

Don't Walk, Skip!

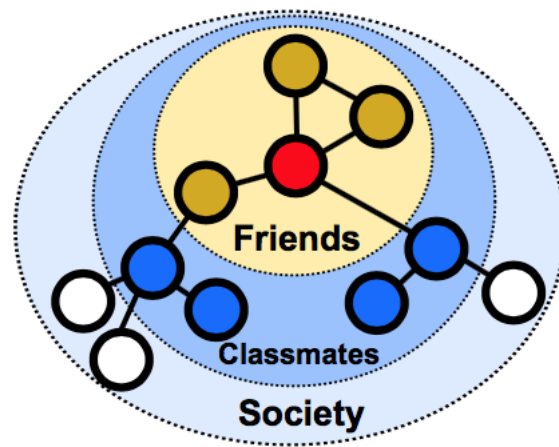
Online Learning of Multi-scale Network Embeddings.

Deepwalk作者对Deepwalk的改进

IEEE/ACM International Conferences on Advances in Social Network Analysis and Mining 2017

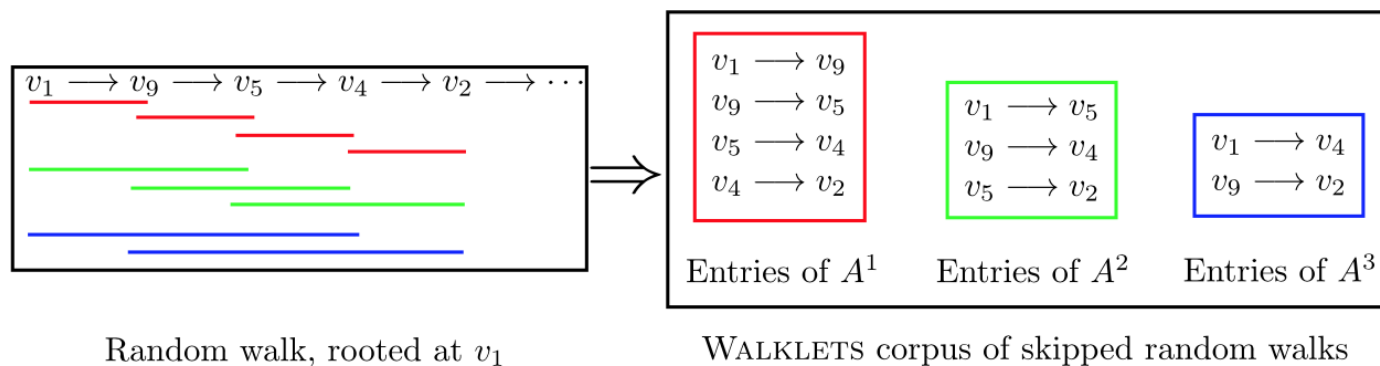
基本假设

- 现实网络是有层次结构的。



Skip!

- 刻画这种层次结构
- 从游走序列中提取训练样本时，用以下形式代替滑窗



目标函数

- 对不同尺度的训练样本，单独优化以下目标函数

$$J_k = - \sum_{v_i, v_j \in C_k} \log \Pr(v_i | v_j)$$

其中 $\Pr(v_i | v_j)$ 表示 v_j 出现在 v_i 的 k 跳上下文中的概率。

- 不同尺度的训练样本是一起训练的。
- 但是文章中并没有说明不同尺度的训练样本的目标函数是如何结合在一起的。
- 我猜是加起来。比如我选择 C_2 和 C_3 的样本来训练，目标函数就是 $J = J_2 + J_3$ 。

实验结果

% Labeled Nodes	10%	50%	90%
WALKLETS (A^2)	37.46**	41.19*	42.59
DeepWalk	34.55	40.77	42.34
LINE	23.65	34.67	37.44
GraRep	37.05**	40.95	42.31
Gain over DeepWalk(%)	8.9	1.0	0.6
Gain over LINE(%)	58.4	18.8	13.8

(a) BlogCatalog

% Labeled Nodes	1%	5%	9%
WALKLETS (A^2)	32.47**	37.41**	38.70**
DeepWalk	31.18	36.64	38.18
LINE	25.06	30.55	32.85
GraRep	-	-	-
Gain over DeepWalk(%)	4.1	2.1	1.4
Gain over LINE(%)	29.6	22.5	17.8

(a) Flickr

% Labeled Nodes	1%	5%	9%
WALKLETS (A^3)	60.49**	65.36**	66.24**
DeepWalk	54.93	63.45	65.16
LINE	45.03	51.69	53.32
GraRep	63.62**	67.47**	68.34**
Gain over DeepWalk(%)	10.1	3.0	1.7
Gain over LINE(%)	34.3	22.6	20.5

(b) DBLP

% Labeled Nodes	1%	5%	9%
WALKLETS (A^2, A^3)	37.19*	40.73**	42.14**
DeepWalk	36.17	39.68	41.49
LINE	33.21	36.94	39.19
GraRep	-	-	-
Gain over DeepWalk(%)	2.8	2.6	1.6
Gain over LINE(%)	12.0	10.2	7.5

(b) YouTube

- 可以看到，作者选择了一些尺度的样本来训练。但是，作者并没有说明为什么要选这些尺度和尺度应该如何选。
- 我猜是作者全部组合都跑了一下然后拿最好的结果放出来的。