

# TasteRank Explorer

## Summary

*Eigenvector Centrality on Sensory Profiles · Plain-Language Overview*

Jure Skarabot · May 2026

*Companion documents: Summary · Technical Appendix · Methods Primer · Data Appendix · Grape Reference*

Vinotheca · published under CC BY-NC 4.0

---

**Abstract.** *Wine grape varieties are conventionally classified by colour, region of origin, or stylistic convention. These categories are useful navigational heuristics, but they obscure a structural reality: grape varieties exist in a continuous space of sensory similarity, in which some occupy central, archetypal positions and others sit at the periphery. TasteRank Explorer is a network-analytic framework that makes this structure explicit. One hundred and one grape varieties are encoded as thirteen-dimensional sensory profile vectors drawn from established tasting methodology. Cosine similarity between profiles defines a dense similarity matrix; a five-nearest-neighbour graph sparsifies it into a network of 341 edges. Eigenvector centrality on the resulting weighted graph defines the TasteRank score, which ranks varieties by structural importance — by how many other varieties resemble them, weighted by the centrality of those resemblances in turn. Greedy modularity optimisation partitions the graph into six communities. The principal findings: centrality concentrates in full-bodied Mediterranean reds; the noble international varieties (Pinot Noir, Nebbiolo, Riesling) rank low precisely because their distinctive profiles place them on the periphery; and the red-white boundary dissolves at the structural level, with ultra-light reds clustering with aromatic whites. PageRank and sensitivity analyses confirm that the structure is intrinsic to the data rather than an artefact of parameter choice.*

---

## 1. Introduction

Wine grape varieties are traditionally classified by colour (red versus white), region of origin, or stylistic convention. While these categories are useful heuristics, they obscure a deeper structural reality: grape varieties exist in a continuous space of sensory similarity, in which some varieties occupy central, archetypal positions and others sit at the periphery.

TasteRank is a network-analytic framework that makes this structure explicit. The framework rests on a simple act of translation. Each grape variety is described by a sensory profile across thirteen dimensions that have, in one form or another, organised structured wine tasting since Peynaud's work at Bordeaux and the analytical traditions that grew up around UC Davis: colour depth, aromatic intensity, the balance of fruit and herb, the architecture of acid, tannin, body, and finish. The choice of these particular thirteen, and the scores assigned to each variety, are the author's own; the underlying vocabulary is the common inheritance of modern oenology. Once a variety becomes a vector in this thirteen-dimensional space, the rest follows as geometry.

Cosine similarity measures how alike two profiles are. A sparse graph of nearest neighbours makes the resulting structure tractable. Eigenvector centrality, applied to that graph, ranks varieties not by reputation or rarity but by how densely they sit at the centre of the tasting universe — how many other varieties resemble them, and how central those resemblances are in turn.

The result is a principled, reproducible ranking that answers a question no wine textbook addresses directly: which grape varieties are the most typical — that is, which varieties sit at the dense centre of sensory space, surrounded by many other similar varieties? High TasteRank identifies the archetypes; low TasteRank identifies the outliers.

## 2. Data

### 2.1 Grape Variety Universe

The dataset comprises 101 grape varieties (53 red, 48 white) selected for global significance, regional diversity, and representation across the canonical literature of wine education. The selection spans the canonical international varieties (Cabernet Sauvignon, Chardonnay, Pinot Noir) through to specialised regional varieties (Xinomavro, Assyrtiko, Blaufränkisch, Plavac Mali).

### 2.2 Sensory Profile Encoding

Each variety is encoded as a thirteen-dimensional vector. The dimensions are drawn from the structured tasting tradition that runs from Peynaud through the modern descriptive frameworks (UC Davis, AWRI, and the various national wine schools that grew from them); they were chosen to span the sensory axes that experienced tasters consistently report as discriminating between varieties, while remaining few enough to be scored reliably on a coarse 0–5 scale. The thirteen dimensions used here are:

