

Penerapan Graf dalam Pemetaan Susunan DNA

Scarletta Julia Yapfrine (13514074)
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13514074@std.stei.itb.ac.id

Abstrak—Pita DNA merupakan suatu susunan ganda yang terdiri dari nukleotida. Nukleotida mengandung sandi-sandi genetika dalam bentuk susunan empat basa nitrogen yang terdiri dari guanina (G), adenina (A), timina (T), dan sitosina (C). Bagi ilmuwan, pemahaman terhadap sandi genetika ini tentunya sangat penting untuk mendapatkan informasi mengenai gen makhluk hidup seperti bakteri dan virus. Namun, suatu sandi genetika tersusun dari basa nitrogen yang sangat banyak sehingga seseorang perlu memahami pola pengurutan basa nitrogen terlebih dahulu untuk memahami sandi genetika. Pola pengurutan basa nitrogen dapat dicari menggunakan pemetaan susunan DNA. Makalah ini membahas penerapan graf dalam pemetaan susunan DNA.

Kata kunci—Benzer, contig, graf interval, pemetaan DNA

I. PENDAHULUAN

Makhluk hidup tumbuh dan berkembang biak. Cara berkembang biak, pertumbuhan, dan semua hal lain yang terjadi pada makhluk hidup telah lama menjadi objek penelitian bagi ilmuwan. Masing-masing jenis makhluk hidup memiliki cara berkembang biak dan pertumbuhan yang berbeda-beda. Hal ini dikarenakan informasi genetik yang disimpan pada DNA suatu makhluk hidup.

DNA merupakan bagian yang sangat kecil dari suatu makhluk hidup namun menyimpan informasi yang sangat penting mengenai makhluk hidup tersebut. Informasi genetika tersebut disimpan hingga tiba saatnya informasi tersebut dibutuhkan, contohnya adalah saat regenerasi sel rusak, saat makhluk hidup tersebut terserang penyakit, dan saat pewarisan sifat ke keturunan. Informasi genetika tersebut tersandi pada nukleotida dalam bentuk susunan empat basa nitrogen.

Seperti yang telah disebutkan, salah satu objek yang diteliti oleh ilmuwan adalah makhluk hidup dan hal-hal ilmiah yang terjadi pada makhluk hidup. Hal ini memang bisa diteliti dengan mengamati perilaku makhluk hidup. Namun, untuk mendapat hasil yang lebih akurat mengenai hal-hal tersebut, maka ilmuwan perlu terlebih dahulu memahami informasi genetika yang tersimpan dalam DNA makhluk hidup tersebut. Cara untuk memahami informasi genetika tersebut adalah dengan memahami pola basa nitrogen pada nukleotida. Pola basa nitrogen dapat dipahami dengan pemetaan susunan DNA.

Menurut Rios dan Vendruscolo, untuk meneliti cara kerja sesuatu yang mikro, akan lebih memungkinkan

untuk fokus pada interaksi antar molekul penyusunnya dibandingkan dengan molekul itu sendiri^[1]. Maka, pemetaan susunan DNA lebih baik dilakukan dengan meneliti interaksi antara fragmen-fragmen DNA. Pemetaan DNA ini menggunakan Pemetaan Benzer.

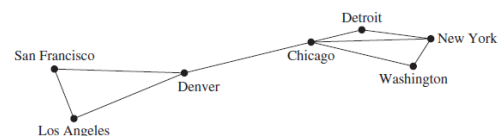
Dalam makalah ini, penulis akan menggunakan graf untuk memodelkan Pemetaan Benzer, untuk mendapat data hubungan antar fragmen-fragmen DNA.

II. DASAR TEORI

2.1. Graf

2.1.1. Definisi Graf

Graf ditulis dalam bentuk $G = (V, E)$, yang dalam hal ini V merupakan himpunan dari simpul-simpul yang ada pada graf $= \{v_1, v_2, \dots, v_n\}$ dan E merupakan himpunan sisi yang menghubungkan dua buah simpul pada graf $= \{e_1, e_2, \dots, e_n\}$. V tidak boleh berupa himpunan kosong, sementara E boleh berupa himpunan kosong. Penamaan simpul pada graf dapat menggunakan huruf, atau bilangan asli, atau gabungan keduanya. Sisi yang menghubungkan simpul v_i dengan simpul v_j dinyatakan dengan pasangan (v_i, v_j) . Bila e merupakan sisi yang menghubungkan simpul v_i dengan simpul v_j , e dapat ditulis sebagai $e = (v_i, v_j)$.



