

JRAM

by Tesdiq Kaloka

Submission date: 02-Feb-2021 10:07AM (UTC+0700)

Submission ID: 1499595204

File name: Tesdiq_Tanpa_Dapus.docx (303.92K)

Word count: 1896

Character count: 11766



ANALISIS HASIL *BICLUSTER* ALGORITMA POLS PADA INTERAKSI PROTEIN MANUSIA DAN HIV-1

TESDIQ PRIGEL KALOKA^{1*}, TITIN SISWANTINING², ALHADI BUSTAMAM³

¹Program Studi Teknik Informatika Politeknik Hasnur, ^{2,3}Departemen Matematika FMIPA Universitas Indonesia

*tesdiq@sci.ui.ac.id

ABSTRAK

AIDS merupakan penyakit yang disebabkan oleh virus HIV-1. Protein merupakan bagian penting dari organisme yang memiliki beragam fungsi. HIV-1 dapat menyerang tubuh manusia karena adanya interaksi protein. Fungsi dan sifat interaksi prote[2] dapat diketahui dengan mengelompokkan protein-protein yang saling berinteraksi. *Bicluster* merupakan metode yang dapat digunakan untuk menyelesaikan permasalahan interaksi protein. Algoritma POLS merupakan algoritma *bicluster* yang menggunakan pendekatan teori graf. Data interaksi protein manusia dan HIV-1 dibagi menjadi dataset 1 dan dataset 2. Setiap dataset dianalisis tingkat kepadatan hasil *bicluster* Algoritma POLS. Hasil penelitian ini diperoleh presentase *bicluster* Algoritma POLS dengan tingkat kepadatan 1 pada dataset 1 sebesar 1,01% dan pada dataset 2 sebesar 4,42%, selain itu *bicluster* Algoritma POLS yang berukuran kecil (2×2 dan 3×3) lebih optimal karena memiliki tingkat kepadatan 1.

Kata Kunci: Interaksi Protein, Algoritma POLS, *Bicluster*

ABSTRACT

AIDS is a disease caused by the HIV-1 virus. Proteins are an essential part of organisms that have various functions. HIV-1 can attack the human body due to protein interactions. T[2] role and nature of protein interactions can be determined by grouping interacting proteins. *Bicluster* is a method that can be used to solve protein interaction problems. The POLS algorithm is a *bicluster* algorithm that uses a graph theory approach. Human protein interaction data and HIV-1 were divided into datasets one and dataset 2. Each dataset was analyzed for the density level of the POLS Algorithm *bicluster*. This study obtained percentage of dataset 1 with density level 1 is 1,01% and dataset 2 with density 2 is 4,42%, beside that, the small *bicluster* from POLS Alogrihtm is more optimal because it has density level of 1.

Kata Kunci: Protein Interaction, POLS Algorithm, Bicluster

1 Pendahuluan

Acquired Immunodeficiency Syndrome (AIDS) merupakan suatu penyakit yang disebabkan oleh virus *Human Immunodeficiency Virus* (HIV). HIV melemahkan kekebalan tubuh manusia secara perlahan, sehingga diagnosa penyakit AIDS terlambat (kekebalan tubuh sudah sangat lemah) d[7] banyak penderita AIDS yang meninggal [1]. Menurut Frankenberg [2] virus HIV terdiri dari 2 jenis, yaitu HIV-1 dan HIV-2.

¹2010 *Mathematics Subject Classification*:
Tanggal Masuk: hh-bb-tt; direvisi: hh-bb-tt; diterima: hh-bb-tt

HIV-1 merupakan virus yang terdiri dari susunan RNA. HIV-1 dapat menyerang kekebalan tubuh manusia karena terjadi proses interaksi protein. Sifat dan fungsi suatu protein dapat diketahui dengan mengelompokkan interaksi protein [3]. Sifat dan fungsi protein yang telah diketahui akan memudahkan pihak kesehatan untuk dapat menemukan cara penyembuhan penderita penyakit AIDS.

Permasalahan yang ditemui pada bidang kesehatan dan biologi adalah proses penelitian yang memakan waktu dan biaya yang cukup besar. Permasalahan tersebut dapat diatasi dengan bantuan Bioinformatika. Bioinformatika merupakan ilmu pengetahuan yang menggabungkan biologi, kedokteran, matematika, statistika, dan teknik komputasi. Secara sederhana, bioinformatika menyelesaikan permasalahan biologi, kedokteran, dan *life science* dengan menggunakan teori-teori matematika dan statistika yang selanjutnya dieksekusi menggunakan bantuan komputer [4]. Salah satu cabang bioinformatika yang dapat menyelesaikan permasalahan pengelompokkan protein adalah *bicluster*. *Bicluster* merupakan suatu metode yang digunakan untuk mengelompokkan data-data ekspresi gen dikarenakan metode clustering biasa tidak dapat mencari pola kesamaan yang cocok [5]. Pada umumnya *bicluster* disamakan dengan *clustering* dua kali.

