复杂网络分析《红楼梦》中心人物

导语：复杂网络分析的主要任务之一是分析网络中各节点的某种特定关系。本文以《红楼梦》中的社交网络为例，讨论度分布、邻近中心性和特征向量中心性等指标，来找到《红楼梦》中最具有代表性的中心人物。

1. 复杂网络的定义

自然界中存在的大量复架系统都可以通过形形色色的网络加以描述。一个典型的网络是由许多节点与连接两个节点之间的一些边组成的，其中节点用来代表真实系统中不同的个体，而边则用来表示个体间的关系，往往是两个节点之间具有某种特定的关系则连一条边，反之则不连边，有边相连的两个节点在网络中被看作是相邻的。例如，神经系统可以看作大量神经细胞通过神经纤维相互连接形成的网络；计算机网络可以看作是自主工作的计算机通过通信介质如光缆、双绞线、同轴电缆等相互连接形成的网络；类似的还有电力网络、社会关系网络、交通网络等等。

关于复杂网络的定义，系统控制论的开创者钱学森给出了一个较严格的表述：具有自组织、自相似、吸引子、小世界、无标度中部分或全部性质的网络称为复杂网络[1]。

1. 研究背景

“《红楼梦》是一部重人、爱人、唯人为中心思想的书，是我们中华文化史上的一部最伟大的著作，《红楼梦》是我们中华民族文化的代表性最强的作品”，可见现代红学家周汝昌对《红楼梦》的评价之重，《红楼梦》也因其异常出色的艺术成就和丰富深刻的思想底蕴而产生了以其为研究对象的专门学问——红学。

但《红楼梦》流传版本众多，主要可分为两个系统：一是仅流传八十回的脂批抄本系统；一是经程伟元、高鹗整理补缀的一百二十回印本系统。在经过对各种版本的收集与整理后，我们选择了由红楼梦研究所校注、人民文学出版社出版的《红楼梦》以庚辰（1760）本《脂砚斋重评石头记》为底本，以甲戌（1754） 本、已卯（1759）本、蒙古王府本、戚蓼生序本、舒元炜序本、郑振铎藏本、红楼梦稿本、列宁格勒藏本（俄藏本）、程甲本、程乙本等众多版本为参校本，是一个博采众长、非常适合大众阅读的本子，以此保证研究底本的权威性和普适性。

本文将结合复杂网络分析和Gephi等分析工具对《红楼梦》的中心人物进行分析，除去人为的主观因素，用数据来探讨谁是《红楼梦》的中心人物，为红学的研究开辟一条新的途径。

1. 《红楼梦》的主要剧情及社交网络

《红楼梦》讲述的是发生在一个未知朝代的封建大家庭中的人事物，其中以贾宝玉、林黛玉、薛宝钗三个人之间的感情纠葛为主线通过对一些日常事件的描述体现了在贾府的大观园中以金陵十二钗为主体的众女子的爱恨情愁。而在这同时又从贾府由富贵堂皇走向没落衰败的次线反映了一个大家族的没落历程和这个看似华丽的家族的丑陋的内在。

主要故事是从贾宝玉和林黛玉在贾府初识,在成长中相知，以"木石前盟"为信念相爱，但最后宝玉却终究在半清醒状态下被骗而娶了长辈眼中"金玉良缘"的薛宝钗，而林黛玉终只有落得焚稿断痴魂归离恨天，故事是以宝玉清醒后终还是以出家来断尘缘酬知己而薛宝钗怀孕守空房为结局的。

在《红楼梦》中，众多角色相互联系，随着剧情的推进，出现更多的新人物，如刘姥姥等。这些角色及其复杂的人物关系集合成了《红楼梦》的社交网络。

1. 《红楼梦》社交复杂网络的建立

社交网络是一种形容人际关系的网络结构，其本身作为一种复杂网络，可以反映出网络中点与点之间的联系。而在小说中，每个人物正是社交网络中的节点，人物与人物之间的关系为社交网络的边[3]。我们计划用Gephi软件构建《红楼梦》的社交复杂网络，所以需要先建立各人物之间的联系表。在一部小说中，两个角色是否存在联系有很多的判断标准，在这里，我们将采用一段中如果两个角色同时出现，那么他们就是存在联系的，即双方都是各自社交网络中的一员。最简单的方法就是直接比较，但是还有一个问题，就是每个人角色都有各自的别称，如贾宝玉也称宝玉、宝二爷等，所以需要在比较的时候将各个角色的别称也考虑在内。

