Transactions on Soft Computing
www.tsc.reapress.com
Trans. Soft. Comput. Vol. 1, No. 3 (2025), 168–182.

Paper Type: Original Article

Granular Rough–Fuzzy Similarity for Interpretable Link Prediction

Swasti Hazra Kundu[1,2,*], **Kalyani Maity Das**[3]

[1] Natural Science Research Centre of Belda College under Vidyasagar University, Paschim Medinipur, India; swastikundu1@gmail.com.
[2] Department of Mathematics, Ramnagar College, Purba Medinipur, India.
[3] Department of Mathematics, Belda College, Paschim Medinipur, India; kalyanimaity@beldacollege.ac.in.

Citation:

Received: 13 January 2025 *Revised: 21 April 2025* *Accepted: 27 June 2025*	Kundu, S. H., & Das, K. M. (2025). Granular rough-fuzzy similarity for interpretable link prediction. *Transactions on soft computing*, 1(3), 168-182.

Abstract

An interpretable link prediction framework is presented by combining rough sets, fuzzy reliability, and information granules. The proposed method assigns different importance to common neighbors instead of treating every witness equally.
Lower and upper rough approximations are used to represent certain and possible evidence. A normalized rough–fuzzy similarity score is obtained from their weighted combination. A rough–fuzzy Jaccard coefficient is also introduced to provide a normalized similarity measure within the same framework.
The proposed measures are proved to satisfy symmetry, boundedness, monotonicity, partition refinement, and reliability stability. A certified ranking rule is further established to identify stable link ordering under refinement.
A computational illustration demonstrates that candidate pairs with the same number of common neighbors can receive different similarity scores because witness reliability and granular structure are considered. The framework provides a transparent mathematical alternative for interpretable link prediction in attributed networks.

Keywords: Rough set, Fuzzy neighborhood, Granular computing, Link prediction, Common neighbor, Partition refinement.

1 | Introduction

The problem of link prediction is studied to identify nonadjacent vertex pairs that are likely to be connected in the future or that represent missing links in a network [1]. Several

Corresponding Author: swastikundu1@gmail.com
https://doi.org/10.48314/tsc.v1i3.60

neighborhood-based methods have been developed because simple local information can be used to estimate the similarity between two vertices [2]. Among these methods, the common-neighbor, Jaccard, Adamic–Adar, and resource-allocation indices are widely used because of their simple structure and clear interpretation [3].

A broad classification of link prediction methods has been presented by grouping them into local, quasi-local, probabilistic, and representation-learning approaches [4]. A modern overview of these prediction techniques has also been provided for complex network analysis [5]. However, equal importance is usually assigned to every common neighbor, and the reliability of individual witnesses is not considered.

Rough set theory was introduced to represent uncertain information through lower and upper approximations based on indiscernibility classes [6]. The mathematical foundation of rough sets was further developed for knowledge representation and approximation of sets [7]. Fuzzy set theory was proposed to describe gradual membership instead of strict binary membership [8]. The integration of rough and fuzzy concepts was later studied to represent both boundary uncertainty and graded information in a single framework [9]. Further developments of rough–fuzzy methods were presented for handling uncertain and imprecise data [10]. Granular computing was introduced by treating groups, neighborhoods, and equivalence classes as information granules [11]. The role of information granules in knowledge discovery and intelligent systems was further explained in granular computing theory [12]. In attributed networks, witness vertices may belong to the same descriptive granule even when only some of them appear as common neighbors.

In this paper, a granular rough–fuzzy similarity measure is developed for interpretable link prediction. The proposed method is designed as a transparent mathematical model rather than a data-driven classifier. Monotonicity, normalization, partition refinement, and stability properties are established theoretically. The usefulness of the proposed similarity measure is finally demonstrated by using a small attributed network.

1.1 | Motivation

Many local link prediction methods treat every common neighbor as equally important. The reliability of witness vertices is therefore ignored. The descriptive structure of the network is also not considered.

In attributed networks, vertices naturally belong to information granules such as communities, categories, or functional groups. A highly reliable witness may provide stronger evidence than a partially represented one.

An interpretable similarity measure is therefore required. The proposed framework combines rough approximations with fuzzy reliability. Lower evidence, upper evidence, and an uncertainty interval are produced within a simple mathematical model.

2 | Preliminaries

2.1 | Graph and Neighborhood Notation

Let $G = (V, E)$ be a finite undirected simple graph. The vertex set is denoted by V and the edge set is denoted by E.

