

WWW.XUANZHI.COM

【雅集】

红楼梦

《红楼梦》中的人物事件及人物关系网络分析

万梦婷

目录

- 数据的基本特征
- 情节线
- 人物关系网络
 - 无权重、有权重
- 人物关系发展
 - 参与人物数量、参与人物结构
 - 对照组：《西游记》、随机模拟数据
- 主成分分析(PCA)
- 社区发现(community discovery)



《红楼梦》人物-事件数据的基本特征

374个人物、475个事件、70个主要人物

- 374个人物
 - 55位没有参与，319位有意义
- 475个事件
 - 前80回：事件1-350
 - 后40回：事件351-475
- 前80回平均每一回包含事件数期望： $350/80 \approx 4$
- 主要人物：参与事件数大于4，即平均参与回目数大于1的人物，共70个人物

人物	参与事件数量	人物	参与事件数量	人物	参与事件数量
贾宝玉	183	林黛玉	100	王熙凤	93
薛宝钗	88	史太君	74	袭人	70
王夫人	67	贾探春	45	平儿	45
李纨	42	史湘云	39	贾珍	36
贾琏	36	薛姨妈	34	邢夫人	33
贾政	31	贾蓉	31	尤氏	31
晴雯	29	麝月	22	鸳鸯	22
贾惜春	21	薛蟠	21	贾赦	20
贾环	20	贾迎春	20	香菱	19
赵姨娘	19	薛宝琴	18	紫鹃	15
刘姥姥	14	秦氏	12	芳官	12
尤二姐	12	秋纹	11	周瑞家的	11
秦钟	11	贾兰	10	贾芸	10
小红	10	贾元春	9	茗烟	9
林之孝家的	9	贾蔷	8	琥珀	8
莺儿	8	玉钏	8	彩云	8
李贵	8				

表 1: 主要人物参与事件数量表



利用主成分分析《红楼梦》中的情节线

主成分分析、贾宝玉、王熙凤

利用主成分分析《红楼梦》中的情节线

- 选取70个主要人物作为变量，前80回发生的事件作为样本，进行主成分分析。
- 第一、第二主成分贡献较大

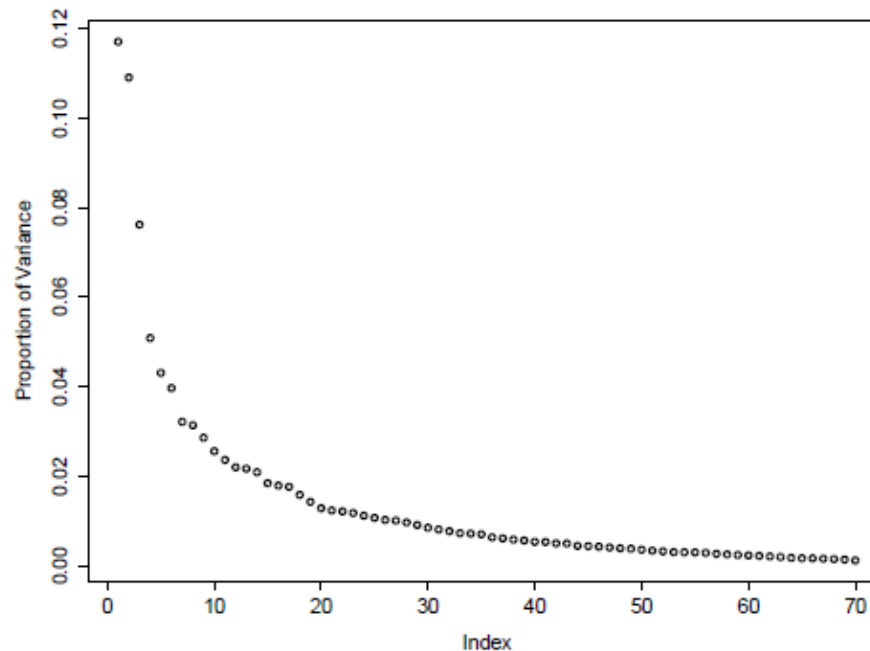


图 2: $CharEve_{main}$ 的不同主成分对于方差贡献比例图

人物	系数	人物	系数
林黛玉	-0.5058839	贾蓉	0.11693321
贾宝玉	-0.4505834	贾璉	0.08943793
薛宝钗	-0.4407956	贾珍	0.08524105
史湘云	-0.2471643	尤氏	0.05100580
贾探春	-0.2354265	平儿	0.04618061
袭人	-0.1965056	尤二姐	0.04521055
李纨	-0.1753417	刘姥姥	0.03583708
史太君	-0.1625484	秦氏	0.03256178
贾迎春	-0.1336803	邢夫人	0.03105103
贾惜春	-0.1265797	贾蔷	0.02871870

表 2: 第一主成分变量系数表

事件	得分	事件	得分
藕香榭聚餐	-2.095548	贾母瞧病	1.0088825
搬进大观园	-2.017665	尤家姐妹	1.0051883
宝玉用功练字	-1.997529	讨银	0.9925885
诗社作诗	-1.958459	宁府	0.9291682
林黛玉追打史湘云	-1.947453	开丧破孝	0.9291682
庆生辰雅座行令	-1.870498	给贾母准备礼物	0.9085889
芦雪庵联诗	-1.761347	贾蔷接事回贾璉	0.9020898
贾政辨凶兆	-1.722533	刘姥姥一进荣国府	0.8949761
众亲戚相认入住大观园	-1.627265	下恩旨	0.8781624
放风筝放晦气	-1.567481	打官司	0.8379200

