基于灰度肖像特征的 NBA 75 球员六类流形结构分析

zqb[1*]

[1*]WormForce.corp, China.

Corresponding author(s). E-mail(s): kobe@helicopter.elbow;

摘要

本文以 NBA 75 名单中可获得官方标准肖像的 74 名球员为样本，建立“像素特征-测地距离-二维投影-距离聚类-六相映射”的完整研究流程。全部图像统一灰度并标准化到 1040x760，进一步下采样到 80x80 后向量化为 6400 维表征；在欧氏距离基础上构建 k 近邻图并计算最短路径，得到测地距离矩阵；采用 Classical MDS 投影到二维命势平面，并在该平面对应距离结构上使用 k-medoids 聚类。六相标签（富相、贵相、寿相、贫贱相、孤苦相、夭相）由簇级分离比 $\boldsymbol{R_c = \mathrm{inter}_c/\mathrm{intra}_c}$ 的降序结果生成。结果表明，Cluster 2 的分离比最高（2.2665），Cluster 4 最低（1.3367）；最近样本对为 Kobe Bryant–Tim Duncan（28.5800），最远样本对为 Clyde Drexler–James Worthy（227.6564）。本文进一步对 14 张图谱进行主题化合并分析，从全局几何、局部扩散、网络骨架、纹理统计和方向分布五个层面构建六相命势叙事，并形成逻辑闭合的结论体系。

Keywords: NBA 75, 肖像流形, 测地距离, Classical MDS, k-medoids, 六相映射

1 引言

在统一来源的人像数据中，像素差异并非随机噪声，而是可以被解释为一类稳定的结构信号。本文将这一结构信号定义为“命势场”中的几何位势：当两个球员肖像在整体轮廓、灰度分布、局部纹理上更接近时，其位势距离更短；反之更长。研究目标不是判断谁“长得像谁”，而是构建一套可计算的六相秩序，使“面相”从形容词转化为结构变量。

本文采用的基本思想是“先建场、后分层、再命名”。先在高维像素空间计算样本间原始距离，再通过图最短路径重建“绕行成本”意义上的测地距离，以保留群体几何中的连通拓扑；随后将该拓扑压缩到二维平面，使全局结构可视化；最后在距离空间中直接聚类，并根据簇级指标完成六相映射。这样，六相并不是先验符号，而是几何场中自然出现的六个统计稳定区。

2 数据与方法

2.1 数据集与特征表示

研究样本为 74 名球员，全部来自官方标准肖像。每张图像统一为灰度（L 模式）并缩放至 1040x760。为保证计算可控且保留主体结构，图像下采样为 80x80，并展平为向量：

$$\mathbf{x}_i \in \mathbb{R}^{6400}.$$

随后执行逐样本标准化（减均值、除标准差），从而将亮度基线差异与结构差异分离。经过该处理后，每个样本的特征都可理解为“形貌起伏曲线”，而非“绝对明暗值”。

2.2 距离构建

首先定义欧氏距离：

$$d_E(i,j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2.$$

欧氏距离作为初始相异度定义，用于建立后续流形重建的局部邻域基础Tenenbaum et al. (2000)。在 d_E 上构建无向 kNN 图（最小连通参数最终为 $k = 8$），并通过全局最短路径获得测地距离：

$$d_G(i,j) = \mathrm{ShortestPath}_{kNN}(d_E).$$

测地距离的计算对应 Floyd 与 Warshall 的最短路径框架，并与流形学习中的全局几何重建思想一致Floyd (1962); Warshall (1962); Tenenbaum et al. (2000)。这一步的意义是把“直线差异”升级为“地形绕行成本”。因此，若两个样本虽然像素差异不小，但处于同一连通团块，则测地距离仍可能较短；反之则显著增大。

