

SUPPLEMENTARY MATERIAL

Description of dataset

Processed data files

File name: **events.tsv**

Description: This file is useful for gathering all information about Bounty events. This file contains all information about a thread. The information includes entries from threads_raw.tsv. Threads also have a category and those with category 'Bounty' have additional information that was derived from the main thread post. Each Bounty event is represented by a thread with the Bounty category.

Column Name	Data Type	Required	Description
thread_id	<i>Integer</i>	True	Unique forum thread ID
page_id	<i>Integer</i>	True	Unique comment page ID of the main thread post (it is always 0)
comment_id	<i>Integer</i>	True	Unique comment ID of the main thread post (it is always 1)
user_id	<i>Integer</i>	True	Unique user ID of thread creator
username	<i>String</i>	True	Username of the thread creator
post_time	<i>Timestamp</i>	True	Time when the thread was created
replies	<i>Integer</i>	True	Number of replies in the thread
views	<i>Integer</i>	True	Number of times the thread was viewed
active	<i>Boolean</i>	True	Is the thread currently active (received comments in the past 7 days)
categories	<i>List of Strings</i>	True	Thread category
title	<i>String</i>	True	Thread title
clean_title	<i>String</i>	True	Thread title without redundant information
last_post_time	<i>Timestamp</i>	True	Time when the latest comment was posted
last_post_user_id	<i>Integer</i>	True	Unique user ID of the latest comment author
last_post_username	<i>String</i>	True	Username of the latest comment author
reward_pool	<i>String</i>	False	The size of the thread reward pool
reward_allocation	<i>List of Strings</i>	False	List of campaigns and percentile reward allocation
token_names	<i>List of Strings</i>	False	List of project token names

social_media_urls	List of Strings	False	List of links with social media platform domains
image_urls	List of Strings	False	List of image links embedded in a thread main post
spreadsheet_urls	List of Strings	False	List of google spreadsheets links
forum_urls	List of Strings	False	List of links with bitcointalk.org domain
other_urls	List of Strings	False	List of links which did not match any other category
post_html	List of Strings	True	Comment in HTML format
post_text	List of Strings	True	Comment in text format (picture text embedded)
general_rules	List of Strings	False	General rules of Bounty event
registration_comment_count	Integer	True	Number of registration type comment in the thread
participation_comment_count	Integer	True	Number of participation type comment in the thread
mixed_comment_count	Integer	True	Number of mixed (registration & participation) type comment in the thread
other_comment_count	Integer	True	Number of other type comment in the thread
author_comment_count	Integer	True	Number of author type comment in the thread (i.e., additional comments by the thread creator)

File name: **rewards_and_rules.tsv**

Description: This file includes information about rewards and rules of campaigns found in Bounty events. Each event can have one or more campaigns (e.g., Facebook campaign, Twitter campaign, Signature campaign)

Column Name	Data Type	Required	Description
thread_id	Integer	True	Unique forum thread ID
campaign_title	String	True	Campaign title
rewards	List of Strings	False	List of rewards for a campaign
rules	List of Strings	False	List of rules for a campaign

File name: **comments_registration.tsv**

Description: This file includes registration and mixed (registration & participation in one) comments which is signified by mixed Boolean. Registration comments are posted by forum users to show intent in participating in an event and providing some information about the user.

The fields *activity*, *merit*, and *forum_rank* refer to Bitcointalk.org-specific metrics for each user. See here for information on how they are calculated: <https://bitcointalk.org/index.php?topic=2766177.0>

