2018年10月25日星期四

下午11时17分

实验记录：

今天重新改写了0.1版本的实验程序，将其中聚类的代码部分改写为了计算RelationshipLevel值的形式，然后针对2000个用户计算离职用户与邮件联系人的交叉匹配结果。

第一次实验的结果排序十分不理想：

Insider\_2: 22 : LVF1626 : 8.5834351709

Insider\_2: 269 : TMT0851 : 5.97386395252

Insider\_2: 654 : HIS1394 : 4.42147050456

Insider\_2: 700 : DCC1119 : 4.33994283548

Insider\_2: 706 : TNB1616 : 4.32819849632

Insider\_2: 952 : OKM1092 : 3.80557278275

Insider\_2: 1094 : ITA0159 : 3.54606875786

Insider\_2: 1252 : BYO1846 : 3.20087504496

Insider\_2: 1304 : HSN0675 : 3.08794915103

Insider\_2: 1479 : HMS1658 : 2.69905980496

Insider\_2: 1496 : GWG0497 : 2.65534256707

Insider\_2: 1543 : KSS1005 : 2.54911846516

Insider\_2: 1548 : JAL0811 : 2.53642282297

Insider\_2: 1611 : SNK1280 : 2.37246826979

Insider\_2: 1620 : CHP1711 : 2.35740214145

Insider\_2: 1687 : VCF1602 : 2.13489915235

Insider\_2: 1787 : TRC1838 : 1.79331055601

Insider\_2: 1832 : RRS0056 : 1.51757551633

Insider\_2: 1849 : ICB1354 : 1.45937586266

Insider\_2: 1890 : MGB1235 : 1.13797245136

Insider\_2: 1894 : CIF1430 : 1.08981239594

Insider\_2: 1895 : HXP0976 : 1.07951987208

Insider\_2: 1902 : OSS1463 : 1.04357673488

Insider\_2: 1904 : MDS0680 : 1.04312885453

Insider\_2: 1907 : SIS0042 : 1.00532266726

Insider\_2: 1918 : ZIE0741 : 0.823083125359

Insider\_2: 1919 : MCP0611 : 0.807553699587

Insider\_2: 1920 : NAH1366 : 0.804531198009

Insider\_2: 1921 : WDT1634 : 0.791127853687

Insider\_2: 1966 : CKP0630 : 0.0949496601764

分析原因，排在第一位的用户有很多是单向邮件，而这些邮件无疑影响了我们的判断，因此需要重新排除掉这些单向邮件，只关注ER在(0,1)之间的有效邮件。