表 3: 第一主成分样本得分表

利用主成分分析《红楼梦》中的情节线

- 第一主成分：基本区分出荣国府和宁国府两条情节线索
- 《红楼梦》的内容重心在荣国府；
- 贾宝玉及荣府女眷
 - 聚餐、联诗、看戏；
- 宁府和刘姥姥
 - 宁府介绍、尤氏姐妹、刘姥姥进大观园；
- 刘姥姥是非常重要的线索人物

人物	系数	人物	系数
王熙凤	-0.4692509	贾宝玉	0.15963493
史太君	-0.4456338	晴雯	0.06729805
王夫人	-0.4130041	林黛玉	0.06171979
邢夫人	-0.2713506	袭人	0.05978446
贾珍	-0.2127553	麝月	0.03474533
尤氏	-0.1986364	香菱	0.03255482
贾琏	-0.1968891	茗烟	0.02359801
贾蓉	-0.1947474	莺儿	0.02221933
薛姨妈	-0.1465684	芳官	0.01922470
鸳鸯	-0.1411202	甄士隐	0.01346875

表 4: 第二主成分变量系数表

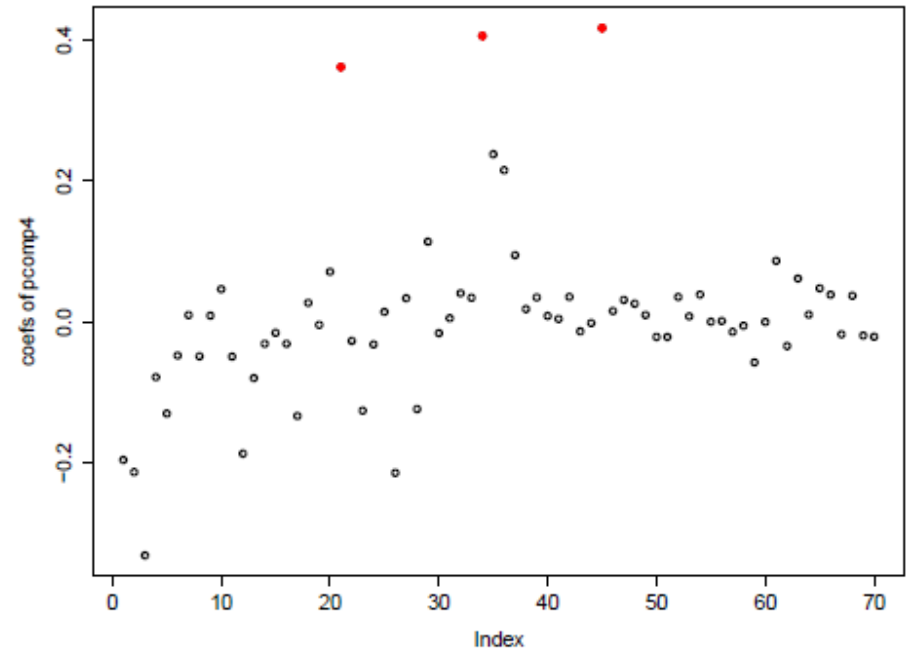
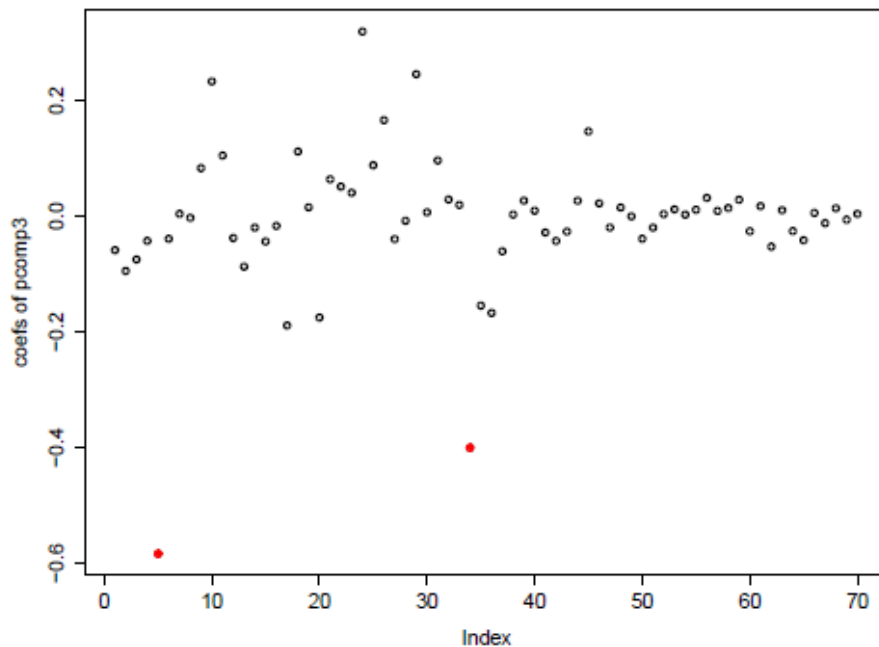
事件	得分	事件	得分
贾琏纳秋桐为妾	-2.352788	晴雯与贾宝玉吵架	0.7245726
春祭恩赏	-2.259318	葬花读会真记听牡丹亭	0.7134274
挑唆张华	-2.115266	袭人房中	0.7095033
秦氏后事	-2.086782	宝玉转话晴雯怒	0.6975981
凤姐生日凑钱	-1.917496	宝玉打闹	0.6975981
贾母携贾氏子侄赏月	-1.882197	梦游太虚幻境	0.6962702
众人斗牌哄贾母高兴	-1.844913	芳官独食	0.6938610
黛玉入府	-1.665450	看字	0.6647881
尤二姐吞金	-1.656763	宝玉命晴雯送手帕给林黛玉	0.6647881
多事之秋众女怨	-1.644008	查夜	0.6628528

