

# Optimizing Consensus Building in vTaiwan Using Integer Programming

Operations Research, Spring 2025 (113-2)

Group M Final Project Proposal

b09607059 蔡宜芳, b12705014 陳泊華, b12705027 徐郁翔, T13H06312

Philip and T13H06303 Kuehnel Paul

Department of Information Management, National Taiwan University

May 24, 2025

## 1 Introduction

The vTaiwan system is an innovative digital democracy platform used in Taiwan that is designed to facilitate public policy deliberation. Despite its successes, challenges such as *group polarization*, *unbalanced participation*, and *inefficiencies in consensus formation* remain.

For example, during discussions on Uber’s legality, tech-savvy users and gig economy supporters dominated the dialogue, creating an echo chamber that marginalized traditional taxi drivers—highlighting the issue of **group polarization**. In the case of online alcohol sales, most participants were industry insiders and libertarian advocates, while key stakeholders like public health experts and concerned parents were largely absent, exposing the problem of **unbalanced participation** and **inefficiencies in consensus formation**. These examples underscore the need for improved design, better stakeholder engagement, and smarter facilitation to ensure vTaiwan remains both inclusive and effective.

Currently, vTaiwan uses **natural clustering** via the Pol.is platform, which groups participants based on opinion similarity. This can lead to echo

chambers where some perspectives dominate discussions. Our project aims to optimize the discussion process using **Integer Programming (IP)**, as a structured alternative to the clustering approach used in **Pol.is**. By enhancing group diversity, our approach seeks to complement Pol.is in cases where more structured deliberation is beneficial and provide a valuable extension to the system.

## 2 Data Process

### 2.1 Data Source

### 2.2 Data Process

## 3 Problem Description

In vTaiwan deliberations, the composition of discussion groups significantly affects the quality and inclusiveness of the discourse. Ideally, groups should be structured to foster constructive discussions that balance consensus and diversity. If groups are too homogeneous, they may lead to echo chambers, reinforcing pre-existing opinions without exposing participants to alternative perspectives. Conversely, if groups are too diverse, disagreements may become unproductive, leading to fragmented discussions where common ground is difficult to find.

Traditional clustering techniques—such as k-means, which partitions data by minimizing intra-cluster variance, or spectral clustering, which leverages graph Laplacians to identify communities—are effective in grouping participants based on similarity. However, these methods suffer from notable limitations in participatory settings: they offer no direct control over group sizes, cannot enforce fair representation across demographic or interest-based subgroups, and are prone to producing imbalanced or homogeneous clusters that may amplify existing biases. These drawbacks hinder inclusive and equitable deliberation. To address these challenges, we propose an optimization-based clustering approach that explicitly incorporates constraints for group size, diversity, and representational balance, enabling more

structured and fair participant grouping for policy discussions and collaborative decision-making.

### 3.1 Mathematical Model

Let  $P = \{1, 2, \dots, n\}$  denote the set of participants and  $G = \{1, 2, \dots, m\}$  the set of groups. Agreement between participants  $i$  and  $j$  is represented by  $A(i, j)$ , which quantifies their level of similarity based on their responses to policy questions or other opinion measures. To avoid excessive homogeneity, we define a diversity measure  $D(i, j)$ , which ensures that groups contain a mix of perspectives.

To model the assignment of participants to groups, we define a binary decision variable:

$$x_{ij} = \begin{cases} 1, & \text{if participant } i \text{ is assigned to group } j, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

### 3.2 Objective Function: Maximizing Agreement

To encourage consensus-driven discussions, we maximize the total agreement within groups:

$$\max \sum_{j \in G} \sum_{i, k \in P} A(i, k) x_{ij} x_{kj}. \quad (2)$$

This ensures that participants who share similar perspectives are grouped together, facilitating constructive and cooperative discussions.

