

# ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ



## Νευρωνικά δίκτυα γράφων

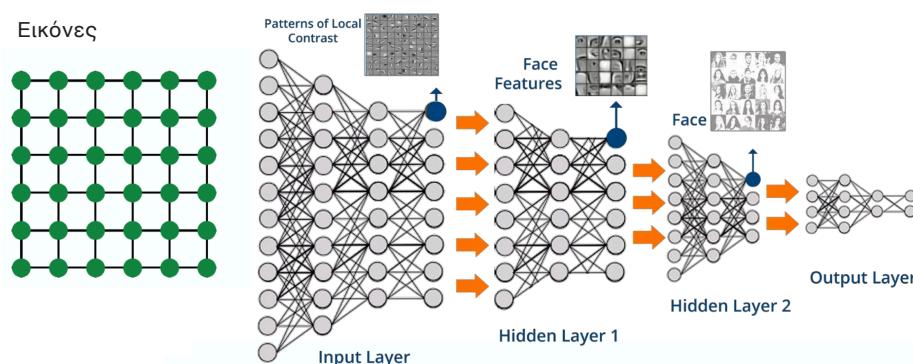
Γιώργος Στάμου

Καθηγητής Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ  
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης - AILS Lab

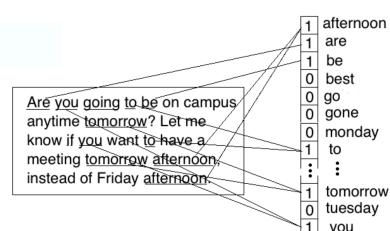
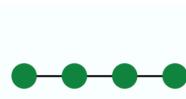
## ΑΝΑΠΑΡΑΣΤΑΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΓΡΑΦΟΥΣ - ΕΙΣΑΓΩΓΗ



### Βαθιά νευρωνικά δίκτυα και αναπαράσταση δεδομένων



Φωνή - Κείμενο

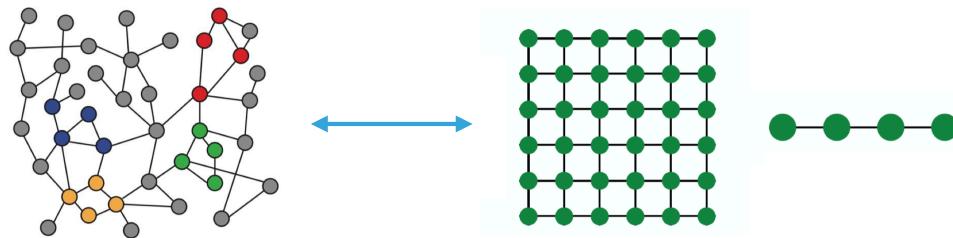




## ΑΝΑΠΑΡΑΣΤΑΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΓΡΑΦΟΥΣ - ΕΙΣΑΓΩΓΗ

### Αναπαράσταση δεδομένων με γράφους

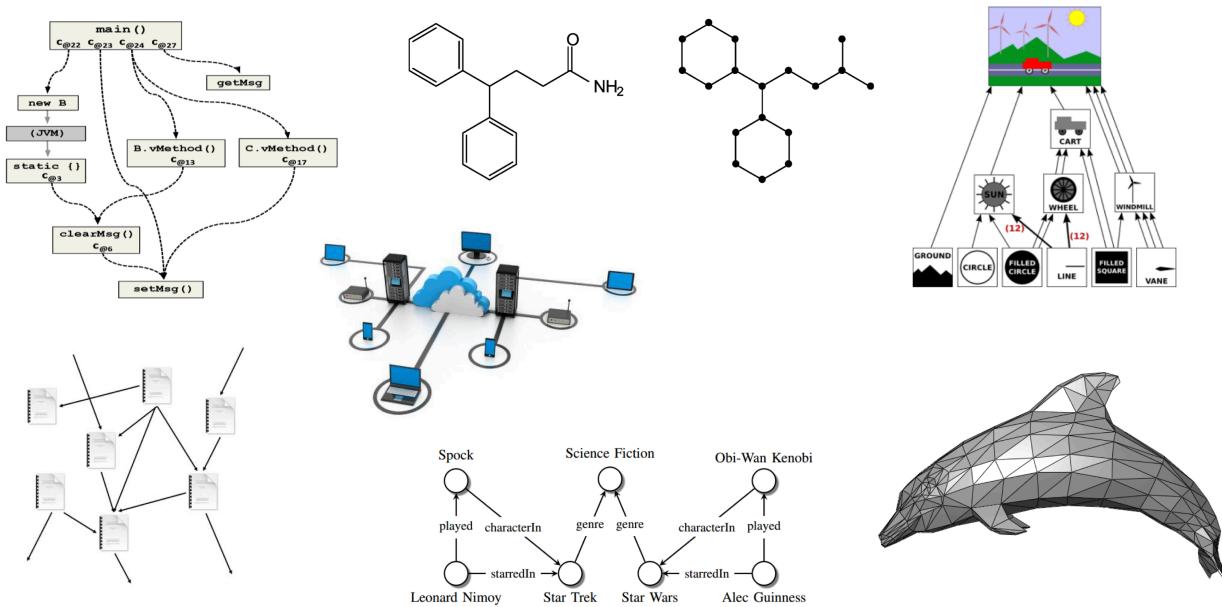
- ▶ Δεν είναι πάντα εφικτή η αναπαράσταση με σταθερή διάταξη-συσχέτιση των οντοτήτων κόμβων
- ▶ Είναι συχνά **δυναμική** η διάταξή τους ενώ έχουν πολλαπλές ιδιότητες που αλληλοσυσχετίζονται
- ▶ Τα **περίπλοκα πεδία** απαιτούν αναπαράσταση των οντοτήτων, αλλά και των συσχετίσεων μεταξύ των οντοτήτων
- ▶ Η αναπαράσταση των συσχετίσεων μπορεί να οδηγήσει σε καλύτερη επίδοση



