

## **Executive Summary: Influential Actors in Communication Networks**

Erdős Institute Data Science Boot Camp, May-Summer-2024

Adam Perhala, Jungbae An

### **1. Overview**

An influential actor can spread information to others in a communication network, and thus change the attitudes of others by doing so. By identifying influential actors, we can track the flow of information about a policy or product and the resulting attitudinal changes, and utilize this influence to intervene in people's attitudes or to undermine abusive interventions. In this project, we show and test a framework for detecting influential actors in the standing committee hearings in the U.S. House of Representatives—one of the communication networks of policymakers utilizing policy-relevant expertise. The influential actors identified by our framework are consistent with the relevant literature. Our detection framework can be used to optimize decision-making that leverages communication networks such as disinformation and mobilizing attention. While the tested example is developed for the academic community and public policy related groups, our framework can also be used in online consumer research, where product-related attitudes spread.

### **2. Dataset**

Our dataset consists of all transcripts of the U.S. House hearings in the 105th-114th Congresses. Hearings are a procedure used by standing committees of Congress to gain external expertise for the policy discussion for further policy agreements. Our dataset contains 1,026,677 speeches, held by 1,045 Representatives, affiliated in the 22 House committees, in the 105-114th Congresses.

### **3. Methods**

Our framework has three steps: We (1) reduce the communication to lower-dimensional issue areas, (2) build weighted networks based on issue similarity, and (3) identify influential actors with high network centrality.

(1) *Topic Classification*: To analyze the unstructured transcripts, we first reduce each speech to a meaningful low dimensionality, e.g., policy issues. We use the Latent Dirichlet Allocation (LDA) algorithm to obtain the proportion of each topic that appears in each speech, after a common preprocessing. The classification quality of our LDA model was evaluated using a perplexity score (Wallach et al., 2009). Although we used a basic topic model, issue classification can be done using different approaches, e.g., sentence embeddings, DNN-based disentanglement learning.

(2) *Similarity Networks*: Actors who can address more general issues would be more influential because they can communicate and collaborate with more colleagues. Our strategy for estimating influence is stable influence across multiple hearings. Therefore, we build a similarity network using the average topic proportion of actors who participated in the same hearing—i.e., members of the same committee. The network weighting is computed as the cosine similarity between two actors.

(3) *Influence Score*: Influential actors in the topic similarity network are identified with mean cosine similarity (edges) of interesting factors

### **4. Results**

Influential actors—i.e., better connected actors in our analysis—are found by comparing the average cosine similarity. We do not find any differences in influence between political parties in the committees, but we do find that members who hold party leadership positions have higher influence than those who do not in the earlier period of the sample analyzed.