### 3.3 Constraints

**Group Assignment Constraints.** Each participant must be assigned to exactly one group:

$$\sum_{j \in G} x_{ij} = 1, \quad \forall i \in P. \quad (3)$$

Additionally, group sizes must remain within predefined limits:

$$S_{\min} \leq \sum_{i \in P} x_{ij} \leq S_{\max}, \quad \forall j \in G. \quad (4)$$

**Diversity Constraint.** To prevent excessive homogeneity, we enforce a lower bound on diversity within each group. Let  $D(i, j)$  represent a measure of disagreement (e.g., ideological distance or differing response patterns). We require that, for each group, the sum of disagreement values among its members exceeds a predefined threshold  $\delta$ :

$$\sum_{i,k \in P} D(i, k) x_{ij} x_{kj} \geq \delta, \quad \forall j \in G. \quad (5)$$

This ensures that every group contains a mix of perspectives, avoiding overly uniform clusters.

## 4 Mathematical Model (V2)

Let:

- $P = \{1, 2, \dots, n\}$ : Set of participants
- $G = \{1, 2, \dots, m\}$ : Set of groups
- $A(i, k)$ : Agreement score between participant  $i$  and  $k$
- $D(i, k)$ : Diversity score (e.g., ideological distance) between participant  $i$  and  $k$
- $E(i)$ : Engagement score of participant  $i$  (e.g., number of comments or votes)
- $S(i) \in \{-1, 0, 1\}$ : Sentiment score of participant  $i$  toward a target topic (disagree, neutral, agree)
- $C(i) \in \{0, 1\}$ : Commentator indicator (1 if participant has authored at least one comment)

Define binary variable:

$$x_{ij} = \begin{cases} 1 & \text{if participant } i \text{ is assigned to group } j \\ 0 & \text{otherwise} \end{cases}$$

—

## 4.1 Objective Function

We define a composite objective to both **\*\*maximize intra-group agreement\*\*** and **\*\*minimize inter-group polarization\*\***, while ensuring active participation and viewpoint diversity:

$$\max \left[ \lambda_1 \sum_{j \in G} \sum_{i, k \in P} A(i, k) x_{ij} x_{kj} - \lambda_2 \sum_{j_1 < j_2} \sum_{i \in P} \sum_{k \in P} D(i, k) x_{ij_1} x_{kj_2} - \lambda_3 \cdot \text{StdDev} \left( \left\{ \sum_{i \in P} E(i) x_{ij} \right\}_{j \in G} \right) \right]$$

Where:

$\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are weights to balance agreement, polarization, and engagement fairness.

The second term penalizes ideological distances across groups to reduce polarization.

The third term minimizes the standard deviation of total engagement scores across groups to encourage equal participation.

—

## 4.2 Constraints

1. Unique Group Assignment Each participant is assigned to exactly one group:

$$\sum_{j \in G} x_{ij} = 1 \quad \forall i \in P$$

2. Group Size Limits and Odd Cardinality Each group must contain an **\*\*odd number\*\*** of participants within a given size range:

$$S_{\min} \leq \sum_{i \in P} x_{ij} \leq S_{\max}, \quad \forall j \in G$$

$$\sum_{i \in P} x_{ij} \bmod 2 = 1 \quad \forall j \in G$$

3. Minimum Intra-group Diversity Ensure sufficient diversity within each group:

$$\sum_{i,k \in P} D(i,k) x_{ij} x_{kj} \geq \delta, \quad \forall j \in G$$

4. At Least One Commentator per Group Ensure every group has at least one active participant who made comments:

$$\sum_{i \in P} C(i) \cdot x_{ij} \geq 1 \quad \forall j \in G$$

5. Sentiment Balance within Group Ensure each group includes a diversity of opinions on a key topic (e.g., Uber), by enforcing sentiment mix: Let  $S_{\text{agree}}, S_{\text{neutral}}, S_{\text{disagree}}$  be the number of members with each sentiment in group  $j$ :

$$\begin{aligned} S_{\text{agree},j} &= \sum_{i \in P, S(i)=+1} x_{ij} \\ S_{\text{disagree},j} &= \sum_{i \in P, S(i)=-1} x_{ij} \\ S_{\text{neutral},j} &= \sum_{i \in P, S(i)=0} x_{ij} \end{aligned}$$

