

MODEL PREDIKSI KOLABORASI ILMIAH DENGAN MEMBANDINGKAN WEIGHTED PREFERENTIAL ATTACHMENT DAN WEIGHTED KATZ INDEX PADA JARINGAN CO-AUTHORSHIP

M. Milky Gazura, Herman Yuliansyah

Teknik Informatika, Universitas Ahmad Dahlan
Jalan Ringroad Selatan, Tamanan, Banguntapan, Bantul, Yogyakarta
2200018396@webmail.uad.ac.id

ABSTRAK

Prediksi kolaborasi ilmiah merupakan permasalahan penting dalam analisis jaringan akademik untuk mengidentifikasi potensi hubungan antar peneliti di masa depan. Namun, sebagian besar pendekatan belum mempertimbangkan frekuensi kolaborasi sebagai bobot hubungan dalam jaringan *co-authorship*. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi keterhubungan pada jaringan *co-authorship* berbobot dengan membandingkan metode Weighted Katz Index (WKatz), Weighted Preferential Attachment (WPA), serta metode dasar Preferential Attachment (PA) dan Katz Index. Dataset yang digunakan merupakan publikasi ilmiah dosen Program Studi Psikologi Universitas Ahmad Dahlan periode 2010–2024 yang dikumpulkan dari *Google Scholar* dan direpresentasikan sebagai graf tak berarah berbobot. Evaluasi dilakukan menggunakan skema *link prediction* berbasis waktu dengan sepuluh iterasi temporal dan diukur menggunakan metrik *Area Under the Curve* (AUC). Hasil penelitian menunjukkan bahwa Katz Index dengan parameter $\beta = 0,01$ memperoleh nilai rata-rata AUC tertinggi sebesar 0,96560, diikuti oleh WKatz sebesar 0,95376, sedangkan WPA dan PA menghasilkan nilai yang lebih rendah, masing-masing sebesar 0,67870 dan 0,64363. Temuan ini menunjukkan bahwa pendekatan berbasis Katz memiliki kinerja prediksi yang lebih unggul dibandingkan metode lainnya dalam konteks jaringan *co-authorship* berbobot.

Kata kunci : *Google Scholar; Link Prediction; Psikologi; Universitas Ahmad Dahlan; Weighted Complex Co-authorship Network.*

1. PENDAHULUAN

Kolaborasi ilmiah memiliki peran penting dalam mendorong perkembangan pengetahuan dan inovasi dalam lingkungan penelitian modern. Interaksi antar peneliti memungkinkan terjadinya pertukaran ide, peningkatan produktivitas penelitian, serta perluasan jaringan keilmuan lintas disiplin. Salah satu indikator yang umum digunakan untuk menganalisis kolaborasi ilmiah adalah hubungan *co-authorship* pada publikasi ilmiah, yang merepresentasikan interaksi dan kerja sama antar peneliti dalam menghasilkan karya ilmiah [1].

Struktur jaringan *co-authorship* diketahui dapat mempengaruhi kinerja akademik dan dampak penelitian seorang peneliti karena hubungan kolaborasi yang terbentuk mencerminkan dinamika interaksi dalam komunitas ilmiah [2].

Hubungan *co-authorship* dapat direpresentasikan dalam bentuk jaringan, di mana penulis direpresentasikan sebagai *node* dan hubungan kolaborasi direpresentasikan sebagai *edge*. Jaringan tersebut umumnya dimodelkan sebagai graf tak berarah yang menggambarkan interaksi kolaboratif antar peneliti dalam suatu komunitas ilmiah [3].

Dalam banyak kasus, jaringan *co-authorship* juga bersifat berbobot, di mana bobot pada suatu *edge* merepresentasikan frekuensi atau intensitas kolaborasi antara dua penulis. Representasi jaringan berbobot memberikan informasi struktural yang lebih kaya dibandingkan jaringan tak berbobot karena mampu menggambarkan kekuatan hubungan antar peneliti

secara lebih akurat serta mencerminkan dinamika kolaborasi dalam jaringan penelitian [4], [5].

Dalam analisis jaringan kolaborasi, salah satu permasalahan penting yang sering dikaji adalah *link prediction*, yaitu proses memperkirakan kemungkinan terbentuknya hubungan baru antara dua *node* dalam suatu jaringan. Dalam konteks kolaborasi ilmiah, *link prediction* dapat digunakan untuk mengidentifikasi potensi kerja sama di masa depan antara peneliti yang sebelumnya belum pernah berkolaborasi. Metode ini banyak digunakan dalam analisis jaringan kompleks untuk memahami pola interaksi serta memprediksi perkembangan struktur jaringan di masa mendatang [6], [7].

Berbagai pendekatan telah dikembangkan untuk menyelesaikan permasalahan *link prediction*, mulai dari metode berbasis kesamaan lokal hingga metode berbasis struktur jaringan global. Metode kesamaan lokal umumnya memanfaatkan informasi tetangga langsung dari suatu *node*, sedangkan metode global mempertimbangkan jalur yang lebih luas dalam jaringan untuk mengukur tingkat keterhubungan antar *node* [6].

Salah satu metode lokal yang banyak digunakan adalah Preferential Attachment (PA) yang didasarkan pada asumsi bahwa *node* dengan derajat koneksi lebih tinggi memiliki peluang lebih besar untuk membentuk hubungan baru dalam jaringan kompleks [8].

