

# Social Network Analysis of Indian Reddit Communities: Identifying Influencers, Echo Chambers, and Misinformation Spread

Amay Dixit  
IIT Bhilai  
Durg, Chhattisgarh  
amayd@iitbhilai.ac.in

## Abstract

Social media platforms like Reddit have become key spaces for public discourse, shaping opinions, influencing narratives, and spreading information. However, these platforms also facilitate the formation of echo chambers, the spread of misinformation, and ideological polarization. This study aims to apply Social Network Analysis (SNA) to Indian-focused Reddit communities to uncover patterns of influence, interconnections, and information dissemination. By analyzing user interactions, we will identify key influencers, detect tightly knit clusters (potential echo chambers), and assess the extent of misinformation propagation.

We will collect data from relevant Indian subreddits using Reddit's API, focusing on metrics such as user engagement, comment-reply structures, and voting patterns. Graph-based techniques and community detection algorithms will be used to analyze network structures, while Natural Language Processing (NLP) will aid in misinformation identification. This research will contribute to understanding how digital communities shape discourse in India, offering insights for media literacy, policy recommendations, and online community management.

## 1 Problem Motivation

With the rapid growth of internet penetration in India, social media platforms like Reddit have become key hubs for discussion, opinion formation, and information dissemination. Unlike other platforms, Reddit's community-driven model allows for in-depth discussions within topic-specific subreddits. While this fosters engagement, it also raises several concerns related to influence dynamics, information silos, and misinformation spread.

### 1.1 Key Motivations for This Study

- (1) **Influence and Power Structures** – Some users play a disproportionately large role in shaping discourse within subreddits. Identifying these key influencers is essential to understanding who drives narratives and how opinions are shaped in digital spaces.
- (2) **Echo Chambers and Polarization** – Many online communities tend to reinforce similar viewpoints, limiting exposure to diverse perspectives. This phenomenon, known as an echo chamber, can lead to confirmation bias and ideological polarization. Understanding subreddit interconnections can help detect and quantify the extent of echo chambers.

- (3) **Misinformation Spread** – The viral nature of online content allows misinformation to propagate rapidly, influencing public perception and decision-making. Analyzing content flow and detecting misinformation patterns can provide early warnings about misleading narratives.
- (4) **Reddit's Role in Indian Discourse** – While platforms like Twitter and Facebook have been extensively studied in the Indian context, Reddit remains relatively unexplored. Given its unique voting and discussion system, understanding Reddit's influence can provide new insights into digital discourse in India.

By addressing these concerns, this study aims to provide a data-driven understanding of online community dynamics in India, with implications for media literacy, content moderation, and policy-making.

## 2 Challenges

Conducting Social Network Analysis (SNA) on Reddit presents several challenges, spanning data collection, network analysis, misinformation detection, and ethical concerns. Understanding these challenges is essential for designing a robust research methodology.

### 2.1 Data Collection Challenges

Reddit provides an API for data retrieval, but several limitations make large-scale data mining difficult:

- **API Rate Limits:** The Reddit API enforces strict rate limits, *restricting the number of requests per minute*. This slows down data collection and increases the time required for large-scale studies.
- **Lack of Historical Data Access:** Reddit's API does not provide full historical access to posts and comments, making it *difficult to analyze long-term trends*. Third-party datasets or web scraping may be required, raising ethical concerns.
- **Deleted and Edited Content:** Users frequently *edit or delete* posts and comments, leading to potential *data inconsistency* and gaps in the dataset.
- **Subreddit Moderation & Bans:** Some subreddits implement *strict content moderation*, auto-deleting posts or banning users. Additionally, subreddits can be *quarantined or banned*, removing access to their data.
- **Bot and Spam Filtering:** Reddit employs anti-spam mechanisms that *shadowban or filter* certain posts and comments. This means some discussions may be *invisible to data collectors*, introducing selection bias.

- **Heterogeneous Data Formats:** Reddit discussions include *text, links, images, videos, and embedded content*. Extracting meaningful insights from non-text data remains a challenge.

## 2.2 Identifying Key Influencers & Community Structures

- **Defining Influence:** Influence on Reddit is complex—it can be measured through *karma, engagement, submission frequency, or cross-subreddit activity*. Choosing an appropriate metric is challenging.
- **Multi-Account Users & Anonymity:** Many users operate multiple accounts (alt accounts) or post anonymously, making it difficult to track *long-term influence*.
- **Cross-Subreddit Interactions:** Users may participate in *multiple subreddits* with *varying ideologies*, making it difficult to model their overall influence in Reddit's network.

## 2.3 Echo Chamber & Polarization Analysis Challenges

- **Dynamic Network Structure:** Subreddit participation changes over time, with users joining and leaving communities. Capturing evolving relationships requires *temporal network modeling*.
- **Lack of Explicit Ideological Labels:** Unlike platforms like Twitter, Reddit does not have self-declared political affiliations. This makes it harder to *quantify ideological divides* without external annotation.
- **Measuring Polarization Accurately:** Traditional NLP-based sentiment analysis may *oversimplify complex political and social discussions*. More sophisticated *stance detection models* are needed.

## 2.4 Misinformation Spread & Content Verification

- **Difficulty in Source Verification:** Reddit allows users to post content without linking to credible sources, making it difficult to assess *information credibility*.
- **Meme & Image-Based Misinformation:** Misinformation is not always textual—it is often spread through *images, memes, and videos*, which require *multimodal analysis techniques*.

## 2.5 Ethical, Privacy, and Bias Concerns

- **Moderation Policies & Data Bias:** Different subreddits enforce *different rules*, which affects which posts remain visible. This creates *moderation bias* in the collected dataset.
- **Algorithmic Bias in NLP Models:** Many NLP models used for sentiment analysis and misinformation detection are trained on *Western datasets*, potentially leading to *bias when applied to Indian discourse*.

## 3 Literature Review

This section reviews relevant research in social network analysis of online communities, with a particular focus on Reddit, echo chambers, influence dynamics, and misinformation spread.

### 3.1 Social Network Analysis of Reddit Communities

Reddit's unique structure of interconnected communities has been the subject of several significant studies in social network analysis. Bower [1] pioneered early work in this domain with an analysis of information flow and social structures on Reddit. Their work established foundational methods for constructing interaction graphs from Reddit data and identified patterns of influence that emerge in these communities.

The platform's structure of segregated yet interconnected communities makes it particularly suitable for studying diverse social phenomena. Reddit communities operate as distinct social units while still allowing cross-community interaction, creating a rich environment for analyzing influence patterns, polarization, and information dissemination [4].

### 3.2 Echo Chambers and Polarization in Online Communities

Recent research has made significant progress in formalizing the concept of echo chambers and developing methodologies for their detection. Cinus et al. [2] proposed a four-stage framework for identifying echo chambers in online social networks that includes: (1) identifying controversial issues, (2) inferring users' ideological positions, (3) constructing user debate networks, and (4) detecting homogeneous meso-scale communities. Their case study covering the first two and a half years of Donald Trump's presidency revealed that while users appear strongly polarized with respect to their ideology, most tend not to insulate themselves completely in echo chambers.

The detection of echo chambers has evolved toward standardized approaches that can be applied across different platforms. La Morgia et al. [5] built on earlier work to propose platform-independent methods for echo chamber detection, emphasizing the importance of both network structure and content analysis in identifying these phenomena. Their research found that polarized communities on Reddit tend to remain stable over time, suggesting persistent divisions between groups with opposing viewpoints.

### 3.3 Controversial Information Spread and Misinformation

Studies examining information dissemination patterns on Reddit have yielded important insights into how controversial content spreads. Salehi et al. [6] analyzed approximately 17 million posts related to cybersecurity issues and found that controversial content tends to travel faster and further from its origin compared to non-controversial content. Their research is particularly valuable as it examined controversiality within typical discussions rather than focusing exclusively on topics known to be divisive.

The mechanics of controversial information spread on Reddit are distinct from other platforms due to the platform's unique features, including the built-in "controversial" tag that indicates mixed reception among users. This feature has enabled researchers to study controversy without relying solely on predetermined controversial topics, offering insights into how everyday discussions can become polarized [6].

### 3.4 Cross-Platform Information Flow and Influence

Research has increasingly focused on cross-platform information flow, examining how content and narratives move between Reddit and other social media platforms. Crupi et al. [3] studied echo chambers in the context of COVID-19 discussions, finding that both network topology and content semantics play crucial roles in the formation of polarized groups. Their methodology of applying community detection to both topology- and content-aware representations offers valuable insights for our approach to analyzing Indian Reddit communities.

### 3.5 Methodological Approaches and Challenges

Various methodological approaches have been developed to study Reddit communities, each with their own strengths and limitations. Graph-based techniques have proven particularly valuable, with researchers employing different types of networks (e.g., user-user interaction graphs, subreddit co-participation graphs) to capture different aspects of community dynamics [4]. Community detection algorithms applied to these networks have been successful in identifying cohesive groups and potential echo chambers.

Content analysis approaches have complemented network-based methods, with researchers employing topic modeling, sentiment analysis, and stance detection to understand the ideological positioning of users and communities [2]. The integration of network and content analysis has emerged as a best practice for comprehensive understanding of online community dynamics.

### 3.6 Research Gaps and Our Contribution

While substantial research has examined Reddit communities in Western contexts, particularly in relation to US politics [2], there remains a significant gap in the literature regarding Reddit communities focused on India. Our research addresses this gap by applying established social network analysis techniques to Indian-focused subreddits, with adaptations to account for India-specific linguistic, cultural, and political contexts.

Additionally, existing research has often focused on either network structure or content analysis, with fewer studies integrating both approaches comprehensively. Our methodology combines graph-based analysis with multilingual content analysis, acknowledging the linguistic diversity of Indian online discourse. This integrated approach allows us to examine not only the structural patterns of influence and echo chambers but also the content characteristics that drive these patterns in Indian Reddit communities.

## 4 Proposed Solution Approach

This section details our methodology for analyzing Indian-focused Reddit communities using Social Network Analysis (SNA). We present a comprehensive six-phase approach covering data collection, graph construction, network analysis, content analysis, evaluation, and dissemination.

### 4.1 Data Collection and Preparation

- **API Configuration and Authentication:** We will establish Reddit API credentials through OAuth2 and configure

API wrappers (PRAW/PMaw) with appropriate rate limiting mechanisms. Our implementation will include retry logic for handling API timeouts and failures.

- **Target Subreddit Selection:** We will identify relevant Indian-focused subreddits using snowball sampling, ensuring diverse representation across regional, political, cultural, and entertainment categories. A metadata repository will track subscriber counts, creation dates, and moderation policies.
- **Data Collection Pipeline:** Our pipeline will collect posts and comments from selected subreddits over a 3-6 month timeframe, extracting metadata such as timestamps, authors, scores, award counts, and comment counts. Data will be stored in MongoDB/PostgreSQL with appropriate indexing, and we will implement incremental collection to address API limitations.
- **Data Preprocessing:** We will remove deleted content and automoderator posts, handle missing values, standardize datetime formats, identify potential bot accounts using activity patterns, and anonymize usernames while preserving network structure.

### 4.2 Graph Construction

We will construct four distinct graph types to capture different dimensions of Reddit interactions:

- **User-User Interaction Graph:** We will define nodes as individual users and edges as interactions (replies, mentions), weighting edges based on interaction frequency and recency. Directed graphs will preserve interaction direction, and we will generate separate graphs for different time periods to enable temporal analysis.
- **Post Propagation Graph:** We will identify posts with shared URLs or crossposted content to track information flow across threads and subreddits. Edges will be weighted based on temporal proximity and engagement metrics, with additional metadata nodes for external domains/sources.
- **Subreddit Co-Participation Graph:** We will calculate user overlap between subreddit pairs, applying threshold filtering to remove noise (minimum shared users). Weighted edges will be created based on Jaccard similarity coefficient, and community structures will be visualized using force-directed layouts.
- **User-Subreddit Affiliation Graph:** We will build a bipartite graph connecting users to their active subreddits, weighting edges by normalized participation frequency. Projection techniques will be applied to derive subreddit similarity networks and identify power users across multiple communities.

### 4.3 Network Analysis

- **Centrality and Influence Analysis:** We will calculate centrality metrics (degree, betweenness, eigenvector) and identify key influencers using PageRank and HITS algorithms. User reach and impact will be quantified through network propagation models, with influence distribution compared across different subreddits.
- **Community Detection:** We will apply community detection algorithms (Louvain, Infomap) to identify tightly-knit

clusters and bridge nodes. Modularity scores will assess community separation, with validation through content similarity analysis.

- **Echo Chamber Quantification:** We will measure homophily within detected communities and calculate E-I index to quantify echo chamber tendencies. Information diversity exposure will be analyzed for different user groups, and opinion polarization will be tracked using stance detection on post content.
- **Temporal Evolution Analysis:** We will construct time-sliced networks to track community evolution, identifying significant shifts in network structure and analyzing user migration patterns between communities. Network changes will be correlated with external events.