Bicluster merupakan suatu metode analisis yang digunakan untuk data-data biologis [6]. Algoritma *bicluster* telah banyak digunakan, salah satu algoritma *bicluster* pertama yang terkenal adalah Algoritma CnC (Cheng and Church), diambil dari nama penemu algoritma tersebut [7]. Beberapa algoritma *bicluster* yang sering digunakan antara lain: *Qualitative Biclustering* (QUBIC), FLOC, dan *Binary Inclusion-Maximal* (BiMax). QUBIC dan BiMax merupakan algoritma *bicluster* yang menggunakan pendekatan teori graf [5]. FLOC merupakan algoritma yang menggunakan konsep probabilitas [8].

Algoritma *Bicluster* terbaru dikemukakan oleh Wang [9] yang disebut dengan Algoritma POLS. Algoritma POLS menggunakan pendekatan teori graf dan teori *biclique* seimbang. Berdasarkan penelitian Bustamam [3], hasil *bicluster* Algoritma POLS menunjukkan adanya protein yang tidak berinteraksi pada satu *bicluster*. Sehingga penelitian ini berfokus pada tingkat kepadatan interaksi setiap *bicluster* Algoritma POLS terhadap interaksi protein manusia dan HIV-1.

2 Tinjauan Pustaka

2.1. Bicluster

Misal diberikan suatu matriks $A(D, C)$ dengan D adalah suatu himpunan data $D = \{I_1, I_2, I_3, \dots, I_D\}$ dan C suatu himpunan fitur $C = \{J_1, J_2, J_3, \dots, J_C\}$. Setiap elemen m_{ij} dari matriks $A(D, C)$ mewakili nilai data ke- i pada fitur ke- j dengan $i \in D$ dan $j \in C$. Sebuah *bicluster* adalah suatu submatriks $M(I, J) = [m_{ij}]$ dari matriks $A(D, C)$, dengan $i \in I$ dan $j \in J$, $i \subseteq D$ dan $j \subseteq C$ [5]. Secara sederhana, penjelasan *bicluster* dapat dilihat pada Persamaan 1.

$$A(D, C) = \begin{bmatrix} m_{11} & \cdots & m_{1C} \\ \vdots & \ddots & \vdots \\ m_{D1} & \cdots & m_{DC} \end{bmatrix} \quad (1)$$

2.2. Algoritma POLS

Algoritma POLS merupakan suatu algoritma pencarian *bicluster* yang menggunakan pendekatan teori graf. Hasil *bicluster* Algoritma POLS merupakan sebuah *biclique* seimbang. *Biclique* seimbang merupakan suatu graf bipartit lengkap dengan jumlah partisi busur yang sama.

Bicluster hasil Algoritma POLS dibentuk dengan menggunakan beberapa tahap. Tahap-tahap tersebut dinamakan kandidat, *addset*, *dropset*, *pscore*, dan *addrule*. Kandidat merupakan tempat untuk menyimpan solusi hingga proses iterasi selesai. *Addset* merupakan himpunan suatu semesta yang akan ditambahkan ke kandidat. *Dropset* merupakan subset dari kandidat. *Pscore* merupakan nilai dari setiap pasang *addset*. *Addrule* merupakan proses pemilihan pasangan busur terbaik dari *addset* [9]. Penjelasan Algoritma POLS dapat lebih dipahami pada Tabel 1.

