Summary of the Replication Data for: "The Internationalization of European Financial Networks: A Quantitative Text Analysis of EU Consultation Responses" by Scott James, Stefano Pagliari, and Kevin Young, Review of International Political Economy", https://doi.org/10.7910/DVN/FBG7I4,

The data repository includes comprises of four documents:

- **Corpus**: text of the consultation responses analysed in the paper
- **Respondents**: characteristics of the groups whose lobbying is captured in the paper.
- **Consultation**: characteristics of the consultations that comprise the dataset
- **Ties**: dyadic data capturing the instance of text reuse and co-signing captured in the dataset

Corpus

This spreadsheet includes the corpus of consultation responses to a broad range of financial policiies that was scraped from the European Commission Website. Letters that are co-signed by different respondents are included multiple times in the document.
The document includes the following variables:

- **ID**: unique numeric identifier of the respondent. This ID is used to merge the data with the demographic information of the respondents;
- **Name**: name of the respondent;
- **File_Name**: unique identifier of the document;
- **Text**: text of the comment letters converted from PDF into plain text;
- **NWords**: number of words;

Respondents

This spreadsheet includes the information about the identity of the groups that signed the letters included in the corpus. Each row is a different group. The coding scheme and information regarding individual respondents is primarily derived from the Inter-Euro project (https://www.intereuro.eu/data.html), by the Transparency Register (https://ec.europa.eu/transparencyregister/public/homePage.do). These are supplemented by coding from the authors. Country-level-variable are constructed using data from the OECD and World Bank.
The document includes the following variables:

- **ID**: unique identifier of the respondent (from Inter-Euro, additional coding from the authors);
- **Name**: Full name of the respondent;
- **Website**: Website of the respondent;
- **Transparency_Register**: identifier of the respondent in the Transparency Register;
- **Num_Letter**: number of letters signed by the group in the dataset;
- **Category**: variable differentiating between different types of organizations, including between financial and non-financial business groups;
- **Finance**: dummy variable coded as 1 when an organization is a financial group;
- **Headquarters**: name of the country where the organization is based (from Inter-Euro, additional coding from the authors);
- **Head_office_city**: name of the city where the organization is based (from Transparency Register);

- **Financial_Centres**: dummy variable coded as 1 if the organization is based in a financial center;
- **LevelofmobilizationLabel**: scale of the organization of the respondent, subnational, national, European, international) (from Inter-Euro, additional coding from the authors);
- **International**: dummy variable when a group is organized above the national level (European or international);
- **Full_time_equivalent_FTE**: number of full-time equivalent staff lobbying European institutions (from the Transparency Register);
- **Corporatism**: level of corporatism of the country where the organization is based (based on Jahn 2016);
- **avg_FZN_Prod**: level of financialization of the country where a group is based by production measures (based on OECD data);
- **avg_FZN_VA**: level of financialization of the country where a group is based by value-added measures (based on OECD data);
- **Domestic_credit_to_private_sector_by_banks**: level of domestic credit to the private sector provided by banks (based on World Bank data);
- **Domestic_credit_provided_by_financial_sector:** level of domestic credit to the private sector provided by financial institutions (based on World Bank data);
- **Num_Finance_Partners_Coordinated**: number of distinct financial groups a respondent is tied to;
- **Num_Dom_Finance_Partners_Coordinated**: number of distinct financial groups based in the same country a respondent is tied to;
- **Num_For_Finance_Partners_Coordinated**: number of distinct financial groups based in another country a respondent is tied to;
- **Num_Partners_Coordinated**: number of distinct groups a respondent is tied to;
- **Num_Dom_Partners_Coordinated**: number of distinct groups based in the same country a respondent is tied to;
- **Num_For_Partners_Coordinated**: number of distinct groups based in another country a respondent is tied to.

Consultation

This spreadsheet includes the information about the policy consultations that comprise the dataset. Each row is a different policy consultation.
The document includes the following variables:

- **Consultation**: unique identifier of the consultation. This value can be used to merge with the other elements of the dataset;
- **Title**: name of the policy;
- **Date_Initial**: starting date of the consultation;
- **Date_End**: end date of the consultation;
- **Link_Page**; url of the webpage of the consultation;
- **Link_Responses**: url of the webpage containing the responses to the consultation;
- **Number_Responses**: number of responses to the consultation;
- **Objective**: short description of the objective of the consultation.

Ties

This spreadsheet includes the different ties among respondents to the same policy consultation. A tie exists between two groups if they have either submitted a co-signed response to the same

consultation or if submitted two separate responses which characterized by significant text reuse. Ties are modelled as undirected because it is not possible to establish which group instigates coordination, or the direction of information flow between them, from the data. The data is dyadic and each row is a different pair of connected groups in a single consultation.
The document includes the following variables (two versions of the same variable capture the two elements of the pair):

- **ID.x** and **ID.y:** unique identifier of the groups connect;
- **Name.x and Name.y:** Full name of the groups;
- **Category.x** and **Category.y:** variable differentiating between different types of organizations, including between financial and non-financial business groups;
- **Headquarters.x** and **Headquarters.y**: name of the country where the organization is based (from Inter-Euro, additional coding from the authors);
- **Edge_Type**: variable differentiating between text-reuse and co-signing ties;
- **Sentences_overlap**: variable capturing the number of sentences shared among the two documents. In the case of co-signed document, this captures the entire length of the document;
- **Consultation:** policy consultation where the tie was identified.