OPEN

Universality in voting behavior: an empirical analysis

Arnab Chatterjee, Marija Mitrović & Santo Fortunato

Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland.

Election data represent a precious source of information to study human behavior at a large scale. In proportional elections with open lists, the number of votes received by a candidate, rescaled by the average performance of all competitors in the same party list, has the same distribution regardless of the country and the year of the election. Here we provide the first thorough assessment of this claim. We analyzed election datasets of 15 countries with proportional systems. We confirm that a class of nations with similar election rules fulfill the universality claim. Discrepancies from this trend in other countries with open-lists elections are always associated with peculiar differences in the election rules, which matter more than differences between countries and historical periods. Our analysis shows that the role of parties in the electoral performance of candidates is crucial: alternative scalings not taking into account party affiliations lead to poor results.

W e know from statistical physics that systems of many particles exhibit, in the aggregate, a behavior which is enforced by a few basic features of the individual particles, but independent of all other characteristics. This result is particularly striking in critical phenomena, like continuous phase transitions and is known as *universality*[1]. Empirical evidence shows that a number of social phenomena are also characterized by simple emergent behavior out of the interactions of many individuals. The most striking example is collective motion[2–4]. Therefore, in the last years a growing community of scholars have been analyzing large-scale social dynamics and proposing simple microscopic models to describe it, alike the minimalistic models used in statistical physics. Such scientific endeavour, initially known by the name of *sociophysics*[5–7], has been meanwhile augmented by scholars and tools of other disciplines, like applied mathematics, social and computer science, and is currently referred to as *computational social science*[8].

Elections are among the largest social phenomena. In India, USA and Brazil hundreds of million voters cast their preferences on election day. Fortunately, datasets can be freely downloaded from institutional sources, like the Ministry of Internal Affairs of many countries. Therefore, it is not surprising that elections have been among the most studied social phenomena of the last decade[9]. By now, several aspects of voting behavior have been examined, like statistics of turnout rates[10,11], detection of election anomalies[12,13], polarization and tactical voting in mayoral elections[14,15], the relation between party size and temporal correlations[16], the relation between number of candidates and number of voters[17], the emergence of third parties in bipartisan systems[18], the correlation between the score of a party and the number of its members[19], the classification of electoral campaigns[20], etc.

The most studied feature is the distribution of the number of votes of candidates[21–35]. In the first analysis by Costa Filho et al.[21], the distribution of the fraction of votes received by candidates in Brazilian federal and state elections seems to decay as a power law with exponent -1 in the central region. Following this finding several similar analyses have been performed on election data of various countries, like India[25], Indonesia[26] and Mexico[28].

However, Fortunato and Castellano observed that the analysis by Costa Filho et al. treats all candidates equally, neglecting the role of the party in the electoral performance[30]. This assumptions appears too strong and unjustified, as the final score of the candidate is likely to depend on whether his/her party is popular or not. For this reason Fortunato and Castellano argued that characterizing and modelling the competition of candidates of the same party is more promising, as the performance of the candidates would be mostly depending on their own activity, rather than on external factors. Such competition occurs in a peculiar type of voting system, viz. proportional elections with open lists, where people may vote for a party and one or more candidates. In this system, people may actually choose their representatives by voting directly for them, whereas the number of candidates entering the Parliament for a given party typically depends on the strength of the party at the national and/or regional level. In these elections, it was found that the distributions of the number of votes of a candidate, divided by the average number of votes of all party competitors in the same list, appear to be the same regardless of

the country and the year of the election[30]. This claim has been recently disputed by Araripe and Costa Filho, who found that the universal curve computed in Ref. 30 does not follow well the profile of the distribution of Brazilian elections, which are also proportional and with open lists.

Here we carry out the first comprehensive analysis of the distribution of candidates' performance, using election results of 15 countries. We focus on proportional elections, as they feature the open-list system that allow voters to choose their representatives, enabling a real competition between candidates. We conclude that the relative performance, i.e. the ratio between the number of votes of a candidate and the average score of his/her party competitors in a given list has indeed the same distribution for countries with similar voting systems, and that discrepancies from the universal distribution emerge when the election has markedly different features (e.g. large districts, compulsory voting and weak role of parties in Brazil). We also show that party affiliations cannot be neglected: statistics of the absolute performance of candidates of different parties, like that investigated in the original analysis by Costa Filho et al., do not compare well between countries.

Results

Proportional elections. The electoral system we wish to study is proportional representation (PR)[36]. We analyze data from parliamentary elections of 15 countries: Italy (until 1992), Poland, Finland, Denmark, Estonia, Sweden, Belgium, Switzerland, Slovenia, Czech Republic, Greece, Slovakia, Netherlands, Uruguay and Brazil. The basic principle is that all voters deserve representation and all political groups deserve to be represented in legislatures in proportion to their strength in the electorate. In order to achieve this 'fair' representation, the country is usually divided into multi-member districts, each district in turn allocating a certain number of seats. Most countries having a PR system use a party list voting scheme to allocate the seats among the parties – each political party presents a list of candidates for each district. On the ballot the voters indicate their preference to a political party by selecting one or more candidates from the list. The number of seats assigned to each party in a district is proportional to the number of votes collected by the party. The party list systems can be categorized into *open*, *semi-open* and *closed*.

Open lists. Open lists enable voters to express their preference not only among parties but also among candidates. A party presents an unordered, random or alphabetically ordered list of candidates. Voters choose one or more candidates, and not the party. The position of each candidate depends entirely on the number of votes that he/she receives. In this category, we have studied data from Italy (before 1994, when a new system was introduced), Poland, Finland, Denmark, Estonia (since 2002), Greece, Switzerland, Slovenia, Brazil, Uruguay.

Semi-open lists. Semi-open lists impose some restrictions on voters directly or indirectly. The voter votes for either a party or a candidate within a party list. The parties usually put up a list of candidates according to their 'initial' preference, which depends on internal party rankings, etc. Candidates conquer parliamentary seats in the order they are ranked in the list, from the first to the last. However, if a candidate gets a number of votes exceeding a threshold, then he/she climbs up the ranking even if he/she was initially at the bottom of the list. The final order of the candidates is decided based on the 'initial' ordering and the actual votes received by the candidates. Sweden, Slovakia, Czech Republic, Belgium, Estonia (until 2002) and Netherlands fall in this category.

