

## 科技前沿

# 基于深度学习的网络表示研究进展

张铭<sup>1</sup>, 尹伊淳<sup>1</sup>, 唐建<sup>2</sup> / 1. 北京大学 2. 微软亚洲研究院

**摘要:** 学习网络的低维网络表示, 在不同应用领域中体现出很好的效率和效果, 近年来受到了学术界和工业界的密切关注。本文对基于神经网络的网络表示方法进行了详细的介绍。这些方法可以处理现实世界中拥有百万级节点和十亿级边的网络结构, 主要考虑了网络结构信息和网络节点自身信息(如文本信息和属性信息等)。最后指出, 未来的研究方向是如何对异质网络和动态网络进行表示学习。

网络结构在现实世界中无处不在(如航线网络、通信网络、论文引用网络、世界万维网和社交网络等), 在此基础之上的应用和研究问题受到了学术界和工业界的广泛关注, 这些研究问题包括链接预测<sup>[1]</sup>、网络节点分类<sup>[2-3]</sup>、推荐<sup>[4]</sup>和异常检测<sup>[5]</sup>等。随着计算机信息技术的高速发展和迅速普及, 现实世界中的网络结构, 尤其是以 Twitter、Facebook 和 Weibo 为代表的大规模社交网络进入了亿级节点时代。除网络结构之外, 网络节点自身也会产生大量的相关信息。如今, 大规模的网络结构数据和丰富的网络节点信息对相关研究方法提出了新的挑战。

表示学习(representation learning)旨在通过无监督的方法, 将符号化的数据(symbolic data)编码到低维、连续、稠密的向量空间中<sup>[6]</sup>。相比传统的独热表示(one-hot representation), 表示学习能够抓住数据之间的相似性, 同时缓解数据稀疏性问题(data sparsity)。在网络表示学习中, 传统的基于特征向量的降维方法<sup>[7-8]</sup>需要构造亲和度矩阵(affinity matrix), 通过特征向量求解的方法获得网络节点的低维表示。特征向量求解的方法有 MDS<sup>[9]</sup>、IsoMap<sup>[10]</sup>、LLE<sup>[11]</sup>和 Laplacian Eigenmap<sup>[12]</sup>等, 这些算法的复杂度通常平方正比于网络结构中节点的数量, 所以并不适用于大规模的网络结构。最近, 基于神经网络的表示学习方法在视觉、语音识别和自然语言处理等领域获得了成功。在自然语言处理中, 文本表示学习模型(如词

表示学习<sup>[13-14]</sup>和文档的表示学习<sup>[15]</sup>)具有算法高效性的特点, 能够处理大规模文本数据, 受到了极大的关注。文献[16]分析了社交网络中随机游走的序列, 发现序列中的网络节点的频率分布类似于自然语言中词频的幂律分布。在最新的科学研究中, 文本表示学习模型被广泛用于网络表示学习模型中。

本文介绍近年内出现的有代表性的基于深度学习的网络表示模型, 其中包括: ① 考虑网络结构信息网络表示方法 DeepWalk<sup>[16]</sup>、LINE<sup>[17]</sup>和 GraRep<sup>[18]</sup>; ② 同时考虑网络结构信息和网络节点多源信息(multi-source)的网络表示方法 TADW<sup>[19]</sup>和 Multi-faceted Representations<sup>[20]</sup>, 并分析了这两类模型的特点, 指明了进一步研究的方向。文献[21]介绍了五类主要的网络表示方法, 包括基于谱方法的网络表示学习、基于最优化的网络表示学习(与任务相关属于半监督学习方法)、概率生成式模型的网络表示学习、基于力导向绘图的网络表示学习和基于深度学习的网络表示学习。与文献[21]不同的是, 本文重点关注近年来基于深度学习的网络表示方法, 所涉及的模型更加全面, 同时给出了详细的推导和分析。

## 基于网络结构信息的网络表示方法

基于网络结构信息的网络表示方法只考虑网络节点之间的链接关系。给定网络图  $G=(V, E)$ 。其中

$V$  表示网络中的节点集合； $E$  是网络中的边集合， $E \subseteq (V \times V)$ 。网络表示学习的目的在于从网络信息中学习得到各个节点的低维表示  $W \in \mathbb{R}^{|V| \times d}$ ， $d \ll |V|$  是向量的维度。

