# Лабораторна робота №6 – Кластеризація на основі матриці подібності

## 1. Вступ

У цій лабораторній роботі розглядається задача кластеризації даних на основі матриці подібності. Для цього використовується графова модель — кожна точка даних представлена як вузол графа, а ребра між вузлами визначаються через функцію подібності. Кластери визначаються як компоненти зв’язності побудованого графа.

## 2. Теоретичні відомості

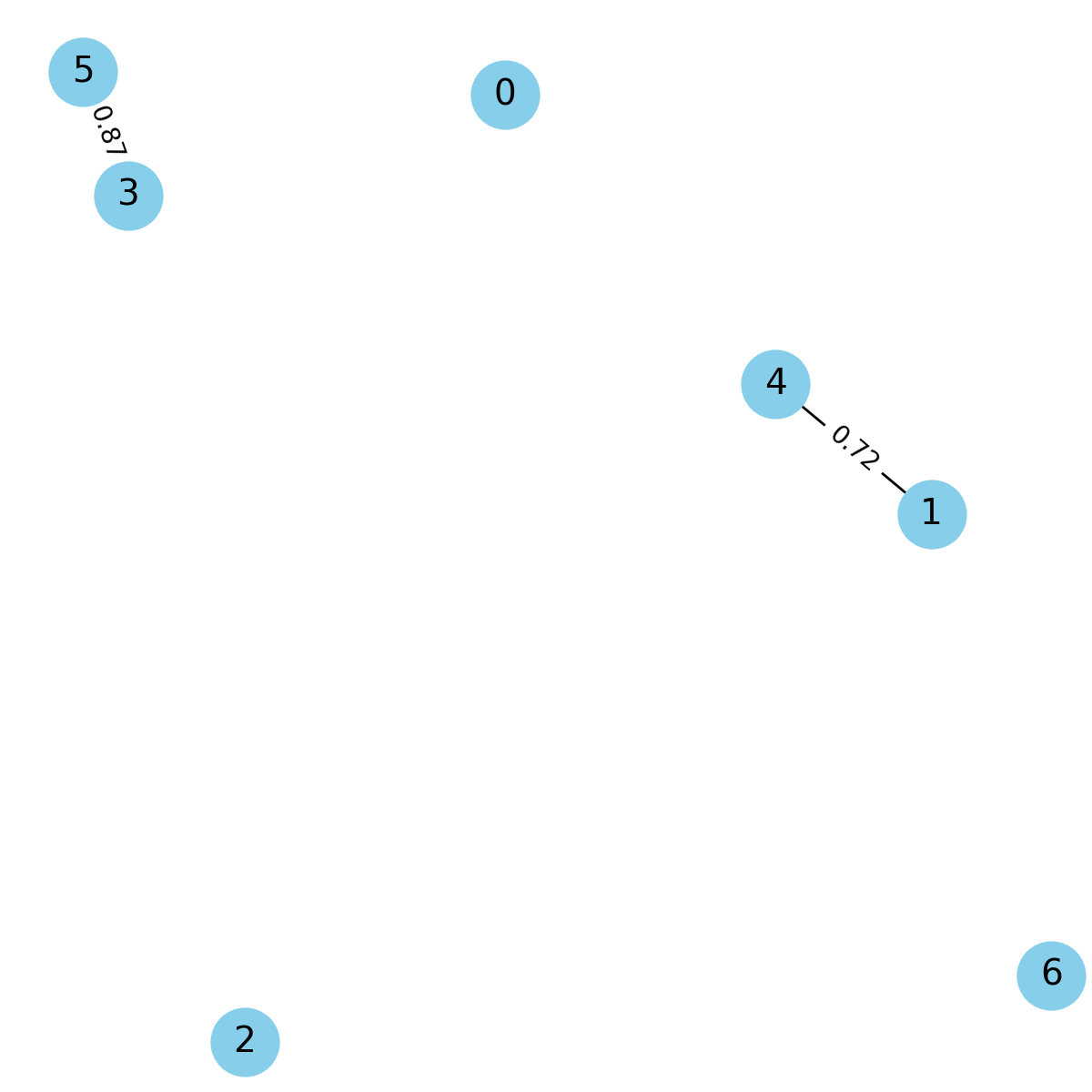
### 2.1. Матриця подібності W

Матриця подібності (W) є симетричною матрицею, яка містить значення схожості між усіма парами об’єктів. У даній роботі використовується функція гаусового ядра для обчислення подібності:  
  
 W(i, j) = exp(-||x\_i - x\_j||² / σ²)  
  
де σ — параметр, що керує чутливістю ядра.

Фрагмент коду на Python для побудови матриці W:

W = np.zeros((n, n))  
for i in range(n):  
 for j in range(n):  
 dist\_sq = np.sum((points[i] - points[j]) \*\* 2)  
 W[i, j] = np.exp(-dist\_sq / (sigma \*\* 2))

Графік графа подібності:



Матриця W:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | P0 | P1 | P2 | P3 | P4 | P5 | P6 |
| P0 | 1.00 | 0.06 | 0.00 | 0.30 | 0.29 | 0.25 | 0.00 |
| P1 | 0.06 | 1.00 | 0.00 | 0.00 | 0.72 | 0.00 | 0.17 |
| P2 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| P3 | 0.30 | 0.00 | 0.00 | 1.00 | 0.03 | 0.87 | 0.00 |
| P4 | 0.29 | 0.72 | 0.00 | 0.03 | 1.00 | 0.01 | 0.04 |
| P5 | 0.25 | 0.00 | 0.00 | 0.87 | 0.01 | 1.00 | 0.00 |
| P6 | 0.00 | 0.17 | 0.01 | 0.00 | 0.04 | 0.00 | 1.00 |

