REASSESSING CONSTRUAL CLUSTERS WITH BIPOLAR DATA: MEASURING SIMILARITIES AND TDA PERSPECTIVES

Manuel Cuerno[1], Sergio Galaz-Garcia[2], Fernando Galaz-García[3] & Telmo Pérez-Izquierdo[4]

CUNEF University[1], Universidad Carlos III de Madrid[2] (UC3M), Durham University[3] & Universidad del País Vasco[4] (UPV)

CUNEF
COLEGIO UNIVERSITARIO DE ESTUDIOS FINANCIEROS

INTRODUCTION

Empirical knowledge on **construal classes**, or "social affinity groups of individuals who share similarities in how they organize their outlooks on a collection of issues", has significantly expanded in recent years. These discoveries have been enabled by a family of methods that make use of community-detection algorithms to cluster respondents from representative surveys according to how similarly their answers change among questions. These clusters are interpreted as empirical estimations of construal classes at a societal level. The use of these methods, which we will refer as **Construal Clustering Methods** (**CCMs**), has produced seminal information about the number, relative population, and characteristics of construal classes for issues as diverse as artistic consumption, political orientations, attitudes towards religion and science, and opinions on the role of the market in social life.

Most of these findings are based on the use of Relational Class Analysis (RCA) [4]. Furthermore, on 2017 Boutyline recently introduced Correlational Class Analysis (CCA) [2]. And, finally, on 2023 Sotoudeh and Dimaggio [5] presented a new version of RCA (N-RCA) that constitutes the current state of the art regarding CCMs. All these methods share the same schema: first, they compute a *similarity function* between pairs of respondents of certain survey, and, second, apply a community-detection algorithm in order to obtain the construal classes. All the previous mentioned methods differ in how they calculate the first step and only N-RCA uses a different community-detection algorithm.

After testing the existent CCMs, we concluded they carry some structural deficiencies. The most important one is that they are not planned for *bipolar questions*, the most common questions in opinion surveys. Bipolar questions are characterized as the union of two sets (*positive and negative opinion semispace*) of possible answers to that question with, at most, non empty intersection in a *neutral option*. For example, an `agree-disagree` question would be a bipolar one: $\mathcal{P} = \{\texttt{Agree}, \texttt{Neither agree nor disagree}\}$, $\mathcal{N} = \{\texttt{Disagree}, \texttt{Neither agree nor disagree}\}$ with neutrality option `Neither agree nor disagree`.

In this poster, we present **Bipolar Class Analysis** (BCA) to address the problems identified in previous CCMs and provide a more accurate method for analyzing bipolar data.

BIPOLAR CLASS ANALYSIS (BCA)

Let $u = (u_1, \ldots, u_Q)$ and $v = (v_1, \ldots, v_Q)$ be two respondents to certain poll and $u_{kl} = (u_k, u_l)$ a pair of answers. Then, we compute the similarity function between u, v following the next formula:

$$B_j(u,v) = \frac{2}{Q(Q-1)} \sum_{k=1}^{Q-1} \sum_{l=k+1}^{Q} \omega(u_{kl}, v_{kl}) \mu_j(u_{kl}, v_{kl}),$$

for $j = 1, 2$ depending on one of the two magnitudes we will introduce.

