



Paper Type: Original Article

Temporal Triadic–Homophily Link Prediction in Fuzzy Citation Networks

Sanchita Guchhait*

Department of Computer and Information Science, Raiganj University, Raiganj, India;
mou.guchhait@gmail.com.

Citation:

Received: 07 February 2025
Revised: 16 June 2025
Accepted: 19 September 2024

Guchhait, S. (2025). Temporal triadic-homophily link prediction in fuzzy citation networks. *Transactions on soft computing*, 1(4), 257-264.

Abstract

Triadic closure and topical homophily provide complementary evidence for new links in citation and collaboration networks. A temporal fuzzy link score is proposed to combine them without converting either source into a crisp preliminary decision. Time-decayed common-neighbor evidence, cosine content homophily, and endpoint activity are aggregated through a weighted complement-geometric operator. The score is bounded, monotone, reducible to each single component, and stable under bounded feature errors. A closed-form condition for retaining a pairwise ranking is obtained. The original chapter results on Cora and CiteSeer are re-examined: the reported Triadic Homophily Clustering AUC values are 0.70 and 0.69, while fixed-threshold precision values are 0.97 and 0.95. These figures are retained as source baselines rather than being relabeled as results of the new temporal model. A four-candidate calculation shows that the highest raw triadic candidate need not rank first once homophily, recency, and uncertainty are considered. The method supports time-respecting validation, validation-selected thresholds, and component-wise interpretation of each prediction.

Keywords: Link prediction, Citation network, Triadic closure, Homophily, Fuzzy evidence, Temporal decay.

1 | Introduction

Link prediction estimates which currently unconnected vertex pairs are likely to form an edge or which observed network entries are missing [1, 2, 3]. Local indices such as common neighbors, Jaccard similarity, Adamic–Adar, and resource allocation are attractive because

Corresponding Author: mou.guchhait@gmail.com

<https://doi.org/10.48314/tsc.v1i4.76>

Licensee System Analytics. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

they are interpretable and inexpensive [4, 5]. Their common premise is triadic closure: two vertices sharing neighbors have structural opportunities to connect [6].

Structural opportunity is not sufficient in attributed networks. Homophily—the tendency of similar actors or documents to connect—can strengthen or weaken the evidence carried by a shared neighbor [7, 8, 9]. Triadic closure and homophily can act cumulatively during network formation [10, 11]. In citation networks, topical similarity, publication time, and existing citation paths interact. A ten-year-old common neighbor and a recent common neighbor should not necessarily have equal effect. Likewise, a strong textual match without any structural opportunity should be distinguished from a pair supported by several recent triangles.

The supplied chapter introduces Triadic Homophily Clustering (THC), uses bag-of-words representations, and reports results on Cora and CiteSeer. Its central idea is useful, but three methodological improvements are needed. First, a fixed threshold of .4 should be selected from training and validation data rather than imposed in advance. Second, random 80:20 edge splitting can leak future information in a temporal problem. Third, triadic and homophily evidence should retain their graded values so uncertainty and rank sensitivity can be reported.

Homophily-based prediction has also been studied with linguistic Z-numbers, retaining uncertainty in the information used to assess possible links [12].

This paper develops a temporal fuzzy extension. It uses a time-decayed triadic score, a content-homophily score, and endpoint activity. A complement-geometric operator treats these as partial evidence sources and permits exact derivatives, perturbation bounds, and component-level interpretation.

2 | Temporal Fuzzy Citation Graph

Let $\mathcal{G}_t = (V_t, \sigma_t, \mu_t)$ be a directed temporal fuzzy graph observed up to time t . Vertex membership $\sigma_t(u)$ represents availability or confidence in the document/actor, while $\mu_t(u, v)$ represents confidence that an observed link from u to v is valid. For an unlinked candidate pair (u, v) , orientation is retained for citation prediction and suppressed for undirected collaboration prediction.

