

A model for French voters: Supplemental Material

Antoine Vendeville*

médialab, Sciences Po, 75007 Paris, France

Complex Systems Institute of Paris Île-de-France (ISC-PIF) CNRS, 75013 Paris, France and

Learning Planet Institute, Research Unit Learning Transitions, 75004 Paris, France

(Dated: January 22, 2025)

I. DETAILS OF COMPUTATIONS

The values of x_i^s and ρ_{ij} are computed with iterative algorithms described in ref. [1]. The argmax of x_i and z_i may lie in several of their coordinates at the same time ; when comparing these argmax with the ground-truth opinion, we consider that the ground-truth is correctly identified if one of the coordinates where the argmax is reached matches the ground-truth. The support vector classifier and the logistic regression are performed using the dedicated functions of Python’s `sklearn` library. We use default parameters ; grid-searching for optimal values did not yield significantly better results. We use `class_weight=balanced` for the logistic regression, because some parties have more supporters than others (cf. Fig 1) and logistic regression is sensitive to class imbalance. Smoothed distributions are plotted using the `kdeplot` function of Python’s `seaborn` library with default parameters.

II. STATISTICS OF NETWORKS

The existence of a unique solution to Eq. 1 requires that every user can be reached by at least one zealot via a path in the network [2, Thm. 2.1]. For each network considered (retweet, Unw., Und., UU, Follow, Mention) we remove all the users who cannot be reached by a zealot. Each network also exhibits a giant (weakly) connected component and several very small components of size one or two. Because our purpose is to study the effect of the user-to-user influence mechanism encapsulated by the voter model, we restrict ourselves to the giant component of each network. After these filtering steps, the system is composed of a number of users and zealots that varies across the different networks considered: 15,607 users and 1,842 zealots for Retweet and Unw. ; 15,996 users and 1,842 zealots for Und. and UU ; 11,920 users and 1,516 zealots for Follow ; 12,786 users and 1,844 zealots for Mention. The distribution of ground-truth labels for users and zealots in the retweet network is shown in Fig. 1. The other networks exhibit highly similar distributions.

* antoine.vendeville@sciencespo.fr

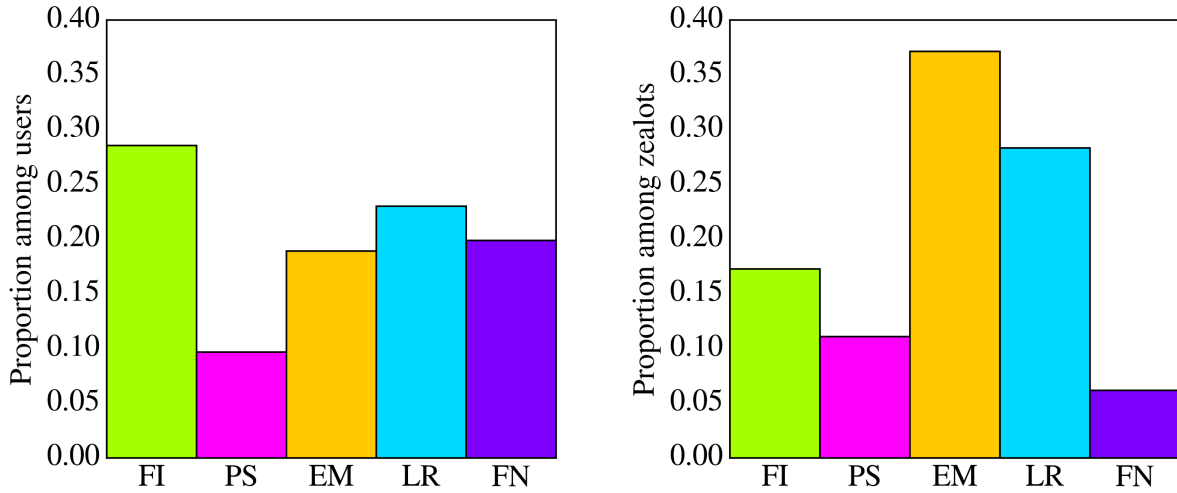


FIG. 1: Distribution of ground-truth opinions in the dataset, for users (left) and zealots (right).