3



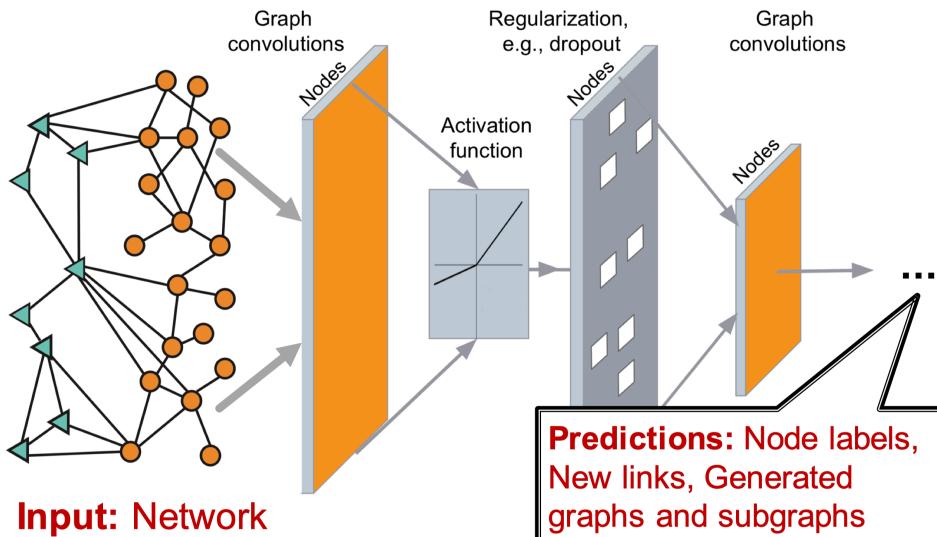
## ΑΝΑΠΑΡΑΣΤΑΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΓΡΑΦΟΥΣ - ΕΙΣΑΓΩΓΗ



4



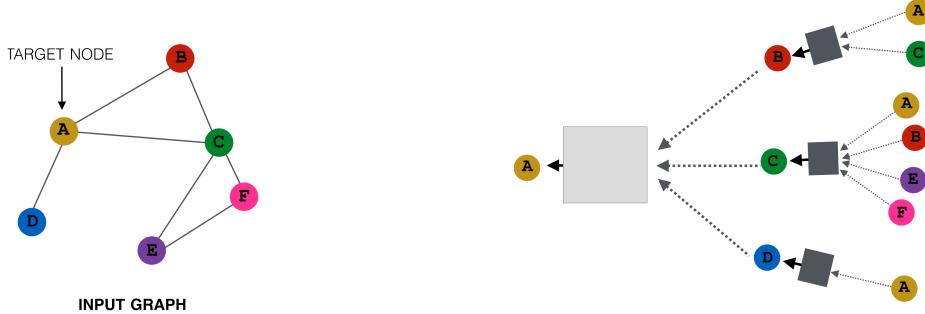
## Τεχνολογίες μηχανικής μάθησης για δεδομένα γράφων



5

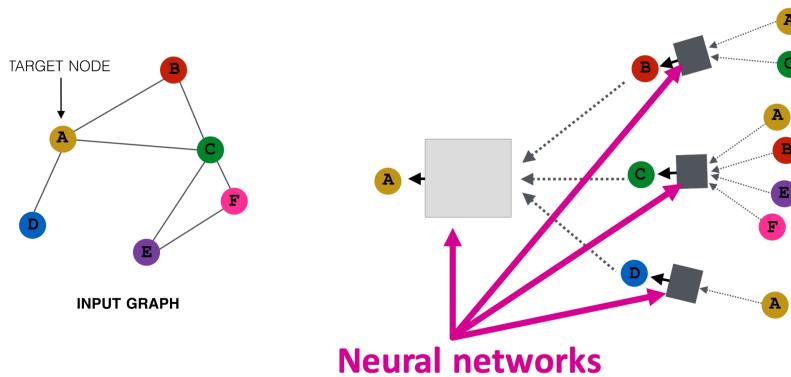


## Τεχνολογίες μηχανικής μάθησης για δεδομένα γράφων

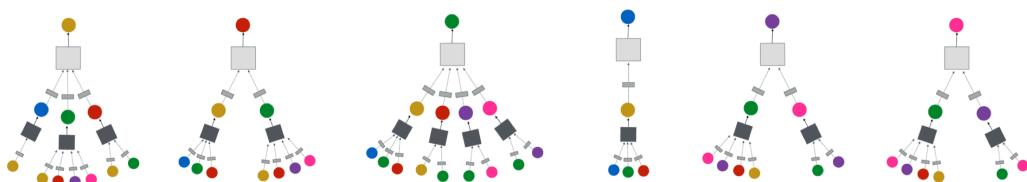




Τεχνολογίες μηχανικής μάθησης για δεδομένα γράφων



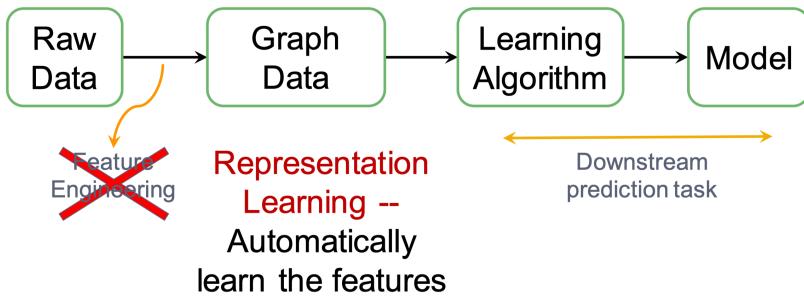
Τεχνολογίες μηχανικής μάθησης για δεδομένα γράφων



Inductive Representation Learning Jure Leskovec, onStanford Large University Graphs . W. Hamilton, R. Ying, J. Leskovec. NIPS, 2017.



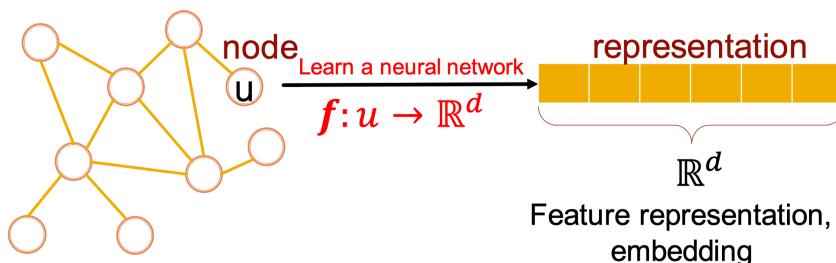
Βαθιά νευρωνικά δίκτυα και αναπαράσταση δεδομένων



9



Βαθιά νευρωνικά δίκτυα και αναπαράσταση δεδομένων

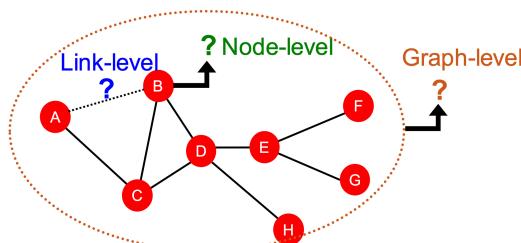


10



### Μηχανική μάθηση για δεδομένα γράφων

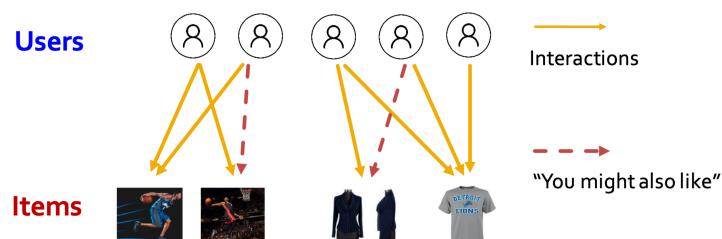
- ▶ Παραδοσιακές μέθοδοι (graphlets, graph kernels)
- ▶ Ενθετικές αναπαραστάσεις γράφων (deepwalk, node2vec)
- ▶ Νευρωνικά δίκτυα γράφων (GCN, GraphSAGE, GAT, Theory of GNNs)
- ▶ Γράφοι γνώσης και αυτόματη συλλογιστική (TransE, BetaE)
- ▶ Βαθιά παραγωγικά μοντέλα γράφων (GraphRNN)
- ▶ Εφαρμογές



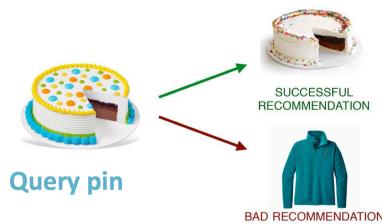
### Ενδιαφέροντα προβλήματα

- ▶ Ταξινόμηση κόμβων (node classification)
  - ▶ Έλεγχες αν μία έννοια θα μπορούσε να ανήκει στην ετικέτα ενός κόμβου
- ▶ Πρόβλεψη ακμής (link prediction)
  - ▶ Έλεγχες αν μία ακμή λείπει από ένα γράφο
- ▶ Ταξινόμηση γράφου
  - ▶ Κατηγοριοποίηση ένα γράφο σε μία έννοια
- ▶ Συσταδοποίηση
  - ▶ Βρες αν μία ομάδα κόμβων ανήκει στην ίδια κατηγορία
- ▶ Παραγωγή γράφου
  - ▶ Κατασκεύασε ένα νέο γράφο, με βάση ένα σύνολο από γράφους
- ▶ Εξέλιξη γράφου
  - ▶ Τροποποίησε ένα γράφο με βάση κάποιο κριτήριο

11

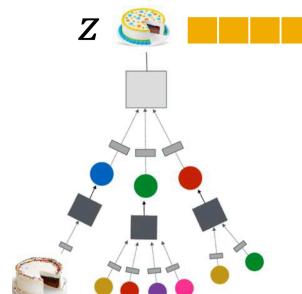
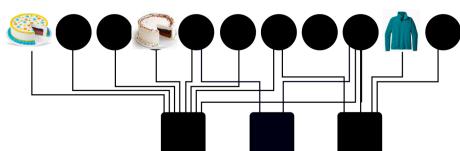


12

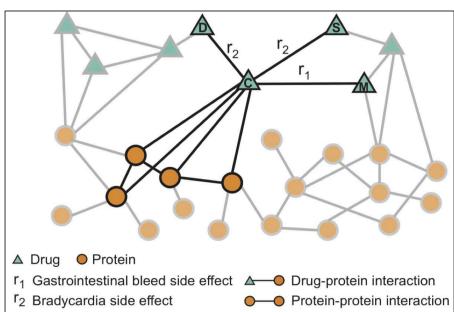


**Task:** Learn node embeddings  $z_i$  such that  

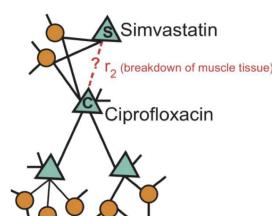
$$d(z_{cake1}, z_{cake2}) < d(z_{cake1}, z_{sweater})$$



13



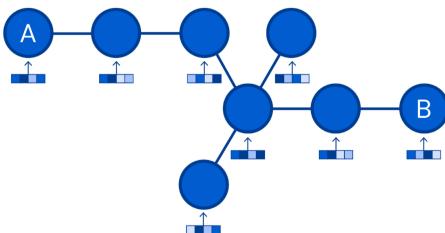
- ▶ Κόμβοι: Φάρμακα και πρωτεΐνες
- ▶ Ακμές: Επιδράσεις
- ▶ Ερώτημα: Πόσο πιθανό είναι η Simvastatin και η Ciprofloxacin μαζί να διασπούν το μυϊκό ιστό



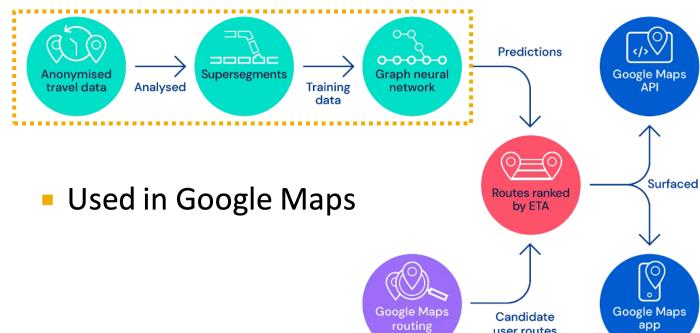
14



## ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΓΡΑΦΩΝ - ΕΦΑΡΜΟΓΕΣ



- ▶ Κόμβοι: Τμήματα του δρόμου
- ▶ Ακμές: Διασυνδέσεις τμημάτων
- ▶ Πρόβλεψη: Χρόνος άφιξης

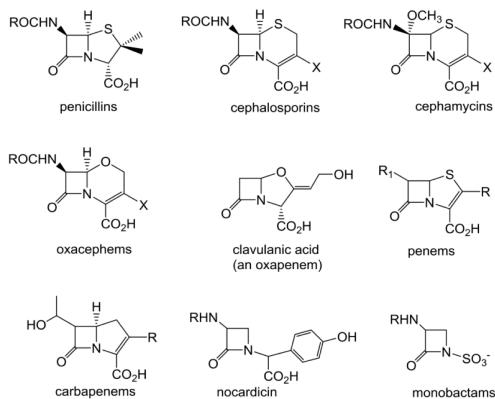


DeepMind

15



## ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΓΡΑΦΩΝ - ΕΦΑΡΜΟΓΕΣ



- ▶ Κόμβοι: Άτομα
- ▶ Ακμές: Χημικοί δεσμοί
- ▶ Πρόβλεψη: Ταξινόμηση αντιβιοτικών

Konaklieva, Monika I. "Molecular targets of β-lactam-based antimicrobials: beyond the usual suspects." *Antibiotics* 3.2 (2014): 128-142.

16



### Γράφος

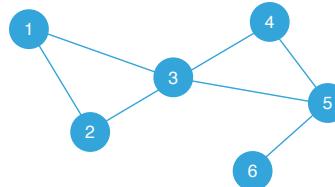
$N$ : σύνολο κόμβων

$E$ : σύνολο ακμών

$G(N, E)$ : γράφος

$T$ : σύνολο τύπων κόμβων

$R$ : σύνολο τύπων ακμών



### Ιδιότητες και είδη γράφων

- ▶ Διασυνδεδεμένοι
- ▶ Κατευθυνόμενοι
- ▶ Διμερείς
- ▶ Βαθμονομημένοι
- ▶ Αλυσίδες
- ▶ ...

### Αναπαράσταση γράφου

$A$ : πίνακας γειτνίασης

0	1	1	0	0	0
1	0	1	0	0	0
1	1	0	1	1	0
0	0	1	0	1	0
0	0	1	1	0	1
0	0	0	0	1	0

$A_L$ : Λαγλασιανός πίνακας

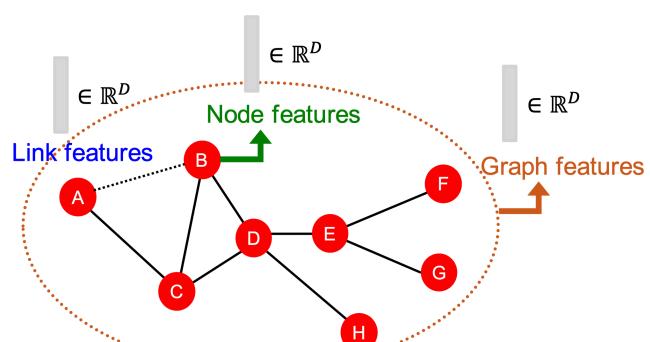
2	-1	-1	0	0	0
-1	2	-1	0	0	0
-1	-1	4	-1	-1	0
0	0	-1	2	-1	0
0	0	-1	-1	3	-1
0	0	0	0	-1	1

17



### Χρήση μοντέλων μηχανικής μάθησης

- ▶ Όρισε ένα σύνολο από χαρακτηριστικά για την αναπαράσταση του γράφου
- ▶ Υπολόγισε τα χαρακτηριστικά για όλους τους γράφους του συνόλου δεδομένων
- ▶ Εκπαίδευσε ένα μοντέλο (γραμμικό ταξινομητή, δέντρο απόφασης, τυχαίο δάσος, νευρωνικό δίκτυο κλπ)
- ▶ Για κάθε νέο γράφο: υπολόγισε τα χαρακτηριστικά και ταξινόμησε με βάση το μοντέλο

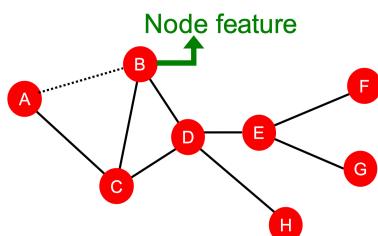
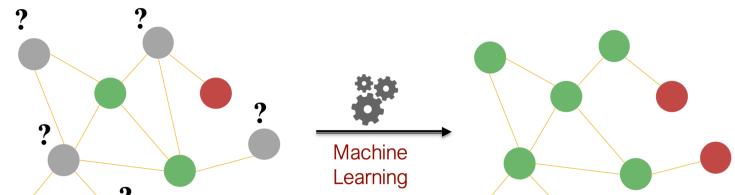


18



### Χαρακτηριστικά κόμβων

- ▶ Περιγράφουν τη δομή και τη θέση του κόμβου στο γράφο
- ▶ Βαθμός κόμβου (node degree)
- ▶ Κεντρικότητα κόμβου (node centrality)
- ▶ Συντελεστής συσταδοποίησης (clustering coefficient)
- ▶ Graphlets

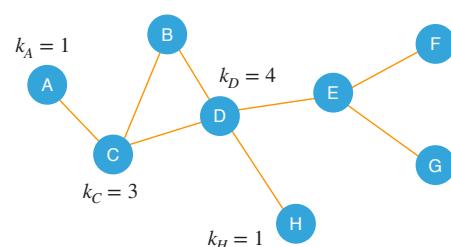


19



### Χαρακτηριστικά κόμβων

- ▶ Βαθμός κόμβου
- ▶ Πλήθος ακμών του κόμβου

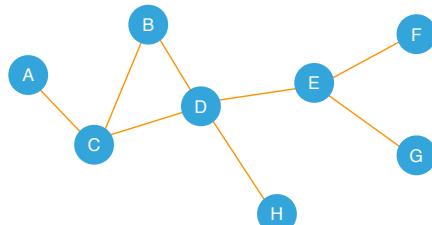


20



### Χαρακτηριστικά κόμβων

- ▶ Κεντρικότητα κόμβου
  - ▶ Σημασία του κόμβου στον γράφο
  - ▶ Πλήθος ακμών προς **σημαντικούς** κόμβους
  - ▶ Σημαντικότητα κόμβου
    - ▶ Ένας κόμβος είναι σημαντικός αν συνδέεται με σημαντικούς κόμβους
- ▶ Μέτρα σημαντικότητας
  - ▶ Ιδιοτιμές (eigenvalues)
  - ▶ Ενδιαμεσότητα (betweenness)
  - ▶ Εγγύτητα (closeness)
  - ▶ Συντελεστής συσταδοποίησης (clustering coefficient)
  - ▶ ...



Ιδιοτιμές

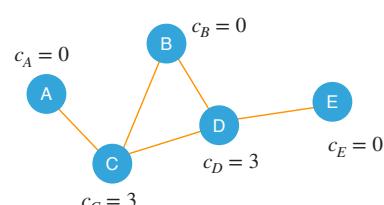
$$c_v = \frac{1}{\lambda} \sum_{u \in N(v)} c_u \quad \lambda c = Ac \quad \text{Ιδιοδιανύσματα του } A$$

Κεντρικότητα: Ιδιοδιανύσμα που αντιστοιχεί στην μέγιστη ιδιοτιμή  $\lambda_{\max}$



### Χαρακτηριστικά κόμβων

- ▶ Κεντρικότητα κόμβου
  - ▶ Σημασία του κόμβου στον γράφο
  - ▶ Πλήθος ακμών προς **σημαντικούς** κόμβους
- ▶ Σημαντικότητα κόμβου
  - ▶ Ένας κόμβος είναι σημαντικός αν συνδέεται με σημαντικούς κόμβους
- ▶ Μέτρα σημαντικότητας
  - ▶ Ιδιοτιμές (eigenvalues)
  - ▶ Ενδιαμεσότητα (betweenness)
  - ▶ Εγγύτητα (closeness)
  - ▶ Συντελεστής συσταδοποίησης (clustering coefficient)
  - ▶ ...



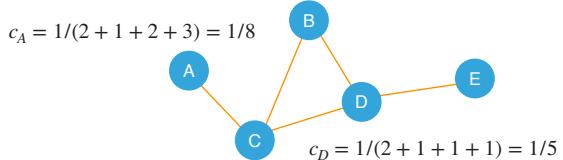
Ενδιαμεσότητα

$$c_v = \sum_{s \neq v \neq t} \frac{\text{πλήθος ελάχιστων μονοπατιών από τον } s \text{ στον } t \text{ που περνούν από τον } v}{\text{πλήθος ελάχιστων μονοπατιών από τον } s \text{ στον } t}$$



### Χαρακτηριστικά κόμβων

- ▶ Κεντρικότητα κόμβου
  - ▶ Σημασία του κόμβου στον γράφο
  - ▶ Πλήθος ακμών προς σημαντικούς κόμβους
  - ▶ Σημαντικότητα κόμβου
    - ▶ Ένας κόμβος είναι σημαντικός αν συνδέεται με σημαντικούς κόμβους
- ▶ Μέτρα σημαντικότητας
  - ▶ Ιδιοτιμές (eigenvalues)
  - ▶ Ενδιαμεσότητα (betweenness)
  - ▶ Εγγύτητα (closeness)
  - ▶ Συντελεστής συσταδοποίησης (clustering coefficient)
  - ▶ ...