Selain metode berbasis kesamaan lokal, metode berbasis jalur global juga banyak digunakan dalam penelitian *link prediction*. Salah satu metode yang

sering digunakan adalah Katz Index, yang menghitung tingkat kedekatan dua *node* dengan mempertimbangkan seluruh jalur yang menghubungkannya dalam jaringan, dengan memberikan bobot lebih besar pada jalur yang lebih pendek. Pendekatan ini memungkinkan metode tersebut menangkap hubungan tidak langsung antar *node* yang sering kali tidak terlihat pada metode berbasis tetangga langsung [9].

Meskipun berbagai penelitian telah mengkaji *link prediction* pada jaringan kompleks, sebagian besar penelitian masih berfokus pada jaringan tak berbobot atau pada jaringan kolaborasi berskala besar seperti jaringan antar institusi atau antar disiplin ilmu. Padahal, pada jaringan kolaborasi ilmiah, frekuensi kolaborasi antar peneliti dapat memberikan informasi penting mengenai kekuatan hubungan dalam jaringan. Penggunaan informasi bobot dalam jaringan diketahui dapat meningkatkan kemampuan model dalam menangkap pola keterhubungan dalam jaringan kompleks [10], [11].

Namun, pengabaian informasi bobot dalam beberapa pendekatan dapat menyebabkan kurang optimalnya hasil prediksi keterhubungan antar peneliti.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengkaji dan membandingkan prediksi keterhubungan pada jaringan *co-authorship* berbobot menggunakan metode Weighted Preferential Attachment (WPA) dan Weighted Katz Index (WKatz). Untuk mencapai tujuan tersebut, evaluasi dilakukan menggunakan pendekatan *time-based link prediction* dengan beberapa iterasi temporal untuk menangkap dinamika perkembangan jaringan kolaborasi ilmiah serta menganalisis pengaruh bobot hubungan terhadap kinerja metode prediksi dalam memprediksi potensi kolaborasi di masa mendatang.

2. TINJAUAN PUSTAKA

Link prediction merupakan salah satu permasalahan penting dalam analisis jaringan kompleks yang bertujuan untuk memperkirakan kemungkinan terbentuknya hubungan baru antara dua *node* dalam suatu jaringan. Teknik ini banyak digunakan untuk memahami dinamika perkembangan jaringan serta memprediksi struktur jaringan di masa mendatang.

Menurut Zhou [4], *link prediction* merupakan metode yang digunakan untuk mengidentifikasi potensi hubungan baru berdasarkan informasi struktur jaringan yang telah ada. Dalam konteks jaringan kolaborasi ilmiah, metode ini dapat digunakan untuk memprediksi kemungkinan kerja sama antara peneliti yang sebelumnya belum pernah berkolaborasi. Lü dan Zhou [7] menjelaskan bahwa pendekatan *link prediction* umumnya memanfaatkan informasi kesamaan struktural antar *node* untuk memperkirakan peluang terbentuknya hubungan baru dalam jaringan.

Dalam jaringan kolaborasi ilmiah, hubungan antar peneliti biasanya direpresentasikan dalam bentuk

jaringan *co-authorship*, di mana penulis direpresentasikan sebagai *node* dan hubungan kolaborasi direpresentasikan sebagai *edge*.

Menurut Hasin dan Hassan [5], jaringan *co-authorship* merupakan salah satu bentuk representasi jaringan sosial yang banyak digunakan untuk menganalisis pola kolaborasi dalam publikasi ilmiah. Pada jaringan ini, hubungan antar peneliti dapat dianalisis untuk memahami struktur kolaborasi, mengidentifikasi komunitas penelitian, serta memprediksi potensi kerja sama di masa depan.

Salah satu metode yang banyak digunakan dalam penelitian *link prediction* adalah Preferential Attachment (PA). Metode ini didasarkan pada asumsi bahwa *node* dengan jumlah koneksi yang lebih tinggi memiliki kecenderungan lebih besar untuk membentuk hubungan baru dalam jaringan.

Menurut Piva dkk. [8], konsep ini dikenal sebagai fenomena *rich-get-richer*, di mana *node* dengan derajat tinggi akan lebih mudah memperoleh koneksi tambahan seiring pertumbuhan jaringan. Pada metode Preferential Attachment, skor kesamaan antara dua *node* dihitung berdasarkan hasil perkalian derajat dari kedua *node* tersebut.

Selain metode berbasis kesamaan lokal, terdapat pula metode berbasis jalur global seperti Katz Index. Metode ini mempertimbangkan seluruh jalur yang menghubungkan dua *node* dalam jaringan, dengan memberikan bobot lebih besar pada jalur yang lebih pendek. Menurut Vidza dkk. [9], Katz Index mampu menangkap hubungan tidak langsung antar *node* yang tidak dapat dideteksi oleh metode berbasis tetangga langsung. Dengan mempertimbangkan informasi jalur dalam jaringan, metode ini sering digunakan untuk meningkatkan akurasi prediksi keterhubungan dalam jaringan kompleks.

Pada jaringan berbobot, metode *link prediction* dapat dikembangkan dengan mempertimbangkan kekuatan hubungan antar *node*.

Deepthi dan Ravi [10] menjelaskan bahwa penggunaan bobot dalam jaringan memungkinkan model prediksi untuk memperhitungkan intensitas interaksi antar *node*. Oleh karena itu, metode Preferential Attachment dan Katz Index dapat dikembangkan menjadi Weighted Preferential Attachment (WPA) dan Weighted Katz Index (WKatz) yang memanfaatkan informasi bobot hubungan dalam jaringan untuk meningkatkan kemampuan prediksi keterhubungan [11].

