

Permodelan *World Wide Web* Menggunakan Teori Graf

Adi Purnama (13514006)¹

Program Studi Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

¹adipurnama@s.itb.ac.id

Abstrak — *World Wide Web* adalah kumpulan halaman web yang saling terhubung menggunakan *hyperlink*. Struktur relasi antar halaman web yang terdapat dalam *World Wide Web* dapat dimodelkan menjadi sebuah graf berarah, dengan halaman web sebagai simpul dan *hyperlink* sebagai sisi. Karena sifat *World Wide Web* yang berukuran besar, tidak memiliki pusat, dan dapat berubah secara dinamis, graf ini memiliki beberapa karakteristik unik. Pengkajian mengenai karakteristik graf ini bermanfaat untuk menentukan algoritma *crawling*, *searching* & *ranking*. Ketiga algoritma ini digunakan pada mesin pencari. Makalah ini akan membahas beberapa teori untuk memodelkan graf *World Wide Web* dan analisis struktur graf *World Wide Web* berdasarkan data empiris.

Kata Kunci — *Web Graph*, *Erdos-Renyi Random Graph*, *Small World Network*, *Scale Free Network*, *Crawler*, *World Wide Web*.

I. PENDAHULUAN

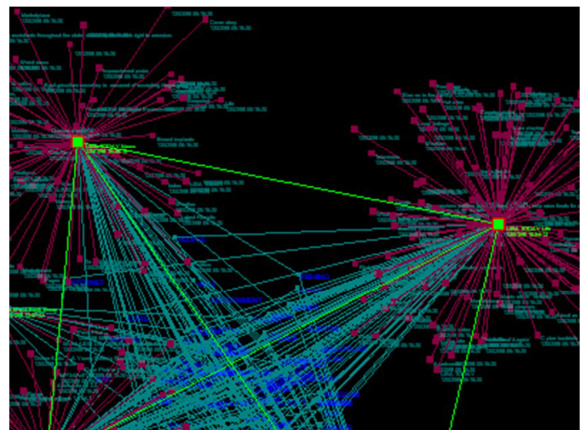
World Wide Web merupakan sebuah sistem informasi global yang terdiri dari halaman web yang memiliki alamat (*Uniform Resource Locator*), dan memiliki keterkaitan dengan halaman web yang lain menggunakan *hyperlink*. Halaman web ini dapat diakses menggunakan internet. *World Wide Web* diciptakan oleh Tim Berners-Lee dan peneliti lainnya di laboratorium energi tinggi CERN Geneva pada tahun 1989. Pada awalnya, WWW berfungsi sebagai sarana berbagi informasi antar peneliti di laboratorium tersebut. Sekarang, WWW telah berkembang menjadi repositori informasi global & sebuah media komunikasi baru. *World Wide Web* menyediakan ruang bagi kita untuk berbagi informasi secara global.

World Wide Web memiliki karakteristik yang unik, diantaranya adalah sifatnya yang terdesentralisasi, tidak memiliki pusat, dan perkembangannya yang cepat. Berbeda dengan beberapa jaringan lainnya (seperti jaringan listrik, telepon & jalan), *World Wide Web* tidak memiliki struktur yang terencana oleh sistem. Jaringan *hyperlink* yang terbentuk antar halaman web merupakan kumpulan aksi yang tidak terkoordinasi dari individu-individu pengguna *World Wide Web*. [3] Sistem yang terdesentralisasi ini merupakan keunggulan yang dimiliki

oleh *World Wide Web*. Berkat sistem ini, pengguna dapat menambahkan informasi baru ke dalam *World Wide Web* secara mudah. Hal ini memungkinkan penyebaran informasi yang cepat.

Meskipun demikian, keunggulan-keunggulan *World Wide Web* ini juga menimbulkan permasalahan baru. Akibat strukturnya yang besar dan terus berubah, proses pencarian informasi juga semakin sulit. Oleh karena itu, diperlukan sebuah strategi untuk melakukan pencarian di dalam *World Wide Web*. Strategi ini dapat dikembangkan setelah kita mengetahui bagaimana struktur relasi yang menghubungkan antar halaman web.

Struktur relasi pada *World Wide Web* dapat dimodelkan sebagai sebuah graf berarah. Graf *World Wide Web* dapat dinyatakan $G = (V, E)$, dimana V adalah himpunan tidak kosong dari halaman web ($V = v_1, v_2, v_3, \dots$). & E adalah himpunan pasangan (v_1, v_2) yang menunjukkan bahwa ada *hyperlink* dari halaman web a ke halaman web b .



Gambar 1. Representasi *World Wide Web* sebagai sebuah graf. Simpul pada graf merepresentasikan sebuah simpul, sementara sisi pada graf merepresentasikan *hyperlink* pada halaman web. Citra ini dihasilkan oleh program *Internet Cartographer*.