2.3 二维投影与坐标生成个例（Kobe Bryant）

为了把“图像如何变成坐标”解释清楚，本文对 Kobe Bryant 给出完整链路。设原始灰度肖像为

$$\mathbf{I}_i \in \mathbb{R}^{1040\times 760},$$

则处理流程可写为：

$$\mathbf{I}_i \xrightarrow{\text{bilinear resize}} \mathbf{I}_i^{80} \in \mathbb{R}^{80\times 80} \xrightarrow{\text{flatten}} \mathbf{x}_i \in \mathbb{R}^{6400} \xrightarrow{\text{standardize}} \mathbf{z}_i.$$

标准化定义为

$$\mathbf{z}_i = \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}.$$

随后在全部样本上计算欧氏距离矩阵 $\mathbf{D}_E$，构建 kNN 图并求最短路径得到测地距离矩阵 $\mathbf{D}_G$。在 $\mathbf{D}_G$ 上执行 Classical MDS：

$$\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{D}_G^2\mathbf{J}, \qquad \mathbf{J} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T,$$

该步骤属于 Classical MDS 的标准双中心化形式，其理论基础可追溯至多维尺度分析经典文献Torgerson (1952)。取前两特征方向得到二维坐标 (x_i, y_i)。

图 1 逐步展示了 Kobe Bryant 的完整转换过程：原始图（1040x760）、下采样图（80x80）、6400 维向量、标准化后的分布，以及在二维流形中的最终坐标点。对应数值为：$\mu(\mathbf{x}_{\text{Kobe}}) = 0.1558$、$\sigma(\mathbf{x}_{\text{Kobe}}) = 0.2405$、$\mu(\mathbf{z}_{\text{Kobe}}) = 0$、$\sigma(\mathbf{z}_{\text{Kobe}}) = 1$，最终投影坐标为

$$(x_{\text{Kobe}}, y_{\text{Kobe}}) = (-29.13,\ 20.51).$$

该点位于中心高密团块边缘，且在统计上归属 Cluster 4（天相），其 closeness 为 0.011998，Q25 邻域内近邻数为 40，说明其在命势场中兼具中心可达性与局部连接活性。

2.4 聚类与六相映射

聚类使用 k-medoids，直接在测地距离空间优化簇内到 medoid 的距离和。k 搜索显示自动最优为 k=2（silhouette=0.2768），本文按照研究任务固定采用 k=6，以获得六相解释粒度。定义簇级分离比：其中 k-medoids 聚类采用 PAM 系列的经典定义与实现框架Kaufman and Rousseeuw (1990)。

$$R_c = \frac{\text{inter}_c}{\text{intra}_c},$$

其中 inter_c 表示簇 c 到其他簇的平均距离，intra_c 表示簇内平均距离。按 R_c 降序依次映射：富相、贵相、寿相、贫贱相、孤苦相、夭相。

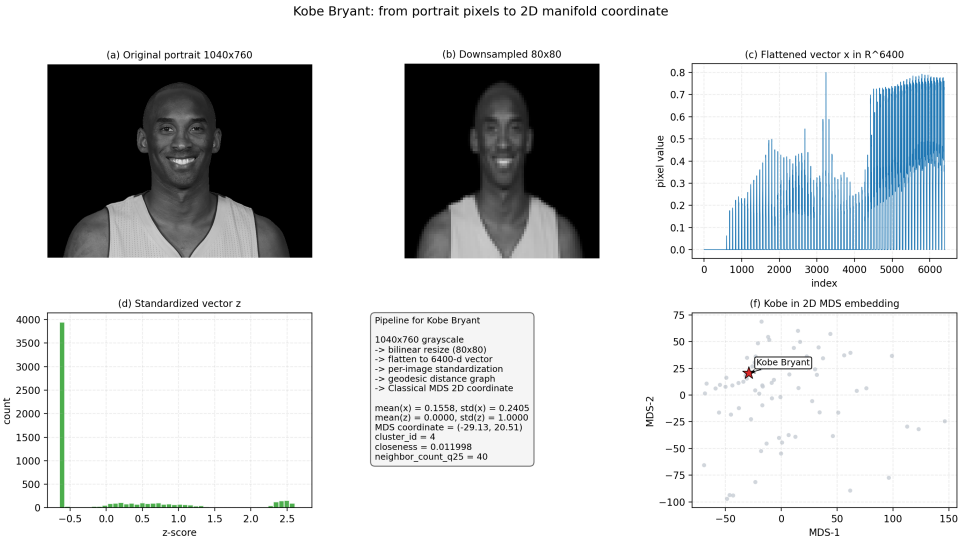

图 1 Kobe Bryant 坐标生成全链路：1040x760 灰度肖像 → 80x80 下采样 → 6400 维向量 → 标准化分布 → 二维 MDS 坐标。