3. METODE PENELITIAN

Penelitian ini menggunakan dataset publikasi ilmiah dosen Program Studi Psikologi Universitas Ahmad Dahlan yang dikumpulkan dari *Google Scholar* pada periode 2010–2024. Pengambilan data dilakukan menggunakan teknik *web crawling* untuk memperoleh informasi publikasi seperti nama penulis, Google Scholar ID, judul publikasi, penulis bersama, tahun publikasi, serta URL publikasi. Proses pengumpulan data menghasilkan 1.496 publikasi yang

kemudian digunakan untuk mengidentifikasi hubungan *co-authorship*. Berdasarkan data tersebut dibangun jaringan kolaborasi ilmiah yang terdiri dari 1.455 *node* dan 23.294 *edge*. Jaringan dimodelkan sebagai graf tak berarah berbobot, di mana bobot pada setiap *edge* merepresentasikan frekuensi kolaborasi antara dua penulis.

Beberapa karakteristik global dari jaringan *co-authorship* yang dihasilkan ditunjukkan pada Tabel 1.

Tabel 1. Statistik jaringan *co-authorship*

Metric	Value
Nodes	1455
Edges	23294
Density	0.022
Avg. Degree	32.02
Avg. Clustering Coefficient	0.778
Avg. Degree Centrality	0.022
Avg. Closeness Centrality	0.258
Avg. Betweenness Centrality	0.002
Avg. Eigenvector Centrality	0.008

Berdasarkan Tabel 1, jaringan *co-authorship* yang dibangun menunjukkan karakteristik jaringan kolaborasi ilmiah yang cukup kompleks dengan jumlah *node* dan *edge* yang relatif besar. Hal ini menunjukkan bahwa jaringan kolaborasi yang terbentuk memiliki hubungan interaksi yang cukup padat antar peneliti.

Jaringan *co-authorship* dibangun dengan merepresentasikan penulis sebagai *node* dan hubungan kolaborasi sebagai *edge*. Jika dua penulis terlibat dalam satu publikasi bersama, maka terbentuk hubungan antara kedua *node*. Untuk merepresentasikan intensitas kolaborasi, jaringan dibangun sebagai jaringan berbobot di mana bobot pada setiap *edge* menunjukkan jumlah publikasi bersama antara dua penulis. Representasi jaringan ini juga dapat dinyatakan dalam bentuk matriks adjacency berbobot yang menunjukkan frekuensi kolaborasi antar pasangan penulis.

Penelitian ini menggunakan empat metode *link prediction*, yaitu Preferential Attachment (PA), Weighted Preferential Attachment (WPA), Katz Index, dan Weighted Katz Index (WKatz). Metode Preferential Attachment menghitung kemungkinan keterhubungan dua *node* berdasarkan derajat koneksi masing-masing *node* [8], dirumuskan pada persamaan (1).

$$PA(x, y) = |\Gamma(x)| \times |\Gamma(y)| \quad (1)$$

Katz Index merupakan metode berbasis jalur global yang menghitung kedekatan dua *node* dengan mempertimbangkan seluruh jalur yang menghubungkannya dalam jaringan menggunakan parameter peluruhan β [9], dirumuskan pada persamaan (2).

$$Katz(x, y) = \sum_{l=1}^{\infty} \beta^l \cdot |paths_{x,y}(l)| \quad (2)$$

Untuk jaringan berbobot digunakan Weighted Preferential Attachment (WPA) yang merupakan pengembangan dari metode Preferential Attachment dengan mempertimbangkan kekuatan *node*. Berbeda dengan PA yang hanya menggunakan derajat *node*, WPA menghitung skor kesamaan berdasarkan total bobot hubungan yang dimiliki oleh setiap *node*, sehingga mampu merepresentasikan intensitas interaksi dalam jaringan secara lebih akurat [12]. Perhitungan skor kesamaan pada metode WPA dirumuskan pada persamaan (3).

$$WPA(x, y) = s(x) \times s(y) \quad (3)$$

Selain itu digunakan Weighted Katz Index yang merupakan pengembangan dari metode Katz Index dengan memasukkan informasi bobot hubungan dalam perhitungan jalur. Metode ini tidak hanya mempertimbangkan keberadaan jalur antar *node*, tetapi juga kekuatan hubungan yang dilalui pada setiap jalur, sehingga mampu menangkap pola keterhubungan yang lebih kompleks dalam jaringan berbobot [9]. Perhitungan WKatz dirumuskan pada persamaan (4).

$$WKI_{(x,y)} = \sum_{l=1}^{\infty} \beta^l (A_w^l)_{(x,y)} \quad (4)$$

Evaluasi performa model dilakukan menggunakan metrik Area Under the Curve (AUC) serta metrik berbasis *confusion matrix* yang meliputi *accuracy*, *precision*, *recall*, dan *F1-score* [13], [14]. Nilai AUC digunakan untuk mengukur kemampuan model dalam membedakan pasangan *node* yang memiliki kemungkinan membentuk hubungan dengan pasangan *node* yang tidak memiliki hubungan [15]. Secara matematis, nilai AUC dapat dirumuskan pada persamaan (5).

$$AUC = \frac{N' + 0.5 \times N''}{N} \quad (5)$$

di mana N' merupakan jumlah pasangan sampel di mana skor prediksi untuk pasangan positif lebih tinggi dibandingkan pasangan negatif, N'' merupakan jumlah pasangan dengan skor yang sama, dan N adalah total perbandingan antara pasangan positif dan negatif. Nilai AUC berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan performa model yang semakin baik dalam melakukan prediksi.

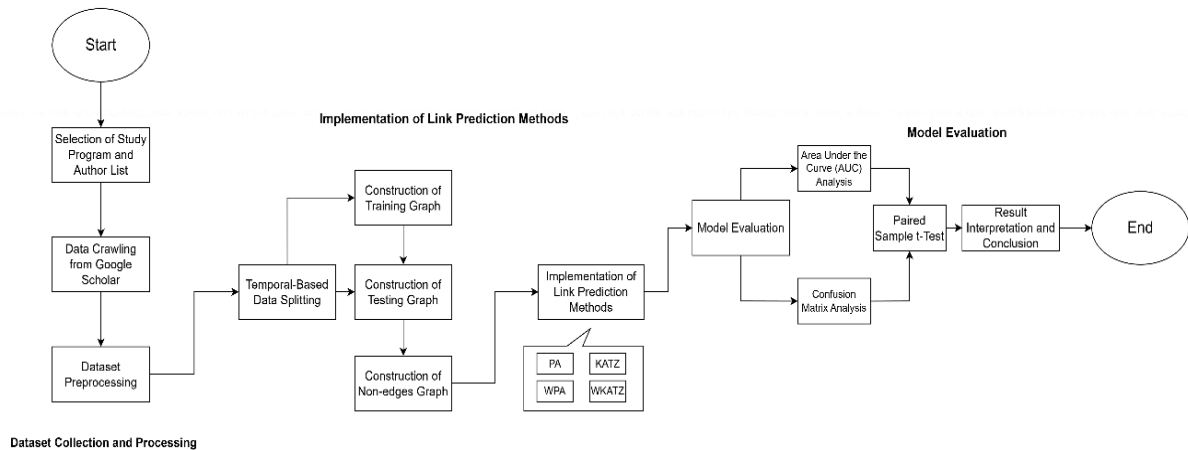
Eksperimen dilakukan menggunakan pendekatan *time-based link prediction* dengan strategi *sliding window*. Dataset dibagi menjadi data pelatihan dan data pengujian berdasarkan tahun publikasi. Data pada periode sebelumnya digunakan untuk membangun jaringan pelatihan, sedangkan hubungan kolaborasi

yang muncul pada tahun berikutnya digunakan sebagai data pengujian.

Pada setiap iterasi, jaringan pelatihan digunakan untuk menghitung skor kesamaan antar pasangan *node* yang belum terhubung menggunakan metode *link prediction* yang diusulkan. Selanjutnya, hasil prediksi dibandingkan dengan data aktual pada jaringan

pengujian untuk mengevaluasi performa model. Proses ini dilakukan dalam sepuluh iterasi temporal dengan menggeser jendela waktu secara bertahap untuk menangkap dinamika perkembangan jaringan kolaborasi dari waktu ke waktu.

Alur penelitian yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian prediksi keterhubungan pada jaringan co-authorship

Berdasarkan Gambar 1, proses penelitian diawali dengan pemilihan program studi dan pengumpulan daftar penulis sebagai sumber data penelitian. Selanjutnya dilakukan proses pengambilan data publikasi dari Google Scholar dan tahap praproses data untuk menghasilkan dataset yang siap digunakan. Dataset kemudian dibagi menggunakan pendekatan *temporal-based data splitting* untuk membentuk data pelatihan dan data pengujian. Pada tahap berikutnya dilakukan implementasi metode *link prediction* yang meliputi Preferential Attachment (PA), Weighted Preferential Attachment (WPA), Katz Index, dan Weighted Katz Index (WKatz). Hasil prediksi kemudian dievaluasi menggunakan metrik AUC serta analisis *confusion matrix*. Karakteristik jaringan pada setiap iterasi temporal ditunjukkan pada Tabel 2.

Tabel 2. Karakteristik temporal jaringan *co-authorship*

No	Train Years	Test Year	Nodes	Edges	Non-edges
1	2010–2014	2015	247	5111	25270
2	2011–2015	2016	266	5092	30153
3	2012–2016	2017	267	5011	30500
4	2013–2017	2018	301	5098	40052
5	2014–2018	2019	435	5456	88939
6	2015–2019	2020	597	6520	171386
7	2016–2020	2021	716	6981	248389
8	2017–2021	2022	820	7554	328326
9	2018–2022	2023	1081	17700	566040
10	2019–2023	2024	1184	18033	682303

Tabel 2 menunjukkan karakteristik jaringan kolaborasi pada setiap iterasi temporal yang digunakan dalam eksperimen. Jumlah *node* dan *edge* meningkat

secara bertahap seiring dengan bertambahnya periode publikasi yang dianalisis. Hal ini menunjukkan bahwa jaringan kolaborasi ilmiah terus berkembang dari waktu ke waktu, sehingga pendekatan evaluasi berbasis waktu diperlukan untuk menganalisis performa metode *link prediction* secara lebih realistis.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen yang dilakukan untuk mengevaluasi performa metode *link prediction* pada jaringan *co-authorship* berbobot. Evaluasi difokuskan pada perbandingan performa empat metode, yaitu Preferential Attachment (PA), Weighted Preferential Attachment (WPA), Katz Index, dan Weighted Katz Index (WKatz). Analisis dilakukan dalam beberapa tahap, yaitu analisis parameter β pada metode Katz dan WKatz, evaluasi performa prediksi pada setiap iterasi temporal menggunakan metrik AUC, serta perbandingan performa rata-rata metode menggunakan beberapa metrik evaluasi.

4.1. Analisis Parameter β

Metode Katz Index dan Weighted Katz Index menggunakan parameter peluruhan β untuk mengontrol kontribusi jalur yang lebih panjang dalam jaringan. Nilai β berperan dalam menentukan seberapa besar pengaruh jalur tidak langsung terhadap perhitungan kesamaan antar *node*. Nilai β yang terlalu kecil akan membatasi kontribusi jalur panjang, sedangkan nilai β yang terlalu besar dapat menyebabkan jalur panjang menjadi terlalu dominan sehingga berpotensi menurunkan akurasi prediksi akibat masuknya informasi yang kurang relevan.

Oleh karena itu, dilakukan pengujian terhadap beberapa variasi nilai β untuk menentukan parameter yang paling optimal dalam menghasilkan performa prediksi yang baik. Pengujian dilakukan dengan menggunakan rentang nilai β dari 0,01 hingga 0,09, di mana setiap nilai diuji menggunakan sepuluh iterasi temporal. Nilai performa kemudian diukur menggunakan metrik *Area Under the Curve* (AUC) dan dirata-ratakan untuk setiap nilai β .

Tabel 3 menunjukkan nilai rata-rata AUC yang diperoleh oleh metode Katz dan WKatz untuk setiap variasi parameter β .

Tabel 3. Nilai rata-rata AUC untuk berbagai nilai parameter β

β	Katz AUC	WKatz AUC
$\beta = 0.01$	0,96560	0,95376
$\beta = 0.02$	0,94186	0,89936
$\beta = 0.03$	0,92455	0,86825
$\beta = 0.04$	0,88263	0,85132
$\beta = 0.05$	0,86705	0,84252
$\beta = 0.06$	0,84471	0,83706
$\beta = 0.07$	0,81968	0,83360
$\beta = 0.08$	0,80601	0,83118
$\beta = 0.09$	0,79793	0,82943

Berdasarkan Tabel 3, nilai $\beta = 0.01$ menghasilkan performa prediksi terbaik baik pada metode Katz maupun WKatz. Seiring dengan meningkatnya nilai β , nilai rata-rata AUC cenderung mengalami penurunan. Hal ini menunjukkan bahwa jalur yang lebih pendek memberikan informasi yang lebih relevan dalam memprediksi potensi kolaborasi dibandingkan jalur yang lebih panjang.

Kondisi tersebut mengindikasikan bahwa hubungan tidak langsung dengan jarak yang dekat dalam jaringan memiliki kontribusi yang lebih signifikan dalam membentuk keterhubungan antar penulis. Sebaliknya, jalur yang terlalu panjang cenderung memasukkan informasi yang kurang relevan sehingga dapat menurunkan kualitas prediksi. Dengan demikian, pemilihan nilai β yang kecil lebih efektif dalam menjaga keseimbangan antara pemanfaatan informasi jalur dan akurasi prediksi pada jaringan kolaborasi ilmiah.

4.2. Performa Prediksi pada Setiap Iterasi

Untuk mengevaluasi stabilitas performa metode *link prediction*, dilakukan eksperimen menggunakan sepuluh iterasi temporal dengan pendekatan *time-based link prediction*. Pada setiap iterasi, model dilatih menggunakan data kolaborasi pada periode tertentu dan diuji pada kolaborasi yang muncul pada tahun berikutnya. Tabel 4 menunjukkan nilai AUC yang

diperoleh oleh masing-masing metode pada setiap iterasi.

Tabel 4. Nilai AUC pada setiap iterasi temporal

No	PA	WPA	Katz ($\beta=0.01$)	WKatz ($\beta=0.01$)
1	0.356520	0.353043	0.924957	0.920389
2	0.349295	0.435363	0.932099	0.946552
3	0.433113	0.518358	0.933595	0.949858
4	0.473276	0.539102	0.943804	0.953768
5	0.641970	0.693532	0.963421	0.957772
6	0.969914	0.971770	0.998437	0.997047
7	0.787715	0.814118	0.982946	0.961159
8	0.824720	0.839471	0.989991	0.960147
9	0.976531	0.975717	0.999092	0.995847
10	0.623306	0.646616	0.987686	0.895230

Berdasarkan Tabel 4, metode berbasis jalur global yaitu Katz dan WKatz secara konsisten menunjukkan performa yang lebih tinggi dibandingkan metode berbasis kesamaan lokal seperti PA dan WPA pada hampir seluruh iterasi. Hal ini menunjukkan bahwa metode berbasis jalur global mampu menangkap hubungan tidak langsung antar penulis yang tidak dapat diidentifikasi oleh metode berbasis kesamaan lokal.

Selain itu, terlihat bahwa nilai AUC pada metode Katz dan WKatz cenderung stabil dan berada pada rentang yang tinggi di sebagian besar iterasi, bahkan mendekati nilai 1 pada beberapa iterasi. Hal ini mengindikasikan bahwa metode tersebut memiliki kemampuan yang baik dalam membedakan pasangan *node* yang berpotensi membentuk hubungan dengan yang tidak.

Sebaliknya, metode PA dan WPA menunjukkan performa yang relatif lebih rendah dan fluktuatif antar iterasi. Hal ini disebabkan karena metode tersebut hanya mempertimbangkan informasi lokal seperti derajat *node* atau kekuatan *node*, sehingga kurang mampu menangkap struktur jaringan yang lebih kompleks.

