
Machine Learning Models for the Measurement of Media Capture

September 20, RTC Talking Methods

*Christopher Barrie (University of Edinburgh), Neil Ketchley (University of Oxford), Alexandra A. Siegel
(University of Colorado at Boulder)*



Research Question(s)

1. What happens to political reporting after media capture?

In plain English: *after media capture, are newspapers less critical of the Executive?*

Why ask this question?

1. Media capture key arm of coercion under authoritarianism
2. Media capture has consequences for attitudes and for possibility of opposition
3. Media criticality is important index of media freedom



Source: <https://www.aljazeera.net/news/humanrights/2015/12/9/>



What do we know already?

Often it's not like 1984...

Media capture rarely *total*:

- a. Strategic benefits to some media liberalization
- b. Media capture takes form of “red lines”
- c. Red line is often the Executive



How do we measure this now?

Two main approaches:

- a. Qualitative studies of media freedom and change
- b. Expert surveys/composite indices of democracy and media freedom (e.g., VDEM; Freedom House)

Case study

- Egypt 2010-2020
- During this time:
 - January 25 Revolution and overthrow of Mubarak (2011)



Source: <https://carnegieendowment.org/sada/48094>

Case study

- Egypt 2010-2020
- During this time:
 - January 25 Revolution and overthrow of Mubarak (2011)
 - period of liberalization (2011-2013)



Source: <https://www.aljazeera.net/news/presstour/2013/7/1>

Case study

الشارع السياسي

التحدي.. "الأهرام" تصدر ملحقاً من 20 صفحة توثق إنجازات الرئيس خلال عامين

15:51 | 5-6-2016

- Egypt 2010-2020
- During this time:
 - January 25 Revolution and overthrow of Mubarak (2011)
 - period of liberalization (2011-2013)
 - period of media (re)capture (2013-)