Add a soft or hard constraint (depending on experiment design) to require:

$$S_{\text{agree},j} \cdot S_{\text{disagree},j} > 0 \quad (\text{at least one } + \text{ and one } - \text{ per group})$$

Or alternatively: ensure weighted sentiment score per group stays near zero:

$$\left| \sum_{i \in P} S(i) \cdot x_{ij} \right| \leq \theta, \quad \forall j \in G$$

—

### 4.3 Optional LP Relaxation

To handle scalability for large-scale deliberation settings, the binary constraint on  $x_{ij}$  may be relaxed:

$$x_{ij} \in [0, 1]$$

This allows fractional assignments, and final groupings can be obtained via rounding heuristics or stochastic sampling.

—

### 4.4 Summary

This mathematical formulation incorporates social, behavioral, and structural constraints into an Integer Programming framework that supports structured, inclusive, and diverse online deliberation. By tuning objective weights  $\lambda_1 \sim \lambda_3$  and experimenting with sentiment and engagement constraints, practitioners can flexibly adapt the model to suit different policy topics and deliberative formats.

## 5 Mathematical Model (V3)

Let:

- $P = \{1, 2, \dots, n\}$ : Set of participants
- $G = \{1, 2, \dots, m\}$ : Set of groups
- $D(i, k)$ : Diversity score between participant  $i$  and  $k$  (e.g., ideological distance)
- $E(i)$ : Engagement score of participant  $i$  (e.g., number of distinct issues voted on)
- $x_{ij} \in \{0, 1\}$ : Binary variable indicating whether participant  $i$  is assigned to group  $j$

## 5.1 Objective Function

We aim to **maximize intra-group diversity** to promote exposure to differing perspectives, while also minimizing variance in engagement across groups to ensure equitable participation.

$$\max \left[ \lambda_1 \sum_{j \in G} \sum_{i, k \in P} D(i, k) x_{ij} x_{kj} - \lambda_2 \cdot \text{Var} \left( \left\{ \sum_{i \in P} E(i) x_{ij} \right\}_{j \in G} \right) \right]$$

- The first term rewards groupings with greater internal diversity.
- The second term penalizes unequal distribution of participant engagement across groups.
- $\lambda_1, \lambda_2 > 0$  are weights that balance diversity and fairness.

## 5.2 Constraints

**(1) Unique Assignment** Each participant must be assigned to exactly one group:

$$\sum_{j \in G} x_{ij} = 1 \quad \forall i \in P$$

**(2) Group Size Bounds and Odd Cardinality** Each group must contain an odd number of participants within specified bounds:

$$S_{\min} \leq \sum_{i \in P} x_{ij} \leq S_{\max}, \quad \forall j \in G$$

$$\sum_{i \in P} x_{ij} \bmod 2 = 1 \quad \forall j \in G$$

**(3) Upper Bound on Intra-group Diversity** To prevent excessive fragmentation or conflict within groups, intra-group diversity must not exceed a predefined threshold:

$$\sum_{i, k \in P} D(i, k) x_{ij} x_{kj} \leq \delta \quad \forall j \in G$$



**(4) Minimum Group Engagement** Ensure that each group contains a sufficient level of active participation:

$$\sum_{i \in P} E(i) x_{ij} \geq \eta \quad \forall j \in G$$

### 5.3 Optional LP Relaxation

To scale to larger datasets, binary constraints on  $x_{ij}$  can be relaxed:

$$x_{ij} \in [0, 1]$$

A fractional solution can then be post-processed through probabilistic rounding or heuristic assignment.

### 5.4 Summary

This optimization model supports structured and inclusive deliberation by explicitly balancing diversity, engagement, and group structure constraints. The use of diversity bounds avoids unproductive polarization, while the engagement-aware objective promotes equitable group participation.

