View Reviews

Paper ID
227

Paper Title
Z-shadow An Efficient Method for Estimating Bicliques in Massive Graphs Using Füredi's Theorem

Track Name
EDBT2025 (Research - 3rd round)

Reviewer #1

Questions

5. Importance (You may indicate more than one option.)
The paper contains controversial ideas and/or will generate interesting discussion
EDBT attendees will learn something interesting from the paper

6. Novelty (You may indicate more than one option.)
Novel solution

7. Three or more strengths of the paper. This can refer to importance, novelty, technical depth and quality of content, experiments or presentation. Please refer to the points as S1, S2, etc.
##

(S1) The paper proposed the Z-shadow approximation algorithm through Z-shadow-find and sampling, eschewing direct counting bicliques on the whole dataset.

(S2) Conducts experiments on both real-world and synthetic datasets. The experiment results show that the Z-S approach achieves more than 500x speedup.

(S3) The paper proposes a novel online sampling method that significantly reduces memory consumption.

8. Three or more weaknesses of the paper. This can refer to importance, novelty, technical depth and quality of content, experiments or presentation. Please refer to the points as W1, W2, etc.
(W1) In algorithm 1, how do other threshold functions α influence the efficiency and accuracy (See D1)?

(W2) The comparison of the experiment is not thorough (See D2, D3, D4, D5).

(W3) The presentation needs to be polished and improved (See D6).

9. Detailed evaluation. Please refer to the points as D1, D2, etc.

(D1) **Alpha function:** In the Z-shadow-Finder approach, the author chose the alpha function based on Füredi's theory. Do other alpha functions affect the result, especially the accuracy? Experiments should be added.

(D2) **Ordering:** The Z-S method is more efficient than Z-S(NS). One question is what kind of vertex ordering strategy is used for BCL and BCL++. The BCL and BCL++ also show more efficiency by ordering vertex by degree. Also, do other ordering strategies influence efficiency? It seems the reorder does not influence the results in the synthetic dataset. Could you please explain?

(D3) **Missing comparison:** Z-S is an approximation algorithm. Yet it is not the only tackling the biclique counting. A recent state-of-the-art [*] approximate algorithm also does that and has results similar to the one in the paper.

[*] Ye, X., Li, R.H., Dai, Q., Qin, H. and Wang, G., 2023. Efficient biclique counting in large bipartite graphs. *Proceedings of the ACM on Management of Data*, *1*(1), pp.1-26.

(D4) **Hyperparameters:** It is not clear how s,t affects accuracy and efficiency since the paper only conducts the experiment for s, t range from 2 to 6. Could the authors please show more experiments with larger p, q such as s>10 and t>10? A larger t requires more iteration of Algorithm 1 and should affect the efficiency.

(D5) **Sparse graphs:** I fully understand that authors are targeting large dense bipartite graphs. However, does the Z-S-based solution also work well for sparse graphs in terms of accuracy and efficiency compared with the exact algorithm? It would be appreciated to see some experiments about the sparse graph.

(D6) **Presentation:** The authors do not discuss the limitations of the previous work, such as [14], BCList++ and other sampling techniques and how the proposed method is better. Instead, the discussion contains general statements such that “these works have certain limitations”. The related work needs to be more specific on such matters. In addition, the paper needs a better structure. For instance, the pseudocode deserves more explanation for the reader.

Reviewer #2

Questions

5. Importance (You may indicate more than one option.)

EDBT attendees will learn something interesting from the paper

6. Novelty (You may indicate more than one option.)

Novel implementation and evaluation of previous solutions to existing problem

7. Three or more strengths of the paper. This can refer to importance, novelty, technical depth and quality of content, experiments or presentation. Please refer to the points as S1, S2, etc.

1.introduces a unique approach to biclique estimation by integrating Füredi's Theorem with a randomized sampling method.
2.High Accuracy and Efficiency in Large-Scale Graphs: The experimental results show that Z-Shadow achieves accuracy within a 1% error margin while providing up to 500x speedups compared to existing methods
3.The paper introduces an online sampling technique that minimizes memory usage by discarding unnecessary data early in the process.

8. Three or more weaknesses of the paper. This can refer to importance, novelty, technical depth and quality of content, experiments or presentation. Please refer to the points as W1, W2, etc.

1.Unclear Presentation of Experimental Results: In the results section, terms like "accuracy within 0.5%" and "500x speedup" are mentioned without clear context on how these metrics were calculated.
2.Inconsistent Terminology and Definitions: The paper sometimes uses terminology that could be better standardized. For example, terms like "sampling threshold" and "shadow construction" appear without an immediate explanation
3.Incomplete Explanation of Figures: Some figures, such as Figures 2 and 3 showing time and accuracy comparisons, lack detailed captions or explanations.

9. Detailed evaluation. Please refer to the points as D1, D2, etc.

D1.Z-Shadow's computational efficiency, which achieves a 500x speedup over baseline methods, is a notable strength.
D2.authors conducted thorough testing on both synthetic and real-world datasets, showcasing the algorithm's effectiveness with high accuracy and substantial speed improvements.

Reviewer #3

Questions

5. Importance (You may indicate more than one option.)

EDBT attendees will learn something interesting from the paper
No impact expected

6. Novelty (You may indicate more than one option.)

Novel solution

7. Three or more strengths of the paper. This can refer to importance, novelty, technical depth and quality of content, experiments or presentation. Please refer to the points as S1, S2, etc.

S1. The technical contribution of the paper is satisfactory.

S2. The paper demonstrates some significant improvements over some competitors at its experimental section.

S3. The paper looks into important things, such as the running time and space complexities, but also evaluates these experimentally against several competitors.

8. Three or more weaknesses of the paper. This can refer to importance, novelty, technical depth and quality of content, experiments or presentation. Please refer to the points as W1, W2, etc.

W1. The paper is not easy to ready and its presentation requires significant improvements in most of its parts.

W2. The motivation of the paper is not strong and could be strengthened.

W3. The presented algorithms are hard to follow with the amount of introduced notation and the lack of adequate set of examples at some points.

9. Detailed evaluation. Please refer to the points as D1, D2, etc.

D1. Related to W2, the paper needs to motivate why the studied poblem is important (and relevant to EDBT), and why approximate solutions for the counting problem suffice. The last 2 sentences of the 2nd paragraph of the introduction do not provide a convincing paper motivation.

D2. While the introduction of the paper nicely presents the contributions of the paper, the presentation of the paper requires significant improvements. Some things that could be improved include:
a) The notation is heavy, and the paper could use a table of notation.
b) Explaining the Zarankiewicz number would be nice.
c) After Definition 4 it would be nice to present an intuitive explanation and an example. If the reader does not understand the notion of a shadow here, he will not be able to follow the rest of the paper.
d) Algorithm 1 is never referenced at the text, nor described line by line.
e) Figure 5 is not adequately described at the text.
f) The paragraph below Algorithm 1 only references Figure 5, but the text mainly refers to Algorithm 1, so this is confusing.
g) Algorithms 2 and 3 are mentioned at the text, but never described there.
h) Explaining the competitive algorithms at Section 6.1, rather than presenting them in an algorithm description at the Appendix.
i) Adding titles all y axes at the figures of the experiments (see Figure 7 and 8).
j) Explaining Z-S(G) better.
k) It is not clear to which table Section 6.6 refers to and where the results that this section mentions can be found.
l) There are several grammatical errors throughout the paper that could be corrected.

D3. On a positive note, the paper has some nice experimental results, demonstrating significant improvements over some competitors.

D4. Moving some proofs to the Appendix was definitely a correct step, as the paper is already difficult to read at its core algorithmic part.