用Scala编程得到所有角色的姓名数组（为了分析方便，将部分属于同一家庭且关联性小的角色合为一个，如：“周瑞家的”、“余信家的”等）；各个角色的别称HashMap表；经过分段处理的《红楼梦》文本文档。然后对文本进行遍历，如果一段中两个角色同时出现那么就建立联系，直到遍历结束。同时，为了计算部分指标，需要计算出每条边的权重，这里将以两个人物在所有段落中共同出现的次数之和作为这条边的权重。最终结果将建立一个保存各个角色联系且带有每条边权重的csv文件。

社交网络图可以分为有向图和无向图，有向图是人物联系有方向的社交网络，如电话、邮件、短信的方向；无向图是人物联系无方向的社交网络，如同学、朋友等。所以我们选择建立无向图预处理得到的csv文件导入到Gephi中，得到《红楼梦》人物关系的复杂网络图见图1。利用该复杂网络可以进行中心性等指标的测量。

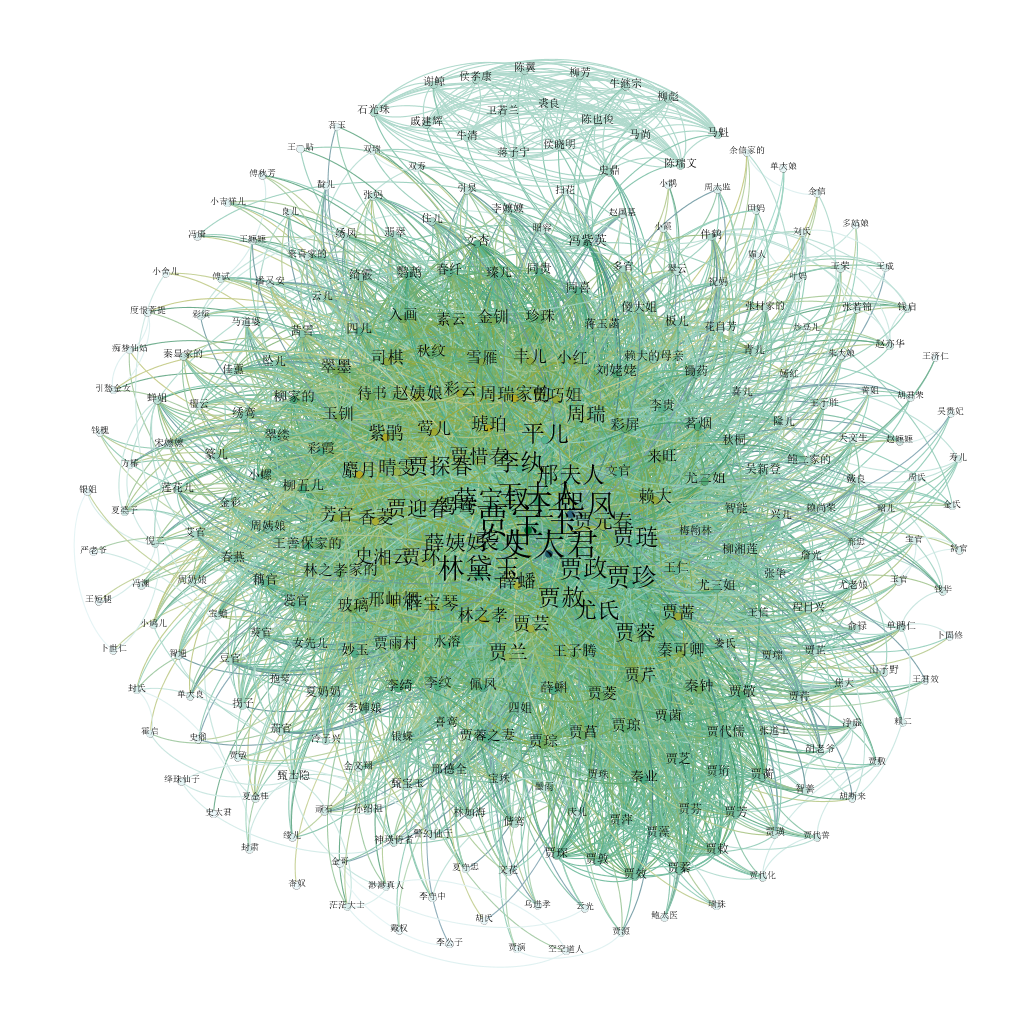


图1：《红楼梦》人物关系网络图

1. 《红楼梦》社交网络的分析指标测量
2. 人物的度中心性分析

