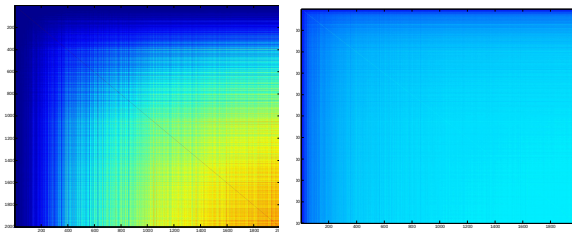


(a) Graph with global scaling and $\alpha = 0.01$.

(b) Graph with local scaling and $\alpha = 0.01$.

Figure 1: Similarity graphs for the Reuters dataset

1 Experiments

In order to measure the performance of spectral networks on real-world data and to explore the effect of the graph estimation procedure, we performed experiments on three datasets from computer vision, text categorization and computational biology.

1.1 Reuters

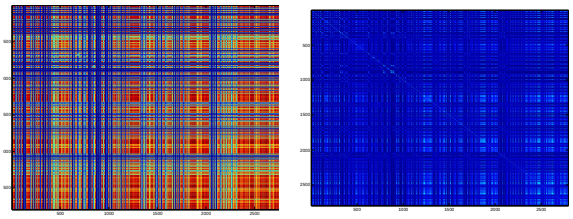
We used the same version of the Reuters dataset as in [**?**], which consists of training and test sets each containing 201,369 documents from 50 mutually exclusive classes. Each document was represented as a log-normalized bag of words for 2000 common non-stop words. The baseline we compared the spectral networks against was the fully-connected network of [**?**] with two hidden layers consisting of 2000 and 1000 hidden units. This network was regularized using dropout.

We based the spectral network architecture on that of a classical convolutional network, namely by interleaving graph convolution and graph pooling layers and ending with a fully connected layer. Performing pooling at the beginning of the network was especially important to reduce the dimensionality in the graph domain and alleviate the expensive Graph Fourier Transform operation. The architectures we explored were the following

Architecture	Parameters
$fc - 2000 - fc - 1000$	$8 \cdot 10^6$
gconv-h-pool-s-fc-1000	2000
gconv-h-pool-s-gconv-h-pool-s-fc-1000	10000

1.2 Merck Molecular Activity Challenge

The Merck Molecular Activity Challenge is a computational biology benchmark where models are required to predict activity levels for various molecules based on the distances in bonds between different atoms. For our experiments we used the DPP4 dataset which has 8193 samples and 2796 descriptors. We chose



(a) Graph with global scaling and $\alpha = 0.01$.

(b) Graph with local scaling and $\alpha = 0.01$.

Figure 2: Similarity graphs for the Merck DPP4 dataset

this dataset because it was one of the more challenging and was of relatively low dimensionality which made the spectral networks tractable. As a baseline architecture, we used the state-of-the-art network of [**?**] which has 4 hidden layers and is regularized using dropout and weight decay.

1.3 Imagenet

In the experiments above our graph construction relied on an approximate estimation from the data. To measure the influence of the graph construction compared to the filter learning in the frequency domain, we performed the same experiments on the ImageNet dataset for which the graph is already known, namely it is the 2-D grid. The spectral network was thus a convolutional network whose weights were defined in the frequency domain. Training was performed exactly as in Figure 1, except that the linear transformation was a Fast Fourier Transform.