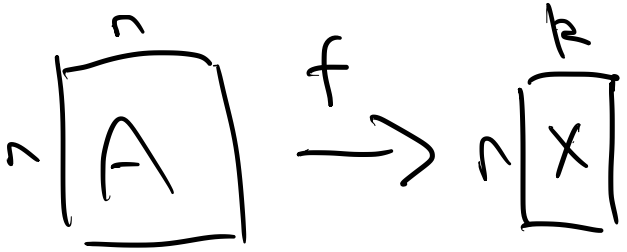
April 14, 2026

Last time: "representation" learning on graphs



- generative modeling
- data interpretation / vis
- "downstream" ML tasks

① Clustering / node prediction → unsupervised

→ semi-supervised → $\begin{bmatrix} X_H \\ X_U \end{bmatrix} = X$ (k)

$$\min_z \|X_H z - y\|_2^2 + \text{regularization } \|z\|_2^2$$

$$z^* \quad X_U z^*$$

Ex: age, page views, traffic flow

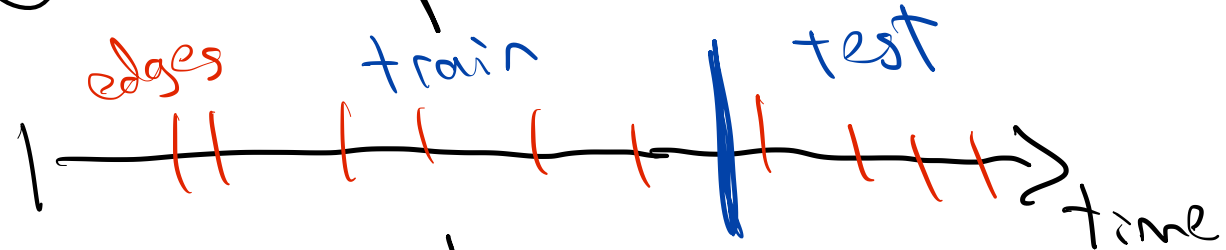
Ex: gender, protein function

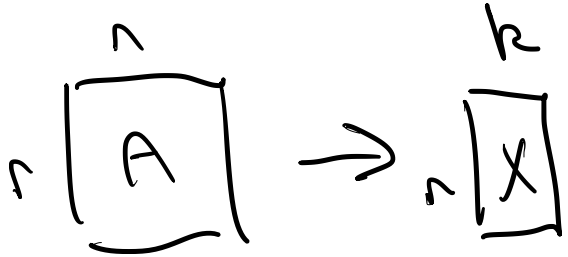
$$\Pr(y=1 \mid x) = \frac{1}{1+\exp(-z^T x)}$$