3 六相映射与全局锚点

表 1 六相映射结果（按 R_c 降序）

排序	Cluster	R_c	对应面相	size	medoid
1	2	2.2665	富相	6	Kareem Abdul-Jabbar
2	0	1.4295	贵相	10	Bill Sharman
3	3	1.4072	寿相	4	Moses Malone
4	1	1.4006	贫贱相	8	Dennis Rodman
5	5	1.3783	孤苦相	16	Shaquille O'Neal
6	4	1.3367	夭相	30	Reggie Miller

全局锚点如下：样本数 74；Q25 距离阈值 89.692893；最近对 Kobe Bryant–Tim Duncan（28.5800）；最远对 Clyde Drexler–James Worthy（227.6564）。在极值意义上，x 轴范围为 [-69.043, 146.329]，y 轴范围为 [-97.215, 68.716]。这些数字为后续所有图谱解释提供统一标尺。

4 结果：面相命势场的多层联合分析

4.1 全局命势平面与六相成形机制

图 2 将初始投影图与 k=6 聚类图按纵向全宽展示。先看上图，点云表现为“中心团块 + 边缘长尾”的典型流形结构，说明样本总体并非均匀分布，而是围绕主核展开。再看下图，六簇并非随机切分，而是沿该流形方向形成主核、过渡层与外缘层三级形态：Cluster 4 和 Cluster 5 构成主连接带，Cluster 2 与 Cluster 3 则更接近外层模块。

这一结构与簇规模一致：Cluster 4=30、Cluster 5=16、Cluster 0=10、Cluster 1=8、Cluster 2=6、Cluster 3=4。由此可见，六相不是六个同体量类别，而是一个“核心大相团 + 外围特征团”的分层系统。具体地，夭相（Cluster 4）提供中心密度，孤苦相（Cluster 5）提供次级包络，富相（Cluster 2）与寿相（Cluster 3）提供高辨识度边缘支点。

4.2 分离比秩序：六相排序的统计底盘

若将六相看作“结构能级”，则分离比 R_c 就是其能级排序依据。图 3 的热力图给出了簇内/簇间块结构，散点图给出了紧致度-分离度联合位置。可直接观察到 Cluster 2 同时满足最低簇内均值（66.7879）和最高簇间均值（151.3769），因此其 $R_2 = 2.2665$ 显著领先。相反，Cluster 4 的内外差值最小，对应最低分离比 1.3367。

从中间层看，Cluster 0（1.4295）、Cluster 3（1.4072）、Cluster 1（1.4006）、Cluster 5（1.3783）形成连续梯度，没有出现断裂。这种“单峰高值 + 平滑中段 + 低值大簇”的排序模式，使六相映射具有明确秩序：富相占据高分离端，夭相占据高可达端，其他四相在二者之间呈阶梯过渡。

