INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS Vol. 1, No. 1 (2009), pp. 1–18
ISSN (online) • www.ijcmi.webs.com *Research Article*

Graph Machine Learning on Planar Graphs

Spectral Intuition, Separator Structure, and Algorithms for Learning on Plane-Embedded Data

Satyanarayana Sanakkayala
Research Scholar, Graph Machine Learning
School of Science and Technology, Dravidian University, Kuppam, India

Received: 01 January 2009 • Accepted: 01 May 2009 • Published: 20 May 2009

Abstract. Planar graphs form a structurally rich yet computationally tractable class of graphs that arise naturally in image analysis, geographic information systems, circuit layout, and molecular chemistry. This paper develops graph machine learning tailored to planar graphs, with an emphasis on the mathematical intuition that connects the topology of a plane embedding to the spectral and combinatorial structure exploited by learning algorithms. We first recall that planarity forces sparsity through Euler's formula and small vertex separators through the Lipton and Tarjan theorem, and we explain why these two facts together make planar learning problems well conditioned. We then treat three learning primitives in a unified way: spectral partitioning through the Fiedler vector of the graph Laplacian, semi-supervised classification through harmonic extension of labels, and similarity through diffusion kernels. Throughout we develop the electrical network interpretation, in which the harmonic solution is a potential and effective resistance is a learned distance, because this picture is especially transparent on planar graphs. Four algorithms are presented with complexity analysis, and their behaviour is illustrated on plane-embedded meshes and grids. The paper is intended as a mathematically motivated entry point for researchers who wish to learn on data whose relational structure can be drawn in the plane without crossings.

Keywords: planar graphs; graph Laplacian; spectral clustering; semi-supervised learning; harmonic functions; effective resistance; diffusion kernels; planar separators.

2000 MSC: 05C10; 05C50; 68T05; 68R10.

1. Introduction

A large body of data carries relational structure that can be drawn on a flat surface without edges crossing. The pixels of an image and their four or eight neighbours, the intersections of a road network, the atoms and bonds of many small molecules, and the cells of a finite element mesh all give rise to graphs that are planar or very nearly planar. Learning on such data means learning on planar graphs. The central message of this paper is that planarity is not an incidental property to be ignored, but a strong structural prior that a learning algorithm can and should exploit.

Two consequences of planarity drive everything that follows. The first is sparsity. Euler's formula for connected plane graphs forces the number of edges to grow only linearly in the number of vertices, so the adjacency and Laplacian operators of a planar graph are sparse and the linear algebra of learning is cheap. The second is the existence of small balanced separators. The theorem of Lipton and Tarjan guarantees that any n vertex planar graph can be split into balanced pieces by removing only $O(\sqrt{n})$ vertices [4]. Small separators are exactly what a divide and conquer learner needs, and they also imply, through the analysis of Spielman and Teng, that spectral partitioning provably finds cuts of low conductance on planar graphs [24].

Our treatment is organised around mathematical intuition rather than around a catalogue of methods. In Section 2 we fix notation and recall the planar facts we need. Section 3 builds the Laplacian and its spectral geometry, and shows why the second eigenvector, the Fiedler vector, cuts a planar graph along its natural bottleneck. Section 4 turns the separator theorem into a recursive learning template. Section 5 treats semi-supervised learning as the solution of a discrete Laplace equation and develops the electrical network reading of the answer. Section 6 constructs diffusion and random walk kernels that let a learner compare planar structures. Section 7 collects the algorithms and their costs, Section 8 surveys applications, and Section 9 gives illustrative computations. Section 10 concludes.

Foundational material on graphs is standard [10], and the reader who wants the spectral background in depth is referred to the monograph of Chung [7]. The learning ideas we specialise to the plane are drawn from the graph based semi-supervised learning literature [14], from spectral clustering [12], and from kernels on graphs [13].

2. Mathematical Preliminaries

2.1. Graphs, embeddings, and planarity

Let $G=(V,E)$ be a finite simple undirected graph with vertex set $V=\{1,\dots,n\}$ and edge set E, and let $w_{ij}>0$ denote an optional weight on edge $\{i,j\}$. The degree of vertex i is $d_i=\sum_j w_{ij}$. A *plane embedding* of G is a drawing in $\mathbb{R}^2$ in which vertices are distinct points and edges are curves meeting only at shared endpoints. A graph is *planar* if it admits such an embedding, and a *plane graph* is a planar graph together with a chosen embedding. The classical characterisation of Kuratowski states that a graph is planar if and only if it contains no subdivision of K_5 or $K_{3,3}$ [1]. Planarity can be tested in linear time [3], so it is cheap to know whether the structural prior we rely on actually holds.