Gambar 2.1. Contoh Graf

Sumber : Rosen, K.H., *Discrete Mathematics and Its Applications*, 2006, New York : McGraw-Hill.

2.1.2. Jenis-Jenis Graf

Berdasarkan ada tidaknya gelang atau sisi ganda pada suatu graf, ada dua jenis graf, yaitu graf sederhana (graf yang tidak mengandung gelang maupun sisi ganda) dan graf tak-sederhana (graf yang mengandung sisi ganda atau gelang).

Berdasarkan jumlah simpul pada suatu graf, ada dua jenis graf, yaitu graf berhingga (graf yang jumlah simpulnya berhingga) dan graf tak-berhingga (graf yang jumlah simpulnya tidak berhingga).

Berdasarkan orientasi arah pada sisi suatu graf, ada dua jenis graf yaitu graf tak-berarah (graf yang isinya tidak mempunyai orientasi arah) dan graf berarah (graf yang setiap sisinya mempunyai orientasi arah). Pada graf tak-berarah, $(v_i, v_j) = (v_j, v_i)$ adalah sisi yang sama. Pada graf berarah, $(v_i, v_j) \neq (v_j, v_i)$. Untuk busur (v_i, v_j) pada graf berarah, simpul v_i dinamakan simpul asal dan simpul v_j dinamakan simpul terminal.

2.1.3. Terminologi Graf

2.1.3.1. Bertetangga

Dua buah simpul pada suatu graf tak-berarah G dikatakan bertetangga jika keduanya terhubung langsung dengan sebuah sisi. Dengan kata lain, v_i bertetangga dengan v_j jika (v_i, v_j) adalah sebuah sisi pada graf G . Pada graf berarah, sisi disebut dengan busur. Jika (v_i, v_j) adalah sebuah busur maka v_i bertetangga dengan v_j dan sebaliknya.

2.1.3.2. Bersisian

Sisi e bersisian dengan v_i dan v_j , jika $e = (v_i, v_j)$.

2.1.3.3. Simpul Terpencil

Simpul terpencil merupakan simpul yang tidak bersisian dengan sisi manapun atau tidak bertetangga dengan simpul lainnya.

2.1.3.4. Graf Kosong

Graf kosong adalah graf yang himpunan sisinya merupakan himpunan kosong. Grafkosong ditulis sebagai N_n , dengan n adalah jumlah simpul.

2.1.3.5. Derajat

Pada graf tak-berarah, derajat suatu simpul adalah jumlah sisi yang bersisian dengan simpul tersebut. Pada graf berarah, derajat dibagi menjadi dua, yaitu derajat masuk (jumlah busur yang masuk ke simpul) dan derajat keluar (jumlah busur yang keluar dari simpul).

2.1.3.6. Lintasan

Lintasan yang panjangnya n dari simpul awal v_0 ke simpul tujuan v_n di dalam graf G ialah barisan berselang-seling simpul-simpul dan sisi-sisi yang berbentuk $v_0, e_1, v_1, e_2, v_2, \dots, v_{n-1}, e_n, v_n$ sedemikian sehingga $e_1 = (v_0, v_1), e_2 = (v_1, v_2), \dots, e_n = (v_{n-1}, v_n)$ adalah sisi-sisi dari graf G . Ada dua jenis lintasan, yaitu lintasan Euler (lintasan yang melalui masing-masing sisi tepat satu kali) dan lintasan Hamilton (lintasan yang melewati masing-masing simpul tepat satu kali).

2.1.3.7. Siklus atau Sirkuit

Siklus atau sirkuit adalah lintasan yang berawal dan berakhir pada suatu simpul yang sama.

2.1.3.8. Terhubung

Graf tak-berarah G disebut graf terhubung jika untuk setiap pasang simpul di dalam graf terdapat lintasan yang menghubungkan kedua buah simpul. Jika tidak, maka G merupakan graf

tak-terhubung. Graf berarah G dikatakan terhubung jika graf tak-berarahnya terhubung (graf tak-berarah dari suatu graf berarah diperoleh dengan menghilangkan arahnya).

