NETWORK SCIENCE

Inference and influence of network structure using snapshot social behavior without network data

Antonia Godoy-Lorite[1,2] and Nick S. Jones[1]*

Population behavior, like voting and vaccination, depends on the structure of social networks. This structure can differ depending on behavior type and is typically hidden. However, we do often have behavioral data, albeit only snapshots taken at one time point. We present a method jointly inferring a model for both network structure and human behavior using only snapshot population-level behavioral data. This exploits the simplicity of a few parameter model, geometric sociodemographic network model, and a spin-based model of behavior. We illustrate, for the European Union referendum and two London mayoral elections, how the model offers both prediction and the interpretation of the homophilic inclinations of the population. Beyond extracting behavior-specific network structure from behavioral datasets, our approach yields a framework linking inequalities and social preferences to behavioral outcomes. We illustrate potential network-sensitive policies: How changes to income inequality, social temperature, and homophilic preferences might have reduced polarization in a recent election.

INTRODUCTION

Human behavior, from voting preferences to vaccine sentiments, can depend on the structure of social networks (*1*). While we have huge, high-quality, social-scientific datasets linking the behavior of individuals to their individual circumstances [from censuses (*2*) through health surveys (*3*) to, in fine spatial aggregate, voting outcomes], it is extremely costly, or even impossible, to have direct access to the social networks on which this behavior is articulated. Here, we do not seek to infer individual links with high accuracy but instead seek a behavior-specific network model that is informative of social network structure and is policy relevant. The need to understand social network structure and how it shapes behavior appears acute: There are concerns about both the role of social networks in health from vaccine refusal to obesity (*4*–*6*), and the recurring notion that our societies are becoming excessively polarized (*7*). By accessing social structure and the behavioral dynamics it supports, we could also improve our perturbative understanding: clarifying how changes in social inequalities might change health and social polarization (*3*).

Given the need to characterize social network structure, it follows that there has been immense scientific excitement about data from large networking platforms from Twitter to mobile phones (*8*, *9*). It is, however, widely acknowledged that technology platform data have numerous practical issues. A leading concern is whether technology-dependent network datasets give a true indication of the social structure on which society-relevant behaviors, like smoking or voting, depend; it is likely, instead, that different behaviors are spread on different aspects of our social networks (*10*). Technology-dependent network datasets are commercially sensitive and so are hard to access and share, and are often available for limited time spans or spatial extents, and specific platforms themselves are unlikely to exist indefinitely: This creates concerns for reproducibility and generalizability. The most substantial issue, however, which must limit all such efforts in the future, is the immense privacy implication of large-scale social network data: Social network data are hard to anonymize

(*11*, *12*). An alternative route to using data from technology platforms is to use conventional surveys. Beyond issues with scalability, it is often a challenge to identify from surveys whether the inferred network structure relates to the true network on which a particular behavior is articulated (*13*–*15*). A third established route is to attempt to infer network models through, e.g., time-series data (*16*–*19*). These approaches typically assume repeated observations of individual-level data; unfortunately, human behavior, such as voting or smoking, is often sampled at a single point in time.

While it might thus seem challenging to access behavior-specific social structure, there is one distinctive feature of social data that assists inference: Unlike many networked systems, censuses provide socially relevant coordinate information for individual nodes. Peter Blau postulated an intuitive and powerful theory for social structure such that each individual in a society can be considered as being a point in a high-dimensional space (with dimensions like age, gender, and income) where the rates of connection between individuals are driven by homophily and depend on their relative separation in the space (*20*–*22*). This homophily suggests that we can consider individuals to form links conditional on their separation in social space. Typically, these networks are modeled by a soft random geometric graph (SRGG), wherein the probability of a link between two individuals decays with the distance between them (*23*) [in contrast to random geometric graphs (RGGs), where connections are deterministic: two nodes are connected if the distance between them is within a given threshold]. Beyond information about the coordinates of individuals, health and voting datasets give us snapshot information about the behavior of individuals. There is well-developed theory to capture discrete choices (*24*), which, in turn, has links to finite-temperature linear threshold models of influence (*25*, *26*) and Ising models (*27*–*29*). This paper exploits a merger between Blau's geometric view of social structure and Ising models of behavior to infer kernels for SRGG models of social structure; we name this a kernel-Blau-Ising (KBI) model.

Our model allows behavior to depend on both social circumstances and the behavior of neighbors in a network. We illustrate this conceptualization in Fig. 1. It takes as input a set of individuals in a social space and invokes both an Ising model for behavior and a simple (SRGG) model for how distances in a social space affect the

[1]EPSRC Centre for the Mathematics of Precision Healthcare, Department of Mathematics, Imperial College London, London SW7 2AZ, UK. [2]Centre for Advanced Spatial Analysis, University College London, 90 Tottenham Court Road, London W1T 4TJ, UK.
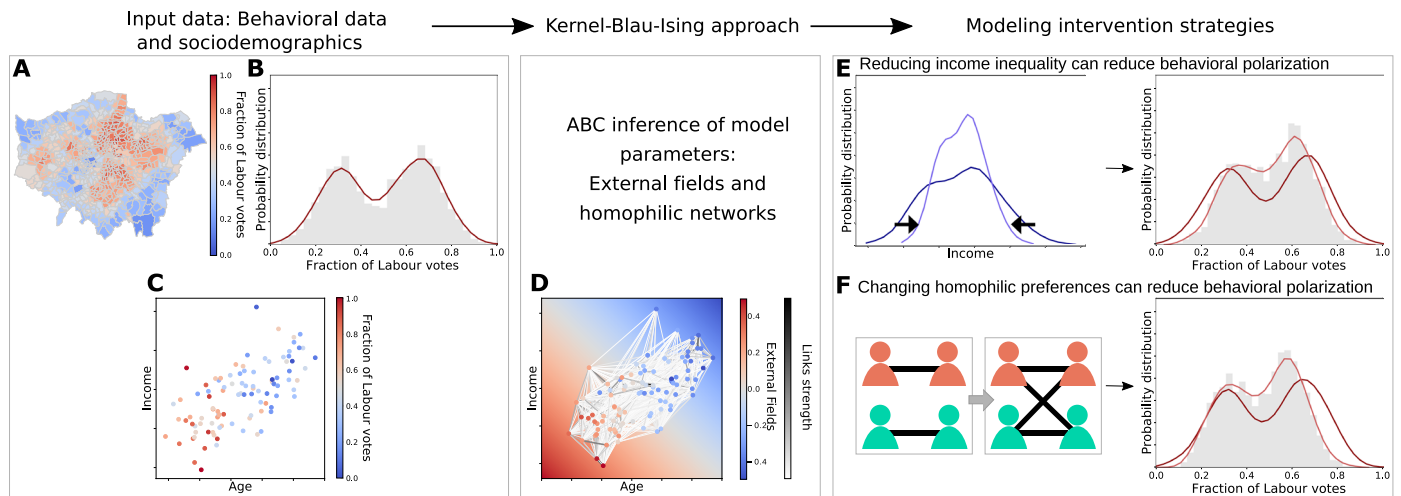*Corresponding author. Email: nick.jones@imperial.ac.uk

Fig. 1. Outline of the KBI methodology. Input data consist of aggregated behavioral data for different geographical areas and sociodemographic variables (age, income, education, etc.) associated to those areas (from census data). (**A**) Heatmap of (hypothetical) behavioral data in Greater London, in this case electoral outcomes, where red represents 100% votes to Labour and blue represents 100% votes to Conservatives. (**B**) Probability distribution of behavioral outcomes in (A). (**C**) Blau space representation of the behavioral outcomes spanned by sociodemographic characteristics (e.g., age and income). (**D**) Blau space representation of KBI approach using input data in (1) and learning parameters: the EFs, which account for the general trends, e.g., older people are more likely to vote Conservatives than younger people, and the network that connects the population according to their distances in the Blau space and their homophilic preferences. Once the model parameters are learnt, we can further estimate how changes and interventions affect behavioral outcomes. Examples of potential network-sensitive intervention strategies: how changes to income distribution (**E**) and homophilic preferences (**F**) can reduce behavioral polarization.

chance of connections. The simplicity of our model means we can use it to infer network parameter values for simulated data, which carries no network information but only a snapshot of system behavioral state and coordinate information. We illustrate our results for the European Union (EU) referendum and two London mayoral elections (MEs), using only census data and voting outcomes, where we infer network parameter values consistent with the literature and we are able to train on a subset of voting data and make accurate predictions on hold-out data. Last, our model allows us to quantify, model-dependently, the potentially depolarizing effects of shifts in social connectivity preferences (e.g., eliminating income or age homophily) and social coordinates (e.g., reducing income inequality).

