

4 - Νευρωνικά δίκτυα και γραφήματα - εισαγωγή

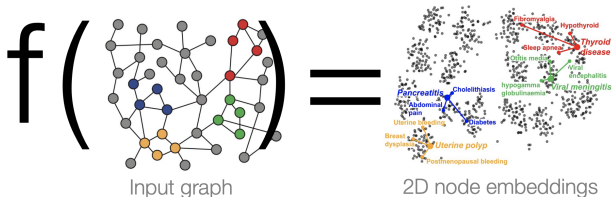
Δημήτριος Κοσμόπουλος

Πανεπιστήμιο Πατρών
Τμήμα Μηχανικών ΗΥ κ Πληροφορικής

14 Δεκεμβρίου 2023

Ενσωματώσεις κόμβων

Απεικόνιση κόμβων σε d -διάστατες ενσωματώσεις (embeddings) ώστε όμοιοι κόμβοι να βρίσκονται κοντά στο χώρο ενσωμάτωσης.

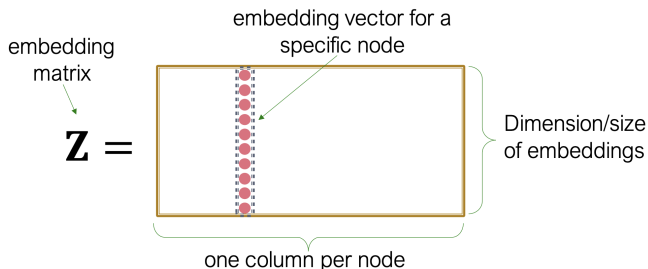


Πρέπει να ορίσουμε τι είναι ομοιότητα κόμβων καθώς και τη συνάρτηση κωδικοποίησης ώστε $similarity(u, v) \approx z^T z$

Ρηχή κωδικοποίηση

Απλούστερη προσέγγιση:

- ▶ Lookup table: $ENC(u) = z_u = Z \cdot u$
- ▶ $Z \in \mathbb{R}^{|V| \times d}$ πίνακας με στήλες τις ενσωματώσεις (αυτό μαθαίνουμε)
- ▶ $u \in \mathbb{K}^{|V|}$ διάνυσμα τύπου indicator με μηδενικά παντού εκτός από τη θέση που αντιστοιχεί στον κόμβο



Πρόβλημα για γραφήματα με πολλούς κόμβους

Προβλήματα ρηχής κωδικοποίησης

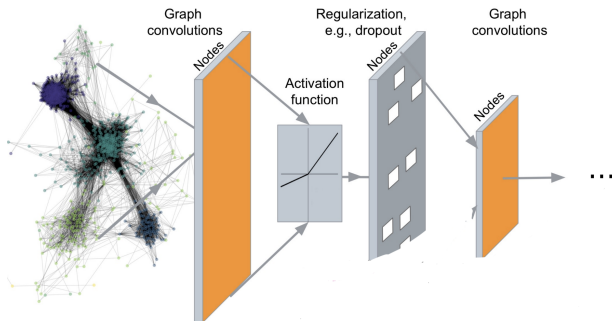
- ▶ $O(|V|d)$ παράμετροι: δεν μοιράζονται οι παράμετροι μεταξύ κόμβων και κάθε κόμβος έχει μια μοναδική ενσωμάτωση
- ▶ δεν μπορούμε να έχουμε την ενσωμάτωση κόμβου που δεν έχει εκπαιδευτεί
- ▶ δεν μπορούμε να χρησιμοποιήσουμε χαρακτηριστικά των κόμβων

Βαθιοί κωδικοποιητές

Οι κωδικοποιητές που θα εξετάσουμε

- ▶ Βασίζονται σε νευρωνικά δίκτυα γραφημάτων (Graph Neural Networks - GNNs)
- ▶ Πρόκειται για πολλαπλά επίπεδα μη γραμμικών μετασχηματισμών που βασίζονται στη δομή του γραφήματος
- ▶ Διάφορες συναρτήσεις ομοιότητας κόμβων μπορούν να χρησιμοποιηθούν

