

A Generalized Voting Game for Categorical Network Choices

Yueh Lin¹, Stefano Nasini², and Martine Labbé³

¹IESEG School of Management, Lille, France, ✉ y.lin@ieseg.fr

²IESEG School of Management, Lille, France, ✉ s.nasini@ieseg.fr

³Free University of Bruxelles, Bruxelles, Belgium, ✉ martine.labbe@ulb.be

Pairwise dependencies among categorical outcomes arise in a wide range of statistical estimation contexts. For instance, countries align or oppose each other’s policy positions, adolescents’ choices are shaped by their peers, and firms imitate or differentiate their products in response to competitors. In all these settings, individual outcomes are not independent but reflect structured patterns of similarity and dissimilarity that are naturally formalized through networks.

From this perspective, several statistical estimation problems involving network dependencies (such as, data classification, network propagation, and network discovery) are often treated separately. However, they can be viewed as different facets of the same underlying phenomenon concerning network dependencies in individual choices. Specifically, when the dependency structure is known, one may predict individual choices, as in data classification, or simulate how choices propagate through the network. Conversely, when choices are observed, one may seek to infer the dependency structure that rationalizes them. These tasks differ only in what is taken as exogenous and what is inferred, rather than in their fundamental logic. Despite this conceptual unity, existing approaches typically address these problems in isolation. Models of influence diffusion, such as linear threshold and cascade models, are widely used to simulate propagation dynamics but often lack a clear behavioral micro-foundation and are weakly connected to estimation theory. Conversely, econometric and statistical approaches to modeling cross-sectional dependence focus on estimating correlations and similarities in categorical data but abstract from the strategic and nonlinear mechanisms underlying influence and imitation. As a result, modern machine learning methods for classification and imputation often operate as black boxes, with limited interpretability or behavioral grounding.

We propose a unified solution by modeling categorical choices as the outcome of strategic interactions over a signed and weighted network. Individuals are connected through relationships that can be either attractive (encouraging alignment) or repulsive (encouraging differentiation), each with a measurable strength. For each categorical feature, an individual aggregates the influence of their neighbors through a voting rule, which determines whether the feature is adopted. The voting rule is governed by an individual-specific threshold that separates perceived gains from losses, providing a natural behavioral interpretation rooted in reference-dependent preferences.

The resulting behavioral mechanism gives rise to two complementary game-theoretic formulations. In the Direct Voting Game for Categorical Network Choices (DVG-CNC), the network structure is taken as given and individuals’ choices are endogenous. This formulation supports data classification and network propagation and generalizes the K -nearest neighbors (KNN) methodology by allowing heterogeneous thresholds, signed relationships, and weighted influence. In contrast, the Inverse Voting Game for Categorical Network Choices (IVG-CNC) takes observed choices as given and treats the network of pairwise relationships as endogenous. This inverse formulation addresses the problem of network discovery by identifying the attractive and repulsive ties that could have generated the observed choice patterns.

On the theoretical side, we provide sufficient conditions for the existence of Nash equilibria in the DVG-CNC and characterize individual best responses in closed form. We show that equilibrium existence depends critically on the interplay between mandatory and forbidden choices and the structure of attractive and repulsive links. We also establish a labeling symmetry property, which implies that payoff and equilibrium properties are invariant to feature relabeling under suitable threshold adjustments. From a computational perspective, we prove that the problem of characterizing Nash equilibria is NP-complete, highlighting the intrinsic difficulty of equilibrium computation in this class of games. To address this challenge, we develop MILP formulations for both the DVG-CNC and the IVG-CNC. These formulations

allow us to compute social welfare solutions and approximate Nash equilibria using state-of-the-art optimization solvers. The MILP approach provides a practical bridge between game-theoretic modeling and large-scale empirical applications.

Empirically, we illustrate the framework through three applications. The first analyzes geopolitical alliances among 184 countries and predicts national policy positions on issues such as sustainability, LGBT legal protection, migration, and international recognition. Out-of-sample results show that the Nash equilibrium solution of the DVG-CNC fits observed choices better than the social welfare benchmark, consistent with strategic behavior by countries. The second application uses the DVG-CNC for missing data imputation, demonstrating substantial performance gains over the traditional KNN approach. The third application employs the IVG-CNC to uncover network influences among secondary school students, revealing interpretable patterns of peer attraction and repulsion.

Taken together, these results highlight the generality of the network choice problem and demonstrate how a single behavioral mechanism can unify classification, propagation, and network discovery. By providing a game-theoretic micro-foundation for commonly used machine learning techniques and enhancing their empirical performance, the proposed framework offers a transparent and flexible approach to modeling categorical data with network dependencies.