

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# A Sociophysical Application of TSP: The Corporate Vote

Hugo Hernández-Saldaña  
*Universidad Autónoma Metropolitana at Azcapotzalco  
 Mexico*

## 1. Introduction

The relevance of the Traveling Salesman Problem (TSP), or the Traveling Salesperson Problem, is large and an indication of that it is the fact the present book is not the first (Applegate et al., 2006; Lawler et al., 1985; Gutin & Punnen, 2007). In this chapter we do not define the problem, neither offer new and faster way of solution, we present, instead, an application of TSP to sociophysics. The specific problem we deal here is to offer a dynamical explanation to the vote distribution of some corporations, i.e. the corporate vote. Recently, this distribution was described for the ruling party in Mexico during the majority of the XX century. With the arrival of the new millennium such a party became opposition, but it keep part of the organization which gave it a large power during seven decades. In reference (Hernández-Saldaña, 2009) the distribution of votes of this political party during the new millennium elections was described very well by the so called daisy models of rank  $r$  (Hernández-Saldaña et al., 1998). However, the physical origins of these models makes hard to establish a direct link with a socio-political phenomena. In order to explore a solution to this problem a TSP approach was proposed in (Hernández-Saldaña, 2009) and analyzed in (Hernández-Saldaña et al., 2010). In the present chapter we offer an integrated and detailed exposition of the subject.

In recent years the analysis of the distribution of votes from the point of view of statistical physics has been of interest. The analysis include the proportional vote in Brazil (Filho et al., 1999; 2003; Bernardes et al., 2002; Lyra et al., 2003); India and Brazil (Araripe et al., 2006); India and Canada (Sinha & Pan, 2006); Mexico (Morales-Matamoros et al., 2006) and Indonesia (Situngkir, 2004). A statistical analysis of election in Mexico (Báez et al., 2010) and Russia (Sadovsky & Gliskov, 2007) has been realized. Several models appeared in order to understand why we vote as we do (Fortunato & Castellano, 2007) or a study of the spatial correlations of the voting patterns (Borghesi & Bouchaud, 2010). The analysis of this problem is only one aspect of two new branches in physical sciences: the sociophysics and the econophysics. For an illuminating exposition of the former topic see the book of P. Ball (Ball, 2004), and for the latter consult the book of R.N. Mantegna (Mantegna & Stanley, 2000).

But, how to use TSP to model votes?. The idea is compare the statistical properties of the number of votes obtained for a political party in each cabin with the distance between cities in a TSP. The way to compare pears(votes) with apples(distances) is to map them to new variables where their statistical properties could be compared. Such a process is named, for historical reason, unfolding and we shall devote subsection 3.1 to explain carefully how this is performed. The idea is to measure in a dimensionless variable with density one. This approach has been successful analyzing fluctuation properties in a large set of problems,

mainly in the spectra of quantum systems with a chaotic classical counterpart. One of the relevant aspects was to find universal features in the fluctuation which depends only of large symmetries present in the system, like the existence of time invariance or not. In fact this statistical approach, named Random Matrix Theory, is currently successful and with an increasing number of applications, see Refs. (Brody et al., 1981; Mehta, 2004; Guhr et al., 1998) for explanation.

The TSP has not been absent of such approach, in reference (Méndez-Sánchez et al., 1996) an attempt to find universal properties of the quasi-optimal paths of TSP on an *ensemble* of randomly distributed cities was performed. However no RMT fluctuations have been found, but a fitting with a novel model, named daisy model of rank  $r$  was found in a posterior work (Hernández-Saldaña et al., 1998). This model raised in the frame of the statistical analysis of spectra of disordered systems at the transition from metal to insulator, i.e. a localization transition of the system quantum states. This model shall be describe below, since in one of its version, the rank 2 reproduces the statistical properties of the quasi-optimal paths in TSP and, describes some cases of the corporate vote distributions.

Since in the present chapter we are concern to statistical properties of quasi-optimal paths, the way and the time we obtain such a solutions is irrelevant. The results presented here have been obtained using simulated annealing according to (Press et al., 2007).

The rest of the chapter is organized as follows: In the next section we describe the TSP models we use to compare with the corporate vote. In the same chapter we discuss about the importance of the initial distribution of cities taken as an example actual country maps. In section 3 we develop the statistical measures we shall use. In section 4 we explain how the daisy models are built up as well as the electoral data. We compare all the models and data in section 5. We summarize the results and offer some conclusions in section 6.

## 2. Models and initial conditions

In order to enlighten the corporate vote we use two models (Hernández-Saldaña et al., 2010), both depart from a master square lattice of size  $b$  in the Euclidean space  $(x, y)$ , i.e. the intersection are localized at  $(nb, mb)$ , with  $n$  and  $m$  integers. The cities shall be positioned in the intersection surroundings according to a probability distribution of width  $\sigma$ ,  $\mathcal{P}(\sigma)$  centered at  $(nb, mb)$ . Hence, the position of the cities is

$$x_i = nb + p_i \quad (1)$$

$$y_i = mb + q_i, \quad (2)$$

being  $(p_i, q_i)$  selected from the distribution  $\mathcal{P}(\sigma)$ , which have zero mean. Up to now, the particular distribution is arbitrary. We choose as our model I an uniform one of total width  $2\sigma_s$  and as the model II a Gaussian one of standard deviation  $\sigma_g$ . We shall use this parameter in order to obtain a transition from the square lattice for  $\sigma_s = \sigma_g = 0$  to a map which looks like a randomly distributed cities map.

As it shall be clear when we discuss the nearest neighbour distribution, the important feature in this transition appears when the distribution width is large enough in order to admit a distribution overlap between nearest sites, i.e., when  $\sigma_s = \sigma_g = b/2$  (see figure 1).

In order to obtain enough statistics we consider an *ensemble* of maps. In the present work we use 500 maps of  $32 \times 31$  cities. The quasi-optimal paths is obtained using simulated annealing (Press et al., 2007). In figure 2 we present four realizations for different values of  $\sigma$  and for both models.

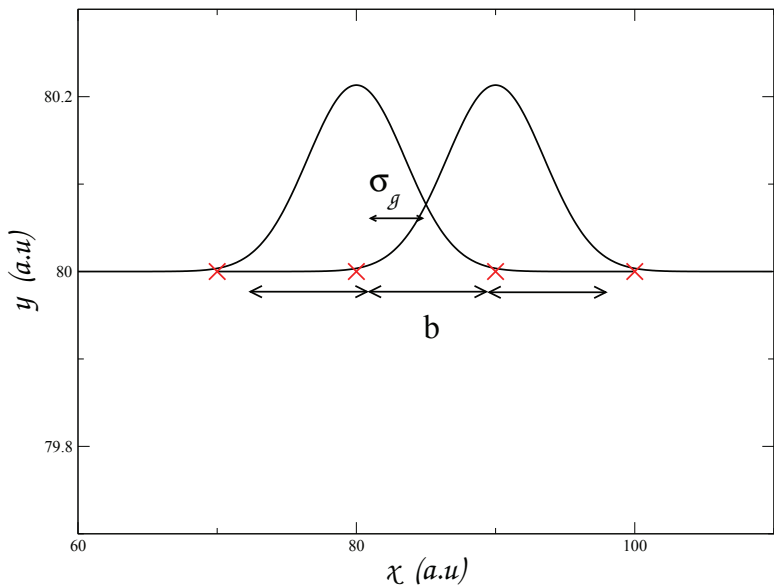


Fig. 1. Schematic construction of models I and II. In the figure we consider model II. The Gaussians are centered at red cross sites and cites are selected from the distribution.