Βαθιοί κωδικοποιητές σε γραφήματα



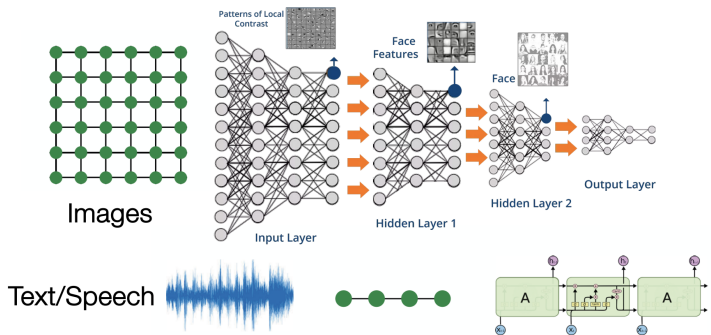
Πιθανές έξοδοι:

- ▶ οι ενσωματώσεις των κόμβων
- ▶ ενσωματώσεις υπο-γραφημάτων
- ▶ ενσωματώσεις γραφημάτων

Πιθανές εργασίες

- ▶ Ταξινόμηση κόμβου: πρόβλεψε τον τύπο (ετικέτα) κόμβου
- ▶ Πρόβλεψη ακμής
- ▶ Ανίχνευση κοινότητας (ομαδοποίηση)
- ▶ Ομοιότητα: πόσο μοιάζουν δυο (υπο)γραφήματα

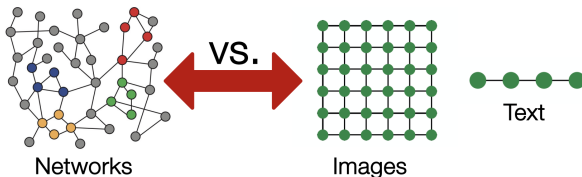
Τυπικές αρχιτεκτονικές μηχανικής μάθησης



- ▶ Ορθογώνιο πλέγμα για εικόνες
- ▶ Σειριακό πλέγμα για ακολουθίες (ήχος, κείμενο)

Δυσκολίες γραφημάτων

- ▶ Δεν υπάρχει καθορισμένη αρίθμηση των κόμβων
- ▶ Αυθαίρετη τοπολογία
- ▶ Πολλές φορές έχουν δυναμικά χαρακτηριστικά



Γενικό πλαίσιο βαθιάς μάθησης

- ▶ Συνάρτηση απώλειας (*loss*) την οποία προσπαθούμε να ελαχιστοποιήσουμε
 $\mathcal{L}(y, f_{\Theta}(x))$
- ▶ Η f μπορεί να είναι ένα απλό επίπεδο *NN*, ένα *MLP*, *CNN* κλπ
- ▶ Δειγμάτισε ένα *batch* των εισόδων x
- ▶ Forward propagation: υπολόγισε το \mathcal{L} δεδομένου του x
- ▶ Back propagation: υπολόγισε το *gradient* $\nabla_{\Theta}\mathcal{L}$ με τον κανόνα της αλυσίδας
- ▶ Χρησιμοποίησε *stochastic gradient descent* για να βελτιστοποιήσεις το \mathcal{L} για τις τιμές του Θ με πολλές επαναλήψεις

Τι θα εξετάσουμε στα γραφήματα

Περιγραφή γειτονιάς

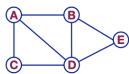
- ▶ Στρατηγικές συνένωσης (aggregation)
- ▶ Ορισμός γραφημάτων για υπολογισμούς

Συσώρευση πολλαπλών επιπέδων

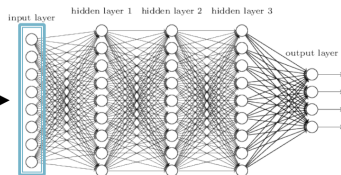
- ▶ περιγραφή μοντέλου, παραμέτρων, εκπαίδευσης
- ▶ πώς ταιριάζουμε το μοντέλο
- ▶ παράδειγμα επιβλεπόμενης και μη επιβλεπόμενης μάθησης

Αφελής προσέγγιση

- ▶ Στη γενική περίπτωση ο κάθε κόμβος έχει χαρακτηριστικά (π.χ, το προφίλ ενός χρήστη κοινωνικού δικτύου)
- ▶ Συνένωσε πίνακα γειτνίασης και χαρακτηριστικά
- ▶ Δώσε το αποτέλεσμα στην είσοδο ΤΝΔ
- ▶ Προβλήματα της προσέγγισης:
 - ▶ $O(|V|)$ παράμετροι
 - ▶ μη εφαρμόσιμο σε γραφήματα διαφορετικού μεγέθους
 - ▶ διαφορετική αρίθμηση των κόμβων δίνει άλλο αποτέλεσμα



