Graph Neural Networks e MinCut Pooling

Daniele Grattarola (@riceasphait)

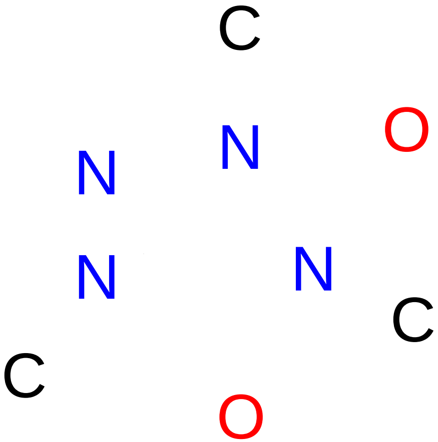
danielegrattarola.github.io

June 29, 2020

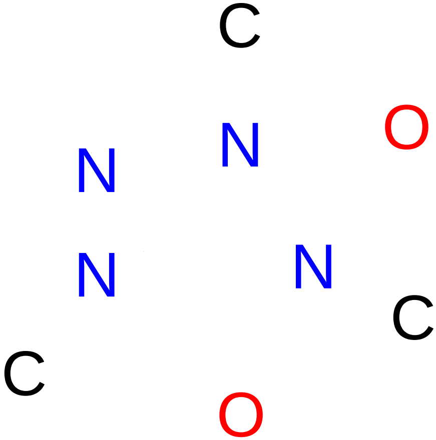
Contenuti

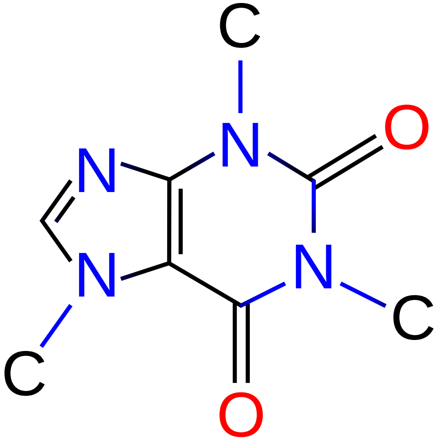
1. Graph Neural Networks.
2. Spectral Clustering with Graph Neural Networks for Graph Pooling (ICML 2020).
3. Esperimenti.