Column Name	Data Type	Required	Description
thread_id	Integer	True	Unique forum thread ID
page_id	Integer	True	Unique comment page ID
comment_id	Integer	True	Unique comment ID

user_id	Integer	True	Unique user ID of comment author
username	String	True	Username of the comment author
posts	Integer	True	Number of posts created by the user in the forum (extracted from the user forum page)
activity	Integer	True	User activity score. A Bitcointalk.org-specific metric for each user
merit	Integer	True	Amount of merit received from other users. A Bitcointalk.org-specific metric for each user
forum_rank	String	True	Forum rank of the author extracted from the forum user page. A Bitcointalk.org-specific metric for each user
registered_time	Timestamp	True	Time when the user registered in the forum
last_active_time	Timestamp	True	Time when the user produced the latest post in the forum
post_time	Timestamp	True	Time when the comment was posted
mixed	Boolean	True	Comment type: registration (FALSE) or registration & participation (TRUE)
campaigns	List of Strings	False	List of campaigns user registered for in this event
telegram_username	String	False	Telegram username
twitter_username	String	False	Twitter username
bitcoin_forum_rank	String	False	Author's forum rank at the time of the post (extracted from the comment when available, rather than the user's profile page)
post_count	Integer	False	Number of posts created by the user (extracted from the comment when available, rather than the user's profile page)
spreadsheet_number	Integer	False	Spreadsheet line number associated with the user (extracted from the comment when available)
wallet_address	String	False	User cryptocurrency wallet address
other_information	List of String	False	Other unprocessed information included in the comment

File names: **comments_participation.tsv**

Description: This file includes participation comments and mixed (participation & registration in one) comments. These comments are used by forum users to prove their participation in a campaign. This is usually done by providing links to performed actions (e.g., link to Twitter retweet or Facebook post).

The fields *activity*, *merit*, and *forum_rank* refer to Bitcointalk.org-specific metrics for each user. See here for information on how they are calculated:

<https://bitcointalk.org/index.php?topic=2766177.0>

Column Name	Data Type	Required	Description
thread_id	Integer	True	Unique forum thread ID
page_id	Integer	True	Unique comment page ID
comment_id	Integer	True	Unique comment ID
user_id	Integer	True	Unique user ID of comment author
username	String	True	Username of the comment author
posts	Integer	True	Number of posts created by the user in the forum (extracted from the user forum page)

activity	<i>Integer</i>	True	User activity score. A Bitcointalk.org-specific metric for each user
merit	<i>Integer</i>	True	Amount of merit received from other users. A Bitcointalk.org-specific metric for each user
forum_rank	<i>String</i>	True	Forum rank of the author extracted from the forum user page. A Bitcointalk.org-specific metric for each user
registered_time	<i>Timestamp</i>	True	Time when the user registered in the forum
last_active_time	<i>Timestamp</i>	True	Time when the user produced the latest post in the forum
post_time	<i>Timestamp</i>	True	Time when the comment was posted
mixed	<i>Boolean</i>	True	Comment type: participation (FALSE) or participation & registration (TRUE)
social_media_handle	<i>List of Strings</i>	False	List of social media platforms and usernames
campaigns_participated	<i>List of Strings</i>	False	List of campaigns user participated in for this event
twitter_ids	<i>List of Strings</i>	False	List of TwitterIDS in the post
twitter_links	<i>List of Strings</i>	False	List of links with twitter domain and no ID
facebook_links	<i>List of Strings</i>	False	List of links with fakebook domain
instagram_links	<i>List of Strings</i>	False	List of links with instagram domain
telegram_links	<i>List of Strings</i>	False	List of links with telegram domain
reddit_links	<i>List of Strings</i>	False	List of links with reddit domain
youtube_links	<i>List of Strings</i>	False	List of links with youtube domain
medium_links	<i>List of Strings</i>	False	List of links with medium domain
linkedin_links	<i>List of Strings</i>	False	List of links with linkedin domain
discord_links	<i>List of Strings</i>	False	List of links with discord domain
tiktok_links	<i>List of Strings</i>	False	List of links with tiktok domain
steemit_links	<i>List of Strings</i>	False	List all links with steemit domain
image_links	<i>List of Strings</i>	False	List all links with image sharing service domains
other_links	<i>List of Strings</i>	False	All links that were not assigned any of the categories above

File names: ***comments_author.tsv, comments_other.tsv***

Description: These files include *other* and *author* type comments. *Author* comments are posts created by the same user who started the thread. *Other* comments are posts which did not get assigned participation/registration/author category but were detected in Bounty events.

