

FENOMENA *SMALL-WORLD* DAN APLIKASINYA DALAM SISTEM PENYEBARAN INFORMASI *PURE P2P*

Joelian Samuel – NIM : 13505092

Program Studi Teknik Informatika, Institut Teknologi Bandung

Jl. Ganesha 10, Bandung

E-mail: if15092@students.if.itb.ac.id

Abstrak

Makalah ini membahas fenomena *small-world* dan aplikasinya dalam sistem penyebaran informasi *pure P2P*. *Small-world* sendiri dapat dirumuskan sebagai berikut fenomena yang mengatakan bahwa hubungan antara dua orang yang tak saling kenal adalah berjarak enam pemisah hubungan sosial.

Small-world pertama kali dikemukakan oleh Karinthy pada tahun 1920-an dan dirumuskan kembali oleh seorang psikolog sosial, James Milgram, melalui percobaannya yang terkenal. Dari eksperimennya tersebut, terkenal istilah “enam-derajat pemisah”. Dalam perkembangannya lebih lanjut, hal ini menarik perhatian banyak pihak mulai dari matematikawan sampai aktor. Aktor Kevin Bacon membuat *Bacon Number*, dan matematikawan Paul Erdos membuat *Erdos number*. Selanjutnya, model pertama dalam graf dikemukakan oleh Strogatz dan Watt. Dan dikembangkan lebih lanjut oleh Kleinberg.

Kleinberg, dalam model yang dibuat, juga membuat sebuah algoritma dalam mencari jalan terpendek antara dua simpul. Dengan metode algoritma desentralisasi, model ini dapat mencari jalan terpendek yang harus dilalui antara dua simpul secara efektif. Keunggulan dari algoritma ini adalah simpul tidak harus tahu semua himpunan simpul dan sisi yang ada pada suatu graf, simpul yang sedang membawa pesan cukup tahu informasi lokal saja. Informasi lokal adalah informasi yang berupa tetangga yang memiliki hubungan lokal, tetangga hubungan jauh, dan alamat simpul tujuan.

Pure P2P arsitektur telah menjadi model yang penting untuk pertukaran informasi diantara kelompok pemakai yang dinamis, dengan biayanya yang rendah dan merupakan model yang sesuai dengan besar kelompok tersebut. Perkembangan terkini dari system ini telah memunculkan pertanyaan baru dan tantangan baru. Dengan mengidentifikasi dua factor kunci di sebuah lingkungan, dapat dibuat sebuah algoritma baru pencarian heuristic yang akan membuat penggunaan fenomena *small-world* lebih baik yang bertujuan mencari “enam-derajat pemisah” lebih baik. Dengan algoritma baru ini, ada peningkatan sekitar sepuluh persen dibandingkan algoritma BFS ketika *querying information related*. Dan meningkat 20 persen ketika berhubungan dengan *a shift of interest*. Algoritma heuristic juga memiliki kontrol yang lebih baik terhadap jumlah simpul yang harus dikunjungi menggunakan fitur Penghitung-Count dibandingkan mesin TTL.

Kata kunci: *Small-World Phenomenon, Milgram’s experiment, Pure P2P information-sharing system, P2P system, six degrees of separation, Watt’s model, long-range contacts, graph network application.*

I. PENDAHULUAN

Teori graf merupakan pokok bahasan dalam mata kuliah matematika diskrit yang dipelajari di tingkat dua program studi informatika ITB. Graf sendiri merupakan cara untuk merepresentasikan objek-objek diskrit dan hubungan antara objek-objek tersebut. Graf merupakan pokok bahasan yang sudah sangat tua, masalah pertama tentang graf yang tercatat dalam sejarah adalah masalah jembatan Konigsberg. Yang nantinya menurut teorema Euler, tidak ada penyelesaian untuk kasus seperti ini. Seiring perkembangan teknologi dan ilmu pengetahuan. Penggunaan graf juga semakin berkembang. Graf dipakai di berbagai disiplin ilmu, antara lain, kelistrikan, kimia, olahraga, dan ilmu komputer. Salah satu pertanyaan ataupun kegunaan yang timbul dari graf adalah fenomena Small-World.

Dari namanya kita bisa menebak kira-kira tentang apakah fenomena tersebut. Fenomena Small-World adalah fenomena yang menyebutkan bahwa semua orang di dunia ini dapat terhubung dengan orang lain di dunia ini yang sama sekali tidak dikenalnya hanya dalam sebuah rantai kenalan sosial yang pendek. Hal ini dikemukakan pertama kali oleh Karinthy pada tahun 1920-an, kemudian diteliti lebih lanjut oleh Stanley Milgram, seorang psikolog sosial, pada tahun 1967. Milgram dalam percobaannya mengatakan bahwa dua warga Negara Amerika Serikat yang diambil secara acak dapat dihubungkan melalui enam mata rantai kenalan mereka. Atau lebih lanjut akan disebut enam-tingkat pemisah. Sesuatu hipotesis yang masih menjadi perdebatan sampai sekarang.

Fenomena ini berkembang dengan sangat pesat dan menemukan penggunaannya pada system yang berhubungan dalam penyebaran informasi seperti system *peer to peer*. Dalam penyebaran informasi, dibutuhkan jalan terpendek untuk mendapatkan informasi tersebut, lalu dikirim kembali kepada sang peminta. Hal ini yang nantinya akan dibahas lebih lanjut.

Makalah ini sendiri terbagi menjadi dua bagian besar. Yang pertama adalah mengenai sejarah dan perkembangan fenomena ini, seperti yang disebut di atas, melalui percobaan Milgram, sampai kepada percobaan Kleinberg. Bagian kedua adalah aplikasi praktis dan langsung dari fenomena ini, yaitu pencarian algoritma heuristik yang memperbaharui algoritma yang lama dalam system *peer to peer information sharing*. Dan

bagian terakhir diluar kedua bagian tersebut adalah berupa kesimpulan.

II. SMALL-WORLD PHENOMENA

II.1 Sejarah

Small-world Phenomenon atau dalam bahasa disebut Fenomen Dunia Kecil, merupakan fenomena yang menyangkut graf didalamnya. Fenomena ini pertama sekali dikemukakan oleh Karinthy pada tahun 1920-an, lalu dirumuskan kembali oleh Stanley Milgram pada tahun 1967 melalui sebuah eksperimen.

