Original Article

e-ISSN: 2774-2016 - https://journal.itera.ac.id/index.php/indojam/
p-ISSN: 2774-2067

Received 19th February 2023
Accepted 1st March 2023
Published 31st July 2023

Open Access

DOI:
10.35472/indojam.v3i1.1288

Penerapan Algoritma Batchelor-Wilkins dalam Pengklasteran Graf

Christyan Tamaro Nadeak [a*]

[a] *Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera*

* *Corresponding E-mail:* christyan.nadeak@sd.itera.ac.id

Abstract: Batchelor-Wilkins Algorithm is a simple and heuristic clustering algorithm used when the number of classes is unknown. In this paper we will use Batchelor-Wilkins algorithm in graph clustering, specifically a Banana Tree Graph $\boldsymbol{B(n,k)}$, a graph obtained by connecting one leaf of each of n copies of a complete bipartite graph $\boldsymbol{K_{1,k-1}}$ to a single root vertex.

Keywords: *Batchelor-Wilkins Algorithm, Graph Clustering, Banana Tree Graph*

Abstrak: Algoritma Batchelor-Wilkins merupakan algoritma sederhana dan heuristik yang digunakan dalam pengklasteran ketika jumlah klaster tidak diketahui. Dalam penelitian ini akan digunakan algoritma Batchelor-Wilkins untuk melakukan pengklasteran pada graf, secara khusus pengklasteran terhadap graf Pohon Pisang. Graf Pohon Pisang $\boldsymbol{B(n,k)}$ diperoleh dengan menghubungkan satu daun dari setiap graf dari n kopi graf bipartit lengkap $\boldsymbol{K_{1,k-1}}$ ke satu titik akar.

Kata Kunci: *Algoritma Batchelor-Wilkins, Pengklasteran Graf, Graf Pohon Pisang*

Pendahuluan

Dalam penelitian ini, suatu graf adalah pasangan himpunan terurut $G = (V, E)$ sehingga $E \subseteq [V]^2$ [1]. Himpunan V disebut dengan himpunan titik, sedangkan himpunan E disebut dengan himpunan sisi. Jika terdapat sebuah sisi $e \in E$ yang menghubungkan dua titik x dan $y \in V$, maka x dikatakan bertetangga dengan y, dapat dituliskan sebagai $e = \{x, y\} = xy$.

Himpunan tetangga dari suatu titik v pada graf G dinotasikan dengan $N(v)$. Kemudian, kardinalitas dari himpunan $N(v)$ disebut sebagai derajat dari titik v pada graf G, dinotasikan sebagai $d(v)$. Jika semua titik pada graf G memiliki derajat yang sama, maka G disebut sebagai graf regular. Notasi $\delta(G) \coloneqq \min_{v}\{d(v) \mid v \in V\}$ dan $\Delta(G) \coloneqq \max_{v}\{d(v) \mid v \in V\}$ berturut-turut menyatakan derajat minimum dan derajat maksimum dari graf G. Lebih jauh, titik dengan derajat 0 pada suatu graf disebut dengan titik terisolasi, kemudian titik dengan derajat 1 disebut sebagai daun.

Suatu jalan pada graf G didefinisikan sebagai barisan bergantian antara titik dan sisi $v_0 e_0 v_1 e_1 \dots e_{k-1} v_k$ sehingga $e_i = \{v_i, v_{i+1}\}$ untuk setiap $i < k$. Jika setiap titik pada jalan ini berbeda, maka jalan ini disebut sebagai lintasan pada graf G. Panjang dari suatu jalan atau lintasan adalah banyaknya sisi yang terdapat dalam barisan pada jalan atau lintasan tersebut.

Jarak antara dua titik x dan y, dinotasikan dengan $d_G(x, y)$, didefinisikan sebagai panjang dari lintasan terpendek yang menghubungkan x dan y. Jarak maksimum antara dua titik yang paling mungkin

Original Article

dalam suatu graf disebut dengan diameter graf. Suatu graf tak-kosong G disebut terhubung jika untuk setiap pasangan titik pada G dihubungkan oleh suatu lintasan.

Definisikan Graf Bipartit Lengkap $K_{a,b}$ sebagai graf dengan himpunan titik sebanyak $a+b$ yang dapat dipartisi menjadi dua subhimpunan titik V_1 dan V_2 berturut-turut dengan kardinalitas a dan b, sehingga setiap titik pada himpunan V_1 bertetangga dengan setiap titik pada himpunan V_2.

Metode

Metode yang digunakan dalam penelitian ini adalah metode penelitian kepustakaan atau kajian pustaka, yakni melakukan penelitian untuk memperoleh data-data dan informasi-informasi serta objek-objek yang digunakan dalam pembahasan masalah tersebut.

Adapun langkah-langkah yang telah dilakukan, diantaranya:

1. Menelaah algoritma Batchelor-Wilkins secara umum untuk penggunaan pengklasteran.
2. Menerapkan algoritma Batchelor-Wilkins ke dalam graf, lebih khusus terhadap graf Pohon Pisang.

Matriks Ketetanggan dan Matriks Jarak

Misalkan graf G memiliki titik sebanyak n. Definisikan matriks ketetanggaan dari G sebagai matriks A berukuran $n \times n$ dengan entri diberikan sebagai berikut:

$$a_{ij} = \begin{cases} 1, & jika\ v_i\ dan\ v_j\ bertetangga \\ 0, & selainnya \end{cases}$$

Definisikan matriks jarak dari G sebagai matriks D berukuran $n \times n$ dengan entrinya adalah d_{ij}, yaitu jarak antara titik v_i dan v_j. Matriks ketetanggaan dan matriks jarak dari suatu graf G merupakan matriks simetri.

Diketahui graf $G = (V, E)$ memiliki matriks ketetanggan A. Beberapa teorema berikut dapat digunakan untuk mendapatkan matriks jarak D jika diketahui matriks A.

Teorema 1. Banyaknya jalan dengan panjang k di G, dari titik v_i ke v_j, adalah entri pada posisi baris ke-i dan kolom ke-j pada matriks A^k [2].

Jika nilai $A_{ij}^k = 0$, artinya pada graf G tidak ada jalan atau lintasan yang menghubungi v_i dan v_j dengan panjang k.Ini dapat memberikan gambaran dalam menentukan matriks jarak D.

Teorema 2. Untuk suatu nilai k terkecil yang memenuhi $A_{ij}^{k-1} = 0$ dan $A_{ij}^k > 0$, maka titik v_i dan titik v_j memiliki jarak k pada graf G, atau dengan kata lain $D_{ij} = k$.

Pada matriks jarak D, jelas bahwa $D_{ij} = 0$ untuk setiap $i = j$. Dibutuhkan algoritma iteratif untuk mendapatkan matriks jarak D berdasarkan matriks ketetanggaan A.

1. Untuk matriks ketetanggaan A yang menjadi input, tetapkan $D = A$, yaitu menetapkan setiap entri pada matriks A ke setiap entri pada matriks D ($d_{ij} = a_{ij}$). Kemudian tetapkan suatu nilai $k = 2$.
2. Hitung matriks A^k.
3. Cek untuk setiap $1 \le i, j \le n$ ketika $D_{ij} = 0$. Jika $A_{ij}^k > 0$, maka tetapkan $D_{ij} = k$.
4. Cek apakah pada matriks D terdapat entri 0 selain pada diagonal utamanya. Jika ya, maka tetapkan nilai $k = k + 1$, lalu kembali ke langkah nomor 2.
5. Jika tidak, maka algoritma selesai, dan diperoleh matriks jarak D.

