REBOUND ethical review #2

Aristides Gionis

KTH Royal Institute of Technology

February 21, 2023

objective of the REBOUND project

- social media provide a means to democratize content and promote diversity
- at the same time, we observe phenomena of information silos, polarization, bias and misinformation
- REBOUND aims to develop computational methods to
  - address deficiencies in today's online media
  - discover structure of segregation and conflict in social media
  - break information silos and create awareness to explore alternative viewpoints

research themes in the REBOUND project

- DISCOVERY
  - can we detect patterns of bias, polarization, conflict, and lack of information flow in online media?
- EXPLORATION
  - can we help users understand the global information landscape, users gain control on their news diet, and explore alternative view points?
- RECOMMENDATION
  - can we provide recommendations that increase diversity and balance information exposure?

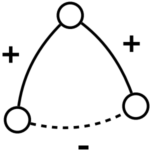
concrete topics we study

- detection of polarization structure
- mining signed networks / mining temporal networks
- data clustering and information summarization
- opinion formation in social networks
- information dissemination in social networks
- recommendations to increase diversity or exposure to information
- recommendations to increase network connectivity

overview of results so far and relevant publications

mining signed networks

- signed networks: networks where edges are labeled with $+$ or $-$
  - model “friends” or “enemies,” alternatively, “endorsement” or “conflict”



$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$$

- we design methods for discovering polarized communities in signed networks

mining signed networks — publications

Florian Adriaens, Simon Apers. Testing properties of signed graphs. To appear in the International Web Conference, 2023.

Stefan Neumann, Pan Peng. Sublinear-time clustering oracle for signed graphs. International Conference on Machine Learning (ICML), 2022.

Ruo-Chun Tzeng, Bruno Ordozgoiti, Aristides Gionis. Discovering conflicting groups in signed networks. Advances in Neural Information Processing Systems (NeurIPS), 2020.

Han Xiao, Bruno Ordozgoiti, Aristides Gionis. Searching for polarization in signed graphs: A local spectral approach. International Web Conference, 2020.

Bruno Ordozgoiti, Antonis Matakos, Aristides Gionis. Finding large balanced subgraphs in signed networks. International Web Conference, 2020.

opinion dynamics in social networks

- Friedkin–Johnsen model is used to study opinion dynamics in social networks
  - users assumed to have innate and expressed opinions
  - expressed opinions formed via minimizing “tension” of users with respect to their inner opinions and the opinions of their neighbors
- the model is used to quantify polarization and disagreement
- it is then used to understand the effects of certain graph “operations,” e.g.,

  how much the network polarization will be reduced after adding a certain edge in the network? or

  what are the best edges to add to a network to reduce disagreement?

opinion dynamics in social networks — publications

Tianyi Zhou, Stefan Neumann, Kiran Garimella, Aristides Gionis, Minimizing polarization and disagreement using topic-based timeline algorithms. Submitted for publication.

Sijing Tu, Stefan Neumann, Aristides Gionis. Adversaries with limited information in the Friedkin–Johnsen model. Submitted for publication.

Sijing Tu, Stefan Neumann. A viral marketing-based model for opinion dynamics in online social networks. International Web Conference, 2022.

recommendations to improve "efficiency" of a social network

- we design recommendation algorithms to optimize measures of "efficiency" of a social network
  - recommendations can be either edges (users to follow) or content (relevant posts)
  - measures of efficiency can be
    - polarization / disagreement
    - diversity / exposure to information
    - probability to be exposed to harmful content
    - number of strong ties
    - distances / diameter of the network
    - . . .

recommendations — publications

Corinna Coupette, Stefan Neumann, Aristides Gionis. Reducing Exposure to Harmful Content via Graph Rewiring. Submitted for publication.

Florian Adriaens, Honglian Wang, Aristides Gionis. Minimizing hitting time between disparate groups with shortcut edges. Submitted for publication.

Florian Adriaens, Aristides Gionis. Diameter minimization by shortcutting with degree constraints. International Conference on Data Mining (ICDM), 2022.

