

# Penerapan Graf pada PageRank

Hartono Sulaiman Wijaya  
13509046

Program Studi Teknik Informatika  
Sekolah Teknik Elektro dan Informatika  
Institut Teknologi Bandung, Jl. Ganessa 10 Bandung 40132, Indonesia  
[hartono.sulaiman@students.itb.ac.id](mailto:hartono.sulaiman@students.itb.ac.id)

## ABSTRAK

Graf merupakan salah satu cabang matematika yang sampai saat ini digunakan dalam banyak aplikasi. Telah terdapat berbagai aplikasi yang menggunakan graf sebagai acuan dan telah diterapkan di berbagai bidang. Salah satu terapan graf saat ini adalah penerapan PageRank oleh Google, sebuah search engine terkenal saat ini. PageRank adalah sebuah algoritma analisis link yang digunakan untuk memberi nilai kepada setiap elemen dari web dengan tujuan mengukur tingkat kepentingan dari sebuah web. Dengan demikian seorang pengguna mesin pencarian dapat menemukan web yang relevan dengan kata kunci yang dicari.

**Kata kunci:** graf, PageRank

## I. PENDAHULUAN

Graf adalah sebuah cara untuk merepresentasikan sebuah himpunan objek. Himpunan objek ini dianalogikan sebuah simpul dan relasi antara objek-objek digambarkan sebagai busur atau sisi yang menghubungkan dua simpul. Simpul ini dapat merepresentasikan sebuah objek unik, misal manusia, tempat tinggal, perusahaan, atau komputer. Sisi dapat menyatakan relasi antar manusia, jalan yang menghubungkan tempat tinggal, hubungan antar perusahaan, bahkan jaringan komputer.

Bila kita menganalogikan jaringan internet sebagai graf dengan halaman web sebagai simpul dan *hyperlink* sebagai sisi yang menghubungkan suatu web dengan web lainnya. Maka dapat diperoleh sebuah graf yang sangat besar yang memungkinkan sebuah simpul dapat mempunyai banyak sisi yang bersisian dengan simpul tersebut (sebuah web dengan banyak *hyperlink* yang mengacu kepada halaman web tersebut) dan begitu pula sebaliknya.

PageRank sendiri adalah sebuah cara untuk memberikan nilai kepada sebuah halaman web dan halaman web yang jumlahnya sangat banyak tersebut diurutkan berdasarkan nilai-nilai atau rank yang telah diproses berdasarkan banyak kriteria. Dengan kata lain, PageRank mengurutkan web-web berdasarkan tingkat

kepentingannya sehingga sebuah pencarian dapat memberikan hasil yang cocok dengan kata kunci yang dicari. Dasar pemikiran PageRank itu sendiri sangat berkorelasi dengan konsep manusia tentang kepentingan atau relevansi sebuah informasi.

Algoritma PageRank ini menganalisis link-link yang terdapat diantara halaman-halaman web dalam jaringan internet. Link yang mengacu ke sebuah halaman web dapat dianggap sebagai sebuah suara yang mendukung halaman web tersebut. Semakin banyak link yang mengacu, semakin tinggi pula nilai dari web itu dan semakin tinggi ranking dari web itu.

Saat ini terdapat banyak makalah ilmiah terkait dengan PageRank sejak makalah asli dari Page dan Brin dipublikasi. Dalam makalah ini hanya akan dijelaskan tentang bagaimana teori graf digunakan untuk membangun sebuah mesin pencarian.

## II. GRAF

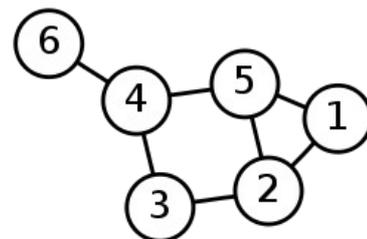
Secara matematis, graf dapat didefinisikan sebagai berikut:

**Definisi 2.1** Graf  $G$  didefinisikan sebagai pasangan himpunan  $(V, E)$ , ditulis dengan notasi  $G = (V, E)$ .  $V$  adalah himpunan tidak kosong dari simpul-simpul dan  $E$  adalah himpunan dari sisi yang menghubungkan sepasang simpul.