这部分分别介绍 DeepWalk、LINE 和 GraRep 三种模型。其中 DeepWalk 是以 Skip-gram 模型为基础，本质上使用了二阶的网络上下文信息；LINE 模型显示地提出了网络表示方法的目标函数，考虑了一阶和二阶的上下文信息；GraRep 对 LINE 模型进行了拓展，可以对节点的任意阶上下文信息建模。

### DeepWalk 模型

DeepWalk 模型首先采用随机游走 (random walk) 的方法产生标准的输入序列，然后使用 Skip-gram 模型对序列建模得到网络节点表示 (具体算法见表 1)。随机游走首先基于均匀分布得到序列的起始点，然后从当前点的邻居节点中随机选择一点作为后续节点，依次迭代直到产生特定长度的序列。

表 1 DeepWalk 算法

DeepWalk( $G, s, d, \gamma, t$ ) 算法
输入: $G = (V, E)$ , 窗口大小 $s$ , 表示向量维度 $d$ , 迭代次数 $\gamma$ , 随机游走步长 $t$ 。
输出: $W \in \mathbb{R}^{ V  \times d}$
1: 基于均匀分布初始化 $W \in \mathbb{R}^{ V  \times d}$
2: <b>for</b> $i = 0$ <b>to</b> $\gamma$ <b>do</b>
3: $O = \text{Shuffle}(V)$
4: <b>for each</b> $v_i \in O$ <b>do</b>
5: $\text{Seq}_{v_i} = \text{RandomWalk}(G, v_i, t)$
6: $\text{Skip-gram}(W, \text{Seq}_{v_i}, s)$
7: <b>end for</b>
8: <b>end for</b>

相比基准的模型方法 (Spectral Clustering<sup>[7]</sup>、Modularity<sup>[8]</sup>、EdgeCluster<sup>[22]</sup>、wvRN<sup>[23]</sup>)，DeepWalk 模型有效地解决了训练数据稀疏的问题，在训练数据较少的情况下， $F_1$  值上有 10% 的提高。在一些标准数据集中，仅使用 60% 的训练数据 DeepWalk 模型就可以超过使用 100% 训练数据的所有基准方法。

### LINE 模型

文献 [17] 提出了一种适用于不同类别网络图

结构 (有向图、无向图和加权图) 的网络学习模型 LINE。具体上，LINE 模型从一阶相似性 (first-order proximity) 和二阶相似性 (second-order proximity) 两方面设计目标函数。基于一阶或者二阶相似性，LINE 模型可以分别学习到一种网络表示。为了同时使用这两种相似性，LINE 模型将一阶节点向量和二阶节点向量拼接起来作为最终的节点表示。

一阶相似性表示网络中两个节点之间的点对相似性，具体为节点之间边的权重 (如果点对不存在边，则其一阶相似性为 0)。为了建模一阶相似性，模型首先定义点对  $v_i$  和  $v_j$  联合概率为

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

其中  $\mathbf{v}_i \in \mathbb{R}^d$  和  $\mathbf{v}_j \in \mathbb{R}^d$  分别是节点  $v_i$  和节点  $v_j$  的向量表示。节点  $v_i$  和  $v_j$  的经验联合概率为  $\hat{p}_1(v_i, v_j) = \frac{\text{wei}_{i,j}}{\text{WEI}}$ ， $\text{wei}_{i,j}$  表示边  $(i, j)$  上的权重， $\text{WEI} = \sum_{(i,j) \in E} \text{wei}_{i,j}$ 。一阶相似性模型通过最小化概率分布  $p_1(*, *)$  与  $\hat{p}_1(*, *)$  的 KL 距离来获得网络表示。

