

Numerical Stability and Perturbation Behaviour in TF-IDF-Based Similarity Systems

Matthew Maksymilian Miezaniec
King's College London
matthewmiezaniec1@gmail.com

March 2026

Abstract

This repository presents a **self-contained implementation and analysis** of a **TF-IDF-based document similarity and ranking system**, written to make the **algebraic structure and perturbation behaviour** of the pipeline explicit.

Documents are embedded using **explicit tokenisation**, **n-gram construction**, **smoothed inverse document frequency (IDF)**, and **sparse TF-IDF vectors**, with similarity measured via **cosine similarity** and rankings induced through a deterministic sorting procedure. The implementation exposes intermediate quantities such as document frequencies, IDF weights, vector norms, and similarity scores. These enable direct examination of how corpus structure and similarity geometry relate to ranking behaviour under perturbation.

Beyond analysing numerical sensitivity of TF-IDF embeddings and cosine similarity, the repository studies **ranking stability** as a first-class object. Stability is characterised empirically using **score-separation margins**, including boundary margins governing top- k membership and adjacent margins controlling ordering within top- k results. These margins yield explicit, sufficient conditions under which rankings remain invariant under bounded perturbations of similarity scores.

The work further isolates **decision-level discontinuities** introduced by deterministic **tie-breaking rules**, demonstrating that ranking outcomes may change even when cosine similarity scores are equal within numerical tolerance. This separation between **numerical stability of similarity scores** and **stability of induced rankings** highlights structural sources of fragility in similarity-based retrieval and content-based recommendation systems.

1. Introduction

This project examines the numerical stability of **TF-IDF-based similarity and ranking systems** under small perturbations in data and preprocessing.

TF-IDF (term frequency-inverse document frequency) is a foundational technique in **information retrieval** and **text-based modelling**. It represents documents as vectors in a **high-dimensional feature space**, where similarity is typically measured using **cosine similarity**. Despite its widespread use, the behaviour of TF-IDF-based pipelines under perturbations—such as changes in corpus composition, token distributions, or preprocessing rules—is rarely analysed in a systematic and explicit manner.

This work studies TF-IDF in a controlled setting where documents arise from **text of interest** and similarity scores are used to induce **content-based k -nearest-neighbour (k -NN) rankings**. The system is implemented in **pure Python**, with explicit control over preprocessing, vocabulary construction, IDF computation, vector formation, and ranking. This explicit formulation enables direct examination of how corpus structure and similarity geometry relate to similarity scores, neighbourhood structure, and ranking behaviour under perturbation.

The central question is one of stability:

How sensitive are TF-IDF weights, cosine similarities, and induced rankings to small changes in the underlying data?

More specifically, this work addresses two closely related aspects of this question:

1. how **score-separation margins** govern the stability of similarity-based rankings under bounded perturbations, and
2. how **deterministic tie-breaking rules** introduce discontinuities in ranking outcomes that are independent of numerical error in similarity computation.

1.1. Purpose and Research Intent

The purpose of this repository is **investigative rather than applicative**. Although TF-IDF is a *classical method*, the aim is to examine the **mathematical structures induced by the TF-IDF pipeline** and to understand how these structures behave under controlled variation of assumptions and parameters, particularly in the context of **content-based similarity** and **k -nearest-neighbour ranking**.

Rather than treating the similarity pipeline as a black box, the implementation explicitly exposes **preprocessing operations**, **n -gram structure**, **document-frequency thresholds**, **IDF scaling**, and **sparse vector geometry**. This level of explicitness supports detailed analysis of **sensitivity** and **stability phenomena** in TF-IDF embeddings, cosine similarity, and induced k -NN neighbourhoods. In particular, it enables investigation of situations in which small perturbations in documents, corpus composition, or user-derived profiles lead to **disproportionate changes** in similarity scores, neighbourhood structure, or ranking outcomes.

In addition to numerical sensitivity at the level of embeddings and similarity scores, the repository treats **ranking stability as a primary object of analysis**. Stability is operationalised using **empirical score-separation margin distributions**, which quantify the tolerance of ranking outcomes to bounded perturbations. Separately, the work isolates **instability arising solely from deterministic tie-breaking rules**, demonstrating how ranking outcomes may change even when similarity scores are equal within numerical tolerance.