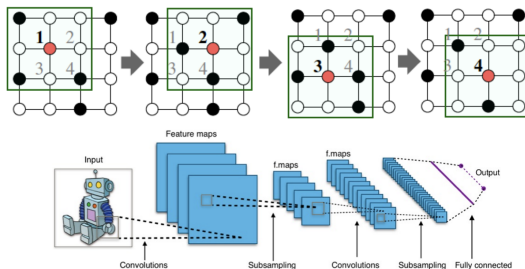
	A	B	C	D	E	Feat
A	0	1	1	1	0	1 0
B	1	0	0	1	1	0 0
C	1	0	0	1	0	0 1
D	1	1	1	0	1	1 1
E	0	1	0	1	0	1 0



?

Συνελικτικά δίκτυα

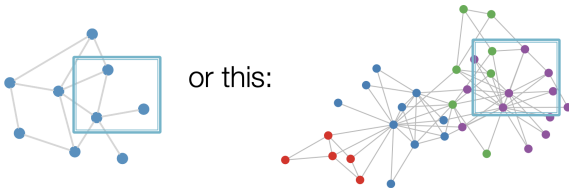
Συνελικτικό δίκτυο σε μια εικόνα:



Στόχος είναι να γενικεύσουμε τις συνελίξεις πέρα από τις βασικές δομές 1- Δ , 2- Δ πλέγματος.

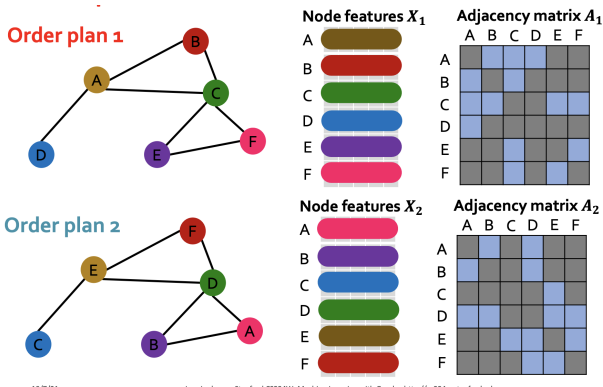
Συνελικτικά δίκτυα

Τα γραφήματα έχουν δομές όπως:



- ▶ δεν υπάρχει η έννοια της τοπικότητας ώστε να οριστεί κυλιόμενο παράθυρο στο γράφημα
- ▶ το γράφημα είναι αναλλοίωτο στις αντιμεταθέσεις (αρίθμηση) κόμβων
- ▶ υπάρχουν $V!$ διαφορετικές μεταθέσεις

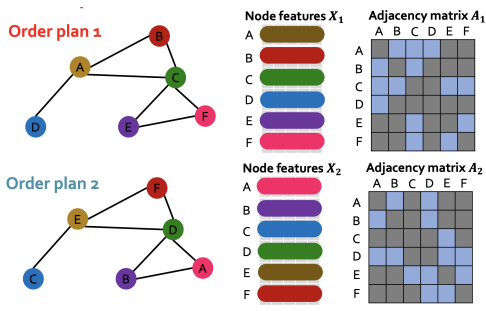
Αναπαράσταση αναλλοίωτη στη μετάθεση



Η αναπαράσταση του γραφήματος θα πρέπει να είναι ίδια (αναλλοίωτη) για και για τις δύο μεταθέσεις.

Αν η f απεικονίζει ένα γράφημα στο χώρο \mathbb{R}^d με A_i τους πίνακες γειτνίασης, X_i τα διανύσματα χαρακτηριστικών, θα πρέπει: $f(A_1, X_1) = f(A_2, X_2)$.

Αναπαράσταση αναλλοίωτη στη μετάθεση



Ορισμός:

Μια συνάρτηση γραφημάτων $f : \mathbb{R}^{|V| \times m} \times \mathbb{R}^{|V| \times |V|} \rightarrow \mathbb{R}^{|V| \times d}$ είναι αναλλοίωτη στη μετάθεση όταν $f(A, X) = f(PAP^T, PX)$ για κάθε μετάθεση P .

