Christos' template - v15: METAPAPER: Spotting Fake Reviews in App Stores

First Name
CS, UCR
first@cs.ucr.edu

Second Name
EE, ASU
second@ee.asu.edu

Christos Faloutsos
SCS CMU
christos@cs.cmu.edu

November 10, 2026

Abstract

How can we find suspicious reviews on, say, epinions.com or yelp.com? Spotting anomalies in who-reviews-what graphs is an important topic. Our METAPAPER method 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. Experiments on **3GB** of real data from epinions.com illustrate the benefits of our method.

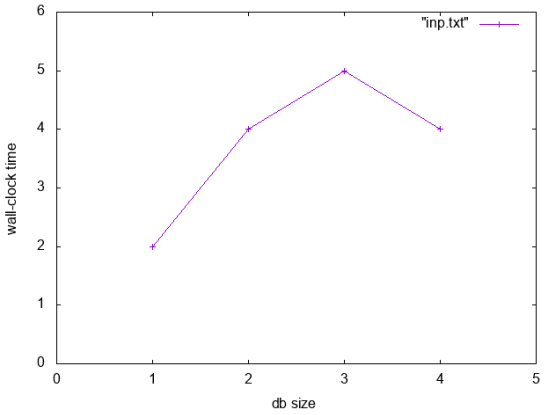


Figure 1: **METAPAPER wins**: Execution time for METAPAPER, on epinions.com

1 Introduction

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 where METAPAPER outperforms the competition by up to **999%**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Informally, the problem we address here is the following:

Informal Problem Definition 1 (Suspicious Re-

views). *As mentioned above, the problem is as follows:*

- *Given a set of reviews, each with*
  - *author-id, item-id, timestamp*
  - *and review-text*
  - *and NO labels*
- *Find suspicious reviews (and authors, and products)*
- *to optimize the error in a labeled, hold-out set of reviews*

The main insight behind METAPAPER is that fraudsters often exhibit lock-step behavior, e.g., too many, too similar reviews, about the same time, for the same product(s), by the same authors.
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.

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

The outline of the paper is typical: we give the survey (Section 2), the proposed method (Section 3), experiments (Section 4), and conclusions (Section 6).

2 Background and Related Work

In short, none of the baselines fullfils all the properties that METAPAPER does, as we show in Table 1 We review these baselines below, grouping them in the following sub-sections.

Fraud detection *Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.*[4] [2]

Anomaly detection *Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.*[1]

Summary However, as we said, none of the above methods fullfils all the specificaitons that our method does (see Table 1).

Property \ Method	*method1*	*method2*	*METAPAPER*
Scalability	✔		✔
Effectiveness		✔	✔
Parameter-free			✔
nimble			✔
explainable			✔

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

3 Proposed Method

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.

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 illustrates the intution

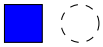


Figure 2: Sample tikz figure

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.2 Algorithm

Based on the intuition above, Algorithm 1 shows the pseudo code.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

```
Data: a graph G
Result: the communities in G
initialization;
put graph G on stack;
pop stack;
while stack is not empty do
    process current graph;
    if has n^2 edges then
        add to output;
    else
        split in two;
        put both on stack;
    end
    pop stack;
end
```

Algorithm 1: METAPAPER$_{CD}$: Proposed community detection algorithm

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus

sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.3 Complexity Analysis

Lemma 1. METAPAPER *requires time* $O(n)$*, that is, it is linear on the input size.*

Proof. All steps in algorithm 1 are linear. Details omitted for brevity (see Appendix A) ■

4 Experiments

Here we report experiments to answer the following questions

- **Q1 Scalability**: How fast is our METAPAPER
- **Q2 Effectiveness**: How well does METAPAPER work on real data?
- **Q3** Ablation study: which of the components is the most influential?

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

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.

4.1 Q1 - Scalability

In figure 3 we present the experimental results for the real-world datasets we used.

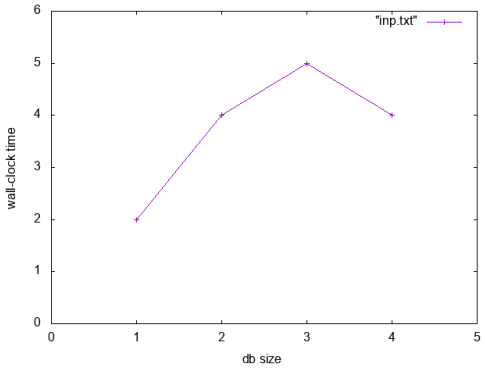


Figure 3: **METAPAPER scales linearly**: response time vs input size

4.2 Q2 - Effectiveness

In Figure bla, we show the precision/recall of METAPAPER for the epinions.com dataset. Notice that METAPAPER outperforms the baselines.

Table 4 shows the accuracy of METAPAPER and top baselines, applied on several datasets. METAPAPER wins or ties in almost all cases.

Method	dataset1	dataset2	dataset3
baseline1	0.3	0.4	0.3
baseline2	0.4	0.1	0.1
baseline3	0.1	0.5	0.2
METAPAPER	0.8	0.7	0.5

Table 4: **METAPAPER wins** on accuracy. Experiments on several datasets. Shades of green indicate winner, 2nd and 3rd place.

4.3 Q3 - Ablation Study

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasel-

lus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5 Discussion

5.1 Discussion - practitioner's guide

Our sensitivity analysis showed that the accuracy of our algorithm has a flat optimal for p in $(0.10-10.0)$. Thus, we recommend $p = 1.0$.

5.2 Discoveries - METAPAPER at work

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

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.*

Our explanation is again that our analysis assumes uniform distributions, which constitute a worst-case setting, while real distributions are skewed.

6 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:

- *Scalability*: it scales linearly with the input size, as shown in Figure 3 and Lemma 1.
- *Effectiveness*: it gives excellent precision, on real world data
- *Parameter-free*: it requires no user intervention - METAPAPER sets all parameters to reasonable defaults, and it is insensitive to the exact choices anyway.

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

Reproducibility: We have already open-sourced our code, at `www.cs.cmu.edu`. The datasets are available at `kaggle.com`.

Acknowledgements: We would like to thank Christos Faloutsos for his *MetaPaper* list of suggestions [3] 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] C. Faloutsos. MetaPaper: Lessons learned from 1k research paper rejections. `https://www.cs.cmu.edu/~christos/MetaPaper/`, March 2023.

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

A Appendix - proofs

The proof for Lemma 1 is as follows:

Proof. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum. ■