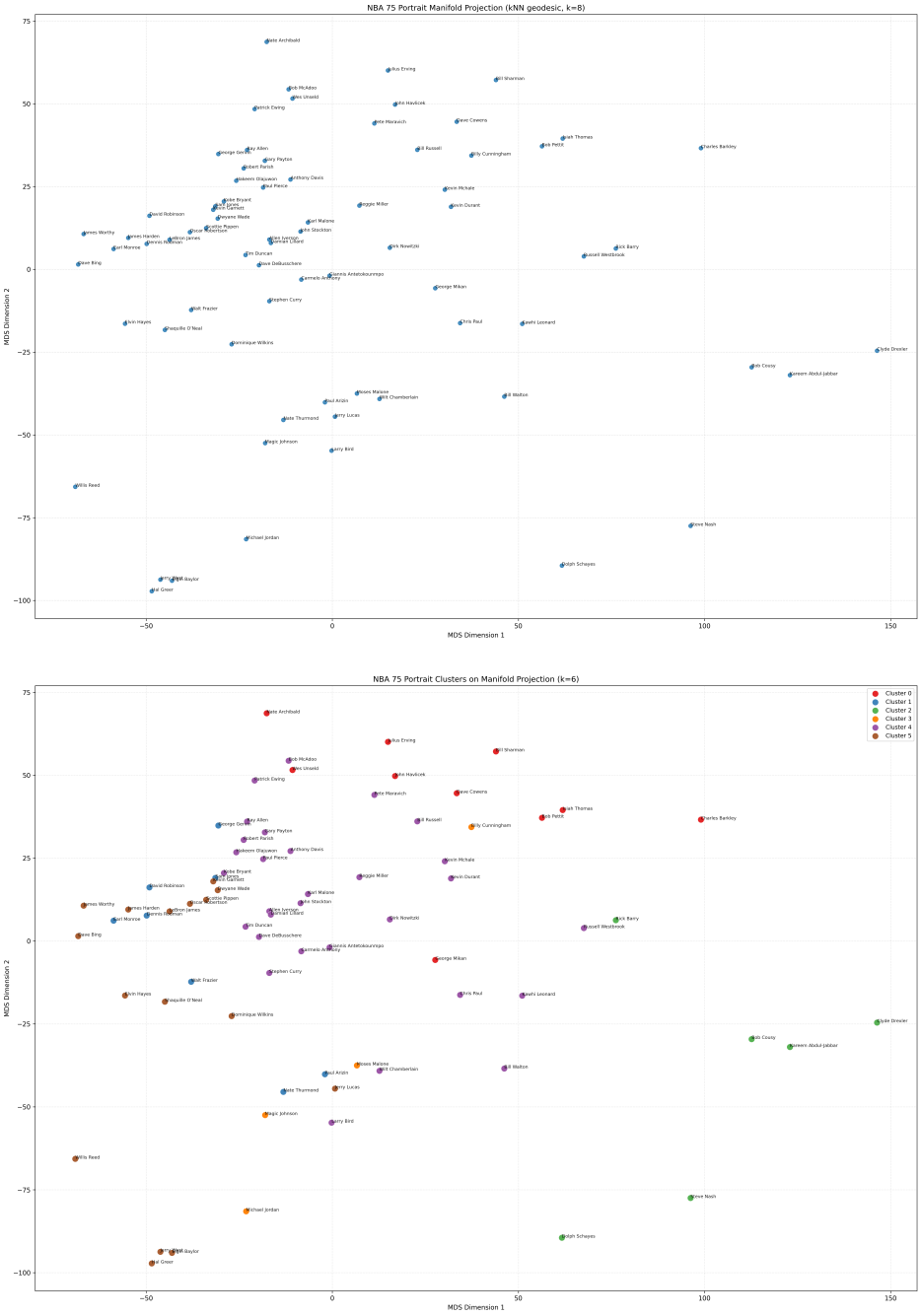

图 2 全局命势场的形成过程：上图为二维流形投影，下图为同平面的 k=6 聚类结果（纵向全宽展示）。

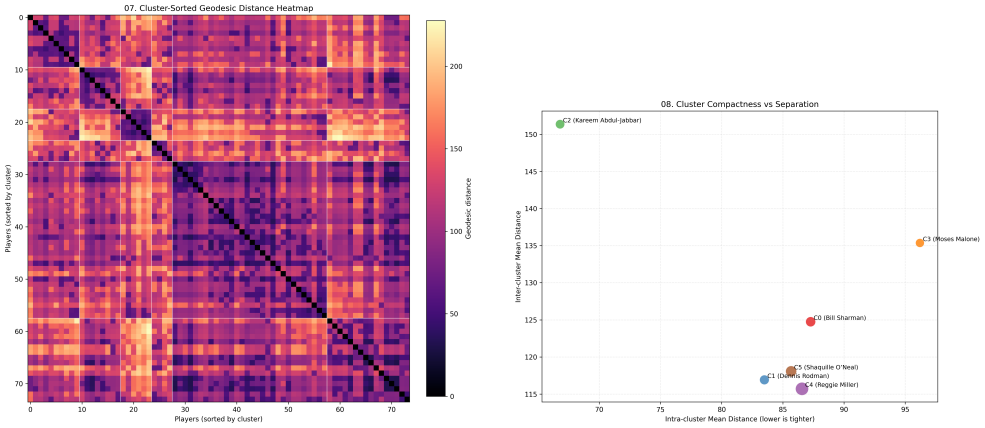

图 3 六相排序的统计底盘：左图为测地距离块结构，右图为簇级紧致度-分离度平面。

4.3 中心-边缘双轨：孤立度与可达度的反向耦合

图 4 显示了最直接的双轨结构。孤立度（avg_distance）前列集中在 Clyde Drexler（162.4059）、Bob Cousy（158.3207）、Moses Malone（145.2187）、Steve Nash（144.6232）、Kareem Abdul-Jabbar（144.2792）；而可达度（closeness）前列集中在 Reggie Miller（0.012913）、Tim Duncan（0.012742）、John Stockton（0.012245）、Kobe Bryant（0.011998）、Dwyane Wade（0.011994）。

这组对照说明：外缘高分离样本并不承担中心传播，中心传播样本也未必具有高边缘辨识度。换言之，命势场不是单轴系统，而是“双轨耦合系统”：一条轨道负责拉开尺度（边缘轨），另一条轨道负责维持连通（核心轨）。六相恰好对应这两条轨道的不同组合位姿。

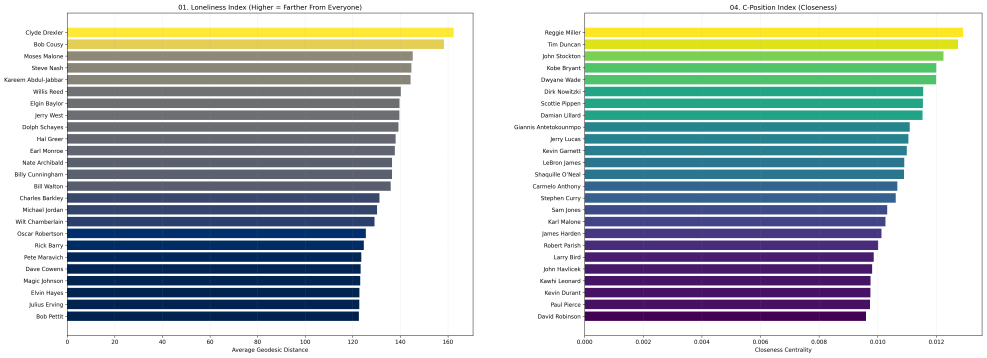

图 4 中心-边缘双轨结构：左图为孤立度，右图为可达度。

4.4 局部扩散与跨簇协商：谁在连接谁

图 5 描述了局部邻域的“扩散协商机制”。在社交气泡图中，近邻数最高的是 Reggie Miller（45），其次为 John Stockton 与 Tim Duncan（均为 43）；最低端则大量聚集在 8 邻域。这意味着核心簇在局部范围内形成高密短链网络，而外缘簇更依赖跨区跳转。

跨簇率图进一步给出簇级分工：C3=0.8000 最高，呈桥接簇；C4=0.2933 最低，呈核心吸附簇；C1=0.6750 与 C0=0.5700 充当中间协商层。由此可以把六相看成一个带“路由权限”的拓扑社会：寿相负责跨域转发，夭相负责中心存储，贵相与贫贱相负责中层交换，富相与孤苦相在外层提供边界信号。