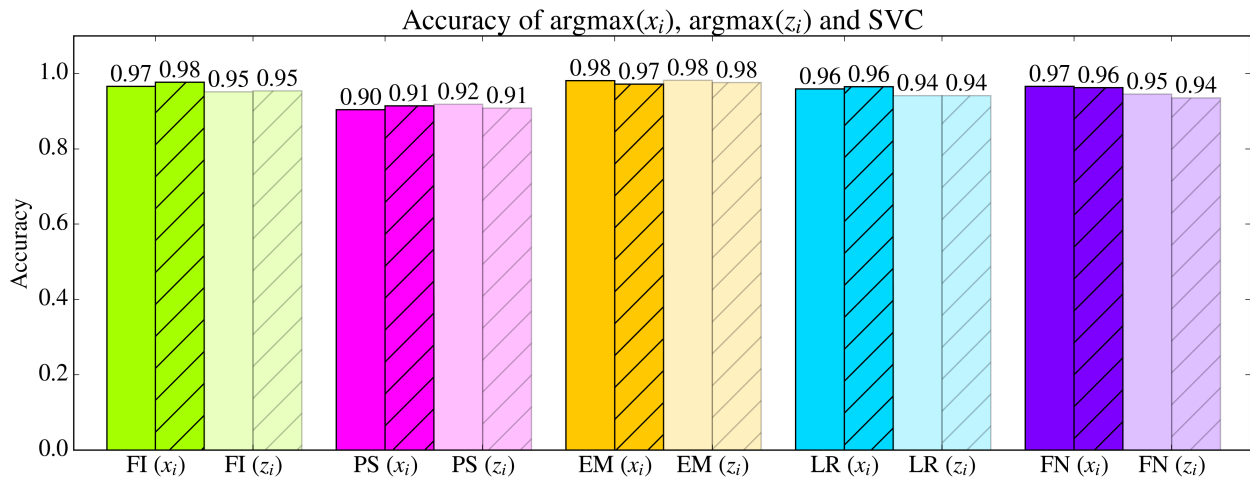


FIG. 2: Party-wise accuracy of $\text{argmax}(x_i)$ (plain bars) and SVC (hatched bars) for the retrieval of ground-truth opinions y_i from x_i and z_i . Dark bars stand for x_i and transparent bars for z_i .

III. ADDITIONAL RESULTS AND FIGURES

A. Direct connections with zealots

Figure 2 and Table I provide more detailed results about the comparison between the performance of x_i and the performance of z_i . As stated in the main text, while users seem to be better separated along party lines in the opinion space by z_i than by x_i (last four lines of Table I), ground-truth opinions are more easily recognized by x_i (first two lines of Table I).

	Voter model ($u_i = x_i$)	Baseline ($u_i = z_i$)
Accuracy $\operatorname{argmax}_{s \in \mathcal{S}} u_i^s$	96.2	95.1
Accuracy SVC	96.4	94.8
$\langle u^s \rangle$	0.826	0.925
$\langle u^{-s} \rangle$	0.044	0.019
$\langle \ u_i - u_j\ \rangle_{\text{within}}$	0.219	0.190
$\langle \ u_i - u_j\ \rangle_{\text{cross}}$	1.146	1.339

TABLE I: Comparison between the equilibrium opinions in the voter model and a baseline that considers solely direct connections with zealots. The notation u refers to x for the voter model and to z for the baseline. The network is restricted to users with at least one direct connection with zealots. Subscripts “within” and “cross” precise whether the average is taken over pairs of users supporting the same party or two different parties. Accuracies are given as percentages.

B. Approximation of discord probabilities

We proposed in ref. [1, Eq. 10] a formula to approximate quickly and efficiently the discord probabilities ρ_{ij} on the sole base of x_i and x_j . The approximation becomes exact under certain conditions of independence between the opinions of i and j . To assess the approximation error, we compare in Fig. 3 its results with those obtained from the exact formula (Eq. 4 in the main text). We disregard user pairs with independent opinions and for which the approximation is exact. We find that the approximation tends to over-estimate, and rarely under-estimate, the discord probabilities. This is not too surprising, given that factors of dependencies between opinions tend to bring those toward the same values. For example, if two users are connected by heavily weighted edges, their opinions will naturally be much more than could be assumed from x_i and x_j taken separately. The average error lies between 0.003 (Retweet network) and 0.025 (Mention network), with maximum errors ranging from 0.447 (unweighted retweet network) to 0.668 (Mention network). Therefore, while the approximation fares well for most users, if one is interested in the discord between two users which highly dependent opinions, it becomes crucial to use the exact formula.