### Εγγύτητα

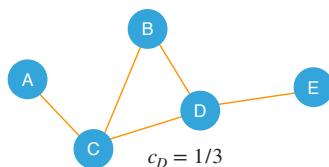
$$c_v = \frac{1}{\sum_{u \neq v} \text{μήκος ελάχιστου μονοπατιού από τον } u \text{ στον } v}$$

23



### Χαρακτηριστικά κόμβων

- ▶ Κεντρικότητα κόμβου
  - ▶ Σημασία του κόμβου στον γράφο
  - ▶ Πλήθος ακμών προς σημαντικούς κόμβους
- ▶ Σημαντικότητα κόμβου
  - ▶ Ένας κόμβος είναι σημαντικός αν συνδέεται με σημαντικούς κόμβους
- ▶ Μέτρα σημαντικότητας
  - ▶ Ιδιοτιμές (eigenvalues)
  - ▶ Ενδιαμεσότητα (betweenness)
  - ▶ Εγγύτητα (closeness)
  - ▶ Συντελεστής συσταδοποίησης (clustering coefficient)
  - ▶ ...



### Συντελεστής συσταδοποίησης

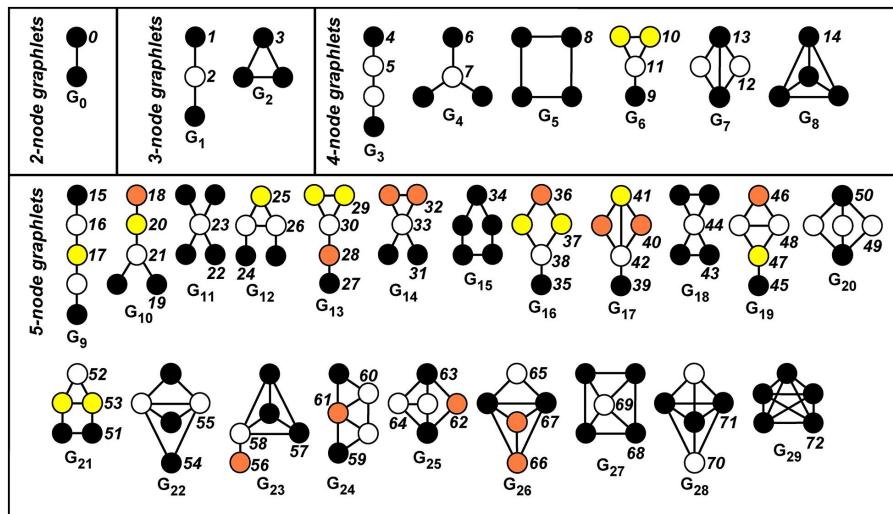
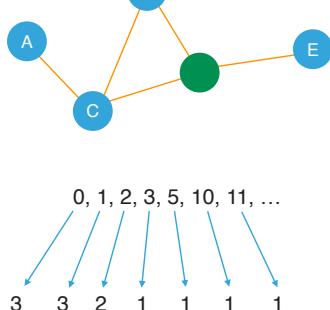
Πόσο διασυνδεδεμένοι είναι μεταξύ τους οι γειτονικοί κόμβοι;

$$c_v = \frac{\text{αριθμός ακμών μεταξύ των γειτονικών κόμβων}}{\text{συνολικός αριθμός πιθανών ακμών μεταξύ των γειτονικών κόμβων}}$$

24

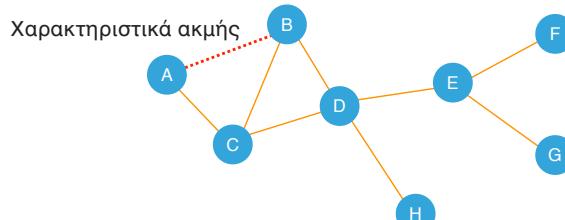
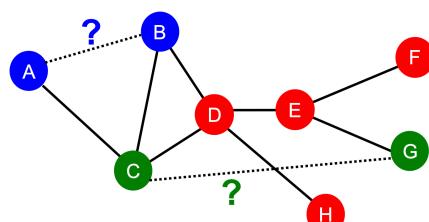


## Graphlets



## Χαρακτηριστικά ακμών

- ▶ Περιγράφουν τη σχέση δύο κόμβων του γράφου
- ▶ Με βάση την απόσταση
- ▶ Με βάση τους κοινούς άμεσους γείτονες
- ▶ Με βάση τους κοινά προσβάσιμους κόμβους

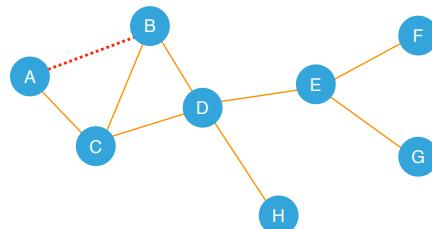




## ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΓΡΑΦΩΝ - ΤΑΞΙΝΟΜΗΣΗ ΜΕ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

### Χαρακτηριστικά ακμών

- ▶ Περιγράφουν τη σχέση δύο κόμβων του γράφου
  - ▶ Με βάση την απόσταση
  - ▶ Με βάση τους κοινούς άμεσους γείτονες
  - ▶ Με βάση τους κοινά προσβάσιμους κόμβους



### Πλήθος κοινών άμεσων γειτόνων

$$|n(A) \cap n(B)| = |\{C\}| = 1$$

$$\frac{|n(A) \cap n(B)|}{|n(A) \cup n(B)|} = \frac{|\{C\}|}{|\{C,D\}|} = 1/2$$

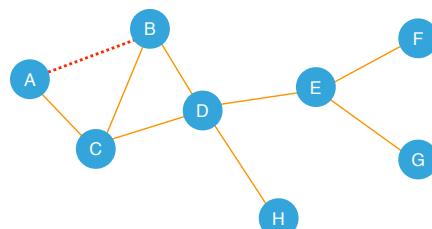
27



## ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΓΡΑΦΩΝ - ΤΑΞΙΝΟΜΗΣΗ ΜΕ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

### Χαρακτηριστικά ακμών

- ▶ Περιγράφουν τη σχέση δύο κόμβων του γράφου
  - ▶ Με βάση την απόσταση
  - ▶ Με βάση τους κοινούς άμεσους γείτονες
  - ▶ Με βάση τους κοινά προσβάσιμους κόμβους



