Is the Voter Model a model for voters?

Juan Fernández-Gracia, 1 Krzysztof Suchecki, 2 José J. Ramasco, 1 Maxi San Miguel, 1 and Víctor M. Eguíluz 1

1 IFISC (CSIC-UIB), 07122 Palma de Mallorca, Spain
2 Faculty of Physics and Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland
(Dated: March 12, 2014)

The voter model has been studied extensively as a paradigmatic opinion dynamics’ model. However, its ability for modeling real opinion dynamics has not been addressed. We introduce a noisy voter model (accounting for social influence) with agents’ recurrent mobility (as a proxy for social context), where the spatial and population diversity are taken as inputs to the model. We show that the dynamics can be described as a noisy diffusive process that contains the proper anisotropic coupling topology given by population and mobility heterogeneity. The model captures statistical features of the US presidential elections as the stationary vote-share fluctuations across counties, and the long-range spatial correlations that decay logarithmically with the distance. Furthermore, it recovers the behavior of these properties when the geographical space is coarse-grained at different scales from the county level through congressional districts and up to states. Finally, we analyze the role of the mobility range and the randomness in decision making which are consistent with the empirical observations.

PACS numbers: 89.65.-s 89.75.Fb 89.75.Hc 05.40.-a

Opinion dynamics focuses on the way different options compete in a population, giving raise to either consensus (every individual holding the same opinion or option) or coexistence of several opinions. Many theoretical efforts have been devoted to clarify the implications on the macroscopic outcome, among other aspects, of different interaction mechanisms, different topologies of the interaction networks, the inclusion of opinion leaders or of zealots, external fields [1, 2]. To advance our understanding on social phenomena these theoretical efforts need to be complemented with empirical [3–5] and experimental results [6–9]. In this context elections offer an opportunity for contrasting opinion dynamics’ models with empirical results [10]. On one hand, the data are publicly available in many countries, with a good level of spatial resolution and several temporal observations. On the other hand, there is evidence that voting behavior is strongly influenced by the social context of the individuals [6, 7, 9, 11–20]. Thus it is natural to model electoral processes as systems of interacting agents with the aim of explaining the statistical regularities [21–38], as for example, the universal scaling of the distribution of votes in proportional elections [21, 22] or signatures of irregularities in the democratic process [23, 26].

In this work we asses the capacity of the voter model to capture real voter choices and propose a microscopic foundation for modeling voting behavior in elections. The model is based on social influence and recurrent mobility (SIRM): social influence will be modeled as a noisy voter model, while recurrent mobility serves as a proxy of the social context. In the voter model each agent updates its state by randomly copying the opinion of one of its neighbors [39–41]. We will consider that agents interact at home and at work locations according to their commuting pattern [42, 43]. We first obtain the statistical features of the US presidential elections and then we introduce and analyze the model.

Statistical regularities in elections.- The analysis is focused on US presidential elections from 1980 to 2012 [44] due to the combination of data availability and an almost bipartisan system. The vote-share per county \(v\), that is, percentage of votes in a county, for any of the two main parties is distributed following approximately a Gaussian distribution (Fig. 1a), consistently with observations in other countries [23]. The average vote-share over all counties changes from election to election but the width remains approximately constant for each year, with a standard deviation of \(\sigma_v \simeq 0.11\). We also find that the spatial correlation of the vote-shares decays logarithmically with the geographical distance (Fig. 1b), as reported previously for turnout and winner party vote-shares [24, 25]. The spatial correlation function is com-

FIG. 1: (Color online) US electoral results 1980-2012. a) County vote-share probability density functions. b) Spatial vote-share correlations as a function of distance. The dashed lines indicate logarithmic decay.

PACS numbers: 89.65.-s 89.75.Fb 89.75.Hc 05.40.-a
puted as
\[ C(r) = \frac{\langle v_i v_j \rangle_d(r, r_i) - \langle v \rangle^2}{\sigma^2(v)}, \]
where \( \langle v \rangle \) is the average vote-share over all the cells, \( \sigma^2(v) \) its standard deviation, and \( \langle v_i v_j \rangle_d(r, r_i) = r \) is averaged over pairs of cells separated a distance \( r \). The stationarity of the vote-share dispersion and the logarithmic decay of the spatial correlations will be considered as generic of the fluctuations in electoral dynamics.

The model.- In the SIRM model \( N \) agents live in a spatial system divided in non-overlapping cells [45]. The agents are distributed among the different cells according to their residence cell. The number of residents in a particular cell \( i \) is \( N_i \). While many of these individuals may work at \( i \), some others will work at different cells. This defines the fluxes \( N_{ij} \) of residents of \( i \) recurrently moving to \( j \) for work. By consistency, \( N_i = \sum_j N_{ij} \). The working population at cell \( i \) is \( N_i' = \sum_j N_{ij} \) and the total population in the country is \( N = \sum_{ij} N_{ij} \).

We describe agents’ opinion by a binary variable (+1 or −1). The number of individuals holding opinion +1, living in county \( i \) and working at \( j \) is \( V_{ij} \), thus \( V_i = \sum_j V_{ij} \) is the number of voters living in \( i \) holding opinion +1; \( V_j' = \sum_i V_{ij} \) is the number of voters working at \( j \) holding opinion +1. We assume that each individual interacts with people living in her own location with a probability \( \alpha \), while with probability \( 1 - \alpha \) she does so with individuals of her work place. Once an individual interacts with others, its opinion is updated following a noisy voter model [15, 39, 40, 46, 47]: an interaction partner is chosen and the original agent copies her opinion imperfectly (with a certain probability of making mistakes). The evolution of the system can be expressed in terms of the transition rates:

\[
\begin{align*}
    r^-_{ij}(V) &= V_{ij} \left[ \frac{\alpha N_i - V_i}{N_i} + (1 - \alpha) \frac{N_j' - V_j'}{N_j'} \right] + N_{ij} D \frac{\eta_i}{2} \delta(t), \\
    r^+_{ij}(V) &= (N_{ij} - V_{ij}) \left[ \frac{\alpha V_i}{N_i} + (1 - \alpha) \frac{V_j'}{N_j'} \right] + N_{ij} D \frac{\eta_j}{2} \delta(t),
\end{align*}
\]

where \( \mathcal{V} = \{V_{ij}\} \) is the configuration of the system according to the set of variables \( V_{ij} \), and \( r^\pm_{ij}(V) \) are the rates of change of \( V_{ij} \) by one unit to \( V_{ij} \pm 1 \). These rates include recurrent mobility and so they are different from those obtained for random diffusion processes [48]. The variables \( \eta_{ij}^\pm(t) \) are noise terms accounting for imperfect imitation, modeled as Gaussian noise with zero mean and \( \langle \eta_{ij}^\pm(t) \eta_{ij}'(t') \rangle = \delta(t - t') \delta_{ab} \delta_{ik} \delta_{jl} \) [49]. At the leading order, which corresponds to taking into account only the external noise coming from imperfect imitation, while the internal noise coming from the finite number of voters is neglected, the set of stochastic differential equations for \( v_{ij} = V_{ij}/N_{ij} \) is

\[
\frac{dv_{ij}}{dt} = \alpha \sum_i A_{ij} v_{il} + (1 - \alpha) \sum_j B_{ij} v_{lj} + D \eta_j(t),
\]

