GAMES REVEAL YOU

PLAYER 1 Michał Ilski
PLAYER 2 Jan Pawłowski
PLAYER 3 Jakub Balicki
Katedra Sztucznej Inteligencji
Politechnika Wrocławska

THE AIM OF THIS PROJECT

is to learn about players' communities based on their Reddit activity. There must be something special about them, right?

SETUP

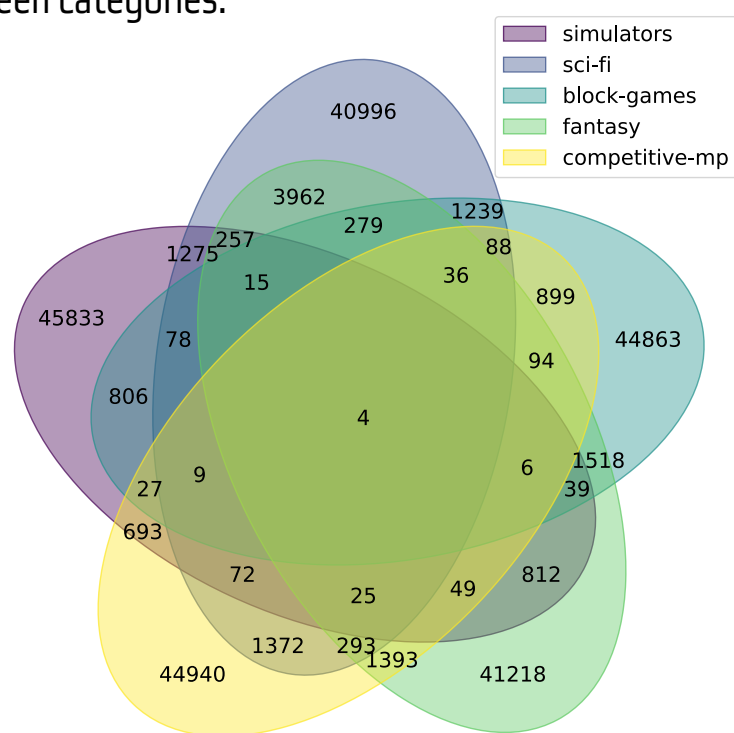Players were divided into groups by the comments they posted in games-related subreddits:

- Block games - e.g. *r/Minecraft, r/roblox, r/Terraria*
- Competitive multiplayer - e.g. *r/counterstrike, r/leagueoflegends*
- Fantasy - e.g. *r/Eldenring, r/skyrim, r/witcher3*
- Science fiction - e.g. *r/StarWars, r/Bioshock, r/masseffect*
- Simulators - e.g. *r/MicrosoftFlightSim , r/FIFA, r/iRacing*

All the data comes from Monthly Reddit Dump Dataset from November 2022. From the whole dataset (~260M comments) we listed users commenting at the specified subreddits (you might be there!) and sampled 50k users from each group. We went through the dataset again to collect their entire activity on Reddit.

DATA PROCESSING PIPELINE

1. *Selecting users* - We iterated through ~260M Reddit comments from November 2022 and gathered all the users posting comments in subreddits related to specified categories. Then we sampled 50k users from each category for further processing. Bots and users with little activity were excluded.
2. *Collecting activity* - For selected users we gathered all of their comments made in November 2022 on Reddit.
3. *Filtering and preprocessing* - We further processed the data for analysis by e.g. filtering out subreddits with little activity before calculating correlations or tokenized the comments for toxicity preditions.
4. *Data analysis* - We analyzed different aspects of collected data. For example we checked intersections between groups, checked the toxicity of the comments and correlations with other subreddits.
5. *Visualization* - Final plots and tables presenting selected results from the data analysis.

SEPERATE OR NOT?