Milgram melakukan eksperimen berupa mengirimkan sebanyak 300 surat dari orang-orang yang berbeda kepada satu nama yang terletak di Boston. Surat-surat tersebut berasal pertama kali dari Nebraska dan Kansas. Dalam surat tersebut berisi instruksi untuk melanjutkan surat tersebut kepada rekan atau relasinya yang dia kenal berdasarkan *first-name basis*. Pada saat pertama kali sang penerima surat mendapat surat, dia mendapatkan informasi berupa alamat dan pekerjaan dari orang yang akan dikirimkan suratnya. Dari 300 surat yang dikirim dalam percobaan ini, yang membentuk rantai terdapat 64 surat yang mencapai target. Dan dari ke-64 surat yang mencapai target tersebut rata-rata memiliki rantai sebanyak enam kali pengiriman kembali. Dalam peta di bawah ini dapat dilihat perjalanan dari surat tersebut.



Gambar 1. Peta mengenai perjalanan surat. Daerah merah merupakan Nebraska dan Kansas, dan daerah hijau merupakan Boston.

Percobaan ini membuat hipotesis “enam-tingkat permissahan” menjadi terkenal.

Dalam perkembangan selanjutnya fenomena ini langsung menggelitik pikiran berbagai kalangan mulai dari aktor film sampai matematikawan.

Pertama kali diterapkan oleh para aktor film dan matematikawan melalui Kevin Bacon dan Paul Erdos. Kevin Bacon mengatakan bahwa setiap aktor memiliki nilai yang dapat dihitung (finite number) Bacon Number, yang merupakan angka yang mewakili seseorang aktor dengan aktor lain pernah bermain di suatu film bersama-sama. Dan setelah melalui percobaan didapatkan rata-rata dari Angka Bacon (Bacon Number) untuk dirinya sendiri adalah 2.943 dan angka maksimalnya adalah 8. Sedangkan Paul Erdos juga memiliki Erdos Number, yaitu hubungan antara para matematikawan yang mewakili kerjasama kedua orang tersebut dalam penerbitan suatu makalah. Paul Erdos memiliki rata-rata Angka Erdos kurang dari 5 dan nilai maksimalnya adalah 5 dan nilai kebanyakan adalah kurang dari 8. Untuk hubungan antara aktor dan aktor, Aktor Rod Steiger merupakan aktor yang paling "terhubung dengan aktor lain" dengan angka Bacon Numbrernya mencapai 2.6666 . Sedangkan angka Bacon untuk Paul Erdos adalah empat, karena dia pernah terlibat dalam film *N is a Number* produksi tahun 1993.

Fenomena ini juga membuat Duncan J Watt dan Steven H Strogatz mencoba untuk memodelkan fenomena ini. Dan hasil yang mereka dapatkan adalah jaringan neuron cacing seperti *C.elegans* merupakan representasi yang baik dari fenomena small-world, yang akan dibahas lebih lanjut.

Sampai 30 tahun setelah pertama kali dikemukakan oleh S.Milgram, masalah ini masih merupakan pertanyaan yang menarik bagi semua matematikawan.

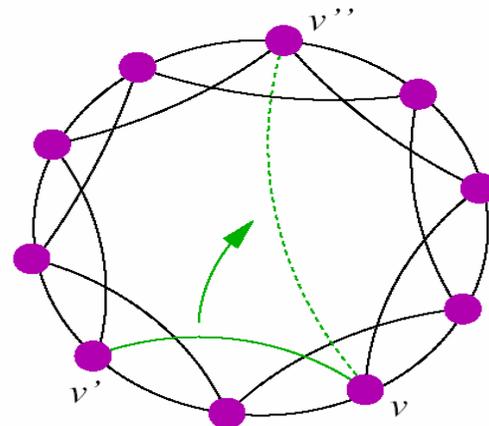
II.2 Model Watt dan Strogatz

Watt dan Strogatz mengajukan sebuah model untuk fenomena ini yang didasarkan pada suatu kelas dari jaringan acak yang menggabungkan kedua keadaan yang sangat bertolak belakang, dimana sisi dari jaringan model ini dibagi menjadi hubungan lokal dan jarak-jauh. Contoh yang dapat dipakai adalah, sebuah "lingkaran kisi-kisi kawat", yang dapat digambarkan sebagai berikut. Pertama kali dimulai dengan himpunan V yang memiliki n buah titik yang berlokasi secara seragam di dalam sebuah lingkaran, kemudian gabungkan setiap titik dengan K tetangga, dengan K adalah nilai konstan kecil. Himpunan inilah yang disebut hubungan lokal dalam sebuah jaringan. Kemudian dari sebanyak K tetangga yang ada ditambahkan pula dalam jumlah kecil tetangga

dari sebuah simpul yang dipilih secara acak dari himpunan V yang dapat disebut sebagai hubungan jarak-jauh. Dari model seperti ini, Watt dan Strogatz berpendapat model mereka menangkap dua parameter yang amat krusial dari jaringan sosial : pertama, ada suatu struktur dasar yang sederhana yang menjelaskan kebanyakan sisi (sesuai dengan model acak-seragam) dan kedua, ada sisi-sisi yang dihasilkan dari proses acak yang tidak mengikuti struktur acak-seragam.

Mereka menunjukkan bahwa model yang mereka punya merupakan representasi yang baik dari banyak jaringan yang ada di alam seperti jaringan saraf di cacing nematode *C.Elegans* dan jaringan *Power-Grid* di Western US dan sudah dipakai juga untuk analisa *graph hyperlink* untuk sebuah situs. Berikut adalah contoh dari model Watt dan Strogatz. Garis hitam menunjukkan hubungan lokal dan garis hijau putus-putus menunjukkan sebuah hubungan jarak-jauh.

Watts and Strogatz Model



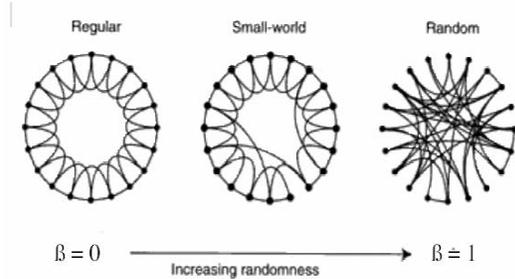
1-lattice with $k = 2$ being rewired.

Gambar 2. Model Watt-Strogatz dengan nilai $K = 2$.

