GP-Cosine: Stationary-Centered Diffusion Geometry for Multiway Community Detection

Duy Hieu DO[a] and Thi Ha Duong PHAN[b]

[a]Institute of Mathematics, Vietnam Academy of Science and Technology
[b]The International Centre for Research and Postgraduate Training in Mathematics under the auspices of UNESCO, Institute of Mathematics, Vietnam Academy of Science and Technology

Email: ddhieu@math.ac.vn (Duy Hieu DO), phanhaduong@math.ac.vn (Thi Ha Duong PHAN).

Abstract

Random-walk diffusion profiles encode mesoscopic graph structure, but their raw directions become less informative as the walk approaches stationarity. We study a stationary-centered representation that first normalizes each t-step transition profile and then subtracts the normalized stationary distribution. In the two-community setting, we show that for a connected non-bipartite graph with a dominant positive second random-walk mode, the centered directions asymptotically recover the sign geometry of that mode. This motivates a simple seed-based binary GP-Cosine rule. For any prescribed $K \geq 2$, including $K = 2$, we then define *K-GP-Cosine* by clustering the normalized stationary-centered directions with spherical K-means, yielding all communities simultaneously without recursive splitting.

Experiments use degree-corrected, hierarchical, and Gaussian latent-position graphs, with comparisons to transition-profile K-means ablations and graph-native spectral baselines. Under a common oracle tuning protocol, it attains the highest mean normalized mutual information (NMI) in most tested regimes across graph sizes up to $N = 4000$. With K fixed and only the walk length varied, it also achieves the strongest accuracy–stability tradeoff among the directly comparable diffusion methods. On ten real networks, modularity-based parameter selection places K-GP-Cosine first on six networks and tied first on one. These results support stationary centering as a useful representation for direct multiway diffusion-based community detection.

1 Introduction

Community detection seeks a partition of a network into vertex groups that are internally cohesive and externally separated. It is a central problem in network science, with applications ranging from social and information networks to biological and technological systems [12]. Spectral methods provide one of the classical approaches: Laplacian eigenvectors reveal low-dimensional geometry associated with graph cuts [11, 33, 26], while modularity spectral methods use the leading eigenstructure of the modularity matrix [23, 24]. These methods are powerful, but they require extracting spectral coordinates and then choosing how those coordinates are discretized

into communities. The diffusion viewpoint developed here provides a different representation rather than a general claim of lower runtime.

Random walks offer a complementary viewpoint. Vertices in the same community tend to retain comparable short- and medium-range diffusion behavior, an idea exploited by Walktrap [28], Infomap [32, 31], Markov Stability [5], personalized-PageRank methods [1], and several recent random-walk-guided community-detection procedures [8, 9, 4, 7]. A direct difficulty is that for a connected aperiodic graph the rows of P^t converge to the same stationary distribution. Consequently, sufficiently long diffusion profiles share a dominant common component and their raw angular similarity becomes progressively less discriminative.

The starting point of this paper is to remove that common stationary direction *after normalizing the diffusion profiles*. Let P denote the random-walk transition matrix and ϕ its stationary distribution. For vertex i and walk length t, we use

$$X_t(i) := \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|_2} - \frac{\phi}{\|\phi\|_2},$$

where ϕ is the stationary distribution. Community information is then read from the angular geometry of the vectors $X_t(i)$. This viewpoint separates the paper into two logically distinct layers. First, the case of two communities exposes the geometry most clearly: after stationary centering, the two groups can appear as two opposing directional clouds, and a seed-based sign rule yields a simple bipartition. Second, the same representation does not intrinsically restrict the graph to two communities. When several directional groups are present, they can be clustered directly on the unit sphere.

The algorithmic development follows this geometry from the simplest case to a unified formulation. For $K = 2$, the two-direction picture leads to a seed-based GP-Cosine rule that separates vertices by the sign of their centered cosine similarity with a reference direction. This binary construction is used only to expose and validate the geometric mechanism. We then replace this special one-seed readout by spherical K-means [6] on the same stationary-centered directions, obtaining the general method **K-GP-Cosine**. K-GP-Cosine is defined for every prescribed $K \geq 2$, including $K = 2$, and obtains all K communities simultaneously without recursive binary splitting.

A central question is whether the stationary-centering step contributes anything beyond simply applying a clustering algorithm to P^t. We therefore distinguish two classes of comparisons. The first class consists of representation-controlled ablations: standard K-means, spherical K-means, and bisecting K-means applied directly to P^t. The second class consists of established graph-native spectral methods, namely normalized spectral clustering in the form analyzed by Rohe, Chatterjee, and Yu [29] and adjacency spectral embedding followed by K-means [35]. This design allows the experiments to separate the effect of stationary-centered diffusion geometry from the effect of the clustering engine itself.

The synthetic evaluation has two main goals. First, we test multiway clustering across degree-corrected stochastic block models, hierarchical stochastic block models, and Gaussian latent-position graphs, including planted community counts $K_{\text{true}} \in \{4, 5, 6, 7, 8\}$, where K_{true} denotes the planted number of communities, and graph sizes up to $N = 4000$. Second, we isolate the sensitivity to the diffusion length t by fixing K to the planted value and varying only the walk length. The latter experiment is particularly important because diffusion methods are meaningful only if their conclusions are not an artifact of a narrowly chosen diffusion scale.

1.1 Contributions

The contributions of the paper are fourfold.

1. We introduce a normalized stationary-centered diffusion representation that removes the common limiting direction of random-walk profiles while retaining their angular deviations. This representation, rather than a particular splitting rule, is the central object of the GP-Cosine framework.

2. We provide a geometric and asymptotic explanation in the two-community setting. Under a dominant nontrivial random-walk mode, the normalized centered profiles become asymptotically collinear up to sign, explaining why cosine polarity can recover a dominant bipartition. The geometric illustration and the explicit 50-vertex example make this mechanism visible, and a small $K = 2$ experiment on the same three graph families used later confirms that the binary rule is competitive with uncentered spherical diffusion clustering and normalized spectral bisection in clearly separated regimes.

3. We turn the two-community geometry into a seed-based GP-Cosine bipartition and then use it to motivate a unified clustering readout. The one-seed sign rule is a special construction for the binary geometry; spherical K-means on the same stationary-centered directions yields **K-GP-Cosine**, a direct algorithm for any prescribed $K \geq 2$, including $K = 2$, without recursive bisection.

4. We evaluate the multiway formulation on several random-graph families, graph sizes, and ten heterogeneous real networks. The comparison includes both P^t-based ablations and established graph spectral baselines. To avoid overstating parameter selection, we explicitly distinguish oracle tuning used in the planted synthetic comparison from modularity-based model selection used in the real-network study; automatic estimation of K is not claimed.

The rest of the paper is organized as follows. Section 2 introduces the random-walk notation. Section 3 develops the stationary-centered diffusion geometry and gives the asymptotic two-community justification. Section 4 follows the same progression algorithmically: it first presents binary GP-Cosine for $K = 2$ as the simplest sign-based readout, and then derives K-GP-Cosine as the unified formulation for every prescribed $K \geq 2$. Section 5 begins with a small binary validation, reports the main multiway synthetic benchmark, then studies walk-length sensitivity and general-K/graph-size robustness before the ten-network real-data evaluation. Section 6 concludes with limitations and open problems.

2 Random-walk preliminaries

Let $G = (V, E)$ be an undirected multigraph with adjacency matrix A, where A_{ij} denotes the number of edges between vertices i and j. Let

$$m = \frac{1}{2} \sum_{i,j} A_{ij}$$

be the total number of edges. For each vertex i, its degree is

$$d_i = \sum_j A_{ij},$$

and D denotes the diagonal degree matrix.

A random walk on G is the Markov chain that, from a current vertex i, chooses one incident edge uniformly at random and traverses it. Equivalently, the transition probability from i to j is

$$P_{ij} = \frac{A_{ij}}{d_i}, \qquad \text{so that} \qquad P = D^{-1}A.$$

For $t \geq 1$, the entry P_{ij}^t denotes the probability of being at vertex j after t steps when starting from i. We denote by $P_{i\bullet}^t$ the i-th row of P^t, namely the full t-step transition profile starting from vertex i.

If G is connected and non-bipartite, then the random walk is irreducible and aperiodic, hence it converges to the unique stationary distribution

$$\phi = (\phi_1, \ldots, \phi_n), \qquad \phi_j = \frac{d_j}{\sum_k d_k} = \frac{d_j}{2m}.$$

Equivalently,

$$\lim_{t\to\infty} P^t = \mathbf{1}\phi,$$

that is, every row of P^t converges to ϕ.

Remark 2.1 *If G is bipartite, then the standard random walk with transition matrix $P = D^{-1}A$ is periodic of period 2, so P^t does not converge as $t \to \infty$. A standard remedy is to replace P by the lazy random walk*

$$\widetilde{P} = \frac{I + P}{2},$$

which preserves the same stationary distribution and makes the chain aperiodic.

More generally, even beyond the strictly bipartite case, the lazy walk is often convenient in practice because it suppresses parity effects and leads to more stable convergence behavior. For simplicity, the theoretical statements below are presented for the connected non-bipartite case.

The finite-t algorithms themselves require only an undirected graph with positive degrees. If the graph is disconnected, the degree-proportional vector $\phi_i = d_i / \sum_j d_j$ is still a stationary distribution, but it is no longer the unique limiting distribution of the full chain. Thus the asymptotic proposition below should be interpreted componentwise only under its stated connectivity assumptions; the finite-t representation and clustering rule remain well defined more broadly.