with \( A_{ij} = \frac{N_i}{N_i'} - \delta_{ij} \) and \( B_{ij} = \frac{N_j'}{N_i} - \delta_{li} \) [44]. The first term on the right hand side describes interactions among agents who live in \( i \) and work elsewhere, while the second term follows from the interactions among agents who work in \( j \) and live elsewhere. The last term is noise coming from a combination of \( \eta_{ij}^\pm(t) \) and \( \eta_{ij}(t) \): \( \eta_{ij}(t) \) is also a Gaussian noise with zero mean and \( \langle \eta_{ij}(t) \eta_{ij}(t') \rangle = \delta(t - t') \delta_{ik} \delta_{jl} \). This term represents imperfect imitation and accounts for the combined effect of all other influences different from the interaction between peers. This includes opinion drift, local...
media or free will of the individuals. When $D \neq 0$ the microscopic rules lead to a noisy diffusive equation, in agreement with previous models of mesoscopic electoral dynamics [24, 25]. The equation corresponds to an Edwards-Wilkinson equation on a disordered medium, described by the coupling matrices $A$ and $B$. In the absence of imperfect imitation ($D = 0$), Eq. (3) can be written as a Laplacian $\frac{d}{dt} \vec{v} = \bar{L} \vec{v}$. This implies a homogeneous asymptotic configuration and the existence of a globally conserved variable, namely the total number of voters holding opinion +1, $V = \sum_{ij} V_{ij}$ [44, 50].

When simulating the model, we integrate the stochastic process by updating the values of the number of agents holding opinion +1 in each cell $ij$, $V_{ij}$, using binomial distributions with the rates in Eq. (2). At each Monte Carlo step we update all cells in a random order. Therefore we simulate the original master equation of the process.

**Model calibration.** We apply the model to the US presidential elections identifying the cells with the counties. The populations and commuting fluxes $N_{ij}$ are obtained from the 2000 census [51] and are input data for the SIRM model. This framework can be applied to any country, from the US, or territorial division (counties, municipalities, provinces, states, etc). Besides these data, there are two free parameters: $D$ and $\alpha$. The parameter $\alpha$ provides a measure of the relative intensity and duration of the social relations at work and at home. According to the survey on the use of the Bureau of Labor Statistics [52], the average individual spends daily almost 8 hours at work and the rest of time at her home location. Out of this home time, close to another 8 hours are spent sleeping. Thus $\alpha$ will be set at 1/2, although other values give rise to qualitatively similar results (Fig. S10 in [44]), as long as $\alpha \neq 0$ or 1 in which cases the system would consist of disconnected patches and thus there would not be any spatial diffusion.

To calibrate the noise intensity $D$, the SIRM model is run for a set of values of $D$ taking as initial condition the results for the elections of the year 2000. The system is evolved for 1000 Monte Carlo steps and then the standard deviation $\sigma$ of the vote-share distribution is measured (panel a) in Fig. 2). Best agreement is obtained for $D = 0.03$ which is taken as the level of noise for the simulations of the model. When the noise intensity is too low we find basically a diffusive process, where the vote share distribution narrows and the correlations grow ($D = 0.005$ in Fig. 2). In contrast, for larger $D$ the noise is dominating the results ($D = 0.35$ in Fig. 2). The vote share distribution widens as time goes by and the spatial correlations vanish. For $D = 0.03$ the standard deviation of the vote-share distribution of the model has the same value as the data. Not only the standard deviation is matched, but also the shape of the vote-share distribution agrees with the empirical one. The distribution, in addition, becomes stationary in time. Furthermore, although we did not take spatial correlations into account for the calibration, they show a stationary logarithmic decay for this value of noise intensity $D$.

Finally, we set the equivalence between the Monte Carlo (MC) steps and the real time between elections (Fig. 2b). Eq. (3) is written in arbitrary time units and is related to the updates by $dt = 1/N$[53]. Sets of electoral results are produced with the model with $D = 0.03$ and with a fixed number of Monte Carlo steps between elections. Then the standard deviation $\delta$ of the vote-share trajectory for each county as a function of a the number of consecutive elections is computed. Averaging over all different counties and comparing with empirical data, we find that both curves grow as $\sqrt{n}$, where $n$ is the number of elections considered (error bars correspond to the dispersion of $\delta$ across counties), reminiscent of a random walk. Both curves have the best overlap when we set
10 MCsteps/election (equivalently 2.5 MCsteps/year).

Results. - The stochasticity of the model introduces uncertainty in the temporal evolution of the vote-shares as can be appreciated for three counties in Fig. 3a. Once the average value is discounted, the shape of the distribution of vote-shares is similar to the one observed in the empirical data (Fig. 3b): the stationarity and the particular functional shape of the distributions are captured by the model. This occurs not only at county level (Fig. 3b) but also at other coarse-grained geographical scales such as congressional districts (Fig. 3c) and states (Fig. 3d). This relates to the ability of the model to properly capture the spatial correlations in the data (Fig. 3e) and Fig. S7 in [44] for a comparison with reshuffled data).

The goodness of the model is also assessed by a direct comparison between model predictions and data for vote-share fluctuations. In Fig. 3f, g, we show the distribution of the ratios between model and data of the vote-shares deviations from the national average, \( v_i - \langle v \rangle \), where \( \langle \cdot \rangle \) denotes spatial average (not average over realizations of the model). We evolve the model for an election, starting with the initial conditions from the electoral results from year 2000, and compare with the electoral results from year 2004, finding that 80% of the ratios fall between 0.6 and 1.5. These numbers become 0.9 and 1.1 at the congressional district level, attesting the quality of the model predictions.

As a final point we investigate the role played by the mobility network on the model results. The links connecting only geographically neighboring counties can be extracted and used as a baseline network. The rest of the links are then added filtering by the distance that separates the centroid of the residence county to that of the work county. The result of performing this operation is a network that includes more and more links as the threshold of the filter is increased. The model has to be calibrated for each new network (Fig. 4a). Once the optimal value for the noise level of the imperfect imitation \( D^* \) is found, the model simulations running on different networks can be compared with the empirical data. In Fig. 4b, we show how the vote-share spatial correlations change when the network is modified. Long links are important to recover correlation values similar to those observed empirically.

Discussion.- We have introduced a microscopic model for opinion dynamics whose main ingredients are social influence (modeled as a noisy voter model), mobility and population heterogeneity. The model can be approximated by a noisy diffusion equation on an anisotropic substrate that is given by the highly heterogeneous population and commuting data. It reproduces generic features of the vote-share fluctuations observed in data coming from three decades of presidential elections. It is important to note that the model is not aimed at predicting the winning party, only the local fluctuations over the national average vote-share. In this sense, it is able to capture the empirical distributions of vote-share fluctuations, the spatial correlations and even the evolution of the local vote-share fluctuations. This agreement between model predictions and empirical data are maintained when the geographical areas considered are coarse-grained, showing thus that the model accounts for the main mechanisms at play in the dynamics of the system at different scales. We have studied, besides, the relevance of the mobility range for the quantitative agreement of the model. Despite the various heterogeneity sources of the system (population, geography, topology and commuter fluxes), the model still displays logarithmic spatial correlations as in a two-dimensional diffusion [54]. This robustness is connected to the spatial decay of the coupling across cell [44, 55]. The field of random walks on heterogeneous media could also provide valuable insight [56].

The present work offers -with the use of demographic data as input- a comparison of a theoretical model with real data, which is used both for calibration and to evaluate the results. It proposes a path for further research in opinion dynamics since it establishes a method to bridge the gap existing between microscopic mechanisms of social interchange and macroscopic results of surveys and electoral processes. One limitation of the work is the use of census data, which translates in a lack of fine structure for the interaction network. The use of digital data will provide the necessary information to fill this gap. Another important issue is the dynamics of the average vote-share. To this end further elements need to be included, as for example the effects of social and mass media.

