

网络表示学习综述：一文理解Network Embedding

本文是一篇来自 DeepWalk 团队的综述文章，对于近几年网络表示学习（Network Representation Learning/Network Embedding）进行了一个阶段性的总结，并对于未来的发展方向进行了研究。

论文摘要

网络嵌入方法（Network Embedding）旨在学习网络中节点的低维度潜在表示，所学习到的特征表示可以用作基于图的各种任务的特征，例如分类，聚类，链路预测和可视化。

在本文中，通过分类和总结本研究领域的最新进展来概述网络嵌入学习相关进展。文章首先讨论网络嵌入的属性特征，并简要介绍了网络嵌入学习的历史。然后文章还讨论了在不同场景下的网络嵌入方法，如监督与无监督学习，同质网络学习嵌入与异构网络等。文章进一步论证了网络嵌入学习方法的具体应用，并总结了该领域未来的工作研究。

Network Embedding 介绍

由于信息网络可能包含数十亿个节点和边缘，因此在整个网络上执行复杂的推理过程可能会非常棘手。因此有人提出了用于解决该问题的一种方法是网络嵌入（Network Embedding）。NE 的中心思想就是找到一种映射函数，该函数将网络中的每个节点转换为低维度的潜在表示。

总的来说，NE 具有如下几个特征：

适应性 (Adaptability) - 现实的网络在不断发展；新的应用算法不应该要求不断地重复学习过程。

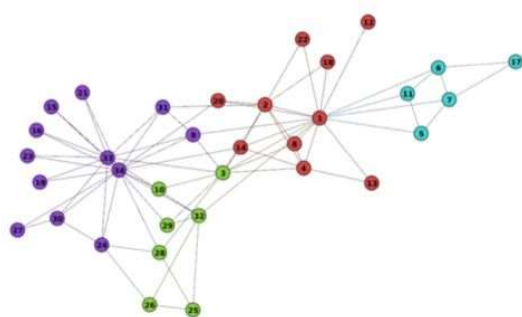
可扩展性 (Scalability) - 真实网络本质上通常很大，因此网络嵌入算法应该能够在短时间内处理大规模网络。

社区感知 (Community aware) - 潜在表示之间的距离应表示用于评估网络的相应成员之间的相似性的度量。这就要求同质网络能够泛化。

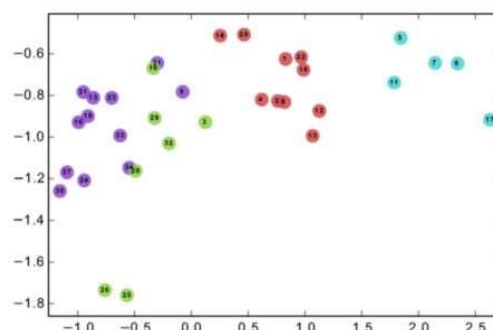
低维 (Low dimensional) - 当标记数据稀缺时，低维模型更好地推广并加速收敛和推理。

持续 (Continuous) - 需要潜在的表示来模拟连续空间中的部分社区成员资格。

一个典型的例子就是 DeepWalk。其学习 Zachary' s Karate network 网络中的拓扑结构信息并转换成一个二维的潜在表示 (latent representation) 。



(a) Input: Karate Graph



(b) Output: Network Embedding

Network Embedding 简史

传统意义上的 Graph Embedding 被看成是一个降维的过程，而主要的方法包括主成分分析 (PCA) 和多维缩放 (MDS) 。所有的方法都可以理解成运用一个 $n \times k$ 的矩阵来表示原始的 $n \times m$ 矩阵，其中 $k \ll n$ 。

在 2000 年代早期，又提出了其他方法，如 IsoMap 和 LLE，以保持非线性流形的整体结构。

总的来说，这些方法都在小型网络上提供了良好的性能。然而这些方法的时间复杂性至少是二次的，这使得它们无法在大规模网络上运行。

另一类流行的降维技术使用可从图中导出的矩阵的光谱特性（例如，特征向量）来嵌入图的节点。拉普拉斯特征映射（Laplacian eigenmaps）通过与其 k 个最小非平凡特征值相关联的特征向量表示图中的每个节点。

Deep Learning

DeepWalk [1] 是第一个被提出来使用表示学习（或深度学习）社区的技术的网络嵌入方法。 DeepWalk 通过将节点视为单词并生成短随机游走作为句子来弥补网络嵌入和单词嵌入之间的差距。然后，可以将诸如 Skip-gram 之类的神经语言模型应用于这些随机游走以获得网络嵌入。

DeepWalk 的优点可以概括为：首先其可以**按需生成随机游走**。由于 Skip-gram 模型也针对每个样本进行了优化，因此随机游走和 Skip-gram 的组合使 DeepWalk 成为在线算法。其次，**DeepWalk 是可扩展的**，生成随机游走和优化 Skip-gram 模型的过程都是高效且平凡的并行化。最重要的是，**DeepWalk 引入了深度学习图形的范例。**

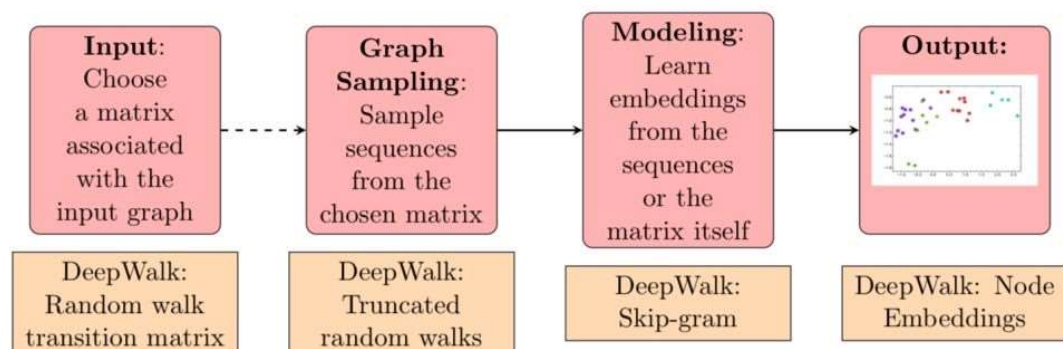


Figure 3: A paradigm for *Deep Learning On Graphs* with DeepWalk's design choice for each building block.

Unsupervised Network Embeddings

Method	Source of Context Nodes	Embedding Learning Method
DeepWalk [33]	Truncated Random Walks	Skip-gram with Hierarchical Softmax
LINE [41]	1-hop and 2-hop Neighbors	Skip-gram with Negative Sampling
Node2vec [21]	Biased Truncated Random Walks	Skip-gram with Negative Sampling
Walklets [34]	A^i where $i = 1, 2, \dots, k$	Skip-gram with Hierarchical Softmax
GraRep [5]	A^i where $i = 1, 2, \dots, k$	Matrix Factorization
GraphAttention [2]	A^i where $i = 1, 2, \dots, k$	Graph Likelihood
SDNE [48]	1-hop and 2-hop Neighbors	Deep Autoencoder
DNGR [6]	Random surfing	Stacked Denoising Autoencoder

Table 1: Unsupervised network embedding methods categorized by source of context nodes and method for representation learning.

LINE [2] 采用广度优先搜索策略来生成上下文节点：只有距离给定节点最多两跳的节点才被视为其相邻节点。此外，与 DeepWalk 中使用的分层 softmax 相比，它使用负采样来优化 Skip-gram 模型。

Node2vec [3] 是 DeepWalk 的扩展，它引入了一个偏向的随机步程序，结合了 BFS 风格和 DFS 风格的邻域探索。

Walklets [4] 显示 DeepWalk 从 $A \sim 1 \sim, A \sim 2 \sim, \dots, A \sim k \sim$ 的加权组合中学习网络嵌入。特别是如果 $i < j$ ，DeepWalk 总是偏向于 $A \sim i \sim$ 而不是 $A \sim j \sim$ 。为了避免上述缺点，Walklets 建议从 $A \sim 1 \sim, A \sim 2 \sim, \dots, A \sim k \sim$ 中学习多尺度网络嵌入。由于计算 $A \sim i \sim$ 的时间复杂度至少是网络中节点数量的二次方，因此 Walklet 通过在短随机游走中跳过节点来近似 $A \sim i \sim$ 。它进一步学习来自 A 的不同权力的网络嵌入，以不同的粒度捕获网络的结构信息。

GraRep [5] 类似地通过将图形邻接矩阵提升到不同的幂来利用不同尺度的节点共现信息。将奇异值分解（SVD）应用于邻接矩阵的幂以获得节点的低维表示。

Walklet 和 GraRep 之间存在两个主要差异。首先，GraRep 计算 $A \sim i \sim$ 的确切内容，而 Walklets 接近它。其次，GraRep 采用 SVD 来获得具有精确分解的节点嵌入，而 Walklet 使用 Skip-gram 模型。

有趣的是，Levy 和 Goldberg 证明带负抽样的跳过法（SGNS）隐含地将节点和各个上下文节点之间的 PMI 矩阵分解。总而言之，**GraRep 使用噪声较小的过程生成网络嵌入，但 Walklet 证明更具可扩展性。**

GraphAttention [6] 提出了一种 attention 模型，它可以学习多尺度表示，最好地预测原始图中的链接。 GraphAttention 不是预先确定超参数来控制上下文节点分布，而是自动学习对图转换矩阵的幂集的 attention。

SDNE [7] 学习节点表示，通过深度自动编码器保持 2 跳邻居之间的接近度。 它通过最小化它们的表示之间的欧几里德距离来进一步保持相邻节点之间的接近度。

DNGR [8] 是另一种基于深度神经网络的学习网络嵌入的方法。 他们采用随机冲浪策略来捕获图形结构信息。他们进一步将这些结构信息转换为 PPMI 矩阵，并训练堆叠去噪自动编码器 (SDAE) 以嵌入节点。

Attributed Network Embeddings

上述无监督网络嵌入方法仅利用网络结构信息来获得低维度的网络特征。但是现实世界网络中的节点和边缘通常与附加特征相关联，这些特征称为属性 (attribute)。例如在诸如 Twitter 的社交网络站点中，用户 (节点) 发布的文本内容是可用的。因此期望网络嵌入方法还从节点属性和边缘属性中的丰富内容中学习。

TADW [9] 研究节点与文本特征相关联的情况。 TADW 的作者首先证明了 DeepWalk 实质上是将转移概率矩阵 M 分解为两个低维矩阵。受此结果的启发，TADW 包含文本特征矩阵 T 通过将 M 分解为 W ， H 和 T 的乘积，进入矩阵分解过程。最后，将 W 和 $H \times T$ 连接起来作为节点的潜在表示。

CENE [10] 是一种网络嵌入方法，它共同模拟节点中的网络结构和文本内容。 CENE 将文本内容视为特殊类型的节点，并利用节点-节点链接和节点内容链接进行节点嵌入。优化目标是共同最小化两种类型链路的损失。

HSCA [11] 是一种归因图的网络嵌入方法，它同时模拟同音，网络拓扑结构和节点特征。

Maxmargin DeepWalk (MMDW) [12] 是一种半监督方法，它学习部分标记网络中的节点表示。 MMDW 由两部分组成：第一部分是基于矩阵分解的节点嵌入模型，第二部分是

学习的表示作为特征来训练标记节点上的最大边缘 SVM 分类器。通过引入偏置梯度，可以联合更新两个部分中的参数。

Heterogeneous Network Embeddings

异构网络具有多类节点或边缘。为了模拟不同类型的节点和边缘，大多数异构网络嵌入方法通过联合最小化每种模态的损失来学习节点嵌入。这些方法要么直接在相同的潜在空间中学习所有节点嵌入，要么事先为每个模态构建嵌入，然后将它们映射到相同的潜在空间。

Chang [13] 提出了异构网络的深度嵌入框架。他们的模型首先为每种模态（如图像，文本）构建一个特征表示，然后将不同模态的嵌入映射到同一个嵌入空间。优化目标是最大化链接节点的嵌入之间的相似性，同时最小化未链接节点的嵌入。注意，边可以在相同模态内的两个节点之间以及来自不同模态的节点之间。

Zhao [14] 是另一种用于在异构网络中构造节点表示的框架。具体来说，他们认为维基百科网络有三种类型的节点：实体，单词和类别。建立相同和不同类型节点之间的共现矩阵，并且使用坐标矩阵分解从所有矩阵联合学习实体，词和类别的表示。

Li [15] 提出了一种神经网络模型，用于学习异构社交网络中的用户表示。他们的方法联合模拟用户生成的文本，用户网络和用户与用户属性之间的多方面关系。

HEBE [16] 是一种嵌入大规模异构事件网络的算法，其中事件被定义为网络中一组节点（可能是不同类型）之间的交互。虽然先前的工作将事件分解为事件中涉及的每对节点之间的成对交互，但 HEBE 将整个事件视为超边界并同时保留所有参与节点之间的接近度。