Model Watt disebut juga dengan *B-graph* , dengan B sebagai lambang dari probabilitas sebuah simpul dapat memiliki tetangga lebih banyak dari nilai K secara acak, meningkatnya nilai B berarti meningkatnya keacakan, dan semakin tidak seragam suatu himpunan V . Nilai B bergerak dari 0 sampai 1. Jika nilai B adalah 0 disebut *regular graf*, nilai 1 disebut *random graf*, dan diantaranya disebut *small-world graf*. Jadi, mekanisme utama dalam membuat sebuah small-world graf adalah penambahan jalan-pintas atau

bisa juga disebut penambahan tetangga melebihi nilai K .

Model ini menjadi model pertama yang dimunculkan matematikawan dalam menjawab permasalahan fenomena ini. Model ini juga nantinya yang menjadi dasar dalam pemakaian fenomena ini berkembang pada pemakaian yang lebih aplikatif.



Gambar 3. B-graph dari 0 sampai 1

II.3 Fenomena Small-world dalam Perspektif Algoritma (Model Kleinberg)

Jon Kleinberg, seorang dosen dari universitas Cornell adalah orang yang cukup peduli tentang fenomena ini. Makalah dia mengenai masalah ini cukup banyak keluar. Dan salah satu isi dari makalah yang dirilis adalah mengenai pemodelan masalah small-world dan hubungannya dalam perspektif algoritma.

Sebenarnya terdapat dua pertanyaan utama yang harus dijawab dalam menjelaskan fenomena ini. Pertama, Mengapa bisa *terdapat* sebuah rantai relasi antara dua orang yang tidak saling kenal. Dan yang kedua, mengapa kedua orang tersebut dapat *menemukan* jalan atau rantai hubungan mereka yang terpendek. Untuk jawaban pertanyaan pertama adalah memang hubungan itu ada dan berada di alam, tidak ada jawaban yang lebih memuaskan selain itu. Dan untuk pertanyaan kedua jawabannya adalah adanya sebuah algoritma yang memungkinkan, hal inilah yang akan dibahas lebih lanjut.

Pada bagian ini, akan dijelaskan sebuah algoritma “decentralized” yang setiap individunya yang hanya mengetahui lokasi relasi dekatnya, berusaha mencari jalan terpendek untuk mencapai targetnya.

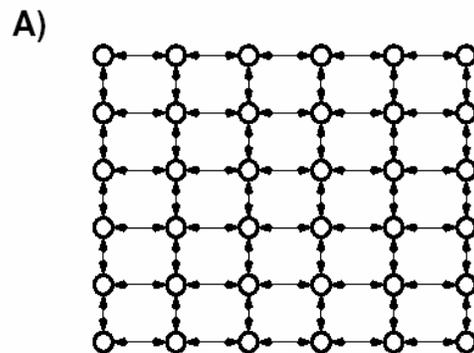
Model Watt dan Strogatz tidak dapat menjawab pertanyaan kedua ini. Dari model ini juga, tidak dapat diaplikasikan algoritma ”decentralized” yang akan membuat jalan tersingkat antara dua

simpul sekaligus menjawab pertanyaan kedua tersebut.

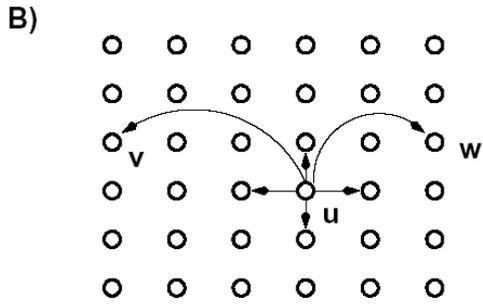
Sebelum membuat algoritmanya, akan dibuat terlebih dahulu model dari graph yang dapat dipakai oleh algoritma “decentralized”. Dalam membuat model dari graphnya, digunakan ruang kerja yang simple, yang mengambil dasarnya dari model Watt-Strogatz, model dengan banyak hubungan lokal dan beberapa hubungan jarak jauh. Akan tetapi, jika pada model Watt-Strogatz memakai bangun lingkaran, pada model algoritma ini yang dipakai adalah grid dua-dimensi yang sisinya berarah. Kemudian, kita dapat memulai dengan sebuah himpunan simpul yang diletakkan pada titik-titik sebuah kisi-kisi persegi berukuran $n \times n$, $\{(i,j): i \in \{1,2,\dots,n\}, j \in \{1,2,\dots,n\}\}$. Dan kita juga mendefinisikan jarak antar kisi adalah jarak antara dua titik (i,j) dan (k,l) disebut juga *langkah-kisi*, yang dirumuskan sebagai berikut;

$$d((i,j), (k,l)) = |i - j| + |k - l|$$

Untuk sebuah konstan $P \geq 1$, jarak kisi antara sebuah simpul u dengan simpul lain sehingga bernilai P , maka antara dua simpul tersebut disebut memiliki *hubungan lokal*. Dan, untuk $Q \geq 0$ dan $R \geq 0$, dapat juga dibentuk suatu hubungan jarak-jauh dari suatu simpul u ke Q banyaknya simpul lainnya dengan menggunakan percobaan *independent random*, besarnya kemungkinan terdapat sebuah sisi berarah dari u ke v mempunyai proporsi kemungkinan sampai $[d(u,v)]^{-P}$. (Untuk mendapatkan distribusi kemungkinannya, kita harus membagi nilainya dengan suatu konstanta penormalan $\sum_v [d(u,v)]^{-P}$, yang kita sebut distribusi invers r -th power). Contoh dari graph ini adalah sebagai berikut,



Gambar 4: (A) Sebuah jaringan grid dua-dimensi dengan $n = 6$, $p = 1$, dan $q = 0$



Gambar 5: (B) Keadaan simpul u dengan $p = 1$ dan $q = 2$. v dan w adalah dua hubungan jarak-jauh yang dimiliki u .

Sebuah analogi geometri sederhana untuk model diatas adalah, seseorang hidup di simpul-simpul graf di atas, dan tahu tetangganya berada beberapa langkah dari dirinya dalam berbagai arah, dia juga memiliki relasi-relasi yang tersebar secara luas di antara simpul-simpul lainnya tersebut.