Sumber gambar :

<http://www.visualcomplexity.com/vc/project.cfm?id=16#>

Diakses 7 Desember 2015

Struktur graf pada *World Wide Web* pertama kali dimodelkan menggunakan teori graf. Meskipun demikian, beberapa penelitian mencoba untuk mengungkap struktur graf *World Wide Web* yang sesungguhnya dengan menjelajahi *World Wide Web* menggunakan *crawler*.

Hasil penelitian ini menunjukkan bahwa secara makroskopis, struktur graf pada *World Wide Web* memiliki karakteristik yang unik.

II. TERMINOLOGI MENGENAI GRAF

Berikut adalah beberapa terminologi mengenai graf yang digunakan pada makalah ini. Sebuah graf berarah didefinisikan sebagai pasangan yang terdiri dari himpunan simpul V yang tidak kosong dan himpunan sisi E . Sebuah graf yang tidak memiliki sisi disebut dengan graf kosong. Sementara itu, sebuah graf yang memiliki sisi ke semua simpul lainnya disebut dengan graf lengkap.

Sisi E didefinisikan sebagai pasangan simpul (v_1, v_2) , dimana v_1 disebut dengan simpul awal, dan v_2 disebut dengan simpul akhir. Derajat masuk v (*indegree*) didefinisikan sebagai jumlah sisi $(v_1, v) \dots (v_n, v)$ (jumlah sisi yang masuk ke simpul v), sementara itu derajat keluar v (*outdegree*) didefinisikan sebagai jumlah sisi $(v, v_1) \dots (v, v_n)$ (jumlah sisi yang keluar dari simpul v).

Simpul u & v disebut terhubung kuat apabila ada sisi yang menghubungkan dari u ke v dan juga sebaliknya (dari v ke u). Sementara itu, simpul u & v disebut terhubung lemah apabila u & v terhubung pada graf tidak berarahnya, namun tidak terhubung kuat.

Sebuah lintasan u ke v adalah rangkaian sisi yang menghubungkan dari u ke v , sementara itu panjang lintasan adalah jumlah sisi dalam lintasan itu. Sebagai contoh, apabila (u, v) adalah sebuah sisi, maka panjang lintasan u ke v adalah 1.

Sementara itu, diameter graf didefinisikan sebagai nilai tertinggi dari seluruh lintasan terpendek (u, v) pada seluruh pasangan simpul (u, v) . Diameter graf juga dapat dipandang sebagai panjang lintasan antar duasinggul yang terjauh.

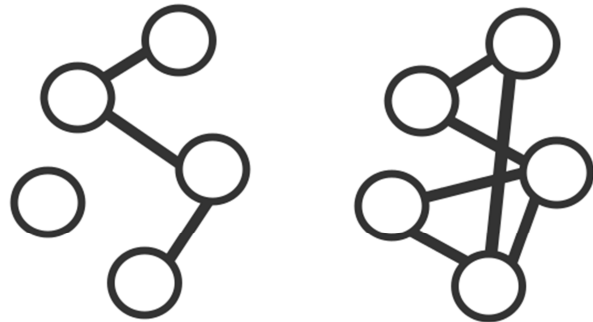
III. PERMODELAN GRAF WORLD WIDE WEB

A. Model Graf Acak (Random Graph) Erdos-Renyi

Relasi yang terbentuk antar halaman – halaman web merupakan murni keputusan individual pengguna *world wide web*, bukan keputusan sistem internal dari *world wide web*. Oleh karena itu, kemunculan hyperlink & halaman web baru yang ada di dalam *world wide web* tidak dapat diprediksi. Atas pertimbangan itu, graf *world wide web* dimodelkan sebagai graf acak (*Random Graph*).

Erdos & Renyi [14] merupakan orang yang pertama kali mengemukakan model graf acak. Graf ini dimulai dengan graf kosong dengan n buah simpul. Setiap $\binom{n}{2}$ pasang simpul dihubungkan oleh sebuah sisi dengan probabilitas p . Nilai p ini disebut juga nilai probabilitas pembentukan

sisi. Sebuah graf disebut dengan graf acak Erdos-Renyi apabila nilai p -nya tidak terlalu dekat dengan 0 atau 1. Ketika p mendekati 1, graf ini cenderung membentuk sebuah graf lengkap. Namun, ketika probabilitas p mendekati 0, graf ini cenderung tetap menjadi graf kosong. Jumlah sisi yang terbentuk pada probabilitas p adalah $p \binom{n}{2}$ atau $\frac{p \times n \times (n-1)}{2}$.



Gambar 2. Graf pada bagian kiri adalah graf acak Erdos-Renyi dengan $n = 5$, $p = 0,3$. Sementara itu, graf pada bagian kanan adalah graf acak Erdos-Renyi dengan $n = 5$, $p = 0,6$. Perhatikan bahwa peningkatan nilai probabilitas pembentukan sisi p akan mengakibatkan meningkatnya jumlah sisi pada graf

Namun, model graf ini tidak cukup untuk memodelkan graf *world wide web* secara keseluruhan. Pada model graf ini, jumlah simpul selalu tetap (n buah simpul). Sementara itu, pada *world wide web*, jumlah simpul (halaman web) dapat bertambah dan berkurang.