Closed lists. In the closed list system the party fixes the order in which the candidates are listed and elected. The voter casts a vote to a party as a whole and cannot express his/her preference for any candidate or

group of candidates. The representatives are then selected as they appear on the list, in the order defined before the elections. Countries voting with this system include Russia, Italy (since 2006), Spain, Angola, South Africa, Israel, Sri Lanka, Hong Kong, Argentina, etc. We did not consider this type of elections in our analysis, as there is no real competition between the candidates.

The allocation of seats to the parties takes place according to some pre-defined method, e.g. *d'Hondt*, *Hagenbach-Bischoff*, *Sainte-Laguë*, or some modified version of these[37].

Distribution of candidates' performance: open lists. In every proportional election, the country is divided into districts and each party presents a list with Q candidates. Voters typically choose one of the parties and express their preference among the candidates of the selected party. The seat allocation depends on the country (see Section I of Supplementary Information online) and has a large influence on how voters choose who they will vote for. The data sets we considered contain information about the number of votes v_i that each candidate i received and the number of candidates Q_i of the party list l_i including candidate i. From this information one can derive the total number of votes N_{l_i} collected by the Q_i candidates of list l_i. By summing over all party lists in the district D^i of candidate i we obtain the number of votes N_{D^i} in the district. The total number of votes cast during the whole election is indicated as N_T.

Our analysis consists in computing the probability distribution of the number of votes received by candidates, suitably normalized. We use the following normalizations:

- The scaling by Fortunato and Castellano[30], where the number of votes v^i of a candidate is divided by the average number of votes $v_0 = N_{l_i}/Q^i$ of all candidates in his/her party list. We shall indicate it as FC scaling.
- The scaling by Costa Filho, Almeida, Andrade and Moreira (CAAM)[21], where one considers the fraction of votes received by a candidate. Since it is unclear to us what one exactly means by that, we consider two possible normalizations: a) the fraction of the total votes in the district, v^i/N_{D^i}; b) the fraction of the total votes in the country v^i/N_T. We shall refer to them as to CAAMd and CAAMn, respectively. We rule out the fraction of votes in the party list because the authors made clear that they do not consider party affiliations. The most sensible definition appears the normalization at the district level, which will be thus reported here. The results for CAAMn are shown in the Supplementary Information (Figs. S1, S2, S3 online).

The universality discovered in Ref. 30 referred to elections held in Finland, Poland and Italy in various years. Here we confirm the result with a larger number of data sets (Fig. 1). Panels A, B and C display the distributions for Italy, Poland and Finland, respectively, in different years. The stability of the curve within the same country is remarkable, especially on the tail. In panel F we compare the distributions across the countries, yielding the collapse found in Ref. 30. Elections data in Denmark and Estonia (detailed in panels D and E), appear to follow the universal curve as well. We indicate this class of countries as Group U in the following. In Ref. 30 it was shown that this universal curve is very well represented by a log-normal function.

Italy (until 1992), Poland, Finland, Denmark and Estonia (after 2002) use open lists[36], in which voters can express their preference toward certain candidates within the party list and have a direct influence on the list ordering. These lists use the plurality method for the allocation of the seats within the party lists: candidates with the largest number of nominative votes are declared elected. There are just small differences in the number of candidates that a voter can indicate, the ordering of the candidates on the ballot, but the systems are basically the same, justifying the observed universality.

Other countries using open lists are Slovenia, Greece, Switzerland, Brazil, Uruguay. The results of the FC scaling are illustrated in Fig 2.
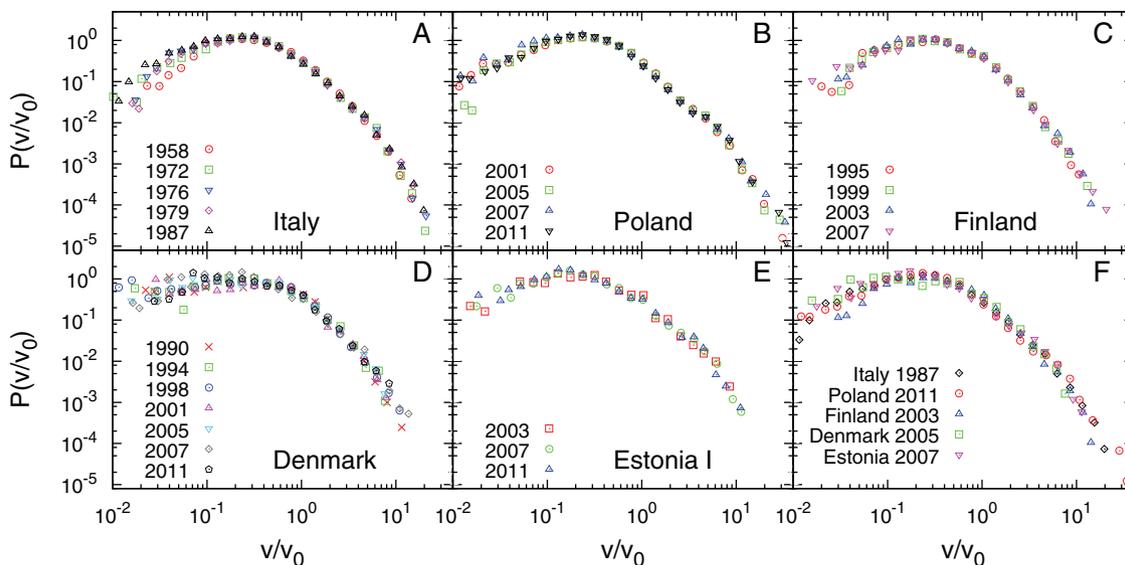
Figure 1 | Distribution of electoral performance of candidates in proportional elections with open lists, according to FC scaling. Italy (until 1992), Poland, Finland, Denmark and Estonia (after 2002) follow essentially the same rules, which is reflected by the data collapse of panel F. The historical evolution of the countries does not seem to affect the shape of the distribution (panels A to E).

While there is a historical persistence of the distribution at the national level, the curves do not really follow a common pattern, and do not match well the behavior of the universal distribution found for Italy, Poland, Finland, Denmark and Estonia. We distinguish here two classes of behaviors: Slovenia, Greece and Switzerland are characterized by a pronounced peak at $v/v_0 = 1$, and their tails match each other quite well. Brazil and Uruguay exhibit a monotonic pattern, quite different from the other three curves. The Brazilian curve follows quite closely the profile of the universal curve of Fig. 1 on the tail ($v/v_0 > 1$).

