Surname Affinity Predicts Political Preference of Campaign Contributors in Chile

Fran Zautzik[1,2], Naim Bro[2,3,4] and Aidan Hogan[1,2]

[1] DCC, Universidad de Chile; Santiago, Chile
[2] IMFD; Santiago, Chile
[3] Universidad Adolfo Ibañez; Santiago, Chile
[4] COES; Santiago, Chile
fzautzik@dcc.uchile.cl, naim.bro.k@uai.cl, ahogan@dcc.uchile.cl

1 Introduction

This study explores whether or not there is an association between surname proximity in the social space [1] and political preference among campaign contributors in Chile. More specifically, are campaign contributors that have socially similar surnames more likely to contribute to the same political candidate?

The study uses data from the Chilean Electoral Service (SERVEL), focusing on the 2017 and 2021 presidential elections.

Individuals in Chile generally possess two surnames: one from the father and the other from the mother. We measure the social proximity of surnames using a surname network built from all the individual's surname pairs present in the electoral rol. Two networks, where each node represents a surname, are constructed. For the first network, the weight of the edges between surnames corresponds to the number of individuals in the dataset that share surnames (we refer to it as "the original network"). The second network is obtained by recalculating the edge weights of the original network as the Jaccard Similarity between two surnames they connect, i.e. the number of individuals with both surnames divided by the number of individuals with the first surname or the second surname , or $\frac{|A \cap B|}{|A \cap B|}$ (we refer to it as the normalized network).

To measure the surname affinity between two individuals, the measures proposed are the average edge weight between surname pairs in the original surname network (***edge-weight***), and the average edge weight between surname pairs in the normalized surname network (***edge-weight-j***). Additional surname affinity measures were devised but are not included in this abstract.

To measure political similarity between two contributors we compare their preference vectors. The preference vector is a vector representation of an individual's preference in an election, where each component of the vector corresponds to a candidate, and the value in a given component represents the amount of money given by the individual to the corresponding candidate. The vectors are normalized and then compared using the Hellinger distance (***hellinger-distance***), a measure used to quantify the similarity between two distributions that goes from 0 to 1. It's important to note that this value is a distance, so smaller values reflect bigger similarity, for this reason the value is then multiplied by -1, this way the value grows together with similarity. Here too, other measures were used for political preference similarity but are not presented.

A dyadic data analysis approach is then employed, constructing pairs of contributors (dyads) and calculating the various surname affinity measures and political preference similarity measures. The surname affinity measures with the political preference similarity measures are then tested for correlations using three different correlation coefficients, Pearson's R, used for linear relations, along with Spearman's Rho and Kendall's Tau, which are rank based correlations.

2 Results

	Pearson's R	**Spearman's Rho**	**Kendall's Tau**
	hellinger-distance	hellinger-distance	hellinger-distance
edge-weight	0.013*	0.056*	0.041*
edge-weight-j	0.016*	0.055*	0.041*

Table 1. Table showing the correlations between the surname affinity measures (***edge-weight*** and ***edge-weight-j***) and political preference similarity measures (***hellinger-distance***) for the 2017 dyads using the three correlation coefficients.

	Pearson's R	**Spearman's Rho**	**Kendall's Tau**
	hellinger-distance	hellinger-distance	hellinger-distance
edge-weight	0.017*	0.104*	0.074*
edge-weight-j	0.018*	0.104*	0.074*

Table 2. Table showing the correlations between the surname affinity measures (***edge-weights*** and ***edge-weights-j***) and political preference similarity measures (***hellinger-distance***) for the 2021 dyads using the three correlation coefficients.

Observing the resulting correlations of both elections in Tables 1 and 2, it can be seen that there is a positive correlation between surname affinity measures and political preference similarity measures, but the correlations are very weak. Also, it can be observed that correlations for ***edge-weight*** and ***edge-weight-j*** are almost identical in both elections. Considering the different correlations coefficients it can be said that the rank-based coefficients yield higher results, suggesting the relationship between the variables can be better described as a monotonically growing function rather than a linear function. It can also be noted that correlations for the 2021 elections are higher than for the previous election, almost twice as high in the case of rank-based coefficients.

As for the statistical significance of the results, the p-values resulting for the correlation tests are all significant (< 0.01). However, this results can be misleading considering the large amount of observations (all contributor pairs) so the Monte Carlo permutation test was used to obtain empirical p-values which also resulted in statistically significant values (< 0.01).

The additional surname affinity measures and political preference similarity measures created, which are not shown in this abstract, were also tested for correlation and yield results which are consistent with those presented here.

We conclude that there is a weak but statistically significant correlation between surname affinity and political preference among campaign contributors in Chile. Future research could refine the measures used and expand the scope to include other types of elections and data sources.

Summary: This study investigates the correlation between surname affinity and political preference among campaign contributors in Chile. By analyzing the publicly available data from the 2017 and 2021 regular presidential elections, measures of surname affinity and political preference similarity were proposed and tested for correlation. The findings indicate predominantly positive correlations between surname affinity and political preference, though the strength of these correlations is generally weak. These results suggest a complex relationship influenced by various factors, warranting further investigation with refined measures and additional data.

References

1. Bro, N., Mendoza, M.: Surname affinity in Santiago, Chile: A network-based approach that uncovers urban segregation. PloS one 16(1), e0244372 (2021)