Comparing Climate Change Policy Networks: Codebook for Twitter Account Collection*

Ted Hsuan Yun Chen[1,2] Mikko Kivelä[1] Arttu Malkamäki[2]
Esa Palosaari[1] Anniina Penttilä[2]

Version: 1.1.0 (October 19, 2021)

Contents

*Contact: Ted Hsuan Yun Chen (ted.hsuanyun.chen@gmail.com). Authors are listed alphabetically by surname as everyone contributed equally to developing the protocol. This work is funded by Academy of Finland grant nos. 320780 and 320781. We thank Otto Schultz and Amirmohammad Ziaei for providing invaluable research assistance in developing this protocol. Project repository: github.com/tedhchen/componMultilayer.

[1]Department of Computer Science, Aalto University
[2]Faculty of Social Sciences, University of Helsinki

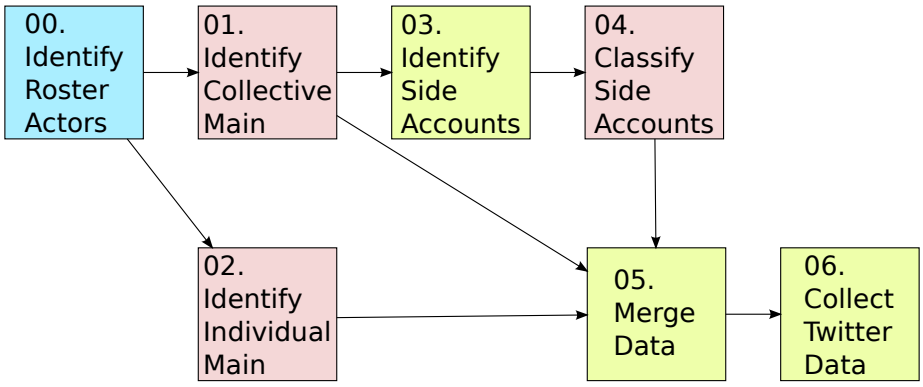

Figure 1: Data collection pipeline. Blue indicates pre-requirement; red indicates human task; yellow indicates code-assisted task.

1 Overview

In this codebook we document the procedures used to collect Twitter account information from organizations in our policy actor roster, including the account information of their members, employees, and other affiliates. We outline a set of standard protocol that aims to guarantee comparability of data collected in different contexts (e.g. countries or policy systems).

We collect the Twitter account usernames of all organizations in our policy actor roster. We also collect the account usernames of the individuals who represent the organizations. This approach affords a more complete picture of an organization's activities on the Twitter platform – individuals from the organization tend to behave differently from formal organizational accounts but still are recognizable to the public as members of the organization (e.g. Greta Thunberg versus Fridays for Future). The relative comprehensiveness of the data allows us to ask theoretically-motivated questions that were otherwise difficult with only survey data, including the role of formal and informal ties in policy systems and the relationship between individuals and the collective within an organization when it comes to organizational goals.

2 Protocol

The pipeline for collecting and labeling the data is summarized in Figure 1. This document is concerned with steps 01 to 04. The project's GitHub repository contains details about the pipeline.

Level	Description	Examples
0) Collective main	Main account of the organization.	`@FT`
1) Individual main	Executive personnel. This level includes only the top-level executive and the chair of the board of directors. The rest of the board of directors are excluded from this category completely.*	`@John_Ridding`; `@khalafroula`
2) Collective side	Accounts of the organization that represent subunits or focused functions of the organization.	`@FTPressOffice`
3) Individual side	Personnel that are not members of the executive branch but work for the organization.*	`@Mikepeeljourno`

*Additional instructions for specific organization types (e.g. universities) are discussed in the main text.

Table 1: Summary of Different Organizational Levels

We specifically identify four "levels" of accounts for each policy actor: 0) collective main, 1) individual main, 2) collective side, and 3) individual side. These levels are summarized in Table 1, using the Financial Times as an example of the roster entry. Detailed instructions for identifying these levels are in the sections below.

2.1 Collective Main

There is only one account in this category for each organization. The account can be identified in a number of ways, including from the organization's website, from a Twitter search, or from a general web search.

In some instances, when an organization is an international one but originated in a specific country, the "national-level" organization will be the international one (e.g. Nokia, which is on the Finnish roster, only has an international account). It is up to the country team to decide whether this is appropriate or to specify that there is no national-level organizational main account.

2.2 Individual Main

This level includes the executive personnel of the organization. We consider these as individuals with decision-making power for the entire organization. This specifically means that we want to collect leaders of the top-level unit in the organization (i.e. before units become parallel). Leaders of lower-level parallel units (i.e. subunits) should not be included.

For example, in a business organization individual main accounts would be members of the top level executive management team. If this team includes leaders of subunits, they are considered to be individual main accounts. If they are not members of the top level team, they would not be considered as individual main accounts.

Individual main accounts can be identified in a number of ways. If the organizations' website does not list the executive personnel, Linkedin serves as an alternative as people tend to publish their role in the organization there. Additionally, Linkedin allows for searching employees of the organizations by relevant keywords (e.g. "executive").

2.3 Collective Side

This level includes the subunits/subfunctions of the main organization. We have developed a Twitter-specific protocol to collect this data. Specifically, we only include Twitter accounts that satisfy all of the following criteria:

1. Follows the organization's collective main Twitter account
2. Is followed by the organization's collective main Twitter account
3. Has at least one of the organization's pre-specified keywords in its Twitter bio; these organization-specific keywords are specified by the country team but usually include at least different variations of the organization's name
4. Twitter bio passes human filtering as the organization's subunit/subfunction

To facilitate this task, we have developed python code that will identify and extract Twitter accounts that satisfy conditions 1–3. The code returns a list of each Twitter account's username and their bio text. Conditions 4 is checked by a human coder using this preprocessed list (i.e. it is up to the country team to decide which subunits are relevant.) The code base and instructions are available at the project's GitHub repository.

2.4 Individual Side

This level includes all identifiable individuals that work or are members of the organization but who are not in the individual main level. Because this is a difficult category to collect, we rely on the same criteria as the collection for collective side accounts. Specifically, we only include Twitter accounts that satisfy all of the following criteria:

1. Follows the organization's collective main Twitter account
2. Is followed by the organization's collective main Twitter account
3. Has at least one of the organization's pre-specified keywords in its Twitter bio; these organization-specific keywords are specified by the country team but usually include at least different variations of the organization's name

4. Twitter bio passes human filtering as the organization's member or employee

5. Is not already included as an individual main account

Again, we have developed python code that will identify and extract Twitter accounts that satisfy conditions 1–3. Conditions 4 is checked by a human coder using this preprocessed list. Condition 5 is automatically checked by code that merges all data files together. Effectively, this means that we can collect the collective side and individual side accounts at the same time by labeling each entry on the list of username-bio text as collective side, individual side, or neither.[1]

Depending on the type of organization, drawing the line between the 'main' and the 'side' at the individual level can be difficult. Some organizations have a relatively flat structure without a clear leader or executive team (e.g. social movement organizations like Extinction Rebellion). In such cases, we label all their representatives as executive personnel (i.e. individual main). The 'flat' organization is not a formal category in our protocol, but a helpful concept. Some other types of organizations have many voices and executive-like authority occurs at multiple levels. For example, universities can have rectors, deans, professors, and directors of various research centers or institutes. For universities, we identify only administrative executives at the university-level as executive personnel, and we exclude students from both individual main and side levels. Specific requirements for certain organization types are described in Table 2.

3 Additional Considerations

3.1 Policy Roster Actors without a Collective Main Account

In some cases, the policy actor will not have a collective main account. In these cases, the individual main accounts can still be collected according to the protocol outlined in Section 2.2. However, the collective side and individual side accounts cannot be collected as there is no way to obtain the potential side account list.