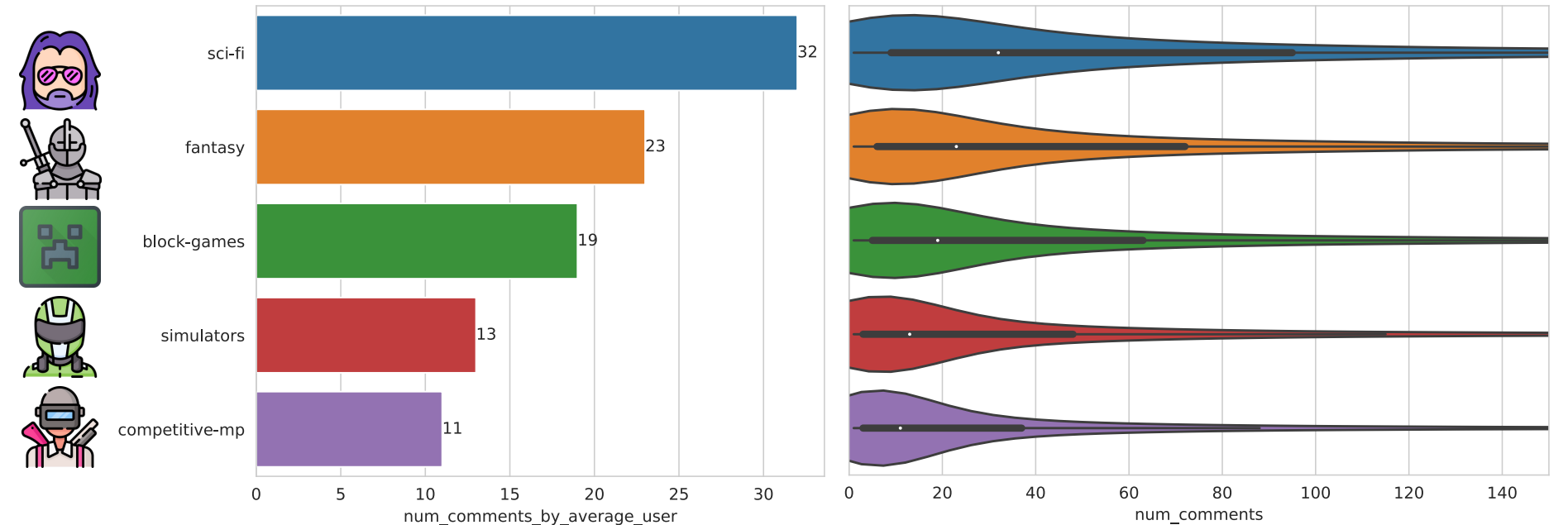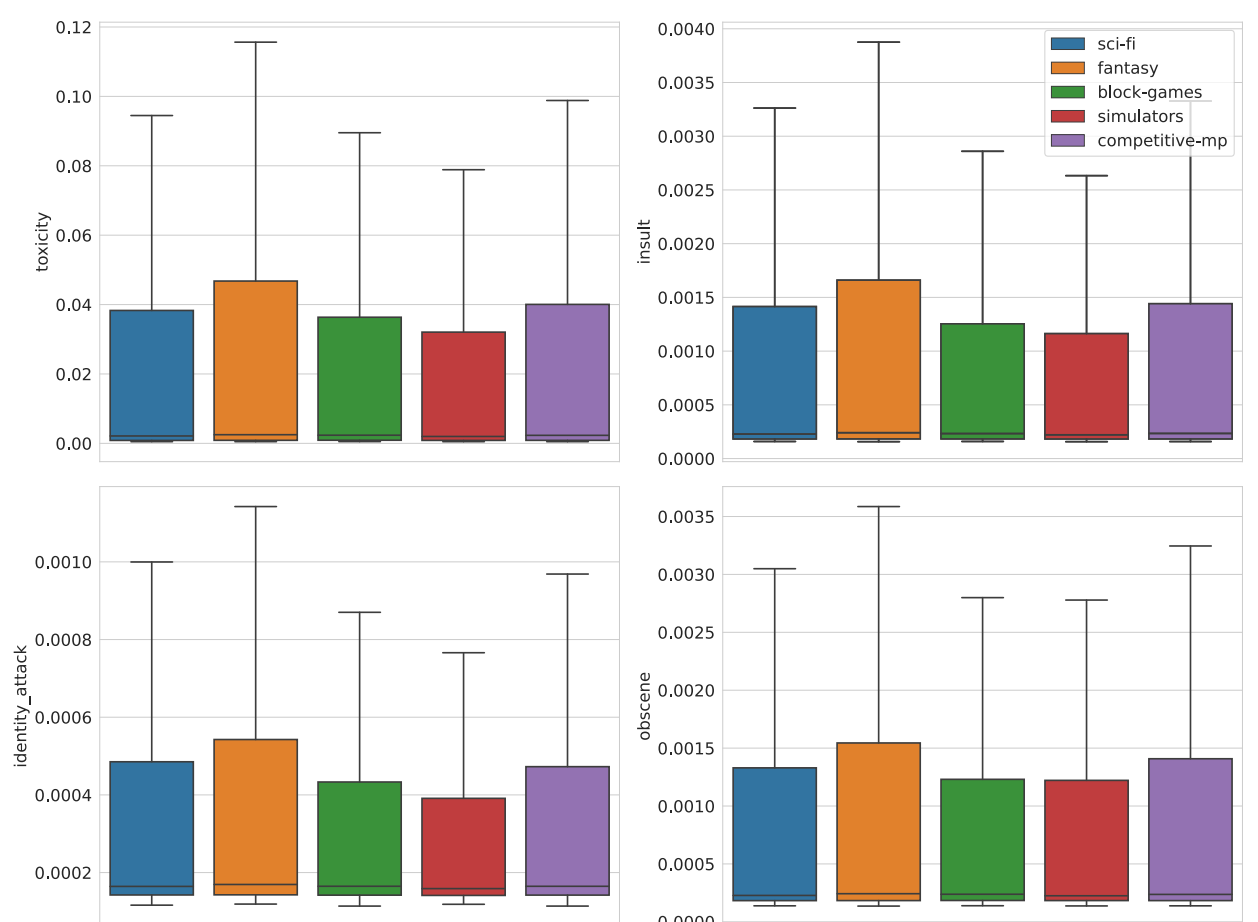Venn diagram showing intersections of the users between categories.