| #  | Dimension          | Scale | Description                                |
|----|--------------------|-------|--|
| 1  | Color Depth        | 0–5   | Intensity of pigmentation (pale to opaque) |
| 2  | Aromatic Intensity | 0–5   | Strength of nose from low to pronounced    |
| 3  | Floral Character   | 0–5   | Degree of floral aromatic contribution     |
| 4  | Fruit Ripeness     | 0–5   | Ripe/dried fruit vs. underripe/green fruit |
| 5  | Herbal/Earthy      | 0–5   | Green herb, earth, mineral notes           |
| 6  | Spice/Oak          | 0–5   | Oak-derived and spice character            |
| 7  | Acidity            | 0–5   | Perceived acidity from low to high         |
| 8  | Tannin             | 0–5   | Tannin level (reds); 0 for most whites     |
| 9  | Body               | 0–5   | Weight and texture on palate               |
| 10 | Alcohol            | 0–5   | Perceived alcohol warmth                   |
| 11 | Flavor Intensity   | 0–5   | Palate flavour concentration               |
| 12 | Finish             | 0–5   | Length and persistence of aftertaste       |
| 13 | Complexity         | 0–5   | Layeredness and evolution of character     |

Scores reflect canonical varietal typicality — the profile a well-made, representative example of the variety would be expected to present — and are the author’s own judgements, calibrated against the standard ampelographic literature (Robinson, Harding & Vouillamoz; d’Agata; Galet) and

against years of structured tasting. They describe the variety, not any particular wine or vintage. A given Burgundian Pinot Noir might score higher in earth and lower in fruit ripeness than the canonical profile suggests; an Oregon Pinot might invert that relationship. The TasteRank score is best read as a statement about the variety’s centre of gravity, not its full range.

### 3. Methodology

#### 3.1 Cosine Similarity

Pairwise similarity between grape varieties is measured using cosine similarity on the thirteen-dimensional sensory profile vectors. For varieties  $i$  and  $j$  with profile vectors  $\mathbf{p}_i$  and  $\mathbf{p}_j$ :

$$\cos(\mathbf{p}_i, \mathbf{p}_j) = \frac{\mathbf{p}_i \cdot \mathbf{p}_j}{\|\mathbf{p}_i\| \|\mathbf{p}_j\|}.$$

Cosine similarity is preferred over Euclidean distance because it measures angular proximity in profile space, making it invariant to the overall magnitude of scores and sensitive only to the shape of the sensory profile. Two varieties with proportionally similar profiles will have high cosine similarity even if one consistently scores higher across all dimensions.

#### 3.2 $k$ -Nearest Neighbour Graph Construction

From the full  $101 \times 101$  similarity matrix we construct a sparse  $k$ -nearest neighbour ( $k$ NN) graph with  $k = 5$ . For each variety, the five most similar varieties are connected by edges weighted by their cosine similarity. The graph is symmetrised by union: an edge exists between varieties  $A$  and  $B$  if  $A$  includes  $B$  in its top-5 or  $B$  includes  $A$  in its top-5. The edge weight is the cosine similarity. This produces a graph with 101 nodes and 341 edges — substantially sparser than the complete graph (which would have 5,050 edges) but dense enough to reveal community structure and centrality patterns.

#### 3.3 Eigenvector Centrality (TasteRank)

TasteRank is defined as the eigenvector centrality of each node in the weighted  $k$ NN graph. For the weighted adjacency matrix  $W$  (a symmetric, non-negative  $101 \times 101$  matrix), the TasteRank score of variety  $i$  is the  $i$ -th component of the leading eigenvector  $\mathbf{x}_1$ , satisfying

$$W\mathbf{x}_1 = \lambda_1\mathbf{x}_1,$$

where  $\lambda_1$  is the largest eigenvalue. Three properties of this definition are essential.

First, TasteRank is an implicit, system-level definition. There is no closed-form formula  $\text{TR}(i) = f(\dots)$  that can be evaluated for a single variety in isolation. The 101 scores are defined by a system of 101 coupled equations that must be solved simultaneously — every variety’s score depends on every other variety’s score. This is exactly analogous to Google’s PageRank, where no individual web page has an independent rank formula. A variety’s TasteRank is not a local property of its profile or its neighbours; it is a global property of its position within the entire network.

Second, only the largest eigenvalue matters. When influence is propagated iteratively through the network ( $\mathbf{z}^t = W^t\mathbf{z}^0$ ), the eigendecomposition shows that all contributions from  $\lambda_2, \lambda_3, \dots$

decay exponentially as  $(\lambda_k/\lambda_1)^t \rightarrow 0$ , leaving only  $\mathbf{x}_1$  as the steady-state distribution. The spectral gap  $|\lambda_2/\lambda_1| \approx 0.72$  for this graph ensures rapid convergence (35–45 iterations) and reflects a clear core–periphery structure.

Third, the logic is recursive. Written component-wise, the eigenvector equation becomes

$$x_i = \frac{1}{\lambda_1} \sum_j W_{ij} x_j.$$

A variety’s score is proportional to the similarity-weighted sum of its neighbours’ scores, which depend on their neighbours’ scores, and so on to infinite depth. The eigenvector is the unique self-consistent solution to this infinite recursion. Sagrantino (rank 1) is connected to five other top-10 varieties, each connected to further high-centrality reds — recursive amplification through a dense, mutually reinforcing cluster. Riesling (rank 85) is connected to peripheral varieties whose neighbours are also low-centrality — recursive dampening through sparse, isolated connections. In TasteRank terms, distinctiveness and centrality are inversely related.

The full mathematical treatment — eigendecomposition, spectral gap analysis, the power iteration algorithm, and the role of higher eigenvalues in spectral embedding — is provided in the Technical Appendix (Sections A.5.1–A.5.6).

### 3.4 PageRank as Structural Counterpoint

PageRank modifies eigenvector centrality by introducing a damping factor  $\alpha = 0.85$  that models a random taster who follows similarity edges 85% of the time and teleports to a random variety 15% of the time. This prevents centrality from concentrating exclusively in the densest cluster and creates a diagnostic contrast: PageRank rewards bridge varieties connecting different communities, while eigenvector centrality rewards varieties deep within the densest core. The Spearman rank correlation between the two measures is  $\rho \approx 0.92$ , confirming robust agreement. The most informative divergence is Sangiovese (TasteRank rank 40, PageRank  $\sim 8$ ) — a bridge spanning Communities C0 and C2. Full derivation in Technical Appendix A.6.

### 3.5 Community Detection

The Clauset–Newman–Moore greedy modularity algorithm partitions the graph into groups more densely connected internally than expected under a random null model, maximising

$$Q = \frac{1}{2m} \sum_{i,j} \left[ W_{ij} - \frac{s_i s_j}{2m} \right] \delta(c_i, c_j).$$

The algorithm detects six communities with  $Q \approx 0.41$ , indicating strong structure. The number of communities is consistent with the eigenvalue spectrum of the modularity matrix, where the first six eigenvalues are clearly separated from the bulk. Full details in Technical Appendix A.7.

## 4. Results

### 4.1 TasteRank Rankings (Top 20)

The top twenty varieties by TasteRank are overwhelmingly full-bodied reds from Community C0, the Mediterranean / Southern Italian cluster. Sagrantino ranks first with a TasteRank of

0.3020 — its extreme profile (maximum colour depth, maximum tannin, maximum body) makes it the prototypical archetype of the densest cluster.

| Rank | Variety            | Type | Community | TasteRank |
|------|--------------------|------|-----------|-----------|
| 1    | Sagrantino         | Red  | C0        | 0.3020    |
| 2    | Nero d’Avola       | Red  | C0        | 0.2849    |
| 3    | Lagrein            | Red  | C0        | 0.2748    |
| 4    | Negroamaro         | Red  | C0        | 0.2380    |
| 5    | Plavac Mali        | Red  | C0        | 0.2251    |
| 6    | Montepulciano      | Red  | C0        | 0.2108    |
| 7    | Petit Verdot       | Red  | C0        | 0.2093    |
| 8    | Monastrell         | Red  | C0        | 0.2060    |
| 9    | Petite Sirah       | Red  | C0        | 0.1768    |
| 10   | Pinotage           | Red  | C0        | 0.1710    |
| 11   | Malbec             | Red  | C0        | 0.1697    |
| 12   | Tannat             | Red  | C0        | 0.1676    |
| 13   | Garnacha Tintorera | Red  | C0        | 0.1670    |
| 14   | Dolcetto           | Red  | C0        | 0.1659    |
| 15   | Mourvèdre          | Red  | C0        | 0.1565    |
| 16   | Zinfandel          | Red  | C0        | 0.1498    |
| 17   | Primitivo          | Red  | C0        | 0.1498    |
| 18   | Merlot             | Red  | C0        | 0.1333    |
| 19   | Carignan           | Red  | C0        | 0.1283    |
| 20   | Limniona           | Red  | C2        | 0.1223    |