Column Name	Data Type	Required	Description
thread_id	<i>Integer</i>	True	Unique forum thread ID
page_id	<i>Integer</i>	True	Unique comment page ID
comment_id	<i>Integer</i>	True	Unique comment ID
user_id	<i>Integer</i>	True	Unique user ID
username	<i>String</i>	True	Username of the comment author
posts	<i>Integer</i>	True	Number of posts made by the user

activity	<i>Integer</i>	True	User activity score
merit	<i>Integer</i>	True	Amount of merit received from other users
forum_rank	<i>String</i>	True	Forum rank of a user
registered_time	<i>Timestamp</i>	True	Time when the user registered in the forum
last_active_time	<i>Timestamp</i>	True	Time when the user produced the latest post in the forum
post_time	<i>Timestamp</i>	True	Time when the comment was posted
post_html	<i>String</i>	True	Comment in HTML format

File name: **users.tsv**

Description: This file contains information about all users in the subforum Bounties(Altcoins). Note that not all of these users were participants of Bounty events since the subtopic can contain other type of events too. Fields that have required = False are unique to Bounty event participants.

The fields *activity*, *merit*, and *forum_rank* refer to Bitcointalk.org-specific metrics for each user. See here for information on how they are calculated: <https://bitcointalk.org/index.php?topic=2766177.0>

Column Name	Data Type	Required	Description
user_id	<i>Integer</i>	True	Unique user ID
username	<i>String</i>	True	Username of the forum user
posts	<i>Integer</i>	True	Number of posts created by the user in the forum (extracted from the user forum page)
activity	<i>Integer</i>	True	User activity score. A Bitcointalk.org-specific metric for each user
merit	<i>Integer</i>	True	Amount of merit received from other users. A Bitcointalk.org-specific metric for each user
forum_rank	<i>String</i>	True	Forum rank of the author extracted from the forum user page. A Bitcointalk.org-specific metric for each user
registered_time	<i>Timestamp</i>	True	Time when the user registered in the forum
last_active_time	<i>Timestamp</i>	True	Time when the user produced the latest post in the forum
telegram_usernames	<i>List of Strings</i>	False	List of all telegram usernames used in proof of registration comments
twitter_usernames	<i>List of Strings</i>	False	List of all twitter usernames used in proof of registration comments
wallet_addresses	<i>List of Strings</i>	False	List of all cryptocurrency wallet addresses used in proof of registration comments
social_media_handles	<i>List of Strings</i>	False	List of all social media handles using in proof of participation comments
bounty_comment_count	<i>Integer</i>	True	Number of comments (events with Bounty category)
bounty_event_count	<i>Integer</i>	True	Number of events commented in (events with Bounty category)
subforum_comment_count	<i>Integer</i>	True	Number of comments (all subforum)
subforum_event_count	<i>Integer</i>	True	Number of events commented in (all subforum)
bounty_participant	<i>Boolean</i>	True	If True participated in at least one Bounty event

File name: **spreadsheets.tsv**

Description: This file contains all rows of Google spreadsheets that were annotated. Google spreadsheets are often used by Bounty event organizers to track additional information / progress of bounty event participants. Note that the information in the spreadsheets is self-reported, so it is not always well-structured.