表 5: 第二主成分样本得分表

- 第二主成分：反映出《红楼梦》中的年龄分布
- 由宝、黛、钗等年轻贵族及仆人进行的活动
 - 作诗结社、感情互动
- 由王熙凤、贾母等年长贵族及仆人进行的活动
 - 丧事、恩赏

利用主成分分析《红楼梦》中的情节线

- 第三主成分：贾宝玉
- 第四主成分：王熙凤
 - 袭人与王熙凤的互动



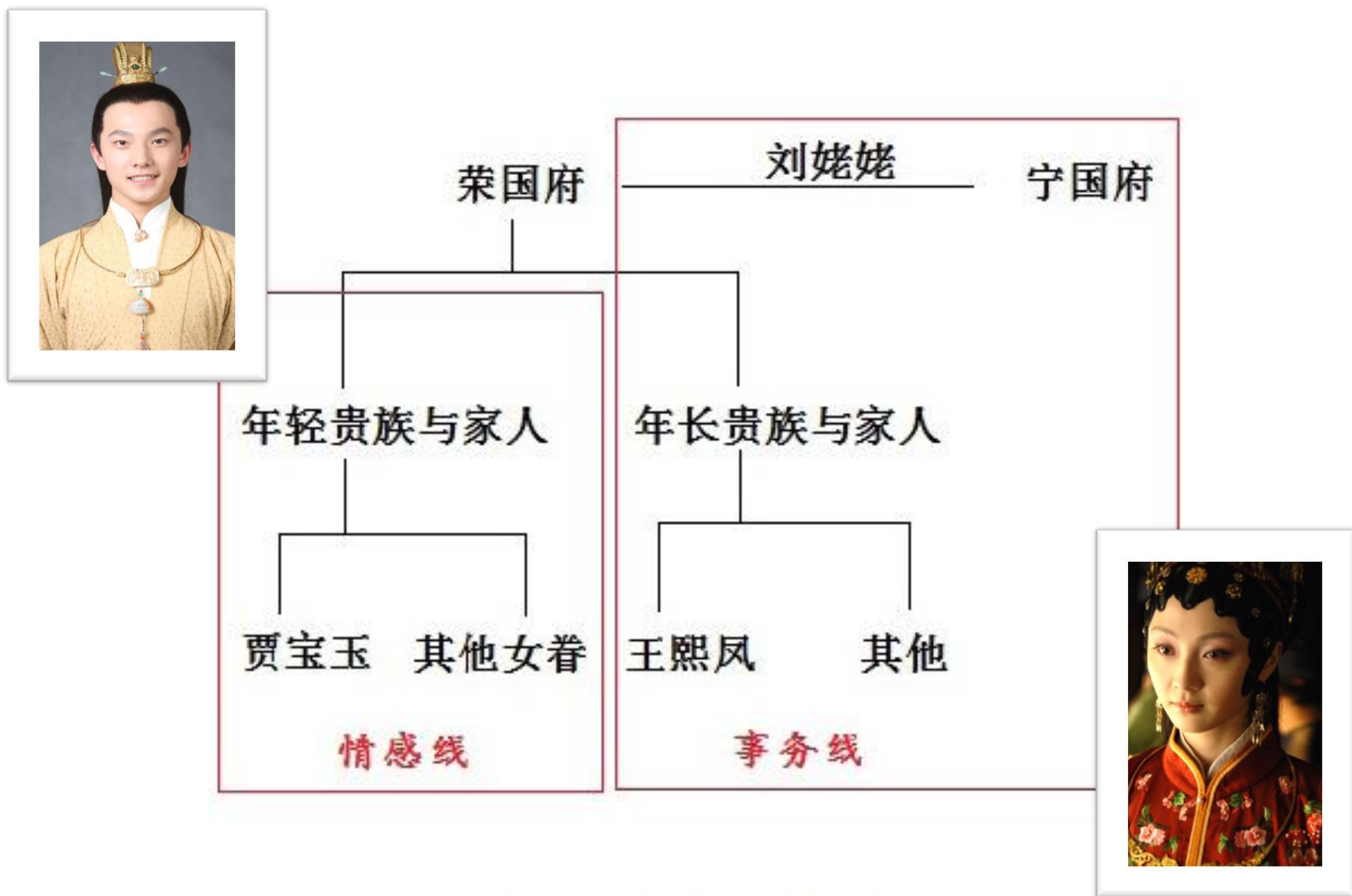


图 4: 《红楼梦》人物情节结构



《红楼梦》中的人物关系网络

无权重无向图、有权重无向图、社区发现

- R: igraph包, Walktrap算法
- 构造连接矩阵 $Adj = \{A_{ij}\}$
 - 无权重
 - 有权重

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j).$$

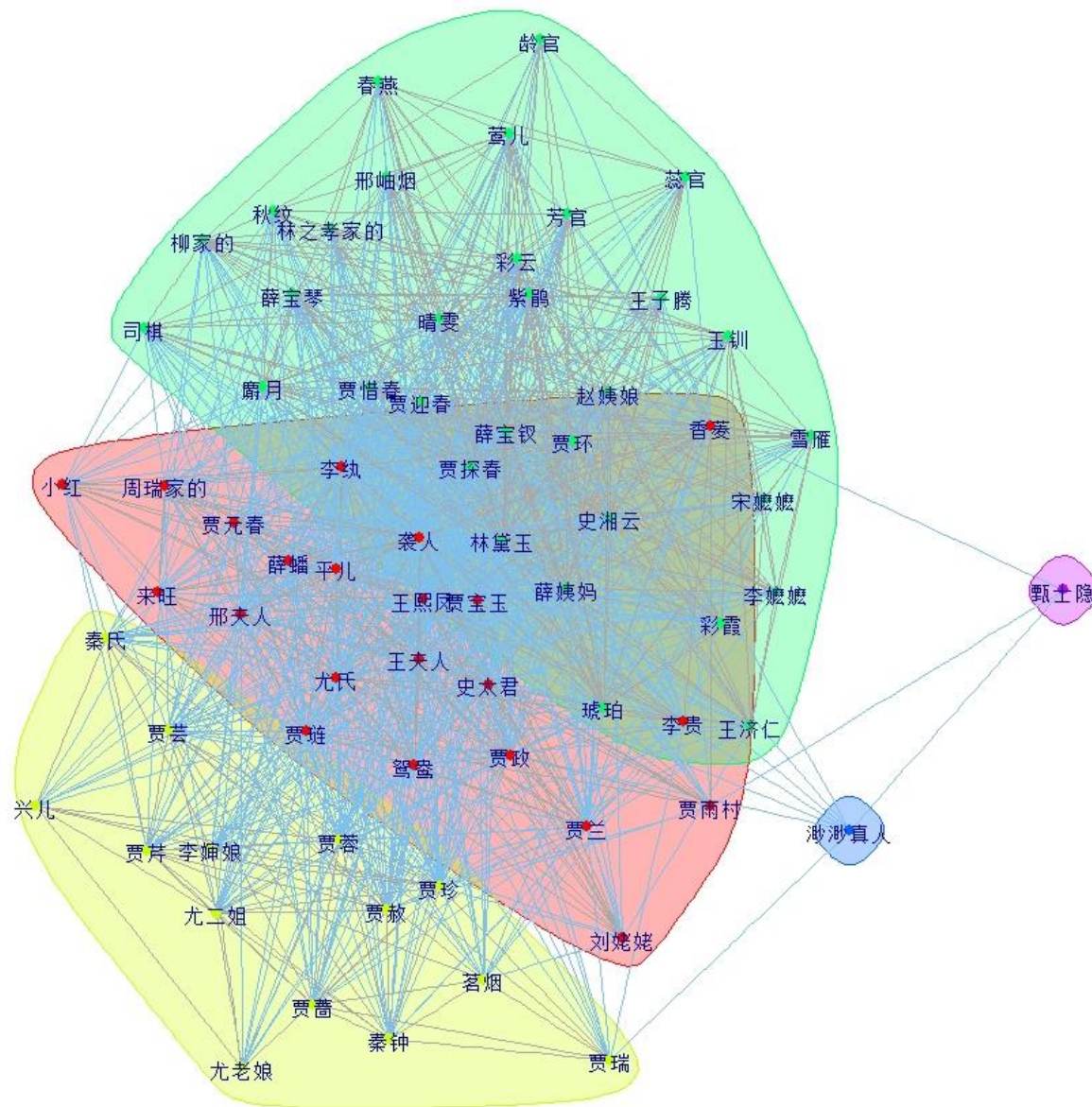
其中 $i, j \in Ch_{main}$, A_{ij} 是连接矩阵 Adj 中的元素; k_i 是顶点 i 的度数, 即 $k_i = \sum_j A_{ij}$; c_i 是顶点 i 属于的社区类别, 若图 G 可以被分为 k 个社区, 则 $c_i \in \{1, 2, \dots, k\}$; 当 $c_i = c_j$ 时 $\delta(c_i, c_j) = 1$, 否则 $\delta(c_i, c_j) = 0$ 。

- 无权重人物网络
 - 反映关系，不反映亲密程度；
 - 主要人物：3个社区（宁府、荣府核心、荣府外戚），
Q=0.05791062

社区类别	规模	代表人物
1	22	贾宝玉、王熙凤、史太君、贾政、贾琏
2	13	贾蓉、贾蔷、秦氏、尤二姐
3	32	贾迎春、贾探春、史湘云、薛宝钗、林黛玉
4	1	渺渺真人
5	1	甄士隐

