Research Design Outline

Detecting Bots on Reddit
Computational Social Science

Matteo Mazzarelli

January 9, 2025

1 Research Question & Relevance

Research Question: Can basic heuristic methods effectively identify bot influence on Reddit discussions?

Hypotheses:

- Account activity heuristics can identify probable bot accounts.
- These accounts may amplify specific narratives or contribute to online polarization.

Relevance: This study is a preliminary investigation into using simple methods for detecting and analyzing potential bot activity on Reddit. The findings will inform the need for more sophisticated methods to understand bots' broader impact on online discourse.

2 Data & Measurement

2.1 Data

- **Source:** Reddit API via the Python Reddit API Wrapper (PRAW).
- **Subreddits:** Selected for high subscriber counts and relevance to bot-targeted topics (e.g., politics, news). Keywords for selection will be generated using a reproducible method with Google Gemini (see Section A.1 for prompt and code). Subreddits will be scored as detailed in Section A.2.
- **Collection:** Posts, comments, timestamps, and available user data like account age and karma.

2.2 Measurement

- **Bot Identification (Heuristics):** Accounts flagged as potential bots if they exhibit:
  - Young account age.
  - Unusually low/high karma.
  - High posting frequency.
  - Repetitive content.

- **Analysis:**
  - **Descriptive:** Estimate bot prevalence, compare content from suspected bots versus humans, and analyze bot engagement patterns.
  - **Inferential:** Explore correlations between bot activity and narrative amplification (via keyword analysis) or polarization (via sentiment analysis).
  - **Accuracy Evaluation:** If feasible, build classification models and assess using cross-validation. Test various heuristic combinations, evaluating their performance with standard classification metrics.

3 Methodology: Challenges, Evaluation, and Scope

3.1 Population & Sample

- **Population:** Reddit users, especially in subreddits related to topics often targeted by bots.
- **Sample:** Data from selected subreddits as outlined in Section 2.

3.2 Challenges & Limitations

Identifying Reddit bots is difficult because of user anonymity, as well as limited API access for data collection, and the inherent inaccuracy of heuristic detection methods. This study prioritizes a feasible approach, rigorously evaluating the accuracy of heuristics while transparently addressing all limitations. Ethical issues are also considered: only publicly available data will be used, and usernames will be anonymized to protect user privacy. The methodology, including limitations of bot identification, will be documented transparently.

3.3 Key Variable Measurement

- **Bot Activity:** Identified using heuristics such as account age, karma, and posting behavior.
- **Narrative Amplification:** Measured by frequency of keywords linked to specific narratives.
- **Polarization:** Assessed using sentiment analysis on discussion content.

3.4 Evaluation of Bot Detection

The effectiveness of the heuristics will be thoroughly evaluated, including testing different combinations. We will explore leveraging Large Language Models to refine detection. If feasible, simple classification models will be developed and their performance assessed using standard accuracy metrics, offering a quantitative evaluation of the bot detection's effectiveness.

3.5 Expected Outcomes

This research aims to establish the feasibility of using basic heuristics for bot detection on Reddit and provide preliminary insights into their potential influence on discussions, particularly regarding narrative amplification and polarization. The primary limitations will be the accuracy of the heuristics and the generalizability of the results to all of Reddit.