We conclude that open list systems do not guarantee identical distributions, but can be grouped in classes of behaviors. A close inspection of the election systems, however, may explain why we observe discrepancies. Slovenia divides its territory into eight districts which in turn are partitioned into 11 electoral units, each giving one candidate in the district. The voters can cast the vote for any of the candidates in the district, but the election of the candidate depends on the number of votes he/she won in his/her unit, i.e. the performance of the candidate in the unit is more important than the number of votes won in the district, which may affect both the candidates' campaigns and the voters' choices.

Greece uses a very complex seat allocation method among party lists and individual candidates. Although the ranking of the candidates on the list and the seats reallocation depends on the number of votes collected by the candidate, if one of the candidates happens to be the head of a party or a current or ex Prime Minister he/she is set automatically at the top of the party list, regardless of his/her electoral performance. Additionally, voting is compulsory, so many people cast a vote because they have to, without an informed opinion and/or motivation to participate in the election.

In Switzerland, voters may cast as many votes as there are seats in the district. They may vote for all members of the list, or for candidates of more than one party. Voters are also allowed to cast two votes per candidate. This type of list is classified as *free list*.

In Brazil, like in Greece, voting is compulsory, and we cannot exclude that this plays a role on the shape of the distribution. In addition, each state is just one district, which then comprises a number of voters orders of magnitude larger than the typical districts in all other elections. This explains why the Brazilian curve spans a much larger range of values for the performance variable than all other curves. The huge number of voters in the same district also

explains why parties present very long lists of candidates (often with over one hundred names). Finally, the role of parties is very weak; the political constellation frequently changes, with new parties being created and old ones being reshaped.

In Uruguay voters cannot choose candidates, but lists of candidates presented by the parties, the so-called sub-lemas. Therefore our analysis focuses on the distribution of performance of sub-lemas, instead of that of single candidates.

Figs. 3 and 4 show the analogues of Figs. 1 and 2 obtained by using CAAMd scaling. The historical stability of the corresponding distributions at the national level holds, however the comparison across countries is poor: curves appear to cross, not to collapse (panel F). According to Costa Filho et al.[21] the central part of the Brazilian curve follows a power law, with exponent close to 1; power law fits of the central region of the other distributions yield exponents sensibly different from each other, which confirms the crossing of the curves (see Supplementary Table S1). In particular, we could not identify any portion of the Polish curve resembling a power law. We conclude that the fraction of votes v/N_D collected by a candidate in his/her electoral district does not follow the same probability distribution in different countries, not even when they have essentially identical voting schemes, as in Figs. 1 and 3.

Distribution of candidates' performance: semi-open lists. The other countries we considered use semi-open lists, with different thresholds for the number of preferences that candidates are required to collect in order to secure a seat in the Parliament. The higher the electoral quota is, the harder is for a candidate to reach the required number of votes. In this case the position of the candidate within the party, as it appears on the ballot, has more influence on his/her final rank than the number of votes he/she collected. This can drastically effect the motivation of the candidate to lead a personal campaign. Also, high quotas diminish the influence of the voter on the final list ordering, which affects both the degree of a candidate's involvement in his/her personal campaign and the way people cast their preference votes. Therefore there is hardly an open competition between candidates, and this may be reflected in the shape of the distribution of performance. Figure 5 shows the probability density distributions for different countries with semi-open lists, according to FC scaling. The elections in Czech Republic held in 2010 had the lowest electoral quota and $P(v/v_0)$ (Fig. 5D) turns out to be very
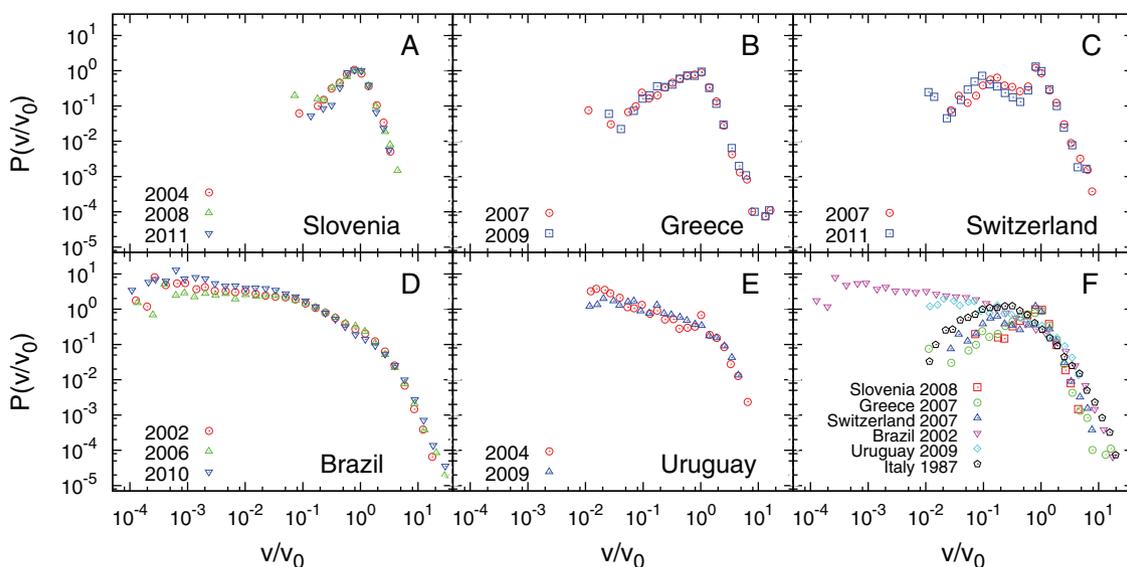
3

Figure 2 | Same analysis as in Fig. 1, for Slovenia, Greece, Switzerland, Brazil and Uruguay. Curves are fairly stable at the national level, but they do not compare well across countries and with the universal curves of Fig. 1 (represented in panel F by the distribution for the Italian elections in 1987). Such discrepancies are likely to be due to the different election rules of these countries as compared to each other and to the ones examined in Fig. 1, although they all adopt open lists.

similar to the curve obtained for Greek elections (Fig. 3B). The country with the highest electoral quota are the Netherlands, where each candidate has to win 10% of votes cast on the national level in order to be directly elected. Voters in Netherlands have little or no influence on the ordering of candidates, which is essentially frozen by the party, and they often vote for the top-ranked candidate and the first several names on the list, as they are the most popular and appreciated members of the party. This resembles the *rich-gets-richer* effect, which is characterized by power-law behavior of the distribution of the relevant quantities[38–42]. Indeed, the distribution of performance of Dutch candidates follows an approximate power-law behavior over most of the range of the performance v/v_0 (Fig. 5F).

