TasteRank Explorer

Eigenvector Centrality on WSET SAT Tasting Profiles

Summary Document

101 grape varieties · 341 similarity edges · 6 communities

Jure Skarabot

April 2026

Table of Contents

1. Introduction

Wine grape varieties are traditionally classified by color (red vs. 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, where some varieties occupy central, archetypal positions and others sit at the periphery.

TasteRank is a network-analytic framework that makes this structure explicit. Starting from the WSET Systematic Approach to Tasting (SAT), we encode each grape variety as a vector in a 13-dimensional sensory space, construct a similarity graph using cosine similarity, and compute eigenvector centrality to rank varieties by their structural importance in the tasting universe.

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 center of sensory space, surrounded by many other similar varieties? High TasteRank identifies the archetypes; low TasteRank identifies the outliers.

2. Data and Feature Construction

2.1 Grape Variety Universe

The dataset comprises 101 grape varieties (53 red, 48 white) selected for global significance, regional diversity, and representation across WSET Diploma-level curricula. The selection spans the canonical international varieties (Cabernet Sauvignon, Chardonnay, Pinot Noir) through to specialized regional varieties (Xinomavro, Assyrtiko, Blaufränkisch, Plavac Mali).

2.2 SAT Profile Encoding

Each variety is encoded as a 13-dimensional vector based on the WSET Systematic Approach to Tasting. The dimensions capture the core sensory axes of wine assessment:

#	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 flavor concentration
12	Finish	0–5	Length and persistence of aftertaste

13	Complexity	0–5	Layeredness and evolution of character

Scores are assigned based on canonical varietal typicity as codified in WSET Level 3 and Diploma study materials, supplemented by structured tasting experience. Each score represents the expected profile of a well-made, representative example of the variety—not any single wine or vintage.

3. Methodology

3.1 Cosine Similarity

Pairwise similarity between grape varieties is measured using cosine similarity on the 13-dimensional SAT profile vectors. For varieties i and j with profile vectors p_i and p_j:

$$\cos(p_i, p_j) = (p_i \cdot p_j) / (||p_i|| \cdot ||p_j||)$$

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 Neighbor Graph Construction

From the full 101 × 101 similarity matrix, we construct a sparse k-nearest neighbor (kNN) graph with $k = 5$. For each variety, the five most similar varieties are connected by edges weighted by their cosine similarity. The graph is symmetrized using maximum-weight union: if variety A includes B in its top-5 or B includes A in its top-5, the edge exists with weight equal to their 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 rich 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 kNN graph. For the weighted adjacency matrix W (a symmetric, non-negative 101 × 101 matrix), the TasteRank score of variety i is the i-th component of the leading eigenvector x_1, satisfying:

$$W \cdot x_1 = \lambda_1 \cdot x_1$$

where λ_1 is the largest eigenvalue. Three properties of this definition are essential to understanding TasteRank.

First, TasteRank is an implicit, system-level definition. There is no closed-form formula TR(i) = f(…) 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 neighbors—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 ($z^t = W^t \cdot z^0$), the eigendecomposition shows that all contributions from λ_2, λ_3, … decay exponentially as $(\lambda_{\square}/\lambda_1)^t \to 0$, leaving only 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:

$$x_i = (1/\lambda_1) \cdot \Sigma_j\, W_{ij} \cdot x_j$$

A variety's score is proportional to the similarity-weighted sum of its neighbors' scores, which depend on their neighbors' 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 neighbors are also low-centrality—recursive dampening through sparse, isolated connections. In TasteRank terms, distinctiveness and centrality are inversely related.

The full mathematical treatment—eigendecomposition proof, spectral gap analysis, 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 ~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, maximizing:

$$Q = (1/2m) \cdot \Sigma_{ij}\, [W_{ij} - (s_i \cdot s_j \,/\, 2m)] \cdot \delta(c_i, c_j)$$

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 20 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 color depth, maximum tannin, maximum body) makes it the prototypical archetype of the densest cluster.

Rank	Variety	Type	Community	TasteRank	Note
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

The modularity optimization 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, Negroamaro
C1	Light / Aromatic Cross-boundary	21	4 / 17	Assyrtiko, Falanghina, Albariño, Riesling
C2	Mid-weight Structured Reds	18	18 / 0	Limniona, St. Laurent, Zweigelt, Blaufränkisch
C3	Rich Textural Whites	13	0 / 13	Marsanne, Fiano, Godello, Pinot Gris
C4	Mineral / Crisp Whites	10	0 / 10	Greco, Friulano, Verdicchio, Verdejo
C5	Lean Neutral & Aromatic Whites	8	0 / 8	Gewürztraminer, Petit Manseng, Malvasia

4.3 Key Structural Findings

Centrality concentration in Southern Italian/Mediterranean reds. The top 9 varieties all belong to Community C0. These are full-bodied, high-tannin, high-color reds with profiles that cluster tightly in the 13-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 center—high complexity and finish, but with unique trait combinations (e.g., Pinot Noir's low color and tannin, Nebbiolo's extreme acidity and tannin with moderate color). 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 sommeliers observe experientially.

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 center 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

As described in Section 3.4, 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 artifact 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 TasteRank > PageRank (recursive amplification within the cluster inflates their eigenvector centrality), while bridge varieties connecting multiple communities have PageRank > 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 |TasteRank rank − 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 neighborhood parameter k from 3 to 7 produces rank correlations of $\rho > 0.94$ for the top 30 varieties. The community structure is stable at $k = 4$ 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 $k = 8+$, 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 $k = 4$–7, confirming that the core-periphery structure is intrinsic to the data rather than an artifact of graph construction. A spectral gap this large indicates a single clearly dominant mode of centrality: the Mediterranean red cluster forms a robust gravitational 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

TasteRank provides a principled, reproducible framework for understanding the structural organization of the wine grape universe through the lens of sensory similarity. Three key 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 center 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 DNA than either group shares with its nominal color peers. The tasting universe does not cleave neatly into red and white hemispheres—it contains a liminal zone where color 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 visualization makes the network navigable and explorable, transforming an abstract mathematical structure into an interactive tool for wine education and discovery.

References

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.

WSET (2023). WSET Level 3 Award in Wines: Study Guide. Wine & Spirit Education Trust, London.

Robinson, J., Harding, J. & Vouillamoz, J. (2012). Wine Grapes: A Complete Guide to 1,368 Vine Varieties. Ecco/HarperCollins.