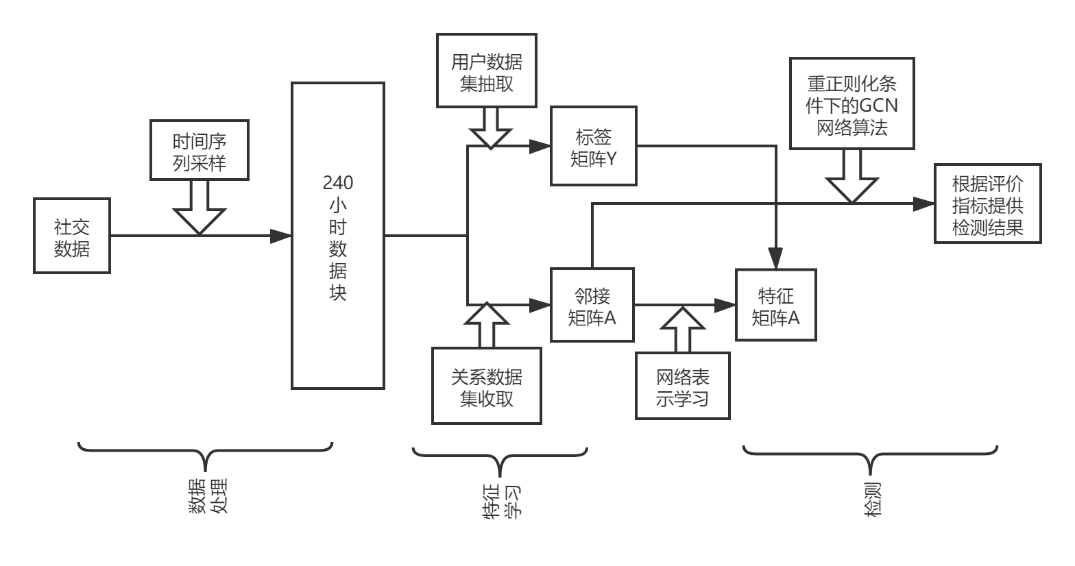
# 基于图卷积网络的社交网络僵尸机器人检测

随着移动互联网技术的广泛应用，在线社交网络由于具有便捷、灵活、内涵丰富的特性而快速成为人们生活重要的组成部分，如Facebook、Twitter、Google、新浪微博、微信等流行社交网络。目前，在线社交网络的用户数量呈指数级别增长，根据腾讯财报数据，截至2019年12月微信和WeChat合并月活就高达11.68亿。由于社交网络蕴含的巨大用户隐私信息及其广阔的商业价值，社交网络成为不法分子图谋不轨的目标，其中，异常用户是不法分子攻击社交网络的主要手段之一。

在社交网络中，僵尸机器人随着网络发展愈发猖獗，它可以未经接收者允许，大量地发送对接收者无用的广告信息，严重地威胁正常用户的信息安全与社交网站的信用体系，且会被用于粉丝售卖。目前研究人员主要采用监督算法、无监督算法以及图算法来检测社交网络僵尸机器人，监督算法需要具有区分度的特征指标，无监督算法需要合理的相似性指标。尽管图算法能够获取深层特征，但是由于图数据量级太大和稀疏性导致计算复杂度过高，效率低。这些方法都存在不小的局限性，比如无法适用高维不规则网格、GCN网络算法需要大量特征字段，在现实情况下很难满足等。

针对现有社交网络僵尸机器人检测方法的提取浅层特征与计算复杂度高的问题，可以考虑基于图卷积网络的社交网络检测技术。基于网络结构信息，通过引入网络表示学习算法提取网络局部结构特征，结合重正则化技术条件下的GCN算法获取网络全局结构特征去检测僵尸机器人，可以有效利用网络结构的局部特征与全局特征，浅层特征与深层特征。在使用重正则化技术的条件下，研究人员[1]发现使用Chebyshev-I多项式的GCN算法具有较高的正确率与较小的时间消耗，优于高阶多项式的实验效果，从直交多项式的角度推导出重正则化技术条件下的GCN算法，可以有效降低图算法的高计算复杂度。

可以设计出流程图如下所示，其中矩形表示不同结构的数据，带箭头的矩形表示不同处理方式。



基于GCN的社交网络僵尸机器人检测技术分为3个模块：数据处理模块、特征学习模块、检测模块。首先先对数据进行处理，提取出关系与特征进行表示，再利用相关算法进行检测，最后返回检测结果。

面对仅含网络结构的社交数据，可以考虑利用网络表示学习提取局部网络结构信息的特征向量，例如DeepWalk、Node2vec、Struc2vec。

根据相关研究，这种方法[2]具有高准确与高效率，并且对于网络图数据结构具有普遍的适用性。

1. KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C] // ICLR. 2017.
2. 曲强, 于洪涛, 黄瑞阳. 基于图卷积网络的社交网络Spammer检测技术[J]. 网络与信息安全学报, 2018, 004(005):P.39-46.