ArCOV-19

The First Arabic COVID-19 Twitter Dataset with Propagation Networks

Fatima Haouari Maram Hasanain Reem Suwaileh
Tamer Elsayed

Qatar University

Webinar - April 20th 2020

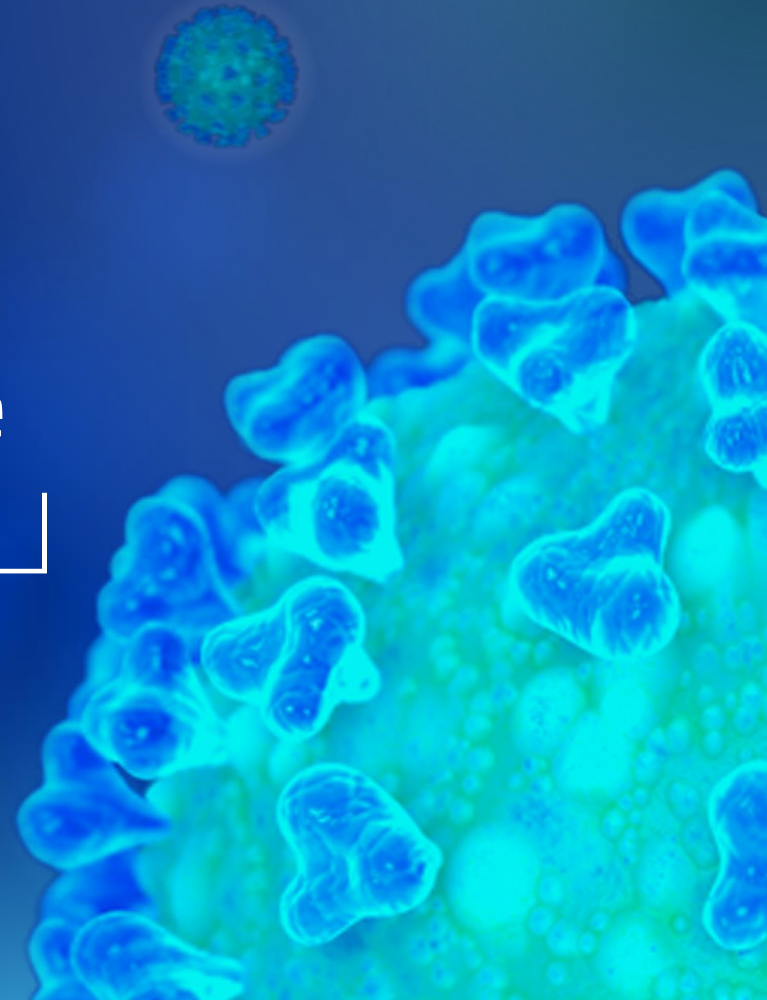
Exceptional moment!

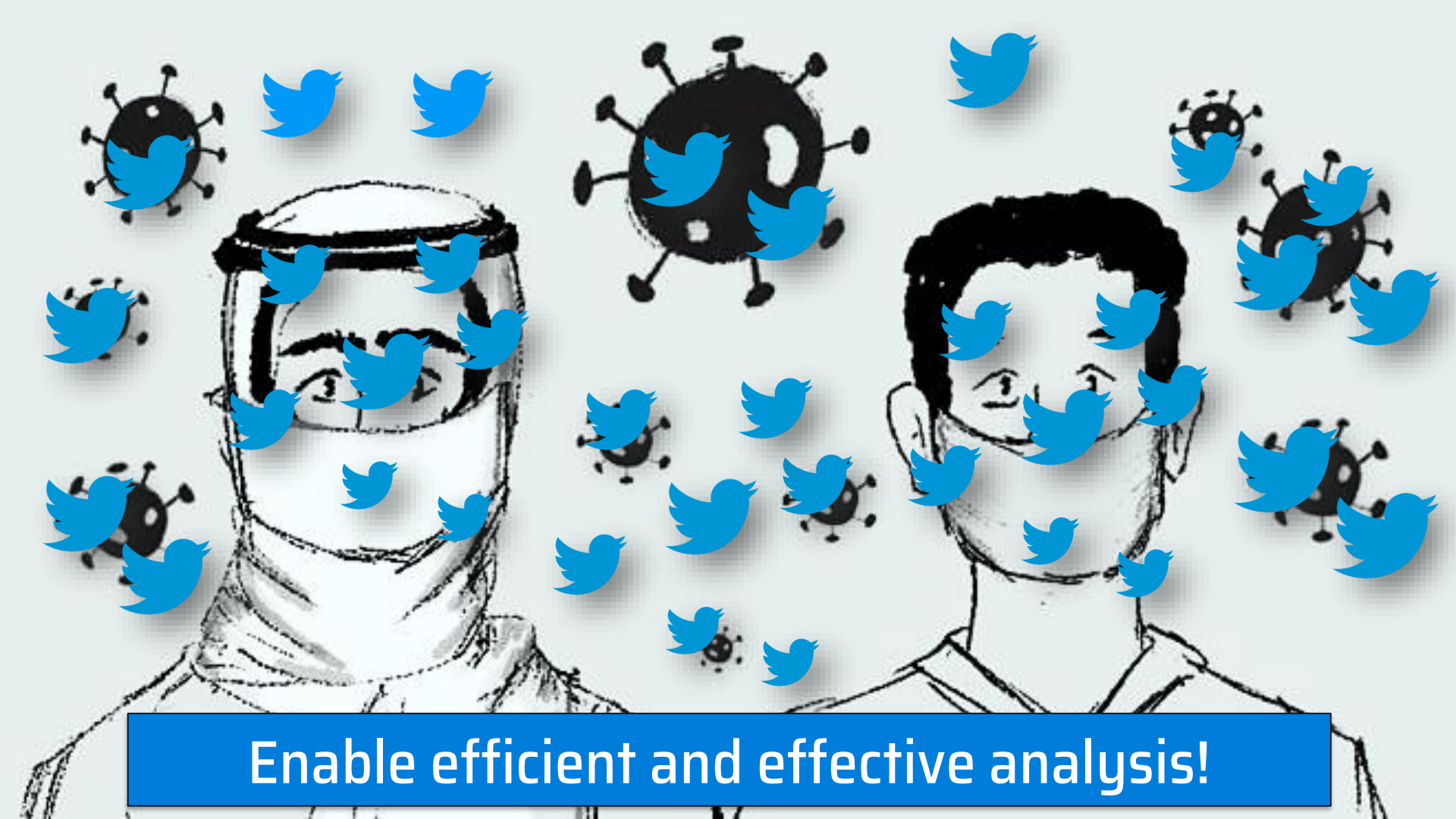
Exceptional history!

History has to be “written” before being “analyzed”!

Contribute to capturing current history in the the Arab world .... to enable efficient and effective analysis!

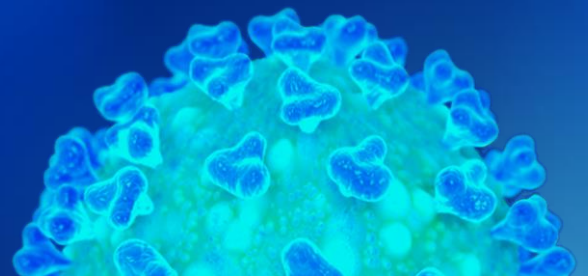
Collect DATA that captures the exceptional history!




Enable efficient and effective analysis!

ArCOV-19

The First Arabic COVID-19 Twitter Dataset
with Propagation Networks



01.

WHY?

Historical moments need historical data!

02.

HOW?

Broad and shallow or deep and narrow?

03.

WHAT RELEASED?

Tweets & code

04.

ArCOV-19?

Tweets, users, topics, ...

05.