Secara umum, hasil ini menunjukkan bahwa pemanfaatan informasi jalur global dalam jaringan memberikan kontribusi yang lebih signifikan dalam meningkatkan akurasi prediksi keterhubungan dibandingkan pendekatan berbasis kesamaan lokal.

4.3. Perbandingan Performa Rata-rata Metode

Untuk memperoleh gambaran performa metode secara keseluruhan, dilakukan perhitungan rata-rata beberapa metrik evaluasi pada seluruh iterasi eksperimen. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, *F1-score*, dan AUC. Hasil perhitungan tersebut disajikan pada Tabel 5.

Tabel 5. Performa rata-rata metode link prediction

Metode	Accuracy	Precision	Recall	F1-score	AUC
PA	0.76805	0.04398	0.37900	0.07137	0.64363
WPA	0.79337	0.04119	0.45600	0.06882	0.67870
Katz ($\beta = 0.01$)	0.91195	0.21570	0.91153	0.22365	0.96560
WKatz ($\beta = 0.01$)	0.91957	0.16035	0.86318	0.19328	0.95376

Hasil pada Tabel 5 menunjukkan bahwa metode Katz Index memiliki performa terbaik dengan nilai rata-rata AUC sebesar 0,96560, diikuti oleh metode Weighted Katz Index (WKatz) dengan nilai AUC sebesar 0,95376. Hal ini menunjukkan bahwa metode berbasis jalur global memiliki kemampuan yang lebih baik dalam membedakan pasangan *node* yang berpotensi membentuk hubungan dibandingkan dengan metode berbasis kesamaan lokal.

Performa tinggi pada metode Katz dan WKatz diperoleh dengan menggunakan parameter $\beta = 0,01$, yang sebelumnya telah ditentukan sebagai nilai optimal berdasarkan hasil analisis parameter. Pemilihan nilai β yang kecil bertujuan untuk memberikan bobot yang lebih besar pada jalur pendek dalam jaringan, sehingga hubungan tidak langsung yang relevan dapat ditangkap dengan lebih baik. Sebaliknya, nilai β yang lebih besar cenderung meningkatkan kontribusi jalur panjang yang berpotensi membawa informasi yang kurang relevan dan dapat menurunkan akurasi prediksi.

Jika ditinjau dari metrik lainnya, metode Katz juga menunjukkan nilai *accuracy* dan *recall* yang tinggi, masing-masing sebesar 0,91195 dan 0,91153, yang mengindikasikan bahwa metode ini mampu mengidentifikasi sebagian besar pasangan *node* yang benar-benar memiliki keterhubungan. Metode WKatz juga memiliki nilai *accuracy* yang sedikit lebih tinggi, yaitu 0,91957, namun nilai *recall*-nya sedikit lebih rendah dibandingkan Katz. Hal ini menunjukkan bahwa meskipun metode berbobot mampu meningkatkan beberapa aspek performa, namun tidak secara konsisten memberikan peningkatan pada seluruh metrik evaluasi.

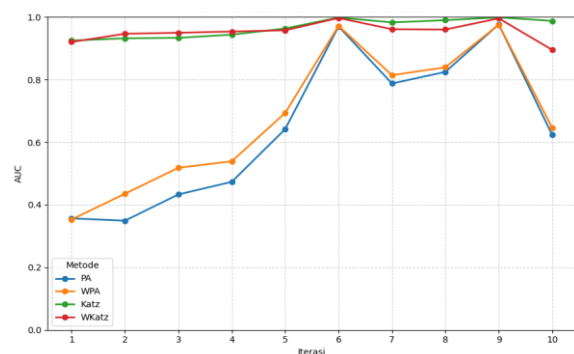
Sebaliknya, metode PA dan WPA menunjukkan performa yang jauh lebih rendah pada seluruh metrik evaluasi. Nilai AUC yang diperoleh PA dan WPA masing-masing hanya sebesar 0,64363 dan 0,67870, yang menunjukkan keterbatasan metode berbasis kesamaan lokal dalam menangkap pola keterhubungan pada jaringan yang kompleks. Selain itu, nilai *precision* dan *F1-score* pada kedua metode tersebut juga relatif rendah, yang mengindikasikan bahwa model kurang akurat dalam mengidentifikasi pasangan *node* yang benar-benar memiliki potensi keterhubungan. Hal ini disebabkan karena metode berbasis lokal hanya mempertimbangkan informasi langsung seperti derajat atau kekuatan *node*, sehingga tidak mampu menangkap hubungan tidak langsung yang sering muncul dalam struktur jaringan kolaborasi ilmiah.

Secara umum, hasil ini menunjukkan bahwa pemanfaatan informasi jalur global dalam jaringan memberikan kontribusi yang lebih signifikan dalam meningkatkan performa prediksi dibandingkan pendekatan berbasis kesamaan lokal. Metode Katz dan WKatz mampu menangkap hubungan tidak langsung antar penulis yang sering kali menjadi indikator penting dalam terbentuknya kolaborasi ilmiah di masa depan, terutama ketika parameter β dipilih secara tepat

untuk menekankan kontribusi jalur yang lebih relevan. Dengan demikian, pendekatan berbasis jalur global lebih sesuai digunakan pada jaringan dengan struktur kompleks dan dinamis seperti jaringan *co-authorship*.

