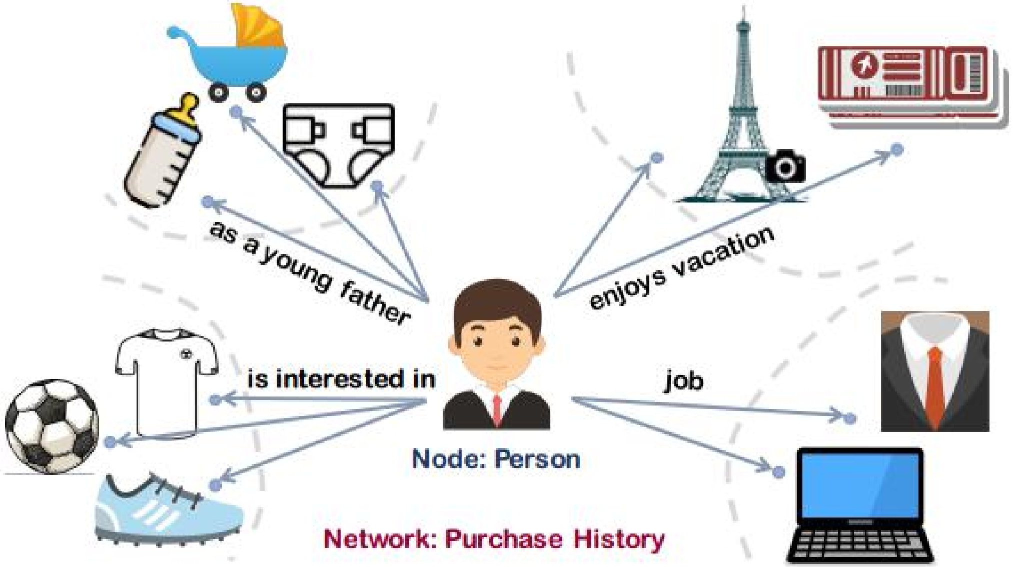
Is a Single Vector Enough? Exploring Node Polysemy for Network Embedding

Ninghao Liu,[1] Qiaoyu Tan,[1] Yuening Li,[1] Hongxia Yang,[2] Jingren Zhou,[2] Xia Hu[1]
[1]Department of Computer Science and Engineering, Texas A&M University, TX, USA
[2]Alibaba Group, Hangzhou, China
{nhliu43,qytan,liyuening,xiahu}@tamu.edu,{yang.yhx,jingren.zhou}@alibaba-inc.com

todo

- propose a polysemous(多义的) network embedding method in order to take into account multiple facets of nodes in network data



challenge

- How to determine the facets of nodes
- How to maintain the correlation among embedding vectors of different facets
- How to make the modeling process adaptive to the existing well-established base models such as Deepwalk, LINE, PTE and GCN

Polysemous Deepwalk

- each node $\mathbf{v}_i$ is associated with a target embedding matrix Ui $\in R^{Ki \times D}$ and a context embedding matrix Hi $\in R^{Ki \times D}$
- The traditional Deepwalk model could be seen as a special case where Ki = 1.

Deepwalk objective function

- Skip-gram

$$\begin{aligned}\mathcal{L}_{DW}(\theta) &= \sum_{o \in O} \log p(o|\theta) = \sum_{o \in O} \log p\big((\mathcal{N}(v_i), v_i)|\theta\big) \\ &= \sum_{o \in O} \sum_{v_j \in \mathcal{N}(v_i)} \log p(v_j|v_i),\end{aligned} \tag{1}$$

- o是random walk路径上的点，O是所有路径集合

Polysemous Deepwalk objective function

- prior knowledge denoted as $\wp$

$$\begin{aligned}\mathcal{L}_{PolyDW}(\theta) &= \sum_{o \in O} \log\ p(o|\mathcal{P}, \theta) \\ &= \sum_{o \in O} \log\ [\sum_{s(o)} p(o|s(o), \mathcal{P}, \theta) \cdot p(s(o)|\mathcal{P}, \theta)].\end{aligned} \tag{2}$$

- o是random walk路径上的点，O是所有路径集合，s(o)是路径o上的点的facet分布

$$s(o) = \{s(v|o)\ |v \in v_i \cup \mathcal{N}(v_i)\}$$

Node-Facet Assignment （get prior knowledge $\wp$）

- perform community discovery on the network

$$\min_{\mathbf{P} \geq 0} \|\mathbf{A} - \mathbf{P} \cdot \mathbf{P}^T\|_F^2 + \alpha \|\mathbf{P}\|_F^2$$