Column Name	Data Type	Required	Description
thread_id	<i>Integer</i>	True	Unique forum thread ID
spreadsheet_id	<i>Integer</i>	True	Unique Google spreadsheet ID
line_id	<i>Integer</i>	True	Google spreadsheet row number
title	<i>String</i>	True	Sheet title
timestamp	<i>Timestamp</i>	False	Time of registration for event dd/mm/YYYY HH:MM:SS
proof_post_url	<i>String</i>	False	Link to registration post
forum_rank	<i>String</i>	False	Forum rank of Bounty event participant
forum_profile_url	<i>String</i>	False	Link to forum profile page
forum_username	<i>String</i>	False	Forum username
wallet_address	<i>String</i>	False	Cryptocurrency wallet address
stakes	<i>String</i>	False	Number of stakes received
twitter_followers	<i>String</i>	False	Number of Twitter followers
twitter_audit	<i>String</i>	False	Twitteraudit.com link
post_count	<i>String</i>	False	Number of posts made by the user in the forum
social_media_username	<i>List of Strings</i>	False	List of social media platforms and username
social_media_profile_url	<i>List of Strings</i>	False	List of all social media platforms and profile links
other_column_names	<i>List of Strings</i>	False	Names of unlabeled columns
other_column_data	<i>List of Strings</i>	False	Data of unlabeled columns

Raw data files

File name: ***threads_raw.tsv***.

Description: This file contains information that was collected from each page of the subforum: <https://bitcointalk.org/index.php?board=238>.

Column Name	Data Type	Required	Description
thread_id	<i>Integer</i>	True	Unique forum thread ID
user_id	<i>Integer</i>	True	Unique user ID of thread creator
username	<i>String</i>	True	Username of the thread creator
replies	<i>Integer</i>	True	Number of replies in the thread
views	<i>Integer</i>	True	Number of times the thread was viewed
title	<i>String</i>	True	Thread title
last_post_time	<i>Timestamp</i>	True	Time when the latest comment was posted
last_post_user_id	<i>Integer</i>	True	Unique user ID of the latest comment author
last_post_username	<i>String</i>	True	Username of the latest comment author

File name: ***comments_raw.tsv***

Description: This file contains comments from each thread in the subforum. Thread IDs were retrieved from table *threads_raw.tsv*. Each comment page of a thread can be accessed with: https://bitcointalk.org/index.php?topic=thread_id.page_id

Column Name	Data Type	Required	Description
thread_id	<i>Integer</i>	True	Unique forum thread ID
page_id	<i>Integer</i>	True	Unique comment page ID
comment_id	<i>Integer</i>	True	Unique comment ID
user_id	<i>Integer</i>	True	Unique user ID
username	<i>String</i>	True	Username of the comment author
post_time	<i>Timestamp</i>	True	Time when the comment was posted
post_html	<i>String</i>	True	Comment in HTML format

File name: ***main_post_images_raw.tsv***

Description: Some events use images to convey information. This file contains image links and the text they were converted to using Tesseract OCR. Image links were extracted from the main post of each thread.

Column Name	Data Type	Required	Description
thread_id	<i>Integer</i>	True	Unique forum thread ID
image_url	<i>String</i>	True	Image links
text	<i>String</i>	False	Image converted to text

File name: *spreadsheets_raw.tsv*

Description: Some events used Google spreadsheets as an additional measure to track information. This file contains each row from Google spreadsheets which were retrieved with unique spreadsheet IDs. Spreadsheet IDs were extracted from the main post of each thread.

Column Name	Data Type	Required	Description
thread_id	Integer	True	Unique forum thread ID
spreadsheet_id	String	True	Unique Google spreadsheet ID
title	String	True	Sheet title
line_id	Integer	True	Google spreadsheet row number
lines	List of Strings	True	List of row data in each column