度中心性是在网络分析中刻画节点中心性的最直接度量指标，一个节点的节点度越大就意味着这个节点的度中心性越高，该节点在网络中就越重要。度中心性指一个点与其他点直接连接的总和除以网络的规模。度中心性衡量节点在整个网络中的地位和作用，有着高度中心性的节点通常是网络中的行为所在地，如果移除的话，对整个社会网络会产生显著影响。

对于有N个结点的无向图中结点i的度中心性：



式中为节点i的度数。

我们可以将与节点相连接的边数作为这个节点的度，度值是判断一个人物在社交网络中重要性的最简单方法。图2展示了《红楼梦》部分主要人物的社交网络，以图2为例，假设每个节点度的初始值为0，即该角色没有和他有联系的人。节点每连接一条边，该节点的度增加1，如下图中贾宝玉的度为8。同理，史太君为7；林黛玉为4。

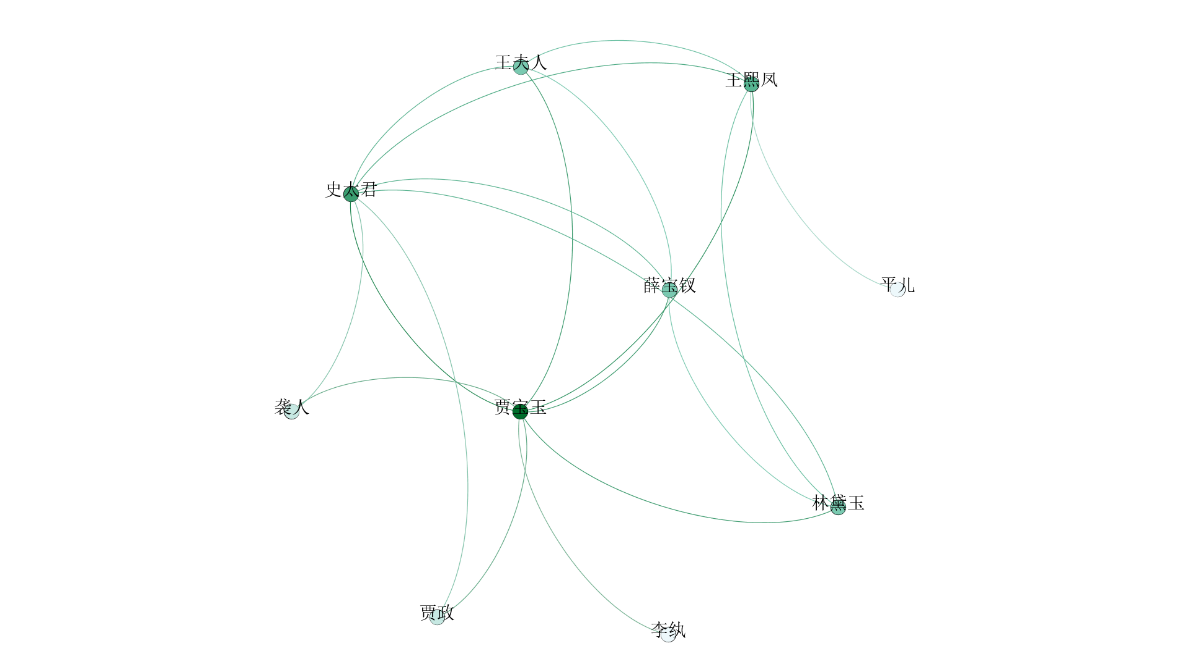


图2：《红楼梦》部分主要人物网络图

利用Gephi计算图1的度分布，得到《红楼梦》中各人物的度分布见图3。其中，度值最高的11位人物的度值情况见图4，然后利用公式可以算出《红楼梦》重要人物度中心性情况见图5。