### Δείκτης Katz

Αριθμός των περιπάτων, μήκους  $l$  μεταξύ των κόμβων  $v_1, v_2$ :

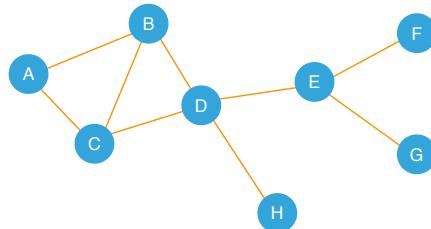
$$s(v_1, v_2) = \sum_{l=1}^{\infty} \beta^l A_{v_1 v_2}^l$$

28



### Χαρακτηριστικά συνολικού γράφου

- ▶ Αναπαριστούν τη δομή του γράφου σε ένα χώρο χαρακτηριστικών, έτσι ώστε η ομοιότητα δύο γράφων να είναι αντιστρόφως ανάλογη της απόστασης τους στο χώρο χαρακτηριστικών



- ▶ Graphlet kernel
- ▶ Weisfeiler-Lehman Kernel
- ▶ Random-walk kernel
- ▶ Shortest-path graph kernel
- ▶ ...

### Πυρήνες (kernels)

$$K(G_1, G_2) \in \mathbb{R}$$

$$\text{similarity}(G_1, G_2) = K(G_1, G_2) = \phi(G_1)^T \phi(G_2)$$

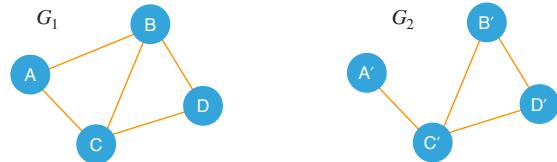
Shervashidze, Nino, et al. "Efficient graphlet kernels for large graph comparison." Artificial Intelligence and Statistics (2009)

Shervashidze, Nino, et al. "Weisfeiler-Lehman graph kernels." Journal of Machine Learning Research 12.9 (2011)



### Τεχνική bag of words (BoW)

- ▶ Διάνυσμα η κάθε διάσταση του οποίου αντιστοιχεί στο πλήθος των κόμβων που έχουν μία συγκεκριμένη ιδιότητα
  - ▶ “λεξιλόγιο” βαθμών κόμβων



1 διάσταση = πλήθος μπλες κόμβων

$$\phi(G_1) = [4]$$

$$\phi(G_2) = [4]$$

$n$  διαστάσεις = πλήθος κόμβων με  $n$  ακμές

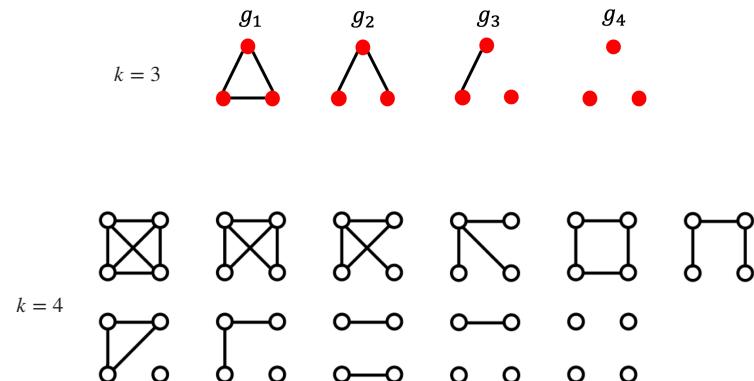
$$\phi(G_1) = [0 \ 2 \ 2]$$

$$\phi(G_2) = [1 \ 2 \ 1]$$



### Τεχνική bag of graphlets

- ▶ Διάνυσμα η κάθε διάσταση του οποίου αντιστοιχεί σε δομικά χαρακτηριστικά του γράφου που περιγράφονται από το πλήθος ενσωματώσεων graphlets στο γράφο
- ▶ στα graphlets συμπεριλαμβάνονται και μη συνδεδεμένοι κόμβοι
- ▶ αφού τα graphlets ενσωματώνονται σε επίπεδο γράφου, δεν θεωρούμε κατά την ενσωμάτωση κάποιον κόμβο-ρίζα



Shervashidze, Nino, et al. "Efficient graphlet kernels for large graph comparison." Artificial Intelligence and Statistics (2009)

31

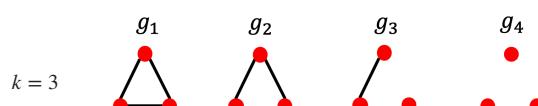


### Τεχνική bag of graphlets

- ▶ Διάνυσμα η κάθε διάσταση του οποίου αντιστοιχεί σε δομικά χαρακτηριστικά του γράφου που περιγράφονται από το πλήθος ενσωματώσεων graphlets στο γράφο
- ▶ στα graphlets συμπεριλαμβάνονται και μη συνδεδεμένοι κόμβοι
- ▶ αφού τα graphlets ενσωματώνονται σε επίπεδο γράφου, δεν θεωρούμε κατά την ενσωμάτωση κάποιον κόμβο-ρίζα



$k = 3 \rightarrow 4$  graphlets  $\rightarrow$  πιστήνας 4 διαστάσεων



### Πρόβλημα 1

- ▶ Σε μεγάλους γράφους τα διανύσματα μπορούν να αποκτήσουν μεγάλες τιμές

### Λύση

- ▶ Μπορούμε να κανονικοποιήσουμε ως προς το άθροισμα των τιμών

32



### Τεχνική bag of graphlets

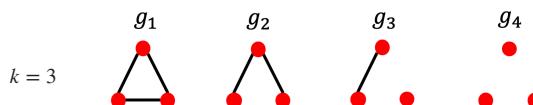
- Διάνυσμα η κάθε διάσταση του οποίου αντιστοιχεί σε δομικά χαρακτηριστικά του γράφου που περιγράφονται από το πλήθος ενσωματώσεων graphlets στο γράφο
- στα graphlets συμπεριλαμβάνονται και μη συνδεδεμένοι κόμβοι
- αφού τα graphlets ενσωματώνονται σε επίπεδο γράφου, δεν θεωρούμε κατά την ενσωμάτωση κάποιον κόμβο-ρίζα



$$\phi(G_1) = [2 \ 2 \ 0 \ 0]$$

$$\phi(G_2) = [1 \ 2 \ 1 \ 0]$$

$k = 3 \rightarrow 4$  graphlets  $\rightarrow$  πυρήνας 4 διαστάσεων



### Λύση

- Πρέπει να βρούμε έναν πυρήνα που να υπολογίζεται σχετικά γρήγορα, ακόμα και για μεγάλους γράφους

33

### Πρόβλημα 2

- Ο υπολογισμός των graphlets είναι πολύ χρονοβόρος (παρόμοιο με το πρόβλημα του graph matching - NP hard)



### Πυρήνας Weisfeiler-Lehman (WL)

- Επέκταση της ιδέας του λεξιλογίου βαθμών κόμβων
- Συνεκτιμούμε και το βαθμό των γειτόνων, το βαθμό των γειτόνων των γειτόνων κλπ.
- Ο υπολογισμός στηρίζεται στον αλγόριθμο color refinement

### Πρόβλημα color refinement

Δίνεται ένας γράφος  $G$  με σύνολο κόμβων  $V$ .

Δώσεις ένα αρχικό χρώμα  $c^{(0)}(v)$  σε κάθε κόμβο  $v$ .

Σε κάθε βήμα τροποποίησε τα χρώματα των κόμβων από την εξίσωση:

$$c^{(k+1)}(v) = \text{HASH} \left( \{c^{(k+1)}(v), \{c^{(k+1)}(u)\}_{u \in N(v)}\} \right) \quad \text{color refinement}$$

όπου:  $N(v)$  οι γείτονες του  $v$

HASH μία αντιστοίχιση διαφορετικών τιμών σε διαφορετικά χρώματα

Μετά από  $k$  βήματα color refinement, το χρώμα:  $c^k(v)$

περιγράφει την  $k$ -βημάτων δομή της γειτονιάς του  $v$

### Υπολογισμός πυρήνα WL

Δεδομένων δύο γράφων  $G_1$  και  $G_2$

'Εστω  $c_1, c_2, \dots, c_n$  όλα τα χρώματα που εμφανίζονται στα βήματα του ταυτόχρονου color refinement των  $G_1, G_2$

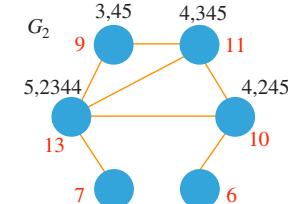
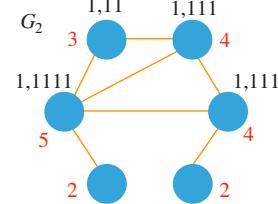
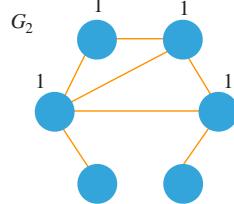
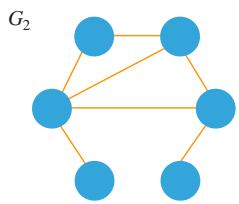
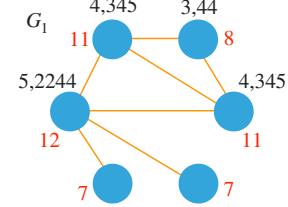
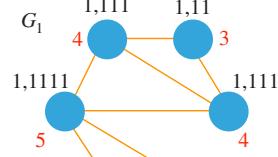
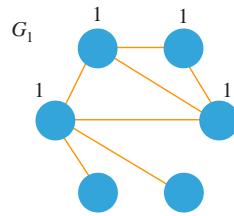
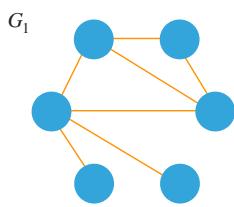
Υπολογίζουμε τον πυρήνα  $K(G_1, G_2) = \phi(G_1)^T \phi(G_2)$   $\phi(G_1) = [t(G_1, c_1), t(G_1, c_2), \dots, t(G_1, c_n)]$   $\phi(G_2) = [t(G_2, c_1), t(G_2, c_2), \dots, t(G_2, c_n)]$

$t(G_i, c_j)$  ο συνολικός αριθμός κόμβων του  $G_i$  που είχαν σε κάποιο βήμα του color refinement το χρώμα  $c_j$

34



## ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΓΡΑΦΩΝ - ΤΑΞΙΝΟΜΗΣΗ ΜΕ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ



$$\phi(G_1) = [t(G_1, c_1), t(G_1, c_2), \dots, t(G_1, c_{13})] \\ = [6, 2, 1, 2, 1, 0, 2, 1, 0, 0, 2, 1, 0]$$

$$\phi(G_2) = [t(G_2, c_1), t(G_2, c_2), \dots, t(G_2, c_{13})] \\ = [6, 2, 1, 2, 1, 1, 1, 0, 1, 1, 1, 0, 1]$$

$$K(G_1, G_2) = \phi(G_1)^T \phi(G_2) = 49$$

HASH

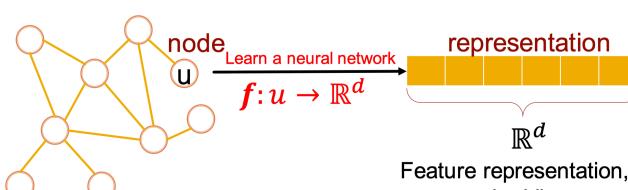
1,1	→	2
1,11	→	3
1,111	→	4
1,1111	→	5

HASH

2,4	→	6
2,5	→	7
3,44	→	8
4,245	→	10
5,2244	→	12
5,2344	→	13

35

## ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΓΡΑΦΩΝ - ΕΝΘΕΤΙΚΕΣ ΑΝΑΠΑΡΑΣΤΑΣΕΙΣ



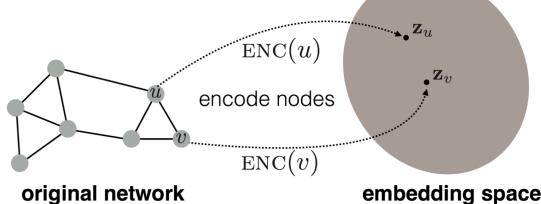
Στόχος:  $\text{sim}(v, u) \approx z_v \cdot z_u$

Ένθεση (embedding)

Απεικονίζει τα κόμβων σε ένα διάνυσμα χαμηλής διάστασης

$v \rightarrow \text{ENC}(v) = z_v$   
Κόμβος του γράφου

Ένθεση διάστασης  $d$

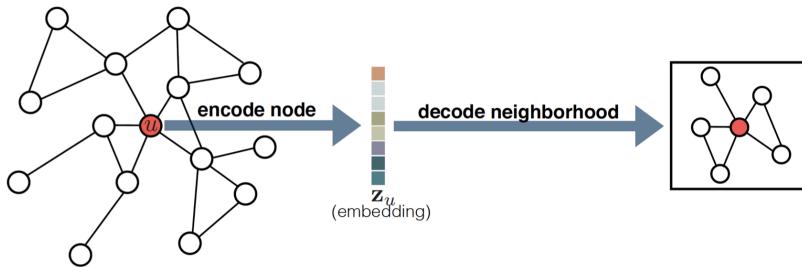


Συνάρτηση ομοιότητας (similarity function)

