

# 用小圖樣描述異質網路：graphlets與軌道

arXiv:2304.14268v3 (2023) | 兩頁中文導讀

## 研究問題：網路不只是有沒有連線

社會、生物及知識網路中，節點可能有不同類別，邊也可能有類型或方向。若忽略這些資訊，相同的拓撲圖樣可能代表不同關係。本文建立異質graphlets與節點軌道的列舉、辨識及計數流程，並介紹GEOMINE工具，目標是用局部小圖樣把複雜網路轉成可比較的特徵。

## graphlet是小型誘導子圖

取少數節點後，保留原網路在這些節點之間的所有邊，所得就是誘導子圖；不能任意刪邊來湊出想要的圖樣。通常只列連通的小圖樣，按同構歸類。異質版本進一步要求同構保留節點顏色、邊類型及方向，所以同一個無色圖形可能分裂成多個不同類型。

## 軌道區分同一圖樣中的角色

無色三點路徑的兩端可由自同構交換，屬同一軌道；中間點則是另一軌道。如果兩端顏色不同，保色自同構不能交換它們，原本同一角色就被拆開。軌道因而不是手動命名的社會角色，而是由圖形自同構群精確決定的等價類。這使特徵具有不依賴節點編號的意義。

## 用矩陣記錄顏色與連線

作者將節點類型與邊資料編碼成矩陣：對角存節點類型，非對角存邊類型或無邊標記；有向圖則必須保留矩陣的不對稱資訊。節點重新編號對應同時置換行列。實作時需清楚區別「無邊」與合法類型編號，否則可能把不同圖樣誤認為相同，或在讀取時遺失一種顏色。

## 標準形式消除重新編號造成的重複

對所有允許節點置換形成的矩陣，選字典序最小者作為canonical form。兩個小圖同構，當且僅當具有相同標準形式。若要識別某個根節點的軌道，就固定根的位置，只置換其他節點。圖樣很小時可直接枚舉置換；這也說明其成本隨圖樣大小迅速上升，不宜把方法理解為任意大子圖都能輕鬆處理。

# 用小圖樣描述異質網路：graphlets與軌道

arXiv:2304.14268v3 (2023) | 兩頁中文導讀

## 列舉所有圖樣與計數是兩個步驟

列舉階段先產生合法顏色及邊配置，以標準形式去重，並快取對應的圖樣與軌道索引。計數階段在目標網路取小型誘導子圖，查表辨認類型，再把各節點對應的軌道計數累加。分開這兩步，可以重用同一組參數的圖樣字典，而不必每次分析新網路都重新列舉。

## 輸出：每個節點的一支特徵向量

向量第*i*欄表示節點參與第*i*種graphlet軌道的次數。它同時反映局部連線結構及類型搭配，可再用於節點比較、聚類或後續模型。特徵本身不是因果效應，也不自動保證分類準確率。不同網路比較時需統一圖樣大小、類型定義及方向規則，必要時處理規模造成的計數差異。

## 來源示例與實際工具定位

文章展示62個節點的海豚網路，以性別資訊引入異質性，也使用10節點、三種節點類型及兩種邊類型的合成例子，說明輸出如何變得更細緻。GEOMINE是作者提供的實作工具；本導讀說明來源中的演算法與介面概念，沒有實際安裝最新版套件或驗證其目前API，亦不宣稱完成獨立效能測試。

## 數量表與複雜度應審慎解讀

圖樣種類會隨節點數、顏色數、邊類型與方向快速增加。來源不同段落有若干列舉數量及參數不一致，例如900對應的邊類型數在文字與表格不同；本導讀因此不以那些矛盾數字作宣傳。計數效率也受輸入圖密度及子圖大小影響，不能只憑概括描述就當作已有對任意輸入的線性時間保證。

## 研究價值與可延伸工作

本篇把異質資訊納入圖同構及自同構軌道，提供可重複使用的局部描述語言。後續可研究更好的同構標準化、快取策略、稀疏網路計數，以及有統計基準的特徵比較。來源：Cleveland、Lee、Tsai、Yu、Lee，arXiv:2304.14268v3（2023年6月），16頁；前兩位為共同第一作者。