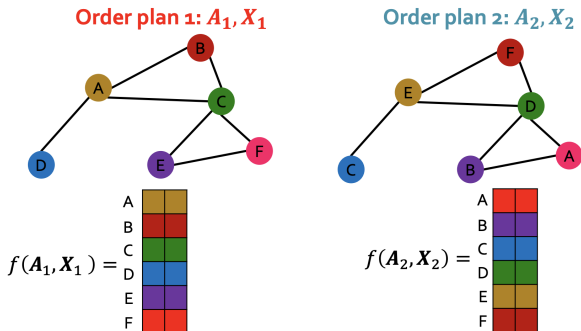
P : αναδιάταξη σειράς των κόμβων π.χ. $(A, B, C) \rightarrow (C, B, A)$

m : η διάσταση των χαρακτηριστικών των κόμβων

Ισοδύναμη αναπαράσταση

Μαθαίνουμε μια συνάρτηση f που απεικονίζει κόμβους του γραφήματος σε ένα πίνακα ενσωματώσεων

$$f : \mathcal{G} \rightarrow \mathbb{R}^{|V| \times d}$$



Στους ίδιους κόμβους του γραφήματος η ενσωμάτωση στην έξοδο είναι η ίδια.

Ισοδύναμη αναπαράσταση

Μια συνάρτηση γραφημάτων

$$f : \mathcal{G}(A, X) \rightarrow \mathbb{R}^{|V| \times d}$$

είναι ισοδύναμη στη μετάθεση όταν η έξοδος (ενσωμάτωση) σε κάθε κόμβο στην ίδια θέση παραμένει η ίδια.

Ορισμός:

Μια συνάρτηση γραφημάτων $f : \mathbb{R}^{|V| \times m} \times \mathbb{R}^{|V| \times |V|} \rightarrow \mathbb{R}^{|V| \times d}$ είναι ισοδύναμη στη μετάθεση όταν $Pf(A, X) = f(PAP^T, PX)$ για κάθε μετάθεση P .

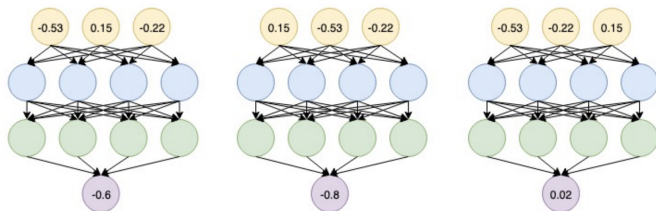
P : αναδιάταξη σειράς των κόμβων π.χ. $(A, B, C) \rightarrow (C, B, A)$

m : η διάσταση των χαρακτηριστικών των κόμβων

Νευρωνικά δίκτυα σε γραφήματα

Η αρχιτεκτονική των Νευρωνικών δικτύων σε γραφήματα *GNN* βασίζεται στις ιδιότητες του αναλλοίωτου στη μετάθεση και της ισοδυναμίας.

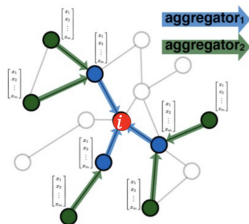
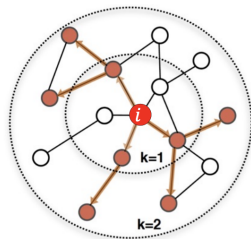
Ισχύουν οι ιδιότητες αυτές για τα κλασσικά ΤΝΔ;



Όχι, αλλάζοντας τη σειρά των εισόδων αλλάζει η έξοδος.

Συνελικτικά νευρωνικά δίκτυα σε γραφήματα - GCN

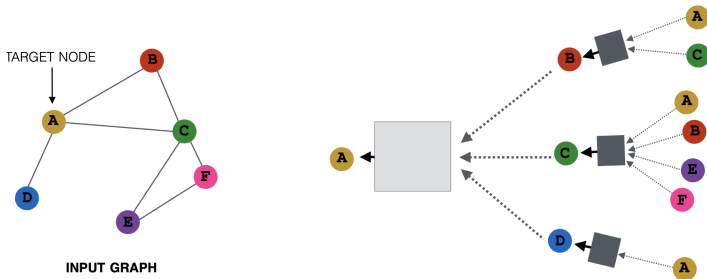
- ▶ Ιδέα: η γειτονιά ενός κόμβου ορίζει ένα γράφημα στο οποίο θα μπορούμε να κάνουμε τους υπολογισμούς (computational graph)
- ▶ Ο αριθμός των βημάτων καθορίζει το βάθος της γειτονιάς
- ▶ Θα δούμε πώς διαδίδονται τα μηνύματα σε αυτή τη γειτονιά ώστε να υπολογίζουμε τα χαρακτηριστικά των κόμβων.