Esempio

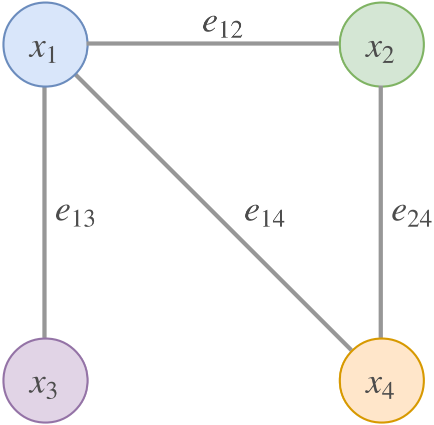


Esempio



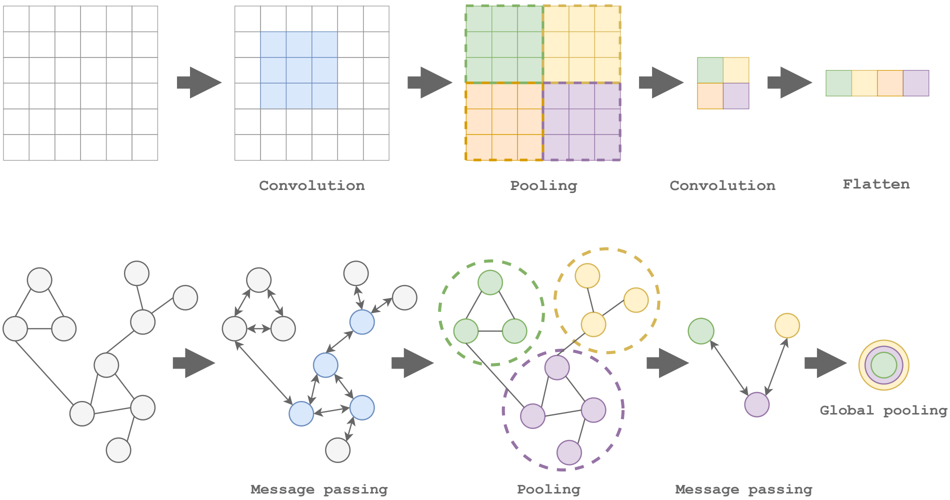


Grafi

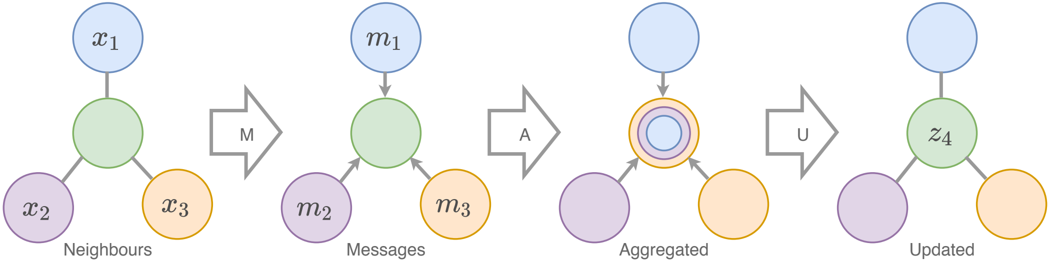


- Nodi ($\mathcal{X}$), archi ($\mathcal{E}$)
- Attributi di nodo ($\mathbf{x}_i \ \forall i \in \mathcal{X}$)
- Attributi di arco ($\mathbf{e}_{i \to j} \ \forall (i, j) \in \mathcal{E}$)

GNN = CNN Generalizzate



Message Passing [1]

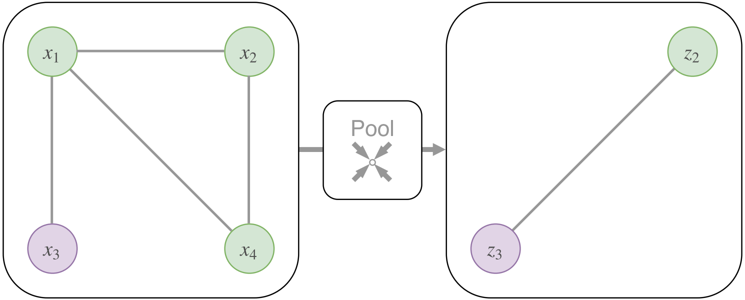


[1] J. Gilmer *et al.*, "Neural message passing for quantum chemistry," 2017.

Strategie per Fare Message Passing

GraphConv *Kipf & Welling*	**ChebConv** *Defferrard et al.*	**GraphSageConv** *Hamilton et al.*	**ARMAConv** *Bianchi et al.*
ECConv *Simonovsky & Komodakis*	**GraphAttention** *Velickovic et al.*	**GraphConvSkip** *Bianchi et al.*	**APPNP** *Klicpera et al.*
GINConv *Xu et al.*	**DiffusionConv** *Li et al.*	**GatedGraphConv** *Li et al.*	**AGNNConv** *Thekumparampil et al.*
TAGConv *Du et al.*	**CrystalConv** *Xie & Grossman*	**EdgeConv** *Wang et al.*	**MessagePassing** *Gilmer et al.*

Pooling nelle Graph Neural Networks



Obiettivo: ridurre il numero di nodi.

Pooling nelle Graph Neural Networks

Model-free

- Indipendenti dal task
- Strategia predefinita
- Teoria dei grafi
- [2], [3]

Model-based

- Dipendenti dal task
- Imparano a ridurre i grafi
- Basati su euristiche
- [4], [5]

[2] I. S. Dhillon *et al.*, "Weighted graph cuts without eigenvectors a multilevel approach," 2007.

[3] F. M. Bianchi *et al.*, "Hierarchical Representation Learning in Graph Neural Networks with Node Decimation Pooling," 2019.

[4] R. Ying *et al.*, "Hierarchical Graph Representation Learning withDifferentiable Pooling," 2018.

[5] H. Gao *et al.*, "Graph U-Nets," 2019.