3 Stationary-centered diffusion geometry and theoretical motivation

3.1 Normalized centered diffusion profiles as vertex representations

Let $G = (V, E)$ be an undirected graph with adjacency matrix A, degree matrix D, and random-walk transition matrix

$$P = D^{-1}A.$$

For a given walk length t, the row vector $P_{i\bullet}^t$ represents the t-step transition probability distribution starting from vertex i.

A basic random-walk principle, used for example in Walktrap [28], is that vertices belonging to the same community tend to have similar t-step transition profiles:

$$P^t_{uw} \simeq P^t_{vw} \qquad \text{for all } w. \tag{3.1}$$

This suggests representing each vertex u by its diffusion profile $P^t_{u\bullet}$.

However, in a connected undirected graph, the profiles $P^t_{u\bullet}$ converge to the same stationary distribution ϕ as t increases. Thus, the raw profiles gradually lose discriminative power. To compare only directions and remove the common stationary component, we introduce the normalized centered representation

$$X_t(u) := \frac{P^t_{u\bullet}}{\|P^t_{u\bullet}\|} - \frac{\phi}{\|\phi\|}. \tag{3.2}$$

The vector $X_t(u)$ measures the deviation of the direction of the normalized diffusion profile from the stationary direction. This normalized centered representation forms the basis of our cosine-based methods: vertices are compared through the cosine similarity between the vectors $X_t(u)$.

3.2 Motivation for normalized centered cosine similarity

Random-walk profiles have long been used in community detection, including the weighted Euclidean distances of Walktrap [28]. Angle-based comparisons have also appeared in random-walk-derived representations; for example, cosine similarity can be applied to degree-weighted transition-profile coordinates, and recent random-walk-guided methods likewise exploit geometric comparisons between diffusion-derived vertex representations [8, 7].

Given this perspective, a natural first idea is to compare two vertices by the raw cosine similarity

$$\cos\big(P^t_{i\bullet}, P^t_{j\bullet}\big).$$

Indeed, if two vertices belong to the same community, then for moderate values of t their random-walk profiles tend to be similar, since the walk explores comparable regions of the graph.

However, this direct comparison becomes less informative as t increases. For a connected undirected graph, the random walk converges to its stationary distribution ϕ, so that

$$P^t_{i\bullet} \to \phi \qquad \text{for every } i.$$

Consequently, the vectors $P^t_{i\bullet}$ gradually align with the same limiting direction ϕ, and the raw cosine similarity tends to become close to 1 for many vertex pairs, regardless of whether they belong to the same community. Thus, cosine similarity applied directly to $P^t_{i\bullet}$ may lose discriminative power.

To remove this common stationary direction and compare only angular deviations, we normalize both the diffusion profiles and the stationary distribution, and then define

$$X_t(i) := \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|} - \frac{\phi}{\|\phi\|}.$$

This transformation removes the common stationary direction shared by all vertices while retaining the angular information associated with the starting position.

We then define the normalized centered cosine similarity by

$$\cos\big(X_t(i), X_t(j)\big) = \frac{\langle X_t(i), X_t(j)\rangle}{\|X_t(i)\|\,\|X_t(j)\|}.$$

Because this comparison depends only on the relative orientation of the normalized centered profiles, it suppresses common radial effects and highlights the angular separation induced by community structure. Therefore, normalized centered cosine similarity can preserve discriminative angular deviations that raw cosine similarity on $P^t_{i\bullet}$ tends to suppress. Whether this yields greater empirical robustness across walk lengths is tested separately in Section 5.5.

3.3 Geometric interpretation of normalized centered cosine similarity

To build geometric intuition, consider a graph with two reasonably well-separated communities. For each vertex i, we associate the normalized diffusion profile

$$\frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|}.$$

In Figure 1, the solid blue vectors correspond to the vertices A, B, C, D, E, while the solid orange vectors correspond to the vertices H, K, I, G. These two groups of vectors form two directional clusters, representing the two communities of the graph. In other words, the vertices A, B, C, D, E may be viewed as belonging to one community, whereas the vertices H, K, I, G belong to the other.

The red vector represents the normalized stationary direction

$$\frac{\phi}{\|\phi\|}.$$

From each normalized diffusion profile, we subtract this common direction and obtain

$$X_t(i) := \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|} - \frac{\phi}{\|\phi\|}.$$

In the figure, the dashed blue vectors are precisely the vectors

$$X_t(A), X_t(B), X_t(C), X_t(D), X_t(E),$$

while the dashed orange vectors are

$$X_t(H), X_t(K), X_t(I), X_t(G).$$

The geometric meaning of the picture is that, after removing the common direction, vectors from the same community tend to point in nearly the same direction. More precisely, the vectors

$$X_t(A), X_t(B), X_t(C), X_t(D), X_t(E)$$

are nearly collinear, and similarly the vectors

$$X_t(H), X_t(K), X_t(I), X_t(G)$$

are nearly collinear. Therefore, if two vertices belong to the same community, the angle between their corresponding vectors is very small, and the cosine similarity is close to 1.

By contrast, the dashed blue group and the dashed orange group tend to point in opposite directions. Hence, if one vector is taken from the blue group and the other from the orange group, the angle between them is close to $180°$, and the cosine similarity is close to -1.

Thus, in this geometric picture, cosine similarity between the vectors $X_t(i)$ provides a natural rule for graph bisection: pairs of vertices with cosine close to 1 tend to belong to the same community, whereas pairs with cosine close to -1 tend to belong to different communities.

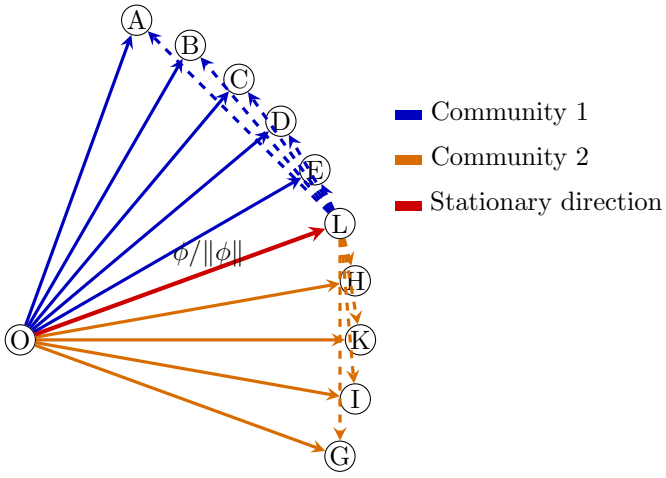


Figure 1: Geometric interpretation of the cosine-similarity rule used in *Graph Partitioning by Cosine Similarity of Normalized Centered Diffusion Profiles* (GP-Cosine). Solid blue and orange vectors denote normalized diffusion profiles from two communities, the red vector denotes the normalized stationary direction, and the dashed vectors denote the corresponding normalized centered profiles. In the figure, dashed vectors of the same color are nearly aligned, whereas the two color groups are nearly opposite.

3.4 Asymptotic justification for normalized centered cosine similarity

The discussion above suggests that normalization and centering should restore directional information that is lost in the raw profiles as t increases. We now make this intuition precise in the asymptotic regime.

Proposition 3.1 *Let G be a connected, undirected, non-bipartite graph, let*

$$P = D^{-1}A$$

be the random-walk transition matrix, and let ϕ be the unique stationary distribution. For each vertex i, define

$$X_t(i) := \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|} - \frac{\phi}{\|\phi\|}.$$

Since the walk is reversible, P is similar to the symmetric matrix

$$S := D^{1/2}PD^{-1/2} = D^{-1/2}AD^{-1/2},$$

hence all eigenvalues of P are real. Order them algebraically as

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n > -1,$$

and assume that the second random-walk eigenmode is also dominant in magnitude among all nontrivial modes,

$$\lambda_2 > \max_{\alpha \geq 3} |\lambda_\alpha|. \tag{3.3}$$

This assumption in particular forces $\lambda_2 > 0$ and rules out a near-bipartite negative mode from dominating the long-time transient. Let w_2 be a right eigenvector of P associated with λ_2. Then the following hold.

1. *For every pair of vertices i, j,*

$$\cos\big(P^t_{i\bullet}, P^t_{j\bullet}\big) \to 1 \qquad (t \to \infty).$$

2. *For every pair of vertices i, j such that $w_2(i) \neq 0$ and $w_2(j) \neq 0$,*

$$\cos\big(X_t(i), X_t(j)\big) \to \operatorname{sign}\big(w_2(i) w_2(j)\big) \qquad (t \to \infty).$$

Consequently, if the sign pattern of w_2 induces the target bipartition, then the normalized centered cosine similarity asymptotically separates vertices according to that bipartition.

Proof Because $S = D^{1/2} P D^{-1/2}$ is symmetric, it admits an orthonormal eigenbasis

$$(s_\alpha)_{\alpha=1}^n$$

such that

$$S s_\alpha = \lambda_\alpha s_\alpha.$$

Hence

$$S^t = \sum_{\alpha=1}^n \lambda_\alpha^t s_\alpha s_\alpha^T, \qquad P^t = D^{-1/2} S^t D^{1/2}.$$

Since G is connected and non-bipartite, the random walk is irreducible and aperiodic, so

$$P^t_{i\bullet} \to \phi \qquad \text{for every } i.$$

Therefore, for any vertices i, j,

$$\cos\big(P^t_{i\bullet}, P^t_{j\bullet}\big) \to \frac{\langle \phi, \phi \rangle}{\|\phi\| \, \|\phi\|} = 1.$$

This proves (1).

Next, write

$$P^t - \mathbf{1}\phi = D^{-1/2}\Big(\sum_{\alpha=2}^n \lambda_\alpha^t s_\alpha s_\alpha^T\Big) D^{1/2}.$$

By the dominance assumption in Eq. (3.3), the $\alpha = 2$ term dominates asymptotically, so

$$P^t - \mathbf{1}\phi = \lambda_2^t D^{-1/2} s_2 s_2^T D^{1/2} + o(|\lambda_2|^t).$$

Define

$$w_2 := D^{-1/2} s_2, \qquad z_2 := (D^{1/2} s_2)^T.$$

Then w_2 is a right eigenvector of P associated with λ_2, and

$$P^t_{i\bullet} = \phi + \lambda_2^t \, w_2(i) \, z_2 + o(|\lambda_2|^t).$$

Now consider the normalization map

$$N(x) := \frac{x}{\|x\|}.$$

Its first-order expansion at ϕ is

$$N(\phi + \delta) = N(\phi) + \frac{1}{\|\phi\|} \left(\delta - \frac{\langle \delta, \phi \rangle}{\|\phi\|^2} \phi \right) + o(\|\delta\|) \qquad (\delta \to 0).$$

Applying this with

$$\delta_i(t) := \lambda_2^t \, w_2(i) \, z_2 + o(|\lambda_2|^t),$$

we obtain

$$\frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|} = \frac{\phi}{\|\phi\|} + \lambda_2^t \, w_2(i) \, \eta + o(|\lambda_2|^t),$$

where

$$\eta := \frac{1}{\|\phi\|} \left(z_2 - \frac{\langle z_2, \phi \rangle}{\|\phi\|^2} \phi \right).$$

Hence

$$X_t(i) = \lambda_2^t \, w_2(i) \, \eta + o(|\lambda_2|^t). \tag{3.4}$$

Similarly,

$$X_t(j) = \lambda_2^t \, w_2(j) \, \eta + o(|\lambda_2|^t).$$

We claim that $\eta \neq 0$. Indeed, if $\eta = 0$, then z_2 would be collinear with ϕ. But z_2 is a left eigenvector associated with the nontrivial eigenvalue $\lambda_2 \neq 1$, whereas ϕ is a left eigenvector associated with the eigenvalue 1. Hence they cannot be collinear. Therefore $\eta \neq 0$.

Thus the two vectors $X_t(i)$ and $X_t(j)$ are asymptotically collinear with the same nonzero direction vector η, and therefore

$$\cos\big(X_t(i), X_t(j)\big) \to \frac{w_2(i) w_2(j)}{|w_2(i)| \, |w_2(j)|} = \operatorname{sign}\big(w_2(i) w_2(j)\big).$$

This proves (2). □

Although Proposition 3.1 gives a directional limit, Eq. (3.4) also shows that $\|X_t(i)\|_2$ decays on the order of $|\lambda_2|^t$ whenever $w_2(i) \neq 0$. Thus the proposition does not imply that arbitrarily large walk lengths are numerically preferable: for large t, the centered signal itself becomes small even though its normalized direction has a limiting sign geometry. This is why the experiments use moderate walk lengths and study sensitivity to t separately.

The following remark can be viewed as a consequence of Proposition 3.1.

Remark 3.2 (Connection with spectral bisection) *For an undirected graph, the matrices*

$$P = D^{-1}A \qquad \text{and} \qquad L_{rw} = I - P$$

share the same right eigenvectors. Under Eq. (3.3)*, w_2 is the second-largest algebraic random-walk eigenvector and therefore corresponds to the second-smallest eigenvector of L_{rw}. Its sign pattern is the standard two-way random-walk/Laplacian spectral direction.*

Under the dominance assumption in Eq. (3.3)*, the normalized centered diffusion profile satisfies*

$$X_t(i) \approx \lambda_2^t \, w_2(i) \, \eta \qquad (t \text{ large}),$$

for a common nonzero direction vector η. Thus, for a fixed seed i_0 with $w_2(i_0) \neq 0$,

$$\cos\big(X_t(i_0), X_t(j)\big) \to \operatorname{sign}\big(w_2(i_0)w_2(j)\big).$$

Hence, in this asymptotic regime, the GP-Cosine rule recovers the same sign partition as the dominant nontrivial random-walk/Laplacian mode.

We emphasize that this connection is not with Newman's modularity spectral method, which is based on the leading eigenvector of

$$B = A - \frac{dd^T}{2m}.$$

Accordingly, the method is a diffusion-based alternative to random-walk/Laplacian spectral bisection, not an approximation to modularity spectral partitioning.

4 From binary GP-Cosine to general K-GP-Cosine

The theory in Section 3 is most transparent for two communities, so we follow the same order algorithmically. We first state the seed-based GP-Cosine bipartition for $K = 2$ and illustrate its sign geometry on the 50-vertex example. We then use this binary construction as the simplest readout of the stationary-centered geometry and pass to K-GP-Cosine, a unified spherical-clustering formulation that is valid for every prescribed $K \geq 2$, including $K = 2$.

4.1 Binary GP-Cosine: the case $K = 2$

The two-community geometry of Section 3 leads directly to the simplest algorithmic form of the method. When the graph contains two dominant directional groups, one stationary-centered direction can be used as a reference: vertices whose centered profiles point to the same side as the reference receive one label, while vertices pointing to the opposite side receive the other. This gives the seed-based GP-Cosine rule below. The construction is presented first because it is the algorithmic realization of Proposition 3.1; it is not used recursively to obtain more than two communities.

Algorithm 1: Binary GP-Cosine bipartition

Input: Undirected graph $G = (V, E)$ with $d_i > 0$, walk length t

Output: A nontrivial bipartition $\{C_1, C_2\}$ when both sign classes are present

Compute $P = D^{-1}A$ and the stationary distribution $\phi_i = d_i / \sum_j d_j$;

Compute the t-step transition profiles and form, for every $i \in V$,

$$X_t(i) = \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|_2} - \frac{\phi}{\|\phi\|_2};$$

If some $X_t(i) = 0$, report a degenerate centered profile; otherwise choose a seed i_0 of maximum degree, breaking ties by the smallest index;

$C_1 \leftarrow \emptyset$, $C_2 \leftarrow \emptyset$;

for *each vertex* $j \in V$ **do**

  compute the cosine similarity

$$s_j = \cos\big(X_t(i_0), X_t(j)\big);$$

  if $s_j \geq 0$ **then**

    add j to C_1;

  else

    add j to C_2;

  end

end

if $C_1 = \emptyset$ ***or*** $C_2 = \emptyset$ **then**

  report that no nontrivial sign bipartition is produced;

end

return $\{C_1, C_2\}$;

The binary rule is intentionally simple: it converts the two-direction geometry into a sign decision relative to one reference direction. The maximum-degree seed is a deterministic practical heuristic, not a consequence of Proposition 3.1; the proposition only requires a reference vertex whose dominant-mode coefficient is nonzero. Because ordinary cosine similarity is undefined for an exact zero centered profile, Algorithm 1 explicitly reports that degenerate case; in numerical code, a small norm tolerance can be used for this check. Its role in this paper is explanatory and is supported by the small $K = 2$ validation in Section 5.3; it is not iterated recursively.

4.2 Illustrative two-community example

To illustrate the binary geometry of the GP-Cosine approach, we consider a synthetic graph of $n = 50$ vertices consisting of two dense communities weakly connected by a few inter-community edges. Figure 2 illustrates this graph, where vertices of the two communities are clearly distinguishable by connectivity. For this graph, the stationary-centered diffusion profiles

$$X_t(i) = \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|_2} - \frac{\phi}{\|\phi\|_2}$$

were computed for a small diffusion step $t = 5$. For this fixed illustration, we take vertex v_0 as the reference seed and measure the cosine similarity between its stationary-centered diffusion profile

and that of every other vertex v_j. This illustrative choice is independent of the deterministic maximum-degree seed heuristic used in Algorithm 1:

$$s_j = \cos\big(X_t(v_0),\, X_t(v_j)\big), \quad j = 0, 1, \ldots, 49.$$

Table 1 lists these cosine values in ascending vertex order. It can be observed that vertices v_0–v_{30} have strongly positive similarities (around 0.999), while vertices v_{31}–v_{49} exhibit negative similarities (around -0.92 to -0.93). This polarity indicates a natural boundary between two vertex groups.

Figure 2 shows the corresponding bipartition obtained directly from the sign of the cosine values: green nodes represent the positive-cosine group (C_1), and red nodes represent the negative-cosine group (C_2). Only a few weak inter-group edges exist; in this constructed example, the cosine-based division aligns exactly with the planted community split.

Observation. This example illustrates that cosine polarity with respect to a single seed can provide a simple signal for separating two cohesive vertex sets. For this graph, GP-Cosine recovers the planted bipartition without using an eigenvector-based partitioning step or global modularity optimization.