3.2 Joint Affiliations

An individual may satisfy the criteria for more than one organization. For example, the Chief Executive Officer of Nokia, Pekka Lundmark, is also the Chair of the Board of the Confederation of Finnish Industries. In our data collection task, each organization is collected

[1] Accounts that should be labeled as individual mains but were missed during the individual main collection step can be labeled as such during this step.

	Main	Side
Collective	Level: 0	Level: 2
	The main account of the organization. If there is only a global account, determine its inclusion based on local knowledge. These accounts are identified through manual search.	Accounts of the organization that are subunits or focused functions of the organization. Do not cross administrative levels (e.g. if the policy actor is a national-level political party, regional branches should not be included here). These accounts are identified from preprocessed lists. **Type-based requirements/clarifications** Political parties: Accounts of party-affiliated think tanks count as subunits. School-based chapters do not count as national party's subunits. Universities: Accounts of research groups and projects count as subunits.
Individual	Level: 1	Level: 3
	Executive personnel. Include only the top-level executives of the organization and the chair of the board of directors. The rest of the board of directors are excluded from this category. This usually translates into members of the executive management team. These accounts are identified through manual search. **Type-based requirements/clarifications:** Governments (e.g. state, city): Include only the top executives (e.g. governors, mayors) and head of the elected legislative bodies (e.g. premiers, council chairs). Government ministries: Include only the ministers, highest-level bureaucratic leadership, and related office accounts (e.g. `@ENERGY`). Political parties: Include only the party chair and vice chairs, the party secretary, and the chair of the parliamentary group/whip. Universities: Include only the top administrative executives (i.e. chancellor, president, rector, or equivalent), and principal deans (or equivalent). Military organizations: Generally, include only the highest ranking military officer, but apply local knowledge.	Personnel that are not members of the executive branch but work for the organization. These accounts are identified from preprocessed lists. **Type-based requirements/clarifications:** Political parties: Include only members of the national legislature or supranational governments legislatures (e.g. European Parliament). Universities: Exclude all types of students, including doctoral students.

Table 2: Summary of Data Collection and Labeling Rules

independently of each other, which means that Lundmark would be first collected as an individual main account when looking at Nokia, then collected as an individual main account when looking at the Confederation of Finnish Industries. This approach ensures we do not miss any joint affiliations, which can be used in different ways depending on the analysis.

This also means that, depending on the policy actor roster, a policy actor (and its collective main account) can be the collective side account of another policy actor.

3.3 "Opinions are my own" and Other Disclaimers

Some Twitter accounts will include disclaimers such as "opinions are my own and do not represent those of my employers". We choose to disregard such statements when collecting account data. While some organizations mandate their employees to add a disclaimer, others may also use it strategically to engage in controversial debates without dragging the entire organization along. Whatever the case, if the accounts can be identified using the protocols described here, they are going to be publicly visible members of their organization, and are therefore included in the data. Whether their opinions align with those of their organization is left for the analysis to decide.

3.4 Temporal Variation

Twitter is changing constantly. Organizations merge and change their names. Individuals change organizations or end up deleting previous content or even their entire account. Sometimes Twitter suspends accounts, making them unviewable or otherwise inaccessible. Our collection protocol is designed for contemporary data. This means that the closer the times between the policy actor roster construction and the Twitter account collection, the more accurate the data.

However, the temporal dimension is one of the benefits of using Twitter data, as policy actor rosters are often constructed prior to the present. If a country team decides to collect additional data on past affiliations, we describe some approaches here.

To ensure the temporal validity of the Twitter data and the policy actor roster, ideally the individuals are assigned an affiliation that they had at the time of roster construction. Checking who worked for the organization in the past can done with the Wayback Machine, which shows certain historical versions of websites. However, it takes a while to open archived websites. Figure 2 shows the Wayback Machine interface.

Using Linkedin is a faster alternative. Linkedin can be used to identify people who currently work for the organization or to check where an individual worked in the past.