Selain itu, pada model ini, nilai probabilitas pembentukan sisi pada setiap pasang simpul selalu seragam. Pada kenyataannya, dalam *world wide web*, probabilitas ini bernilai acak & tidak seragam.

Model graf ini juga hanya dapat memodelkan keadaan *world wide web* secara statis. Model graf ini tidak dapat menggambarkan keadaan *world wide web* yang berubah. Pada kenyataannya, halaman web dan *hyperlink* pada *world wide web* dapat bertambah & berkurang. Oleh karena itu, dibutuhkan model graf yang lain untuk memodelkan graf *World Wide Web* ini. Selain itu, dibutuhkan data empiris mengenai struktur graf *World Wide Web* yang sebenarnya untuk menyusun sebuah model graf yang akurat.

B. Model Graf Small World Network (Watts-Strogats)

Fenomena “Dunia Sempit” (*Small World Phenomenon*) adalah sebuah fenomena yang berasal dari kajian jejaring sosial, bahwa tiap orang di dunia mempunyai keterkaitan dekat dengan orang lain berdasarkan kesamaan kenalannya. Kata *Small World* – “Dunia itu sempit” berasal dari kata yang biasa kita ungkapkan ketika kita bertemu dengan orang asing, dan ternyata orang asing tersebut memiliki kenalan dengan salah satu teman kita.

Secara formal, *Small World Network* adalah sebuah graf yang mayoritas simpulnya tidak saling bertetangga satu sama lain, namun setiap pasangan simpulnya memiliki panjang lintasan yang pendek. Pada graf ini, panjang lintasan antar dua buah simpul sembarang (L) berbanding lurus dengan logaritma dari jumlah simpul pada graf (n). $L \propto \log(n)$.

Berbagai macam jaringan yang terdapat di alam dapat dimodelkan menggunakan graf ini, seperti jaringan syaraf pada cacing *C.elegans*, jaringan kabel transmisi listrik, graf kolaborasi peneliti, dan graf pada jejaring sosial.

Pada model graf ini, sekumpulan simpul cenderung untuk membentuk kelompok-kelompok (*cluster*) yang saling terhubung satu sama lain. Model graf ini cocok untuk memodelkan *world wide web*, karena sekumpulan halaman web yang saling berkaitan cenderung membentuk sebuah kelompok halaman web yang disebut dengan *website*.

Selain itu, salah satu alasan pemilihan model graf *small world network* sebagai model graf *world wide web* adalah bukti eksperimen yang dilakukan oleh Albert, Jeong & Barabasi pada tahun 1999. Berdasarkan penelitian yang mereka lakukan, diameter graf *world wide web* dapat ditentukan dengan persamaan $d = 0.35 + 2.06 \log(n)$, dimana d adalah nilai maksimal panjang lintasan antar dua buah simpul sembarang, dan n adalah jumlah simpul pada graf. Berdasarkan persamaan tersebut, terlihat bahwa nilai d berbanding lurus dengan $\log(n)$. Hal ini mengindikasikan sifat "*Small World Network*" yang dimiliki oleh graf *World Wide Web*.

Berdasarkan persamaan di atas, dengan memasukkan nilai estimasi jumlah halaman web yang ada di *World Wide Web* ($N = 8 \times 10^8$), maka d adalah 18,59. Jadi, berdasarkan teori ini, kita bisa menjelajahi *World Wide Web* cukup dengan maksimal 19 klik pada *hyperlink*.

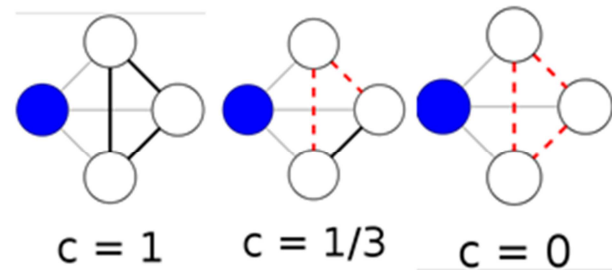
Terdapat dua karakteristik yang terdapat di dalam model graf *Small World Network*, yaitu *characteristic path length* (L) dan *clustering coefficient* (C).

Characteristic path length adalah ukuran seberapa jauh dua buah simpul terpisah. Untuk menghitungnya, pertama hitung jarak lintasan rata-rata dari sebuah simpul v ke seluruh simpul lainnya.

$$d_v = \frac{\sum_{d \neq w} d(v, w)}{|V(G)| - 1}$$

Setelah panjang lintasan rata-rata dari seluruh simpul $v \in V(G)$, hitung median dari seluruh d_v , nilai median ini adalah *characteristic path length*. Secara sederhana, *characteristic path length* adalah panjang lintasan rata-rata untuk mencapai sembarang pasangan simpul pada suatu graf.