$$\max_z \prod_{h \in H} \Pr(y_h \mid z, x_h)$$

② Link prediction / recommendation

Ex:
friend rec FB/Twitter
co-purchase
co-view



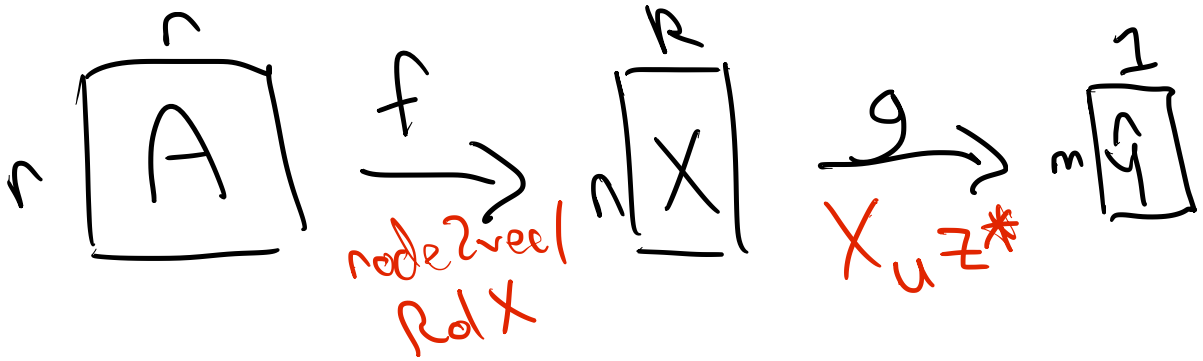


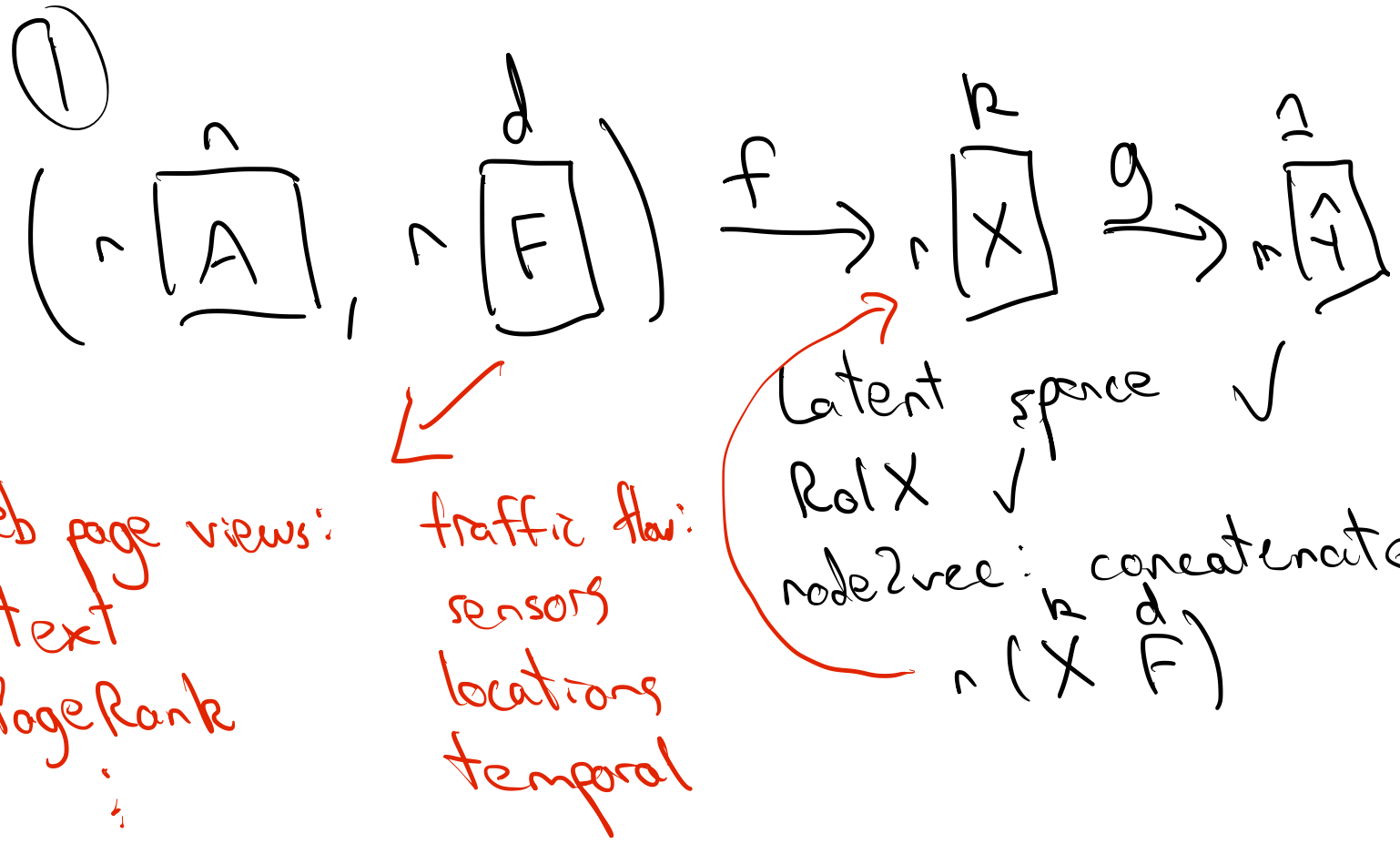
$$Pr((i,j) \text{ forms edge}) = \frac{1}{1+\exp(\|x_i - x_j\|_2^2)}$$