具体而言，对于超边缘中的每个节点，HEBE 将其视为目标节点，并将超边界中的其余节点视为上下文节点。因此，基础优化目标是在给定所有上下文节点的情况下预测目标节点。

EOE [17] 是用于耦合异构网络的网络嵌入方法，其中两个同构网络与网络间边缘连接。EOE 学习两个网络的潜在节点表示，并利用和谐嵌入矩阵将不同网络的表示转换为相同的空间。

Metapath2vec [18] 是 **DeepWalk** 的扩展，适用于异构网络。为了构建随机漫游，Metapath2vec 使用基于元路径的漫游来捕获不同类型节点之间的关系。对于来自随机游走序列的学习表示，他们提出异构 Skip-gram，其在模型优化期间考虑节点类型信息。

总结

该 Network Embedding 综述文章较为系统地阐述了目前 NE 的发展现状，并从 Unsupervised Network Embeddings, Attributed Network Embeddings 和 Heterogeneous Network Embeddings 等几个部分进行了介绍。本笔记主要着重于介绍现有的一系列方法，对于其在不同场景的应用不做详细阐述。

参考文献

- [1] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 701–710. ACM, 2014.
- [2] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Largescale information network embedding. In Proceedings of the 24th International Conference on World Wide Web, pages 1067–1077. ACM, 2015.
- [3] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 855–864. ACM, 2016.
- [4] Bryan Perozzi, Vivek Kulkarni, Haochen Chen, and Steven Skiena. Don' t walk, skip! online learning of multi-scale network embeddings. In 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE/ACM, 2017.
- [5] Shaosheng Cao, Wei Lu, and Qionghai Xu. Grarep: Learning graph representations with global structural information. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pages 891–900. ACM, 2015.
- [6] Sami Abu-El-Hajja, Bryan Perozzi, Rami Al-Rfou, and Alex Alemi. Watch your step: Learning graph embeddings through attention. arXiv preprint arXiv:1710.09599, 2017.
- [7] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1225–1234. ACM, 2016.

- [8] Shaosheng Cao, Wei Lu, and Qionghai Xu. Deep neural networks for learning graph representations. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, pages 1145–1152. AAAI Press, 2016.
- [9] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information. In IJCAI, pages 2111–2117, 2015.
- [10] Xiaofei Sun, Jiang Guo, Xiao Ding, and Ting Liu. A general framework for content-enhanced network representation learning. arXiv preprint arXiv:1610.02906, 2016.
- [11] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. Homophily, structure, and content augmented network representation learning. In Data Mining (ICDM), 2016 IEEE 16th International Conference on, pages 609–618. IEEE, 2016.
- [12] Cunchao Tu, Weicheng Zhang, Zhiyuan Liu, and Maosong Sun. Max-margin deepwalk: discriminative learning of network representation. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI 2016), pages 3889–3895, 2016.
- [13] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. Heterogeneous network embedding via deep architectures. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 119–128. ACM, 2015.
- [14] Yu Zhao, Zhiyuan Liu, and Maosong Sun. Representation learning for measuring entity relatedness with rich information. In IJCAI, pages 1412–1418, 2015.
- [15] Jiwei Li, Alan Ritter, and Dan Jurafsky. Learning multi-faceted representations of individuals from heterogeneous evidence using neural networks. arXiv preprint arXiv:1510.05198, 2015.
- [16] Huan Gui, Jialu Liu, Fangbo Tao, Meng Jiang, Brandon Norick, and Jiawei Han. Large-scale embedding learning in heterogeneous event data. 2016.
- [17] Linchuan Xu, Xiaokai Wei, Jiannong Cao, and Philip S Yu. Embedding of embedding (eoe): Joint embedding for coupled heterogeneous networks. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pages 741–749. ACM, 2017.
- [18] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 135–144. ACM, 2017.

● **END** ●

关于作者：薛寒生，澳大利亚国立大学博士生，研究方向为人工智能与计算生物学。

■ 论文 | A Tutorial on Network Embeddings

■ 链接 | <https://www.paperweekly.site/papers/2203>

■ 作者 | Haochen Chen / Bryan Perozzi / Rami Al-Rfou / Steven Skiena