Tabel 1: Algoritma POLS

Masukan	:	Graf Bipartit
1	:	
2	:	Inisialisasi <i>addset</i> , <i>dropset</i> , dan $S^* = S$
Ulangi	:	
3	:	Pilih sepasang simpul (u, v) dari <i>addset</i> menggunakan <i>addrule</i>
4	:	Masukkan pasangan simpul (u, v) ke $S = S \cup (u, v)$
Hingga	:	<i>Addset</i> = $\emptyset$
5	:	Jika $ S > S^* $, maka $S^* = S$
6	:	Hilangkan pasangan simpul (u, v) pada <i>dropset</i>
Keluaran	:	<i>Bicluster</i>

Jika kandidat solusi dinotasikan sebagai $S = (U^S, V^S, E^S)$. Misal Diberikan suatu graf bipartit $G = G(U, V, E)$ dengan simpul U dan V adalah $U = \{u_1, u_2, u_3, \dots, u_n\}$ dan $V = \{v_1, v_2, v_3, \dots, v_n\}$ serta busur E adalah $E = \{e_1, e_2, e_3, \dots, e_n\}$ maka menurut Wang [9] definisi *addset*, *dropset*, dan *pscore* adalah

$$Addset = \{(u, v) \in |u \notin U^S, u \in N(v'); \forall v' \in V^S, v \notin V^S, v \in N(u); \forall u' \in U^S\} \quad (2)$$

$$Dropset = \{(u, v) \in E | u \in U^S, v \in V^S\} \quad (3)$$

$$pscore(u, v) = score_{lb}(u, v) + \left\lfloor \frac{score_{ub}}{2} \right\rfloor \quad (4)$$

dengan $score_{lb}(u, v) = 1$ dan $score_{ub}(u, v) = \min\{|N(v)| - |N(v) \cap S|, |N(u)| - |N(u) \cap S|\}$. Nilai *pscore* terbesar disebut *add rule*.

2.3. Data

Data interaksi protein manusia dan HIV-1 merujuk pada [10]. Data tersebut menyajikan gen manusia, gen HIV-1, protein manusia, protein HIV-1, dan jenis interaksi. Terdapat 16215 interaksi protein manusia dan HIV-1, 3797 protein manusia, 23 protein HIV, dan 130 jenis interaksi pada data tersebut.

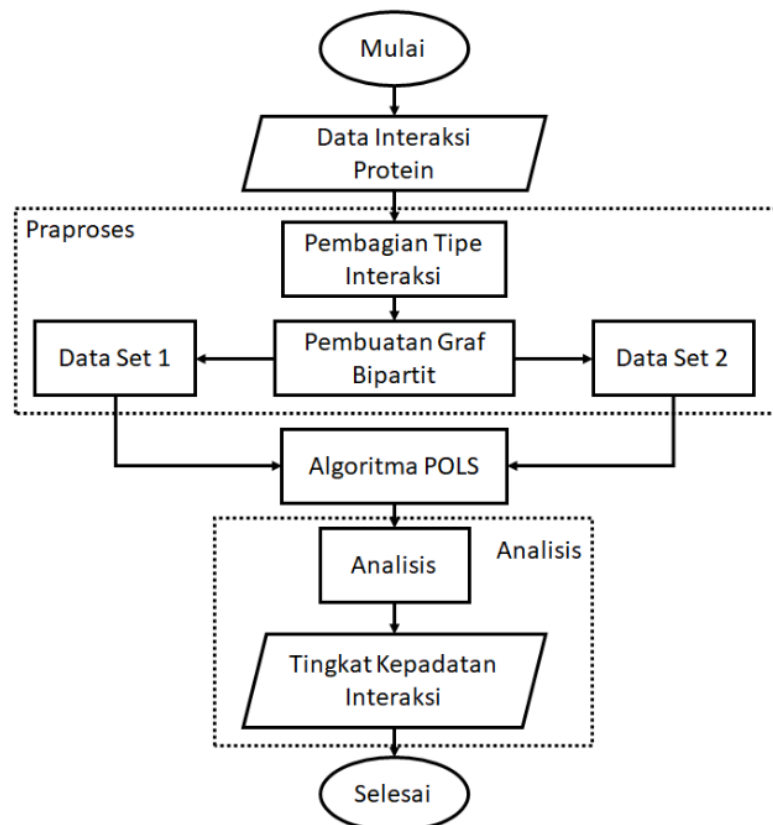
5

3 Metode Penelitian

Proses penelitian dibagi menjadi tiga tahap utama, yaitu: praproses data, implementasi, dan analisis. Praproses data merupakan tahapan transformasi data yang dimiliki menjadi sebuah

dataset yang siap untuk diimplementasikan pada program. Tahap implementasi merupakan proses penggunaan Algoritma POLS pada dataset yang telah dimiliki. Tahap analisis merupakan proses untuk mengetahui tingkat kepadatan interaksi hasil *bicluster*. Keseluruhan proses penelitian dapat dilihat pada Gambar 1.