Υπολόγισε γειτονιά – Διάδοσε/μετασχημάτισε μηνύματα

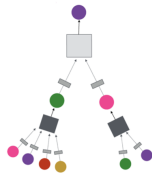
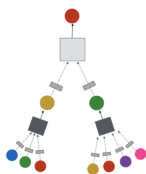
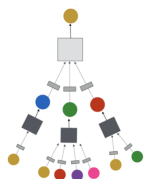
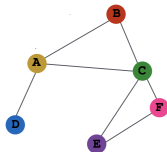
Συνελικτικά νευρωνικά δίκτυα σε γραφήματα - GCN

- ▶ Βασική ιδέα: υπολόγισε ενσωματώσεις κόμβων βασισμένες στις τοπικές γειτονιές
- ▶ Δημιούργησε επίπεδα ανάμεσα στα οποία υπάρχουν νευρωνικά δίκτυα.
- ▶ Τα νευρωνικά δίκτυα συγκεντρώνουν και μετασχηματίζουν τα μηνύματα από τους γειτονικούς κόμβους.



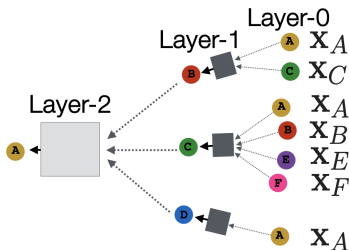
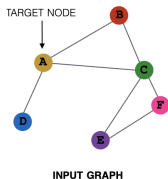
Συνελικτικά νευρωνικά δίκτυα σε γραφήματα - GCN

Για κάθε κόμβο έχουμε ένα γράφημα γειτονιάς



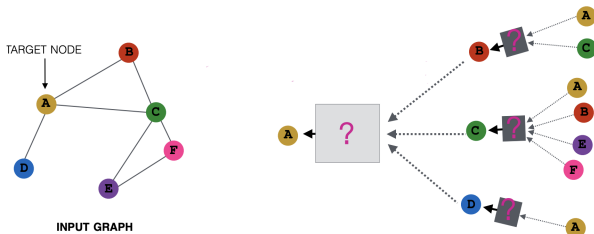
Συνελικτικά νευρωνικά δίκτυα σε γραφήματα - GCN

- ▶ Το μοντέλο μπορεί να έχει οποιοδήποτε βάθος.
- ▶ Το επίπεδο 0 του κόμβου v έχει είσοδο x_v .
- ▶ Το επίπεδο k ενσωματώνει τα μηνύματα από κόμβους που βρίσκονται k βήματα μακριά.



Συνελικτικά νευρωνικά δίκτυα σε γραφήματα - GCN

Οι βασικές διαφοροποιήσεις στους τύπους νευρωνικών δικτύων αφορούν τον τρόπο με τον οποίο γίνεται ο συγκερασμός των μηνυμάτων στα διάφορα επίπεδα



Τι υπάρχει στα κουτιά ;

Βασική προσέγγιση:

1. βγάλε μέση τιμή των μηνυμάτων από τους γείτονες
2. το αποτέλεσμα πέρνα το μέσα από ένα νευρωνικό δίκτυο

Τα νευρωνικά δίκτυα στα διάφορα επίπεδα

Οι ενσωματώσεις στο επίπεδο μηδέν είναι ίσες με τα χαρακτηριστικά των κόμβων: $\mathbf{h}_v^0 = \mathbf{x}_v$

Στο επίπεδο $k + 1$ προκύπτουν από τις εισόδους του επιπέδου k :

$$\mathbf{h}_v^{k+1} = \sigma(\mathbf{W}_k \sum_{u \in \mathcal{N}(v)} \frac{\mathbf{h}_u^k}{|\mathcal{N}(v)|} + \mathbf{B}_k \mathbf{h}_v^k)$$