$\|x_i - x_j\|_2^2 \rightarrow x_i^T x_j$

Test: (i,j) positive
(k,l) negative (artificial)

$\Rightarrow$ accuracy


n
A
f
node2vec
Rd X
R
n
X
g
X u Z*
1
m
Ŷ


①
n
A
d
n
F
f
R
n
X
g
1
m
Ŷ
Latent space ✓
Rd X ✓
node2vec: concatenate
n (X F)
R d
web page views:
text
PageRank
traffic flow:
sensors
locations
temporal

② Couple f, g, loss ℓ

$$\ell \circ g \circ f\left(\hat{A}, \hat{F}\right) \to \ell\left(\hat{Y}\right)$$

Ex: $\|\hat{y} - y_H\|_2^2$

So far: decoupled

$$\text{Rel}X\ (A, F) \to X$$

$$\log(X) = \|\underbrace{X_H z}_{g(X)} - y_H\|_2^2 \qquad \hat{y} = X_H z^*$$

Latent space model

node2vec

implicitly defining log

- $\Pr((i,j) \mid x_i, x_j)$
- $\Pr(N_S(i) \mid x_i)$

Enabled by systems

TensorFlow

PyTorch

Back to April 7...

$\min_x$ "unsmoothness" $f(x)$ + "stay close to known labels"

$x^T L x$ $x^T N x$

$\sum_{(i,j)\in E} (x_i - x_j)^2$

$\lambda \left\| \begin{pmatrix} x_H \\ x_U \end{pmatrix} - \begin{pmatrix} y_H \\ 0 \end{pmatrix} \right\|_2^2$ y

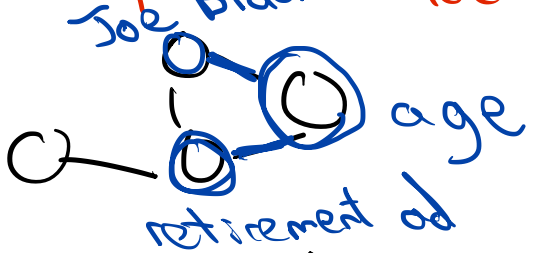
$x_i \Leftrightarrow$ node i

GD step size $\alpha < 1$, $\lambda = \frac{1-\alpha}{\alpha}$, $x^T N x$

$$x^{(t+1)} = \alpha \left(D^{-1/2} A D^{-1/2} \right) x^{(t)} + (1-\alpha) y$$

"global consistency" "local consistency"

Set $y = F(:, i)$ i^{th} feature



$$X^{(t+1)} = \alpha S X^{(t)} + (1-\alpha) F \qquad S = D^{-1/2} A D^{-1/2}$$

$$X^{(0)} = 0 \Rightarrow X^{(1)} = (1-\alpha) F$$

$$X^{(2)} = \alpha S X^{(1)} + (1-\alpha)F = \alpha S X^{(1)} + X^{(1)}$$

$$= [\alpha S + I] X^{(1)} \qquad [\alpha S + I] X^{(t)}$$

$$X^{(3)} = \alpha S X^{(2)} + (1-\alpha) F X^{(2)}$$

$$= [\alpha S + (1-\alpha) I] X^{(2)}$$

$$\bar{X}^{(t+1)} = [\alpha S + (1-\alpha) I] X^{(t)} \qquad X^{(0)} = (1-\alpha)F$$

$X^{(2)} = \bar{X}^{(2)}$

$X^{(1)} = \bar{X}^{(1)}$

Embedding: $\bar{X}^{(3)}$

Started with $N = D^{-1/2} L D^{-1/2} = I - S$

Could use L, $D^{-1}L = I - P$

$$\bar{X}^{(t+1)} = [\alpha P + (1-\alpha) I] \bar{X}^{(t)}$$

① no coupling b/w f, g, l

② $\bar{X}^{(t+1)} = [\alpha S + (1-\alpha) I]\, \bar{X}^{(t)}$ $\quad t \to \infty$?

$\tilde{X}_i^{(t)} = \bar{X}_i^{(t)} / \|\bar{X}_i^{(t)}\|_2$ (per column)

$x^{(t+1)} = \dfrac{A x^{(t)}}{\|A x^{(t)}\|_2}$ $\quad \bar{X} \to [v_1, \dots, v_l]$

$\alpha = 1/2$

$$X^{(0)} = \frac{1}{2} F \qquad M = D^{-1/2} A D^{-1/2} + I$$

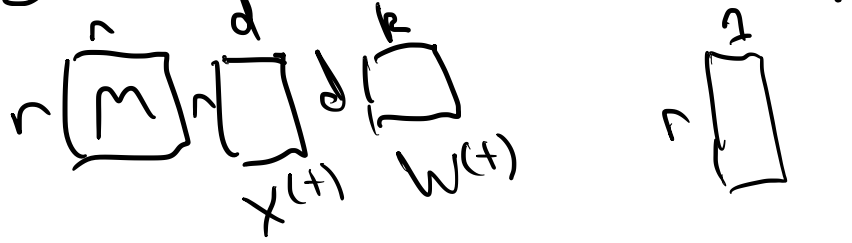
$$X^{(1)} = \frac{1}{2} [S + I] X^{(0)} \qquad \tilde{A} = A + I$$

$$M = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}$$

$$X^{(2)} = \frac{1}{2} M X^{(1)}$$

Graph convolutional network: (GCN)

$$X^{(t+1)} = \sigma\left(\frac{1}{2} M X^{(t)} W^{(t)} + b^{(t)}\right)$$



elementwise nonlinearity

$\sigma(x) = \max(x, 0)$

$\sigma(x) = \tanh(x)$

GCN (two "layers")

$$M_{(W^{(t)}, b^{(t)})}(X^{(t)}) = \sigma\left(\boxed{M} X^{(t)} W^{(t)} + b^{(t)}\right)$$

$S+I$ $\quad D^{-1/2} A D^{-1/2} + I$

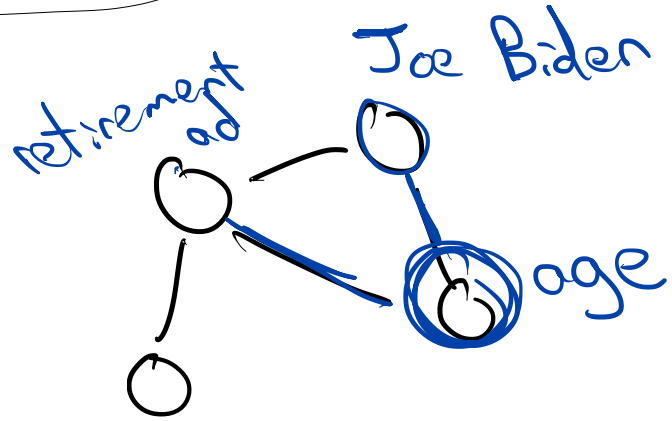
$r \boxed{F}$ features, d

$X^{(0)} = F$

$$\hat{y} = M_{(W_1, b_1)}\left(M_{(W_0, b_0)}(F)\right) Z$$

$r \times k$ $\quad k \times 1$

$$\min_{W_1, b_1, W_0, b_0, Z} \|\hat{y} - y\|_2^2$$



Why not use age as a feature itself?

$$X^{(1)} = \sigma(MFW^{(0)} + b^{(0)})$$

$F = I$ or $\hat{A}$ $\quad MFW^{(0)} = MW^{(0)}$

learn embedding directly

What if no objective?

Make one up: example: link prediction

self-supervision

Neighborhood agg. embedding