Antonis Matakos, Aristides Gionis. Strengthening ties towards a highly-connected world. Data Mining and Knowledge Discovery, 2022.

Antonis Matakos, Sijing Tu, Aristides Gionis. Tell me something my friends do not know: Diversity maximization in social networks. Knowledge and Information Systems, Volume 62, issue 9, 2020.

Antonis Matakos, Cigdem Aslay, Esther Galbrun, Aristides Gionis. Maximizing the Diversity of Exposure in a Social Network. IEEE Transactions on Knowledge and Data Engineering, 2020.

Sijing Tu, Cigdem Aslay, Aristides Gionis. Co-exposure maximization in online social networks. Advances in Neural Information Processing Systems (NeurIPS), 2020.

clustering, summarization, ranking

- when we summarize information we want to enforce diversity or fairness constraints in the summarized data
- this helps to present a *“balanced”* view of the available information
- for example, we introduced the problem of diversity-aware clustering:

  given data with protected attributes, find a high-quality clustering so that all categories along the protected attributes are well represented by the cluster centers

clustering, summarization, ranking — publications

Benedikt Riegel, Joachim Spoerhase, Kamyar Khodamoradi, Bruno Ordozgoiti, Aristides Gionis. A constant-factor approximation algorithm for reconciliation k-median via sparsified k-facility location. International Conference on Artificial Intelligence and Statistics (AISTATS) 2023.

Guangyi Zhang, Nikolaj Tatti, Aristides Gionis. Ranking with submodular functions on the fly. To appear in the SIAM International Conference on Data Mining, 2023.

Guangyi Zhang, Nikolaj Tatti, Aristides Gionis. Ranking with submodular functions on a budget. Data Mining and Knowledge Discovery, 2022.

Bruno Ordozgoiti, Ananth Mahadevan, Antonis Matakos, Aristides Gionis. Provable randomized rounding for minimum-similarity diversification. Data Mining and Knowledge Discovery, 2022.

Suhas Thejaswi, Bruno Ordozgoiti, Aristides Gionis. Diversity-aware k-median: Clustering with fair center representation. European Conference on Machine Learning and Knowledge Discovery in Databases ECML PKDD, 2021.

ethical considerations of REBOUND

1. management of personal data
2. misuse of methods
3. ethical reflections
4. ethical training

datasets discussed in our ethical-review process

A. social-media data (twitter, reddit, etc.)

B. data from user studies to evaluate our methods

C. data from Amazon mechanical turk for training machine-learning models

- so far, we have collected data only from social media (A)
  - in addition to publicly-available benchmark datasets
- project has focused mostly on theoretical aspects
  - most REBOUND researchers are mainly interested in theory and less in data-driven studies

A. social-media data

- information notice in project’s website
  - objectives of research, lawful basis, rights of data subjects, contact info
- data collection according to *data minimization principle*
  - no sensitive attributes
  - but political view can be inferred
- user IDs replaced immediately with pseudo-random IDs
- data stored securely in password-protected data centers
- data not shared with anyone
- results are presented always as aggregate statistics

misuse of methods

Q. how to mitigate the risk that our research results are misused?

- e.g., an adversary may use our methods to manipulate public opinion, or to sow disagreement

A. risk of misuse is not high compared to the benefits

A. manipulation can be easily done without the results of our research

- e.g., creating filter bubbles is much easier than breaking them

A. studying the power of adversaries to sow disagreement is also valuable for increasing awareness about danger of social media

- such studies need to be placed in the proper framework

ethical reflections

Q. is it ethical to make recommendations or “tamper” with user news feed?

A. existing platforms make recommendations already

- often aiming to optimize opaque or not-ethically-driven objectives, e.g., user engagement or monetization

additionally, consent can be obtained by the users for recommendations

ethical training

- training on ethical issues for all new members of the project
- actively thinking and discussing about ethical issues during the development of our projects
- reflections on ethical aspects are becoming increasingly common in the publication venues we are targeting
  - often it is a requirement

ethical review reports required by ERC

	project month	actual date	status
#0	M06	Aug 2020	SERA, approved
#1	M18	Aug 2021	submitted and approved
#2	M36	Feb 2023	ongoing, this meeting
#3	M48*	Feb 2024*	
#3	M60*	Feb 2025*	

project currently on the 3rd year out of 5, and will apply for a 1-year extension