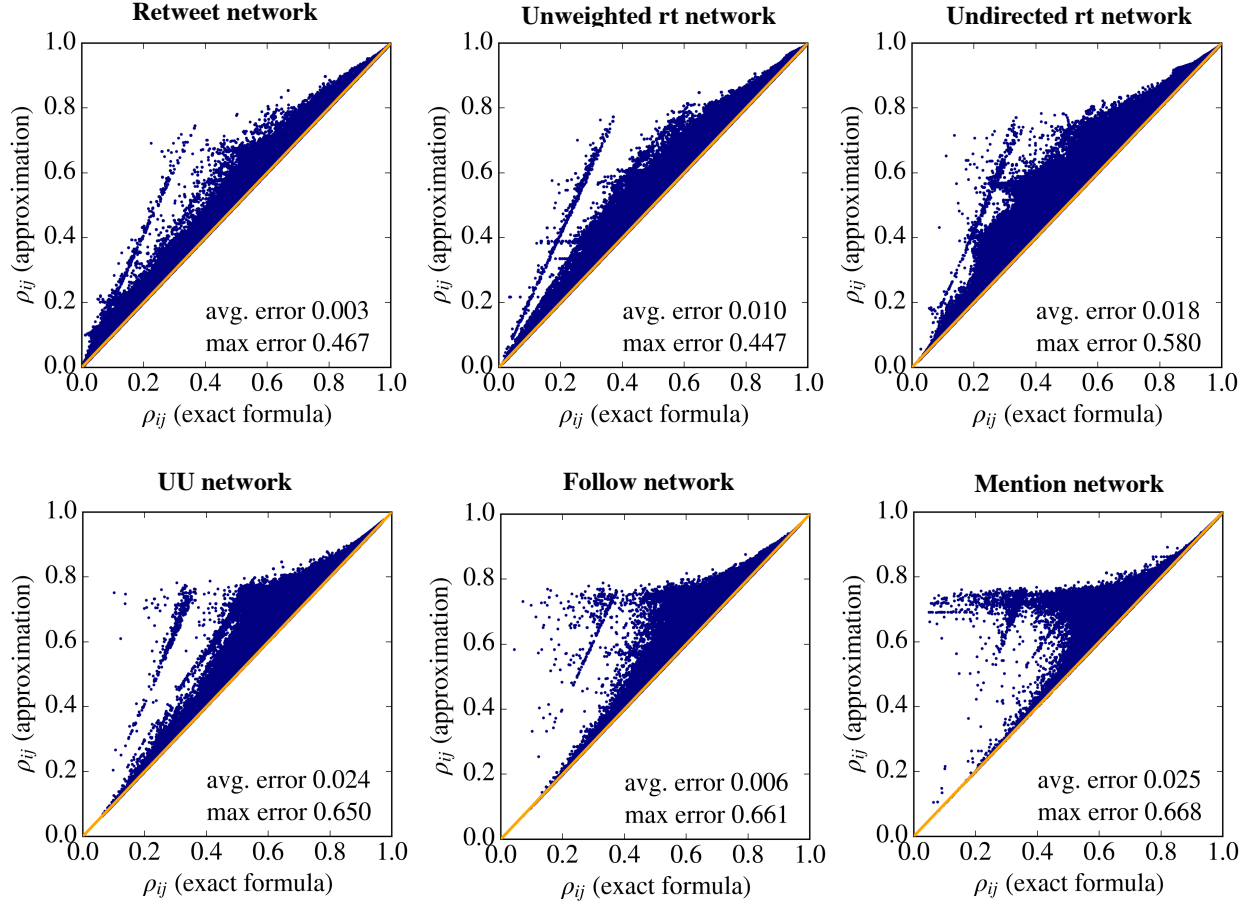


FIG. 3: Comparison between ρ_{ij} values computed using the exact formula (Eq. 4 in the main text, x-axis) and the approximation proposed in [1, Eq. 10] (y-axis) for each network. The diagonal orange lines indicate the identity.

C. Figures for additional networks

We show the equivalent of Fig. 1 of the main text for the other networks: Unw. (Figs. 4), Und. (Fig. 5), UU (Fig. 6), Follow (Fig. 7), Mention (Fig. 8).

[1] A. Vendeville, S. Zhou, and B. Guedj, *Phys. Rev. E* **109**, 024312 (2024).

[2] M. E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione, *Trans. Econ. Comput.* **1**, 10.1145/2538508 (2013).

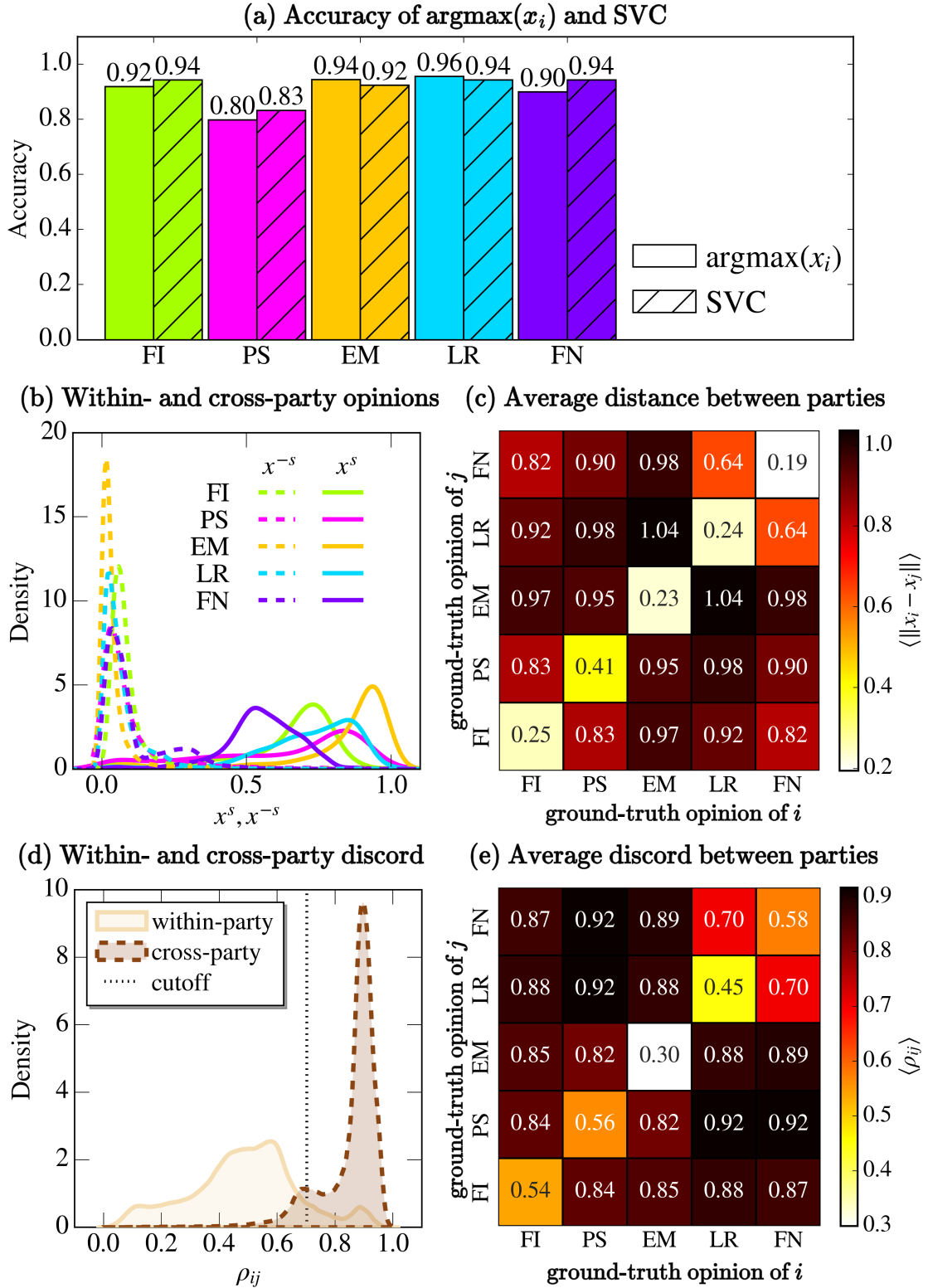


FIG. 4: Correspondence between theoretical opinion distribution x_i and ground-truth opinions y_i for the unweighted retweet network.

