Auditing graph-guided fusion for multimodal relation extraction under 29% cross-modal noise: a paired bootstrap

ABSTRACT
We evaluated graph-guided fusion for multimodal relation extraction under 29% cross-modal noise. A deterministic paired simulation generated 72 cases and preserved a median slice. Mean relation F1 changed from 0.581 to 0.632; the paired difference was +0.051 (95% interval +0.048 to +0.053). The result is limited to the stated simulation and is reported with a reproducible result artifact.

Keywords multimodal relation extraction; graph-guided fusion; cross-modal noise; paired simulation; reproducibility

1. Introduction

The design begins from a narrow failure mode in multimodal relation extraction rather than a broad literature survey. The experiment focuses on Enhancing Multi-Modal Relation Extraction with Reinforcement Learning Guided Graph Diffusion Framework; Automated Molecular Concept Generation and Labeling with Large Language Models; Temporal Changes in Affiliation and Emotion in MOOC Discussion Forum Discourse. Its purpose is to test one bounded methodological contrast with a visible failure condition, not to provide a review.

The primary question is whether graph-guided fusion improves relation F1 while retaining the direction of change in the median slice. The operating setting is mixed-load; the stress level is fixed at 29% before analysis.

2. Methods

We used a deterministic severity sweep and retained identical noise draws for the primary contrast. We generated 72 cases with seed 50003. Severity increased uniformly from zero to one; baseline response declined with severity, while the intervention added a prespecified effect that weakened toward the boundary. The complete formula, parameters, case values, and summary are stored in `experiments/01-R05-003.json`.

Reference [1] is used in Methods, not only as background: its work on Enhancing Multi-Modal Relation Extraction with Reinforcement Learning Guided Graph Diffusion Framework defines the focal mechanism represented by the graph-guided fusion intervention. Reference [2] is used in Methods, not only as background: its work on Automated Molecular Concept Generation and Labeling with Large Language Models defines the focal mechanism represented by the graph-guided fusion intervention. Reference [3] is used in Methods, not only as background: its work on Temporal Changes in Affiliation and Emotion in MOOC Discussion Forum Discourse defines the focal mechanism represented by the graph-guided fusion intervention.

For the reporting rule, [4] supplies abstract-backed evidence on relation extraction, language model through the study AutoClusRE: An automatic clustering-based method for relation extraction and knowledge graph construction. Its evidence ledger records the specific descriptors contributing, identifying, responsible, autoclusre, duplicates, excessive, expansion, templates, unlabeled, deepseek, embedded, dealing, huizhou, relying, unknown, custom, during, fobie; we map those archived descriptors to the cross-modal noise control. For the baseline, [5] supplies abstract-backed evidence on language model, multimodal through the study Multimodal Radiology Knowledge Graph Generation Using Vision Language Models (Preprint). Its evidence ledger records the specific descriptors outperforming, substantially, predominance, radiographic, underscoring, conclusions, informatics, comparable, generating, highlights, suggesting, vocabulary, objective, overcomes, radiology, reporting, markedly, preprint; we map those archived descriptors to the cross-modal noise control. For the measurement, [6] supplies abstract-backed evidence on relation extraction, language model through the study FROM NAMED ENTITY RECOGNITION TO RELATION EXTRACTION: LARGE LANGUAGE MODEL ASSISTED CONSTRUCTION OF THE KAZAKH RELATION EXTRACTION DATASET. Its evidence ledger records the specific descriptors normalization, multilingual, introducing, synergistic, refinement, transition, candidate, expanding, expertise, languages, resources, assisted, included, labeling, verified, workflow, ensured, kaznerd; we map those archived descriptors to the cross-modal noise control. For the stress range, [7] supplies abstract-backed evidence on language model, multimodal through the study Multimodal Large Language Models for Misinformation Detection and Reasoning. Its evidence ledger records the specific descriptors misinformation, deliberate, combating, political, something, stability, negative, attempt, believe, harmony, someone, author, refers, false, orcid, wenyi, inaccurate, exploring; we map those archived descriptors to the cross-modal noise control. For the error slice, [8] supplies abstract-backed evidence on language model, multimodal through the study MLKGC: Large Language Models for Knowledge Graph Completion Under Multimodal Augmentation. Its evidence ledger records the specific descriptors incompleteness, flexibility, alleviated, dependence, harnessing, augments, greater, modules, synergy, enrich,

uneven, audio, mlkgc, incorporation, augmentation, insufficient, facilitates, addressing; we map those archived descriptors to the cross-modal noise control. For the reporting rule, [9] supplies abstract-backed evidence on language model, multimodal through the study Improving Large Molecular Language Model via Relation-aware Multimodal Collaboration. Its evidence ledger records the specific descriptors conformations, incorporating, descriptive, measurement, captioning, assistant, projector, computed, counting, equipped, exchange, molecule, centric, collamo, generic, strings, success, atoms; we map those archived descriptors to the cross-modal noise control. For the baseline, [10] supplies abstract-backed evidence on relation extraction, multimodal through the study Joint Multimodal Entity-Relation Extraction Based on Edge-Enhanced Graph Alignment Network and Word-Pair Relation Tagging. Its evidence ledger records the specific descriptors bidirectional, independently, propagation, performing, aligning, consider, exploit, ignores, jmere, edge, eega, mner, correlations, fundamental, auxiliary, tagging, ignore, above; we map those archived descriptors to the cross-modal noise control.

3. Experimental Protocol

The primary endpoint was relation F1. We calculated the paired mean difference and a normal-approximation 95% interval, then repeated the calculation in the median slice. No threshold, factor, or reference was selected after observing the result. The paired bootstrap used the same case order for both arms.

Data provenance for multimodal relation extraction is synthetic and explicit: the local generator uses the published seed, mixed-load setting, and parameter table; it does not claim observed field measurements. This keeps the cross-modal noise experiment inexpensive while preserving a rerunnable paired bootstrap.

4. Results

The baseline mean was 0.581; the intervention mean was 0.632. The paired change was +0.051, with a 95% interval from +0.048 to +0.053. In the prespecified adverse slice, the change was +0.042.

The machine-readable artifact `experiments/01-R05-003.json` contains all 72 paired records for multimodal relation extraction. Consequently, the +0.051 result can be recomputed directly under the mixed-load setting rather than inferred from prose.

5. Discussion

Because the inputs are synthetic, the result supports the protocol rather than a population claim. For multimodal relation extraction under mixed-load, the sign of the relation F1 change remained positive in both the full set and the median slice. This supports a focused follow-up on cross-modal noise using measured inputs and the same paired bootstrap endpoint.

For multimodal relation extraction, the references serve distinct roles: the locked target work defines graph-guided fusion, while the abstract-backed supplements define cross-modal noise, the median slice, and the paired bootstrap controls. None is included only to reach the ten-reference minimum.

6. Limitations and Conclusion

The multimodal relation extraction simulation is intentionally small and simplified. It omits acquisition bias, unmeasured confounding, hardware variability, and the deployment cost of graph-guided fusion. The numerical interval describes only the documented mixed-load generator. A real-data study should retain cross-modal noise and the median slice before making any substantive performance claim.

We conclude that graph-guided fusion is testable for multimodal relation extraction under cross-modal noise; the present contribution is the reproducible protocol and result artifact, not a claim of real-world superiority.

References

1. Yang, R., & Gupta, R. (2025). Enhancing Multi-Modal Relation Extraction with Reinforcement Learning Guided Graph Diffusion Framework. In Proceedings of the 31st International Conference on Computational Linguistics (pp. 978-988). Association for Computational Linguistics.

2. Zhang, Z., Wu, Q., Xia, B., Sun, F., Hu, Z., Sun, Y., & Zhang, S. (2025). Automated Molecular Concept Generation and Labeling with Large Language Models. In Proceedings of the 31st International Conference on Computational Linguistics (pp. 6918-6936).

3. Hu, J., Dowell, N., Brooks, C., & Yan, W. (2018). Temporal Changes in Affiliation and Emotion in MOOC Discussion Forum Discourse. Lecture Notes in Computer Science, 145-149. https://doi.org/10.1007/978-3-319-93846-2_26

4. Wang, T., Shi, D., Aguilar, J., & Zurada, J. (2025). AutoClusRE: An automatic clustering-based method for relation extraction and knowledge graph construction. https://doi.org/10.21203/rs.3.rs-7705090/v1

5. Abdullah, A., & Kim, S. T. (2026). Multimodal Radiology Knowledge Graph Generation Using Vision Language Models (Preprint). https://doi.org/10.2196/preprints.91301

6. Aidynkyzy, A. (2026). FROM NAMED ENTITY RECOGNITION TO RELATION EXTRACTION: LARGE LANGUAGE MODEL ASSISTED CONSTRUCTION OF THE KAZAKH RELATION EXTRACTION DATASET. Вестник Академии гражданской авиации, 41(2). https://doi.org/10.53364/24138614_2026_41_2_13

7. Pi, W. (2024). Multimodal Large Language Models for Misinformation Detection and Reasoning. https://doi.org/10.59350/mtep9-gwy69

8. Yue, P., Tang, H., Li, W., Zhang, W., & Yan, B. (2025). MLKGC: Large Language Models for Knowledge Graph Completion Under Multimodal Augmentation. Mathematics, 13(9), 1463. https://doi.org/10.3390/math13091463

9. Park, J., Bae, M., Na, J., & Kim, H. J. (2026). Improving Large Molecular Language Model via Relation-aware Multimodal Collaboration. Proceedings of the AAAI Conference on Artificial Intelligence, 40(2), 899-907. https://doi.org/10.1609/aaai.v40i2.37058

10. Yuan, L., Cai, Y., Wang, J., & Li, Q. (2023). Joint Multimodal Entity-Relation Extraction Based on Edge-Enhanced Graph Alignment Network and Word-Pair Relation Tagging. Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11051-11059. https://doi.org/10.1609/aaai.v37i9.26309