MATERIALS AND METHODS

An interpretable generative model for both population behavior and social network structure in Blau space

We deploy a generative model for population behavior where the behavior of individuals is partly determined by their social coordinates (as would be standard in logistic regression from survey data, e.g., regressing vaccine refusal on age and income) and partly determined by the behavior of their neighbors on a network as would be standard in socio-physics models (28, 30). Regarding the network, we also deploy a generative model for social structure, where the chance that nodes have a connection depends on their proximity in social space. In our model, we represent individuals' binary social outcomes as binary spins (for example, voting Conservative/Labour or being smoker/nonsmoker), but it is possible to extend the model to a discrete set of possible outcomes by using a Potts instead of an Ising model (see text S2 for details). Throughout the manuscript, we use the terms individual social outcomes and spins interchangeably. We use an Ising-like model (or binary Markov random field) to

model population social outcomes, but instead of locating spins representing individual outcomes in a regular grid, in our approach spins will be embedded in a multidimensional Blau space (where dimensions are sociodemographic variables and geographical coordinates) and social links between individuals occur with a probability depending on their separation in the Blau space. While our methods are phrased in terms of individuals, social connections, and Blau space, they apply to wider settings with Ising dynamics on unobserved RGGs in the presence of spatially embedded applied fields, given information about vertex location and vertex spin state.

Network model

We have N individuals each embedded in a K-dimensional Blau space, where vector $\mathbf{z_i} \in \mathbb{R}^K$ encodes the ith individual's coordinates in the Blau space representing her age, income, residential coordinates, etc. (see Fig. 2A showing random coordinates of spins in a two-dimensional Blau space). The heterogeneity comes from population distribution in the Blau space, that is, it depends on social segregation in the Blau space. We connect individuals through an SRGG (23) according to a connectivity kernel function, which depends on distances in the Blau space and the kernel parameters (see Fig. 2B for an example of an SRGG). This model makes it easy to simulate realistic networks with clustering (31), although it does not explicitly build in other real social network properties such as heavy-tailed degree distributions. Nonetheless, it is a well-established model for generating social networks that provides interpretable results (31–33). The larger class of conditionally independent link models, e.g., graphons/stochastic block models, could be substituted for the SRGG.

We coded connections between individuals in an adjacency matrix, where $A_{ij} = 1$ if i and j are connected and 0 otherwise, with A_{ij} Bernoulli distributed with a connectivity kernel ρ

$$P(A_{ij} = 1 \mid \mathbf{z_i}, \mathbf{z_j}, \theta) = \rho(z_i, z_j, \theta) \qquad (1)$$
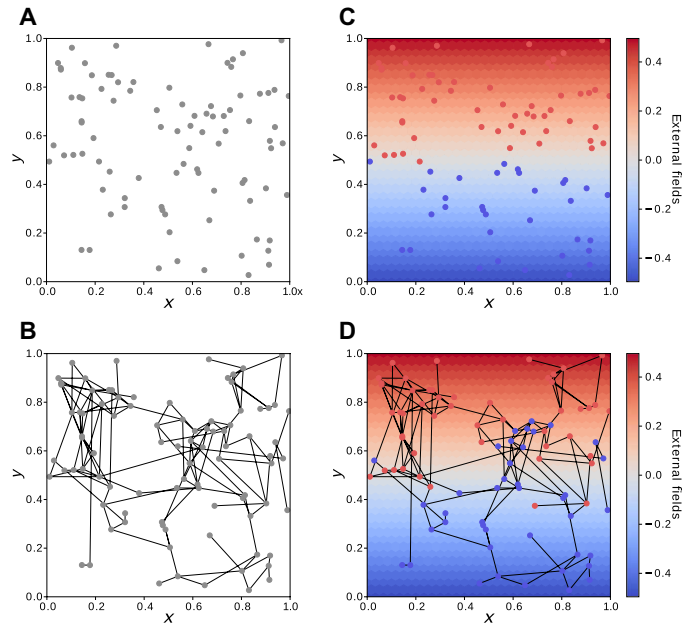
Fig. 2. Generative process for spin configurations balances social/spatial fields and network effects. (**A**) Coordinates of spins in a two-dimensional Blau space (e.g., x, y is age versus income). In (**B**), we show a realization of an SRGG from the connectivity kernel parameters $\theta_x = \theta_y = 10$ ($\theta_0 = 0$). (**C**) Spin configuration for a linear EF in the y axis and low thermal noise ($\beta = 100$), where the spins are aligned with the local EFs. Last, in (**D**), we show the spin configuration under the same EFs as in (C), also with low thermal noise ($\beta = 100$), but now the spins are connected according to the SRGG in (B). We see how spins that in (C) aligned with the EFs have now changed their orientation to align more with their neighbors.

We choose our connectivity kernel to be a logistic sigmoid function, as they have been successfully used for the inference of connectivity kernels on ego networks (*15*) and in latent space inference (*31–33*)

$$\rho(z_i, z_j, \theta) = \frac{1}{1 + \exp(d_{ij})}; d_{ij} = \theta_0 + \sum_{k=1}^{K} \theta_k |z_{ik} - z_{jk}| \qquad (2)$$

where d_{ij} is interpreted as the distance in the Blau space, θ_0 is a bias term that accounts for the overall connectivity density regardless of the distances in the Blau space, and θ_k is the connectivity coefficient of Blau dimension k, which weights the contribution of distances in the k dimension to the overall distance. The connectivity coefficients [θ_k, $k = (1, \ldots, K)$] measure homophily in the Blau space so that the larger θ_k becomes, the stronger the homophily in that dimension (connections becoming more localized in that dimension). The constant bias term θ_0 allows rescaling of system size, because it accounts for density changes without modifying values of the connectivity kernel parameters (see text S3). The connectivity kernels induce an interpretable semimetric for connections on the Blau space (*15*) and can be used to generate particular realizations of SRGGs.

Behavioral model
Our Ising model is to generate spin configurations as follows. Each individual i embedded in the Blau space (with coordinates z_i) is represented by a spin $\sigma_i = [-1,1]$ encoding her binary social outcome so that the population spin configuration is $\sigma \in [-1,1]^N$. The spin orientation depends on the external fields (EFs) and the other spins they are connected to in the network. As is common for conventional social statistics (logistic regression with linear dependence on the

covariates), we model the EFs as linear fields in each dimension of the Blau space, where the linear coefficient in each dimension of the Blau space k is h_k so that the individual spin interaction with the EFs is the scalar product $\boldsymbol{h} \cdot \mathbf{z_i} = \sum_k h_k z_{ik}$. The spins interact with the EFs depending only on their coordinates so that they tend to align with the EFs (see Fig. 2C).

The energy of a spin configuration σ is given by the Hamiltonian function

$$H(\sigma, \mathbf{h}, J, \mathbf{A}) = -\sum_i \left(\sum_k h_k z_{ik}\right)\sigma_i - J\sum_{ij} A_{ij}\sigma_i\sigma_j \qquad (3)$$

where h_k is the EF linear coefficient in dimension k of the Blau space, J is the scale factor between spin-spin interactions and energy known as the connection strength, and A_{ij} is the adjacency matrix. Notice that the connectivity kernel parameter θ governs the adjacency matrix $\mathbf{A}$ such that the only contribution of the kernel parameters to the Hamiltonian is through $\mathbf{A}$. Also, from the Hamiltonian, there is a coupling between the connection strength J and the bias term θ_0 in Eq. 2, which implies that, in some regimes, the same spin configurations can be generated by different combinations of these two factors; this is addressed below in the inference section. We can add a homogeneous field h_0, which is felt by the whole population regardless of their coordinates. However, in the cases we consider in the following, it is reasonable to set h_0 to zero. The configuration probability is given by the Boltzmann distribution with inverse temperature $\beta = 1/T$, $\beta \geq 0$

$$p(\sigma \,|\, \beta, \mathbf{h}, J, \mathbf{A}) = \frac{e^{-\beta H(\sigma, \mathbf{h}, J, \mathbf{A})}}{Z(\beta, \mathbf{h}, J, \mathbf{A})} \qquad (4)$$

and the normalization constant

$$Z(\beta, \mathbf{h}, J, \mathbf{A}) = \sum_{\sigma' \in \Omega} e^{-\beta H(\sigma', \mathbf{h}, J, \mathbf{A})} \qquad (5)$$

is the partition function, where the sum is over all possible spin configurations, Ω, which for an Ising model are 2^N terms. The configuration probabilities $p(\sigma \,|\, \beta, \mathbf{h}, J, \mathbf{A})$ represent the probability that, in equilibrium, the system is in a state with configuration σ. Figure 2D gives an example of a spin assignment that has been generated conditional on a particular network structure—in our case, an SRGG from a particular connectivity kernel. Vitally, the spins are not exclusively determined by either the EFs or network structure.