- We define the facet distribution of a node as

$$\mathbf{p}(v) = [p(1|v), \dots, p(K|v)]$$

Node-Facet Assignment of o

- $\mathbf{p}(o) = (\mathbf{p}(v_i) + \sum_{v_j \in N(v_i)} \mathbf{p}(v_j))/(|N(v_i)|+1)$

$$\mathbf{p}(v|o) = \min(\mathbf{p}(v), \mathbf{p}(o))$$

- s(o)是p(o)分布上sample的结果

Polysemous Deepwalk objective function

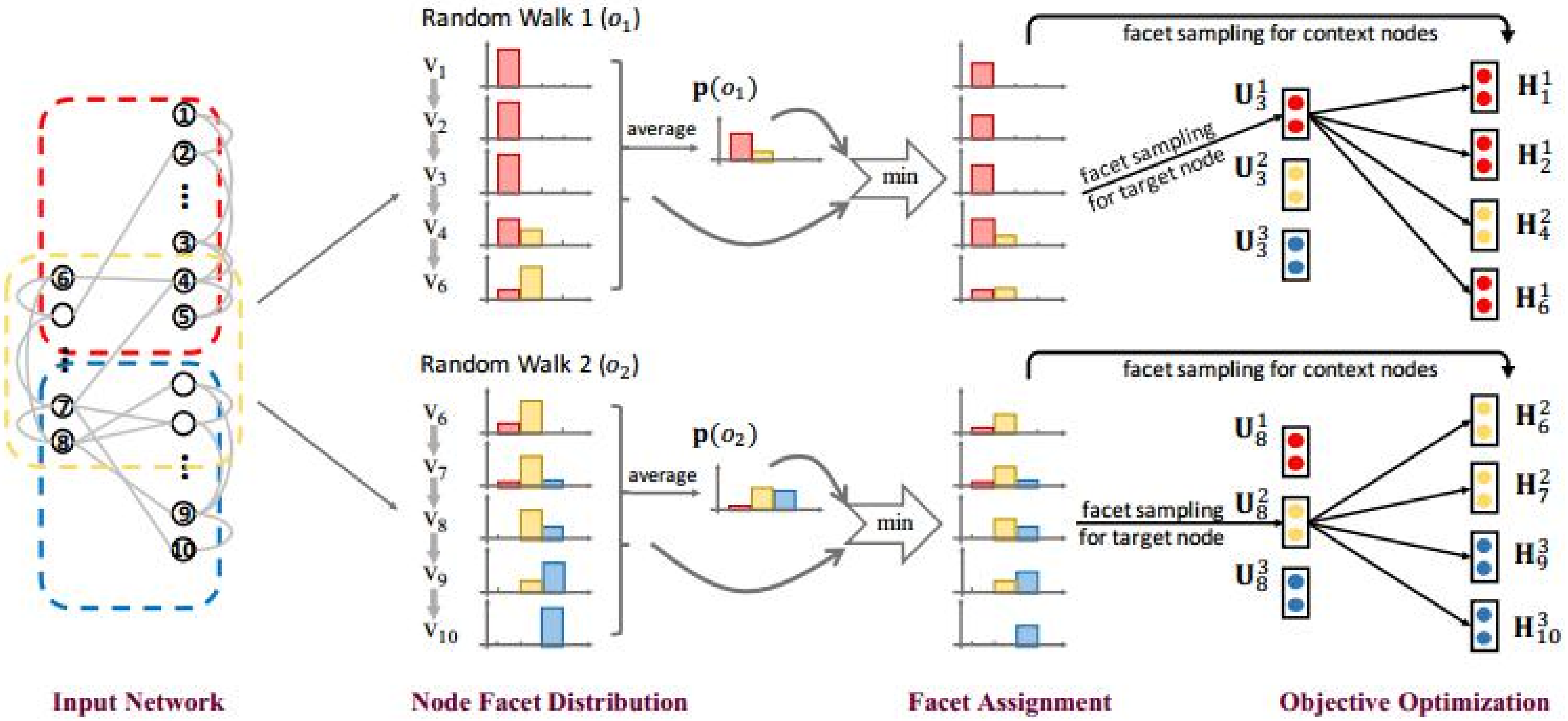
- 。。。。

$$p(o|s(o), \mathcal{P}, \theta) = \prod_{v_j \in \mathcal{N}(v_i)} p(v_j|v_i, s(o)),$$

- similar to the softmax function in traditional skip-gram models

$$p(v_j|v_i, s(o)) = \frac{\exp(< \mathbf{H}_j^{k_j}, \mathbf{U}_i^{k_i} >)}{\sum_{v,k} \exp(< \mathbf{H}_v^{k}, \mathbf{U}_i^{k_i} >)},$$

Polysemous Deepwalk



Polysemous Embedding with GCN

- 传统GCN

$$\mathbf{u}_d^k(i) \leftarrow \sigma\left(\mathbf{W}_d^k \cdot \mathrm{MEAN}(\mathbf{u}_{d-1}^k(i) \cup \{\mathbf{u}_{d-1}^k(j), \forall v_j \in \mathcal{N}^k(v_i)\})\right),$$

- 加上Polysemous Embedding

$$\mathbf{u}_d^k(i) \leftarrow \sigma\left(\mathbf{W}_d^k \cdot \mathrm{MEAN}(\mathbf{u}_{d-1}^k(i) \cup \{\mathbf{u}_{d-1}^k(j), \forall v_j \in \mathcal{N}^k(v_i)\})\right),$$