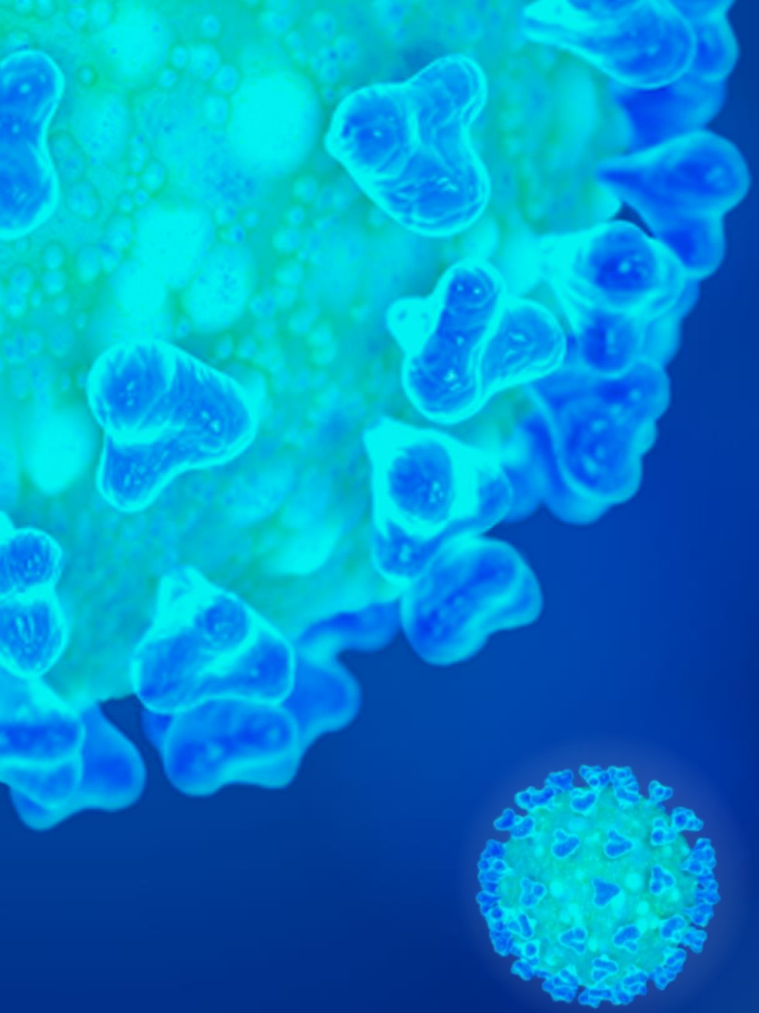
ENABLING RESEARCH

NLP, Text mining, Social Sciences, ...

06.

NEXT?

How to go from here ...



02.

HOW?

Source tweets & Propagation networks

Daily Search ...

since Jan 27th

كوفيد
#فيروس_كورونا
كورونا المستجد
...

#hashtags

keywords

phrases

by Twitter Search API

Get popular tweets

Popular tweets

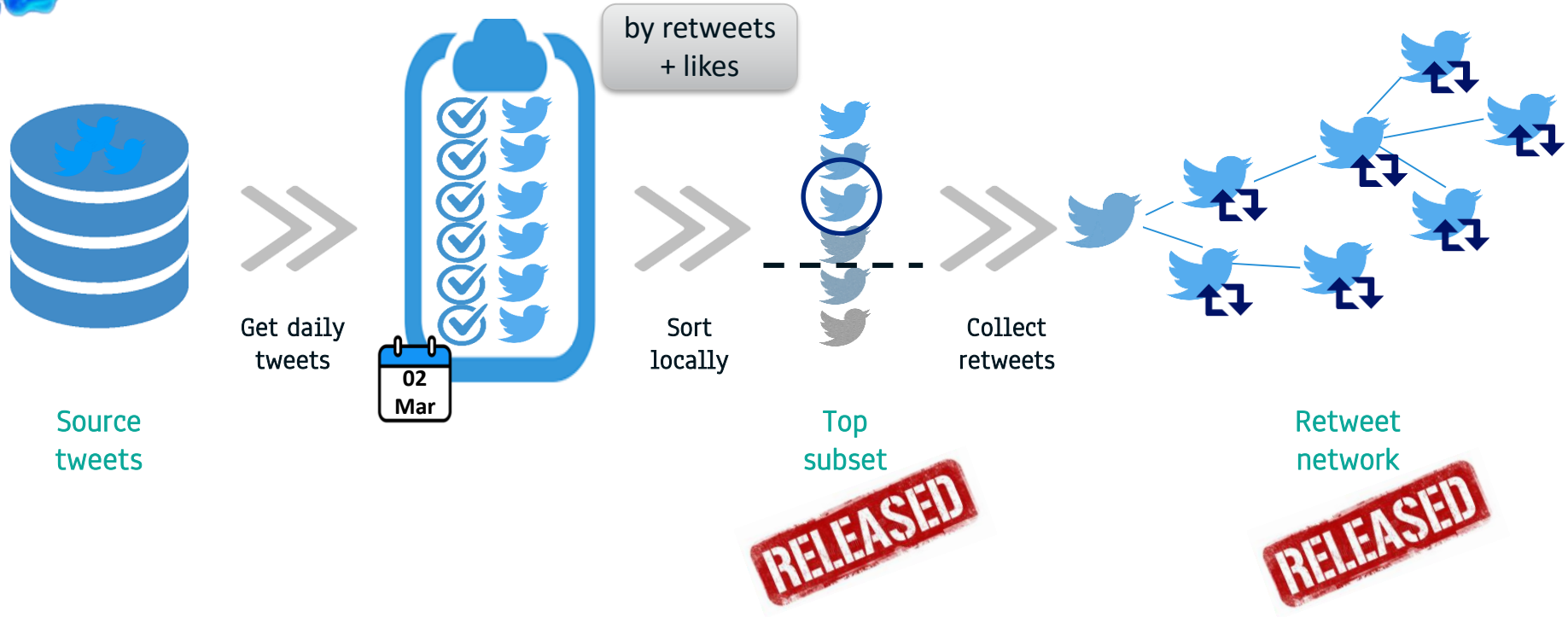
Deduplicate

Sort

Source tweets

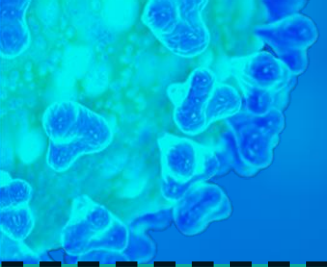
daily ...

Propagation Networks (Retweets)



Propagation Networks (Replies)

Narrow & Deep



All Tweets

Popular Tweets

Top Subset

Retweets

Conversations

Is that typical ?

- No!
  - Typical crawling uses tracking API.
  - Getting “any” tweet that matches the keywords.
  - Subject to spam and “electronic flies”
- Captures influential content
  - Getting “popular” tweets (as defined by Twitter), not “any” tweets.
- Captures propagation networks
  - Retweet networks: spread analysis
  - Reply networks: topical analysis
- Side effects?
  - Sure!

أخبار السعودية
@SaudiNews50
عاجل:
تعليق الدخول إلى السعودية لأغراض العمرة وزيارة المسجد النبوي الشريف مؤقتاً بسبب #كورونا .
Translate Tweet
1:44 AM · Feb 27, 2020 · Twitter for iPhone
6.7K Retweets
10.1K Likes

@Abdullahlbrhim · Feb 27
Replying to @SaudiNews50
استغفر الله
الحمدلله
لا اله الا الله
سبحان الله وبحمده
سبحان الله العظيم
لا حول ولا قوة الا بالله
10
22
Abdulaziz @i6mov · Feb 27
Replying to @SaudiNews50
يلا الحين يجونك المعارضين وشلتهم انظرو السعوديه تمنع المسلمين من دخول مكه
30
16
151
Abdulaziz @i6mov · Feb 27
كما توقعت twitter.com/mhmeyemen/stat...
@nasryemen0086 · Feb 27
السعوديه تمنع تأشيرات العمره وزيارة البيت الحرام ومدينة الرسول الاعظم
للمسلمين بحجة فيروس #كورونا
وتستثني من المنع اليهود والامريكان والاوربيين
21
4
24
@Koz_344 · Feb 27
بموووت
3
@nawaf_subaie · Feb 27
Replying to @SaudiNews50
اللهم انا نعوذ بك من البرص والجنون والجذام ومن سيئ الأسقام
اللهم انا استودعناك أنفسنا وأهالينا وأولادنا وبلادنا وسائر المسلمين
نعوذ بكلماتك التامات من شر ماخلقت
#كورونا
24
150
330
@o_6il · Feb 27
اللهم امين
1
15

@_TH43 · Feb 27
Replying to @SaudiNews50
الضرورات تبيح المحرمات قرار سليم للحفاظ على صحة المواطن والمقيم داخل السعودية والحد من انتشار الفايروس ..
الله يحفظنا
8
12
119
@Almrri97 · Feb 27
أنا بروح العمره الاسبوع الجاي ادخل ولا
23
4
@aj4aj41 · Feb 27
اعتمر عنك خلك بديارك بس
6
@jeejjii · Feb 27
Replying to @SaudiNews50
اللهم إن هذا المرض هو جند من جنودك تصيبُ به من تشاء وتصرفهُ عن من تشاء،
اللهم أصرفه عنا وعن بيوتنا وعن بلادنا وبلاد المسلمين وقنا شر الداء ونجنا من الطعن والطاعون والبلاء بلطفك ياطيف ياخبير إنك على كل شيء قدير، لا حول لنا ولا قوة إلا بك..
آمين
4
24
94
@0m0m70 · Feb 27
يارب امين
2
@7moo_1818 · Feb 27
Replying to @SaudiNews50
اي الله يعطيكم العافيه قرار في محله
1
24
@li_vov · Feb 27
Replying to @SaudiNews50
الله يحمينا استودعت الله جميعنا
1
28

نـورالـمـطـيـري
@Noura_Almoteari
أعتقد أن تركيبة ونشر فيروس #كورونا قطرية بامتياز.. وأن الدوحة دفعت مليارات لزراعة هذا الفيروس المخيف في الصين، بهدف ضرب العام 2020 الذي كان معدا له بدء تحقيق رؤية السعودية 2030 ودبي اكسبو 2020 ونهاية الخلافة العثمانية وتحقيق اتفاق الرياض وعودة السلام للشرق الأوسط.
#قطر_هي_كورونا
Translate Tweet
8:07 PM · Mar 1, 2020 · Twitter for iPhone
1.4K Retweets 5.1K Likes

Ania El Afandi آنيا @Ania27El · Mar 2
Replying to @Noura_Almoteari
بعد هذه التغريدة كورونا يقرر اعتزال كوكب الأرض والانتقال إلى زحل ..اين لا يوجد مثل هؤلاء البشر ..فقد اعترف أخيرا ان هناك أوبئة أكثر فتكا على وجه الأرض..فاستسلم ورحل
20 17 481
نـورالـمـطـيـري @Noura_Almoteari · Mar 2
كتبت لك رد على تغريدتك الصباح ديموزيل آنيا وقمت بمسحها، وربي مشغولة، ما أقدر كل شوي ارد عليك، ما عندكم شغل ؟!؟
26 6 32
Ania El Afandi آنيا @Ania27El · Mar 2
عزيزتي نورا ..انا لو مكانك امسح التغريدة..لانها جعلت الجميع يتعاطف مع قطر ويسخر من كلامك ...يا بنت الحلال(ليس المؤمن بالطعان، ولا اللعان، ولا الفاحش، ولا البذيء)...صدقيني بالله أنني اكتب لك بكل محبة ...
14 3 152
نـورالـمـطـيـري @Noura_Almoteari · Mar 2
ولا يهمك 🙏
6 9
Ania El Afandi آنيا @Ania27El · Mar 2
يومك طيب
4 33
Elsaid Fetih @ElsaidFetih · Mar 3
كل إناء ينضح بما فيه
5
علا الفارس @OlaAlfares · Mar 2
Replying to @Noura_Almoteari
شنهوووووووو 🤔 الذكاء والقدرات الخارقة هذه بالفطرة أو كيف طورتي موهبتك استاذة نورا!!!!

د. زايد العمري @dr_zayedalamri · Mar 1
Replying to @Noura_Almoteari
الكاتبة القديرة | نورا المطيري
لا اوافقك على هذا الاستنتاج فنظام الحمدين اقل بكثير من ان يقوم بمثل هذا العمل . اكبر عمل ممكن ان يقوم به هو تحويل الاموال القطرية الى تركيا او أيران، كجزية لا غير .
370 48 304
نـورالـمـطـيـري @Noura_Almoteari · Mar 1
قطر ليس لديها شخص واحد مؤهل للقيام بالأعمال التخريبية، فعلا هم عاجزون، لكنهم يستأجرون الخبثاء والمرتزقة الذين يقدمون المشروعات تلو الأخرى، والدليل أن عزمي بشارة هو حاكم قطر الفعلي ، الأمر الناهي بسياساتها، لا أستبعد #تورط_قطر_في_تصنيع_كورونا
451 88 344
د. زايد العمري @dr_zayedalamri · Mar 1
طالما اتفقنا انهم اقل من ان يقوموا بذلك فالصين اكبر من ان تخترق مهما ضخ المال القطري للعملاء الرخيصين، ولكن لك الحق ان تشكي في هذا النظام الأفاك الأثيم، حتى تبين مع الأيام الحقائق . تحياتي لك .
155 10 47
محمد بن احمد @malobidy · Mar 1
خوش نقاش 😂😂
44 4 197
JadZain @Bahraini1961 · Mar 1
لا و دكتور بعد ! 😅🙄
8 1 97
Walid Ghanem @walideas · Mar 1
الله لا يعطيكن العافية على هالنقاش.. قال دكتور وكاتبة قال
10 2 150
Romessinho @habbosh909 · Mar 2
😂😂😂😂😂
1 8

خليل المقداد
@Kalmuqdad

تسترت #الصين على حجم وباء #كورونا ففضحهم إعلان يطلب موظفين لجمع الجثث من المنازل ونقلها للمحارق.

الراتب ٥٧٠ دولار لكل وردية عمل ليلية مدتها ٤ ساعات فقط.

في ووهان ٧ مراكز دفن و ٨٤ محرقة تحتوي كل منها على ١١ فرنا تحرق ٣٠٠٠ جثة يوميا.

الحرق على مدار الساعة ولكم تخيل حجم الوباء!

Translate Tweet

7:15 PM · Feb 19, 2020 · Twitter for Android

480 Retweets **1.5K** Likes

خليل المقداد @Kalmuqdad · Feb 19
Replying to @Kalmuqdad

راديو أسيا الحرة أنجز تقريرا تحدث عن حجم الوباء وعمل المحارق على مدار الساعة وإعلان السلطات عن حاجتها لتوظيف أشخاص جدد عدا عن المتطوعين.

Funeral Homes in China's Wuhan 'Working 24/7 to Cremate Bodies'
Funeral homes in the city are overwhelmed amid the coronavirus epidemic, and are advertising for round-the-clock workers at premium...
rfa.org

6 · 55 · 169

خليل المقداد @Kalmuqdad · Feb 20

ليس بعيدا عن الصين!

كوريا الجنوبية تعزل ٢،٥ مليون شخص في دايغو جنوب شرق البلاد بسبب #كورونا بعد ارتفاع حالات الإصابة في المدينة ووفاة مواطنين كوريين كانا على متن السفينة السياحية دايموند برنسيس.

الصحة العراقية تطالب بإغلاق منافذها مع ايران ومنع دخول إيرانيين

وقم الإيرانية تعطل

5 · 17 · 76

🦋بـــوح ♥ آليـآسـمين🦋 @atrnadah · Feb 19
Replying to @Kalmuqdad

اللهم عافينا مما ابتلاهم

18

Emhemed Beneyad @EmhemedBeneyad · Feb 19
Replying to @Kalmuqdad and @RQ_QATAR

يمهل ولا يهمل سبحانه ما اعدله

15

Aizawa Shota 🇪🇬 @QJOb2ix9l5lzI0F · Feb 19
Replying to @Kalmuqdad

اذا 2000 جثة يوميا
الصين قالت وصل عدد الوفيات 2000 لحد الان فقط هههه

5 · 2 · 7

melmelatha&Citroen @melmelatha · Feb 19

متى صدقت الصين حتى تصدق في عدد فطائسها،الم تنكر المرض في بدايانه، الم تنذر طبيب العيون الذي شعر بالمرض ووقعته على محاظر عدم الحديث عن المرض وقد اعتذرت له بعد ان ادرك العالم بتسترها،بل وصل بالهندبالحديث عن حربيبولجية قادمة ،وان الفيروز تسرب من المختبر وان كان اصاه من الخفافيش.

2 · 3 · 25

1 more reply

FSG فراس @AF16454362 · Feb 19
Replying to @Kalmuqdad

العمى كنت عم دور ع شغل لسا عم قول لمرتي ٥٧٠ دولار قالتلي لك رووووج قلتلا بتعرفي لوين قالت وين ماكان قلتلا ووهان قالتلي اذا رحت لاترجع بلاها 🤣🤣🤣