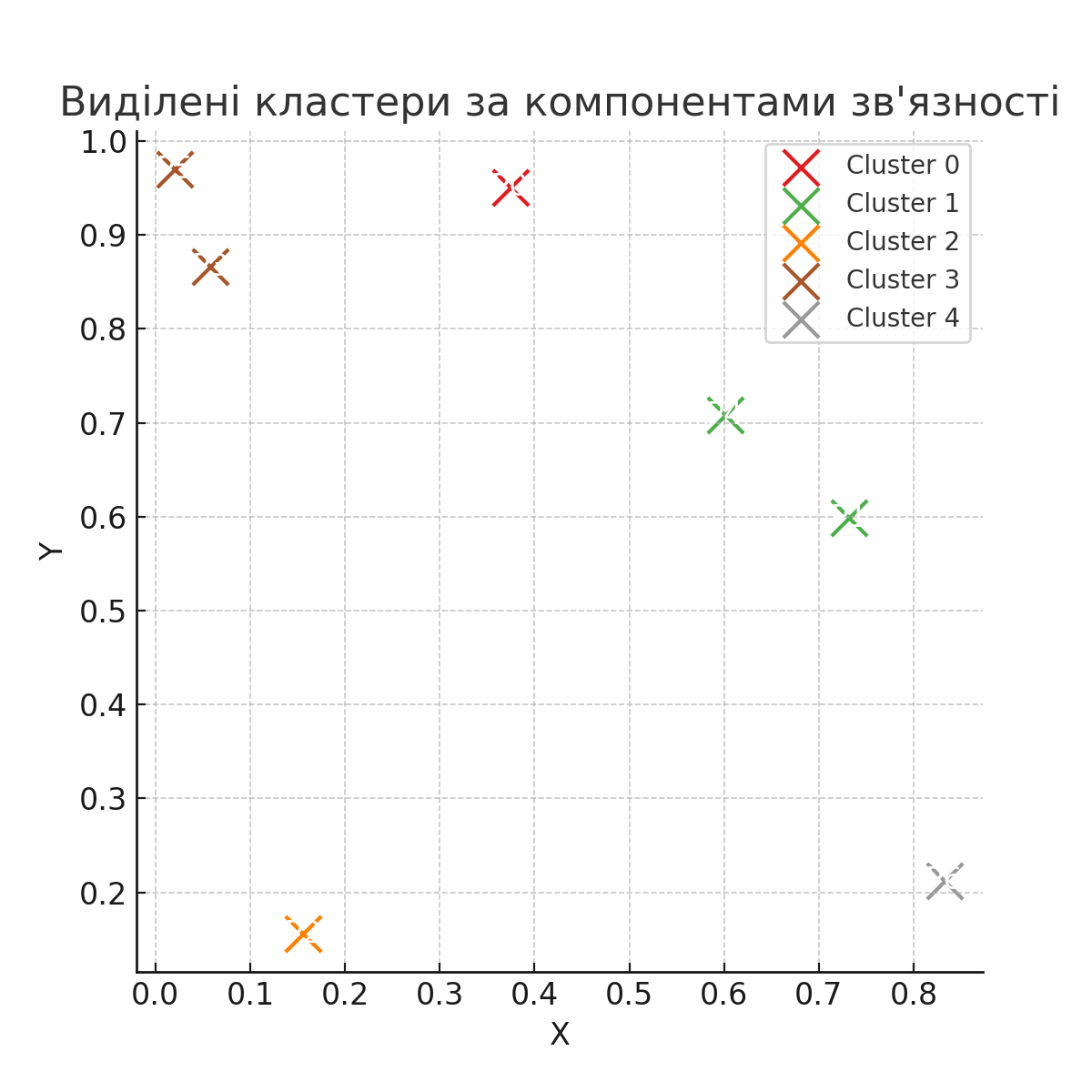
### 2.2. Виявлення кластерів у графі

Після побудови графа за матрицею W, виділяються його компоненти зв’язності. Кожна компонента представляє окремий кластер. Це дозволяє знаходити природні групи у даних без вказівки кількості кластерів.

Фрагмент коду на Python для виділення кластерів:

components = list(nx.connected\_components(G))  
cluster\_assignment = {}  
for cluster\_id, comp in enumerate(components):  
 for node in comp:  
 cluster\_assignment[node] = cluster\_id

Графік кластерів (компоненти зв’язності):



Таблиця кластеризації точок:

|  |  |  |  |
| --- | --- | --- | --- |
| X | Y | Point | Cluster |
| 0.3745401188473625 | 0.9507143064099162 | P0 | 0 |
| 0.7319939418114051 | 0.5986584841970366 | P1 | 1 |
| 0.1560186404424365 | 0.1559945203362026 | P2 | 2 |
| 0.05808361216819946 | 0.8661761457749352 | P3 | 3 |
| 0.6011150117432088 | 0.7080725777960455 | P4 | 1 |
| 0.02058449429580245 | 0.9699098521619943 | P5 | 3 |
| 0.8324426408004217 | 0.2123391106782762 | P6 | 4 |

## 3. Висновки

У результаті виконання лабораторної роботи було реалізовано кластеризацію на основі матриці подібності. Побудовано граф подібності та визначено кластери як компоненти зв’язності. Отримані результати демонструють ефективність підходу для виявлення групування об’єктів без потреби у заданій кількості кластерів.