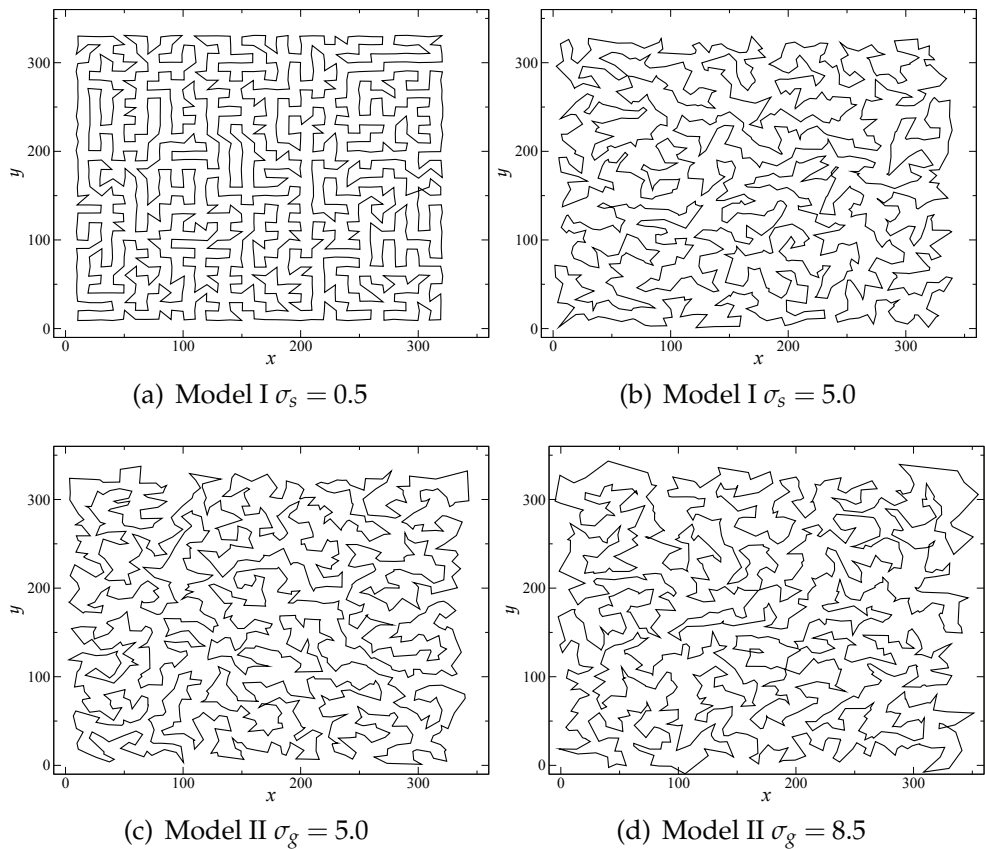


Fig. 2. Realizations of maps for both models with the parameter values indicated.

The statistical properties of distances between cities in the quasi-optimal path will be the point of comparison with votes for a corporate party. Note that this simple models are based on *standard* TSP, the peculiarity resides in the initial city distributions used for the calculations. This point is important since the initial set of cities, even in actual distribution

of cities, rules what kind of distances distributions could appear. If the initial distribution of cities has no large distances the quasi-optimal path can not create them. In reference (Hernández-Saldaña et al., 2010) the distribution of cities was done and a large dependence of the country was found. Here we present a sample made for 16 countries with data taken from (Applegate et al., 2006). In figure 3 we present the histogram of all the possible distances in the maps. Even when all the countries, exception of Tanzania and Oman, present almost the same power law behaviour at the beginning, the tail and the bulk in the histograms present large variations, from exponential, like Sweden, to large fluctuations like Canada and China. The distribution of normalized distances (or unfolded) with the procedure described below shows similar fluctuations (see figure 1 and 2 in (Hernández-Saldaña et al., 2010)). The search of universal properties in the TSP of actual countries is important even when figure 3 shows large fluctuations.

### 3. Statistical properties

The statistical characterization we used is commonly named spectral analysis since it is applied to spectra of quantum, optical or acoustical systems. But, for completeness, we add here a wide explanation about it together with some computational details usually not considered in the literature. The applications of this statistical analysis is completely general, see references (Bohigas, 1991; Guhr et al., 1998; Mehta, 2004) for a general introduction.

#### 3.1 Unfolding

The statistical analysis starts with a crucial step: separate the fluctuating properties from the secular ones in the data. This procedure fulfills two goals. First one, we take into account that the data could have a non constant density and a smooth variation could exist. The second goal is to have normalized fluctuations that can be compared between different members of an ensemble or even between different systems. This aspect played a crucial role in the analysis

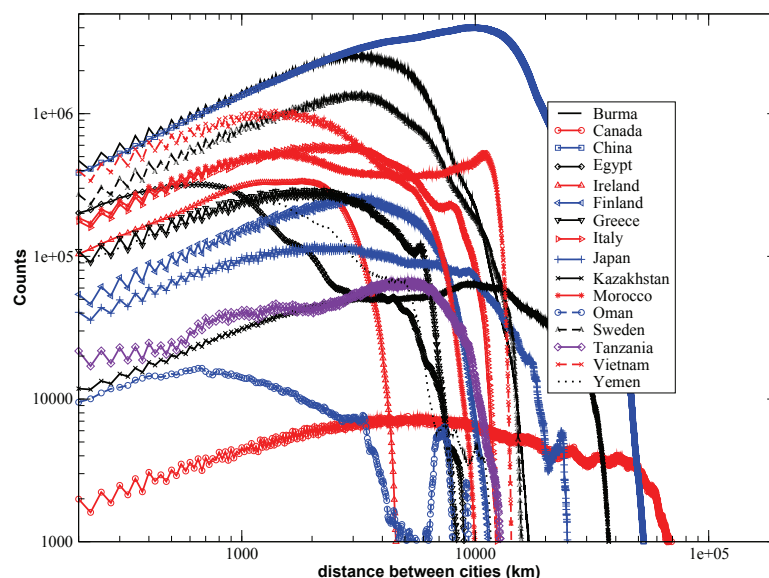


Fig. 3. Histogram of distances between cities for the countries shown in the inset. The data have been taken from (Applegate et al., 2006). Notice that all the countries, except Tanzania and Oman, start with a power law.

