MetaPaper: Lessons from 1K rejected research papers

Christos Faloutsos
CMU

https://www.cs.cmu.edu/~christos/MetaPaper
Disclaimer – meta-advice

Mainly, for ML / Data Science / DB venues

For other venues:

Some of the suggestions may still hold, but:

👉 Team up with seasoned authors from those venues
Self-introduction

Qualifications:
https://www.cs.cmu.edu/~christos/

- Over 30+ years of research-paper submissions from CMU (and UofT, UMD, AT&T, IBM, MSR, etc)
- 3-digit acceptances
- 4-digit (~1K) rejections
- …
Self-introduction

Qualifications:
https://www.cs.cmu.edu/~christos/

• Over 30+ years of research-paper submissions from CMU (and UofT, UMD, AT&T, IBM, MSR, etc)
• 3-digit acceptances
• 4-digit (~1K) rejections
• … and counting (sigh!)
On the bright side:

You may learn much more from a game you lose than from a game you win.

José Raúl Capablanca
World chess champion 1921-27
(‘the Cuban chess machine’)
On the bright side:

You may learn much more from a game you lose than from a game you win.

Only a fool learns from his own mistakes. The wise man learns from the mistakes of others.

(attr. to Bismarck (‘Iron Chancellor’) - wikiquote.org)
Only a fool learns from his own mistakes. The wise man learns from the mistakes of others.

(attr. to Bismarck (‘Iron Chancellor’) - wikiquote.org)

germanculture.com.ua
(ouch!)

Only a fool learns from his own mistakes. The wise man learns from the mistakes of others.

(attr. to Bismarck (‘Iron Chancellor’) - wikiquote.org)
Outline

• Top 3 lessons: ‘F.A.N.’
  – F
  – A
  – N

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
Outline

• Top 3 lessons: ‘F.A.N.’
  – F
  – A
  – N

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
‘F’ – which one?

- Pick one to review
- Q1) Which one?
- Q2) Why?

(a)

(b)

(c)
‘F’ – which one?

- Pick one to review
- Q1) Which one?
- Q2) Why?
‘F’ – which one?

- Pick one to review
- Q1) Which one? -> ‘c’
- Q2) Why?
‘F’ — which one?

• Pick one to review
• Q1) Which one? -> ‘c’
• Q2) Why?

• A2.1: ‘an image is worth a thousand words’
• A2.2: the author respects reviewer’s time
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – A
  – N

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
What makes a good ‘crown jewel’ figure?

- [Consider a paper that tries to find near-duplicate escort ads -> human trafficking ]
'F': What makes a good ‘crown jewel’ figure?

- [Consider a paper that tries to find near-duplicate escort ads -> human trafficking]
What makes a good ‘crown jewel’ figure?

- [Consider a paper that tries to find near-duplicate escort ads -> human trafficking ]
'F': What makes a good ‘crown jewel’ figure?

- Consider a paper that tries to find near-duplicate escort ads -> human trafficking.

If interested call 123456
email: xyzw@

If interested call 987654
email: xyzw@

If interested call 456789
email: xyzw@
'F': What makes a good ‘crown jewel’ figure?

• [Consider a paper that tries to find near-duplicate escort ads -> human trafficking ]

Mock-up text, for victims’ safety

If interested
call 123456
email: xyzw@

If interested
call 987654
email: xyzw@

If interested
call 456789
email: xyzw@
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

A) [Graph showing time vs. number of ads]

B) [Graph showing precision vs. percentage of non-singleton clusters]

C) [Map showing distribution of escort ads]

D) [Diagram showing data flow]

E) [Table showing statistics]

Lessons Learned © 2021-2023 Christos Faloutsos
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

\[
\text{Message??}
\]
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

(A) scalable

\[ f(x) = \frac{3x}{400} \]
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?
What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?
'F': What makes a good ‘crown jewel’ figure?

• Which one a reviewer will find most impressive?

Fig. 1: Pervasiveness of Issue: Distribution of escort ads in

Message?
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

Fig. 1: Pervasiveness of Issue: Distribution of escort ads in
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?
'F': What makes a good ‘crown jewel’ figure?

