论文阅读报告

一、论文基本情况

论文名	LINE: Large-scale Information Network Embedding				
作者	Jian Tang(1),Meng Qu(2),Mingzhe Wang(2), Ming Zhang(2), Jun Yan(1), Qiaozhu Mei(3)				
作者机构	1 Microsoft Research Asia 2 School of EECS, Peking University 3 School of Information, University of Michigan				
发表年份	2015				
来源（刊物名、会议名）	WWW				
论文可信力	期刊	EI/SCI		影响因子	
		Jcr 分区		CCF 分类	
	会议	CCF 分类	A	领域影响力	
	引用次数	325			
阅读人	房慧文	阅读日期		2019.01.10	

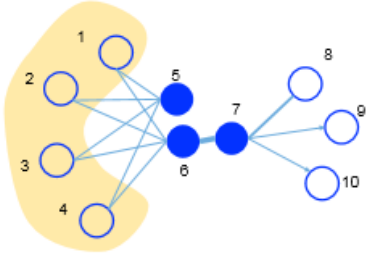
二、论文主要工作

1. 论文概述

该论文针对大多数现有的 graph embedding 方法不能用于通常包含数百万个节点的真实世界信息网络的问题，提出一种新的 network embedding 方法，称为“LINE”，适用于任意类型的信息网络：无向、有向和有权、无权。该方法优化了精心设计的目标函数，能够保留局部和全局网络结构。提出了边缘采样算法，解决了经典随机梯度下降的局限性，提高了算法的有效性和效率。实验证明了 LINE 在各种现实世界信息网络（包括语言网络，社交网络和引用网络）上的有效性。该算法非常高效，能够在典型的单机上在几个小时内学习数百万个顶点和数十亿条边的网络。

2. 论文提出的方法

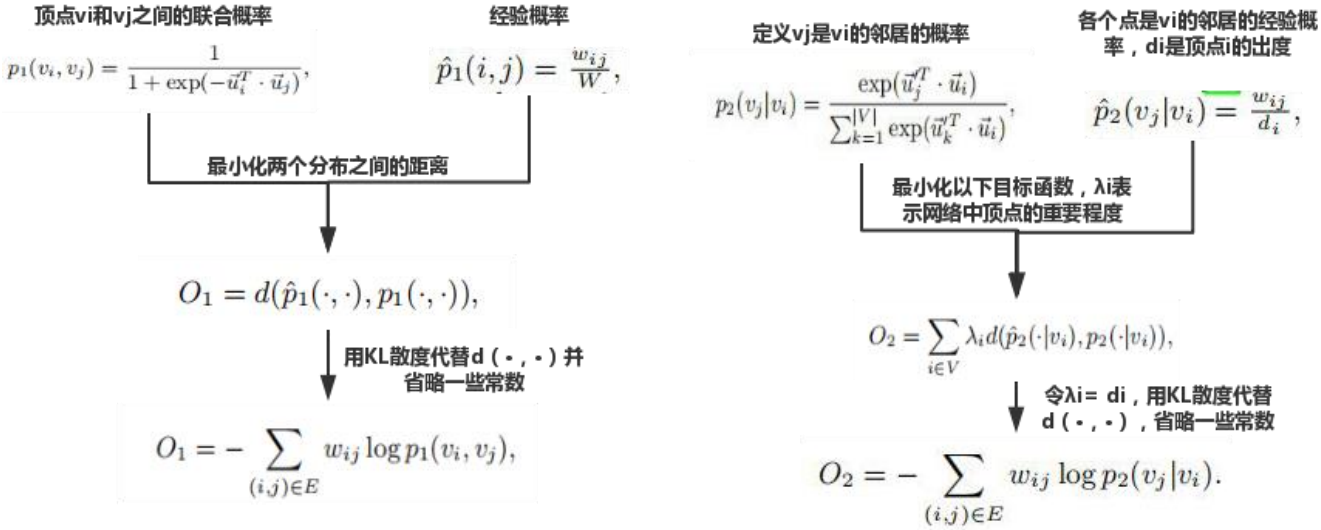
首先提出一阶相似以及二阶相似的定义，如下图，顶点 6 和 7 之间的边的权重大，即 6 和 7 有高一阶相似性，所以他们在被嵌入的低维空间中应该接近；另一方面，虽然顶点 5 和 6 之间没有联系，但他们有着许多共同的邻居，即他们有高二阶相似性，因此，在嵌入的空间中也应该接近。



输入输出：输入是网络图，输出是网络图中节点的向量表示。
适用范围：大规模（百万的顶点和数十亿的边）的任意类型的网络：有向或无向、有权或无权只保留一阶相似度的 LINE 模型：最小化 $O_1 = -\sum_{(i,j)\in E} w_{ij} \log p_1(v_i, v_j),$

