

## Data document – Group 3

To evaluate the quality of Twitter accounts, we wish to create networks of users. Our idea is to provide some measures and insights given their accounts' information and their relationship in those networks. In order to accomplish that, first, we use several keywords and query them in Twitter's search with constraints on number of retweets and likes. By doing so, adding the date to the queries provided us a huge number of tweet by the authors who are "so-called" KOLs on that field. After that, we will retrieve the information from their followers in order to build the networks.

In short, our data will contain 3 main parts

1. Crawled tweets
2. Author information from crawled tweets
3. Follower information from authors from part 2.

### 1. Crawled tweets

#### Fields

Name	Datatype	Description	Select?
original_tweet	Dict	Information on the tweet, containing some later fields	
id	Int -> String	Id of the tweet	x
created_on/ date	datetime.datetime	Time when tweet is created	x
author	User	Author of the tweet, details on part 2. For this part, only basic information: <ul style="list-style-type: none"> <li>- ID</li> <li>- Username, Name</li> <li>- Verification status</li> </ul>	x
is_retweet	boolean	If this tweet is a retweet of other	x
retweeted_tweet	Tweet	If this tweet is a retweet of other, then it is a Tweet identity, else NoneType	x
rich_text	Text	rich text included	
text/tweet_body	String	Content of the tweet	x
is_quoted	Same as is_retweet, retweeted_tweet		
quoted_tweet			
is_reply			
replied_to			
is_sensitive	boolean	Sensitive filtered?	x
reply_counts	Int	As named	x
quote_counts	Int		x
bookmark_count	Int		x
vibe/views	String	Vibe of tweet (if any, declared by author)	
language	String	Detected language	
likes	Int	Number of likes	x
place	String/None	Author's geographical location	
retweet_counts	Int	As named	
source	String	Source of quote/media	
audio_space_id	Int -> String	Information of audio space, if available	
voice info	dict		
media	List	List of media items, along with their attributes: <ul style="list-style-type: none"> <li>- URL</li> <li>- ID</li> <li>- Type</li> </ul>	
pool	dict	Viewer pool, if available: question & choices & number of votes	
user_mentions	List	List of mentioned users	

urls	List	List of urls used in tweet	
hashtags	List	List of hashtags used in tweet	x
symbols	List	List of symbols / emotes used in tweet	
community_note / community		Community (Twitter feature)-related posting	
url	String	URL of the tweet	x
threads	List	List, related to tweet	
comments	List		

### Example:

← Post



Microsoft ✓  
@Microsoft

...

Nostalgic, but trendy.