2.1.3.9. Upagraf dan Komplemen Upagraf

$G_1 (V_1, E_1)$ adalah sebuah upagraf dari $G (V, E)$ jika $V_1 \subseteq V$ dan $E_1 \subseteq E$. Komplemen dari upagraf G_1 terhadap graf G adalah graf $G_2 = (V_2, E_2)$ sedemikian sehingga $E_2 = E - E_1$ dan V_2 adalah himpunan simpul yang bersisian dengan anggota-anggota E_2 .

2.1.3.10. Upagraf Merentang

Upagraf $G_1 = (V_1, E_1)$ dari $G (V, E)$ merupakan upagraf merentang jika $V_1 = V$ (yaitu G_1 mengandung semua simpul dari G).

2.1.3.11. Cut-Set

Cut-set dari suatu graf terhubung adalah himpunan sisi yang bila dibuang dari graf tersebut menyebabkan graf menjadi graf tak-terhubung.

2.1.3.12. Graf Berbobot

Graf berbobot merupakan graf yang setiap sisinya diberi harga (bobot).

2.1.4. Matriks Ketetanggaan

Misalkan $G = (V, E)$ adalah graf dengan n simpul, $n \geq 1$. Matriks ketetanggaan G adalah matriks dwimatris yang berukuran $n \times n$. Jika matriks tersebut dinamakan $A = [a_{ij}]$, maka

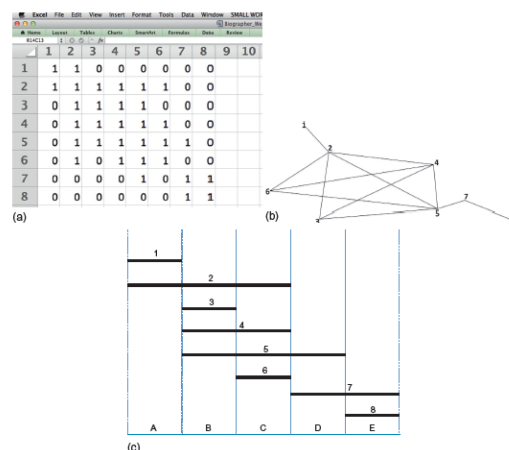
$$a_{ij} = \begin{cases} 1, & \text{jika simpul } i \text{ dan } j \text{ bertetangga} \\ 0, & \text{jika simpul } i \text{ dan } j \text{ tidak bertetangga} \end{cases}$$

2.1.5. Golongan

Golongan merupakan sebuah upagraf yang semua simpulnya terhubung oleh sebuah sisi ke setiap simpul yang lain. Golongan maksimum merupakan golongan yang bukan merupakan himpunan bagian dari golongan yang lebih besar.

2.1.6. Graf Interval

Graf interval merupakan graf khusus yang bisa dijadikan kelompok interval sepanjang garis sesungguhnya.



Gambar 2.2. (a) Matriks Ketetanggaan dari G (b) Graf G

(c) Graf Interval dari G

Graf interval memiliki kolom sejumlah golongan maksimum yang dimiliki oleh suatu graf. Graf G pada gambar 2.2 di atas memiliki lima golongan maksimum yaitu golongan $A = \{1, 2\}$, golongan $B = \{2, 3, 4, 5\}$, golongan $C = \{2, 4, 5, 6\}$, dan golongan $D = \{5, 7\}$, dan golongan $E = \{7, 8\}$. Setiap kolom pada graf interval dinamakan sesuai dengan nama-nama golongan maksimum pada graf. Setiap simpul pada graf digambar di graf interval sebagai garis lurus sepanjang kolom golongan maksimum yang mengandung simpul tersebut.

2.2. DNA

DNA, atau asam deoksiribonukleat, merupakan komponen turun-temurun pada manusia dan hampir semua makhluk hidup lain. Hampir semua sel pada tubuh suatu makhluk hidup memiliki DNA yang sama. Sebagian besar DNA berada pada inti sel, namun ada sebagian kecil DNA yang bisa ditemukan di mitokondria.

Informasi pada DNA tersimpan sebagai sandi yang tersusun dari empat basa nitrogen, yaitu adenina (A), guanina (G), sitosina (C), dan timina (T). DNA manusia terdiri dari tiga miliar basa. Urutan dari basa tersebut menyimpan informasi untuk membangun dan mempertahankan suatu makhluk hidup.

III. PEMETAAN BENZER

Pada tahun 1950-an, Seymour Benzer mempelajari struktur linear dari gen. Berkat penelitiannya, gen didefinisikan sebagai satuan dari struktur, fungsi, dan rekombinasi. Penelitiannya berlangsung tidak lama setelah ditemukannya struktur DNA berupa dua pita nukleotida yang saling berkomplemen dalam struktur *double helix* yang menyimpan informasi genetika dalam bentuk susunan empat basa nitrogen, yaitu A, C, G, dan T. Mutasi dapat terjadi dalam bentuk pertukaran, penghapusan, atau penambahan salah satu basa nitrogen pada DNA. Rekombinasi dapat terjadi melalui pemutusan atau penggabungan fragmen, antara dua basa nitrogen. Suatu fungsi mungkin tersimpan dalam susunan basa nitrogen yang sangat panjang. Benzer menemukan hubungan tersebut dengan memetakan DNA virus yang menyerang bakteri melalui eksperimen. Virus yang digunakan merupakan sepasang mutan T4 dan interval tertentu pada genom virus tersebut telah dihilangkan. Interval yang dihilangkan pada masing-masing mutan tidak diketahui. Jika dua interval tersebut sama, pasangan T4 tersebut kehilangan bagian dari genomnya dan menjadi tidak aktif sehingga bakteri yang diserang selamat. Jika dua interval tersebut berbeda, genom pasangan T4 tersebut saling melengkapi dan aktif sehingga bakteri yang diserang mati.

Dengan metode ini, bisa dilakukan penyusunan DNA untuk seluruh kromosom. Pertama, ditentukan fragmen mana yang meliputi satu sama lain. Jika fragmen telah

tersusun, maka dilihat susunan secara keseluruhannya apakah ada kesamaan susunan (ini metode yang sedang digunakan). Fragmen DNA yang saling meliputi ini dikenal dengan nama contig. Atau, fragmen bisa dipotong dengan restriksi endonuklease kedua dan didapat anak fragmen, lalu dilihat apakah anak fragmen tersebut adalah fragmen yang sudah diidentifikasi. Logika biner untuk menentukan secara kualitatif apakah dua fragmen saling meliputi lebih terlihat pada metode dua. Oleh karena itu, penulis akan menggunakan metode dua untuk selanjutnya.

Pemetaan Benzer merupakan metode yang susah untuk divisualisasi. Oleh karena itu, diperlukan graf interval untuk menggambarkan contig dan membantu proses pemetaan DNA.

IV. PENERAPAN GRAF INTERVAL DALAM PEMETAAN SUSUNAN DNA

Ini adalah contoh sederhana yang menggunakan dua belas fragmen yang dinamai dengan huruf Yunani dan empat endonuklease, yaitu EcoR1, Pst1, Sp1, dan XmaIII. Berikut adalah contoh hasil percobaan untuk menentukan fragmen mana yang saling meliputi satu sama lain.

No	Fragmen Awal	Endonuklease	Fragmen yang dihasilkan
1	Lambda	-	Lambda
2	Lambda	EcoR1	Alpha, theta, gamma
3	Lambda	Pst1	Delta, mu
4	Lambda	Pst1 dan Spq	Epsilon, kappa, mu
5	Lambda	EcoR1 dan Sp1	Alpha, epsilon, eta, theta
6	Lambda	XmaIII dan Sp1	Alpha, beta, epsilon
7	Lambda	EcoR1, XmaIII, dan Pst1	Alpha, iota, theta
8	Lambda	Keempat enzim	Alpha, epsilon, kappa, theta, zeta
9	Delta	Sp1	Epsilon, kappa
10	Mu	EcoR1	Alpha, theta, zeta
11	Gamma	Pst1	Epsilon, kappa, zeta
12	Iota	Pst1	Epsilon, kappa,

			zeta
13	Gamma	Sp1	Epsilon, eta
14	Iota	Sp1	Epsilon, eta
15	Gamma	Pst1	Zeta, delta
16	Iota	Pst1	Zeta, delta
17	Beta	EcoR1	Eta, theta
18	Beta	EcoR1 dan Pst1	Kappa, theta, zeta
19	Eta	Pst1	Kappa, zeta
20	Alpha	Keempat enzim	Alpha
21	Epsilon	Keempat enzim	Epsilon
22	Kappa	Keempat enzim	Kappa
23	Theta	Keempat enzim	Theta
24	Zeta	Keempat enzim	Zeta
25	Lambda	Keempat enzim	Alpha, epsilon, kappa, theta, zeta

Tabel 4.1. Hasil Percobaan

Sumber : Jungck, John R., Rama Viswanathan, *Algebraic and Discrete Mathematical Methods for Modern Biology*, 2015, Burlington : Academic Press

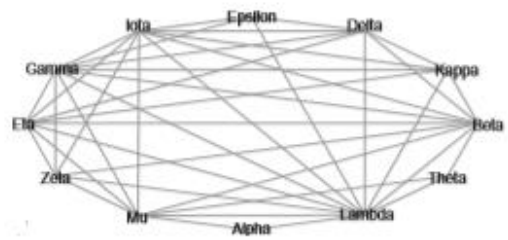
Tabel di atas diubah ke dalam bentuk matriks ketetangaan dan tiap sel matriks bernilai 0 jika suatu fragmen tidak saling meliputi dan bernilai 1 jika suatu fragmen saling meliputi. Fragmen dikatakan saling meliputi jika fragmen tersebut menghasilkan fragmen lainnya pada tabel di atas (kolom dua dan kolom tiga). Berikut adalah matriks ketetangaan yang dihasilkan

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	0	0	0	0	0	0	0	0	0	1	1			1 A	Alpha
2	0	1	1	1	0	1	1	1	1	1	1	1			2 B	Beta
3	0	1	1	1	1	1	1	0	1	1	1	1			3 C	Gamma
4	0	1	1	1	1	0	1	0	1	1	1	0			4 D	Delta
5	0	0	1	1	1	0	0	0	1	0	1	0			5 E	Epsilon
6	0	1	1	0	0	1	1	0	1	0	1	1			6 F	Zeta
7	0	1	1	1	0	1	1	0	1	1	1	1			7 G	Eta
8	0	1	0	0	0	0	0	1	0	0	1	1			8 H	Theta
9	0	1	1	1	1	1	1	0	1	1	1	1			9 I	Iota
10	0	1	1	1	0	0	1	0	1	1	1	0			10 J	Kappa
11	1	1	1	1	1	1	1	1	1	1	1	1			11 K	Lambda
12	1	1	1	0	0	1	1	1	1	0	1	1			12 L	Mu

Gambar 4.1. Matriks Ketetangaan yang Dihasilkan

Sumber : Buatan Penulis

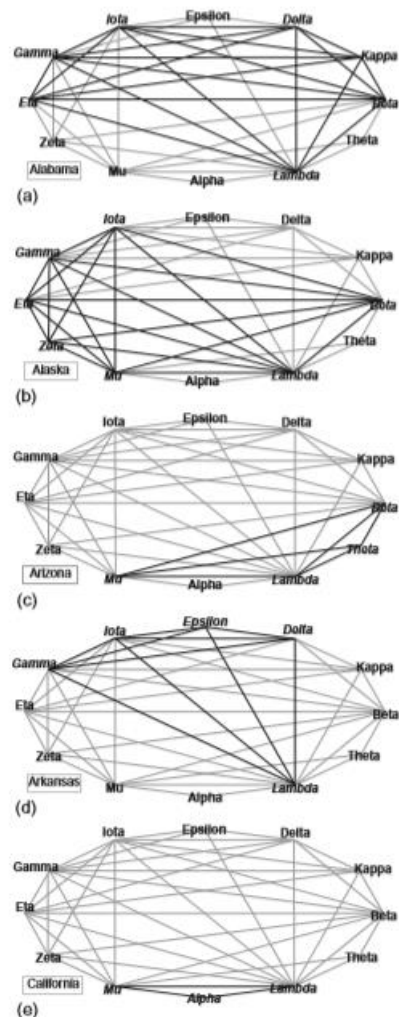
Dari matriks ketetangaan tersebut, maka dapat dibuat graf sebagai berikut