Polarity function ω

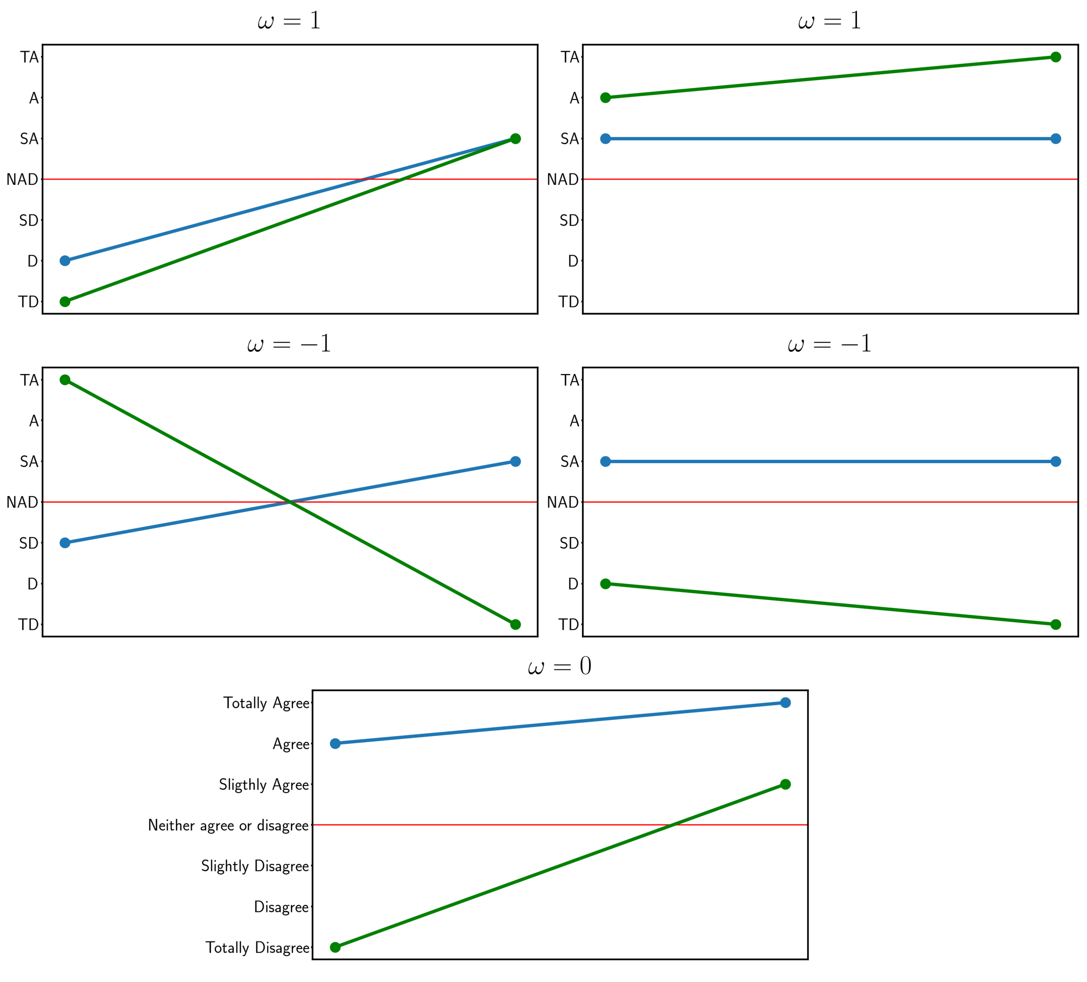

Fig. 1: Values of $\omega(u_{kl}, v_{kl})$ when $u_{kl}, v_{kl} \neq n_{kl} = (n_k, n_l)$

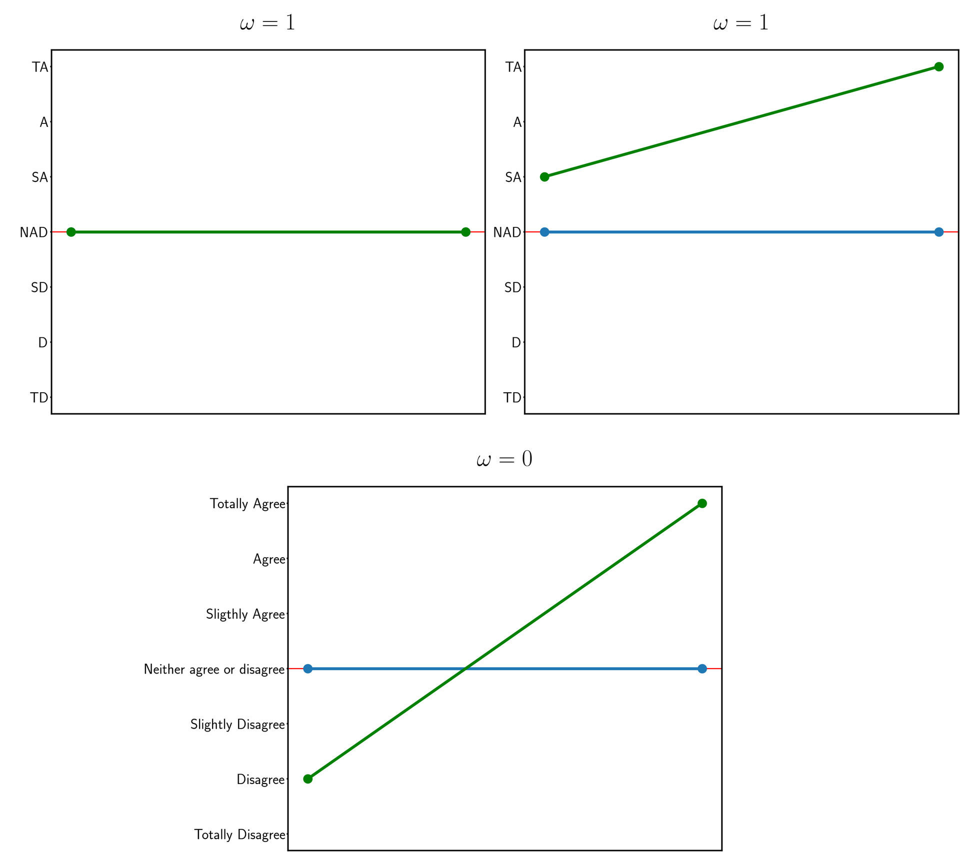

Fig. 2: Values of $\omega(u_{kl}, v_{kl})$ when u_{kl} or $v_{kl} = n_{kl} = (n_k, n_l)$

MAGNITUDE TERM μ_j

We fix two diferent magnitudes μ_1 and μ_2.