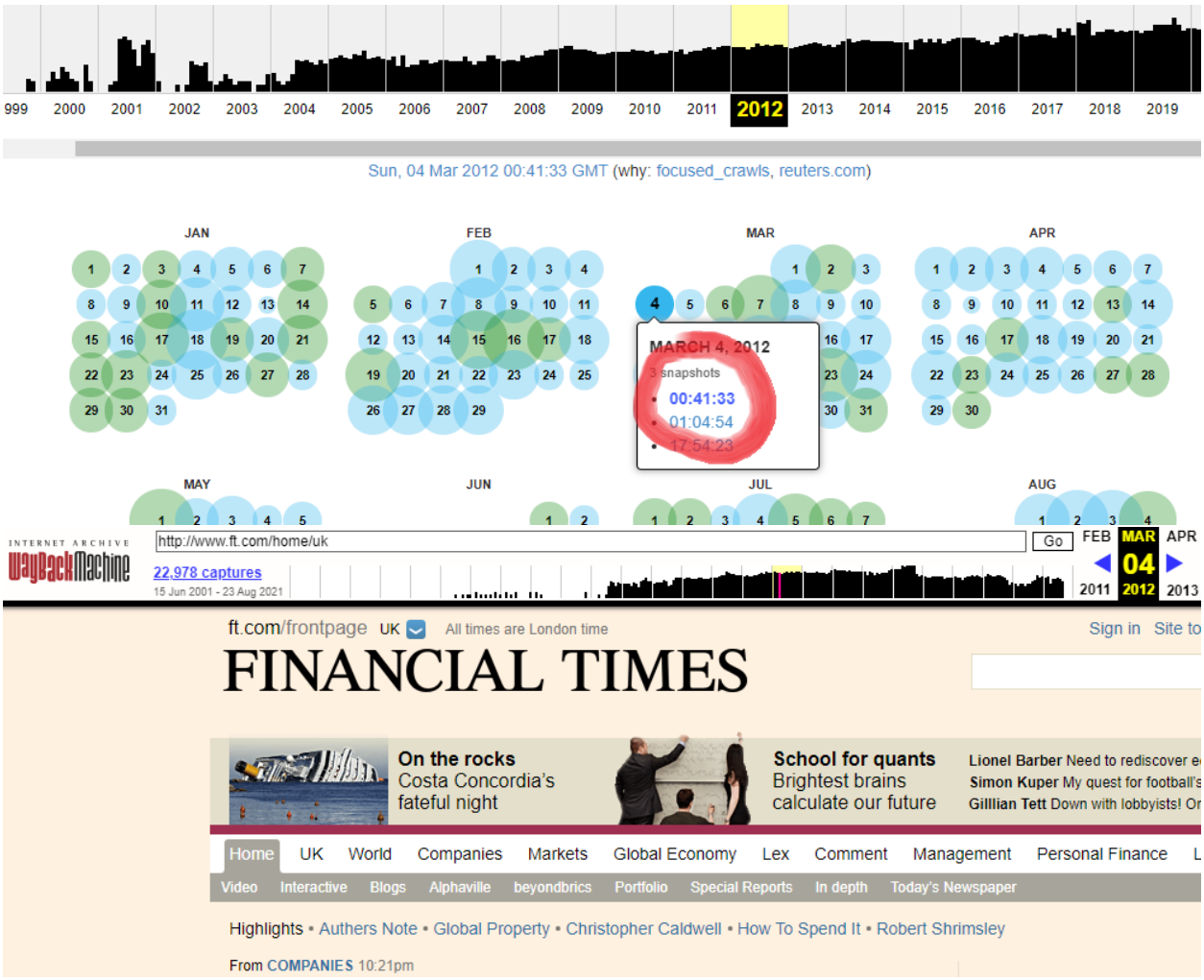

Figure 2: Wayback Machine's Interface

Neither Wayback Machine nor Linkedin contain all relevant information, so country teams can also use local knowledge to fill this information.

4 Frequently Asked Questions

Should an individual who indicates in their bio that they vote for a certain political party or that they share the views of a certain political party be coded as individual sides?

For political parties, stricter rules apply (see Table 2). Specifically, only members of the national legislature (or supranational governments like the European Union) should be coded as individual sides. This means that supporters do not count as individual side accounts. Similarly, even regular members of political parties are not coded as individual sides.