Gambar 4.2. Graf yang Dihasilkan

Sumber : Jungck, John R., Rama Viswanathan, *Algebraic and Discrete Mathematical Methods for Modern Biology*, 2015, Burlington : Academic Press

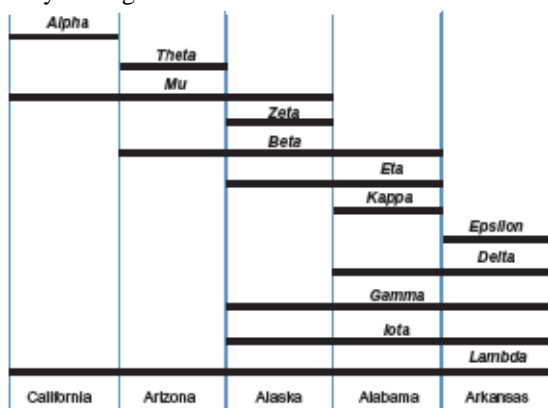
Graf di atas kemudian dibagi menjadi upagraf golongan maksimum. Ada lima buah upagraf yang terbentuk yang masing-masing dinamai Alabama, Alaska, Arizona, Arkansas, dan California.



Gambar 4.3. Upagraf yang Terbentuk

Sumber : Jungck, John R., Rama Viswanathan, *Algebraic and Discrete Mathematical Methods for Modern Biology*, 2015, Burlington : Academic Press

Berdasarkan lima upagraf tersebut, dibuat graf intervalnya sebagai berikut



Gambar 4.4. Graf Interval yang Terbentuk

Sumber : Jungck, John R., Rama Viswanathan, *Algebraic and Discrete Mathematical Methods for Modern Biology*, 2015, Burlington : Academic Press

Graf interval membuat konsep Pemetaan Benzer menjadi lebih mudah untuk dimengerti. Hal ini disebabkan oleh bentuknya yang sederhana dan tidak terlalu banyak garis tumpang-tindih. Oleh karena itu, mahasiswa yang mempelajari Pemetaan Benzer menggunakan konsep graf interval ini untuk memudahkan pemahaman konsep mereka.

V. KESIMPULAN

Graf dapat diterapkan dalam pemetaan susunan DNA, tepatnya dalam Pemetaan Benzer. Graf digunakan untuk menentukan fragmen DNA yang saling meliputi satu sama lain. Simpul-simpul graf menandakan fragmen-fragmen DNA dan sisi-sisi pada graf menandakan hubungan antar fragmen (apakah saling meliputi atau tidak). Setelah didapat pola melalui graf, pola tersebut dibuat menjadi graf interval sebagai visualisasi dari Pemetaan Benzer.

VI. UCAPAN TERIMA KASIH

Pertama-tama, penulis ingin menyampaikan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan berkah-Nya sehingga penulis bisa menyelesaikan makalah ini. Penulis juga ingin menyampaikan terima kasih kepada Dr. Rinaldi Munir S.T. dan Dra. Harlili sebagai Dosen Mata Kuliah Matematika Diskrit yang telah membimbing penulis.

DAFTAR PUSTAKA

- [1] De Los Rios P, Vendruscolo M. Network views of the cell. In: Buchanan M, Caldarelli C, De Los Rios P, Rao F, Vendruscolo M, editors. *Networks in cell biology*. New York, NY: Cambridge University Press; 2010. p. 4–13.
- [2] <http://ghr.nlm.nih.gov/handbook/basics/dna>, diakses pada 9 Desember 2015
- [3] Jungck, John R., Rama Viswanathan, *Algebraic and Discrete Mathematical Methods for Modern Biology*. 2015. Burlington : Academic Press

- [4] Munir, Rinaldi. "Diktat Kuliah IF2120 Matematika Diskrit". 2006. Bandung : Institut Teknologi Bandung
- [5] Rosen, K.H.. *Discrete Mathematics and Its Applications*. 2006. New York : McGraw-Hill

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 8 Desember 2015

Scarletta Julia Yapfrine - 13514074