UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA
DOUTORADO EM FÍSICA

LUDWING FERNEY MARENCO CAMACHO

ON THE DESCRIPTION OF LEGISLATIVE CONSENSUAL REGIMES BASED EXTENDED BOLTZMANN MACHINE AND NEAREST CORRELATED CLUSTERS ALGORITHM

FORTALEZA
2023

LUDWING FERNEY MARENCO CAMACHO

ON THE DESCRIPTION OF LEGISLATIVE CONSENSUAL REGIMES BASED EXTENDED BOLTZMANN MACHINE AND NEAREST CORRELATED CLUSTERS ALGORITHM

Tese apresentada ao Programa de Pós-Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Física. Área de Concentração: Física da Matéria Condensada.

Orientador: Prof. Dr. Carlos Lenz Cesar

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

M279o Marenco Camacho, Ludwing Ferney.
On the description of legislative consensual regimes based Extended Boltzmann Machine and nearest correlated clusters algorithm / Ludwing Ferney Marenco Camacho. – 2023.
142 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Física , Fortaleza, 2023.
Orientação: Prof. Dr. Carlos Lenz Cesar.

1. Regime de consenso legislativo. 2. Máquina de Boltzmann estendida. 3. Nearest correlated cluster algorithm. I. Título.

CDD 530

LUDWING FERNEY MARENCO CAMACHO

ON THE DESCRIPTION OF LEGISLATIVE CONSENSUAL REGIMES BASED EXTENDED BOLTZMANN MACHINE AND NEAREST CORRELATED CLUSTERS ALGORITHM

Tese apresentada ao Programa de Pós-Graduação em Física do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Física. Área de Concentração: Física da Matéria Condensada.

Aprovada em: 16/06/2023.

BANCA EXAMINADORA

Prof. Dr. Carlos Lenz Cesar (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. André Auto Moreira
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Nepomuceno Filho
Universidade Federal do Ceará (UFC)

Prof. Dr. Rilder de Sousa Pires
Universidade de Fortaleza (UNIFOR)

Prof. Dr. Jose Maria Ferreira Jardim da Silveira
Universidade Estadual de Campinas (UNICAMP)

To Maria for giving the light of life and to Carla for keeping it alive

ACKNOWLEDGEMENTS

Quiero agradecer al Prof. Dr. Carlos Lenz por el fructífero tiempo que dedicamos a divertirnos con este trabajo. A mi familia y a mis amigos en Colombia por el apoyo incondicional. A Carla por la paciencia durante todos estos años. A los muchachos del grupo de sistemas complejos de la universidad, en especial a Israel y Marciel por las risas inmortales. Al Prof. Dr. Diego Mauricio Gallego Mahecha por sus acertados consejos cuando era una semilla queriendo crecer y finalmente, al Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por el apoyo financiero.

ABSTRACT

A new effective method to analyze Legislative Systems is presented. Methodology is based on two main approaches: The interactionism and the statistical. For the first approach, a set of algorithms to gather the interaction parameters of a like spin glass system was developed. Using the Linear Response Approximation, interaction and local field values are analytically obtained. In addition, for the first time, an algorithm to estimate the temperature based on the iterative scaling of the partition function is exposed. Set of procedures are condensed in the Extended Boltzmann Machine. In the statistical approach, an algorithm for clustering data based on the maximization of its correlation is presented. Procedure employs a like percolation process to compute first and second giant components of a complete connected network. Clusters emerge by analyzing plateaus of giant components and can be visualized on the Minimal Spanning Tree (MST) ordered correlation matrix. Procedure is named as the Nearest Correlated Clusters Algorithm. Available free roll-call vote data for three Legislative Lower Houses were acquired. Specifically, roll-call vote data collected are from the United States Houses of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil. By extracting the political parties' majority opinion matrix and using it in the interactionism approach, consensual and dissensual Legislative zones appear. These zones are gathered by comparing the average political parties' opinion with the degree of political interaction in which the transition from dissensus to consensus happened. On the other hand, using the Lower House members' roll-call vote data into the statistical approach, consensual and dissensual Legislative states emerge. These states are characterized by analyzing the time evolution of MST ordered correlation matrices and its probability distribution function. By joining results of both approaches, Legislative consensual regimes are proposed. This methodology can be used to understand profoundly, collective behavior in Legislative systems and to foresee political storms.

Keywords: legislative consensual regimes; extended Boltzmann machine; nearest correlated cluster algorithm.

RESUMO

Apresenta-se um novo método para analisar Sistemas Legislativos. A metodologia é baseada em duas abordagens: A baseada em interações e a estatística. Para a primeira, foram desenvolvidos um conjunto de algoritmos para obter os parâmetros de interação de um sistema semelhante ao spin glass. Usando a Aproximação de Resposta Linear, expressões analíticas para os valores de interação e campo local são apresentadas. Junto com isso, pela primeira vez é exposto um algoritmo para estimar a temperatura baseada no escalamento iterativo da função de partição. O conjunto de algoritmos são condensados na Máquina de Boltzmann Estendida. Na abordagem estatística, mostra-se um algoritmo para clusterizar os dados baseado na maximização da sua correlação. Este procedimento usa um processo semelhante à percolação para obter a primeira e segunda componente gigante de uma rede completamente conectada. Esse algoritmo é nomeado como o Algoritmo dos Clusters mais Correlacionados. Dados de votação nominal disponíveis e de graça para três casas baixas Legislativas foram coletados. Em específico, foram coletados dados de votação nominal da Casa de Representantes dos Estados Unidos, a Casa dos Comuns do Reino Unido e a Câmara dos Deputados do Brasil. Extraindo a matriz de posição política dos partidos e usando ela dentro da abordagem baseada em interações, zonas de consenso e dissenso legislativo aparecem. Essas zonas aparecem comparando a opinião média dos partidos políticos junto com o valor de interação política necessária para que uma transição de dissenso para consenso aconteça. Por outra parte, usando os dados de votação nominal dos membros das casas baixas dentro da abordagem estatística, estados de consenso e dissenso são obtidos. Esses estados são caracterizados pela análise temporal das matrizes de correlação organizada pelo MST e a sua distribuição de probabilidade. Juntando os resultados das duas abordagens, os regimes de consenso Legislativo são propostos. Essa metodologia pode ser usada para entender profundamente as características de comportamento coletivo dos sistemas Legislativos e pode prever tormentas políticas.

Palavras-chave: regime de consenso legislativo; maquina de Boltzmann estendida; nearest correlated cluster algorithm.

LIST OF FIGURES

Figure 1 – General overview 15
Figure 2 – Boltzmann machine with visible spins 24
Figure 3 – Flowchart for procedure 1 42
Figure 4 – Flowchart for procedure 2 46
Figure 5 – Probability distribution function for observed parameters $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$ . . 47
Figure 6 – Comparison of quantities computed using observed and inferred parameters 49
Figure 7 – Discrepancy among observed and inferred quantities in function of β_{O} . 50
Figure 8 – Discrepancy of parameters in function of β_{O} 51
Figure 9 – Linear fitting between inferred and observed parameters 52
Figure 10 – Average angular coefficients in function of β_{O} 53
Figure 11 – Scaled discrepancy of parameters in function of β_{O} 53
Figure 12 – Dependence of parameter discrepancy with respect to β 54
Figure 13 – Average minimum of scaled discrepancy in function of β_{O} 55
Figure 14 – Discrepancy of expected value $\langle \sigma_i \rangle$ as a function of β 56
Figure 15 – Flowchart for procedure 3 59
Figure 16 – Convergence condition of procedure 3 60
Figure 17 – Comparison between $\hat{\beta}$ and β_{O} obtained using proposed procedure . . . 61
Figure 18 – Final comparison for $\hat{\beta}$ 61
Figure 19 – Partition function characterization of observed and inferred parameters. 63
Figure 20 – Average error value of maximum values of $C(\mathbb{W}, \mathbb{H}, \beta)$. 64
Figure 21 – Comparison of observed and predicted data. 65
Figure 22 – Input-output success prediction of Boltzmann machine 66
Figure 23 – Legislative periods to be studied 68
Figure 24 – Quantity of bills voted in Lower Houses databases 69
Figure 25 – Pictorial representation of Lower House datasets 70
Figure 26 – Names of Chief Executive power for each Legislative database 70
Figure 27 – Quantity of voting Members in Lower House databases 71
Figure 28 – Flowchart for procedure 4 75
Figure 29 – Quantity of political parties in Lower Houses 76
Figure 30 – Flowchart for procedure 5 79
Figure 31 – Flowchart for procedure 6 83

Figure 32 – Normalized giant components 84
Figure 33 – NECO cluster identification 85
Figure 34 – Expected value of input quantities for learning process. 87
Figure 35 – Probability distribution function for learned parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ 89
Figure 36 – Significance of $\hat{\beta}$ in Legislative process 90
Figure 37 – $\hat{\beta}$ learned in Lower Houses Databases. 92
Figure 38 – Consensual and dissensual states regions. 94
Figure 39 – Consensual and dissensual zones for the three Lower Houses. 95
Figure 40 – Characterization of $Z_B\left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$ 98
Figure 41 – Maximum values of $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ and $\beta_{\max}\left(C(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ 99
Figure 42 – Correlation matrices for the United States House of Representatives . . 103
Figure 43 – Correlation matrices for the House of Commons of the United Kingdom 104
Figure 44 – Correlation matrices for the Chamber of Deputies of Brazil 105
Figure 45 – Correlation matrices colored by political parties for the United States House of Representatives 107
Figure 46 – Correlation matrices colored by political parties for the House of Commons of the United kingdom 108
Figure 47 – Correlation matrices colored by political parties for the Chamber of Deputies of Brazil 109
Figure 48 – Ordered correlation matrices colored by learned clusters for the United States House of Representatives. 111
Figure 49 – Ordered correlation matrices colored by learned clusters for the House of Commons of the United Kingdom 112
Figure 50 – Ordered correlation matrices colored by learned clusters for the Chamber of Deputies of Brazil 113
Figure 51 – Cluster learned by NECO algorithm 115
Figure 52 – Probability distribution function for correlation matrices for the United States House of Representatives 116
Figure 53 – Probability distribution function for correlation matrices for the House of Commons of the United Kingdom 117
Figure 54 – Probability distribution function for correlation matrices for the Chamber of Deputies of Brazil 118

Figure 55 – Political states for correlation probability distributions. 119
Figure 56 – Cumulative distribution function for correlation matrices for the United States House of Representatives 121
Figure 57 – Cumulative distribution function for correlation matrices for the House of Commons of the United Kingdom 122
Figure 58 – Cumulative distribution function for correlation matrices for the Chamber of Deputies of Brazil 123
Figure 59 – Bimodal index for Lower Houses. 129
Figure 60 – Legislative consensual-dissensual regimes 131
Figure 61 – Legislative consensual regimes for Lower Houses 132

LIST OF TABLES

Table 1 – Expected possibilitites for three spins system 25
Table 2 – Identified vote options in Lower House databases 69
Table 3 – Fitted parameters for cumulative distribution function for the United States House of Representatives 125
Table 4 – Fitted parameters for cumulative distribution function for the House of Commons of the United Kingdom 126
Table 5 – Fitted parameters for cumulative distribution function for the Chamber of Deputies of Brazil 127

LIST OF ALGORITHMS

Algorithm 1 – Boltzmann machine algorithm using LRA. 41
Algorithm 2 – Boltzmann machine algorithm using LRA and Diagonal Trick. . . 45
Algorithm 3 – Boltzmann machine algorithm for learning $\hat{\beta}$ 57
Algorithm 4 – Political parties' majority opinion algorithm. 74
Algorithm 5 – Political party majority random data 77
Algorithm 6 – Nearest correlated cluster algorithm (NECO) 81

CONTENTS

1	**INTRODUCTION**	14
2	**REVIEW ON STATISTICAL PHYSICS**	17
2.1	**Information Theory – Shannon/Boltzmann Entropy**	17
2.2	**Joint Entropy**	17
2.3	**Thermodynamic Ensembles**	18
2.3.1	***Microcanonical Ensemble***	18
2.3.2	***Canonical Ensemble and Helmholtz Free Energy***	19
2.3.3	***P-T Ensemble and Gibbs Free Energy***	20
2.4	**Thermodynamics of magnetic systems**	22
2.4.1	***Ising Model***	22
3	**THEORETICAL ASPECTS**	23
3.1	**Analogy to a bill voting system**	23
3.2	**Formulation of the Boltzmann Machine Learning Problem**	23
3.2.1	***Partition Functions***	26
3.3	**Learning Algorithm using Linear Response Approximation and Bethe Free Energy**	29
3.3.1	***Introducing Bethe free energy***	37
3.4	**Improvement of Learning Algorithm using Diagonal Trick**	43
3.5	**Checking if the Learning Algorithm is Working Properly**	47
3.6	**Learning β: A hidden feature in Boltzmann machines**	51
4	**DATA MINING AND MANIPULATION**	67
4.1	**Setting up the data**	67
4.2	**Legislative Lower House Databases**	68
4.3	**Lower House Databases Manipulation**	71
4.4	**Creating political party majority opinion vote matrix**	73
4.5	**Generating political party majority random data**	76
4.6	**Nearest correlated cluster algorithm (NECO)**	79
5	**RESULTS AND ANALYSIS**	86
5.1	**Input quantities for extended Boltzmann machine learning process**	86
5.2	**Learned parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$: From statistics to interactions**	88

5.3 $\hat{\beta}$ **learned: A new way for measuring the degree of political interaction** 89
5.4 **Consensual and Dissensual zones: Analysing** $\hat{\beta}$ **and** $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ . . 94
5.5 **Finding** $\hat{\beta}_{\max}$**: An estimator to know the transition from dissensual to consensual state** 96
5.6 **Lower Houses Members analysis via nearest correlated cluster algorithm** 101
5.7 **Lower Houses members correlation distributions** 115
5.8 **Bimodality index: Another way to characterize dissensual and consensual states** 128
5.9 **Legislative consensual-dissensual regimes** 130
6 **CONCLUSIONS** 134
BIBLIOGRAPHY 136
APPENDIX A – Axioms of the information theory 139
APPENDIX B – Maximizing entropy 141
APPENDIX C – Inequality relation for entropy 142

1 INTRODUCTION

Legislative systems deals and mediate over the best and most adequate government proposals that determines the social and economic development in democratic societies. Under the premise of being elected by citizens, Legislative Representatives must be able for listening society necessities, search a solution for those demands and in the ideal fashion, ensures its effective application. All this is done in a well-established flow of legislative activity, beginning with the proposal of bills and ending in its discussion in plenary sessions. Bills are discussed in order to made amendments to specific issues that ensures that if a bill is accepted to become a law, it is because its content has been carefully checked and agreed by all Representatives not matter its political position. However, Legislative activity is full of political features that given its complexity, is hard to being identified and analysed in a quantitative way. Besides that, despite some attempts to capture these political features using statistical approximation based network science based [1, 2] and machine learning framework [3], a complete method mixing statistical and numerical approach is still unknown.