Table 1: Cosine similarity between vertex v_0 and all other vertices v_j ($n = 50$). Positive values correspond to cluster C_1, negative values to cluster C_2.

j	s_j	**Cluster**	j	s_j	**Cluster**	j	s_j	**Cluster**
0	1.000000	C_1	17	0.9995128	C_1	34	-0.9094001	C_2
1	0.9999580	C_1	18	0.9995151	C_1	35	-0.9342221	C_2
2	0.9993037	C_1	19	0.9994580	C_1	36	-0.9080570	C_2
3	0.9993289	C_1	20	0.9994794	C_1	37	-0.9111114	C_2
4	0.9994273	C_1	21	0.9999886	C_1	38	-0.9104676	C_2
5	0.9994614	C_1	22	0.9999354	C_1	39	-0.9126136	C_2
6	0.9998928	C_1	23	0.9999428	C_1	40	-0.9199061	C_2
7	0.9993725	C_1	24	0.9998708	C_1	41	-0.9290714	C_2
8	0.9995108	C_1	25	0.9998381	C_1	42	-0.9116720	C_2
9	0.9995592	C_1	26	0.9993225	C_1	43	-0.9301532	C_2
10	0.9995682	C_1	27	0.9994733	C_1	44	-0.9110874	C_2
11	0.9993926	C_1	28	0.9994913	C_1	45	-0.9308717	C_2
12	0.9999866	C_1	29	0.9994189	C_1	46	-0.9220995	C_2
13	0.9993939	C_1	30	0.9999561	C_1	47	-0.9089199	C_2
14	0.9993539	C_1	31	-0.9203909	C_2	48	-0.9233262	C_2
15	0.9992501	C_1	32	-0.9252084	C_2	49	-0.9233262	C_2
16	0.9994136	C_1	33	-0.9290044	C_2			

4.3 From the binary sign rule to a general K-directional formulation

Algorithm 1 relies on a feature that is special to the two-community picture: after stationary centering, the dominant structure can be represented by two nearly opposite directions. A single reference vector then suffices, because the sign of one cosine similarity distinguishes the two sides. This is a useful geometric readout, but it is not the general clustering rule of the framework.

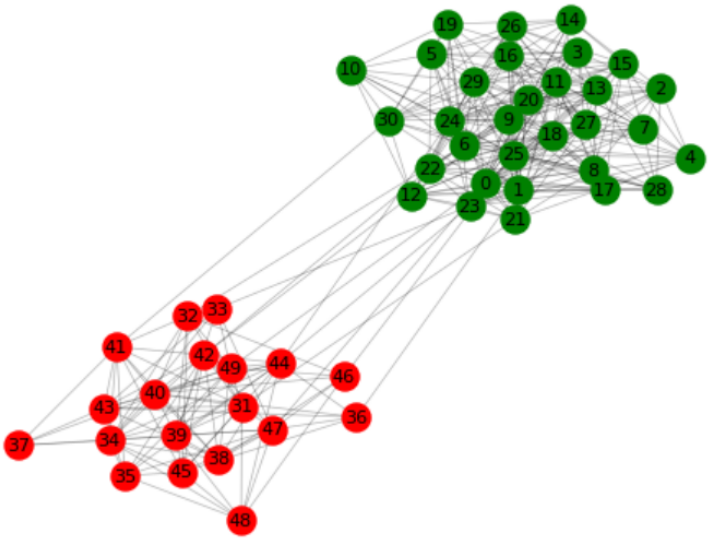


Figure 2: Example of GP-Cosine bipartition on a 50-node synthetic graph. Green and red colors correspond to the two detected communities C_1 and C_2, respectively.

To obtain one formulation that applies uniformly for any prescribed $K \geq 2$, including $K = 2$, we treat the stationary-centered profiles as directional data and cluster them directly on the unit sphere. When several directional groups are present, the collection

$$\mathcal{X}_t = \{X_t(i) : i \in V\}$$

may contain several recurrent directions. At the population level of an ideal block model, vertices in the same block can share a common block-level diffusion pattern; in degree-corrected and latent-position settings the corresponding profiles need only concentrate around community-dependent directional structure rather than be identical. Normalization followed by subtraction of the stationary direction therefore suggests a collection of directional representatives rather than a single binary axis. Spherical clustering provides a uniform readout of this geometry for both $K = 2$ and larger prescribed values of K.

For every nonzero $X_t(i)$, define its unit direction

$$Y_t(i) := \frac{X_t(i)}{\|X_t(i)\|_2}. \tag{4.1}$$

Given K, we seek unit centroids $c_1, \ldots, c_K$ and labels $z_i \in \{1, \ldots, K\}$ that maximize

$$\sum_{i=1}^{n} \langle Y_t(i), c_{z_i} \rangle, \qquad \|c_k\|_2 = 1. \tag{4.2}$$

This is the spherical K-means objective [6]. The seed-sign rule is therefore best viewed as a special binary readout of the same stationary-centered geometry, whereas K-GP-Cosine provides the uniform clustering rule for any prescribed $K \geq 2$. At $K = 2$, spherical 2-means learns two representative directions rather than fixing a single seed; for larger K, the same objective extends without any change in formulation. This motivates the general K-GP-Cosine algorithm below.

We emphasize that Proposition 3.1 is a one-mode asymptotic statement and does not constitute a recovery theorem for arbitrary K. The general K construction is motivated by the directional geometry of the representation; a recovery theory for multiple stationary-centered directions remains open.

4.4 General K-GP-Cosine

Let $K \geq 2$ be a prescribed number of communities and let t be a walk length. K-GP-Cosine keeps the same stationary-centered diffusion representation used by binary GP-Cosine, normalizes each nonzero centered profile to its unit direction as in Eq. (4.1), and clusters those directions by spherical K-means. This is the general algorithm of the framework, including when $K = 2$; the seed-based sign rule above is retained only as the simplest geometric realization of the binary case. The resulting K communities are obtained simultaneously in one clustering step, and no recursive bisection is used.

Algorithm 2: K-GP-Cosine: direct multiway GP-Cosine

Input: Undirected graph $G = (V, E)$ with $d_i > 0$, number of communities $K \geq 2$, walk length t

Output: Partition $\mathcal{P} = \{C_1, \ldots, C_K\}$

Compute $P = D^{-1}A$ and the stationary distribution $\phi_i = d_i / \sum_j d_j$;

Compute the t-step transition profiles $P^t_{i\bullet}$ for all $i \in V$;

for *each vertex* $i \in V$ **do**

form

$$X_t(i) = \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|_2} - \frac{\phi}{\|\phi\|_2};$$

if $X_t(i) = 0$, report a degenerate centered profile; otherwise set $Y_t(i) = X_t(i)/\|X_t(i)\|_2$;

end

Apply spherical K-means to the rows $Y_t(i)$, using cosine similarity and unit-norm centroids;

Return the K clusters induced by the spherical K-means labels;

The extra normalization from $X_t(i)$ to $Y_t(i)$ does not change pairwise cosine similarities for nonzero profiles; it only places the inputs explicitly on the unit sphere used by spherical K-means. Exact zero centered profiles contain no directional information. They did not occur in the reported experiments; in finite precision, a small tolerance can be used to detect a numerically degenerate profile before normalization. We use K-means++ seeding [2] for numerical initialization, but the clustering objective remains spherical K-means; the initialization strategy is not a separate community-detection method.

Remark 4.1 (Relation to the binary construction) *When $K = 2$, K-GP-Cosine and the seed-based binary GP-Cosine in Algorithm 1 use exactly the same stationary-centered representation but different partitioning rules. The binary method fixes one seed direction and separates vertices according to the sign of the cosine with that seed. K-GP-Cosine instead optimizes two spherical centroids. Therefore the binary rule should be viewed as the most transparent geometric realization of the two-direction mechanism, not as an algebraically identical implementation of spherical 2-means.*

Lemma 4.2 *([28, Claim 1]) For a sparse transition matrix $P = D^{-1}A$, a single t-step probability vector $P^t_{i\bullet}$ can be computed by t sparse matrix–vector multiplications in time $O(tm)$ and space $O(n)$.*

Theorem 4.3 (Complexity of K-GP-Cosine) *Let $n = |V|$ and $m = |E|$. Assume all n length-t profiles are computed by sparse matrix–vector propagation. If spherical K-means uses at most I_K iterations, then K-GP-Cosine has running time*

$$O\big(ntm + I_K K n^2\big)$$

and working memory

$$O(n^2 + m + Kn).$$

For fixed small K and t, the explicit profile representation therefore has a quadratic-memory bottleneck, even though the method requires no eigendecomposition.

The bound above is stated for per-source sparse propagation through P. If an implementation instead materializes successive sparse matrix powers $P^2, P^3, \ldots$, substantial fill-in can occur as t grows, especially on fast-mixing graphs; in that case the practical time and memory cost can approach dense behavior well before the formal $O(n^2)$ profile-storage limit is reached.

Proof By Lemma 4.2, one t-step diffusion profile can be propagated in $O(tm)$ time, hence all n profiles cost $O(ntm)$. Explicitly storing and normalizing the n profiles of length n costs $O(n^2)$ time and memory. During one spherical K-means iteration, assigning n vectors in $\mathbb{R}^n$ to K centroids costs $O(Kn^2)$, while centroid accumulation and normalization do not increase this order. Multiplying by at most I_K iterations gives the stated time bound. The stored profiles require $O(n^2)$ memory, the sparse graph requires $O(m)$, and the K centroids require $O(Kn)$. □

Complexity of the binary rule. Under the same explicit-profile assumption, binary GP-Cosine requires $O(ntm + n^2)$ time and $O(n^2 + m)$ working memory: the n diffusion profiles cost $O(ntm)$ to propagate, while their explicit storage and the seed-to-all cosine comparisons cost at most quadratic time and memory. As with K-GP-Cosine, this statement concerns the explicit profile implementation and does not imply a runtime advantage over optimized spectral eigensolvers.