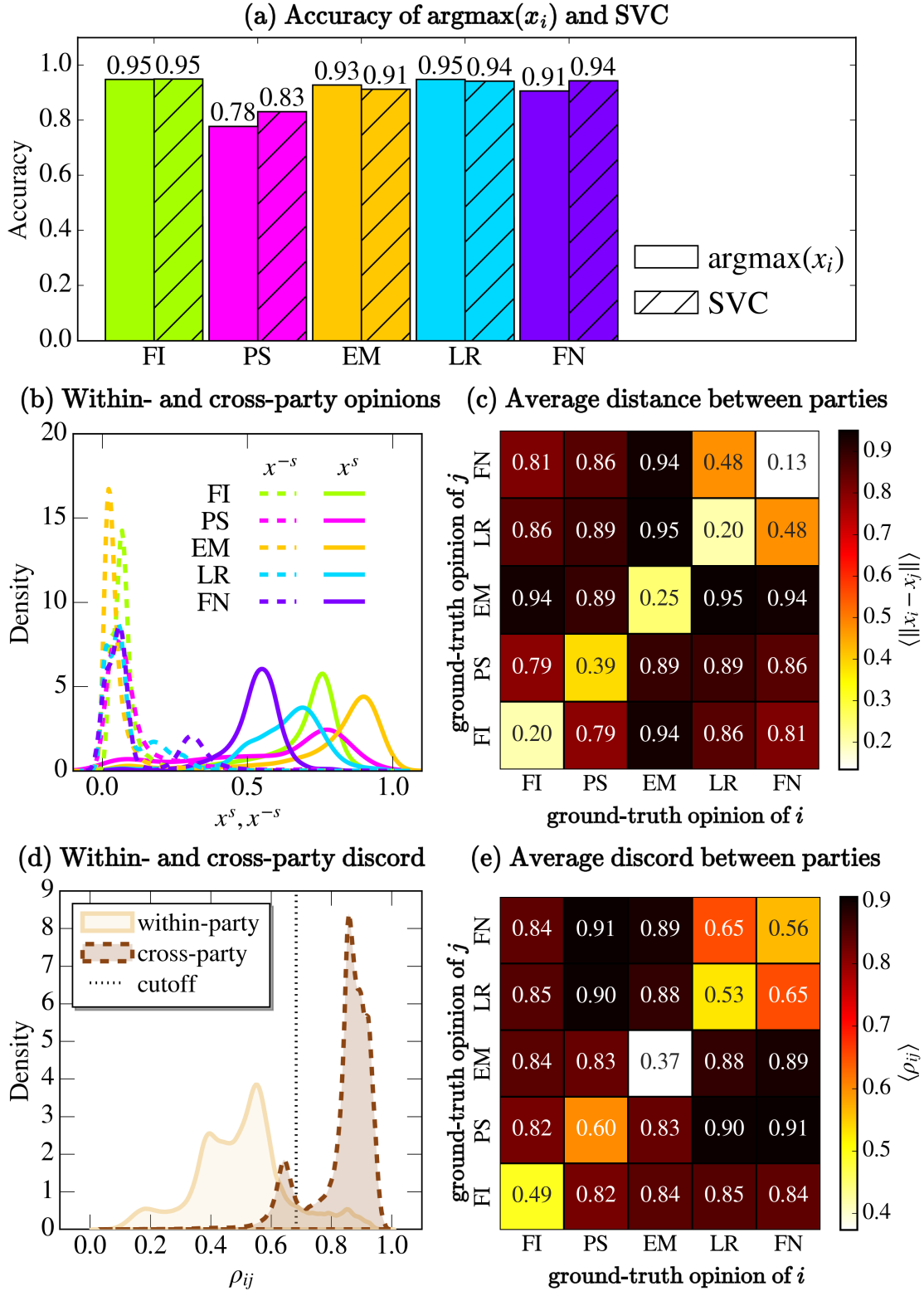


FIG. 5: Correspondence between theoretical opinion distribution x_i and ground-truth opinions y_i for the undirected retweet network.