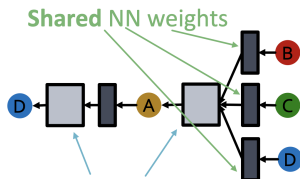
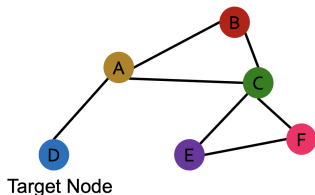
όπου $k \in \{0, \dots, K - 1\}$, και:

- ▶ σ : η μη γραμμικότητα π.χ. *ReLU*
- ▶ \mathbf{W}_k ο πίνακας βαρών στο επίπεδο k που συσχετίζεται με τους γείτονες
- ▶ \mathbf{B}_k ο πίνακας βαρών στο επίπεδο k που συσχετίζεται με τον τρέχοντα κόμβο
- ▶ το άθροισμα είναι η μέση τιμή των γειτόνων (αναλλοίωτο ως προς τη μετάθεση)

Τελικά έπειτα από k επίπεδα θα έχουμε $\mathbf{z}_v = \mathbf{h}_v^k$

Αναλλοίωτο και ισοδύναμο ως προς μετάθεση

Δοθέντος ενός κόμβου το *GCN* που υπολογίζει την ενσωμάτωση είναι αναλλοίωτο ως προς τη μετάθεση

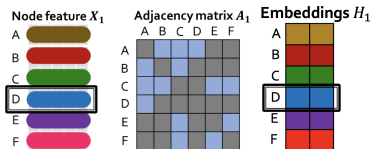
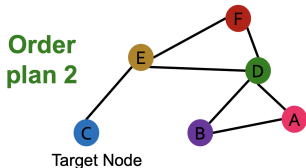
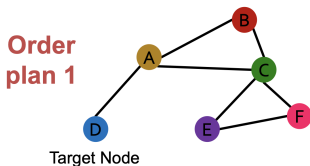


Average of neighbor's previous layer embeddings - **Permutation invariant**

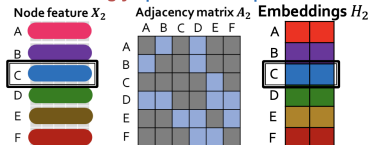
Αναλλοίωτο και ισοδύναμο ως προς μετάθεση

Δοθέντων όλων των κόμβων σε ένα γράφημα ο υπολογισμός για το GCN είναι ισοδύναμος ως προς τη μετάθεση.

Οι έξοδοι ευθυγραμμίζονται με τους κόμβους.



Permute the input, the output also permutes accordingly - permutation equivariant



Το μοντέλο

Οι ενσωματώσεις στο επίπεδο 0: $\mathbf{h}_v^0 = \mathbf{x}_v$

Στο επίπεδο $k + 1$:

$$\mathbf{h}_v^{k+1} = \sigma(\mathbf{W}_k \sum_{u \in \mathcal{N}(v)} \frac{\mathbf{h}_u^k}{|\mathcal{N}(v)|} + \mathbf{B}_k \mathbf{h}_v^k)$$

Έπειτα από k επίπεδα: $\mathbf{z}_v = \mathbf{h}_v^k$

- ▶ Προσπαθούμε να μάθουμε τους πίνακες βαρών $\mathbf{W}_k, \mathbf{B}_k$.
- ▶ Εισάγουμε τα διανύσματα των ενσωματώσεων σε μια συνάρτηση απώλειας την οποία προσπαθούμε να ελαχιστοποιήσουμε με *SGD* και να μάθουμε τα ζητούμενα βάρη.

Αναπαράσταση με μητρώα

Η συνένωση των διανυσμάτων αναπαράστασης μπορεί να γίνει με χρήση πινάκων

- ▶ Έστω $H^k = [h_1^k, \dots, h_{|V|}^k]^T$
- ▶ Τότε το άθροισμα των γειτόνων του v μπορεί να γραφτεί
$$\sum_{u \in N(v)} h_u^k = A_{v,:} H^k$$
- ▶ Έστω ότι D διαγώνιος πίνακας έτσι ώστε
$$D_{v,v} = \text{deg}(v) = |N(v)|$$
- ▶ Ο αντίστροφος D^{-1} είναι επίσης διαγώνιος και μάλιστα
$$D_{v,v}^{-1} = 1/|N(v)|$$
- ▶ Άρα η μέση τιμή της γειτονιάς του v μπορεί να γραφτεί ως γινόμενο πινάκων:
$$\sum_{u \in N(v)} \frac{h_u^k}{|N(v)|} = D^{-1} A H^k$$