5 Experiments

The empirical study begins with a deliberately small validation of the seed-based binary GP-Cosine rule on the same three random-graph families used later, specialized to $K_{\text{true}} = 2$. We then turn to the principal direct multiway formulation and compare K-GP-Cosine with representation-controlled and graph-native baselines on random graphs. Immediately after the main synthetic benchmark, we isolate sensitivity to the walk length t while holding K fixed; we then study general-K and graph-size robustness. Finally, we evaluate the six multiway methods on ten real networks using modularity-based selection over a prespecified parameter grid.

5.1 Methods, metrics, and tuning protocol

The small binary validation in Section 5.3 uses only three methods: the seed-based GP-Cosine rule of Algorithm 1, spherical 2-means on row-normalized P^t without stationary centering, and normalized spectral bisection based on the second nontrivial random-walk/Laplacian direction [33, 26]. All subsequent multiway experiments compare the following six methods.

1. **K-GP-Cosine**: spherical K-means applied to the normalized stationary-centered directions $Y_t(i) = X_t(i)/\|X_t(i)\|_2$.
2. **K-means on P^t**: standard Lloyd K-means [20] applied directly to the rows of P^t.
3. **Spherical K-means on P^t**: spherical K-means [6] applied directly to row-normalized transition profiles, without stationary centering.
4. **Bisecting K-means on P^t**: the bisecting K-means procedure [34] applied to the transition profiles.
5. **Rohe spectral clustering**: normalized graph spectral clustering based on $D^{-1/2}AD^{-1/2}$ as analyzed for stochastic block models by Rohe, Chatterjee, and Yu [29], followed by K-means.
6. **Adjacency spectral embedding (ASE)**: the scaled embedding $U_K\Sigma_K^{1/2}$ followed by K-means, following Sussman et al. [35].

The first three P^t baselines are deliberately close to K-GP-Cosine. In particular, spherical K-means on P^t uses the same angular clustering engine but omits the stationary-centering transformation, making it the most direct ablation of the proposed representation. Rohe spectral clustering and ASE provide graph-native embeddings followed by a prescribed-K clustering step and do not depend on a walk length t. The comparison is intentionally centered on methods that produce a K-controlled partition through a clustering stage, so direct modularity-optimization algorithms are outside the scope of this representation-focused benchmark.

For planted graphs we report Normalized Mutual Information (NMI) [36] and Adjusted Rand Index (ARI) [17]. When the number of detected clusters is also inspected, we report whether the selected K equals the planted value and the absolute error $|K - K_{\text{true}}|$. For the real-network experiment we report modularity [25], defined for a partition $\mathcal{P}$ by

$$Q(\mathcal{P}, G) = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \delta_{C(i),C(j)}. \tag{5.1}$$

For the K-means-based implementations, standard and bisecting K-means use K-means++ initialization [2] with 10 restarts; spherical K-means uses K-means++ directions with 5 restarts and at most 80 assignment/update iterations. Fixed random seeds are used for each combination of graph, method, and candidate parameters so that repeated comparisons are reproducible.

The main synthetic tables use an *oracle tuning protocol* in order to compare the best partition quality attainable by each method over a common candidate grid. For K-GP-Cosine and the three P^t baselines, each generated graph is evaluated over candidate pairs (K, t) and the pair with maximum ground-truth NMI is retained *before* averaging across repetitions. Rohe spectral clustering and ASE scan only K. The larger candidate set for diffusion-based methods reflects

their additional walk-length parameter, so this oracle study compares best attainable quality rather than equal-size hyperparameter searches. This protocol is intentionally optimistic and is used only for controlled comparison; it is not an automatic estimator of K. Consequently, a reported K-exact rate means that the NMI-optimal candidate coincides with the planted K, not that the method estimates K without ground truth.

5.2 Random-graph families

We use three graph families chosen to stress different structural features. All graphs are undirected and isolated vertices, if any, are connected to a uniformly chosen vertex in the same planted community. We use the common density scale $c = 20$ throughout the synthetic generators. In the Bernoulli block-model generators, pairwise edge probabilities are clipped at 0.95 when necessary.

Degree-corrected stochastic block model (DCSBM). Following the degree-corrected block-model principle [18], each vertex receives a multiplicative propensity

$$\theta_i = \exp(Z_i - \sigma^2/2), \qquad Z_i \sim N(0, \sigma^2),$$

renormalized to unit mean inside each block and capped at 3 in the implementation. For a mixing level μ, the base within- and between-block probabilities are set from the common density scale c so that the corresponding internal and external contributions are controlled by $(1-\mu)c$ and μc, respectively. We use balanced blocks with $\sigma = 0.8$, unbalanced blocks with $\sigma = 0.8$, and balanced blocks with stronger degree heterogeneity $\sigma = 1.2$, each at $\mu \in \{0.40, 0.50, 0.55\}$. In the unbalanced case, block sizes are proportional to linearly spaced weights from 0.55 to 1.45 before rounding to integers.

Hierarchical stochastic block model. Motivated by hierarchical and nested block structures [27], we use the following two-level synthetic construction. Each of the K coarse planted communities is divided into two latent subcommunities. Seventy percent of the expected within-community connectivity is concentrated inside the latent subcommunity and the remaining thirty percent connects the paired subcommunities; the external expected degree is controlled by the same mixing parameter $\mu \in \{0.40, 0.50, 0.55\}$. Evaluation uses the K coarse communities as ground truth.

Gaussian latent-position graph. Motivated by latent-space network models [16], we use the following synthetic construction to test a geometry that is not block-constant at the vertex level. We generate balanced communities from Gaussian latent positions. The K community centers are equally spaced on a circle in $\mathbb{R}^2$ with adjacent-center separation $\Delta \in \{2.0, 2.5, 3.0\}$. Given community g_i,

$$z_i \sim N(c_{g_i}, I_2).$$

Pairs receive radial-basis affinity

$$w_{ij} = \exp\left(-\frac{\|z_i - z_j\|^2}{2h^2}\right), \qquad h = 1.5,$$

and edge probability

$$p_{ij} = 1 - \exp(-\lambda w_{ij}),$$

where λ is calibrated numerically so that the expected number of edges matches the common density scale, i.e. approximately $nc/2$ edges. Larger Δ corresponds to easier separation.

5.3 Small binary validation of GP-Cosine

Before the main multiway benchmark, we perform a deliberately small experiment to validate the seed-based binary rule in Algorithm 1. The purpose is not to establish a separate state-of-the-art binary clustering method, but to check whether the two-direction geometry developed in Section 3 translates into competitive partitions on the same three graph families used in the multiway study.

We specialize the generators to $K_{\text{true}} = 2$ and $N = 500$, using five independently generated graphs for each family. Because this experiment is intended to illustrate a clearly separated binary regime rather than behavior near a detectability threshold, we use DCSBM with $\sigma = 0.8$ and $\mu = 0.20$, hierarchical SBM with $\mu = 0.20$ and the same 70%/30% internal hierarchy split as above, and a Gaussian latent-position graph with center separation $\Delta = 4.5$, latent covariance I_2, and bandwidth $h = 1.5$. The same density scale $c = 20$ is used. These are the same three generator families as in the main study, but the binary parameters are intentionally easier than the multiway difficulty grid.

We compare GP-Cosine with two natural binary references. The first is spherical 2-means applied to the row-normalized but uncentered transition profiles P^t, which directly isolates the effect of stationary centering. The second is normalized spectral bisection using the second non-trivial direction of $D^{-1/2}AD^{-1/2}$. For GP-Cosine and spherical 2-means we evaluate $t \in \{3, 4, 5\}$. On each generated graph and for each diffusion-based method, the walk length with the largest ground-truth NMI is retained before averaging across the five repetitions; the corresponding ARI is reported. Thus, as in the main planted synthetic benchmark, this is an oracle comparison over the prespecified t grid rather than a data-driven selection rule.

Table 2: Binary validation on the three synthetic graph families used in the main experiments, specialized to $K_{\text{true}} = 2$. Entries are mean $\pm$ standard deviation over five independently generated graphs. For GP-Cosine and spherical 2-means on P^t, $t \in \{3, 4, 5\}$ is selected graphwise by maximum NMI before averaging.

Family	Method	NMI	ARI
DCSBM	GP-Cosine	0.8322 ± 0.0362	0.8971 ± 0.0272
	Spherical 2-means on P^t	0.8295 ± 0.0363	0.8956 ± 0.0267
	Normalized spectral	0.8273 ± 0.0375	0.8941 ± 0.0267
Hierarchical SBM	GP-Cosine	0.9643 ± 0.0192	0.9841 ± 0.0097
	Spherical 2-means on P^t	0.9753 ± 0.0251	0.9888 ± 0.0121
	Normalized spectral	0.9685 ± 0.0249	0.9857 ± 0.0118
Gaussian latent	GP-Cosine	0.9199 ± 0.0314	0.9588 ± 0.0179
	Spherical 2-means on P^t	0.9170 ± 0.0331	0.9573 ± 0.0204
	Normalized spectral	0.9144 ± 0.0346	0.9557 ± 0.0212

The three methods are close on all three graph families. GP-Cosine has the largest mean NMI on the DCSBM and Gaussian latent-position regimes and is within 0.011 NMI of the best method on the hierarchical SBM. The corresponding ARI values show the same pattern. We

therefore use this experiment only as empirical support for the binary geometric mechanism: in clearly separated two-community regimes, the seed-based sign rule can match the accuracy of a normalized spectral bisection and of the closest uncentered diffusion ablation. No runtime advantage over spectral clustering is claimed.