只保留二阶相似度的 LINE 模型：最小化 $O_2 = -\sum_{(i,j)\in E} w_{ij} \log p_2(v_j|v_i).$

结合一阶相似度和二阶相似度：采用分别训练一阶相似度模型和二阶相似度模型，然后将学习的两个向量表示连接成一个更长的向量。



模型优化：
1. 负采样。在计算条件概率 p2（ • | vi）时需要对整个顶点集进行求和，计算量太大。采用负采样方法，根据每个边（i，j）的噪声分布对多个负边进行采样。
2. 边采样。采用异步随机梯度下降算法（ASGD）优化 。梯度计算过程中，要乘以边的权重，当权重的变化范围很大时，梯度变化大。为了解决这一问题，可以将有权边展开为多个无权边。展开后，会显著增加内存需求。解决方案为先在网络中，对边进行采样，采样的概率与边的权重成正比；再将采样后的边展开成无权边。

3. 论文仿真实验及结果

论文分别对语言网络、社交网络、引用网络进行了仿真，在这里只分析了社交网络的仿真实验及结果。

数据集：使用两个社交网络：Flickr 和 Youtube。 其中，Flickr 网络比 Youtube 网络（DeepWalk 中使用的网络）更密集。
比较算法：
Graph factorization (GF)：只适用于无向网络。
DeepWalk：只适用于无权边，仅利用二阶相似度。

LINE-SGD：没有经过模型优化的 LINE 算法，包括只考虑一阶相似度（LINE-SGD(1st)）和只考虑二阶相似度（LINE-SGD(2nd)）的两个变体。其中，LINE-SGD(1st)只适用于无向图，LINE-SGD(2nd)适用于各种图。
LINE：经过模型优化后的 LINE 算法，同样包括 LINE(1st)和 LINE(2nd)两个变体。其中，LINE(1st)：只适用于无向图，LINE(2nd)适用于各种图。
LINE (1st+2nd)：同时考虑一阶相似度和二阶相似度。将由 LINE（1st）和 LINE（2nd）学习得到的两个向量表示，连接成一个更长的向量。在连接之后，对维度重新加权以平衡两个表示。因为在无监督的任务中，设定权重很困难，所以只应用于监督学习的场景。
Skip-Gram：只能用于语言网络，因此只在对语言网络进行实验时参与对比。
实验结果：

Table 5: Results of multi-label classification on the FLICKR network.

Metric	Algorithm	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1	GF	53.23	53.68	53.98	54.14	54.32	54.38	54.43	54.50	54.48
	DeepWalk	60.38	60.77	60.90	61.05	61.13	61.18	61.19	61.29	61.22
	DeepWalk(256dim)	60.41	61.09	61.35	61.52	61.69	61.76	61.80	61.91	61.83
	LINE(1st)	63.27	63.69	63.82	63.92	63.96	64.03	64.06	64.17	64.10
	LINE(2nd)	62.83	63.24	63.34	63.44	63.55	63.55	63.59	63.66	63.69
	LINE(1st+2nd)	**63.20****	**63.97****	**64.25****	**64.39****	**64.53****	**64.55****	**64.61****	**64.75****	**64.74****
Macro-F1	GF	48.66	48.73	48.84	48.91	49.03	49.03	49.07	49.08	49.02
	DeepWalk	58.60	58.93	59.04	59.18	59.26	59.29	59.28	59.39	59.30
	DeepWalk(256dim)	59.00	59.59	59.80	59.94	60.09	60.17	60.18	60.27	60.18
	LINE(1st)	62.14	62.53	62.64	62.74	62.78	62.82	62.86	62.96	62.89
	LINE(2nd)	61.46	61.82	61.92	62.02	62.13	62.12	62.17	62.23	62.25
	LINE(1st+2nd)	**62.23****	**62.95****	**63.20****	**63.35****	**63.48****	**63.48****	**63.55****	**63.69****	**63.68****

Significantly outperforms DeepWalk at the: ** 0.01 and * 0.05 level, paired t-test.

社交网络（Flicker 网络）：
1.评估方法：与语言网络相比，社交网络更加稀疏，因此，将每个节点分配到一个或多个社区，通过多标签分类任务来评估算法。 实验中，随机抽样不同比例的顶点进行训练，其余部分用于评估，结果在 10 次不同的运行中取平均值。
2.实验结果分析：
LINE（1st + 2nd）再次胜过所有其他方法。通过连接 LINE(1st)和 LINE(2nd)的表示，性能进一步提高，证实两个相似是互补的
LINE(1st)略好于 LINE(2nd)，这与语言网络上的结果相反。原因有两点：(1)社交网络中的一阶相似性比二阶相似性更重要; (2)当网络太稀疏,节点的平均邻居数量太小时，二阶邻近可能变得不准确。
LINE(1st)胜过图分解。这表明更好的模拟一阶相似性。
LINE(2nd)胜过 DeepWalk。这表明更好的建模二阶相似。

Table 6: Results of multi-label classification on the YOUTUBE network. The results in the brackets are on the reconstructed network, which adds second-order neighbors (i.e., neighbors of neighbors) as neighbors for vertices with a low degree.

Metric	Algorithm	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Micro-F1	GF	25.43 (24.97)	26.16 (26.48)	26.60 (27.25)	26.91 (27.87)	27.32 (28.31)	27.61 (28.68)	27.88 (29.01)	28.13 (29.21)	28.30 (29.36)	28.51 (29.63)
	DeepWalk	39.68	41.78	42.78	43.55	43.96	44.31	44.61	44.89	45.06	45.23
	DeepWalk(256dim)	39.94	42.17	43.19	44.05	44.47	44.84	45.17	45.43	45.65	45.81
	LINE(1st)	35.43 (36.47)	38.08 (38.87)	39.33 (40.01)	40.21 (40.85)	40.77 (41.33)	41.24 (41.73)	41.53 (42.05)	41.89 (42.34)	42.07 (42.57)	42.21 (42.73)
	LINE(2nd)	32.98 (36.78)	36.70 (40.37)	38.93 (42.10)	40.26 (43.25)	41.08 (43.90)	41.79 (44.44)	42.28 (44.83)	42.70 (45.18)	43.04 (45.50)	43.34 (45.67)
	LINE(1st+2nd)	39.01* (**40.20**)	41.89 (**42.70**)	43.14 (**43.94****)	44.04 (**44.71****)	44.62 (**45.19****)	45.06 (**45.55****)	45.34 (**45.87****)	45.69** (**46.15****)	45.91** (**46.33****)	46.08** (**46.43****)
Macro-F1	GF	7.38 (11.01)	8.44 (13.55)	9.35 (14.93)	9.80 (15.90)	10.38 (16.45)	10.79 (16.93)	11.21 (17.38)	11.55 (17.64)	11.81 (17.80)	12.08 (18.09)
	DeepWalk	28.39	30.96	32.28	33.43	33.92	34.32	34.83	35.27	35.54	35.86
	DeepWalk (256dim)	28.95	31.79	33.16	34.42	34.93	35.44	35.99	36.41	36.78	37.11
	LINE(1st)	28.74 (29.40)	31.24 (31.75)	32.26 (32.74)	33.05 (33.41)	33.30 (33.70)	33.60 (33.99)	33.86 (34.26)	34.18 (34.52)	34.33 (34.77)	34.44 (34.92)
	LINE(2nd)	17.06 (22.18)	21.73 (27.25)	25.28 (29.87)	27.36 (31.88)	28.50 (32.86)	29.59 (33.73)	30.43 (34.50)	31.14 (35.15)	31.81 (35.76)	32.32 (36.19)
	LINE(1st+2nd)	**29.85** (29.24)	31.93 (**33.16****)	33.96 (**35.08****)	35.46** (**36.45****)	36.25** (**37.14****)	36.90** (**37.69****)	37.48** (**38.30****)	38.10** (**38.80****)	38.46** (**39.15****)	38.82** (**39.40****)

Significantly outperforms DeepWalk at the: ** 0.01 and * 0.05 level, paired t-test.

社交网络（YouTube 网络）：
因为 YouTube 网络非常稀疏，除了在原图上进行实验外，还重构了新的网络进行实验。在重构图中，为了丰富节点的邻居，使用广度优先搜索策略扩展每个顶点的邻域，即递归地添加邻居的邻居，直到扩展邻域的大小达到 1000 个节点。
实验结果分析：
原图中，LINE(1st)大多数情况下优于 LINE(2nd)。原因与 Flicker 网络中一样。
原图中，LINE(2nd)的性能逊色于 DeepWalk.这是因为网络太稀疏。
原图中，LINE(1st+2nd)的性能优于 DeepWalk。表明两个近似是相互补充的，能够解决网络问题稀疏。
重构网络中，GF，LINE(1st)和 LINE(2nd)的表现都有所提高，特别是 LINE(2nd)，LINE（1st + 2nd）的性能并没有太大的提高。这意味着原始网络上一阶和二阶相似的组合已经获得了大部分信息，LINE(1st + 2nd)方法对于网络嵌入是一个非常有效和高效的方法，适用于密集和稀疏网络。
重构网络中，LINE（2nd）在大多数情况下胜过 DeepWalk。

三、存在问题

1. 自己对该论文不太理解的地方

目标函数的推算过程（使用 KL 散度部分）

2. 该论文方法本身的问题

文章没有实现共同训练一阶相似度和二阶相似度的目标函数，只是分别训练后连接成一个更长的向量。

四、论文评价

1. 论文亮点

同时考虑了一阶相似度和二阶相似度；适用于各种网络

2. 评价

领域内经典论文