Αναπαράσταση με μητρώα

Η αναδρομική συνάρτηση που υπολογίζει το H μπορεί να γραφεί συναρτήσει πινάκων:

$$H^{k+1} = \sigma(W_k D^{-1} A H^k + B_k H^k)$$

- ▶ Ο πρώτος όρος αφορά τη συνεισφορά των γειτόνων, ο δεύτερος όρος αφορά τη συνεισφορά του ίδιου κόμβου.
- ▶ Η παράσταση $D^{-1}A$ δίνει αραιό πίνακα, άρα μπορούν να χρησιμοποιηθούν βιβλιοθήκες που εκμεταλλεύονται την αραιότητα.
- ▶ Δεν μπορούν όλα τα *GNN* να δώσουν αναπαράσταση γινομένου πινάκων, ειδικότερα εάν η συνάρτηση συνένωσης της γειτονιάς γίνει πιο πολύπλοκη από τη μέση τιμή.

Εκπαίδευση

- ▶ Η ενσωμάτωση z_v κάθε κόμβου v εξαρτάται από το γράφημα
- ▶ Επιβλεπόμενη μάθηση: Θεωρούμε συνάρτηση απώλειας \mathcal{L} την οποία προσπαθούμε να ελαχιστοποιήσουμε

$$\min_{\Theta} \mathcal{L}(\mathbf{y}, f_{\Theta}(z_v))$$

- ▶ \mathbf{y} : η επιθυμητή τιμή του κόμβου
- ▶ \mathcal{L} μπορεί να είναι η $L2$ αν το \mathbf{y} είναι πραγματικός (*regression*) ή *cross – entropy* αν είναι κατηγορικό (*classification*).
- ▶ Μη επιβλεπόμενη μάθηση
 - ▶ δεν είναι διαθέσιμη η επιθυμητή τιμή του κόμβου κατά την εκπαίδευση
 - ▶ χρησιμοποιούμε τη δομή του γραφήματος

Μη επιβλεπόμενη Εκπαίδευση

- ▶ Ιδέα: όμοιοι κόμβοι έχουν όμοιες ενσωματώσεις:

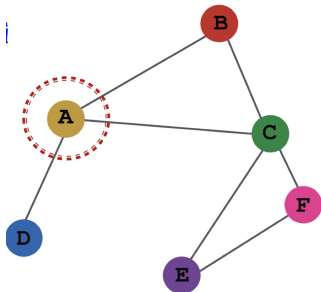
$$\min_{\Theta} \mathcal{L} = \sum_{z_u, z_v} \text{CE}(y_{u,v}, \text{DEC}(z_u, z_v))$$

- ▶ CE είναι η συνάρτηση απώλειας Cross Entropy:
$$\text{CE}(\mathbf{y}, f(\mathbf{x})) = - \sum_{i=1}^C (y_i \log f_{\Theta}(x)_i)$$
- ▶ y_i και $f_{\Theta}(x)_i$ είναι οι πραγματικές και οι προβλεπόμενες τιμές της i -κλάσσεως
- ▶ όσο χαμηλότερη είναι η CE τόσο πλησιάζουμε το *one – hot* διάνυσμα
- ▶ Η ομοιότητα των κόμβων μπορεί να εξαχθεί με βάση τις γνωστές μεθόδους π.χ. *DeepWalk*, *node2vec*, παραγοντοποίηση πίνακα

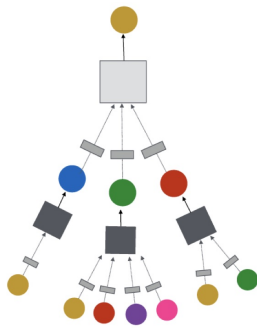
Επιβλεπόμενη Εκπαίδευση

Εκπαίδευσε απευθείας τον μοντέλο για επιβλεπόμενη εργασία
π.χ. ταξινόμηση κόμβου

Ερώτημα: ο κόμβος του ΚΔ είναι τοξικός ή όχι:



(α') αρχικό γράφημα



(β') ανάπτυγμα γειτονιάς

Επιβλεπόμενη Εκπαίδευση

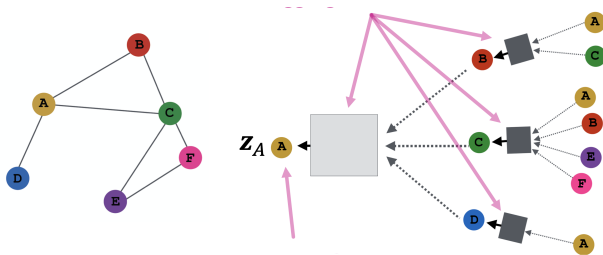
- ▶ Εκπαίδευουμε απευθείας το μοντέλο σύμφωνα με το υπόδειγμα της επιβλεπόμενης μάθησης.
- ▶ Στην έξοδο παίρνουμε το διάνυσμα z_v , το οποίο χρησιμοποιούμε σε συνδυασμό με κατάλληλη συνάρτηση απώλειας για να κάνουμε την εκπαίδευση
- ▶ Δυαδική ταξινόμηση άρα πρακτικά προσθέτουμε ένα ακόμη επίπεδο με βάρη θ και χρησιμοποιούμε την συνάρτηση απώλειας CE :

$$\mathcal{L} = - \sum_{v \in V} y_v \log(\sigma(z_v^T \theta)) + (1 - y_v) \log(1 - \sigma(z_v^T \theta))$$

- ▶ y_v η ετικέτα του κόμβου v (0 ή 1)
- ▶ θ τα βάρη του τελικού επιπέδου του δικτύου
- ▶ z_v η ενσωμάτωση του κόμβου v

Επιβλεπόμενη Εκπαίδευση

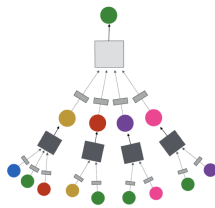
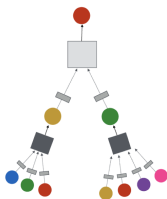
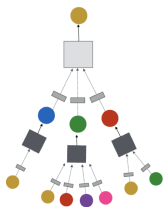
(1) Ορίζουμε μια συνάρτηση συγκερασμού των εισόδων



(2) Ορίζουμε μια συνάρτηση απώλειας για τις ενσωματώσεις των κόμβων

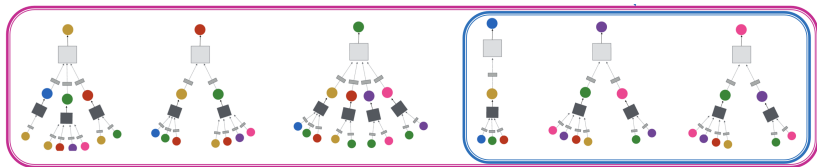
Επιβλεπόμενη Εκπαίδευση

(3) Εκπαιδύουμε σε σύνολο κόμβων για τους οποίους έχουμε την αντίστοιχη ετικέτα



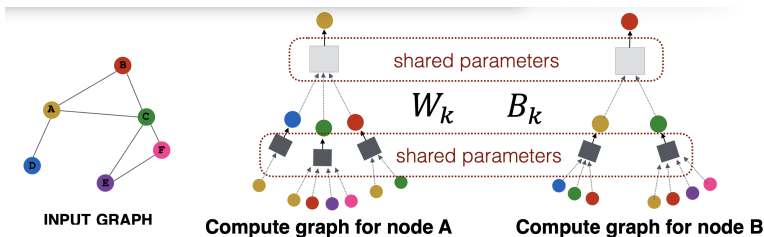
Επιβλεπόμενη Εκπαίδευση

- (4) Υπολογίζουμε τις ενσωματώσεις για τους **κόμβους που μας ενδιαφέρουν**
- (5) Γενικεύουμε και για εκείνους **που δεν συμπεριλάβαμε στην εκπαίδευση**
- (6) Υπολογίζουμε για κάθε κόμβο την ετικέτα παίρνοντας το μεγαλύτερο από τα $\log(\sigma(z_V^T \theta))$, $\log(1 - \sigma(z_V^T \theta))$



Επιβλεπόμενη Εκπαίδευση

- ▶ Τα ίδια βάρη μοιράζονται σε όλους τους κόμβους
- ▶ Ο αριθμός των βαρών είναι υπο-γραμμικό σε σχέση με το $|V|$



- ▶ Το δίκτυο γενικεύει σε νέους κόμβους και νέα γραφήματα (επαγωγική ικανότητα)
 - ▶ δημιουργία ενσωματώσεων σε πραγματικό χρόνο