Inference method for model parameters
Given a record of a social outcome $\sigma \in [-1,1]^N$ together with the population Blau space coordinates $\mathbf{z} \in \mathbb{R}^{N \times K}$, our goal is to infer the model parameters $\Theta = [\beta, \mathbf{h}, J, \theta]$. For the Ising model, the partition function Z cannot be computed even for small systems, because it requires the computation of 2^N terms. A further challenge is that Z itself depends on the model parameters and would need to be recomputed for each possible parameter set. In this case, the inference is called doubly intractable (*34*) and Markov chain Monte Carlo is challenging.

As an alternative likelihood-free method (that nonetheless avoids mean-field approximation), we use approximate Bayesian computation (ABC), which has been applied to a wide spectrum of problems with intractable likelihoods (*35*, *36*). In Algorithm 1, we show our rejection-based ABC inference method for the model parameters. We suppose we have priors, $\pi(\Theta)$, on possible parameter values. On lines 4 and 5, the generation of a spin configuration requires two steps: (i) the generation of an SRGG from the connectivity kernel

---

Algorithm 1 ABC rejection algorithm

1: **for** i = 1 to N **do**

2:      **repeat**

3:          Draw independent proposal $\Theta' \sim \pi(\Theta)$

4:          Generate $\mathbf{A}' \sim \rho(\boldsymbol{z}, \boldsymbol{\theta}')$

5:          Generate $\boldsymbol{\sigma}' \sim p(\cdot | \mathbf{A}', \Theta')$

6:      **until** $\|\eta(\boldsymbol{\sigma}', \boldsymbol{\sigma})\| < \epsilon$

7:      Set $\Theta_i \leftarrow \Theta'$

(Eq. 2), given the spin coordinates, and (ii) the generation of the spin configuration conditional on the graph generated in (i) and for the parameters of the Hamiltonian in Eq. 3.

We use Glauber dynamics to generate spin configurations $\boldsymbol{\sigma}'$ for any combination of the model parameters Θ' from Boltzmann distribution. To improve the efficiency of the ABC rejection algorithm, we define a set of lower-dimensional summary statistics. We summarize spins that share the same coordinates z by the fraction of spins down (or up) $S_z = \frac{1}{n_z} \sum_{i|z_i=z} \delta(\sigma_i, -1)$, where n_z is the number of individual spins at coordinate z of Blau space. When there are C different Blau space coordinates populated with spins, the summary statistic would be such that $\mathbf{S}(\boldsymbol{\sigma}) \in [0,1]^C$ (also denoted as S for simplicity of notation). For the method that we proposed to succeed, we require $C \ll N$. Therefore, we approximate our posteriors by $p(\boldsymbol{\sigma} | \{\|\eta[\mathbf{S}(\boldsymbol{\sigma}'), \mathbf{S}(\boldsymbol{\sigma})]\| < \epsilon\})$, where $\eta[\mathbf{S}(\boldsymbol{\sigma}'), \mathbf{S}(\boldsymbol{\sigma})]$ measures the discrepancy between $\boldsymbol{\sigma}'$ and $\boldsymbol{\sigma}$ after they are summarized with function $\mathbf{S}(\boldsymbol{\sigma}')$. We define the distance between $\mathbf{S}(\boldsymbol{\sigma})$ (summary statistics of the original social outcome data) and $\mathbf{S}(\boldsymbol{\sigma}')$ (observational data of the spin configuration generated from Θ') as

$$\eta[\mathbf{S}(\boldsymbol{\sigma}'), \mathbf{S}(\boldsymbol{\sigma})] = \frac{1}{N}\sum_{z=1}^{C} n_z |S_z - S_z'| \qquad (6)$$

where $|S_z - S_z'|$ is the absolute difference between the fraction of spins down in the observed spin data and the generated spins so that the distance is zero only if $S_z = S_z'$ $\forall z$. The distance $\eta(\mathbf{S}', \mathbf{S})$ can also be considered as a weighted mean absolute error (WMAE). Notably, our observational data itself will be aggregated in the form of fractions of spin-up in different small spatial regions (e.g., proportions of smokers or voters in different small spatial patches). Because we are precisely attempting to simulate our observational data, and given that all spins with the same coordinates are statistically indistinguishable (Eq. 3), vitally, $\mathbf{S}(\boldsymbol{\sigma})$ are sufficient statistics and the ABC posteriors tend to exact Bayesian posteriors in the limit of $\epsilon \to 0$. If, e.g., continuous microdata are available for all individuals, some binning would be required to take advantage of our ABC inference, with a corresponding effect on sufficiency. For some applications (as in our voting illustration), population size can be rescaled through rescaling the connectivity kernel bias term θ_0 in Eq. 2 (see text S3). Last, because the ABC approximate posteriors of the model parameters are obtained through independent samples, it is straightforward to parallelize the algorithm so that the samples

are obtained concurrently. For a more efficient sampling procedure, sequential sampling schemes for ABC can also be used (37).

RESULTS

ABC inference can estimate model parameters from synthetic data

We tested the ability of our ABC inference method to recover the known parameters of both the network and behavioral processes for synthetic snapshot data. We note that, in a manner distinct from other inverse Ising or network inference approaches (38, 39), we are not seeking to recover unique network links, we do not observe each spin state (only coarsened observations), and we will not use time-series data (although it is possible to extend the model to repeated observations; see text S2). As is reasonable for social data, we suppose that we are given access to information about the social coordinates of individuals. Our simple, but justified, model structure allows us to extract information from very limited datasets composed of snapshot behavioral data and census information. Given that survey and census demographic variables are typically ordinal or categorical, we use ordinal data in our experiments with synthetic data. We performed the ABC rejection method for two different combinations of connection strengths at a temperature corresponding to ordered spin states (compared to the disordered spin states, when the system is in a "hotter" state showing no overall magnetization; see fig. S3). We found that for a given temperature β, there is a value of J_{aligned} where every pair of spins that are connected is aligned (J_{aligned} changes for the different β values). This phenomenon corresponds to a strong alignment regime, where the contribution of the interaction term to the Hamiltonian dominates the dynamics of the systems. Therefore, any $J > J_{\text{aligned}}$ has similar distributions over spin configurations, where connected spins are aligned. We test our inference method in two different synthetic data scenarios: one with strong connection strength $J_s > J_{\text{aligned}}$, and another with weak connection strength $J_w < J_{\text{aligned}}$.