Sementara itu *clustering coefficient* (C) adalah ukuran seberapa besar kecenderungan sebuah simpul untuk membentuk kelompok. Nilai C adalah perbandingan jumlah sisi yang terbentuk antar tetangga simpul n dengan jumlah sisi maksimum yang terbentuk antar tetangga simpul n . Sebagai contoh, lihat ilustrasi di bawah ini.



Gambar 3. Ilustrasi local clustering coefficient pada simpul berwarna biru
 Sumber gambar : https://en.wikipedia.org/wiki/Clustering_coefficient
 Diakses 8 Desember 2015

Pada gambar pertama (terkiri), nilai C adalah satu. Jumlah maksimal sisi yang terbentuk antar tetangga simpul biru adalah 3, sementara jumlah sisi sebenarnya yang terbentuk antar tetangga simpul biru adalah 3. Jadi $C = 3 / 3 = 1$. Sementara itu, pada gambar kedua (tengah), nilai C adalah $1/3$. Jumlah maksimal sisi yang terbentuk antar tetangga simpul biru adalah 3, sementara jumlah sisi sebenarnya yang terbentuk adalah 1. Jadi nilai C adalah $1/3$. Nilai C ini adalah *local clustering coefficient* untuk simpul berwarna biru. Untuk mendapatkan nilai C untuk keseluruhan graf, dapat dihitung dengan merata-ratakan seluruh nilai C pada setiap simpul.

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$

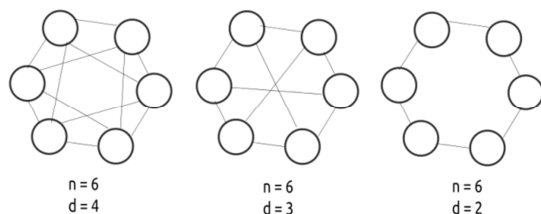
Berdasarkan kedua karakteristik ini, Adamic [4] telah membuktikan sifat *Small World Network* dari graf *World Wide Web* berdasarkan definisi dari Watts & Strogatz. Watts & Strogatz mendefinisikan sifat-sifat dari graf *Small World Network* sebagai berikut.

1. Nilai *clustering coefficient* (C) pada *Small World Network* lebih besar daripada graf acak Erdos Renyi (*Erdos Renyi Random Graph*) dengan jumlah simpul yang sama.
2. Nilai *characteristic path length* (L) pada *Small World Network* kurang lebih sama kecilnya dengan L pada graf acak Erdos Renyi (*Erdos Renyi Random Graph*) dengan jumlah simpul yang sama.

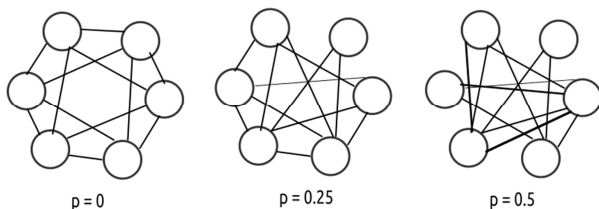
Adamic [4] melakukan penelitian dengan menggunakan data uji sekitar 50 juta halaman web dan

269.794 website. Data ini dikumpulkan menggunakan crawler Alexa . Dengan menganalisis graf dengan website berdomain .edu , dia menemukan bahwa nilai L graf .edu adalah 4,062 , sementara itu nilai L graf acak Erdos-Renyi adalah 4,048. Nilai C pada graf .edu adalah 0.156 , sementara itu nilai C pada graf acak Erdos-Renyi adalah 0.0012 . Berdasarkan hasil penelitian tersebut, dapat disimpulkan bahwa graf world wide web memiliki karakteristik Small World Network , sesuai dengan definisi yang ditetapkan oleh Watts & Strogatz.

Watts & Strogatz [5] menjelaskan mengenai prosedur pembentukan graf Small World Network yang disebut dengan Random Rewiring Procedure. Pertama , buat graf dengan n simpul , setiap simpul terhubung dengan d buah simpul yang terdekat. Jumlah sisi yang terbentuk pada graf ini adalah $m = \frac{n \times d}{2}$. Graf ini disebut juga graf teratur.



Gambar 4. Beberapa contoh graf teratur



Gambar 5. Graf "Small World Network" disusun dari graf teratur dengan n=6 dan d = 4 dengan berbagai probabilitas p.

Saat nilai probabilitas $p = 0$, graf masih berupa graf teratur. Saat nilai $p = 0.25$, terdapat $m = \frac{0.25 \times 6 \times 4}{2} = 3$ buah sisi yang dipindahkan secara acak. Saat nilai $p = 0.5$, terdapat 6 buah sisi yang dipindahkan secara acak. Ketika nilai p mendekati 1 , graf ini akan cenderung membentuk graf acak Erdos-Renyi .