Empirical investigations in this repository focus primarily on score-level and decision-level perturbations, while corpus- and embedding-level effects are analysed analytically to characterise their influence on downstream stability.

1.2. Investigative Scope

By retaining access to **intermediate quantities** that are typically abstracted away in higher-level libraries, the repository supports systematic investigation of questions including:

- the effect of corpus perturbations on **document frequency** and **smoothed IDF values**,
- the response of **TF-IDF embeddings** to token-level edits or preprocessing changes,
- the sensitivity of **cosine similarity** to perturbations in *sparse, non-negative vectors*,
- conditions under which **similarity-based rankings** and **k -NN neighbourhoods** remain invariant under bounded perturbations,
- the influence of **user-profile construction** on personalised similarity and neighbourhood structure,
- **margin distributions** governing the stability of top- k membership and ordering within top- k results,

- **tie-group analysis**, quantifying ranking changes induced by deterministic tie-breaking rules when similarity scores fall within numerical tolerance.

This level of access connects **formal derivation** with **empirical observation** and makes the framework suitable for studying **conditioning behaviour**, **perturbation effects**, and **ranking stability** in TF-IDF-based similarity systems.

1.3. Position Within a Broader Mathematical Context

TF-IDF-based similarity lies at the intersection of **information retrieval**, **numerical linear algebra**, **finite-dimensional functional analysis**, and **probabilistic models of text**. The construction provides a concrete setting in which abstract numerical phenomena—such as **sparsity**, **scaling behaviour**, **angular distortion**, and **perturbation amplification**—can be observed and analysed directly.

From a decision-theoretic perspective, similarity-based ranking may be viewed as a **piecewise-constant functional of similarity scores**, with **discontinuities induced by secondary ordering and tie-breaking rules**. This perspective clarifies the distinction between numerical stability of the underlying similarity computation and stability of the induced ranking outcomes.

These issues recur in **high-dimensional feature representations** used in machine learning, **content-based similarity and recommendation systems**, and **retrieval pipelines**, where numerical sensitivity and decision discontinuities can have practical consequences. Accordingly, this work is positioned as an **expository and exploratory study** of established mathematical constructions within a TF-IDF setting, with particular attention to their behaviour under perturbation and their implications for ranking stability and interpretability.

1.4. Intended Use

This repository is **not intended as a production-grade recommendation library**. Instead, it serves as a **reference implementation**, an **exploratory mathematical environment**, and a foundation for *small-scale research investigations* into TF-IDF representations, cosine similarity, content-based k -NN ranking, and ranking stability.

The emphasis throughout is on understanding **mathematical behaviour**, particularly **conditioning**, **sensitivity**, and **perturbation effects** in sparse vector spaces, as well as the relationship between numerical similarity computation and downstream ranking decisions. These considerations are central to the **reliability**, **reproducibility**, and **interpretability** of similarity-based and content-based recommendation models.

2. Mathematical Formulation

Let

$$\mathcal{D} = \{d_1, \dots, d_n\}$$

be a finite corpus of preprocessed documents, and define

$$N := |\mathcal{D}|.$$

Each document d_i is a finite sequence of tokens obtained from raw text via a fixed preprocessing map (normalisation, stopword removal, lemmatisation, and n -gram generation). Throughout, **n -grams are treated as atomic tokens**.

All preprocessing operations are assumed to be deterministic and fixed across all perturbation experiments.

2.1. Vocabulary, Document Frequency, and IDF

From the corpus $\mathcal{D}$, a vocabulary V is constructed by collecting all tokens (including n -grams) that satisfy a minimum document-frequency threshold and, optionally, a maximum-feature constraint. For each token $t \in V$, the document frequency is defined as

$$\text{df}(t) = |\{i : t \text{ appears at least once in } d_i\}|.$$

Let $N = |\mathcal{D}|$. The smoothed inverse document frequency used throughout the implementation is

$$\text{idf}(t) = \log \left(\frac{1 + N}{1 + \text{df}(t)} \right) + 1.$$

This choice ensures monotonic decay of $\text{idf}(t)$ as $\text{df}(t)$ increases. The additive constant ensures strict positivity even in the limiting case $\text{df}(t) = N$.

(Here and throughout, $\log$ denotes the natural logarithm.)

2.2. TF-IDF Embedding

For each document d_i , let $\text{count}_i(t)$ denote the number of occurrences of token $t \in V$ in d_i . The term frequency is defined as

$$\text{tf}_i(t) = \frac{\text{count}_i(t)}{\sum_{s \in V} \text{count}_i(s)}.$$

That is, term frequencies are normalised with respect to **in-vocabulary tokens**, with out-of-vocabulary tokens ignored after vocabulary construction. Documents whose in-vocabulary token count is zero are mapped to the zero vector.

The TF-IDF weight of token t in the document d_i is then

$$w_i(t) = \text{tf}_i(t) \cdot \text{idf}(t),$$

and the document is represented as a sparse vector

$$w_i \in \mathbb{R}_{\geq 0}^{|V|},$$

with coordinates indexed by the vocabulary V .

2.3. Cosine Similarity, Ranking, and Stability Quantities

Given two non-zero TF-IDF vectors $u, v \in \mathbb{R}_{\geq 0}^{|V|}$, the cosine similarity is defined as

$$\cos(u, v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2},$$

with the convention that the similarity is set to zero if either vector is the zero vector. Since all coordinates are non-negative, it follows that

$$\cos(u, v) \in [0, 1].$$

Given a query vector $q \in \mathbb{R}_{\geq 0}^{|V|}$, embedded using the same vocabulary V and IDF mapping as the corpus documents, and a collection of document vectors $\{w_i\}$, similarity scores

$$s_i = \cos(q, w_i)$$

are computed for each document.

2.3.1. Ranking Operator

Similarity scores alone do not define a total order when ties occur. Accordingly, the final ranking is defined via a deterministic sorting operator

$$\pi = \text{Sort}((s_i, a_i)_i),$$

where Sort orders documents by decreasing similarity score s_i and resolves ties lexicographically using the fixed attribute tuple a_i (e.g. popularity, rating, engagement, identifier). This yields a total ordering $(r_1, r_2, \dots, r_n)$, where r_j denotes the document at rank j .

As a consequence, the mapping from similarity scores to rankings is not continuous globally; it is locally constant away from tie hyperplanes.

2.3.2. Score-Separation Margins (A1)

Let $\text{score}(r_j)$ denote the similarity score of the document at rank j . The **boundary margin** at rank k is defined as

$$m_k = \text{score}(r_k) - \text{score}(r_{k+1}).$$

This quantity governs the stability of top- k membership under bounded perturbations of similarity scores.

The **minimum adjacent margin within the top- k** is defined as

$$m_{\min}^{\text{top}} = \min_{1 \leq j < k} (\text{score}(r_j) - \text{score}(r_{j+1})),$$

This quantity controls the stability of the ordering within the top- k set.

For convenience, the corresponding **flip radius** at rank k is defined as

$$\varepsilon_k^{\text{flip}} = \frac{m_k}{2},$$

representing the largest uniform perturbation magnitude (in score space) under which the relative ordering of ranks k and $k + 1$ is preserved.

2.3.3. Tie Groups and Decision Discontinuities (A2)

Fix the ranking

$$(r_1, r_2, \dots, r_n)$$

induced by the deterministic ranking operator π .

To formalise near-ties, fix a numerical tolerance $\tau > 0$. For a given rank position j , the associated **tie group** is defined as

$$G_\tau(j) = \{i : |s_i - \text{score}(r_j)| \leq \tau\}.$$

Here the scores s_i are the same similarity scores used to produce the ranking (r_j) .

Documents within a tie group are indistinguishable at the level of similarity scores up to numerical tolerance. In such cases, the final ordering within $G_\tau(j)$ is determined entirely by the deterministic tie-breaking rules embedded in the ranking operator π .

This construction makes explicit the distinction between **numerical stability of similarity scores** and **stability of the induced ranking**, and provides a formal basis for analysing decision-level discontinuities arising from secondary ordering criteria.

3. Solution Procedure and Implementation Structure

The repository implements the TF-IDF pipeline and associated perturbation-theoretic investigations in a **fully explicit and reproducible** manner. Each stage of the pipeline is designed to expose **intermediate quantities** and **algebraic structure**, rather than to optimise performance or abstract away implementation details.

- **Preprocessing and Corpus Construction**

A fixed, deterministic preprocessing map is applied to raw **text of interest**, including normalisation, tokenisation, stopword removal, lemmatisation, and n -gram generation. This yields a reproducible corpus suitable for **controlled perturbation analysis** via explicit inspection of intermediate quantities.