J.F.-G. receives economic support from the Conselleria d’Educació, Cultura i Universitats of the Government of the Balearic Islands and the ESF. J.J.R. acknowledges funding from the Ramón y Cajal program of the Spanish Ministry of Economy (MINECO). Partial support was received from MINECO and FEDER through projects, (FIS2011-24785), (FIS2012-30634) and from the EU Commission through the FP7 projects EUNOIA and LASAGNE.
[1] C. Castellano, S. Fortunato and V. Loreto, Rev. Mod. Phys. 81, 591-646 (2009).
[2] M. San Miguel, V.M. Eguíluz, R. Toral and K. Klemm, Computing in Science and Engineering 7, 67-73 (2005).
[3] L. Lizana, N. Mitarai, K. Sneppen and H. Nakanishi Phys. Rev. E 83, 066116 (2011).
[4] A. Kandler, R. Unger and J. Steele, Phil. Trans. R. Soc. B 365 3855-3864 (2010).
[5] D.M. Abrams and S.H. Strogatz, Nature 424, 900 (2003).
[6] R.M. Bond, C.J. Fariss, J.J. Jones, A.D.I. Kramer, C. Marlow, J.E. Settle and J.H. Fowler, Nature 489, 295-298 (2012).
[7] D. Centola, Science 329, 1194-1197 (2010).
[8] A.C. Gallagher, J.J. Haleb, D.J.T. Sumpter, S. Garnierra, A. Kacelnikb, J.R. Krebsb, I.D. Couzin, Proc. Natl. Acad. Sci. USA 109, 72457250 (2011).
[9] J. Lorenz, H. Rauhut, F. Schweitzer and D. Helbing, Proc. Natl. Acad. Sci. USA 108, 9020-9025 (2011).
[10] S. Fortunato and C. Castellano, Phys. Today 65, 74 (2012).
[11] B.C. Straits, Public Opin. Q. 54, 64-73 (1990).
[12] C.B. Kenny, Am. J. Pol. Sci. 36, 259-267 (1992).
[13] P.A. Allen, R.J. Dalton, S. Greene and R. Huckfeldt, Am. Polit. Sci. Rev. 96, 57-73 (2002).
[14] A.S. Zuckerman The Social Logic Of Politics: Personal Networks As Contexts For Political Behavior (Temple University Press) (2005).
[15] J.H. Fowler, Turnout in a small world in Social Logic of Politics, ed. A. Zuckerman (Temple University Press), pp 269-287 (2005).
[16] P.F. Lazarsfeld, B. Berelson and H. Gaudet, The People’s Choice: How the Voter Makes Up His Mind in a Presidential Campaign (Columbia University Press) (1948).
[17] E.M. Rogers, in Diffusion of Innovations (Free Press, New York) (1995).
[18] N.A. Christakis and J.H. Fowler, N. Engl. J. Med. 357, 370-379 (2007).
[19] A.L. Hill, D.G. Rand, M.A. Nowak and N.A. Christakis, Proc. Biol. Sci. 277, 3827-3835 (2010).
[20] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksen, M.W. Feldman, L. Fogarty, S. Ghirlanda, T. Lilli-crap and K.N. Laland, Science 328, 208-213 (2010).
[21] S. Fortunato and C. Castellano, Phys. Rev. Lett. 99, 138701 (2007).
[22] A. Chatterjee, M. Mitrović and S. Fortunato, Sci. Rep. 3, 1049 (2013).
[23] P. Klimek, Y. Yegorov, R. Hanel and S. Thurner, Proc. Natl. Acad. Sci. USA 109, 16469-16473 (2012).
[24] C. Borghesi and J.-P. Bouchaud, Eur. Phys. J. B 75, 395-404 (2010).
[25] C. Borghesi, J.-C. Raynal and J.-P. Bouchaud, PLoS ONE 7, e36289 (2012).
[26] R. Enikolopov, V. Korovkin, M. Petrova, K. Sonin and A. Zakharov A, Proc. Natl. Acad. Sci. USA 110, 448-452 (2012).
[27] R.N. Costa Filho, M.P. Almeida, J.S. Andrade and J.E. Moreira, Phys. Rev. E 60, 1067-1068 (1999).
[28] R.N. Costa Filho, M.P. Almeida, J.E. Moreira and J.S. Andrade Jr., Phys. A 322, 698-700 (2003).
[29] L.E. Araripe and R.N. Costa Filho, Phys. A 388, 41674170 (2009).
[30] A.T. Bernardes, D. Stauffer and J. Kertész, Eur. Phys. J. B 25, 123-127 (2002).
