LBK: A Heuristic Approach to Meta-Based Automated Reasoning in Knowledge Graphs

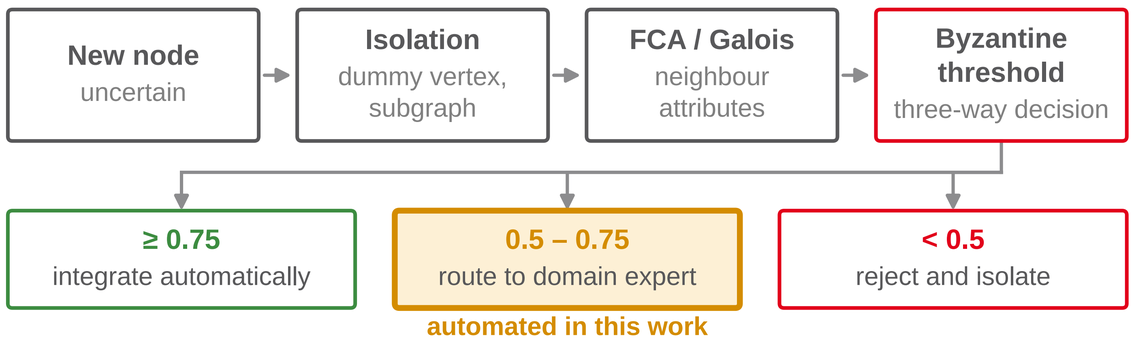
Simon von Oppenkowski · Benedikt Lerch · Klemens Schnattinger

Data Science and Artificial Intelligence Department, Baden-Wuerttemberg Cooperative State University (DHBW), Lörrach, Germany
{vonoppes, lerchb, schnattinger}@dhbw-loerrach.de

KDIR 2026
Paper #131

Phase 1 leaves a manual middle band

New nodes enter a knowledge graph as uncertain. Phase 1 of LBK [1] isolates each node in a local subgraph, abstracts its neighbours' attributes by Galois derivation and applies a Byzantine consensus threshold. Nodes scoring from 0.5 to below 0.75 wait for a domain expert, which ties up experts and blocks dependent processes.

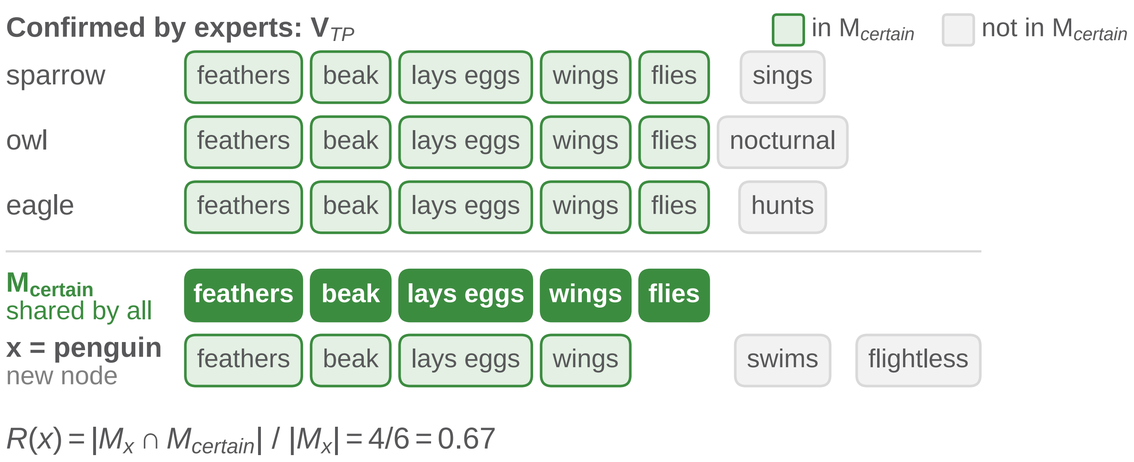


Research question

How can semi-certain knowledge streams be automatically validated and seamlessly integrated into or isolated from an evolving topology?

Expert decisions define the reference set

Every node an expert confirms joins V_{TP}. The attributes shared by all of them form $M_{certain}$. A new node x is scored by the share of its own attributes that lie in $M_{certain}$.



$$R(x) = |M_x \cap M_{certain}| \,/\, |M_x| = 4/6 = 0.67$$

$R(x)$ measures structural consistency with the verified population, not factual correctness.

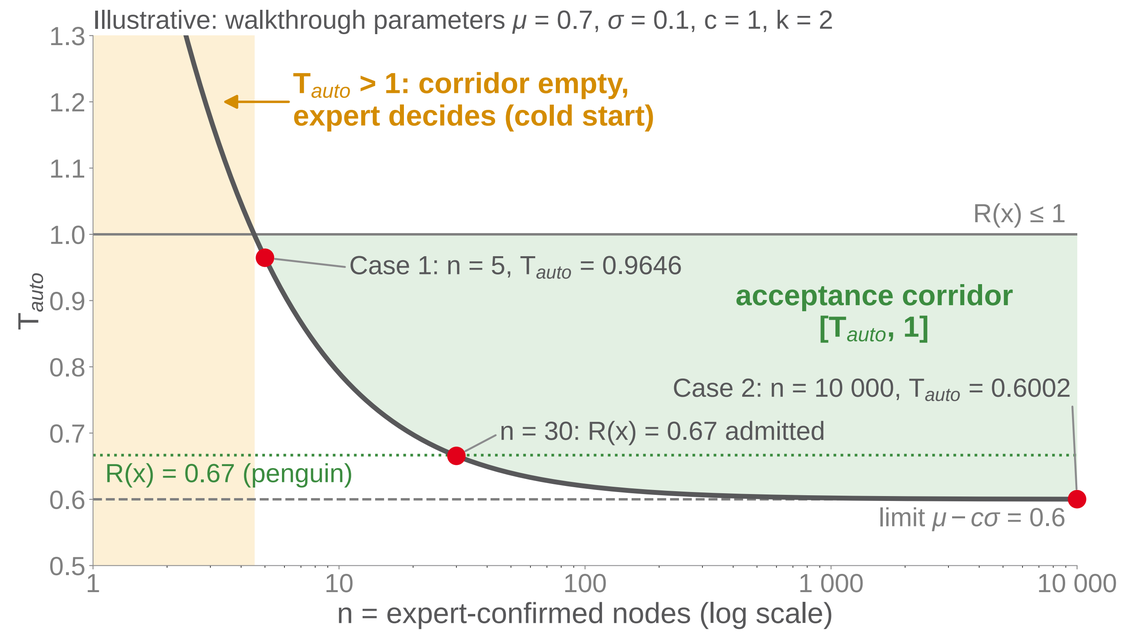
The corridor opens as evidence grows

$$T_{auto} = (\mu - c\cdot\sigma) + k\cdot\ln(1 + 1/n)$$

accept x if $R(x) \geq T_{auto}$

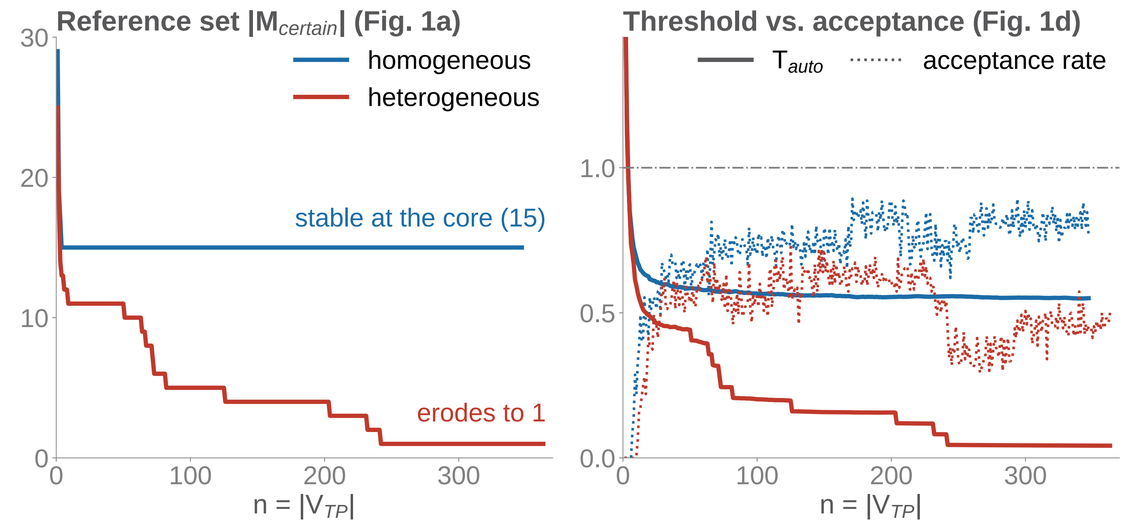
- μ, σ: mean and standard deviation of $R(v)$ over V_{TP}; c: declared tolerance, default 1.
- For small $n = |V_{TP}|$ the penalty dominates; $T_{auto} > 1$ empties the corridor and the expert decides (cold start).
- The activation point n_0 fixes $k = (1 - (\mu - c\cdot\sigma)) / \ln(1 + 1/n_0)$: when automation becomes available, not when it becomes reliable.

Illustrative: walkthrough parameters $\mu = 0.7$, $\sigma = 0.1$, c = 1, k = 2



Two regimes: stabilisation or erosion

The decaying penalty widens the corridor; the baseline $\mu - c\cdot\sigma$ tightens it under homogeneity. Synthetic sensitivity analysis: 60 attributes, core of 15, 4–14 peripheral attributes per node, attributes independent.



- **Homogeneous:** $M_{certain}$ stabilises at the core after fewer than 50 nodes; μ settles at 0.63, σ at 0.09, T_{auto} at 0.55; acceptance rises from 0.68 to 0.77.
- **Heterogeneous** (core attribute absent with probability 0.005): $|M_{certain}|$ decays from 15 to 1, μ and T_{auto} fall to 0.04, acceptance stays near 0.5. Threshold and acceptance decouple.
- **Monitoring:** a monotone decline of $|M_{certain}|$ suspends the autonomous mode in favour of expert validation.

Sensitivity study under controlled assumptions, not an empirical evaluation on real-world data.

Positioning

Approach	Representation	Threshold origin	Human in the loop	What is validated
SHACL / rule-based	declarative schema	manually specified	rule maintenance	schema conformance
KGE / GNN confidence	dense vectors	calibrated on training data	retraining	plausibility in embedding space
Credible intervals [3]	sample statistics	estimated from sample	sampling	KG accuracy estimate
LBK (this work)	discrete attribute sets	empirical, from V_{TP}	below activation point n_0	structural consistency

Scope and next steps

- Uncertainty at the attribute level only; entity identity, relation semantics and source trust need other mechanisms.
- Homophily assumption: autonomous mode only in domain-homogeneous regions.



- Phase 3: GDELT benchmark against stochastic and embedding baselines.
- Scale: streaming architecture with incremental updates of $M_{certain}$.

Take-home message

- Expert decisions become the reference set.
- Closed during cold start, the corridor then adapts to the verified evidence.
- Autonomy only in homogeneous domains; erosion of $M_{certain}$ hands control back to experts.

[1] S. von Oppenkowski, B. Lerch, K. Schnattinger (2026). An approach for deriving uncertain knowledge in knowledge graphs under the theoretical mapping of heuristics. ECKM 2026, Rome. In press.

[2] S. D. von Oppenkowski (2026). Supplementary material and algorithmic validation report for the Let's Build Knowledge (LBK) framework. DHBW Lörrach. researchgate.net/publication/407442333

[3] S. Marchesin, G. Silvello (2025). Credible intervals for knowledge graph accuracy estimation. Proceedings of the ACM on Management of Data.

Report [2]