Analysis Of The Jury's Work

1. Introduction

Since its launch on May 07, 2020, Free TON has come a long way. The token reward system for participating in contests has created unique conditions for community members.

Many Free TON contests have generated a lot of interest inside and outside the community, and some of them actively discussed on the forum. With contests in the spotlight, it's no surprise that community members are watching the voting process closely. Doubts about the competence and responsibility of some of the jury have already been expressed many times.

2. Judging problems

The judging of the contests is carried out by the jury members giving points from 1 to 10 for each submission, and there are also opportunities to reject the submission and abstain. In addition, judges can leave comments explaining their decisions. The rating of submissions is formed on the basis of jury points.

The jury receives 5% of the prize fund of the contest, regardless of the number of entries.

One of the problems is that professional qualifications and experience are not taken into account in the selection of jury members, while main Governance contests involve many professional areas. For example, design, software and concept development and even a literary competition (Free TON positioning essay) are all judged by the same panel of judges.

One of the most discussed problems is the high degree of subjectivity in the jury's assessments. In addition, community members suggested that many of the jury did not devote enough time to studying the entries. Perhaps that's why the first works were rated higher and by a large number of jury members. It was noticed that in recent works only the visual design was assessed, and the marks were given by analogy with the estimates of other juries.

3. Studying the problem

The object of my research was the results of voting in the main Governance №42 Free TON Positioning Essay contest (70 submissions). This is an example of a contest without clear criteria for evaluation.

I used data such as: jury ID, application number, jury points and jury comment on the assessment. Data from the gov.freeton.org contest web page was added to the Excel spreadsheet. The data was analyzed according to the following criteria:

- the ratio of unrated submissions to the total number of submissions;
- number of reject;
- mean score;
- standard deviation of the score;
- data offset relative to the average;
- number of unique comments.

Calculation table for jury members of the №42 Free TON Positioning Essay contest:

Jurors	Unrated submissions/ all submissions	Average rates	Standard deviation	Average rates	Skew
0:00000ad ···· 1d2e928f7	0,3	3,257142857	1,854622832	0,5712639509	3
0:0109697 ···· 5b9967b60	0	4,614285714	2,783454517	0,3076406816	3
0:3427ab9 ···· d5ed00940	0	5,185714286	2,578016635	-0,2577081318	4
0:557ac80 ···· abef8093e	0	5,285714286	2,86984753	-0,01585056334	5
0:7288a67 ···· 2d7639c3e	0	4,857142857	2,75726032	0,1566911858	4
0:7b19105 ···· 79b327430	0	4,271428571	2,723602839	0,5656870592	5
0:7dd2b7e ···· b8d6115e0	0,01428571429	5,371428571	2,323386422	0,02286376253	3
0:913b2da ···· edf83a312	0	5,385714286	2,121613121	0,09885012587	4
0:9197e61 ···· a122e5680	0	5,1	2,698093798	0,042264248	4
0:9999999 ···· b205990c4	0	7,6	2,504488724	-0,7373090104	4
0:b1af91c ···· ee44983e7	0	5,371428571	2,456410888	-0,4125100022	5
0:b909fb7 ···· d864f8af0	0,6	2,028571429	2,159021922	-0,1929739947	1
0:e755447 ···· f0c449015	0	3,642857143	2,513936311	-0,06603064944	16
0:f4f2113 ···· a36834c05	0	5,042857143	2,115749871	0,1461486358	4
-1:046374 ···· a8f25cba0	0	5,557142857	2,790437679	-0,1823681112	5
-1:36cd2a ···· 5186d8585	0	5,757142857	2,634419411	-0,1518088321	5
-1:8b0ea5 ···· ff8249946	0	5,3	3,032911262	-0,4344050057	4
		Average submission rating			
		5,206764706			

How can the obtained results be used to evaluate the work of the jury?

First, the percentage of works not assessed by a jury member (unrated submissions / all submissions) is a clear indicator of the jury's participation and fulfillment of their duties.

Second, the number of "rejects" can indirectly indicate how strictly the judge evaluated the competition works. Especially the ratio of refusals of each judge with the total number of refusals for one specific submission and for all submissions.

Third, figures related to the average score of a jury member and variance (Average rates, Standard deviation, Skew) cannot directly assess the performance of a jury member. They can show high or low scores, whether they judged below or above the average score, and how wide the range of scores for judging was used. It depends on the opinion of the jury, whether they consider the level of work to be high or low. However, the standard deviation shows to what extent the jury member influenced the distribution of marks and the formation of the rating of the entries.

Fourth, the analysis of comments provides information about uniqueness and argumentation. Uniqueness indicates the independence and versatility of the judge's assessments. However, uniqueness does not mean information content. Argumentation with reference to the contest requirements is a much more important indicator.