Secara umum, graf dapat digolongkan menjadi dua jenis graf:

### a. Graf sederhana

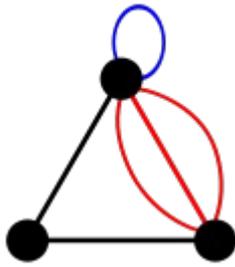
Graf sederhana adalah graf yang tidak memiliki gelang maupun simpul ganda.



**Gambar 2.1** Graf sederhana dengan 6 simpul dan 7 sisi

b. Graf tak sederhana

Graf tak sederhana adalah graf yang memiliki sisi ganda atau gelang.



**Gambar 2.2** Graf tak sederhana

Selain itu, sisi sebuah graf dapat mempunyai orientasi arah, oleh karena itu berdasarkan orientasi pada sisi secara umum, graf dapat dibedakan menjadi dua jenis:

a. Graf tak berarah

Graf yang sisinya tidak mempunyai orientasi arah disebut graf tak berarah. Jadi urutan pasangan simpul yang dihubungkan oleh sisi tidak diperhatikan.  $(u,v) = (v,u)$ .

b. Graf berarah

Graf yang tiap sisinya diberikan orientasi arah disebut sebagai graf berarah. Sisi dari graf berarah sering disebut sebagai busur. Pada graf berarah,  $(u,v)$  dan  $(v,u)$  menyatakan dua busur yang berbeda. Untuk busur  $(u,v)$ , simpul  $u$  dinamakan simpul asal dan simpul  $v$  dinamakan simpul terminal.

Selain penjelasan di atas, ada beberapa terminologi dasar graf yang perlu diketahui:

a. Lintasan (path)

**Definisi 2.2** 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 terbentuk  $v_0, e_0, v_1, e_1, v_2, e_2, \dots, e_n, v_n$ , sedemikian hingga  $e_1 = (v_0, v_1), e_2 = (v_1, v_2), \dots, e_n = (v_{n-1}, v_n)$  adalah sisi-sisi dari graf  $G$ .

b. Terhubung (connected)

Dua buah simpul  $u$  dan  $v$  dikatakan terhubung jika terdapat lintasan dari  $u$  ke  $v$ . Jika dua buah simpul terhubung, maka pasti simpul yang pertamadapat dicapai dari simpul yang kedua.

**Definisi 2.3** graf tak berarah  $G$  disebut graf terhubung jika untuk setiap pasang simpul  $u$  dan  $v$  di dalam himpunan  $V$  terdapat lintasan dari  $u$  ke  $v$  (yang juga harus berarti ada lintasan dari  $v$  ke  $u$ ). Jika tidak, maka graf  $G$  disebut graf tak terhubung.

**Definisi 2.4** graf berarah  $G$  dikatakan terhubung jika graf tak berarahnya terhubung (graf tak berarah dari  $G$  dapat diperoleh

dengan menghilangkan arahnya).

Dua buah simpul,  $u$  dan  $v$  pada graf berarah  $G$  disebut terhubung kuat jika terdapat lintasan berarah dari  $u$  ke  $v$  dan juga sebaliknya. Jika  $u$  dan  $v$  tidak terhubung kuat, tetapi tetap terhubung pada graf tak berarahnya maka  $u$  dan  $v$  dikatakan terhubung lemah.

**Definisi 2.5** graf berarah  $G$  disebut graf terhubung kuat apabila untuk setiap pasang simpul sembarang  $v_i$  dan  $v_j$  di  $G$  terhubung kuat, kalau tidak,  $G$  disebut graf terhubung lemah.

c. Graf berbobot (weighted graph)

Definisi 2.6 graf berbobot adalah graf yang setiap sisinya diberi sebuah harga (bobot).

