Geometric deep learning for business classification

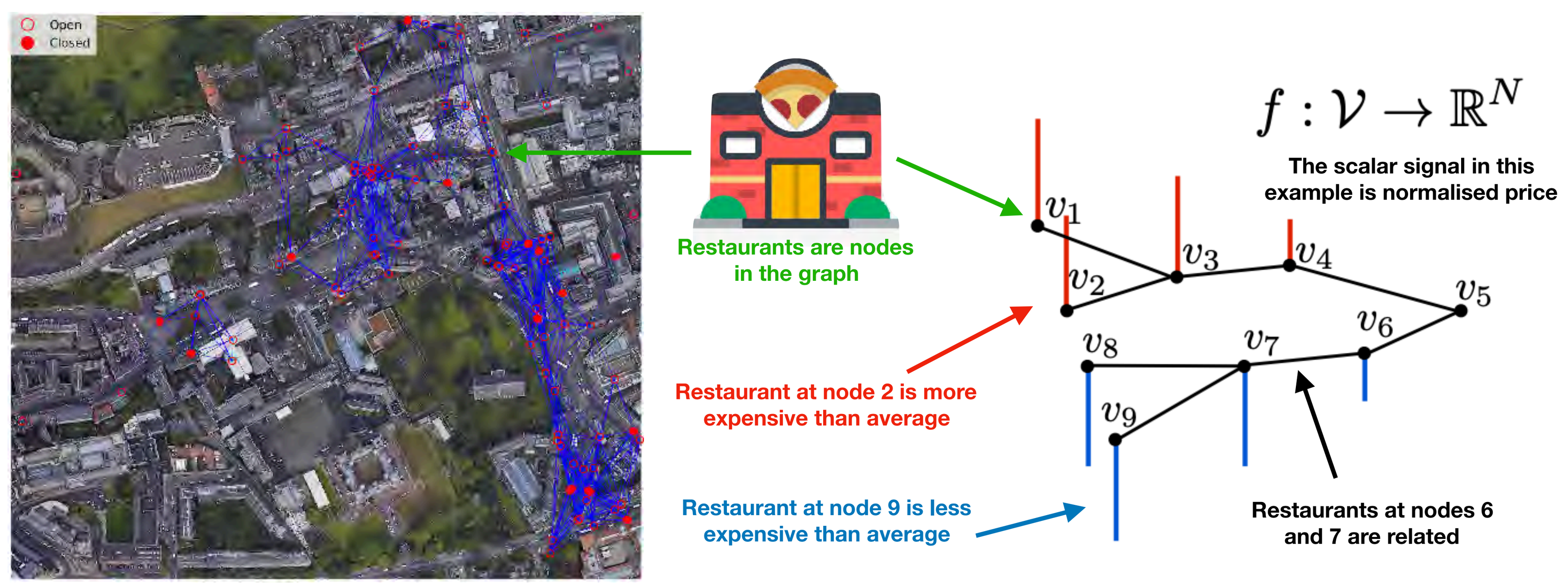
Henry Kenlay, Xiaowen Dong, Stephen Roberts

Introduction

Graph signal processing
- Processing of signals defined on vertex set of weighted and undirected graphs

Geometric deep learning
- Deep learning on data residing on graphs
- Spectral approaches take inspiration from graph signal processing

Business classification
- Given attributes of a business and a graph topology relating businesses in Edinburgh, can we predict if the restaurant has closed down?
- We use a graph convolutional neural network for classification hopefully exploiting pairwise relationships between businesses



Graph signal processing and convolution on graphs

Laplace operator: $-\nabla^2$

Eigenfunction basis: $e^{i\omega x}$

Classical FT:

$$\hat{f}(\omega) = \langle e^{i\omega x}, f \rangle = \int (e^{i\omega x})^* f(x) dx$$

$$f(x) = \frac{1}{2\pi} \int \hat{f}(\omega) e^{i\omega x} d\omega$$

Classical convolution:

$$(f * h)(x) = \int f(\tau) h(x - \tau) d\tau = \frac{1}{2\pi} \int \hat{f}(\omega) \hat{h}(\omega) e^{i\omega x} d\omega$$

Graph Laplacian: $L = D - W$

Eigenvector basis: χ_ℓ

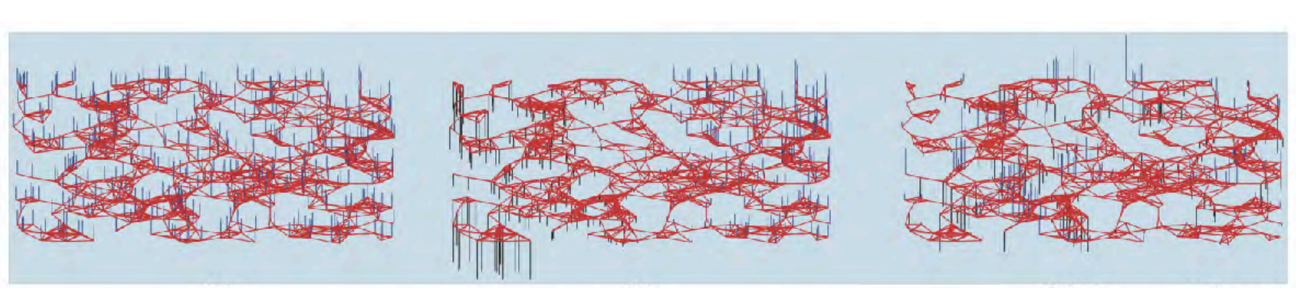
Graph FT:

$$\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell(i)^* f(i)$$

$$f(i) = \sum_{\ell=1}^{N} \hat{f}(\ell) \chi_\ell(i)$$

Graph convolution:

$$(f * h)(i) = \sum_{\ell=1}^{N} \hat{f}(\ell) \hat{h}(\ell) \chi_\ell(i)$$



χ_0 χ_1 χ_{50} [Shuman13]

$L = \chi \Lambda \chi^T$ Low frequency $\chi_0^T L \chi_0 = \lambda_0 = 0$ High frequency $\chi_{50}^T L \chi_{50} = \lambda_{50}$

Graph CNN framework

Filtering in the graph spectral domain
- Eigenvalue decomposition is $\mathcal{O}(n^3)$
- Change of basis is also $\mathcal{O}(n^3)$
- Can approximate filtering using Chebyshev polynomials [1]
- Further simplification by considering polynomials of order one [2]

Propagation rule:

$$H^{(l+1)} = \mathrm{ReLU}\left(\boxed{\tilde{D}}^{-\frac{1}{2}} \boxed{\tilde{A}} \tilde{D}^{-\frac{1}{2}} H^{(l)} \boxed{W}^{(l)}\right)$$

Degree matrix Adjacency matrix Parameter matrix

Layer properties - CNN correspondence
- Signal value on node i in layer K+1 is a function of the value at i as well as that at i's 1-hop neighbours in layer K - localised filters
- Weight sharing across nodes - filters independent of input size
- Sparse multiplications give low computational cost

Classification problem

Problem set up
- Yelp dataset - multimodal data about businesses
- Each restaurant is represented by a node in the graph with edges capturing relationship between businesses
- Each restaurant has an associated feature vector
- Label is if the restaurant is still operating (binary classification)
- Elastic net model used to select subset of features

