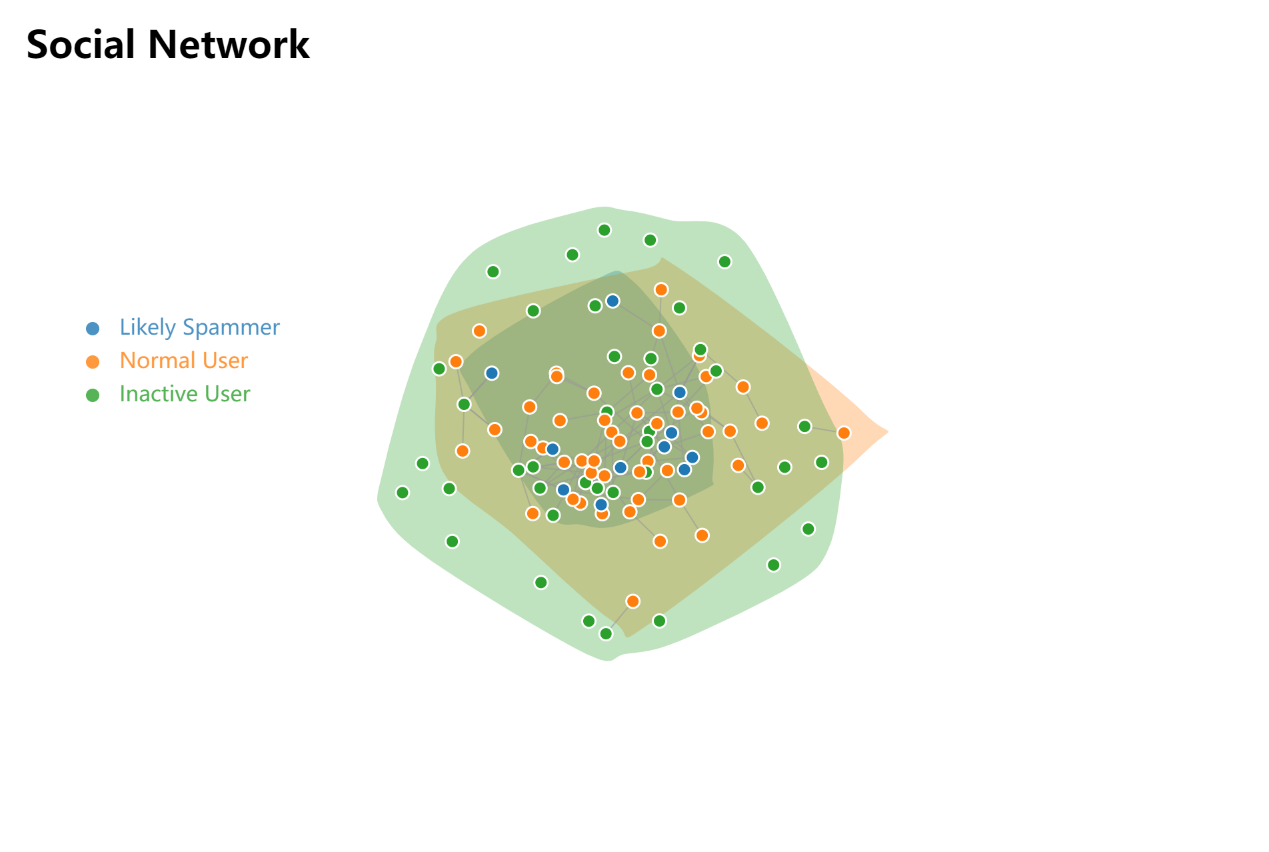
社交网络中垃圾邮件商的分析

在我们的日常生活中，相必大家已经对QQ、微信、微博等社交软件十分熟悉了。在使用这些社交软件时，我们常常会遇到打广告的垃圾邮件商，他们发给我们的垃圾邮件和消息非常影响我们的使用体验。所以，我想通过对社交网络的数据可视化分析，判断社交网络中有哪些用户更有可能是垃圾邮件商，他们到底有什么特性。

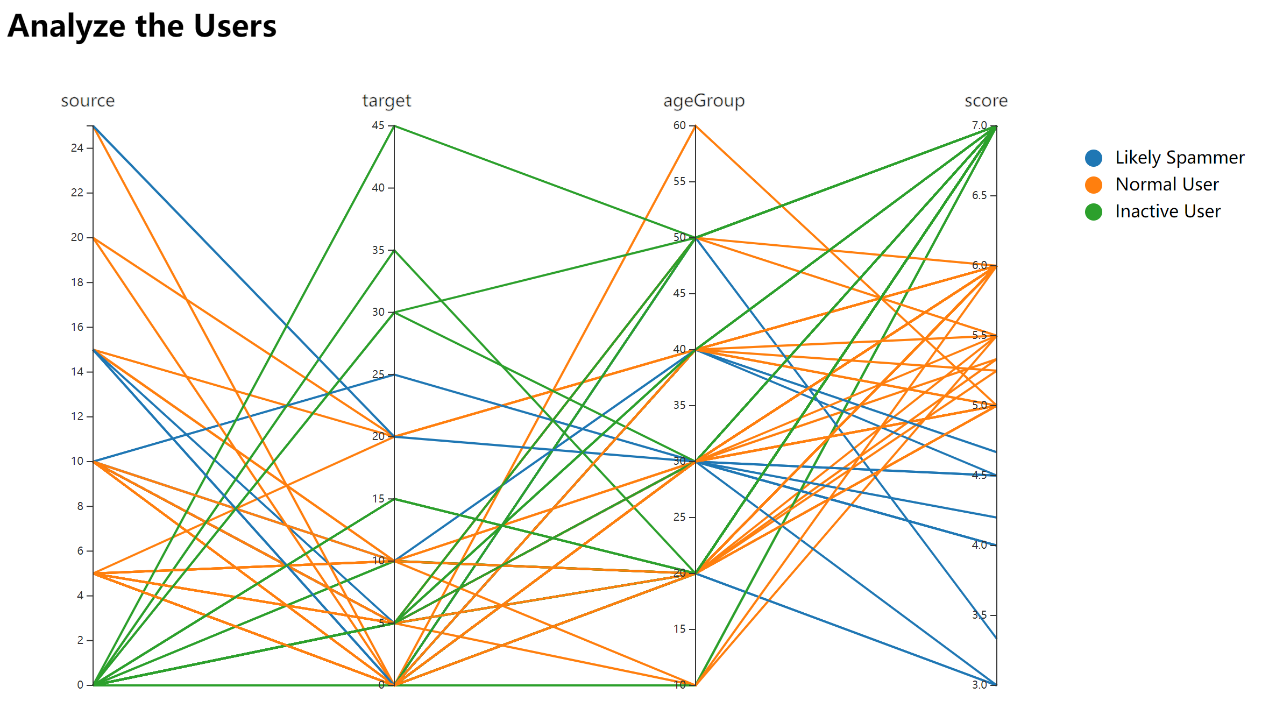
在数据可视化绘图之前，我首先利用社交网络课程中的知识点，把各个用户分为了3类：可能的垃圾邮件商（Likely Spammer）、正常用户（Normal User）、不活跃用户（Inactive User）。其中不活跃用户是没有发出过消息的用户，垃圾邮件商则通过得分判定。由此进行可视化展示，效果可以更加清晰。

分析一个社交网络，我们可以把先把它画出来再看看其中的特性：



在图中，我们可以明显地看到，不活跃的用户集中在外围，因为他们大多数都与其他用户没有联系，所以没有受到多少吸引力。此外，蓝色结点更多集中在最内部，这是因为垃圾邮件商通常发的信息较多，所以收到的连接力也就较大。

此外，我还使用平行坐标图对每个用户的各个属性进行了分析：



在平行坐标图中，可以明显看到，各个不活跃用户发出消息数均为0，而三类用户通过得分分分门别类。在平行坐标图中，我们可以看到三种用户的许多特点：垃圾邮件商接收到的信息数量普遍很少——毕竟发出的垃圾邮件，很少有用户愿意理你；但是不活跃用户收到的信息数量反而较多——我认为这可能是因为该用户本身是个比较出名的人物，所以很少在社交软件上回复信息，但是该明星人物会有众多粉丝，对其发消息。

同时，我们可以看到，垃圾邮件商的平均年龄明显更小，这是因为垃圾邮件商本身确实是做的这项工作，而这种工作需要大量精力，更适合年轻人。至于其他用户，年龄分布就比较平均。

通过这次使用数据可视化方式对社交网络的分析，我确实感受到了不一样的统计方式。如果单纯采用各种数据处理的方法，那通常得到的结果不过就是冷冰冰的数字，即使可以列表或画柱状图、折线图来展示结果，表达出来的东西仍然远不如这种可视化方法得到的直观。所以我觉得，完全可以把数据可视化方法融入到我们日常对大数据的研究中，这样对观众产生的效果绝对是非常有意义的。