表 6: 《红楼梦》前80回主要人物无权重网络节点分类

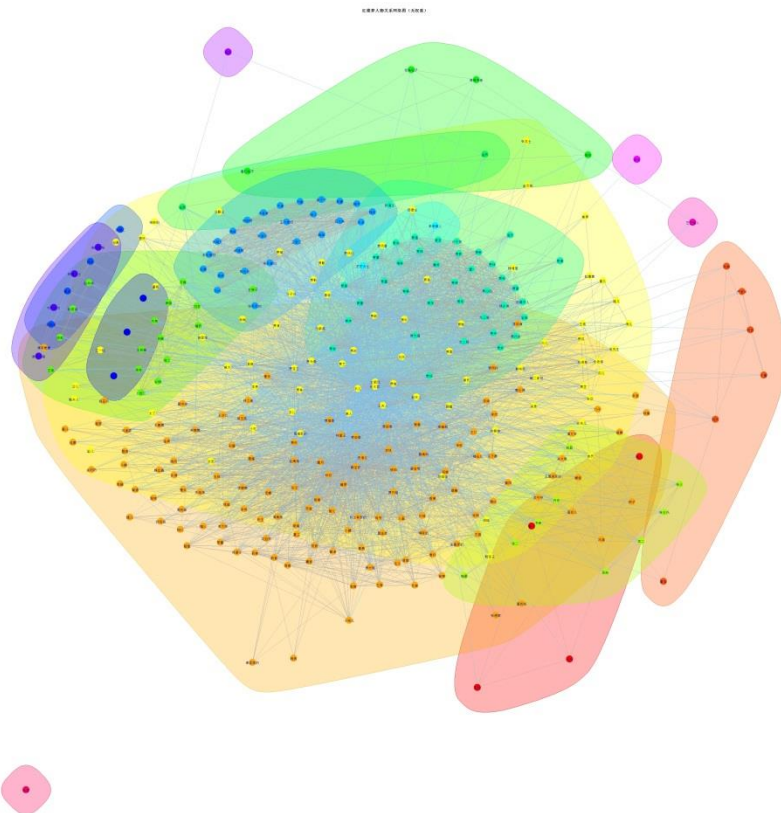
红楼梦主要人物关系网络图（无权重）



- 无权重人物网络

- 全部人物：可以发现不活跃社区，但是主要人物的结构被掩盖，

$Q=0.3363465$

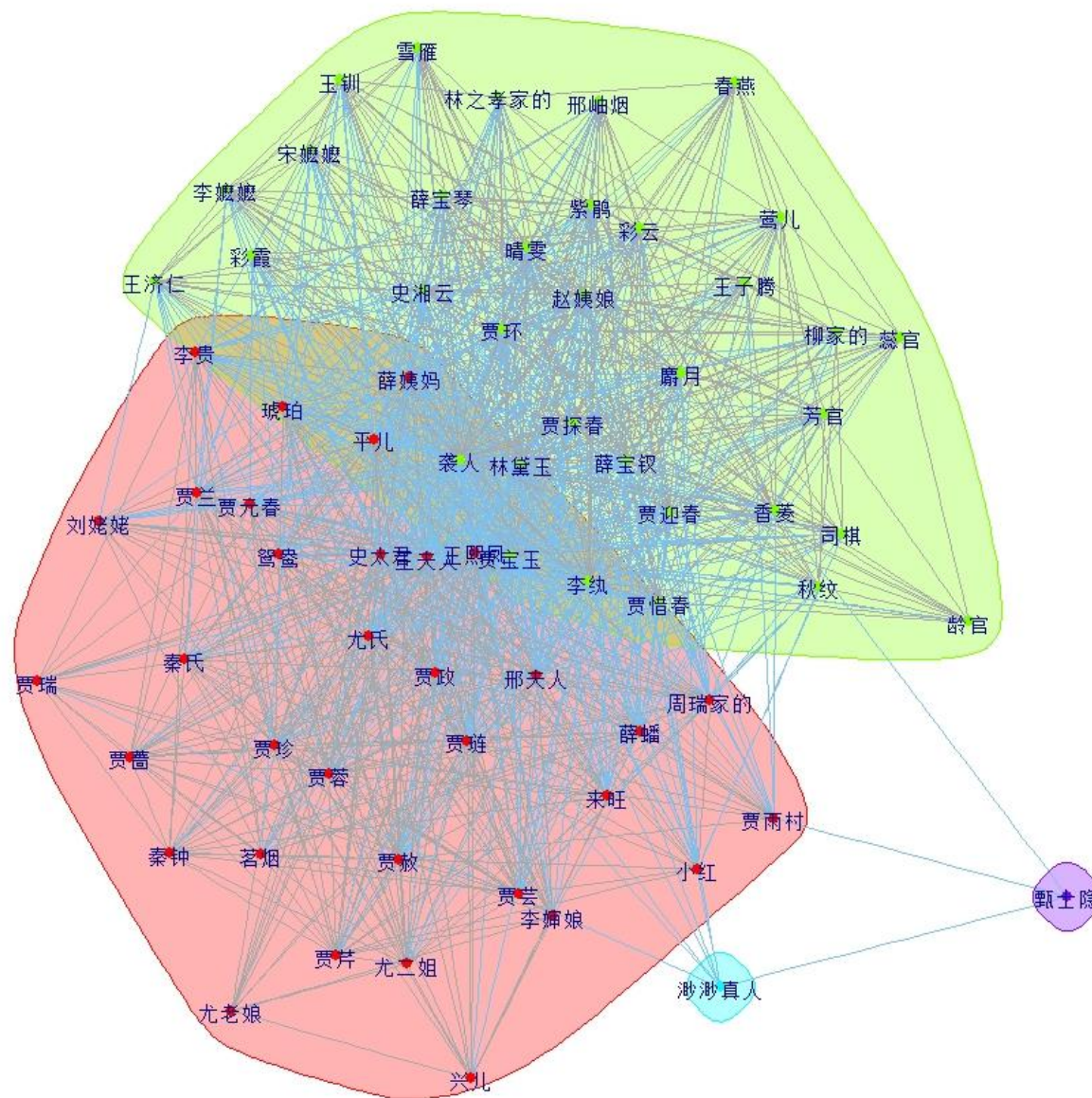


- 有权重人物网络
 - 反映关系，也反映亲密程度
 - 主要人物：2个社区（以贾宝玉为中心的年轻贵族及仆人，以王熙凤为中心的年长贵族及仆人），
Q=0.1448009

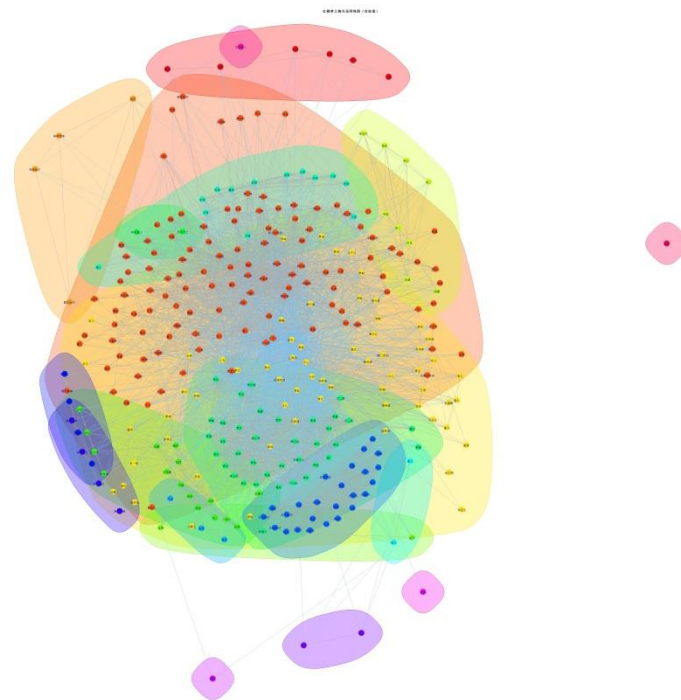
社区类别	规模	代表人物
1	34	王熙凤、史太君、贾政、贾琏、贾蓉、秦氏
2	34	贾宝玉、贾探春、史湘云、薛宝钗、林黛玉
3	1	渺渺真人
4	1	甄士隐