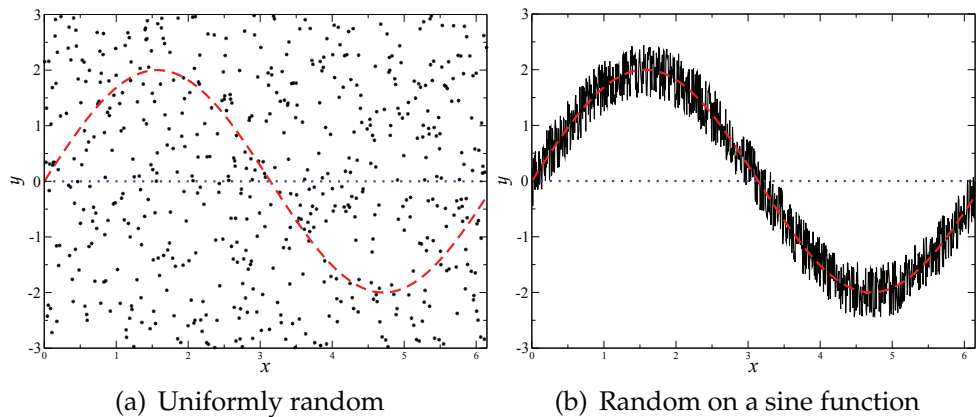


Fig. 4. Different averages of a random sequence, in (a) an uniformly random signal is shown in stars, its average is zero (blue dotted line). In (b) a signal with arithmetical average equal to zero, but with fluctuation around  $2\sin(x)$ , in red dashed line. The units are arbitrary.

of quantum chaotic systems and several of its applications. Including TSP and electoral processes. This procedure is called, for historical reasons, unfolding.

It is common to consider fluctuations, or the statistics like the variance, of a sequence which present a constant average or density like the presented in figure 4(a) in stars. In that case it is clear that the variance, for instance, is calculated in the usual way. However, there exists a large number of cases where the average is not longer constant but a smooth function like a polynomial or a sine function. In figure 4(b) we show this case. The arithmetical average is the same as in the previous case and indicated in dotted blue line, but the fluctuations, in fact, are around a sine function (red dashed line) and they are what we wish to analyze. That is clear that in the latter case the true average is a function and not longer a single number. In fact, an arithmetical analysis says that the function  $\langle y \rangle(x) = 0$  is a good result in both cases and the variance is smaller in case (b). A simple view of the data shows how wrong is this procedure and what we require to analyze are the fluctuations around the sine function.

The nature of the average function depends on the particular system and could have different forms and contributions. A common example is the case of the quantum energy levels,  $E_i$ , in a Hamiltonian system where there exists a good approximation to the average density. In such a case the density is well represented by the Weyl formulae,

$$\langle \rho(E) \rangle = \frac{1}{(2\pi\hbar)^d} \int \delta(H - E) dp dq. \tag{3}$$

Where  $\delta(\cdot)$  is the Dirac delta function,  $H$  corresponds to the Hamiltonian of  $d$  degrees of freedom and the integral is perform on all the phase space variables  $(p, q)$ . The basic quantum area is represented by  $\hbar$ , the Planck's constant. This approximation is valid in the semiclassical limit. Even in the case where the system dynamics is well established an accurate calculation of the average is a delicate task (Guhr et al., 1998).

It is customary to deal with the integrated density  $\mathcal{N}(E)$  instead of the density itself and the split is denoted as

$$\mathcal{N}(E) = \mathcal{N}_{\text{Secular}}(E) + \mathcal{N}_{\text{fluctuations}}(E). \tag{4}$$

As we explained previously, a second goal is to compare fluctuations of several members of an ensemble or even different systems. The reason, as we shall see below, is after unfolding the new variables will have mean equal to one. In the case where the system has a constant



density this procedure can be performed by dividing the spectrum by the density. In fact, the name itself “unfolding” comes from the density of energy in nuclear systems: For complex nuclei, the density is almost a semicircle and the idea is to *unfold it* to create a new, constant, density. Notice that this procedure is *always* of local character.

Now, we proceed to explain carefully how the unfolding is performed for distances in the TSP, however, *mutatis mutandis*, the present explanation can be applied to quantum energy levels, acoustic resonances, votes, DNA basis distances, etc. In order to perform the separation of equation (4) we require the analogue of the spectrum  $E_i$ , which is ordered as

$$E_1 < E_2 < E_3 < \dots < E_k < \dots \quad (5)$$

Hence we consider the cumulative of lengths for a quasi-optimal tour,  $d_k = \sum_{i=1}^k l_i$  with  $l_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$ . The cities are positioned at the points  $(x_i, y_i)$  and the cumulative lengths are ordered in increasing index. Notice that in this case the tour is periodic but in the case of energies not. The density of  $d$ 's (or of energies) is defined as

$$\rho(d) = \sum_k \delta(d - d_k), \quad (6)$$

The corresponding cumulative, or integrated, density is

$$\mathcal{N}(d) = \sum_k \Theta(d - d_k), \quad (7)$$

with  $\Theta(x)$ , the Heaviside or step function. This creates a staircase function with the pair of numbers  $(d_k, k)$  and an example is shown in figure 5(b). The secular part could be evaluated by a polynomial fitting of degree  $n$ . Numerical experience shows better results if, instead starting at  $\mathcal{N}(d_1) = 1$  we start at one half (see figure 5) (Bohigas, 1991).

We shall perform the statistical analysis on the variable transformed as

$$\tilde{\zeta}_k = \mathcal{N}_{\text{secular}}(d_k), \quad (8)$$

The secular part depends on the particular system taken into account or if we are considering a particular window. Notice that this transformation makes the mean level spacing to be unity, i.e.  $\Delta \equiv \langle \tilde{\zeta}_k - \tilde{\zeta}_{k-1} \rangle = 1$ . The new variables are expressed in units of mean level spacing.

The analysis is performed on windows of different size, this kind of analysis is always of local character. The polynomial fitted does not produce a good match in the window extremes (see figure 5(a)) hence, the extreme data are not considered for the statistical analysis. It is convenient to divide the data set in three parts and use only the middle third of the data. In order to avoid a data loss, the window is moved from third to third in order to only loss the extremes of the whole data. The subsets of data  $\tilde{\zeta}_k^i$  of window  $i$  are considered as members of an *ensemble* and the statistics will consider averaging on these subsets, i.e., an *ensemble* average. We shall return to this point later. In figure 5 we show a fitting on a large window of a very complex set of data (the quasi-optimal path obtained for Sweden). The larger the window the larger the polynomial degree used to fit it, hence, it is important to consider smaller windows in case the integrated density presents similar long range fluctuations. For practical purposes we consider no less than 300 data per window in order to keep 100 fitted data. Another important question is which degree in the polynomial must be considered. A large value of  $n$  could consider part of the fluctuations as the secular term. The answer is not trivial and requires of knowledge of the system itself. Any way, it is important to look directly

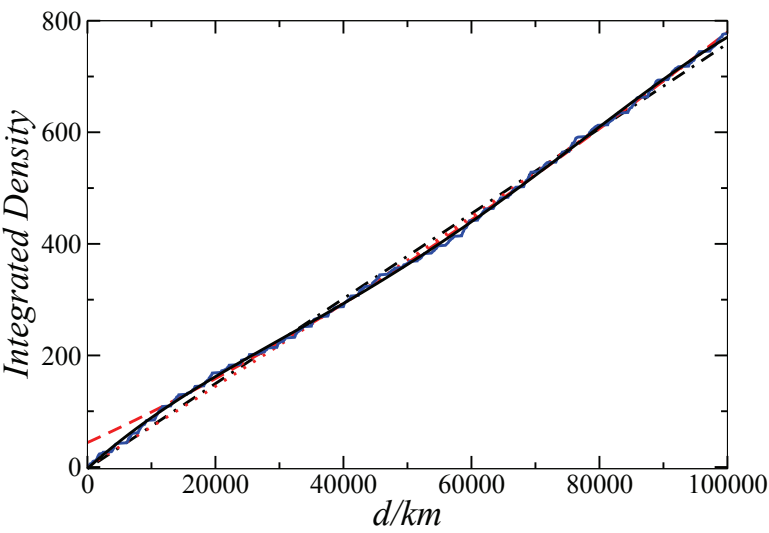


Fig. 5. Integrated density  $\mathcal{N}(d)$  calculated for the quasi-optimal path obtained for Sweden (blue histogram) and different polynomial fittings. We plot a window from 0 to 100000 km in the length variable  $d$ . The fitting was performed in the window from  $d = 10000$  to 100000 km. The linear approach is shown in black dot-dash line, the 2nd order in red dotted line, the 3rd order in dashed red line and the 4th order polynomial in black line. Note that even the fourth degree polynomial does not fit at all this large window but follows nicely the general shape.

the data fitting and do not believe the correlation coefficient value a priori(see figure 4). Many statistical estimators could give false results.

