



北京大学
PEKING UNIVERSITY

当《西游记》遇见大数据：用网络科学解锁人物关系密码

姓名 马唯硕 学号 2200013081 院系 信息科学技术学院

摘 要

作为中国古典文学的巅峰之作，《西游记》以其庞杂的人物体系和错综复杂的关系网络，为历代读者与研究者提供了无尽的解读空间。传统的人物分析多依赖于文本细读与感性归纳，而本文尝试引入一种全新的研究范式——计算社会网络分析。本文收集了三百多名小说中出现的主次角色，将全书分割成四百多个场景，利用人物在场景中的共同出现关系，构建出一个客观、可度量的人物关系网络。本文通过对网络拓扑结构、核心节点功能及“社群”派阀的综合分析，旨在揭示《西游记》隐藏在文本之下的叙事动力学与权力结构。研究发现，《西游记》的人物关系网络在宏观上呈现为一个由紧密“圈子”构成的、高度连通的“小世界”结构。在此结构中，我们确定了孙悟空作为情节推动与网络连接的绝对核心地位；而精细的社群检测则进一步揭示了取经团队内部“行动者”与“在场者”的功能分工，以及神魔世界中并存的多个权力派阀。总体来说，本文从数据科学的新视角，将传统学术观点与文化印象转化为可被检验的命题，从而对《西游记》的人物关系的宏观结构做出全新的、有数据支持的诠释。代码已开源至<https://github.com/yiming421/Journey-to-the-west-figure-graph>。

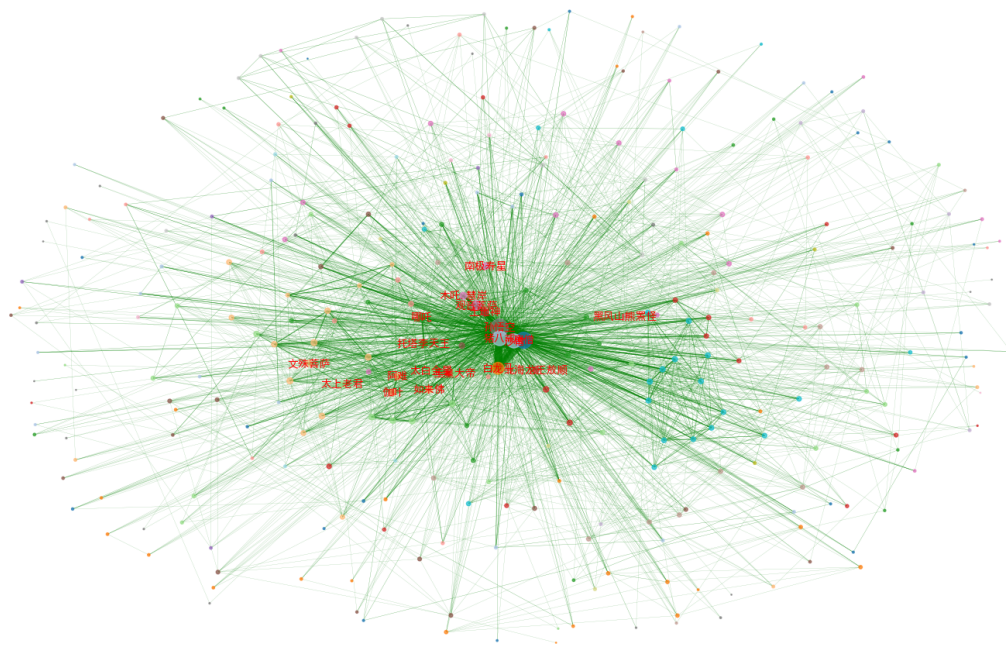


图 1: 《西游记》中的人物关系网。

1 引言

百回本《西游记》中，有名有姓的人物达数百名之多，遍及天、地、人三界，彼此之间相互联系，构成了错综复杂的人物网络。基于文本细读的传统方法，对重要人物之间的微妙关系进行了细致入微的分析，深刻地塑造了我们对这部煌煌巨著的理解。但面对数百名角色在贯穿百章回的大量互动时，只依赖文本细读，可能不能有效地捕捉到宏观人物关系中复杂而系统的结构特征，而且常常停留在不能被精确验证的感性归纳层面上。

近年来，随着数据分析技术的进步，“远读”开始成为受数字人文学者关注的新潮流 [1]。“远读”主张研究者跳出对文本逐字逐句地“精读”，转而对大规模的文本数据做统计学分析，以发现那些只有在宏观尺度下才能浮现的规律。受这种观点启发，本文引入社会网络分析中的技术，将每一个角色抽象为网络中的“节点”，如果角色之间在同一个场景中出现，就在两个角色间连边 [2, 3]。通过这种方法，本文从《西游记》中提取出了可供量化分析的完整关系网络。我们希望通过网络的分析，揭示出《西游记》中的人物关系在宏观层面上呈现出的规律。