The time of vertex z is t_z , its observed degree at t is $d_t(z)$, and the common-neighbor set is

$$N_t(u, v) = N_t^+(u) \cap N_t^-(v)$$

for a two-step directed citation opportunity. Other applications may use $N_t(u) \cap N_t(v)$.

Time-decayed triadic evidence.

Each witness $z \in N_t(u, v)$ contributes

$$e_z(u, v; t) = \exp[-\lambda(t - t_z)] \frac{\min\{\mu_t(u, z), \mu_t(z, v)\}}{\log(2 + d_t(z))}. \quad (1)$$

The bottleneck grade requires both sides of the triangle, exponential decay values recent witnesses more highly, and the logarithmic divisor reduces the dominance of very high-degree witnesses. Aggregate evidence is

$$T_{uv}(t) = 1 - \exp \left[- \sum_{z \in N_t(u, v)} e_z(u, v; t) \right]. \quad (2)$$

This saturating form keeps the score in $[0, 1)$ while allowing several moderate witnesses to accumulate.

Content homophily.

Let $x_u, x_v \in \mathbb{R}_+^p$ be nonzero TF-IDF or contextual embeddings. The nonnegative cosine homophily is

$$H_{uv} = \frac{x_u^\top x_v}{\|x_u\|_2 \|x_v\|_2}. \quad (3)$$

When contextual embeddings have negative cosine values, they are mapped to $[0, 1]$ only after the normalization rule is stated. Bag-of-words can be retained for transparency, but a comparison against TF-IDF and contextual features is recommended.

Endpoint activity.

Let $n_u(\Delta)$ be the number of links incident to u in the most recent window $[t - \Delta, t]$. Define

$$A_{uv} = 1 - \exp\{-\xi[n_u(\Delta) + n_v(\Delta)]/2\}. \quad (4)$$

This component prevents an old structural pattern from being treated identically to an active developing pair.

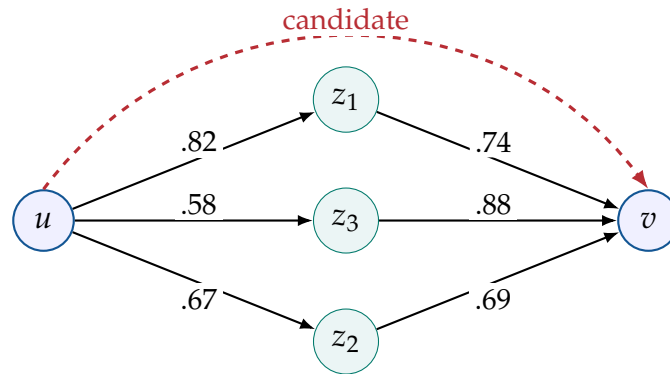


Figure 1. A candidate link supported by three graded, time-stamped triangular witnesses.

3 | Complement-Geometric Aggregation

Let $w_T, w_H, w_A \geq 0$ and $w_T + w_H + w_A = 1$. Define

$$S_{uv} = 1 - (1 - T_{uv})^{w_T} (1 - H_{uv})^{w_H} (1 - A_{uv})^{w_A}. \quad (5)$$

The operator is a weighted geometric mean of evidence complements followed by complementation. It behaves like a graded fuzzy union: a high component can support the prediction, but low values in the remaining components still affect the result. Unlike an unnormalized sum, the score cannot exceed one.

Proposition 3.1 (Basic properties). *The score in Eq. (5) is bounded in $[0, 1]$, continuous, nondecreasing in each component, and symmetric in any components having equal weights. If $w_T = 1$, then $S = T$; analogous reductions hold for H and A .*

Proof. Each complement belongs to $[0, 1]$. Their weighted product also belongs to $[0, 1]$. Differentiation gives

$$\frac{\partial S}{\partial X_k} = w_k \frac{1 - S}{1 - X_k} \geq 0, \quad (6)$$

for $X_k < 1$. The limiting and symmetry statements follow directly.

Equation (6) is also an explanation coefficient. The locally most influential component is the one with largest

$$I_k = w_k(1 - S)/(1 - X_k).$$

This is more informative than stating only that all three inputs were used.

4 | Uncertainty and Stability

Suppose the calculated evidence vector is $X = (T, H, A)$ and the admissible error is $|\tilde{X}_k - X_k| \leq \varepsilon_k$. If $X_k + \varepsilon_k \leq 1 - \rho$ for some $\rho \in [0, 1]$, then Eq. (6) gives the following result.

Theorem 4.1 (Score stability).

$$|\tilde{S} - S| \leq \frac{1}{\rho} \sum_{k \in \{T, H, A\}} w_k \varepsilon_k =: B. \quad (7)$$

Proof. Apply the multivariate mean-value theorem along the line between X and $\tilde{X}$. On the admissible box, $(1 - X_k)^{-1} \leq \rho^{-1}$ and $1 - S \leq 1$. Summing the coordinate changes proves the bound.

Corollary 4.2 (Pairwise rank stability condition). *For candidates p and q with bounds B_p, B_q , the order $S_p \geq S_q$ is preserved whenever*

$$S_p - S_q \geq B_p + B_q.$$

When component intervals are not small, monotonicity yields an exact interval evaluation without calculus:

$$S^- = S(T^-, H^-, A^-), \quad S^+ = S(T^+, H^+, A^+).$$

Candidate p certainly outranks q if $S_p^- \geq S_q^+$.

5 | Temporal Estimation Algorithm

Candidate generation begins with unlinked pairs having at least one recent two-step path. Witness terms in Eq. (1) are accumulated, content vectors are compared, and activity is calculated over a preselected window. Weights and decision threshold are chosen only on training/validation time windows.

For m observed edges and c candidates, wedge enumeration costs $O(\sum_z d_z^- d_z^+)$ in the directed case. Sparse-vector cosine computation costs $O(\text{nnz}(x_u) + \text{nnz}(x_v))$ per candidate. Candidate scores and explanations are then linear in c . Time-decay values can be updated recursively between snapshots.

1. Split by time: train before t_1 , validate on $[t_1, t_2)$, test on $[t_2, t_3)$.

2. Generate an equal or stated ratio of negative candidates from pairs observable at each time.
3. Tune $\lambda, \xi, w_T, w_H, w_A$ and the operating threshold on validation data.
4. Report AUC and average precision together with Brier score and threshold metrics.
5. Recompute with feature intervals and report stable pairwise ranks or ambiguous sets.

The protocol avoids using citations that occur after the prediction date. This point is essential because random edge splits can present future neighborhoods to the model.

6 | Worked Candidate Ranking

Consider four candidate pairs with their evidence values shown in Table 1. Use $(w_T, w_H, w_A) = (.45, .35, .20)$.

Table 1. Component values and calculated link scores.

Candidate	T	H	A	S	Rank
uv	.72	.81	.60	.7375	2
xy	.83	.52	.70	.7261	3
rs	.55	.90	.45	.7233	4
pq	.76	.74	.80	.7620	1

For uv ,

$$\begin{aligned}
 S_{uv} &= 1 - (.28)^{.45} (.19)^{.35} (.40)^{.20} \\
 &= .73746.
 \end{aligned}$$

The pair xy has the strongest triadic value but only moderate homophily, so it does not rank first. Pair rs is highly homophilic but has weaker structural and activity evidence. Pair pq is not maximal in either T or H ; it leads because it is consistently strong across all three sources.

If each component of pq has error at most .015, then its exact monotone interval is obtained by evaluating $(.745, .725, .785)$ and $(.775, .755, .815)$ in Eq. (5). The same calculation for uv determines whether the observed difference .0246 remains positive under the stated errors. Reporting this interval is preferable to rounding both scores to two decimals and declaring a certain order.

7 | Source Results and Methodological Comparison

The supplied chapter reports precision and accuracy at a fixed threshold .4 as shown in Table 2. It also supplies the AUC comparison in Table 3. A malformed Jaccard entry is interpreted as 0.50; no other values are altered.