In Fig. 3, we show the ABC posteriors for the different scenarios, with $J = 5$ (weak) and $J = 25$ (strong). We have set two model parameters in the inference: the x axis EF h_x ($h_x = 1$) and the connectivity bias term θ_0. Regarding h_x, from Eq. 4, we see that the inverse temperature parameter β multiplies the linear EFs $\mathbf{h}$; thus, there is one degree of freedom that we choose to reduce by setting $h_x = 1$ without
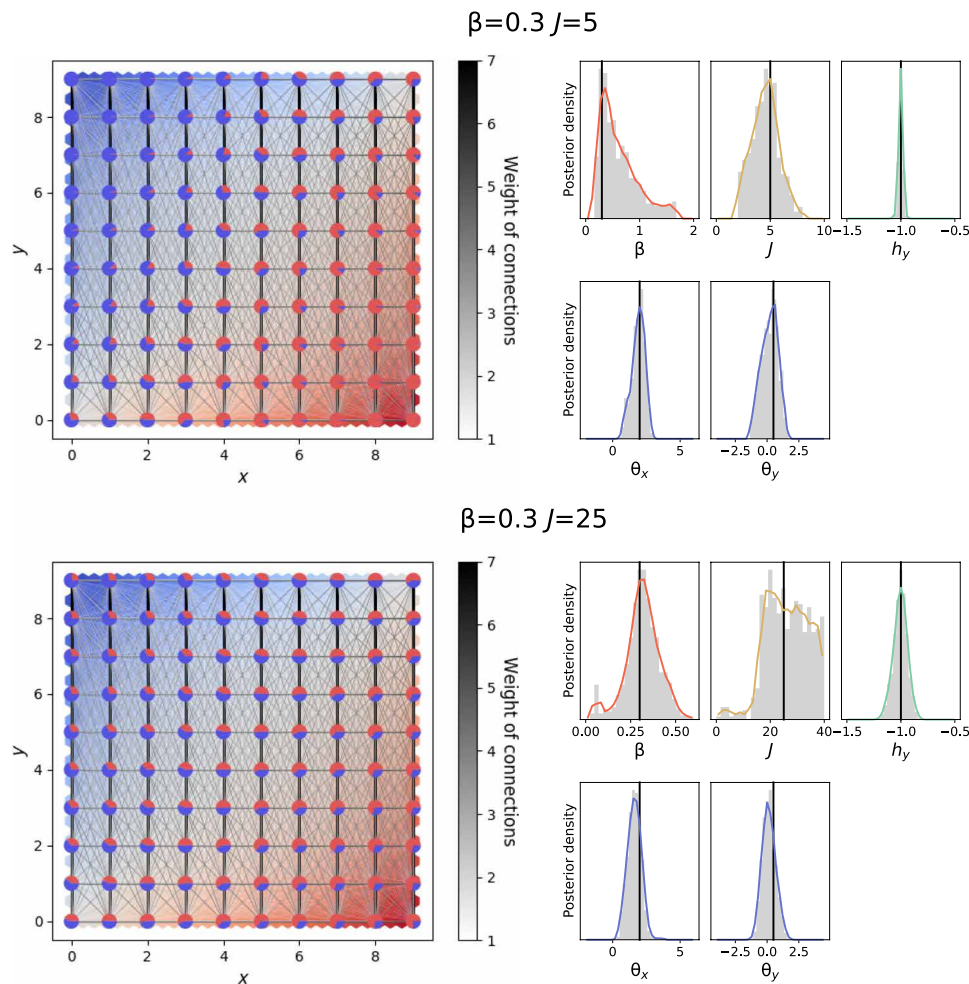
Fig. 3. Inference allows the recovery of model parameters for synthetic snapshot data. On the top, synthetic data with a weak connectivity strength spin configuration ($J = 5$), and on the bottom, a strong connectivity strength spin configuration ($J = 25$); both for $\beta = 0.3$ (see fig. S3) and kernel parameters $\theta_0 = 9$, $\theta_x = 2$, and $\theta_y = 0.5$. There are a total of $N = 10,000$ spins, with 100 spins on each of the discrete coordinates on the grid where $x, y = (0,1,\ldots,9)$ (for visualization purposes, links are aggregated at the coordinate level). To avoid coupling between certain model parameters, we choose to set $h_x = 1$ and $\theta_0 = 9$ for the ABC inference (see main text). We use uniform priors for $\beta \in [0,2]$, $h_y \in [-1.5,0.5]$, $\theta_x \in [-0.5,4.5]$, $\theta_y \in [-2.5,2.5]$, and $J \in [0,10]$ in the $J = 5$ scenario and $J \in [0,40]$ in the $J = 25$ scenario. We show the ABC marginal posterior distributions for $\approx$500 samples under tolerance thresholds of $\epsilon = 0.03$ and $\epsilon = 0.036$ for $J = 5$ and $J = 25$, respectively, from a large sampling procedure of over $6M$ i.i.d samples. The samples are visualized using histograms in gray and moving averages as solid lines, and vertical lines correspond to the real values used to generate the synthetic spin configuration. On the right, for $J = 25$, it shows that the ABC inference is not able to distinguish between configurations above given values of $J > J_{\text{aligned}}$ because all spins that are connected are already aligned. The ABC inference algorithm accurately estimates the connectivity kernel parameters without using network data for synthetic systems.

violating any constraint. Regarding the connectivity bias term θ_0 in Eq. 2, there is a coupling between the connection strength J and the bias term θ_0 in the Hamiltonian (Eq. 3). This implies that, in some regimes, the same spin configurations can be generated by different combinations of these two factors, i.e., a spin configuration could be a result of a certain combination of low connectivity density (large θ_0) and strong connection strength J and vice versa, high connectivity density (small θ_0) and weak connection strength—this trade-off does not hold when the link density is very low, with almost no connections, or when it is very high, such that it become almost a complete network (see fig. S2). Therefore, as a practical step, we set the connectivity bias term $\theta_0 = 9$ in the inference (such that the average degree is $\kappa \approx 2$, a parameter choice discussed in the next section). For the ABC inference in Fig. 3, we show the ABC marginal posteriors for $\approx$500 samples under tolerance thresholds of $\epsilon = 0.03$

and $\epsilon = 0.036$ for $J = 5$ and $J = 25$, respectively, from a large sampling procedure of over $6M$ i.i.d (independent and identically distributed random variables) samples. Results show that all inferred parameters are consistent with the original values for both weak and strong connection scenarios. As expected, the posterior of the connection strength for $J = 25$ accepts all values for $J > J_{\text{aligned}}$. Considering the connectivity kernel parameters, we show that we can recover posteriors consistent with θ_x and θ_y from one single observation of a spin configuration without using network data.

Inferred parameters for MEs and EU referendum are consonant with homophilic tendencies and voting preferences
To demonstrate our model and inference approach beyond synthetic data, we apply it to three electoral datasets in Greater London: 2012 London ME (2012 ME), 2016 ME (2016 ME), and EU (Brexit)

referendum in 2016. Our KBI model allows us to compare the different electoral outcomes using readily interpretable parameters, to visualize and compare the social connectivity structures, and to further estimate interventions.

The electoral results are given as aggregated outcomes, where we know the total outcome for a specific area but we do not know the votes of individuals. The smallest areas in spatial resolution that we can get for electoral outcomes are the electoral wards—630 electoral wards in Greater London from which, due to data mismatches, we can only use 608 ward outcomes; for the EU referendum, ward-level data are missing for 18 boroughs, so we use a combination of 280 ward-level outcomes and 18 borough electoral outcome data (see text S1). While census microdata are available at the ward level, we define the Blau space coordinates of each electoral ward as the average value from census data in each Blau space dimension, which are education, age, gender, wards centroid spatial coordinates, and income (see text S1 for details). This coarsening overestimates proximity within wards and could conceal heterogeneities among individuals within the same wards—favoring more homogeneous behavior. However, we note that although we lose information about individuals' connectivity by using aggregated data instead of individual's microdata coordinates, we see that average coordinates at the ward level differ appreciably from ward to ward (in fig. S4, we show evidence of heterogeneous distribution of average wards' coordinates by bootstrapping census microdata for each ward). Using a single Blau space coordinate for members of each ward induces a type of metapopulation structure (one population per ward), although fluctuations of links and spins within each ward are still possible. We finally note that census data are from 2011, and London is a fast-changing city; nonetheless, we consider this 1- to 5-year gap adequate for our illustrative aim.

We now define distances in the Blau space and how we can return reasonable posteriors for real data. We define distance between two wards in the Blau space (Eq. 2) as the absolute difference of their coordinate values for education, age, gender, and income dimensions, and as the distance between centroid coordinates for the spatial distance dimension (see text S1). Before passing the distances to the inference algorithm, we standardize them by subtracting the mean distance and then dividing by twice their standard deviation for each Blau dimension (40). The standardization allows us to compare homophily kernel parameters among them and makes interpretation easier. Last, for the ABC rejection algorithm, we simulate a representative sample of spins in each electoral ward instead of simulating the whole population. Specifically, we keep the relative size of the ward population according to the census data (from a population of $N \sim 8{,}800{,}000$ individuals in Greater London, we rescale the system to $N = 60{,}683$ with an average of 100 spins per ward). The population rescaling affects the bias term θ_0 in Eq. 2, but for sufficiently large samples, it can provide an adequate approximation for the homophily kernel parameters (see text S3). Given the coupling between the connection strength J and the connectivity bias term θ_0, we set $\theta_0 = 14$, which corresponds to a network with an average degree of $\kappa \approx 2$ comparable with real data on the estimated number of ego-confidants (41); see also fig. S2 where we show evidence of insensitivity of dimension-specific kernel parameters to changes in the degree of the network. As per the synthetic data, we fixed one of the EFs $h_{\text{education}}$ to remove the degree of freedom in the Hamiltonian (Eq. 4) between the inverse temperature and the EFs. We choose to set $h_{\text{education}} = 0.45$ based on the results for a simple multilinear

regression (see text S6), where we find that the education linear coefficient has the least variability when comparing values among the three elections. Last, we used uniform priors for all model parameters, because we did not have any other information on them and they are all of similar order of magnitude. We take a three-step approach (see text S7 for details). First, we use preexisting values and fits to the model to establish an approximate value of the parameters: multiple linear regression for the EFs (see text S6), and for the connectivity kernel parameters, we use as reference prior estimates from Hoffmann's homophilous kernels (15). Then, in our second and third steps, from these preexisting values, we define very wide priors; while using the range of accepted values with a very permissive tolerance ϵ, we yield more specific ranges from which we sample intensely (see caption Fig. 4). There was only one scenario where we found that our posterior had significant mass at the boundary of the prior (this can be an indicator of having priors that are too narrow), and this was for the connectivity strength parameter J. However, in this case, there were clear physical reasons why we could expect to see posteriors with mass at the boundary as we see below. In Fig. 4, we show the ABC marginal posteriors for ≈ 400 samples for each election from a sampling procedure of over $150M$ i.i.d samples with tolerance thresholds being $\epsilon = 0.088$ for 2012 ME, $\epsilon = 0.086$ for 2016 ME, and $\epsilon = 0.050$ for the EU referendum (see text S8 for details on how larger tolerance thresholds have no effect on the ABC marginal posteriors other than having very slightly wider posterior distributions).