Besides the values of the electoral threshold, these countries also differ in the number of nominative preferences a voter can cast, in the size and number of multi-member districts, as well as in the electoral formula that determines the final rankings (see Section I of Supplementary Information). Any change in the electoral system, i.e. these several factors, might influence the shape of $P(v/v_0)$. For instance, in 1994 Slovakia changed the number of multi-member districts, leading to appreciable changes in the shape of the distribution (Fig. 5C). The change in the electoral quota and the number of nominative votes decided in Czech Republic in 2006, may be the responsible for the variation of the curve before and after that year (Fig. 5D). The transition from semi-open to open lists introduced in Estonia in 2002, might explain why the curves before and after that year look different (Fig. 1E versus Fig. 5E). Interestingly, after the introduction of open lists in Estonia, the distribution of performance matches the universal distribution of the other countries with similar election systems (Fig. 1F), while before 2002 we find clear discrepancies.
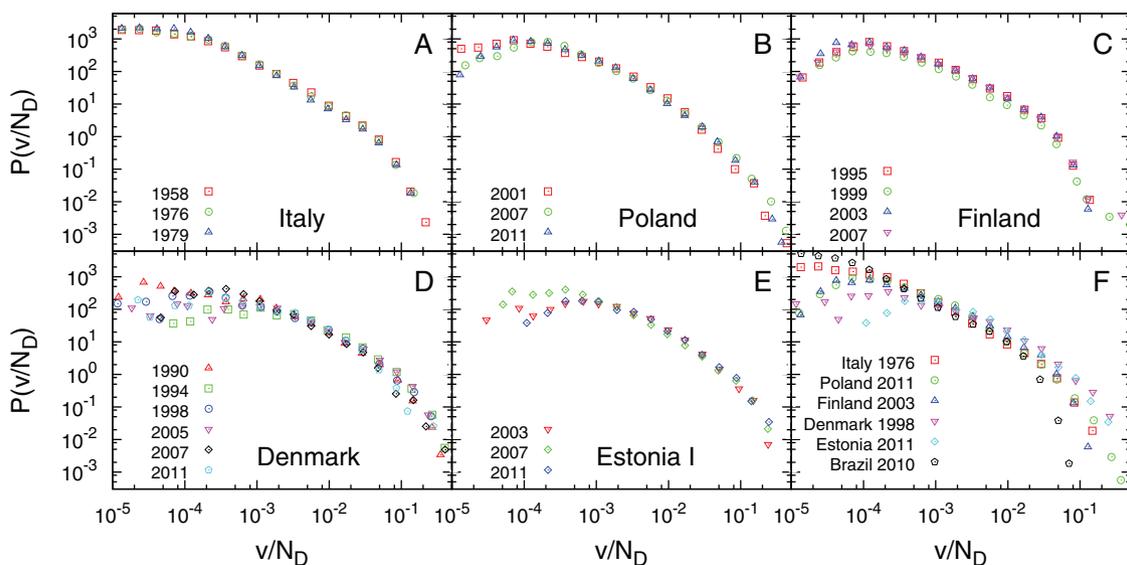

Figure 3 | Same analysis as in Fig. 1, with CAAMd scaling. Curves are stable at the national level, but they do not compare well across countries.
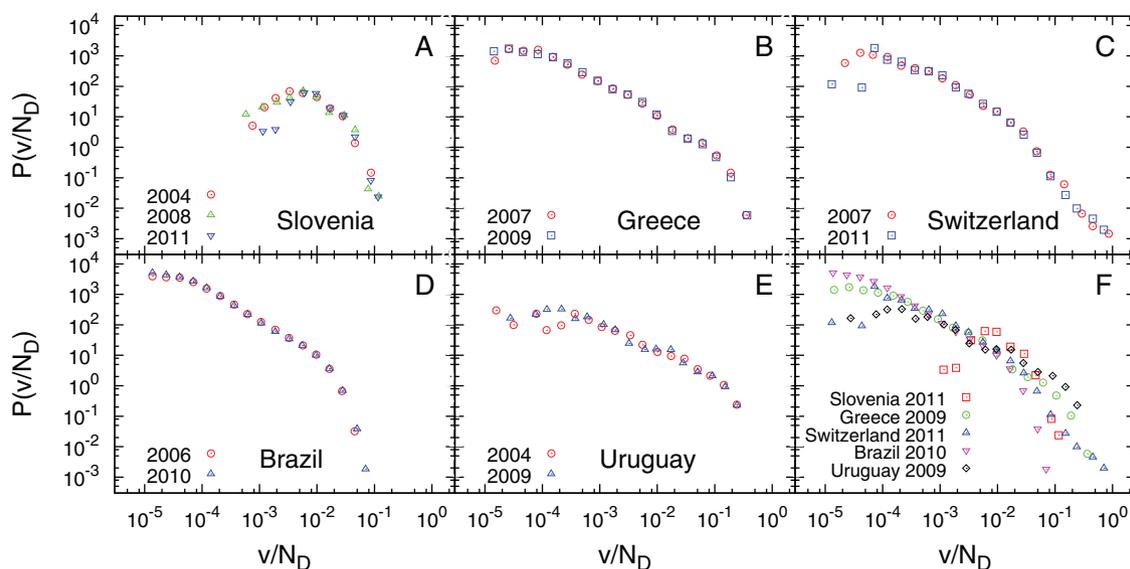
Figure 4 | Same analysis as in Fig. 2, with CAAMd scaling. Curves are stable at the national level, but they do not compare well across countries.

The corresponding distributions with CAAMd scaling also show marked differences between different countries (Fig. 6).

Estimating the similarity of the distributions. So far the estimation of the agreement or disagreement of different curves has been basically visual. In this section we would like to attempt a quantitative assessment of this issue. We build two matrices, whose entries are the values of the average distance D_{avg} and the maximum distance D_{max} between the distributions for any pair of countries for which we gathered election data (see Methods). The dissimilarity values for elections in the same country are reported on one diagonal of the matrix. Since we have adopted three different types of scaling for the electoral performance of candidates, FC, CAAMd

and CAAMn, we end up with six matrices, which are illustrated in Fig. 7. In each column we display the pair of matrices corresponding to one type of scaling, the first row contains the average distances, the second row the maximum distances. We built 16×16 matrices, even if we studied 15 countries, because we considered two sets of elections for Estonia, because of their transition from semi-open lists (Ee II) to open lists (Ee II), which took place in 2002.

