(Appendix) Robust Regular Path Queries over Streaming Graphs via Load-Aware Windowing

Mauro Famà
INSA Lyon
Lyon, France
mauro.fama@insa-lyon.fr

Riccardo Tommasini
INSA Lyon
Lyon, France
riccardo.tommasini@insa-lyon.fr

Angela Bonifati
Lyon 1 University
Lyon, France
angela.bonifati@univ-lyon1.fr

PVLDB Reference Format:
Mauro Famà, Riccardo Tommasini, and Angela Bonifati. (Appendix) Robust Regular Path Queries over Streaming Graphs via Load-Aware Windowing. PVLDB, 14(1): XXX-XXX, 2020.
doi:XX.XX/XXX.XX

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at URL_TO_YOUR_ARTIFACTS.

1 EXTENDED PROOF OF INFEASIBILITY OF LOSSLESS ONLINE RESIZING

In this section, we report the extend proof of Theorem 4.4, that for the sake of space, we reported as a sketch in the full paper.

Theorem 4.4 (Infeasibility of Lossless Online Resizing). *There is no online policy ρ^{early} that, given only the stream seen so far, selects the subset W^{early} of $\widehat{W}_{A,\beta}$ that yields $\widehat{Ans}^{early}$.*

Proof. Without loss of generality, we make the following *assumptions*: we fix a slide $\beta > 0$ and let window sizes be integer multiples of β. Let $A = [\alpha_{\min}, \alpha_{\max}] \cap \{\beta, 2\beta, \ldots\}$ and let $\widehat{W}_{A,\beta}$ be the corresponding family of sliding windows. For the proof, we can consider any RPQ that admits a finite accepting path of length at least 2 (e.g., $a \cdot b$).

Indistinguishable prefix. We pick two admissible window sizes $\alpha_S, \alpha_L \in A$ with $\beta \leq \alpha_S < \alpha_L \leq \alpha_{\max}$. Construct two streams that are identical up to time α_S: each has a single a at time 0 and no b in $(0, \alpha_S)$ (arbitrary neutral edges may occur). They diverge after α_S as follows, for a fixed $0 < \varepsilon < \beta$:

$$S^{\text{early}}: \ b \text{ occurs at } \alpha_S - \varepsilon \quad \text{and} \quad S^{\text{late}}: \ b \text{ occurs at } \alpha_L - \varepsilon.$$

In S^{early}, the earliest emission for matching $a \cdot b$ occurs at $c = \alpha_S$ using the window $[0, \alpha_S)$; any larger window closing later is not part of $\widehat{W}_{early}$ by minimality. In S^{late}, completeness requires the window $[0, \alpha_L)$; smaller windows cannot contain both a and b.

Online indistinguishability. Any online selector sees identical history up to α_S on both streams and must irrevocably decide which windows to maintain going forward. There are only two possibilities at (or before) α_S:

(1) It *does not* maintain the α_L-window that will close at $c = \alpha_L$ (equivalently, it evicts data prior to $\alpha_L - \alpha_L = 0$). Then, on S^{late}, the pair created by a at 0 and b at $\alpha_L - \varepsilon$ cannot be produced at $c = \alpha_L$, violating completeness w.r.t. $\widehat{Ans}^{early}$.
(2) It *does* maintain that α_L-window. Then, on S^{early}, the offline-optimal set W^{early} *does not* include the larger window closing at α_L (it is redundant for earliest emission). Hence the online selector cannot be outputting exactly W^{early} on S^{early}.

In both cases, the online selector fails to output the offline-optimal window set: Either it loses completeness (case 1) or it selects a strict superset of W^{early} (case 2), which by definition does not "yield $\widehat{Ans}^{early}$" as the *selected set*. Therefore, no online procedure can always compute W^{early}. The argument extends to randomised algorithms [1] against an adaptive adversary (who picks S^{early} or S^{late} after observing the decision at α_S). □

REFERENCES

[1] Andrew Chi-Chin Yao. 1977. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*. 222–227. https://doi.org/10.1109/SFCS.1977.24

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX