Christos' template - v14: METAPAPER: Spotting Fake Reviews in App Stores

first name
SCS CMU
first@cs.cmu.edu

second name
SCS CMU
second@andrew.cmu.edu

Christos Faloutsos
SCS CMU
christos@cs.cmu.edu

March 16, 2023

Abstract

Ask a rhetorical question: - What is the best rhetorical question you can start with?

How can you find strange nodes in a who-calls-whom graph? Spotting anomalies in graphs is an important topic.

'what' - NOT 'how': List the benefits of the approach - NOT the details of how you do it!

Our METAPAPER method

baptize: Give a NAME to the method - ideal name should (a) be an english-like word, but NOT a vocabulary word (b) easy to pronounce ('say it three times, quickly') (c) should emphasize the main idea/insight/advantage of your method (NOT the steps you took - 'FraudSpot' is good, 'DeepLearnFraud' is not) (d) should have positive connotation ('eagle', 'lion', 'safe', 'guard', 'spot', 'alert')

has the following properties (a) *Scalability*, being linear on the input size (b) *Effectiveness*, spotting 90% of the anomalies in real data (c) *Parameter-free*, requiring no user-defined parameters.

numbers: Mention some performance numbers

Experiments on *3GB* of real data from epinions.com illustrate the benefits of our method.

1 Introduction

Again, a rhetorical question

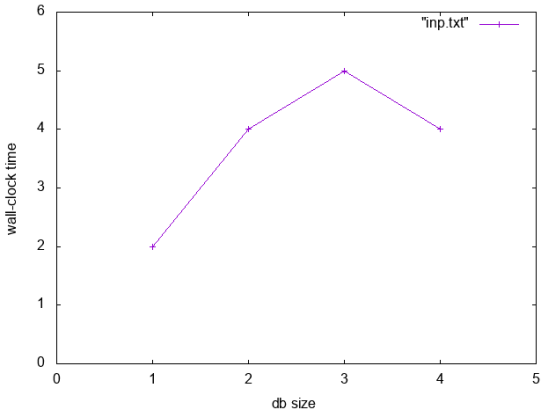

Figure 1: **METAPAPER wins**: Execution time for METAPAPER, on epinions.com

Given a large count of reviews for products, how can one spot the fake ones. On line reviews are important. They are often faked, for monetary gain. How to spot the truth?

Here we propose METAPAPER, a method to spot fake reviews. The main idea behind our method is a principled way to merge several warning signals. Figure 1 shows the results of our method

Crown-jewel figure: show-case our very best results, easy to understand

2word-tag: for each figure caption, give the 2-word conclusion in bold

where METAPAPER outperforms the competition by up to 999%

The advantages of our method are

- **Scalability** : it scales linearly with the input

size

- **Effectiveness**: it gives very good reconstruction error, on real data
- **Parameter-free** it requires no user-defined parameters.

bullets: 2-3 bullet items; make 'tags' into macros, and REPEAT in abstract AND conclusions

Reproducibility: we publish our data and our code, at `www.cs.cmu.edu`

reproducibility: repeat, in conclusions, too

The outline of the paper is typical: we give the survey (Section 2), the proposed method (Section 3), experiments (Section 4), and conclusions (Section 5).

2 Background and Related Work

There is a lot of work on app reviews, and we group it in the following sub-sections.

2.1 Fraud detection

Bla-bla-bla - fake citation: [3] [2]

2.2 Anomaly detection

bla-bla - oddball, subdue etc [1]

However none of the above methods fullfils all the specs of our method: (a) scalability (b) effectiveness.

salesman matrix: last *column*: YOUR method

Table 1 contrasts METAPAPER against the state of the art competitors.

3 Proposed Method

novelty: NO citations, outside the 'survey' - they make METAPAPER seem incremental.

In this section we present the proposed method, we analyze it and provide the reader with several interesting -at least in our opinion- observations. Table 2 gives the list of symbols we use.

Give table of symbols and dfns

Method / Property	*method1*	*method2*	METAPAPER
Scalability			✔
Effectiveness			✔
Parameter-free			✔
other-stuff			✔

Table 1: **METAPAPER matches all specs**, while competitors miss one or more of the features.

Symbols	Definitions
G	a graph
A	adjacency matrix

Table 2: Symbols and Definitions

3.1 Intuition

The main idea behind our METAPAPER is to exploit network effects: if we see a bi-partite core, then we suspect fraud.

Figure 2 and 3 illustrates the intution comment them out, if the tikz package is missing

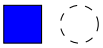

Figure 2: Sample tikz figure

3.2 Algorithm

Algorithm 1 shows the pseudo code and Algorithm 2 shows a variation

3.3 Complexity Analysis

Theorem 1 METAPAPER *requires time linear on the input size*

Proof 1 *from Eq *bla*, with lagrange multipliers*

3.4 SQL implementation

In fact, we can use SQL to implement our algorithm:

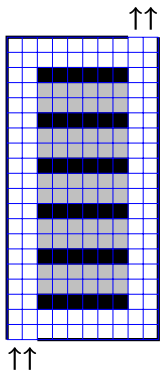

Figure 3: Another Sample tikz figure

Data: this text
Result: how to write algorithm with LaTeX2e
initialization;
while *not at end of this document* **do**
    read current;
    if *understand* **then** `// nice!`
        go to next section;
        current section becomes this one;
    else
        go back to the beginning of current section;

Algorithm 2: How to write algorithms

```
select name, ssn
from student
where name = 'smith'
  and grade = 'A';
```

Data: a graph G
Result: the communities in G
1 initialization;
2 put graph G on stack;
3 pop stack;
4 **while** *stack is not empty* **do**
5 process current graph;
6 **if** *has* n^2 *edges* **then**
7 add to output;
8 **else**
9 split in two;
10 put both on stack;
11 **end**
12 pop stack;
13 **end**

Algorithm 1: FakeCom: Community detection algorithm

4 Experiments

Here we report experiments to answer the following questions

"bullets": should correspond to the 'advantages' bullets in the intro and conclusions, as well as to the headers of the upcoming subsections

Q1. **Scalability**: How fast is our METAPAPER
Q2. **Effectiveness**: How well does METAPAPER work on real data?

Q3. some other question, if any

The graphs we used in our experiments are described in the table 3.

4.1 Q1 - Scalability

In figure 4

2-word-tag: For each caption, give a 2-3 word summary.

we present the experimental results for the real-world datasets we used.

Nodes	Edges	Description
Real-world Networks		
13,579	37,448	AS Oregon
23,389	47,448	CAIDA AS 2004 to 2008

Table 3: Summary of real-world networks used.

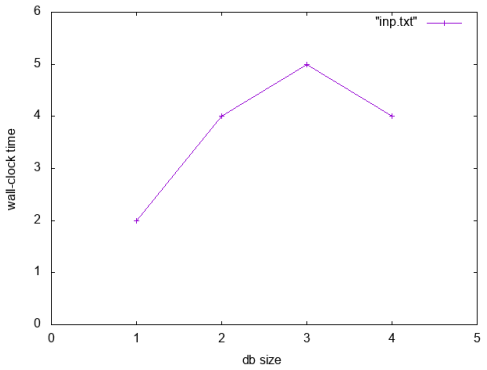

Figure 4: **METAPAPER scales lineary**: response time vs input size

4.2 Q2 - Effectiveness

In Figure bla, we show the precision/recall of METAPAPER for the epinions.com dataset. Notice that the results are great

4.3 Discussion - practitioner's guide

Given the above, we recommend that practitioners choose bla for the parameter values of METAPAPER, and do preprocessing using bla-bla for their datasets.

4.4 Discoveries - METAPAPER at work

Thanks to our method, we processed real data and noticed the following observations

success stories: elaborate... - if too many nice observations, promote this to a full section

Observation 1 METAPAPER *works better than expected, on real data*

The reason is that we assumed uniformity, while real data have skewed distributions (Pareto-like), and thus favor our approach.

Observation 2 METAPAPER *works faster than expected*

bla-bla

5 Conclusions

We presented METAPAPER, which addresses the fake-review problem, using network effects. The main idea is to spot anomalous graph substructures, using belief propagation

The advantages of the method are

1. **Scalability**: it scales linearly with the input size, as shown in Figure 4 and Lemma +++

   "bullets": exact replica of the bullet items of the intro, plus pointers to the figures or lemmas or section numbers

2. **Effectiveness**: it gives excellent precision, on real world data
3. **Parameter-free**: it requires no user intervention - METAPAPER sets all parameters to reasonable defaults,

   whatever else we said in the intro and experiments - slightly modify the wording, but keep the 'tags'/macros

   and it is insensitive to the exact choices anyway

We also presented experiments on *3GB* of real data, where METAPAPER outperformed the competitors by 10 percentage points of accuracy, and 3x faster execution time.

reproducibility: repeat, from intro:

Reproducibility: We have already open-sourced our code, at `www.cs.cmu.edu`

Acknowledgements: We would like to thank Christos Faloutsos for his *MetaPaper* list of suggestions on the presentation.

References

[1] L. Akoglu, M. McGlohon, and C. Faloutsos. oddball: Spotting anomalies in weighted graphs. In *PAKDD (2)*, volume 6119 of *Lecture Notes in Computer Science*, pages 410–421. Springer, 2010.

[2] C. Faloutsos. Signature files. In *Information Retrieval: Data Structures & Algorithms*, pages 44–65. 1992.

[3] H. Tong and U. Kang. Big data clustering. In *Data Clustering: Algorithms and Applications*, pages 259–276. 2013.

A Stealth appendix

[here we put self notes etc, that we will NOT include in the final paper]

Todo list

Ask a rhetorical question: - What is the best rhetorical question you can start with? 1

'what' - NOT 'how': List the benefits of the approach - NOT the details of how you do it! 1

baptize: Give a NAME to the method - ideal name should (a) be an english-like word, but NOT a vocabulary word (b) easy to pronounce ('say it three times, quickly') (c) should emphasize the main idea/insight/advantage of your method (NOT the steps you took - 'FraudSpot' is good, 'DeepLearnFraud' is not) (d) should have positive connotation ('eagle', 'lion', 'safe', 'guard', 'spot', 'alert') . . 1

numbers: Mention some performance numbers 1

Again, a rhetorical question 1

Crown-jewel figure: show-case our very best results, easy to understand . . 1

2word-tag: for each figure caption, give the 2-word conclusion in bold 1

bullets: 2-3 bullet items; make 'tags' into macros, and REPEAT in abstract AND conclusions 2

reproducibility: repeat, in conclusions, too 2

salesman matrix: last *column*: YOUR method 2

novelty: NO citations, outside the 'survey' - they make METAPAPER seem incremental. 2

Give table of symbols and dfns 2

"bullets": should correspond to the 'advantages' bullets in the intro and conclusions, as well as to the headers of the upcoming subsections 3

2-word-tag: For each caption, give a 2-3 word summary. 3

success stories: elaborate... - if too many nice observations, promote this to a full section 4

"bullets": exact replica of the bullet items of the intro, plus pointers to the figures or lemmas or section numbers 4

whatever else we said in the intro and experiments - slightly modify the wording, but keep the 'tags'/macros . . . 4

reproducibility: repeat, from intro: 4