Potential data collapses are indicated by low values of D_{avg} and D_{max}, which are easier to spot by using a color code, as we did in the figure. Numerical values are listed in the Supplementary Tables S2, S3, S4, S5, S6, S7 online. Dark squares (black-blue) correspond to the lowest values of D_{avg} and D_{max}, so to very similar distributions. The data collapse for the countries of Group U (Fig. 1F) is illustrated by
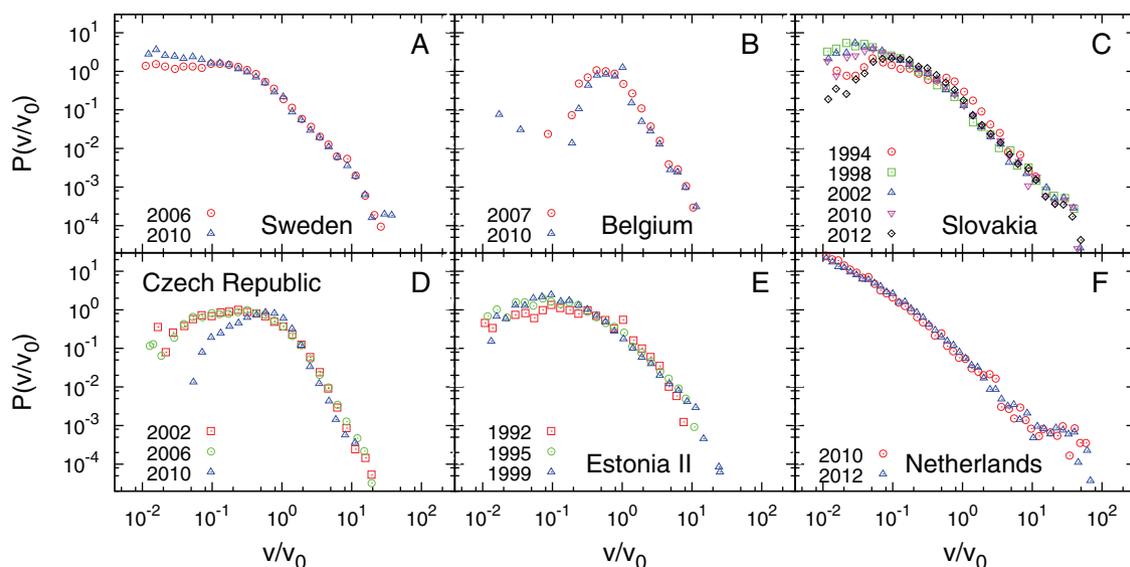

Figure 5 | Distribution of electoral performance of candidates in proportional elections with semi-open lists, according to FC scaling. Voters may express preferences for the candidates, but this plays a role for the final seat assignments only if the number of votes obtained by a candidate exceeds a given threshold, which varies from a country to another. At the national level curves are mostly stable, significant discrepancies correspond to changes in the election rules, like in Slovakia (C), Czech Republic (D) and Estonia (E). The apparent power law of the Dutch curve (F) might be generated by a rich-gets-richer mechanism, since the threshold is very high (10% at the national level) and voters typically tend to support the candidates based on their popularity. We stress that Estonia since 2002 has adopted open lists, which is why distributions of Estonian elections after 2002 are illustrated in Figs. 1 and 2 (labeled as Estonia I).
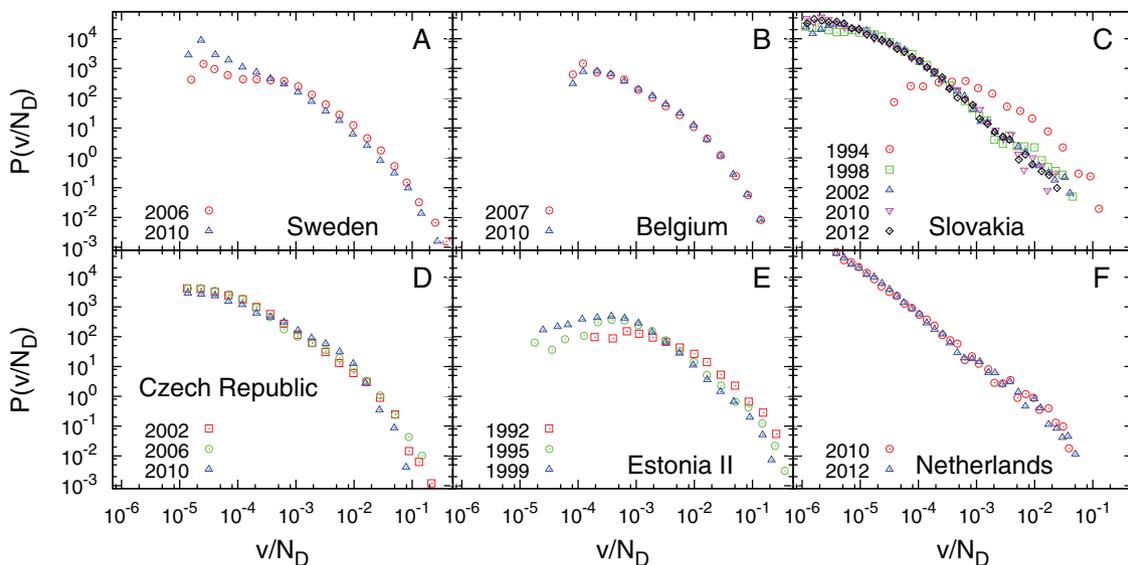
Figure 6 | Same as Fig. 5, with CAAMd scaling.

the bottom left block of A and B. Interestingly, we see that only the Estonian elections held after 2002 (Ee I) are very similar to the other curves of Group U; before 2002 Estonians used semi-open lists, the corresponding curves do not match well with the universal distribution.

We see that also the Brazilian and the Uruguayan distributions are fairly similar, on average, to the universal curve, mostly on the tail, although they considerably differ in the initial part, especially the Brazilian distributions. The strong similarity between the results of elections in the nations of Group U persists even if we consider the
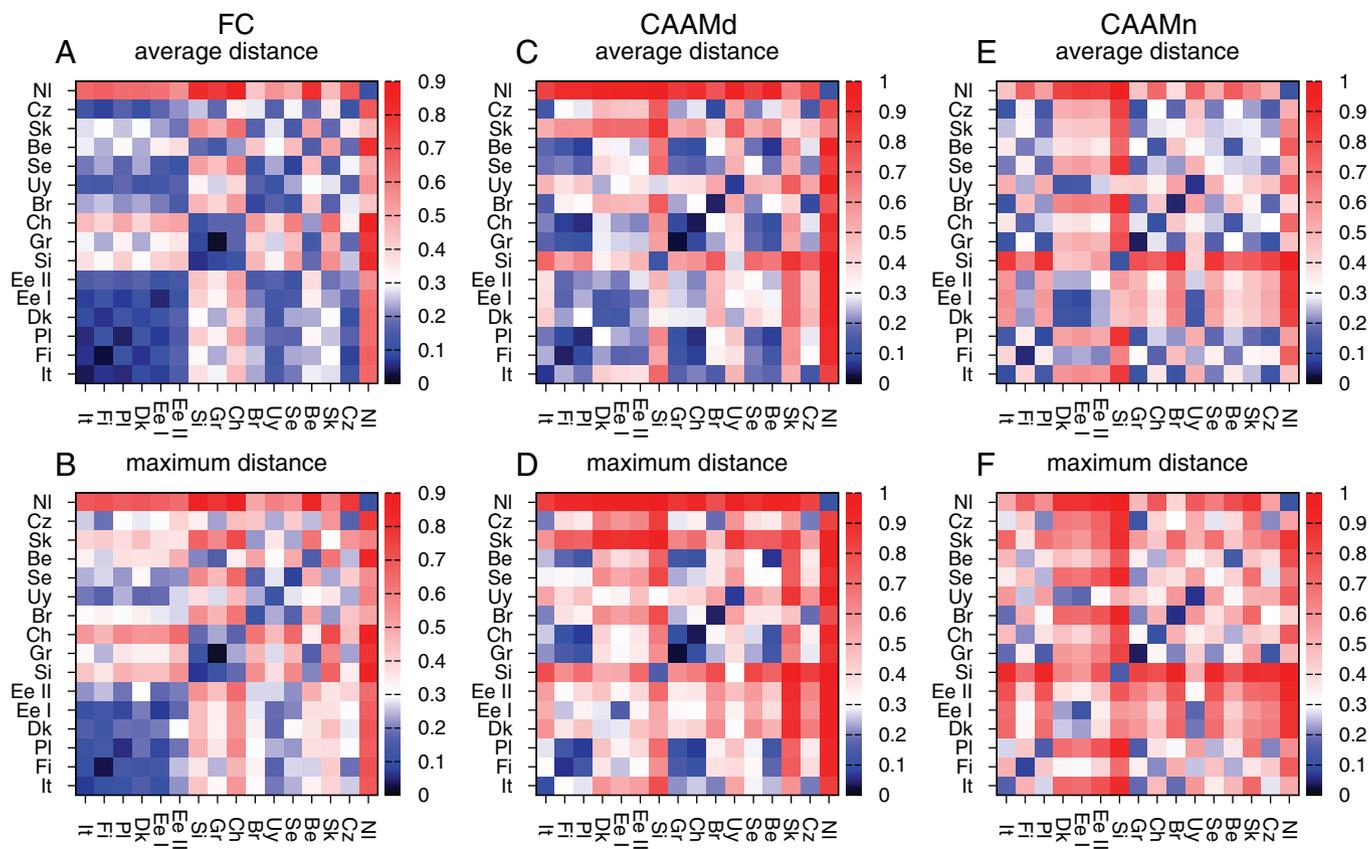

Figure 7 | Quantitative assessment of the similarity between distributions at the national level and between countries. The matrices in the top row indicate the values of the average K-S distance between datasets of different countries. On the bottom row the maximum distances are reported. Each column corresponds to one of the three types of distributions we have examined, by using FC, CAAMd and CAAMn scaling. A color code is adopted to better distinguish the low values of the distance (indicated by dark colours), indicating a big similarity between the curves, from the larger values, corresponding to poor collapses. The dark square on the bottom left of the matrices obtained via FC scaling confirms that the distributions of those countries are pretty close to each other, as illustrated in Fig. 1F. Conversely, the similarity between distributions obtained via CAAMd and CAAMn scaling appears rather modest for most pairs of countries.

maximum distance (panel B), as the dark block is still there, though blurred. Slovenia and Greece appear very similar to each other but sensibly different from the other countries. The diagonal from bottom left to top right shows the values of the distance for datasets in the same country. In general, the distances are pretty low, but we also find fairly large values. These correspond to countries which introduced changes in the election rules, reflected in the shape of the distributions, as described above.

If we move to CAAMd scaling (panels C and D) the scenario is considerably worse, in that the curves are much more dissimilar to each other than the ones obtained with FC scaling. In panel C, the average distance between the countries of Group U is still low, though higher than for FC scaling (panel A), but when one moves to the maximum distance the block disappears (panel D). For CAAMn (panels E and F) the curves are even more dissimilar to each other.

We are not giving here any indication on the significance of the measured values of the K-S distance. Large values indicate with certainty that the corresponding distributions are really different curves, but low values could still have high significance. As a matter of fact, all values that we found, for all types of scaling, indicate a significant discrepancy between the corresponding distributions. However, we stress that here we are considering the whole profile of the distribution, from the lowest to the highest value of the performance variable. The most interesting part of the distributions, and the one which is likely to reflect collective social dynamics, is certainly the tail, because it is where one has the largest cascades of votes for the same individual. On the contrary, the initial part of the curve corresponds to poorly voted candidates, and there are many ways to get to such modest outcomes (like being voted solely by closest family members and friends), hardly susceptible of a mathematical modelling. But at this stage we did not want to identify the most "interesting" part of the distribution by constraining the range of the variable, which is always tricky. Therefore we decided to compare the full distributions.

We finally remark that in social dynamics one can hardly get the same striking data collapses obtained in physical systems and models. Even if the social atom hypothesis implies that just a few features of the social actors and their interactions determine the large-scale behavior, the complexity of human nature and context-dependent factors may still have some influence, albeit small. For instance, in the Polish distributions of Fig. 1B there is a hump for $\nu/\nu_0 \approx 5$, which occurs systematically at the national level, but which is absent in the other distributions of the same class. Therefore, obtaining the agreement of the distributions shown in Fig. 1F, despite all differences between countries and historical ages, is truly remarkable.

Discussion

We have performed an empirical analysis of elections held in 15 countries in various years. We focused on the competition between candidates, which is a truly open competition when the voters can indicate their favourite representatives in the ballot and candidates with the largest number of votes are ranked the highest. This occurs in proportional elections with open lists. Of the countries for which we found data, 10 adopt open lists. Five of them (Group U), Italy, Finland, Poland, Denmark and Estonia (since 2002) have very similar election rules, the other five are characterized by important differences (e.g. compulsory vote, huge districts and weak role of parties in Brazil), which are likely to affect the behavior of voters and candidates, leading to measurable differences in the statistical properties of the electoral outcomes. Indeed, the distribution of the number of votes received by a candidate, normalized by the average number of votes gained by his/her competitors in the same party list, seems to be the same for the nations of Group U, while there are marked differences from the curves obtained from the other countries. This result, originally found by Fortunato and Castellano for

Italy, Finland and Poland[30], is confirmed here on a much larger data collections and holds for Denmark and Estonia as well.

Different patterns are found for countries adopting semi-open lists, in which in principle voters can choose the candidates, but the main ranking criterion is still imposed by their party, regardless of the final electoral score of the candidate, unless it exceeds a given threshold. In this system the competition among the candidates is therefore not really open, and it is no wonder that the distribution of electoral performance does not follow the profile of the curves of Group U.

In general we found that the shape of the distribution is much more sensitive to the specific election rules adopted in the countries than to the historical and cultural context where the election took place. This is evident when one considers the evolution in time of distributions of any given country, which remain essentially identical even after many years, if the voting system does not change, but display visible variations following the introduction and/or modification of election rules as it happened in Estonia in 2002, Slovakia in 1994, Czech Republic in 2006. The case of Estonia is spectacular: before 2002 it used semi-open lists, and the distributions of relative performance of a candidate with respect to his/her party competitors did not compare well with the curves of the other countries of Group U. After the introduction of open lists, instead, the distributions became very similar to the universal curve. Such sensitivity of the distributions might allow to detect anomalies, e.g. large-scale fraud, in future elections[12,13].

Our analysis proves that the success of a candidate, measured by the number of votes, strongly depends on the party he/she belongs to, and that only when one considers the competition among candidates of the same party universal signatures may emerge. Indeed, neglecting the party affiliation does not seem to take us very far: the two party-independent normalizations we have considered, following the procedure by Costa Filho et al.[21,23,32], do not seem to reveal strong common features among distributions of different countries, not even when the latter follow nearly identical election schemes (e.g. the nations of Group U).

Methods

Election data. Here we consider the data sets for parliamentary elections from 15 countries with open and semi-open lists: Italy (1958, 1972, 1976, 1979 and 1987)[43], Poland (2001, 2005, 2007 and 2011)[44], Finland (1995, 1999, 2003 and 2007)[45], Denmark (1990, 1994, 1998, 2001, 2005, 2007 and 2011)[46], Estonia (1992, 1995, 1999, 2003, 2007 and 2011)[47], Slovenia (2004, 2008 and 2011)[48], Greece (2007 and 2009)[49], Switzerland (2007 and 2011)[50], Brazil (elections for state deputies in 2002, 2006 and 2010)[51], Uruguay (2004 and 2009)[52], Sweden (2006 and 2010)[53], Belgium (2007 and 2010)[54], Slovakia (1994, 1998, 2002, 2010 and 2012)[55,56], Czech Republic (2002, 2006 and 2010)[57] and the Netherlands (2010 and 2012)[58]. Further details and sources for each file are given in Supplementary Table S8 online.

All datasets can be freely downloaded from the Website http://becs.aalto.fi/en/research/complex_systems/elections/.

Comparing distributions. We use the Kolmogorov-Smirnov (K-S) distance[59] to measure the dissimilarity of two empirical distributions. The K-S distance D is defined as the maximum value of the absolute difference between the corresponding cumulative distribution functions, i.e.

$$D = \max_x |S_{N_1}(x) - S_{N_2}(x)| \qquad (1)$$

where $S_{N_1}(x)$ and $S_{N_2}(x)$ are the cumulative distributions for two data sets of size N_1 and N_2.

Since we have multiple datasets for each country, in order to compute the dissimilarity of the distributions at the national level and across countries we proceed as follows. For a given country X we compute the distance between any two distributions for elections of X. For a pair of countries X and Y we compute the distance between any pair of distributions P_X and P_Y, corresponding to one dataset of X and one of Y, respectively. In both cases we take the average D_{avg} and the maximum D_{max} of the resulting values. In this way we estimate the average and the maximum distance between distributions of the same country and between distributions of two different countries.

1. Binney, J., Dowrick, N., Fisher, A. & Newman, M. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Oxford University Press, Oxford, UK, 1992).

2. Helbing, D. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067–1141 (2001).
3. Helbing, D., Farkas, I., Molnár, P. & Vicsek, T. *Pedestrian and Evacuation Dynamics* Eds. Schreckenberg, M. & Sharma, S. D. 19–58 (Springer Verlag, Berlin, Germany, 2002).
4. Vicsek, T. & Zafeiris, A. Collective motion. *Physics Reports* **517**, 71–140 (2012).
5. Ball, P. *Critical mass* (Farrar, Straus and Giroux, New York, USA, 2004).
6. Buchanan, M. *The social atom* (Marshall Cavendish Business, London, UK, 2007).
7. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).
8. Lazer, D. *et al.* Computational social science. *Science* **323**, 721–723 (2009).
9. Fortunato, S. & Castellano, C. Physics peeks into the ballot box. *Physics Today* **65**, 74–75 (2012).
10. Borghesi, C. & Bouchaud, J-P. Spatial correlations in vote statistics: a diffusive field model for decision-making. *Eur. Phys. J. B* **75**, 395–404 (2010).
11. Borghesi, C., Raynal, J. C. & Bouchaud, J-P. Election turnout statistics in many countries: Similarities, differences, and a diffusive field model for decision-making. *PLoS ONE* **7**, e36289 (2012).
12. Báez, G., Hernández-Saldaña, H. & Méndez-Sánchez, R. A. On the reliability of voting processes: the mexican case. Eprint arxiv:physics/0609114 (2006).
13. Klimek, P., Yegorov, Y., Hanel, R. & Thurner, S. Statistical detection of systematic election irregularities. *Proc. Natl. Acad. Sci.* **109**, 16469–16473 (2012).
14. Araripe, L. E., Costa Filho, R. N., Herrmann, H. J. & Andrade, J. S. Plurality Voting: the Statistical Laws of Democracy in Brazil. *Int. J. Mod. Phys. C* **17**, 1809–1813 (2006).
15. Araújo, N. A. M., Andrade, Jr. J. S. & Herrmann, H. J. Tactical voting in plurality elections. *PLoS ONE* **5**, e12446 (2010).
16. Andresen, C. A., Hansen, H. F., Hansen, A., Vasconcelos, G. L. & Andrade, Jr. J. S. Correlations between political party size and voter memory: A statistical analysis of opinion polls. *Physica A* **19**, 1647–1658 (2008).
17. Mantovani, M. C., Ribeiro, H. V., Moro, M. V. & Mendes, R. S. Scaling laws and universality in the choice of election candidates. *Europhys. Lett.* **96**, 48001 (2011).
18. Romero, D., Kribs-Zaleta, C., Mubayi, A. & Orbe, C. An epidemiological approach to the spread of political third parties. *Discrete and Continuous Dynamical Systems-Series B (DCDS-B)* **15**, 707–738 (2011).
19. Schneider, J. J. & Hirtreiter, C. The Impact of Election Results on the Member Numbers of the Large Parties in Bavaria and Germany. *Int. J. Mod. Phys. C* **16**, 1165–1215 (2005).
20. Sadovsky, M. G. & Gliskov, A. A. Towards a typology of elections at Russia. Eprint arxiv:0706.3521 (2007).
21. Costa Filho, R. N., Almeida, M. P., Andrade, Jr. J. S. & Moreira, J. E. Scaling behavior in a proportional voting process. *Phys. Rev. E* **60**, 1067–1068 (1999).
22. Bernardes, A. T., Stauffer, D. & Kertész, J. Election results and the Sznajd model on Barabási network. *Eur. Phys. J. B* **25**, 123–127 (2002).
23. Costa Filho, R. N., Almeida, M. P., Moreira, J. E. & Andrade, Jr. J. S. Brazilian elections: voting for a scaling democracy. *Physica A* **322**, 698–700 (2003).
24. Lyra, M. L., Costa, U. M. S. & Costa Filho, R. N. Generalized zipf's law in proportional voting processes. *Europhys. Lett.* **62**, 131–137 (2003).
25. González, M. C., Sousa, A. O. & Herrmann, H. J. Opinion formation on a deterministic pseudo-fractal network. *Int. J. Mod. Phys. C* **15**, 45–57 (2004).
26. Situngkir, H. Power Law Signature in Indonesian Legislative Election 1999–2004. Eprint arxiv:nlin/0405002 (2004).
27. Sousa, A. Consensus formation on a triad scale-free network. *Physica A* **348**, 701–710 (2005).
28. Morales, O., Martinez, M. & Tejeida, R. *Mexican voter network as a dynamic complex system.* Proceedings of the 50th Annual Meeting of the ISSS (2006).
29. Travieso, G. & da Fontoura Costa, L. Spread of opinions and proportional voting. *Phys. Rev. E* **74**, 036112 (2006).
30. Fortunato, S. & Castellano, C. Scaling and universality in proportional elections. *Phys. Rev. Lett.* **99**, 138701 (2007).
31. Gradowski, T. M. & Kosiński, R. A. Statistical properties of the proportional voting process. *Acta Phys. Pol.* **114**, 575–580 (2008).
32. Araripe, L. E. & Costa Filho, R. N. Role of parties in the vote distribution of proportional elections. *Physica A* **388**, 4167–4170 (2009).
33. Hernández-Saldaña, H. On the corporate votes and their relation with daisy models. *Physica A* **388**, 2699–2704 (2009).
34. Chou, C. I. & Li, S. P. Growth model for vote distributions in elections. Eprint arXiv:0911.1404 (2009).
35. Banisch, S. & Araújo, T. On the empirical relevance of the transient in opinion models. *Phys. Lett. A* **374**, 3197–3200 (2010).
36. Ortega Villodes, C. *Preference voting systems and their impact on the personalisation of politics.* in *Comparative studies of electoral systems meeting (CSES), Sevilla, Spain*, 2006.
37. Colomer, J. M. in *Handbook of Electoral System Choice,* Ed. Colomer, J. M. pp 3–78 (Palgrave-Macmillan, London, 2004).
38. Eggenberger, F. & Pólya, G. Uber die Statistik verketter Vorgange. *Zeitschrift fur Angewandte Mathematik und Mechanik* **3**, 279–289 (1923).
39. Simon, H. A. On a class of skew distribution functions. *Biometrika* **42**, 425–440 (1955).
40. Merton, R. The Matthew Effect in Science. *Science* **159**, 56–63 (1968).
41. de Solla Price, D. A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inform. Sci.* **27**, 292–306 (1976).
42. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
43. Ministero dell'Interno, Archivio Storico delle Elezioni http://elezionistorico.interno.it. 7 December 2012.
44. Państwowa Komisja Wyborcza (National Election Commision) http://pkw.gov.pl. 7 December 2012.
45. Statistics Finland http://tilastokeskus.fi. 7 December 2012.
46. Statistics Denmark http://www.dst.dk. 7 December 2012.
47. Estonian National Electoral Committee http://vvk.ee. 7 December 2012.
48. Republika Slovenija - odločanje državljank in državljanov na volitvah in referendumih (The Republic of Slovenia - decisions of citizens in elections and referendums) http://volitve.gov.si. 7 December 2012.
49. Hellenic Republic, Ministry of Interior: ELection Results http://www.ypes.gr/en/Elections/NationalElections/Results. 7 December 2012.
50. Swiss Statistics - Swiss Federal Statistical Office http://www.bfs.admin.ch. 7 December 2012.
51. Tribunal Superior Eleitoral - Brazil (Superior Electoral Court - Brazil) http://www.tse.jus.br. 7 December 2012.
52. Corte Electoral, Republica Oriental del Uruguay (Electoral Court, Oriental Republic of Uruguay) http://www.corteelectoral.gub.uy. 7 December 2012.
53. Valmyndigheten (Election Authority) http://www.val.se. 7 December 2012.
54. General Directorate of Institutions and Population http://www.ibz.rrn.fgov.be. 7 December 2012.
55. Political Transformation and the Electoral Process in Post-Communist Europe http://www.essex.ac.uk/elections. 7 December 2012.
56. Statistical Office of the Slovak Repulic http://portal.statistics.sk. 7 December 2012.
57. Czech Statistical Office http://www.volby.cz. 7 December 2012.
58. Databank Verkiezingsuitslagen (Election Results Database) http://www.verkiezingsuitslagen.nl. 7 December 2012.
59. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes 3rd Edition: The Art of Scientific Computing.* 3rd ed. (Cambridge University Press, New York, NY, USA, 2007).

Acknowledgments

Author contributions

All authors designed the research, collected and analyzed the data and participated in the writing of the manuscript.

Additional information

ERRATUM: Universality in voting behavior: an empirical analysis

Arnab Chatterjee, Marija Mitrović & Santo Fortunato

Due to a technical error, Figure 7 was published incorrectly in the html version of this article. This has now been corrected online.