#### 4.4 Content Analysis

- **NLP-based Topic Modeling:** We will extract topics using LDA/BERTopic on subreddit content and compare topic distributions across different communities. Contentious topics that drive polarization will be identified and mapped to detected network communities.
- **Sentiment and Stance Analysis:** We will apply specialized NLP models for Indian languages/contexts to measure sentiment polarity and strength in discussions. Stance on controversial topics will be detected and correlated with network structures.
- **Misinformation Detection and Analysis:** We will train fact-checking models on controversial content and track propagation paths of flagged misinformation. Super-spreaders of misleading content will be identified, and community response to misinformation (reinforcement vs. correction) will be analyzed.
- **Cross-Platform Content Flow:** We will track URLs from external sources (Twitter, news sites) and analyze how external content shapes internal discussions. Bridging users who import external narratives will be identified, and time lag between external events and Reddit discussions will be measured.

#### 4.5 Evaluation and Interpretation

- **Statistical Validation:** We will apply bootstrapping techniques to validate network metrics and perform significance testing on community structures. Null models will benchmark observed network properties, and confidence intervals for influence measures will be calculated.
- **Case Studies:** We will select 3-5 significant events for detailed analysis, tracking information flow during these events and identifying narrative formation and competing perspectives. Exemplary cases of echo chambers or bridge building will be documented. Examples of such cases include the analysis of posts and comments after Lok Sabha Elections in 2024, or during the Maha Kumbh Mela in 2025
- **Limitations Analysis:** We will document sampling biases and data collection constraints, assess the impact of missing

data on network analysis, address ethical considerations regarding privacy, and acknowledge NLP limitations for Indian languages.

- **Insights and Recommendations:** We will synthesize findings on influence patterns and echo chambers, develop recommendations for platform design to reduce polarization, create guidelines for digital literacy in the Indian context, and propose moderation approaches to address misinformation.

This comprehensive approach combines network analysis techniques with content analysis to provide a holistic understanding of Indian Reddit communities, their influence structures, and information flow patterns. By implementing this methodology, we aim to address the challenges outlined in Section 2 while advancing the research objectives stated in our problem motivation.

### 5 Data Collection Strategy

To ensure comprehensive and methodologically sound data collection for our social network analysis of Indian Reddit communities, we will employ the following structured approach:

#### 5.1 Target Community Selection

We will collect data from a diverse set of India-centric subreddits, selected based on user engagement levels, ideological diversity, and topical relevance. These include:

- **General Discussions:** r/india, r/indiaspeaks, r/indiadiscussion, r/indiasocial, r/indiatrending
- **Regional Subreddits:** r/delhi, r/mumbai, r/chennai, r/bangalore
- **News & Politics:** r/indianews, r/indiamoderate, r/unitedstatesofindia
- **Cultural & Nostalgia:** r/incredibleindia, r/indianostalgia
- **Youth-Oriented:** r/teenindia
- **Q&A & General Inquiry:** r/askindia

This selection ensures a diverse representation of discussions across different domains, capturing community interaction patterns and the dissemination of information.

#### 5.2 Data Scope and Temporal Coverage

- **Initial Historical Period:** 12-month primary analysis window (January 2024 - December 2024) to establish baseline discourse patterns.
- **Expansion Strategy:** Following initial analysis, we will incrementally extend the historical data collection to cover earlier periods, eventually encompassing a 3-year historical window.
- **Temporal Segmentation:** Data will be organized into monthly and quarterly segments to facilitate longitudinal analysis of community evolution and discourse trends.

This approach enables both comprehensive historical analysis and identification of long-term trends in community dynamics.

#### 5.3 Data Collection Methods

- **Historical Data Acquisition:**
  - Primary historical data retrieval via Pushshift API and Reddit archive services.
  - Supplemental data collection from available dumps and archived snapshots.

- Batch processing of historical data with optimized storage and retrieval mechanisms.
- **Collection Parameters:** To ensure a rich dataset, we will extract multiple dimensions of data:
  - **Content Features:**
    - \* Posts – Full text, timestamps, scores, upvote ratios, associated flairs, embedded media.
    - \* Comments – Full text, timestamps, scores, parent-child relationships, award metrics.
  - **User Interaction Patterns:**
    - \* Community participation frequency and cross-subreddit activity (anonymized).
    - \* Network structures based on reply chains, co-participation matrices, and cross-posting behavior.
  - **Metadata Extraction:**
    - \* Author activity level (e.g., comment/post frequency), subreddit engagement trends.
    - \* Thread-level engagement metrics (e.g., conversation depth, average response time).