Permodelan graf World Wide Web menggunakan "Small World Network" lebih baik daripada permodelan menggunakan graf acak Erdos Renyi karena permodelan ini telah didukung menggunakan bukti data empiris. [4] Meskipun demikian , model graf ini masih belum dapat menjelaskan sifat dinamis dari world wide web. Model ini tidak menjelaskan proses penghapusan/penciptaan halaman web & hyperlink baru. Jumlah simpul dan sisi pada model ini selalu tetap.

C. Model Graf "Scale Free Networks" (Barabasi, Albert)

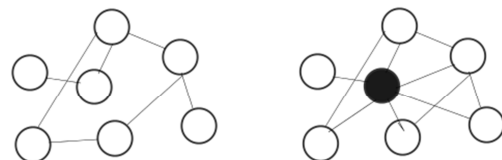
Seperti yang telah dijelaskan, beberapa model graf sebelumnya masih belum mencukupi untuk memodelkan graf World Wide Web. Pada model ER (Erdos-Renyi Random Graph Model) , sisi-sisi diletakkan secara acak , pada model SW (Watts-Strogatz Small World Network) , sisi-sisi hanya dipindahkan dari graf teratur . Kedua model graf ini tidak menggambarkan graf di dunia nyata, yang memungkinkan penambahan / pengurangan simpul baru.

Selain itu, pada model ER & SW , probabilitas pasangan simpul membentuk sebuah sisi (p) diasumsikan sama dan seragam. Kenyataan di dunia nyata tidak demikian. Sebuah simpul baru cenderung membentuk sisi dengan simpul yang telah memiliki banyak sisi.[6].

Berdasarkan dua kelemahan ini, Barabasi & Albert [6] menyusun sebuah model graf "Scale Free Network" , yang memiliki karakteristik sebagai berikut :

1. Jaringan graf dapat berkembang dengan penambahan simpul-simpul baru.
2. Simpul yang baru ditambahkan cenderung membuat sisi dengan simpul yang telah memiliki banyak sisi.

Ciri dari graf Scale Free Network persebaran derajat pada setiap simpulnya yang tidak merata. Terdapat simpul yang berperan sebagai hub . Simpul ini terhubung dengan banyak simpul lainnya sekaligus.[7]

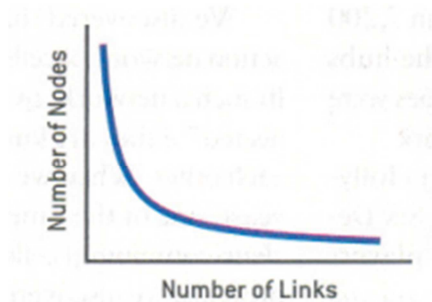


Gambar 6. Perbandingan graf acak Erdos-Renyi (kiri) dengan graf "Scale Free Network" (kanan). Pada graf Scale Free Network , terdapat simpul hub yang terhubung dengan banyak simpul sekaligus. Simpul hub ditandai dengan warna hitam

Kemunculan simpul hub ini disebabkan oleh kecenderungan simpul baru untuk membentuk sisi dengan simpul yang memiliki sisi terbanyak. Ungkapan "yang kaya semakin kaya (rich get richer) " cocok untuk menggambarkan situasi ini. Simpul yang memiliki banyak sisi akan cenderung memiliki lebih banyak sisi seiring bertambahnya simpul baru.

Distribusi derajat sisi pada simpul graf Scale Free Network mengikuti hukum pangkat (power law). Probabilitas sebuah simpul memiliki derajat k berbanding

lurus dengan $\frac{1}{k^n}$, dimana n adalah sebuah konstanta



Gambar 7. Grafik Distribusi Hukum Pangkat.
Sumber Gambar : Albert-Laszlo Barabasi, Eric Bonabeau, "Scale Free Networks", in Scientific American, 2003

Grafik diatas menunjukkan bahwa mayoritas simpul memiliki jumlah sisi yang sedikit, namun ada simpul *hub* yang memiliki jumlah sisi yang banyak.

Proses pembentukan graf *Scale Free Network* dapat dijelaskan menggunakan mekanisme *Preferential Attachment Model*. Pertama, buat sebuah graf kosong dengan n buah simpul. Lalu, setiap satu satuan waktu, tambahkan sebuah simpul baru, dengan m buah sisi. Probabilitas sebuah simpul baru membentuk sisi dengan simpul i adalah

$$p = \frac{d_i}{\sum_j d_j}$$

Dimana d_i adalah derajat simpul i, $\sum_j d_j$ adalah jumlah derajat seluruh simpul pada graf. Dengan rumus ini, simpul baru akan memiliki kecenderungan untuk membentuk sisi dengan simpul yang memiliki derajat tertinggi.

IV. ANALISIS STRUKTUR GRAF *WORLD WIDE WEB* BERDASARKAN DATA EMPIRIS

Pada bab sebelumnya telah dijelaskan beberapa permodelan *world wide web* menggunakan teori graf. Namun, untuk mendapatkan gambaran nyata mengenai struktur graf *world wide web*, kita harus menjelajahi seluruh halaman web di dalam *world wide web*, mengunduh setiap halamannya & menganalisis keterkaitan antar halaman-halaman web tersebut.

