

网络表示学习模型汇总

1. 早期

早期一般作为降维算法的一种，很多用来可视化。典型的有[局部线性嵌入](#)算法，基于[图的矩阵分解](#)等。

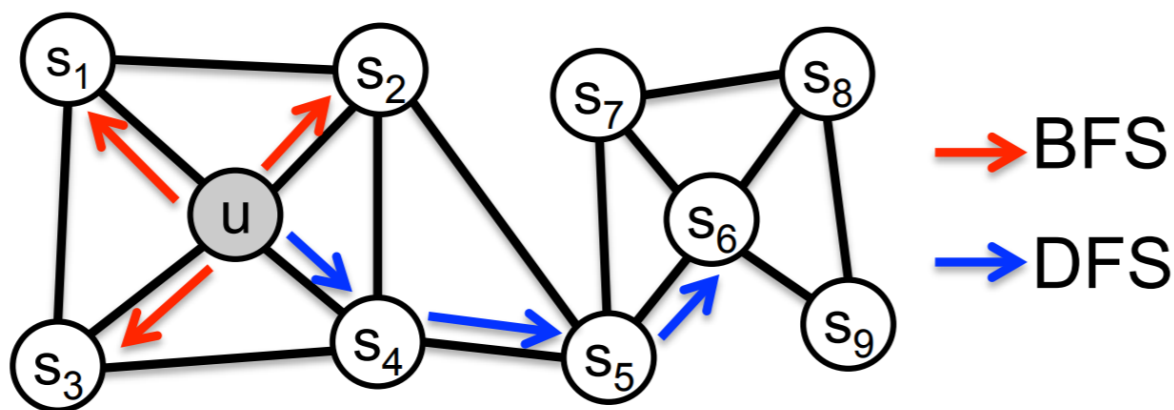
2. 发展

后来，受到自然语言处理领域词嵌入技术的启发，在2014年提出了[DeepWalk](#)算法，之后这个领域开始百花齐放。

2.1 考虑网络结构信息

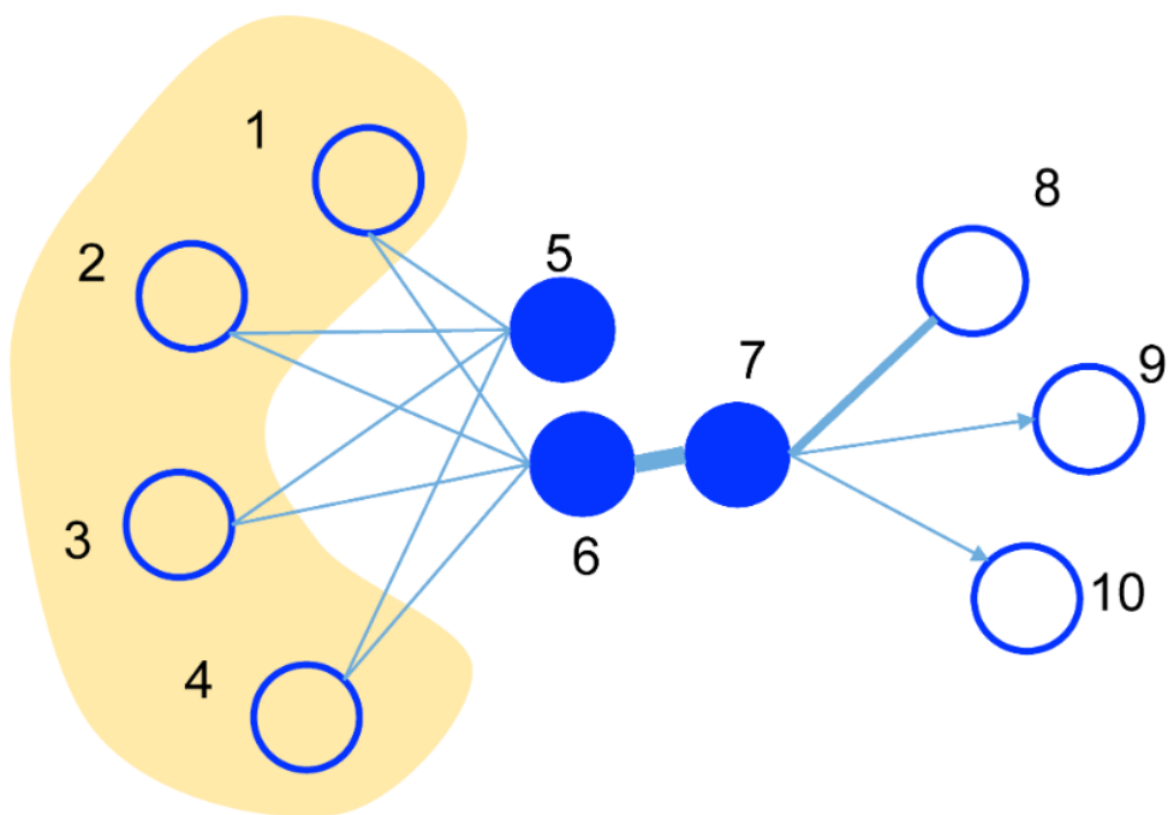
[DeepWalk, 2014](#)的灵感来自于自然语言处理中的词嵌入技术，作者将图中节点类比于句子中的单词，采用随机游走的策略构造一个共现节点的序列，然后类比词嵌入技术中的成本函数，对节点的嵌入也就是向量表示进行优化，最终得到每个节点的嵌入表示，即节点向量。DeepWalk的基本假设是高概率共同出现的节点在嵌入空间中的向量表示也更相似或者说它们的位置更靠近。DeepWalk第一次将自然语言处理中的词嵌入思想注入到图嵌入技术中，DeepWalk的策略实际上是一种深度优先搜索。