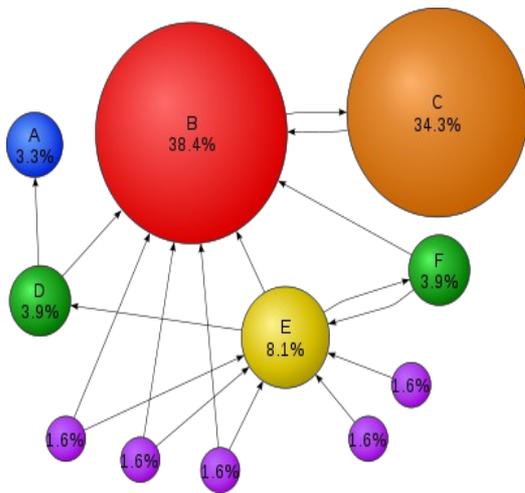
Namun graf berbobot sesungguhnya lebih luas lagi definisinya. Bobot tidak hanya dapat diberikan pada sisi, tetapi juga pada simpul.

### III. PAGERANK

Bila dilihat dari namanya, PageRank itu sendiri dapat diartikan ranking dari sebuah halaman web. PageRank suatu halaman web berkaitan erat dengan nilai dan peringkat halaman itu. Namun hal ini tidak sederhana itu, PageRank adalah sebuah distribusi peluang yang digunakan merepresentasikan keadaan ketika seseorang mengakses link secara random dan tiba di suatu halaman web tertentu.

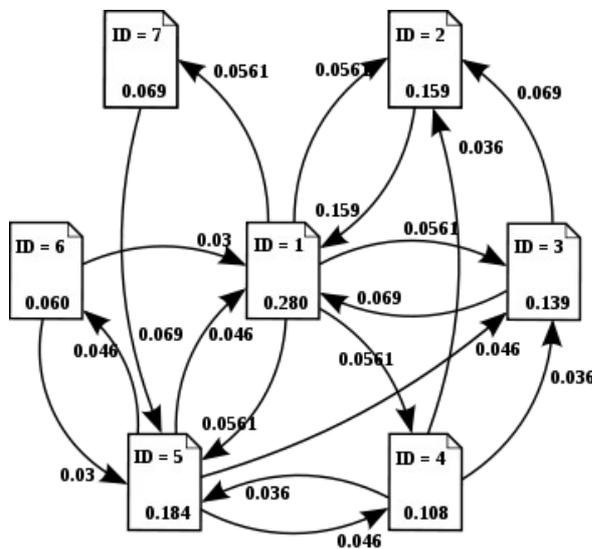
Secara sederhana, PageRank merupakan peluang yang mempunyai nilai antara 0 dan 1. Sebagai contoh, bila PageRank suatu halaman web adalah 0.5, maka peluang terjadinya kejadian dimana seseorang secara random mengklik link dan akan terarahkan ke halaman itu sebesar 50%.

Pada proses perhitungan PageRank untuk sekumpulan dokumen diasumsikan bahwa distribusi ini terbagi secara merata ke seluruh dokumen yang ada. Tentang bagaimana perhitungan peluang dan proses-proses yang dilalui tidak akan begitu dibahas pada makalah ini.



**Gambar 3.1** Model matematik PageRank untuk jaringan sederhana

Gambar 3.1 di atas mengilustrasikan sebuah jaringan internet yang disederhanakan dan dianalogikan sebagai graf berarah dan berbobot. Dengan simpul sebagai halaman web dan busur merupakan hyperlink dengan tujuan alamat halaman web tersebut. Bobot menyatakan nilai atau peringkat dari halaman web tersebut. Halaman C memiliki bobot yang lebih tinggi dari halaman E, namun memiliki link yang lebih sedikit. Peluang seseorang mengklik secara sembarang dan akan terarahkan ke halaman web E adalah 8.1 %.



**Gambar 3.2** Prinsip Kerja PageRank

Bagaimana sebenarnya PageRank bekerja akan dijelaskan sebagai berikut. Lihat gambar 3.2 di atas. Kita dapat menggunakan analogi seperti pada penjelasan sebelumnya, simpul mewakili halaman web dan busur mewakili link-link yang menghubungkan antara halaman web yang satu dengan halaman web yang lain.

