

解决的问题

如何在图结构的社交数据中预测顶点的影响力。

在图中，给定顶点 v 与它的邻域以及一个时间段，通过对开始时各顶点的状态进行建模，来对结束时顶点 v 的状态进行预测（是否被激活）。

问题定义

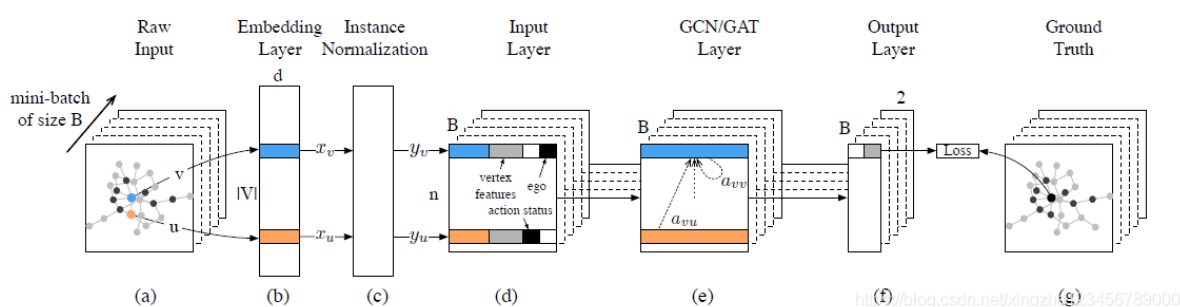
- 邻域：给定图 $G = (V, E)$ ，顶点 v 的邻域定义为 $N_v^r = \{u : d(u, v) \leq r\}$ ，是一个顶点集合，不包含顶点 v 自身
- 中心网络：由邻域中的顶点及边所组成的网络，以 G_v^r 表示
- 用户行为：以 s_v^t 表示，用户对应于图中的顶点，对于一个时刻 t ，如果顶点 v 有产生动作，例如转发、引用等，则 $s_v^t = 1$

给定用户 v 的中心网络、邻域中用户的行为集合 $S_v^t = \{s_i^t : i \in N_v^r\}$ ，论文想解决的问题是，在一段时间 Δt 后，对用户 v 的行为的预测：

$$P(s_v^{t+\Delta t} | G_v^r, S_v^t)$$

做法及创新

数据预处理



数据预处理方面，论文通过带重启的随机漫步来为图中的每个顶点 v 获取固定大小 n 的中心网络 G_v^r ，接着使用DeepWalk来得到图中顶点的embedding，最后进行归一化。通过这几个步骤对图中的特征进行提取后，论文还进一步添加了几种人工提取的特征，包括用户是否活跃等等：

Name	Description
Vertex	Coreness [4].
	Pagerank [35].
	Hub score and authority score [9].
	Eigenvector Centrality [6].
	Clustering Coefficient [50].
	Rarity (reciprocal of ego user's degree) [1].
Embedding	Pre-trained network embedding (DeepWalk [36], 64-dim).
Ego	The number/ratio of active neighbors [2].
	Density of subnetwork induced by active neighbors [46].
	#Connected components formed by active neighbors [46].

摘要里说传统的影响力建模方法都是人工提取图中顶点及结构的特征，论文的出发点就是自动学习这种特征表示，结果在预处理的最后还是添加了几种人工提取的特征，这不是自相矛盾吗？

经过上面的步骤后，最后得到包含所有用户特征的一个特征矩阵 $H \in \mathbb{R}^{n \times F}$ ，每一行 h_i^T 表示一个用户的特征， F 等同于 DeepWalk 长度加上人工特征长度。

影响力计算

这一步纯粹是在套 GAT 的框架，没什么可以说的，计算如下：

$$H' = \text{GAT}(H) = g(A_{\text{GAT}}(G)HW^T + b)$$
$$A_{\text{GAT}}(G) = [a_{ij}]_{n \times n}$$

其中 $W \in \mathbb{R}^{F' \times F}$, $b \in \mathbb{R}^{F'}$ 是模型的参数， a_{ij} 的计算在 GAT 论文的笔记中有记录，不再赘述。

数据集

OAG、Digg、Twitter、Weibo