The Cora table does not support a claim that THC outperforms every listed method, because Adamic–Adar has AUC .71 versus .70. On CiteSeer, THC is highest among the displayed values. These are honest, dataset-specific conclusions. The new temporal score has not been assigned invented AUC values; it requires the time-respecting experiment described above.

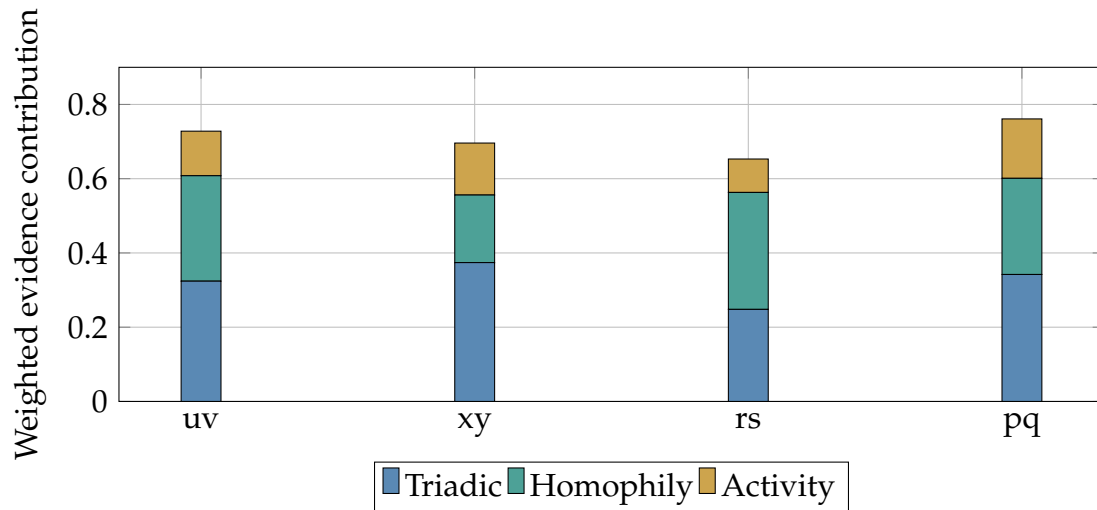


Figure 2. Weighted component evidence. The final nonlinear score is not the bar sum; the bars explain the relative support pattern.

Table 2. Fixed-threshold results reproduced from the supplied chapter.

Dataset	Precision	Accuracy
Cora	.97	.86
CiteSeer	.95	.83

Comparison with related methods.

Classical methods treat evidence differently. Common neighbors counts witnesses; Adamic-Adar discounts high-degree witnesses; Jaccard normalizes by neighborhood union; THC combines triadic closure with content; graph neural methods learn representations but may behave differently under heterophily [13]. The proposed method adds explicit time decay, fuzzy path confidence, endpoint activity, and interval stability.

At minimum, an applied report should contain:

- the temporal split dates and negative-sampling procedure;
- content representation and its preprocessing;
- component weights and the validation-selected threshold;
- AUC, average precision, precision, recall, Brier score, and class balance;
- score intervals or perturbation radii; and
- examples where triadic and homophily evidence disagree.

Homophily is not universally beneficial. Recent analyses show that feature heterophily changes which link-prediction architectures perform well [13], while adversarial studies demonstrate that homophily assumptions can also be exploited [14]. Therefore, w_H should be validated rather than fixed by convention.

Table 3. AUC values reproduced and notation-corrected from the supplied chapter.

Dataset	Common neighbors	Adamic-Adar	Jaccard	THC
Cora	.68	.71	.50	.70
CiteSeer	.66	.67	.60	.69

8| Conclusion

A temporal fuzzy triadic-homophily score has been developed for citation and collaboration link prediction. The method combines time-decayed two-step evidence, content similarity, and recent activity through a bounded complement-geometric operator. Monotonicity, derivative expressions, interval evaluation, and pairwise stability bounds show how the score changes. The source Cora and CiteSeer figures were retained and corrected without claiming that they validate the new extension. A chronological experiment with validation-selected thresholds, current embeddings, and interval-coverage results is required before comparative empirical claims can be made.

Author Contributions

The author developed the theoretical extension, numerical reexamination, and interpretation.

Funding

No external funding was reported.

Data Availability

Cora and CiteSeer are public benchmark datasets. Values reproduced from the supplied chapter are identified as such.

Competing Interests

The author declares no competing interests.

Generative-AI Assistance

Generative AI assisted with language, LaTeX preparation and technical representation of the paper.

References

- [1] Liben-Nowell, D., & Kleinberg, J. (2003). The link prediction problem for social networks. Proceedings of the twelfth international conference on information and knowledge management (pp. 556-559). ACM Digital Library. <https://doi.org/10.1145/956863.956972>
- [2] Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. Physica A: Statistical mechanics and its applications, 390(6), 1150-1170. <https://doi.org/10.1016/j.physa.2010.11.027>
- [3] Kumar, A., Singh, S. S., Singh, K., & Biswas, B. (2020). Link prediction techniques, applications, and performance: A survey. Physica A: Statistical mechanics and its applications, 553, 124289. <https://doi.org/10.1016/j.physa.2020.124289>

- [4] Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social networks*, 25(3), 211-230. [https://doi.org/10.1016/S0378-8733\(03\)00009-1](https://doi.org/10.1016/S0378-8733(03)00009-1)
- [5] Zhou, T., Lü, L., & Zhang, Y. C. (2009). Predicting missing links via local information. *The European physical journal B*, 71(4), 623-630. <https://doi.org/10.1140/epjb/e2009-00335-8>
- [6] Huang, H., Tang, J., Wu, S., Liu, L., & Fu, X. (2014). Mining triadic closure patterns in social networks. *Proceedings of the 23rd international conference on World wide web* (pp. 499-504). ACM Digital Library. <https://doi.org/10.1145/2567948.2576940>
- [7] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- [8] Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, 6(2), 1-33. <https://doi.org/10.1145/2180861.2180866>
- [9] Ciotti, V., Bonaventura, M., Nicosia, V., Panzarasa, P., & Latora, V. (2016). Homophily and missing links in citation networks. *EPJ data science*, 5(1), 7. <https://doi.org/10.1140/epjds/s13688-016-0068-2>
- [10] Asikainen, A., Iñiguez, G., Ureña-Carrión, J., Kaski, K., & Kivelä, M. (2020). Cumulative effects of triadic closure and homophily in social networks. *Science advances*, 6(19), eaax7310. <https://doi.org/10.1126/sciadv.aax7310>
- [11] Talaga, S., & Nowak, A. (2019). Homophily as a process generating social networks: Insights from social distance attachment model. <https://doi.org/10.48550/arXiv.1907.07055>
- [12] Guchhait, S., Sahoo, L., Samanta, S., Dubey, V. K., & Allahviranloo, T. (2024). Homophily-based link prediction within a social network using linguistic z-number. In *Management of uncertainty using linguistic z-numbers: Applications for decision-making, granular computing and social networks* (pp. 179-199). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65854-9_11
- [13] Zhu, J., Li, G., Yang, Y. A., Zhu, J., Cui, X., & Koutra, D. (2024). On the impact of feature heterophily on link prediction with graph neural networks. *Advances in neural information processing systems*, 37, 65823-65851. https://proceedings.neurips.cc/paper_files/paper/2024/hash/79353864175b6b8c3d073cde84d7014a-AbstractConference.html
- [14] Wang, Y., Zhang, Z., Qiao, P., Yuan, Y., & Wang, G. (2025). Homophily-guided backdoor attacks on GNN-based link prediction. *Applied sciences*, 15(17), 9651. <https://doi.org/10.3390/app15179651>