Also, studying the comments reveals a number of facts about refereeing. For example, in the course of analyzing the comments, it was found that judges -1: 046374 ⋯ a8f25cba0 and 0: 557ac80 ⋯ abef8093e left identical comments on submissions 1 through 46.

Detailed comparsion sheet

https://docs.google.com/spreadsheets/d/14na0HvpbPh3bYwaCDhAH_A13DkGXTn9gur3hq0pzCp0/edit?usp=sharing

4. Judging efficiency indicator

To assess the performance of the jury, I used the following formula:

Unrated submissions/all submissions (%) =
(COUNTBLANK(range)+COUNTIF(range;
"=abstain"))/submissions_count*100

Average rates = SUM(range)/submissions_count
Standard deviation = STDEVA(range)
Rejects count = =COUNTIF(range; "=reject")

Unique comments /all submissions (%) =
=COUNTUNIQUE(range)/(submissions_count-COUNTBLANK(range)) *100

Jury involvement calculations were made for the following contests: (lighter color - better, darker - vice versa).

Main Governanace:

27. Slashing condition specification - Developers Contest (11 submissions)
39. Free TON Landing Design (61 submissions);
42. Free TON Positioning Essay (70 submissions)
72. Free TON nomination to the YPO Global Student Entrepreneurship Awards program (9 submissions);
81. Virtual Hero / Spokesperson of Free TON (122 submissions)

Expert	Unrated submissions/ all submissions (%)	Average rates	Standard deviation	Rejects coun	Unique comments / all submissions %
0:00000ad ···· 1d2e928f7	17.55	3.17	2.47	6.25	67.00
0:0109697 ···· 5b9967b60	0.00	4.20	2.74	4.50	59.00
0:08b3740 ···· bf2e9c355	0.00	4.66	2.58	8.00	71.00
0:0d21110 ···· 0dd50f079	30.30	2.85	1.79	10.00	43.67
0:104bfaf ···· ec848ad61	0.00	6.71	2.79	4.67	71.67
0:15bba35 ···· 5db9c5dd9	39.70	2.89	3.80	5.00	57.50
0:3427ab9 ···· d5ed00940	0.73	5.04	2.64	6.25	66.25
0:557ac80 ···· abef8093e	0.00	4.52	2.76	5.50	59.00
0:6833bb2 ···· 70143a1aa	90.90	0.00	0.00	1.00	18.00
0:7288a67 ···· 2d7639c3e	0.00	4.40	2.70	4.50	59.00
0:7b19105 ···· 79b327430	0.00	3.96	2.70	5.50	58.50
0:7dd2b7e ···· b8d6115e0	25.53	3.93	1.84	3.75	79.50
0:8eca6c0 ···· cd447cd8b	90.55	0.39	1.33	5.50	42.00
0:913b2da ···· edf83a312	0.00	5.30	2.53	3.00	69.75
0:9197e61 ···· a122e5680	0.00	6.10	2.94	2.00	84.00
0:9999999 ···· b205990c4	36.88	3.48	1.66	3.00	50.00
0:a2c66fb ···· 43b2c0b0b	30.30	3.14	1.65	6.33	39.33
0:a851b46 ···· 8d2258097	30.83	3.25	2.15	2.67	42.33
0:b1af91c ···· ee44983e7	8.60	4.14	2.75	8.50	78.50
0:b909fb7 ···· d864f8af0	50.02	2.42	1.96	3.60	64.80
0:e755447 ···· f0c449015	18.18	4.02	2.11	9.00	71.00
0:f4f2113 ···· a36834c05	18.18	3.72	2.15	6.00	66.60
-1:046374 ···· a8f25cba0	0.00	4.73	2.71	6.40	76.40
-1:36cd2a ···· 5186d8585	44.45	3.43	1.32	2.50	85.00
-1:8b0ea5 ···· ff8249946	0.00	4.77	2.75	7.40	75.20
-1:9e8683 ···· 5ca52d653	0.00	3.64	2.66	2.00	36.00
-1:de8568 ···· 613e33e86	0.00	3.82	2.96	2.00	45.00
-1:e70a86 ···· 5c1532990	1.25	4.71	2.85	7.50	98.00
-1:f18a64 ···· ef1601efd	0.00	3.82	2.96	2.00	45.00
-1:f6967e ···· 465cd62af	33.33	3.45	1.64	1.67	42.33

DeFi Subgovernance:

