NODE2MOTIF: Hierarchical Invariant Embeddings of Structured Graphs Using the Bispectrum

Sophia Sanborn[1,2,3]
sanborn@berkeley.edu

Ram Mehta[5]
ram.mehta@gmail.com

Noah Shutty[1,4]
noaj@stanford.edu

Christopher Hillar[2,5]
hillarmath@gmail.com

[1]Intel AI Research; [2]Redwood Center for Theoretical Neuroscience
[3]Berkeley AI Research; [4]Stanford Institute for Theoretical Physics; [5]Awecom, Inc.

Abstract

We present an unsupervised algorithm for embedding graphs into a hierarchy of continuous features that separate local isomorphism classes. The approach incorporates group representation theory and the bispectrum, a third-order polyspectrum that is a complete invariant for small graphs, although any graph invariant can be used in the method. As a direct application, we compute a structural distance metric between organic molecules from the QM9 dataset.

1 Introduction

Many graphs in nature contain structural similarities at multiple scales. Organic molecules, for example, are composed of smaller distinctive fragments referred to as functional groups. Proteins, similarly, are compositions of amino acids. Following (Milo et al., 2002), we refer to these distinctive connectivity patterns as graph *motifs*.

Determining the exact presence and distribution of motifs in a set of graphs is tantamount to the NP-complete *subgraph isomorphism problem*. However, for real-world applications, the subgraph isomorphism problem is neither a suitable nor desirable formulation. Graphs obtained from data typically involve both noise on edge weights and labelling ambiguities (permutations of vertex labels), which make computing exact matches untenable. Additionally, it is significantly more powerful in practice to construct a continuous representational space whose metric reflects structural similarities rather than obtain a binary solution to a decision problem. An unsupervised approach is also often desired, as labeled data is highly limited for many tasks involving graph-structured data.

We advocate the structural graph embedding approach: Given a graph, compute a smooth map β to a hierarchical feature space Φ that collapses isomorphic neighborhoods and maps nodes with similar surrounding motif structure to similar features. The particular embedding we propose here incorporates the bispectrum from harmonic analysis and group representation theory (Tukey, 1953; Gourd, Gauthier, and Younes, 1989; Sadler and Giannakis, 1992; Kakarala, 1993; Risi Kondor and Borgwardt, 2008) although other graph invariants (e.g. Laplacian spectrum) can be used.

Our algorithm, NODE2MOTIF, combines graph invariants with message passing to define a continuous node embedding that identifies isomorphic sub-structures and provides a hierarchical, invariant metric over graphs. We illustrate the method's utility on pedagogical examples and the QM9 dataset.

2 Background

An undirected weighted graph G on v nodes $V = \{1, 2, \ldots, v\}$ is defined by a real symmetric *adjacency matrix* $A_G = W = [W_{ij}]_{i,j=1}^{v} \in \mathbb{R}^{v \times v}$ that encodes the strength of edges between pairs

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

of nodes. Consider the adjacency matrix $A = A_{N_i}$ corresponding to a neighborhood N_i of node i. We seek a map $\beta(A) = \Phi \in \mathbb{R}^d$ that sends isomorphic A to the same point in feature space Φ (i.e., $\beta(A) = \beta(PAP^\top)$ for all permutation matrices $P \in \{0,1\}^{v \times v}$); such a map is called a *graph invariant*. Additionally, we desire this map to send non-isomorphic subgraphs to distinct points; i.e., the map is a *complete* invariant. Finally, we wish for a metric in feature space Φ to reflect *structural similarity*; that is, nearby points in Φ should correspond to graphs that are nearly isomorphic.

One widely used graph invariant is the *graph spectrum*, which consists of the eigenvalues of a matrix representation of a graph. Commonly, the matrices used are W or the graph Laplacian $L = D - W$, where D is the diagonal matrix of vertex degrees. The spectrum is invariant to permutation similarity, as eigenvalues are invariant to change of basis. However, it is not a complete invariant as many sets of graphs are non-isomorphic but cospectral; see (Van Dam and Haemers, 2003) and Fig. 1.

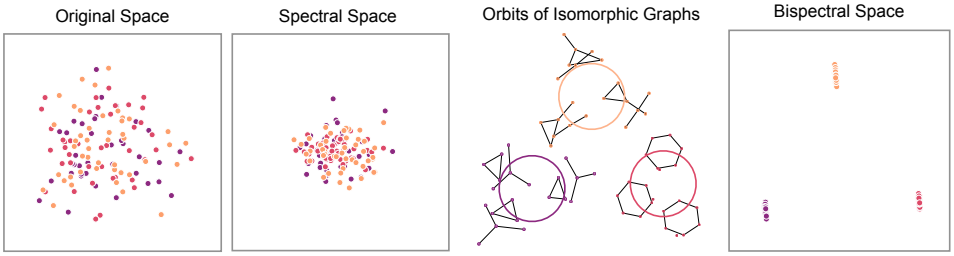


Figure 1: **Graph Spectra**. Noise and random relabelling applied to 9 graphs (center right) with same Laplacian spectrum, projected into PCA space using features: (far left) adjacency matrices, (center left) Laplacian spectrum, and (far right) graph bispectrum. Spectral features collapse the nine example graphs to one cluster, whereas the bispectrum separates non-isomorphic classes.

We also onsider an alternative invariant—the *graph bispectrum*—which is defined using the representation theory of the symmetric group S_v acting on graphs (R. Kondor, 2007; R. Kondor, 2008). Let $Irr(S_v)$ denote inequivalent irreducible representations of S_v. We assume that $\rho \in Irr(S_v)$ are also *unitary* representations (i.e., $\rho^{-1} = \rho^*$, with ρ^* the conjugate transpose), which can be made the case for any finite group. Intuitively, these ρ are the "fundamental frequencies" of the group. In particular, for $\mathbb{Z}/n\mathbb{Z}$ (the group of *integers modulo* n) these are the classical discrete Fourier frequencies.

Definition 2.1. *Graph bispectrum.* Given a weight matrix W and $\rho_i, \rho_j \in Irr(S_v)$, the graph bispectrum $\beta_{ij}(W)$ is the matrix: $\beta_{ij}(W) = C_{ij}^*(\widehat{W}(\rho_i) \otimes \widehat{W}(\rho_j))^* C_{ij} \left(\bigoplus_{\phi \in I_{ij}} \widehat{W}(\phi)\right)$.

Here, $\widehat{W}$ is the generalized Fourier transform of W, the multiset I_{ij} consists of those representations in $Irr(S_v)$ that appear in the decomposition of $\rho_i \otimes \rho_j$ into irreducible representations, and matrices C_{ij} are the corresponding change of basis (usually called the *Clebsch-Gordan matrices*; important also in physics). For each pair i and j, the bispectrum β_{ij} is a graph invariant.

All unweighted, undirected graphs on 6 nodes or fewer are identified up to isomorphism using the graph bispectrum; that is, the bispectrum is a complete invariant for graphs up to six nodes. Notably, all cospectral, non-isomorphic graphs on 7 nodes have distinct bispectra. Thus, the bispectrum provides a powerful alternative representational space to distinguish motifs.

3 NODE2MOTIF

Using a local graph invariant, we construct an algorithm for embedding nodes from a labelled weighted graph into a feature space that is invariant to permutations of neighborhood structure. The basic object we compute over is a generalization of the adjacency matrix that incorporates node and edge labels. For a graph G with nodes i having labels $L_i \in \mathbb{C}$ (e.g. atom type) and *labeling function* $\Gamma(L_i, L_j) \in \mathbb{C}$, we define the *labeled adjacency matrix* $\overline{A_G}(i,j) = \Gamma(L_i, L_j) \cdot W_{ij}$.

Given a node i in graph G, the $1-$hop *neighborhood* N_i is the subgraph of vertices that are connected to i, as well as edges between all such vertices. We compute first an embedding for i as the graph invariant of the labeled adjacency matrix of N_i; that is, $\Phi_i^0 = \beta(\overline{A_{N_i}})$ (for unlabelled graphs, we set $\Gamma = 1$). Embeddings are also computed for k iterations, where, after the zeroth iteration, labeling is a

function of the previous embedding: $\Gamma(\Phi_i^{k-1}, \Phi_j^{k-1})$. This iterative strategy resembles "message-passing" (Hansen, 1970); as k increases, the embeddings of more distant nodes are incorporated.

Algorithm 1: Node2Motif

Input:
Labeled graph $G = (V, E)$; Labeling function Γ; Labeled adjacency matrix $\overline{A_G}$; Iterations K.
Output:
Node embeddings Φ_i^k for $i \in V$ and $k \in \{0, 1, \ldots, K\}$.

for k *in* $\{0, 1, \ldots, K\}$ **do**
  for i *in* V **do**
    N_i = 1-hop neighborhood of vertex i in $\overline{A_G}$;
    $\Phi_i^k = \beta(\overline{A_{N_i}})$;
  end
  for all i, j **do**: $\overline{A_G}(i, j) = \Gamma(\Phi_i^k, \Phi_j^k) \cdot W_{ij}$;
end

For the purposes of this work, we use the bispectrum graph invariant and choose the following initial labeling function. Node and edge types are assigned uniform random numbers in $[-1, 1]$. For a connected pair of nodes, the real component of Γ is the product of the node labels, and the imaginary component is the edge label. For the next iterations, the labeling function is the product of the first (complex) principal component (over all graphs in a dataset) of the embeddings of node i and node j.

4 Related Work

The worst-case time complexity for identifying isomorphisms of unweighted graphs is unknown although several algorithms exist (Darga, Sakallah, and Markov, 2008; McKay and Piperno, 2014; Babai, 2016). These are limited for data analysis, though, as they do not directly capture approximate iso-/automorphisms and do not provide a metric for computing similarity on graphs.

Several methods also exist for embedding graphs. Most are based on random walks or global graph spectra, and are suited for identifying community structure, e.g. (Perozzi, Al-Rfou, and Skiena, 2014; Risi Kondor and Pan, 2016; Grover and Leskovec, 2016). Somewhat similar to our approach are (Shervashidze et al., 2011; Neumann et al., 2016), which also propagate information iteratively. GraphWave (Donnat et al., 2018) is notable for emphasizing *structural similarity* between vertices.

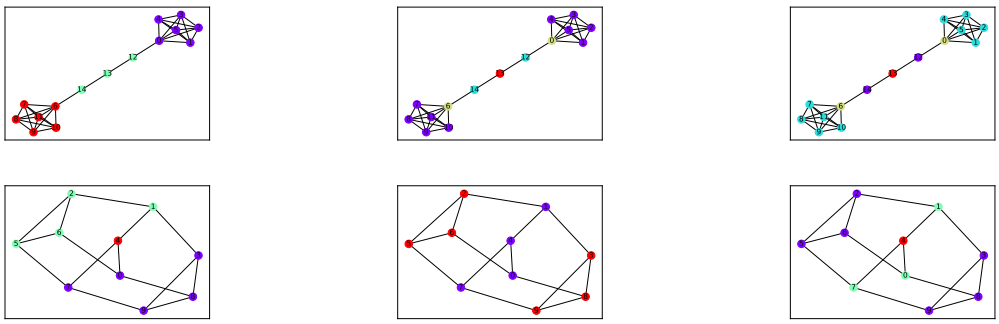
Graph convolutional neural networks (GCNNs) provide a promising alternative for learning continuous representations on graphs. In many approaches to GCNNs, invariance to node relabeling is achieved through summation or pooling of the vector representations of neighboring nodes (Wu et al., 2019). Summation is indeed invariant to relabeling. However, it is *too* invariant, as it does not preserve local connectivity structure. In line with recent contributions (Kondor et al., 2018), we incorporate group representation theory to find feature spaces that preserve symmetries.

5 Experiments & Results

We present example embeddings on two small graphs and close with an application to molecules.

Orbit Partitions. Structural similarity between vertices in a graph is formalized by the *orbit partition*. The automorphism group of edge-preserving bijections acts on a graph to permute its vertex labels. Two vertices u, v lie in the same orbit iff there is an automorphism which maps $u \mapsto v$. The orbits of a graph form its orbit partition: an equivalence relation for structurally identical vertices.

There is no known algorithm that computes the orbit partition of a graph in polynomial time. Computing the orbit partition is closely related to two important problems whose computational complexity is unknown: the problem of determining whether two graphs are isomorphic, and the problem of determining whether a graph has a nontrivial automorphism (Lubiw, 1981). We relax the problem of computing the orbit partition of a graph into a structural embedding problem and ask whether NODE2MOTIF embeddings distinguish orbit partitions in small graphs.



(a) NODE2VEC (b) GRAPHWAVE (c) NODE2MOTIF

Figure 2: **Identifying Orbit Partitions**. Node embeddings for (a) NODE2VEC, (b) GRAPHWAVE, and (c) NODE2MOTIF. Embeddings are clustered and nodes are colored accordingly.

Fig. 2 compares embeddings learned by NODE2VEC, GRAPHWAVE, and NODE2MOTIF on two example graphs. For each node, we compute a 3-layer hierarchical embedding, perform hierarchical clustering, and assign vertices colors corresponding to cluster labels. Across both graphs, NODE2VEC identifies communities rather than structural similarities. GRAPHWAVE and NODE2MOTIF both learn structural embeddings of the "barbell" graph (above). For the "trianglion" graph (below), GRAPHWAVE fails to separate the unique central vertex from neighbors (over many hyperparameters). This vertex is stabilized by the automorphism group of the graph and is identified by NODE2MOTIF.

Organic Molecules. We examine the performance of our embedding method on a dataset of 133,885 organic molecules (QM9; Ruddigkeit et al., 2012; Ramakrishnan et al., 2014). For this dataset, each node is an atom, and each edge is a bond. Atom type and bond type are encoded with the labeling function described in section 3. For each node in each molecule, we compute a 3-layer hierarchical embedding. Figure 3 shows the results of querying the embedded dataset for a single node in the molecule displayed in the bottom left cell. Over the entire dataset, NODE2MOTIF yields a distance metric that captures structural similarity and identifies near isomorphic molecules.

Most methods for molecular database query rely on hand designed features or string-based representations (Willett, Barnard, and Downs, 1998; Weininger, 1988). More recent methods adapt features to the statistics of molecular graphs (Duvenaud et al., 2015); however, most convolutional approaches lose important structural information through pooling. Our method is fully unsupervised and provides an invariant metric over molecules, which may be used for similarity queries in molecular databases. We see potential for our method as unsupervised pre-training for downstream supervised learning.

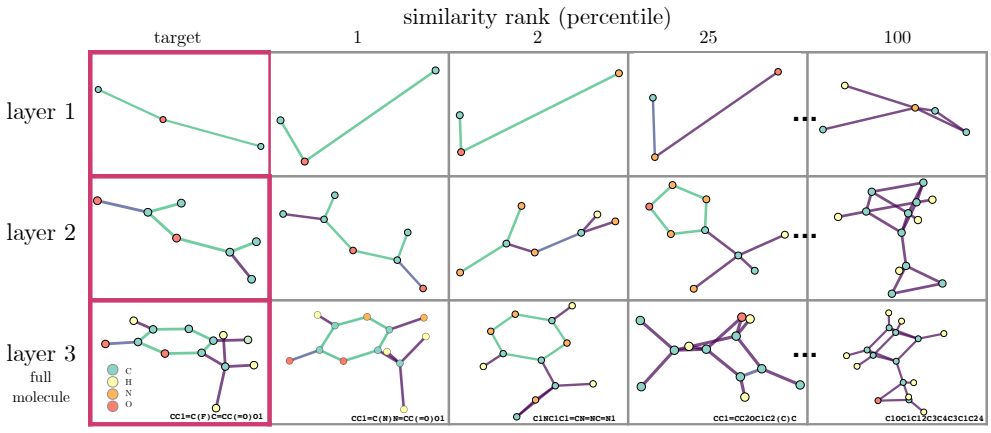


Figure 3: **Molecules.** Molecular similarity in NODE2MOTIF embedding space. The target molecule is depicted in the bottom left cell. For each embedding layer, we show the molecular fragment (motif) that is 'visible' to the target node embedding, and a set of molecules/fragments that are rank-ordered by Euclidean distance from the target. Rightmost column shows fragments least similar to target.

References

Tukey, J. (1953). "The spectral representation and transformation properties of the higher moments of stationary time series". In: *Reprinted in The Collected Works of John W. Tukey* 1, pp. 165–184.

Hansen, Per Brinch (1970). "The nucleus of a multiprogramming system". In: *Communications of the ACM* 13.4, pp. 238–241.

Lubiw, Anna (1981). "Some NP-complete problems similar to graph isomorphism". In: *SIAM Journal on Computing* 10.1, pp. 11–21.

Weininger, David (1988). "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules". In: *Journal of chemical information and computer sciences* 28.1, pp. 31–36.

Gourd, F., J. Gauthier, and H. Younes (1989). "Une méthode d'invariants de l'analyse harmonique en reconnaissance de formes". In: *Traitement du signal* 6.3, pp. 161–178.

Sadler, B. and G. Giannakis (1992). "Shift-and rotation-invariant object reconstruction using the bispectrum". In: *JOSA A* 9.1, pp. 57–69.

Kakarala, R. (1993). "A group theoretic approach to the triple correlation". In: *IEEE Workshop on higher order statistics*, pp. 28–32.

Willett, Peter, John M Barnard, and Geoffrey M Downs (1998). "Chemical similarity searching". In: *Journal of chemical information and computer sciences* 38.6, pp. 983–996.

Milo, Ron et al. (2002). "Network motifs: simple building blocks of complex networks". In: *Science* 298.5594, pp. 824–827.

Van Dam, Edwin R and Willem H Haemers (2003). "Which graphs are determined by their spectrum?" In: *Linear Algebra and its applications* 373, pp. 241–272.

Kondor, R. (2007). "A complete set of rotationally and translationally invariant features for images". In: *Arxiv preprint cs/0701127*.

Darga, Paul T, Karem A Sakallah, and Igor L Markov (2008). "Faster symmetry discovery using sparsity of symmetries". In: *2008 45th ACM/IEEE Design Automation Conference*. IEEE.

Kondor, R. (2008). *Group theoretical methods in machine learning*. Columbia University, PhD Thesis.

Kondor, Risi and Karsten M Borgwardt (2008). "The skew spectrum of graphs". In: *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 496–503.

Shervashidze, Nino et al. (2011). "Weisfeiler-lehman graph kernels". In: *Journal of Machine Learning Research* 12.Sep, pp. 2539–2561.

Ruddigkeit, Lars et al. (2012). "Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17". In: *Journal of chemical information and modeling* 52.11.

McKay, Brendan D and Adolfo Piperno (2014). "Practical graph isomorphism, II". In: *Journal of Symbolic Computation* 60, pp. 94–112.

Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena (2014). "Deepwalk: Online learning of social representations". In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 701–710.

Ramakrishnan, Raghunathan et al. (2014). "Quantum chemistry structures and properties of 134 kilo molecules". In: *Scientific Data* 1.

Duvenaud, David K et al. (2015). "Convolutional networks on graphs for learning molecular fingerprints". In: *Advances in neural information processing systems*, pp. 2224–2232.

Babai, László (2016). "Graph isomorphism in quasipolynomial time". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. ACM, pp. 684–697.

Grover, Aditya and Jure Leskovec (2016). "node2vec: Scalable feature learning for networks". In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 855–864.

Kondor, Risi and Horace Pan (2016). "The multiscale laplacian graph kernel". In: *Advances in Neural Information Processing Systems*, pp. 2990–2998.

Neumann, Marion et al. (2016). "Propagation kernels: efficient graph kernels from propagated information". In: *Machine Learning* 102.2, pp. 209–245.

Donnat, Claire et al. (2018). "Learning structural node embeddings via diffusion wavelets". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, pp. 1320–1329.

Kondor, R et al. (2018). "Covariant compositional networks for learning graphs". In: *arXiv preprint arXiv:1801.02144*.

Wu, Zonghan et al. (2019). "A comprehensive survey on graph neural networks". In: *arXiv preprint arXiv:1901.00596*.