Figure 4 shows that the marginal posteriors for the two London MEs are qualitatively similar (except possibly the level of noise, indicated by β), with the Conservative party winning the 2012 ME, while the Labour party won the 2016 ME (see fig. S9 for the detailed distribution of outcomes for the three elections). However, the EU referendum ABC marginal posteriors present some differences. In terms of the parameters of the model, the temperatures for the three elections are such that the spin states present magnetic order (see fig. S3). Regarding the connection strength J, for the three elections, the value of J is large enough that all spins connected are aligned and the inference cannot distinguish J values larger than J_{aligned} as in the strong connection scenario in Fig. 3 (our choice of θ_0 is partly determining our inferred J). For the EFs, the three elections show the same sign for the linear coefficients in the different dimensions, with the difference that EU referendum EFs are closer to zero than for the two MEs. Notice that $h_{\text{education}}$ is set to 0.45 so that more educated people tend to vote Labour/remain (see text S6). The observed EFs are in agreement with traditional two-party partisan voting sociodemographic tendencies in the United Kingdom: h_{age} is negative, meaning that older voters prefer Conservative/leave; h_{gender} is positive, meaning that men vote more Conservative/leave than women; and h_{income} is negative, meaning that higher income voters prefer Conservative rather than Labour, although for the EU referendum income EF is peaked around zero.

We do not seek an identity claim between our inferred model (or samples from it) and the true social network: Although microscopically interpretable, our model is clearly a substantial simplification; nonetheless, it is meaningful to consider the parameters of the network model we infer and the homophilic preferences they imply. The inferred connectivity kernels are nonnegative for the three electoral datasets, which is in agreement with the homophilous tendency of social relations (21, 42, 43). The connectivity kernel parameters for the two MEs are similar but differ from EU referendum kernel
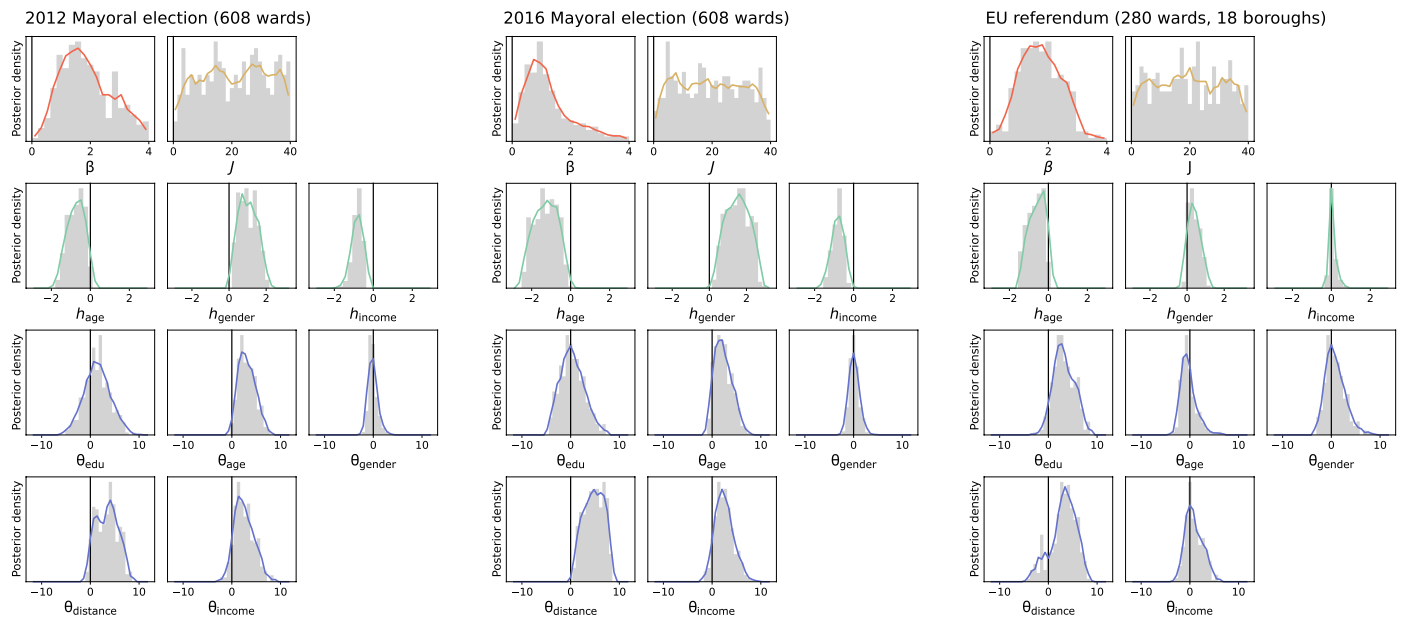
Fig. 4. ABC marginal posteriors for London MEs 2012 and 2016 and EU (Brexit) referendum are consistent with known homophilic and political preferences. We show ABC marginal posteriors for London MEs 2012 and 2016 for 608 electoral wards (see ABC marginal posteriors for MEs 2012 and 2016 for 280 electoral wards and 18 boroughs in fig. S6). We show estimates for parameters: inverse temperature β in red; connectivity strength J in yellow; EF for age, gender, and income, h_i; i = (age, gender, income); and connectivity kernel parameters θ_i; i = (education, age, gender, distance, income). $h_{education}$ and θ_0 as fixed (see main text). We use uniform priors for $\beta \in [0,4]$, $J \in [0,40]$, $h_{age} \in [-3,0.15]$, $h_{gender} \in [-0.3,3]$, $h_{income} \in [-1.6,0.7]$, $\theta_{education} \in [-7,12]$, $\theta_{age} \in [-5,12]$, $\theta_{gender} \in [-6,12]$, $\theta_{distance} \in [-7,11]$, and $\theta_{income} \in [-5,12]$. The ABC marginal posterior distributions are shown for ≈400 samples under tolerance thresholds of ϵ = 0.088 for 2012 ME, ϵ = 0.086 for 2016 ME, and ϵ = 0.050 for EU referendum, from a large sampling procedure of over 150M i.i.d samples (see text S8). We show the histograms of the ABC marginal posteriors in gray, as a solid line the moving averages of the histograms, and the vertical lines corresponding to 0 value. The parameter estimates show that the connectivity kernel parameters of the two MEs are similar but considerably different from EU referendum connectivity kernel. Unlike the MEs, the EU referendum does not show homophilic signal for age, gender, and income but only for distance and education, although the EU referendum took place a month after ME 2016.

parameters. Regarding the similarities, distance homophily is persistently the strongest homophily signal in the three electoral outcomes, consistent with other work (*15*, *22*) (a more detailed comparison can be found in text S10). Apart from spatial distance homophily, MEs' kernels show positive signal (homophily) for age and income dimensions, while for the EU referendum age and income ABC marginal posteriors are peaked around zero, but the education kernel parameter shows a positive signal that is not present in the MEs' kernels. Notably, the 2016 ME was held on 5 May 2016, only 49 days before the EU referendum on 23 June. Therefore, the changes in the social kernels we infer are not due to temporary evolution of the social connectivity structure but rather indicate that different social ties were at play for the EU referendum compared to MEs.

The differences in the connectivity kernels between MEs and the EU referendum indicate that different aspects of network structure were at play in the MEs versus at the EU referendum (see fig. S7 for maps of KBI electoral outcomes and inferred connectivity networks of the three elections). We also computed the median distance for each election using average ward-ward connections generated from parameter sets in Fig. 4. We find that the median distance between notionally connected individuals is similar for the three elections—5.75 km for 2012 ME, 4.90 km for 2016 ME, and 6.31 km for EU referendum—where the median borough diameter is 6.22 km (median ward diameter is 1.38 km); this is consonant with spatial homophily observed by others (*15*, *21*). Therefore, most of the social connectivities we infer are inside the borough and to first neighbors at the borough level.