There is a significant intersection between science fiction and fantasy (~5k) in compare to other pairs (~2k and less). There are also only 4 users that commented in all of the categories-related subreddits in November 2022.

GENERAL REDDIT ACTIVITY

The distribution of the number of comments made during the month of November 2022 by the studied users.

There is a disproportion in general activity on reddit between the groups. An average user from science fiction group posted ~32 comments on average during November 2022, which is more than twice as many as an average user from competitive multiplayer group who posted only 11 posts on average. This shows that science fiction group is full of NERDS, no surprises here.

CORRELATIONS

between subreddits from specified categories and the others. Subreddits that were self-related to specified categories and the one with low users' activity were filtreed out, leaving ~1k subreddits. The first table presents top correlating subreddits in general. The subreddits are sorted in descending order by the Pearson correlation coefficient.

block-games	competitive-mp	fantasy	sci-fi	simulators
PhoenixSC	Games	fromsoftware	movies	fut
Dinosaurs	LivestreamFail	teslore	PrequelMemes	soccer
feedthebeast	summonerschool	bloodborne	marvelstudios	simracing
smashbros	nfl	Sekiro	MawInstallation	lakers
StardewValley	atheism	badredman	Edgerunners	ModernWarfareII
memes	PS5	gaming	gaming	Simracingstewards
tf2	CFB	Rainbow6	SpidermanPS4	chelseafc
btd6	Cooking	PatchesEmporium	WhitePeopleTwitter	dating_advice
shitposting	LOTR_on_Prime	BeyondTheFog	dragonage	battlefield2042
Bossfight	Paladins	antinatalism	saltierthancrait	PremierLeague

The second table presents the highest correlating subreddits (the same sorting strategy) related to political and social topics.

block-games	competitive-mp	fantasy	sci-fi	simulators
lgbt	wallstreetbets	antinatalism	antiwork	ukpolitics
politics	politics	antiwork	politics	conspiracy
asktransgender	antinatalism	PoliticalHumor	neoliberal	Wallstreetsilver
PoliticalCompassMemes	Economics	asktransgender	PoliticalHumor	Anarcho_Capitalism
antinatalism	PoliticalCompassMemes	WorkReform	PurplePillDebate	conspiracy_commons
AskALiberal	antiwork	Wallstreetsilver	moderatepolitics	PurplePillDebate
PoliticalCompass	neoliberal	AskALiberal	WorkReform	moderatepolitics
Anarcho_Capitalism	Wallstreetsilver	lgbt	Economics	wallstreetbets
PurplePillDebate	PurplePillDebate	conspiracy	ukpolitics	AskALiberal
Wallstreetsilver	Anarcho_Capitalism	PoliticalCompassMemes	PoliticalCompassMemes	neoliberal

The naive *average number of posts per subreddits* approach did not show any significant differences between studied groups, as they mainly post on the popular subreddits. However, the correlation of the number of posts per user shows, that the groups differ in general, political and social activity.

TOXICITY SCORES

were calculated on 1M comments using *toxic-bert* model from *huggingface*, delivered by *Unitary AI*. For each toxicity category scores are in the [0, 1] range.

All the toxicity-related median results are close to zero, however there are still comments with relatively higher score which can be seen in the third quartile and maximum values in each case. Fantasy fans seems to be the most toxic and simulators' players to be the least. Nevertheless, there are no communities that seem to be significantly toxic.