• Which one a reviewer will find most impressive?

System Overview
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

Result, on real ‘tweets’
**F**: What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

It works.

---

<table>
<thead>
<tr>
<th></th>
<th>Constant</th>
<th>Slot</th>
<th>Insertion</th>
<th>Deletion</th>
<th>Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$ today stats</td>
<td>*</td>
<td>one follower</td>
<td>no unfollowers via <a href="http://apexdiversity.com">http://apexdiversity.com</a></td>
<td>no unfollowers via <a href="http://apexdiversity.com">http://apexdiversity.com</a></td>
<td>no unfollowers via <a href="http://apexdiversity.com">http://apexdiversity.com</a></td>
</tr>
<tr>
<td>#1 today stats</td>
<td>no new followers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2 today stats</td>
<td>followers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#3 today stats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#4 today stats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Result, on real ‘tweets’
'F': What makes a good ‘crown jewel’ figure?

- Which one a reviewer will find most impressive?

A) scalable

B) accurate

C) important

D) elaborate

E) It works

Lessons Learned © 2021-2023 Christos Faloutsos
'F': What makes a good ‘crown jewel’ figure?

• Which one a reviewer will find most impressive?

A) scalable

B) accurate

C) important

Fig. 1: Pervasiveness of Issue: Distribution of escort ads in

D) elaborate

E) It works
What makes a good ‘crown jewel’ figure?

Q: Which one to choose?

A) scalable

B) accurate

C) important

D) elaborate

E) It works
F: What makes a good ‘crown jewel’ figure?

- A: all 3 - it’s OK to have >1 figures

A) scalable

B) accurate

E) It works
'F': What makes a good ‘crown jewel’ figure?

- Emphasize ‘what’ we can achieve – not ‘how’

‘How’

‘What’

D) elaborate

E) It works
'F': What makes a good ‘crown jewel’ figure?

- Emphasize ‘what’ we can achieve – not ‘how’

‘How’
Embed -> cluster -> rank”
“*anybody* could do that – what’s so novel?”

D) elaborate

E) It works
'F': What makes a good ‘crown jewel’ figure?

- Emphasize ‘what’ we can achieve – not ‘how’

Embed: ‘> rank’ "*anybody* could do that – what’s so novel?"

D) elaborate

E) It works
'F': What makes a good ‘crown jewel’ figure?

• Emphasize ‘what’ we can achieve – not ‘how’

E) It works
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – A
  – N

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
'A': Best way to start the abstract?

- Consider a paper that tries to find near-duplicate escort ads -> human trafficking

If interested
call 123456
e-mail: xyzw@

If interested
call 987654
e-mail: xyzw@

If interested
call 456789
e-mail: xyzw@
A): Best way to start the abstract?

A)

Human trafficking is an age old problem that continues to affect 25 million people worldwide.

......
'A': Best way to start the abstract?

A) Human trafficking is an age-old problem that continues to affect 25 million people worldwide.

B) Given a million escort advertisements, how can we spot near-duplicates?

……
'A': Best way to start the abstract?

A) Human trafficking is an age old problem that continues to affect 25 million people worldwide.

B) Given a million escort advertisements, how can we spot near-duplicates?

……
Human trafficking is an age old problem that continues to affect 25 million people worldwide.

Given a million escort advertisements, how can we spot near-duplicates?

......
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – Ask (ask a rhetorical question)
  – N

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
Human trafficking is an age old problem that continues to affect 25 million people worldwide.

Given a million escort advertisements, how can we spot near-duplicates?

‘Rhetorical question’: = you know the answer
'A': Best way to start the abstract?

A) Human trafficking is an age old problem that continues to affect 25 million people worldwide.

B) Given a million escort advertisements, how can we spot near-duplicates?

......

‘Rhetorical question’: = you know the answer

rhetor
orator
'A': Best ‘rhetorical question’:

C)

How can we help law enforcement fight human trafficking?

......

B)

Given a million escort advertisements, how can we spot near-duplicates?

......

‘Rhetorical question’:
= you know the answer
'A': Best ‘rhetorical question’:

C) How can we help law enforcement fight human trafficking?

......

Wrong question. (‘call your congressperson’, ‘deploy face recognition s/w’, ...)

B) Given a million escort advertisements, how can we spot near-duplicates?

......

‘Rhetorical question’: = you know the answer
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – Ask (ask a rhetorical question)
  – N

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
Reminder: our sample problem

- [Consider a paper that tries to find near-duplicate escort ads -> human trafficking ]

If interested
call 123456
e-mail: xyzw@

If interested
call 987654
e-mail: xyzw@

If interested
call 456789
e-mail: xyzw@
'N': Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

~ A. On human trafficking detection  Too general

B. Embedding and clustering for human trafficking detection

C. Fast and accurate human trafficking detection

D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

A. On human trafficking detection

B. Embedding and clustering for human trafficking detection

C. Fast and accurate human trafficking detection

D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

A. On human trafficking detection

B. Embedding and clustering for human trafficking detection

C. Fast and accurate human trafficking detection

D. TrafficSpot: Fast and accurate human trafficking detection

‘What’ – not ‘How’
'N': Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection
 Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection

C. Fast and accurate human trafficking detection
Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection

✓✓ D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

Q: Benefits of a name?

✅ ✅ D. TrafficSpot: Fast and accurate human trafficking detection
'N': Which title is best?

Q: Benefits of a name?
- Easy reference
- .

✓✓ D. TrafficSpot: Fast and accurate human trafficking detection
Which title is best?

A. On human trafficking detection
B. Embedding and clustering for human trafficking detection
C. Fast and accurate human trafficking detection
D. TrafficSpot: Fast and accurate human trafficking detection

Q: Benefits of a name?
- Easy reference
- ‘novelty’

D. TrafficSpot: Fast and accurate human trafficking detection
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – Ask (ask a rhetorical question)
  – Name (give a name to your method/system)

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
'N': How to pick good names?

a) CAE (-> Cluster And Embed)
b) CUBE (-> Cluster and Embed)
c) Spot
d) TrafficSpot
e) InfoShield
'N': How to pick good names?

a) CAE (-> Cluster And Embed)
b) CUBE (-> CLuster and emBEEd)
c) Spot
d) TrafficSpot
e) InfoShield

• Meaningless
• ‘how’
'N': How to pick good names?

a) CAE (-> Cluster And Embed)
b) CUBE (-> CLuster and emBED)
c) Spot
d) TrafficSpot
e) InfoShield
'N': How to pick good names?

a) CAE (-> Cluster And Embed)

b) CUBE (-> CLuster and embEEd)

c) Spot

d) TrafficSpot

e) InfoShield

- English word
- irrelevant
- ‘how’
'N': How to pick good names?

a) CAE (-> Cluster And Embed)
b) CUBE (-> Cluster and emBEEd)
c) Spot
d) TrafficSpot
e) InfoShield
"N": How to pick good names?

a) CAE (-> Cluster And Embed)
b) CUBE (-> ClUster and emBEEd)
c) Spot
d) TrafficSpot
e) InfoShield
'N': How to pick good names?

a) CAE (-> Cluster And Embed)
b) CUBE (-> CLuster and emBEEd)
c) Spot
d) TrafficSpot
e) InfoShield
'N': How to pick good names?

a) CAE (-> Cluster And Embed)

b) CUBE (-> CLUSTER and embed)

c) Spot

d) TrafficSpot

e) InfoShield
'N': How to pick good names?

- a) CAE (-> Cluster And Embed)
- b) CUBE (-> CLuster and emBEEd)
- c) Spot
- d) TrafficSpot
- e) InfoShield
'N': How to pick good names?

- a) CAE (-> Cluster And Embed)
- b) CUBE (-> ClUster and emBEEd)
- c) Spot
- d) TrafficSpot
- e) InfoShield
Recipes for ‘good names’ – part 1/4

clusterEmbed

TrafficLight
Recipes for ‘good names’ – part1/4

1) ‘What’ – not ‘how’

clusterEmbed TrafficLight
Recipes for ‘good names’ – part 1/4

1) ‘What’ – not ‘how’

clusterEmbed  TrafficLight
Recipes for ‘good names’ – part2/4

Heraclitus
TripWire
WireShark
ZX23-S
Let’s take a small break

youtube channels for math/physics, with *superb* presentations//animations:

- 3blue1brown
- Mathologer

… ditto, for python+ coding

- mCoding

Declaration: I am not being funded nor any other way related to the above channels
Recipes for ‘good names’ – part2/4

- Heraclitus
- TripWire
- WireShark
- ZX23-S
Recipes for ‘good names’ – part2/4

Heraclitus

TripWire
WireShark
ZX23-S
Recipes for ‘good names’ – part2/4

Heraclitus
TripWire
WireShark
ZX23-S
Recipes for ‘good names’ – part 2/4

Heraclitus
TripWire
WireShark
ZX23-S
Recipes for ‘good names’ – part2/4

2) Easy to pronounce/remember

Heraclitus
TripWire
WireShark
ZX23-S
Recipes for ‘good names’ – part 2/4

2) Easy to pronounce/remember

- Heraclitus
- TripWire
- ZX23-S
- WireShark
Recipes for ‘good names’ – part3/4

SOAR
SkySoar
3) But NOT an English word (google collisions..)

SOAR

SkySoar
Recipes for ‘good names’ – part3/4

3) But NOT an English word (google collisions..)

SOAR    SkySoar
Recipes for ‘good names’ – part4/4

EaglesEye
MAFIA
Poltergeist
Sherlock
Recipes for ‘good names’ – part 4/4

4) Positive connotation

EaglesEye
MAFIA
Poltergeist
Sherlock
Recipes for ‘good names’ – part4/4

4) Positive connotation

MAFIA
Poltergeist

EaglesEye
Sherlock
Recipes for ‘good names’ – part4/4

4) Positive connotation

MAFIA  EaglesEye
Poltergeist  Sherlock
Recipes for ‘good names’

1) ‘What’ – not ‘how’
   - clusterEmbed
   - TrafficLight

2) Easy to pronounce/remember
   - Heraclitus
   - TripWire
   - ZX23
   - WireShark

3) But NOT an English word (google collisions..)
   - SOAR
   - SkySoar

4) Positive connotation
   - MAFIA
   - EaglesEye
   - Poltergeist
   - Sherlock
Recipes for ‘good names’

1) ‘What’ – not ‘how’
   - clusterEmbed
   - TrafficLight

2) Easy to pronounce/remember
   - Heraclitus
   - TripWire
   - ZX23
   - WireShark

3) But NOT an English word (google collisions..)
   - SOAR
   - SkySoar

4) Positive connotation
   - MAFIA
   - EaglesEye
   - Poltergeist
   - Sherlock
Drill: pick a name!

Suppose that you have developed a method that detects Twitter users that exhibit similar behavior. You use random seeds, and belief propagation, to spot such groups.

- **SSTU** (‘spotting similar twitter users’)
- **BPSim** (Belief Propagation for similar user detection)
- **TwinSpot**
Suppose that you have developed a method that detects Twitter users that exhibit similar behavior. You use random seeds, and belief propagation, to spot such groups.

- **SSTU** (‘spotting similar twitter users’)
- **BPSim** (Belief Propagation for similar user detection)
- **TwinSpot**
Drill: pick a name!

✓✓• CopyCatch (!!) [Alex Beutel+, WWW13]
Drill: pick a name!

✅✅ • CopyCatch (!!) [Alex Beutel+, WWW13]

👉 team up with a native speaker 😊
Summary of summary: F.A.N.

- **Figure** (one or more; ‘what’, not ‘how’)

- **Ask** a ‘rhetorical’ question (and answer it!)