Selain itu, graf di atas merupakan graf berbobot dengan bobot simpul merupakan PageRank dari halaman web tersebut. Kemudian busur juga mempunyai bobot tertentu yang mewakili nilai dukungan ke halaman web yang dituju.

Nilai dukungan ini adalah nilai dari PageRank yang “disumbangkan” ke halaman web lain bila ada suatu link dari web “penyumbang” ke web yang dituju. Dengan kata lain, PageRank suatu halaman web ditentukan oleh link-link yang menuju halaman web tersebut beserta dengan PageRank dari halaman web tempat link berasal.

Sebagai contoh, bila terdapat empat buah halaman web A, B, C, dan D. Asumsikan bahwa PageRank awal adalah sama dengan 0.25. Halaman B, C, dan D mempunyai link yang mengacu kepada A, maka PageRank A ( $PR(A)$ ) adalah

$$PR(A) = PR(B) + PR(C) + PR(D).$$

Andai saja B mempunyai link ke halaman C dan A sekaligus, dan halaman D mempunyai link ke A, B, dan C. Maka nilai dukungan yang diberikan akan dibagi rata ke semua link yang keluar dari halaman itu. Maka B akan memberikan nilai 0.125 ke A dan C, kemudian hanya sepertiga dari PageRank D diberikan untuk A.

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}.$$

Secara umum, PageRank dapat dirumuskan sebagai berikut:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)},$$

Dengan  $PR(v)$  adalah PageRank halaman yang mempunyai link ke halaman u, dan  $L(v)$  adalah jumlah link keluar dari halaman v.

Lihat gambar 3.2 di samping. Ada tujuh buah halaman web dengan PageRank masing-masing. Total dari semua PageRank itu akan mendekati satu. Hal ini berarti total dari PageRank semesta (dengan kata lain = internet) sama dengan satu. Lihat web dengan ID = 7,  $PR(7) \approx PR(1) / 5$ . Lihat pula web dengan ID = 4,  $PR(4) = PR(1) / 5 + PR(5) / 4$ . Begitu pula dengan halaman web lainnya. Mengapa tidak terjadi ketepatan misal  $PR(7) = 0.069$ , sedangkan link yang masuk bernilai 0.0561, hal ini disebabkan oleh adanya faktor *dumping*.

Dengan adanya PageRank ini, dapat diketahui bagaimana peluang terbukanya suatu halaman web bila seseorang membuka atau mengklik suatu link secara random dan akhirnya tiba di suatu halaman web tertentu.

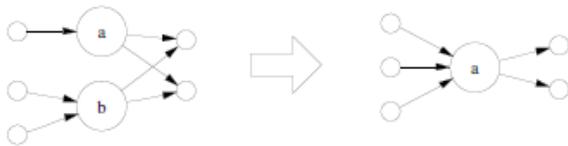
Hal ini yang mendasari sebuah mesin pencarian seperti Google untuk menemukan halaman web yang terkait dengan kata kunci pencarian.

Selain penggambaran cara kerja PageRank itu, graf

masih digunakan untuk berbagai hal dalam sistem pencarian ini. Misal *vote by committee*, *collapsing*, *proxy*, *delete*, *duplicate*, dan sebagainya. Penjelasan mengenai beberapa hal berikut tidak akan begitu dibahas dalam makalah ini.