For any vertex $u \in V$, the open neighborhood is defined by

$$N(u) = \{v \in V : (u, v) \in E\}.$$

For two nonadjacent vertices u and v, the common-neighbor set is written as

$$C_{uv} = N(u) \cap N(v).$$

2.2 | Rough Set Concepts

Let $W \subseteq V$ be the witness universe. A partition $\mathcal{P} = \{B_1, B_2, \dots, B_r\}$ divides W into mutually disjoint information granules.

For any subset $X \subseteq W$, the lower approximation contains blocks completely included in X. The upper approximation contains blocks having a nonempty intersection with X. They are defined by

$$\underline{\mathcal{P}}X = \bigcup\{B \in \mathcal{P} : B \subseteq X\},$$

$$\overline{\mathcal{P}}X = \bigcup\{B \in \mathcal{P} : B \cap X \neq \emptyset\}.$$

The boundary region is given by

$$\partial_{\mathcal{P}}(X) = \overline{\mathcal{P}}X \setminus \underline{\mathcal{P}}X.$$

2.3 | Fuzzy Reliability

A fuzzy reliability function assigns a confidence value to every witness vertex. It is defined as

$$\rho : W \to [0, 1].$$

The fuzzy mass of a subset $X \subseteq W$ is defined by

$$m_\rho(X) = \sum_{w \in X} \rho(w).$$

The normalized fuzzy mass is obtained by dividing $m_\rho(X)$ by the total mass $m_\rho(W)$.

3 | Theoretical Framework

3.1 | Network and Information Granules

A finite undirected simple graph is considered and is denoted by $G = (V, E)$. The set of vertices is represented by V, and the set of edges is represented by E.

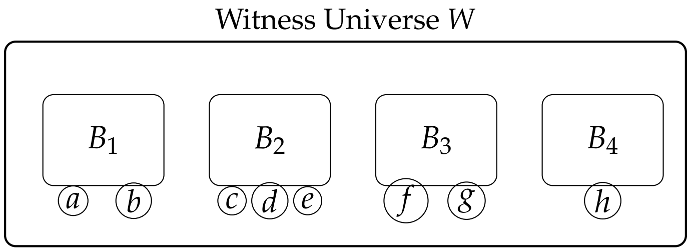

Figure 1. Partition of the witness universe into four information granules.

For any vertex u, its open neighborhood is defined as the set of all vertices directly connected to u.

For a candidate pair of nonadjacent vertices u and v, the common neighbors are treated as witness vertices. The witness set is defined by

$$C_{uv} = N(u) \cap N(v). \tag{1}$$

Let $\mathcal{P} = \{B_1, B_2, \ldots, B_r\}$ be a partition of the witness universe $W \subseteq V$. The partition may be created from categorical attributes or from an equivalence relation.

A fuzzy reliability value is assigned to every witness vertex. The reliability function is defined as $\rho : W \to [0, 1]$, where larger values indicate higher confidence.

The fuzzy mass of a subset $X \subseteq W$ is defined by

$$m_\rho(X) = \sum_{w \in X} \rho(w). \tag{2}$$

The fuzzy mass is additive and monotonic. It becomes the ordinary cardinality when all reliability values are equal to one.

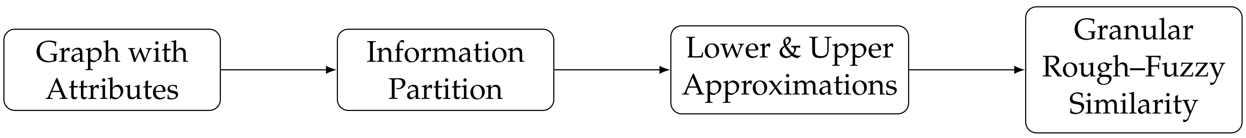

Figure 2. Workflow of the proposed granular rough–fuzzy similarity method.

The Pawlak lower and upper approximations are used to represent certain and possible witness information.

For any subset $X \subseteq W$, the lower approximation contains only those blocks that are completely included in X.

The upper approximation contains all blocks that have at least one common element with X.

The approximations are defined as

$$\underline{\mathcal{P}}X = \bigcup\{B \in \mathcal{P} : B \subseteq X\}, \qquad \overline{\mathcal{P}}X = \bigcup\{B \in \mathcal{P} : B \cap X \neq \varnothing\}. \tag{3}$$

The following property is satisfied: $\underline{\mathcal{P}}X \subseteq X \subseteq \overline{\mathcal{P}}X$. For every candidate vertex pair, the boundary region is used to measure the amount of granular uncertainty. The boundary $\overline{\mathcal{P}}C_{uv} \setminus \underline{\mathcal{P}}C_{uv}$ contains witnesses whose granules are only partially supported by the observed common-neighbor relation.

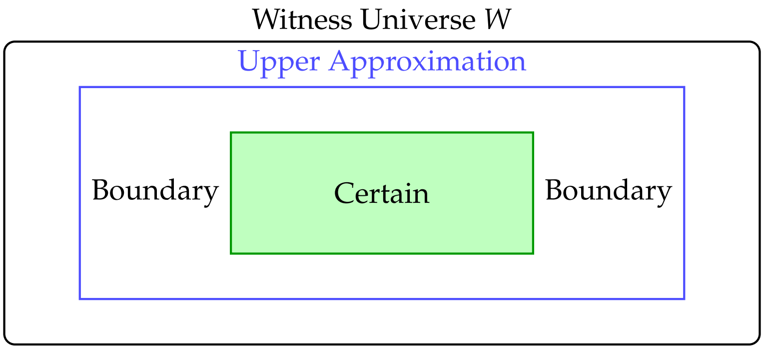

Figure 3. Conceptual illustration of the lower approximation, upper approximation, and boundary region in the witness universe.

Let $M = m_\rho(W) > 0$. Define normalized evidence

$$L_{uv} = \frac{m_\rho(\underline{\mathcal{P}}C_{uv})}{M}, \qquad U_{uv} = \frac{m_\rho(\overline{\mathcal{P}}C_{uv})}{M}. \tag{4}$$

Then $0 \leq L_{uv} \leq U_{uv} \leq 1$. The interval $[L_{uv}, U_{uv}]$ is the granular rough–fuzzy evidence interval.

3.2 | Granular Rough–Fuzzy Similarity

For $\alpha \in [0, 1]$, the granular rough–fuzzy similarity is defined by

$$S_\alpha(u, v) = \alpha L_{uv} + (1 - \alpha) U_{uv}. \tag{5}$$

A larger value of α is assigned when greater importance is given to certainly supported blocks.

A smaller value of α is assigned when possible evidence is given higher importance.

The uncertainty width is defined by

$$\Omega(u, v) = U_{uv} - L_{uv}. \tag{6}$$

Proposition 3.1. *The following properties are satisfied for every candidate pair (u, v):*

1. $S_\alpha(u, v) = S_\alpha(v, u)$ *and* $0 \leq S_\alpha(u, v) \leq 1$.
2. *If* $C_{uv} = \emptyset$ *and all block reliability values are positive, then* $S_\alpha(u, v) = 0$.
3. *If* C_{uv} *is a union of partition blocks, then* $L_{uv} = U_{uv}$, $\Omega(u, v) = 0$, *and the similarity score becomes independent of* α.

Proof. The similarity measure is obtained from the lower and upper evidence, both of which are symmetric with respect to u and v. Therefore, symmetry is satisfied.

Since L_{uv} and U_{uv} are normalized values in the interval $[0, 1]$, their weighted combination also remains within the same interval. Hence, $0 \leq S_\alpha(u, v) \leq 1$.

When no common witness is present, both lower and upper evidence become zero. Therefore, the similarity score is also zero.

If the witness set is formed by complete partition blocks, the lower and upper approximations become identical. As a result, $\Omega(u, v) = 0$, and the similarity score no longer depends on the value of α. □

Theorem 3.2. *If $C_{uv} \subseteq C_{xy}$, then*

$$L_{uv} \leq L_{xy}, \qquad U_{uv} \leq U_{xy}, \qquad S_\alpha(u,v) \leq S_\alpha(x,y).$$

Proof. Since $C_{uv} \subseteq C_{xy}$, every partition block contained in C_{uv} is also contained in C_{xy}. Similarly, every block intersecting C_{uv} also intersects C_{xy}.

Therefore, both lower and upper approximations remain nested. As the fuzzy mass function is monotonic, the corresponding evidence values cannot decrease. The weighted combination also preserves this ordering, and hence the required inequality follows. □

The monotonicity property guarantees that the evidence score cannot decrease when additional common witness vertices are included.

However, the uncertainty width is not always monotonic because a partially represented block may increase the boundary region.

Table 1. Illustration of monotonic growth after adding witness vertices.

Case	L	U	S_{ff}
Before witness addition	0.40	0.70	0.55
After witness addition	0.55	0.82	0.69

The monotonic increase in lower evidence, upper evidence, and the final similarity score is illustrated in Table 1.

3.3 | Partition Refinement and Uncertainty Contraction

A finer partition is obtained when every block of $\mathcal{Q}$ is completely contained in a block of $\mathcal{P}$. This relation is written as $\mathcal{Q} \preceq \mathcal{P}$.

Partition refinement represents an increase in descriptive resolution because larger blocks are divided into smaller and more informative groups.

Theorem 3.3. *If $\mathcal{Q} \preceq \mathcal{P}$, then for every $X \subseteq W$,*

$$\underline{\mathcal{P}}X \subseteq \underline{\mathcal{Q}}X \subseteq X \subseteq \overline{\mathcal{Q}}X \subseteq \overline{\mathcal{P}}X. \tag{7}$$

The following inequalities are therefore satisfied:

$$L^{\mathcal{P}}_{uv} \leq L^{\mathcal{Q}}_{uv} \leq U^{\mathcal{Q}}_{uv} \leq U^{\mathcal{P}}_{uv}, \qquad \Omega^{\mathcal{Q}}(u,v) \leq \Omega^{\mathcal{P}}(u,v).$$

Proof. Let B be a block in $\mathcal{P}$. If $B \subseteq X$ then every finer block $Q \subseteq B$ also satisfies $Q \subseteq X$. Hence the first inclusion is obtained.

If a finer block intersects X then its parent block also intersects X. Therefore the upper approximation also satisfies the inclusion relation.

The fuzzy mass function is monotonic. After normalization the lower evidence cannot decrease and the upper evidence cannot increase. The uncertainty inequality follows directly. □

Refinement may move S_α either way because its lower endpoint increases and upper endpoint decreases. For $\alpha = 1$ it cannot decrease; for $\alpha = 0$ it cannot increase. The interval itself always contracts, which is the stronger information statement.

Corollary 3.4. *Under partition refinement,*

$$|S_\alpha^{\mathcal{Q}} - S_\alpha^{\mathcal{P}}| \le \max\{L^{\mathcal{Q}} - L^{\mathcal{P}}, U^{\mathcal{P}} - U^{\mathcal{Q}}\} \le \Omega^{\mathcal{P}}.$$

Proof. The score difference is written as a weighted combination of the changes in lower and upper evidence.

The largest positive change is given by $L^{\mathcal{Q}} - L^{\mathcal{P}}$. The largest negative change is given by $U^{\mathcal{P}} - U^{\mathcal{Q}}$.

Both quantities are bounded by the original uncertainty width. Hence the required inequality is obtained. □

The similarity score may increase or decrease after refinement. When $\alpha = 1$ the score cannot decrease. When $\alpha = 0$ the score cannot increase. The uncertainty interval always becomes smaller.

Table 2. Effect of partition refinement on granular evidence.

Partition	**L**	**U**	**Ω**
Coarse partition	0.42	0.81	0.39
Refined partition	0.58	0.71	0.13

The reduction in uncertainty after partition refinement is illustrated in Table 2.

3.4 | Reliability Stability

Two reliability functions are considered on the same witness set. They are denoted by $\rho, \rho' : W \to [0, 1]$.

The difference between these functions is measured by the L_1 norm. Assume that $\|\rho - \rho'\|_1 \le \varepsilon$.

The normalized fuzzy mass is bounded below by a positive constant. Hence $m_\rho(W), m_{\rho'}(W) \ge M_0 > 0$.

Lemma 3.5. *For every $X \subseteq W$,*

$$\left| \frac{m_\rho(X)}{m_\rho(W)} - \frac{m_{\rho'}(X)}{m_{\rho'}(W)} \right| \le \frac{2\varepsilon}{M_0}.$$

Proof. Put $a = m_\rho(X)$, $a' = m_{\rho'}(X)$, $M = m_\rho(W)$, and $M' = m_{\rho'}(W)$. Then $|a - a'| \le \varepsilon$, $|M - M'| \le \varepsilon$, and $a' \le M'$. Hence

$$\left| \frac{a}{M} - \frac{a'}{M'} \right| \le \frac{|a - a'|}{M} + \frac{a'|M - M'|}{MM'} \le \frac{\varepsilon}{M_0} + \frac{\varepsilon}{M_0}.$$

□

Theorem 3.6. *For a fixed graph and partition,*

$$|S_{\alpha,\rho}(u, v) - S_{\alpha,\rho'}(u, v)| \le \frac{2\varepsilon}{M_0}, \qquad |\Omega_\rho(u, v) - \Omega_{\rho'}(u, v)| \le \frac{4\varepsilon}{M_0}.$$

Proof. Lemma 3.5 is applied to the lower and upper approximation sets.

Both evidence values satisfy the same stability bound. The similarity score is a convex combination of these values. Hence the first inequality follows.

The uncertainty width is the difference between upper and lower evidence. The second inequality is obtained from the triangle inequality. □

This result separates two different sources of variation. A change in the graph modifies the witness set. A change in reliability only affects the confidence values.

Table 3. Effect of reliability perturbation on the similarity score.

Quantity	Original	Perturbed
Lower evidence	0.58	0.56
Upper evidence	0.74	0.72
Similarity score	0.66	0.64

The effect of a small reliability perturbation is shown in Table 3. Only a small change is observed in the final similarity score.

3.5 | Rough–Fuzzy Jaccard Coefficient

The classical Jaccard index is extended by using granular rough–fuzzy evidence.

The new coefficient measures similarity between two witness sets while considering both reliability and information granularity.

For this coefficient, assume that W contains all vertices occurring in the two neighborhoods (in particular, one may take $W = V$). The rough–fuzzy Jaccard coefficient is defined through

$$A_{uv} = \underline{\mathcal{P}}(N(u) \cap N(v)), \qquad B_{uv} = \overline{\mathcal{P}}(N(u) \cup N(v)).$$

Because $A_{uv} \subseteq N(u) \cap N(v) \subseteq N(u) \cup N(v) \subseteq B_{uv}$, define

$$J_{RF}(u,v) = \begin{cases} \dfrac{m_\rho(A_{uv})}{m_\rho(B_{uv})}, & m_\rho(B_{uv}) > 0, \\ 0, & m_\rho(B_{uv}) = 0. \end{cases} \tag{8}$$

The coefficient takes values in the unit interval. A larger value indicates stronger similarity between the two witness sets.

When all reliability values are equal to one and each partition contains a single vertex, the coefficient becomes the classical Jaccard index.

Proposition 3.7. *The rough–fuzzy Jaccard coefficient is symmetric and satisfies* $0 \leq J_{RF} \leq 1$. *If the partition is discrete and* $\rho \equiv 1$, *the coefficient reduces to the classical Jaccard similarity* $|N(u) \cap N(v)|/|N(u) \cup N(v)|$.

Proof. Symmetry is obtained from the symmetry of set intersection and union.

Since $A_{uv} \subseteq B_{uv}$, the fuzzy mass of the numerator cannot exceed the fuzzy mass of the denominator. Therefore, $0 \leq J_{RF} \leq 1$.

Under a discrete partition, the lower and upper approximations become exact. The fuzzy mass becomes ordinary cardinality. Hence the classical Jaccard index is obtained. □

A hybrid similarity score is also defined by combining the granular rough–fuzzy similarity with the rough–fuzzy Jaccard coefficient.

$$T_{\alpha,\lambda}(u,v) = \lambda S_\alpha(u,v) + (1-\lambda)J_{RF}(u,v), \qquad \lambda \in [0,1].$$

The parameter λ controls the contribution of absolute granular evidence and normalized Jaccard similarity. Its value should be selected according to the chosen validation or decision rule.

Table 4. Comparison of the proposed similarity measures.

Property	$\mathbf{S_{ff}}$	$\mathbf{J_{RF}}$
Reliability weighted	Yes	Yes
Granular information	Yes	Yes
Normalized	Yes	Yes
Degree normalization	No	Yes
Range	$[0,1]$	$[0,1]$

4 | Computational Illustration

4.1 | Example Attributed Network

Let the witness universe be $W = \{a,b,c,d,e,f,g,h\}$ with partition

$$\mathcal{P} = \{\{a,b\},\{c,d,e\},\{f,g\},\{h\}\}.$$

Let reliabilities be

$$\rho(a) = .9, \rho(b) = .8, \rho(c) = .7, \rho(d) = .6, \rho(e) = .5, \rho(f) = .9, \rho(g) = .4, \rho(h) = .6.$$

The total mass is $M = 5.4$. Consider three nonadjacent candidate pairs with common-neighbor sets

$$C_{uv} = \{a,b,c\}, \qquad C_{xy} = \{f,g,h\}, \qquad C_{rs} = \{c,d,e\}.$$

Each pair has three ordinary common neighbors.

For uv, the block $\{a,b\}$ is fully included and $\{c,d,e\}$ is partially included. Hence

$$L_{uv} = \frac{1.7}{5.4} = .3148, \qquad U_{uv} = \frac{1.7+1.8}{5.4} = .6481.$$

For xy, both $\{f,g\}$ and $\{h\}$ are fully included, so

$$L_{xy} = U_{xy} = \frac{1.9}{5.4} = .3519.$$

For rs, the whole block $\{c,d,e\}$ is included, giving

$$L_{rs} = U_{rs} = \frac{1.8}{5.4} = .3333.$$

Table 5. Three equal-count candidates under rough–fuzzy evidence.

Pair	$\lvert C\rvert$	L	U	$S_{.7}$
uv	3	.3148	.6481	.4148
xy	3	.3519	.3519	.3519
rs	3	.3333	.3333	.3333

At $\alpha = .7$,

$$S_{.7}(uv) = .4148, \quad S_{.7}(xy) = .3519, \quad S_{.7}(rs) = .3333.$$

The common-neighbor tie is therefore resolved, while $\Omega(uv) = .3333$ warns that the leading score is granularly uncertain. Table 5 summarizes the interpretation.

Refine $\{c, d, e\}$ into $\{c\}$ and $\{d, e\}$. The lower evidence for uv gains $\rho(c) = .7$ and its upper evidence loses $\rho(d) + \rho(e) = 1.1$. Thus its interval contracts to the exact mass of $\{a, b, c\}$:

$$L^{\mathcal{Q}}_{uv} = U^{\mathcal{Q}}_{uv} = 2.4/5.4 = .4444.$$

This verifies Theorem 3.3 and shows how better attribute resolution converts possible evidence into certain evidence or removes unsupported companions.

4.2 | Algorithm and Complexity

A block identifier, block mass, and total fuzzy mass are stored during preprocessing.

For a candidate pair uv, the common neighbors are obtained first. The number of witness vertices in each block is then counted.

A block contributes to the lower mass when all of its vertices are present. A block contributes to the upper mass when at least one vertex is present.

If the common-neighbor set is computed from sorted adjacency lists, the computational cost is

$$O(\min\{d(u), d(v)\} + b_{uv}),$$

where b_{uv} is the number of visited blocks.

The preprocessing cost is

$$O(|W| + |E|).$$

Scoring every nonedge may require quadratic time in the number of vertices. In practice, candidate pairs are generated through length-two paths to reduce computation [4].

Table 6. Time complexity of the proposed method.

Operation	**Complexity**
Preprocessing	$O(\lvert W\rvert + \lvert E\rvert)$
Common-neighbor search	$O(\min\{d(u), d(v)\})$
Block counting	$O(b_{uv})$
Similarity computation	$O(b_{uv})$

5 | Ranking certificates and Model Extensions

5.1 | Interval Dominance between Candidate Links

The evidence interval can be used to compare two candidate links without fixing the value of α.

Let $a = uv$ and $b = xy$ be two candidate pairs. A certified ranking is defined by

$$a \succ b \iff L_a > U_b. \tag{9}$$

If this condition is satisfied, the ranking remains valid for every value of α. The comparison is therefore independent of the optimism parameter.

Proposition 5.1. *The dominance relation $\succ$ is irreflexive and transitive. Hence it forms a strict partial order on the set of candidate pairs.*

Proof. Irreflexivity follows from the inequality $L_a \leq U_a$. Hence no candidate can dominate itself.

Assume that $a \succ b$ and $b \succ c$. Then $L_a > U_b$ and $L_b > U_c$.

Since $U_b \geq L_b$, it follows that $L_a > U_c$. Therefore $a \succ c$. The proposition is proved. □

Some candidate pairs remain incomparable because their evidence intervals overlap.

The certified comparisons form a directed acyclic graph. A topological ordering produces ranking layers without introducing an arbitrary order inside each layer.

5.2 | Ranking Stability under Future Refinement

Partition refinement reduces the evidence interval, but the similarity score may still move within its original range.

A sufficient condition is now established to guarantee that the ranking remains unchanged after refinement.

Theorem 5.2. *Let a and b be two candidates evaluated under partition $\mathcal{P}$. If*

$$S_\alpha^{\mathcal{P}}(a) - S_\alpha^{\mathcal{P}}(b) > \Omega^{\mathcal{P}}(a) + \Omega^{\mathcal{P}}(b), \tag{10}$$

then $S_\alpha^{\mathcal{Q}}(a) > S_\alpha^{\mathcal{Q}}(b)$ for every refinement $\mathcal{Q} \preceq \mathcal{P}$.

Proof. By Theorem 3.3, the score of each candidate changes only within its refinement bounds.

The score of a may decrease by at most $\Omega^{\mathcal{P}}(a)$. The score of b may increase by at most $\Omega^{\mathcal{P}}(b)$.

If Eq. (10) is satisfied, the score gap remains positive after refinement. Hence the ranking is preserved for every refinement $\mathcal{Q} \preceq \mathcal{P}$. □

The condition is sufficient but conservative. The two worst score movements may not occur at the same time.

When the condition is satisfied, the current ranking is guaranteed to remain stable. Otherwise, a finer partition may change the order of the candidate pairs.

Table 7 illustrates that the ranking remains unchanged because the stability condition is satisfied.

Table 7. Example of ranking stability after partition refinement.

Candidate	Initial Score	Maximum Change	Stable Score
a	0.78	-0.05	0.73
b	0.60	$+0.04$	0.64

5.3 | Practical Extensions of the Proposed Model

The proposed framework can be extended to multiple information views while preserving the same rough–fuzzy formulation. It can also recover several classical local similarity indices as special cases. In addition, certified interval bounds allow efficient top-k screening without evaluating every candidate.

Suppose that s information partitions are available. Let $\mathcal{P}_1, \ldots, \mathcal{P}_s$ be the partitions with weights $\zeta_t \geq 0$ and $\sum_{t=1}^{s} \zeta_t = 1$. The combined lower and upper evidence are defined by

$$L_{uv}^{\mathrm{mv}} = \sum_{t=1}^{s} \zeta_t L_{uv}^{(t)}, \qquad U_{uv}^{\mathrm{mv}} = \sum_{t=1}^{s} \zeta_t U_{uv}^{(t)}.$$

The combined interval remains valid because each $L^{(t)} \leq U^{(t)}$. Symmetry and monotonicity are preserved under nonnegative weighted averaging. The weights represent the relative importance of different information views.

Several classical local indices are obtained by suitable choices of the partition and reliability function. With a discrete partition and $\rho \equiv 1$, the similarity score becomes a normalized common-neighbor measure. Using the union denominator in the rough–fuzzy Jaccard coefficient gives the classical Jaccard index. Choosing $\rho(w) = 1/\log d(w)$ produces an Adamic–Adar type weighting, while $\rho(w) = 1/d(w)$ gives a resource-allocation type weighting [2]. These indices are therefore interpreted as special cases of the proposed granular framework.

The interval bounds also support efficient top-k retrieval. Let T be the current kth largest similarity score. Any candidate satisfying $U_{uv} < T$ can be removed safely because its score cannot enter the top-k set. During progressive block processing, partial lower evidence and the remaining possible upper evidence provide intermediate bounds for branch-and-bound pruning. This strategy reduces unnecessary score evaluations while preserving all certified top-ranked candidates.

6 | Calibration and Temporal Validation

The similarity score S_α is an evidence score in the interval $[0, 1]$. It is not interpreted directly as the probability of future link formation. A probabilistic interpretation is obtained only after calibration on observed future links.

Let $t_0 < t_1 < t_2$ be three observation times. The network, partitions, and reliability values are constructed by using only the information available at t_0. The links appearing in $(t_0, t_1]$ are used to select α, reliability rules, or a calibration map. The period $(t_1, t_2]$ is reserved for the final evaluation. This protocol prevents future information from entering the witness sets.

An isotonic calibration map $g : [0, 1] \to [0, 1]$ may be fitted on the validation set. The map is obtained by minimizing the squared error under the monotonicity constraint. Since g

is nondecreasing, the ranking produced by S_α is preserved. The calibrated value $g(S_\alpha)$ estimates an empirical link frequency only under similar temporal conditions.

The rough interval may also be calibrated by using two monotone maps g_L and g_U. The calibrated interval is reported as

$$[\min\{g_L(L), g_U(U)\}, \max\{g_L(L), g_U(U)\}].$$

This interval is descriptive and should not be interpreted as a formal statistical confidence interval.

For evaluation, Precision at k measures the quality of the highest-ranked recommendations. Recall at k measures the proportion of new links that are correctly recovered. When future links are rare, the precision–recall curve is preferred over the ROC curve [13]. If negative sampling is used, the sampling ratio should also be reported because precision and calibration depend on class prevalence.

Tie handling should also be specified. The interval order in Eq. (9) may leave several candidate pairs incomparable. A practical report may classify candidates into three groups: certainly above the decision threshold, boundary candidates, and certainly below the threshold. If a complete top-k list is required, secondary ordering may be performed by using S_α, followed by the smaller uncertainty width Ω, and finally a fixed deterministic identifier.

The partition and reliability function should be updated at every forecasting time. Using information from the complete observation period would produce a temporally invalid evaluation even though the mathematical index remains correct.

7 | Results and Interpretation

The computational example shows that the proposed method distinguishes candidate links even when they have the same number of common neighbors. All three candidate pairs contain three common neighbors, but different granular structures produce different evidence intervals and similarity scores.

Table 5 shows that the pair uv receives the highest score because the highly reliable block $\{a, b\}$ is fully supported. However, its uncertainty width is also the largest because the block $\{c, d, e\}$ is only partially represented. The pair xy has a smaller score, but its evidence is exact since the lower and upper values are identical. The pair rs also has an exact interval, although its final score is slightly lower.

The refinement experiment confirms Theorem 3.3. After the block $\{c, d, e\}$ is divided into smaller granules, the uncertain part disappears and the interval for uv contracts to a single value. The similarity score therefore becomes more reliable without changing the underlying graph structure.

The stability analysis shows that small perturbations in witness reliability produce only limited changes in the similarity score. This behaviour makes the framework suitable for networks where confidence values are estimated from noisy observations.

The interval representation also provides additional information that is unavailable in classical local indices. Instead of producing only one score, the proposed method reports lower evidence, upper evidence, and uncertainty width. These three quantities allow robust ranking, certified comparisons, and efficient top-k screening.

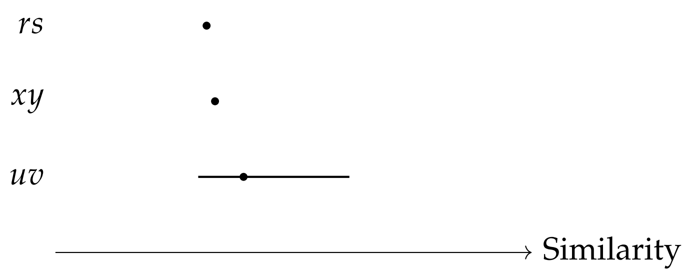

Figure 4. Evidence intervals and similarity scores for the three candidate pairs. The filled point denotes $S_{0.7}$ and the horizontal line denotes the interval $[L, U]$.

An interpretable link prediction framework has been presented by combining rough sets, fuzzy reliability, and information granules. Lower and upper approximations were used to represent certain and possible evidence, and a normalized similarity score was obtained from their weighted combination.

Theoretical analysis established symmetry, monotonicity, partition refinement, reliability stability, and boundedness of the proposed measures. A rough–fuzzy Jaccard coefficient was also introduced to provide a normalized similarity measure within the same framework.

The computational example showed that candidate pairs with the same number of common neighbors can receive different similarity scores because witness reliability and granular structure are considered. The interval representation also supported certified ranking and efficient top-k screening.

The proposed framework provides a transparent mathematical alternative to black-box similarity models and may be extended to multi-view attributed networks and adaptive reliability estimation in future work.

Author Contributions

All authors contributed equally.

Funding

The funding statement must be confirmed by all listed authors before publication.

Data Availability

No external data were used. The full synthetic partition, reliabilities, and candidate witness sets are printed in the article.

Competing Interests

The competing-interests statement must be confirmed by all listed authors before publication.

Generative-AI Assistance

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

References

[1] Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology,* 58(7), 1019-1031. https://doi.org/10.1002/asi.20591

[2] Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social networks, 2*5(3), 211-230. https://doi.org/10.1016/S0378-8733(03)00009-1

[3] Zhou, T., Lü, L., & Zhang, Y. C. (2009). Predicting missing links via local information. *The European physical* j*ournal B, 71*(4), 623-630. https://doi.org/10.1140/epjb/e2009-00335-8

[4] Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical mechanics and its applications, 3*90(6), 1150-1170. https://doi.org/10.1016/j.physa.2010.11.027

[5] 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,* 55*3,* 124289. https://doi.org/10.1016/j.physa.2020.124289

[6] Pawlak, Z. (1982). Rough sets. *International* j*ournal of computer* & *information sciences, 11*(5), 341-356. https://doi.org/10.1007/BF01001956

[7] Pawlak, Z. (1991). *Rough set: Theoretical aspects of reasoning about data*. Kluwer Academic Publishers. https://books.google.com/books/about/Rough_Sets.html?hl=tr&id=MJ-PLCqIniGsC

[8] Zadeh, L. A. (1965). Fuzzy sets. *Information and control,* 8(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X

[9] Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International* j*ournal of general system, 1*7(2-3), 191-209. https://doi.org/10.1080/03081079008935107

[10] Radzikowska, A. M., & Kerre, E. E. (2002). A comparative study of fuzzy rough sets. *Fuzzy sets and systems, 12*6(2), 137-155. https://doi.org/10.1016/S0165-0114(01)00032-X

[11] Zadeh, L. A. (1997). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy sets and systems,* 90(2), 111-127. https://doi.org/10.1016/S0165-0114(97)00077-8

[12] Yao, Y. Y. (2001). Information granulation and rough set approximation. *International* j*ournal of intelligent systems, 1*6(1), 87-104. https://doi.org/10.1002/1098-111X (200101)16:1%3C87::AID-INT7%3E3.0.CO;2-S

[13] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one, 1*0(3), e0118432. https://doi.org/10.1371/journal.pone.0118432