The KBI model accurately predicts unobserved voting outcomes

The distance measure between generated and real electoral outcomes $\eta(\mathbf{S}', \mathbf{S})$ in Eq. 6 for the three electoral datasets gives us a measure of how accurately we can reproduce electoral outcomes. In the previous section, we showed that the KBI is able to accurately reproduce the electoral outcomes for the three elections, in particular with a weighted mean absolute percent error (WMAPE) below 8.8% for MEs and below 5% for the EU referendum. In addition to these general results, we measure the predictive power of our approach.

Motivated by the fact that we have missing ward data for the EU referendum, we define as training data the 280 ward outcome data available for the three elections and as the test data for the MEs as the 328 ward outcomes that are missing for the EU referendum, which correspond to 53.94% of overall wards (the training data are thus smaller than the test data). Because the EU referendum outcomes are not available at the ward level for the test wards, we thus evaluate the prediction on the corresponding 18 borough outcomes. Instead of using just the single best parameterization, we perform the predictions with the T = 100 parameter sets with lowest distance in the training data such that our results are not excessively sensitive to a single choice, and evaluate the prediction performance in terms of the average (over 100) distance on the test data $\langle \eta(\mathbf{S}'_{\text{test}}, \mathbf{S}_{\text{test}}) \rangle = \frac{1}{T}\sum_t^T \frac{\sum_{c \in \text{test}} n_c |S'^t_c - S_c|}{\sum_{c \in \text{test}} n_c}$, where S_c and S'_c are the real and predicted outcomes of c ward/borough, and n_c is the number of spins in c. We

compare the predictions of our KBI model ($J > 0$) with the predictions of a model with only EFs and without social connectivity kernel ($J = 0$, only EFs) (see results of ABC marginal posteriors of the model parameters with $J = 0$ for the three elections in figs. S8 and S9). The top 100 accepted simulations had an error of $\eta(\mathbf{S}', \mathbf{S}) < 0.078$ ($J > 0$) versus $\eta(\mathbf{S}', \mathbf{S}) < 0.083$ ($J = 0$) in the MEs and $\eta(\mathbf{S}', \mathbf{S}) < 0.062$ ($J > 0$) versus $\eta(\mathbf{S}', \mathbf{S}) < 0.067$ ($J = 0$) in the EU referendum; we thus find support favoring the richer KBI model for both types of votes. We also compared the predictions with a baseline null model that assigns the same prediction for all outcomes in the test, which is the weighted average outcome on the training data $S_{\text{pred}} = \frac{\sum_{c \in \text{training}} n_c S_c}{\sum_{c \in \text{training}} n_c}$.

Results in Fig. 5 show that the KBI model outperforms predictions with only EFs in all the three training tests. In particular, we observe larger improvements for MEs than for the EU referendum, where the relative advantage when considering the connectivity kernel is smaller. We believe that this is due to differences in the ward polarization of the electoral outcomes, where we define polarization as the tendency of a distribution of ward-level vote shares to have separated peaks (44). In fig. S10, we show that MEs have larger ward outcome polarization than the EU referendum. According to our experiments also in fig. S10, for the same model parameters by removing the connectivity kernel, the ME outcome distribution markedly increases their polarization, while for EU referendum outcome distribution remains significantly less affected. Therefore, this suggests that for Brexit, the kernel adds less new information above the EFs compared to the MEs, as we would expect with outcomes that do not have clearly defined niches in the Blau space (14). Overall, we demonstrate that the model can successfully predict missing electoral data.

Reducing homophilic tendencies and social inequality can reduce polarization in silico
The interpretability of our model parameters allows us to explore different notional scenarios for public intervention while illuminating the role that the different elements of the model play in producing social outcomes. In this section, we illustrate how possible interventions might reduce electoral polarization. This is, of course, an in silico exercise, which markedly supposes the microscopic and causal relevance of the model we fit. We define electoral ward outcome polarization as the tendency of the distribution of ward outcomes S_{ward} (as the fraction of votes to one of the two options) to be peaked toward the extremes $S_{\text{ward}} = 0$ and $S_{\text{ward}} = 1$ of the distribution as $P(\mathbf{S}) = \frac{1}{N_L} \sum_{(i,j) \in L} |S_i - S_j|$, where S_i is the outcomes of ward i, and L and N_L are all pairs and the number of pairs of electoral wards, respectively. In Fig. 6, we use the ward outcome distribution only for 2016 London ME for simplicity, although the observed effects in Fig. 6 should be qualitatively similar for 2012 ME and the EU referendum elections, given that they also present homophilous kernels and similar thermal noise and connection strengths. As a starting point, in Fig. 6 gray histograms showing the observed data, we see that far from a Gaussian distribution, the ward outcome distribution is polarized, with a group of electoral wards voting by majority Labour, then a decrease in the number of wards voting 50/50, and again a higher number of wards voting by majority Conservatives (see fig. S10 for the probability distributions of S_{ward} for the three elections and their polarization). In what follows, we show that social outcome polarization can be reduced in silico following very different strategies: by reducing inequality (changing the Blau space coordinates of individuals) (Fig. 6A), by reducing homophily (e.g., changing the connectivity kernel) (Fig. 6, B and G) ,by increasing thermal noise (Fig. 6E), by shifting the EFs (Fig. 6F), or by shifting single/multiple ward outcomes (directly altering the configuration of the spins) (Fig. 6H).

Figure 6A shows how redistribution of income—changes in the coordinates of wards in the Blau space—for those wards with lower and higher income reduces polarization. Selected wards' income are set to their weighted average so that the total income is conserved. Only redistribution of income of the 30 wards with the lowest income and the 30 wards with the highest income in Greater London (corresponding to the 5% lowest and 5% highest income wards) results in a marked drop in polarization. In silico results indicate that even small efforts reducing inequality can have a big impact in terms of social outcomes. In Fig. 6B, we eliminate homophily in the
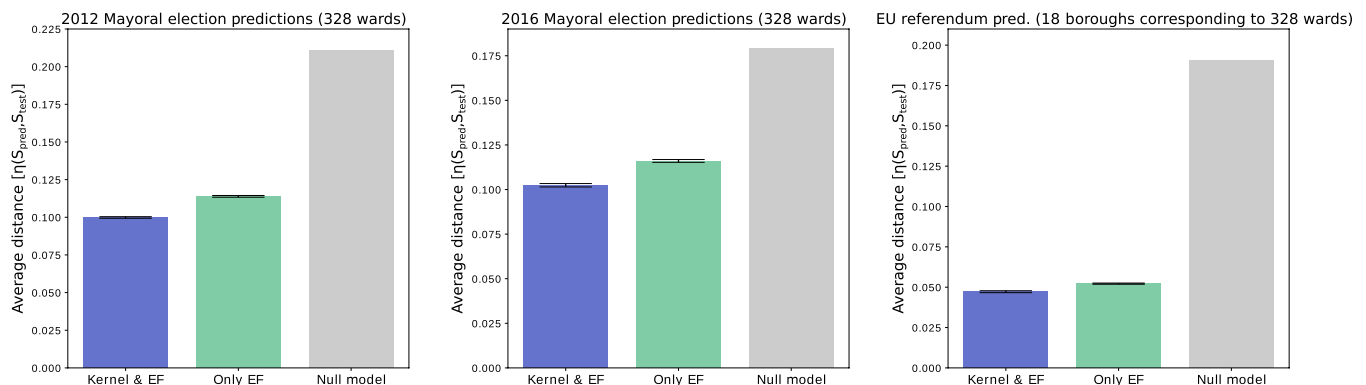
Fig. 5. KBI models, which include the connectivity kernel, outperform those using only EFs in the three elections. We measure prediction on 2012 and 2016 MEs and EU referendum, where we train the models with 280 observed wards' outcomes and make the predictions on 328 test wards' outcomes (corresponding to 18 boroughs in EU referendum; see text S1). Each bar represents the average distance [$\eta(\mathbf{S}'_{\text{pred}}, \mathbf{S}_{\text{test}})$] between predicted wards' (boroughs') outcomes and test wards' (boroughs') outcomes over 100 predictions from those parameters sets with the lowest distance in the training data. The error bars show the error of the mean. We show the predictions for the KBI in blue, for the model only using EFs (without social kernel) in green, and for a null model that predicts for all wards (boroughs) the weighted average computed from the training data in gray. KBI accurately predicts unobserved outcomes in the three elections, with the relative effect of social connections being lower in the EU referendum.
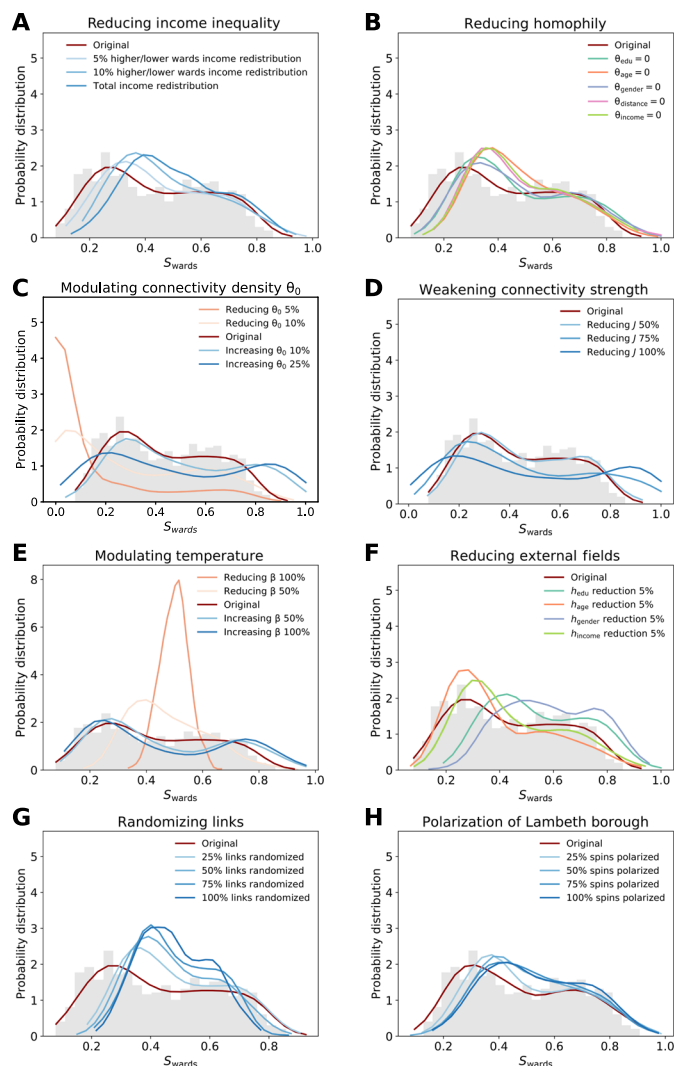
Fig. 6. Model polarization can be reduced by reducing inequalities and encouraging social mixing. (A to H) For simplicity, we only show results for the 2016 London ME. On the x axis, S_{ward} is the fraction of Conservative spins in the wards, being $S_{ward} = 0$ all votes for Labour and $S_{ward} = 1$ all votes for Conservatives. We show in gray the histogram of the real electoral outcomes in the 608 electoral wards and in dark red the original Gaussian kernel density of electoral outcomes from multivariate Gaussian of ABC marginal posteriors in Fig. 4. In solid lines, the different Gaussian kernel density estimates for the different interventions in the system are shown.

