

**Analisis Keahlian dalam Model Jejaring Koneksi LinkedIn
(Studi pada Mahasiswa Ilmu Komputer Angkatan 2016 Fakultas Ilmu Komputer Universitas
Indonesia)**

Wisnu Pramadhitya Ramadhan (1606918055)

Fakultas Ilmu Komputer

Universitas Indonesia

Depok, Indonesia

1. PENDAHULUAN

LinkedIn merupakan media sosial yang menghubungkan para profesional dan bisnis. Pengguna menggunakan LinkedIn untuk membuat informasi profesional, seperti jabatan, pendidikan, keahlian, pencapaian dan lainnya sehingga dapat dibentuk menjadi sebuah *resume*. Hal ini bermanfaat menjadi sarana publikasi diri secara profesional. Selain itu, LinkedIn dapat digunakan untuk membuat koneksi sesama pengguna di dalamnya dan melakukan *recruitment* atau melamar pekerjaan.

LinkedIn menjadi tempat yang tepat di dunia maya untuk mengembangkan karir dan memperluas jaringan perkenalan profesional. Berdasarkan ide tersebut, muncul sebuah kesempatan untuk dilakukan analisis lebih mendalam mengenai jejaring profesional dan perkenalan. Oleh karenanya, Saya tertarik untuk membuat sebuah model jejaring koneksi LinkedIn untuk dianalisis.

Analisis yang dilakukan berupa seberapa penting seseorang dalam jejaring koneksi LinkedIn dilihat dari kesamaan keahlian orang tersebut terhadap koneksinya. Analisis ini akan melihat apakah seseorang yang menjadi sosok penting memiliki keahlian atau minat yang mirip dengan koneksinya. Hal ini akan memperlihatkan bagaimana suatu koneksi atau pertemanan terbentuk atas suatu kesamaan. Dalam hal ini, kesamaan tersebut adalah keahlian profesional.

Selain topik sosok penting, Saya juga tertarik dalam menganalisis apakah dua atau lebih orang yang belum terhubung dan memiliki keahlian yang sama akan terhubung di masa mendatang. Dua orang akan memiliki interaksi yang lebih kuat ketika memiliki kesamaan dalam dasar pemahaman (Z. Xiao, 2018). Contohnya, seseorang akan lebih suka membahas suatu masalah dengan teman yang ikut kelas khusus yang sama dengannya. Hal ini menjadi landasan bahwa sebuah pertemanan dapat terbentuk atas kesamaan minat. Pada dasarnya pertemanan atau koneksi bisa terbentuk ketika terdapat interaksi di antaranya (K. Campbell, 2015).

2. LITERATUR

2.1. KONEKSI DAN KESAMAAN MINAT

Terdapat setidaknya lima komponen yang dapat membentuk sebuah ikatan pertemanan. Kesamaan dalam minat dan bakat merupakan salah satu komponen yang signifikan untuk membentuk sebuah ikatan pertemanan (K. Campbell, 2015). Di tingkat komunitas pelajar, banyak pelajar yang ingin berteman yang berbagi minat yang sama (Z. Xiao, 2018). Hal ini disebabkan mereka dapat berbagi topik permasalahan yang sama, keterbukaan, memiliki kecocokan dan pribadi yang perhatian dapat membentuk ikatan, meningkatkan keharmonisan dan kekuatan koneksi antar anggota di sebuah komunitas (Z. Xiao, 2018; K. Campbell, 2015).

Penelitian (Z. Xiao, 2018) memiliki pendekatan dalam menentukan sebuah komunitas memiliki minat yang sama adalah dengan deteksi komunitas dan *clustering coefficient*. Penelitian tersebut menyarankan bahwa pembagian kelompok tempat tinggal asrama dapat dibagi atas minat anggotanya (Z. Xiao, 2018). Penelitian ini menunjukkan bahwa asrama yang anggotanya tidak memiliki minat yang sama memiliki kesusahan dalam mencari topik berkomunikasi sehingga ikatan antar anggota tidak erat (Z. Xiao, 2018).

2.2. SOSOK PENTING

Sudah banyak pendekatan yang ada untuk mencari sosok penting dalam sebuah jejaring. Metode yang sering dilakukan adalah pendekatan *centrality*. *Centrality* merupakan gambaran akan bentuk kepentingan atau keunggulan (S. P. Borgatti, 1995). Freeman melahirkan tiga pendekatan, namun dalam penelitian ini hanya menggunakan *Betweenness* dan *Closeness* (M. Kaur and S. Singh, 2017). *Closeness* adalah jumlah dari jarak dari semua node, yaitu jarak dari satu node ke yang lain adalah panjang dari jalur terpendek dari satu ke yang lain – semakin sentral *node-node* tersebut, maka semakin cepat mereka dapat menggapai *node-node* lain (S. P. Borgatti, 2005). Berbeda dengan *betweenness*, dimana sosok penting dilihat dari seberapa penting suatu *nodes* mencapai *nodes* lainnya melalui jalur terpendek (S. P. Borgatti, 2005).

Pendekatan *centrality* populer lainnya adalah *Eigenvector centrality* (Bonacich, P., 1972). *Eigenvector centrality* memberikan pandangan baru bahwa *node* penting adalah yang terhubung pada *node* lain yang memiliki pengaruh penting juga (S. P. Borgatti, 2005). Sehingga pada kasus ini, kemungkinan mahasiswa menjadi sosok paling sentral adalah mahasiswa yang memiliki koneksi dengan mahasiswa-mahasiswa penting. *Eigenvector centrality* menjawab masalah bahwa memiliki banyak koneksi bukan berarti memastikan bahwa sosok tersebut paling penting dalam sebuah jejaring.

Sebagai pembanding, Saya menggunakan algoritma *PageRank* untuk menghitung sosok penting. *PageRank* pada dasarnya adalah algoritma untuk mencari *website* populer atau penting. Secara dasar, sebenarnya *PageRank* digunakan pada jejaring berarah (*directed graph*), namun pada beberapa kasus khusus *PageRank* digunakan pada jejaring tidak berarah (*undirected graph*). Namun, penghitungan tetap berdasarkan jejaring berarah. Pada jejaring tidak berarah, *PageRank* menghitung secara *bidirectional*.

2.3. PREDIKSI KONEKSI

Prediksi koneksi merupakan salah satu metode untuk memprediksi koneksi antara dua *nodes* di masa mendatang ketika belum adanya koneksi tersebut diantaranya (H. A. Deylami, 2015). Dengan metode ini, dapat diperkirakan kemungkinan adanya *edge* baru antara dua *nodes*. Beberapa pengukuran yang populer untuk melakukan prediksi koneksi adalah *Jaccard* dan *Preferential*.

3. METODOLOGI

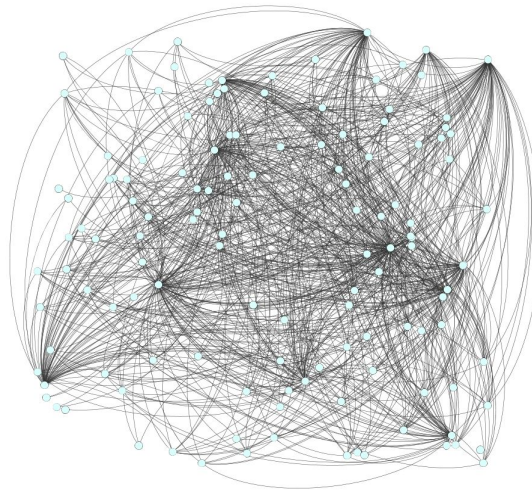
3.1. DATA JEJARING

Jejaring yang diambil merupakan suatu *undirected social network graph*. Jejaring menggambarkan koneksi antar mahasiswa-mahasiswa Ilmu Komputer angkatan 2016 Fakultas Ilmu Komputer Universitas Indonesia. Setiap mahasiswa memiliki ID khusus yang unique untuk identifikasi.



Gambar 1. Node dan edge jejaring koneksi LinkedIn

Pada jejaring koneksi LinkedIn, himpunan mahasiswa yang memiliki akun LinkedIn menjadi *node*. Sedangkan *edge* memiliki makna memiliki koneksi, sehingga *edge* merupakan himpunan koneksi antara dua pengguna LinkedIn. *Node* memiliki atribut keahlian, riwayat pekerjaan, riwayat pendidikan, tanggal lahir, nama depan, nama belakang, dan atribut pengguna lainnya. Tiap-tiap riwayat pekerjaan mengandung posisi pekerjaan, perusahaan, lokasi, dan rentang waktu bekerja. Sedangkan pendidikan memiliki atribut bidang studi, tempat bersekolah dan rentang waktu.



Gambar 2. Anggota dari mahasiswa Ilmu Komputer angkatan 2016

3.2. METODE

Untuk dapat menganalisis jejaring koneksi LinkedIn mahasiswa ini, sebuah pendekatan dibangun untuk mengidentifikasi sosok penting dan prediksi koneksi baru. Data yang digunakan merupakan populasi dari mahasiswa Ilmu Komputer angkatan 2016 Fakultas Ilmu Komputer Universitas Indonesia. Data mahasiswa didapatkan dari penghimpunan yang dilakukan oleh pengurus angkatan 2016. Dari jumlah populasi mahasiswa sebanyak 181 orang, ditemukan 130 mahasiswa mempunyai akun LinkedIn.

3.2.1. PENGUMPULAN DATA

Untuk merancang jejaring koneksi antar pengguna dalam LinkedIn, Saya melakukan *scraping* dan menggunakan perangkat lunak Gephi untuk visualisasi dan pengukuran. Kode sumber *scraping* dibuat dengan menggunakan bahasa pemrograman Python. *Scraping* dilakukan dengan memberikan URL profil target pengguna. Data pengguna yang dapat diambil berupa *user identifier*, data diri, keahlian, pekerjaan, pendidikan dan koneksi antar pengguna.

Syarat melakukan *scraping* harus masuk dan ter-*authorized*. Akun yang digunakan dapat sembarang akun, namun disarankan yang sudah memiliki jejaring koneksi yang luas dan banyak. Pada uji coba ini, tim menggunakan akun pribadi dan mengambil *cookies* dari *browser* untuk disisipkan ke dalam program.

Secara khusus untuk melakukan prediksi koneksi baru, maka menggunakan dua dataset. Dataset tersebut adalah *dataset* populasi yang digunakan di dalam Gephi dan *dataset actual*, yaitu dataset populasi yang sama namun hanya node-node yang memiliki keahlian yang sama yang terhubung yang sebelumnya belum terhubung.

3.2.2. PENGUKURAN

Perangkat lunak Gephi digunakan untuk melakukan visualisasi dan pengukuran. Data yang didapatkan pada langkah 1 diimpor ke Gephi. Untuk mendapatkan sosok penting atau sentral dalam model ini, Saya menggunakan pengukuran sentralitas menggunakan metode *betweenness*, *eigenvector*, dan *closeness centrality*. Selain itu, digunakan pengukuran dengan algoritma *PageRank* untuk membandingkan hasil pengukuran sentralitas.

Untuk mendapatkan prediksi koneksi baru Saya menggunakan metode *Jaccard Similarity* dan *Preferential Attachment*. Pada dasarnya pendekatan-pendekatan tersebut akan mengukur *similarity* dari dua buah *nodes* dimana kesamaan tersebut dapat digunakan sebagai dasar prediksi kemunculan *edge* baru antara dua *nodes* yang belum terhubung.

3.2.3. ANALISIS

Pada langkah ini, diambil N sosok penting tertinggi untuk dianalisis signifikan jumlah seseorang yang menjadi sosok penting memiliki keahlian atau minat yang mirip dengan koneksinya.

$$Ratio_{Vi} = \frac{\text{jumlah similar } Vi}{\text{jumlah neighbor } Vi} \quad (1)$$

Karena terdapat banyak *node* yang atribut keahliannya kosong, maka analisis hanya menghitung *node-node* yang keahliannya terisi.

Untuk setiap hasil metode analisis prediksi koneksi baru, *edge* baru hasil prediksi akan dibandingkan dengan *dataset actual*. Hasil akan dihitung menggunakan nilai *F1-score test* dengan formula berikut:

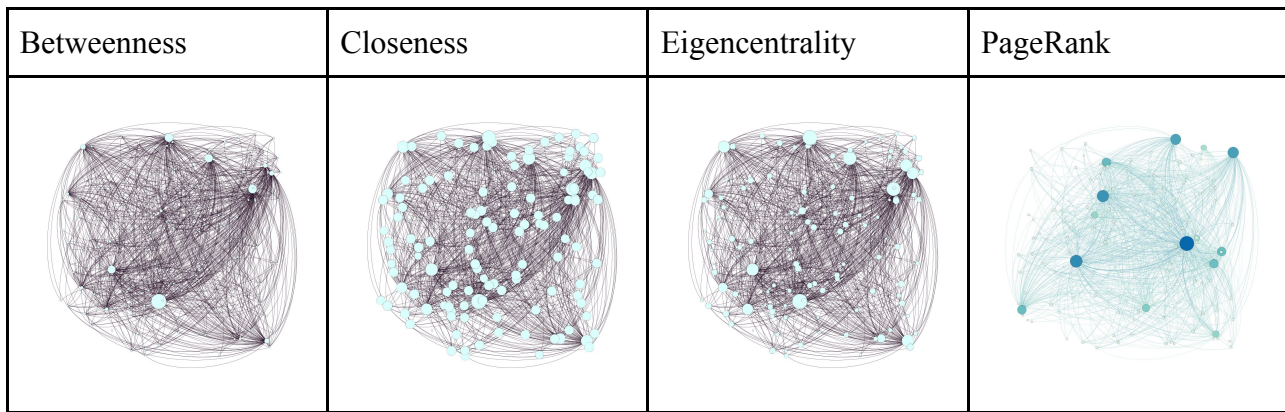
$$F1_{score} = \frac{2 \times \text{TruePositives}}{2 \times \text{TruePositives} + \text{FalsePositives} + \text{FalseNegatives}} \quad (2)$$

4. HASIL

4.1. SOSOK SENTRAL

Setelah melakukan pengukuran *betweenness*, *closeness*, *eigencentrality*, dan *pageranks*, dihasilkan tabel berikut. Tabel 1 berikut berisi sampel berjumlah 11 mahasiswa (yang ditampilkan) yang diambil berdasarkan *centrality* tertinggi dan telah mengisi keahlian. 11 orang tersebut dicocokkan dengan *node-node* tetangganya apakah memiliki keahlian sama. Dari keempat pengukuran, ternyata memiliki hasil yang stabil dimana mahasiswa-mahasiswa berikut yang memiliki peranan penting di model jejaring ini.

Tabel 1. Gambar jejaring setelah pengukuran *centrality*

Tabel 2. Tabel penghitungan *centrality*

ID	betweenness	closeness	eigencentrality	pageranks
ichlasul-affan	0.177927	0.902256	1	0.057837
a-hasan-gar	0.09947	0.794702	0.907654	0.046248
rayzaarasj	0.098964	0.833333	0.969562	0.048672
agasyan	0.091139	0.759494	0.873636	0.042942
albertusangga	0.045196	0.701754	0.793923	0.035006
alsabila-shakina-prasetyo-636887155	0.041259	0.693642	0.753992	0.033839
satryaji	0.038137	0.697674	0.795056	0.034445
donny-samuel	0.033811	0.685714	0.748215	0.032285
aldobima	0.024969	0.659341	0.695085	0.028714
al dihilmanr	0.016469	0.625	0.580655	0.024094

Selanjutnya, berdasarkan penghitungan rasio kesamaan keahlian dan jumlah tetangga didapatkan tabel 2 berikut. Antara *node* terpenting dan tetangganya ternyata memiliki kesamaan keahlian yang rendah. Hal ini berbanding terbalik dengan penelitian (Z. Xiao, 2018) yang menggunakan deteksi komunitas dan menemukan bahwa sebuah komunitas memiliki kesamaan minat. Hasil berikut juga berbeda dengan hasil penelitian bahwa sebuah pertemanan dibentuk juga berdasarkan komponen kesamaan/kecocokan (K. Campbell, 2015).

Tabel 3. Tabel penghitungan rasio jumlah yang memiliki minat sama dan jumlah tetangga

Id	neighbor	similar	result
ichlasul-affan	66	45	0.6818
a-hasan-gar	68	38	0.5588
rayzaarasj	63	39	0.619
agasyan	63	0	0
albertusangga	58	29	0.5
alsabila-shakina-prasetyo-636887155	51	2	0.0392
satryaji	52	1	0.0192
donny-samuel	42	4	0.0952

aldobima	46	30	0.6522
al dihilmanr	37	21	0.5676

4.2. PREDIKSI KONEKSI

Berdasarkan hasil prediksi menggunakan metode prediksi koneksi menggunakan pengukuran *Jaccard* dan *Preferential* menemukan bahwa F1 skor kedua pengukuran rendah dan memiliki presisi yang rendah. Namun dilihat dari *recall*, ternyata banyak yang berhasil terprediksi.

Tabel 4. Tabel penghitungan F1 skor dari hasil prediksi *edge* dengan *edge* aktual

Pengukuran	Presisi	Recall	F1
Jaccard	0.3573	0.8304	0.4997
Preferential	0.3625	0.8598	0.5101

5. KESIMPULAN

Kesimpulan yang didapatkan adalah bahwa seorang mahasiswa yang penting di lingkungannya belum tentu memiliki koneksi dikarenakan memiliki kesamaan keahlian. Hal ini dapat terlihat dari hasil analisis *centrality* dan dibandingkan dengan *node-node* tetangganya yang ternyata cenderung tidak memiliki keahlian yang sama. Selain itu, dua orang yang memiliki keahlian yang sama, berteman di masa mendatang dengan orang yang memiliki keahlian yang sama. Hal ini memang berbanding terbalik dengan penemuan yang telah ada. Hal ini dapat disebabkan semisal memang keahlian bukan suatu komponen yang signifikan dalam kesamaan untuk membentuk pertemanan.

6. REFERENSI

- P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *The Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.
- S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.
- S. P. Borgatti, "Centrality and AIDS", *CONNECTIONS* 18(1), pp. 112-115, 1995.
- K. Campbell, N. Holderness, and M. Riggs, "Friendship chemistry: An examination of underlying factors," *The Social Science Journal*, vol. 52, no. 2, pp. 239–247, 2015.
- H. A. Deylami and M. Asadpour, "Link prediction in social networks using hierarchical community detection," *2015 7th Conference on Information and Knowledge Technology (IKT)*, 2015.

M. Kaur and S. Singh, “Ranking based comparative analysis of graph centrality measures to detect negative nodes in online social networks,” *Journal of Computational Science*, vol. 23, pp. 91–108, 2017.

Z. Xiao, J. Li, and G. Zhou, “Do common interests of students play a role in friendship?,” *Procedia Computer Science*, vol. 131, pp. 733–738, 2018.

7. LAMPIRAN

“Database Omega 2016.” Google Docs, bit.ly/DataOmega.

Source code untuk melakukan scraping, <https://github.com/wisnuprama/pyscrap>

<https://remote-lib.ui.ac.id:2053/science/article/pii/S1877750316303945>

<https://remote-lib.ui.ac.id:2053/science/article/pii/S0378873314000197>