Untuk melakukan penjelajahan & pengunduhan halaman web di dalam *world wide web*, digunakanlah program *web crawler*. *Web crawler* adalah program untuk menjelajahi

isi *world wide web* dengan otomatis dan sistematis. [10] *Web crawler* menjelajahi *world wide web* dengan mengikuti struktur *hyperlink* yang terdapat di dalam halaman web

Pertama-tama, *web crawler* diberikan sebuah daftar URL yang harus ia kunjungi. Lalu, *crawler* akan mengunjungi URL tersebut, mengunduh halaman web pada URL, mengidentifikasi seluruh *hyperlink* yang terdapat di dalam halaman web dan memasukkan daftar *hyperlink* tersebut ke daftar URL yang harus ia kunjungi. Proses ini dilakukan berulang-ulang hingga *crawler* memutuskan kapan untuk berhenti. [9]

Menurut Trupti, dkk [10] ada beberapa komponen dasar yang dimiliki oleh *web crawler*, yaitu.

1. *Crawler Frontier*
Sebuah daftar URL yang akan dikunjungi oleh *web crawler*
2. *Page Downloader*
Program untuk mengunduh halaman web berdasarkan URL yang ada di dalam *Crawler Frontier*
3. *Web Repository*
Sebuah tempat penyimpanan halaman web yang telah berhasil diunduh oleh *web crawler*

Web crawler digunakan oleh peneliti graf *world wide web* untuk menganalisis struktur graf pada *world wide web*. Selain itu, *web crawler* juga digunakan dalam mesin pencari. *Crawling* (proses pengunduhan halaman-halaman web untuk dianalisis lebih lanjut) merupakan proses yang pertama kali mesin pencari lakukan dalam pencarian informasi di dalam *world wide web*.

Karena struktur *world wide web* yang begitu besar, sampai saat ini, belum ada yang berhasil menjelajahi seluruh isi *world wide web*. Pada banyak kasus, *crawler* tidak mampu untuk mengunduh seluruh halaman web yang ada di *world wide web*, bahkan mesin pencari-pun hanya berhasil mengunduh sebagian kecil dari keseluruhan web [9]. Oleh karena itu, beberapa analisis struktur graf yang akan dijelaskan berikut ini bergantung pada jumlah halaman web yang berhasil diunduh oleh *web crawler*. Perbedaan metode *crawling* yang digunakan oleh *crawler* juga mempengaruhi hasil analisis struktur graf.

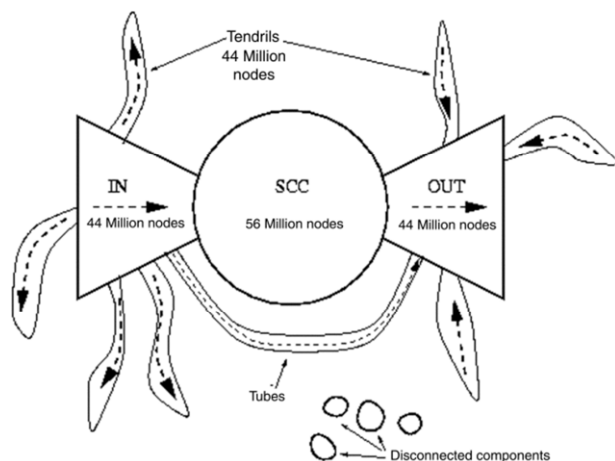
A. Struktur Graf Dasi Kupu-Kupu, Broder, dkk

Broder, dkk [8] menjelaskan struktur graf *world wide web* berdasarkan data empiris. Data ini dikumpulkan menggunakan *crawler* Altavista untuk menjelajahi *hyperlink* yang terdapat di dalam halaman web. Terdapat sekitar 200 juta halaman web dan 1,5 milyar *hyperlink* yang dianalisis di dalam penelitian ini.

Penelitian ini mengonfirmasi berlakunya hukum pangkat

(*power law*) pada derajat masuk & derajat keluar. Nilai konstanta hukum pangkat yang ditemukan adalah 2,1. Hal ini menunjukkan bahwa graf *World Wide Web* dapat dimodelkan menggunakan graf *Scale Free Network*.

Selain itu, graf ini memiliki struktur makroskopik yang unik. Apabila graf ini dianggap sebagai graf yang tidak berarah, sekitar 90% simpul graf membentuk sebuah kesatuan komponen yang terhubung. Komponen terhubung ini terbagi menjadi empat bagian.



Gambar 8. Bow Tie Structure, sebuah alternatif struktur graf *world wide web* yang diajukan oleh Broder,dkk [8]
 Sumber gambar : A.Broder, R. Kumar, F.Maghouli, P.Raghavan, S. Rajagopalan, S. Stata, A. Tomkins, J. Wiener, "Graph structure in the web", in *Computer Networks*, 2000

1. SCC (27,74 %)

SCC merupakan singkatan dari *Strongly Connected Component*. Setiap pasangan simpul di dalam SCC terhubung kuat. Andaikan u & v adalah simpul, maka terdapat sisi dari u ke v dan sebaliknya.
2. IN (21,29 %)

IN terdiri dari simpul yang memiliki sisi menuju komponen SCC, namun tidak sebaliknya. Halaman web pada IN dapat mengakses SCC, namun SCC tidak memiliki link menuju IN.
3. OUT (21,21 %)

OUT terdiri dari simpul yang memiliki sisi dari komponen SCC, namun tidak sebaliknya. Komponen SCC dapat mengakses OUT, namun OUT tidak memiliki link menuju SCC.
4. TENDRILS (21,52%)

Terdiri dari simpul yang tidak dapat diakses dari SCC dan tidak dapat mengakses SCC. IN dapat mengakses TENDRILS atau TENDRILS dapat mengakses komponen OUT.
5. TUBES¹

Terdiri dari simpul yang dapat diakses dari IN

¹ Dalam penelitian yang dilakukan Broder,dkk [8] jumlah TUBES & TENDRILS digabung menjadi satu, yaitu 21,52 %

dan dapat mengakses OUT, tetapi bukan bagian dari SCC.

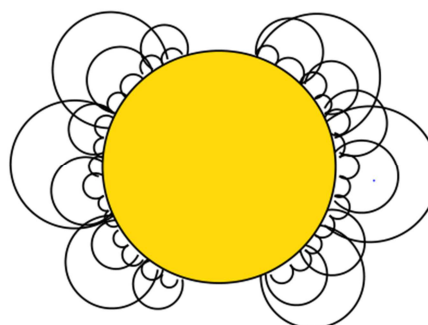
6. DISCONNECTED (8,24 %)

Terdiri dari simpul yang tidak memiliki hubungan dengan komponen-komponen di atas.

B. Struktur Graf Bunga Aster, Donato,dkk

Penelitian mengenai struktur graf *world wide web* yang dilakukan oleh Broder,dkk [8] dilanjutkan oleh Donato,dkk[11]. Data graf uji yang digunakan di dalam penelitian ini terdiri dari 360 juta simpul dan 1,5 milyar sisi, dikumpulkan menggunakan *crawler* WebBase Project, Stanford.

Penelitian ini mengulang kembali eksperimen yang dilakukan oleh Broder,dkk, namun dengan data *crawler* yang lebih baru. Secara umum, mereka kembali menemukan struktur graf dasi kupu-kupu seperti yang telah dikemukakan oleh Broder,dkk[8]. Namun, mereka menemukan sebuah fakta baru bahwa komponen IN dan OUT pada graf ini memiliki kedalaman graf yang rendah dan bersifat terfragmentasi satu sama lain. [11] Mayoritas simpul pada IN dan OUT bertetangga secara langsung dengan simpul yang ada di dalam SCC. Berdasarkan hasil temuan ini, Donato,dkk [11] mengajukan struktur graf *world wide web* baru yang disebut dengan struktur graf bunga aster (*daisy*).



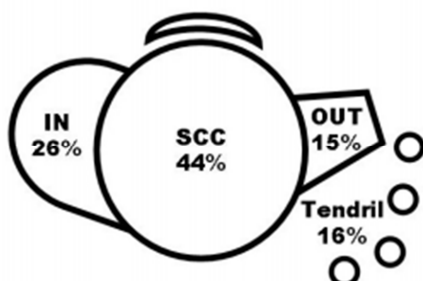
Gambar 9. Daisy Structure, sebuah alternatif struktur graf *world wide web* berdasarkan pengembangan dari *Bow Tie Structure*. Bagian pusat yang berwarna kuning menggambarkan komponen SCC, sementara itu, bagian kelopak menggambarkan komponen IN & OUT yang terfragmentasi.

Sumber gambar : Debora Donato, Stefano Leonardi, Stefano Milozzi, Panayiotis Tsaparas, "Mining the Inner Structure of the Web Graph" in *Eighth International Workshop on Web and Databases*, 2005

C. Struktur Graf Teko Teh, Zhu,dkk

Berdasarkan hasil penelitian Broder,dkk [8] dan Donato,dkk[11], Zhu,dkk [12] melakukan sebuah penelitian untuk membandingkan struktur bunga aster dengan struktur dasi kupu-kupu. Berbeda dengan dua penelitian sebelumnya, data di dalam penelitian ini berfokus kepada halaman web dari Tiongkok. Data graf uji yang digunakan pada penelitian ini adalah 837 juta halaman web dan 43 miliar *hyperlink*.

Berdasarkan hasil penelitian, graf web tersebut tidak membentuk struktur graf bunga aster maupun struktur dasi kupu-kupu, melainkan membentuk struktur graf baru yang disebut dengan struktur graf teko teh (*teapot*). Hal ini ditandai oleh ukuran relatif komponen IN yang jauh lebih besar daripada komponen OUT.