A Appendix

A.1 Keyword Generation

The list of keywords can be generated with the help of a large language model API (e.g. Gemini)
tuned for reproducibility (stochastic parameters such as Temperature turned to 0) and includes:

```python
import google.generativeai as genai
import os

genai.configure(api_key=os.environ["GEMINI_API_KEY"])

# Create the model
generation_config = {
  "temperature": 0,
  # controls randomness.
  # 0 = most deterministic
  # (always selects highest probability token).
  "top_p": 0,
  # nucleus sampling:
  # limits token selection to the most probable.
  # 0 = most deterministic (used when temperature > 0).
  "top_k": 1,
  # restricts to top 'k' tokens.
  # 1 = most deterministic (used when temperature > 0).
  "max_output_tokens": 8192,
  "response_mime_type": "text/plain",
}

model = genai.GenerativeModel(
  model_name="gemini-exp-1206",
  # best performing gemini model (see: https://livebench.ai)
  # api usage is basically unlimited and free of charge
  generation_config=generation_config,
)

import praw
import pandas as pd

reddit = praw.Reddit(
    client_id=os.environ["PRAW_CLIENT_ID"],
    client_secret=os.environ["PRAW_CLIENT_SECRET"],
    user_agent=os.environ["PRAW_USER_AGENT"],
    username=os.environ["PRAW_USERNAME"],
    password=os.environ["PRAW_PASSWORD"],
)


# Fetch a large subset of popular subreddits
```

```python
# (large limit makes this representative of
# the largest overall subreddits by subscribers,
# check: https://gummysearch.com/tools/top-subreddits/)
subreddits = list(reddit.subreddits.popular(limit=1000))

# Create a DataFrame containing generated subs list
subs_df = pd.DataFrame([{
    "Name": subreddit.display_name,
    "Subscribers": subreddit.subscribers,
    "Description": subreddit.public_description,
    "Over 18": subreddit.over18,
    "Submission Type": subreddit.submission_type
} for subreddit in subreddits]).sort_values(
    by="Subscribers", ascending=False, ignore_index=True)

import ast

chat_session = model.start_chat()

response = chat_session.send_message("What are some keywords I can use to create a
↪    list of subreddits which are likely to be influenced by bots because of their
↪    controversial nature? These are keywords that I would look for within a
↪    subreddit's name or description. For example: \"news\", \"politics\",
↪    \"discussion\", \"war\", \"vaccines\", \"controversial\", \"conflict\",
↪    etc.\n\nKeep the answer short, only including 50 keywords and saving them in a
↪    python list as follows [\"key1\",\"key2\",...]. Send the output as text not as
↪    code.")

bot_influence_keywords =
↪    ast.literal_eval(response.candidates[0].content.parts[0].text.replace("\n",
↪    ""))

for i in range(0, len(bot_influence_keywords), 5):
    print(*bot_influence_keywords[i:i+5])
```

```
news politics discussion war vaccines
controversial conflict debate opinion world
current events election government policy social
media conspiracy activism protest rights
freedom censorship propaganda misinformation disinformation
fake news bias ideology partisan left
right conservative liberal progressive republican
democrat socialism capitalism communism fascism
anarchy global national security privacy
surveillance technology science health
```

A.2 Selected Subreddits

The top 10 subreddits based on this scoring system are:

```python
# Score subreddits based on subscribers and keywords in description
def calculate_bot_influence_score(row):
    score = 0

    # Large subscriber base increases potential for bot activity
    if row['Subscribers'] > 10000000:
        score += 5
    elif row['Subscribers'] > 5000000:
        score += 4
    elif row['Subscribers'] > 1000000:
        score += 3

    # Check for keywords in description and subreddit name
    description = row['Description'].lower()
    sub_name = row['Name'].lower()
    for keyword in bot_influence_keywords:
        if keyword in description:
            score += 1
        if keyword in sub_name:
            score += 1

    return score

subs_df['Bot Score'] = subs_df.apply(calculate_bot_influence_score, axis=1)

# Get top 10 most vulnerable subreddits
top_vulnerable = subs_df.nlargest(10, 'Bot Score')[['Name', 'Subscribers',
  ↪  'Submission Type', 'Bot Score']].reset_index(drop=True)
top_vulnerable
```

	Name	Subscribers	Submission Type	Bot Score
0	worldnews	43568694	link	9
1	news	29225132	link	9
2	technology	17726871	any	9
3	science	33423695	link	8
4	AmItheAsshole	23088618	self	8
5	politics	8694945	link	8
6	movies	34170875	any	7
7	askscience	25977459	self	7
8	sports	21709169	link	7
9	UpliftingNews	20190302	link	7