5.4 Main multiway benchmark at $N = 500$

We first use $N = 500$ and $K_{\text{true}} = 5$ to compare the six methods across all 15 graph regimes: nine DCSBM settings, three hierarchical settings, and three Gaussian latent-position settings. Each regime contains 10 independently generated graphs. The P^t methods use candidate $K = 2, \ldots, 8$ and $t \in \{3, 4, 5, 6, 8, 10\}$; the two graph spectral methods use the same K grid and no t parameter.

Table 3 averages the regime-level means within each graph family. K-GP-Cosine achieves the largest mean NMI in 12 of the 15 individual regimes. The three regimes in which another method has the largest mean are all very close: the deficit of K-GP-Cosine from the best method is below 0.002 NMI in each case. The strongest competitor is spherical K-means on P^t, which is also the most informative ablation because it differs from K-GP-Cosine primarily by the stationary-centering step.

Table 3: Mean oracle-best NMI and ARI at $N = 500$, averaged over the regimes of each random-graph family. Each regime contains 10 independent graphs.

	DCSBM		Hierarchical SBM		Gaussian latent	
Method	NMI	ARI	NMI	ARI	NMI	ARI
K-GP-Cosine	**0.5927**	**0.6227**	**0.7996**	**0.8231**	**0.5397**	**0.4918**
Spherical K-means on P^t	0.5507	0.5742	0.7953	0.8172	0.5289	0.4801
K-means on P^t	0.4671	0.4692	0.7976	0.8195	0.5285	0.4815
Bisecting K-means on P^t	0.4158	0.4181	0.7262	0.7339	0.5164	0.4520
Rohe spectral	0.4222	0.4194	0.7701	0.7886	0.4990	0.4513
ASE	0.3815	0.0986	0.7553	0.7660	0.4809	0.4305

The family-level results also clarify where stationary centering is most useful. On the hierarchical SBM, several diffusion-based methods are nearly tied. In contrast, the gap between K-GP-Cosine and uncentered spherical clustering becomes substantial in the more difficult degree-corrected regimes, particularly for the stronger-heterogeneity settings and larger μ. This pattern is consistent with the intended role of stationary centering: removing the stationary direction can preserve community-dependent angular structure that is less visible in the raw diffusion profiles.

5.5 Sensitivity to the walk length t

To isolate the effect of diffusion length from the effect of choosing the number of clusters, we conduct a separate experiment with the general K-GP-Cosine algorithm at a fixed planted value $K = K_{\text{true}} = 5$ for every method. We vary

$$t \in \{3, 4, 5, 6, 8, 10\}$$

on six representative regimes: three DCSBM settings at $\mu = 0.50$ (balanced $\sigma = 0.8$, unbalanced $\sigma = 0.8$, and balanced $\sigma = 1.2$), hierarchical SBM at $\mu = 0.50$, and Gaussian latent-position graphs with $\Delta = 2.0$ and 2.5. Each regime contains 10 independently generated graphs. Because

Rohe spectral clustering and ASE have no walk-length parameter, this experiment focuses on the four directly comparable diffusion methods: K-GP-Cosine, K-means on P^t, spherical K-means on P^t, and bisecting K-means on P^t. Thus only t varies; neither the graph generator nor the number of clusters is retuned across walk lengths.

Absolute variation alone can be misleading: a method with uniformly poor NMI can appear stable simply because all of its scores are near zero. We therefore compute stability on each graph first and then average over repetitions. In addition to the mean NMI over the tested t values, we use the relative range

$$R_{\text{rel}} = \frac{\max_t \text{NMI}(t) - \min_t \text{NMI}(t)}{\max_t \text{NMI}(t)} \tag{5.2}$$

and the coefficient of variation

$$CV_t = \frac{\text{SD}_t(\text{NMI})}{\text{mean}_t(\text{NMI})}. \tag{5.3}$$

Smaller values indicate less relative sensitivity. We also report $R90$, the fraction of tested t values whose NMI is at least 90% of the graphwise maximum.

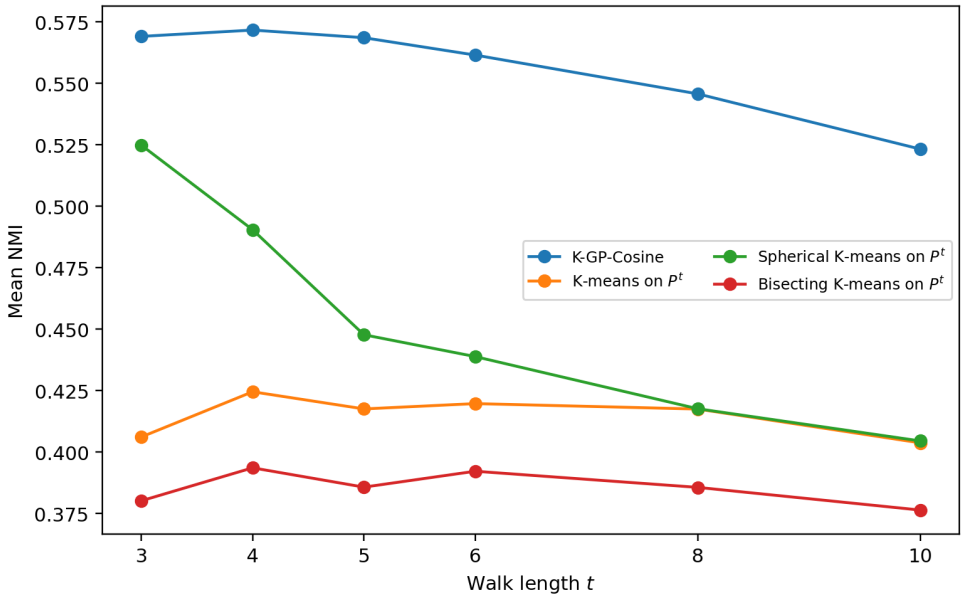


Figure 3: Mean NMI as a function of walk length t, averaged over the six sensitivity regimes. Here $K = K_{\text{true}} = 5$ is fixed for every method, so the curves isolate sensitivity to diffusion length.

Table 4: Walk-length sensitivity, computed graphwise and then averaged over the six regimes and 10 repetitions. Lower R_{rel} and CV_t indicate greater relative stability.

Method	Mean NMI	R_{rel}	CV_t	$R90$
K-GP-Cosine	**0.5566**	**0.1112**	**0.0461**	**0.8972**
K-means on P^t	0.4149	0.1961	0.1312	0.7889
Spherical K-means on P^t	0.4540	0.3068	0.2774	0.6194
Bisecting K-means on P^t	0.3856	0.1943	0.1031	0.7167

The general K-GP-Cosine algorithm therefore exhibits the best overall performance–stability profile in this study: it has both the highest mean NMI and the smallest average relative range and coefficient of variation among the four directly comparable diffusion methods. The direct comparison with uncentered spherical K-means is particularly informative. Since both methods use cosine geometry and the same transition profiles, the substantially smaller sensitivity of K-GP-Cosine supports the interpretation that subtracting the stationary direction stabilizes the useful angular structure across diffusion scales. This is empirical evidence rather than a general theorem, and individual regimes can contain near-ties or low-accuracy baselines whose small absolute fluctuations should not be over-interpreted.

5.6 General K and graph-size robustness

The next study varies both the planted number of communities and graph size. We use

$$K_{\text{true}} \in \{4, 5, 6, 7, 8\}, \qquad N \in \{1000, 2000, 4000\}.$$

For every pair of graph size and regime, 10 graphs are generated with planted counts cycling through $4, 5, 6, 7, 8$ twice, so each value of K_{true} occurs equally often. Candidate values are restricted to $K \in \{4, 5, 6, 7, 8\}$, and the P^t methods use the reduced walk-length grid $t \in \{3, 4, 5\}$. The full 15-regime difficulty grid from Section 5.4 is retained. Thus this study contains $3 \times 15 \times 10 = 450$ independently generated graphs.

Figure 4 summarizes the mean oracle-best NMI over the 15 regimes. K-GP-Cosine remains the highest-mean method at every tested size. Moreover, it is the regime winner in 11 of 15 regimes at $N = 1000$, 11 of 15 at $N = 2000$, and 11 of 15 at $N = 4000$. The mean NMI changes from 0.6357 to 0.6204 and 0.6094 as the graph grows, while spherical K-means on P^t changes from 0.6131 to 0.6025 and 0.5957.

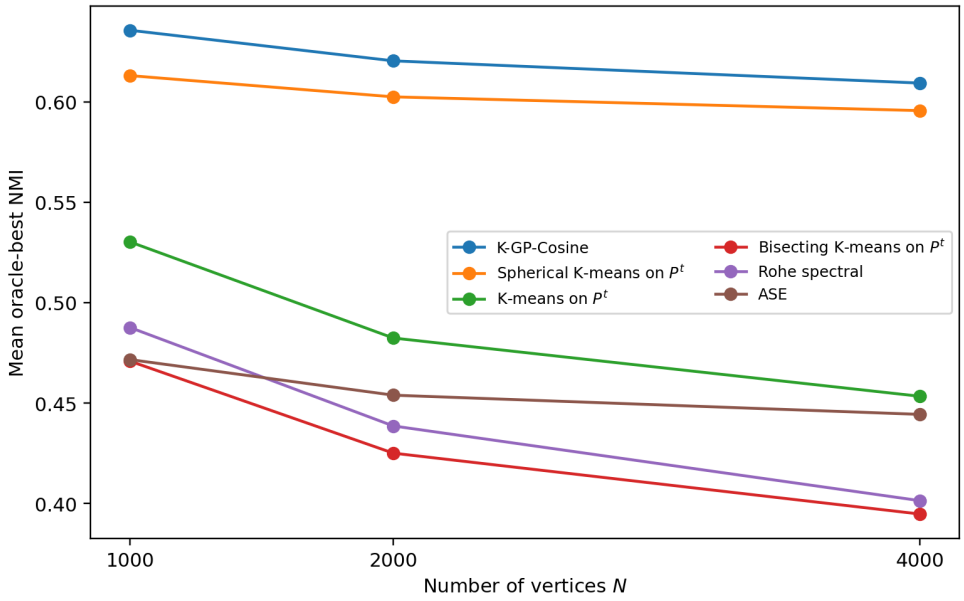


Figure 4: Mean oracle-best NMI versus graph size, averaged over the 15 random-graph regimes. K-GP-Cosine remains the highest-mean method for $N = 1000$, 2000, and 4000.

Table 5: Mean oracle-best NMI and ARI averaged over all 15 regimes at each graph size.

	$N = 1000$		$N = 2000$		$N = 4000$	
Method	NMI	ARI	NMI	ARI	NMI	ARI
K-GP-Cosine	**0.6357**	**0.6360**	**0.6204**	**0.6235**	**0.6094**	0.6021
Spherical K-means on P^t	0.6131	0.6161	0.6025	0.6124	0.5957	**0.6070**
K-means on P^t	0.5303	0.5084	0.4824	0.4512	0.4534	0.4221
Bisecting K-means on P^t	0.4709	0.4470	0.4251	0.3979	0.3947	0.3650
Rohe spectral	0.4876	0.4671	0.4386	0.4123	0.4014	0.3783
ASE	0.4717	0.2909	0.4539	0.2842	0.4444	0.2823

The ARI summary is broadly consistent with the NMI comparison but is not identical. K-GP-Cosine has the largest mean ARI at $N = 1000$ and $N = 2000$, whereas at $N = 4000$ spherical K-means on P^t is marginally higher (0.6070 versus 0.6021). We therefore interpret the size experiment as evidence of a strong and persistent overall performance profile, rather than as uniform dominance under every metric.

Figure 5 reorganizes the same 450-graph experiment by planted community count. K-GP-Cosine has the largest mean NMI at every $K_{\text{true}} \in \{4, 5, 6, 7, 8\}$, with corresponding means 0.4582, 0.5835, 0.6521, 0.6924, and 0.7230. Together with the $K = 2$ formulation in Section 4.4, this confirms empirically that the same K-GP-Cosine algorithm applies across a range of prescribed community counts rather than only to the two-direction illustration used in the theoretical motivation.

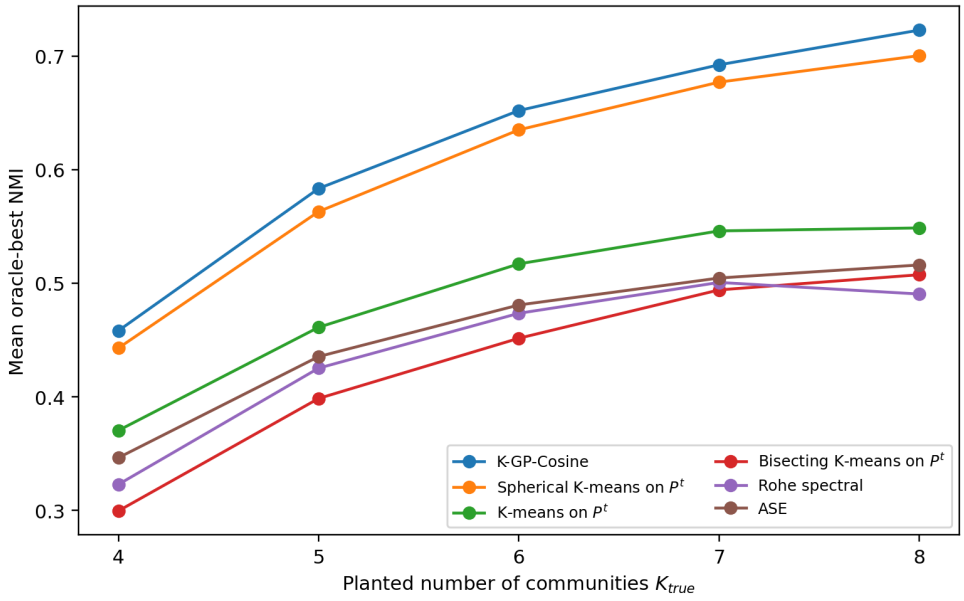


Figure 5: Mean oracle-best NMI versus the planted number of communities, aggregated over graph sizes, regimes, and repetitions. The experiment uses planted $K_{\text{true}} \in \{4, 5, 6, 7, 8\}$.

As a diagnostic, the NMI-optimal K-GP-Cosine candidate coincides with K_{true} in 90.0%, 82.2%, 80.0%, 80.0%, and 96.7% of the cases for planted $K = 4, 5, 6, 7, 8$, respectively. These percentages should not be interpreted as an automatic K-estimation result because the candidate is selected with access to ground-truth NMI. They simply indicate that the best K-GP-Cosine partition over the common candidate grid is frequently attained at the planted number of groups.

5.7 Real-world experiments

We next evaluate K-GP-Cosine on ten real networks drawn from social, communication, informational, biological, and ecological domains. The source-network sizes and descriptions are summarized in Table 6. We use the full source networks rather than sampled or induced subgraphs. Before running any method, every dataset is converted to a simple undirected unweighted graph: edge weights are ignored, reciprocal directed arcs and parallel edges are merged, and self-loops are removed, while all source vertices are retained. Consequently, for directed or multigraph sources the raw edge count in Table 6 can exceed the number of edges in the simple benchmark graph used by the algorithms. The table reports raw counts only to document source provenance; all algorithmic calculations use the post-processed simple graph.

Network	Source $\lvert V\rvert$	Raw $\lvert E\rvert$	Description
College football [13]	115	613	Games among U.S. college football teams during the 2000 season.
Dolphins [21]	62	159	Social association network of bottlenose dolphins in Doubtful Sound, New Zealand.
Asoiaf [3]	796	2,823	Character co-occurrence network from the first five *A Song of Ice and Fire* novels; characters are linked when their names occur within 15 words.
Jazz [14]	198	2,742	Collaboration network of jazz musicians represented by joint performances.
Arenas-Email [15]	1,133	5,451	Institutional email communication network; vertices represent users and edges represent observed communication.
Political books [19]	105	441	Books on U.S. politics around the 2004 presidential election; edges represent frequent co-purchasing.
Moreno-names [30]	1,773	9,131	Weighted co-occurrence network of noun phrases (places and names) in verses of the King James Bible.
C. elegans metabolic [10]	453	4,596	Metabolic interaction network of *C. elegans*; the source network is distributed with the Arenas-group benchmark data.
C. elegans neural [38, 37]	297	2,359	Directed weighted neuronal connectivity network compiled from the *C. elegans* connectome.
Little Rock food web [22]	183	2,494	Directed food web of taxa in Little Rock Lake, Wisconsin; edges represent feeding relations.

Table 6: Ten real networks used in the multiway evaluation. The reported edge counts are raw source counts before the common simple-undirected-unweighted preprocessing described in the text; all methods are run on the corresponding post-processed simple graphs.

The six methods are exactly the same as in Section 5.1. For K-GP-Cosine, standard K-means on P^t, spherical K-means on P^t, and bisecting K-means on P^t, we scan

$$t \in \{3, 4, 5\}.$$

The candidate community-count grid depends only on graph size and is shared by all methods on the same dataset. For $n < 500$ we use

$$K \in \{2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20\},$$

and for $500 \leq n < 2000$ we use

$$K \in \{4, 5, 6, 8, 10, 12, 15, 20, 30, 40\}.$$

For each P^t-based method we retain the candidate (K, t) with maximum modularity Q; Rohe spectral clustering and ASE do not use t and therefore retain the best candidate over the same K grid. This is a ground-truth-free model-selection protocol, but because modularity is used both for parameter selection and reporting, the values in Table 7 should be interpreted as the best modularity attained over the prespecified search grid, not as evidence of automatic recovery of a uniquely correct K.

Network	K-GP-Cosine	K-means P^t	Spherical P^t	Bisecting P^t	Rohe	ASE
College football	0.603566	**0.604429**	0.603566	0.592831	0.600517	0.594541
Dolphins	**0.523496**	0.516020	**0.523496**	0.483308	0.500514	0.387781
Asoiaf	**0.581724**	0.560486	0.572175	0.563876	0.463726	0.269312
Jazz	**0.442944**	0.442395	0.442507	0.442395	0.441397	0.379182
Arenas-Email	**0.559550**	0.537525	0.557398	0.534843	0.473846	0.321826
Political books	0.526123	**0.526229**	0.526008	0.513027	0.522681	0.482636
Moreno-names	0.487719	0.460333	**0.495730**	0.472609	0.371025	0.234565
C. elegans metabolic	**0.425646**	0.386192	0.399112	0.381187	0.348465	0.115716
C. elegans neural	**0.382016**	0.370181	0.376608	0.368419	0.357126	0.301456
Little Rock food web	**0.327088**	0.125724	0.125724	0.134642	0.313896	0.266894

Table 7: Best modularity over the prespecified parameter grid on ten real networks. For the four P^t-based methods, each entry is $\max_{K,t} Q$ with $t \in \{3, 4, 5\}$; Rohe spectral clustering and ASE maximize only over K. Bold values indicate the highest modularity in each row, with ties bolded jointly.

On the ten networks reported in Table 7, K-GP-Cosine attains the highest modularity on six networks and ties for the highest value on Dolphins. Relative to the closest representation-controlled ablation, spherical K-means on raw P^t, K-GP-Cosine is higher on seven networks, tied on two, and lower on one. Its mean rank across the six compared methods is 1.40, compared with 2.35 for spherical K-means on P^t, 3.10 for standard K-means on P^t, 4.15 for bisecting K-means, 4.40 for Rohe spectral clustering, and 5.60 for ASE. We interpret these modularity values only as comparative performance over the stated candidate grids, not as evidence that one representation uniformly dominates all network structures.

The real-network results also show why the stationary-centering step is not merely cosmetic. On the Little Rock food web, K-GP-Cosine reaches $Q = 0.327088$, whereas both standard and spherical K-means on raw P^t reach only 0.125724; the closest competitor there is Rohe spectral clustering at 0.313896. At the same time, the near-ties on College football, Dolphins, Jazz, and Political books indicate that stationary centering does not uniformly dominate the raw diffusion geometry. The appropriate conclusion is that K-GP-Cosine provides a strong and consistently competitive direct multiway partition across heterogeneous networks, with the clearest gains appearing on a subset of structures rather than on every dataset.

6 Conclusion and limitations

This paper develops community detection around a stationary-centered diffusion geometry. The central representation

$$X_t(i) = \frac{P^t_{i\bullet}}{\|P^t_{i\bullet}\|_2} - \frac{\phi}{\|\phi\|_2}$$

removes the common stationary direction of random-walk profiles and retains their angular deviations. The two-community case provides the clearest geometric explanation: under a dominant nontrivial random-walk mode, the centered profiles become asymptotically aligned up to sign. We use that result as motivation rather than as a claim that the full multiway problem reduces to a single spectral mode.

The algorithmic construction follows the same progression. In the binary case, GP-Cosine turns the two-direction geometry into a one-seed sign rule. K-GP-Cosine then provides the unified clustering rule: spherical K-means is applied directly to the same stationary-centered

directions to obtain all K communities in one step for any prescribed $K \geq 2$, including $K = 2$. Thus the binary rule explains the mechanism, while K-GP-Cosine is the general algorithm of the framework; no recursive bisection is required.

The small binary validation complements the geometric discussion: on clearly separated $K_{\text{true}} = 2$ instances from the same three graph families, GP-Cosine attains mean NMI and ARI comparable to both normalized spectral bisection and spherical clustering of the uncentered diffusion profiles. This experiment is deliberately illustrative and does not support a claim of runtime superiority over spectral clustering. The main multiway synthetic experiments support three further conclusions. First, stationary centering improves the strongest direct ablation, spherical K-means on P^t, most clearly in difficult degree-corrected regimes. Second, K-GP-Cosine remains competitive across planted $K \in \{4, 5, 6, 7, 8\}$ and graph sizes up to $N = 4000$, with the highest mean NMI among the compared methods at each tested size. Third, when K is fixed and only t varies, K-GP-Cosine has the best overall combination of clustering accuracy and relative stability across the tested diffusion lengths.

The real-network study reinforces this picture. On the ten networks reported here, K-GP-Cosine has the highest modularity on six networks and ties for the highest on one under the common parameter-search protocol. Its mean rank is 1.40, the best among the six compared methods. The strongest direct competitor is spherical K-means on raw P^t, which reinforces that the main empirical question in this benchmark is the effect of stationary centering rather than the choice of a fundamentally different clustering engine.

Several limitations are important. First, the present multiway theory is incomplete: Proposition 3.1 requires a connected non-bipartite graph and a dominant positive second random-walk eigenmode, and it explains only a binary asymptotic sign geometry; a finite-sample recovery theory for multiple stationary-centered directions remains open. The finite-t algorithm can be run on disconnected graphs with positive degrees, but the global asymptotic interpretation is then not covered by the proposition. Second, the synthetic comparison uses oracle selection over candidate K and, for diffusion methods, t, whereas the real-network experiment selects parameters by maximizing modularity; neither protocol establishes an intrinsic automatic estimator of K. Third, explicitly storing all diffusion profiles requires $O(n^2)$ memory, and materialized powers P^t can exhibit substantial fill-in on fast-mixing graphs. In addition, the present real-network benchmark discards edge directions and weights when they are present because the current formulation is for undirected unweighted graphs; extending the stationary-centered construction directly to directed and weighted networks is left for future work. Sparse, truncated, compressed, or on-the-fly approximations of the profile representation are natural next steps.

These limitations also indicate the most useful directions for future work: theoretical recovery guarantees beyond the dominant binary mode, principled joint selection of K and t without planted labels, broader stress tests on heterogeneous graph families, and scalable approximations that preserve the stationary-centered angular geometry on much larger networks.

References

[1] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 475–486, 2006.

[2] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[3] Andrew Beveridge. Character interaction networks for george r. r. martin's a song of ice and fire saga. GitHub data repository, `https://github.com/mathbeveridge/asoiaf`. The combined first-five-books edge list is distributed as `asoiaf-all-edges.csv`; accessed September 2026.

[4] Tien Dat Dang, Duy Hieu Do, and Thi Ha Duong Phan. Community detection in directed graphs using stationary distribution and hitting times methods. *Social Network Analysis and Mining*, 13, 2023. Article 80.

[5] Jean-Charles Delvenne, Sophia N. Yaliraki, and Mauricio Barahona. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, 107(29):12755–12760, 2010.

[6] Inderjit S. Dhillon and Dharmendra S. Modha. Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42:143–175, 2001.

[7] Duy Hieu Do and Thi Ha Duong Phan. Detecting communities in large networks using the extended walktrap algorithm. In *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 100–105. IEEE, 2022.

[8] Duy Hieu Do and Thi Ha Duong Phan. An improvement on the louvain algorithm using random walks. *Journal of Combinatorial Optimization*, 50(2), 2025. Article 14.

[9] Duy Hieu Do and Thi Ha Duong Phan. Overlapping community detection algorithms using modularity and the cosine. *Advances in Complex Systems*, 28(03):2550006, 2025.

[10] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical Review E*, 72(2):027104, 2005.

[11] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[12] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, 2010.

[13] Michelle Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[14] Pablo M. Gleiser and Leon Danon. Community structure in jazz. *Advances in Complex Systems*, 6(4):565–573, 2003.

[15] Roger Guimerà, Leon Danon, Albert Díaz-Guilera, Francesc Giralt, and Alex Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):065103, 2003.

[16] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[17] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

[18] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

[19] Valdis Krebs. The political books network. Unpublished network dataset distributed through network-data archives, `https://networks.skewed.de/net/polbooks`. Compiled by Valdis Krebs; accessed September 2026.

[20] Stuart P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[21] David Lusseau, Karsten Schneider, Oliver Boisseau, Patti Haase, Elisabeth Slooten, and Stephen M. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

[22] Neo D. Martinez. Artifacts or attributes? effects of resolution on the little rock lake food web. *Ecological Monographs*, 61(4):367–392, 1991.

[23] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, 2006.

[24] M. E. J. Newman. Spectral methods for community detection and graph partitioning. *Physical Review E*, 88(4):042822, 2013.

[25] M. E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

[26] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems (NIPS)*, volume 14, pages 849–856, 2001.

[27] Tiago P. Peixoto. Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1):011047, 2014.

[28] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191–218, 2006.

[29] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

[30] Christoph Römhild and Chris Harrison. Bible cross-references. Dataset distributed as the `bible_nouns/moreno_names` network, `https://networks.skewed.de/net/bible_nouns`. Original data/visualization credited to the authors by the dataset catalogue; accessed September 2026.

[31] Martin Rosvall, Daniel Axelsson, and Carl T. Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, 2010.

[32] Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

[33] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[34] Michael Steinbach, George Karypis, and Vipin Kumar. A comparison of document clustering techniques. In *KDD Workshop on Text Mining*, 2000.

[35] Daniel L. Sussman, Minh Tang, Donniell E. Fishkind, and Carey E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

[36] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(Oct):2837–2854, 2010.

[37] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[38] John G. White, Eileen Southgate, J. N. Thomson, and Sydney Brenner. The structure of the nervous system of the nematode caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 314(1165):1–340, 1986.