Spectral Clustering with Graph Neural Networks for Graph Pooling

Filippo Maria Bianchi [*1] **Daniele Grattarola** [*2] **Cesare Alippi** [2 3]

Abstract

Spectral clustering (SC) is a popular clustering technique to find strongly connected communities on a graph. SC can be used in Graph Neural Networks (GNNs) to implement pooling operations that aggregate nodes belonging to the same cluster. However, the eigendecomposition of the Laplacian is expensive and, since clustering results are graph-specific, pooling methods based on SC must perform a new optimization for each new sample. In this paper, we propose a graph clustering approach that addresses these limitations of SC. We formulate a continuous re-

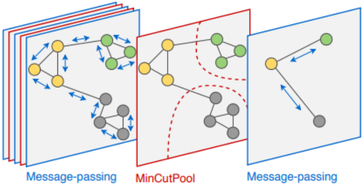
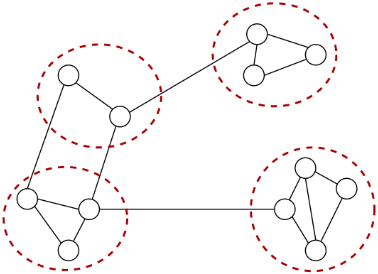
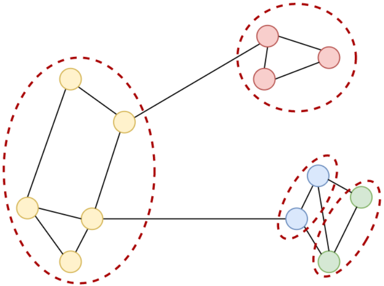


Figure 1. A deep GNN architecture where message-passing is followed by the MinCutPool layer.

Spectral Clustering



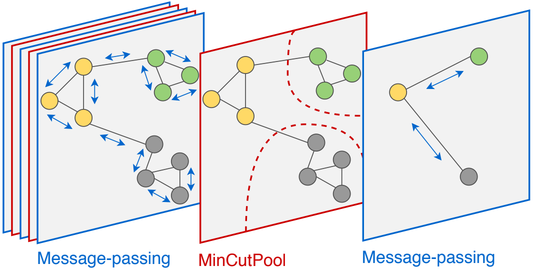
Spectral clustering standard.



Possiamo migliorarlo?

MinCut Pooling

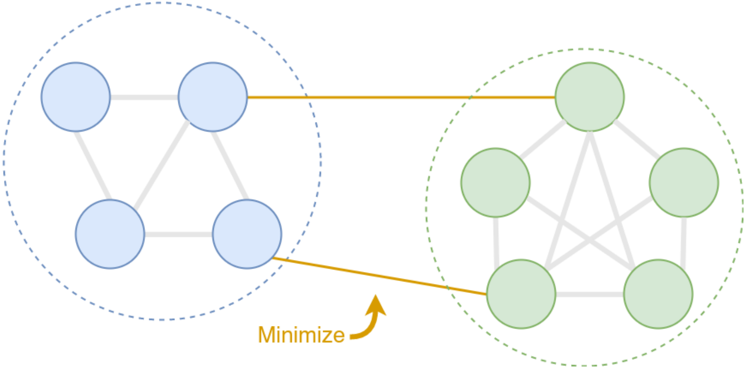
- Problemi di SC: non differenziabile, costoso, usa solo gli solo archi.
- Idea: generare i cluster con una rete neurale.
- Cerchiamo cluster simili a SC: usiamo il **taglio minimo** (MinCut) come loss.
- La NN può trovare l'equilibrio tra la MinCut loss e la task loss.



Spectral clustering $\leftrightarrow$ MinCut

MinCut: trovare K gruppi di nodi per cui

- il volume tra i cluster è minimizzato;
- il volume all'interno dei cluster è massimizzato;



Spectral clustering $\leftrightarrow$ MinCut

Possiamo scrivere il problema MinCut come

$$\text{maximize} \ \ \frac{1}{K}\sum_{k=1}^{K}\frac{\mathbf{C}_k^T\mathbf{A}\mathbf{C}_k}{\mathbf{C}_k^T\mathbf{D}\mathbf{C}_k}, \ \ \text{s.t.} \ \ \underbrace{\mathbf{C}\mathbf{1}_K = \mathbf{1}_N}_{\text{1 nodo} \longleftrightarrow \text{1 cluster}}$$

$\mathbf{C} \in \{0,1\}^{N\times K}$ è una matrice di clustering discreta.

Spectral clustering $\leftrightarrow$ MinCut

Formulazione rilassata

$$\underset{\mathbf{Q}\in\mathbb{R}^{N\times K}}{\arg\max}\ \frac{1}{K}\sum_{k=1}^{K}\mathbf{Q}_k^T\mathbf{A}\mathbf{Q}_k,$$
$$\text{s.t. } \mathbf{Q}=\mathbf{C}(\mathbf{C}^T\mathbf{D}\mathbf{C})^{-\frac{1}{2}},\ \underbrace{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}_K,}_{\text{Ortogonale}}\ \underbrace{\mathbf{C}\mathbf{1}_K=\mathbf{1}_N}_{\substack{\text{Nodi divisi}\\ \text{equamente}}}$$

$\mathbf{C}\in\mathbb{R}^{N\times K}$ è una matrice di clustering continua.
$\mathbf{D}$ è la matrice di grado.

Spectral clustering $\leftrightarrow$ MinCut

Soluzione ottimale:

$$\mathbf{Q}^* = \mathbf{U}_K \mathbf{O}$$

$\mathbf{U}_K$ è la base degli autovettori associati ai primi K autovalori
$\mathbf{O}$ è una trasformazione ortogonale.

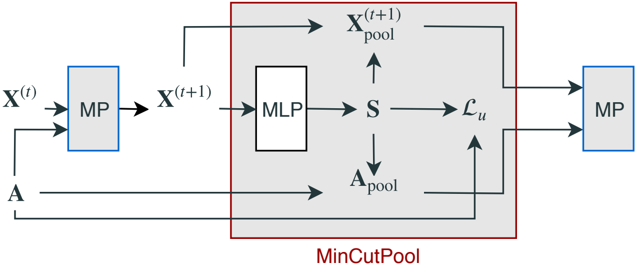
Spectral clustering

K-means sulle righe di $\mathbf{U}_K$ per ottenere C discreta.

MinCut pooling

Impara a generare i cluster:

- $\mathbf{S} = \mathrm{MLP}(\mathbf{X})$
- Softmax $\implies$ $\mathbf{S}\mathbf{1}_K = \mathbf{1}_N$



MinCut pooling

MinCut loss:

$$\mathcal{L}_c = -\frac{Tr(\mathbf{S}^T\mathbf{A}\mathbf{S})}{Tr(\mathbf{S}^T\mathbf{D}\mathbf{S})}$$

Clustering ottimale ($\mathcal{L}_c = -1$):

- $\mathbf{s}_i = [0.25, 0.25, 0.25, 0.25] \leftarrow$ non va bene
- $\mathbf{s}_i = [1.00, 0.00, 0.00, 0.00]$

MinCut pooling

Orthogonality loss (previene i minimi degeneri di $\mathcal{L}_c$):

$$\mathcal{L}_o = \left\| \frac{\mathbf{S}^T\mathbf{S}}{\|\mathbf{S}^T\mathbf{S}\|_F} - \frac{\mathbf{I}_K}{\sqrt{K}} \right\|_F$$

MinCut pooling

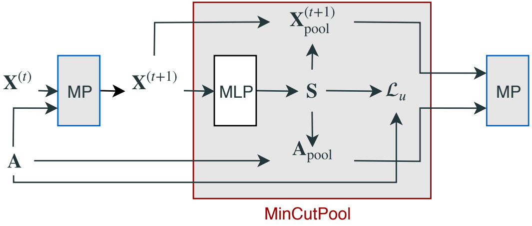
$$\mathcal{L}_u = \mathcal{L}_c + \mathcal{L}_o = \underbrace{-\frac{Tr(\mathbf{S}^T\tilde{\mathbf{A}}\mathbf{S})}{Tr(\mathbf{S}^T\tilde{\mathbf{D}}\mathbf{S})}}_{\mathcal{L}_c} + \underbrace{\left\| \frac{\mathbf{S}^T\mathbf{S}}{\|\mathbf{S}^T\mathbf{S}\|_F} - \frac{\mathbf{I}_K}{\sqrt{K}} \right\|_F}_{\mathcal{L}_o}$$

Loss non supervisionata di MinCutPool.

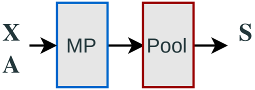
MinCut pooling

Pooling:

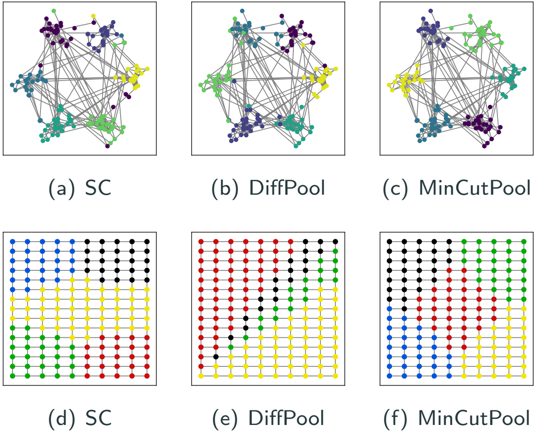
- $\mathbf{A}' = \mathbf{S}^\top \mathbf{A}\mathbf{S}$
- $\mathbf{X}' = \mathbf{S}^\top \mathbf{X}$
- Sommiamo la loss ausiliaria $\mathcal{L}_u$ alla loss supervisionata.



Esperimenti



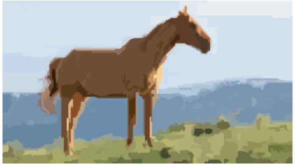
GNN per clustering and segmentazione.



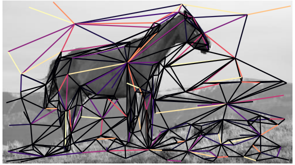
Clustering di grafi semplici.

Segmentazione di Immagini

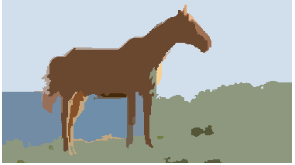
(a) Original



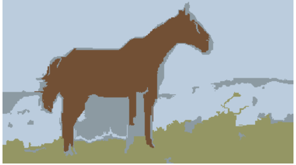
(b) Oversegmentation



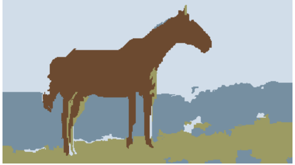
(c) Region Adjacency Graph



(d) SC



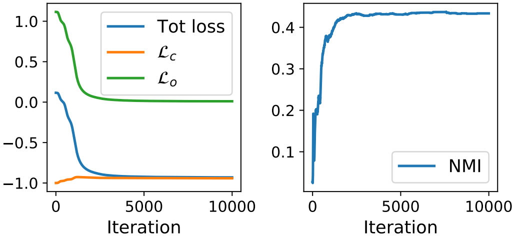
(e) DiffPool ($K = 4$)



(f) MinCutPool ($K = 4$)

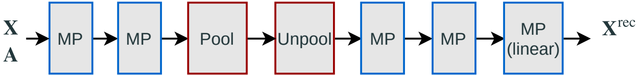
Segmentazione di un'immagine tramite il clustering del suo Region Adjacency Graph.

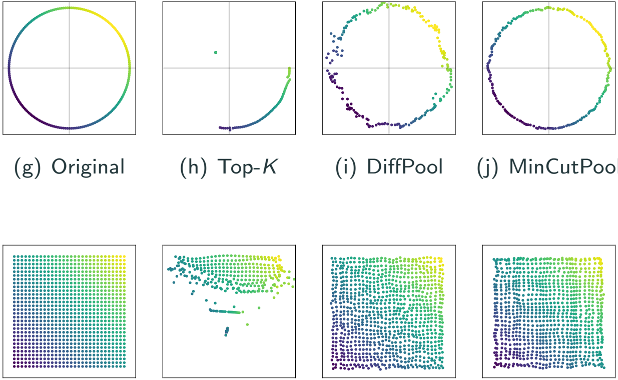
Clustering (citation networks)