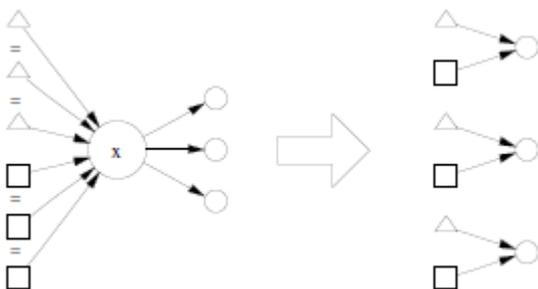
Gambar 3.3 *vote by committee*



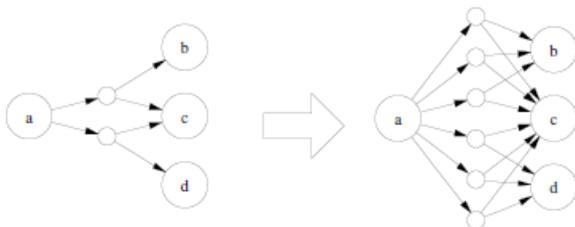
Gambar 3.4 *collapsing*



Gambar 3.5 *proxy*



Gambar 3.6 *delete*



Gambar 3.7 *duplicate*

Tentunya untuk membangun sebuah mesin pencarian

yang baik tidaklah sesederhana dengan pemaparan-pemaparan sebelumnya. Teori graf dapat digunakan dan digabung dengan algoritma-algoritma lainnya agar pembangunan graf menjadi lebih efektif. Misal pada gambar 3.6 memperlihatkan proses penghapusan sebuah simpul dari graf. Sedangkan gambar 3.7 memperlihatkan sebuah graf yang diduplikasi simpul-simpulnya.

Penggunaan graf pun dapat menjadi semakin kompleks, sebagai ilustrasi akan dilampirkan gambar dari makalah orisinal buatan Page dan Bin [3]:

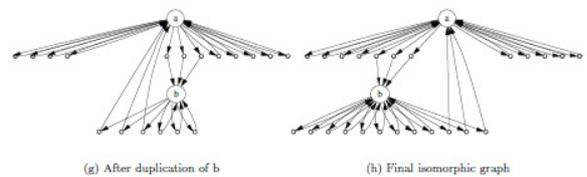
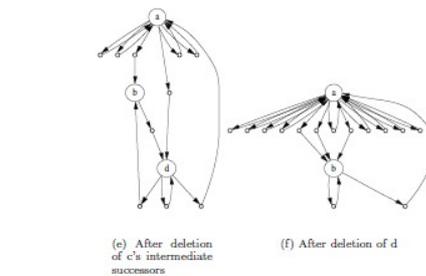
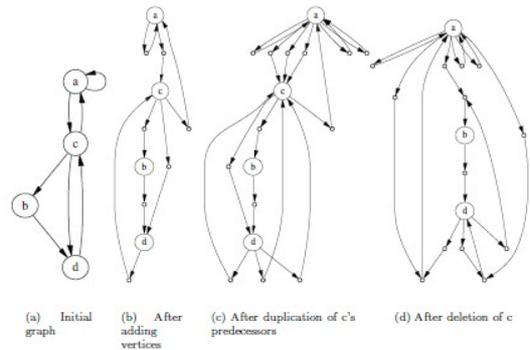


Figure 4: Example run of the completeness algorithm. Here  $a \neq b$ .

Gambar 3.8 Ilustrasi yang dibuat dalam makalah Page dan Bin

#### IV. KESIMPULAN

Penerapan graf yang sudah sangat meluas ini menunjukkan kemampuan graf sebagai sebuah alat yang mempermudah kehidupan manusia. Konsep graf amat mudah diterapkan dan digunakan dalam berbagai bidang kehidupan manusia. Penggunaan graf dalam algoritma PageRank ini juga telah mendorong semakin berkembangnya algoritma baru yang lebih efektif. Pemaparan mengenai PageRank dalam makalah ini pun hanya sebagai dasar, tentunya saat ini PageRank sudah makin dikembangkan dan dimodifikasi sehingga menjadi jauh lebih baik.

#### REFERENSI

- [1] Munir, Rinaldi. 2006. Diktat Kuliah IF 2153 Matematika Diskrit, edisi keempat. Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung.
- [2] <http://en.wikipedia.org/wiki/PageRank> (waktu akses 14 Desember 2010 20:10)
- [3] <http://stanford.edu/~epsalon/pagerank.pdf> (waktu akses 14 Desember 2010 20:25)
- [4] [http://en.wikipedia.org/wiki/Graph\\_%28mathematics%29](http://en.wikipedia.org/wiki/Graph_%28mathematics%29) (waktu akses 14 Desember 2010 20: 35)

#### 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, 17 Desember 2010

ttd

Hartono Sulaiman Wijaya  
13509046