[node2vec, 2016](#)则是对DeepWalk的进一步改良。深度优先搜索的方法可以找到并捕获节点的邻域中某种潜在的相似关系，但是广度优先搜索的方法更容易发现和捕获图中结构上更加相似的节点。node2vec模型中综合考虑了深度优先搜索和广度优先搜索的优势，设计了一种二阶随机游走的方式。它通过 p, q 两个参数来平衡深度优先和广度优先两种搜索策略。相比DeepWalk，它能够通过调节 p, q 两个权重参数的大小捕获图中节点的局部相似性。



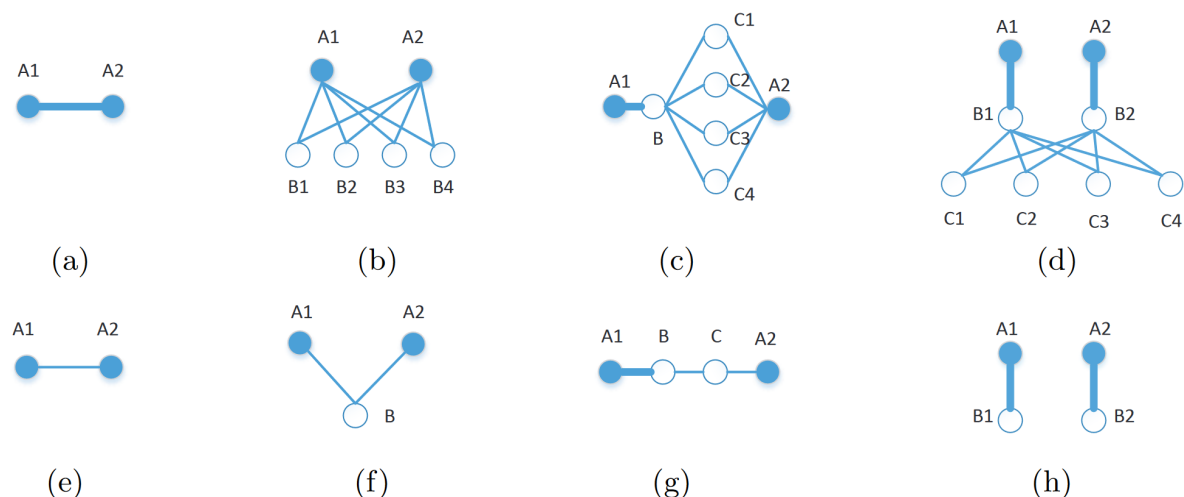
从节点 u 开始的BFS和DFS策略 ($k=3$) 图来自论文[node2vec](#)

[LINE, 2015](#)是在2015年提出的一种大规模信息网络嵌入方法，作者在论文中提出了两个度量节点局部结构的概念：一阶相似度和二阶相似度。一阶相似度用来描述直接相连的节点之间的相似性；二阶相似度则考虑两个节点之间共享的一阶邻居节点的数量，以此描述两个节点间的二阶相似度。论文中提出了一种有效的负采样优化技术，在之后的很多网络嵌入方法中都有使用。LINE也基于邻域相似的假设，某种角度上说，捕获二阶相似性时使用了广度优先搜索的方法。LINE可以应用于大规模网络，对于有向图和无向图都有不错的效果。



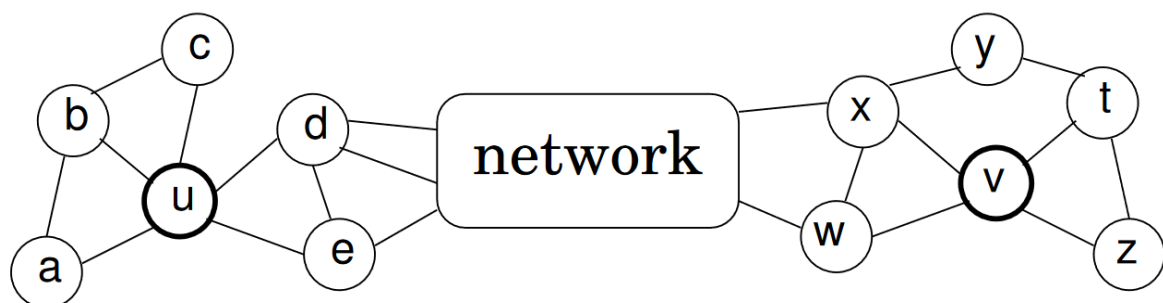
边既可以是有向的也可以是无向的，既可以是加权的也可以是无权的。节点6和节点7应该映射到低维空间靠近的位置上，因为它们是强连接的。节点5和节点6也应该映射到相靠近的位置，因为它们共享了许多相同的邻居节点。图来自论文[LINE](#)

[GraRep, 2015](#)在LINE定义的一阶和二阶相似性的启发下，将这种相似性推广到更高阶，定义了 k 阶相似性。GraRep中也使用了像LINE中一样将高阶的图中的相似关系在低维向量空间中用条件概率表示，并且使用了负采样优化的策略。



在图表示问题中捕获不同的 k 步信息的重要性 ($k=1,2,3,4$) 图来自论文[GraRep](#)

[struc2vec, 2017](#)考虑了在图中相距很远的节点。如果它们的邻域空间结构很相似，那么它们在嵌入空间中的向量表示也应该很相似。基于这种想法，作者设计了一个用于描述两个节点的邻域节点组成的序列相似性的函数。作者使用分层定义结构相似的方法，在近距离的两个节点之间关系紧密的情况下，函数会扩大感受的范围，在更高阶的层次度量两个节点之间的相似性。还是通过随机游走构建序列，通过参数自动调节函数是否扩大感受域，以在整个图中寻找和捕获结构相似的节点。struc2vec是在node2vec的基础上，向前迈出的一大步。



两个节点 (u和v) 的结构相似 (度数分别为5和4, 分别连接3和2个三元环, 都通过两个节点连接到网络的其余部分), 但是在这个网络中相距很远。图来自论文[struc2vec](#)

小结: 以上方法均为具有代表性的网络嵌入方法, 这些方法最早从自然语言处理领域的词嵌入技术和Skip-gram策略中获得灵感, 同时期还有很多网络嵌入算法也受到这些思想的启发。以上模型都只考虑了网络中的结构信息, 即只需要网络的结构信息就能学习到网络的嵌入表示, 但是当网络的结构发生变化时, 需要重新学习新的嵌入表示。

2.2 在网络结构信息的基础上考虑网络的属性信息

为什么考虑网络的属性信息? 真实世界中的网络通常在节点和边上具有丰富的属性信息。对于某些网络信息挖掘任务, 节点和边的上的属性信息不可或缺。如果能够高效地利用这些属性信息, 将其考虑到网络表示学习的模型中, 应该能得到更优质的网络表示。