- **Name** your method

  CopyCatch ClusterEmbed
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – Ask (ask a rhetorical question)
  – Name (give a name to your method/system)

• A step back: ‘think like a reviewer’
• More battle scars; and ~10 remedies
• Conclusions
Think like a reviewer

Papers are graded on:

• Relevance
• Novelty
• Technical quality
• Presentation
Think like a reviewer

Papers are graded on:

- **Relevance**
- **Novelty**
  - ‘what’, not ‘how’
  - name
- **Technical quality**
- **Presentation**
Think like a reviewer

Papers are graded on:
• Relevance
• **Novelty**
• Technical quality
• Presentation

Reviewers:
- Pressed for time
- Un-paid
Think like a reviewer

Papers are graded on:

• Relevance
• Novelty
• Technical quality
• Presentation

Reviewers:
- Pressed for time
- Un-paid

Q: what should authors do?
Think like a reviewer

Papers are graded on:
• Relevance
• **Novelty**
• Technical quality
• **Presentation**

Reviewers:
- Pressed for time
- Un-paid

Q: what should authors do?
• A1: save their time (using figures, tables, **emphasis**)
• A2: attention routing
Think like a reviewer

Papers are graded on:
• Relevance
• Novelty
• Technical quality
• Presentation

Reviewers:
- Pressed for time
- Un-paid

Q: what should authors do?
• A1: save their time (using figures, tables, emphasis)
• A2: attention routing

All F.A.N. recipes do that
Outline

• Top 3 lessons: ‘F.A.N.’
  – Figure (‘crown jewel’ figure)
  – Ask (ask a rhetorical question)
  – Name (give a name to your method/system)

• A step back: ‘think like a reviewer’

• More battle scars; and ~10 remedies

• Conclusions
More recipes

• Check template tar-file with ‘orange suggestions’:
  https://www.cs.cmu.edu/~christos/MetaPaper

• `tar xfv; make`
More recipes

- Check template tar-file with ‘orange suggestions’: https://www.cs.cmu.edu/~christos/MetaPaper
- `tar xfv`; make

*rhetorical question:* - What is the best rhetorical question you can start with?
More recipes – part 1 of 4

1. Title:
   ✓ a. **Name your method** (‘what’, not ‘how’)

2. Abstract
   ✓ a. **Ask a rhetorical question**
   b. **Give performance numbers**
More recipes - part 2 of 4

3. Intro:
   a. Again, rhetorical question
   b. ‘crown jewel’ figure
   c. List (bullets) 2-4 contributions
   d. (Informal) problem definition
   e. Give two-word summary for each contribution

4. Literature survey
   a. ‘salesman matrix’: rows are features; columns are baselines
More recipes - part 3 of 4

5. Proposed method
   a. No citations from now on
   b. Clear problem definition
   c. Add theorems/lemmas and proofs (‘QED’)

6. Experiments
   a. Each sub-section should confirm each of the contributions
More recipes - part 3 of 4

7. Conclusions
   a. Repeat the contributions from the intro

8. Globally:
   a. ‘two-word tag’, for every figure/table caption
F.A.Q.

• Are these guidelines mandatory/necessary?

• Are they enough/sufficient?
F.A.Q.

• Are these guidelines mandatory/necessary?
  – NO (but help: + epsilon)

• Are they enough/sufficient?
F.A.Q.

• Are these guidelines mandatory/necessary?
  – NO (but help: + epsilon)

• Are they enough/sufficient?
  – NO (more, every year)
Resources

1. Human-trafficking detection paper:
   http://catvajiac.me/files/infoshield.pdf
   INFOSHIELD: Generalizable Information-Theoretic Human-Trafficking Detection, Meng-Chieh Lee, Catalina Vajiac, et al, ICDE 2021, Chania, Greece

2. Check ‘orange suggestions’ at:
   https://www.cs.cmu.edu/~christos/MetaPaper/
Conclusions – high level

1) Respect reviewers’ time

2) ‘what’, not ‘how’
Conclusions - detailed

- F.A.N. (Figure; Ask question; Name your method)

- (check orange suggestions)
Conclusions - detailed

• F.A.N. (Figure; Ask question; Name your method)

‘ACK: … we used the MetaPaper …’

christos@cs.cmu.edu
https://www.cs.cmu.edu/~christos/MetaPaper/
Conclusions - detailed

- F.A.N. (Figure; Ask question; Name your method)
Conclusions

• F.A.N. (Figure; Ask question; Name your method)

Questions?

Lessons Learned © 2021-2023 Christos Faloutsos