## 5.4 Data Processing and Storage

- **Database Implementation:**
  - Primary storage in PostgreSQL database with specialized indexing for network queries.
  - Graph-optimized secondary storage using Neo4j for relationship modeling.
  - Distributed processing framework for handling large-scale historical data retrieval.
- **Data Integrity and Validation:**
  - Rigorous deduplication and normalization of historical data.
  - Automated checks for temporal consistency and completeness.
  - Documentation of known gaps in archived data and statistical adjustments.

## 6 Data Cleaning, Pre-processing, and Modeling Strategy

After collecting the historical Reddit data, we will implement a systematic workflow for transforming raw data into analysis-ready datasets and building appropriate network models. This section outlines our comprehensive approach to data cleaning, pre-processing, and modeling.

### 6.1 Data Cleaning

- **Content Normalization:**
  - **Text Standardization:** Conversion of all textual content to UTF-8 encoding with consistent line ending formats
  - **URL Normalization:** Standardization of URLs through canonical form conversion, UTM parameter removal, and domain extraction
  - **Temporal Alignment:** Conversion of all timestamps to UTC with standardized ISO 8601 format and handling of daylight saving time transitions

- **Special Character Handling:** Processing of markdown formatting, HTML entities, Unicode special characters, and emoji
- **Missing Data Management:**
  - **Deleted Content Handling:** Preservation of thread structure with appropriate markers for deleted/removed content
  - **Banned User Content:** Special processing for content from subsequently banned accounts
  - **Partial Record Reconstruction:** Implementation of interpolation techniques for salvaging incomplete comment trees
  - **Metadata Imputation:** Statistical techniques for estimating missing non-critical metadata based on contextual information

### 6.2 Data Pre-processing

- **Text Processing Pipeline:**
  - **Language Detection:** Application of fastText-based language identification to separate English, Hindi, and other Indian languages
  - **Tokenization:** Implementation of language-specific tokenization using spaCy models optimized for Indian English and transliterated text
  - **Named Entity Recognition:** Application of custom NER models trained on Indian entity types (political parties, states, public figures)
  - **Code-Mixing Analysis:** Special handling for Hindi-English code-mixing common in Indian online discourse
- **Feature Engineering:**
  - **Temporal Features:** Extraction of cyclical features (hour of day, day of week, month), post time relative to India time zones
  - **Content-Based Features:** Calculation of text statistics (length, unique words, readability scores, language complexity)
  - **Engagement Features:** Derivation of normalized engagement metrics (comments per view, upvote velocity, controversy scores)
  - **User-Level Features:** Construction of user activity profiles including engagement patterns and subreddit participation distributions
- **Network Feature Extraction:**
  - **Interaction Graph Features:** Calculation of reply frequencies, interaction strengths, and user-user connection weights
  - **Community Membership Features:** Quantification of user activity distribution across ideologically diverse subreddits
  - **Temporal Network Features:** Extraction of user interaction evolution patterns and community participation shifts
  - **Content Sharing Features:** Identification of URL sharing patterns, cross-posting behaviors, and information sourcing habits
- **External Context Integration:**

- **News Event Correlation:** Mapping of discussion peaks to major news events from curated event timeline
- **Cross-Platform Reference Extraction:** Identification of references to external platforms (Twitter, YouTube, Instagram)
- **Domain Credibility Scoring:** Assignment of reliability scores to linked domains based on established media credibility databases
- **Political Alignment Markers:** Application of semi-supervised learning to estimate political alignment of shared content

### 6.3 Network Modeling Strategy

- **Multi-layer Network Construction:**
  - **User-User Interaction Graph:**
    - \* **Node Definition:** Individual Reddit users (anonymized)
    - \* **Edge Creation:** Direct interactions through replies and mentions
    - \* **Edge Weighting:** Frequency and recency of interactions with time decay function
    - \* **Directionality:** Directed edges preserving replier-recipient relationship
    - \* **Temporal Dimension:** Time-stamped edges enabling dynamic network analysis
  - **Post Propagation Graph:**
    - \* **Node Definition:** Individual posts and shared URLs
    - \* **Edge Creation:** Content sharing, cross-posting, and explicit reference relationships
    - \* **Edge Weighting:** Temporal proximity and similarity strength
    - \* **Directionality:** Directed edges representing information flow
    - \* **Metadata Incorporation:** Node attributes including content type, subreddit, and engagement metrics
  - **Subreddit Co-Participation Graph:**
    - \* **Node Definition:** Individual subreddits
    - \* **Edge Creation:** Shared active user base between subreddit pairs
    - \* **Edge Weighting:** Jaccard and cosine similarity indices of user bases
    - \* **Thresholding:** Minimum user overlap requirements to filter noise
    - \* **Node Attributes:** Subreddit size, activity levels, topic classification
  - **User-Subreddit Affiliation Graph:**
    - \* **Node Definition:** Users and subreddits (bipartite graph)
    - \* **Edge Creation:** User participation in subreddits
    - \* **Edge Weighting:** Normalized participation frequency and engagement levels
    - \* **Temporal Evolution:** Monthly snapshots to track affiliation changes
    - \* **Projection Techniques:** Derivation of user similarity and subreddit similarity networks
- **Dynamic Network Modeling:**
  - **Temporal Slicing:** Creation of network snapshots at regular intervals (weekly, monthly)