7 · 2 · 68

خليل المقداد @Kalmuqdad · Feb 19
😂😂😂

17

محمد حسان @xYcbbEO2fr9nPwH · Feb 20
Replying to @Kalmuqdad

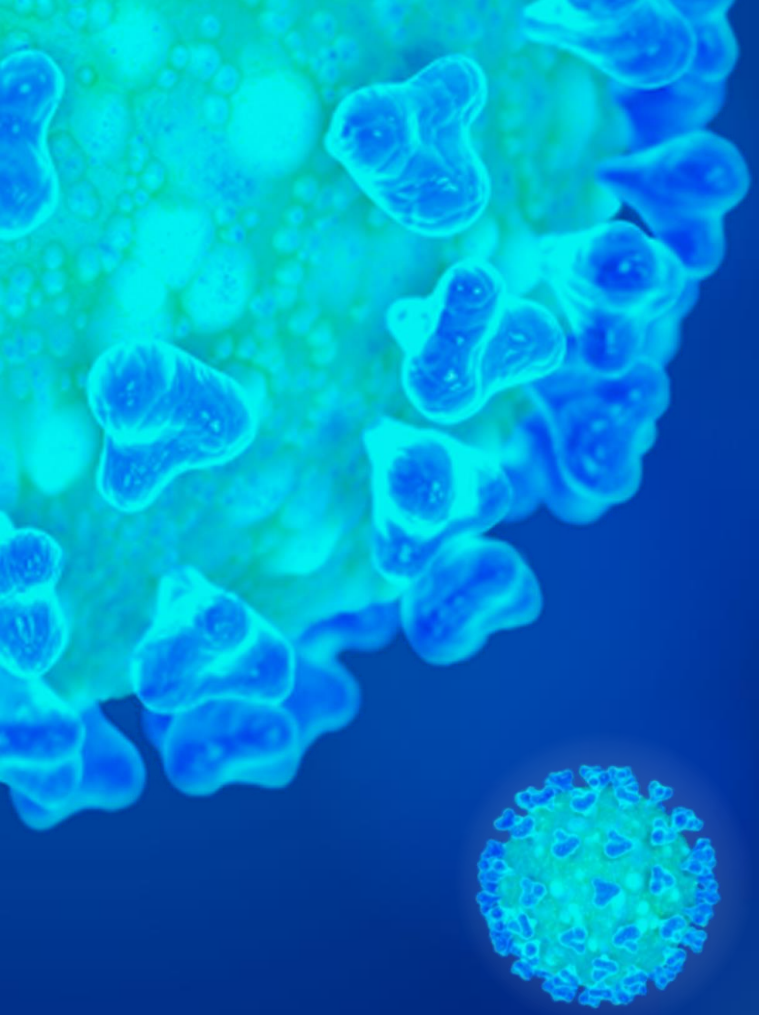
الخوف كل الخوف أن تكون الصين هي السبب في كورونا لأبادة المسلمين الاوغور ولن يلومها أحد باعتبار أنه وباء، نسأل الله أن يرد كيدهم في نحورهم

1 · 10

إبن الثورة @IbnAlthaura · Feb 19
Replying to @Kalmuqdad

اللهم لاتبقى ولا تذر

6



03.

WHAT RELEASED?

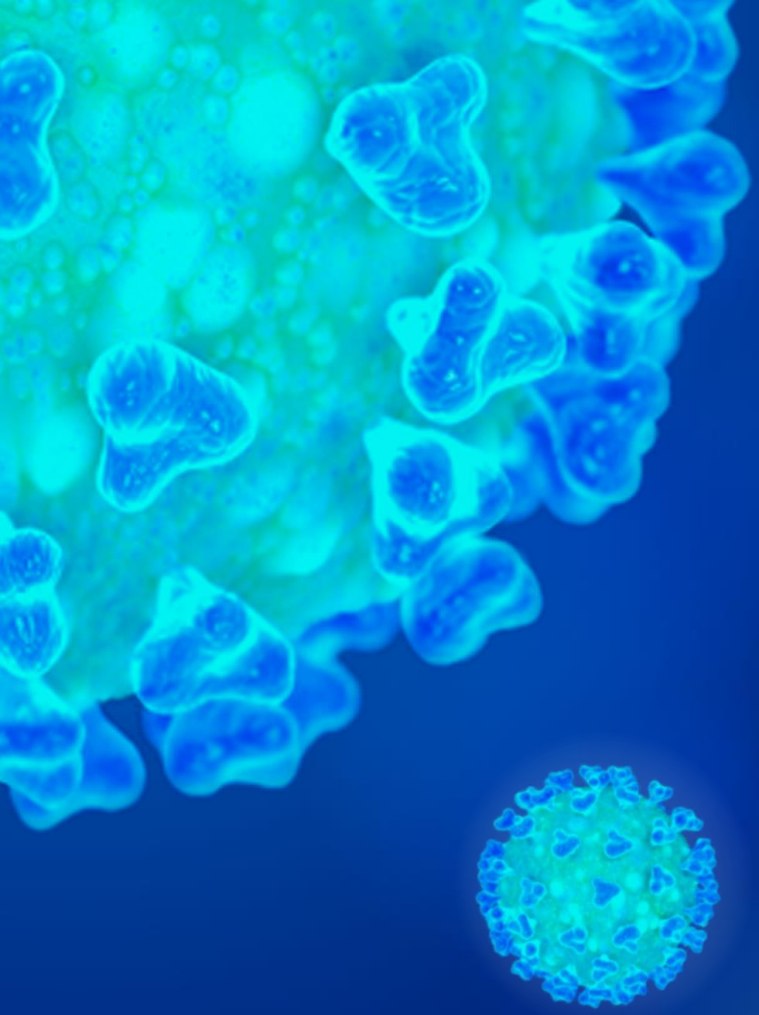
What is released?

- **Dataset**
  - Source tweets (tweet IDs)
  - Top subset (tweet IDs)
    - + retweet networks (tweet IDs)
    - + conversational threads (tweet IDs)
  - Search queries
- **Statistics**
  - Frequent words, hashtags, users, etc.
- **Code**
  - Open-source Crawler

How to get ArCOV-19?

- **Repo**:
  https://gitlab.com/bigirqu/ArCOV-19

- **Documentation & Preliminary Analysis**:
  https://arxiv.org/abs/2004.05861



04.

ArCOV-19?

Tweets Overview

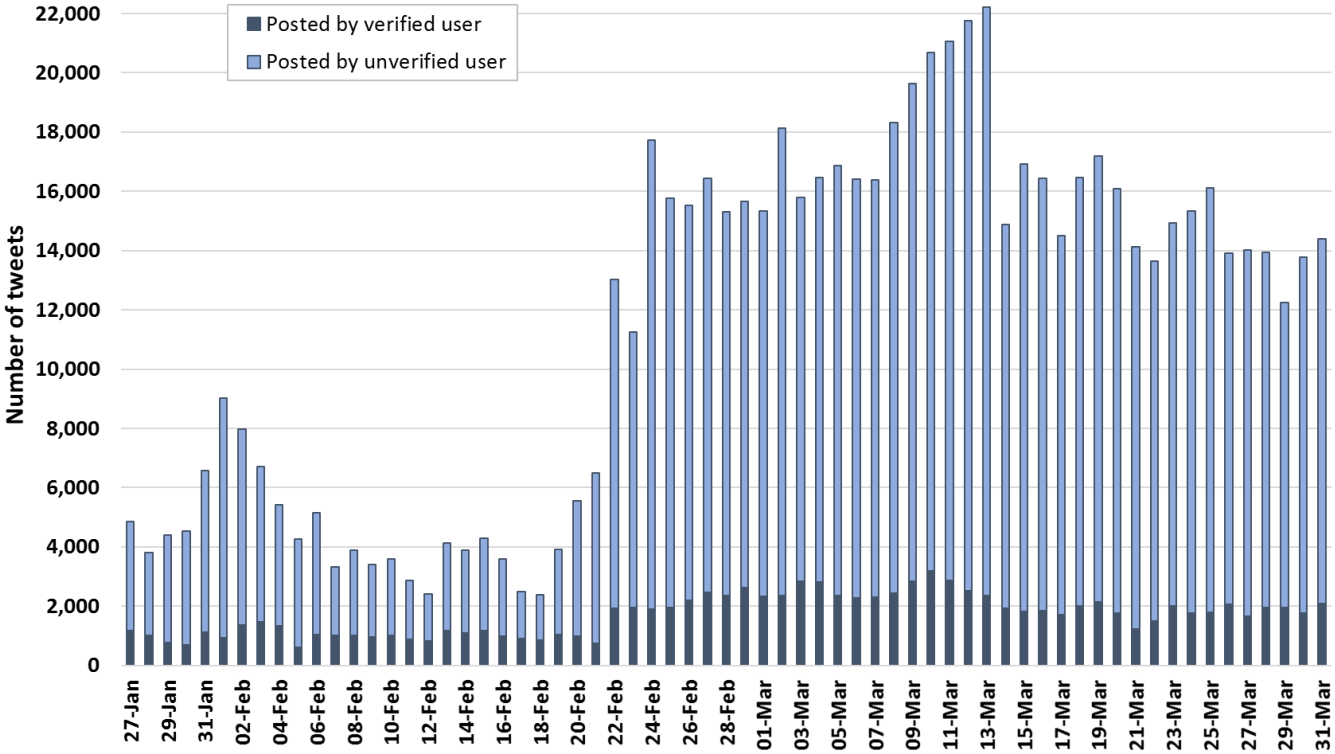
Tweets		
Duration	Jan 27 – Mar 31*	
Source Tweets	747,599	
Geolocated	20,031	(2.68%)
Verified	110,141	(14.7%)
Linked	159,239	(21.3%)
Top Subset	65,000	(8.69%)