4.4. Analisis Kurva ROC

Selain menggunakan nilai AUC, performa metode juga dapat dianalisis melalui kurva ROC (*Receiver Operating Characteristic*) yang menggambarkan hubungan antara *true positive rate* (TPR) dan *false positive rate* (FPR). Kurva ROC memberikan gambaran visual mengenai kemampuan model dalam membedakan antara pasangan *node* yang memiliki hubungan dan yang tidak. Kurva ROC untuk masing-masing metode ditunjukkan pada Gambar 2.



Gambar 2. Kurva ROC metode link prediction

Berdasarkan Gambar 2, metode Katz dan WKatz yang digunakan pada visualisasi ini merupakan konfigurasi dengan parameter $\beta = 0,01$, yaitu nilai yang telah ditentukan sebagai parameter optimal berdasarkan hasil analisis sebelumnya. Kedua metode tersebut memiliki kurva ROC yang berada lebih dekat dengan sudut kiri atas grafik, yang menunjukkan performa klasifikasi yang lebih baik dibandingkan metode PA dan WPA. Posisi kurva yang mendekati sudut kiri atas mengindikasikan bahwa metode tersebut mampu mencapai nilai *true positive rate* yang tinggi dengan *false positive rate* yang relatif rendah.

Selain itu, terlihat bahwa kurva metode Katz cenderung lebih stabil dan konsisten berada di atas metode lainnya pada sebagian besar iterasi, yang menunjukkan keunggulan dalam menangkap pola keterhubungan dalam jaringan. Metode WKatz juga menunjukkan performa yang tinggi, meskipun pada beberapa titik sedikit berada di bawah Katz. Hal ini menunjukkan bahwa penambahan bobot hubungan tidak selalu memberikan peningkatan signifikan dalam performa klasifikasi, terutama ketika struktur topologi jaringan sudah mampu merepresentasikan pola keterhubungan dengan baik.

Sebaliknya, metode PA dan WPA memiliki kurva ROC yang berada lebih jauh dari kondisi ideal, yang menandakan performa klasifikasi yang lebih rendah. Hal ini menunjukkan bahwa metode tersebut belum mampu mencapai keseimbangan yang baik antara *true positive rate* yang tinggi dan *false positive rate* yang rendah. Dengan kata lain, model masih

menghasilkan prediksi yang kurang akurat dalam membedakan pasangan *node* yang benar-benar memiliki potensi keterhubungan dengan yang tidak.

Kondisi ini mengindikasikan bahwa metode berbasis kesamaan lokal kurang mampu menangkap pola keterhubungan yang kompleks dalam jaringan, karena hanya mengandalkan informasi langsung seperti derajat atau kekuatan *node*. Akibatnya, hubungan tidak langsung yang sebenarnya penting dalam jaringan kolaborasi ilmiah menjadi tidak teridentifikasi dengan baik.

4.5. Uji Signifikansi Statistik

Untuk mengetahui apakah perbedaan performa antar metode bersifat signifikan secara statistik, dilakukan uji *paired sample t-test* menggunakan nilai AUC pada sepuluh iterasi eksperimen. Uji ini digunakan karena setiap metode dievaluasi pada kondisi eksperimen yang sama, sehingga memungkinkan perbandingan yang lebih adil antar metode. Selain itu, uji ini juga membantu dalam mengidentifikasi apakah perbedaan performa yang diamati benar-benar disebabkan oleh perbedaan metode yang digunakan, atau hanya merupakan hasil dari variasi data pada setiap iterasi.

Pengujian dilakukan dengan tingkat signifikansi $\alpha = 0,05$, di mana nilai *p-value* yang lebih kecil dari α menunjukkan adanya perbedaan yang signifikan secara statistik antara dua metode yang dibandingkan [16]. Hasil pengujian *paired sample t-test* ditunjukkan pada Tabel 6.

Tabel 6. Hasil uji *paired sample t-test*

Pasangan Metode	t-value	p-value	Keputusan
PA vs WPA	-3.196	0.011	Signifikan
Katz vs WKatz	1.173	0.271	Tidak signifikan

Berdasarkan Tabel 6, perbandingan antara metode PA dan WPA menghasilkan nilai *p-value* sebesar 0,011, yang lebih kecil dari tingkat signifikansi $\alpha = 0,05$. Hal ini menunjukkan bahwa perbedaan performa antara kedua metode tersebut bersifat signifikan secara statistik. Dengan demikian, penggunaan bobot pada metode WPA memberikan peningkatan performa yang nyata dibandingkan metode PA dalam memprediksi keterhubungan.

Sebaliknya, perbandingan antara metode Katz dan WKatz menghasilkan nilai *p-value* sebesar 0,271, yang lebih besar dari tingkat signifikansi $\alpha = 0,05$. Hal ini menunjukkan bahwa perbedaan performa antara kedua metode tersebut tidak signifikan secara statistik. Meskipun Katz memiliki nilai AUC yang sedikit lebih tinggi, peningkatan tersebut tidak cukup kuat untuk dianggap berbeda secara signifikan.