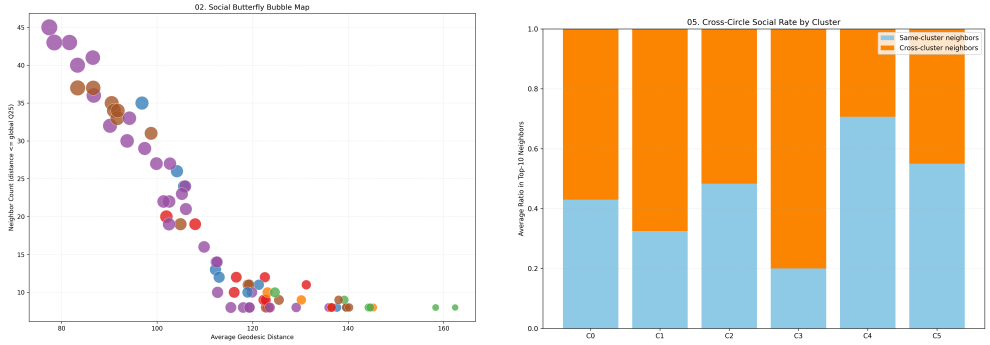

图 5 局部扩散机制：近邻规模-可达度联合分布与簇级跨簇连接比例。

4.5 骨架网络与外轮廓：命势场的张力来源

图 6 的左图仅保留最短 60 条边，右图给出每个样本的最大测地距离。网络侧显示，主骨架由 C4-C5 连接构成，体现核心簇与次核心簇的耦合支撑；边缘侧显示，最大离散张力主要来自 Clyde Drexler 与 James Worthy（227.6564 并列）、Dave Bing（224.6149）、Earl Monroe（223.9978）。

当把二者合并解读时，可以得到“骨架-外轮廓互补律”：骨架负责连通稳定，外轮廓负责尺度张开。若无骨架，命势场会分裂成局部孤岛；若无外轮廓，命势场会塌缩为单一中心团块。六相的稳定存在正是这两种力量平衡的结果。

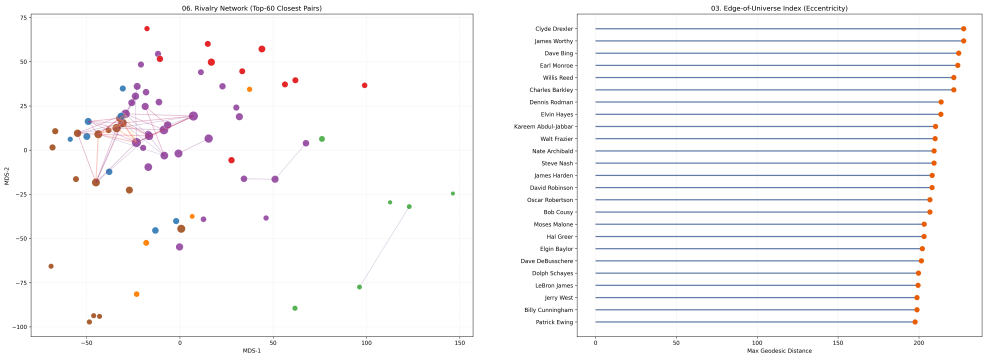

图 6 命势场张力结构：左图为最短边网络骨架，右图为边缘指数分布。

4.6 面相气候层 I：对称性与色调谱

在几何层之外，图 7 进入“气候层”分析。对称性方面，簇均值最高为 C4=0.7854，其次 C2=0.7520，最低 C3=0.5040；个体上界为 Damian Lillard（0.9603），下界为 Jerry West（0.1948）。这表明核心簇具有较强形貌稳定性，而桥接簇保留更高结构波动。

色调方面，亮度均值最高 C0=0.2129、最低 C2=0.1365；对比度最高 C1=0.2987、最低 C2=0.2357；熵最高 C3=4.1343、最低 C2=3.4536。与分离比联合观察可见：高分离并不等于高亮度或高熵，而是由多指标共同配平得到。于是六相不再是单维标签，而是“几何分层 + 气候风格”的双系统编码。