[TADW, 2015](#), [CENE, 2016](#), [SIGNet, 2017](#), [ANRL, 2018](#)和[SANE, 2018](#)等是针对属性网络的表示学习方法, 它们结合网络结构和节点的属性信息使学习到的嵌入表示更具有信息性。[BiasedWalk, 2018](#)则考虑兼顾有偏随机游走的优势以及网络节点的属性信息, 以学习到信息更加丰富的网络嵌入表示。

[ABRW, 2018](#)考虑了真实世界的网络结构信息完整性问题, 提出一种属性偏差随机游走的算法, 以利用属性信息通过转移矩阵补偿不完整结构信息。而[attri2vec, 2019](#)考虑到真实世界的网络中节点属性不完整的可能以及节点属性特征空间和网络结构空间之间的差异, 通过在原始属性空间上执行网络结构引导变换以更一致的方式表示网络结构, 以学习更高质量的节点表示。

针对于现实世界中广泛存在的异构网络, 也有很多人做出了探索, 如[HNE, 2015](#), [HEBE, 2016](#), [EOE, 2017](#)等。

2.3 其他一些方法

[SDNE, 2016](#)是一种半监督的深度学习模型, 它具有很多层非线性函数, 从而可以捕获高度非线性的网络结构。它利用了深度学习中的自编码器方法保持节点的二阶邻居节点的接近程度, 通过最小化其表示的欧几里得空间距离保证相邻接点的接近程度。

[ANG, 2017](#), [GraphGAN, 2017](#)和[GraphSGAN, 2018](#)的灵感来自近几年非常流行的[GAN, 2014](#), 以对抗式的方式生成更接近原始网络结构信息的表示。

[NetMF, 2018](#)把DeepWalk, LINE, PTE和node2vec统一到了具有封闭形式的矩阵分解框架中。其作者首先阐述了以上四种方法的关联性, 然后进一步提供了基于skip-gram的网络嵌入算法和拉普拉斯矩阵算法之间的理论联系, 最后提出了NetMF方法用于网络嵌入。

[Splitter, 2019](#)通过学习得到每个节点的多个向量表示, 以更好地描述网络节点的信息。[SNE, 2017](#)和[SINE, 2018](#)提出了对单个图的嵌入表示方法。

[BANE, 2018](#)和[MCNE, 2019](#)考虑了大型网络对计算机存储和计算成本带来的挑战, 以从输入网络中学习紧凑的嵌入表示。[RandNE, 2018](#)以高斯随机投影的方式将网络映射到低维的嵌入空间中, 同时保留节点之间的高阶邻接。该方法减少了计算的时间复杂度, 同时可以使用分布式计算方案加速以处理十亿级的网络。

3. 图神经网络

图神经网络是一类基于深度学习的处理图结构数据的方法, 图神经网络的动机源于卷积神经网络 (CNNs) 和图嵌入 (Graph Embedding) [38]。一方面, CNN的广泛应用带来了机器学习领域的突破, 并开启了深度学习的新时代, 图神经网络的重点研究问题之一就是如何将卷积操作应用到图结构数据上。另一方面, DeepWalk、LINE、SDNE等方法在网络表示学习领域取得了很大的成功。然而, 这些方法在计算上较为复杂并且在大规模的图上往往并不是最优的, 图神经网络旨在解决这些问题。

——[Graph Neural Networks: A Review of Methods and Applications](#)

图神经网络是一种基于深度学习方法的神经网络, 用于处理图结构的数据。

神经网络是一种端到端的模型, 图神经网络即可以处理图数据的神经网络, 可以取出神经网络的中间层特征作为网络的表示。

图神经网络可以充分利用网络的结构信息和属性信息, 从而在图域中提取深层的 (高级的) 网络特征。

网络嵌入和图神经网络有着非常紧密的关系。某种角度来说，图神经网络的研究源于网络嵌入的研究，但是图神经网络展现出了越来越强大的表示和学习能力。可以说，图神经网络是从网络表示学习领域成长起来的一个极具潜力的新方向。