Hasil ini mengindikasikan bahwa penambahan bobot hubungan memberikan dampak yang lebih terlihat pada metode berbasis kesamaan lokal, namun tidak memberikan pengaruh yang signifikan pada

metode berbasis jalur global. Secara keseluruhan, metode Katz tetap menunjukkan performa yang stabil dan unggul dalam memprediksi keterhubungan pada jaringan *co-authorship*.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil eksperimen pada jaringan *co-authorship* berbobot menggunakan pendekatan *time-based link prediction*, metode berbasis jalur global yaitu Katz Index dan Weighted Katz Index menunjukkan performa yang lebih baik dibandingkan metode berbasis kesamaan lokal seperti Preferential Attachment (PA) dan Weighted Preferential Attachment (WPA). Katz Index memperoleh nilai AUC rata-rata tertinggi sebesar 0,96560 dengan parameter $\beta = 0,01$, diikuti oleh Weighted Katz Index dengan nilai AUC sebesar 0,95376, sedangkan WPA dan PA masing-masing hanya mencapai 0,67870 dan 0,64363. Hasil ini menunjukkan bahwa metode berbasis jalur global lebih mampu menangkap hubungan tidak langsung antar penulis dibandingkan metode berbasis lokal. Selain itu, analisis statistik menunjukkan bahwa perbedaan performa antara PA dan WPA signifikan dengan nilai *p-value* sebesar 0,011, sedangkan perbedaan antara Katz dan WKatz tidak signifikan dengan *p-value* sebesar 0,271. Temuan ini mengindikasikan bahwa struktur topologi jaringan memiliki pengaruh yang lebih dominan dibandingkan intensitas kolaborasi dalam memprediksi keterhubungan. Secara keseluruhan, metode berbasis jalur global terbukti lebih efektif dalam memprediksi kolaborasi pada jaringan ilmiah yang kompleks. Untuk penelitian selanjutnya, dapat dilakukan pengembangan dengan mengintegrasikan metode berbasis pembelajaran mesin atau memanfaatkan informasi tambahan seperti topik penelitian, afiliasi institusi, atau kesamaan semantik publikasi guna meningkatkan akurasi prediksi keterhubungan.

DAFTAR PUSTAKA

- [1] H. Yuliansyah, Z. A. Othman, and A. A. Bakar, "Co-authorship prediction method based on degree of gravity and article keywords similarity," *Phys. A Stat. Mech. its Appl.*, vol. 665, no. May 2024, p. 130511, 2025, doi: 10.1016/j.physa.2025.130511.
- [2] A. Abbasi, J. Altmann, and L. Hossain, "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures," *J. Informetr.*, vol. 5, no. 4, pp. 594–607, 2011, doi: 10.1016/j.joi.2011.05.007.
- [3] H. Yuliansyah, Z. A. Othman, and A. A. Bakar, "Extending adamic adar for cold-start problem in link prediction based on network metrics," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, pp. 271–284, 2022, doi: 10.26555/ijain.v8i3.882.

- [4] T. Zhou, "Progresses and challenges in link prediction," *iScience*, vol. 24, no. 11, p. 103217, 2021, doi: 10.1016/j.isci.2021.103217.
- [5] H. A. Hasin and D. Hassan, "Link Prediction in Co-authorship Networks," *Sci. J. Univ. Zakho*, vol. 10, no. 4, pp. 235–257, 2022, doi: 10.25271/sjuoz.2022.10.4.1040.
- [6] H. Wu, C. Song, Y. Ge, and T. Ge, "Link Prediction on Complex Networks: An Experimental Survey," *Data Sci. Eng.*, vol. 7, no. 3, pp. 253–278, 2022, doi: 10.1007/s41019-022-00188-2.
- [7] L. L. Linyuan and T. Zhou, "Link prediction in complex networks: A survey," *Phys. A Stat. Mech. its Appl.*, vol. 390, no. 6, pp. 1150–1170, 2011, doi: 10.1016/j.physa.2010.11.027.
- [8] G. G. Piva, F. L. Ribeiro, and A. S. Mata, "Networks with growth and preferential attachment: Modelling and applications," *J. Complex Networks*, vol. 9, no. 1, pp. 1–23, 2021, doi: 10.1093/comnet/cnab008.
- [9] M.-S. Vidza, M. Budka, W. K. Chai, M. Thrush, and M. T. Alves, "Expanding the Katz Index for Link Prediction: A Case Study on a Live Fish Movement Network," pp. 1–14, 2024, [Online]. Available: <http://arxiv.org/abs/2404.12871>
- [10] R. D. P and L. R. Deepthi, "Analysis of Link Prediction Methods in Weighted and Unweighted Citation Network," *2022 Int. Conf. Connect. Syst. Intell.*, pp. 1–6, 2022, doi: 10.1109/CSI54720.2022.9924105.
- [11] M. Liu, Y. Wang, J. Chen, and Y. Zhang, "Link Prediction Model for Weighted Networks Based on Evidence Theory and the Influence of Common Neighbours," *Complexity*, vol. 2022, 2022, doi: 10.1155/2022/9151340.
- [12] L. Yang, D. Wei, and C. Xiang, "Link Prediction for Weighted Collaboration Networks Based on Nodal Attributes." SSRN (Social Science Research Network), 2025. doi: 10.2139/ssrn.5147849.
- [13] O. Rainio, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.
- [14] I. Safder *et al.*, "Parsing AUC Result-Figures in Machine Learning Specific Scholarly Documents for Semantically-enriched Summarization," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2021.2004347.
- [15] C. Pu, J. Li, J. Wang, and T. Q. S. Quek, "The Node-Similarity Distribution of Complex Networks and Its Applications in Link Prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 4011–4023, 2022, doi: 10.1109/TKDE.2020.3026311.
- [16] D. Diez, *OpenIntro Statistics Fourth Edition*, Fourth Edi. OpenIntro, 2019.