

基于 PageRank 算法的中国菜系食材核心度分析报告

摘要

本项目旨在应用网络科学的分析方法，对中国主流菜系的食材结构进行一次创新的量化探索，以期超越传统的感性认知。我们通过改编 Google 的 PageRank 算法，构建了基于经典菜谱的食材共现网络模型，将食材间的搭配关系转化为网络中的连接，从而计算出每种食材在菜系中的核心度（PageRank 值）。研究选取了鲁、粤、川、淮扬、本帮、东北六大菜系共 48 道经典菜品作为分析样本。

分析结果成功地识别并量化了各大菜系的“风味 DNA”：例如，模型清晰地揭示了辣椒、花椒和豆瓣酱是构成川菜“麻辣”灵魂的绝对核心；高汤与火腿则是淮扬菜追求“鲜和”味觉体验的骨架。此外，通过对“糖”这一特定食材的专题分析，我们量化了其在以“浓油赤酱”为特点的本帮菜和以“咸鲜”为主的鲁菜中的枢纽地位。

本报告的结论证明，该模型为理解、对比和传承中国博大精深的烹饪文化提供了一个新颖、客观且富有洞察力的数据化视角。

1. 引言

1.1. 研究背景

中国饮食文化源远流长，各大菜系以其独特的风味和烹饪技法闻名于世。从川菜的麻辣到粤菜的清鲜，从鲁菜的醇厚到淮扬菜的平和，这些风味特点早已成为人们的共识。然而，长期以来，对菜系的分析与描述大多停留在经验总结和定性描述的层面，如“一菜一格，百菜百味”或“食不厌精，脍不厌细”。这些感性认知虽然精准，但缺乏一个统一的、量化的结构性对比框架，难以精确地回答“一个菜系的风味究竟是由哪些食材决定的？”这一根本问题。

1.2. 问题提出

随着数据科学的发展，我们有机会用全新的视角来解构这一传统文化命题。本研究的核心问题是：

- 我们能否建立一个科学、客观的数学模型，来定义并识别出一个菜系的“核心食材”或“风味 DNA”？
- 我们能否通过这个模型，精确地量化不同菜系在食材运用上的异同点，而不仅仅是停留在模糊的印象层面？

1.3. 项目目标

为解决上述问题，本项目设定了三个明确的目标：

- 建立可复现的分析模型：**基于网络科学理论，改编 PageRank 算法，建立一个能够量化分析菜系中食材重要性的数学模型。

- 2. **识别六大菜系的核心食材：**应用该模型，对鲁、粤、川、淮扬、本帮、东北六大主流菜系的经典菜品进行分析，输出各自的核心食材排行。
- 3. **进行专题对比分析：**以“糖”作为“示踪剂”，深入探索同一种食材在不同菜系文化中所扮演的角色差异，并进行地理位置可视化。

1.4. 报告结构

本报告将首先介绍模型构建的理论基础与方法，随后展示详细的分析结果，包括各菜系的内部核心度分析和跨菜系的专题对比。最后，我们将对模型的洞察与局限性进行讨论，并对未来研究方向进行展望。

2. 模型与方法

2.1. 核心算法：PageRank 简介

PageRank 是 Google 创始人拉里·佩奇（Larry Page）和谢尔盖·布林（Sergey Brin）早期用于解决网页重要性排序的核心算法。其革命性的思想在于，一个网页的重要性（PR 值）并非由其自身内容决定，而是由指向它的其他网页的“数量”和“质量”共同决定的。

其核心理念可以通俗地理解为：“**高质量的投票更有价值**”。假设有四个网页：A（知名大学官网）、B（该校教授主页）、C 和 D（两名新生的个人博客）。如果 A 链接到 B，而 D 链接到 C，尽管 B 和 C 都只获得了一票，但来自权威页面 A 的“投票”显然比来自普通页面 D 的“投票”分量更重。因此，PageRank 算法会判定 B 的重要性远高于 C。正是这种模拟现实世界声望传递的机制，使其能够有效地识别出网络中的核心节点。



图片 1

计算 PageRank 值的公式为：

$$PR(u) = \frac{(1 - d)}{N} + d \sum \left(\frac{PR(v)}{Out(v)} \right)$$

其中， $PR(u)$ 是节点 u 的 PageRank 值， d 是阻尼系数， N 是网络中节点的总数， v 是链接到 u 的节点， $Out(v)$ 是节点 v 的出链数。

2.2. 模型构建的核心假设

在本项目中，我们创造性地将 PageRank 算法从其原始的有向图（网页链接）应用场景，改编至无向图（食材搭配）的分析中。为此，我们建立了一个核心假设：在一道经典菜品中共同出现的食材，可以被视为它们在“烹饪重要性”上进行了一次相互投票。

基于此假设，我们模型计算出的 PageRank 值，其物理意义也发生了转变。它不再是衡量单向的“权威度”，而是量化了食材在整个菜系食材网络中的“**网络中心度**”或“**枢纽性**”。一个食材的 PR 值高，意味着它不仅与众多食材产生搭配，更重要的是，与它搭配的食材本身也都是网络中的“核心成员”。这使得高 PR 值的食材成为支撑起整个菜系风味结构的基础或枢纽。

2.3. 数据集构建

为保证分析的代表性与权威性，我们选取了鲁、粤、川、淮扬、本帮、东北六大主流菜系。我们为每个菜系精心挑选了 8 道公认的、经典的代表菜品，构成了一个总计 48 道菜的分析样本。随后，我们对每道菜品进行拆解，提取其核心原材料，忽略如盐、油、水等过于基础的调味品，最终形成用于模型计算的结构化数据集。（完整数据集见附录 A）

2.4. 实现流程

整个模型的实现流程清晰，主要分为以下几步：

1. **网络构建**：针对每个菜系，我们创建一个独立的无向图。将每种食材作为图中的一个节点。对于任意一道菜，其中出现的所有食材，在图中两两之间建立一条边。如果某两个食材在多道菜中都出现了，则它们之间边的权重相应增加。
2. **算法执行**：我们使用 Python 的 NetworkX 库，这是一个强大的网络分析工具包。我们调用其内置的 pagerank 函数，并将边的权重考虑在内，在每个菜系的网络图上独立运行算法，得出每种食材的 PR 值。
3. **可视化呈现**：为清晰展示结果，我们使用了两个 Python 库。Matplotlib 用于绘制各菜系内部所有原材料的 PR 值条形图，Folium 则用于创建交互式的地理信息地图，将特定食材的分析结果与地理空间相结合。

3. 结果与分析

3.1. 各菜系内部食材核心度分析

通过对六大菜系分别建模计算，我们得到了各自的食材核心度排行，其结果与我们的烹饪直觉高度吻合。

- **川菜**：辣椒、花椒、豆瓣酱的 PR 值遥遥领先，这三种食材构成了川菜“麻辣”风味的铁三角，是几乎所有经典川菜都离不开的灵魂。

- **粤菜：**姜、葱、蒜位列前茅。这清晰地反映了粤菜烹饪哲学——注重食材本味，多用简单的香辛料来“提鲜”而非覆盖。
- **鲁菜：**姜、葱、酱油、糖、醋均表现出极高的核心度。这揭示了鲁菜味型丰富、咸鲜为基、善用糖醋的特点。
- **淮扬菜：**高汤、火腿、笋的核心地位凸显。这三种食材都是“鲜味”的重要来源，印证了淮扬菜对“和、精、清、新”的极致追求。
- **本帮菜：**酱油和糖的 PR 值极高，完美地量化了其“浓油赤酱、咸中带甜”的标志性风味特征。
- **东北菜：**酱油、酸菜、八角成为核心。这反映了东北菜味重咸香、偏爱炖菜和发酵风味的烹饪习惯。



图片 2

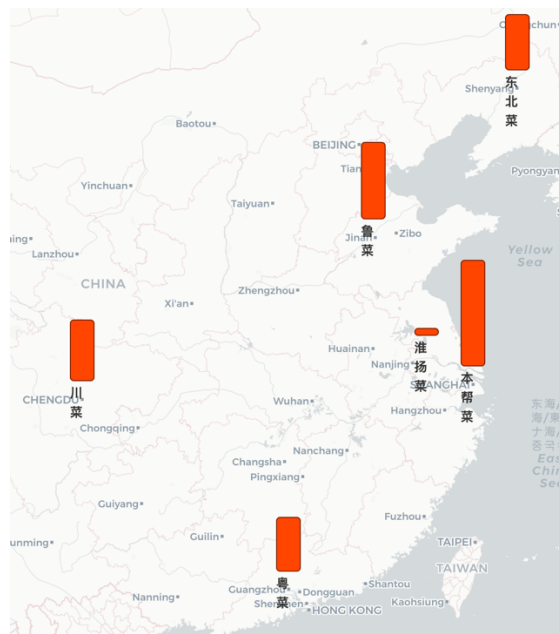
3.2. 专题分析：以“糖”为示踪剂的跨菜系对比

为了更深入地进行横向对比，我们选取“糖”作为特定分析对象。量化结果显示，“糖”在六大菜系中的核心度（PR 值）排序为：

本帮菜 > 鲁菜 > 东北菜 > 粤菜 > 川菜 > 淮扬菜。

我们将此结果绘制在地理信息地图上。可以清晰地看到，代表 PR 值的柱状图在上海（本帮菜）和济南（鲁菜）最高，这与“南甜北咸”的传统认知有所不同，揭示了“糖”在鲁菜复杂的味型平衡中扮演着与本帮菜同等重要的基础角色。而在成都（川菜）和广州（粤

菜），柱状图的高度中等，说明糖在这些菜系中多作为平衡辣味（如宫保鸡丁）或提鲜（如咕咾肉）的辅助角色出现。最引人注目的是，在扬州（淮扬菜），代表“糖”的柱子几乎为零，有力地证明了传统淮扬菜对“本味”的尊重，极少使用糖来直接调味。



图片 3

4. 讨论

4.1. 模型揭示的关键洞察

本模型不仅成功量化了各大菜系的“风味名片”（如川菜的花椒），也清晰地揭示了中餐烹饪的“通用语”（如姜、葱、蒜、酱油在多个菜系中均有较高的 PR 值）。它证明了，尽管风味各异，但中国主流烹饪体系共享着一个共同的鲜香基底。此外，模型也为我们打破一些刻板印象提供了数据支持，例如“糖”在北方菜系鲁菜中的重要性。

4.2. 模型的局限性与批判性思考

我们必须承认，本模型存在一些局限性：

- 数据样本的局限：**48 道菜虽具代表性，但对于一个庞大的菜系而言，样本量依然有限。更广泛的数据集可能会带来更精确的结果。
- 算法改编的讨论：**如前所述，我们将 PageRank 应用于无向图是一种功能性的改编。因此，结果必须被严谨地解读为食材在网络结构中的“中心度”。
- 原材料拆解的主观性：**将一道菜拆解为核心原材料的过程，不可避免地带有主观判断，这可能会对结果产生细微影响。

5. 结论与展望

5.1. 主要结论

本项目成功地构建并验证了一个基于 PageRank 算法的创新模型，用于量化分析和对比中国菜系的食材核心度。模型结果不仅与传统烹饪知识和美食家的直觉高度吻合，还提供了一些超越感性认知的数据洞察。我们证明了，网络科学方法可以作为一种强大的新工具，用于解构和理解复杂的烹饪文化体系。

5.2. 未来工作展望

未来的研究可以从以下三个方向进行深化：

- 1. **扩大数据集：**通过爬虫技术获取大型美食网站的全量菜谱数据，构建一个包含成千上万节点的超大型食材网络，其分析结果将更具统计意义。
- 2. **引入更多算法：**除了 PageRank，可以引入网络科学中其他的中心度指标（如介数中心性、特征向量中心性）进行交叉验证，从不同维度更全面地评估食材的重要性。
- 3. **构建有向图：**这是一个更具挑战性的方向。通过自然语言处理技术分析烹饪步骤，定义食材间“风味赋予”的方向关系，构建一个有向图。这将使模型在概念上更贴合 PageRank 的原始设计，可能会带来更深层次的发现。

附录

附录 A：完整数据集：

菜系	代表菜品	核心原材料
鲁菜	九转大肠	猪大肠 糖 醋 葱 姜
	葱烧海参	海参 大葱 酱油 姜 高汤
	糖醋鲤鱼	鲤鱼 糖 醋 酱油 姜
	油爆双脆	猪肚 鸡胗 葱 姜 蒜
	德州扒鸡	鸡 酱油 糖 八角 桂皮
	四喜丸子	猪肉 鸡蛋 葱 姜 酱油
	爆炒腰花	猪腰 蒜 姜 醋 酱油
	糟溜鱼片	鱼 酒酿 鸡蛋 木耳 高汤
粤菜	白切鸡	鸡 姜 葱 蒜 香油
	烤乳猪	乳猪 糖 醋 芝麻酱 五香粉
	老火靚汤	猪骨 鸡爪 药材 蜜枣 姜
	清蒸石斑鱼	石斑鱼 姜 葱 酱油 蒸鱼豉油

	深井烧鹅	鹅 梅子酱 糖 醋 五香粉
	菠萝咕咾肉	猪肉 菠萝 鸡蛋 番茄酱 醋
	豉汁排骨	排骨 豆豉 蒜 辣椒 葱
	菜心炒牛肉	牛肉 菜心 姜 蒜 蚝油
川菜	麻婆豆腐	豆腐 牛肉 豆瓣酱 花椒 辣椒
	回锅肉	猪肉 蒜苗 豆瓣酱 甜面酱 辣椒
	宫保鸡丁	鸡肉 花生 辣椒 花椒 糖 醋
	夫妻肺片	牛肉 牛杂 辣椒油 花椒 花生
	水煮牛肉	牛肉 豆芽 豆瓣酱 辣椒 花椒
	鱼香肉丝	猪肉 木耳 笋 豆瓣酱 糖 醋
	酸菜鱼	鱼 酸菜 辣椒 花椒 姜
	辣子鸡	鸡肉 辣椒 花椒 蒜 姜
淮扬菜	清炖蟹粉狮子头	猪肉 蟹粉 白菜 高汤 姜
	大煮干丝	豆腐干 鸡丝 火腿 虾仁 高汤
	三套鸭	鸭 鸽子 火腿 笋 高汤
	软兜长鱼	鳝鱼 蒜 醋 胡椒粉 酱油
	扬州炒饭	米饭 鸡蛋 虾仁 火腿 海参
	文思豆腐	豆腐 火腿 鸡丝 笋 高汤
	水晶肴肉	猪蹄 八角 葱 姜 香醋
	平桥豆腐羹	豆腐 鲫鱼 虾仁 香菇 火腿
本帮菜	本帮红烧肉	猪肉 酱油 糖 黄酒 姜
	油爆虾	河虾 糖 醋 葱 姜
	腌笃鲜	咸肉 鲜猪肉 笋 百叶结 黄酒
	响油鳝糊	鳝鱼 蒜 姜 酱油 胡椒粉
	糖醋小排	排骨 糖 醋 酱油 姜
	草头圈子	猪大肠 草头 酱油 糖 蒜
	八宝辣酱	鸡胗 猪肚 花生 豆腐干 笋 甜面酱
	上海熏鱼	鱼 酱油 糖 八角 葱

东北菜	锅包肉	猪肉 糖 醋 姜 葱
	猪肉炖粉条	猪肉 粉条 酸菜 酱油 八角
	小鸡炖蘑菇	鸡 蘑菇 粉条 酱油 八角
	地三鲜	土豆 茄子 青椒 蒜 酱油
	酸菜白肉血肠	猪肉 酸菜 血肠 姜 葱
	东北大拉皮	拉皮 黄瓜 肉丝 蒜 麻酱
	杀猪菜	猪肉 猪肝 猪大肠 酸菜 血肠
	酱大骨	猪骨 酱油 糖 八角 桂皮

表格 1