The Graphlet Spectrum

Risi Kondor RISI@GATSBY.UCL.AC.UK
Gatsby Computational Neuroscience Unit, UCL, 17 Queen Square, London, WC1N 3AR, U.K.

Nino Shervashidze NINO.SHERVASHIDZE@TUEBINGEN.MPG.DE
Max Planck Institute for Biological Cybernetics and Max Planck Institute for Developmental Biology, Spemannstr. 41, 72076 Tübingen, Germany

Abstract

Current graph kernels suffer from two limitations: graph kernels based on counting particular types of subgraphs ignore the relative position of these subgraphs to each other, while graph kernels based on algebraic methods are limited to graphs without node labels. In this paper we present the *graphlet spectrum*, a system of graph invariants derived by means of group representation theory that capture information about the number as well as the position of labeled subgraphs in a given graph. In our experimental evaluation the graphlet spectrum outperforms state-of-the-art graph kernels.

1. Introduction

Over recent years, graph kernels have grown to become an important branch of graph mining. Their fundamental purpose is to represent a graph by features in a reproducing kernel Hilbert space. While most graph kernels derive these features by counting particular types of subgraphs, such as walks, shortest paths, subgraphs of fixed size k, or subtrees (Kashima et al., 2003; Gärtner et al., 2003; Borgwardt & Kriegel, 2005; Borgwardt et al., 2007; Bach, 2008), recently, a group theoretical approach was proposed and shown to have state-of-the-art performance (Kondor & Borgwardt, 2008). However, both approaches have limitations: in counting subgraphs, the graph-theoretic approach completely ignores the relative position of subgraphs within the graph, while the algebraic approach suffers from the fact that it is restricted to unlabeled graphs, which are rare in applications.

Presented at *ILP-MLG-SRL*, Leuven, Belgium, 2009.

In this work, we overcome these two limitations by defining a new group-theoretic approach that allows both for labeled subgraphs and considers the relative position of subgraphs.

2. Graph invariants

Let us first introduce some notations. Let $\mathcal{G}$ be a directed weighted graph of n vertices. We represent $\mathcal{G}$ by its adjacency matrix $A \in \mathbb{R}^{n\times n}$, where $[A]_{i,j} \in \mathbb{R}$ is the weight of the edge from vertex i to vertex j. $\mathbb{S}_n$ denotes the **symmetric group** of degree n. Given a function $f\colon \mathbb{S}_n \to \mathbb{R}$, the group structure suggests defining the **left–translate** of f by $\pi \in \mathbb{S}_n$ as

$$f^\pi\colon \mathbb{S}_n \to \mathbb{R}, \qquad f^\pi(\sigma) = f(\pi^{-1}\sigma).$$

In terms of any complete set of inequivalent irreducible representations $\{\rho_\lambda\}_{\lambda\vdash n}$ of $\mathbb{S}_n$ the **Fourier transform** of a function $f\colon \mathbb{S}_n \to \mathbb{R}$ is defined as the sequence of matrices

$$\widehat{f}(\lambda) = \sum_{\sigma\in\mathbb{S}_n} f(\sigma)\,\rho_\lambda(\sigma) \qquad \lambda \vdash n.$$

Of the several properties of ordinary Fourier transformation inherited by such generalized Fourier transforms, we are particularly interested in the **translation theorem**, which, coupled with the unitarity of $\rho_\lambda(\pi)$, tells us that the matrices

$$\widehat{a}(\lambda) = \widehat{f}(\lambda)^\dagger \cdot \widehat{f}(\lambda), \quad \lambda \vdash n$$

are translation invariant.

Kondor and Borgwardt (2008) show that if we encode the adjacency matrix in the function

$$f_A(\sigma) = A_{\sigma(n),\sigma(n-1)}, \tag{1}$$

then permuting the vertices of $\mathcal{G}$ by π transforms f_A exactly into $(f_A)^\pi$.

3. The graphlet spectrum

A common alternative to the algebraic approach proposed in the above work is to characterize graphs in terms of the frequency or position of certain elementary subgraphs embedded within them. Depending on the context these small subgraphs are usually called **graphlets** or **motifs**. Given a graphlet g of $k < n$ vertices whose adjacency matrix we denote with the same letter g, the **indicator**

$$\mu_g(v_1, v_2, \ldots, v_k) = \begin{cases} 1 & \text{if } \; g_{i,j} \leq A_{v_i, v_j} \quad \forall\, i, j, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

captures whether g is a subgraph of $\mathcal{G}$ at position $(v_1, v_2, \ldots, v_k)$. If we replace $\leq$ by $=$ in (2), then the corresponding indicator μ_g^{ind} captures whether g is an induced subgraph at the same position.

The fundamental observation motivating the present work is that (at least for unweighted graphs), f_A, as defined in (1), can be re-written as

$$f_A(\sigma) = \mu_e(\sigma(n), \sigma(n-1)),$$

where e stands for the elementary graphlet of two vertices and a single directed edge. In other words, f_A encodes where the edge e occurs in $\mathcal{G}$ as a subgraph. It is easy to extend this idea to larger graphlets by

$$f_{A,g}(\sigma) = \mu_g(\sigma(n), \sigma(n-1), \ldots, \sigma(n-k+1)), \tag{3}$$

or the same with μ_g^{ind}. Crucially, $f_{A,g}$ will still obey the same transformation property as f_A did, since if μ_g^π is the indicator of the permuted adjacency matrix A^π, then

$$\mu_g^\pi(\pi(v_1), \pi(v_2), \ldots, \pi(v_k)) = \mu_g(v_1, v_2, \ldots, v_k), \tag{4}$$

hence $\mu_g^\pi(v_1, \ldots, v_k) = \mu_g(\pi^{-1}(v_1), \ldots, \pi^{-1}(v_k))$,and therefore

$$f_{A^\pi,g}(\sigma) = \mu_g(\pi^{-1}\sigma(n), \ldots, \pi^{-1}\sigma(n-k+1)) = f_{A,g}(\pi^{-1}\sigma) = (f_{A,g})^\pi(\sigma). \tag{5}$$

This means that we can invoke the machinery of power spectra, skew spectra, etc. to derive graph invariants, but these new invariants will be sensitive to the presence of entire subgraphs in $\mathcal{G}$ and not just individual edges.

An attractive feature of our approach is that given a small library $g_1, g_2, \ldots, g_m$ of graphlets we can compute a separate f_{A,g_i} function for each graphlet, and then form invariants from all possible combinations of these functions, capturing information about the relative position of different types of subgraphs as well as different subgraphs of the same type. Since in this case second order invariants already yield a rich set of features, we forgo computing higher order, more expensive invariants, such as the skew spectrum. Our exact definition of the graphlet spectrum is as follows.

Definition 1 *Given a graph $\mathcal{G}$ of n vertices and adjacency matrix A, relative to a collection $g_1, g_2, \ldots, g_m$ of graphlets and an indicator function such as (2), the* ***graphlet spectrum*** *of $\mathcal{G}$ is defined to be the sequence of matrices*

$$\widehat{q}_{i,j}(\lambda) = \big(\widehat{f}_{A,g_i}(\lambda)\big)^\dagger \cdot \widehat{f}_{A,g_j}(\lambda), \qquad j \leq i, \qquad \lambda \vdash n, \tag{6}$$

where f_{A,g_i} is defined as in (3).

Proposition 1 *Each scalar component $[\widehat{q}_{i,j}(\lambda)]_{a,b}$ of the graphlet spectrum is a graph invariant.*

4. Computational considerations

More often than not, the biggest challenge in applying representation theoretical ideas to real world problems is making the necessary computations scalable. In the case of the graphlet spectrum at first sight it appears that computing the Fourier transform (2) already demands $O((n!)^2)$ time, which is clearly forbiddingly expensive. There are two key ingredients to reducing this computational burden to a level that is feasible in a practical algorithm: sparsity and the theory of fast Fourier transforms.

Since the Fourier transform $f \mapsto \widehat{f}$ is a unitary transformation, the combined size of the $\widehat{f}(\lambda)$ matrices appearing in (2) is $n!$. However, any f defined by (3) is a so-called right $\mathbb{S}_{n-k}$–invariant function. For such functions most $\widehat{f}(\lambda)$ Fourier components turn out to be identically zero, and even the remaining components will have a characteristic column-sparse structure.

The reason that $\widehat{f}$ can be efficiently computed is not just that it is sparse, but that its sparsity structure is closely matched to the structure of the non-commutative fast Fourier transforms that are gaining popularity in the non-commutative harmonic analysis community (Rockmore, 1997; Clausen, 1989). We provide the following result without proof:

Proposition 2 *If $f\colon \mathbb{S}_n \to \mathbb{R}$ is defined as in (3), then in Young's Orthogonal Representation its Fourier transform can be computed in*

$$\left[\frac{(n+1)\,n\,(n-1)}{3} - \frac{(m+1)\,m\,(m-1)}{3}\right]\frac{n!}{m!} \tag{7}$$

scalar operations, where m is a shorthand for $n-k$. For fixed k this expression grows as $O(n^{2+k})$.

	MUTAG	ENZYMES	NCI1	NCI109
Number of instances/classes	188/2	600/6	4110/2	4127/2
Max. number of nodes	28	126	111	111
Graphlet spectrum	88.11 (0.46)	**35.42** (0.58)	**65.0** (0.09)	**65.31** (0.08)
Reduced skew spectrum	**88.61** (0.21)	25.83 (0.34)	62.72 (0.05)	62.62(0.03)
Graphlet count kernel	81.7 (0.67)	23.94 (0.4)	54.34 (0.04)	52.39 (0.09)

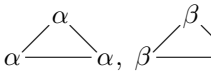
Table 1. Prediction accuracy in percent for the graphlet spectrum features and state of the art graph kernels on four classification benchmarks in 10 repetitions of 10-fold cross-validation. Standard errors are indicated in parentheses. Best results for each datasets are in bold.

We aim for applications involving medium sized graphs (few hundred nodes), and a handful of graphlets with k in the range 2 to 6.

5. Experiments and discussion

We assess the performance of the graphlet spectrum features on several benchmark datasets of chemical structures of molecules. The experiments consisted of running SVMs on the above data using a linear kernel on top of the the graphlet spectrum features. For comparison, we applied a linear kernel on the reduced skew spectrum features from (Kondor & Borgwardt, 2008) and a graphlet count kernel that counts the number of common graphlets in two graphs (Shervashidze et al., 2009). Both these kernels had been shown to outperform the classic random walk kernel (Gärtner et al., 2003) in earlier studies.

One of the strengths of the graphlet spectrum is that it allows the practitioner to use graphlets specifically designed to pick out salient features, such as functional groups in molecules. In our experiments we started with a minimal set of graphlets and saw performance increase as we added further ones one by one. The graphlets used in our experiments were the following:

- MUTAG: C−C, C−C−C, C−C−C−C, C−N, O−N, ∗−∗;
- NCI1 and NCI109: C−C, C−N, C−O, O−N, O−O, N−N;
- ENZYMES: ∗−∗, 

α−α, α−β, β−β;

where $* - *$ denotes an edge with arbitrary node labels. For fair comparison in the graphlet count kernel we used the same graphlets. Further experimentation and incorporating more knowledge from chemistry could lead to a siginificantly more powerful system of graphlets for organic molecules. It is important to stress that computational time, while a constraint, was not the limiting factor here. For enzymes α and β denote α-helices and β-sheets, respectively.

Experiments show that on graphs of medium size (up to a few hundred vertices) the graphlet spectrum is compatible with state of the art graph kernels, and in several cases outperforms all other methods. Theoretical results from non-commutative harmonic analysis and the representation theory of $\mathbb{S}_n$, together with a custom-built FFT library allow the graphlet spectrum to scale up to real-world problems with relative ease.

References

Bach, F. (2008). Graph kernels between point clouds. *Proc. Intl. Conf. Machine Learning* (pp. 25–32).

Borgwardt, K. M., & Kriegel, H.-P. (2005). Shortest-path kernels on graphs. *Proc. Intl. Conf. Data Mining* (pp. 74–81).

Borgwardt, K. M., Petri, T., Vishwanathan, S. V. N., & Kriegel, H.-P. (2007). An efficient sampling scheme for comparison of large graphs. *MLG*.

Clausen, M. (1989). Fast generalized Fourier transforms. *Theor. Comput. Sci.*, 55–63.

Gärtner, T., Flach, P., & Wrobel, S. (2003). On graph kernels: Hardness results and efficient alternatives. *Proc. Annual Conf. Computational Learning Theory* (pp. 129–143). Springer.

Kashima, H., Tsuda, K., & Inokuchi, A. (2003). Marginalized kernels between labeled graphs. *Proc. Intl. Conf. Machine Learning* (pp. 321–328). San Francisco, CA: Morgan Kaufmann.

Kondor, R., & Borgwardt, K. (2008). The skew spectrum of graphs. *Proc. Intl. Conf. Machine Learning* (pp. 496–503).

Rockmore, D. N. (1997). Some applications of generalized FFTs. *Proceedings of the DIMACS workshop on groups and computation.*

Shervashidze, N., Vishwanathan, S. V. N., Petri, T., Mehlhorn, K., & Borgwardt, K. (2009). Efficient graphlet kernels for large graph comparison. *Proceedings of International Conference on Artificial Intelligence and Statistics.*