所以，贾宝玉的度中心性最大，也就是说在《红楼梦》中，贾宝玉人脉最广且最受欢迎。

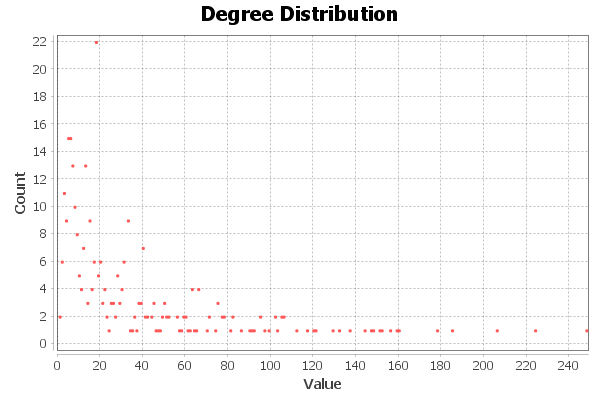


图3：《红楼梦》人物度分布

图4：《红楼梦》主要人物度分布

图5：《红楼梦》主要人物度中心性

1. 人物节点的介数中心性分析

节点介数中心性是网络中所有节点对的最短路径中经过该节点的路径数目占最短路径总数的比例。介数中心性能够反映某个节点或链路在拥塞情况下对网络流的控制，反映了相应的节点或者边在整个网络中的作用和影响力，是一个重要的全局几何量[4]。高介中心性的节点如果被移除，可能会使得某些节点间无法建立连接。公式如下：

式中，表示经过节点v的s→t的最短路径条数，表示s→t的最短路径总条数。

用Gephi对图1进行介数中心性分析，得到《红楼梦》主要人物介数中心性见图6可以看出贾宝玉介数中心度最大，即贾宝玉在人物网络中的“中介”能力最强。

图6：《红楼梦》主要人物介数中心性

1. 特征向量中心度

对于度中心性单纯以度的数量来度量，特征向量中心度则是认为图中各个节点的度并不具有完全相同的重要性，一个顶点的权值既受到连接到该节点的邻居的数量影响，同时还受到这些邻居的质量的影响。综合考虑每个顶点邻居的质量和数量才能够更加客观的评价一个顶点的相对重要性，这就是特征向量中心度。[5]

利用Gephi迭代100次后算出《红楼梦》各人物的特征向量中心度分布情况见图7，取度值最高的前11位见图8所示。

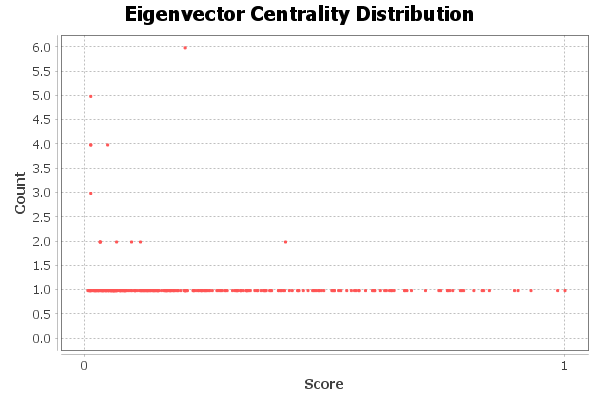


图7：《红楼梦》各人物特征向量中心度分布

图8：《红楼梦》重要人物特征向量中心度

参考文献：

1. 陈关荣, 汪小帆, 李翔. 复杂网络引论 : 模型、结构与动力学= Introduction to Complex Networks:Models,Structuresand Dynamics : 英文 / 陈关荣, 汪小帆, 李翔编著.[M]. 高等教育出版社, 2012.
2. 潘欣源. 计算机时代下的大观园——通过复杂网络研究红楼梦人物关系[J]. 科技资讯, 2015, 013(026):253-254.
3. 楼锴毅, 霸元婕, 李绍昂. 基于社交网络的小说聚类[J]. 软件工程, 2018, 21(10):18-20.
4. 王文钊, 汪斌强, 王志明, et al. 基于网络中心性分析的虚拟网络映射算法[J]. 计算机应用研究, 2015(2):565-568.
5. Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual Web search engine[J]. Computer Networks and ISDN Systems, 30(1-7):107-117.