- **TF-IDF Vectorisation**

A pure-Python TF-IDF vectoriser constructs the vocabulary, computes document frequencies and smoothed inverse document frequency (IDF) values, and embeds documents as vectors in $\mathbb{R}_{\geq 0}^{|V|}$. In typical use, these embeddings are sparse due to the size of the vocabulary relative to document length.

- **Similarity, Ranking, and k -NN Structure**

Cosine similarity is computed between query vectors and corpus vectors, followed by deterministic ranking. In the context of **content-based k -nearest-neighbour recommendation**, the top- k elements of this ranking define the neighbourhood associated with a query. Secondary attributes, such as auxiliary metadata or identifiers, are applied via lexicographic tie-breaking, yielding a **total order** on candidate items.

In addition to score computation, the implementation explicitly computes **score-separation margins** governing top- k membership and within-top- k ordering, and supports stability profiling of rankings under bounded perturbations. Observed score-separation margins are used as empirical certificates of stability, rather than relying on explicit noise injection. The ranking procedure is further instrumented to enable **tie-breaking ablation experiments**, in which ordering is recomputed under alternate tie-break priorities or using score-only sorting with a fixed, attribute-independent identifier as a final deterministic tie-break, thereby isolating decision-level effects attributable to secondary attributes from numerical similarity computation.

- **User-Profile Documents**

User-specific documents are constructed from interactions such as liked, viewed, or favoured items. These user-profile documents are embedded using the **same vocabulary and**

IDF mapping as the corpus, ensuring that all similarity computations take place in a common vector space.

- **Perturbation Analysis**

The explicit structure of the implementation supports analytical study and targeted empirical inspection of how small perturbations in documents, corpus composition, or user interactions propagate through document frequencies, IDF values, TF-IDF embeddings, cosine similarities, induced k -NN neighbourhoods, and ranking outcomes. In particular, the framework enables examination of mechanisms that can amplify small upstream perturbations, arising from interactions between the sparse vector geometry, IDF scaling, angular similarity, and deterministic decision rules.

The code is organised to preserve **algebraic clarity** and to expose intermediate quantities at every stage. As such, it is well suited as a basis for further mathematical analysis of **stability**, **sensitivity**, and **decision-level fragility** in TF-IDF-based similarity systems.

All stages of the pipeline are deterministic given a fixed corpus, configuration, and software environment (including library versions), ensuring reproducible similarity scores and rankings across runs.

4. Error Analysis and Perturbation Quantities

Let w_i denote the TF-IDF vector associated with document d_i , and let

$$s_i = \cos(q, w_i)$$

denote the similarity score between a query vector q and the i -th document. This section introduces quantitative measures for analysing how perturbations affect intermediate quantities in the TF-IDF pipeline and how these effects translate to similarity scores and induced rankings.

Throughout, perturbations are treated as **bounded changes** in intermediate numerical quantities. No probabilistic or adversarial noise model is assumed unless stated explicitly. In the implementation, such perturbations are analysed primarily at the level of similarity scores and induced rankings, with upstream effects treated analytically.

4.1. Perturbations in Document Frequency and IDF

Consider a perturbation of the corpus induced by adding or removing a document, or by modifying the token content of an existing document. Let $\text{df}(t)$ and $\text{df}'(t)$ denote the document frequencies of token t before and after perturbation, and let N and N' denote the corresponding corpus sizes.

Using the smoothed IDF definition employed throughout the implementation,

$$\text{idf}(t) = \log \left(\frac{1 + N}{1 + \text{df}(t)} \right) + 1,$$

the corresponding change in IDF is

$$\begin{aligned} \Delta \text{idf}(t) &= \text{idf}'(t) - \text{idf}(t) \\ &= \log \left(\frac{1 + N'}{1 + \text{df}'(t)} \right) - \log \left(\frac{1 + N}{1 + \text{df}(t)} \right). \end{aligned}$$

This expression makes explicit the competing effects of changes in corpus size and document-frequency distribution on IDF values. In particular, tokens with low document frequency remain sensitive to corpus perturbations even under smoothing.