different Blau space dimensions—effectively eliminating social preference based on these features. While the EFs alone (without considering network effects) are strong predictors of voting outcomes (Fig. 5), we find that modulating homophily can also significantly shift outcomes in the model. As could be expected, the distributions are less polarized compared to the original, for age, distance, and income, which are the dimensions with stronger homophilous signal (original distribution polarization, in dark red, is of $P = 0.227$, and then removing homophily in the different dimension from less polarized to more polarized: removing homophily in age $P = 0.190$, income $P = 0.197$, distance $P = 0.203$, and then gender and education $P = 0.223$). In Fig. 6C, we also show the changes in the 2016 ME

outcomes when changing link density and keeping the homophilous kernel parameters. What we see is that a small decrease of the bias term θ_0 (5 to 10% reduction) results in a network where connectivity is high enough that almost all spins align toward Labour vote. On the other hand, when links' density decreased, the mixing effect of the network is reduced, resulting in a more polarized scenario. We observe a similar effect in Fig. 6D for decreasing link strength. Reducing the network effect results in a decrease of the social mixing of outcomes, giving rise to more polarized outcomes. Therefore, although the connectivity network is homophilous, and reducing homophily decreases polarization, the network also serves to mix population voting behavior, and thus, weakening the network increases polarization. Modulating thermal noise in Fig. 6E has a strong effect on the outcomes. Increasing thermal noise (reducing inverse temperature β) decreases polarization, and, the other way around, cooling down the system increases polarization. As might be anticipated from conventional social science models, we observe big changes when manipulating the linear EFs in Fig. 6F. The changes in the EFs shifted the outcome distribution toward more negative or more positive outcomes while reducing polarization. In Fig. 6G, we reduce homophily by randomizing links. The randomization of a homophilous network increases the mixing of outcomes and reduces polarization. Last, we can also change polarization by shifting opinions locally. In Fig. 6H, we fix the alignment of increasing fractions of spins in the borough of Lambeth, a traditionally Labour borough (72% vote to Labour) toward Conservative vote, and measure the effect on the other ward outcomes (Lambeth borough has 21 wards, corresponding to ~4% of all spins in the system). For the outcome distribution in Fig. 6H, we remove outcomes from Lambeth borough wards; thus, we only measure the changes on the remaining wards. The changes in the ward outcome distributions are only due to the fact that wards are connected. We see that polarization decreases and the outcomes are shifted toward more Conservative outcomes.

DISCUSSION

We presented the KBI model that allows us to accurately reproduce population-level behavior and known homophilic tendencies only using snapshot population behavioral data, without using network data. By effectively inferring the metric for behaviorally relevant social connections in the Blau space, we are able to improve not only our understanding of the behavioral process (and its links to inequalities and social preferences) but also predictability of unobserved social outcomes.

Our KBI model exploits the spatial character of homophilic preferences to reduce the number of parameters needed, thus allowing us to perform the inference from snapshot data and to avoid needing time-series data as in (17, 18, 38). Our efforts have much in common with other recent works that seek to merge the notion of behavior in Blau space with social networks through social network survey data (14, 15). In our case, we avoid using any social network survey data, instead recovering a network model tuned to each behavior type. The model is designed to avoid confounding of homophily and behavior (45) by making strong model assumptions about the form of the EFs and their interplay with social connections. Investigating its appropriate generalizations is for further work.

Our illustrative results on voting data in Greater London corroborate known homophilic tendencies for social ties as regards distance, age, income, and education (21, 42, 43). Spatial distance is consistently

the most homophilic dimension for the formation of social ties for the three votes, which is in agreement with previous studies (*15, 22*), and further, the spatial length scales we infer are consonant with spatial homophily observed by others (*14, 15*). We found large similarities between the inferred network structures for 2012 and 2016 London MEs but a different social structure for the EU (Brexit) referendum. This is consistent with the observation that traditional left-right politics do not help explain the Brexit referendum vote (*46–48*), perhaps suggesting that people discussed Brexit with a different type of person from the ones with whom they discussed the MEs (*10*). Regarding the apparent differences between votes, the EU referendum fit (unlike the MEs) shows educational homophily: the models that were selected suppressed inter-educational communication. Moreover, we do not infer the significant age or income homophily that we observe for both MEs. Education has been particularly highlighted as a significant explanatory variable (though treated as an EF) in other studies (*49, 50*). We also demonstrate that the model can successfully predict missing electoral data. The relative improvement when considering the connectivity kernel is significant for MEs but less relevant for the EU referendum. We believe that this is due to less clearly defined niches in the Blau space for Brexit compared to MEs (*14*).

Because the model and its parameters are interpretable, it enables us to explore in silico different intervention strategies in a novel framework. Beyond observing that an in silico reduction in income inequality reduces polarization, we also find unexpected strategies: In the model of the 2016 ME, we achieved a larger depolarizing effect by allowing more intergenerational connectivity—eliminating age homophily—than by eliminating income or distance homophily. This hypothesis is suggestive, because reducing age homophily might be less contentious than other social interventions. We found that, although the inferred network is homophilous, it nonetheless helps to mix population outcomes: Weakening the network effect—by decreasing connectivity density or connection strength—results in a more polarized scenario.

Given its simplicity and interpretability, we believe that the KBI model will be useful in a wide variety of behavioral/attitudinal processes, where social connections appear to have an effect, with special focus on health risk behaviors such as smoking, alcohol consumption, or vaccine refusal. We also believe that the KBI model is relevant because it makes explicit, in a simple and interpretable manner, how social inequalities (from income to educational inequities) interplay with our social preferences to shape social network structure and then how, in turn, our social network structure and social tendencies (EFs) shape our behaviors.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/23/eabb8762/DC1

REFERENCES AND NOTES
1. J. Zhang, D. Centola, Social networks and health: New developments in diffusion, online and offline. *Annu. Rev. Sociol.* **45**, 91–109 (2019).
2. Office for National Statistics; National Records of Scotland; Northern Ireland Statistics and Research Agency; UK Data Service, 2011 Census aggregate data (2017). DOI: 10.5257/census/aggregate-2011-2.
3. University of Essex, Institute for Social and Economic Research, *Understanding Society: Waves 1–10, 2009–2019 and Harmonised BHPS: Waves 1–18, 1991–2009* (UK Data Service, ed. 13, 2020). SN: 6614, DOI: 10.5255/ukda-sn-6614-14.
4. J.-P. Onnela, F. Reed-Tsochas, Spontaneous emergence of social influence in online systems. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18375–18380 (2010).
5. H. J. Larson, A. De Figueiredo, Z. Xiahong, W. S. Schulz, P. Verger, I. G. Johnston, A. R. Cook, N. S. Jones, The state of vaccine confidence 2016: Global insights through a 67-country survey. *EBioMedicine* **12**, 295–301 (2016).
6. N. A. Christakis, J. H. Fowler, The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).
7. R. Johnston, K. Jones, R. Sarker, C. Propper, S. Burgess, A. Bolster, Party support and the neighbourhood effect: Spatial polarisation of the British electorate, 1991–2001. *Polit. Geogr.* **23**, 367–402 (2004).
8. D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M. Van Alstyne, Computational social science. *Science* **323**, 721–723 (2009).
9. M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, D. Stillwell, Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *Am. Psychol.* **70**, 543–556 (2015).
10. S. Cowan, D. Baldassarri, "It could turn ugly": Selective disclosure of attitudes in political discussion networks. *Soc. Netw.* **52**, 1–17 (2018).
11. A. Narayanan, V. Shmatikov, De-anonymizing social networks, in *2009 30th IEEE Symposium on Security and Privacy* (IEEE, 2009).
12. Y.-A. de Montjoye, S. Gambs, V. Blondel, G. Canright, N. de Cordes, S. Deletaille, K. Engø-Monsen, M. Garcia-Herranz, J. Kendall, C. Kerry, G. Krings, E. Letouzé, M. Luengo-Oroz, N. Oliver, L. Rocher, A. Rutherford, Z. Smoreda, J. Steele, E. Wetter, A. Pentland, L. Bengtsson, On the privacy-conscientious use of mobile phone data. *Sci. Data* **5**, 180286 (2018).
13. C. McCarty, M. J. Lubbers, R. Vacca, J. L. Molina, *Conducting Personal Network Research: A Practical Guide* (Guilford Publications, 2019).
14. M. McPherson, J. A. Smith, Network effects in blau space: Imputing social context from survey data. *Socius* **5**, 2378023119868591, (2019).
15. T. Hoffmann, N. S. Jones, Inference of a universal social scale and segregation measures using social connectivity kernels. *J. R. Soc. Interface* **17**, 20200638 (2020).
16. M. Timme, J. Casadiego, Revealing networks from dynamics: An introduction. *J. Phys. A Math. Theor.* **47**, 343001 (2014).
17. T. P. Peixoto, Network reconstruction and community detection from dynamics. *Phys. Rev. Lett.* **123**, 128301 (2019).
18. M. T. Schaub, S. Segarra, J. N. Tsitsiklis, Blind identification of stochastic block models from dynamical observations. *SIAM J. Math. Data Sci.* **2**, 335–367 (2019).
19. T. Hoffmann, L. Peel, R. Lambiotte, N. S. Jones, Community detection in networks without observing edges. *Sci. Adv.* **6**, eaav1478 (2020).
20. P. M. Blau, A macrosociological theory of social structure. *Am. J. Sociol.* **83**, 26–54 (1977).
21. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
22. J. R. Hipp, A. J. Perrin, The simultaneous effect of social distance and physical distance on the formation of neighborhood ties. *City Community* **8**, 5–25 (2009).
23. M. D. Penrose, Connectivity of soft random geometric graphs. *Ann. Appl. Probab.* **26**, 986–1028 (2016).
24. D. McFadden, Economic choices. *Am. Econ. Rev.* **91**, 351–378 (2001).
25. C. Lynn, D. D. Lee, *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates Inc., 2016), pp. 2495–2503.
26. D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network. *Theory Comput.* **11**, 105–147 (2015).
27. W. A. Brock, S. N. Durlauf, Discrete choice with social interactions. *Rev. Econ. Stud.* **68**, 235–260 (2001).
28. I. Gallo, A. Barra, P. Contucci, Parameter evaluation of a simple mean-field model of social interaction. *Math. Mod. Meth. Appl. Sci.* **19**, 1427–1439 (2009).
29. S. Galam, S. Moscovici, Towards a theory of collective phenomena: Consensus and attitude changes in groups. *Eur. J. Soc. Psychol.* **21**, 49–74 (1991).
30. M. Fedele, Rescaled magnetization for critical bipartite mean-fields models. arXiv:1310.7615 [math-ph] (28 October 2013).
31. M. S. Handcock, A. E. Raftery, J. M. Tantrum, Model-based clustering for social networks. *J. R. Stat. Soc. A. Stat. Soc.* **170**, 301–354 (2007).
32. P. D. Hoff, A. E. Raftery, M. S. Handcock, Latent space approaches to social network analysis. *J. Am. Stat. Assoc.* **97**, 1090–1098 (2002).
33. P. D. Hoff, Multiplicative latent factor models for description and prediction of social networks. *Comput. Math. Organ. Theory* **15**, 261 (2009).
34. I. Murray, Z. Ghahramani, D. J. C. MacKay, in *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06 (AUAI Press, 2006), pp. 359–366.
35. J.-M. Marin, P. Pudlo, C. P. Robert, R. J. Ryder, Approximate Bayesian computational methods. *Stat. Comput.* **22**, 1167–1180 (2012).

36. M. A. Beaumont, W. Zhang, D. J. Balding, Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002).

37. R. Dutta, M. Schoengens, A. Ummadisingu, N. Widmer, J.-P. Onnela, A. Mira, ABCpy: A high-performance computing perspective to approximate bayesian computation. arXiv:1711.04694 [stat.CO] (13 November 2017).

38. M. Fedele, C. Vernia, P. Contucci, Inverse problem robustness for multi-species mean-field spin models. *J. Phys. A Math. Theor.* **46**, 065001 (2013).

39. M. Fedele, C. Vernia, Inverse problem for multispecies ferromagneticlike mean-field models in phase space with many states. *Phys. Rev. E* **96**, 042135 (2017).

40. A. Gelman, Scaling regression inputs by dividing by two standard deviations. *Stat. Med.* **27**, 2865–2873 (2008).

41. M. McPherson, L. Smith-Lovin, M. E. Brashears, Social isolation in America: Changes in core discussion networks over two decades. *Am. Sociol. Rev.* **71**, 353–375 (2006).

42. J. A. Smith, M. McPherson, L. Smith-Lovin, Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *Am. Sociol. Rev.* **79**, 432–456 (2014).

43. Y. Wang, H. Zang, M. Faloutsos, in *2013 Proceedings IEEE INFOCOM* (IEEE, 2013), pp. 3363–3368.

44. J.-M. Esteban, D. Ray, On the measurement of polarization. *Econometrica* **62**, 819–851 (1994).

45. C. R. Shalizi, A. C. Thomas, Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* **40**, 211–239 (2011).

46. K. Swales, *Understanding the Leave Vote* (NatCen Social Research, 2016), p. 7.

47. R. Johnston, D. Manley, C. Pattie, K. Jones, Geographies of Brexit and its aftermath: Voting in England at the 2016 referendum and the 2017 general election. *Space Polity* **22**, 162–187 (2018).

48. S. B. Hobolt, T. Leeper, J. Tilley, *Divided by the Vote: Affective Polarization in the Wake of Brexit* (American Political Science Association, 2018).

49. S. O. Becker, T. Fetzer, D. Novy, Who voted for Brexit? A comprehensive district-level analysis. *Econ. Policy* **32**, 601–650 (2017).

50. E. Alabrese, S. O. Becker, T. Fetzer, D. Novy, Who voted for Brexit? Individual and regional data combined. *Eur. J. Polit. Econ.* **56**, 132–150 (2019).

Citation: A. Godoy-Lorite, N. S. Jones, Inference and influence of network structure using snapshot social behavior without network data. *Sci. Adv.* **7**, eabb8762 (2021).

Science Advances

Inference and influence of network structure using snapshot social behavior without network data

Antonia Godoy-Lorite and Nick S. Jones

ARTICLE TOOLS	http://advances.sciencemag.org/content/7/23/eabb8762
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2021/05/28/7.23.eabb8762.DC1
REFERENCES	This article cites 38 articles, 4 of which you can access for free http://advances.sciencemag.org/content/7/23/eabb8762#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service