Απεικονίζει τη σχέση των ενθέσεων των κόμβων στη σχέση των κόμβων του γράφου

$\text{sim}: z_v^T z_u \approx \text{sim}(v, u)$   
dot-product των ενθέσεων      Ομοιότητα των κόμβων

36



#### Προσέγγιση κωδικοποιητή-αποκωδικοποιητή (encoder-decoder)

- ▶ Ο κωδικοποιητής αναπαριστά υπολογίζει το διάνυσμα της ενθετικής αναπαράστασης του κόμβου
- ▶ Ο αποκωδικοποιητής ανακατασκευάζει μία γειτονιά του κόμβου, δηλαδή ένα μέρος του αρχικού γράφου
- ▶ Με τον τρόπο αυτό μπορούμε να κατασκευάσουμε εύκολα ένα σύνολο δεδομένων εισόδου-εξόδου για την εκμάθηση της ενθετικής αναπαράστασης, αρκεί να έχουμε ένα μεγάλο σύνολο από γράφους

37



#### Κωδικοποιητής - ρηχές αναπαραστάσεις

Χρησιμοποιούμε απλά ένα πίνακα αναζήτησης  
(lookup table)

$$\text{ENC}(v) = z_v = \mathbf{Z} \cdot v$$

$\in \mathbb{R}^{d \times |V|}$        $v \in \mathbb{I}^{|V|}$

Μοναδιαίο διάνυσμα στη θέση  $i$  που αντιστοιχεί στον  $v_i$

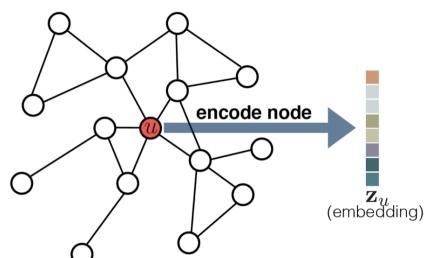
Η  $i$  στήλη του είναι το ενθετικό διάνυσμα του  $v$

$$\text{ENC}(v_i) \longrightarrow$$

$$\mathbf{Z} = \begin{bmatrix} z_1^{(v_1)} & \dots & z_1^{(v_i)} & \dots & z_1^{(v_n)} \\ z_2^{(v_1)} & \dots & z_2^{(v_i)} & \dots & z_2^{(v_n)} \\ \vdots & & \vdots & & \vdots \\ z_d^{(v_1)} & \dots & z_d^{(v_i)} & \dots & z_d^{(v_n)} \end{bmatrix}$$

$$\mathbf{I}^{(v_i)} = \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{bmatrix}$$

$i$ -θέση



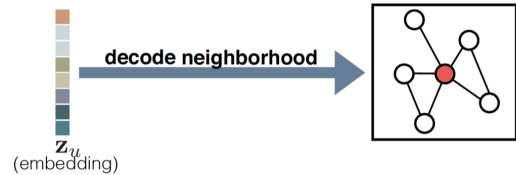
$$\begin{bmatrix} z_1^{(v_i)} \\ z_2^{(v_i)} \\ \vdots \\ z_d^{(v_i)} \end{bmatrix} \xrightarrow{z_{v_i}}$$

38



### Αποκωδικοποιητής

Δεδομένης της ένθεσης  $z_v$  ενός κόμβου  $v$ ,  
βρες ένα σύνολο γειτόνων  $N(v)$  του  $v$ ,  
δηλαδή μία εκτίμηση της γραμμής  $A[v]$  του πίνακα γειτνίασης  $A$ .



Η εκτίμηση χρησιμοποιεί έναν κατά ζεύγη αποκωδικοποιητή (pairwise decoder)

$$\text{DEC} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$$

Με βάση τον αποκωδικοποιητή αυτό έχουμε:

$$\text{DEC}(\text{ENC}(v), \text{ENC}(u)) = \text{DEC}(z_v, z_u) \approx S[v, u] \doteq A[v, u]$$

**Inner product decoder**

$$\text{DEC}(z_v, z_u) = z_v \cdot z_u$$

Συνεπώς, ανακατασκευάζεται ο πίνακας γειτνίασης.



### Μέθοδος παραγοντοποίησης (factorization)

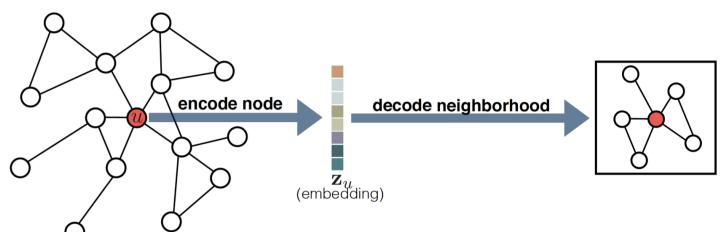
Εμπειρικό σφάλμα ανακατασκευής:

$$L_{\text{ID}} = \sum_{(v,u) \in \mathbb{D}} \text{loss}(\text{DEC}(z_v, z_u), A[v, u])$$

$$\text{loss} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$$

για παράδειγμα:

$$\text{loss} = \|\text{DEC}(z_v, z_u) - A[v, u]\|_2^2$$



### Laplacian eigenmaps

Αν  $A$  ο Λαπλασιανός πίνακας του γράφου

$$\text{Έχουμε: } \text{DEC}(z_v, z_u) = \|z_v - z_u\|_2^2 \quad L = \sum_{(v,u) \in \mathbb{D}} \text{DEC}(z_v, z_u) \cdot A[v, u]$$

Τότε οι  $d$  μικρότερες ιδιοτιμές  
ελαχιστοποιούν το σφάλμα  $L$

Το σφάλμα είναι μεγάλο όταν ταυτόχρονα τα  $\text{DEC}(z_v, z_u)$  και  $A[v, u]$  είναι μεγάλα



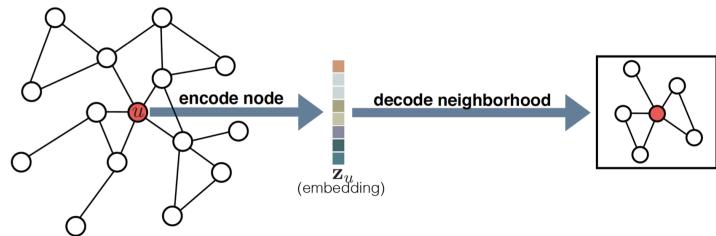
### Deep walk

Κωδικοποιητής ρηχής αναπαράστασης

$$\text{ENC}(v) = z_v = \mathbf{Z} \cdot v$$

Αποκωδικοποιητής εσωτερικού γινομένου

$$\text{DEC}(z_v, z_u) = z_v \cdot z_u$$



Ορισμός ομοιότητας κόμβων

$$\text{sim}(z_v, z_u) = \{\eta \text{ πιθανότητα να περνάει από τον } v \text{ ένας τυχαίος περίπατος } k \text{ βημάτων, που ξεκινά από τον } u\}$$