Jika kita pandang P dan Q sebagai konstanta yang tak dapat diganti, dengan demikian dapat didapatkan sebuah himpunan satu-parameter dari model jaringan di atas dengan merubah nilai R . Jika $R = 0$, kita mendapatkan sebuah distribusi seragam untuk hubungan jarak-jauh, yang dipakai di dalam model Watt-Strogatz. Jadi dari model ini pun, kita bisa mendapatkan model Watt-Strogatz - suatu hubungan jarak-jauh dalam model Watt-Strogatz dipilih secara independen dari posisi mereka pada grid. Semakin besar nilai R , hubungan jarak-jauh suatu simpul akan semakin mengumpul atau berdekatan di sekitar grid tersebut. Dan R , kemudian akan menjadi alat ukur atau patokan seberapa luas jaringan suatu himpunan simpul tersebut

Komponen algoritma dari model ini kita dapatkan berdasarkan eksperimen Milgram. Kita mulai dari dua simpul yang berubah-ubah, sebut s dan t . Tujuannya adalah menyampaikan pesan dari s ke t dalam langkah yang sesedikit mungkin. Melalui algoritma “decentralized”, kita dapat memperoleh mekanisme dimana sebuah pesan disampaikan secara sekuensial dari pemegang pesan ke salah satu relasinya (lokal atau jarak-jauh) dengan hanya menggunakan informasi lokal Informasi lokal ialah :

1. Semua himpunan yang membentuk hubungan lokal

2. Lokasi t berada
3. Semua hubungan jarak-jauh dan lokasi tujuan yang berhubungan dengan tempat pesan berada sekarang

Dengan diberikan informasi ini simpul s harus dapat menentukan kemana dia akan mengirim pesannya. Waktu perkiraan dari algoritma desentralisasi ini adalah angka perkiraan banyaknya langkah yang dibuat untuk mengirim pesan tersebut melalui sebuah jaringan yang didapat berdasar distribusi invers r^{th} -power. Tentu saja, penggunaan informasi yang hanya berupa informasi lokal pada algoritma ini menjadi aspek yang sangat penting, karena bila sebuah simpul mengetahui semua hubungan lokal maupun hubungan jarak-jauh, akan sangat mudah mendapatkan jalan terpendek. Tetapi hal ini sangat susah terjadi di dunia sebenarnya.

Hasil dari algoritma ini, memperlihatkan struktur dari model jaringan sangat mempengaruhi pembentukan jalan tersingkat dari satu simpul ke simpul lain.

Ketika $R = 0$ – distribusi hubungan jarak-jauh seragam – hasil dari teori graf acak dapat dipakai yaitu dengan probabilitas yang tinggi terdapat jalan antara tiap simpul dengan panjang sebesar $\log n$. Sedangkan algoritma desentralisasi tidak dapat menyelesaikan permasalahan ini. Algoritma ini mempunyai teorema sebagai berikut,

Teorema 1. Jika terdapat konstanta α_0 , yang bergantung kepada P dan Q , dan tidak bergantung terhadap n , maka waktu perkiraan pengiriman adalah paling sedikit $\alpha_0 n^{2/3}$

Dengan meningkatnya nilai R , algoritma desentralisasi dapat mengambil banyak keuntungan dari struktur model hubungan jarak-jauh, diwaktu yang bersamaan, hubungan jarak-jauh menjadi semakin tidak berguna dalam memindahkan pesan untuk jarak yang jauh. Karena itu juga, nilai R dapat memberikan perkiraan waktu yang lebih akurat. Untuk $R = 2$, distribusi yang terjadi adalah invers-kuadrat, sehingga menghasilkan teorema berikut.

Teorema 2. Terdapat sebuah algoritma desentralisasi A dan konstanta α_2 , yang independent terhadap n , sehingga ketika $R = 2$ dan $P = Q = 1$, waktu perkiraan maksimal dari A menjadi $\alpha_2 (\log n)^2$.

Kedua teorema diatas merefleksikan konsekuensi fundamental dari model yang dibuat. Ketika hubungan jarak-jauh dibentuk secara independen, rantai pendek akan terbentuk, tetapi simpul, yang bekerja di level lokal, tidak akan menghasilkan rantai itu. Ketika hubungan jarak-jauh terbentuk dari proses yang berhubungan dengan geometri dari grid dengan jalan yang spesifik, rantai pendek akan tetap terbentuk dan simpul yang bekerja dengan pengetahuan lokal dapat membentuknya juga.

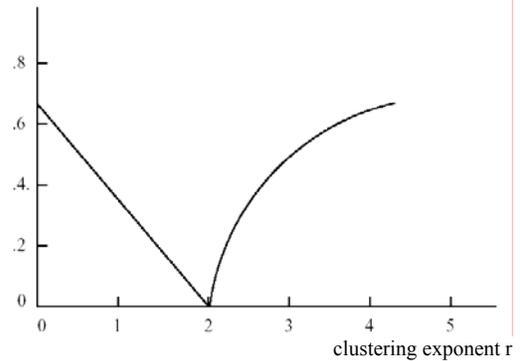
Pembuktiannya seperti ini, Algoritma desentralisasi A mengikuti aturan yang simple sebagai berikut : di setiap langkah yang dilakukan, simpul pemegang pesan akan mengirimkan pesan kepada simpul yang terdekat dengan simpul tujuan, yang berdasarkan jarak kisi yang tersisa. Untuk tambahan, algoritma ini tidak memerlukan himpunan yang telah dilalui simpul pesan sebelumnya. Untuk menganalisa apa yang dilakukan algoritma ini, kita katakan pesan terletak di dalam *fasa j* jika jarak antara simpul pembawa pesan terakhir dan simpul tujuan terletak antara 2^j dan 2^{j+1} . Dapat ditunjukkan, dalam *fasa j*, Waktu perkiraan sebelum simpul pembawa pesan terakhir memiliki jarak dengan simpul tujuan dalam hubungan jarak-jauh sebesar 2^j , akan meloncat sebesar $\log n$. Pada titik itu, *fasa j* akan menjadi fase terakhir. Selama langkah terbanyak yang dilakukan oleh algoritma ini adalah $1+\log n$, akan terjadi batas proporsional sampai $(\log n)^2$. Ketidakmungkinan hasil dari teorema 1 didasarkan pada fakta bahwa distribusi seragam menghalangi algoritma ini untuk memakai petunjuk yang disediakan oleh grid. Secara kasar, dapat kita katakan semua simpul dalam himpunan U terletak dengan jarak $n^{2/3}$ dari simpul tujuan. Dengan probabilitas yang tinggi, simpul sumber akan terletak diluar himpunan U , dan bila pesan tidak pernah melewati hubungan jarak-jauh menuju U , langkah yang diperlukan untuk mencapai simpul tujuan paling tidak adalah $n^{2/3}$. Akan tetapi bila ada hubungan jarak-jauh yang menghubungkan antara simpul sumber dan himpunan U , maka jarak yang ditempuh menjadi $n^{2/3}$, sehingga kita dapat menyebut langkah maksimal untuk mencapai simpul tujuan adalah $n^{2/3}$.