An important point is the integrated density could involve large numbers and we can loss accuracy in the fitting. A way to solve it is to use the translational invariance of the data and translate the windows to the origin each time. Since we are interested in differences of  $\zeta_k$  this translation does not

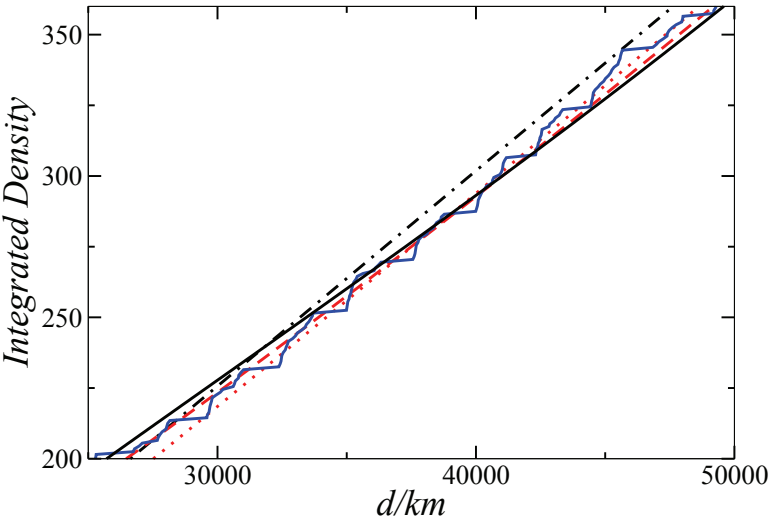


Fig. 6. A close up to previous figure, in order to see the fitted polynomial. Notice the staircase structure.



### 3.2 Fluctuations

Given the new sequence

$$\xi_1 < \xi_2 < \xi_3 < \dots < \xi_k < \dots < \xi_N, \quad (9)$$

we shall characterize their statistical properties. The probability of have a value at  $\xi_1$  and  $\xi_1 + d\xi_1$ , another between  $\xi_2$  and  $\xi_2 + d\xi_2$  and, in general a number between  $\xi_k$  and  $\xi_k + d\xi_k$ , is

$$P(\xi_1, \dots, \xi_N) d\xi_1 \dots d\xi_N, \quad (10)$$

regardless the labelling. Hence the statistical properties are characterized by the  $n$ -point correlation function,

$$R_n(\xi_1, \dots, \xi_n) = \frac{N!}{(N-n)!} \int P(\xi_1, \dots, \xi_N) d\xi_{n+1} \dots d\xi_N. \quad (11)$$

However, for practical purposes, in the literature it is considered two measures, the nearest-neighbor distribution  $p(1,s)$  and the number variance  $\Sigma^2(L)$ . The first one is the probability to have a separation  $s_i = \xi_i - \xi_{i-1}$  between  $s$  and  $s + ds$  and measures short range correlations. The latter is the variance of the number of levels in a box of size  $L$  and measures long range correlations. Several other statistics could be used, like Fourier transform, skewness, kurtosis and the  $n$ -th neighbor distribution width. For larger explanation, the reader could see references (Bohigas, 1991; Guhr et al., 1998; Mehta, 2004).

The nearest-neighbour distribution is built up by a histogram of  $s_i = \xi_i - \xi_{i-1}$ . The mean value is, by construction, one. Since, in general, we are considering an analysis on windows which are taken as statistically equivalent, the final distribution is an average on the distributions of all the windows, i.e., we consider an *ensemble* average. However could be possible to build up an ensemble of systems, as those considered in model I and II, this average is considered in a natural way. In both cases we consider the average of all the distributions, say the  $p(1;s)$  of each window or system.

In a similar way the  $n$ -th neighbour distribution is measured in the variable  $s_i = \xi_{i+n} - \xi_i$ . The number variance is calculated directly counting the number of level or numbers in boxes of size  $L$  and considering the variance. In order to obtain theoretical predictions it is useful to consider the 2-level cluster function  $Y_2(s)$ , defined as

$$Y_2(s) = 1 - R_2(s). \quad (12)$$

Where  $R_2(s)$  is the two point correlation function as defined in (11). This function can be evaluated if we have information about the neighbours distribution for all order, i.e.,

$$R_2(s) = \sum_{n=1}^{\infty} p(n;s). \quad (13)$$

Using equations (12) and (13) we can obtain the number variance by

$$\Sigma^2(L) = L - 2 \int_0^L ds (L-s) Y_2(s). \quad (14)$$

Hence, if we know all the  $n$ -th neighbour distribution it is possible to calculate an analytical expression for  $\Sigma^2$ . Even when we shall not use the  $\Sigma^2$  statistics for comparison with the electoral results is useful to calculate it. In figure 7 we show the result for several values of the TSP models and the asymptotic behavior in daisy models, where the grow is  $\sim L/(r+1)$ . The correlations are slightly different but all them remain linear.

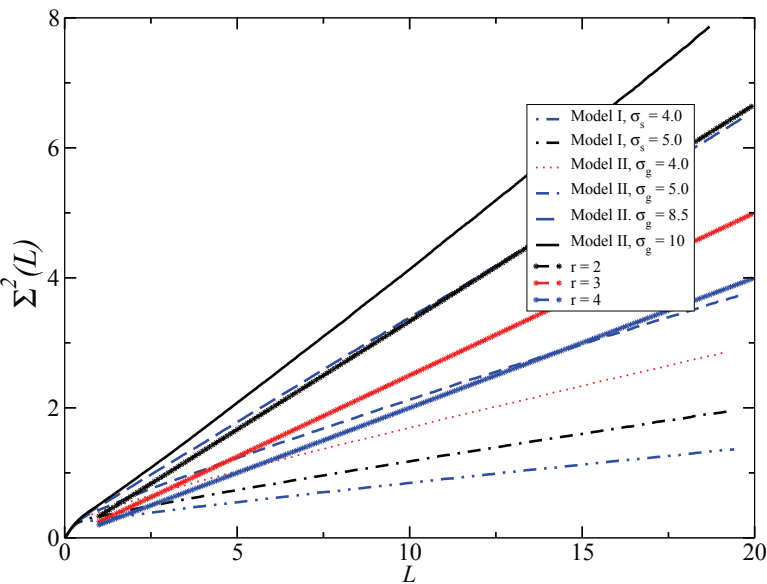


Fig. 7. Number variance  $\Sigma^2$  for the models and values indicated in the inset. For the daisy models only the asymptotic values are reported.