Expert	Unrated submissions/ all submissions (%)	Average rates	Standard deviation	Rejects coun	Unique comments / all submissions %
0:0109697 ···· 5b9967b60	86.5	0.92	2.17	0	100
0:1c8de90 ···· a752afabb	51.4	1.11	3.4	12	76
0:1db4191 ···· 7c1833105	97.3	0	0	1	100
0:3433762 ···· 052ee5c5c	86.5	0.65	4.09	1	100
0:345309e ···· 2b54dcd88	91.9	0.81	0	0	33
0:418d174 ···· 209b8b91e	45.9	2.14	4.41	9	95
0:583228a ···· 0adaed702	89.2	0	0	4	18
0:7a63873 ···· 641109c97	94.6	0.43	1.41	0	100
0:7dd2b7e ···· b8d6115e0	62.2	2.51	3.93	3	86
0:93696e3 ···· 111c5c480	86.5	0.81	2.45	0	100
0:a6ae6b3 ···· 10bb6be66	86.5	0.76	2.7	0	100
0:deb918b ···· 489b052be	51.4	1.54	3.45	9	95
0:e2b874e ···· a5ea895c3	89.2	0.38	4.73	2	75
0:e755447 ···· f0c449015	45.9	2.24	3.25	6	100

SMM Subgovernance:

Expert	Unrated submissions/ all submissions (%)	Average rates	Standard deviation	Rejects coun	Unique comments / all submissions %
0:28f1f2b ···· 1ab126f49	16.4	6.02	3.52	4	90
0:30d4123 ···· 93185b9ea	70.5	2.03	1.88	0	100
0:389cba3 ···· 5c2a70672	85.2	1.13	4.98	2	100
0:4cb8ec1 ···· 52102739f	34.4	5.74	3.13	1	73
0:85414bd ···· 7fbede6c7	57.4	2.9	3.77	1	73
0:9fa3e49 ···· bb09443ec	49.2	3.49	3.2	0	68
0:a734f38 ···· b9545e920	98.4	0.16	0	0	100
0:ad30271 ···· e81b128cd	32.8	4.56	3.97	5	77
0:b9ec5a0 ···· c8d12ec14	60.7	3.11	4.11	3	96

DevEx Subgovernance:

Expert	Unrated submissions/ all submissions (%)	Average rates	Standard deviation	Rejects coun	Unique comments / all submissions %
0:4e4649f ···· 47e72e29c	87.7	0.46	2.42	5	14
0:583228a ···· 0adaed702	80	1.26	3.86	3	100
0:5b833e3 ···· 1417b3279	78.5	1.31	4.3	4	100
0:625e7dd ···· 9327ec7b9	93.8	0.31	5.48	2	80
0:7075902 ···· 290b41d6a	76.9	1.31	4.33	5	56
0:842e953 ···· cf4355332	92.3	0.46	5.48	2	100
0:a03a62d ···· 6147c9cdb	92.3	0.31	2.97	2	40
0:b6ed2e8 ···· d4247820b	84.6	0.92	4.24	3	100
0:d2cd1ff ···· 97465bed2	78.5	1.48	4.03	1	100
0:e613ada ···· 78d5d8236	92.3	0.46	4.39	2	46
0:ef0602f ···· bc5b0e3af	93.8	0.62	0	0	75

Link to Google Drive sheet:

https://drive.google.com/file/d/1FpqMZi5AbcezqLYPwpNm1F6Sm0mpeCPh/view?usp=sharing

5. Conclusions and suggestions

I can conclude that on the basis of the available data, there is no way to assess the effectiveness of the jury's work, but only indirectly to demonstrate the degree of their participation in the voting process.

To improve the efficiency and transparency of the jury's work, I suggest the following:

- not to conduct competitions related to issues of other subGovernances within the framework of main Governance;

- clearly define the set of criteria for judging in each contest;

- to establish the dependence of the jury's remuneration on the number of evaluated works, and to increase the fund of judges' reward with a large number of submissions;

- use scoring tables for each of the previously defined criteria. The data must be public;

- hide for the jury the evaluations of other juries during the voting. Ratings and comments should be

made public only after the end of the voting;

- develop a mechanism for exclusion from jury members in case of systematic violation of the rules or the establishment of facts of dishonest voting.

My suggestions are based on data analysis and personal voting experience. As a member of the SMM subGovernance jury, I developed a grading table for entries in the Free TON Blog Contest. This table was used by two judges. Subsequently, we posted the results on the forum and received positive feedback from community members.

Detailed post with description & sheet for judges voting for Blog contest

https://forum.freeton.org/t/free-ton-blog-contest/2917/76

Full raw data used for my analysis of the jury's work:

https://docs.google.com/spreadsheets/d/1_aMjteI19fMFayaFyfO9T__FEPiTcBuQkNR0PKz7BFk/edit#gid=768455046

Yours faithfully,
Sergey Potemkin
@exch_1001btc_com