二阶相似性模型假设如果节点间共享相似的邻居节点，那么两者就趋于相似。具体上，点对之间的二阶相似性表示两个节点在整个网络上的一阶相似性的分布相似度 (如果点对没有共同的相邻节点，则二阶相似性为 0)。在这种情况下，每个节点有目标节点和其他节点的上下文两个角色。形式上，用  $\mathbf{v}_i$  和  $\mathbf{v}'_i$  分别指  $v_i$  作为目标节点的表示和  $v_i$  作为其他节点上下文的表示。二阶相似性模型首先定义节点  $v_i$  和  $v_j$  的条件概率为

$$p_2(v_j | v_i) = \frac{\exp(\mathbf{v}'_j{}^T \cdot \mathbf{v}_i)}{\sum_{k=1}^{|V|} \exp(\mathbf{v}'_k{}^T \cdot \mathbf{v}_i)} \quad (2)$$

节点  $v_i$  和  $v_j$  的经验条件概率  $\hat{p}_2(v_j | v_i) = \frac{\text{wei}_{i,j}}{d_i}$ ，其中  $d_i$  是节点  $v_i$  的出度。通过最小化概率分布  $p_2(* | *)$  与  $\hat{p}_2(* | *)$  之间的 KL 距离来获得二阶相似性模型的网络表示。

一阶相似性和二阶相似性模型都采用了基于边的负采样优化方法来得到网络节点表示。实验表明 LINE 模型在语言网络、社交网络和论文引用网络的数据集上均超过了 DeepWalk 模型和基于随机梯度的矩阵分解方法<sup>[24]</sup>。

### GraRep 模型

文献 [18] 指出 LINE 模型中的一阶相似性和二阶相似性分别捕捉到节点间一阶和二阶的局部信息 (如图 1 (a) 和 (b) 所示), 并在此基础上提出更一般化的模型 GraRep。GraRep 模型可以捕捉更高阶的网络信息 (如图 1 (c) 和 (d) 所示), 并对每一阶的局部信息分别建模, 最后串接各阶网络表示得到最终节点表示。

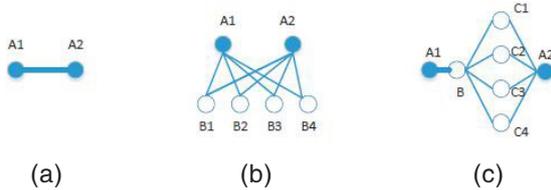


图 1 网络结构中节点 A1 与 A2 之间  $k$  阶信息 [18]。图 (a) ~ (d) 分别表示  $k=1, 2, 3$  和 4, 线的粗细表示边权重大小

GraRep 模型基于概率转移矩阵来获得网络表示。首先定义一阶概率转移矩阵  $A=D^{-1}S$ , 其中  $S$  为邻接矩阵 ( $S_{ij}=wei_{ij}$ )、 $D$  为度对角矩阵 (degree matrix)。所得到的  $A_{ij}$  是节点  $v_i$  到  $v_j$  的一阶转移概率。进一步, 通过计算  $A^k = \underbrace{A \cdots A}_k$  可以得到  $k$  阶概率转移矩阵。GraRep 模型优化目标在于最大化  $(c, w)$  对的出现概率, 同时最小化随机产生的  $(c', w)$  出现的概率, 其中  $w$  为目标词、 $c$  是  $w$  的上下文词、 $c'$  是随机得到的上下文词。采用负采样的方法建模  $k$  阶信息, 考虑  $(c, w)$  的出现概率, 最大化的目标函数为

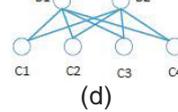
$$L_{k\text{-step}}(c, w) = p_k(c|w) \log \sigma(c \cdot w) + \lambda p_k(c) \log \sigma(-c \cdot w) \quad (3)$$

其中  $p_k(c|w) = A_{wc}^k$  表示从  $w$  到  $c$  的  $k$  步转移概率;  $\sigma(\cdot)$  是 sigmoid 函数;  $\lambda$  是负例的个数; 上下文词  $c$  出现的概率为  $p_k(c) = \frac{1}{N} \sum_w A_{wc}^k$ 。根据文献 [25], 优化上述式子本质上等价于将矩阵  $Y$  分解成  $W$  和  $C$ , 其中  $W$  的每一行代表节点的表示, 而  $C$  中的每一列表示节点作为上下文的表示。

GraRep 模型采用 SVD 矩阵分解的方法来得到网络节点的表示。相比 DeepWalk 和 LINE 模型, GraRep 模型考虑了更高阶的上下文信息, 在网络结构数据上得到了更好的效果。值得一提的是, 虽然在文献 [18] 中, GraRep 模型使用了复杂度较高的 SVD 矩阵分解的方法, 但它也可以采用随机梯度下降的优化方法, 因此该模型同样适用于大规模的网络结构。

### 基于网络结构和节点信息的网络表示方法

除了节点之间的网络结构信息, 网络节点本身往往存在丰富的信息。比如, 在维基百科中的文章连接形成的信息网络中, 每篇文章作为一个节点, 节点包含了丰富的文本信息; 在社交网络中 (如图 2 所示), 每个用户节点包含用户产生的文本内容及用户属性 (如性别、学校、地点、公司等)。



这部分介绍两种同时考虑网络结构和节点信息的模型: TADW 和 Multi-faceted Representations。Multi-faceted Representations

模型考虑与节点  $v_i$  相关的信息为  $I_i = \{\text{text}_i, G_i, R_i, M_i\}$ 。其中  $\text{text}_i$  表示节点  $v_i$  产生的文本内容, 由单词序列组成  $\text{text}_i = \{w_1, w_2, \dots\}$ ;  $G_i$  表示与  $v_i$  相关的网络结构;  $R_i$  表示  $v_i$  与属性实体之间的关系集合 (如图 2 中的 Like、LivesIn 和 StudyAt 等);  $M_i$  表示用户  $v_i$  拥有的属性实体集合 (如图 2 中 StarTrek、Boston 和 Harvard 等)。TADW 模型则只考虑了网络结构信息和节点的文本信息  $I_i = \{\text{text}_i, G_i\}$ 。

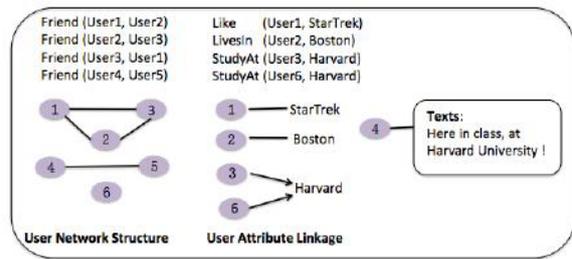


图 2 社交网络中与网络节点相关的信息 [20]

### TADW 模型

文献 [19] 证明 DeepWalk 模型等价于矩阵分解并给出了待分解矩阵的具体形式。为了建模网络节点自身的文本特征, TADW(text-associated DeepWalk) 采用了文献 [26] 中诱导矩阵填充 (inductive matrix completion) 的方法。

从文献 [25] 可知, 使用 hierarchical softmax 优化 Skip-gram 模型等价于分解矩阵  $Y$ ,  $Y$  中的元素为

$$Y_{ij} = W_i \cdot C_j = \log \frac{N(v_i, c_j)}{N(v_i)} \quad (4)$$

其中  $N(v_i, c_j)$ 、 $N(v_i)$  和  $N(c_j)$  分别表示  $(v_i, c_j)$ 、节点  $v_i$ 、上下文节点  $c_j$  出现在训练语料中的个数。在 DeepWalk 上下文节点  $c_j$  出现在  $v_i$  的左边或者右边的上下文期望次数为  $\frac{2s \cdot N(v_i, c_j)}{N(v_i)}$ ，其中  $s$  是设定的窗口大小。文献 [19] 指出 DeepWalk 模型的优化过程本质上是在分解矩阵  $Y$ ，其中  $Y$  中每个元素为

$$Y_{ij} = W_i \cdot C_j = \log \frac{N(v_i, c_j)}{N(v_i)} = \log \frac{[e_i(A + A^2 + \dots + A^s)]_j}{s} \quad (5)$$

其中  $e_i$  是一个  $|V|$  维的向量，在  $i$  维上为 1，其余为 0。进而可知， $e_i A$  是从节点  $v_i$  出发经过一步到达网络上各节点的概率分布， $[e_i(A + A^2 + \dots + A^s)]_j$  是上下文节点  $c_j$  出现在  $v_i$  一边窗口（左边或者右边）的期望次数。

在此基础上，TADW 采用了诱导矩阵填充方法，同时对文本特征和网络结构建模，求解的目标函数为

$$\min_{W, H} \|Y - W^T H T\|_F^2 + \frac{\alpha}{2} (\|W\|_F^2 + \|H\|_F^2) \quad (6)$$

其中  $T$  的每一列代表一个节点的文本特征（例如 TF-IDF 值），在求解中保持不变。模型得到的  $W$  和  $HT$  作为网络节点和上下文节点的低维表示串接起来，作为最终的网络表示。在进行矩阵分解之前首先要获得  $Y$ ，当  $s$  变大矩阵分解时间复杂度为  $O(|V|^3)$ 。一般情况下， $t$  越大模型效果越好，但计算越费时。在 TADW 中， $t$  取 2。

TADW 模型在 Cora、Citeseer、Wiki 三个标准数据集的表现超过了不用文本特征的 DeepWalk 模型、只用文本特征的 PLSA<sup>[27]</sup> 模型，以及简单将两模型所得表示串接起来的方法。这说明 TADW 模型能够有效地融合文本信息获得更好的节点表示。

### Multi-faceted Representations(多方面表示)模型

文献 [20] 关注社交网络中节点表示学习，除了节点的网络链接信息之外，还考虑真实存在的多源信息（见图 2），包括用户产生的文本内容和用户本身的属性背景等。多方面表示模型目标在于学习得到每个用户、每个属性实体和每个用户属性关系的低维向量表示。给定用户  $v_i$ ，整个模型最大化  $I_i = \{\text{text}_i, G_i, R_i, M_i\}$  的出现概率为

$$P(I_i | v_i; \Theta) = P(\text{text}_i | v_i; \Theta_v) \times P(G_i | v_i; \Theta_v) \times P(R_i, M_i | v_i; \Theta_v, \Theta_r, \Theta_m) \quad (7)$$

其中  $\Theta_v, \Theta_r, \Theta_m$  分别表示用户、用户属性实体关系和实体的潜在表示矩阵。模型假设词的代表已经提前学习得到，因此整个模型的参数为  $\Theta = \{\Theta_v, \Theta_r, \Theta_m\}$ 。

$P(\text{text}_i | v_i; \Theta_v)$  借助于段落向量 (paragraph vector)<sup>[15]</sup> 的思想建模，使用用户级别的表示和邻接词来预测文本中的目标词。公式如下

$$P(\text{text}_i | v_i; \Theta_v) = \sum_{w \in \text{text}_i} \log p(w | v_i) + \sum_{w \in \text{text}, w' \in C(w)} \log p(w | w') \quad (8)$$

模型假设词级别的表示是预先学习好的，因此式 (8) 后一项被看作常数项。模型使用 AUC 损失函数（间隔排序损失 margin ranking loss）的方法对函数的第一部分进行优化，使得  $(w, v_i)$  配对的分值大于随机产生的  $(w', v_i)$  配对分值，其中  $w'$  是从词典  $C(w)$  中负采样得到的词。

$P(G_i | v_i; \Theta_v)$  定义为用户  $v_i$  的链接节点的似然，借助于 Skip-gram 模型的思想，使用用户节点来预测其链接的节点。最后，采用 AUC 损失函数的方法进行优化。

$P(R_i, M_i | v_i; \Theta_v, \Theta_r, \Theta_m)$  目标在于预测用户  $v_i$  是否与属性实体  $m \in M_i$  具有  $r \in R_i$  的关系。首先使用双线性模型对每个元组  $r(v_i, m_i)$  建模打分，然后采用 AUC 损失函数对模型进行优化。

多方面表示模型使用随机梯度下降的方法对整个模型进行优化。通过将不同信息建模到统一空间，模型学习得到的网络表示，不仅可以对用户之间的链接关系预测，同时也可以预测用户本身的属性（比如性别、工作、位置等）。

## 网络表示学习的应用

由于基于神经网络的网络表示方法的高效性，它被越来越多地应用到除传统应用场景（如网络节点分类、推荐和链接预测等）之外的其他方面（如文本建模<sup>[17,28]</sup> 和可视化处理<sup>[29]</sup>）。

文本语料可以表示为一个带权重的网络结构（节点表示词，边权重表示词与词之间共现的程度），因此网络的表示方法同时也可以应用到文本建模中。LINE<sup>[17]</sup> 模型一个重要的应用就是学习词的向量表示，相比现在流行的 Skip-gram 词向量模型，它具有效率更高和效果更好的特点。在文献 [17] 中，作者指出 LINE 模型使用的语言网络结构相比 Skip-

gram 模型使用的文本原始的词序列结构, 能够更好地抓住词之间的全局共现信息。另外, 当数据量很大时, LINE 模型的训练效率更高。为了引入标签信息来指导文本表示的学习, 文献 [28] 在 LINE 模型的基础上, 提出了针对文本的半监督表示学习模型 PTE ( Predictive Text Embedding )。除了使用基本的词与词的共现网络信息, PTE 模型还考虑了词与文档和词与标签的网络信息。相比有监督的神经网络模型 CNN(Convolutional neural network) 和无监督表示模型 Paragraph vector 和 LINE, PTE 在长文本语料和短文本语料都取得了更好的效果; 并且, PTE 模型的参数在不同数据集上更加稳定。文献 [29] 基于 LINE 模型获得数据表示, 提出了对高维和大规模的数据进行可视化处理的新方法。

## 总结与展望

本文从建模的信息种类上, 论述了近年来有代表性的基于深度学习的网络表示学习模型。

基于网络上下文信息的网络表示方法中, LINE 模型使用了一阶和二阶网络局部信息, 拓展的 GraRep 模型可以对任意阶局部信息建模来获得网络表示。值得注意的是, 在得到最终的网络表示方

法上, LINE 和 GraRep 模型都是对每一阶的局部信息分别建模, 然后再将各阶表示向量拼接起来作为最终的表示。这两个模型都没有采用对各阶局部信息统一建模的思路, 同时也没有比较各阶网络上上下文信息对网络表示的重要性。

基于网络结构和节点信息的网络表示方法中, 这些模型可以看作是第一部分模型的拓展。从节点内部信息的种类来说, TADW 模型考虑了节点自身产生的文本信息; Multi-faceted Representations 模型考虑了文本信息和节点自身的属性信息。

由于基于深度学习的网络表示还处于起步阶段, 建模的信息, 相比实际网络结构中存在的信息相当有限。一方面, 在实际的网络结构 ( 尤其是社交网络 ) 中, 网络节点还存在图像、视频和音频等内容信息。如何统一地建模多模态 ( multi-modal ) 信息以及探讨不同的信息对网络表示的重要性会是网络表示学习的一个重要的研究方向。更进一步, 实际网络往往由不同的类型的节点 ( 如文本、图像和视频 ) 构成, 同时网络节点具有动态变化的特点, 如何对异质网络和动态网络进行表示学习将会是一个非常具有挑战性的研究方向。

## 参考文献

- [1] Liben-Nowell D and Kleinberg J. The link prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019-1031, 2007.
- [2] Bhagat S, Cormode G, and Muthukrishnan S. Node classification in social networks. In *Social Network Data Analytics*, pages 115-148. Springer, 2011.
- [3] Sen P, Namata G, Bilgic M, Getoor L, Galligher B, and Eliassi-Rad T. Collective classification in network data. *AI magazine*, 29(3):93, 2008.
- [4] Yu X, Ren X, Sun Y, Gu Q, Sturt B, Khandelwal U, Norick B and Han J. Personalized entity recommendation: A heterogeneous information network approach. In *Proceedings the 7th ACM international conference on Web search and data mining*, pages 283-292. ACM, 2014.
- [5] Bhuyan M, Bhattacharyya D, and Kalita J. Network anomaly detection: methods, systems and tools. *IEEE Communications Surveys*, 16(1):303-336, 2014.
- [6] Bengio Y, Courville A, and Vincent P. Representation learning: A review and new perspectives. 2013.
- [7] Tang L and Liu H. Leveraging social media networks for classification. *Data Mining and Knowledge Discovery*, 23(3):447-478, 2011.
- [8] Tang L and Liu H. Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD, KDD '09*, pages 817-826, New York, NY, USA, 2009. ACM.
- [9] Cox T F and Cox M A. *Multidimensional scaling*. CRC Press, 2000.
- [10] Tenenbaum J B, De-Silva V, and Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319-2323, 2000.
- [11] Roweis S T and Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323-2326, 2000.
- [12] Belkin M and Niyogi P. Laplacian Eigen maps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585-591, 2001.

- [13] Mikolov T, Chen K, Corrado G, and Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- [14] Mikolov T, Sutskever I, Chen K, Corrado G S, and Dean J. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems, pages 3111-3119, 2013.
- [15] Le Q V, Mikolov T. Distributed Representations of Sentences and Documents. ICML, 2014.
- [16] Perozzi B, Al-Rfou R, and Skiena S. Deepwalk: Online learning of social representations. In Proceedings of SIGKDD, pages 701-710, 2014.
- [17] Tang J, Qu M, Wang M, Zhang M, Yan J, and Mei Q. LINE: Large-scale information network embedding. In WWW. ACM, 2015.
- [18] Cao S, Lu W and Xu Q. GraRep: Learning Graph Representations with Global Structural Information. CIKM, pages 891-900, 2015.
- [19] Yang C, Liu Z, Zhao D, Sun M and Chang E. Network Representation Learning with Rich Text Information. International Joint Conference on Artificial Intelligence (IJCAI) 2015.
- [20] Li J, Ritter A and Jurafsky D. Learning multi-faceted representations of individuals from heterogeneous evidence using neural networks. arXiv: <http://arxiv.org/pdf/1510.05198v1.pdf>.
- [21] 陈维政, 张岩, and 李晓明. “网络表示学习.” 大数据 1.3 (2015): 2015025.
- [22] Tang L and Liu H. Scalable learning of collective behavior based on sparse social dimensions. In Proceedings of the 18th ACM conference on Information and knowledge management, pages 1107-1116. ACM, 2009.
- [23] Macskassy S A and Provost F. A simple relational classifier. In Proceedings of the Second Workshop on Multi-Relational Data Mining (MRDM-2003) at KDD-2003, pages 64-76, 2003.
- [24] Ahmed A, Shervashidze N, Narayanamurthy S, Josifovski V, and Smola A J. Distributed large-scale natural graph factorization. In Proceedings of the 22nd international conference on World Wide Web, pages 37-48. International World Wide Web Conferences Steering Committee, 2013.
- [25] Levy O and Goldberg Y. Neural word embedding as implicit matrix factorization. In Proceedings of NIPS, pages 2177-2185, 2014.
- [26] Natarajan N and Dhillon I S. Inductive matrix completion for predicting gene-disease associations. Bioinformatics, 30(12):i60-i68, 2014.
- [27] Hofmann T. Probabilistic latent semantic indexing. In Proceedings of SIGIR, pages 50-57, 1999.
- [28] Tang J, Qu M, Mei Q Z. PTE: predictive text embedding through large-scale heterogeneous text networks. Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 2015.
- [29] Tang J, Liu J, Zhang M and Mei Q. Visualizing Large-scale and High-dimensional Data. WWW'16.

基金项目：国家自然科学基金项目（61272343，61472006）；“973”项目（2014CB340405）。



### 张铭

北京大学信息科学技术学院教授，ACM Education Council 中国委员，兼任中国 ACM 教育专委会主席。目前主持两项国家自然科学基金、一项教育部博士点基金在研项目，合作发表学术论文 100 多篇，与博士生唐建一起获得 ICML2014 最佳论文奖。获得 ICML2014 最佳论文奖。出版学术专著 1 部，获软件著作权 6 项，获发明专利 3 项。主持的《数据结构与算法》被评选为国家级和北京市级精品课程，获北京市精品教材奖并得到国家“十二五”规划教材支持。主要研究方向为文本挖掘、社会网络分析、教育大数据等。



### 尹伊淳

北京大学信息科学技术学院博士研究生，主要研究方向为自然语言处理和深度学习。



### 唐建

2014 年获北京大学博士学位，目前为微软亚洲研究院机器学习组的副研究员。主要研究方向为深度学习、统计主题模型，以及这些方法在自然语言理解、网络分析、用户行为分析等领域的应用。