12:33 AM · Sep 28, 2021



```
{'original_tweet': {'bookmark_count': 6, 'bookmarked': False, 'created_at': 'Mon
Sep 27 17:33:02 +0000 2021', 'conversation_id_str': '1442542812197801985',
'display_text_range': [0, 22], 'entities': {'media': [{'display_url':
'pic.twitter.com/2vFhwtBWHL', 'expanded_url':
'https://twitter.com/Microsoft/status/1442542812197801985/photo/1', 'ext_alt_text':
'White background with a bright yellow baseball hat with the MICrosoft Windows 95
logo.', 'id_str': '1442542808548716546', 'indices': [23, 46], 'media_key':
'3_1442542808548716546', 'media_url_https':
'https://pbs.twimg.com/media/FATxCpzXMAI6w69.jpg', 'type': 'photo', 'url':
'https://t.co/2vFhwtBWHL', 'ext_media_availability': {'status': 'Available'},
'features': {'large': {'faces': []}, 'medium': {'faces': []}, 'small': {'faces':
[]}, 'orig': {'faces': []}}, 'sizes': {'large': {'h': 1080, 'w': 1920, 'resize':
'fit'}, 'medium': {'h': 675, 'w': 1200, 'resize': 'fit'}, 'small': {'h': 383, 'w':
680, 'resize': 'fit'}, 'thumb': {'h': 150, 'w': 150, 'resize': 'crop'}}},
'original_info': {'height': 1080, 'width': 1920, 'focus_rects': [{'x': 0, 'y': 0,
'w': 1920, 'h': 1075}, {'x': 276, 'y': 0, 'w': 1080, 'h': 1080}, {'x': 343, 'y': 0,
'w': 947, 'h': 1080}, {'x': 546, 'y': 0, 'w': 540, 'h': 1080}, {'x': 0, 'y': 0,
'w': 1920, 'h': 1080}]}], 'user_mentions': [], 'urls': [], 'hashtags': [],
'symbols': [], 'extended_entities': {'media': [{'display_url':
'pic.twitter.com/2vFhwtBWHL', 'expanded_url':
'https://twitter.com/Microsoft/status/1442542812197801985/photo/1', 'ext_alt_text':
'White background with a bright yellow baseball hat with the MICrosoft Windows 95
logo.', 'id_str': '1442542808548716546', 'indices': [23, 46], 'media_key':
```

```
'3_1442542808548716546', 'media_url_https':
'https://pbs.twimg.com/media/FATxCpzXMAI6w69.jpg', 'type': 'photo', 'url':
'https://t.co/2vFhwtBWHL', 'ext_media_availability': {'status': 'Available'},
'features': {'large': {'faces': []}, 'medium': {'faces': []}, 'small': {'faces':
[]}, 'orig': {'faces': []}}, 'sizes': {'large': {'h': 1080, 'w': 1920, 'resize':
'fit'}, 'medium': {'h': 675, 'w': 1200, 'resize': 'fit'}, 'small': {'h': 383, 'w':
680, 'resize': 'fit'}, 'thumb': {'h': 150, 'w': 150, 'resize': 'crop'}},
'original_info': {'height': 1080, 'width': 1920, 'focus_rects': [{'x': 0, 'y': 0,
'w': 1920, 'h': 1075}, {'x': 276, 'y': 0, 'w': 1080, 'h': 1080}, {'x': 343, 'y': 0,
'w': 947, 'h': 1080}, {'x': 546, 'y': 0, 'w': 540, 'h': 1080}, {'x': 0, 'y': 0,
'w': 1920, 'h': 1080}]}], 'favorite_count': 1753, 'favorited': False,
'full_text': 'Nostalgic, but trendy. https://t.co/2vFhwtBWHL', 'is_quote_status':
False, 'lang': 'en', 'possibly_sensitive': False, 'possibly_sensitive_editable':
True, 'quote_count': 12, 'reply_count': 64, 'retweet_count': 102, 'retweeted':
False, 'user_id_str': '74286565', 'id_str': '1442542812197801985', 'id':
'1442542812197801985', 'created_on': datetime.datetime(2021, 9, 27, 17, 33, 2,
tzinfo=tzlocal()), 'date': datetime.datetime(2021, 9, 27, 17, 33, 2,
tzinfo=tzlocal()), 'author': User(id=74286565, username=Microsoft, name=Microsoft,
verified=True), 'is_retweet': False, 'retweeted_tweet': None, 'rich_text': None,
'text': 'Nostalgic, but trendy. https://t.co/2vFhwtBWHL', 'tweet_body': 'Nostalgic,
but trendy. https://t.co/2vFhwtBWHL', 'is_quoted': False, 'quoted_tweet': None,
'is_reply': False, 'is_sensitive': False, 'reply_counts': 64, 'quote_counts': 12,
'replied_to': None, 'bookmark_count': 6, 'vibe': '', 'views': 'Unavailable',
'language': 'en', 'likes': 1753, 'place': None, 'retweet_counts': 102, 'source':
'Sprinklr Publishing', 'audio_space_id': None, 'voice_info': None, 'media':
[Media(id=1442542808548716546, type=photo)], 'pool': None, 'user_mentions': [],
'urls': [], 'hashtags': [], 'symbols': [], 'community_note': None, 'community':
None, 'url': 'https://twitter.com/Microsoft/status/1442542812197801985', 'threads':
[], 'comments': []}
```

## 2 and 3. User information

Since Twitter API doesn't provide any methods for crawling followers and post likes anymore, we integrated APIs provided on RapidAPI in order to perform the task.

In general, the data used in both these part (2. and 3.) are not different from each other. The only difference is that an extra field about which KOL the user is following is used in part 3.

### Fields:

Name	Datatype	Description	Select?
original_user	Dict	Information on the user, containing some later fields in this table	
id/rest_id	Int	ID & alternate IDs of the user	x
created_on/ date	datetime.datetime	Time when account is created	x
entities	dict	Attached items/URL to the account	
description/bio	String	User short description in bio	
favourites_count	Int	As named	x
fast_followers_count	Int		x
followers_count	Int		x
friends_count	Int		x
has_custom_timelines	boolean	Timeline has customize themes?	
is_translator	boolean/None		
listed_count	Int	Number of list items posted	x
location	String	Location of user	
media_count	Int	Number of media posted	x
name/screen_name/username	String	Real name of user; Username for login (diff)	x

