Challenges of capturing engagement on Facebook for Altmetrics

Asura Enkhbayar  |  @AsuraEnkhbayar   |  #ScholCommLab
Introduction
“At the same time, the higher percentage of non-English posts on Facebook overall indicates that people, including non-English speakers, perceive the two platforms differently, with Twitter as a place for discussions with a global public and Facebook a place where more targeted (potentially locally relevant) discussions take place.”

(Barata et al., 2017, p.13)
Introduction

| Study                           | Twitter coverage | Facebook Coverage |
|---------------------------------|------------------|-------------------|
| Thelwall et al., 2013           | 27.7%            | 11.3%             |
| Hammarfelt, 2014                | 20%              | 2.9%              |
| Zahedi and Costas (2018)*       | 57%              | 16.3%             |

Why?

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. PLOS ONE, 13(5), e0197326. https://doi.org/10/gdkbgc

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. PLOS ONE, 8(5), e64841.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. Scientometrics, 101(2), 1419-1430.
Collecting Facebook Metrics

- Two Approaches:
  1. Public posts on public pages
  2. Private engagement through FB API
Facebook’s Graph API

• FB converts URLs to Open Graph Objects (OGB)
  – <meta tags> or heuristics

• URL node to access og_object, engagement

• Each OGB tracks Engagement
  – Shares
  – Likes
  – Comments
  – Comments Plugin
If there were no challenges

1) A document would be identified by a Digital Object Identifier (DOI);
2) Crossref would provide the most recent URL associated with that DOI;
3) the Graph API would be queried with the URL;
4) Facebook would map this URL to their internal identifier system; and
5) it would simultaneously return the number of its engagements
A more complete (but still idealized) scenario

Mapping articles to URLs
- Scientific Article
  - DOI
  - PMC
  - PubMed
- URL for preprint
- Depreciated URL at old domain
- Current URL registered at CrossRef
- Alt. URL 1 (PMC)
- Alt. URL 2 (PubMed)

Mapping URLs to OG objects
- Graph API
  - OpenGraph Object
  - no result
- Graph API
  - OpenGraph Object
  - no result
- FB Engagement
Quantifying the challenges

• **Challenge Area 1: Mapping articles to URLs**
  • Problem Case 1: *Identifying the landing page from any given DOI*

• **Challenge Area 2: Mapping URLs to OG Objects**
  • Problem Case 2: *Equivalent URLs mapped to different OG Objects*
  • Problem Case 3: *Different articles are mapped onto the same Graph Object*

• Dataset: 103k random DOIs from the Web of Science (Piwowar et al, 2017)
Problem Case 1: Identifying the landing page from any given DOI

Dealing with URLs and DOIs is hard (Wass, 2016)

Wass, J. (2016, November 4). URLs and DOIs: a complicated relationship. Crossref website. Retrieved March 21, 2018, from https://www.crossref.org/blog/urls-and-dois-a-complicated-relationship/

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., et al. (2017). The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles. Peerj.
We assume to have determined a set of relevant URLs for each article.
Graph API returns two entities for each URL:
- Open Graph Object
- Engagement Object
Ideally, all URLs should correspond to a single OG object with one identifier and canonical URL.
**Problem Case 2:** Equivalent URLs mapped to different OG Objects

- the URL where the DOI resolved,
- the “opposite” protocol URL (http vs https, and vice versa),
- the currently recommended syntax by Crossref `https://doi.org/[doi]`, and
- the older syntax `http://dx.doi.org/[doi]`. 
CA2: Mapping URLs to OG objects

- **Problem Case 2:** Equivalent URLs mapped to different OG Objects

| Variant | Description                                           | Responses with at least one Ob_ID (N=91490) | Responses positive engagement (N=91490) |
|---------|-------------------------------------------------------|--------------------------------------------|----------------------------------------|
| 1       | URL where DOI resolved*                               | 8,452                                      | 1,426                                  |
| 2       | The “opposite” protocol URL*                          | 13,305                                     | 2,458                                  |
| 3       | The current recommended DOI syntax (https://doi.org/[doi]) | 179                                        | 74                                      |
| 4       | The older DOI syntax (http://dx.doi.org/[doi])       | 10,124                                     | 2,612                                  |
| **All** | Any of the above variants                            | 26,775                                     | 5,498                                  |

*21,871 (23.9%) DOIs resolved to http and 69,619 (76.1%) resolved to https

**Facebook Oddities:**
- The API **always** returns an engagement object
- Some OG objects return engagement of 0
### Problem Case 2: Equivalent URLs mapped to different OG Objects

| Case description                              | Number | Not matching IDs | Matching ID (matching shares) | Matching ID (not matching engagements) |
|-----------------------------------------------|--------|------------------|-------------------------------|------------------------------------------|
| No variant returned an Ob_ID*                 | 106    | -                | -                             | -                                        |
| One variant returns an Ob_ID                  | 3,687  | -                | -                             | -                                        |
| Two variants return an Ob_ID                  | 1,535  | 769              | 620                           | 146                                      |
| Three variants return an Ob_ID**              | 161    | 131              | 99                            | 43                                       |
| Four variants return an Ob_ID**               | 9      | 8                | 6                             | 3                                        |
| **Total**                                     | **5,498** | **908**        | **725**                       | **192**                                  |

*Although it should not be possible to have engagements without having an Ob_ID, we found some instances where this was the case.*

**In some cases, two or three of the Ob_IDS matched, but one or two did not; such cases are counted under all of the appropriate columns.*
CA2: Mapping URLs to OG objects

- Problem Case 3: *Different articles are mapped onto the same Graph Object*

- 66 Ob_IDs (0.2% of 28711) linked to multiple DOIs
- Linked to 507 articles;
- Including 482 of the 5,498 (8.8%)
Summary

• Problem Case 1

12,049 (11.6% of all DOIs)

• Problem Case 2+3

648 (11.8% of those with engagements)

• Total

12,722 (12.3%) of the 103,539 DOIs
Conclusion and Outlook

• First attempt to quantify the difference between public and private engagement on FB
• Further research is needed (and underway) to explore the impact of different URL selections and datasets
• Collaboration of PKP, CrossRef, ImpactStory to build a tool that collects private engagement for
Collecting, Calculating and Displaying Altmetrics with Open Source
Juan Pablo Alperin, Asura Enkhbayar, Heather Piwowar, Jason Priem & Joe Wass

Web mentions of research collected by Crossref

FB mentions of research collected by PKP and the ScholCommLab (under development)

Crossref Event Data turns mentions into events

event stream API

turns events into metrics

ALTMETRICS WITH OPEN SOURCE OPEN DATA FREE FOR EVERYONE

Paperbuzz metrics API
Easy to use and includes aggregated metrics by day, month, year, total

PaperBuzzViz
PKP | PUBLIC KNOWLEDGE PROJECT

drop-in Javascript stand-alone module that pulls metrics from Paperbuzz API and generates graphs that can be inserted anywhere

turns metrics into visualization with d3.js

5AM: Altmetrics Conference, September 28, 2018
Thank you!