## 4.2 Community Structure

Modularity optimisation detects six communities. The most structurally significant finding is Community C1, which bridges the red–white divide: its four red members (Schiava, Poulsard, Frappato, Kadarka) are ultra-light reds with minimal tannin, positioning them closer to aromatic whites than to their nominal red peers.

| ID | Character                       | Size | R / W  | Hub varieties                     |
|----|---------------------------------|------|--------|-----------------------------------|
| C0 | Full-bodied Mediterranean Reds  | 31   | 31 / 0 | Sagrantino, Nero d’Avola, Lagrein |
| C1 | Light / Aromatic Cross-boundary | 21   | 4 / 17 | Assyrtiko, Falanghina, Albariño   |
| C2 | Mid-weight Structured Reds      | 18   | 18 / 0 | Limniona, St. Laurent, Zweigelt   |
| C3 | Rich Textural Whites            | 13   | 0 / 13 | Marsanne, Fiano, Godello          |
| C4 | Mineral / Crisp Whites          | 10   | 0 / 10 | Greco, Friulano, Verdicchio       |
| C5 | Lean Neutral & Aromatic Whites  | 8    | 0 / 8  | Gewürztraminer, Petit Manseng     |

## 4.3 Key Structural Findings

**Centrality concentration in Southern Italian / Mediterranean reds.** The top nine varieties all belong to Community C0. These are full-bodied, high-tannin, high-colour reds with profiles that cluster tightly in the thirteen-dimensional space. Sagrantino, Nero d’Avola, and Lagrein form the innermost core.

**Noble varieties rank lower than expected.** Cabernet Sauvignon ranks 30th, Pinot Noir 47th, Nebbiolo 42nd. These varieties have distinctive profiles that set them apart from the dense centre — high complexity and finish, but with unique trait combinations (e.g. Pinot Noir’s low colour and tannin; Nebbiolo’s extreme acidity and tannin with moderate colour). TasteRank measures centrality, not quality.

**The red–white boundary dissolves in Community C1.** Four light reds (Schiava, Poulsard, Frappato, Kadarka) cluster with aromatic whites. The cosine similarity between Frappato and Loureiro exceeds 0.98 — higher than Frappato’s similarity to most other reds. This confirms algorithmically what experienced tasters observe.

**White wine centrality peaks with medium-bodied textural varieties.** Among whites, Marsanne (rank 54) and Fiano (rank 57) lead. These medium-bodied, moderately complex whites occupy the centre of the white-grape subspace. The most distinctive whites — Riesling (rank 85), Sauvignon Blanc (rank 90) — rank near the bottom, precisely because their extreme acidity and aromatic profiles set them apart.

## 5. Robustness Checks

### 5.1 TasteRank vs. PageRank: Empirical Comparison

PageRank provides a structural counterpoint to eigenvector centrality by introducing a teleportation mechanism that prevents extreme concentration. The Spearman rank correlation between TasteRank and PageRank across all 101 varieties is  $\rho \approx 0.92$  — strong enough to confirm that the centrality structure is not an artefact of the choice of spectral measure, but with enough divergence to be informative.

The divergence pattern is systematic: varieties in the dense core of Community C0 have  $\text{TasteRank} > \text{PageRank}$  (recursive amplification within the cluster inflates their eigenvector centrality), while bridge varieties connecting multiple communities have  $\text{PageRank} > \text{TasteRank}$  (the random taster passes through them frequently when transitioning between clusters). Sangiovese is the most dramatic case: rank 40 by TasteRank but approximately rank 8 by PageRank, reflecting its 7 edges spanning Communities C0 and C2. The magnitude of the gap  $|\text{TasteRank rank} - \text{PageRank rank}|$  thus serves as a diagnostic for structural role: large positive gaps identify core members; large negative gaps identify bridges.