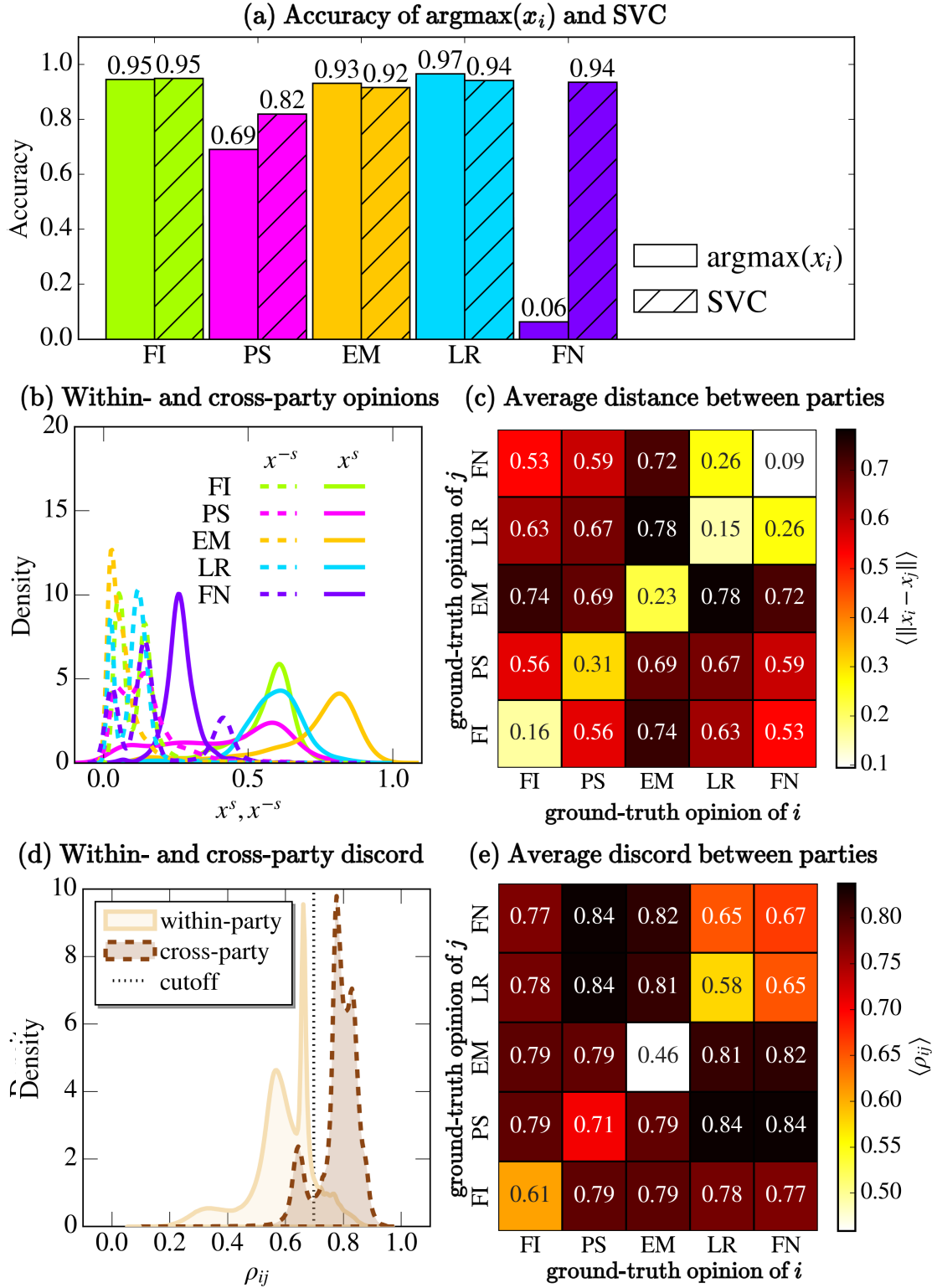


FIG. 6: Correspondence between theoretical opinion distribution x_i and ground-truth opinions y_i for the undirected and unweighted (UU) retweet network.