Source: <https://gate.ahram.org.eg/News/1056266.aspx>

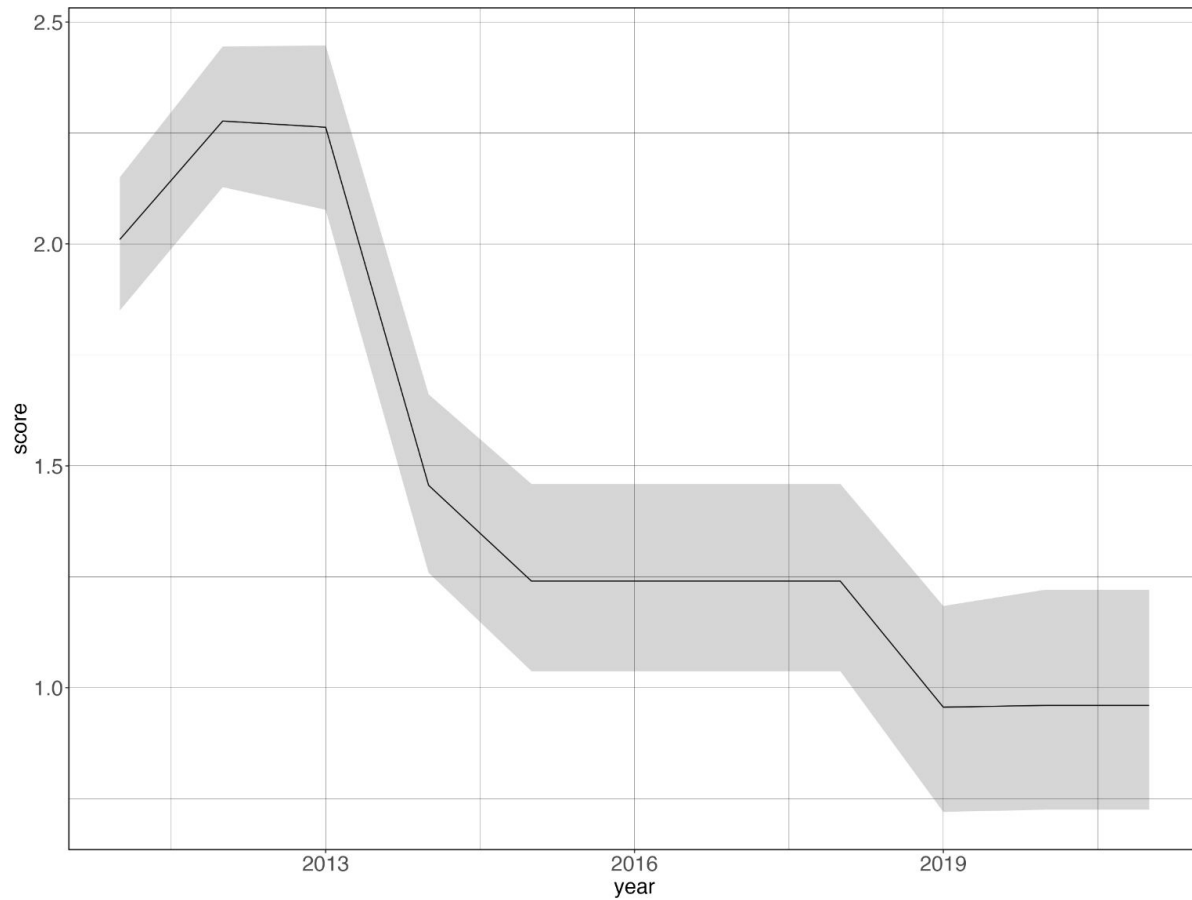


Figure 1: V-Dem “Print/broadcast media critical” measure

Data and Method



Data

- Egyptian news media
- Observation period: January 2010–January 2020
- 79 different newspapers
- All articles from news aggregation platform <https://www.masress.com/>
- ~33m articles in total

Data source

- News aggregation platform Masress
 - English and Arabic versions
 - >90 publications
 - 79 usable
 - Machine labelled
 - Crawled since ~2007 for seven countries in MENA including Egypt



Source: <https://www.masress.com/>



Method:

1. Sample 10% of all publications and train embedding layer
2. Take 10% sample of all articles from valid publications
3. Filter for politics-labelled articles
4. Pre-process (stop words, non-Arabic words, numbers, punctuation)
5. Filter with NER (CaMeL Tools) for articles with a byline
6. Generate set of “Executive” seed words

Table A.4: Dictionary of political terms.

<i>V-Dem theme</i>	<i>Arabic terms</i>
Elections	(مرشح); (تصويت); (انتخاب); (انتخابات)
Political parties	(احزاب); (حزب)
Civil society	(منظمات غير حكومية); (منظمة غير حكومية); (نشطاء); (ناشط); (المجتمع المدني)
Civil liberties	(حقوق); (حرية)
Democracy	(استفتاء); (ديمقراطية)
Executive	(وزراء); (وزير); (المجلس الاعلى للقوات المسلحة); (الإتحادية); (تنفيذي); (رئيس)
Legislature	(مجلس الشيوخ); (مجلس الشورى); (مجلس الشعب); (برلمان)
Judiciary	(قاضي); (مجلس الدولة); (القضاة); (وزارة العدل); (القضاء)

Table A.4: Dictionary of political terms.

<i>V-Dem theme</i>	<i>Arabic terms</i>
Elections	(مرشح); (تصويت); (انتخاب); (انتخابات)
Political parties	(احزاب); (حزب)
Civil society	(منظمات غير حكومية); (منظمة غير حكومية); (نشطاء); (ناشط); (المجتمع المدني)
Civil liberties	(حقوق); (حرية)
Democracy	(استفتاء); (ديمقراطية)
Executive	(وزراء); (وزير); (المجلس الاعلى للقوات المسلحة); (الائتلافية); (تنفيذي); (رئيس)
Legislature	(مجلس الشيوخ); (مجلس الشورى); (مجلس الشعب); (برلمان)
Judiciary	(قاضي); (مجلس الدولة); (القضاة); (وزارة العدل); (القضاء)



Method:

1. Sample 10% of all publications and train embedding layer
2. Take 10% sample of all articles from valid publications
3. Filter for politics-labelled articles
4. Pre-process (stop words, non-Arabic words, numbers, punctuation)
5. Filter with NER (CaMeL Tools) for articles with a byline
6. Generate set of “Executive” seed words
7. Calculate opposition-support vector (معارضة - دعم)
8. Estimate criticality of reporting toward Executive over time by projecting Executive kwics onto criticality index with `get_seq_cos_sim()` in `conText` package



How does this work?

ALC embedding regression approach following Rodriguez et al. (2021)

Uses **corpus-specific embedding layer** + **local transformation matrix** for:

- Efficient estimation of differences in word meaning/context in regression context
- Efficient estimation of over-time change in word meaning

No but actually how does it work?

 : Center Word

 : Context Word

c=0 The cute  jumps over the lazy dog.

c=1 The    over the lazy dog.

c=2      the lazy dog.

Source:

<https://cbail.github.io/textasdata/word2vec/rmarkdown/word2vec.html>


No but actually how
does it work?

A 4-dimensional embedding

cat =>	1.2	-0.1	4.3	3.2
mat =>	0.4	2.5	-0.9	0.5
on =>	2.1	0.3	0.1	0.4

...

...



How does this work?

1. Split observation period into year-week slices;
2. Get context words around our target words for each week.
3. Using ALC approach, then estimate a time-period-specific embedding, by
 - a. taking the average of the vectors of surrounding context words from **pre-trained** embedding layer;
 - b. combine context words and;
 - c. apply transformation matrix to downweight commonly appearing words

Results

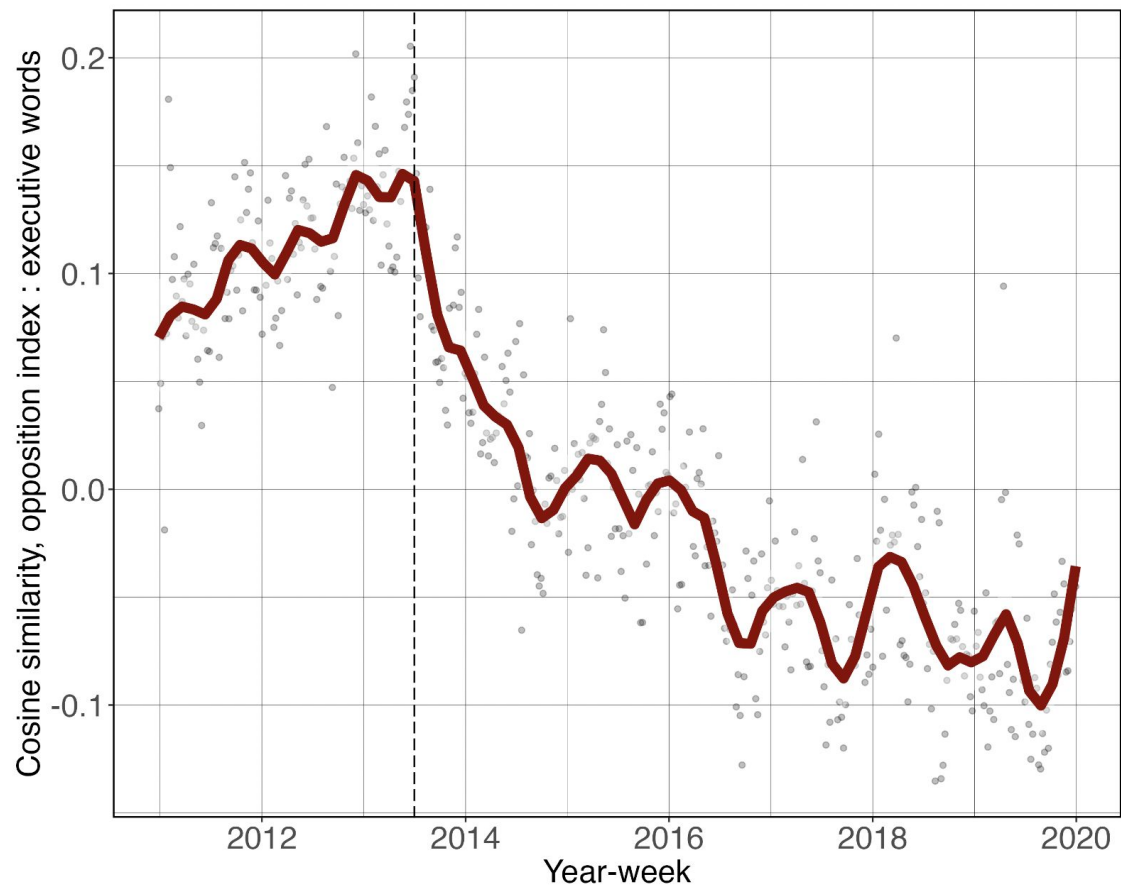


Figure 2: Cosine similarity of Executive words to criticality index over time

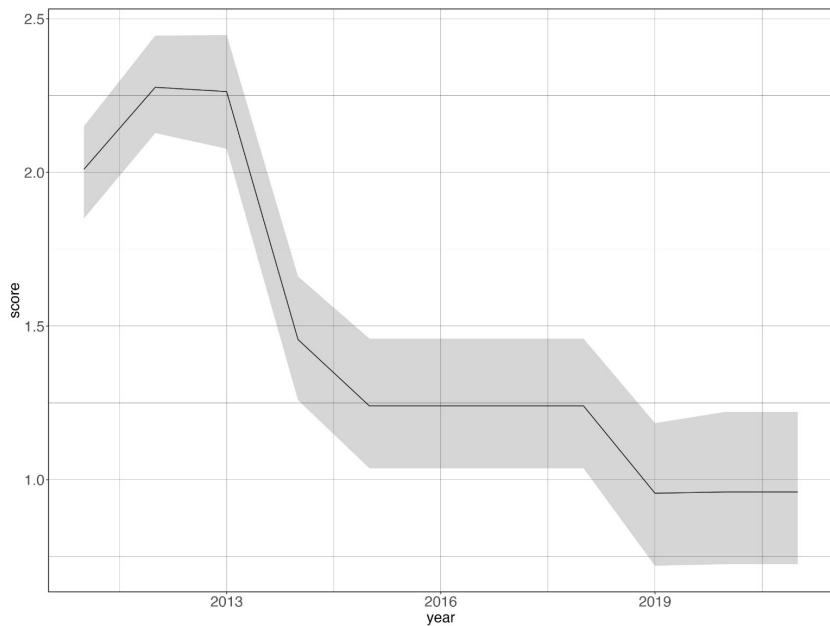


Figure 1: V-Dem “Print/broadcast media critical” measure

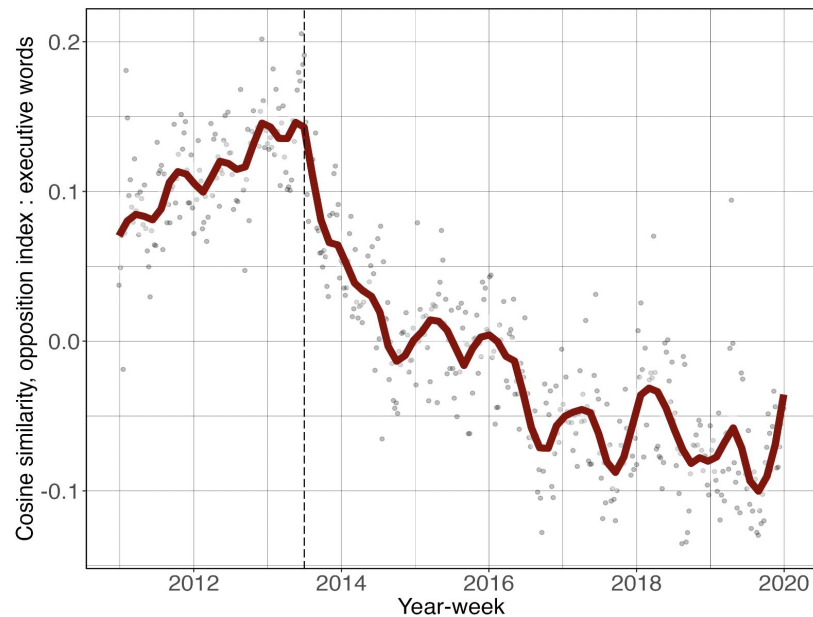


Figure 2: Cosine similarity of Executive words to criticality index over time



In summary

1. Large decrease in media criticality starting at date of coup
2. Efficient method of estimating without expensive surveys
3. Granular and responsive



Next steps



#TODO

- Human coder validation tests
- Using transformer-based model to train e.g. AraBERT



Thank you!

Contact:

-  @cbarrie
-  <https://www.cjbarrie.com/>
-  christopher.barrie@ed.ac.uk