$$\mu_1(u_{kl}, v_{kl}) = 1.$$

Given the bipolar structure of the data, this option is always feasible without requiring additional structural impositions. Therefore, we consider it a suitable benchmark. Comparing results obtained using μ_1 with those from other functions can help assess their robustness. We will denote BCA with this magnitude as **BCA$_1$**.

$$\mu_2(u_{kl}, v_{kl}) = 1 - \frac{|\Delta(u_{kl}) - \Delta(v_{kl})|}{\max_q(D_q)},$$

where

$$\Delta(u_{kl}) = |dist_k(u_k, n_k) - dist_l(u_l, n_l)|,$$

$dist_i(\cdot,\cdot)$ denotes a metric function that measures the distance from answer u_i to the neutrality element n_i and $D_q = \max_{q \in Q}\{dist(a_q, n_q)\}$ where a_q is a possible answer of question q.

μ_2 needs the answers of each question to follow a *total order* and neutrality must be an explicit option. This method avoids comparing answers across different questions and provides an implicit normalization. This approach allows BCA to address more effectively the variation in response numbers compared to RCA. We will denote BCA with this magnitude as **BCA$_2$**.

SIMULATIONS

We define a new simulation method for this kind of data (opinion polls with bipolar questions). These simulations adhere to a modeling strategy that incorporates **ordered choice and multivariate dependency**. Respondents hold a specific latent position towards each question, which they use to select opinion values from a finite set of alternatives. The positions of respondents across questions are constrained through a specific dependency structure. This structure, interpretable as a construal, is predetermined by design.

We present three different *performance metrics* where BCA performs 10 times better than the other CCMs in terms of *right number of classes* and at least as well as them in the other two metrics.

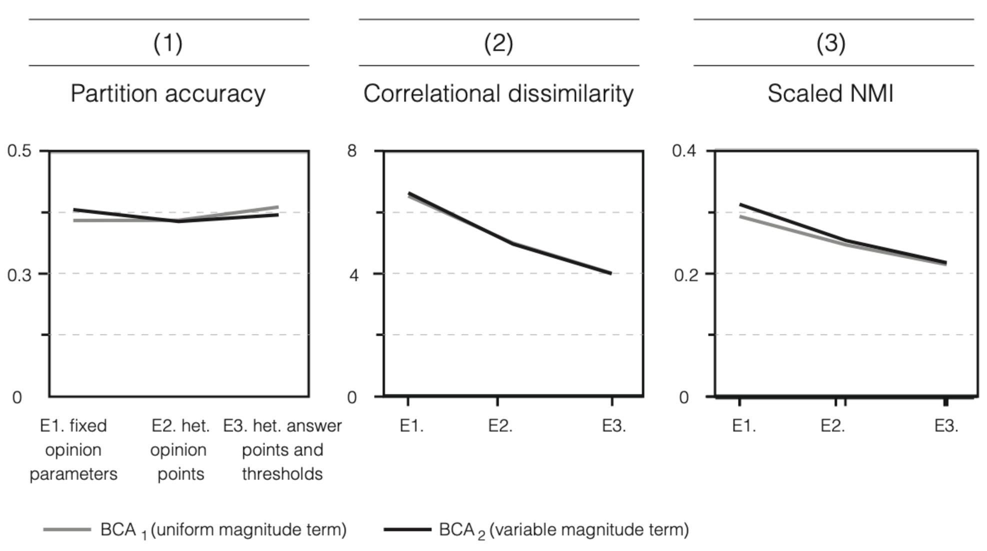

Fig. 3: BCA$_1$ and BCA$_2$' performance metrics

Real-data results and TDA

We also show that BCA's construal clusters are not only more accurate than those from other CCMs but also substantively different. We do so by applying BCA to re-analyze the *ANES survey data* used in [1] to analyze **groups of political opinion similarity** among American voters.

In future studies, we also plan to use **TDA** to focus on political areas such as *polarization*. Using the simulation model we defined and the correlation matrices of each construal extracted from real data, we will model a synthetic construal with extreme opinions and measure, via **Persistence Diagrams and the bottleneck distance**, how polarized a certain construal is and how its polarization has evolved historically.

References

[1] Delia Baldassarri et al. "Neither Ideologues nor Agnostics: Alternative Voters' Belief System in an Age of Partisan Politics". In: American Journal of Sociology 120.1 (2014), pp. 45–95. ISSN: 00029602, 15375390. (Visited on 11/17/2023).

[2] Andrei Boutyline. "Improving the measurement of shared cultural schemas with correlational class analysis: Theory and method". In: Sociological Science 4 (2017), pp. 353–393.

[3] Manuel Cuerno et al. "Reassessing Relationality for Bipolar Data". In: (2024). arXiv: 2404.17042.

[4] Amir Goldberg. "Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined". In: American journal of sociology 116.5 (2011), pp. 1397–1436.

[5] Ramina Sotoudeh et al. "Coping with plenitude: A computational approach to selecting the right algorithm". In: Sociological Methods & Research 52.4 (2023), pp. 1838–1882.

Contact: manuel.mellado@cunef.edu Webpage: https://manuelcuerno.github.io/personalpage/