Algoritma Batchelor-Wilkins

Algoritma Batchelor-Wilkins merupakan algoritma yang digunakan ketika banyaknya klaster tidak diketahui. Langkah yang dilakukan dalam algoritma pengklasteran ini sederhana dan heuristik, dan disebut juga dengan algoritma jarak maksimum (Maximum distance algorithm). [3]

1. Algoritma Batchelor-Wilkins dimulai dari menentukan banyaknya klaster yang diinginkan dan memilih satu titik pada data sebagai pusat klaster. Misalkan klaster ini adalah klaster 1.
2. Kemudian pilih titik pada data yang jaraknya paling jauh dengan titik yang sudah ditetapkan pada klaster pertama. Dalam hal ini, titik ini akan masuk pada klaster 2.
3. Untuk menentukan titik yang akan berada pada klaster selanjutnya, lakukan step berikut:

Original Article

- Untuk setiap titik data yang masih belum termuat pada klaster, hitung jaraknya terhadap titik pada klaster 1 dan titik pada klaster 2. Pilih yang paling minimum: $\min\left(d(v_j, v_1), d(v_j, v_2)\right)$
- Kemudian tentukan indeks j yang memiliki nilai maksimum dari semua nilai $\min\left(d(v_j, v_1), d(v_j, v_2)\right)$, yaitu
$$\max_j\left(\min\left(d(v_j, v_1), d(v_j, v_2)\right)\right)$$
- Maka titik j inilah yang akan masuk ke dalam centroid dari kluster yang baru, misalkan klaster 3.

4. Proses ini (Langkah 3) berlanjut sampai mendapatkan jumlah klaster yang diinginkan.
5. Jika sudah ditentukan jumlah klaster yang diinginkan, untuk titik yang bukan merupakan pusat klaster dapat dimasukkan ke dalam klaster sesuai dengan jarak yang paling dekat dengan pusat klaster.

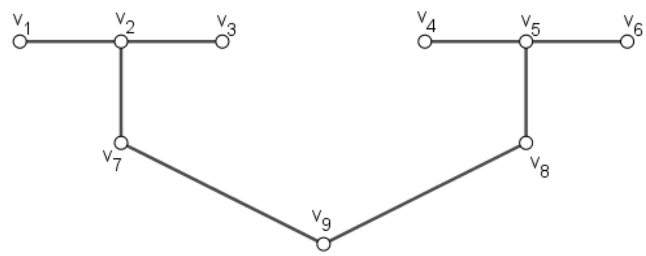
Hasil dan Pembahasan

Dalam penelitian ini, untuk pengklasteran graf, pemilihan kedua titik awal yang akan menjadi pusat klasternya merupakan sesuatu yang masih perlu dibatasi, dibandingkan hanya memilih secara bebas. Ada beberapa langkah pemilihan dua titik awal yang dapat menjadi pusat klaster graf.

1. Pemilihan dua titik dengan jarak maksimum pada graf tersebut, atau diameter dari graf. Ini dapat diperiksa melalui matriks jarak yang diperoleh dari suatu graf. Jika terdapat lebih dari satu pasangan yang memiliki jarak maksimum, dapat dipilih salah satu saja.
2. Pemilihan titik awal (titik pertama) yang merupakan derajat maksimum dari graf tersebut. Pemilihan titik dengan derajat maksimum sebagai pusat klaster dapat dikatakan sesuai, karena ini berarti setiap tetangga dari titik tersebut dapat dimasukkan ke dalam klaster yang sama.

Berikut adalah salah satu keluarga graf dengan karakteristik memiliki beberapa titik dengan derajat tinggi, yaitu graf Pohon Pisang.

Graf Pohon Pisang $B(n,k)$, dengan $n,k \geq 2$ adalah graf yang diperoleh dengan menghubungkan satu daun dari setiap graf dari n kopi graf bipartit lengkap $K_{1,k-1}$ ke satu titik akar. [4]



Gambar 1. Graf Pohon Pisang B(2,4)

Dari contoh graf pohon pisang pada gambar, diperoleh matriks ketetanggan sebagai berikut:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Dari matriks ketetanggaan A di atas, dapat diperoleh matriks jarak dari graf B(2,4)

$$D = \begin{pmatrix} 0 & 1 & 2 & 6 & 5 & 6 & 2 & 4 & 3 \\ 1 & 0 & 1 & 5 & 4 & 5 & 1 & 3 & 2 \\ 2 & 1 & 0 & 6 & 5 & 6 & 2 & 4 & 3 \\ 6 & 5 & 6 & 0 & 1 & 2 & 4 & 2 & 3 \\ 5 & 4 & 5 & 1 & 0 & 1 & 3 & 1 & 2 \\ 6 & 5 & 6 & 2 & 1 & 0 & 4 & 2 & 3 \\ 2 & 1 & 2 & 4 & 3 & 4 & 0 & 2 & 1 \\ 4 & 3 & 4 & 2 & 1 & 2 & 2 & 0 & 1 \\ 3 & 2 & 3 & 3 & 2 & 3 & 1 & 1 & 0 \end{pmatrix}$$

Dengan algoritma Batchelor-Wilkins, akan dilihat pengklasteran graf pohon pisang B(2,4)

Published by: Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)
Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Original Article

untuk pemilihan pusat klaster yang sudah ditentukan.

Untuk pemilihan berdasarkan jarak maksimum atau diameter, dapat dilihat bahwa diameter dari graf Pohon Pisang $B(2,4)$ adalah 6. Kemudian terdapat lebih dari satu pasangan titik dengan jarak 6. Secara acak dapat dipilih titik v_1 dan v_4, sehingga kedua titik ini akan menjadi pusat klaster. Lalu akan dihitung jarak minimum dari setiap titik lain terhadap masing-masing klaster.

TABEL 1. Tabel jarak antara semua titik terhadap v_1 dan v_4

v_i	$d(v_i, v_1)$	$d(v_i, v_4)$	$min(d(v_i, v_1), d(v_i, v_4))$
v_2	1	5	1
v_3	2	6	2
v_5	5	1	1
v_6	6	2	2
v_7	2	4	2
v_8	4	2	2
v_9	3	3	3

Dari semua nilai $\min(d(v_i, v_1), d(v_i, v_4))$, nilai maksimum yang mungkin adalah 3, ketika $v_i = v_9$. Sehingga v_9 adalah titik berikutnya yang dijadikan pusat klaster. Dengan pembagian graf menjadi tiga klaster, maka titik lain yang bukan merupakan pusat klaster dapat dimasukkan ke dalam klaster sesuai dengan jarak terdekatnya terhadap pusat klaster.

TABEL 2. Tabel penentuan klaster berdasarkan jarak terdekat terhadap jarak antara semua titik pusat klaster.

v_i	$d(v_i, v_1)$	$d(v_i, v_4)$	$d(v_i, v_9)$	$Klaster$
v_2	1	5	2	v_1
v_3	2	6	3	v_1
v_5	5	1	2	v_4
v_6	6	2	3	v_4
v_7	2	4	1	v_9
v_8	4	2	1	v_9

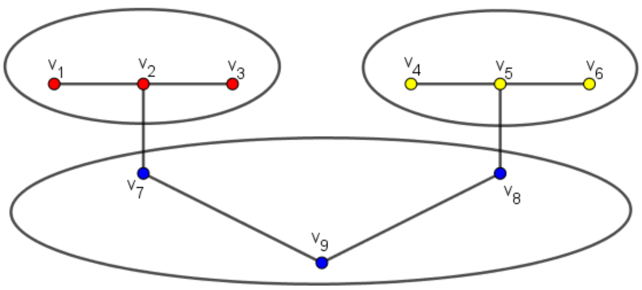
Untuk pemilihan berdasarkan derajat maksimum, dapat dilihat bahwa derajat maksimum dari graf adalah 3, dengan v_2 dan v_5 memiliki derajat sebesar 3. Pilih titik v_2 sebagai titik pusat klaster awal. Kemudian, titik dengan jarak terjauh terhadap v_2 adalah titik v_4 dan v_6, sebesar 5. Pilih titik v_6 sebagai pusat klaster berikutnya. Kemudian, hitung jarak minimum dari setiap titik lain terhadap masing-masing klaster.

TABEL 3. Tabel jarak antara semua titik terhadap v_2 dan v_6.

v_i	$d(v_i, v_2)$	$d(v_i, v_6)$	$min(d(v_i, v_1), d(v_i, v_4))$
v_1	1	6	1
v_3	1	6	1
v_4	5	2	2
v_5	4	1	1
v_7	1	4	1
v_8	3	2	2
v_9	2	3	2

Dari Tabel 3, dapat dilihat ada tiga kemungkinan titik yang dapat menjadi pusat klaster selanjutnya berdasarkan algoritma Batchelor-Wilkins, yaitu v_4, v_8, dan v_9.

1. Untuk pusat klasternya adalah v_2, v_4, dan v_6, perhatikan bahwa titik v_1, v_3, v_7, dan v_9 termasuk ke dalam klaster v_2. Sedangkan v_5 dan v_8 dapat masuk ke dalam klaster v_4 atau v_6. Ini dikarenakan pada graf Pohon Pisang B(2,4), $d(v_5, v_4) = d(v_5, v_6)$ dan $d(v_8, v_4) = d(v_8, v_6)$. Bahkan lebih lanjut $d(v_i, v_4) = d(v_i, v_6)$, untuk setiap $i \neq 4{,}6$. Ini dapat dikaji lebih lanjut terkait pembagian klaster yang tepat, atau titik v_4 dan v_6 dapat dijadikan satu klaster.
2. Untuk pusat klasternya adalah v_2, v_4, dan v_8, titik v_1, v_3, dan v_7 termasuk ke dalam klaster v_2. Kemudian titik v_9 termasuk ke dalam klaster v_8. Sedangkan v_5 dan v_6 dapat masuk ke dalam klaster v_4 atau v_8. Ini dikarenakan pada graf Pohon Pisang B(2,4), $d(v_5, v_4) = d(v_5, v_8)$ dan $d(v_6, v_4) = d(v_6, v_8)$.
3. Untuk pusat klasternya adalah v_2, v_4, dan v_9, titik v_1 dan v_3 termasuk ke dalam klaster v_2. Kemudian titik v_8 termasuk ke dalam klaster v_9. Sedangkan v_5 dan v_6 termasuk ke dalam klaster v_4. Namun v_7 dapat masuk ke dalam klaster v_2 atau v_9 dikarenakan pada graf Pohon Pisang B(2,4), $d(v_7, v_2) = d(v_7, v_9)$.



Gambar 2. Contoh Pengklasteran Graf Pohon Pisang B(2,4)

Secara umum, dapat diperlihatkan melalui algoritma Batchelor-Wilkins bahwa graf Pohon Pisang $B(n,k)$ dapat dibagi menjadi $n+1$ klaster dengan pemilihan pusat klasternya adalah masing-masing satu titik daun pada graf bipartit lengkap $K_{1,k-1}$, total sebanyak n titik, ditambah dengan satu titik pusat klaster dari titik akar pada graf.

Kesimpulan

Algoritma Batchelor-Wilkins dapat digunakan untuk melakukan pengklasteran graf, namun perlu dilakukan penelitian lebih lanjut terutama dalam menentukan penghentian iterasi atau langkah pada algoritma dalam menentukan jumlah klaster. Kemudian perlu dikaji pemilihan penempatan klaster pada suatu titik yang memiliki jarak yang sama terhadap dua titik pusat klaster.

Konflik Kepentingan

Tidak ada konflik kepentingan yang dinyatakan dalam penulisan artikel ini.

Referensi

[1] R. Diestel, Graph Theory. Springer-Verlag Heidelberg, 2005.

[2] N. Biggs, Algebraic Graph Theory. Cambridge University Press, 1993.

[3] S. Bow, Pattern Recognition and Image Preprocessing. Marcel Dekker, 2002.

[4] W. C. Chen, H. I. Lu, dan Y. N. Neh, "Operations of Interlaced Trees and Graceful Trees," Southeast Asian Bull, Math, Vol. 21: 337-348, 1997.

Published by: Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)
Institut Teknologi Sumatera, Lampung Selatan, Indonesia