- **Continuous-Time Models:** Implementation of temporal edge models for fine-grained evolution analysis
- **Change Point Detection:** Statistical methods to identify significant structural shifts in networks
- **Stability Analysis:** Measurement of community persistence and evolution over time
- **Multilingual Content Modeling:**
  - **Language-Specific Embedding:** Application of language-specific word embedding models
  - **Transliteration Handling:** Special processing for Romanized Indian language content
  - **Cross-Lingual Alignment:** Mapping of topics and entities across language boundaries
  - **Code-Mixing Models:** Implementation of specialized NLP pipelines for Hindi-English mixed content

### 6.4 Implementation Technologies

- **Data Processing Framework:**
  - **Core Technologies:** Python ecosystem (pandas, NumPy, SciPy) for data manipulation
- **Network Analysis Stack:**
  - **Core Libraries:** NetworkX for prototype development, graph-tool for performance-critical operations
- **NLP and Content Analysis:**
  - **Text Processing:** spaCy and NLTK for core NLP tasks
  - **Embedding Models:** Application of IndicBERT and MuRIL for Indian language context
  - **Topic Modeling:** Implementation of BERTopic and contextualized topic models
- **Visualization and Reporting:**
  - **Network Visualization:** Implementation using Gephi for static visualizations.
  - **Report Generation:** Automated notebook-to-report pipeline using Jupyter Book
  - **Data Stories:** Development of interactive narratives using Observable framework

This comprehensive data cleaning, pre-processing, and modeling strategy addresses the unique challenges of analyzing Indian Reddit communities while establishing a foundation for rigorous social network analysis. The approach balances methodological rigor with context-specific adaptations to ensure valid insights into online community dynamics.

## 7 References

- [1] T. Bower. 2011. Information and social analysis of Reddit. *Stanford Network Analysis Project*. Class Project CS224W. [http://snap.stanford.edu/class/cs224w-2011/proj/tbower\\_Finalwriteup\\_v1.pdf](http://snap.stanford.edu/class/cs224w-2011/proj/tbower_Finalwriteup_v1.pdf).
- [2] Filippo Cinus, Salvatore Micciché, Michele Tumminello, and Fabrizio Lillo. 2021. The echo chamber effect in financial Twitter: a network approach. *Scientific Reports*, 11, 1, 1–15.
- [3] Antonio Crupi, Mattia Zago Pasta, and Corrado Monti. 2021. Echoes in the COVID-19 bubble: the effects of polarized news on social media. *Social Network Analysis and Mining*, 11, 1, 1–20. <https://link.springer.com/article/10.1007/s13278-021-00779-3>.
- [4] Sam Griesemer. 2019. Detection and analysis of subreddit communities. Research Paper. [https://samgriesemer.com/docs/research/subreddit\\_communities.pdf](https://samgriesemer.com/docs/research/subreddit_communities.pdf).
- [5] Massimo La Morgia, Alessandro Mei, Francesco Mongardini, and Julinda Stefa. 2021. Toward a standard approach for echo chamber detection: Reddit case study. *Applied Network Science*, 6, 1, 1–20. [https://www.researchgate.net/publication/352303162\\_Toward\\_a\\_Standard\\_Approach\\_for\\_Echo\\_Chamber\\_Detection\\_Reddit\\_Case\\_Study](https://www.researchgate.net/publication/352303162_Toward_a_Standard_Approach_for_Echo_Chamber_Detection_Reddit_Case_Study).

- [6] Seyedmostafa Salehi, Mehrdad Malekzadeh, and Mehrdad Koosha. 2021. Controversial information spread: an empirical study on Reddit. *Journal of Computational Social Science*, 4, 2, 667–679. <https://link.springer.com/article/10.1007/s42001-021-00121-z>.