Crowdsourced Information Authentication: A Graph-based Model from the Science of Hadith

Ayoub Ghriss

University of Colorado, Boulder

Presented at the 4th MusIML Workshop - ICML 2025

Motivation: A Crisis of Trust in Information

Motivation: A Crisis of Trust in Information

The Core Problem

Information is often evaluated based on the perceived authority of the final publisher, while ignoring **how** the information was sourced, transmitted, and shaped along the way.

A Solution: Learning from Hadith Scholarship

A Solution: Learning from Hadith Scholarship

Islamic scholarship developed **'Ilm al-Hadith** to address similar challenges.

A Solution: Learning from Hadith Scholarship

Islamic scholarship developed **'Ilm al-Hadith** to address similar challenges.

A Two-Fold Verification System

1. **The Content (Matn):** Is it consistent with established truths?
2. **The Chain of Transmission (Isnad):** Who transmitted it? Is every narrator in the chain reliable?

A Solution: Learning from Hadith Scholarship

Islamic scholarship developed **'Ilm al-Hadith** to address similar challenges.

A Two-Fold Verification System

1. **The Content (Matn):** Is it consistent with established truths?
2. **The Chain of Transmission (Isnad):** Who transmitted it? Is every narrator in the chain reliable?

The Power of the Isnad

The authenticity of information depends critically on the integrity of its transmission path.

'Ilm al-Hadith meets Graph Theory

Each narration i consists of:

- A transmitted piece of information, T_i.
- A transmission path, $P_i = (N_{i,1} \rightarrow N_{i,2} \rightarrow \cdots \rightarrow N_{i,m_i})$.

'Ilm al-Hadith meets Graph Theory

Each narration i consists of:

- A transmitted piece of information, T_i.
- A transmission path, $P_i = (N_{i,1} \to N_{i,2} \to \cdots \to N_{i,m_i})$.

With $\mathcal{N} = \{N_j \mid 1 \leq j \leq n\}$ as the set of narrators, a narration is an element of:

$$\{\text{Set of Contents}\} \times \{\text{Paths over } \mathcal{N}\}$$

'Ilm al-Hadith meets Graph Theory

Each narration i consists of:

- A transmitted piece of information, T_i.
- A transmission path, $P_i = (N_{i,1} \to N_{i,2} \to \cdots \to N_{i,m_i})$.

With $\mathcal{N} = \{N_j \mid 1 \leq j \leq n\}$ as the set of narrators, a narration is an element of:

$$\{\text{Set of Contents}\} \times \{\text{Paths over } \mathcal{N}\}$$

The Analogy

- Narrators ($\mathcal{N}$) are the nodes in a graph.
- Transmissions are directed edges.
- An Isnad (chain) is a path in the graph.

'Ilm al-Hadith meets Graph Theory

Each narration i consists of:

- A transmitted piece of information, T_i.
- A transmission path, $P_i = (N_{i,1} \to N_{i,2} \to \cdots \to N_{i,m_i})$.

With $\mathcal{N} = \{N_j \mid 1 \leq j \leq n\}$ as the set of narrators, a narration is an element of:

$$\{\text{Set of Contents}\} \times \{\text{Paths over } \mathcal{N}\}$$

The Analogy

- Narrators ($\mathcal{N}$) are the nodes in a graph.
- Transmissions are directed edges.
- An Isnad (chain) is a path in the graph.