Gambar 10. Struktur graf teko teh, struktur graf ini merepresentasikan graf web di Tiongkok.

Sumber gambar : Jonathan J. H. Zhu, Tao Meng, Zhengmao Xie, Geng Li, Xiaoming Li, "A Teapot Graph and Its Hierarchical Structure of the Chinese Web", in WWW, 2008

Berdasarkan ketiga hasil penelitian ini, secara umum struktur graf *world wide web* terdiri dari beberapa komponen, yaitu SCC, IN, OUT, TENDRIL, TUBES & DISCONNECTED. Namun ukuran relatif masing-masing komponen bergantung & berubah secara dinamik terhadap ruang dan waktu. Hal ini disebabkan oleh sifat alami *world wide web* yang dapat berubah-ubah secara dinamis.

V. KESIMPULAN

Struktur *World Wide Web* dapat dimodelkan menjadi sebuah graf berarah, dimana halaman web merupakan simpul pada graf dan *hyperlink* merupakan sisi pada graf. Karena sifat strukturnya yang terdesentralisasi dan acak, graf ini dapat dimodelkan menggunakan model graf acak, seperti graf acak Erdos-Renyi (*Erdos-Renyi Random Graph Model*), graf *Small World Network Watts-Stogatz*, dan graf *Scale Free Networks*. Dari ketiga model graf ini, model graf *Scale Free Networks* lebih cocok untuk memodelkan graf *World Wide Web* karena sudah dikonfirmasi kebenarannya berdasarkan data empiris.[8]

Sementara itu, struktur graf *World Wide Web* juga dapat ditentukan menggunakan data empiris. Halaman-halaman web dalam jumlah besar diunduh dari *world wide web* menggunakan *web crawler* untuk dianalisis lebih lanjut. Berdasarkan hasil analisis beberapa penelitian [8,11,12], struktur graf ini terbagi menjadi beberapa komponen, yaitu SCC, IN, OUT, TENDRILS, TUBES & DISCONNECTED. Ukuran relatif masing-masing komponen bergantung pada ruang (wilayah analisis struktur web) dan waktu (kapan proses *crawling* itu dilakukan). Hal ini disebabkan oleh sifat alami *world wide web* yang dapat berubah secara dinamis

VII. UCAPAN TERIMA KASIH

Penulis mengucapkan syukur kepada Allah S.W.T, karena atas rahmat-Nya, penulis dapat menyelesaikan makalah ini. Selain itu, penulis juga mengucapkan terima kasih kepada Bapak Dr. Ir. Rinaldi Munir, M.T dan Ibu Harlili selaku dosen pengampu kuliah IF 2120 Matematika Diskrit yang telah memberikan ilmu yang bermanfaat kepada penulis.

REFERENSI

- [1] Narsigh Deo, Panjaj Gupta, "World Wide Web : Graph-Theoretic Perspective", University of Central Florida, 2001
- [2] Reka Albert, Hawoong Jeong, Albert-Laszlo Barabasi, "Diameter of World Wide Web" in Nature, Vol 401, 1999
- [3] Jon Kleinberg, Steve Lawrence, "The Structure of Web" in Science Vol 294, 2001
- [4] Lada Adamic, "The Small World Web" in Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries, 1999
- [5] D.Watts, S. Strogatz, "Collective dynamics of 'small world' networks" in Nature, Vol 393, 1998
- [6] Albert-Laszlo Barabasi, Reka Albert, "Emergence of Scaling in Random Networks", in Science, Vol 286, 1999
- [7] Albert-Laszlo Barabasi, Eric Bonabeau, "Scale Free Networks", in Scientific American, 2003
- [8] A.Broder, R. Kumar, F.Maghouli, P.Raghavan, S. Rajagopalan, S. Stata, A. Tomkins, J. Wiener, "Graph structure in the web", in Computer Networks, 2000
- [9] Jungho Choo, "Crawling the Web : Discovery and Maintenance of Large Scale Web Data", Stanford University, 2001
- [10] Trupti V. Udupure, Ravindra D. Kali, Rajesh C. Dharmik, "Study of Web Crawler and its Different Types", in IOSR Journal of Computer Engineering, 2014
- [11] Debora Donato, Stefano Leonardi, Stefano Milozzi, Panayiotis Tsaparas, "Mining the Inner Structure of the Web Graph" in Eighth International Workshop on Web and Databases, 2005
- [12] Jonathan J. H. Zhu, Tao Meng, Zhengmao Xie, Geng Li, Xiaoming Li, "A Teapot Graph and Its Hierarchical Structure of the Chinese Web", in WWW, 2008
- [13] Rinaldi Munir, "Diktat Kuliah IF2120 Matematika Diskrit", Program Studi Teknik Informatika, Institut Teknologi Bandung, 2006
- [14] P Erdos, A.Renyi, "On Random Graphs I", 1958

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

Adi Purnama