Berdasarkan Gambar 1, tahap praproses terdiri dari empat bagian. Pembagian tipe interaksi merupakan proses pembagian 130 jenis interaksi menjadi tiga tipe. Tipe pertama merupakan interaksi dari HIV-1 ke manusia. Tipe kedua merupakan interaksi dari manusia ke HIV-1. Tipe ketiga merupakan interaksi dua arah. Pembuatan graf bipartite dilakukan dengan memberikan nilai pada tipe interaksi. Berikan nilai 1 apabila tipe pertama, nilai -1 apabila tipe kedua dan X apabila tipe ketiga dan 0 apabila tidak ada interaksi. Dataset 1 merupakan kumpulan nilai 0, 1 dan X. Dataset 2 merupakan kumpulan nilai 0, -1, dan X.



Gambar 1. Alur Penelitian

Tahapan Algoritma POLS merupakan implementasi dari langkah-langka Algoritma POLS dan teori graf. Tahapan analisis merupakan perhitungan kepadatan interaksi dengan menggunakan perbandingan sederhana.

4 Hasil dan Pembahasan

Hasil *bicluster* Algoritma POLS merupakan sebuah *biclique* seimbang. Pada penerapan *bicluster*, *biclique* seimbang mengharuskan terpenuhinya dua syarat: jumlah baris dan kolom sama; serta nilai baris dan kolom harus terisi (bukan 0). Sehingga submatriks merupakan matriks persegi. Selanjutnya entri-entri submatriks tersebut akan dianalisis tingkat kepadatan

interaksi protein. Kepadatan interaksi merupakan perbandingan nilai 1 dan 0 pada submatriks hasil *bicluster*. Semakin tinggi tingkat kepadatan, maka hasil *bicluster* semakin baik.

Data penelitian ini dibagi menjadi 2, yaitu Dataset 1 dan Dataset 2, sehingga diperoleh hasil untuk dataset 1 dan dataset 2. Selanjutnya akan dijelaskan untuk hasil *bicluster* setiap dataset.

4.1. Dataset 1

Dataset 1 menghasilkan 297 *bicluster* dengan *bicluster* berukuran 2×2 sebanyak 171, 3×3 sebanyak 35, 4×4 sebanyak 78, 5×5 sebanyak 5, 6×6 sebanyak 7, dan satu *bicluster* berukuran 7×7 . Hasil *bicluster* dataset 1 dapat dilihat pada Tabel 2. Terlihat bahwa seluruh hasil *bicluster* merupakan submatriks dari dataset 1 dengan jumlah baris dan kolom sama (matriks persegi).

Tabel 2: Hasil *Bicluster* Dataset 1

Ukuran <i>Bicluster</i>	Jumlah
2×2	171
3×3	35
4×4	78
5×5	5
6×6	7
7×7	1

Seluruh 297 *bicluster* kemudian dilihat tingkat kepadatan *bicluster*. Tingkat kepadatan *bicluster* diperoleh dengan cara:

$$D = \frac{n}{N} \quad (5)$$

dengan D merupakan tingkat kepadatan *bicluster*, n adalah jumlah entri *bicluster* yang tak 0 dan N adalah total entri *bicluster*.

Tabel 3: Tingkat Kepadatan *Bicluster* Dataset 1

Jumlah <i>Bicluster</i>	Presentase	Tingkat Kepadatan
228	76,77%	$> 0,5$
22	7,41%	$> 0,75$
3	1,01%	1

Berdasarkan Persamaan 5 dan data Tabel 2, terdapat 228 *bicluster* dengan tingkat kepadatan lebih dari 0,5. Jumlah ini setara dengan 76,77% dari total hasil *bicluster*. Dari 297 *bicluster*, terdapat 22 *bicluster* (7,41%) yang memiliki tingkat kepadatan lebih dari 0,75 dan

hanya terdapat 3 (1,01%) *bicluster* yang memiliki tingkat kepadatan 1. Rangkuman tingkat kepadatan *bicluster* dapat dilihat pada Tabel 3.

21 dari 22 *bicluster* dengan tingkat kepadatan lebih dari 0,75 masing-masing merupakan matriks 2×2 dan hanya terdapat satu *bicluster* yang berukuran 3×3 . Selain itu, *bicluster* berukuran 3×3 tersebut juga merupakan *bicluster* dengan tingkat kepadatan 1. Sehingga pada *bicluster* dengan tingkat kepadatan 1, terdapat 2 *bicluster* 2×2 dan 1 *bicluster* 3×3 .