4.2. Perturbations in TF-IDF Vectors

Let w_i and w'_i denote the TF-IDF vectors of a document before and after perturbation. Writing

$$\begin{aligned} w_i &= \text{tf}_i \odot \text{idf}, \\ w'_i &= \text{tf}'_i \odot \text{idf}', \end{aligned}$$

where $\odot$ denotes pointwise (Hadamard) multiplication, we obtain

$$w'_i - w_i = (\Delta \text{tf}_i) \odot \text{idf} + \text{tf}_i \odot (\Delta \text{idf}) + (\Delta \text{tf}_i) \odot (\Delta \text{idf}),$$

with $\Delta \text{tf}_i := \text{tf}'_i - \text{tf}_i$ and $\Delta \text{idf} := \text{idf}' - \text{idf}$.

Applying the inequality $\|a \odot b\|_2 \leq \|a\|_2 \|b\|_\infty$ termwise yields the bound

$$\begin{aligned} \|w'_i - w_i\|_2 &\leq \|\Delta \text{tf}_i\|_2 \cdot \|\text{idf}\|_\infty \\ &\quad + \|\text{tf}_i\|_2 \cdot \|\Delta \text{idf}\|_\infty \\ &\quad + \|\Delta \text{tf}_i\|_2 \cdot \|\Delta \text{idf}\|_\infty. \end{aligned}$$

This decomposition separates perturbations arising from **local document edits** (Δtf_i), **global corpus changes** (Δidf), and their interaction. In sparse high-dimensional embeddings, the interaction between local changes and globally scaled IDF weights provides a natural mechanism for perturbation amplification.

4.3. Perturbations in Cosine Similarity

Let u, v be TF-IDF vectors and let u', v' denote their perturbed counterparts. Under mild assumptions on the norms of these vectors, a standard inequality yields

$$|\cos(u', v') - \cos(u, v)| \leq C (\|u' - u\|_2 + \|v' - v\|_2),$$

for a constant C depending on lower and upper bounds on $\|u\|_2$, $\|v\|_2$, $\|u'\|_2$, and $\|v'\|_2$.

This establishes a Lipschitz-type bound governing **score stability** under bounded perturbations of TF-IDF vectors. Such bounds control the magnitude of numerical changes in similarity scores but do not, by themselves, determine the stability of induced rankings when deterministic tie-breaking rules are present (see §4.5), particularly in regimes where score-separation margins are small or ties are present.

4.4. Ranking Stability via Score-Separation Margins (A1)

Let

$$(r_1, r_2, \dots, r_n)$$

denote the ranking induced by sorting similarity scores in decreasing order, and let $\text{score}(r_j)$ denote the similarity score of the document at rank j .

The **boundary margin** at rank k is defined as

$$m_k = \text{score}(r_k) - \text{score}(r_{k+1}).$$

If similarity scores are subject to a uniform perturbation bounded by ε , i.e.

$$|\Delta s_i| \leq \varepsilon \quad \text{for all } i,$$

then the top- k set is invariant under perturbation whenever

$$\varepsilon < \frac{m_k}{2}.$$

Similarly, the **minimum adjacent margin within the top- k** is defined as

$$m_{\min}^{\text{top}} = \min_{1 \leq j < k} (\text{score}(r_j) - \text{score}(r_{j+1})).$$

Under the same uniform bound, the ordering within the top- k set is preserved whenever

$$\varepsilon < \frac{m_{\min}^{\text{top}}}{2}.$$

These conditions provide explicit, sufficient criteria for ranking stability in terms of score-separation margins. They depend only on similarity scores and do not account for secondary ordering rules applied in the presence of ties.

4.5. Tie-Breaking Discontinuities and Decision Sensitivity (A2)

Let a_i denote a vector of secondary attributes associated with document i (e.g. popularity, rating, engagement, identifier). The final ranking operator is defined as

$$\pi = \text{Sort}(s_i, a_i),$$

where similarity scores form the primary key and secondary attributes are applied lexicographically to resolve ties.

To isolate the effect of tie-breaking, define a **score-only ranking**

$$\pi_{\text{score}} = \text{Sort}(s_i, id_i),$$

where id_i is a fixed, stable identifier used solely to impose a deterministic but attribute-independent order among equal scores.

An **alternate tie-break ranking** is defined by reordering the priority of secondary attributes:

$$\pi_{\text{alt}} = \text{Sort}(s_i, a_i \text{ with reordered priority}).$$

Fix a numerical tolerance $\tau > 0$ and consider cases where the boundary margin satisfies

$$m_k \leq \tau.$$

In this regime, documents near the top- k boundary may form a tie group in which

$$|s_i - s_j| \leq \tau.$$

Even when

$$\Delta s_i \approx 0,$$

the top- k set or ordering may differ between π , π_{score} , and π_{alt} , due solely to the choice of deterministic tie-breaking rules.

This motivates a notion of **tie-break sensitivity**, which can be measured, for example, by:

- an indicator of whether the top- k set differs between π and π_{score} , and
- a distance between orderings restricted to tie groups (e.g. inversion count or Kendall τ distance).

These quantities capture **decision-level instability** that is independent of numerical error in similarity computation and arises purely from secondary ordering criteria.

5. Interpretation and Scope

The preceding formulation highlights several structural features of **TF-IDF-based similarity systems** that become especially clear when the pipeline is expressed in **operator-level form**:

- **IDF sensitivity is governed by explicit logarithmic dependence** on corpus size and document-frequency counts, as seen in

$$\Delta \text{idf}(t) = \log\left(\frac{1 + N'}{1 + \text{df}'(t)}\right) - \log\left(\frac{1 + N}{1 + \text{df}(t)}\right)$$

This makes the stability of IDF directly traceable to perturbations in corpus composition.

- **TF-IDF perturbations admit an explicit decomposition** into *local* (TF), *global* (IDF), and *second-order interaction* terms, providing a transparent mechanism for understanding how small edits propagate through the embedding.
- **Cosine similarity admits a geometric interpretation** as the cosine of the angle between *sparse, non-negative vectors*. This framing clarifies how sparsity patterns and IDF scaling influence angular distortion under perturbation.
- **Ranking robustness can be characterised in terms of score-separation margins**, with explicit sufficient conditions ensuring invariance under bounded perturbations. In the context of **content-based k -nearest-neighbour recommendation**, such ranking stability directly governs the stability of induced neighbourhoods.

- **Ranking stability is governed primarily by margin distributions, rather than by aggregate or average score changes alone.** In particular, the presence of small score-separation margins near decision boundaries dominates stability behaviour, even when aggregate similarity scores are numerically well-conditioned.
- **A long tail of near-zero margins can imply rare but extreme fragility.** While most rankings may be stable under small perturbations, documents near top- k boundaries with vanishing margins can induce abrupt changes in neighbourhood structure under otherwise negligible score variation.
- **Deterministic tie-breaking introduces non-perturbative discontinuities.** Even in the absence of meaningful numerical perturbation ($\Delta s \approx 0$), ranking outcomes may change solely due to secondary ordering rules, underscoring that stability of computed similarities does not guarantee stability of downstream decisions.

The emphasis throughout is on **derivational transparency** rather than algorithmic optimisation. No dimensionality reduction, latent-semantic modelling, or neural embeddings are introduced. The goal is to expose the **algebraic and geometric structure** of the TF-IDF pipeline in a form suitable for **perturbation analysis, stability reasoning, and controlled experimentation**.

6. Limitations

Several limitations of the present framework should be noted:

- **High-dimensional sparsity complicates geometric intuition.**
TF-IDF vectors inhabit a large, sparse subset of $\mathbb{R}^n$, where small changes in support can produce disproportionately large angular effects. This can lead to instability in similarity scores and, by extension, in induced k -NN neighbourhoods.
- **Cosine similarity becomes unstable for low-norm vectors.**
When documents are short or contain few in-vocabulary tokens, the norms $\|u\|_2$ and $\|v\|_2$ may become small, increasing sensitivity to perturbations in similarity values. In such regimes, score-separation margins tend to shrink, making ranking outcomes increasingly susceptible to domination by deterministic tie-breaking rules rather than by similarity geometry alone.
- **Smoothed IDF reduces but does not eliminate volatility associated with rare tokens.**
Tokens with very low document frequency remain highly sensitive to corpus perturbations, even under smoothing, and can dominate similarity computations in sparse settings.
- **Deterministic tie-breaking can dominate outcomes under near-ties.**
When score-separation margins fall below numerical tolerance, ranking outcomes may be determined primarily by secondary ordering rules. In such cases, stability results based solely on similarity scores or margin conditions do not fully capture decision-level instability in the induced rankings.
- **No numerical optimisation techniques are employed.**
The implementation intentionally avoids stabilising transformations, such as sublinear TF scaling, vector normalisation variants, or dimensionality reduction, in order to preserve **analytical clarity** and direct interpretability of perturbation effects.

Accordingly, the framework is intended for **analytical, educational, and controlled experimental use**, rather than production deployment or large-scale retrieval tasks.

7. Experimental Protocol and Results

This section presents the evaluation protocol and empirical results corresponding to the two stability questions analysed throughout this work:

- **(A1)** How score-separation margins govern the stability of similarity-based rankings under bounded perturbations.
- **(A2)** How deterministic tie-breaking rules induce decision-level discontinuities in the presence of near-ties.

The purpose of the experiments is not to benchmark retrieval quality against external baselines, but to characterise stability and fragility properties of the implemented TF-IDF + cosine ranking pipeline under a fixed, deterministic preprocessing, embedding, and ranking configuration.

7.1. Query Set and Evaluation Setup

Corpus and Representation.

All documents are embedded using the fixed preprocessing map, vocabulary construction, smoothed IDF definition, and TF normalisation specified in §§2.1–2.2. Similarity scores are computed exclusively using cosine similarity as defined in §2.3.

Query Construction.

Experiments are conducted using the following query construction strategies:

- **User-profile queries:** a query document is constructed by aggregating text from a user’s interacted items (e.g. liked, viewed, or favourited), and embedded into the same TF-IDF space as the corpus.
- **Leave-one-out evaluation:** for users with multiple interactions, one interacted item is removed and treated as a held-out target; the remaining interactions form the query profile.

Other supported query modes (e.g. item-as-query) are part of the implementation but are not evaluated in the present experiments.

k values.

All stability metrics are evaluated for

$$k \in \{5, 10, 20, 50\}.$$

Near-tie tolerance.

A fixed numerical tolerance $\tau > 0$ defines the near-tie regime. Two similarity scores s_i and s_j are treated as indistinguishable whenever

$$|s_i - s_j| \leq \tau,$$

and tie groups are defined as in §2.3.3. The value of τ is chosen to exceed floating-point noise while remaining small relative to typical score separations. All tie-break sensitivity results are explicitly conditional on this choice of τ .

All reported distributions are computed over the full set of evaluated queries; the exact number of queries and users depends on the dataset configuration.

Unless stated otherwise, all reported results use a fixed τ across queries.

7.2. Margin Distributions and Ranking Stability (A1)

For each query, let

$$(r_1, r_2, \dots)$$

denote the ranking induced by sorting similarity scores in decreasing order. Boundary and within-top- k margins are defined as in §§2.3.2 and §4.4:

$$m_k = \text{score}(r_k) - \text{score}(r_{k+1}),$$

$$m_{\min}^{\text{top}} = \min_{1 \leq j < k} (\text{score}(r_j) - \text{score}(r_{j+1})).$$

For each

$$k \in \{5, 10, 20, 50\},$$

empirical distributions of m_k and $m_{\min}^{\text{top}}$ are computed across all evaluated queries.

Reported Statistics.

- Percentile summaries of m_k and $m_{\min}^{\text{top}}$ (including lower-tail, median, and upper-tail behaviour).
- Corresponding flip radii

$$\varepsilon_k^{\text{flip}} = \frac{m_k}{2},$$

which provide sufficient bounds for top- k stability under uniform score perturbations.

Visualisation.

- Empirical cumulative distribution functions (ECDFs) or histograms of m_k .
- ECDFs or histograms of $m_{\min}^{\text{top}}$.

These distributions characterise whether ranking stability is typical (margins well separated from zero) or fragile (substantial mass concentrated near zero), particularly at decision boundaries.

7.3. Tie-Break Ablations and Decision Sensitivity (A2)

To isolate decision-level effects arising from deterministic secondary ordering, rankings are recomputed under the following sorting operators (§4.5):

- **Full Ranking:**

$$\pi = \text{Sort}(s_i, a_i)$$

- **Score-only Ranking:**

$$\pi_{\text{score}} = \text{Sort}(s_i, id_i)$$

- **Alternate Tie-Break Ranking:**

$$\pi_{\text{alt}} = \text{Sort}(s_i, a_i \text{ with reordered priority})$$

For each query and each k , the following quantities are measured:

- **Top- k disagreement rate:** the fraction of queries for which the top- k set differs between π and π_{score} , and between π and π_{alt} .
- **Within-top- k reordering:** an ordering distance restricted to tie-affected subsets (e.g. inversion count or Kendall τ distance).