4. Electoral data and daisy models

4.1 The electoral data

The electoral data were taken from the Mexican electoral authorities official web page (Instituto Federal Electoral, 2006) and they can be obtained on request as well. We consider three elections in the new millennium: the federal elections in 2000, 2003 and 2006. The selection was made since, in the three cases, the corporate party, the Partido Revolucionario Institucional (PRI), arrived as opposition with only their corporate members (in fact, this party loose the presidential election in 2000). This offer the opportunity to explore the corporate vote only. Additionally, the database of the last elections are available in electronic format. Analysis on different Mexican elections is matter of current research.

Federal elections in Mexico are organized by the Instituto Federal Electoral (IFE) and they are made by direct vote. In such an election, people choose the republic president and the members of both chambers. The country is divided in electoral districts and the cabins are localized according to the number of registered electors. Each cabin admits a maximum of 750 votes, except the special cabin which could admit a larger number of votes. Such cabins are devoted to people who have the right to vote but who are in transit or live in a different place where they are registered for electoral matters. The number of cabins in whole country is around 117,000.

The distribution of votes is built up by counting in how many cabins exists 1 vote, 2 votes and so on and considering a histogram of them. In symbols, be  $n_i$  the number of votes in cabin  $i$ , built the sequence  $x_i$  by

$$x_{i+1} = x_i + n_{i+1},$$

(15)

and define  $x_1 = n_1$ . The corresponding histograms for the crude data can be seen in reference (Hernández-Saldaña, 2009). In order to obtain the statistical properties the unfolding must be done, however, a natural way to order the votes counts like in (9) is not obvious or, even, does not exist. This is clear since the order of cabins is assigned alphabetically into the database (Instituto Federal Electoral, 2006) and with no relation to the social or political distribution of votes of this particular party. Hence, the density of vote could vary tremendously or,

even, could not exist in the database. A way to handle this problem and in order to break the geographical correlations is to consider a randomization of the votes and construct an ordered sequence like in equation (9). Clearly, this new sequence should have an uniform distribution and we can use it in order to unfold the sequence since its density is constant. Notice that the transformed nearest neighbour distribution of votes is the mapped vote  $n_i$  to the unfolded new variable  $s_i$ . This procedure gives only information about the nearest neighbour distribution and it puts limits to our analysis. However, opens the research of a natural order in the vote being it geographical, corporative, social, etc.

In TSP the problem does not appear since the quasi optimal path offers a natural way to make the ordering. In the case of actual cities maps, however, the unfolding must be carefully performed since there exists a lot of variations (Hernández-Saldaña et al., 2010).

#### 4.2 Daisy model of rank $r$

The daisy model is constructed by retaining each  $r + 1$  level in a sequence of random numbers increasingly ordered as in (9). Since the original sequence of numbers have the  $n$ th-neighbour distribution given by

$$p(n;s) = \frac{s^{n-1}}{(n-1)!} \exp(-s), \quad (16)$$

and being  $n = 1$  the first or nearest neighbour, the new sequence has the  $(r + 1)$ th-neighbours distribution. But is must be renormalized in order to recover the standard average values, i.e.

$$\int_0^\infty s p(n;s) ds = n. \quad (17)$$

With the appropriate change of variable we obtain the  $n$ th-neighbour distribution for the daisy model of rank  $r$ :

$$p_r(n;s) = \frac{(r+1)^{(r+1)n}}{\Gamma((r+1)n)} s^{(r+1)n-1} \exp(-(r+1)s). \quad (18)$$

Where  $\Gamma(\cdot)$  is the gamma function. See reference (Hernández-Saldaña et al., 1998) for the whole derivation. The name of the models becomes clear for the case of rank  $r = 1$ : Here we label each level in the sequence  $\xi_i$ , one with the label “she loves me” and the next with the label “she loves me not”. Since we are optimistic we retain only the love sequence. Meanwhile the rank 1 model applies to the metal to insulator transition the rank 2 fits very well the TSP with an ensemble of randomly distributed cities (Hernández-Saldaña et al., 1998).

### 5. Results

We can compare the nearest neighbour distribution  $p(s)$  obtained from model I and II with the electoral results. The comparison is done on three cases: the presidential election of 2006, the senators election in 2000 and the low chamber election in 2006. The selection was done in order to present different cases. We shall use the notation  $p(s)$  instead of  $p(1;s)$  for shorthand. The TSP models were built up on rectangles of size  $b = 10 \text{ km}$  and 500 maps of  $32 \times 31$  cities. The  $p(s)$  for 2006 presidential election is presented in figure 8 in black histogram. The daisy model of rank 3, in black dashed line, presents a remarkable fitting to the distribution main bulk. The explicit expression for this daisy model is

$$p_3(s) = \frac{4^4}{3!} s^3 \exp(-4s). \quad (19)$$

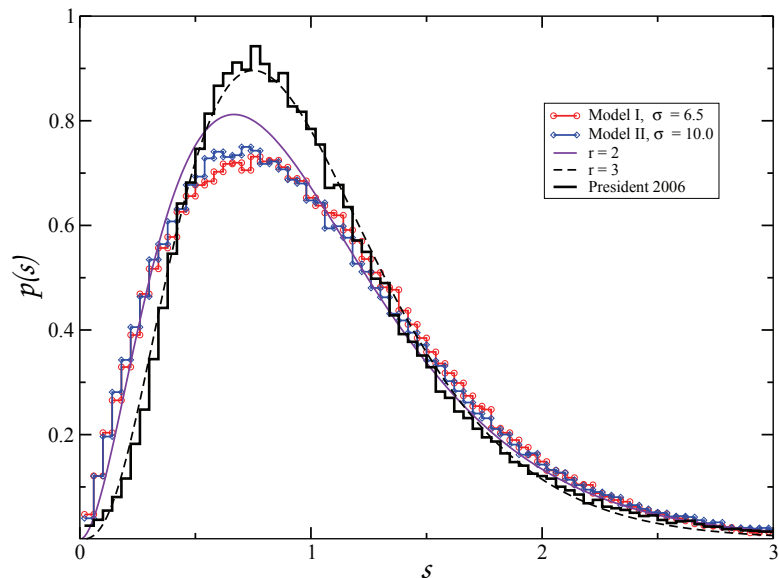


Fig. 8. Nearest neighbour distribution for the Presidential election in 2006 compared with models I and II and daisy models. The parameter values are indicated in the inset. Notice that the daisy model of rank 3 fits remarkably well the vote distribution. The variable  $s$  is expressed in mean spacings units, see subsection 3.1 for explanation.

However the tail does not longer fit, being fitted instead by a rank 2 daisy (see figure 9 ). Neither model I or II could fit the distribution bulk, but the tail is well fitted from widths,  $\sigma$ , departing from  $\sigma_s \sim 6.5$  (in red line with circles) and  $\sigma_g \sim 8.5$  (in blue line with diamonds)) for the respective model. The numerical exploration in both model shows that the decay reaches a limit compatible with an  $\exp(-3s)$  decay (see figure 9). This limit corresponds to the daisy model of rank 2,

$$p_2(s) = \frac{3^3}{2!} s^2 \exp(-3s), \tag{20}$$

and it is compatible with a TSP with cities selected from an uniform random distribution (Hernández-Saldaña et al., 1998). As we have noticed before, for this values of  $\sigma$  we have overlapping city distributions, since for model I this start at  $\sigma_s = b/2$  and for model II at  $\sigma_g = b/2$ , the distributions centers are two standard deviations apart. Hence, a tail compatible with slower decay  $\sim 3$  does not appear in the analyzed electoral data. It will be interesting to analyze the data for the incoming election in 2012. In the case of our models, model II presents slower decay for values larger than  $3b$ .

Referring the fit at the beginning of the vote distribution, we compare with other  $\sigma$  values for both models, see figure 10. For model I, maps with a  $\sigma_s \approx 2.75$  fit well, meanwhile for model II the value is  $\sigma_g = 3.5$  and the close up is reported in figure 11. As a general feature, the TSP models presented here are unable to produce a distribution with a start different from a power law of order 2 for values of  $\sigma$  larger than  $\sim b/2$ . Hence, this models do not reproduce this feature present in both, the electoral data and in the daisy models.

The rest of the comparisons are reported in figures 12 and 13. For the case of Low chamber election during 2006, figure 12, the value of the fitting parameters at the tail are compatible with  $b/2$  and rank 3 daisy model. The values are  $\sigma_s \approx 3.90$  and  $\sigma_g \approx 4.75$ . For the distribution beginning the values are equal to those reported for Presidential election in 2006. In the case of Senator in 2000, figure 13, the tail is fitted by  $r = 4$  and  $\sigma_s \approx 3.13$  and  $\sigma_g = 3.5$ . The beginning of the vote distribution presents a linear grow that makes it incompatible with all the models

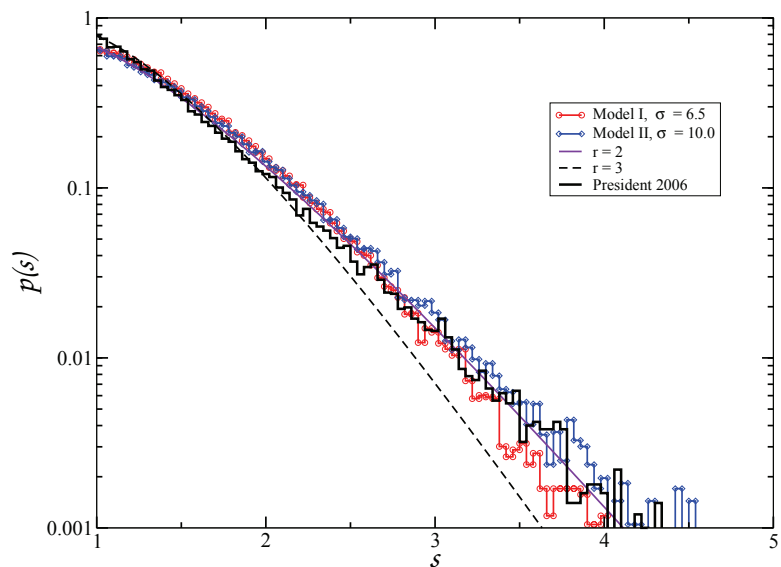


Fig. 9. Same as previous figure but in semilog scale. Notice that daisy of rank 3 does not fit the tail, it is better fitted by rank 2 daisy and for models I and II with the values indicated in the inset.

presented here. Notwithstanding such a result appears clearly as a deviation into the crude data reported previously in (Hernández-Saldaña, 2009). From the analysis it is clear the correct tail for electoral results requires some overlapping between the initial cities distribution. The model II, the Gaussian one, presents a better fitting in all the cases. The TSP models fails in to obtain correctly the maxima of the distribution. This fail is common in all the cases and that happens even in the case of daisy models. It is a truism that the maxima and the average in a probability distribution are not the same. Since the distribution is normalized a different value of position of maxima causes a different rate decay at the tail. Hence, a search on new TSP models which could present a different

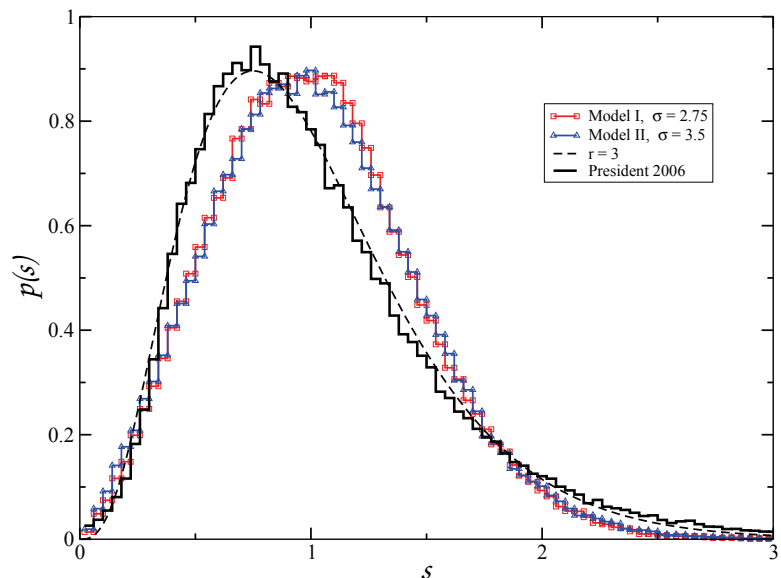


Fig. 10. Nearest neighbour distribution for the Presidential election in 2006 compared with models I and II and daisy models in order to fit the distribution at the beginning. The parameter values are indicated in the inset.

position of the maxima could give a better fitting of electoral data. A remarkable exception is Presidential case in 2006.

In reference to the long range correlation for the TSP models they grow linearly, but with a slope that slightly differs from daisy models. It is important to remark that long range correlations are extremely sensitive to unfolding procedure, but knowledge of longer correlations than first neighbour helps to understand the dynamics. Such has been the experience in quantum chaos and recently in the study of DNA sequences.

As a final point for this section is in reference to the reliability of electoral data. Even when we assume a fair play election and that the data are correctly collected and expressed, such a phenomenon is of a large complexity. One, which can be determined, is the self consistency in the data. In the Mexican 2006 election such a reliability is unclear (Báez et al., 2010) and must be taken into account when the analysis is performed. The 2006 data for the corporate party was particularly clear since they were, by far, the third position in the whole election. Many of the usual allies that sum their vote to this party were split with the two other large parties.

6. Conclusion

In this chapter we expose an application of TSP to a socio-political problem: explain the distribution of votes of a corporate party. The link between TSP and vote processes is made considering the distances between cities as the amount of votes received for a party in a cabin. In order to compare their statistical properties we perform a deconvolution process in order to separate the fluctuation from the average properties, such a procedure is called unfolding. The comparison was made for the nearest neighbour distribution of the unfolded variables. The corporate vote distribution for Mexican elections of 2000, 2003 and 2006 show that the distribution presents an exponential decay and start with a power law. These distributions are well described by a daisy model of rank  $r$  but there is not a clear interpretation for these models and the vote distribution. However, since the TSP with a cities distribution randomly selected is well described by a daisy model or frank 2, we decide to explore this option. To this end, we consider two TSP models with different initial cities distribution. We start with a square master lattice with a probability distribution centered in the intersection. The city

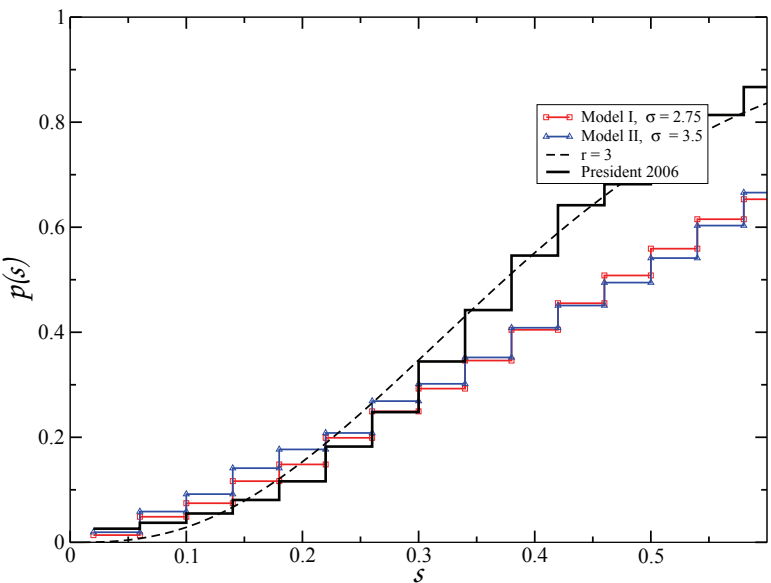


Fig. 11. Same as previous figure. The TSP models does not fit departing from  $s \approx 0.35$



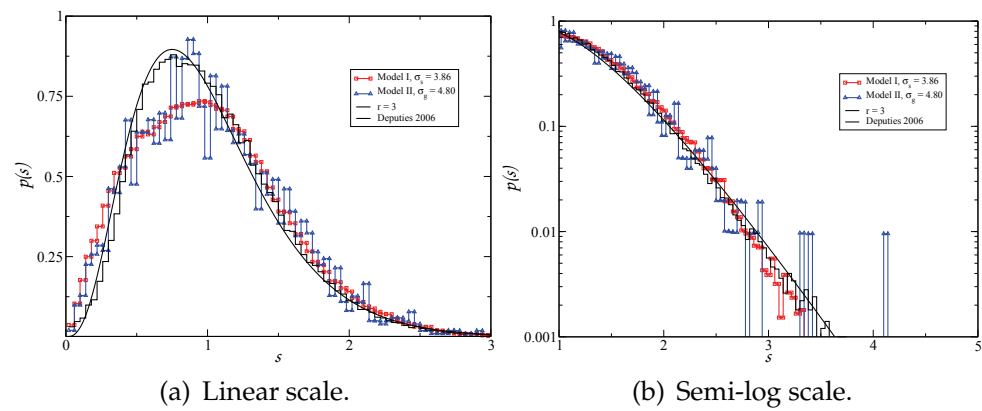


Fig. 12. Nearest neighbour distribution for the cases indicated in the inset. (a) Linear scale and (b) the semilog scale in order to show the behaviour at the tail.

location is selected from this probability distribution. We use two models, model I consists of an uniform distribution of total width  $2\sigma_s$  and, as model II, a Gaussian distribution of width  $\sigma_g$ . Both models present a transition from a distribution of cities, departing from a square lattice for  $\sigma = 0$  to a distribution of cities that resembles one selected randomly. For large enough width both models reproduce the rank 2 daisy model tail. With this two models we analyzed the electoral problem. The result obtained shows that the tail behavior could be described by such a models but the behaviour for small distances is not. The vote data best situated for analysis is the federal election in 2006. For the Presidential election of 2006, the extreme case, the decaying behaviour is well fitted by the models and the rank 2 daisy model. For an intermediate case, as it is represented by the Deputy and Senator elections in 2006 the model reproduces the tail by width values that are near from the value  $b/2$ , being  $b$  the size of the master square lattice. In all the vote distribution of 2006 the small distance behaviour is well described by our models with width of the order of  $(3/5)b$ . The problem in the description for small distances with our models is that they present a limit for the beginning of the distribution. This limit appears when the width of the models admits overlapping of the different sites in the master lattice. This limit appears not only in our model but in actual cities distribution( see figure 3). If this behaviour is universal or not it is matter of current research.

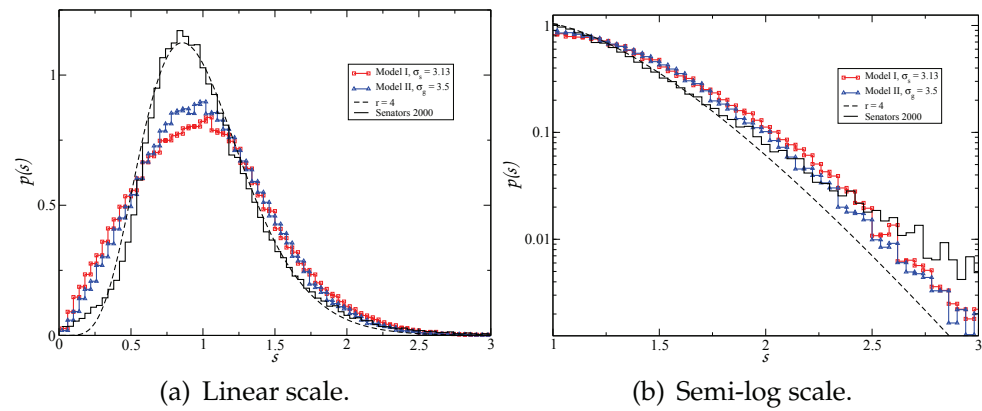


Fig. 13. Nearest neighbour distribution for the cases indicated in the inset. (a) Linear and (b) semilog scales in order to show the behaviour at the tail.

The existence of overlapping or not in the models becomes crucial in the description. This could be interpreted, in the vote case, as the corporate party having a similar distribution of voters in each cabin instead of single agents. In terms of a political description this means that agents of the party (the members of a national syndicate like the elementary school teachers, for instance) have a sphere of influence (with their relatives, for instance). This kind of behaviour has been observed, see (Crespo, 2008). How far this analogy goes is matter of current research and opens new questions about how we vote, beyond of the opinion models currently investigated by (Fortunato & Castellano, 2007), for instance. The peculiarities and deviations in the actual data can be taken into account. In the case of 2000 elections all the vote distributions presented a linear grow for small number of votes. This behaviour could not be explained by the models, this is a characteristic of such election. The electoral data itself can be noisy and subject to a lot of influences no related to fair-game. To this subject see (Crespo, 2008; Báez et al., 2010).

Even when the distribution of votes of the corporate party are better described by daisy models than the presented models of TSP, the latter offers a better dynamical insight of the voters behaviour. The search of new TSP models that present a better description of the actual data is open as well as models that reproduce the dynamics of the corporate vote. Human societies are complex, indeed, but simple models are discovering and enlightening features that could be explanation of the large consequences of our every day behaviour.