(a) Nodes (b) As the crow flys (c) Jaccard (d) Kernel

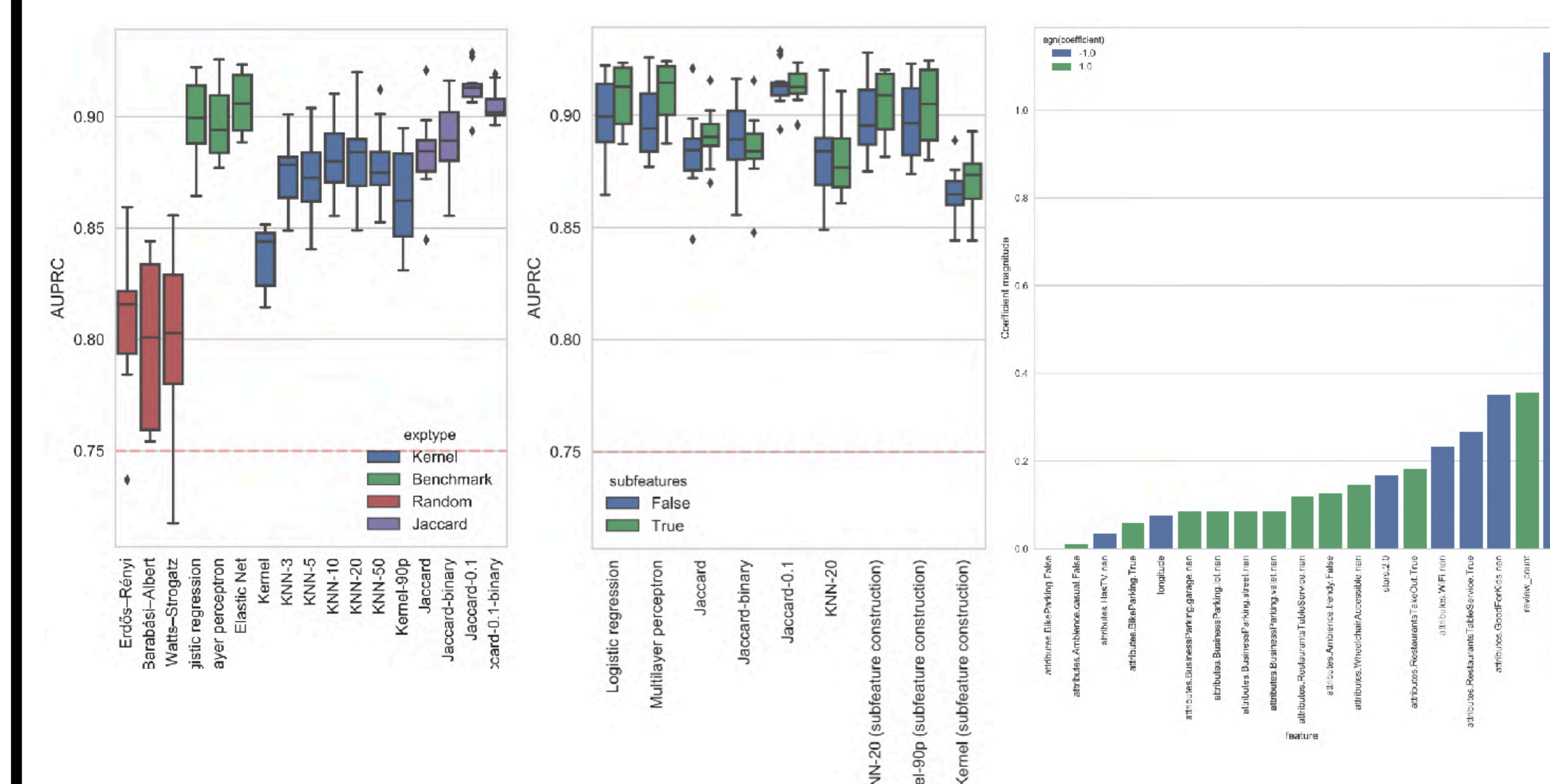
Importance of graph topology
- Edges provide a relational inductive bias
- Using no edges is equivalent to a multilayer perceptron
- We experiment with different ways to generate edge weights:
  - Randomly
  - Kernel-based methods (e.g., Gaussian RBF kernel)
  - Restaurant distance
  - Reviewer data - Jaccard index between two sets of reviewers (one for each restaurant)

Results

Results
- Benchmarks were competitive
- Review data gave slight advantage when using full set of features
- Random graphs were detrimental to performance

Feature selection
- Using elastic net selected features improved performance for many approaches
- The benefit is less significant in graph convolutional neural networks

Thresholding edges
- Thresholding the Jaccard graph improved performance despite disconnecting the graph and isolating 10.6% of the nodes



Discussion

Graph constructions
- Social data
- Statistical methods

Longitudinal prediction
- Can we predict business closure in the future?

Other frameworks
- Explore alternative graph convolutional neural network frameworks

Theoretical approach to the importance of graph topology
- Theoretical bound on the output given a small perturbation of the input graph topology

Contact: kenlay@robots.ox.ac.uk

[1] Defferrard, Michaël, Xavier Bresson, and Pierre Vandergheynst. "Convolutional neural networks on graphs with fast localized spectral filtering." *Advances in Neural Information Processing Systems*. 2016.
[2] Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." *arXiv preprint arXiv:1609.02907* (2016).