Results are stratified by the boundary margin m_k relative to τ , since tie-break effects are expected to concentrate in the near-tie regime.

$$m_k \leq \tau.$$

Visualisation.

- Probability of top- k disagreement as a function of m_k , highlighting the transition region around $m_k \approx \tau$.

This analysis separates numerical stability of similarity scores from instability introduced purely by deterministic decision rules.

7.4. Constructed Near-Tie Case Study

To illustrate decision-level discontinuities explicitly, a constructed near-tie case is reported for a fixed query:

- Two documents A and B are identified such that

$$|s_A - s_B| \leq \tau$$

near the top- k boundary (or as adjacent elements within the top- k).

- The tuple

$$(s_A, s_B, m_k, \tau)$$

and the associated tie-break attributes

$$(a_A, a_B)$$

are reported.

- Ranking outcomes under π , π_{score} , and π_{alt} are compared.

This case study is **deliberately illustrative rather than representative**, and serves to make concrete the fact that ranking outcomes can change with

$$\Delta s \approx 0,$$

driven solely by deterministic tie-breaking rules. The example directly instantiates the decision-level discontinuities analysed abstractly in §§2.3.3 and 4.5.

8. References

The following sources provide the theoretical, numerical, and conceptual background that informs the present work.

Classical Information Retrieval

- [3] Thomas M. Cover and Peter E. Hart. “Nearest neighbor pattern classification”. In: *IEEE Transactions on Information Theory* 13.1 (1967).
- [7] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [8] Gerard Salton and Christopher Buckley. “Term-weighting approaches in automatic text retrieval”. In: *Information Processing & Management* 24.5 (1988).
- [13] Justin Zobel and Alistair Moffat. “Inverted files for text search engines”. In: *ACM Computing Surveys* 38.2 (2006).

Numerical Linear Algebra and Stability

- [5] Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. 2nd ed. SIAM, 2002.
- [11] Lloyd N. Trefethen and David Bau. *Numerical Linear Algebra*. SIAM, 1997.

Sparse Vector Geometry and Similarity

- [1] Charu C. Aggarwal. *Data Mining: The Textbook*. Springer, 2015.
- [6] Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman. *Mining of Massive Datasets*. Cambridge University Press, 2020.

Statistical Learning Context

- [4] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer, 2009.
- [9] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

Online Resources

- [2] *Cosine similarity*. Accessed 15 December 2025. 2025. URL: https://en.wikipedia.org/wiki/Cosine_similarity.
- [10] *TF-IDF*. Accessed 15 December 2025. 2025. URL: <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>.
- [12] *Vector space model*. Accessed 15 December 2025. 2025. URL: https://en.wikipedia.org/wiki/Vector_space_model.

9. Authorship

Implementation and exposition

Matthew Maksymilian Miezaniec
Email: matthewmiezaniec1@gmail.com

Implementation includes the full TF-IDF similarity pipeline, explicit perturbation analysis tooling, **stability profiling via score-separation margins**, and a **tie-break ablation framework** for isolating decision-level discontinuities in ranking outcomes.

Mathematical and theoretical foundations

This work draws on **classical information retrieval methodology**, including TF-IDF weighting and vector-space models [8, 7], and is supported by established treatments of **numerical stability, conditioning**, and **perturbation behaviour** in high-dimensional vector spaces [5, 11].

The interpretation of similarity scores in terms of **content-based k -nearest-neighbour ranking and neighbourhood structure** follows classical nearest-neighbour and similarity-search perspectives [3], without introducing learning-based methods.

Broader contextual connections to feature representations and similarity-based reasoning draw on standard statistical learning references [4, 9]. Supplementary intuition and terminology are informed by widely used online references on **TF-IDF, cosine similarity**, and **vector-space models**.

10. Acknowledgements

The author thanks colleagues and peers for informal discussions that helped clarify aspects of numerical stability, sparse vector geometry, and similarity-based reasoning. The work also benefited from exposure to standard academic treatments of information retrieval and numerical linear algebra through coursework, independent study, and open-source documentation.

Any remaining errors or omissions are the responsibility of the author alone. This acknowledgment does not imply endorsement or direct contribution by any individual or institution.