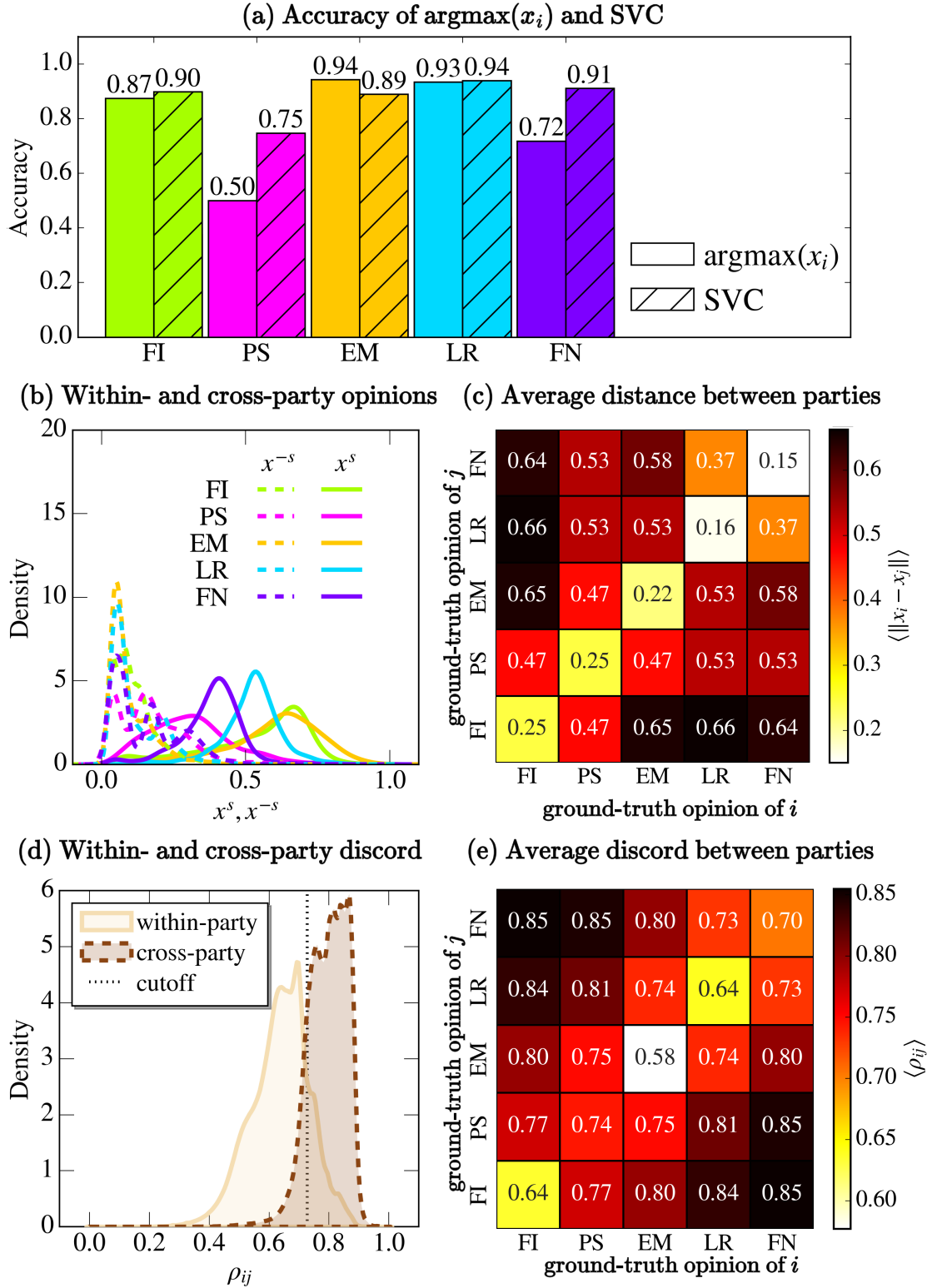


FIG. 7: Correspondence between theoretical opinion distribution x_i and ground-truth opinions y_i for the Follow network.