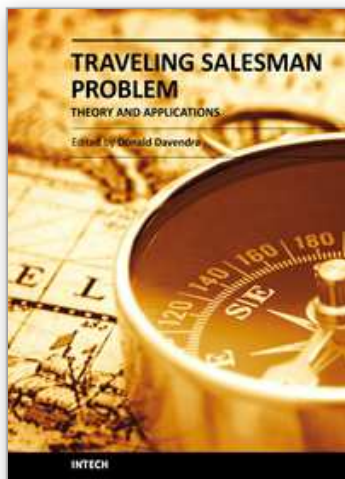
## 7. Acknowledgements

We thank support from PROMEP 2115/35621, Mexican agency.

## 8. References

- Applegate, D. L., Bixby, R. E., Chvatal, V. & Cook, W. (2006). *The Traveling Salesman Problem: A Computational Study*, Princeton Univ. Press, U.S.A.
- Araripe, L., Costa-Filho, R., Herrmann, H. & Andrade-Jr, J. (2006). Plurality voting: the statistical laws of democracy in Brazil, *Int. J. Mod. Phys. C* Vol. 17: 1809.
- Báez, G., Hernández-Saldaña, H. & Méndez-Sánchez, R. (2010). On the reliability of electoral processes: the mexican case. [arxiv:physics/0609114v2](https://arxiv.org/abs/physics/0609114v2).
- Ball, P. (2004). *Critical mass*, Farrar, Straus and Giroux, London U.K.
- Bernardes, A., Stauffer, D. & Kertész, J. (2002). Election results and the Sznajd model on Barabási network, *Eur. Phys. J. B* Vol. 25: 123–127.
- Bohigas, O. (1991). Random matrices and chaotic dynamics, in A. V. J. Z.-J. M.-J. Giannoni (ed.), *Chaos and Quantum physics*. Les Houches, session LII, 1989, North-Holland, Amsterdam, pp. 87–200.
- Borghesi, C. & Bouchaud, J.-P. (2010). Partial correlations in vote statistics: a diffusive field model for decision-making, *Eur. Phys. J. B* Vol. 75: 395404.
- Brody, T. A., Flores, J., French, J., Pandey, A. & Wong, S. S. M. (1981). Random-matrix physics: spectrum and strength fluctuations, *Rev Mod. Phys.* Vol. 53(3): 385–480.
- Crespo, J. A. (2008). 2006: *Hablan las actas. Las debilidades de la autoridad electoral mexicana*, Debate, Mexico.
- Filho, R. C., Almeida, M., Andrade, J. & Moreira, J. (1999). Scaling behavior in a proportional voting process, *Phys. Rev. E* Vol. 60: 1067–1068.
- Filho, R. C., Almeida, M. & Jr., J. A. (2003). Brazilian elections: voting for a scaling democracy, *Physica A* Vol. 322: 698–700.

- Fortunato, S. & Castellano, C. (2007). Scaling and universality in proportional elections, *Phys. Rev. Lett.* Vol. 99: 138701–138704.
- Guhr, T., Müller-Groeling, A. & Weidenmüller, H. (1998). Random matrix theories in quantum physics: Common concepts, *Phys. Rep.* Vol. 229: 189–425.
- Gutin, G. & Punnen, A. P. (2007). *The Traveling Salesman Problem and its variations*, Springer, New York.
- Hernández-Saldaña, H. (2009). On the corporate votes and their relation with daisy models, *Physica A* Vol. 388: 2699–2704.
- Hernández-Saldaña, H., Flores, J. & Seligman, T. (1998). Semi-poisson statistics and beyond, *Physical Review E* Vol 60: 449–452.
- Hernández-Saldaña, H., Gómez-Quezada, M. & Hernández-Zapata, E. (2010). Statistical distributions of quasi-optimal paths in the traveling salesman problem: the role of the initial distribution of cities, *Rev. Mex Phys. S* Accepted paper: 6 pages. [arxiv.org: \[cond-mat.stat-mech\] 1003.3913](http://arxiv.org: [cond-mat.stat-mech] 1003.3913).
- Instituto Federal Electoral, I. F. E. (2006). Cómputo distrital de las elecciones federales de 2006. [www.ife.org.mx](http://www.ife.org.mx).
- Lawler, E. L., Lenstra, J. K., Kan, A. H. G. R. & Shmoy, D. B. (1985). *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, Wiley, U.S.A.
- Lyra, M., Costa, U., Filho, R. C. & Andrade, J. (2003). Generalized Zipf's law in proportional voting processes, *Europhys. Lett.* Vol. 62: 131.
- Mantegna, R. N. & Stanley, H. E. (2000). *An introduction to Econophysics*, Cambridge Univ. Press, U.K.
- Mehta, M. L. (2004). *Random Matrices*, Elsevier, Amsterdam.
- Méndez-Sánchez, R., Valladares, A., Flores, J., Seligman, T. & Bohigas, O. (1996). Universal fluctuations of quasi-optimal paths of the travelling salesman problem, *Physica A* Vol.232: 554–562.
- Morales-Matamoros, O., Martínez-Cruz, M. & Tejeida-Padilla, R. (2006). Mexican voter network as a dynamic complex system, *50th Annual Meeting of the ISSS*.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical Recipes. The art of scientific computation*, Cambridge Univ. Press, U.K.
- Sadovsky, M. & Gliskov, A. (2007). Towards the typology of elections at Russia. [arXiv:0706.3521 \[physics.soc-ph\]](http://arXiv:0706.3521 [physics.soc-ph]).
- Sinha, S. & Pan, R. (2006). How a hit is born: the emergence of popularity from the dynamics of collective choice, in A. C. B.K. Chakrabarti, A. Chakraborti (ed.), *Econophysics and Sociophysics: Trends and Perspectives*, Wiley-VCH, Berlin, pp. 417–447. [arXiv:0704.2955 \[physics.soc-ph\]](http://arXiv:0704.2955 [physics.soc-ph]).
- Situngkir, H. (2004). Power-law signature in Indonesian legislative election 1999 and 2004. [arXiv:nlin/0405002](http://arXiv:nlin/0405002).



## **Traveling Salesman Problem, Theory and Applications**

Edited by Prof. Donald Davendra

ISBN 978-953-307-426-9

Hard cover, 298 pages

**Publisher** InTech

**Published online** 30, November, 2010

**Published in print edition** November, 2010

This book is a collection of current research in the application of evolutionary algorithms and other optimal algorithms to solving the TSP problem. It brings together researchers with applications in Artificial Immune Systems, Genetic Algorithms, Neural Networks and Differential Evolution Algorithm. Hybrid systems, like Fuzzy Maps, Chaotic Maps and Parallelized TSP are also presented. Most importantly, this book presents both theoretical as well as practical applications of TSP, which will be a vital tool for researchers and graduate entry students in the field of applied Mathematics, Computing Science and Engineering.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hugo Hernandez-Saldana (2010). A Sociophysical Application of TSP: the Corporate Vote, Traveling Salesman Problem, Theory and Applications, Prof. Donald Davendra (Ed.), ISBN: 978-953-307-426-9, InTech, Available from: <http://www.intechopen.com/books/traveling-salesman-problem-theory-and-applications/a-sociophysical-application-of-tsp-the-corporate-vote->

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen