Assessment of Answer Reliability of a Cybersecurity Knowledge Graph Under Software Vulnerability Abstraction Perturbation

Dovydas Marius Zapkus, Asta Slotkienė

Institute of Data Science and Digital Technologies, Vilnius University, marius.zapkus@mif.vu.lt, asta.slotkiene@mif.vu.lt

INTRODUCTION

The prevention of and defense against cybersecurity threats increasingly rely on structured, openly maintained security knowledge. Knowledge about software vulnerabilities (CVE), their root-cause weaknesses (CWE), and the associated attack patterns (CAPEC) is increasingly represented as a knowledge graph. The reliability of these graphs is, however, fundamentally constrained by the quality of the CVE-to-CWE assignments upon which they are constructed. Şimşek et al. showed that a large proportion of NVD records exhibit invalid, placeholder, or insufficiently specific CWE mappings [1]. Further studies [2,3] show that, contrary to MITRE's mapping guidance, vulnerabilities are frequently assigned to overly general (higher-level) weaknesses, with classification errors concentrated among hierarchically adjacent parent–child and sibling pairs. According to Mahmud et al. [3], the absence or ambiguity of descriptive CWE properties degrades knowledge-mining results; notably, the evaluation in [2] treats a same-branch substitute as approximately correct, implicitly presupposing those annotations are inherited along the CWE hierarchy. We empirically evaluated this assumption on 207 CWE chains (class ⇒ base ⇒ variant) across 54 sibling groups, measuring semantic proximity with sentence embeddings. A same-branch sibling was closer to the true weakness than its parent in about 60% of comparisons, and in 36% after excluding three groups of near-duplicate weaknesses (CWE-41). A key factor was intra-group semantic homogeneity: siblings outperformed the parent in 60% to 81% of comparisons within homogeneous groups, but only in 17% to 41% within heterogeneous groups. Same-branch substitution should therefore not be treated as approximately correct by default. It is justified only for semantically homogeneous sibling sets, and under classifier uncertainty, the parent constitutes the safer prediction.

AIM

To assess how two common expert deviations in CWE assignment affect the semantic correctness of the CWE knowledge graph: parent perturbation, where the general parent weakness is assigned instead of the specific variant, and sibling perturbation, where a different variant from the same branch is assigned.

METHODOLOGY

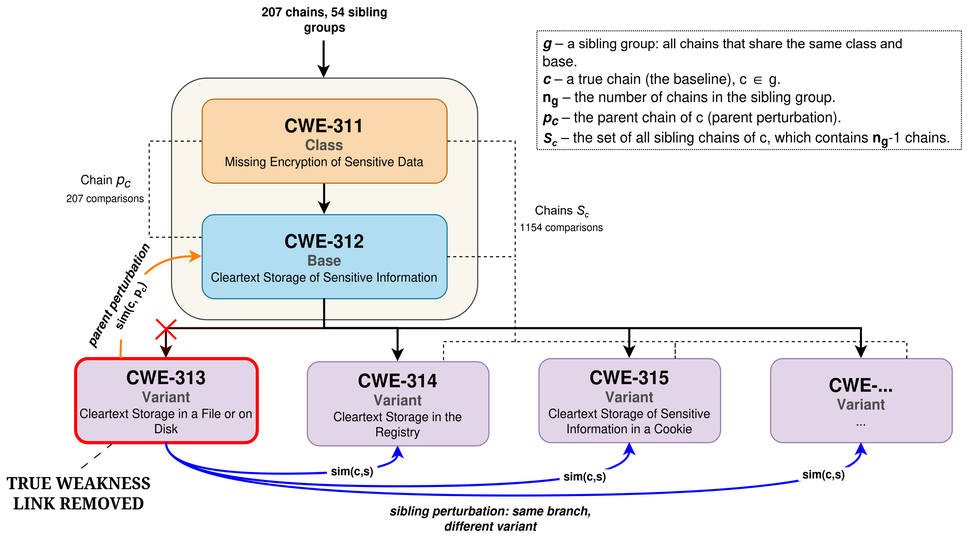

Figure 1. Parent and sibling perturbations of a true chain c (example)

$$\mathrm{PS}_g^{\mathrm{sibling}} = \frac{1}{n_g(n_g-1)} \sum_{c \in g} \sum_{s \in S_c} \mathbf{1}[\mathrm{sim}(c,s) > \mathrm{sim}(c,p_c)] \quad (1)$$

$$\mathrm{PS}_g^{\mathrm{parent}} = \frac{1}{n_g(n_g-1)} \sum_{c \in g} \sum_{s \in S_c} \mathbf{1}[\mathrm{sim}(c,p_c) > \mathrm{sim}(c,s)] \quad (2)$$

where $\mathrm{sim}(c,s) = \frac{\vec{c}\cdot\vec{s}}{\|\vec{c}\|\|\vec{s}\|}$ is the cosine similarity of the embedding vectors $\vec{c}$ and $\vec{s}$ of the texts of chains c and s.

PS - Probability of Superiority

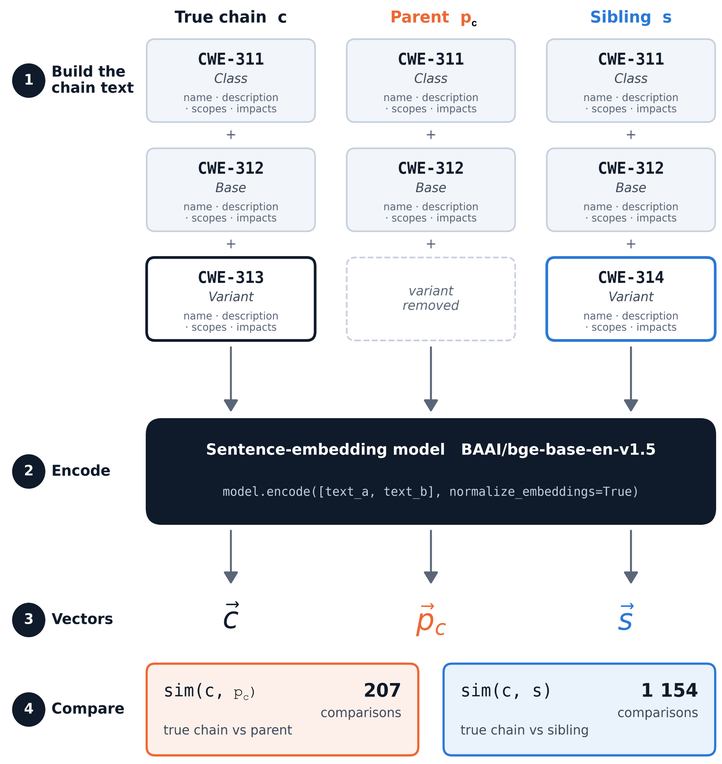

Figure 2. Chain texts are encoded into embedding vectors and compared by the cosine similarity sim

RESULTS

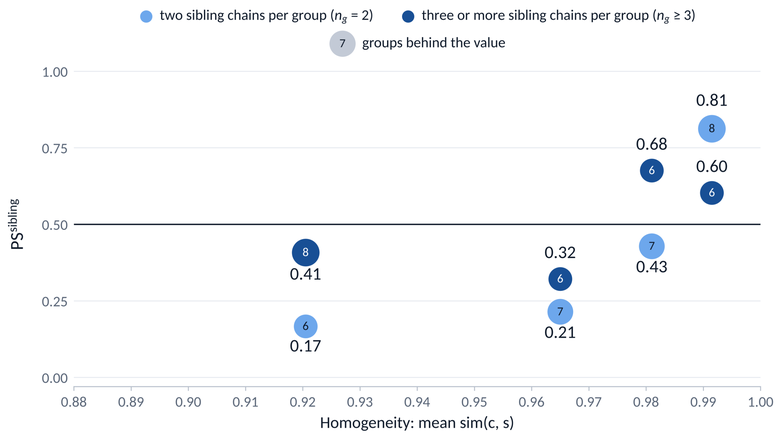

Figure 3. The sibling's PS by homogeneity of the sibling group

Whether a same-branch sibling preserves the meaning of the true weakness depends largely on the homogeneity of its sibling set, not on the hierarchy level alone.

When the exact variant is uncertain, assigning the parent is the safer deviation (closer in 31 of 54 groups).

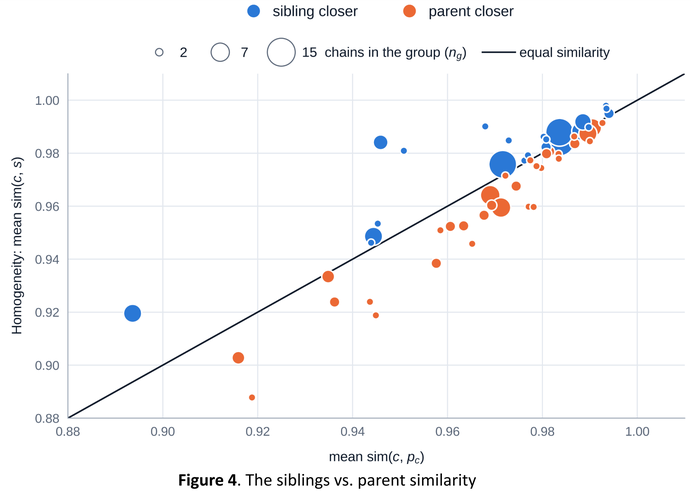

Figure 4. The siblings vs. parent similarity

A sibling is closer in most comparisons, and in all 31 groups below the line, the parent is closer.

The outcome is determined not by homogeneity alone, but by whether the siblings are closer to the chain than their parent is.

CONCLUSIONS AND LIMITATIONS

The semantic homogeneity of a sibling group indicates that siblings outperformed the parent in 60–81% of comparisons within the most homogeneous quartile, but in only 17–41% of comparisons within the most heterogeneous quartile.

Of the two perturbations performed, the parent deviation is the safer one: in 31 of 54 groups, the parent was closer to the true weakness than the siblings in most comparisons. The sibling's 60% share of all comparisons comes from three large near-duplicate groups (CWE-41), which contribute a disproportionate number of pairwise comparisons; without them it falls to 36%. A sibling prediction should be accepted only when the sibling weaknesses are semantically alike at ≥ 0.985. A limitation of this experiment is that it uses only variant-level CWE chain texts that share the class and base descriptions.

REFERENCES

[1] Şimşek, Ş., Xia, H., Gluck, J., Medina, D. S., & Starobinski, D. (2025, July). Fixing Invalid CVE-CWE Mappings in Threat Databases. In 2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 950-960). IEEE.

[2] Schwengber Kelm, A., Bockermann, C., & Frochte, J. (2026). Multi-Class vs. Multi-Label BERT for CVE-to-CWE Mapping: How Taxonomy Structure Shapes the Errors. arXiv e-prints, arXiv:2607.07573.

[3] Mahmud, A., Rawajfih, Y., Shahriar, H., & Wu, F. (2025, December). Contamination-aware, taxonomy-driven vulnerability classification for CPS: Evidence and insights at scale. In 2025 IEEE International Conference on Big Data (BigData) (pp. 4298-4304). IEEE.