Should an individual who includes the organization's name in their bio, but their relationship to the organization is unclear, be coded as individual side?

If there is no certainty that the account is an employee or member of the organization, then they should not be included.

What if the individual's bio is in a language the coder does not understand?

When coding the individual side accounts, include only individuals whose bios are in English or a reasonable working language of the country. It is up to the country team to decide what are considered reasonable working languages.

If an individual indicates in their bio that they previously worked for the organization, should they be included as an individual side account?

No.

Should temporary employees be counted when labeling individual side accounts?

Yes. The general approach is to include everyone who satisfies the five-point criteria (i.e. identifies with the organization and is important enough to be followed by the organization).

Do regional party organizations' accounts count as collective sides?

Regional party organizations (e.g. state or provincial chapters) should not be included because they operate on a different level of public administration than the national party.

When coding university accounts, are PhD candidates included as individual side accounts?

Students should not be included as individual side accounts even if they are doctoral students employed by the university.

In terms of universities and research institutions, there are often sub-accounts for research projects or study programs. Should they be included as collective side accounts? How about accounts for joint projects where multiple organizations are mentioned in the bio?

The account should be coded as collective side for all the organizations mentioned in the bio. As discussed in subsection 3.2, each organization is collected independently of each other, so if a research project account shows up for multiple organizations, it should be included as an individual side account for every organization they are affiliated with.

What if our policy actor is a subunit of another unit in the policy actor roster?

If a policy actor a is a subunit of another actor in the data b, it will enter the data set in two ways. First, when labeling the actor a, b will be entered as a collective side. Then, it will be treated like any other policy actor on the roster and be its own collective main.

How should national branches of international organizations be coded?

The national account of the policy actor should be coded as collective main, except in cases where the national and international accounts are the same (e.g. Nokia in Finland). Only subunits/subfunctions should be coded as collective sides, meaning that other national units should never be included.

For example, when coding Greenpeace Finland's accounts, Greenpeace International or Greenpeace Spain should not be counted as collective sides, nor should be their employees or members be counted as individual sides.

Should I manually add additional side accounts?

No. Following the four- or five-point criteria for side accounts will help maintain consistency. However, if you find that the keywords you specified in point 3 do not include enough relevant individual side accounts, you can adjust those keywords then redo the side account labeling.

What if an individual (main or side) has multiple accounts (e.g. personal and official)?

All accounts belonging to an individual should be collected as long as they satisfy the conditions for that level. For individual main accounts, this means all accounts belonging to the individual main would count. For individual side accounts, the account should be included as long as it satisfies the 5-point condition in subsection 2.4

Should directors of institutes be included as individual mains for universities?

An university's individual main accounts include the top administrative executives and principal deans (or equivalent). The latter, and specifically the "or equivalent", is key to this question. If the institute is an academic institute that is at the same organizational level as a faculty or school (i.e. it is not below a faculty or school), then directors hold equivalent positions to deans, and should be included. Directors of institutes that fall under faculties or schools should not be included.

5 Programming and Other Requirements

5.1 Programming Languages

In the pipeline outlined in Figure 1, the yellow tasks require basic working knowledge of Python (steps 03 and 05) and R (step 06), which means being able to run scripts and diagnose basic errors that might arise. No code development knowledge is required. The code base for this project is stored at its GitHub repository, along with step-by-step instructions for proceeding through the pipeline.

5.2 Labour Intensity

There are three primary data collection and labeling tasks. Collecting collective main accounts should be very quick; if an organization has a main account, it will be easy to find.

Collecting the individual main accounts is the most difficult task of the three. Estimating the time that it takes to collect these accounts is difficult. It depends on the number of organizations, organization types, and the ease of locating the relevant persons on the websites and matching their names to their Twitter profiles, which depends on the context-specific culture of disclosing information about individuals working for an organization.

Finally, hand-labeling the side accounts into collective side, individual side, and unrelated is quick. A trained research assistant who is very familiar with the protocols outlined in this document should be able to classify 1000-1500 accounts in an hour.