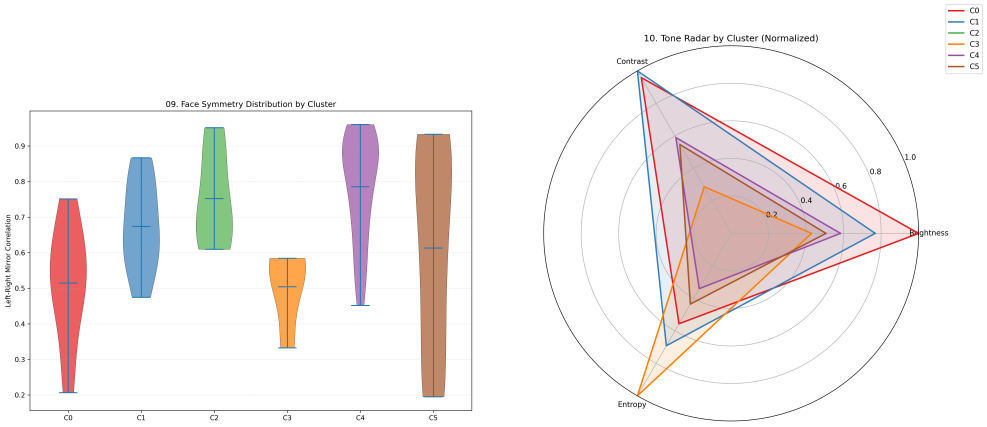

图 7 面相气候层 I：左图为对称性分布，右图为亮度/对比度/熵轮廓。

4.7 面相气候层 II：纹理强度与方向扇区

图 8 给出气候层的第二组证据。纹理方面，edge_density 均值最高 C3=0.09698，最低 C5=0.06526；laplacian_variance 均值最高 C3=0.01963，最低 C5=0.00566。个体极值上，Dave DeBusschere 在 edge_density 上更突出，George Gervin 在 laplacian_variance 上达到最高值（0.05805）。这意味着寿相簇在局部纹理起伏上具备高活性特征。

方向方面，玫瑰图 12 扇区计数为 [4, 8, 4, 7, 8, 15, 5, 2, 9, 3, 3, 6]，主峰位于 H6（15），低谷位于 H8（2）。方向分布显著非均匀，说明命势场在方位上具有主轴偏置。纹理强度与方向偏置叠加后，六相获得了“局部起伏-全局指向”的双重解释：某些相类通过纹理增强辨识，某些相类通过方位聚集形成群体势能。