Dataset	***K***	**Spectral clustering**		**DiffPool**		**MinCutPool**	
		NMI	CS	NMI	CS	NMI	CS
Cora	7	0.025 ± 0.014	0.126 ± 0.042	0.315 ± 0.005	0.309 ± 0.005	**0.404** ± 0.018	**0.392** ± 0.018
Citeseer	6	0.014 ± 0.003	0.033 ± 0.000	0.139 ± 0.016	0.153 ± 0.020	**0.287** ± 0.047	**0.283** ± 0.046
Pubmed	3	0.182 ± 0.000	**0.261** ± 0.000	0.079 ± 0.001	0.085 ± 0.001	**0.200** ± 0.020	0.197 ± 0.019

Autoencoder





(g) Original (h) Top-K (i) DiffPool (j) MinCutPool

(k) Original (l) Top-K (m) DiffPool (n) MinCutPool

Classificazione di Grafi

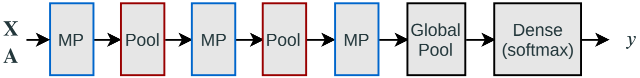


Table 1: Accuracy di classification. I risultati significativamente migliori ($p < 0.05$) sono in grassetto.

Dataset	WL	Dense	No-pool	Graclus	NDP	DiffPool	Top-K	SAGpool	MinCutPool
Bench-easy	92.6	29.3±0.3	98.5±0.3	97.5±0.5	97.9±0.5	98.6±0.4	82.4±8.9	84.2±2.3	**99.0**±0.0
Bench-hard	60.0	29.4±0.3	67.6±2.8	69.0±1.5	**72.6**±0.9	69.9±1.9	42.7±15.2	37.7±14.5	**73.8**±1.9
Mutagenicity	**81.7**±1.1	68.4±0.3	78.0±1.3	74.4±1.8	77.8±2.3	77.6±2.7	71.9±3.7	72.4±2.4	79.9±2.1
Proteins	71.2±2.6	68.7±3.3	72.6±4.8	68.6±4.6	73.3±3.7	72.7±3.8	69.6±3.5	70.5±2.6	**76.5**±2.6
DD	78.6±2.7	70.6±5.2	76.8±1.5	70.5±4.8	72.0±3.1	**79.3**±2.4	69.4±7.8	71.5±4.5	**80.8**±2.3
COLLAB	74.8±1.3	79.3±1.6	**82.1**±1.8	77.1±2.1	79.1±1.5	81.8±1.4	79.3±1.8	79.2±2.0	**83.4**±1.7
Reddit-Binary	68.2±1.7	48.5±2.6	80.3±2.6	79.2±0.4	84.3±2.4	86.8±2.1	74.7±4.5	73.9±5.1	**91.4**±1.5

Conclusioni

- Nuovo metodo di pooling per le GNN.
- Impara a fare il pooling, ma ha una base teorica.
- Risolve problemi di spectral clustering.
- Funziona molto bene in pratica.

Speaker

Daniele Grattarola (@riceasphait)

Contatti

daniele.grattarola@usi.ch

filippombianchi@gmail.com

Codice

Disponibile su Spektral e Pytorch Geometric.

https://github.com/FilippoMB/Spectral-Clustering-with-Graph-Neural-Networks-for-Graph-Pooling

[1] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," *arXiv preprint arXiv:1704.01212*, 2017.

[2] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 11, pp. 1944–1957, 2007.

[3] F. M. Bianchi, D. Grattarola, L. Livi, and C. Alippi, "Hierarchical representation learning in graph neural networks with node decimation pooling," *arXiv preprint arXiv:1910.11436*, 2019.

[4] R. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec, "Hierarchical graph representation learning withdifferentiable pooling," *arXiv preprint arXiv:1806.08804*, 2018.

[5] H. Gao and S. Ji, "Graph u-nets," *CoRR*, vol. abs/1905.05178, 2019. arXiv: `1905.05178`. [Online]. Available: `http://arxiv.org/abs/1905.05178`.