Goal: To create a model that can jointly and iteratively learn:

1. The authenticity score of each transmission (S_T).
2. The reliability score of each narrator (R_N).

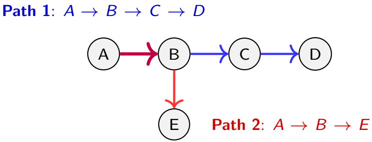
Scoring Functions

- **Path Score**: The "weakest link" principle.

$$S_{path}(P_i) = \min_i R(N_{i,m_i})$$

- **Path Overlap**: The Edge Jaccard Index.

$$O(P_i, P_j) = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}$$



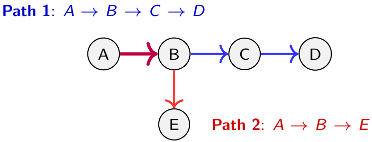
Scoring Functions

- **Path Score**: The "weakest link" principle.

$$S_{path}(P_i) = \min_i R(N_{i,m_i})$$

- **Path Overlap**: The Edge Jaccard Index.

$$O(P_i, P_j) = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}$$



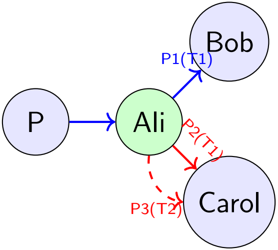
- **Information Consistency**: Uses expert knowledge, $m(T_i)$.

Example: A Trusted Narrator and a False Report

- **T1 (Weak):** *"Seek knowledge, even if you need to travel to China."*
- **T2 (False):** *"The Prophet used a compass for prayer."* $\rightarrow$ Content score $M(T2) = 0$.

Example: A Trusted Narrator and a False Report

- **T1 (Weak):** *"Seek knowledge, even if you need to travel to China."*
- **T2 (False):** *"The Prophet used a compass for prayer."* → Content score $M(T2) = 0$.



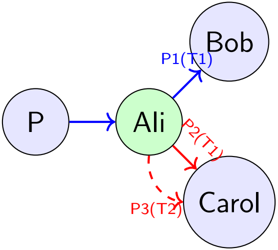
Priors:

- $R_N(Ali) = 0.9$ (Trusted)
- $R_N(Bob) = 0.6$
- $R_N(Carol) = 0.5$ (Neutral)

An Iterative Co-Update Algorithm

Step A: Update Authenticity $S_T(T_i)$

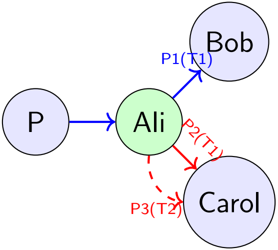
- Sort paths for T_i by their S_{path} score.



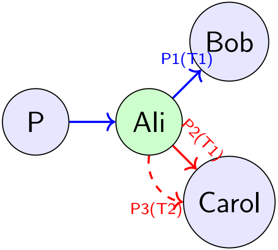
An Iterative Co-Update Algorithm

Step A: Update Authenticity $S_T(T_i)$

- Sort paths for T_i by their S_{path} score. $(S_{path}(P_1) = 0.6,\ S_{path}(P_2) = 0.5)$



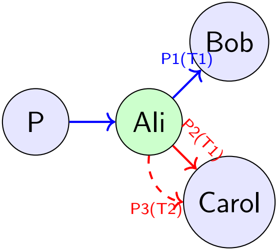
An Iterative Co-Update Algorithm



Step A: Update Authenticity $S_T(T_i)$

- Sort paths for T_i by their S_{path} score. $(S_{path}(P_1) = 0.6,\ S_{path}(P_2) = 0.5)$
- Iteratively compute discounts d_j:
$$d_j = 1 - \max_{P_k \in P_{buffer}} O(P_j, P_k)$$

An Iterative Co-Update Algorithm



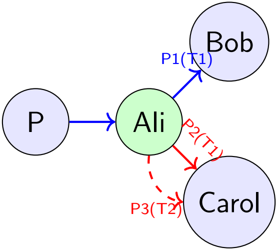
Step A: Update Authenticity $S_T(T_i)$

- Sort paths for T_i by their S_{path} score. $(S_{path}(P_1) = 0.6,\ S_{path}(P_2) = 0.5)$
- Iteratively compute discounts d_j:

$$d_j = 1 - \max_{P_k \in P_{buffer}} O(P_j, P_k)$$

$(d_1 = 1,\ d_2 = 2/3)$

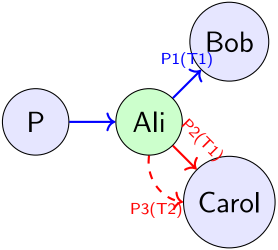
An Iterative Co-Update Algorithm



Step A: Update Authenticity $S_T(T_i)$

- Sort paths for T_i by their S_{path} score. ($S_{path}(P_1) = 0.6$, $S_{path}(P_2) = 0.5$)
- Iteratively compute discounts d_j:
$$d_j = 1 - \max_{P_k \in P_{buffer}} O(P_j, P_k)$$
($d_1 = 1$, $d_2 = 2/3$)
- Compute interim authenticity score:
$$S_{interim} = m(T_i) \cdot \sum_j d_j S_{path}(P_j) = 0.93$$

An Iterative Co-Update Algorithm



Step A: Update Authenticity $S_T(T_i)$

- Sort paths for T_i by their S_{path} score.
  $(S_{path}(P_1) = 0.6,\ S_{path}(P_2) = 0.5)$
- Iteratively compute discounts d_j:
  $$d_j = 1 - \max_{P_k \in P_{buffer}} O(P_j, P_k)$$
  $(d_1 = 1,\ d_2 = 2/3)$
- Compute interim authenticity score:
  $$S_{interim} = m(T_i) \cdot \sum_j d_j S_{path}(P_j) = 0.93$$
- Dampen with $\alpha = 0.5$:
  $$S_T(T_1) = \sigma(S_{interim}, \alpha) = \frac{S_{interim}}{S_{interim} + \alpha} \approx 0.65$$
  $$S_T(T_2) = 0$$

Step B: Update Narrator Reliability (R_N)

Principle

A narrator who consistently transmits information deemed authentic is considered reliable.

Step B: Update Narrator Reliability (R_N)

Principle

A narrator who consistently transmits information deemed authentic is considered reliable.

$$R_N(N_i) = \frac{\sum(\text{Auth. of Info } c_j) \times (\text{Contrib. of } N_i \text{ to } c_j)}{\sum(\text{Total Contrib. of } N_i) + \delta}$$

- The smoothing factor δ stabilizes scores, especially for narrators with few transmissions.

Step B: Update Narrator Reliability (R_N)

Principle

A narrator who consistently transmits information deemed authentic is considered reliable.

$$R_N(N_i) = \frac{\sum(\text{Auth. of Info } c_j) \times (\text{Contrib. of } N_i \text{ to } c_j)}{\sum(\text{Total Contrib. of } N_i) + \delta}$$

- The smoothing factor δ stabilizes scores, especially for narrators with few transmissions.

Final Scores after Convergence ($\alpha = 0.5, \delta = 1$)

Entity	Initial Trust	Final Score
Reliability (Ali)	0.9 (Trusted)	0.21
Reliability (Bob)	0.6	0.21
Reliability (Carol)	0.5 (Neutral)	0.10 (Untrustworthy)

Final Scores and Interpretation

Final Scores after Convergence ($\alpha = 0.5, \delta = 1$)

Entity	**Initial Trust**	**Final Score**
Reliability (Ali)	0.9 (Trusted)	0.21
Reliability (Bob)	0.6	0.21
Reliability (Carol)	0.5 (Neutral)	0.10 (Untrustworthy)

Interpretation: The Algorithm's Judgment

- **Punishment Fits the Crime:** Assigns lowest reliability to the narrator of the false report.
- **Trust is Not Absolute:** Revises the trust in the initially reliable Ali downwards.
- **Truth Discovery:** Distinguishes between a weak report and a false one, reflecting the collective evidence.

A Volatile and Extreme Reaction

With: $\alpha = \delta = 0.1$.

A Volatile and Extreme Reaction

With: $\alpha = \delta = 0.1$.

Step A: Update Authenticity (S_T)

- **For T2 (False):** Score remains 0 due to the content score $M(T2) = 0$.
- **For T1 (Weak):**
  - $S_{interim} \approx 0.93$ (same as before).
  - With a tiny α, the score is barely squashed:

$$S_T^{(1)}(T1) = \frac{0.93}{0.93 + 0.1} \approx 0.9$$

A Volatile and Extreme Reaction

With: $\alpha = \delta = 0.1$.

Step A: Update Authenticity (S_T)

- **For T2 (False):** Score remains 0 due to the content score $M(T2) = 0$.
- **For T1 (Weak):**
  - $S_{interim} \approx 0.93$ (same as before).
  - With a tiny α, the score is barely squashed:

$$S_T^{(1)}(T1) = \frac{0.93}{0.93 + 0.1} \approx 0.9$$

Step B: Update Narrator Reliability (R_N)

With low dampening, the narrator updates are severe:

- $R_N^{(1)}(Bob)$: Reliability **skyrockets** to 0.81.
- $R_N^{(1)}(Ali)$: Trust is tarnished, dropping to 0.58.
- $R_N^{(1)}(Carol)$: Reliability **plummets** to 0.42.

Paper's Argument for Convergence

Compactness and Continuity

The analysis relies on two key properties of the ICUA operator, F:

- **Boundedness:** All authenticity (S_T) and reliability (R_N) scores are constrained to the interval [0, 1].
- **Continuity:** The update functions (min, sum, division, σ) are continuous.

Paper's Argument for Convergence

Compactness and Continuity

The analysis relies on two key properties of the ICUA operator, F:

- **Boundedness:** All authenticity (S_T) and reliability (R_N) scores are constrained to the interval [0, 1].
- **Continuity:** The update functions (min, sum, division, σ) are continuous.

Core Argument: Brouwer's Fixed-Point Theorem

Since F is a continuous function mapping a compact, convex set to itself, the theorem **guarantees that at least one fixed point X^* exists**, where $X^* = F(X^*)$.

Paper's Argument for Convergence

Compactness and Continuity

The analysis relies on two key properties of the ICUA operator, F:

- **Boundedness:** All authenticity (S_T) and reliability (R_N) scores are constrained to the interval [0, 1].
- **Continuity:** The update functions (min, sum, division, σ) are continuous.

Core Argument: Brouwer's Fixed-Point Theorem

Since F is a continuous function mapping a compact, convex set to itself, the theorem **guarantees that at least one fixed point X^* exists**, where $X^* = F(X^*)$.

The Crucial Caveat: Existence is Not Convergence

A proof of convergence would require showing F is a **contraction mapping**.

Conclusion and Future Work

Contributions:

- A novel, graph-based framework (ICUA) formalizing principles from traditional Hadith science.
- A method for jointly assessing information authenticity and source reliability.
- Incorporates nuances like the "weakest link" principle and path overlap discounting.

Future Work:

- Developing a full mathematical proof of convergence.
- Making model parameters (α, δ) learnable from data.
- Enhancing the analysis of textual content using advanced NLP.
- Applying the ICUA framework to other domains like fake news detection or peer review.

Thank You

Questions?

Convergence Analysis: The Role of α and δ

Sufficient Condition for Convergence

ICUA is guaranteed to converge if its iterative update function is a **contraction mapping**, meaning the change in scores shrinks with each iteration.

This leads to the following condition:

$$\left(\max_n \frac{W_n}{W_n + \delta}\right) \cdot \left(\frac{C_{max}}{\alpha}\right) < 1$$

- A **trade-off** between the dampening effects of α and δ.
- α (**Transmission Dampening**): Counters signal amplification from transmissions with many paths (C_{max}).
- δ (**Narrator Dampening**): Stabilizes narrator scores.
- **Convergence is achieved** when dampening effects (α, δ) overcome the amplification effects inherent in the graph.

ICUA Algorithm: A Flowchart