*crawling is on going beyond that

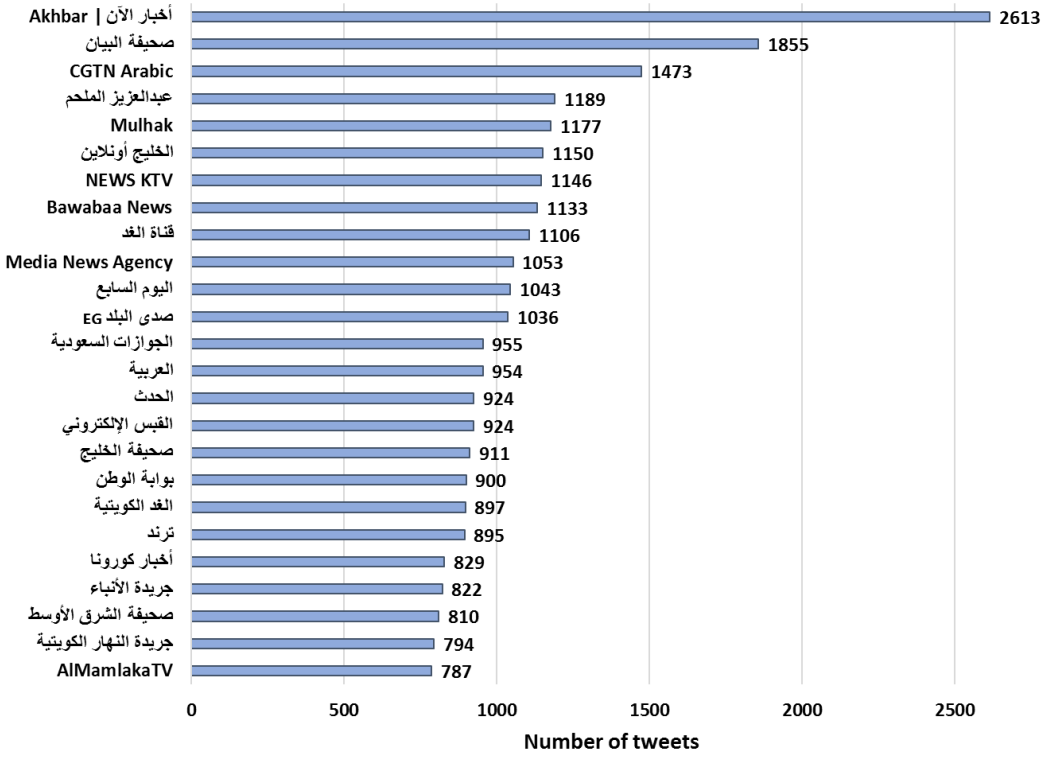
Users Overview

Users		
Unique	248,513	
Verified	3,537	(1.42%)
Average followers	8,334	
Average friends	1,274	
Average statuses	11,201	

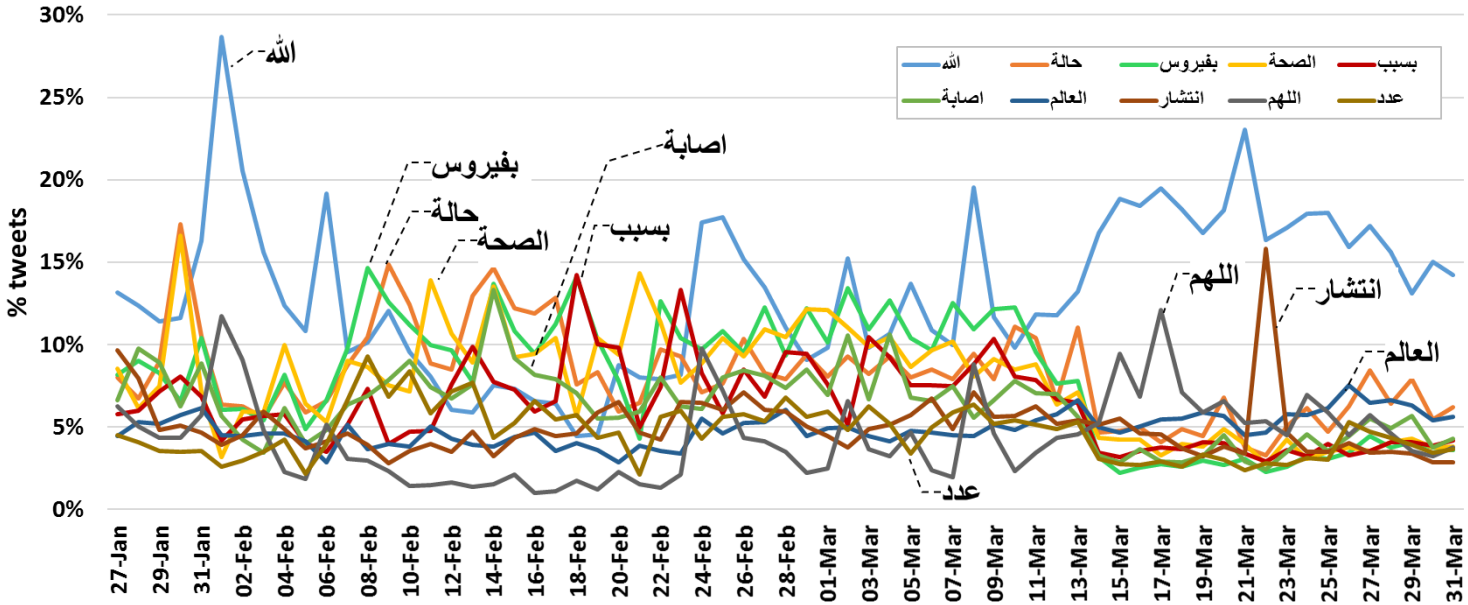
Tweets


Posted by verified user
Posted by unverified user
Number of tweets
22,000
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
27-Jan
29-Jan
31-Jan
02-Feb
04-Feb
06-Feb
08-Feb
10-Feb
12-Feb
14-Feb
16-Feb
18-Feb
20-Feb
22-Feb
24-Feb
26-Feb
28-Feb
01-Mar
03-Mar
05-Mar
07-Mar
09-Mar
11-Mar
13-Mar
15-Mar
17-Mar
19-Mar
21-Mar
23-Mar
25-Mar
27-Mar
29-Mar
31-Mar

Top Tweeters

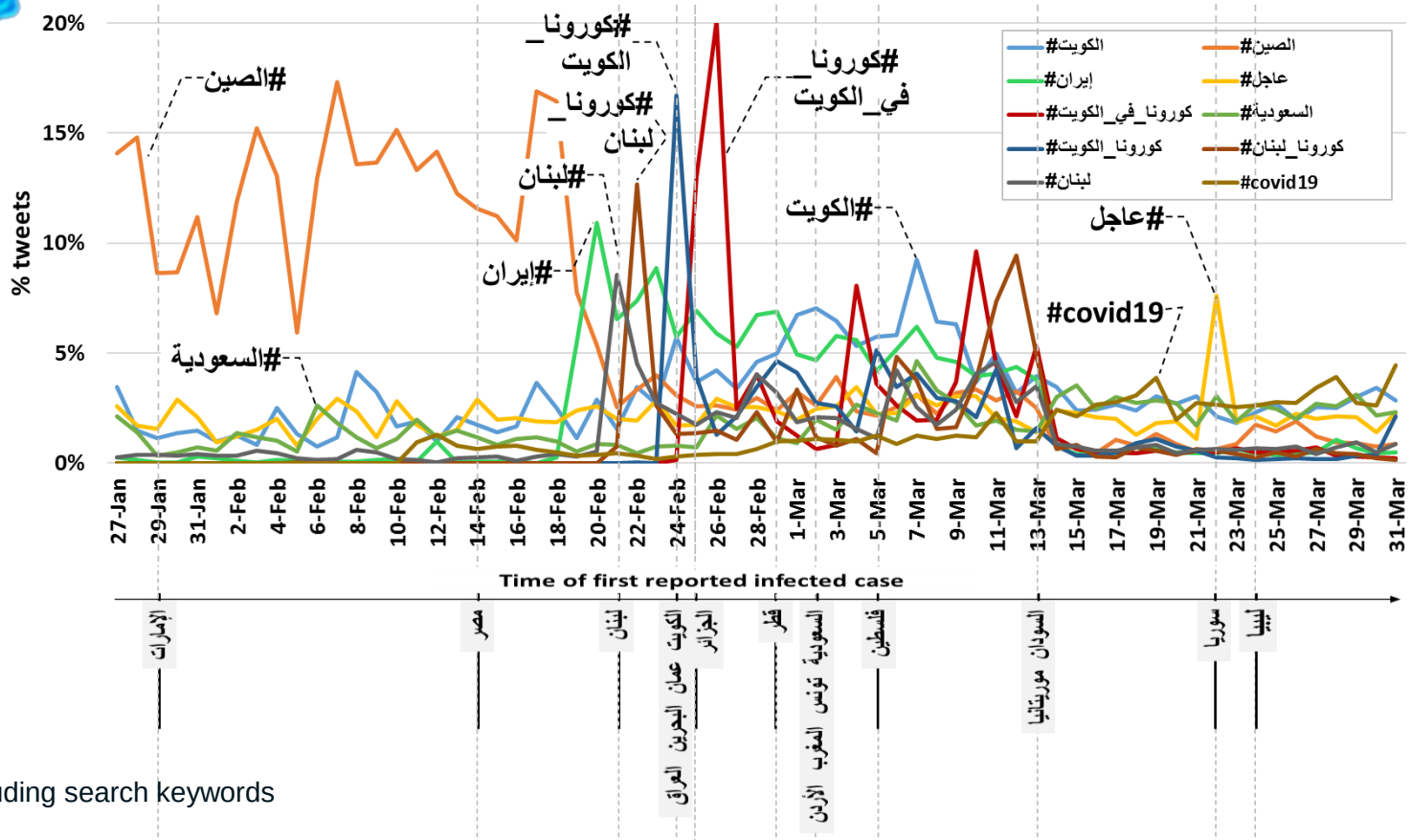


Top Frequent Words*



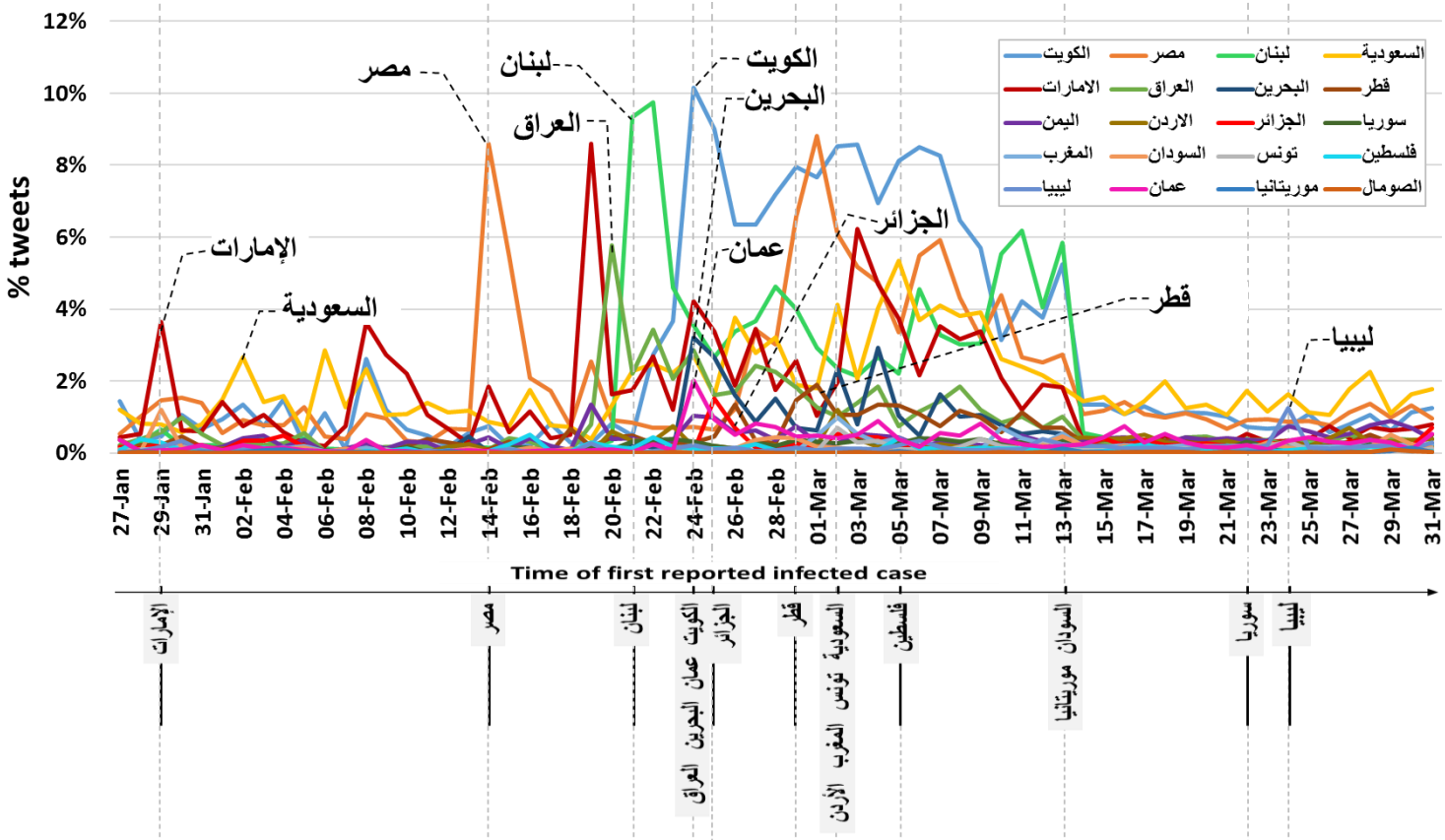
*after excluding search keywords

Top Frequent Hashtags*

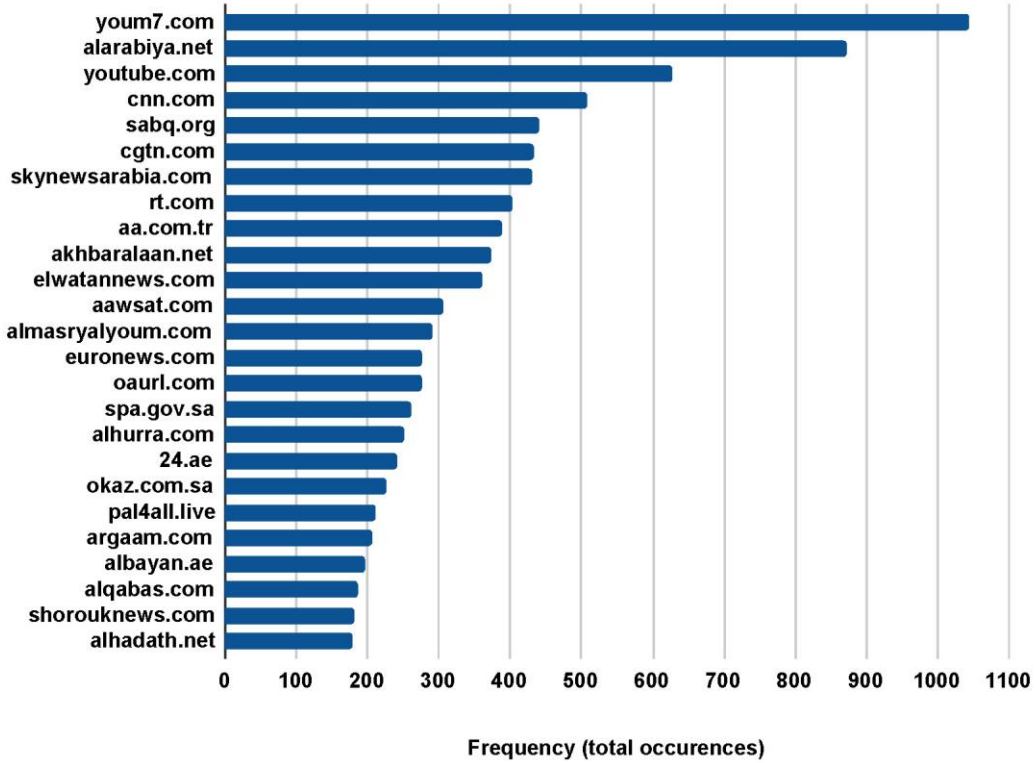


*after excluding search keywords

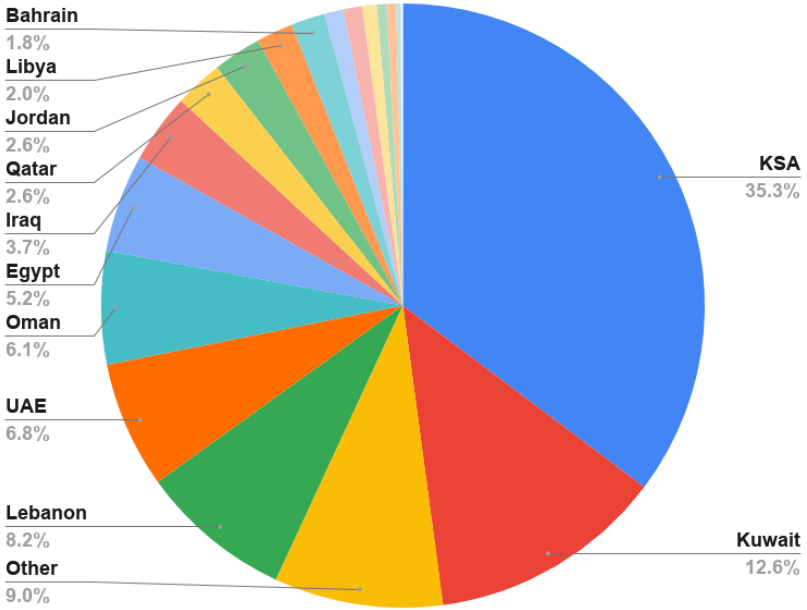
Arab Country Names

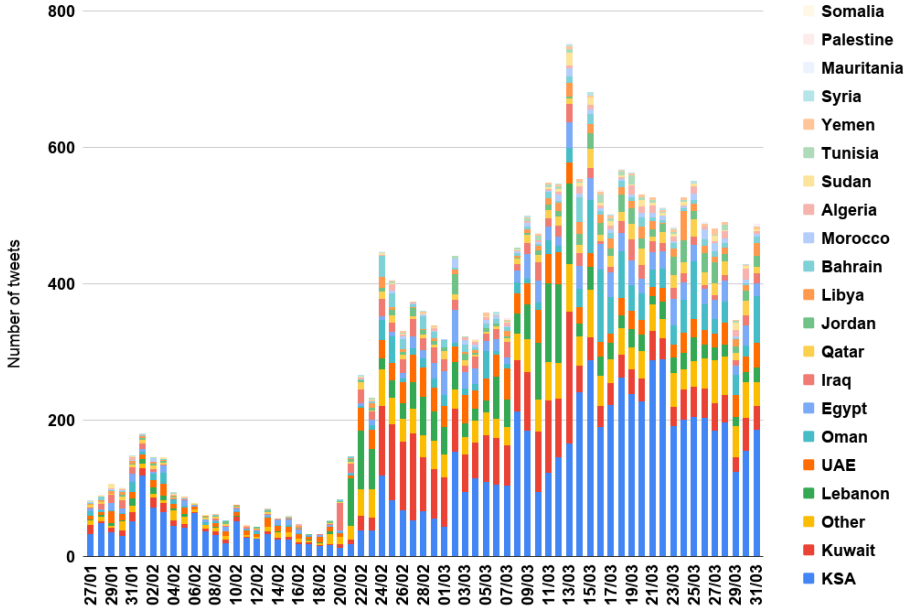


Top Domains in Top Subset

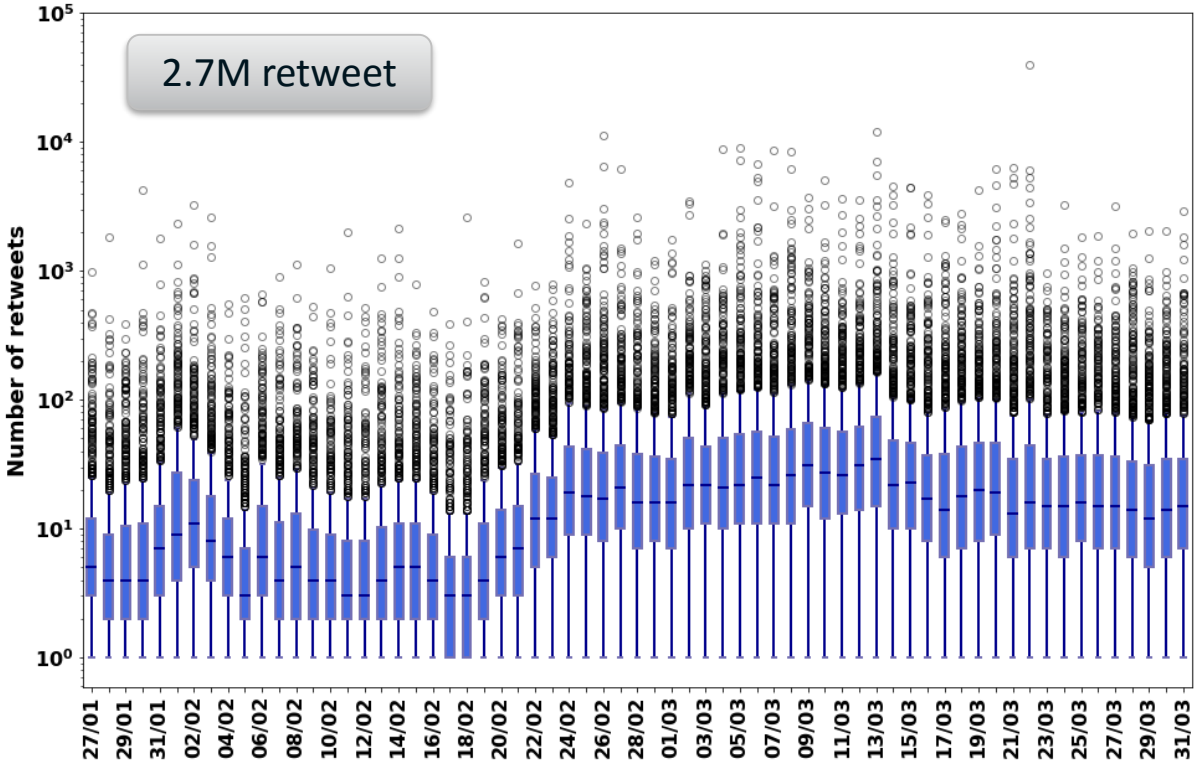


Geographic

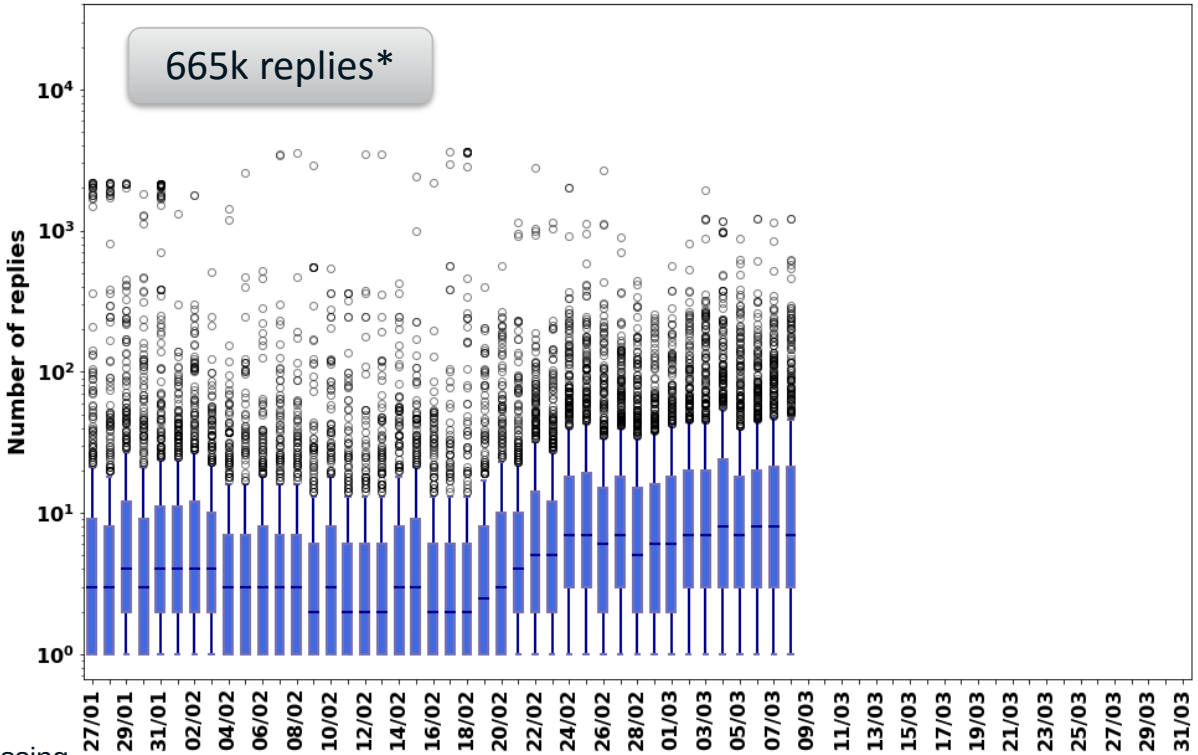

Bahrain
1.8%
Libya
2.0%
Jordan
2.6%
Qatar
2.6%
Iraq
3.7%
Egypt
5.2%
Oman
6.1%
UAE
6.8%
Lebanon
8.2%
Other
9.0%
KSA
35.3%
Kuwait
12.6%


Number of tweets
800
600
400
200
0
27/01
29/01
31/01
02/02
04/02
06/02
08/02
10/02
12/02
14/02
16/02
18/02
20/02
22/02
24/02
26/02
28/02
01/03
03/03
05/03
07/03
09/03
11/03
13/03
15/03
17/03
19/03
21/03
23/03
25/03
27/03
29/03
31/03
Somalia
Palestine
Mauritania
Syria
Yemen
Tunisia
Sudan
Algeria
Morocco
Bahrain
Libya
Jordan
Qatar
Iraq
Egypt
Oman
UAE
Lebanon
Other
Kuwait
KSA

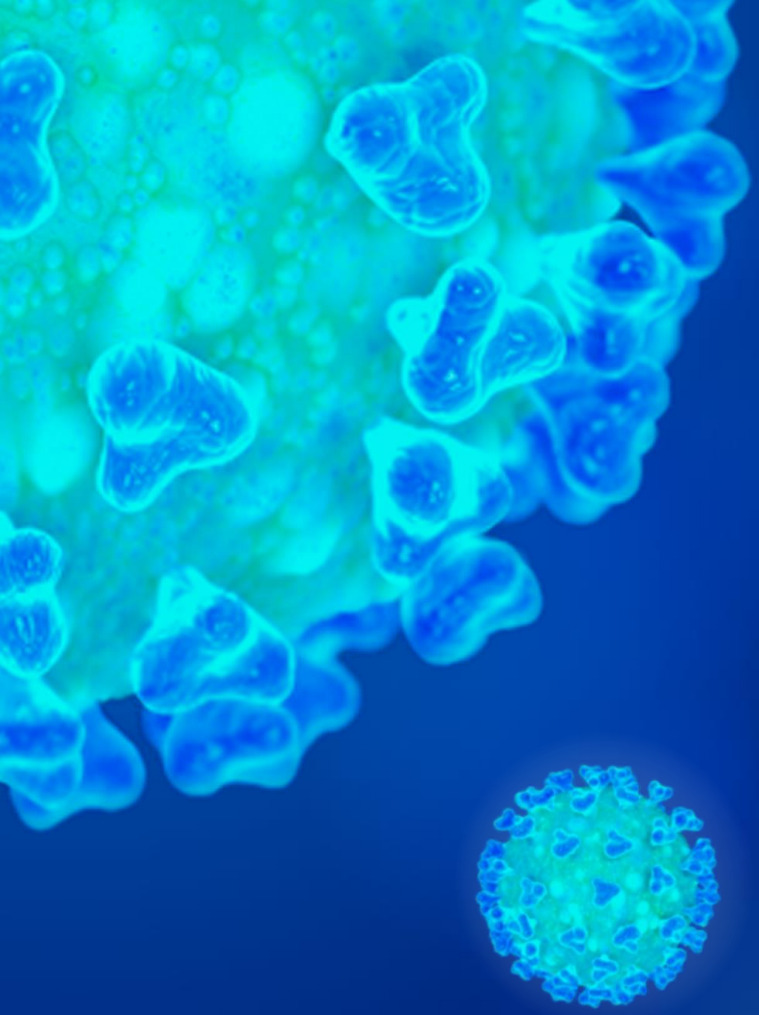
Retweets in Top Subset



Replies in Top Subset



*22 days are still missing

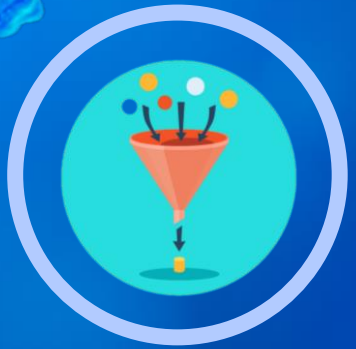


05.

ENABLING RESEARCH ...

Just “examples” of use cases ...

Text Mining



Topic detection
& tracking

Summarization

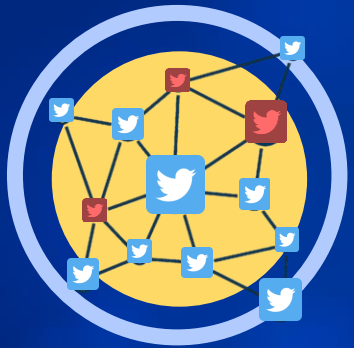
Eye-witnesses
identification

Event & subevent
detection

Geo-location

Misinformation Detection

Fake news / rumor detection



Misinformation propagation

Claim Verification

Propagators identification

Natural Language Processing

Stance Detection

Sentiment Analysis

Offensive Language Detection

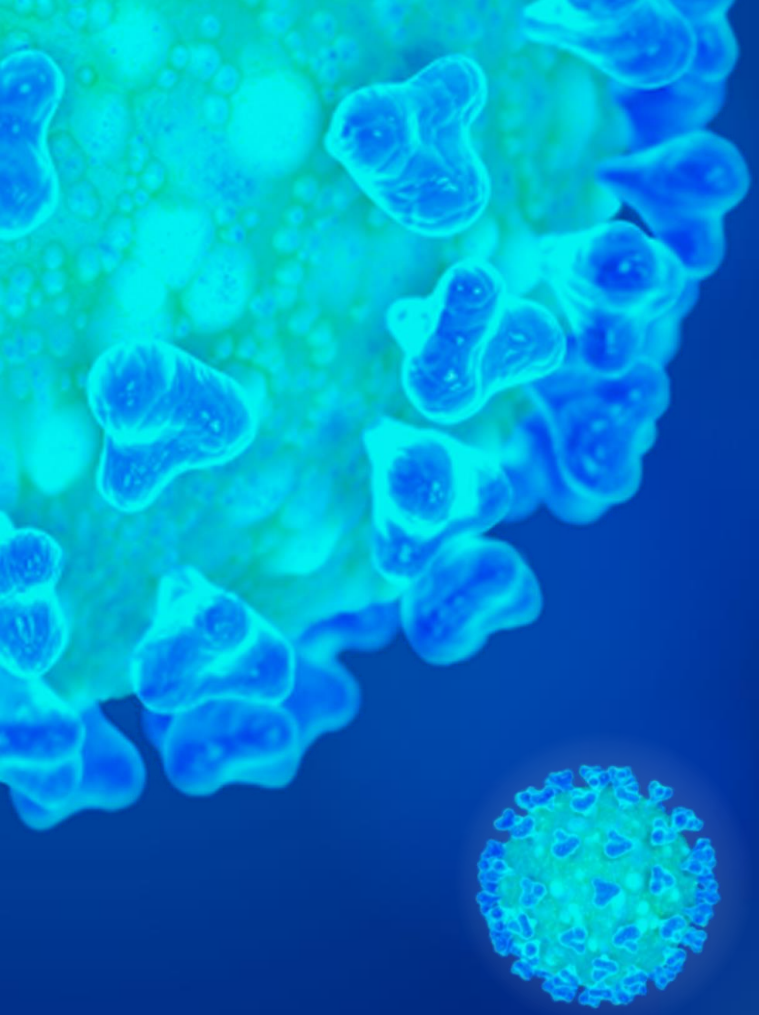
Hate Speech Detection

Social Sciences

Public opinion
& reactions

Topic Identification

Gender Gap



06.

NEXT?

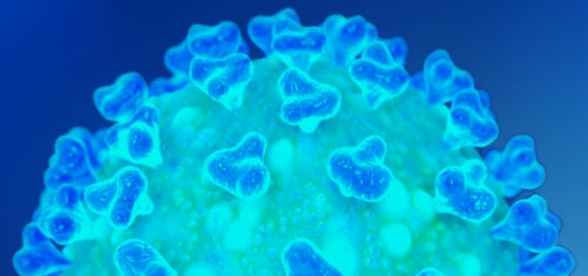
Next …

- Still crawling!
  - Source tweets daily
  - Top subset
  - Propagation networks
- Also tracking ☺
  - Since Mar 7th only
- Web Search interface
- Labeling

- Collaboration …

ArCOV-19

The First Arabic COVID-19 Twitter Dataset
with Propagation Networks



Conclusion

First Arabic COVID-19 Twitter Dataset

With propagation networks

~750k “popular” tweets

Deep and narrow

Good coverage

Matching with events

Enables research in several domains ...

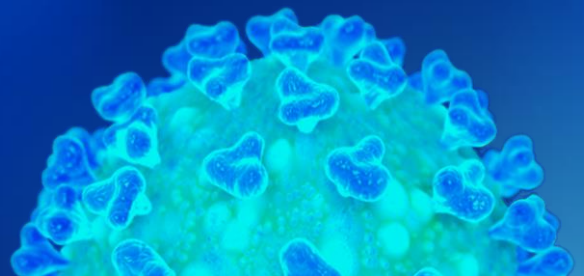
Text mining, NLP, Social Sciences, ...

Public, open source

Tweets & code

On going ...

Check out our repo!



Acknowledgments

الصندوق القطري لرعاية البحث العلمي
Qatar National Research Fund
Member of Qatar Foundation

Thanks!

Do you have any questions?

@bigIR_group
https://sites.google.com/view/bigir
@tamer_elsayed
telsayed@qu.edu.qa
http://qufaculty.qu.edu.qa/telsayed/

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik