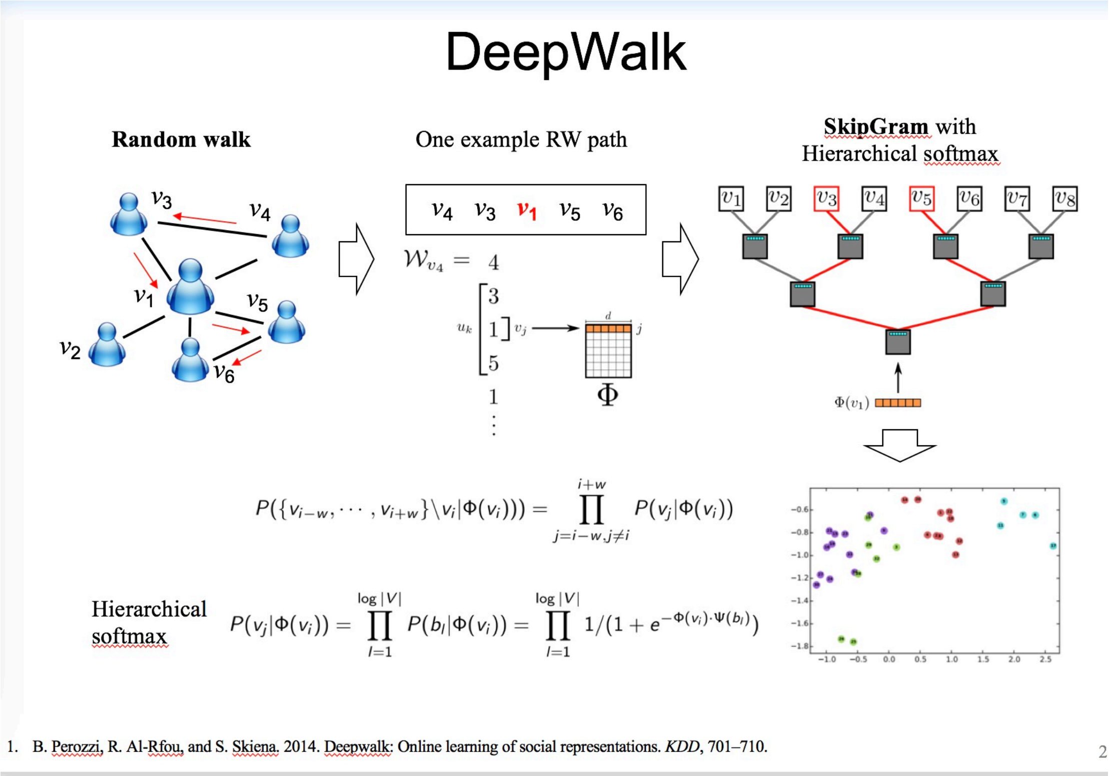
DeepWalk



Algorithm 1: DeepWalk

for $n = 1, 2, \ldots, N$ **do**
  Pick w_1^n according to a probability distribution $P(w_1)$;
  Generate a vertex sequence $(w_1^n, \cdots, w_L^n)$ of length L by a random walk on network G;
  for $j = 1, 2, \ldots, L - T$ **do**
    for $r = 1, \ldots, T$ **do**
      Add vertex-context pair (w_j^n, w_{j+r}^n) to multiset $\mathcal{D}$;
      Add vertex-context pair (w_{j+r}^n, w_j^n) to multiset $\mathcal{D}$;
Run SGNS on $\mathcal{D}$ with b negative samples.

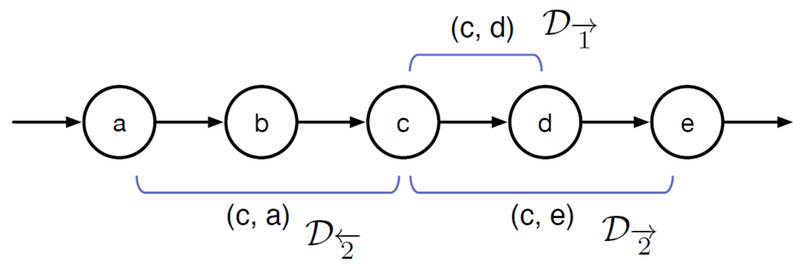
- SGNS maintains a multiset $\boldsymbol{\mathcal{D}}$ that counts the occurrence of each word-context pair (w, c)
- Objective

$$\mathcal{L} = \sum_w \sum_c (\#(w,c) \log g(x_w^T x_c) + \frac{b\#(w)\#(c)}{|\mathcal{D}|} \log g(-x_w^T x_c))$$

where x_w and x_c are d-dimenational vector

- For sufficiently large dimension d, the objective above is equivalent to factorizing the PMI matrix[1]

$$\log \frac{\#(w,c)|\mathcal{D}|}{b\#(w)\#(c)}$$



DeepWalk is factorizing a matrix

DeepWalk is asymptotically and implicitly factorizing

$$\boxed{\log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)} \qquad vol(G)=\sum_i\sum_j A_{ij}$$

$\boldsymbol{A}$ Adjacency matrix

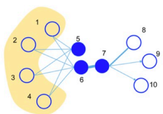
$\boldsymbol{D}$ Degree matrix

b: #negative samples

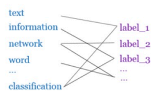
T: context window size

Later…

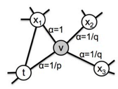
- LINE[1]: explicitly preserves both *first-order* and *second-order* proximities.



- PTE[2]: learn heterogeneous text network embedding via a semi-supervised manner.

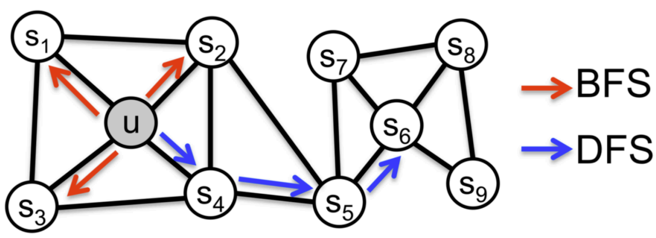


- Node2vec[3]: use a biased random walk to better explore node's neighborhood.



Node2vec

Idea: use flexible, biased random walks that can trade off between local and global views of the network.

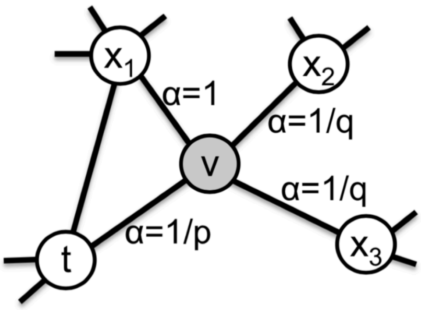


Biased random walk R that given a node u generates neighborhood $N_R(u)$

- Two parameters:
  - Return parameter p: Return back to the previous node.
  - In-out parameter q: Moving outwards (DFS) vs inwards (BFS)
  - Consider a random walk that just traversed edge (t, v) and now resides at node v. The transition probability from v to the next node x is defined as:

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

Walker is at v, where to go next?



- BFS-like walk: Low value of p
- DFS-like walk: Low value of q

LINE

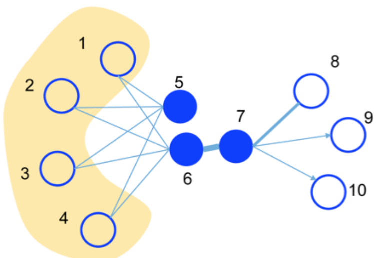


Figure 1: A toy example of information network. Edges can be undirected, directed, and/or weighted. Vertex 6 and 7 should be placed closely in the low-dimensional space as they are connected through a strong tie. Vertex 5 and 6 should also be placed closely as they share similar neighbors.

LINE with First-order Proximity

- For each undirected e_{ij}, we define the joint probability between vertex v_i and v_j as follows:

$$p_1(v_i, v_j) = \frac{1}{1 + \exp(-\vec{u}_i^T \cdot \vec{u}_j)}$$

- It defines distribution $p(\cdot, \cdot)$ over the space $V \times V$, and its empirical probability can be defined as $\hat{p}_1(i, j) = \frac{w_{i,j}}{W}$ where $W = \sum_{i,j \in E} w_{i,j}$
- Objective Function: Minimize the following **objective function**, where $d(\cdot, \cdot)$ is the distance between two distributions.

$$O_1 = d(\hat{p}_1(\cdot, \cdot), p_1(\cdot, \cdot))$$

- Instantiate $d(\cdot, \cdot)$ as KL-divergence:

$$O_1 = -\sum_{(i,j) \in E} w_{ij} \log p_1(v_i, v_j)$$

LINE with Second-order Proximity

- We first define the probability of "context" v_j generated by vertex v_i as:

$$p_2(v_j \mid v_i) = \frac{\exp(\vec{u'}_j^T \cdot \vec{u}_i)}{\sum_{k=1}^{|V|} \exp(\vec{u'}_k^T \cdot \vec{u}_i)}$$

- We minimize the following objective function:

$$O_2 = \sum_{i \in V} \lambda_i d(\hat{p}_2(\cdot \mid v_i), p_2(\cdot \mid v_i))$$

  The empirical probability $\hat{p}_2(v_j|v_i)$ defined as: $\hat{p}_2(v_j|v_i) = \frac{w_{ij}}{d_i}$ where d_i is the out-degree of vertex i.
- Replacing $d(\cdot, \cdot)$ with KL-divergence, setting $\lambda_i = d_i$, we have:

$$O_2 = -\sum_{(i,j) \in E} w_{ij} \log p_2(v_j \mid v_i)$$