Secara lebih umum dapat ditunjukkan sebuah karakteristik kuat dari teorema berikut untuk keluarga model jaringan berikut : $R = 2$ adalah satu-satunya nilai yang memungkinkan algoritma desentralisasi membuat rantai dengan

panjangnya berupa fungsi polynomial dalam $\log n$.

Teorema 3. (a) untuk $0 \leq r < 2$, terdapat α_r , yang bergantung terhadap P, Q, R , tetapi tidak bergantung terhadap n , sehingga memiliki waktu perkiraan pengiriman minimal $\alpha_r n^{(2-r)/3}$.
 (b) untuk $r > 2$, terdapat α_r , yang bergantung terhadap P, Q, R , tetapi tidak bergantung terhadap n , sehingga memiliki waktu perkiraan pengiriman minimal $\alpha_r n^{(r-2)/(r-1)}$.

lower bound T on delivery time (given as $\log T$) n



Gambar 6: Batas bawah yang diimplikasikan oleh teorema 3. Sumbu x nilai r; sumbu y hasil ekponen r

Meskipun pembahasan mengenai algoritma pencarian jalan terpendek ini terfokus hanya pada dimensi-2, sebenarnya teorema ini berlaku juga untuk dimensi-k. Untuk dimensi-k, algoritma desentralisasi dapat membentuk jalan dalam bentuk polynomial $\log n$ jika dan hanya jika $R = K$. Untuk bukti dari teorema di atas, tidak diberikan pada makalah ini.

Algoritma yang dapat bekerja pada situasi yang berbeda telah mempertimbangkan masalah pencarian jalan dengan informasi lokal, dapat diambil sebagai contoh adalah problem dalam membuat tabel pencari jejak yang ringkas bagi komunikasi jaringan ataupun problem dalam membuat navigasi robot di daerah yang tidak dia kenal. Memang, algoritma ini jauh berbeda dengan penyelesaian masalah-masalah di atas. Tetapi masih ada persamaan dalam tujuan umum dalam mengidentifikasi komponen jaringan yang kualitatif yang dapat membuat pencari jejak memakai informasi lokal mudah dilakukan dan menghasilkan sebuah model yang efektif dalam

mencari jejak. Kesimpulan yang lain yang dapat diambil dari algoritma ini adalah, Hubungan antara struktur lokal dan hubungan jarak-jauh memberikan petunjuk utama untuk mencari jalan dalam sebuah jaringan. Ketika hubungan ini terletak dekat dengan batas kritis, struktur dari hubungan jarak-jauh akan membentuk sejenis gradient yang akan menuntun pesan secara efisien menuju targetnya. Ketika hubungan terletak di bawah nilai kritis dan jaringan sosial menjadi homogen, petunjuk tersebut akan hilang dan model algoritma ini tidak akan menemukan jalan tersebut.

III Fenomena Small-world dalam Sistem Penyebaran Informasi Pure P2P

Arsitektur Pure P2P telah menjadi model yang penting untuk pertukaran informasi diantara kelompok pemakai yang dinamis, dengan biayanya yang rendah dan merupakan model yang sesuai dengan besar kelompok tersebut. Perkembangan terkini dari system ini telah memunculkan pertanyaan baru dan tantangan baru. Dengan mengidentifikasi dua factor kunci di sebuah lingkungan, dapat dibuat sebuah algoritma baru pencarian heuristic yang akan membuat penggunaan fenomena small-world lebih baik yang bertujuan mencari “enam-tingkat pemisahan” lebih baik. Dengan algoritma baru ini, ada peningkatan sekitar sepuluh persen dibandingkan algoritma BFS untuk *querying information related*. Dan meningkat 20 persen ketika berhubungan dengan *a shift of interest*. Algoritma heuristic juga memiliki kontrol yang lebih baik terhadap jumlah simpul yang harus dikunjungi menggunakan fitur Penghitung-Count dibandingkan mesin TTL.

System P2P yang dipunyai sekarang, memiliki dua batasan :pertama, hanya mendukung pencarian dengan sebuah pengenalan atau katakunci. Di dalam system P2P file-sharing yang terstruktur file atau lokasi informasi disimpan berdasarkan pengenalan *hashed* dan dikembalikan dengan pengenalan itu menggunakan algoritma pencari jejak DHT yang berbeda. Dan didalam P2P yang tidak terstruktur seperti P2P file-sharing system, baik global index maupun skema global digunakan untuk melaksanakan fasilitas searching. Dan kedua, P2P file-sharing system hanya memiliki topic atau tipe data yang terbatas. Seperti Napster, Kazaa, dan Gnutella file yang dibagikan kebanyakan file musik dan film. Sehingga, hanya sedikit orang yang menggunakan P2P file-sharing system

dibandingkan pengguna mesin-pencari seperti Google ataupun Yahoo. Karena mereka menyediakan lebih banyak topic yang dapat dibagikan. System P2P information-sharing akan memberikan kesempatan untuk menerapkan algoritma baru yang akan mengeksplorasi fenomena small-world lebih dalam.

Beberapa teknik yang diaplikasikan pada system P2P yang sekarang ada memang telah menggunakan fenomena ini, seperti di system yang dimiliki oleh Gnutella yang memilih tetangga yang mengembalikan hasil terbesar, akan diasumsikan memiliki kedekatan dengan kata kunci yang dicari, sehingga mengefektifkan pencarian. Atau, di dalam jaringan yang memakai distribusi *power-law*, dengan memilih tetangga yang memiliki derajat tertinggi. Akan tetapi, dengan pencarian yang diterapkan oleh kebanyakan system P2P file-sharing saat ini, yang hanya mengidentifikasi kata kunci atau pengidentifikasi, tidak lah cukup untuk menetapkan dimana seharusnya mencari pesan dalam sebuah system P2P information-sharing. Sehingga perlu dibuat suatu algoritma yang memenuhi semua keinginan yang disebut diatas.

III.1 Deskripsi Persoalan

Dua factor utama di dalam P2P file-sharing system telah mendorong pembuatan algoritma yang baru. Yaitu, peningkatan yang dramatis tentang topik-topik yang berkembang saat ini berbanding dengan terbatasnya minat para pemakai, dan tingkat pergeseran minat pemakai.

a. Peningkatan Topik

Perbedaan utama antara system tradisional P2P file-sharing dengan sistem P2P information-sharing adalah peningkatan topik yang ada. Hal ini nantinya akan menimbulkan masalah dalam pencarian.

Sebagai contoh, perbedaan pada topik film dan musik biasanya dibatasi dan karenanya dapat diidentifikasi hanya dengan skema global seperti pencipta, penyanyi, dan sebagainya. Akan tetapi, banyak hal lain, yang tidak dapat diselesaikan hanya dengan melihat global skemanya

Akan tetapi, walaupun topik yang ada semakin banyak dan terus berkembang, minat dari pemakai biasanya terbatas. Sebagai contoh, pada Google newsgroup. Menurut penelitian yang dilakukan, hanya terdapat lima minat atau jenis berita dalam 800 newsgroup.

b. Pergeseran Minat Pemakai

Pergeseran minat pemakai menjadi hal yang sangat penting, apalagi saat era informasi yang dinamis saat-saat ini. Di dalam P2P information-sharing, orang akan dengan sangat mudah merubah minat dan kegemaran mereka. Dua jenis perubahan minat dapat diidentifikasi sebagai berikut :

1. Perubahan minat dalam waktu lama
Setelah berubah, orang ini akan tetap mendalami topik yang baru ini
2. Perubahan minat dalam waktu singkat
Setelah berubah, biasanya hanya sebentar mendalaminya, lalu akan berubah lagi

Kedua jenis perubahan ini akan menimbulkan dua kebutuhan dalam algoritma. Yang pertama, membutuhkan algoritma yang dapat mengeluarkan semua topik dalam minat yang baru. Dan yang kedua, hanya membutuhkan keefisienan dalam mencari, karena akan mudah berpindah.

III.2 Algoritma

III.2.1 Algoritma Heuristik

Dalam algoritma heuristik yang dipakai adalah penggabungan dua algoritma yang telah dipakai pada system terdahulu yaitu Breadth-first Traversal (BFS) dan Deep-first Traversal (DFS).

Ide utama dalam pembuatan algoritma ini adalah mengambil tingkat kesamaan yang akan merefleksikan kesamaan antara dua kata kunci, sehingga dapat dihasilkan pencarian yang efektif. Ketika tingkatnya rendah, menghasilkan tingkat kemungkinan yang rendah pada titik tersebut, sehingga tetangga yang memiliki kata kunci yang berkaitan atau berhubungan dengan kata kunci yang dicari juga berjumlah sedikit . Sehingga, pesannya harus dikirim kepada jumlah tetangga yang sedikit itu Ketika tingkatnya tinggi, kemungkinan titik itu dan probabilitas tetanggannya memiliki kata kunci yang berkaitan tinggi pula, sehingga tetangga yang harus dikunjungi pun harus semakin banyak. Sehingga pesan akan dapat sampai dengan efektif. Berikut *pseudo-codenya*.

Table 1. Symbols used in our heuristic algorithm

Symbol	Meaning
T_{low}	Lower threshold of result rankings
T_{high}	Higher threshold of result rankings
N_{result}	Number of results to be considered
$M_{neighbor}$	Number of neighbors to send query
$Query$	Current Query
$Resultlist$	Linked List of query results
$Peer$	Peer being queried
$Similarity$	Value based on the Resultlist

```

Procedure 1 Peer.Main()
1: Query=Peer.GetQuery()
2: Resultlist=Peer.ProcessQuery(Query)
3: Resultlist.Sort()
4: Similarity=Peer.EvalSimilarity(Resultlist)
5: if Similarity>Thigh then
6:   Mneighbor=Peer.GetNumOfNeighbors()
7: else if Similarity<Tlow then
8:   Mneighbor=1
9:   else
10:    Mneighbor=Peer.GetNumOfNeighbors()*
      (Similarity-Tlow)/(Thigh-Tlow)
11:   end if
12: end if
13: if Mneighbor=0 then
14:   Mneighbor=1
15: end if
16: Send Query to Mneighbor neighbors
17: return
  
```

```

Procedure 2 Peer.EvalSimilarity()
1: count=0
2: totalrank=0
3: while not Resultlist.EOF() do
4:   result=Resultlist.GetCurrentResult()
5:   if count>Nresult then
6:     break
7:   else
8:     count=count+1
9:     totalrank=totalrank+result.GetRank()
10:    Resultlist.NextResult()
11:   end if
12: return totalrank/Nresult
  
```

```

Procedure 3 Peer.GetNumOfNeighbors()
Return the number of the peer's neighbors
  
```

```

Procedure 4 Peer.GetQuery()
Get the query the peer just received
  
```

```

Procedure 5 Peer.ProcessQuery(Query)
Process the query and return a linked list of results with rankings
  
```

Gambar 7. Tabel dan pseudo-code algoritma heuristik

III.2.2 Fitur Kontrol Node-to-Visit

Masalah utama penggunaan Time-to-live (TTL) adalah banyaknya simpul yang harus dikunjungi bergerak secara eksponensial. Karena tidak dapat diketahui banyaknya tetangga sebelumnya, sehingga banyaknya simpul yang dikunjungi dapat diluar kontrol.

Pada algoritma ini digunakan fungsi *NodeCount*, untuk mengetahui jumlah simpul yang dikunjungi. Terdapat empat langkah yang harus diikuti.

Untuk Simpul yang Dikunjungi :

1. Jika, simpul telah dikunjungi sebelumnya, kembalikan *NodeCount* ke simpul pengirim, dan terminasi fungsinya, jika tidak lakukan proses dua.
2. Kurangi *NodeCount*. Jika *NodeCount* mencapai nol, lakukan proses tiga. Jika tidak, Bagi *NodeCount* ke $M_{neighbour}$ bagian dan kirim *NodeCount* yang baru. Terminasi Algoritma.
3. Daftarkan Node ID ke simpul pengirim untuk *update*. Terminasi Algoritma

Untuk Simpul yang Mengirim

4. Jika terdapat, ekstra *feedback* dan Simpul yang terregister, distribusikan *NodeCount* ke simpul terregister dan mulai langkah dua. Lakukan terus sampai di terminasi oleh program

Jelas, algoritma ini lebih baik dari pada algoritma TTL. Tanpa mengetahui jumlah tetangga setiap simpul dan model dari jaringannya, tetap dapat mengontrol simpul-yang-dikunjungi.

III.3 Hasil Pengujian

Setelah dilakukan percobaan dengan algoritma ini dapat dihasilkan grafik sebagai berikut,

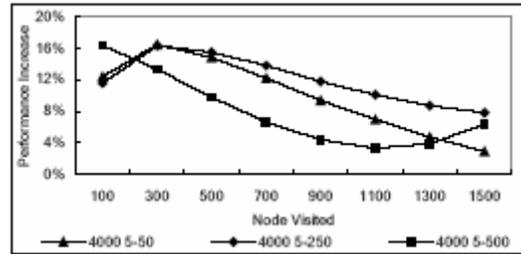
Tabel berikut menggambarkan peningkatan presentase mendapatkan topik yang relevan yang ada hubungannya antara dua topik yang berbeda sama sekali.

	10%-30%	30%-50%	10%-50%
4000 5-50	15.86%	10.79%	13.04%
4000 5-250	14.77%	12.78%	14.17%
4000 5-500	14.89%	9.13%	10.76%
Average	15.17%	10.90%	12.66%
4000 5-50 S	15.06%	9.96%	12.22%
4000 5-250 S	22.77%	17.55%	19.87%
4000 5-500 S	29.67%	22.87%	26.27%
Average S	22.50%	16.79%	19.45%

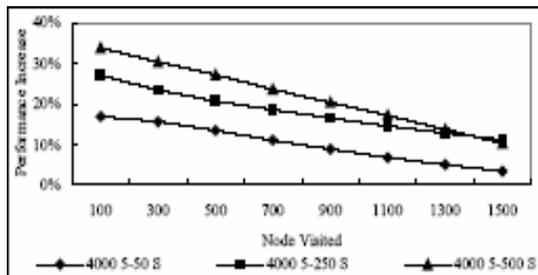
Table 1. Kenaikan berdasarkan kepada persentase mendapatkan dokumen yang relevan untuk jumlah total topic yang tersedia berbeda-beda

dan berikut ini adalah grafik yang memperlihatkan kenaikan performansi. Untuk gambar 8, memperlihatkan grafik kenaikan untuk tidak adanya pergeseran minat, sedangkan gambar 9 memperlihatkan grafik peningkatan bila adanya pergeseran minat. Di dalam gambar 8 dan 9 dipakai tiga jenis jumlah total topic (50 250 500) yang disimulasikan ke lebih dari 4000 *peers* dan tanpa pergeseran minat. Sangat jelas terlihat algoritma yang baru ini lebih unggul dibandingkan algoritma BFS pada 1500 simpul pertama yang dikunjungi, dengan kenaikan 10 %

pada *Related Query* dan kenaikan 17 % pada *Non-Related Query*.



Gambar 8 Kenaikan performansi dengan TIDAK adanya pergeseran minat. Total topic yang dihasilkan berkisar antara 50 sampai 500.



Gambar 9. Kenaikan performansi dengan adanya pergeseran minat. Total topic yang dihasilkan berkisar antara 50 sampai 500.

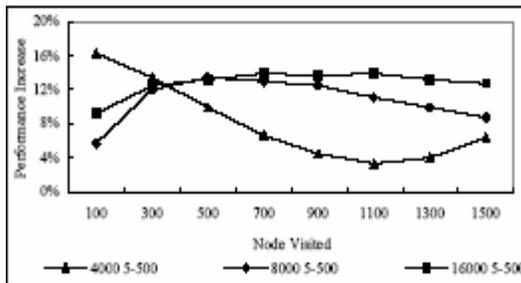
Rata-rata kenaikan performansi yang terjadi oleh pemakaian algoritma heuristik ditunjukkan oleh table 3. Kebanyakan pengguna hanya membutuhkan sedikit dari total dokumen yang disediakan. 10-30 % dari jumlah semua total dokumen sudah cukup untuk memenuhi kebutuhan pengguna. Dan heuristik algoritma juga mengalami kenaikan sangat tinggi dalam performansi, 15% untuk *related-query* dan 23% untuk *non-related query*. Dengan penggunaan fenomena small-world yang lebih baik akan membuat kinerja algoritma ini lebih baik untuk semua kondisi.

Di tabel dan grafik sebelumnya dapat ditemukan indikasi keefektifan algoritma ini dengan meningkatnya jumlah topik yang dapat disediakan dan pergeseran minat pemakai. Akan tetapi, pengujian yang menghasilkan tabel dan grafik di atas berjalan dengan hanya 4000 *peers* dan pada tabel dan grafik di bawah ini dapat ditunjukkan algoritma heuristik ini dapat diperbesar skala penggunaannya. Dalam gambar 10 dan gambar 11, digambarkan grafik untuk tiga jenis skala yang berbeda (4000 8000 16000) dan rata-rata minat adalah lima. Untuk 1500 simpul pertama, didapatkan, rata-rata kenaikan

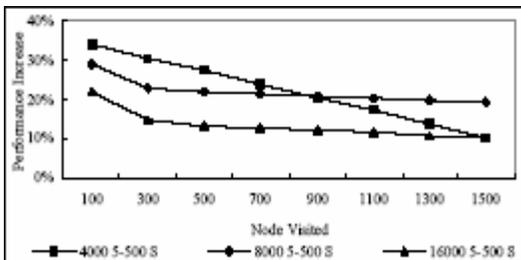
performansi sebesar 10 % untuk *Related Query* dan 18 % untuk *non-Related Query*

	10%-30%	30%-50%	10%-50%
4000 5-500	14.89%	9.13%	10.76%
8000 5-500	12.48%	9.68%	10.83%
16000 5-500	12.36%	8.37%	10.21%
Average	13.24%	9.06%	10.60%
4000 5-500S	29.67%	22.87%	26.27%
8000 5-500S	21.07%	18.63%	19.97%
16000 5-500S	10.09%	6.43%	8.13%
Average S	20.28%	15.97%	18.12%

Table 2. Kenaikan berdasarkan kepada persentase mendapatkan dokumen yang relevan untuk jumlah skala yang berbeda



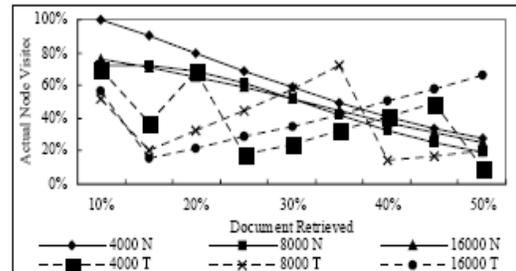
Gambar 10 Kenaikan performansi dengan TIDAK adanya pergeseran minat. Skala berkisar dari 4000 sampai 16000 peers