表 7: 《红楼梦》前80回主要人物有权重网络节点分类

红楼梦主要人物关系网络图（有权重）



- 有权重人物网络
 - 全部人物：可以发现不活跃社区，主要人物的结构不会被掩盖
- 验证了上一节所得结果



《红楼梦》中的人物关系发展

参与人物数量、人物关系结构、后40回作者



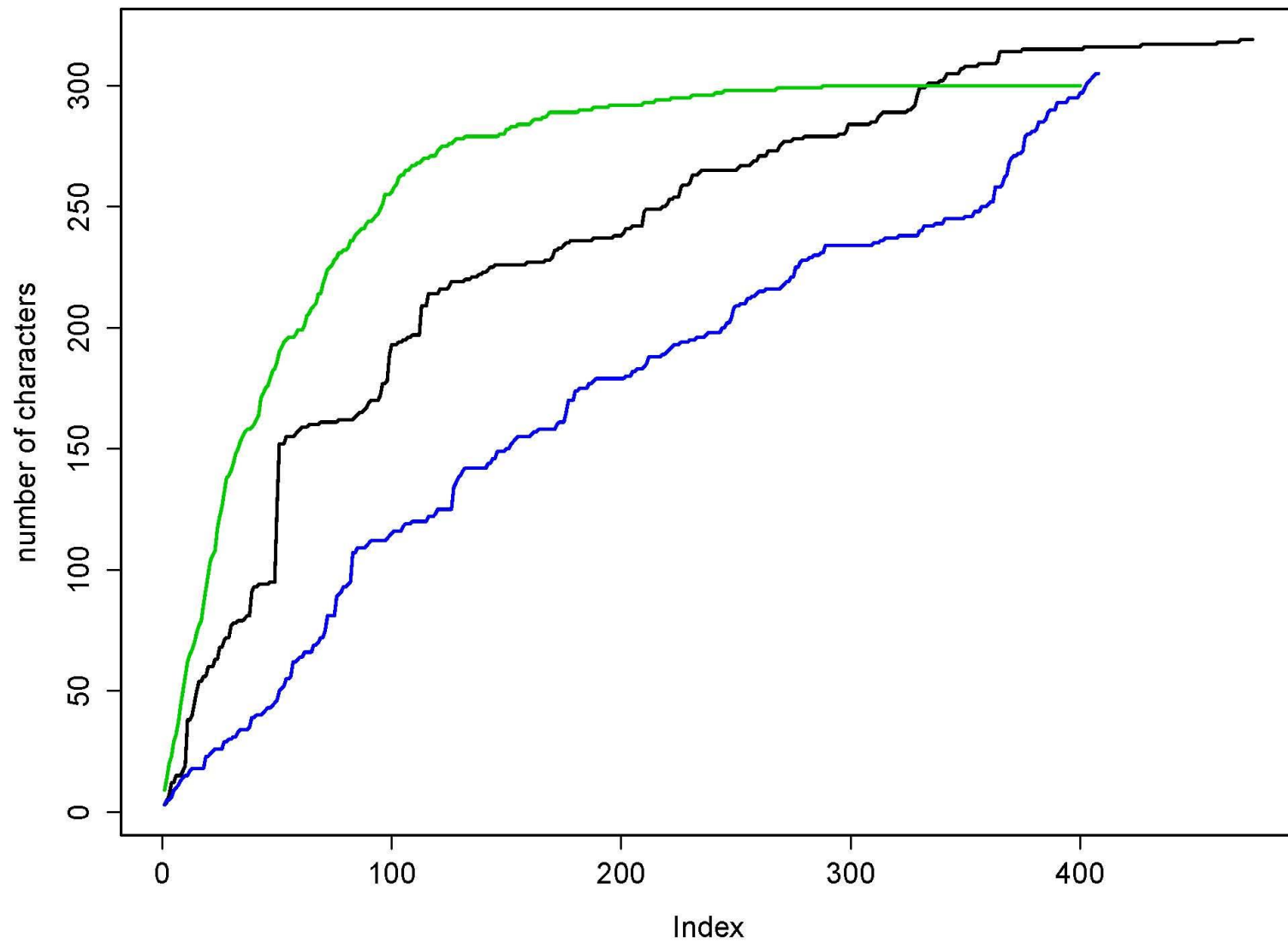
- 数据

- 《红楼梦》所有事件、所有人物形成的人物-事件数据
- 对照：《西游记》、随机模拟数据

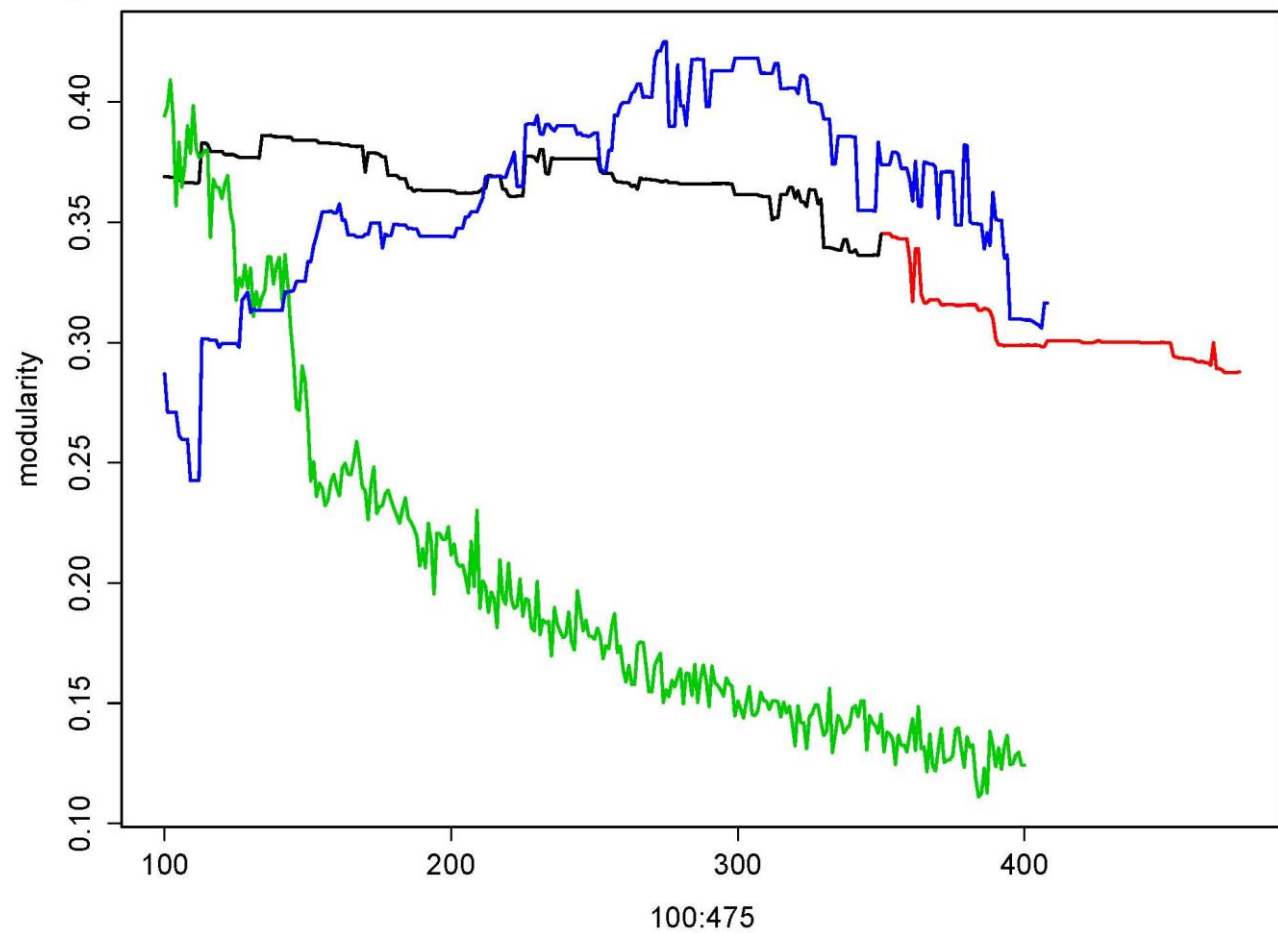
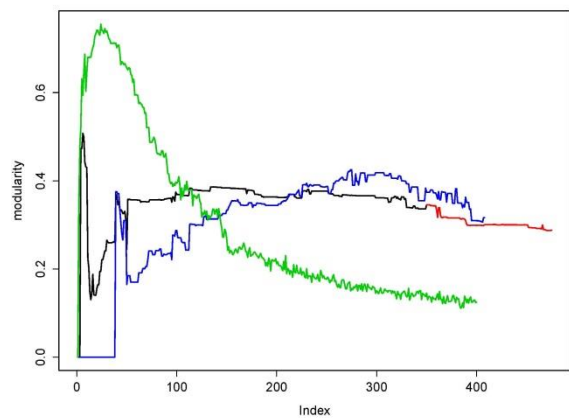
- 随机数据产生机制

1. 估计出《红楼梦》参与每个情节的人数期望 $\lambda = \sum_{i \in Event} \frac{\sum_{j \in Ch} ce_{ij}}{|Event|}$, 其中 $CharEve = \{ce_{ij}\}$;
2. 以 $Poisson(\lambda)$ 产生一个长为400的向量 $\vec{s} = \{s_1, s_2, \dots, s_{400}\}$ 作为每个情节的参与人数;
3. 从1, 2, ..., 300这一序列中抽取的 s_i 个数字, 作为在第 i 个事件中出现的人物, 并重复400次, 得到一个随机产生的 300×400 的人物事件矩阵;

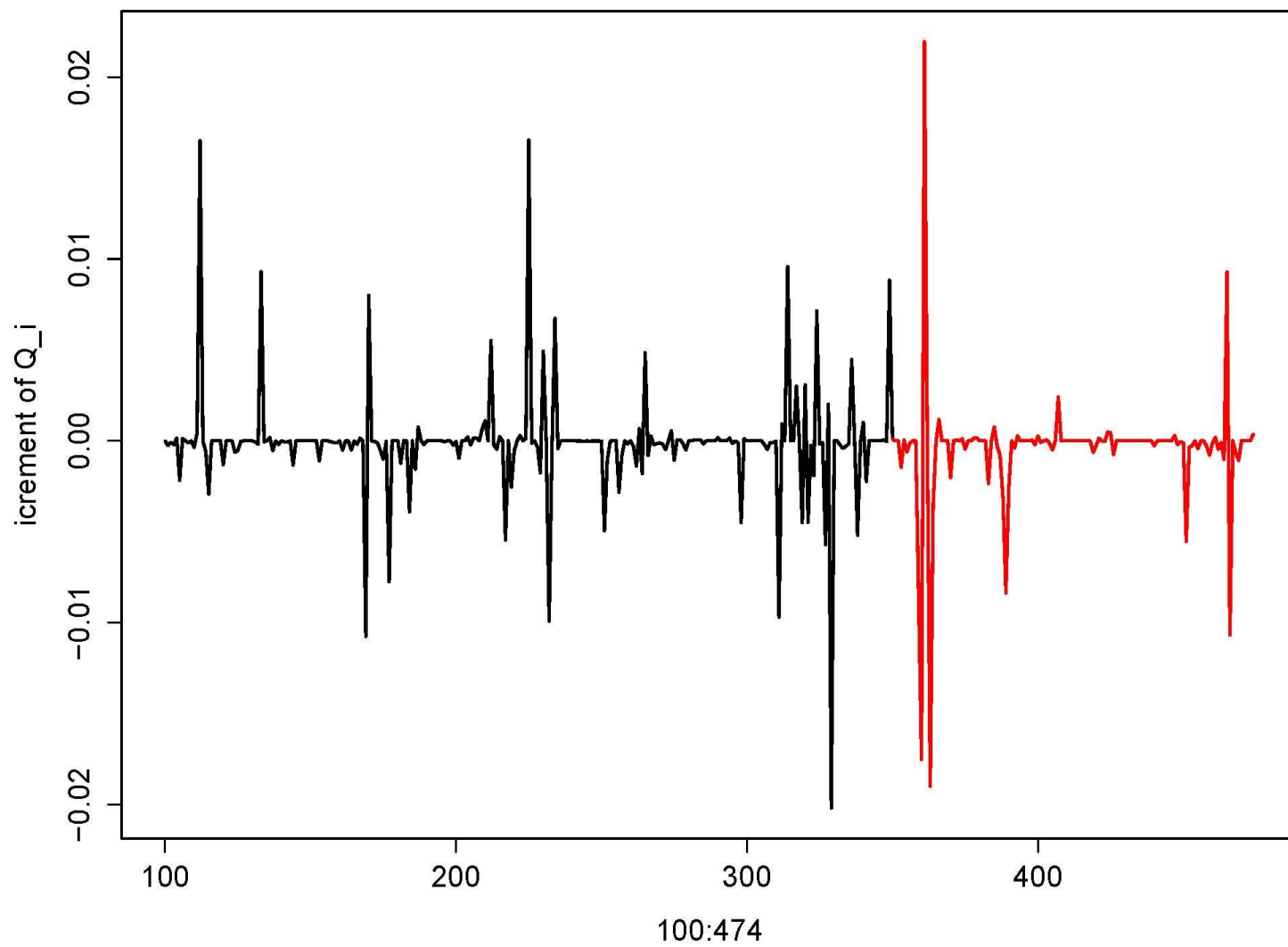
- 计算参与了前 i 个事件的人物总数 s_i
- 《红楼梦》：多线索交叉发生机制
- 《西游记》：单线索递增发生机制
- 线索越混乱越接近随机模拟结果？



- 计算由所有人物和前 i 个事件生成的人物关系网络的模块化指标 Q_i
- 基本假设：当情节发展到一定程度后，人物关系趋于稳定，即新的人物互动对既有结构破坏不大
- 随机数据：情节发展到一定程度后， Q_i 开始下降，推测若不断试验下去 Q_i 接近于0
- 真实数据：情节发展到一定程度后， Q_i 趋于一个大于0的值



- 观察《红楼梦》 Q_i 的变化情况，做出增量 dQ_i 的变化趋势图
- Q_i 波动最大的事件多在前80回与后40回接壤处（是否与后40回作者非曹雪芹有关？）
 - 矢孤介杜绝宁国府（第七十四回）
 - 贾政升官（第八十五回）
 - 贾家庆贺（第八十五回）
 - 薛姨妈营救薛蟠（第八十六回）



- 荣府和宁府
 - 小说重心在荣国府
 - 刘姥姥是重要线索人物
- 以贾宝玉为中心的情感线索和以王熙凤为中心的事务线索
 - 年龄分布（年轻、年长）
 - 活动类型（嬉戏玩闹、家族政治）
- 多线索交叉结构
- 小说后期结构变化，一定程度上暗示了后40回的作者并非曹雪芹



2HI.COM

谢谢！

以此项工作纪念曹雪芹逝世250周年

文件中所有图片均来自豆瓣“新红楼梦”词条相册