

Link Prediction Task and Features

- 根据已有的连接预测新的连接

Two formulations of the link prediction task:

1) Links missing at random:

- Remove a random set of links and then aim to predict them

2) Links over time

- Given $G[t_0, t_0']$ a graph on edges up to time t_0' , output a ranked list L of links (not in $G[t_0, t_0']$) that are predicted to appear in $G[t_0, t_0']$
Take top n elements of L and count correct edges.

- Link Prediction via Proximity

如何为给定的一对节点提供描述符?

$$(x, y) \rightarrow c(x, y)$$

例如 $c(x, y)$ 可以是 x, y 的共同邻居数

日期： /

• Distance - Based Features

1. Shortest-path distance between two nodes

不衡量邻域重叠的程度或连接强度

2. Local Neighborhood Overlap

当2个节点无关

同邻居时总是为0捕捉2个节点 v_1, v_2 的共有邻居数

1) Common neighbors : $|N(v_1) \cap N(v_2)|$

2) Jaccard's coefficient : $\frac{|N(v_1) \cap N(v_2)|}{|N(v_1) \cup N(v_2)|}$

3) Adamic-Adar index : $\sum_{u \in N(v_1) \cap N(v_2)} \frac{1}{\text{deg}(u)}$

3. Global Neighborhood Overlap

- Katz Index : 给定节点对之间所有不同长度的路径数目 How to compute ?

令 $P_{uv}^{(k)}$ 表示 u, v 之间长为 k 的路径数目

$$P^{(k)} = A^k$$

v_1 与 v_2 之间的 Katz Index 表示为：

$$S_{v_1 v_2} = \sum_{k=1}^{\infty} \beta^k P_{v_1 v_2}^k \quad (0 < \beta < 1)$$

折扣因子，降低长序列重要性

Katz Index 有闭式解： $S = \sum_{i=1}^{\infty} \beta^i A^i = (I - \beta A)^{-1} - I$