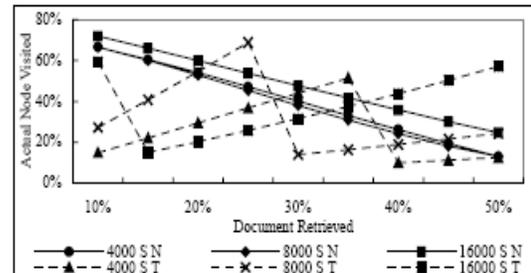
Gambar 11 Kenaikan performansi dengan adanya pergeseran minat. Skala berkisar dari 4000 sampai 16000 peers

Sekarang mengenai keefektifan kontrol. Semua hasil dari eksperimen yang didapatkan dan digambarkan pada gambar 12 dan 13 serta pada tabel 3 dilakukan menggunakan 500 topik dan rata-rata minat sebesar 5. N adalah *NodeCount* dan T untuk mekanisme *TTL*. Seperti yang terlihat pada gambar 12 dan 13. Mekanisme *TTL* memiliki lebih banyak bentuk rigid dibandingkan *NodeCount*, dimana hasil dari *TTL* mengendalikan jumlah simpul yang dikunjungi. Di lain pihak, keefektifan kontrol dari *NodeCount* menurun sejalan dengan meningkatnya persentase penerimaan dokumen yang relevan. Akan tetapi, seperti yang sudah disebutkan di atas, hanya sekitar 10-30% saja

dari total topik sudah dapat mencukupi kebutuhan dari user. Dan, untuk tingkat tersebut mekanisme *NodeCount* berhasil mengunjungi 70 % untuk simpul yang *RelatedQuery* dan 55% untuk *Non-Related Query*, berbanding dengan 39% dan 34% dengan mekanisme *TTL* tradisional.



Gambar 12. Keefektifan kontrol dalam penerimaan dokumen yang relevan dengan TIDAK ada pergeseran minat



Gambar 13. Keefektifan kontrol dalam penerimaan dokumen yang relevan dengan ada pergeseran minat

	16000	8000	4000	Average
10%-30% N	64.68%	65.26%	79.32%	69.75%
10%-30% T	31.65%	41.15%	43.37%	38.72%
30%-50% N	38.04%	34.34%	42.02%	38.13%
30%-50% T	50.53%	36.35%	31.41%	39.43%
10%-50% N	51.34%	49.52%	60.89%	53.92%
10%-50% T	41.73%	36.63%	38.82%	39.06%

	16000 S	8000 S	4000 S	Average S
10%-30% N	59.78%	52.56%	53.62%	55.32%
10%-30% T	30.38%	40.99%	29.39%	33.59%
30%-50% N	35.90%	24.78%	26.36%	29.01%
30%-50% T	44.07%	18.95%	25.78%	29.60%
10%-50% N	47.87%	38.74%	39.98%	42.20%
10%-50% T	37.86%	31.76%	25.76%	31.79%

Table 3. Detail Kefektifan kontrol .

III.4 Kesimpulan

Kebingungan dalam pencarian di system P2P telah menyelesaikan banyak masalah. Dengan system yang baru ini, pencarian informasi di P2P system akan menjadi lebih baik. Dari apa yang dipaparkan di atas, setelah mengidentifikasi dua factor utama dalam P2P file-sharing yang menjadi kendala, lalu menunjukkan algoritma pencarian yang lebih baik, dapat ditunjukkan algoritma ini memakai fenomena small-world dengan lebih baik.

kenaikan lebih dari 10 % untuk *Related-Query* dan 20 % ketika berhubungan dengan pergeseran minat. Dan *NodeCount* terlihat lebih baik performansinya dibandingkan mekanisme *TTL*.

IV. KESIMPULAN

Kesimpulan yang dapat diambil dari fenomena *small-world* dan aplikasinya dalam system penyebaran informasi *pure P2P* :

1. Fenomena Small-World adalah fenomena yang menyebutkan bahwa semua orang di dunia ini dapat terhubung dengan orang lain di dunia ini yang sama sekali tidak dikenalnya hanya dalam sebuah rantai kenalan sosial yang pendek.
2. Model Watt-Strogatz merupakan model pertama dari fenomena small-world yang menjadi dasar perkembangan fenomena ini dalam penggunaan yang lebih aplikatif.
3. Algoritma desentralisasi dapat diterapkan dalam pencarian jalan terpendek antara dua simpul dengan memakai model small-world yang dikembangkan oleh Kleinberg dengan mekanisme yang hanya menggunakan informasi lokal yang dimiliki oleh simpul. Algoritma ini bekerja untuk semua dimensi pekerjaan.
4. Algoritma desentralisasi memperlihatkan struktur dari model jaringan sangat mempengaruhi pembentukan jalan tersingkat dari satu simpul ke simpul lain.
5. Dua faktor utama di dalam P2P file-sharing system telah mendorong pembuatan algoritma yang baru. Yaitu, peningkatan yang dramatis tentang topik-topik yang berkembang saat ini berbanding dengan terbatasnya minat para pemakai, dan tingkat pergeseran minat pemakai.
6. Melalui pengujian, algoritma heuristic yang baru dapat menggunakan fenomena small-world dengan lebih baik. Dengan

DAFTAR PUSTAKA

- [1] *Characteristics of Small World Networks* (2006) <http://www.tp.umu.se/~kim/Network/>
Tanggal akses : 26 Desember 2006 pukul 20:21

- [2] Explore the “Small World Phenomena” in Pure P2P Information Sharing System (2006)
http://xena1.ddns.comp.nus.edu.sg/p2p/yr_ccgrid03.pdf
Tanggal akses : 26 Desember 2006 pukul 20:21

- [3] Munir, Rinaldi. (2004).Diktat Kuliah IF2151 Matematika Diskrit, ed 4. Departemen Teknik Informatika, Institut Teknologi Bandung.

- [4] The Small-World Phenomenon: An Algorithmic Perspective (2006)
<http://www.cs.cornell.edu/home/kleinber/swn> Tanggal akses : 26 Desember 2006 pukul 20:21

- [5] The Small World Phenomenon (2006).
<http://www.wikipedia.org> Tanggal akses : 26 Desember 2006. pukul 20:21