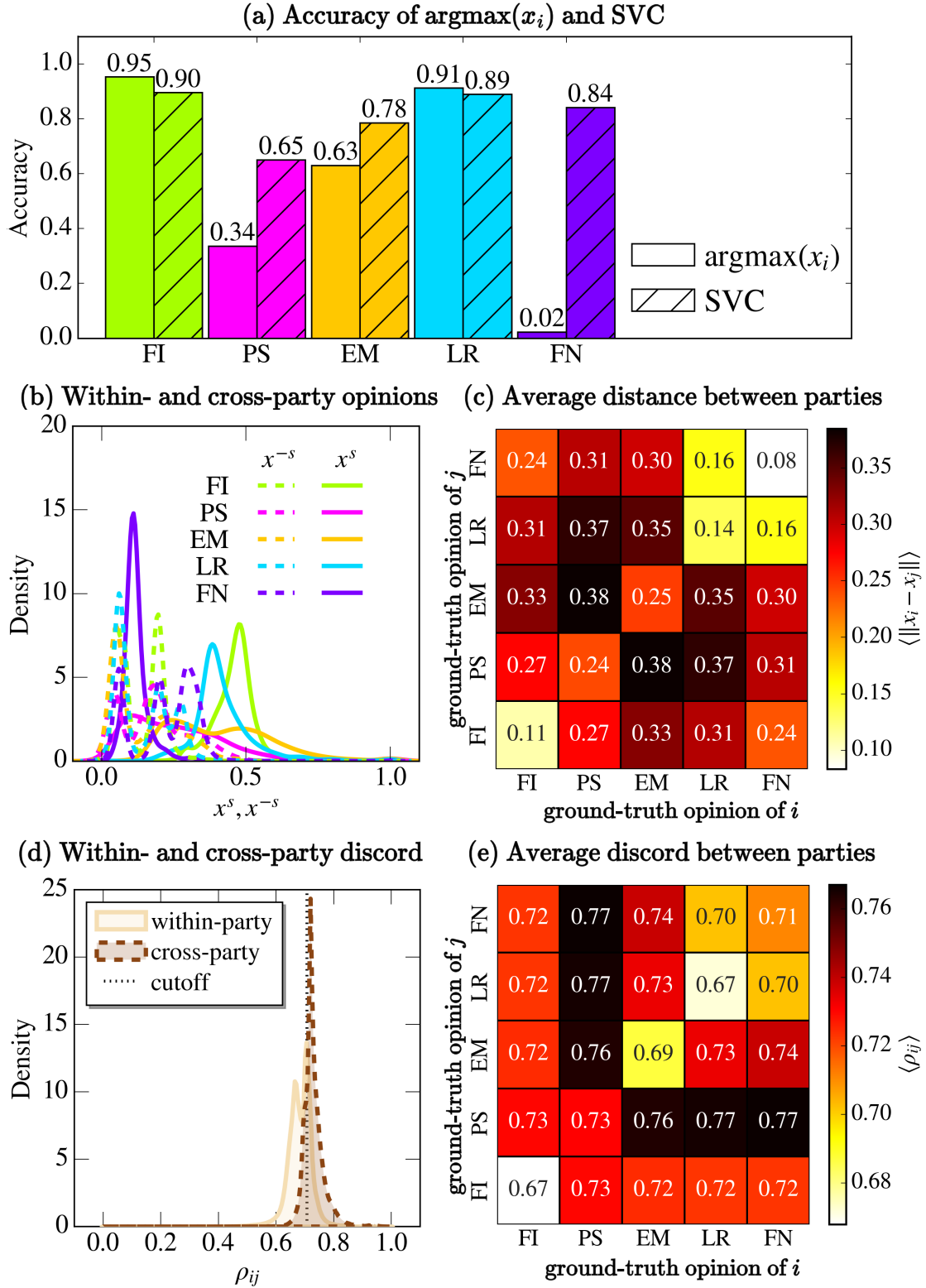


FIG. 8: Correspondence between theoretical opinion distribution x_i and ground-truth opinions y_i for the Mention network.