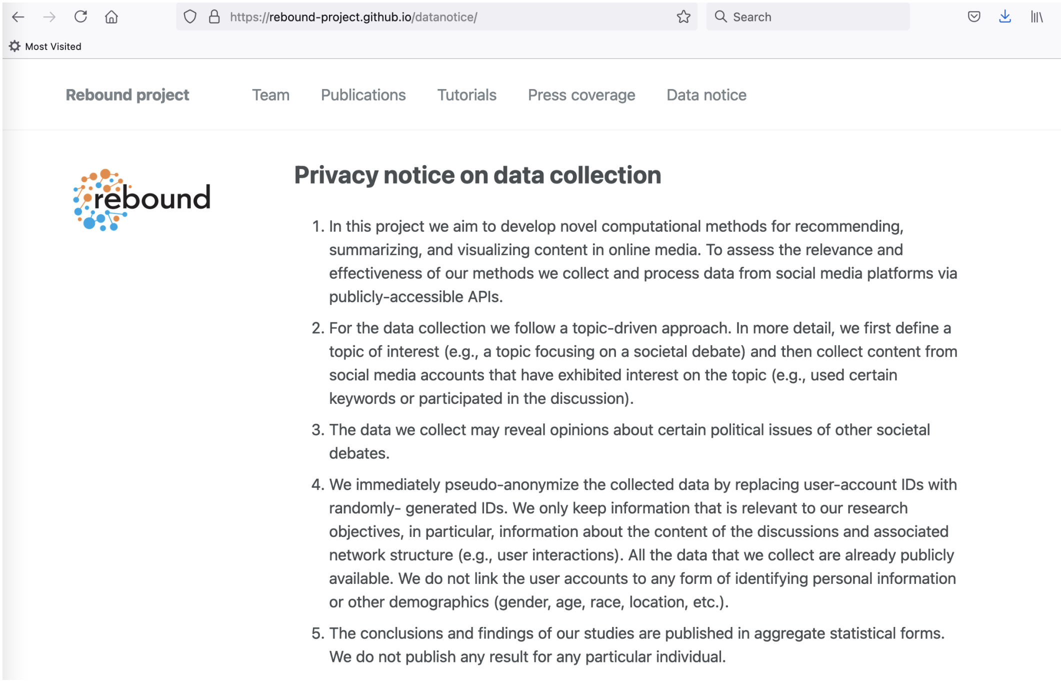
* dates likely to change, if extension is granted

summary

- research has focused mostly on theoretical aspects
- datasets used are either public or collected from social media
- we believe that risks of misuse are small compared to benefits
- we are continuously reflecting on ethical aspects
- global research landscape is evolving towards giving stronger emphasis on ethical considerations

thank you!

q & a



Privacy notice on data collection

1. In this project we aim to develop novel computational methods for recommending, summarizing, and visualizing content in online media. To assess the relevance and effectiveness of our methods we collect and process data from social media platforms via publicly-accessible APIs.
2. For the data collection we follow a topic-driven approach. In more detail, we first define a topic of interest (e.g., a topic focusing on a societal debate) and then collect content from social media accounts that have exhibited interest on the topic (e.g., used certain keywords or participated in the discussion).
3. The data we collect may reveal opinions about certain political issues of other societal debates.
4. We immediately pseudo-anonymize the collected data by replacing user-account IDs with randomly- generated IDs. We only keep information that is relevant to our research objectives, in particular, information about the content of the discussions and associated network structure (e.g., user interactions). All the data that we collect are already publicly available. We do not link the user accounts to any form of identifying personal information or other demographics (gender, age, race, location, etc.).
5. The conclusions and findings of our studies are published in aggregate statistical forms. We do not publish any result for any particular individual.