SOCIUS

American Sociological Association

Heatmaps for Patterns of Association in log-Linear Models

Mauricio Bucca[1] (iD)

Abstract
Log-linear models offer a detailed characterization of the association between categorical variables, but the breadth of their outputs is difficult to grasp because of the large number of parameters these models entail. Revisiting seminal findings and data from sociological work on social mobility, the author illustrates the use of heatmaps as a visualization technique to convey the complex patterns of association captured by log-linear models. In particular, turning log odds ratios derived from a model's predicted counts into heatmaps makes it possible to summarize large amounts of information and facilitates comparison across models' outcomes.

Keywords
log-linear models, heatmaps, social mobility

Log-linear models for contingency tables play a crucial role in the sociological study of social mobility and assortative mating. The basic goal of these models is to describe the association between categorical variables as a function of two distinct quantities: the marginal distribution of the variables and the net association between them (Agresti 2002). Mobility scholars, for example, want to distinguish temporal changes or cross-country differences in relative mobility from differences in the occupational structure across time and place. Another reason why log-linear models are appealing is that they capture *patterns* of association between variables, without reducing them to a single summary measure (e.g., correlation coefficients). In this vein, a key finding in mobility research is that patterns of social mobility are remarkably similar across industrialized countries (Erikson and Goldthorpe 1992).

Describing complex patterns, however, comes at the cost of parsimony. Log-linear models typically involve a large number of parameters, making it difficult for researchers to directly examine a model's outcomes or to compare across model candidates. Moreover, the multiple ways these models can represent association in contingency tables (e.g., "topological" vs. "ordinal" models, log-linear vs. log-multiplicative models; Powers and Xie 2000) preclude a common meaning for parameters across different models. For these reasons, it is customary to first decide on a preferred model (e.g., via the Akaike information criterion or the Bayesian information criterion) and then draw substantive conclusions based on it. Once a preferred model has been decided upon, the researcher will likely focus on the subset of parameters

relevant to test the theories of interest, relegating the remaining results to a secondary role. For the same reasons, researchers rarely compare patterns of association derived from their preferred model against *observed* patterns or those yielded by alternative models.

In sum, log-linear models are able to provide a rich characterization of patterns of association between variables, but the full picture is often missed because of practical constraints. I propose the use of heatmaps, a type of graph that maps values contained in a matrix into colors of different intensity, as a simple way to visualize these patterns. In particular, by turning log odds ratios derived from a model's predicted counts into a heatmap, it is possible to visualize complex patterns of association that are otherwise hard to convey. Moreover, visualizing log odds ratios implied by different models would make it possible to compare outputs that are not always readily comparable.

Revisiting canonical work by Erikson, Goldthorpe, and Portocarero (1982) and Xie (1992), Figure 1A reports the margins-free association between class of origin and destination in England, France, and Sweden under the unidiff model, the authors' preferred model. This figure effectively conveys the main findings from these works, namely, that patterns of

[1]European University Institute, Florence, Italy

Corresponding Author:
Mauricio Bucca, European University Institute, Department of Political and Social Sciences, Via dei Roccettini, 9, 50014 San Domenico di Fiesole, Fiesole, Firenze 50014, Italy.
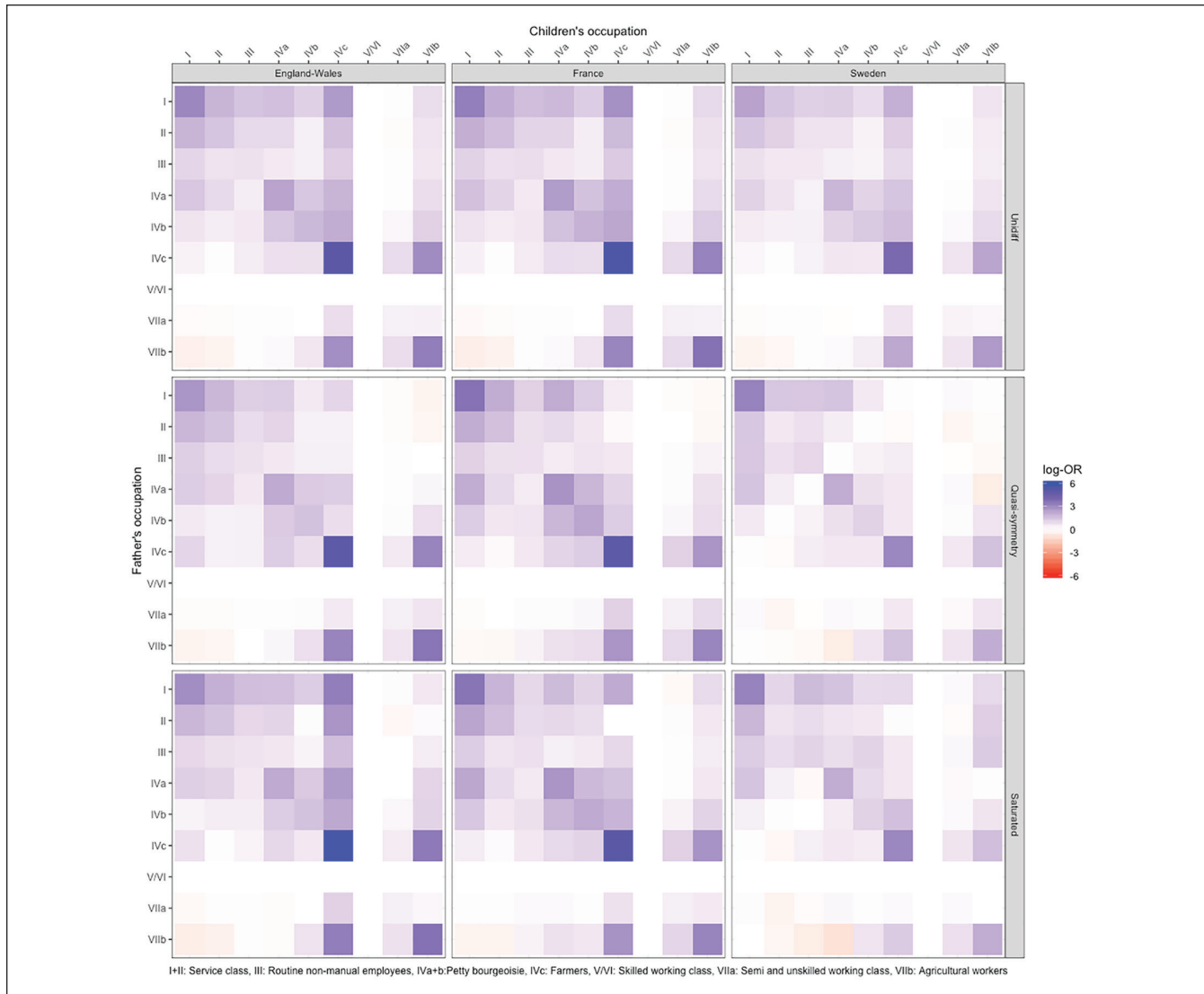E-mail: mauricio.bucca@eui.eu

Figure 1. Patterns of relative social (im)mobility in three countries under three models. The figure displays patterns of association in the three-way contingency table, cross-classifying class of origin, class of destination, and country. Panel A plots results from the unidiff model, panel B shows results from the quasi-symmetry model, and panel C corresponds to the saturated model. Each cell in these heatmaps represents the margins-free log odds ratio of every origin-destination combination in each country, with respect to a reference category (origin and destination "V/VI: Skilled working class" in "England-Wales"). The coloring of the cells indicates both the sign and the size of the log odds ratio. White coloring means that under a given model, the predicted count in the cell is not different from what would be expected by the marginal distributions of the variables, indicated by main effects and two-way interactions in the case of a three-way contingency table. Blue coloring means that the predicted count is larger than expected, while red coloring indicates that the predicted count is smaller than expected. As for size, darker coloring represents a larger log odds ratio.

relative mobility have a similar structure in all three countries, but Sweden features greater fluidity compared with France and England. Moreover, by transposing the unidiff's 84 parameters into a single, intuitive visualization, the plots in Figure 1's first row make it possible not only to know that there is a common (im)mobility structure but to explore its topology. In addition, comparison across all rows of Figure 1 shows that patterns implied by the unidiff model share important commonalities with those yielded by the quasi-symmetry model—Figure 1's second row, the second best

model according to the Bayesian information criterion (see the Supplemental Materials)—and the observed patters, as described by the saturated model (Figure 1's third row).

ORCID iD

Mauricio Bucca ![iD] https://orcid.org/0000-0003-3566-5953

Supplemental Material

Supplemental material for this article is available online.

References

Agresti, Alan. 2002. *Categorical Data Analysis*. Hoboken, NJ: Wiley-Interscience.

Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford, UK: Clarendon.

Erikson, Robert, John H. Goldthorpe, and Lucienne Portocarero. 1982. "Social Fluidity in Industrial Nations: England, France and Sweden." *British Journal of Sociology* 33(1):1–34

Powers, Daniel A., and Yu Xie. 2000. *Statistical Methods for Categorical Data Analysis*. San Diego, CA: Academic Press.

Xie, Yu. 1992. "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57(3):380–95.

Author Biography

Mauricio Bucca is a Max Weber Postdoctoral Fellow at the European University Institute and previously earned a PhD in sociology at Cornell University. He studies labor market inequalities, intergenerational mobility, and beliefs about inequality using a combination of statistical modeling, empirical strategies for causal inference, and experimental and computational methods. His scholarly agenda is articulated around two core, complementary research streams. One series of projects examines key structural aspects of social inequality: intergenerational mobility, gender, and ethnoracial gaps in the labor market and assortative mating. Another line of research focuses on the cultural facets of social stratification as they crystallize in specific belief systems about inequalities and fairness. His work has been published in academic journals such as *Science Advances, Sociological Methods and Research, RSF: The Russell Sage Foundation Journal of the Social Sciences*, and *Research in Social Stratification and Mobility*.