In this work it is proposed for the first time, the most complete method for analyse emergence political collective behaviour. Our method analyses the available free information of roll-call vote data. Specifically, we use roll-call vote data related to three Lower Houses: The United States House of Representatives, The House of Commons of the United Kingdom and the Chamber of Deputies of Brazil. In order to analyse roll-call vote data, proposed method relies on two different approximations namely, the interactionism and the statistical.

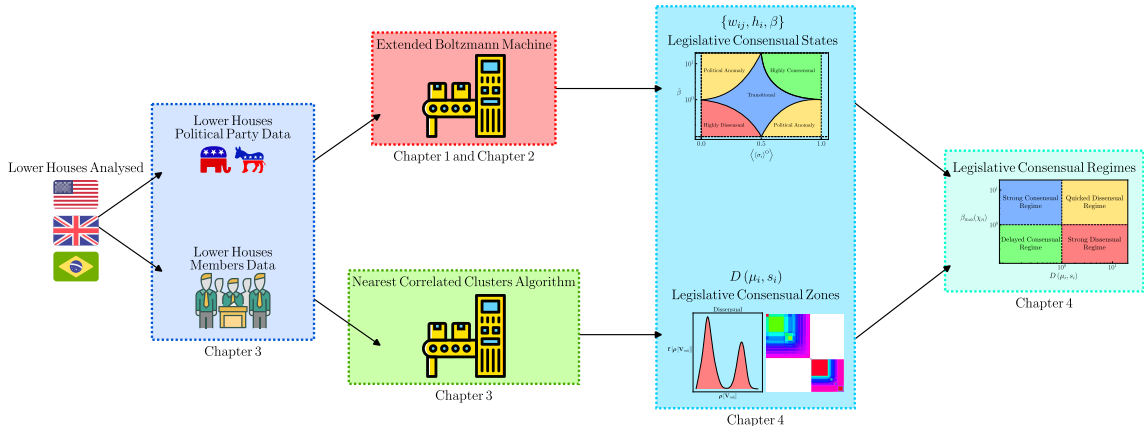
In the interactionism approach, political parties' roll-call vote data is used to extract a set of parameters $\{w_{ij}, h_i\}$ used to emulate the dynamics of like spin glass lattice system. In this case, spins in the lattice will be associated to political parties. Extraction of parameters is carried out using a Boltzmann machine in which parameters are derived by finding its analytical expression using Linear Response Approximation and Bethe Free Energy. Besides that, for the best of our knowledge, an extension in the Boltzmann Machine to obtain inverse of temperature parameter β is presented. This extension proposes a new iteratively algorithm able to estimate β, using $\{w_{ij}, h_i\}$ already gathered. This method is named as the Extended Boltzmann Machine. In the political context, $\{w_{ij}, h_i, \beta\}$ will be associated to the political parties' pairwise interaction value,

the capacity for a political party to maintain its political position and the average value of political interaction among all political parties, respectively.

In the statistical approach, it is analyse Lower Houses Members roll-call vote data. Lower Houses Members are Representatives, Commons and Deputies for the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively. Analysis of Lower Houses Members roll-call vote data is carried using method presented in [1] which reorder Lower House Members correlation matrix based on the result of Minimal Spanning Tree (MST) technique. However, in this work, for the first time, an extension of this method to obtain clusters of Lower Houses Members only taking into account its correlation matrix is exposed. Clusters are obtained from a like percolation algorithm. This method is named as the Nearest Correlated Cluster Algorithm.

Figure 1 exposes a summarized plot of both approximations. Upper part of plot is dedicated to the interactionism approach whereas lower part is for the statistical approach.

Figure 1 – General overview of the content of this work. In Chapter 1 and 2 is developed the theoretical aspects related to the extended Boltzmann Machine. In Chapter 3 is exposed the tranformation of Lower Houses data and the explanation concerned to the Nearest Correlated Clusters Algorithm. Finally, Chapter 4 is dedicated to the results found using the interactionism and statistical approach.



Source: Author.

Taking into account Figure 1, this work is divided as follows. Chapter 1 and 2 are entirely dedicated to the theoretical aspects related to the extended Boltzmann Machine. Chapter 3 shows the procedures used for gathering Lower Houses roll-call vote data, its processing and transformation along all simulated data created to compare results and the Nearest Correlated Cluster Algorithm. Chapter 4 shows the results obtained from the two aforementioned approximations. In this chapter, we begin by exposing results

coming from the Extended Boltzmann Machine and subsequently, results of the Nearest Correlated Cluster Algorithm are uncovered. Results obtained from both approaches are condensed in the Legislative Consensual Regimes plot. Finally, last Chapter is dedicated to the conclusion learned.

2 REVIEW ON STATISTICAL PHYSICS

It is presented a small review of the most important concepts of Statistical Physics mandatory to understand the analogy with voting systems.

2.1 Information Theory – Shannon/Boltzmann Entropy

Shannon extracted his information entropy under only four axioms on a probability set $\vec{p} = (p_1, p_2, \cdots, p_n) \quad \rightarrow \quad p_i \geq 0 \quad \rightarrow \quad \sum\limits_i p_i = 1$ given by $S = -c \sum\limits_i p_i \log_a p_i \quad \rightarrow \quad a > 1,\ c > 1$, (See Appendix A). Von Neumann pointed out to Shannon that the entropy Shannon found had the same functional form as the Boltzmann thermodynamic entropy, therefore it should be also called an Entropy. The maximization of entropy is a well-known process to find probability distributions. For example, if we find the maximum of $S = -c \sum\limits_i p_i \log_a p_i$with the only constraint that $\sum\limits_i p_i = 1$ we obtain equal probabilities $p_i = \frac{1}{n}$ (See Appendix B). The equal probabilities ensemble is called Microcanonical Ensemble. When Boltzmann solved the ideal gas problem in this ensemble he discovered that the constants must be given by $c = k_B$ and $a = e$, therefore $S = -k_B \sum\limits_i p_i \ln p_i$, for the theory to agree with ideal gas law $PV = nRT$. The Boltzmann constant k_B is just $k_B = \frac{R}{N_{\text{av}}}$ the universal gas constant R divided by the Avogadro number N_{av}.

2.2 Joint Entropy

Let x and y be two random variables, and $p(x_i, x_j) = \text{Prob}[x = x_i \,\text{and}\, y = y_j]$. The joint entropy is given by $S(x, y) = -c \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p(x_i, x_j) \log_a p(x_i, x_j)$, while the marginal entropies are given by:

$$\begin{aligned} S(x) &= -c \sum_{i=1}^{n} p(x_i) \log_a p(x_i) \quad \rightarrow \quad p_x(x_i) = \sum_{j=1}^{m} p(x_i, y_j,), \\ S(y) &= -c \sum_{i=1}^{n} p(x_i) \log_a p(x_i) \quad \rightarrow \quad p(y_j) = \sum_{i=1}^{n} p(x_i, y_j). \end{aligned} \tag{2.1}$$

These quantities obey the following inequalities equations (See Appendix C):

1. Let $\vec{q} \neq \vec{p} \quad \rightarrow \quad q_i \geq 0\ \forall i \quad \rightarrow \quad \sum\limits_i q_i = 1$ then:

$$-c \sum_{i=1}^{n} p_i \ln p_i \leq -c \sum_{i=1}^{n} p_i \ln q_i, \tag{2.2}$$

or

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \geq 0; \tag{2.3}$$

2. $S(x) + S(y) \geq S(x, y)$ known as the subadditivity of entropies.

Taking into account that, it is possible to define the **Cross Entropy** as

$$S(\vec{p}, \vec{q}) = -c \sum_{i=1}^{n} p_i \log_a q_i, \tag{2.4}$$

and the **Kullback–Leibler divergence**, or relative entropy as:

$$D_{\mathrm{KL}}(\vec{p}, \vec{q}) = -c \sum_{i=1}^{n} p_i \log_a \left(\frac{q_i}{p_i} \right) = c \sum_{i=1}^{n} p_i \log_a \left(\frac{p_i}{q_i} \right). \tag{2.5}$$

Note that $D_{\mathrm{KL}}(\vec{p}, \vec{q} = \vec{p}) = 0$, therefore, inferences over a probability distribution $\vec{p}$ expected to be similar to a known distribution $\vec{q}$ can be performed by finding the minimum of $D_{\mathrm{KL}}(\vec{p}, \vec{q})$.

2.3 Thermodynamic Ensembles

2.3.1 Microcanonical Ensemble

For a gas system with extensive variables (U, V, N) the intensive variables at thermodynamic equilibrium are given by the maximization of Boltzmann entropy with only the sum of probabilities constraint. From thermodynamics we know that $dU = TdS - PdV + \mu dN$ from which we obtain:

$$dS = \frac{1}{T} dU + \frac{P}{T} dV - \frac{\mu}{T} dN \quad \rightarrow \quad \frac{\partial S}{\partial U} = \frac{1}{T} \quad \rightarrow \quad \frac{\partial S}{\partial V} = \frac{P}{T} \quad \rightarrow \quad \frac{\partial S}{\partial N} = -\frac{\mu}{T}. \tag{2.6}$$

This problem can be written as a maximization problem

$$\begin{aligned}
\max S(U,V,N) = -c\sum_i p_i \ln p_i \quad &\rightarrow \quad \text{sc} \quad \sum_i p_i = 1, \\
\mathcal{L} &= -c\sum_i p_i \ln p_i - \lambda\left(\sum_i p_i - 1\right), \\
\frac{\partial \mathcal{L}}{\partial p_j} &= -c\ln p_j - c - \lambda = 0, \\
p_j &= p_o = \frac{1}{\Omega}, \\
\Omega(U,V,N) &= \text{number of states compatible with } U,V,N, \\
S &= -c\sum_i p_o \ln p_o = c\ln\Omega.
\end{aligned} \tag{2.7}$$

For the ideal gas the energy depends only on the velocities, but not the position, therefore the number of states is proportional to $\Omega = V^N \omega(U,N)$. The entropy $S = k_B \ln\left[V^N \omega(U,N)\right] = Nc\ln V + c\ln\omega(U,N) \quad \rightarrow \quad \frac{\partial S}{\partial V} = \frac{Nc}{V} = \frac{P}{T} \quad \rightarrow \quad PV = NcT = n(N_{\text{av}}c)T$. To be compatible with ideal gas law $PV = nRT \quad \rightarrow \quad c = k_B = \frac{R}{N_{av}}$ and the Boltzmann entropy becomes:

$$S = -k_B \sum_i p_i \ln p_i. \tag{2.8}$$

For Boltzmann, however, the logarithm is the Neperian one, and the connection with the ideal gas law of thermodynamics require it to be multiplied by the Boltzmann constant, which is the Avogadro number multiplied by the universal gas constant.

2.3.2 Canonical Ensemble and Helmholtz Free Energy

Now let the gas be in contact with an infinite source of temperature and kept at a constant temperature T. Now we know (T,V,N) instead of (U,V,N) and the internal energy U is obtained by the expected value $\sum_i p_i \varepsilon_i = U$ where ε_i are the energy of each

state, or configuration. This case calls, then, for the optimization problem of the form:

$$\begin{aligned}
\max S(U,V,N) &= -k_B \sum_i p_i \ln p_i \quad \rightarrow \quad \text{sc} \quad \sum_i p_i = 1 \quad and \quad \sum_i p_i \varepsilon_i = U, \\
\mathcal{L} &= -k_B \sum_i p_i \ln p_i - \lambda_1 \left(\sum_i p_i - 1 \right) - \lambda_2 \left(\sum_i p_i \varepsilon_i - U \right), \\
\frac{\partial \mathcal{L}}{\partial p_j} &= -k_B \ln p_j - c k_B - \lambda_1 - \lambda_2 \varepsilon_j = 0 \quad \rightarrow \quad p_j = p_o e^{-\frac{\lambda_2 \varepsilon_j}{k_B}}, \\
\sum_i p_i &= p_o \sum_i e^{-\frac{\lambda_2 \varepsilon_i}{k_B}} = 1 \quad \rightarrow \quad p_o = \frac{1}{\sum_i e^{-\frac{\lambda_2 \varepsilon_i}{k_B}}}, \\
S &= -k_B \sum_i p_i \ln \left(p_o e^{-\frac{\lambda_2 \varepsilon_i}{k_B}} \right), \\
&= -k_B \ln p_o \sum_i p_i + \lambda_2 \sum_i p_i \varepsilon_i = -k_B \ln p_o + \lambda_2 U, \\
\frac{\partial S}{\partial U} &= \lambda_2 = \frac{1}{T} \quad \rightarrow \quad p_o = \frac{1}{\sum_i e^{e^{-\frac{\varepsilon_i}{k_B T}}}} = \frac{1}{Z_B} \quad \rightarrow \quad Z_B = \sum_i e^{e^{-\frac{\varepsilon_i}{k_B T}}},
\end{aligned} \tag{2.9}$$

where the partition function is given by $Z_B = \sum_i e^{e^{-\frac{\varepsilon_i}{k_B T}}}$. In terms of thermodynamic potential the entropy $S(U,V,N)$ must be changed to a function of $F(T,V,N)$, the Helmholtz Free Energy, obtained by the Legendre Transformation:

$$\begin{aligned}
F = U - TS \quad &\rightarrow \quad dF = dU - TdS - SdT = -SdT - PdV + \mu dN, \\
\frac{\partial F}{\partial T} = -S \quad &\rightarrow \quad \frac{\partial F}{\partial V} = -P \quad \rightarrow \quad \frac{\partial F}{\partial N} = \mu.
\end{aligned} \tag{2.10}$$

From which we obtained the Helmholtz Free Energy as:

$$\begin{aligned}
p_i &= p_o e^{-\frac{\varepsilon_i}{k_B T}}, \\
S &= -k_B \sum_i p_i \ln \left(p_o e^{-\frac{\varepsilon_i}{k_B T}} \right), \\
&= -k_B \ln p_o + \frac{1}{T} \sum_i p_i \varepsilon_i = -k_B \ln p_o + \frac{1}{T} U, \\
F &= U - TS = U - U + k_B T \ln p_o = -k_B T \ln Z_B, \\
Z_B &= e^{-\frac{F}{k_B T}} = \sum_i e^{-\frac{\varepsilon_i}{k_B T}}.
\end{aligned} \tag{2.11}$$

2.3.3 P-T Ensemble and Gibbs Free Energy

Now the gas is in contact with a thermal reservoir at temperature T and a piston to control the pressure always at P. The volume is given by the expected value

$\sum_i p_i V_i = V$. The optimization problem becomes:

$$
\begin{aligned}
&\max\, S(U,V,N) = -k_B \sum_i p_i \ln p_i \;\rightarrow\; \text{sc} \sum_i p_i = 1 \;\rightarrow\; \sum_i p_i \varepsilon_i = U \;\rightarrow\; \sum_i p_i V_i = V,\\
&\mathcal{L} = -k_B \sum_i p_i \ln p_i - \lambda_1 \left(\sum_i p_i - 1 \right) - \lambda_2 \left(\sum_i p_i \varepsilon_i - U \right) - \lambda_3 \left(\sum_i p_i V_i - V \right),\\
&\frac{\partial \mathcal{L}}{\partial p_j} = -k_B \ln p_j - c k_B - \lambda_1 - \lambda_2 \varepsilon_j - \lambda_3 V_j = 0 \quad \rightarrow \quad p_j = p_o e^{-\frac{\lambda_2 \varepsilon_j + \lambda_3 V_j}{k_B}},\\
&\sum_i p_i = p_o \sum_i e^{-\frac{\lambda_2 \varepsilon_i}{k_B}} = 1 \quad \rightarrow \quad p_o = \frac{1}{\sum_i e^{-\frac{\lambda_2 \varepsilon_j + \lambda_3 V_j}{k_B}}},\\
&S = -k_B \sum_i q_i \ln \left(q_o e^{-\frac{\lambda_2 \varepsilon_i + \lambda_3 V_i}{k_B}} \right),\\
&= -k_B \ln p_o + \sum_i \lambda_2 \varepsilon_i + \sum_i \lambda_3 V_i,\\
&= -k_B \ln p_o + \lambda_2 U + \lambda_3 V,\\
&\frac{\partial S}{\partial U} = \lambda_2 = \frac{1}{T} \quad \rightarrow \quad \frac{\partial S}{\partial V} = \lambda_3 = \frac{P}{T},\\
&p_o = \frac{1}{\sum_i e^{e^{-\frac{\varepsilon_i + P V_i}{k_B T}}}} = \frac{1}{Z_G},\\
&Z_G = \sum_i e^{e^{-\frac{\varepsilon_i + P V_i}{k_B T}}}.
\end{aligned}
\tag{2.12}
$$

Now the Free Energy must exchange $U \rightarrow T$ and $V \rightarrow P$ therefore the Gibbs Free Energy is given by the Legendre transformation $G = U - TS + PV = F + PV$:

$$
dG = -S dT + V dP + \mu dN \quad \rightarrow \quad \frac{\partial G}{\partial T} = -S \quad \rightarrow \quad \frac{\partial G}{\partial P} = V \quad \rightarrow \quad \frac{\partial G}{\partial N} = \mu. \tag{2.13}
$$

$$
\begin{aligned}
S &= -k_B \sum_i p_i \ln \left(p_o e^{-\frac{\varepsilon_i + P V_i}{k_B T}} \right),\\
&= -k_B \ln p_o + \frac{1}{T} \sum_i p_i \varepsilon_i + \frac{P}{T} \sum_i p_i V_i = -k_B \ln p_o + \frac{1}{T} U + \frac{P}{T} V,\\
G &= U - TS + PV = U - T \left(-k_B \ln p_o + \frac{1}{T} U + \frac{P}{T} V \right) + PV,\\
&= k_B T \ln p_o = -k_B T \ln Z_G,\\
G &= -k_B T \ln Z_G \quad \rightarrow \quad Z_G = e^{-\frac{G}{k_B T}} = \sum_i e^{-\frac{\varepsilon_i + P V_i}{k_B T}}.
\end{aligned}
\tag{2.14}
$$

2.4 Thermodynamics of magnetic systems

For the magnetic system with a fixed number of dipoles the pressure, volume and number of dipoles do not play any role and the work must be changed to $W = -PdV \quad \rightarrow \quad W = -\mu H dm = -hdm$. Therefore, all we have to do is to perform the substitutions:

$$\begin{aligned} &P \rightarrow h \quad \text{and} \quad V \rightarrow m, \\ &\frac{\partial F}{\partial T} = -S \quad \rightarrow \quad \frac{\partial F}{\partial m_i} = -h_i \quad \rightarrow \quad Z_B = \sum_i e^{-\frac{\varepsilon_i}{k_B T}} \quad \rightarrow \quad F = -k_B T \ln Z_B, \\ &\frac{\partial G}{\partial T} = -S \quad \rightarrow \quad \frac{\partial G}{\partial h_i} = m_i \quad \rightarrow \quad Z_G = \sum_i e^{-\frac{\varepsilon_i + h_i m_i}{k_B T}} \quad \rightarrow \quad G = -k_B T \ln Z_G. \end{aligned} \tag{2.15}$$

2.4.1 Ising Model

In the Ising model only the spins, which can only be up $(+1)$ and down (-1), generate the magnetization. The spins are fixed in space, but they interact with each other and with the local field in their localization. The energy of the spins is given by $E = -\sum_{i=1}^{n} \sum_{j>i}^{n} w_{ij}\sigma_i\sigma_j - \sum_{i=1}^{n} h_i\sigma_i$, where w_{ij} is the interaction between the spins, that can be positive or negative, and h_i is the local field in the i^{th} spin.

3 THEORETICAL ASPECTS

It is presented the formulation of the Boltzmann Machine Learning problem and its solution using the Linear Response Approximation based on the Bethe Free Energy. After that, we disclose the algorithms for learning interaction parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$. Subsequently, for the first time, a new algorithm for learning β parameter in which original data was generated is exposed. Finally, a detailed description of simulations that ensures that proposed algorithms work properly in reconstructing statistical properties and original data is displayed.

3.1 Analogy to a bill voting system

In a voting system, the number of voting agents is fixed and they can only approve $(+1)$or disapprove (-1) the bills proposed in a given period, analogous to spin up $(\uparrow)$ and down $(\downarrow)$ magnetic system. Moreover, each agent interacts with each other, that have influence over each one votes, represented by the exchange energy $E = -\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_i\sigma_j$. Finally, each agent feel the influence of the average of its peers as well as its voters, and the political environment in general, which also have influence over its votes. Therefore, the energy $E = -\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_i\sigma_j - \sum_{i=1}^{n} h_i\sigma_i$ is a very good analogy to the influences over each agent votes. The political environment can change the probability of each one votes and could be viewed as an environmental temperature.

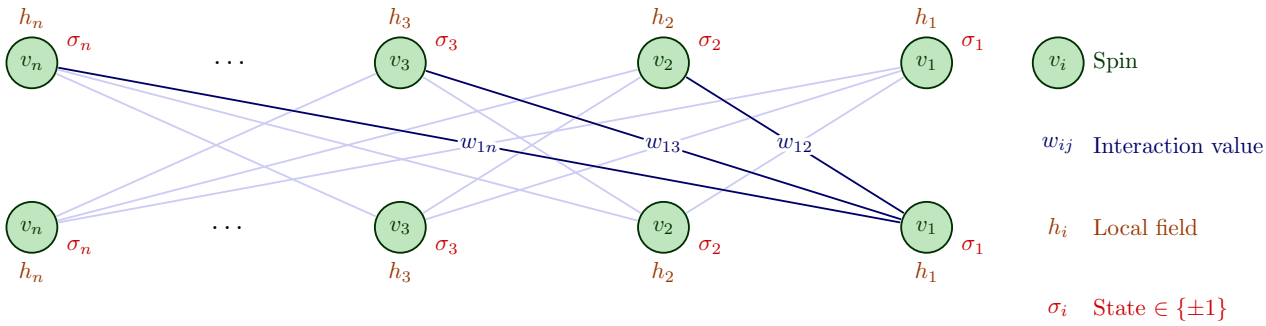
3.2 Formulation of the Boltzmann Machine Learning Problem

We will use the Ising model for the political parties' votes in an effort to forecast how they will vote in the future. For that, we need to infer reliable values to spin-spin interactions w_{ij} and the local fields h_i using past observations of the system behavior. The point is that for the past observation we obtain only the averaged votes $\langle\sigma_i\rangle$ of each agent and the covariance matrix between the agents $c_{ij} = \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle$ for a given period, preferable in a legislature period when all the agents are the same. In terms of magnetic system we know the magnetization and temperature, therefore the best partition function is the Gibbs, instead of the Helmholtz. For that we use the formulation of the Boltzmann Machine Learning Problem.

A Boltzmann machine can be defined as a neural network represented as an

undirected weighted graph $\mathbf{G}(\mathbb{V}, \mathbb{E}, \mathbb{W}, f)$. Vertex set $\mathbb{V} \equiv \{v_1, v_2, v_3, \ldots, v_n\}$ represents all visible or hidden spins. However, for the sake of simplicity, in the remainder of this work only Boltzmann machines with visible spins will be used. Visible spins are regarded as the input - output states $\sigma_i \in \{\pm 1\}$ from all possible configurations of the state set $\vec{\sigma} \equiv \{\sigma_1 \in \{\pm 1\}, \sigma_2 \in \{\pm 1\}\ ,\ \sigma_3 \in \{\pm 1\}, \ldots, \sigma_n \in \{\pm 1\}\}$ [4]. If $\mathbb{V} \otimes \mathbb{V}$ is the set of all ordered pairs $\{v_i, v_j\}$, the edge set $\mathbb{E} \subseteq \mathbb{V} \otimes \mathbb{V}$ is then a relation $\mathbb{E}$ on the set $\mathbb{V}$, meaning that spins v_i and v_j interacts each other in a specific configuration $\{\sigma_i, \sigma_j\}$. Interaction set $\mathbb{W} \equiv \{w_{ij}\}$ is a finite non-empty set of real numbers associated to how strong is the "force" of interplay of each element of the edge set $\mathbb{E}$ that is assigned through the surjective function $f : \{v_i, v_j\} \in \mathbb{E} \longrightarrow w_{ij} \in \mathbb{W} \mid w_{ij} \in \Re$. A graphical representation of a Boltzmann machine with n visible spins is shown at Figure 2.

Figure 2 – Pictorial Representation of a complete connected Boltzmann machine with n visible spins. Green circles represent each spin of the network. Spins are considered as the input - output states σ_i (red label) for all possible configurations of the set $\vec{\sigma}$. Besides the state, all spins possess a local field value h_i (orange label). Note that self-connections are not allowed. Edges for the first spin were dark blue highlighted in order to expose that for each edge an interaction value w_{ij} is assigned.



Source: Author.