## 6 Performance Analysis

Solving the integer program exactly may be computationally expensive for large-scale deliberation settings. To improve efficiency, we consider a Linear Programming (LP) relaxation, where the binary constraint on  $x_{ij}$  is replaced with:

$$x_{ij} \in [0, 1]. \tag{6}$$

This allows fractional assignments, which can be interpreted probabilistically or rounded to obtain integer solutions. While this relaxation sacrifices strict guarantees on discrete assignments, it provides a computationally efficient approximation that can guide group formation.

## 7 Conclusion and Expected Results

By formulating the group formation problem as an Integer Program, we introduce a structured and flexible approach to participant assignment in deliberative settings. Our model maximizes intra-group agreement to foster constructive discussions while enforcing a diversity constraint to prevent excessive homogeneity. This formulation overcomes the limitations of traditional clustering methods, which do not provide explicit control over group size or diversity. The inclusion of an LP relaxation offers a computationally efficient alternative, making the approach scalable for large deliberative processes. The proposed optimization model provides a practical and theoretically grounded method for improving discussion dynamics in vTaiwan and similar participatory platforms. We expect the following results. First, we anticipate **more balanced groups** in terms of both consensus and diversity, compared to the natural clustering method currently used in Pol.is. Second, we expect **improved discussion efficiency** through optimal facilitator assignment. Third, we foresee **more representative decisions**, as the structured approach ensures the incorporation of diverse perspectives into discussions.

## 8 Data Collection and Implementation Plan

vTaiwan currently utilizes the open-source platform Pol.is, which powers policy deliberation using machine learning algorithms. For our project, we will run our Operations Research (OR) model using the open datasets provided by Pol.is on GitHub.

### 8.1 Data Sources

**participants-votes.csv:** This dataset includes participant voting behavior and group assignments, such as the number of comments written, the number of votes cast, and the group each participant belongs to based on K-means clustering.

**comments.csv:** This dataset contains details about each comment made during the deliberation, including the author’s ID, vote counts (agrees, dis-

agrees), moderation status, and the text of the comment.

## 8.2 Pre-moderation Statistics

Before moderation, the dataset contains the following summary:

- Total Votes: 49,997 votes cast in the conversation.
- Total Comments: 197 comments written by participants.
- Total Visitors: 4,668 unique visitors to the conversation.
- Total Voters: 49,997 participants who voted on the comments.
- Total Commenters: 197 participants who contributed written comments.

These values represent the raw data before any moderation is applied. In Pol.is, moderation involves filtering or removing content that violates platform guidelines (e.g., abusive or irrelevant comments). Comments are flagged as allowed (1), unmoderated (0), or moderated out (-1), ensuring the discussion remains focused and respectful.

## 8.3 Implementation Plan

We will apply an Integer Programming (IP) model to optimize the group formation process in vTaiwan deliberations. The goal is to create groups that strike a balance between intra-group consensus and inter-group diversity while minimizing group polarization.

The initial step involves **data preparation**, where datasets will be preprocessed to extract relevant features such as participant votes, comment data, and group assignments. This stage also includes cleaning and formatting the data to ensure compatibility with the optimization model.

Next, we will develop an **Integer Programming Model**. This involves defining the optimization problem using participant voting data and group assignments as inputs. The objectives will be set to maximize consensus within groups while ensuring diversity across groups. Constraints

will be introduced to maintain the desired group size, balance, and reduce polarization.

Following model development, **model execution** will be performed. The preprocessed data will serve as input to the IP model, and the optimization problem will be solved to obtain the best possible group configurations.

The performance of the optimized groups will then undergo **evaluation and benchmarking**. This will involve assessing predefined metrics such as group diversity, consensus, and polarization. The model’s performance will be benchmarked against baseline clustering methods (e.g., K-means) to assess the improvement in deliberation quality.

Finally, **implementation and refinement** will be conducted. Based on the evaluation results, the model will be refined as necessary to improve group formation. We will also explore the impact of different constraints and parameters on the optimization process to enhance the flexibility and robustness of the model.

By applying this structured approach, we aim to improve the deliberation process on vTaiwan, making it more inclusive, balanced, and conducive to informed decision-making.