MODIFIED KERNEL POLYNOMIAL METHOD FOR ESTIMATING GRAPH SPECTRA

David S. Bindel, Kun Dong

SIAM Workshop on Network Science 2015
May 15-16 · Snowbird

Summary

The *kernel polynomial method* (KPM) is a standard tool in condensed matter physics to estimate the density of states for a quantum system. We use the KPM to instead estimate the eigenvalue densities of the normalized adjacency matrices of "natural" graphs. Because natural graph spectra often include high-multiplicity eigenvalues corresponding to certain motifs in the graph, we introduce a pre-processing phase that counts just these special eigenvalues, leaving the rest of the eigenvalue distribution to be estimated by the standard KPM.

Additional details

Spectral information plays two broad roles in problems of geometry and physics. First, individual (extremal) eigenvalues and their corresponding eigenvectors reflect features such as metastable states in diffusion processes, or bounds on isoperimetric constants, e.g. via Cheeger's inequality. Second, the distribution of eigenvalues or density of states reflects features such as electronic band structure in quantum systems, or the volume and dimension of a manifold, e.g. via Weyl's law. In contrast, in spectral graph theory the extremal eigenvalues and their corresponding eigenvectors play a starring role, while the distribution of eigenvalues has been less studied. The shape of the eigenvalue density is known for some families of random graphs: the normalized adjacency spectra of Erdos-Renyi random graphs have a semicircular density [4], while scale-free and small-world model graphs have more complicated eigenvalue distributions [3, 5, 6, 7, 8]. However, with a few exceptions [1, 2], spectral densities for specific real-world graphs are rarely computed.

The Kernel Polynomial Method Exploring real-world graph spectra by computing all eigenvalues of large matrices is expensive. But an efficient alternative, the *kernel polynomial method* (KPM) [9, 10], is widely used for similar spectral density estimates in condensed matter physics. In the KPM, the spectrum of the normalized adjacency matrix $\bar{A}$ of an undirected graph is written as a generalized function and estimated by truncating and filtering a formal (dual) Chebshev series:

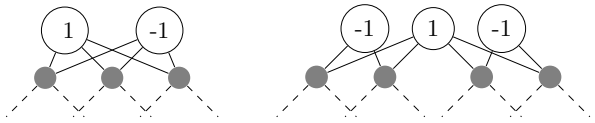

Figure 1: Nonzeros in graph adjacency matrix null vectors with two or three elements.

$$\mu(x) = \sum_{j=1}^{N} \delta(x - \lambda_j) \sim \sum_{n=0}^{\infty} d_n \phi_n(x) \approx \sum_{n=0}^{k} g_{n,d} d_n \phi_n(x),$$

where $\lambda_j \in [-1, 1]$ are the eigenvalues of $\bar{A}$, $\phi_n(x) = (2T_n(x))/\left((1+\delta_{0n})\pi\sqrt{1-x^2}\right)$ is a dual basis to the Chebyshev polynomials T_n, $d_n = \mathrm{tr}(T_n(\bar{A}))$ are the Chebyshev moments of the spectral distribution, and the the constants $g_{n,d}$ are smoothing factors that damp Gibbs oscillations in the expansion. The first two Chebyshev moments, $d_0 = N$ and $d_1 = \mathrm{tr}(\bar{A}) = 0$, can be computed exactly; the remaining moments are computed by the stochastic trace estimator $d_n = E[z^T T_n(\bar{A}) z]$ where z is a vector of independent standard normals; thus, the moments can be estimated by a three-term recurrence involving multiplications by $\bar{A}$ and vector operations.

Filtering eigenvalues through structural motifs Structural motifs in graphs produce particular eigenvalues associated with localized eigenvectors [1]. For example, when two nodes have identical neighbors, the normalized adjacency matrix both has a zero eigenvalue and a corresponding eigenvector with only two nonzeros. These eigenvalues often have high multiplicity, leading to "spikes" in the histogram that are misleadingly blurred by the use of smoothing factors; see, for example, the zero eigenvalue in Figure 2. We therefore propose a two-step process in which we first count some motif-related eigenvalues, then use the KPM to estimate the distribution of the remaining eigenvalues. We focus on null vectors of the normalized adjacency matrix, which are in one-to-one correspondence with null vectors of the standard adjacency matrix.

We use two approaches to remove zero eigenvalues. First,

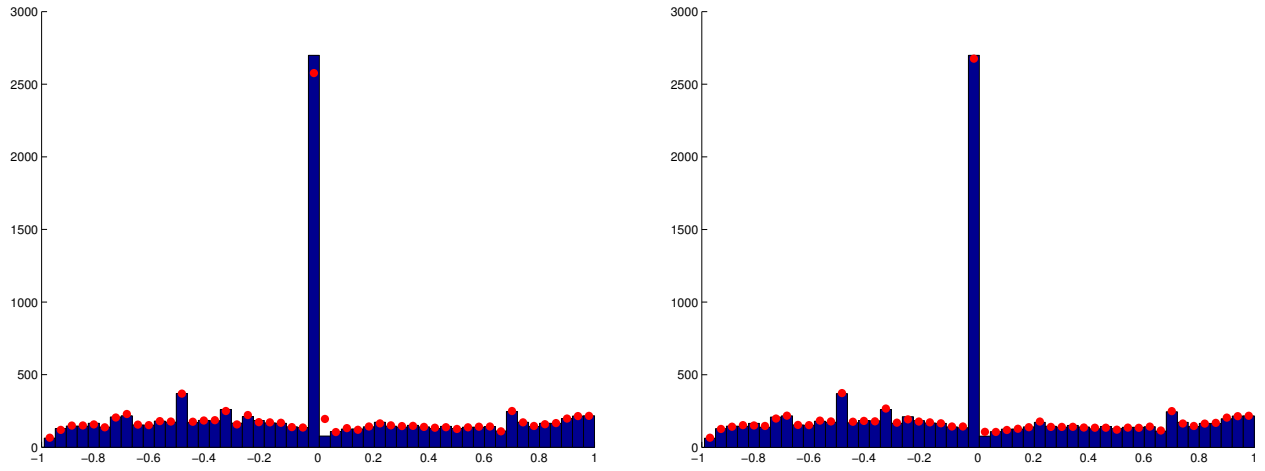

Figure 2: Spectral histogram for a PGP privacy network. Bars indicate exact eigenvalues (courtesy David Gleich); dots indicate a KPM estimate using 10 probe vectors and 1000 moments without preprocessing (left) and with preprocessing (right). Preprocessing reduces the problem size from 10680 to 8673, and visibly improves the artificial smoothing of the "spike" at zero.

in the case of zero eigenvalues associated with k nodes with the same neighbor set (Figure 1 (left)), we modify the (normalized) graph, replacing the k copies of the node with a single copy in which the edge weights are scaled by $\sqrt{k}$. The modified graph has the same normalized adjacency eigenvalues as the original graph, but with $k-1$ fewer zero eigenvalues. In more complicated cases, such as when the neighborhood of one node is a disjoint union of the neighborhoods of two other nodes (Figure 1 (right)), we remove the influence of a known eigenvector v from the kernel polynomial estimate by replacing the random vectors z used in the stochastic trace estimate by the projected vectors $(I - vv^T/\|v\|^2)z$. Subsequent to this projection, no changes to the KPM are needed, though the moments d_0 and d_1 for the modified distribution will need to be modified if they are computed exactly rather than via a stochastic trace estimator. This preprocessing visibly improves the smoothing artifacts near the high-multiplicity zero eigenvalue present in many graphs (Figure 2).

To find the patterns shown in Figure 1, we assign each node a value $v_i = (Aw)_i$ where A is the ordinary adjacency matrix and w is a random vector (cost $O(|E|)$). To find pairs of nodes with identical neighborhoods, we sort the nodes by the v_i and scan for duplicates (cost $O(|V| \log |V|)$). To find null vectors with three nonzeros, for each node i we compute $|v_j - v_k|$ for each j and k in the neighborhood of i, then check whether there is some v_l such that $\pm v_l = |v_j - v_k|$ (cost $O(|V|(d_{\max}^2 + \log |V|))$ where $d_{\max}$ is the maximum degree).

References

[1] A. Banerjee. *The Spectrum of the Graph Laplacian as a Tool for Analyzing Structure and Evolution of Networks*. PhD thesis, Universität Leipzig, 2008.

[2] O. Bhardwaj, Y. Ineichen, C. Bekas, and A. Curioni. Highly scalable linear time estimation of spectrograms – a tool for very large scale data analysis. In *Supercomputing*, 2013.

[3] S. Chauhan, M. Girvan, and E. Ott. Spectral properties of networks with community structure. *Phys. Rev. E*, 80:056114, Nov 2009.

[4] F. Chung, L. Lu, and V. Vu. Spectra of random graphs with given expected degrees. *PNAS*, 100(11):6313–6318, 2003.

[5] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, and A. N. Samukhin. Spectra of complex networks. *Phys. Rev. E (3)*, 68(4):046109, 10, 2003.

[6] I. Farkas. Spectra of "real-world" graphs: Beyond the semicircle law. *Physical Review E*, 64(2), 2001.

[7] I. Farkas, I. Derényi, H. Jeong, Z. Néda, Z. N. Oltvai, E. Ravasz, A. Schubert, A.-L. Barabási, and T. Vicsek. Networks in life: scaling properties and eigenvalue spectra. *Physica A: Statistical Mechanics and its Applications*, 314(1–4):25–34, 2002.

[8] R. Kühn and J. van Mourik. Spectra of modular and small-world matrices. *Journal of Physics A: Mathematical and Theoretical*, 44(16):165205, 2011.

[9] R. N. Silver, H. Roeder, A. F. Voter, and J. D. Kress. Kernel polynomial approximations for densities of states and spectral functions. *Journal of Computational Physics*, 124(1):115–130, 1996.

[10] A. Weisse, G. Wellein, A. Alvermann, and H. Fehske. The kernel polynomial method. *Review of Modern Physics*, 78, 2006.