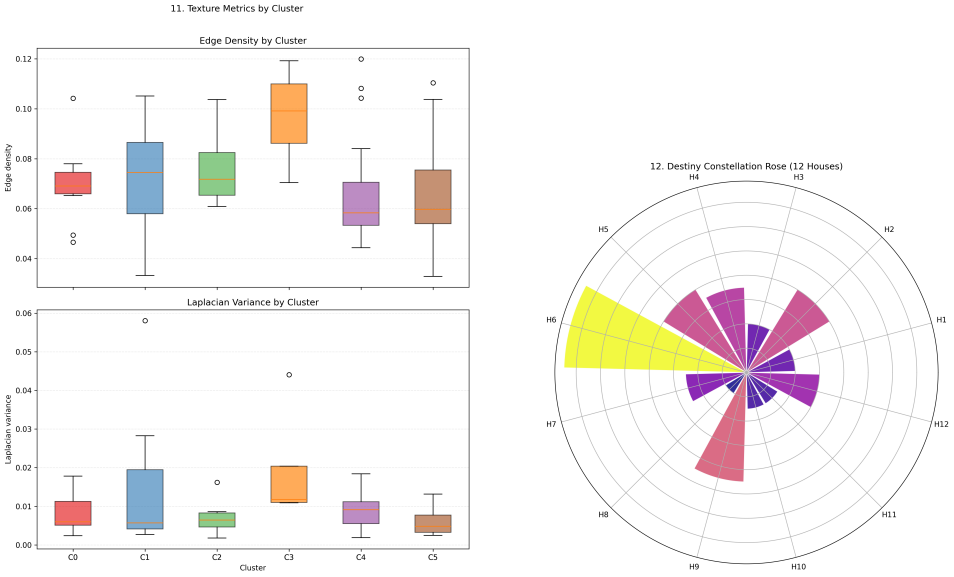

图 8 面相气候层 II：左图为纹理强度，右图为极角方向分布。

5 结论

本文完成了一个可复算、可投影、可聚类、可叙事的六相命势场分析框架，主要结论如下：

1. 六相来源于距离结构而非先验命名，映射规则由 R_c 排序唯一确定。
2. 全局命势场呈“核心大团 + 外缘小团”分层格局，其中夭相占据中心主核，富相占据高分离外层。

3. 命势场存在稳定的双轨机制：孤立度轨负责尺度张开，可达度轨负责连通维持。
4. 局部扩散层显示明确簇级分工：桥接簇负责跨域协商，核心簇负责局部吸附，外围簇负责边界识别。
5. 纹理与方向两类气候指标为六相提供附加结构解释，使结论从“距离分类”提升为“几何-网络-气候”联合分类。

附录 A：主要公式

$$\mathbf{x}_i \in \mathbb{R}^{6400}, \tag{1}$$

$$d_E(i,j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2, \tag{2}$$

$$d_G(i,j) = \mathrm{ShortestPath}_{kNN}(d_E), \tag{3}$$

$$\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{D}^2\mathbf{J}, \tag{4}$$

$$R_c = \frac{\mathrm{inter}_c}{\mathrm{intra}_c}. \tag{5}$$

附录 B：六簇成员列表

表 2: Cluster 成员明细

Cluster	Members
C0（贵相）	Bill Sharman; Bob Pettit; Charles Barkley; Dave Cowens; George Mikan; Isiah Thomas; John Havlicek; Julius Erving; Nate Archibald; Wes Unseld.
C1（贫贱相）	David Robinson; Dennis Rodman; Earl Monroe; George Gervin; Nate Thurmond; Paul Arizin; Sam Jones; Walt Frazier.
C2（富相）	Bob Cousy; Clyde Drexler; Dolph Schayes; Kareem Abdul-Jabbar; Rick Barry; Steve Nash.
C3（寿相）	Billy Cunningham; Magic Johnson; Michael Jordan; Moses Malone.
C4（夭相）	Allen Iverson; Anthony Davis; Bill Russell; Bill Walton; Bob McAdoo; Carmelo Anthony; Chris Paul; Damian Lillard; Dave DeBusschere; Dirk Nowitzki; Gary Payton; Giannis Antetokounmpo; Hakeem Olajuwon; John Stockton; Karl Malone; Kawhi Leonard; Kevin Durant; Kevin Mchale; Kobe Bryant; Larry Bird; Patrick Ewing; Paul Pierce; Pete Maravich; Ray Allen; Reggie Miller; Robert Parish; Russell Westbrook; Stephen Curry; Tim Duncan; Wilt Chamberlain.

Cluster	Members
C5（孤 苦相）	Dave Bing; Dominique Wilkins; Dwyane Wade; Elgin Baylor; Elvin Hayes; Hal Greer; James Harden; James Worthy; Jerry Lucas; Jerry West; Kevin Garnett; LeBron James; Oscar Robertson; Scottie Pippen; Shaquille O'Neal; Willis Reed.

References

Floyd RW (1962) Algorithm 97: Shortest path. Communications of the ACM 5(6):345. https://doi.org/10.1145/367766.368168

Kaufman L, Rousseeuw PJ (1990) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York

Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323. https://doi.org/10.1126/science.290.5500.2319

Torgerson WS (1952) Multidimensional scaling: I. theory and method. Psychometrika 17(4):401–419. https://doi.org/10.1007/BF02288916

Warshall S (1962) A theorem on boolean matrices. Journal of the ACM 9(1):11–12. https://doi.org/10.1145/321105.321107