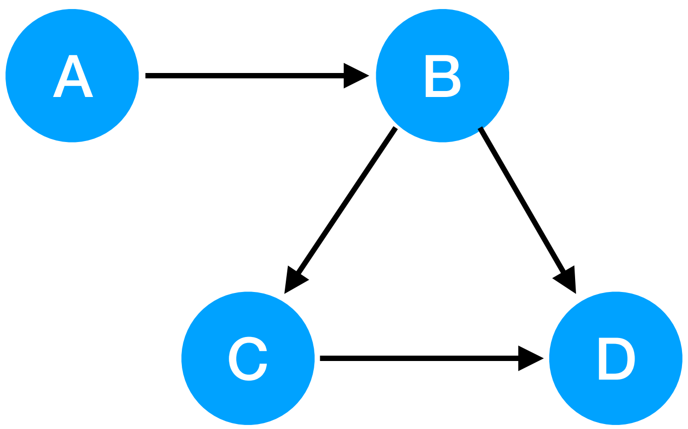
Semi-Supervised Classification with Graph Convolutional Networks

Thomas N. Kipf
University of Amsterdam
T.N.Kipf@uva.nl

Max Welling
University of Amsterdam
Canadian Institute for Advanced Research (CIFAR)
M.Welling@uva.nl

Slides by Nolan Dey

Graph Notation



A =

	A	B	C	D
A	0	1	0	0
B	0	0	1	1
C	0	0	0	1
D	0	0	0	0

D =

	A	B	C	D
A	1	0	0	0
B	0	2	0	0
C	0	0	1	0
D	0	0	0	0

- A = adjacency matrix —> defines graph edges
- D = degree matrix —> defines number of edges per node
- $\hat{A} = A + I$
- $\hat{D} = D + I$

Network Notation

- $N =$ Number of nodes
- $d^l =$ Number of node features at l^{th} layer
- $F^l =$ Hidden representation at l^{th} layer
- $F^l \rightarrow (N \times d^l)$
- $F^0 = X$
- $A \rightarrow (N \times N)$

GCN Layer

- Fully connected layer: $F^l = \sigma(F^{l-1}W^l + b)$

- GCN layer: $F^l = \sigma(\text{transform}(\text{aggregate}(A, F^{l-1}), W^l))$

- aggregate purpose: Take a weighted sum of features from adjacent nodes (analog of convolution)

- transform purpose : Transform aggregated features using a weight matrix

  - $\text{transform}(M, W^l) = MW^l$

  - GCN layer: $F^l = \sigma(\text{aggregate}(A, F^{l-1})W^l)$

Sum Aggregation

- aggregate$(A, F^{l-1}) = AF^{l-1}$

```
A                X              AX
[[0 1 0 0]     [[ 0   0]      [[ 1 -1]
 [0 0 1 1]      [ 1 -1]        [ 5 -5]
 [0 1 0 0]      [ 2 -2]        [ 1 -1]
 [1 0 1 0]]     [ 3 -3]]       [ 2 -2]]
```

- Pros: Aggregated features are the sum of the features of neighbouring nodes

- Cons: A node’s own features do not get propagated

Sum Aggregation 2

- aggregate$(A, F^{l-1}) = \hat{A}F^{l-1}$

- $\hat{A} = A + I$

```
A_hat                 X            A_hat @ X
[[1. 1. 0. 0.]    [[ 0   0]    [[ 1. -1.]
 [0. 1. 1. 1.]     [ 1 -1]      [ 6. -6.]
 [0. 1. 1. 0.]     [ 2 -2]      [ 3. -3.]
 [1. 0. 1. 1.]]    [ 3 -3]]     [ 5. -5.]]
```

- Pros: Aggregated features are the sum of a node's own features and the features of neighbouring nodes

- Cons: Nodes with more connections have features of higher magnitude

Mean Aggregation

- aggregate$(A, F^{l-1}) = \hat{D}^{-1}\hat{A}F^{l-1}$
- $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$

```
D_hat               A_hat               X          inv(D_hat) @ A_hat @ X
[[2. 0. 0. 0.]      [[1. 1. 0. 0.]      [[ 0  0]   [[ 0.5        -0.5       ]
 [0. 3. 0. 0.]       [0. 1. 1. 1.]       [ 1 -1]    [ 2.         -2.        ]
 [0. 0. 2. 0.]       [0. 1. 1. 0.]       [ 2 -2]    [ 1.5        -1.5       ]
 [0. 0. 0. 3.]]      [1. 0. 1. 1.]]      [ 3 -3]]   [ 1.66666667 -1.66666667]]
```

- Pros: Aggregated features are the average of a node's own features and the features of neighbouring nodes
- Cons: Dynamics are "not interesting enough"

Spectral Aggregation

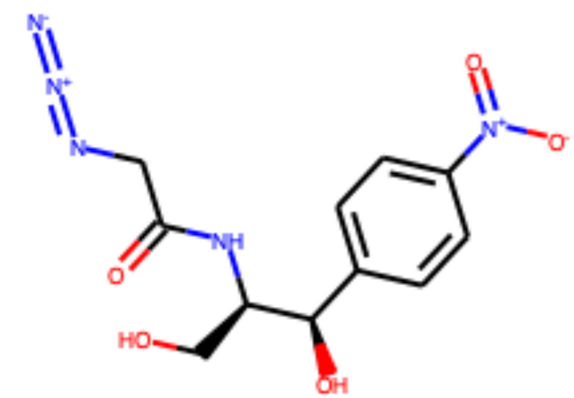
- aggregate$(A, F^{l-1}) = \hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}F^{l-1}$
- First order approximation of a spectral graph convolution

What are GCNs?

- GCN layer: $F^l = \sigma(\text{transform}(\text{aggregate}(A, F^{l-1}), W^l))$
- $\text{transform}(M, W^l) = MW^l$
- $\text{aggregate}(A, F^{l-1}) = \hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}F^{l-1}$
- GCN layer output: $F^l = ReLU(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}F^{l-1}W^l)$

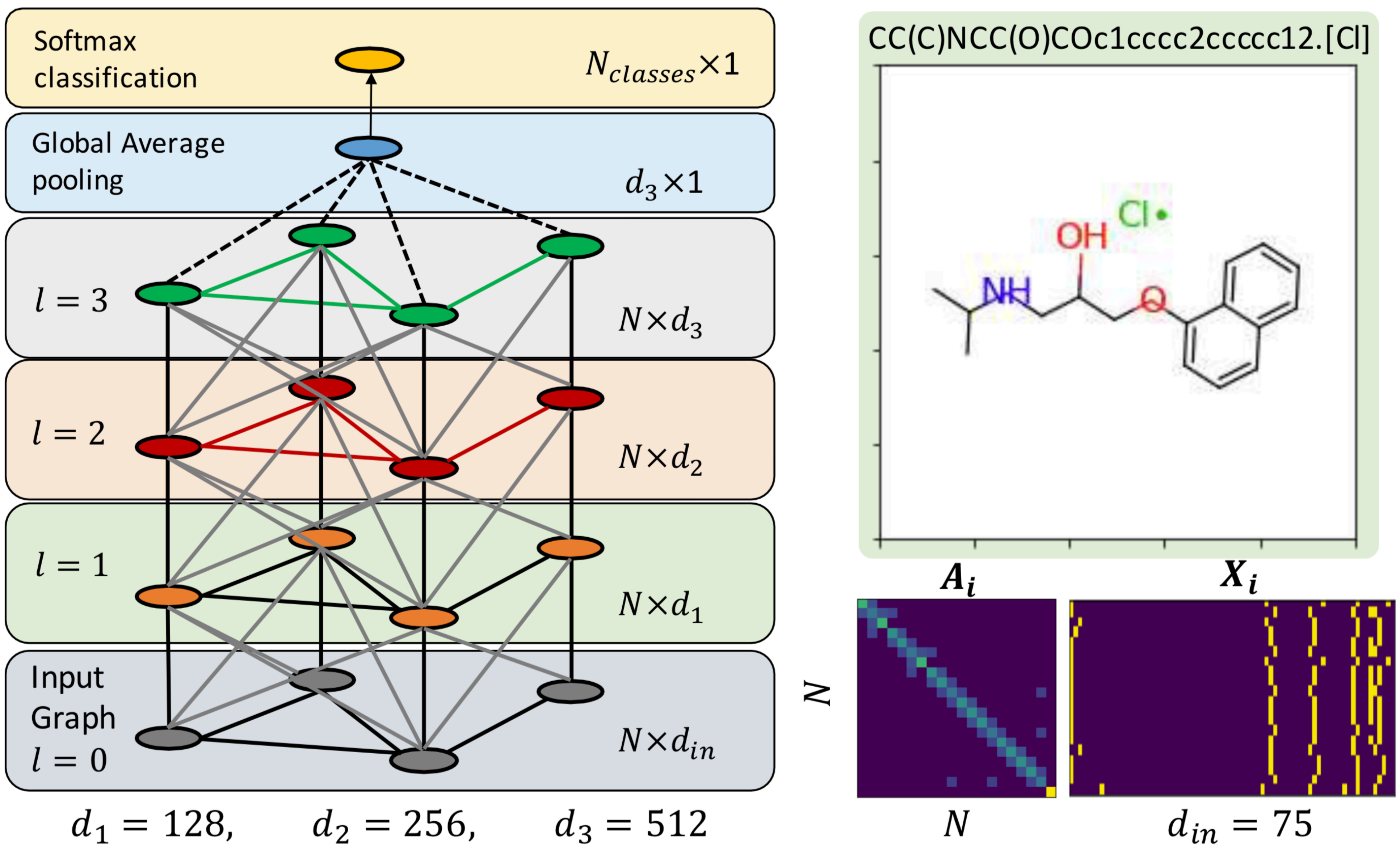
Sample Dataset: Blood Brain Barrier Penetration (BBBP)

num	name	p_np	smiles
1	Propanolol	1	[Cl].CC(C)NCC(O)COc1cccc2ccccc12
2	Terbutylchlorambucil	1	C(=O)(OC(C)(C)C)CCCc1ccc(cc1)N(CCCl)CCCl
3	40730	1	c12c3c(N4CCN(C)CC4)c(F)cc1c(c(C(O)=O)cn2C(C)CO...
4	24	1	C1CCN(CC1)Cc1cccc(c1)OCCCNC(=O)C
5	cloxacillin	1	Cc1onc(c2ccccc2Cl)c1C(=O)N[C@H]3[C@H]4SC(C)(C)...



- Binary Classification
- 2050 molecules
  - 1567 penetrate the blood brain barrier
  - 483 do not penetrate the blood brain barrier
- Applications in drug design

Sample Architecture



Applications

- Image classification
- Recommender systems
- Path planning
- 3D point cloud segmentation and classification
- Molecular classification

Thank you!