[31] M.L. Lyra, U.M.S. Costa, R.N. Costa-Filho and J.S. Andrade, Europhys. Lett. 62, 131-137 (2003).
[32] G. Travieso and L. da Fontoura Costa, Phys. Rev. E 74, 036112 (2006).
[33] L.E. Araripe, R.N. Costa Filho, H.J. Herrmann and J.S. Andrade, Int. J. Mod. Phys. C 17, 1809 (2006).
[34] M.C. Gonzalez, A.O. Sousa and H.J. Herrmann, Int. J. Mod. Phys. C 15, 45-57 (2004).
[35] C.A. Andersen, H.F. Hansen, G.L. Vasconcelos and J.S. Andrade, Int. J. Mod. Phys. 19, 1647 (2008).
[36] H. Fernández-Saldana, Phys. A 388, 2699-2704 (2009).
[37] N.A. Araújo, J.S. Andrade and H.J. Herrmann, PLoS ONE 5, e12446 (2010).
[38] J. Kim, E. Elliott and D.-M. Wang, Electoral Studies 22, 741-761 (2003).
[39] R. Holley and T.M. Liggett, Annals of Probability 3, 643-663 (1975).
[40] F. Vazquez and V.M. Eguíluz, New. J. Phys. 10, 063011 (2008).
[41] K. Suchecki, V.M. Eguíluz and M. San Miguel, Phys. Rev. E 72, 036132 (2005).
[42] M.C. González, C. Hidalgo and A.-L. Barabási, Nature 453, 779-782 (2008).
[43] C. Song, Z. Qu, N. Blumm and A.-L. Barabási, Science 327, 1018-1021 (2010).
[44] See Supplemental Material at [URL will be inserted by publisher] for more details about the datasets and the derivation of the equations.
[45] This approach is based on a division of the space in metapopulations. L. Sattenspiel and K. Dietz, Math. Biosci. 128, 71-91 (1995). D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco and A. Vespignani, Proc. Natl. Acad. Sci. USA, 106, 21484-21489 (2009); M. Tizzoni, P. Bajardi, C. Poletto, J.J. Ramasco, D. Balcan, B. Gonçalves, N. Perra, V. Colizza and A. Vespignani, BMC Medicine 10, 165 (2012).
[46] V. Sood and S. Redner, Phys Rev L 94, 178701 (2005).
[47] P.L. Krapivsky, S. Redner and E. Ben-Naim, in A Kinetic View of Statistical Physics, (Cambridge University Press), (2010).
[48] A. Baronchelli and R. Pastor-Satorras, J. Stat. Mech. 2009, L11001 (2009).
[49] M.A. Rodriguez, L. Pesquera, M. San Miguel and J.M. Sancho, J. Stat. Phys. 40, 669-724 (1985).
[50] K. Klemm, M.A. Serrano, V.M. Eguíluz and M. San Miguel, Scientific Reports 2, 292 (2012).
[51] United States’ Census Bureau, Web page: http://www.census.gov
[52] The American Time Use Survey (ATUS) for 2012, Bureau of Labor Statistics, available online at http://www.bls.gov/tus
[53] In the derivation of the master equation we set the differential of time to be $\delta t = 1/N$. This leads to Eq.(3) which is written in arbitrary units, as indicated in the revised manuscript. The choice of $\delta t = 1/N$ makes the arbitrary time unit identifiable to the MC step. Thus in the case of the full commuting network, where 2.5 MC steps correspond to a year, the arbitrary time unit in the Langevin equation corresponds to 0.4 years. In order to have the Langevin equation written such that the time unit is one year, one should rescale time by a factor $a = 2.5$ and both the deterministic part of the equation and the noise
term should be properly rescaled.

[54] S.-K. Ma, in *Statistical Mechanics*, published by World Scientific Publishing Co. Pte. Ltd. (Singapore 1985).

[55] S.A. Marvel, T. Martin, C.R. Doering, D. Lusseau, M.E.J. Newman, arXiv:1310.2636 [physics.soc-ph] (2013).

[56] D. Brockmann, L. Hufnagel, T. Geisel, Nature **439**, 462-465 (2006).