4.2. Dataset 2

Proses dan cara implementasi pada dataset 2 sama dengan dataset 1, sehingga diperoleh: 203 *bicluster* dengan *bicluster* berukuran 2×2 sebanyak 110, 3×3 sebanyak 24, 4×4 sebanyak 59, 5×5 sebanyak 5, 6×6 sebanyak 4, dan satu *bicluster* berukuran 7×7 . Ukuran dan jumlah *bicluster* dapat lebih mudah dipahami melalui Tabel 4.

Tabel 4: Hasil *Bicluster* Dataset 2

Ukuran <i>Bicluster</i>	Jumlah
2×2	110
3×3	24
4×4	59
5×5	5
6×6	4
7×7	1

Terdapat 11 *bicluster* dengan tingkat kepadatan 1, 17 *bicluster* dengan tingkat kepadatan lebih dari 0,75, dan 182 *bicluster* dengan tingkat kepadatan lebih dari 0,5. Presentase *bicluster* dengan tingkat kepadatan lebih dari 0,5, lebih dari 0,75, dan 1 secara berturut-turut adalah 89,65%, 8,37%, dan 5,42%. Presentase tingkat kepadatan *bicluster* dapat pula dilihat pada Tabel 5. Seluruh *bicluster* dengan tingkat kepadatan 1 dan lebih dari 0,75 berukuran 2×2 .

Tabel 5: Tingkat Kepadatan *Bicluster* Dataset 2

Jumlah <i>Bicluster</i>	Presentase	Tingkat Kepadatan
182	89,65%	$> 0,5$
17	8,37%	$> 0,75$
11	5,42%	1

4.3. Analisis

Berdasarkan Tabel 3 dan 5, presentase untuk $D = 1$ tidak lebih dari 6%, bahkan pada Dataset 1 (Tabel 3) hanya mencapai 1,01%. Selain itu, berdasarkan penjelasan *bicluster* Dataset 1 dan 2, hampir seluruh *bicluster* dengan $D = 1$ berukuran 2×2 . Hanya terdapat 1

bicluster berukuran 3×3 pada Dataset 1. Sehingga dapat dikatakan bahwa, dari total 400 *bicluster*, hanya terdapat 14 *bicluster* yang memenuhi kriteria submatriks persegi dan memiliki tingkat kepadatan sama dengan 1. Berdasarkan Tabel 2 dan 4, hanya terdapat satu *bicluster* yang berukuran 7×7 pada masing-masing Dataset 1 dan 2. Tingkat kepadatan *bicluster* tersebut tidak mencapai 0,5 sehingga tidak dapat dikatakan *bicluster* yang baik.

Pada permasalahan interaksi protein manusia dan HIV-1, jumlah (ukuran) protein manusia dan HIV-1 jauh berbeda. Protein manusia berjumlah 3797 dan HIV-1 berjumlah 23. Perbedaan jumlah yang signifikan tersebut menyebabkan hasil Algoritma POLS didominasi dengan *bicluster* berukuran 2×2 . Dilihat dari beberapa pemaparan, maka Algoritma POLS lebih baik menghasilkan *bicluster* berukuran kecil (2×2 dan 3×3). Hal ini dikarenakan hanya *bicluster* berukuran kecil yang memiliki tingkat kepadatan 1. Selain itu Algoritma POLS lebih cocok untuk permasalahan dengan data yang seimbang (jumlah simpul yang sama).

5 Kesimpulan

Berdasarkan hasil penelitian, dari 400 *bicluster*, terdapat 14 *bicluster* yang memiliki tingkat kepadatan 1. Kemudian presentase hasil *bicluster* Algoritma POLS dengan tingkat kepadatan 1 pada dataset 1 sebesar 1,01% dan pada dataset 2 sebesar 4,42%. *Bicluster* Algoritma POLS yang berukuran kecil (2×2 dan 3×3) lebih optimal karena memiliki tingkat kepadatan 1.

ORIGINALITY REPORT

6%

SIMILARITY INDEX

4%

INTERNET SOURCES

0%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Universitas Negeri Surabaya The
State University of Surabaya

Student Paper

2%

2

ejournal.undip.ac.id

Internet Source

1%

3

repository.usu.ac.id

Internet Source

1%

4

jurnalpolitanipyk.ac.id

Internet Source

1%

5

docplayer.info

Internet Source

<1%

6

media.neliti.com

Internet Source

<1%

7

resapanbiopori.blogspot.com

Internet Source

<1%

8

surabaya.liputan6.com

Internet Source

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography On