larger than $\mathrm{Var}_i[f_t(v_i)]$ whenever Φ_t has a non-negligible edge-spectrum component—that is, whenever the unfiltered field is being partially driven by the centroid attractor. This variance increase directly improves the discriminative resolution of the preference field, allowing meaningful alignment differences to emerge earlier in the navigation trajectory.

Empirically, Wu et al. report up to 14% improvement in MTEB downstream task performance across Qwen, Llama, and Mistral backbones at $\tau = 2$ (halving the embedding dimension), confirming that the bulk subspace carries the semantically relevant structure [3].

5. The Indistinguishability Threshold and Regime Transitions

5.1. *The Signed Triangle Count as a Geometric Statistic*

The central statistic for distinguishing an RGG from $G_{n,p}$ is the *signed triangle count*. For a graph H on n vertices with edge-probability p and adjacency matrix A, define

$$T(H) = \sum_{\{i,j,k\}\subset[n]} (A_{ij} - p)(A_{jk} - p)(A_{ik} - p). \tag{16}$$

Proposition 5.1 (Signed triangle count discriminates RGG from $G_{n,p}$)**.** *For $G \sim G_{n,p}$ (independent edges): $\mathbb{E}[T(G)] = 0$. For $G \sim$ spherical RGG with the same marginal edge probability:*

$$\mathbb{E}[T(G)] \approx \frac{n^3p^3}{\sqrt{d}}. \tag{17}$$

In both models $\mathrm{Var}[T(G)] \approx n^3p^3$.

The signal-to-noise ratio for distinguishing the two distributions is therefore $n^3p^3/\sqrt{d}\,/\,\sqrt{n^3p^3} = \sqrt{n^3p^3/d}$. The geometric signal is detectable when this ratio exceeds a constant, i.e. when $d < n^3p^3$.

5.2. The Regime Indicator

Definition 5.1 (Preference field regime indicator)**.** At state t, for threshold θ, let $n_\theta = |A_t(\theta)|$ and let p_θ be the effective edge probability in $G_t(\theta)$. The *regime indicator* is

$$\Lambda_t(\theta) = \frac{n_\theta^3 p_\theta^3}{d}. \tag{18}$$

Λ_t is the natural order parameter of the theory. Every other quantity—alignment variance, clique number, triangle density, spectral discriminability—is monotone in Λ_t. When Λ_t crosses its critical value the system undergoes a phase transition from a regime in which the corpus geometry is hidden to one in which it is geometrically structured and navigable. In this respect Λ_t plays the same role in preference field theory that the Reynolds number plays in fluid mechanics, the inverse temperature β plays in statistical mechanics, and the basic reproduction number R_0 plays in epidemiology: it is the single dimensionless parameter that separates qualitatively distinct phases of the system's behavior.

Proposition 5.2 (Field detectability criterion)**.** *Let $G_t(\theta)$ be the threshold graph induced by the preference field at threshold θ. Geometric information about the underlying corpus manifold is recoverable from $G_t(\theta)$ if and only if $\Lambda_t(\theta)$ exceeds a constant-order threshold. More precisely:*

- *As $\Lambda_t \to 0$: the distribution of $G_t(\theta)$ converges in total variation to G_{n_θ, p_θ}; no statistical test can distinguish the threshold graph from an Erdős–Rényi random graph; the preference field carries no recoverable geometric information.*
- *As $\Lambda_t \to \infty$: the signed triangle count $T(G_t(\theta))$ diverges from its $G_{n,p}$ expectation of zero by $\Theta(\sqrt{n_\theta^3 p_\theta^3/d}) \cdot \sqrt{n_\theta^3 p_\theta^3}$ standard deviations; geometric clustering is detectable with power approaching one; the preference field is resolving genuine semantic structure.*
- *At $\Lambda_t \asymp 1$: the signed triangle count achieves signal-to-noise ratio $\Theta(1)$; this is the critical threshold separating detectable from undetectable geometry.*

Sketch. The total variation convergence for $\Lambda_t \to 0$ follows from the HMS indistinguishability theorem [2]: when $d > n^3 p^3$ the total variation distance between the spherical RGG and $G_{n,p}$ goes to zero. The detectability for $\Lambda_t \to \infty$ follows from Proposition 5.1: the signal-to-noise ratio of the signed triangle count is $\sqrt{n_\theta^3 p_\theta^3/d} = \sqrt{\Lambda_t}$, which diverges. The critical threshold follows from setting signal equal to noise. $\square$

Theorem 5.1 (Phase transition in preference field geometry)**.** *Let C be a corpus with embeddings approximately distributed as a spherical RGG in dimension d. Then:*

- *If $\Lambda_t(\theta) \ll 1$: the sub-corpus $A_t(\theta)$ is statistically indistinguishable from G_{n_θ,p_θ}; the preference field cannot resolve geometric structure; the admissible neighborhood is a uniformly random subset of size n_θ.*
- *If $\Lambda_t(\theta) \gg 1$: the geometry of S^{d-1} is detectable in $A_t(\theta)$; triangles close at a rate exceeding $G_{n,p}$ prediction; the admissible neighborhood has genuine cluster structure that the preference field is resolving.*

The transition between regimes occurs at $\Lambda_t(\theta) \asymp 1$, i.e. $n_\theta \asymp (d/p_\theta^3)^{1/3}$.

Sketch. The indistinguishability claim for $\Lambda_t \ll 1$ follows from Milojević, Hunter, and Sudakov's theorem that the total variation distance between a spherical RGG and $G_{n,p}$ converges to zero when $d > n^3p^3$ (see [2] and the companion indistinguishability paper). The detectability claim for $\Lambda_t \gg 1$ follows from Proposition 5.1: the signed triangle count provides a statistically powerful test whenever $\sqrt{n^3p^3/d} \gg 1$. $\square$

5.3. *Implications for Navigation*

The regime indicator $\Lambda_t(\theta)$ gives a principled diagnostic of preference field quality. When $\Lambda_t \ll 1$, the field is not resolving genuine semantic structure: the admissible neighborhood looks random, and additional preference signals cannot improve geometric discrimination until n_θ grows. When $\Lambda_t \gg 1$, the field is operating in the geometric regime: the coherent cluster identified by $A_t(\theta)$ is a real semantic neighborhood, and further preference refinement can sharpen its boundaries.

This has a practical consequence: the transition $\Lambda_t(\theta) \asymp 1$ marks a natural threshold for switching from exploratory navigation (increasing n_θ to find a geometric regime) to refinement (decreasing θ to identify the tightest coherent sub-cluster). The mathematics does not prescribe exploration strategy, but it identifies where in the parameter space exploration is necessary and where refinement becomes productive.

6. Discussion

The central contribution of the preceding sections is that preference-guided navigation naturally generates a geometric field theory. The preference vector

Φ_t is not merely a weighted sum of ratings; it is a direction in S^{d-1} whose induced scalar field has measurable geometric properties. The admissible neighborhood is not merely a ranking threshold; it is a subgraph of a random geometric graph whose clique structure and triangle density encode the semantic organization of the corpus. And spectral admissibility is not merely an engineering post-processing step; it is the removal of an inadmissible subspace that systematically distorts the field's discriminative geometry.

The order parameter Λ_t unifies these observations. It is a single dimensionless quantity computed from the admissible neighborhood size, the effective edge probability, and the embedding dimension. Every other observable quantity in the theory is monotone in Λ_t, and the phase transition at $\Lambda_t \asymp 1$ separates qualitatively distinct navigation regimes.

6.1. *Connections to Constraint-Based Field Frameworks*

The mathematical structures developed here bear structural similarity to several frameworks that study the geometry of accessibility and constraint in different settings. We record these briefly without claiming formal equivalence.

The preference field $f_t(v) = \langle \Phi_t, v \rangle$ plays the same mathematical role as the scalar admissibility field in RSVP-based formulations [5]: a time-dependent scalar that organizes which regions of a state space are accessible and shapes the competitive landscape of local configurations. The spectral admissibility operation $(I - \Pi_{\text{edge}})$ acts as a directional entropy-reduction operator analogous to RSVP's admissibility projection, removing high-entropy directions that carry no differential semantic information. The present work neither assumes RSVP nor claims equivalence; the correspondence is at the level of mathematical role.

The nested sequence of admissible neighborhoods $A_t(\theta_1) \supseteq A_{t'}(\theta_2) \supseteq \cdots$ as the field sharpens resembles the recursive, constraint-preserving tiling structure of TARTAN, in which each refinement step inherits and tightens the structure of the previous level. Whether the preference field sequence can be expressed as a TARTAN tiling in a precise categorical sense is an open question.

Preference trajectories $\Phi_1, \ldots, \Phi_T$ induce ordered chains of admissible projections. Items that persist in $A_t(\theta)$ across many time steps maintain relevance through continued compatibility with the evolving field, rather than through static storage—a structure that resembles Chain of Memory models of persistence [6]. Whether this resemblance can be formalized as a dynamical equivalence or requires a categorical treatment is also open.

7. Future Directions

7.1. Adaptive Spectral Admissibility

The EmbedFilter transformation uses a static admissibility criterion: the edge-spectrum subspace is identified once from the global unembedding matrix SVD. A more principled formulation allows the criterion to adapt to the current field state. Define a parameterized family of admissibility projectors $\Pi_\theta : \mathbb{R}^d \to \mathbb{R}^k$ and update θ to maximize the discriminative variance of the filtered field over the rated set:

$$\theta_t^* = \arg\max_\theta \; \mathrm{Var}_{i \sim \mathcal{R}_t}[\langle \Pi_\theta(\Phi_t),\, \Pi_\theta(v_i) \rangle] \,. \tag{19}$$

This is online principal component analysis under the preference-weighted distribution over $\mathcal{R}_t$, initialized at the static EmbedFilter solution. The effective embedding space specializes to the semantic cluster the field currently inhabits, discarding spectral dimensions that are globally informative but locally irrelevant to the current Φ_t.

7.2. Continuous Field Theory: From Projector to Connection

The present framework is fundamentally corpus-relative: the admissibility projector acts on a finite set of vectors and all objectives are defined through empirical variance over rated examples. Moving to a continuous field theory requires the projector to become a geometric object.

The natural formulation replaces the fixed projector $\Pi : \mathbb{R}^d \to \mathbb{R}^k$ with a smoothly varying family

$$\Pi_v : T_v\mathcal{M} \to A_v, \tag{20}$$

where $A_v \subset T_v\mathcal{M}$ is the admissible tangent subspace at $v \in \mathcal{M}$. The admissible directions form a distribution $A \subset T\mathcal{M}$. The preference field is no longer a single vector $\Phi_t \in S^{d-1}$ but a section

$$\Phi : \mathcal{M} \to T\mathcal{M} \tag{21}$$

whose evolution is constrained to remain inside A. The adaptive projector becomes a *connection* on A, specifying how admissible directions transport as one moves through semantic space:

$$\frac{D\Pi}{dt} = F(\mathcal{R}_t, \Phi_t), \tag{22}$$

where F encodes the local curvature of $\mathcal{M}$.

The fundamental question at this level is whether the distribution A is *integrable*. By the Frobenius theorem, integrability is equivalent to the existence of admissible leaves $\mathcal{L} \subset \mathcal{M}$ such that navigation remains within a semantic foliation. An integrable admissibility distribution implies that semantic concepts correspond to closed leaves and navigation is path-independent within a leaf. A non-integrable distribution implies path dependence: the accessible region depends on the trajectory, not merely the current position. This distinction separates structurally different classes of semantic corpora: a highly-organized scientific corpus may possess nearly integrable admissibility leaves; a heterogeneous internet-scale corpus almost certainly does not.

7.3. *The Continuous Regime Indicator*

The discrete order parameter $\Lambda_t = n_\theta^3 p_\theta^3 / d$ is derived from graph statistics over a finite neighborhood. In the continuous limit we expect it to become a local geometric quantity:

$$\Lambda(x) = \frac{\rho(x)^3 \, p(x)^3}{\kappa(x)}, \tag{23}$$

where $\rho(x)$ is the local semantic density at $x \in \mathcal{M}$, $p(x)$ is the local alignment probability, and $\kappa(x)$ is an effective curvature scale replacing the ambient dimension d. The discrete dimension d measures global indistinguishability; the curvature $\kappa(x)$ measures local geometric obstruction. The phase transition then becomes not "enough vertices versus too many dimensions" but "enough local density to overcome curvature-induced indistinguishability." This is a genuine field-theoretic statement: the transition point is a local property of the manifold, not a global property of the embedding space.

7.4. *Sheaf-Theoretic Admissibility and Global Sections*

In the continuous formulation, each open neighborhood $U \subset \mathcal{M}$ possesses its own admissible subspace $A(U) \subset T_U\mathcal{M}$. These local assignments define a *presheaf of admissible directions*. The central question becomes whether local admissibility assignments are consistent: whether sections defined on overlapping neighborhoods $U_\alpha \cap U_\beta$ agree on the overlap, and whether local sections can be extended to global ones.

If the admissibility presheaf satisfies the sheaf gluing axioms, then a global semantic concept exists whenever its local restrictions are consistent. If it does not—if there are obstructions to gluing—then the concept is only locally definable and cannot be extended across the full manifold. The obstruction to gluing is captured by the sheaf cohomology group $H^1(\mathcal{M}, \mathcal{A})$, where $\mathcal{A}$ is the sheaf of admissible sections.

Viewed this way, the adaptive admissibility projector is the discrete precursor of a sheaf-gluing operator. Its discrete version selects admissible dimensions; its continuous version determines whether local semantic sections extend into global ones. Concepts whose admissibility sections glue globally correspond to $H^1 = 0$; concepts with non-trivial obstruction class correspond to $H^1 \neq 0$ and exist only locally. This is the natural home of the framework: a geometric theory in which admissibility is a distribution on a manifold, the projector is a connection on that distribution, and semantic concepts correspond to globally gluable sections.

7.5. *Ramsey-Theoretic Stopping Criteria*

Corollary 3.1 gives a Ramsey-theoretic bound on the field clique number $\omega_t(\theta)$. Once ω_t stabilizes, the field has extracted the maximum coherent cluster available at the current threshold: further preference refinement within the same θ cannot produce a larger coherent neighborhood. The Frobenius integrability condition provides a deeper interpretation: ω_t stabilization signals that the field has reached the boundary of the current admissible leaf. Continued navigation requires either expanding θ (moving to a coarser leaf) or generating new corpus points by interpolation in $\mathcal{E}$ (probing the leaf's interior by synthesis rather than selection).

7.6. *Formal Functorial Unification*

The structural correspondences of Section 6 suggest a formal program. Define the *preference field category* $\mathbf{PF}$ whose objects are corpus manifolds equipped with preference fields and whose morphisms are field-preserving maps. The natural question is whether there exist functors

$$F_{\text{RSVP}} : \mathbf{PF} \to \mathbf{RSVP}, \quad F_{\text{TARTAN}} : \mathbf{PF} \to \mathbf{TARTAN}, \quad F_{\text{CoM}} : \mathbf{PF} \to \mathbf{CoM}, \tag{24}$$

where $\mathbf{RSVP}$, $\mathbf{TARTAN}$, $\mathbf{CoM}$ are the categories of the respective frameworks. If such functors exist and are faithful, the frameworks are provably instances of a single underlying mathematical structure rather than analogous constructions developed independently. This program requires careful categorical development of each target framework and constitutes a natural long-range goal.

8. Mathematical Summary

For reference, we collect the central definitions and results.

Preference field. $\Phi_t \in S^{d-1}$; updated by $\Phi_t = \tilde{\Phi}_t / \|\tilde{\Phi}_t\|$ where $\tilde{\Phi}_t = \sum_{(i,r_i)\in\mathcal{R}_t} w(r_i) v_i$; induces $f_t(v) = \langle \Phi_t, v \rangle$.

Admissible neighborhood. $A_t(\theta) = \{i : f_t(v_i) \geq \theta\}$; converges under consistent preference (Proposition 2.1).

Triangle corrections. $\delta_r = 1 - a^3 p^3/\sqrt{d}$, $\delta_b = 1 + a^3(1-p)^3/\sqrt{d}$; clique probabilities in the threshold graph are governed by Theorem 3.1.

Spectral admissibility. $(I - \Pi_{\text{edge}})$ removes the centroid-biased subspace; EmbedFilter enforces admissibility (Proposition 4.1); variance of $\tilde{f}_t$ over the corpus increases after filtering.

Regime indicator. $\Lambda_t(\theta) = n_\theta^3 p_\theta^3/d$; $\Lambda_t \ll 1$ is the geometric regime transition (Theorem 5.1).

Field clique bound. $\mathbb{E}[\omega_t(\theta)] \lesssim 2\log_{1/(1-p_\theta)} n_\theta$ with $O(1/\sqrt{d})$ correction (Corollary 3.1).

References

[1] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, et al. Learning transferable visual models from natural language supervision. In *Proceedings of ICML*, 2021.

[2] Zach Hunter, Aleksa Milojević, and Benny Sudakov. Gaussian random graphs and Ramsey numbers. arXiv:2512.17718 [math.CO], 2026.

[3] Songhao Wu, Zhongxin Chen, Yuxuan Liu, Heng Cui, Cong Li, and Rui Yan. Your unembedding matrix is secretly a feature lens for text embeddings. arXiv:2606.07502 [cs.CL], 2026.

[4] Kawin Ethayarajh. How contextual are contextualized word representations? In *Proceedings of EMNLP-IJCNLP*, 2019.

[5] Flyxion. *RSVP: Scalar-Vector-Entropic Field Cosmology*. Monograph, 2024.

[6] Flyxion. *Chain of Memory: Selective Persistence and Iterative Refinement*. Essay, 2024.

[7] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[8] Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.