Regular expressions used to identify social media platforms

```
twitter_url = "(https?:\V|(www\.)|https?:\V(www\.))(mobile\.)?(mobil\.)?(m\.)?twitter\.com\V.+"
twitter_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))(mobile\.)?(mobil\.)?(m\.)?twitter\.com\V[^V]+(V)?$"
twitter_audit_url = "(https?:\V|(www\.)|https?:\V(www\.)twitteraudit\.com\V[^V]+)$"
facebook_url = "(https?:\V|(www\.)|https?:\V(www\.))(web\.)?(m\.)?(mobile\.)?(mbasic\.)?(touch\.)?(free\.)?(lm\.)?facebook\.com\V.+"
facebook_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))(web\.)?(m\.)?(mobile\.)?(mbasic\.)?(touch\.)?(free\.)?(lm\.)?facebook\.com\V[a-zA-Z-0-9]+$"
facebook_watch_url = "(https?:\V|(www\.)|https?:\V(www\.)fb\.watch\V[a-z0-9- _]+(V)?$"
telegram_url = "(https?:\V|(www\.)|https?:\V(www\.)t\.me\V.+"
telegram_profile_url = "(https?:\V|(www\.)|https?:\V(www\.)t\.me\V[a-zA-Z0-9 _@]+)"
instagram_url = "(https?:\V|(www\.)|https?:\V(www\.)instagram\.com\V.+"
instagram_profile_url = "(https?:\V|(www\.)|https?:\V(www\.)instagram\.com\V[^V]+)$"
reddit_url = "(https?:\V|(www\.)|https?:\V(www\.))(new\.)?(old\.)?reddit\.com\V.+"
reddit_short_url = "(https?:\V|(www\.)|https?:\V(www\.)redd\.it\V.+$"
reddit_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))(new\.)?(old\.)?reddit\.com\V(r(?:=V))?(user(?:=V))?(u(?:=V))?\V[^V]+)$"
youtube_url = "(https?:\V|(www\.)|https?:\V(www\.))(m\.)?youtube\.com\V.+"
youtube_short_url = "(https?:\V|(www\.)|https?:\V(www\.))(m\.)?youtu\.be\V.+"
youtube_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))(m\.)?youtube\.com\Vchannel\V[^V]+)$"
medium_url = "(https?:\V|(www\.)|https?:\V(www\.))(link\.)?medium\.com\V.+"
medium_profile_url = "(https?:\V|(www\.)|https?:\V(www\.)medium\.com\V@\V[^V]+)$"
medium_profile_subdomain_url = "(https?:\V|(www\.)|https?:\V(www\.))([a-z0-9]+\.)medium\.com\V.+"
linkedin_url = "(https?:\V|(www\.)|https?:\V(www\.))linkedin(\.com|\.cn)\V.+"
linkedin_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))linkedin(\.com|\.cn)\Vin\V[a-z0-9-]+\V$"
discord_url = "(https?:\V|(www\.)|https?:\V(www\.))discord(\.com|\.gg)\V.+"
discord_users_url = "(https?:\V|(www\.)|https?:\V(www\.))discord(\.com|\.gg)\Vusers\V[^V]+)$"
discordapp_channels_url = "(https?:\V|(www\.)|https?:\V(www\.))discordapp\.com\V.+"
discordapp_users_url = "(https?:\V|(www\.)|https?:\V(www\.))discordapp\.com\Vusers\V[^V]+)$"
tiktok_url = "(https?:\V|(www\.)|https?:\V(www\.))(vt\.)?(vm\.)?tiktok\.com\V.+"
tiktok_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))(vm\.)?(vt\.)??tiktok\.com\V@\V[^V]+)$"
steemit_url = "(https?:\V|(www\.)|https?:\V(www\.))steemit\.com\V.+"
steemit_profile_url = "(https?:\V|(www\.)|https?:\V(www\.))steemit\.com\V[^V]+)$"
```

Tests of threshold values

These are the results of the threshold tests for identifying and distinguishing the different comment types. We conducted a manual check of a random sample of 100 comments of each category for different threshold values. Note: In the *Other* column we have also deducted points for registration/participation comments which could not be identified due to the models limitations (for example were written in a language other than English, had many spelling mistakes or did not use any common keywords), however could be identified by visual inspection.

Threshold tests

	Registration	Participation	Other
Threshold + 1	96	100/100	79/100
Current threshold (4)	100/100	100/100	79/100
Threshold - 1	100/100	97/100	81/100