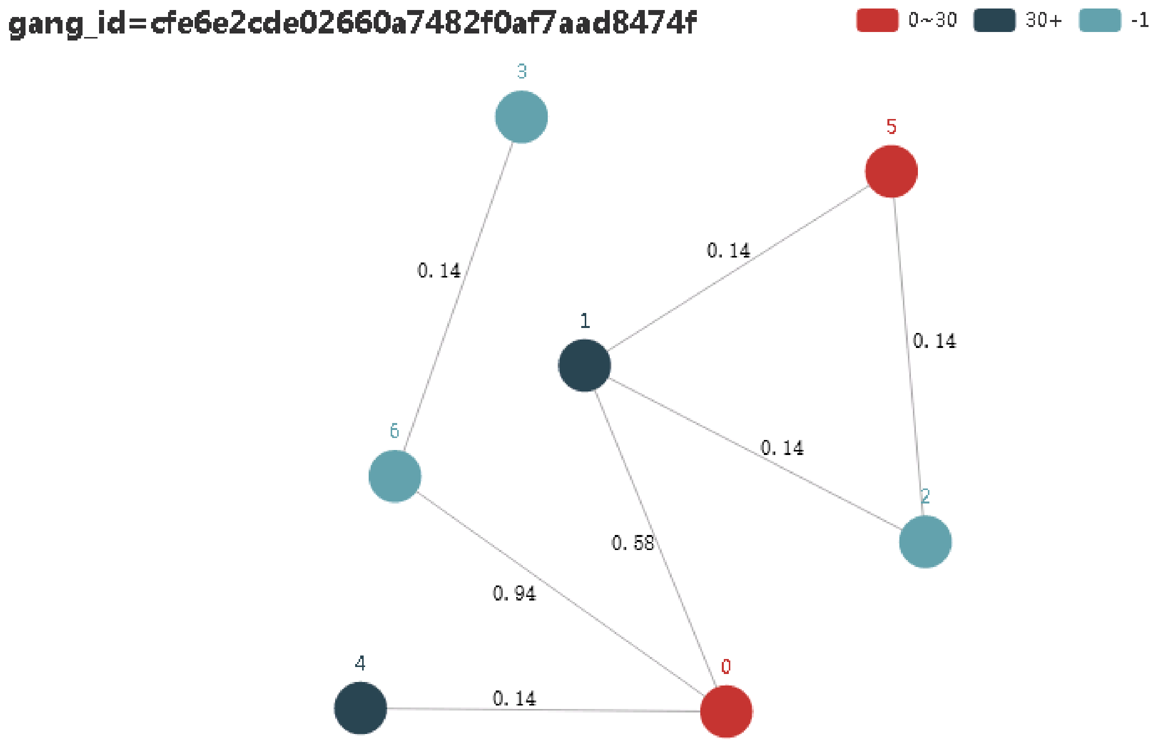
基于关系网络和Personal Rank 随机游走算法的欺诈分计算



一、求图的邻接矩阵

以如上无向加权图为例，其$n \times n$的邻接矩阵W为：

$n = 7$

$$W_{n\times n} = \begin{bmatrix} 0 & 0.58 & 0 & 0 & 0.14 & 0 & 0.94 \\ 0.58 & 0 & 0.14 & 0 & 0 & 0.14 & 0 \\ 0 & 0.14 & 0 & 0 & 0 & 0.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.14 \\ 0.14 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.14 & 0.14 & 0 & 0 & 0 & 0 \\ 0.94 & 0 & 0 & 0.14 & 0 & 0 & 0 \end{bmatrix}$$

二、求图的转移概率矩阵

p_i点的转移概率为：

$$\frac{w_{ij}}{degree(p_i)}$$

其中$degree(p_i)$为p_i点的度数，是该点连接的所有边的权重之和：

$$degree(p_i) = \sum_{j=0}^{n-1} w_{ij}$$

则上图的转移概率矩阵P为：

$$P_{n\times n}=\begin{bmatrix} 0 & 0.35 & 0 & 0 & 0.08 & 0 & 0.57 \\ 0.68 & 0 & 0.16 & 0 & 0 & 0.16 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.87 & 0 & 0 & 0.13 & 0 & 0 & 0 \end{bmatrix}$$

三、求一个点游走到其他点的概率向量

p_i点第一次游走到其他点的概率列向量的计算公式为：

$v_{i,1}=(1-\alpha)P^T v_{i,0}+\alpha r_i$

其中：

1. α为随机转移概率系数，一般取0.15
2. r_i是一个n维的列向量（上图$n=7$），$r_i[k]=\begin{cases}1, & k=i\\0, & k\neq i\end{cases}$，

   即假如$i=0$，$r_i=r_0=\begin{bmatrix}1\\0\\0\\0\\0\\0\\0\end{bmatrix}$
3. $v_{i,0}$为p_i点游走到其他点的初始概率，$v_{i,0}=r_i$
4. P^T表示为矩阵P的转置
5. v_i向量的所有元素之和等于1

经过n次迭代后，概率会收敛于：

$v_{i,n}=(1-\alpha)P^T v_{i,n-1}+\alpha r$

即$v_{i,n}=v_{i,n-1}$时或迭代次数达到60次时停止迭代，取最后收敛的$v_{i,n}$为p_i点游走到各点的概率，且同样满足上述第5点的条件。

p_0点的概率列向量计算结果为

$$v_0=\begin{bmatrix}0.45\\0.17\\0.04\\0.03\\0.03\\0.04\\0.24\end{bmatrix}$$

即p_0点会留在自己本身的概率为0.45，p_0点会游走到p_1点的概率为0.17，……，p_0点会游走到p_6点的概率为0.24

四、求所有点游走到各点的概率矩阵

其他点的计算方法同理，最后得到所有点游走到各点的概率矩阵V：

$$V_{n\times n}=\begin{bmatrix}0.45 & 0.33 & 0.24 & 0.31 & 0.39 & 0.24 & 0.37\\ 0.17 & 0.31 & 0.23 & 0.12 & 0.14 & 0.23 & 0.14\\ 0.04 & 0.07 & 0.24 & 0.03 & 0.03 & 0.13 & 0.03\\ 0.03 & 0.02 & 0.01 & 0.18 & 0.02 & 0.01 & 0.04\\ 0.03 & 0.02 & 0.02 & 0.02 & 0.18 & 0.02 & 0.03\\ 0.04 & 0.07 & 0.13 & 0.03 & 0.03 & 0.24 & 0.03\\ 0.24 & 0.18 & 0.13 & 0.31 & 0.21 & 0.13 & 0.36\end{bmatrix}$$

五、求各点会游走到逾期30天以上点的概率得分

以上图为例，逾期30天以上的点为p_1和p_4，则标签向量为：

$$l_n=\begin{bmatrix}0 & 1 & 0 & 0 & 1 & 0 & 0\end{bmatrix}$$

各点会游走到逾期30天以上点的概率得分为：

$$score=l_nV_{n\times n}=\begin{bmatrix}0.2 & 0.33 & 0.25 & 0.14 & 0.32 & 0.25 & 0.17\end{bmatrix}$$

即除去本身已逾期30天的用户p_1点和p_4点，其他点得分从高至低为：

p_2点有风险的概率为0.25，p_5点有风险的概率为0.25，p_0点有风险的概率为0.2，p_6点有风险的概率为0.17，p_3点有风险的概率为0.14