normal_followers_count	Int	As named	x
profile_banner_url	String	Author's profile URLs	
profile_image_url_https	String		
profile_interstitial_type	String		
protected	bool	Protected account	
statuses_count	Int	As named	x
translator_type	String	As named, mostly 'Regular'	
verified	boolean	Verification status	x
can_dm	boolean	How logged in user can interact with this account	
following	boolean		
notifications/notifications_enabled	boolean		
symbols	List	List of symbols / emotes used in tweet	
community_note / community		Community (Twitter feature)-related posting	
possibly_sensitive	boolean/None	Account does post sensitive content	x
pinned_tweets	Tweet/None	As named	
profile_url	String		x

### Example:

←

Microsoft

30.2K posts

...

✉

Follow

Microsoft

@Microsoft

We're on a mission to empower every person and every organization on the planet to achieve more. Support: [@MicrosoftHelps](#)

📍 Redmond, WA
🔗 [aka.ms/ABKCloseXboxWi...](#)
🕒 Born April 4, 1975

📅 Joined September 2009

2,431 Following

13.6M Followers

```
{'original_user': {'can_dm': True, 'can_media_tag': True, 'created_at': 'Mon Sep 14
22:35:42 +0000 2009', 'default_profile': False, 'default_profile_image': False,
'description': "We're on a mission to empower every person and every organization
on the planet to achieve more. Support: @MicrosoftHelps", 'entities':
{'description': {'urls': []}, 'url': {'urls': [{'display_url':
'aka.ms/ABKCloseXboxWi...', 'expanded_url': 'http://aka.ms/ABKCloseXboxWire', 'url':
'https://t.co/YzffF8RFZbI', 'indices': [0, 23]}]}}, 'fast_followers_count': 0,
'favourites_count': 14842, 'followers_count': 13638543, 'friends_count': 2431,
'has_custom_timelines': True, 'is_translator': False, 'listed_count': 27305,
'location': 'Redmond, WA', 'media_count': 3050, 'name': 'Microsoft',
'normal_followers_count': 13638543, 'pinned_tweet_ids_str': [],
'possibly_sensitive': False, 'profile_banner_url':
'https://pbs.twimg.com/profile_banners/74286565/1681173585',
'profile_image_url_https':
'https://pbs.twimg.com/profile_images/1701324249624150016/tcox7Ax3_normal.jpg',
'profile_interstitial_type': '', 'screen_name': 'Microsoft', 'statuses_count':
30288, 'translator_type': 'regular', 'url': 'https://t.co/YzffF8RFZbI', 'verified':
False, 'verified_type': 'Business', 'want_retweets': False,
'withheld_in_countries': []}, {'id': 74286565, 'rest_id': 74286565, 'created_at':
datetime.datetime(2009, 9, 14, 22, 35, 42, tzinfo=tzlocal()), 'date':
datetime.datetime(2009, 9, 14, 22, 35, 42, tzinfo=tzlocal()), 'entities':
{'description': {'urls': []}, 'url': {'urls': [{'display_url':
'aka.ms/ABKCloseXboxWi...', 'expanded_url': 'http://aka.ms/ABKCloseXboxWire', 'url':
'https://t.co/YzffF8RFZbI', 'indices': [0, 23]}]}}, 'description': "We're on a
mission to empower every person and every organization on the planet to achieve
more. Support: @MicrosoftHelps", 'bio': "We're on a mission to empower every person
and every organization on the planet to achieve more. Support: @MicrosoftHelps",
'fast_followers_count': None, 'favourites_count': 14842, 'followers_count':
13638543, 'friends_count': 2431, 'has_custom_timelines': True, 'is_translator':
None, 'listed_count': 27305, 'location': 'Redmond, WA', 'media_count': 3050,
'name': 'Microsoft', 'normal_followers_count': 13638543, 'profile_banner_url':
'https://pbs.twimg.com/profile_banners/74286565/1681173585',
'profile_image_url_https':
'https://pbs.twimg.com/profile_images/1701324249624150016/tcox7Ax3_normal.jpg',
'profile_interstitial_type': None, 'protected': None, 'screen_name': 'Microsoft',
'username': 'Microsoft', 'statuses_count': 30288, 'translator_type': 'regular',
'verified': True, 'can_dm': True, 'following': False, 'community_role': None,
'notifications_enabled': False, 'notifications': False, 'possibly_sensitive': None,
'pinned_tweets': None, 'profile_url': 'https://twitter.com/Microsoft'}
```