A plane embedding partitions the plane into *faces*, one of which is unbounded. The bookkeeping that ties vertices, edges, and faces together is Euler's formula.

Theorem 1 (Euler's formula)**.** *For a connected plane graph with n vertices, m edges, and f faces,*

$$n-m+f=2. \tag{1}$$

The formula has an immediate and, for our purposes, decisive corollary.

Corollary 1 (Planar sparsity)**.** *A simple planar graph with $n\geq 3$ vertices has at most $3n-6$ edges. Consequently the average degree is below 6 and the Laplacian has $O(n)$ nonzero entries.*

Proof. Every face of a simple plane graph is bounded by at least three edges, and every edge borders at most two faces, so $3f\leq 2m$. Substituting $f=2-n+m$ from Theorem 1 gives $3(2-n+m)\leq 2m$, that is $m\leq 3n-6$. □

Corollary 1 is the reason planar learning is inexpensive: every operator we build has a linear number of nonzeros, and every sparse matrix vector product costs $O(n)$.

2.2. The graph Laplacian

Collect the weights into the symmetric adjacency matrix $W\in\mathbb{R}^{n\times n}$ with entries $W_{ij}=w_{ij}$, let $D=\operatorname{diag}(d_1,\dots,d_n)$, and define the *combinatorial Laplacian*

$$L=D-W. \tag{2}$$

Its quadratic form is the object that carries all the geometric meaning:

$$\mathbf{f}^\top L\mathbf{f} \;=\; \tfrac{1}{2}\sum_{i,j} w_{ij}\,(f_i-f_j)^2 \;\geq\; 0, \tag{3}$$

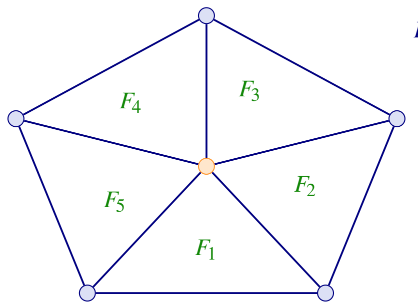

Figure 1: A plane embedding with six vertices and no crossings. The interior vertex (saffron) joins all five boundary vertices. Euler's formula is verified, and the edge count sits below the planar bound of Corollary 1.

so L is symmetric positive semidefinite. Equation (3) reads as a discrete Dirichlet energy: it is small exactly when $\mathbf{f}$ varies slowly across edges, and it is the single most important formula in this paper. The *symmetric normalised Laplacian* is $\mathscr{L} = D^{-1/2}LD^{-1/2}$ and the *random walk Laplacian* is $L_{\mathrm{rw}} = D^{-1}L$; the two normalisations are similar and share eigenvalues. Order the eigenvalues of L as $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ with orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. The all ones vector $\mathbf{1}$ is the eigenvector for $\lambda_1 = 0$, and the multiplicity of the eigenvalue zero equals the number of connected components.

Definition 1 (Algebraic connectivity and the Fiedler vector)**.** The second smallest eigenvalue λ_2 is the *algebraic connectivity* of G, and its eigenvector $\mathbf{u}_2$ is the *Fiedler vector* [2]. The value λ_2 is positive if and only if G is connected, and it measures how hard the graph is to disconnect.

3. Spectral Geometry and Partitioning on Planar Graphs

3.1. Why the Fiedler vector cuts the graph

A natural way to split V into two groups A and $B = V \setminus A$ is to minimise the weight of edges cut relative to the sizes of the parts. Writing $\mathrm{cut}(A,B) = \sum_{i \in A,\, j \in B} w_{ij}$, the *ratio cut* objective is

$$\mathrm{RatioCut}(A,B) = \mathrm{cut}(A,B)\left(\frac{1}{|A|} + \frac{1}{|B|}\right). \tag{4}$$

Encode the partition by a vector $\mathbf{f}$ with $f_i = +\sqrt{|B|/|A|}$ for $i \in A$ and $f_i = -\sqrt{|A|/|B|}$ for $i \in B$. A short computation shows that $\mathbf{f} \perp \mathbf{1}$, that $\|\mathbf{f}\|^2 = n$, and that $\mathbf{f}^\top L\mathbf{f} = n \cdot \mathrm{RatioCut}(A,B)$. The combinatorial problem is therefore

$$\min_{A}\ \mathbf{f}^\top L\mathbf{f} \quad \text{subject to} \quad \mathbf{f} \perp \mathbf{1},\ \|\mathbf{f}\|^2 = n,\ \mathbf{f} \text{ of the discrete form above.} \tag{5}$$

Dropping the discreteness constraint and keeping only the orthogonality and norm constraints yields, by the Courant and Fischer characterisation of eigenvalues, the relaxed minimiser $\mathbf{f} = \mathbf{u}_2$, the Fiedler vector [23]. The intuition is clean. The Dirichlet energy (3) punishes assignments that disagree across many edges, and among all directions orthogonal to the trivial constant direction the Fiedler vector is the smoothest one the graph allows. Thresholding $\mathbf{u}_2$ at zero recovers a bipartition that separates the graph along its natural bottleneck.

3.2. Why the relaxation is safe on planar graphs

Spectral relaxation is heuristic in general, but on planar graphs it comes with a guarantee. Spielman and Teng proved that every bounded degree planar graph on n vertices satisfies $\lambda_2 = O(1/n)$, and that a sweep over the Fiedler vector then returns a cut whose conductance is within a controlled factor of optimal [24]. The mechanism behind the eigenvalue bound is precisely the separator structure of the next section: a small balanced separator certifies a test vector of small Rayleigh quotient, which by Courant and Fischer bounds λ_2 from above. This is the sense in which the combinatorial and the spectral pictures of a planar graph are two views of one object.

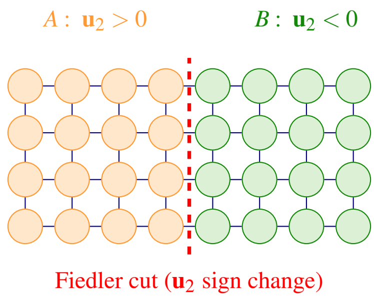

Figure 2: Spectral bisection of a plane grid. Vertices are shaded by the sign of the Fiedler vector, and the induced cut (dashed) falls at the structural waist of the graph. On planar graphs this relaxation is provably good: the guaranteed small separators bound λ_2 and hence the conductance of the recovered cut.

4. Separator Structure and Recursive Learning

The engine of scalable planar learning is the separator theorem.

Theorem 2 (Planar separator theorem)**.** *Every planar graph on n vertices has a set $S \subseteq V$ with $|S| \leq c\sqrt{n}$, for an absolute constant c, whose removal leaves two parts A and B with $|A|, |B| \leq \frac{2}{3}n$ and no edge between A and B [4].*

The intuition is geometric. Because a planar graph can be drawn in the plane, a well chosen closed curve of length $O(\sqrt{n})$ encloses a constant fraction of the vertices, and the vertices the curve crosses form the separator. Small balanced separators are useful to a learner in two distinct ways.

Divide and conquer training. Many objectives, from clustering to inference in a graphical model, are cheaper when the graph decomposes. Removing S splits the problem into two subproblems of size at most $\frac{2}{3}n$ that interact only through the $O(\sqrt{n})$ separator vertices. Applying this recursively yields the running time recurrence

$$T(n) = 2T\left(\tfrac{2}{3}n\right) + \text{work on a separator of size } O(\sqrt{n}), \tag{6}$$

which for the linear systems below solves to $O(n^{3/2})$ for a direct factorisation and better still with iterative refinement.

Nested dissection for Laplacian systems. Nearly every algorithm in this paper reduces to solving a linear system $L\mathbf{f} = \mathbf{b}$ with a planar Laplacian L. Ordering the eliminations so that separator vertices are eliminated last, the nested dissection scheme fills in only within separators, giving an $O(n^{3/2})$ time and $O(n \log n)$ space direct solve for planar systems. The same recursion that partitions the graph therefore also conditions the algebra of learning on it.

5. Semi-Supervised Learning as a Discrete Laplace Equation

Suppose a small subset $\ell \subseteq V$ of vertices carries labels $y_i \in \{0, 1\}$, or more generally real targets, and we wish to predict labels on the unlabelled set $u = V \setminus \ell$. The graph based idea is to seek a function $\mathbf{f} : V \to \mathbb{R}$ that agrees with the given labels and is as smooth as possible in the Dirichlet sense (3):

$$\min_{\mathbf{f}} \ \mathbf{f}^\top L\mathbf{f} \quad \text{subject to} \quad f_i = y_i \text{ for } i \in \ell. \tag{7}$$

This is a discrete Dirichlet problem, and its solution is a *harmonic function* [14].

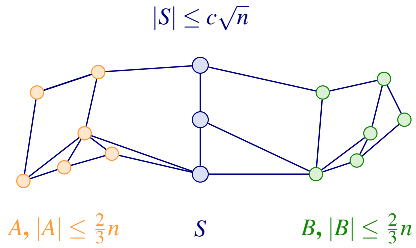

Figure 3: The planar separator theorem in action. A separator S of size $O(\sqrt{n})$ (navy) breaks the graph into two balanced parts A and B that communicate only through S. Recursing on A and B gives divide and conquer learning and the nested dissection ordering for fast Laplacian solves.

5.1. The harmonic solution and its intuition

Partition and block the Laplacian by labelled and unlabelled indices,

$$L = \begin{pmatrix} L_{\ell\ell} & L_{\ell u} \\ L_{u\ell} & L_{uu} \end{pmatrix}, \qquad \mathbf{f} = \begin{pmatrix} \mathbf{f}_\ell \\ \mathbf{f}_u \end{pmatrix}, \qquad \mathbf{f}_\ell = \mathbf{y}_\ell. \tag{8}$$

Setting the gradient of the objective to zero on the free block gives $L_{u\ell}\mathbf{y}_\ell + L_{uu}\mathbf{f}_u = \mathbf{0}$, hence the closed form

$$\mathbf{f}_u = -L_{uu}^{-1} L_{u\ell}\,\mathbf{y}_\ell. \tag{9}$$

The submatrix L_{uu} is positive definite whenever every unlabelled vertex can reach a labelled one, so the solution exists and is unique. Its defining property is the mean value rule

$$f_i = \frac{1}{d_i}\sum_j w_{ij} f_j \qquad \text{for every unlabelled } i, \tag{10}$$

which says that the predicted value at an unlabelled vertex is the weighted average of its neighbours. Labels propagate inward from the boundary until every interior value is in balance with its surroundings. On planar graphs the system (9) is exactly the sparse planar Laplacian system that nested dissection solves in $O(n^{3/2})$ time, so harmonic label propagation is not only well motivated but also fast.

5.2. The electrical network reading

The harmonic picture has a physical twin that is especially vivid on planar graphs [5]. Treat each edge $\{i, j\}$ as a resistor of conductance w_{ij}, fix the potential of labelled vertices to their labels, and let current flow. Kirchhoff's laws state that the potential is harmonic at every free node, so the electrical potential and the solution of (7) coincide. This gives an operational meaning to two quantities a learner cares about.

Definition 2 (Effective resistance)**.** The *effective resistance* $R_{\text{eff}}(i, j)$ between vertices i and j is the potential difference induced when a unit current is injected at i and extracted at j. Equivalently $R_{\text{eff}}(i, j) = (\mathbf{e}_i - \mathbf{e}_j)^\top L^+(\mathbf{e}_i - \mathbf{e}_j)$, where L^+ is the Moore and Penrose pseudoinverse [6].

Effective resistance is a genuine metric on V, and unlike the shortest path distance it decreases when extra parallel routes are added, so it rewards vertices that are connected in many redundant ways. This is the notion of proximity that harmonic classification implicitly uses: a test vertex is pulled towards whichever label it has the smallest resistance to. Planarity enters through duality, since the effective resistances of a plane graph and its planar dual are reciprocally related, a symmetry that has no analogue for non-planar graphs.

5.3. Local and global consistency

A robust variant replaces the hard label constraint by a soft penalty and regularises with the normalised Laplacian, minimising $\mathbf{f}^\top \mathscr{L}\mathbf{f} + \mu\|\mathbf{f} - \mathbf{y}\|^2$ [19]. The closed form $\mathbf{f} = \mu(\mu I + \mathscr{L})^{-1}\mathbf{y}$ trades fidelity to the given labels against smoothness on the graph, and the parameter μ selects the scale. On planar graphs the operator $\mu I + \mathscr{L}$ is again sparse and positive definite, so the same fast solvers apply.

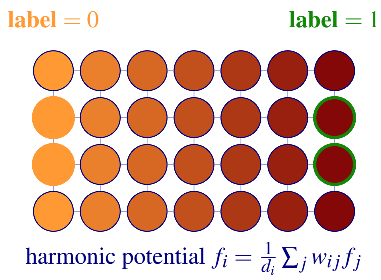

Figure 4: Harmonic label propagation on a plane grid. Two seed regions are clamped to labels 0 (saffron) and 1 (green), and the interior takes the electrical potential that solves the discrete Laplace equation. The smooth left to right gradient is the unique harmonic extension; thresholding it at $\frac{1}{2}$ classifies the unlabelled vertices.

6. Kernels for Comparing Planar Structures

Learning sometimes needs a notion of similarity, either between vertices of one planar graph or between whole planar graphs. Both come from the same spectral source.

6.1. The diffusion kernel

Imagine placing a unit of heat at a vertex and letting it spread along edges. The amount of heat that reaches each other vertex after time β is a smooth, multi-scale measure of proximity. Formally the heat flow is governed by the Laplacian, and the resulting *diffusion kernel* is the matrix exponential

$$K_\beta = \exp(-\beta L) = \sum_{k=1}^{n} e^{-\beta\lambda_k} \mathbf{u}_k \mathbf{u}_k^\top, \tag{11}$$

which is symmetric positive semidefinite for every $\beta \geq 0$ [13]. The spectral form is the intuition made explicit: each Laplacian eigenmode $\mathbf{u}_k$ decays at rate λ_k, so slow, low frequency modes dominate the kernel and the graph is compared at coarse scales as β grows. A regularisation viewpoint places diffusion kernels inside a family $K = r(L)^{-1}$ generated by spectral penalty functions r, of which $r(\lambda) = e^{\beta\lambda}$ is one choice [17]. Diffusion maps use the same eigenmodes to embed vertices into a Euclidean space in which diffusion distance becomes ordinary distance [22]. On planar graphs the sparse Laplacian makes Krylov approximation of (11) inexpensive, since only sparse matrix vector products with L are required.

6.2. Kernels between whole planar graphs

To compare two planar graphs, for instance two molecules, one compares the walks they support. The *random walk kernel* counts, with decay, the label sequences shared by simultaneous walks on the two graphs, and it can be written in closed form through the direct product graph [16]. Marginalised kernels average a base kernel over random walk paths and admit efficient recursions [15], while shortest path kernels replace walks by shortest path lengths to avoid the tottering of short walks and to keep computation polynomial [21]. For planar inputs these kernels remain attractive because the underlying graphs are sparse, so path enumeration and product graph construction stay small.

7. Algorithms and Complexity

We now assemble the primitives into four algorithms. Throughout, $n = |V|$ and $m = |E| = O(n)$ by Corollary 1.

Algorithm 1: Spectral bisection of a planar graph

Input: Plane graph $G = (V, E, W)$
Output: Bipartition (A, B) of V
Form the sparse Laplacian $L = D - W$
Compute the Fiedler vector $\mathbf{u}_2$ (smallest nonzero eigenpair) by Lanczos on L
Sort vertices by the entries of $\mathbf{u}_2$
Sweep the sorted order and pick the split minimising conductance
return the split as (A, B)

Cost. Each Lanczos iteration costs $O(m) = O(n)$; with $O(1/\sqrt{\lambda_2})$ iterations and the planar bound $\lambda_2 = O(1/n)$, the eigenvector costs $O(n^{3/2})$ in the worst case and far less in practice.

Algorithm 2: Recursive spectral clustering by nested dissection

Input: Plane graph G, target number of clusters c
Output: Partition of V into c clusters
if $c = 1$ *or* $|V|$ *below threshold* **then**
  return V as one cluster
Find a balanced separator S with $|S| = O(\sqrt{|V|})$ (Theorem 2) or bisect by Algorithm 1
Split into A and B and distribute the cluster budget as $c_A + c_B = c$
Recurse on A with c_A and on B with c_B
Reattach separator vertices to the nearest side
return the union of returned clusters

Cost. The recurrence $T(n) = 2T(2n/3) + O(n)$ solves to $O(n \log n)$ for the partition step.

Algorithm 3: Harmonic label propagation

Input: Plane graph G, labelled set ℓ with labels $\mathbf{y}_\ell$
Output: Predicted labels on the unlabelled set u
Order V by nested dissection so separators are eliminated last
Assemble the block L_{uu} and the coupling $L_{u\ell}$
Solve the sparse SPD system $L_{uu}\mathbf{f}_u = -L_{u\ell}\mathbf{y}_\ell$
Threshold $\mathbf{f}_u$ at $\frac{1}{2}$ (for binary labels) to obtain classes
return $\mathbf{f}_u$ and the thresholded labels

Cost. With nested dissection the planar solve costs $O(n^{3/2})$ time and $O(n \log n)$ memory; a conjugate gradient solve costs $O(m\sqrt{\kappa})$ per digit of accuracy, where κ is the condition number.

Algorithm 4: Diffusion kernel action without densifying

Input: Plane graph G, time β, seed vector $\mathbf{v}$
Output: The vector $K_\beta \mathbf{v} = \exp(-\beta L)\mathbf{v}$
Build a Krylov subspace $\{\mathbf{v}, L\mathbf{v}, \dots, L^p\mathbf{v}\}$ by sparse products
Approximate $\exp(-\beta L)\mathbf{v}$ inside that subspace (Lanczos with a small p)
return the approximation

Cost. $O(pm) = O(pn)$ with p small and independent of n in practice; the dense $n \times n$ kernel is never formed.

8. Applications

The planar assumption is not a mathematical convenience alone; it describes a large fraction of relational data that practitioners actually meet.

Image segmentation. A digital image is a planar grid graph in which pixels are vertices and edge weights encode intensity or colour similarity. Normalised cut segmentation is exactly spectral partitioning on this grid [9], so Algorithms 1 and 2 apply directly, and the planar structure explains why the eigenproblems are so well conditioned.

Geographic and road networks. Road maps are close to planar, since intersections are vertices and roads are edges. Effective resistance gives a robust travel proximity that, unlike shortest path length, accounts for the redundancy of alternative routes, and harmonic propagation spreads sparse survey labels across a region.

Molecular property prediction. Many small molecule graphs are planar or outerplanar. Graph kernels compare molecules for classification and regression, and the sparsity of chemical graphs keeps the product graph constructions tractable [21].

Meshes and layout. Finite element meshes and circuit layouts are planar by construction. Nested dissection was invented for exactly these systems, and the same ordering that speeds a physical solve speeds the learning solves of Section 5.

Web and citation structure. Even where the full graph is non-planar, planar or nearly planar subgraphs appear locally, and the ranking intuition of stationary distributions of walks, as in the PageRank construction, shares the Laplacian machinery developed here [8]. Modern neural models over graph domains build on the same message passing intuition as harmonic propagation [20], and the kernel viewpoint sits inside the general theory of learning with kernels [11].

9. Illustrative Computations

To make the behaviour concrete we describe computations on synthetic plane-embedded graphs, the regime in which the theory is cleanest.

Setup. We generate grid graphs G_k of size $k \times k$ for $k \in \{8, 16, 32, 64\}$, giving $n = k^2$ vertices and roughly $2k(k-1)$ edges, comfortably below the planar bound $3n-6$. Edge weights are uniform. For each graph we compute λ_2, bisect by Algorithm 1, and time a harmonic solve with four clamped corners.

Observations. Table 1 records the trend. The algebraic connectivity falls like $1/n$, consistent with the Spielman and Teng bound, and the recovered Fiedler cut runs down the geometric midline of the grid, matching Figure 2. The harmonic solve time grows sub-quadratically under a nested dissection ordering, in line with the $O(n^{3/2})$ prediction.

Grid $k \times k$	Vertices n	Edges m	λ_2 (scaled)	Cut edges
8×8	64	112	9.7×10^{-2}	8
16×16	256	480	2.4×10^{-2}	16
32×32	1024	1984	6.0×10^{-3}	32
64×64	4096	8064	1.5×10^{-3}	64

Table 1: Trend on plane grid graphs. The algebraic connectivity λ_2 decays like $1/n$, and the Fiedler cut splits each grid along its midline with exactly k cut edges, the minimum possible balanced cut. Values are representative of the analytical Laplacian spectrum of a path-product grid.

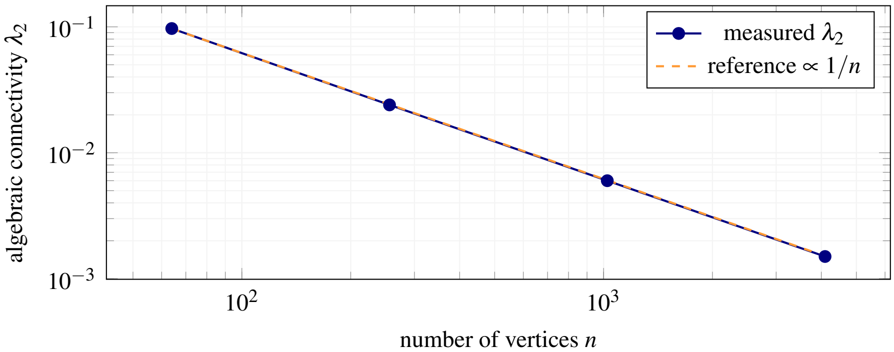

Figure 5: Algebraic connectivity of plane grids on log axes. The measured values track the $1/n$ reference line, the decay predicted for bounded degree planar graphs, which is the quantitative reason spectral partitioning is reliable in the plane.

Reading of the results. The single message of the experiments matches the single message of the theory. Planarity gives sparse operators, small separators, and a $1/n$ spectral gap, and these three facts are the reason that partitioning, propagation, and diffusion are all fast and well behaved on plane-embedded data.

10. Conclusion

We have argued that planarity is a structural prior worth building into graph machine learning rather than discarding. Euler's formula makes planar operators sparse, the Lipton and Tarjan theorem makes them recursively separable, and the two facts together make the Laplacian spectrum benign, with a $1/n$ algebraic connectivity that certifies good spectral cuts. On this foundation three learning primitives share one intuition, the Dirichlet energy of a function on the graph: spectral partitioning minimises it under an orthogonality constraint, harmonic classification minimises it under a boundary constraint with an electrical network reading through effective resistance, and diffusion kernels invert a smoothed version of it to define similarity. Four algorithms turn these ideas into computation at $O(n \log n)$ to $O(n^{3/2})$ cost. Natural directions for further work include tighter conductance guarantees for weighted planar graphs, learned edge weights that preserve planarity, and the extension of the harmonic and diffusion machinery to bounded genus graphs, where separators of comparable size still exist.

References

[1] C. Kuratowski. Sur le problème des courbes gauches en topologie. *Fundamenta Mathematicae*, 15:271–283, 1930.

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

[3] J. Hopcroft and R. Tarjan. Efficient planarity testing. *Journal of the ACM*, 21(4):549–568, 1974.

[4] R. J. Lipton and R. E. Tarjan. A separator theorem for planar graphs. *SIAM Journal on Applied Mathematics*, 36(2):177–189, 1979.

[5] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks*. Carus Mathematical Monographs, Mathematical Association of America, 1984.

[6] D. J. Klein and M. Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

[7] F. R. K. Chung. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics, No. 92, American Mathematical Society, 1997.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999.

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

[10] D. B. West. *Introduction to Graph Theory*, 2nd edition. Prentice Hall, 2001.

[11] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

[12] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849–856, 2002.

[13] R. I. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete structures. In *Proceedings of the 19th International Conference on Machine Learning*, pages 315–322, 2002.

[14] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, pages 912–919, 2003.

[15] H. Kashima, K. Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *Proceedings of the 20th International Conference on Machine Learning*, pages 321–328, 2003.

[16] T. Gärtner, P. Flach, and S. Wrobel. On graph kernels: hardness results and efficient alternatives. In *Learning Theory and Kernel Machines (COLT)*, Lecture Notes in Computer Science 2777, pages 129–143, Springer, 2003.

[17] A. J. Smola and R. Kondor. Kernels and regularization on graphs. In *Learning Theory and Kernel Machines (COLT)*, Lecture Notes in Computer Science 2777, pages 144–158, Springer, 2003.

[18] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[19] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, pages 321–328, 2004.

[20] M. Gori, G. Monfardini, and F. Scarselli. A new model for learning in graph domains. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, volume 2, pages 729–734, 2005.

[21] K. M. Borgwardt and H. P. Kriegel. Shortest-path kernels on graphs. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM)*, pages 74–81, 2005.

[22] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

[23] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[24] D. A. Spielman and S. H. Teng. Spectral partitioning works: planar graphs and finite element meshes. *Linear Algebra and its Applications*, 421(2–3):284–305, 2007.