### 5.2 Sensitivity to $k$

Varying the neighbourhood parameter  $k$  from 3 to 7 produces rank correlations of  $\rho > 0.94$  for the top thirty varieties. The community structure is stable for  $k = 4$  through  $k = 7$  (six communities detected in each case). At  $k = 3$ , the graph fragments into seven communities as some white-grape components disconnect; at  $k \geq 8$ , the graph becomes too dense and community resolution degrades. The choice of  $k = 5$  represents a principled balance between sparsity and connectedness.

### 5.3 Spectral Gap Stability

The spectral gap  $|\lambda_2/\lambda_1| \approx 0.72$  is stable across  $k = 4$  to  $k = 7$ , confirming that the core–periphery structure is intrinsic to the data rather than an artefact of graph construction. A

spectral gap this large indicates a single clearly dominant mode of centrality: the Mediterranean red cluster forms a robust core that persists regardless of parameter choices. The gap also ensures rapid convergence of power iteration (35–45 iterations at  $\varepsilon = 10^{-6}$ ), making the computation numerically stable and reproducible.

## 6. Conclusion

### 6.1 Key Findings

TasteRank provides a principled, reproducible framework for understanding the structural organisation of the wine grape universe through the lens of sensory similarity. Three findings emerge from the analysis.

First, centrality and quality are orthogonal. The varieties with the highest TasteRank — Sagrantino, Nero d’Avola, Lagrein — are not the world’s most celebrated grapes. They are the most typical: their sensory profiles sit at the gravitational centre of the red-wine subspace, surrounded by many similar varieties. The noble varieties (Pinot Noir, Nebbiolo, Riesling) rank low precisely because their distinctive profiles push them to the periphery. In the TasteRank framework, distinctiveness and centrality are inversely related.

Second, the red–white boundary is permeable. Community C1 demonstrates that at the structural level, ultra-light reds (Schiava, Frappato, Poulsard) and aromatic whites (Loureiro, Müller-Thurgau, Welschriesling) share more sensory profile structure than either group shares with its nominal colour peers. The tasting universe does not cleave neatly into red and white hemispheres — it contains a liminal zone in which colour is secondary to structure.

Third, the framework is extensible. The same architecture can incorporate weighted expert similarity, regional terroir dimensions, or empirical tasting data from structured experiments. The TasteRank Explorer visualisation makes the network navigable and explorable, transforming an abstract mathematical structure into an interactive tool for wine education and discovery.

### 6.2 Limitations and Future Work

Three limitations of the present analysis bear emphasis.

The sensory profiles are scored by a single annotator. Even calibrated against the standard ampelographic literature, single-annotator scoring carries the risk of systematic bias. A principled extension would aggregate scores across multiple expert tasters and treat the inter-annotator variance as a regularisation signal in the cosine similarity computation.

The thirteen-dimensional encoding is fixed. The dimensions were chosen for parsimony and interpretability rather than for empirical optimality. An alternative encoding — derived, for example, from factor analysis of larger structured tasting datasets — might reveal centrality structure that the present encoding suppresses or amplifies structure that is largely an artefact of the chosen dimensions.

The framework is presently static. It treats the variety universe as a closed set of 101 grapes. Extending TasteRank to admit new varieties (incremental updates without recomputing the full eigenvector) and to track structural change over time (vintage-to-vintage shifts in profile, response to climate change) is a natural next direction. The same machinery applies; only the data layer needs to evolve.

## References

- Amerine, M. A. & Roessler, E. B. (1976). *Wines: Their Sensory Evaluation*. W. H. Freeman.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1), 113–120.
- Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1–7), 107–117.
- Clauset, A., Newman, M. E. J. & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.
- Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
- Peynaud, É. (1980). *Le Goût du Vin*. Dunod, Paris.
- Robinson, J., Harding, J. & Vouillamoz, J. (2012). *Wine Grapes: A Complete Guide to 1,368 Vine Varieties*. Ecco / HarperCollins.