2 网络的宏观图景：高度抱团的“小世界”

图 1展示了本文构建出的人物关系表：其中节点的大小表示其连边的数目，节点的颜色代表角色属于哪个“社区”（如天庭，灵山）；而边的粗细则代表共同出现次数的多少；出于视觉效果考虑，图中只标注出了一些重要角色。在深入分析具体角色与社群之前，我们首先来鸟瞰网络的整体宏观特性。这是一个由 302 位核心角色与超过 2200 条关系构成的世界。其全局指标揭示了三个核心特征：

1. 高度稠密。我们将每个节点与其他节点的连边数，称为该节点的“度数”。那么这个网络的每个节点，平均度数为 14.8，说明核心角色平均都与近 15 位其他角色有频繁互动，社交关系丰富。

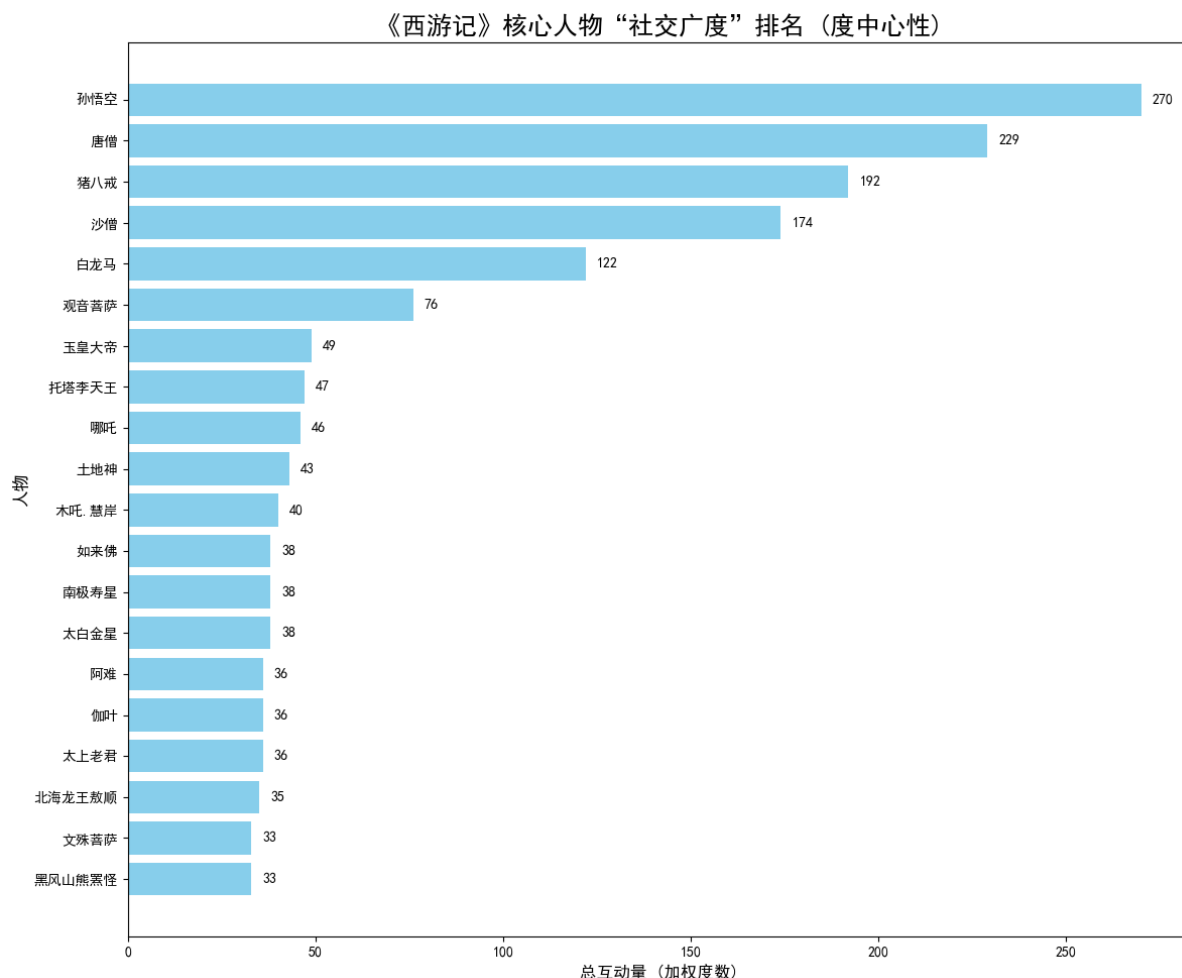


图 2:《西游记》人物关系网的度中心性排名

其中，孙悟空与唐僧的度数更是达到了 200 以上，几乎与每一位出场角色都有联系。

2. 极端抱团。聚类系数是衡量网络中节点之间是否倾向“抱团”的指标，即“朋友的朋友是不是也是朋友”。经过计算，《西游记》中的人物关系网的聚类系数达到惊人的 0.85，这说明人物之间倾向于形成紧密的“小圈子”，在“小圈子”内部几乎每个人都相互认识。

3. “小世界”特征。社交网络中有一个著名的“六度分隔”理论，认为任意两个互不相关的人，至多可以通过六个中间人联系起来，这种现象被称为“小世界”[4, 5]。虽然《西游记》的人物倾向于根据不同场景与派系形成紧密的集团，但取经队伍作为“桥梁”，打通了这些集团之间的关系。根据计算得到，《西游记》的人物网络的直径仅为 4，而平均最短路径更是低至 2.1，这意味着在这个庞大的神魔体系中，任何两个核心角色之间想要建立起联系，只需要通过一位中间人就好（通常来自取经队伍）；网络中两个最遥远的角色，也仅需通过四步就能建立联系。例如凡间虎妖寅将军与阴曹地府的鬼判，只需要三位“中间人”就能建立联系（[寅将军, 唐僧, 孙悟空, 二郎神, 鬼判]）。

3 核心人物的作用：中心性分析

在分析整个人物网络的宏观特征之后，我们将注意力移回那些决定了故事发展的具体角色上。《西游记》中的角色中，哪几位才是最“重要”的？哪几位在整个人物关系网中居于核心位置？网络分析中的“中心性”概念，正是回答这个问题的有力工具[6]。它将“重要性”这个单一概念，折射成

了四个不同但互为补充的维度：社交的广度（度中心性）、结构中的关键性（间介中心性）、传递信息的效率（接近度中心性）以及社会地位上的影响力（特征向量中心性）。通过这四个不同的角度，我们将用数据来回答这个常常为读者争论的问题。

首先，让我们聚焦于最直观的指标——度中心性：这个指标直接衡量一个节点上有多少条连边，即一个人物与多少其他角色有关。正如图 2 所示，取经团队五人组——孙悟空（270）、唐僧（229）、猪八戒（192）、沙僧（174）和白龙马（122）——以绝对优势占据了榜单前五位，形成了无可撼动的“第一梯队”。这一数据有力地证明，他们是整个人物网络的绝对中心，几乎所有的故事线都必须围绕他们展开。紧随其后的是以观音菩萨（76）、玉皇大帝（49）、托塔李天王（47）等为代表的“第二梯队”。他们是神界的核心掌权者和取经任务的关键支持者与干预者。虽然他们不全程参与旅途，但在诸如“大闹天宫”、“搬救兵”等关键节点与主角团的频繁互动，使他们成为了网络中重要的次级中心。再往后，则是其他重要的神佛、以及像黑风山熊黑怪这样与主角团有深入纠缠的妖怪，共同构成了这个神魔世界的社交版图。

然而，度中心性仅仅告诉了我们谁是“社交舞台”上最繁忙的演员。为了更加全面地考察角色们在网络中的功能，我们有必要引入其他维度的中心性指标。四项中心性指标最显著的共同点，在于取经团队的绝对核心地位。无论是社交的广度、信息的中枢、还是影响力的层级，师徒四人始终牢牢占据着网络的前四位，构成了整个故事稳固的结构性骨架。值得注意的是，孙悟空在所有四项中心性指标中全部位列第一。这说明，无论从任何一个网络结构的角度衡量，孙悟空都是《西游记》无可争议的“第一主角”。而但是，其他三项指标，仍提供了与度中心性不同的其他视角。如衡量结构中的“枢纽”性的间介中心性，识别到了一个度中心性并不高的角色：唐太宗李世民。李世民在间介中心性的榜单上位列第五位，紧随师徒四人之后，甚至超过了观音菩萨。这证明，李世民和他相关的“取经缘起”故事线，是整个《西游记》人物关系网中一个至关重要的“结构性桥梁”，正是因为他经历了“游地府”的事件，才引出了“水陆大会”，随后才有了唐僧西天取经的整个故事。任何从“大唐王朝”的角色（如魏徵、秦叔宝）到“取经路上”的角色（如熊黑怪）的最短路径，都极大概率经过李世民这个节点。算法精准地捕捉到了他这一独一无二的、超越其自身出场频率的结构重要性。与之类似，“灭法国国王”、“乌鸡国太子”等将劫难支线连接到取经主线的关键人物，也具有了相对较高的间介中心性。

4 社群的结构：派阀林立的“江湖”版图

在厘清了核心人物的多元重要性之后，我们将目光从“点”转向“面”，探索这些角色是如何“抱团”形成社群。我们使用 Louvain 社群检测算法，将这张复杂的网络剖析为 18 个粒度适中的“人物派阀”[7]。该算法不依赖任何文本，仅仅通过人物之间的关系，就能成功识别出大量独立的故事单元，更揭示了主角团队的功能分工与神魔世界的权力格局，为我们提供了一幅清晰的“社会关系地图”。

值得注意的是，本次社群检测将取经团队进行了功能性的“一分为二”。孙悟空被独立划分至一个独特的社群（社群 15），与他的授业恩师菩提祖师、早期敌人混世魔王以及后期强敌黄眉大王等人聚在一起。这个社群的成员构成，在相当程度上勾勒出了孙悟空的个人英雄轨迹——从学艺、立威到面对重大挑战。这表明，从数据网络的角度看，孙悟空的互动模式与取经队伍其他成员存在显著差异，使他形成了以自我为核心的“行动者”网络。与之相对，唐僧、猪八戒、沙僧与白龙马则被共同归入了另一个庞大的社群（社群 1）。在这个“旅程经历圈”中，与他们紧密相连的，是大量他们在路上被动遭遇的国王、凡人以及作为劫难主体出现的妖怪（如乌鸡国国王、朱紫国国王、盘丝洞蜘蛛精等）。这种划分，从数据层面清晰地反映了团队内部“行动者”与“经历者”的功能定位差异。

除了主角团队的内部分工，这张“地图”也为我们呈现了一幅清晰的神界“政治版图”。庞杂的神仙体系被有效地解构为几个不同的权力中心。例如，以如来佛为核心的“西天佛教集团”（社群 2）囊括了主要的菩萨与罗汉（如文殊菩萨，普贤菩萨）；以玉皇大帝为核心的“天庭道教政府”（社群 4）汇集了天庭的主要官僚和将领（如太白金星，二郎神）；而以观音菩萨为核心的“取经后勤部”则独立成派，其成员多为她的下属、收服的弟子及各路盟友（如红孩儿，木吒），这也在一定程度上印证了她作为取经项目“总负责人”的特殊地位。在更精细的尺度上，该算法依然能够识别出大量独立的故事单元，例如“碧波潭”（社群 7）、“荆棘岭”（社群 8）、“取经缘起”（社群 9）等章节的核心角色都被较为准确地聚类，这客观地反映了小说“单元剧”式的叙事特征。

5 结语

综上所述，本文通过基于人物关系网络的计算性分析，描绘了一幅立体的《西游记》人物关系图景。从宏观上看，这是一个既高度“抱团”，又通过取经主线高效连接起来的“小世界”；在此结构中，中心性分析不仅用数据分析证明了孙悟空在网络中社交、枢纽、信息与影响力四项全能的绝对核心地位，也揭示了如唐太宗这样的“结构性枢纽”人物的隐藏功能；而精细的社群检测，更是从功能上将取经团队一分为二，并清晰地勾勒出神魔世界中派阀林立的政治版图。这种“远读”的方法，将读者凭直觉得到的模糊结论，转化为可以被检验、被量化的命题。它让我们得以从“身在此山中”的沉浸式阅读中抽离，去审视整个叙事大陆的板块构造与山川走向，为这部不朽的古典文学巨著，提供了一种全新的、富有洞见的解读。

参考文献

- [1] Labatut, Vincent, and Xavier Bost. "Extraction and analysis of fictional character networks: A survey." *ACM Computing Surveys (CSUR)* 52.5 (2019): 1-40.
- [2] Yao, Yuan. "journey-to-the-west." GitHub, 2017, github.com/yuany-pku/journey-to-the-west.
- [3] Moretti, Franco. *Distant Reading*. Verso, 2013.
- [4] Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *Nature*, vol. 393, no. 6684, 4 June 1998, pp. 440-42. DOI: 10.1038/30918.
- [5] Milgram, Stanley. "The Small World Problem." *Psychology Today*, vol. 1, May 1967, pp. 60-67.
- [6] Freeman, Linton C. "Centrality in Social Networks: Conceptual Clarification." *Social Networks*, vol. 1, no. 3, 1978, pp. 215-39. DOI: 10.1016/0378-8733(78)90021-7.
- [7] Blondel, Vincent D., et al. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, 17 Oct. 2008, P10008. DOI: 10.1088/1742-5468/2008/10/p10008.