Modeling Graphs with Vertex Replacement Grammars

Satyaki Sikdar, Justus Hibshman, and Tim Weninger
Department of Computer Science and Engineering
University of Notre Dame
{ssikdar,jhibshma,tweninge}@nd.edu

Introduction

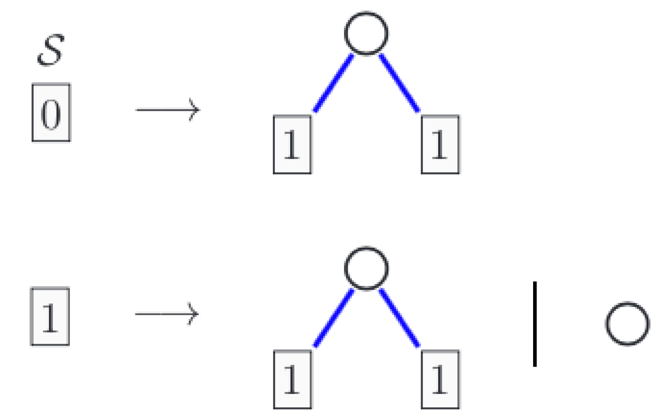
Arguably the most relevant task in the study of graphs is the identification, extraction, and representation of the small substructures that, in aggregate, describe the underlying phenomenon encoded by the graph. From an extracted model, containing the LEGO-like building blocks of real-world graphs, we expect to perform deep scientific analysis and make predictions about the data.

Context Free Grammars

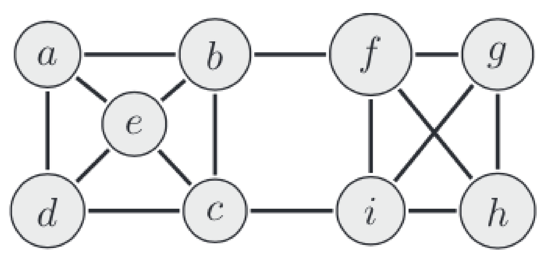
String Grammars

$S \rightarrow$ NP VP
NP $\rightarrow$ `the` N
VP $\rightarrow$ V NP
V $\rightarrow$ `sings` | `eats`
N $\rightarrow$ `cat` | `song` | `canary`

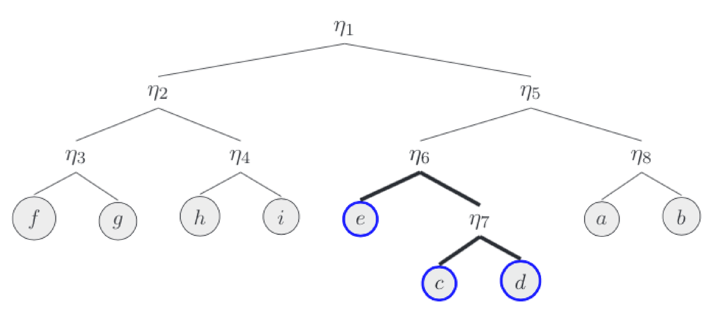
Vertex Replacement Grammars



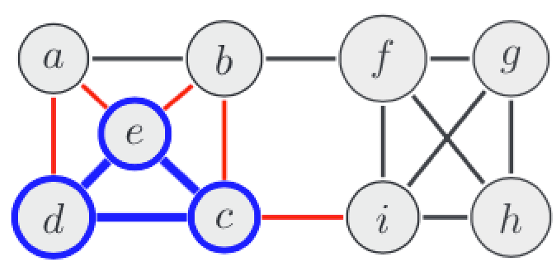
One Iteration of Extraction with λ = 3



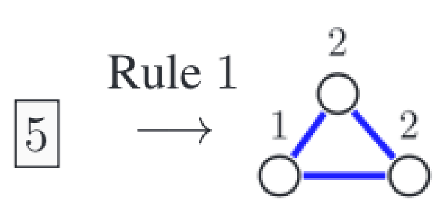
(a) Example graph H with 9 nodes and 16 edges.



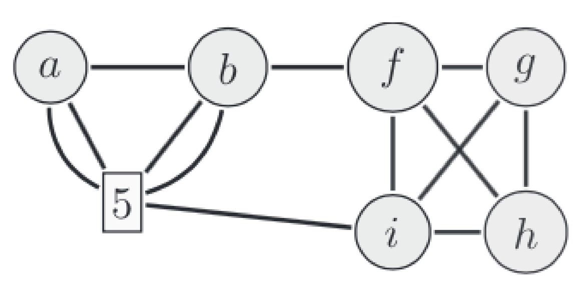
(b) Dendrogram D of H. Node η_6 has exactly λ leaf nodes $\{c, d, e\}$.



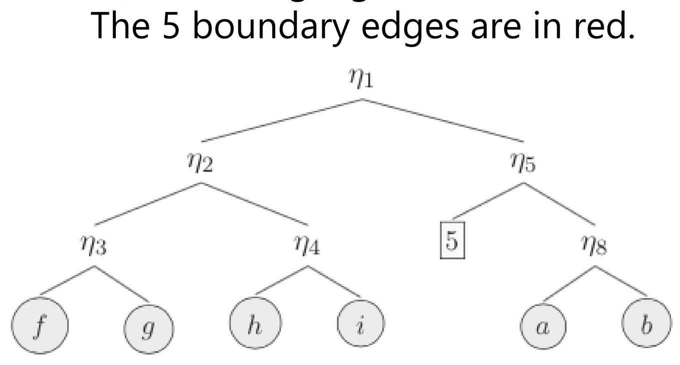
(c) H with highlighted nodes $\{c, d, e\}$. The 5 boundary edges are in red.



(d) The RHS is a non-terminal of size 5 corresponding to the 5 boundary edges in (c).

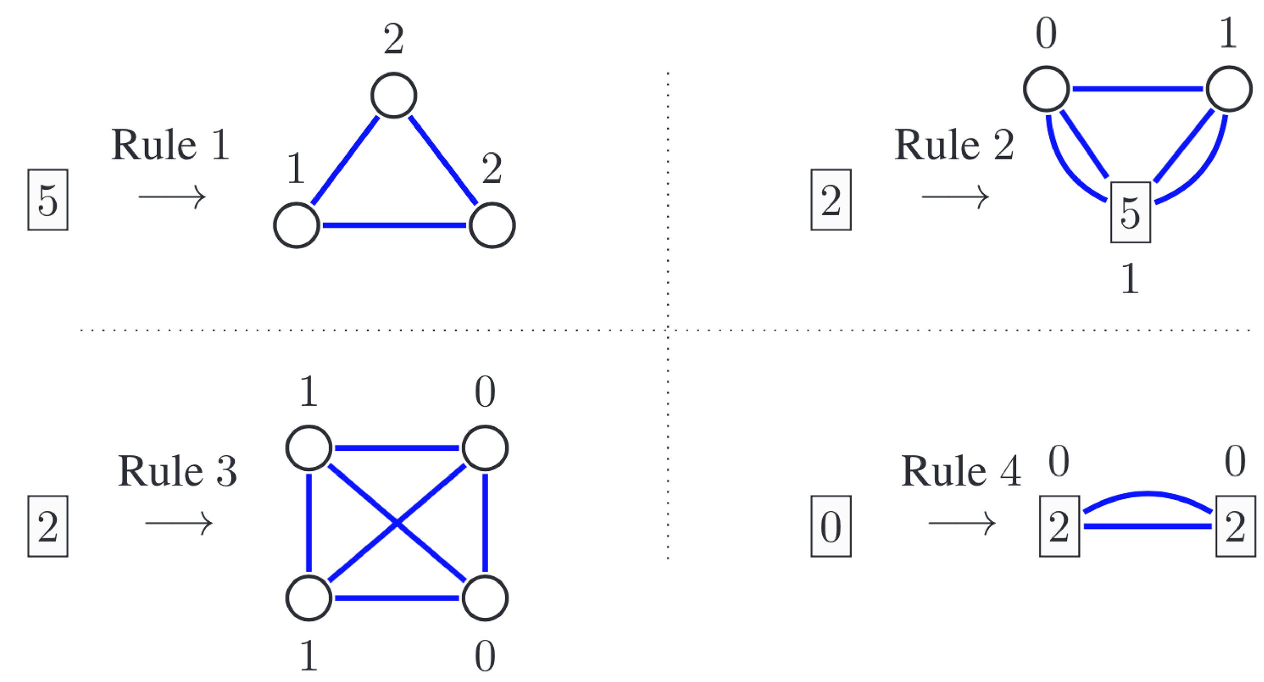


(e) Updated graph with $\{c, d, e\}$ replaced by the non-terminal of size 5.

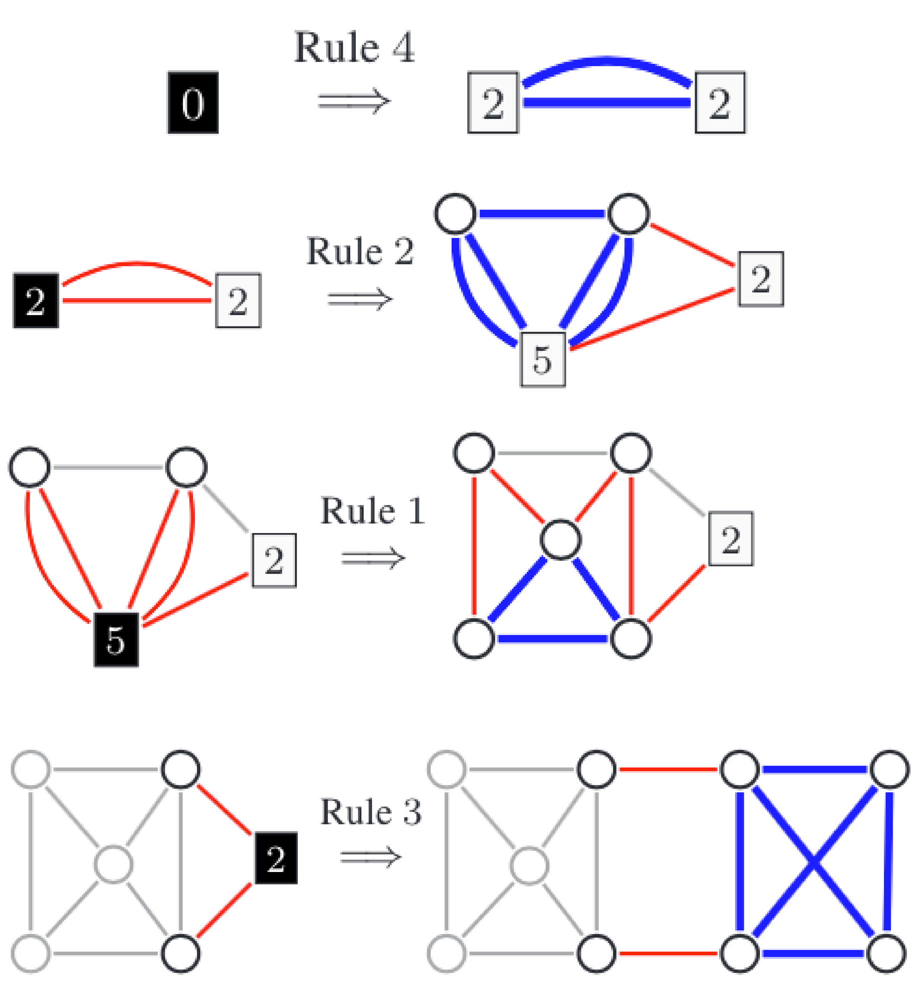


(f) Updated dendrogram with subtree rooted at η_6 replaced by the new non-terminal.

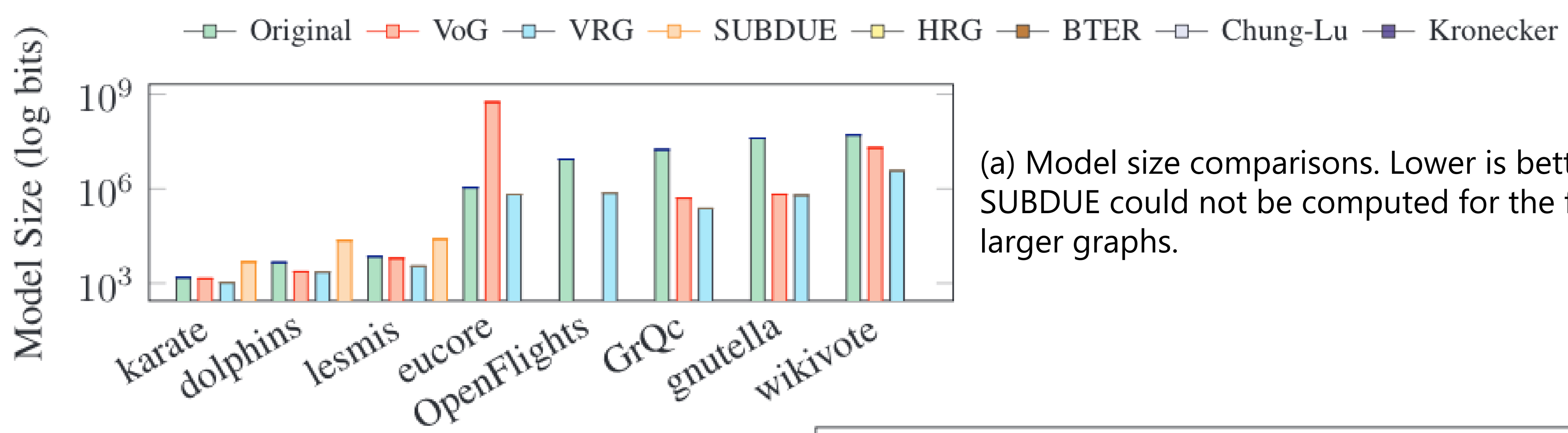
Extracted Rules



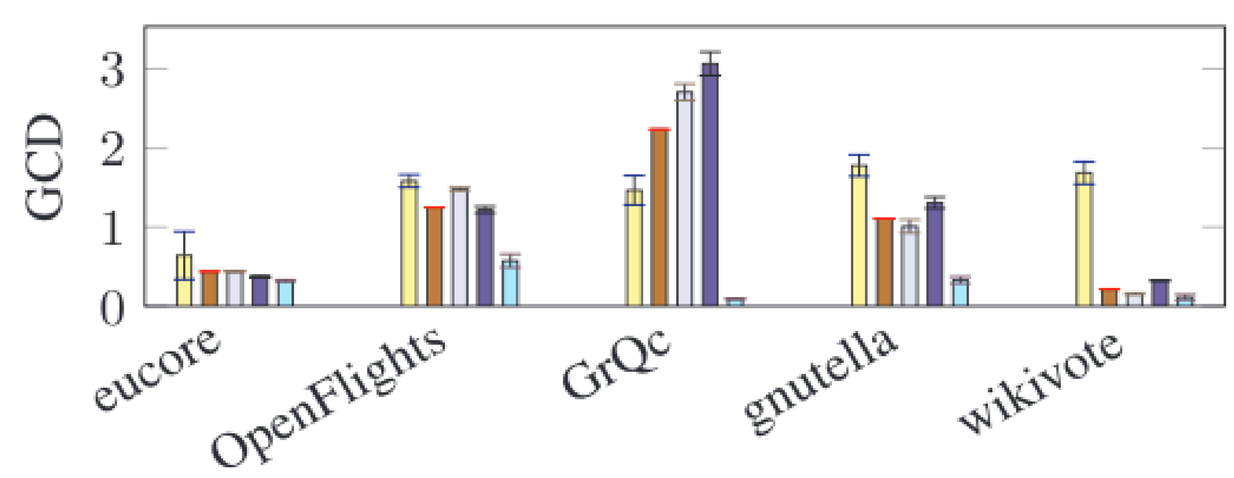
Graph Generation



Main Results



(a) Model size comparisons. Lower is better. SUBDUE could not be computed for the five larger graphs.



(b) GCD comparisons for graph generators. Lower is better. Error bars indicate the 95% confidence interval around the mean.

Conclusion

A potentially significant benefit from the VRG model stems from its ability to directly encode local substructures and patterns in the RHSs of the grammar rules. Forward applications of VRGs may allow scientists to identify previously unknown patterns in graph datasets representing important natural or physical phenomena. Further investigation into the nature of the extracted rules and their meaning (if any) is a top priority.

Related Works

Aguinaga S, Chiang D, Weninger T (2018) *Learning hyperedge replacement grammars for graph generation*. IEEE Transactions on Pattern Analysis and Machine Intelligence pp 1–1, DOI 10.1109/TPAMI.2018.2810877

Pennycuff C, Sikdar S, Vajiac C, Chiang D, Weninger T (2018) *Synchronous hyperedge replacement graph grammars*. In: International Conference on Graph Transformation, Springer, pp 20–36

The work has been submitted to the Journal Track of ECML-PKDD 2019. The authors are supported by a grant from the US National Science Foundation (#1652492).

UNIVERSITY OF NOTRE DAME

WENINGER LAB