Figure 2 shows a complete connected Boltzmann machine. Each node (green circle) of the network represents a visible spin and it possesses two attributes. The first one is the possible input - output state σ_i (red label) and the second one is the local field value h_i (orange label), i.e., $\forall\ v_i\ \exists\ h_i \in \Re$. local field values will play an important role at the time that Boltzmann distribution is formulated. Note that self-connections are not allowed, that is, $w_{ij} = 0$ if $i = j$ and interactions are symmetric $w_{ij} = w_{ji}\ \forall\ i,\ j$ due to the fact that Boltzmann machine is arisen as an undirected graph. Connections for the first spin were dark blue highlighted in order to better visualize that for each edge $\{v_i, v_j\}$ there exists a interaction value w_{ij}.

All the expected values of the spins must be performed over the number of

all possible system realizations. For example, Table 1 gives the possible realizations, and their energies, for a three spins system:

Table 1 – Expected possibilitites for three spins system. For each possible combination k is shown the possible states $\vec{\sigma}$, the energy value and the probability of that state.

k	$\vec{\sigma}$	E_k	p_k
1	$\vec{\sigma}_1 = (1,1,1)$	$E_1 = w_{12} + w_{13} + w_{23} + h_1 + h_2 + h_3$	$p_1 = p_o e^{-\frac{w_{12}+w_{13}+w_{23}+h_1+h_2+h_3}{k_B T}}$
2	$\vec{\sigma}_1 = (1,1,-1)$	$E_2 = w_{12} - w_{13} - w_{23} + h_1 + h_2 - h_3$	$p_2 = p_o e^{-\frac{w_{12}-w_{13}-w_{23}+h_1+h_2-h_3}{k_B T}}$
3	$\vec{\sigma}_1 = (1,-1,1)$	$E_3 = -w_{12} + w_{13} - w_{23} + h_1 - h_2 + h_3$	$p_3 = p_o e^{-\frac{-w_{12}+w_{13}-w_{23}+h_1-h_2+h_3}{k_B T}}$
4	$\vec{\sigma}_1 = (1,-1,-1)$	$E_4 = -w_{12} - w_{13} + w_{23} + h_1 - h_2 - h_3$	$p_4 = p_o e^{-\frac{-w_{12}-w_{13}+w_{23}+h_1-h_2-h_3}{k_B T}}$
5	$\vec{\sigma}_1 = (-1,1,1)$	$E_5 = -w_{12} - w_{13} + w_{23} - h_1 + h_2 + h_3$	$p_5 = p_o e^{-\frac{-w_{12}-w_{13}+w_{23}-h_1+h_2+h_3}{k_B T}}$
6	$\vec{\sigma}_1 = (-1,1,-1)$	$E_6 = -w_{12} + w_{13} - w_{23} - h_1 + h_2 - h_3$	$p_6 = p_o e^{-\frac{-w_{12}+w_{13}-w_{23}-h_1+h_2-h_3}{k_B T}}$
7	$\vec{\sigma}_1 = (-1,-1,1)$	$E_7 = w_{12} - w_{13} - w_{23} - h_1 - h_2 + h_3$	$p_7 = p_o e^{-\frac{w_{12}-w_{13}-w_{23}-h_1-h_2+h_3}{k_B T}}$
8	$\vec{\sigma}_1 = (-1,-1,-1)$	$E_8 = w_{12} + w_{13} + w_{23} - h_1 - h_2 - h_3$	$p_8 = p_o e^{-\frac{w_{12}+w_{13}+w_{23}-h_1-h_2-h_3}{k_B T}}$

Source: Author.

For a n spins the number of possible realizations is $N = 2^n$. The expected values are calculated accordingly by:

$$
\begin{aligned}
N = 2^n \quad &\rightarrow \quad \sigma_{ik} = \text{value of spin } i \text{ in the } k^{th} \text{ realization,} \\
p_k &= \text{probability of } k^{th} \text{ realization,} \\
\langle \sigma_i \rangle = \sum_{k=1}^{N} p_k \sigma_{ik} \quad &\rightarrow \quad \langle \sigma_i \sigma_j \rangle = \sum_{k=1}^{N} p_k \sigma_{ik} \sigma_{jk}, \\
C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle &= \sum_{k=1}^{N} p_k \sigma_{ik} \sigma_{jk} - \left(\sum_{k=1}^{N} p_k \sigma_{ik} \right) \left(\sum_{k=1}^{N} p_k \sigma_{jk} \right), \\
\langle E \rangle = -\sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j>i}^{n} w_{ij} \sigma_i \sigma_j &- \sum_{k=1}^{N} \sum_{j=1}^{n} h_j \sigma_j.
\end{aligned}
\tag{3.1}
$$

The Free Energies for Helmholtz and Gibbs situations are given by:

$$
\begin{aligned}
&Q = TdS \quad \rightarrow \quad W = -\sum_{i=1}^{n} h_i d\sigma_i \quad \rightarrow \quad dU = TdS - \sum_{i=1}^{n} h_i d\sigma_i, \\
&dS = \frac{1}{T}dU + \frac{1}{T}\sum_{i=1}^{n} h_i d\sigma_i \quad \rightarrow \quad \frac{\partial S}{\partial U} = \frac{1}{T} \quad \rightarrow \quad \frac{\partial S}{\partial \sigma_j} = \frac{h_j}{T}. \\
&\text{Helmholtz: } F\left(\overleftrightarrow{w}, \vec{h}; T\right) = U - TS, \\
&dF = -SdT - \sum_{i=1}^{n} h_i d\sigma_i \quad \rightarrow \quad \frac{\partial F}{\partial T} = -S \quad \rightarrow \quad \frac{\partial F}{\partial \sigma_j} = -h_j, \\
&\text{Gibbs: } G\left(\overleftrightarrow{w}, \vec{h}; T\right) = F + \sum_{i=1}^{n} h_i \sigma_i, \\
&dG = -SdT + \sum_{i=1}^{n} \sigma_i dh_i \quad \rightarrow \quad \frac{\partial G}{\partial T} = -S \quad \rightarrow \quad \frac{\partial G}{\partial h_i} = \sigma_i.
\end{aligned}
\tag{3.2}
$$

3.2.1 Partition Functions

$$
\begin{aligned}
&\text{Helmholtz: } Z_B = \sum_{\ell=1}^{N} e^{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{ij}\sigma_{i\ell}\sigma_{j\ell} + \sum_{i=1}^{n} h_i \sigma_{i\ell}}{k_B T}}, \\
&F\left(\overleftrightarrow{w}, \vec{h}; T\right) = -k_B T \ln Z_B \quad \rightarrow \quad S = -\frac{\partial F}{\partial T} \quad \rightarrow \quad h_j = -\frac{\partial F}{\partial \sigma_j}.
\end{aligned}
\tag{3.3}
$$

For the Gibbs situation we know the expected value for each spin, therefore,

we have n constraints given by $\sum_{k=1}^{N} p_k \sigma_{ik} = \eta_i$. The maximization problem is written as:

$$
\begin{aligned}
&\max: \; S = -k_B \sum_{k=1}^{N} p_k \ln p_k \quad \text{subjected to } \; (n+2) \; \text{ constraints}: \\
&(1) \sum_{k=1}^{N} p_k = 1; \\
&(2) \; -\sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_k w_{ij} \sigma S_{ik} S_{jk} - \sum_{k=1}^{N} \sum_{i=1}^{n} p_k h_i \sigma_{ik} = U; \\
&(n) \sum_{k=1}^{N} p_k \sigma_{ik} = \eta_i. \\
&\mathcal{L} = -k_B \sum_{k=1}^{N} p_k \ln p_k - \lambda_1 \left(\sum_{k=1}^{N} p_k - 1 \right) + \\
&\lambda_2 \left[\sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_k w_{ij} \sigma_{ik} \sigma_{jk} + \sum_{k=1}^{N} \sum_{i=1}^{n} p_k h_i \sigma_{ik} + U \right] + \lambda_2 \sum_{i=1}^{n} \nu_i \left(\sum_{k=1}^{N} p_k \sigma_{ik} - \eta_i \right), \\
&= -k_B \sum_{k=1}^{N} p_k \ln p_k - \lambda_1 \left(\sum_{k=1}^{N} p_k - 1 \right) + \\
&\lambda_2 \left[\sum_{k=1}^{N} \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_k w_{ij} \sigma_{ik} \sigma_{jk} + \sum_{k=1}^{N} \sum_{i=1}^{n} p_k \left(h_i + \nu_i \right) \sigma_{ik} + U - \sum_{i=1}^{n} \nu_i \eta_i \right].
\end{aligned}
\tag{3.4}
$$

We need n Lagrange multipliers ν_i for the constraints over each spin expected value. We incorporated, without any loss of generality, the λ_2 Lagrange multiplier in the ν_i. multipliers that can be positive or negative. Following standard procedure:

$$
\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial p_\ell} = -k_B \ln p_\ell - k_B - \lambda_1 + \lambda_2 \left(\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} \left(h_i + \nu_i \right) \sigma_{i\ell} \right) = 0, \\
&p_\ell = p_o e^{\frac{\lambda_2 \left(\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} (h_i + \nu_i) \sigma_{i\ell} \right)}{k_B}}, \\
&\sum_{\ell=1}^{N} p_\ell = 1 \quad \rightarrow \quad p_o = \frac{1}{\sum_{\ell=1}^{N} e^{\frac{\lambda_2 \left(\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} (h_i + \nu_i) \sigma_{i\ell} \right)}{k_B}}}.
\end{aligned}
\tag{3.5}
$$

Again, the λ_2 Lagrange multiplier will be $\lambda_2 = \frac{1}{T}$:

$$\begin{aligned} S &= -k_B \sum_{\ell=1}^{N} p_\ell \ln q_\ell, \\ &= -k_B \ln p_o - \lambda_2 \sum_{\ell=1}^{N} p_\ell \left(\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} (h_i + \nu_i) \sigma_{i\ell} \right), \\ &= -k_B \ln p_o - \lambda_2 \left(U + \sum_{i=1}^{n} \nu_i \langle \sigma_i \rangle \right), \\ \frac{\partial S}{\partial U} &= \lambda_2 = \frac{1}{T}. \end{aligned} \tag{3.6}$$

Therefore, the $\vec{p}$ probability set is given by:

$$\begin{aligned} p_\ell &= p_o e^{\frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} (h_i + \nu_i) \sigma_{i\ell}}{k_B T}}, \\ p_o &= \frac{1}{Z_G} \quad \rightarrow \quad Z_G \left(w_{ij}, h_i, \nu_i; T \right) = \sum_{\ell=1}^{N} e^{\frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} (h_i + \nu_i) \sigma_{i\ell}}{k_B T}}. \end{aligned} \tag{3.7}$$

There is, then, a relationship between the Gibbs and Helmholtz partition functions given by:

$$Z_G \left(w_{ij}, h_i, \nu_i; T \right) = Z_B \left(w_{ij}, h_i + \nu_i; T \right), \tag{3.8}$$

and therefore the Gibbs Free energy (3.2) is given by

$$\begin{aligned} G_B \left(\overleftrightarrow{w}, \vec{h}, \vec{\nu}, \vec{\eta}; T \right) &= F_B \left(\overleftrightarrow{w}, \vec{h}, \vec{\nu}; T \right) + \sum_{i=1}^{n} \nu_i \eta_i, \\ &= -k_B T \ln Z_G \left(\overleftrightarrow{w}, \vec{h}, \vec{\nu} \right) + \sum_{i=1}^{n} \nu_i \eta_i, \\ &= -k_B T \ln \sum_{\ell=1}^{N} e^{\frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} \sigma_{i\ell} \sigma_{j\ell} + \sum_{i=1}^{n} (h_i + \nu_i) \sigma_{i\ell}}{k_B T}} + \sum_{i=1}^{n} \nu_i \eta_i. \end{aligned} \tag{3.9}$$

Take attention that Kullback–Leibler divergence (KLD) (2.5) is

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(\vec{p},\vec{q}\right)=\sum_{k=1}^{N}q_k\ln\left(\frac{q_k}{p_k}\right),\\
&=\sum_{k=1}^{N}q_k\ln q_k-\sum_{k=1}^{N}q_k\ln p_k=\sum_{k=1}^{N}q_k\ln q_k-\sum_{k=1}^{N}q_k\ln\left[\frac{1}{Z_B}e^{\frac{\sum\limits_{i=1}^{n}\sum\limits_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum\limits_{i=1}^{n}h_i\sigma_{ik}}{k_BT}}\right],\\
&=\sum_{k=1}^{N}q_k\ln q_k-\frac{1}{k_BT}\sum_{k=1}^{N}q_k\left[\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}h_i\sigma_{ik}\right]+\ln Z_B\sum_{k=1}^{N}q_k,\\
&=\sum_{k=1}^{N}q_k\ln q_k-\frac{1}{k_BT}\left[\sum_{k=1}^{N}\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}q_k\sigma_{ik}\sigma_{jk}+\sum_{k=1}^{N}\sum_{i=1}^{n}h_iq_k\sigma_{ik}\right]+\ln Z_B,\\
&=\sum_{k=1}^{N}q_k\ln q_k-\frac{1}{k_BT}\left[\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\left\langle\sigma_i\sigma_j\right\rangle+\sum_{i=1}^{n}h_i\left\langle\sigma_i\right\rangle\right]+\ln Z_B,\\
&=\sum_{k=1}^{N}q_k\ln q_k-\sum_{k=1}^{N}\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}q_k\sigma_{ik}\sigma_{jk}-\sum_{k=1}^{N}\sum_{i=1}^{n}h_iq_k\sigma_{ik}+\ln Z_B,\\
&=\sum_{k=1}^{N}q_k\ln q_k-\frac{1}{k_BT}\left[\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\left\langle\sigma_i\sigma_j\right\rangle+\sum_{i=1}^{n}h_i\left\langle\sigma_i\right\rangle\right]-\frac{1}{k_BT}F\left(\overleftrightarrow{w},\vec{h};T\right),
\end{aligned}
\tag{3.10}
$$

where $F\left(\overleftrightarrow{w},\vec{h};T\right)=-k_BT\ln Z_B$ is the Helmholtz energy.

3.3 Learning Algorithm using Linear Response Approximation and Bethe Free Energy

Exact solution to the minimization problem of KLD (3.10) can be achieved through the linear response approximation (LRA) [4, 5, 6]. Linear Response Approximation is a powerful technique used to improve accuracy in estimation of correlations in ensembles that are away or toward equilibrium under perturbative conditions [7]. Physically speaking, LRA means that it is possible to use the **Bethe - Gibbs** free energy as an approximation of the Gibbs free energy. Mathematically, LRA can be successfully implemented if it is possible to prove that inverse Hessian of the **Gibbs free energy** is equal to Hessian of the free energy. Implementation of LRA is as follows.

The condition $\sum\limits_{k=1}^{N}p_k\sigma_{ik}=\eta_i$ on the values ν_i in (3.4) can be introduced as the

dual variation of $G_B\left(\overleftrightarrow{w},\vec{h},\vec{\nu},\vec{\eta};T\right)$ respect to ν_i, i.e.

$$G_B\left(\overleftrightarrow{w},\vec{h},\vec{\nu},\vec{\eta};T\right)=\max_{\vec{\nu}}\left\{-k_BT\ln\sum_{\ell=1}^{N}e^{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}w_{ij}\sigma_{i\ell}\sigma_{j\ell}+\sum_{i=1}^{n}(h_i+\nu_i)\sigma_{i\ell}}{k_BT}}+\sum_{i=1}^{n}\nu_i\eta_i\right\}. \tag{3.11}$$

By shifting Lagrange operator $\nu_i=\gamma_i-h_i$ we have

$$\begin{aligned}
&G_B\left(\overleftrightarrow{w},\vec{h},\vec{\nu},\vec{\eta}\right)\\
&=\max_{\vec{\nu}}\left\{-k_BT\ln\sum_{\ell=1}^{N}e^{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}w_{ij}\sigma_{i\ell}\sigma_{j\ell}+\sum_{i=1}^{n}h_i\sigma_{i\ell}+\sum_{i=1}^{n}(\gamma_i-h_i)\sigma_{i\ell}}{k_BT}}+\sum_{i=1}^{n}(\gamma_i-h_i)\eta_i\right\},\\
&=\max_{\vec{\nu}}\left\{-k_BT\ln\sum_{\ell=1}^{N}e^{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}w_{ij}\sigma_{i\ell}\sigma_{j\ell}+\sum_{i=1}^{n}h_i\sigma_{i\ell}-\sum_{i=1}^{n}h_i\sigma_{i\ell}+\sum_{i=1}^{n}\gamma_i\sigma_{i\ell}}{k_BT}}+\sum_{i=1}^{n}\gamma_i\eta_i-\sum_{i=1}^{n}h_i\eta_i\right\},\\
&=\max_{\vec{\gamma}}\left\{-k_BT\ln\sum_{\ell=1}^{N}e^{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}w_{ij}\sigma_{i\ell}\sigma_{j\ell}+\sum_{i=1}^{n}\gamma_i\sigma_{i\ell}}{k_BT}}+\sum_{i=1}^{n}\gamma_i\eta_i\right\}-\sum_{i=1}^{n}h_i\eta_i,\\
&=\max_{\vec{\gamma}}\left\{-k_BT\ln Z_B\left(\overleftrightarrow{w},\vec{\gamma};T\right)+\sum_{i=1}^{n}\gamma_i\eta_i\right\}-\sum_{i=1}^{n}h_i\eta_i.
\end{aligned} \tag{3.12}$$

Optimal values of $\vec{\gamma}\left(\overleftrightarrow{w},\vec{\eta}\right)$ namely $\vec{\vartheta}\left(\overleftrightarrow{w},\vec{\eta}\right)$ must be chosen such that $\sum_{k=1}^{N}p_k\sigma_{ik}=\eta_i$, or in other words, $\vec{\vartheta}\left(\overleftrightarrow{w},\vec{\eta}\right)$ must satisfy

$$\eta_i=\sum_{k=1}^{N}\frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\sigma_{ik}, \tag{3.13}$$

such that Gibbs free energy becomes

$$G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)=-k_BT\ln Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)+\sum_{i=1}^{n}\vartheta_i\eta_i-\sum_{i=1}^{n}h_i\eta_i. \tag{3.14}$$

Deriving $G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)$ respect to η_j

$$\begin{aligned}\frac{\partial G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)}{\partial\eta_j} &= \frac{\partial}{\partial\eta_j}\left[-k_BT\ln Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)+\sum_{i=1}^{n}\vartheta_i\eta_i-\sum_{i=1}^{n}h_i\eta_i\right],\\ &= \vartheta_j-h_j=0 \quad\rightarrow\quad \vartheta_j=h_j\end{aligned} \tag{3.15}$$

meaning that

$$F_B\left(\overleftrightarrow{w},\vec{h};T\right)=\min_{\vec{\eta}}G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)=G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta}^{*};T\right), \tag{3.16}$$

where $\vec{\eta}^{*}$ is the optimal values minimizing $G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)$. Now deriving (3.13) respect to η_j

$$\begin{aligned}&h_i=\vartheta_i,\\ &Z_B\left(\overleftrightarrow{w},\vec{h};T\right)=\sum_{q=1}^{N}e^{\frac{\sum_{r=1}^{n}\sum_{s>q}^{n}w_{rs}\sigma_{rq}\sigma_{sq}+\sum_{i=1}^{n}\vartheta_r\sigma_{rq}}{k_BT}},\\ &\eta_\ell=\sum_{k=1}^{N}\frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\sigma_{\ell k},\\ &\frac{\partial\eta_\ell}{\partial\eta_m}=\frac{\partial}{\partial\eta_m}\sum_{k=1}^{N}\frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\sigma_{\ell k},\\ &\frac{\partial\eta_\ell}{\partial\eta_m}=\sum_{p=1}^{n}\frac{\partial\vartheta_p}{\partial\eta_m}\frac{\partial}{\partial\vartheta_p}\sum_{k=1}^{N}\frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\sigma_{\ell k},\\ &=\sum_{p=1}^{n}\frac{\partial\vartheta_p}{\partial\eta_m}\sum_{k=1}^{N}\sigma_{\ell k}\frac{\frac{\partial}{\partial\vartheta_p}\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\\ &-\sum_{p=1}^{n}\frac{\partial\vartheta_p}{\partial\eta_m}\sum_{k=1}^{N}\sigma_{\ell k}\frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B^2\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\frac{\partial}{\partial\vartheta_p}Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right),\\ &=\sum_{p=1}^{n}\frac{\partial\vartheta_p}{\partial\eta_m}\sum_{k=1}^{N}\sigma_{\ell k}\frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n}w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}\frac{\partial}{\partial\vartheta_p}\frac{1}{k_BT}\sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{Z_B\left(\overleftrightarrow{w},\vec{\vartheta};T\right)}\end{aligned} \tag{3.17}$$

$$
\begin{aligned}
&- \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \sum_{k=1}^{N} \sigma_{\ell k} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B^2\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sum_{q=1}^{N} e^{\frac{\sum_{r=1}^{n}\sum_{s>q}^{n} w_{rs}\sigma_{rq}\sigma_{sq} + \sum_{i=1}^{n}\vartheta_r\sigma_{rq}}{k_BT}} \frac{1}{k_BT} \frac{\partial}{\partial \vartheta_p} \sum_{i=1}^{n} \vartheta_r \sigma_{rq}, \\
&= \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \sum_{k=1}^{N} \sigma_{\ell k} \frac{1}{k_BT} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sigma_{pk} \\
&- \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \sum_{k=1}^{N} \sigma_{\ell k} \frac{1}{k_BT} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sum_{q=1}^{N} \frac{e^{\frac{\sum_{r=1}^{n}\sum_{s>q}^{n} w_{rs}\sigma_{rq}\sigma_{sq} + \sum_{i=1}^{n}\vartheta_r\sigma_{rq}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sum_{i=1}^{n} \delta_{rp}\sigma_{rq}, \\
&= \frac{1}{k_BT} \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \sum_{k=1}^{N} \sigma_{\ell k} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sigma_{pk} \\
&- \frac{1}{k_BT} \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \sum_{k=1}^{N} \sigma_{\ell k} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sum_{q=1}^{N} \frac{e^{\frac{\sum_{r=1}^{n}\sum_{s>q}^{n} w_{rs}\sigma_{rq}\sigma_{sq} + \sum_{i=1}^{n}\vartheta_r\sigma_{rq}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sigma_{pq}, \\
&= \frac{1}{k_BT} \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \sum_{k=1}^{N} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sigma_{\ell k}\sigma_{pk} \\
&- \frac{1}{k_BT} \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \left(\sum_{k=1}^{N} \frac{\mathrm{e}^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n}\vartheta_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sigma_{\ell k} \right) \left(\sum_{q=1}^{N} \frac{e^{\frac{\sum_{r=1}^{n}\sum_{s>q}^{n} w_{rs}\sigma_{rq}\sigma_{sq} + \sum_{i=1}^{n}\vartheta_r\sigma_{rq}}{k_BT}}}{Z_B\left(\overleftrightarrow{w}, \vec{\vartheta}; T\right)} \sigma_{pq} \right), \\
&= \frac{1}{k_BT} \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \left(\langle \sigma_\ell \sigma_p \rangle - \langle \sigma_\ell \rangle \langle \sigma_p \rangle \right).
\end{aligned}
$$

Therefore, we have

$$
\frac{\partial \eta_\ell}{\partial \eta_m} = \frac{1}{k_BT} \sum_{p=1}^{n} \frac{\partial \vartheta_p}{\partial \eta_m} \left(\langle \sigma_\ell \sigma_p \rangle - \langle \sigma_\ell \rangle \langle \sigma_p \rangle \right) = \delta_{\ell m}. \tag{3.18}
$$

From (3.15) we have $\vartheta_p = \frac{\partial G_B\left(\overleftrightarrow{w}, \vec{h}, \vec{\vartheta}, \vec{\eta}; T\right)}{\partial \eta_p} + h_p$ therefore $\frac{\partial \vartheta_p}{\partial \eta_m} = \frac{\partial^2 G_B\left(\overleftrightarrow{w}, \vec{h}, \vec{\vartheta}, \vec{\eta}; T\right)}{\partial \eta_p \partial \eta_m} =$

$[\mathbf{H}_{G_B}]_{pm}$, where is the Hessian matrix of Gibbs potential. This means that:

$$\begin{aligned} \Sigma_{p\ell} &= \langle \sigma_\ell \sigma_p \rangle - \langle \sigma_\ell \rangle \langle \sigma_p \rangle , \\ \frac{\partial \eta_\ell}{\partial \eta_m} &= \frac{1}{k_B T} \sum_{p=1}^{n} [\mathbf{H}_{G_B}]_{mp} \Sigma_{p\ell} = \frac{1}{k_B T} \delta_{\ell m}, \\ \frac{1}{k_B T} [\mathbf{H}_{G_B}]_{mp} \Sigma_{mp} &= I, \\ \frac{1}{k_B T} \Sigma_{mp} &= [\mathbf{H}_{G_B}]^{-1}_{mp} . \end{aligned} \tag{3.19}$$

Now we must find out what is the matrix Σ. By taking the derivative of Helmholtz potential respect h_ℓ

$$\begin{aligned} \frac{\partial F_B\left(\overleftrightarrow{w}, \vec{h}; T\right)}{\partial h_\ell} &= -k_B T \frac{\partial}{\partial h_\ell} \ln Z_B\left(\overleftrightarrow{w}, \vec{h}; T\right), \\ &= -k_B T \frac{1}{Z_B\left(\overleftrightarrow{w}, \vec{h}; T\right)} \frac{\partial}{\partial h_\ell} \left[\sum_{k=1}^{N} e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n} h_i \sigma_{ik}}{k_B T}} \right], \\ &= -k_B T \frac{1}{Z_B\left(\overleftrightarrow{w}, \vec{h}; T\right)} \left[\sum_{k=1}^{N} e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n} h_i \sigma_{ik}}{k_B T}} \frac{1}{k_B T} \frac{\partial}{\partial h_\ell} \sum_{i=1}^{n} h_i \sigma_{ik} \right], \\ &= -\frac{1}{Z_B\left(\overleftrightarrow{w}, \vec{h}; T\right)} \left[\sum_{k=1}^{N} e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n} h_i \sigma_{ik}}{k_B T}} \sum_{i=1}^{n} \delta_{i\ell} \sigma_{ik} \right], \\ &= -\frac{1}{Z_B\left(\overleftrightarrow{w}, \vec{h}; T\right)} \left[\sum_{k=1}^{N} e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk} + \sum_{i=1}^{n} h_i \sigma_{ik}}{k_B T}} \sigma_{\ell k} \right]. \end{aligned} \tag{3.20}$$

The second derivative is

$$
\begin{aligned}
\frac{\partial^2 F_B\left(\overleftrightarrow{w},\vec{h};T\right)}{\partial h_p \partial h_\ell} &= -\frac{\partial}{\partial h_p}\sum_{k=1}^{N}\frac{e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{h};T\right)}\sigma_{\ell k},\\
&= -\sum_{k=1}^{N}\frac{\frac{\partial}{\partial h_p}e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{h};T\right)}\sigma_{\ell k}\\
&+\sum_{k=1}^{N}\frac{e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B^2\left(\overleftrightarrow{w},\vec{h};T\right)}\sigma_{\ell k}\frac{\partial}{\partial h_p}Z_B\left(\overleftrightarrow{w},\vec{h};T\right),\\
&= -\sum_{k=1}^{N}\frac{e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{h};T\right)}\sigma_{\ell k}\frac{\partial}{\partial h_p}\sum_{i=1}^{n} h_i\sigma_{ik}\\
&+\sum_{k=1}^{N}\frac{e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B^2\left(\overleftrightarrow{w},\vec{h};T\right)}\sigma_{\ell k}\frac{\partial}{\partial h_p}\sum_{q=1}^{N}e^{\frac{\sum_{r=1}^{n}\sum_{s>i}^{n} w_{rs}\sigma_{rq}\sigma_{sq}+\sum_{i=1}^{n} h_r\sigma_{rq}}{k_BT}},\\
&= -\sum_{k=1}^{N}\frac{e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{\ell k}\frac{1}{k_BT}\sum_{i=1}^{n}\delta_{ip}\sigma_{ik} \qquad (3.21)\\
&+\sum_{k=1}^{N}\frac{e^{\frac{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}{k_BT}}}{Z_B^2\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{\ell k}\sum_{q=1}^{N}e^{\frac{\sum_{r=1}^{n}\sum_{s>i}^{n} w_{rs}\sigma_{rq}\sigma_{sq}+\sum_{i=1}^{n} h_r\sigma_{rq}}{k_BT}}\frac{\partial}{\partial h_p}\frac{1}{k_BT}\sum_{i=1}^{n} h_r\sigma_{rq},\\
&= -\frac{1}{k_BT}\sum_{k=1}^{N}\frac{e^{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}}{Z_B\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{\ell k}\sigma_{pk}\\
&+\frac{1}{k_BT}\sum_{k=1}^{N}\frac{e^{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}}{Z_B^2\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{\ell k}\sum_{q=1}^{N}e^{\sum_{r=1}^{n}\sum_{s>i}^{n} w_{rs}\sigma_{rq}\sigma_{sq}+\sum_{i=1}^{n} h_r\sigma_{rq}}\sum_{i=1}^{n}\delta_{rp}\sigma_{rq},\\
&= -\frac{1}{k_BT}\sum_{k=1}^{N}\frac{e^{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}}{Z_B\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{\ell k}\sigma_{pk}\\
&+\frac{1}{k_BT}\left(\sum_{k=1}^{N}\frac{e^{\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_{ik}\sigma_{jk}+\sum_{i=1}^{n} h_i\sigma_{ik}}}{Z_B\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{\ell k}\right)\left(\sum_{q=1}^{N}\frac{e^{r=1}^{n}\sum_{s>i}^{n} w_{rs}\sigma_{rq}\sigma_{sq}+\sum_{i=1}^{n} h_r\sigma_{rq}}{Z_B\left(\overleftrightarrow{w},\vec{h}\right)}\sigma_{pq}\right),\\
&= -\frac{1}{k_BT}\left[\langle\sigma_\ell\sigma_p\rangle-\langle\sigma_\ell\rangle\langle\sigma_p\rangle\right] = -\frac{1}{k_BT}\Sigma_{\ell p}.
\end{aligned}
$$

That is to say

$$\begin{aligned}\frac{\partial^2 F_B\left(\overleftrightarrow{w},\vec{h};T\right)}{\partial h_p \partial h_\ell} &= -\frac{1}{k_B T}\Sigma_{\ell p} \quad \rightarrow \quad \frac{1}{k_B T}\Sigma_{\ell p} = -[\mathbf{H}_{F_B}]_{\ell p}, \\ [\mathbf{H}_{G_B}]^{-1}_{\ell p} &= -[\mathbf{H}_{F_B}]_{\ell p},\end{aligned} \tag{3.22}$$

therefore

$$[\mathbf{H}_{G_B}]^{-1}_{ij} = -[\mathbf{H}_{F_B}]_{ij}. \tag{3.23}$$

Equation above is known as the Linear Response Relation (LRR) [4]. This relation opens up the way to implement the linear response approximation. In order to implement LRA lets compute extremal conditions for KLD (3.10). Extremal conditions for $\vec{h}$ gives

$$\begin{aligned}\frac{\partial D_{\mathrm{KL}}\left(\vec{p},\vec{q}\right)}{\partial h_\ell} &= \frac{\partial}{\partial h_\ell}\left[\sum_{k=1}^{N} q_k \ln q_k - \frac{1}{k_B T}\left[\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\left\langle\sigma_i\sigma_j\right\rangle + \sum_{i=1}^{n} h_i\left\langle\sigma_i\right\rangle\right] - \frac{1}{k_B T}F\left(\overleftrightarrow{w},\vec{h};T\right)\right], \\ &= \frac{\partial}{\partial h_\ell}\left[\sum_{k=1}^{N} q_k \ln q_k\right]\grave{} - \frac{1}{k_B T}\frac{\partial}{\partial h_\ell}\left[\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\left\langle\sigma_i\sigma_j\right\rangle\right], \\ &- \frac{1}{k_B T}\frac{\partial}{\partial h_\ell}\left[\sum_{i=1}^{n} h_i\left\langle\sigma_i\right\rangle\right] - \frac{1}{k_B T}\frac{\partial}{\partial h_\ell}\left[F\left(\overleftrightarrow{w},\vec{h};T\right)\right], \\ &= -\frac{1}{k_B T}\sum_{i=1}^{n}\delta_{i\ell}\left\langle\sigma_i\right\rangle - \frac{1}{k_B T}\frac{\partial}{\partial h_\ell}\left[F\left(\overleftrightarrow{w},\vec{h};T\right)\right], \\ &= -\frac{1}{k_B T}\left[\left\langle\sigma_\ell\right\rangle + \frac{\partial}{\partial h_\ell}\left[F\left(\overleftrightarrow{w},\vec{h};T\right)\right]\right], \\ 0 &= -\frac{1}{k_B T}\left[\left\langle\sigma_\ell\right\rangle + \frac{\partial}{\partial h_\ell}\left[\min_{\vec{\eta}} G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)\right]\right], \\ 0 &= -\frac{1}{k_B T}\left[\left\langle\sigma_\ell\right\rangle - \eta_i^*\right].\end{aligned} \tag{3.24}$$

Extremal conditions for $\overleftrightarrow{w}$

$$\begin{aligned}
\frac{\partial D_{\mathrm{KL}}\left(\vec{p},\vec{q}\right)}{\partial w_{\ell m}} &= \left[\sum_{k=1}^{N} q_k \ln q_k\right]` - \frac{1}{k_B T}\frac{\partial}{\partial w_{\ell m}}\left[\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\left\langle \sigma_i \sigma_j\right\rangle\right], \\
&- \frac{1}{k_B T}\frac{\partial}{\partial w_{\ell m}}\left[\sum_{i=1}^{n} h_i \left\langle \sigma_i\right\rangle\right] - \frac{1}{k_B T}\frac{\partial}{\partial w_{\ell m}}\left[F\left(\overleftrightarrow{w},\vec{h};T\right)\right], \\
&= -\frac{1}{k_B T}\left[\sum_{i=1}^{n}\sum_{j>i}^{n}\delta_{i\ell}\delta_{jm}\left\langle \sigma_i\sigma_j\right\rangle\right] - \frac{1}{k_B T}\frac{\partial}{\partial w_{\ell m}}\left[F\left(\overleftrightarrow{w},\vec{h};T\right)\right], \\
&= -\frac{1}{k_B T}\left[\left\langle \sigma_\ell \sigma_m\right\rangle + \frac{\partial}{\partial w_{\ell m}}\left[F\left(\overleftrightarrow{w},\vec{h};T\right)\right]\right], \\
0 &= -\frac{1}{k_B T}\left[\left\langle \sigma_\ell \sigma_m\right\rangle + \frac{\partial^2 F\left(\overleftrightarrow{w},\vec{h};T\right)}{\partial h_\ell h_m} + \frac{\partial F\left(\overleftrightarrow{w},\vec{h};T\right)}{\partial h_\ell}\frac{\partial F\left(\overleftrightarrow{w},\vec{h};T\right)}{\partial h_m}\right], \\
0 &= -\frac{1}{k_B T}\left[\left\langle \sigma_\ell \sigma_m\right\rangle + \left[\mathbf{H}_{G_B}^{-1}\right]_{\ell m}\Big|_{\vec{\eta}=\vec{\eta}^*} + \eta_\ell^* \eta_m^*\right].
\end{aligned} \tag{3.25}$$

As already mentioned, it was proven that optimal values of η_i^* are precisely the expected values of σ_i given by (3.1). So, the exact learning algorithm in Boltzmann machine requires for finding desired parameters $\left\{\overleftrightarrow{w},\vec{h}\right\}$ that satisfy the conditions

$$\begin{aligned}
\left\langle \sigma_i\right\rangle &= \eta_i^*, \\
\left[\mathbf{H}_{G_B}\right]_{ij}\Big|_{\vec{\eta}=\{\langle\sigma_i\rangle\}} &= \left[\Sigma^{-1}\right]_{ij},
\end{aligned} \tag{3.26}$$

where $[\Sigma]_{ij} = \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle$ is the covariance matrix of the data. Note that, the inverse operator this time was applied over the covariance matrix. This is allowed mainly because we are working with a complete connected Boltzmann machine. Besides, $[\mathbf{H}_{G_B}]_{ij}$ and $[\Sigma]_{ij}$ are symmetrical and it is possible to conclude that invertible operator change holds for all diagonal elements[1]. Main concern now is that the exact value of Gibbs free energy is computationally, difficult to obtain. Therefore, in order to overcome this situation, we use the Bethe - Gibbs free energy as an approximation of the Gibbs free energy, that is, we applied the linear response approximation. To do that, let's introduce the **Bethe free energy** [4, 6]

[1] To clarify, this just works in the exact case. When $G_B\left(\overleftrightarrow{w},\vec{h},\vec{\vartheta},\vec{\eta};T\right)$ is approximated, invertible operator change does not work with diagonal elements. This is known as the diagonal problem. However, this difficulty can be better understood when loop contributions of Gibbs free energy are introduced, as it will be exposed below.

3.3.1 Introducing Bethe free energy

Let's consider an operator $\hat{A}$ and an arbitrary base $|i\rangle$, not necessarily the eigenvectors of $\hat{A}$, and define the matrix $A_{ij} = \langle j|\,\hat{A}\,|i\rangle$. The expected value, or expectancy, of this operator is given by $\left\langle \hat{A} \right\rangle = \sum_i \langle i|\,\hat{A}\,|i\rangle = \sum_i A_{ii} = \mathrm{Tr}\,(A)$, where $\mathrm{Tr}\,(.)$ stands for trace, applied only to square matrices. In the operator eigenvector basis $\hat{A}\,|a_i\rangle = \lambda_i\,|a_i\rangle \quad \rightarrow \quad \langle a_j|\,\hat{A}\,|a_i\rangle = \lambda_i\,\langle a_j|a_i\rangle = \lambda_i\delta_{ij} \quad \rightarrow \quad \mathrm{Tr}\,(A) = \sum_i \lambda_i$. For a complete set of eigenvectors any ket can be projected in that basis as $|\psi\rangle = \sum_i |a_i\rangle\,\langle a_i|\psi\rangle = \left(\sum_i |a_i\rangle\,\langle a_i|\right)|\psi\rangle \quad \rightarrow \quad \sum_i |a_i\rangle\,\langle a_i| = I \quad \rightarrow \quad \sum_i |i\rangle\,\langle i| = I$. Therefore, the trace is independent of the chosen base:

$$\begin{aligned}
\mathrm{Tr}\,(A) &= \sum_i \langle i|\,\hat{A}\,|i\rangle, \\
&= \sum_i \langle i|\left(\sum_k |a_k\rangle\,\langle a_k|\right)\hat{A}\,|i\rangle, \\
&= \sum_i \left(\sum_k \langle i|a_k\rangle\,\langle a_k|\,\hat{A}\,|i\rangle\right), \\
&= \sum_k \left(\sum_i \langle a_k|\,\hat{A}\,|i\rangle\,\langle i|a_k\rangle\right) = \sum_k \langle a_k|\,\hat{A}\,|a_k\rangle.
\end{aligned} \tag{3.27}$$

The trace of a multiplication of matrices is commutative, that is, $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$:

$$\begin{aligned}
(AB)_{ij} &= \sum_k A_{ik}B_{kj} \quad \rightarrow \quad \mathrm{Tr}\,(AB) = \sum_i \sum_k A_{ik}B_{ki}, \\
(BA)_{ij} &= \sum_k B_{ik}A_{kj} \quad \rightarrow \quad \mathrm{Tr}\,(BA) = \sum_i \left(\sum_k B_{ik}A_{ki}\right) = \sum_k \left(\sum_i A_{ki}B_{ik}\right) = \mathrm{Tr}\,(AB).
\end{aligned} \tag{3.28}$$

Therefore, the cyclic multiplication holds: $\mathrm{Tr}\,(ABC) = \mathrm{Tr}\,(CAB) = \mathrm{Tr}\,(BCA)$, in particular $\mathrm{Tr}\,(S^{-1}AS) = \mathrm{Tr}\,(ASS^{-1}) = \mathrm{Tr}\,(AI) = \mathrm{Tr}\,(A)$.

Consider the probability vector $\vec{p} = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \end{pmatrix} \quad \rightarrow \quad p_i \geq 0 \quad \rightarrow \sum_i p_i = 1$ and all possible states of a system $|\psi_i\rangle$ with probability p_i. The expected value for $\hat{A}$ is, therefore, given by $\left\langle \hat{A} \right\rangle = \sum_i p_i\,\langle \psi_i|\,\hat{A}\,|\psi_i\rangle = \sum_i p_i\,\mathrm{Tr}\left(\langle \psi_i|\,\hat{A}\,|\psi_i\rangle\right) = \sum_i p_i\,\mathrm{Tr}\left(\hat{A}\,|\psi_i\rangle\,\langle \psi_i|\right) = \mathrm{Tr}\left(\hat{A}\sum_i p_i\,|\psi_i\rangle\,\langle \psi_i|\right)$. The density matrix, therefore, is defined as: $\rho = \sum_i p_i\,|\psi_i\rangle\,\langle \psi_i|$, and the expected value of $\hat{A} = \mathrm{Tr}\left(\hat{A}\rho\right)$. An important property of the density matrix is that $\mathrm{Tr}\,(\rho) = 1 \quad \rightarrow \quad \mathrm{Tr}\,(\rho) = \sum_i p_i\ \mathrm{Tr}\,(|\psi_i\rangle\,\langle \psi_i|) = \sum_i p_i\ \mathrm{Tr}\,(\langle \psi_i|\psi_i\rangle) =$

$\sum_i p_i = 1$. The Bethe free energy can be defined as follows:

$$\begin{aligned} H &= -\sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\sigma_i\sigma_j - \sum_{i=1}^{n} h_i\sigma_i, \\ F^{(n)} &= \mathrm{Tr}\left[\rho^{(n)}\left(H + k_B T \ln \rho^{(n)}\right)\right] \quad \rightarrow \quad \mathrm{Tr}\left(\rho^{(n)}\right) = 1. \end{aligned} \tag{3.29}$$

The Bethe Free energy only considers terms up to the second order in the form:

$$\mathcal{F}_{\mathrm{Bethe}} = -\sum_{i=1}^{n} h_i \,\mathrm{Tr}\left(\rho^{(1)}\sigma_i\right) - \sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\,\mathrm{Tr}\left(\rho^{(2)}\sigma_i\sigma_j\right) + k_B T\,\mathrm{Tr}\left(\rho^{(1)}\ln\rho^{(1)}\right) + k_B T\,\mathrm{Tr}\left(\rho^{(2)}\ln\rho^{(2)}\right),$$

$$\mathrm{Tr}\left(\rho^{(1)}\right) = 1 \quad \rightarrow \quad \mathrm{Tr}\left(\rho^{(2)}\right) = 1,$$

$$\rho^{(1)`} = \frac{1}{2}\begin{pmatrix} 1+m_i & 0 \\ 0 & 1-m_i \end{pmatrix} \quad \mathrm{Tr}\left(\rho^{(1)}\right) = \frac{1}{2}\left(1+m_i+1-m_i\right) = 1,$$

$$\rho^{(1)`} = \frac{1}{4}\begin{pmatrix} 1+m_i+m_j+\xi_{ij} & 0 & 0 & 0 \\ 0 & 1-m_i+m_j-\xi_{ij} & 0 & 0 \\ 0 & 0 & 1+m_i-m_j-\xi_{ij} & 0 \\ 0 & 0 & 0 & 1-m_i-m_j+\xi_{ij} \end{pmatrix},$$

$$\mathrm{Tr}\left(\rho^{(2)}\right) = \frac{1}{4}\left(1+m_i+m_j+\xi_{ij}+1-m_i+m_j-\xi_{ij}+1+m_i-m_j-\xi_{ij}+1-m_i-m_j+\xi_{ij}\right) = 1,$$

where function m_i and ξ_{ij} are defined as

$$\begin{aligned} m_i &= \mathrm{Tr}\left(\rho^{(1)}\sigma_i\right), \\ \xi_{ij} &= \mathrm{Tr}\left(\rho^{(2)}\sigma_i\sigma_j\right). \end{aligned} \tag{3.30}$$

Therefore Bethe free energy is given by

$$\begin{aligned} \mathcal{F}_{\mathrm{Bethe}} = &-\sum_{i=1}^{n} h_i m_i - \sum_{i=1}^{n}\sum_{j>i}^{n} w_{ij}\xi_{ij} + k_B T\sum_{i=1}^{n}(1-k_i)\sum_{\sigma_i=-1}^{+1}\frac{1+\sigma_i m_i}{2}\ln\left(\frac{1+\sigma_i m_i}{2}\right) \\ &+ k_B T\sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{\sigma_i=-1}^{+1}\sum_{\sigma_j=-1}^{+1}\frac{1+\sigma_i m_i+\sigma_j m_j+\sigma_i\sigma_j\xi_{ij}}{4}\ln\left(\frac{1+\sigma_i m_i+\sigma_j m_j+\sigma_i\sigma_j\xi_{ij}}{4}\right), \end{aligned} \tag{3.31}$$

where $k_i \geq 1$ is the degree of each visible spin. In the case of complete connected Boltzmann machine $k_i = n-1$. Bethe free energy can be understood as the free energy up to pairwise interactions among visible spins. Observe that Bethe free energy depends on two new parameters $\{\mathbf{m}, \boldsymbol{\xi}\}$. These two new parameter sets are employed to ensure that

Bethe free energy should be taken as an approximation of free energy only over optimal values of $\{\mathbf{m}, \boldsymbol{\xi}\}$, that is,

$$F_B\left(\overleftrightarrow{w}, \vec{h}; T\right) := F_B\left(\mathbb{W}, \mathbb{H}; T\right) \approx \min_{\mathbf{m},\boldsymbol{\xi}} \left\{\mathcal{F}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \mathbf{m}, \boldsymbol{\xi}); T\right\}, \tag{3.32}$$

where $\left(\overleftrightarrow{w}, \vec{h}\right)$ were re-defined in term of set notation $(\mathbb{W}, \mathbb{H})$. From now, for the sake of simplicity we assume $k_B T = 1$[2]. Performing the minimization process over $\boldsymbol{\xi}$ and solving the resulting equation for the same variable, we obtain that optimal values of $\boldsymbol{\xi}$ has the form

$$\xi_{ij} = \coth(2w_{ij}) \left(1 - \sqrt{1 - (1 - m_i^2 - m_j^2)\tanh(2w_{ij}) - 2m_i m_j \tanh(2w_{ij})}\right). \tag{3.33}$$

Already defined the Bethe free energy, relation (3.16) can be rewritten as

$$\begin{aligned}\mathcal{G}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) = &\max_{\nu_i} \left\{\min_{\mathbf{m}} \left\{\mathcal{F}_{\text{Bethe}}(\mathbb{W}, \nu, \mathbf{m}, \boldsymbol{\xi}(\mathbb{W}, \mathbf{m}))\right\} + \sum_i \nu_i \eta_i\right\} \\ &- \sum_i h_i \eta_i,\end{aligned} \tag{3.34}$$

where the dependence of $\boldsymbol{\xi}$ on $\mathbb{W}$ and $\mathbf{m}$ was explicitly mentioned. Performing maximization over $\boldsymbol{\nu}$ implies that values of $\mathbf{m}$ that minimizes $\mathcal{F}_{\text{Bethe}}(\mathbb{W}, \nu, \mathbf{m}, \boldsymbol{\xi}(\mathbb{W}, \boldsymbol{\eta}))$ must be the expectation values imposed as a constraint in equation (3.1). That is, using the approximation $\mathbf{m}$ is equivalent to parameter $\boldsymbol{\eta}$. Therefore, Bethe - Gibbs free energy results in

$$\begin{aligned}\mathcal{G}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) = &\mathcal{F}_{\text{Bethe}}(\mathbb{W}, \nu_i^*(\mathbb{W}, \boldsymbol{\eta}), \boldsymbol{\eta}, \boldsymbol{\xi}(\mathbb{W}, \boldsymbol{\eta})) \\ &+ \sum_i \nu_i^*(\mathbb{W}, \boldsymbol{\eta})\eta_i - \sum_i h_i \eta_i.\end{aligned} \tag{3.35}$$

Using again relation (3.15) is simple to realize that $\nu_i^*(\mathbb{W}, \boldsymbol{\eta}) = h_i$ leading us to conclude

$$\mathcal{F}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}, \boldsymbol{\xi}(\mathbb{W}, \boldsymbol{\eta})) = \min_{\boldsymbol{\eta}} \mathcal{G}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}). \tag{3.36}$$

[2] Mainly because temperature is a free paramater.

Using $\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})$ instead of $G_B(\mathbb{W},\mathbb{H},\boldsymbol{\eta})$ in extremal conditions for KLD, equations (3.24) and (3.25) turn out assuming again $k_BT=1$ in

$$\begin{aligned}\frac{\partial D_{\text{KL}}\left(\vec{p},\vec{q}\right)}{\partial h_i} &\approx -\left\langle\sigma_i\right\rangle - \frac{\partial}{\partial h_i}\mathcal{F}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta},\boldsymbol{\xi}(\mathbb{W},\boldsymbol{\eta})),\\ 0 &= -\left\langle\sigma_i\right\rangle - \frac{\partial}{\partial h_i}\left(\min_{\boldsymbol{\eta}}\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})\right),\\ 0 &= -\left\langle\sigma_i\right\rangle + \eta_i^*,\end{aligned} \tag{3.37}$$

for $\mathbb{H}$ set. Extremal conditions for $\mathbb{W}$ holds as

$$\begin{aligned}\frac{\partial D_{\text{KL}}\left(\vec{p},\vec{q}\right)}{\partial w_i} &\approx -\left\langle\sigma_i\sigma_j\right\rangle - \frac{\partial}{\partial w_i}\mathcal{F}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta},\boldsymbol{\xi}(\mathbb{W},\boldsymbol{\eta})),\\ 0 &= -\left\langle\sigma_i\sigma_j\right\rangle - \frac{\partial^2}{\partial h_i\partial h_j}\left(\min_{\boldsymbol{\eta}}\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})\right)\\ &\quad + \frac{\partial}{\partial h_i}\left(\min_{\boldsymbol{\eta}}\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})\right)\frac{\partial}{\partial h_j}\left(\min_{\boldsymbol{\eta}}\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})\right),\\ 0 &= -\left\langle\sigma_i\sigma_j\right\rangle + \left[\mathbf{H}^{-1}_{\mathcal{G}_{\text{Bethe}}}\right]_{ij}\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} + \eta_i^*\eta_j^*.\end{aligned} \tag{3.38}$$

In order to obtain value of η_i^* that minimizes $\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})$, we perform minimization of Bethe free energy (3.31) respect to η_i. By solving the resulting equation, optimal values η_i^* must satisfy

$$\eta_i^* = \tanh\left(\frac{1}{2-n}\left[h_i - \frac{1}{2}\sum_{j/i}\sum_{\sigma_j=\pm1}\operatorname{arctanh}\left(\frac{\eta_i^*+\sigma_j\xi_{ij}\left(\eta_i^*,\eta_j^*,w_{ij}\right)}{1+\eta_j^*\sigma_j}\right)\right]\right), \tag{3.39}$$

where j/i means to sum over all spins except spins i. Now, Bethe - Gibbs Hessian matrix is given by

$$\left[\mathbf{H}_{\mathcal{G}_{\text{Bethe}}}\right]_{ij} \equiv \frac{\partial^2\mathcal{G}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j} = \begin{cases}\boldsymbol{\Phi}_i(\mathbb{W},\boldsymbol{\eta}) & i=j,\\ \boldsymbol{\Psi}_{ij}(w_{ij},\eta_i,\eta_j) & i\neq j,\end{cases} \tag{3.40}$$

where $\boldsymbol{\Phi}_i(\mathbb{W},\boldsymbol{\eta})$ and $\boldsymbol{\Psi}_{ij}(w_{ij},\eta_i,\eta_j)$ are given by

$$\begin{aligned}\boldsymbol{\Phi}_i(\mathbb{W},\boldsymbol{\eta}) &= \frac{2-n}{1-\eta_i^2} + \frac{1}{4}\sum_{j/i}\sum_{\sigma_i,\sigma_j=\pm1}\frac{\left(\sigma_j+\Gamma\left(w_{ij},\eta_i,\eta_j\right)\right)^2}{\alpha_{ij}\left(w_{ij},\eta_i,\eta_j,\sigma_i,\sigma_j\right)},\\ \boldsymbol{\Psi}_{ij}(w_{ij},\eta_i,\eta_j) &= \frac{1}{4}\sum_{\sigma_i,\sigma_j=\pm1}\frac{\left(\sigma_j+\Gamma\left(w_{ij},\eta_i,\eta_j\right)\right)\left(\sigma_i+\Gamma\left(w_{ij},\eta_j,\eta_i\right)\right)}{\alpha_{ij}\left(w_{ij},\eta_i,\eta_j,\sigma_i,\sigma_j\right)},\end{aligned} \tag{3.41}$$

where $\Gamma\left(w_{ij},\eta_i,\eta_j\right)$ and $\alpha_{ij}\left(w_{ij},\eta_i,\eta_j,\sigma_i,\sigma_j\right)$ are defined by

$$\begin{aligned}\Gamma\left(w_{ij},\eta_i,\eta_j\right) &= \frac{\eta_j-\eta_i\tanh\left(2w_{ij}\right)}{1-\tanh\left(2w_{ij}\right)\xi_{ij}\left(w_{ij},\eta_i,\eta_j\right)},\\ \alpha_{ij}\left(w_{ij},\eta_i,\eta_j,\sigma_i,\sigma_j\right) &= 1+\sigma_i\eta_i+\sigma_j\eta_j+\sigma_i\sigma_j\xi_{ij}\left(w_{ij},\eta_i,\eta_j\right).\end{aligned} \tag{3.42}$$

Finally, condition (3.26) now arises as

$$\langle\sigma_i\rangle = \eta_i^*, \tag{3.43}$$

$$\left[\mathbf{H}_{\mathcal{G}_{\text{Bethe}}}\right]_{ij}\Big|_{\boldsymbol{\eta}=\{\langle\sigma_i\rangle\}} = \left[\mathbf{C}^{-1}\right]_{ij}. \tag{3.44}$$

With these new relations, desired parameters $\{\hat{\mathbb{W}},\hat{\mathbb{H}}\}$ can be learned just by solving (3.44) respect to $\mathbb{W}$. To summarize learning process in Boltzmann machines using LRA is necessary to follow the following procedure:

Procedure 1 Boltzmann machine algorithm using LRA.

Require: $\langle\sigma_i\rangle$ and $[\mathbf{C}]_{ij}$.

Ensure: $[\mathbf{C}]_{ij}$ is not a singular matrix.

for $i,j=1$ **to** n **do**

From (3.44), solve non-linear equations

$$\boldsymbol{\Psi}_{ij}\left(w_{ij},\langle\sigma_i\rangle,\langle\sigma_j\rangle\right) = \left[\mathbf{C}^{-1}\right]_{ij}, \tag{3.45}$$

respect to w_{ij} for obtaining $\hat{w}_{ij}$.

end for

for $i=1$ **to** n **do**

Use solutions of $\hat{\mathbb{W}}$ to achieve $\hat{h}_i$ using

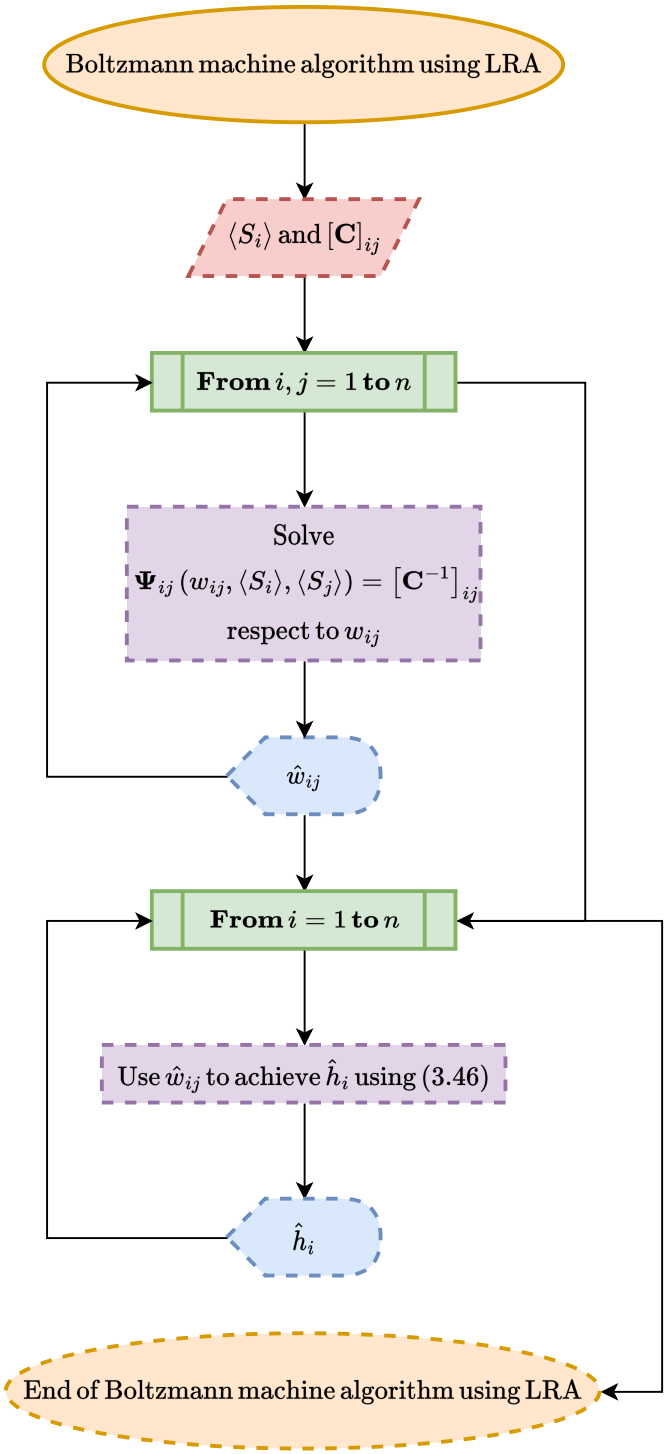
$$\hat{h}_i = (2-n)\,\text{arctanh}\left(\langle\sigma_i\rangle\right)+\frac{1}{2}\sum_{j/i}\sum_{c_j=\pm1}\text{arctanh}\left(\frac{\langle\sigma_i\rangle+c_j\xi_{ij}\left(\langle\sigma_i\rangle,\langle\sigma_j\rangle,\hat{w}_{ij}\right)}{1+c_j\langle\sigma_j\rangle}\right). \tag{3.46}$$

end for

return $\left\{\hat{\mathbb{W}},\hat{\mathbb{H}}\right\}$

Figure 3 resumes the procedure aformentioned exposed.

Figure 3 – Flowchart for procedure 1



Source: Author.

Advantage of learning parameters with the above procedure is that it does not require any iteration. Computationally, the cost of the learning algorithm is $\mathrm{O}(n^2)$ because of the computation of $[\mathbf{C}^{-1}]_{ij}$ [4]. However, in order to assure unique existence of

solutions, $\mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j)$ has the following properties:

1. $\mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j) = 0$ if and only if $[\mathbf{C}^{-1}]_{ij} = 0$ and therefore $\hat{w}_{ij} = 0$;
2. $\mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j)$ is a monotonic decreasing function. Meaning that solution can be found either deterministic or computationally;
3. Taking in consideration monotonicity of $\mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j)$, equation (3.45) possess solution for $\hat{w}_{ij}$ if and only if

$$-\frac{1}{2|\langle\sigma_i\rangle - \langle\sigma_j\rangle|} < \left[\mathbf{C}^{-1}\right]_{ij} < \frac{1}{2|\langle\sigma_i\rangle + \langle\sigma_j\rangle|}. \tag{3.47}$$

In principle, if the third condition is violated, algorithm (1) could not be used for learning desired parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$. However this difficulty can be overcome if we note that (3.47) could be written as

$$\lim_{w_{ij}\to\infty} \mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j) < \frac{\partial^2 \mathcal{G}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta})}{\partial\eta_i \partial\eta_j} < \lim_{w_{ij}\to-\infty} \mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j). \tag{3.48}$$

If we decompose Bethe - Gibbs free energy as $\mathcal{G}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) = \tilde{\mathcal{G}}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) + \mathcal{L}_B(\mathbb{W}, \boldsymbol{\eta})$ relation (3.48) turns

$$\lim_{w_{ij}\to\infty} \mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j) < \frac{\partial^2 \tilde{\mathcal{G}}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) + \mathcal{L}_B(\mathbb{W}, \boldsymbol{\eta})}{\partial\eta_i \partial\eta_j} < \lim_{w_{ij}\to-\infty} \mathbf{\Psi}_{ij}(w_{ij}, \eta_i, \eta_j), \tag{3.49}$$

where $\mathcal{L}_B(\mathbb{W}, \boldsymbol{\eta})$ is the Gibbs free energy associated with contributions coming from loops, that is, from diagonal elements of Hessian and $\tilde{\mathcal{G}}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta})$ is the naive Bethe - Gibbs free energy. Equation (3.49) tells us that (3.47) is satisfied if loop contributions are introduced in the learning algorithm. Modification of the learning algorithm taking into account the role of $\mathcal{L}_B(\mathbb{W}, \boldsymbol{\eta})$ is known as the diagonal problem [4, 5].

3.4 Improvement of Learning Algorithm using Diagonal Trick

Diagonal problem aforementioned is solved as follows. Let's introduce a new set of parameters $\boldsymbol{\gamma} \equiv \{\gamma_i\}$ to the Bethe - Gibbs free energy to redefine it as

$$\check{\mathcal{G}}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) = \mathcal{G}_{\text{Bethe}}(\mathbb{W}, \mathbb{H}, \boldsymbol{\eta}) - \frac{1}{2}\sum_i \gamma_i^2. \tag{3.50}$$

Due to the increment of parameters it is necessary to introduce n new constraints to determine $\boldsymbol{\gamma}$. New constraints must satisfy that values of the diagonal inverse of the Hessian matrix are fixed to the values of the exact ones. That is to say, possess the form

$$\left[\tilde{\mathbf{H}}_{\check{\mathcal{G}}_{\text{Bethe}}}\right]_{ii}\Bigg|_{\boldsymbol{\eta}=\{\eta_i^*\}} = 1-(\eta_i^*)^2, \tag{3.51}$$

where $\left[\tilde{\mathbf{H}}_{\check{\mathcal{G}}_{\text{Bethe}}}\right]_{ij}$ is defined by

$$\left[\tilde{\mathbf{H}}_{\check{\mathcal{G}}_{\text{Bethe}}}\right]_{ij} = \frac{\partial^2 \check{\mathcal{G}}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})}{\partial\eta_i\partial\eta_j} = \left[\mathbf{H}_{\mathcal{G}_{\text{Bethe}}}\right]_{ij} - \gamma_i\delta_{ij}. \tag{3.52}$$

Here η_i^* are the values minimizing $\check{\mathcal{G}}_{\text{Bethe}}(\mathbb{W},\mathbb{H},\boldsymbol{\eta})$ that must satisfy

$$\eta_i^* = \tanh\left(\frac{1}{2-n}\left[h_i - \frac{1}{2}\sum_{j/i}\sum_{\sigma_j=\pm1}\operatorname{arctanh}\left(\frac{\eta_i^*+\sigma_j\xi_{ij}\left(\eta_i^*,\eta_j^*,w_{ij}\right)}{1+\eta_j^*\sigma_j}\right) + \gamma_i\eta_i^*\right]\right). \tag{3.53}$$

Therefore conditions (3.43) and (3.44) are transformed into

$$\begin{aligned} \langle\sigma_i\rangle &= \eta_i^*, \\ \left[\tilde{\mathbf{H}}_{\check{\mathcal{G}}_{\text{Bethe}}}\right]_{ij}\Bigg|_{\boldsymbol{\eta}=\{\langle\sigma_i\rangle\}} &= \boldsymbol{\Psi}_{ij}(w_{ij},\langle\sigma_i\rangle,\langle\sigma_j\rangle) \\ &\quad + \left(\boldsymbol{\Phi}_i\left(w_{ij},\langle\sigma_i\rangle,\langle\sigma_j\rangle\right) - \gamma_i\right)\delta_{ij} = \left[\mathbf{C}^{-1}\right]_{ij}. \end{aligned} \tag{3.54}$$

With these relations the extended learning procedure can be established as follows:

Procedure 2 Boltzmann machine algorithm using LRA and Diagonal Trick.

Require: $\langle \sigma_i \rangle$ and $[\mathbf{C}]_{ij}$.

Ensure: $[\mathbf{C}]_{ij}$ is not a singular matrix.

for $i, j = 1$ **to** n **do**

From (3.44), solve non-linear equations

$$\mathbf{\Psi}_{ij}(w_{ij}, \langle \sigma_i \rangle, \langle \sigma_j \rangle) = \left[\mathbf{C}^{-1}\right]_{ij}, \tag{3.55}$$

respect to w_{ij} for obtaining $\hat{w}_{ij}$.

end for

for $i = 1$ **to** n **do**

Use $\hat{\mathbb{W}}$ to achieve diagonal parameters $\hat{\gamma}_i$ using

$$\hat{\gamma}_i = \mathbf{\Phi}_i \left(\hat{w}_{ij}, \langle \sigma_i \rangle, \langle \sigma_j \rangle\right) - \left[\mathbf{C}^{-1}\right]_{ii}. \tag{3.56}$$

end for

for $i = 1$ **to** n **do**

Use solutions of $\hat{\mathbb{W}}, \hat{\boldsymbol{\gamma}}$ to achieve $\hat{h}_i$ using

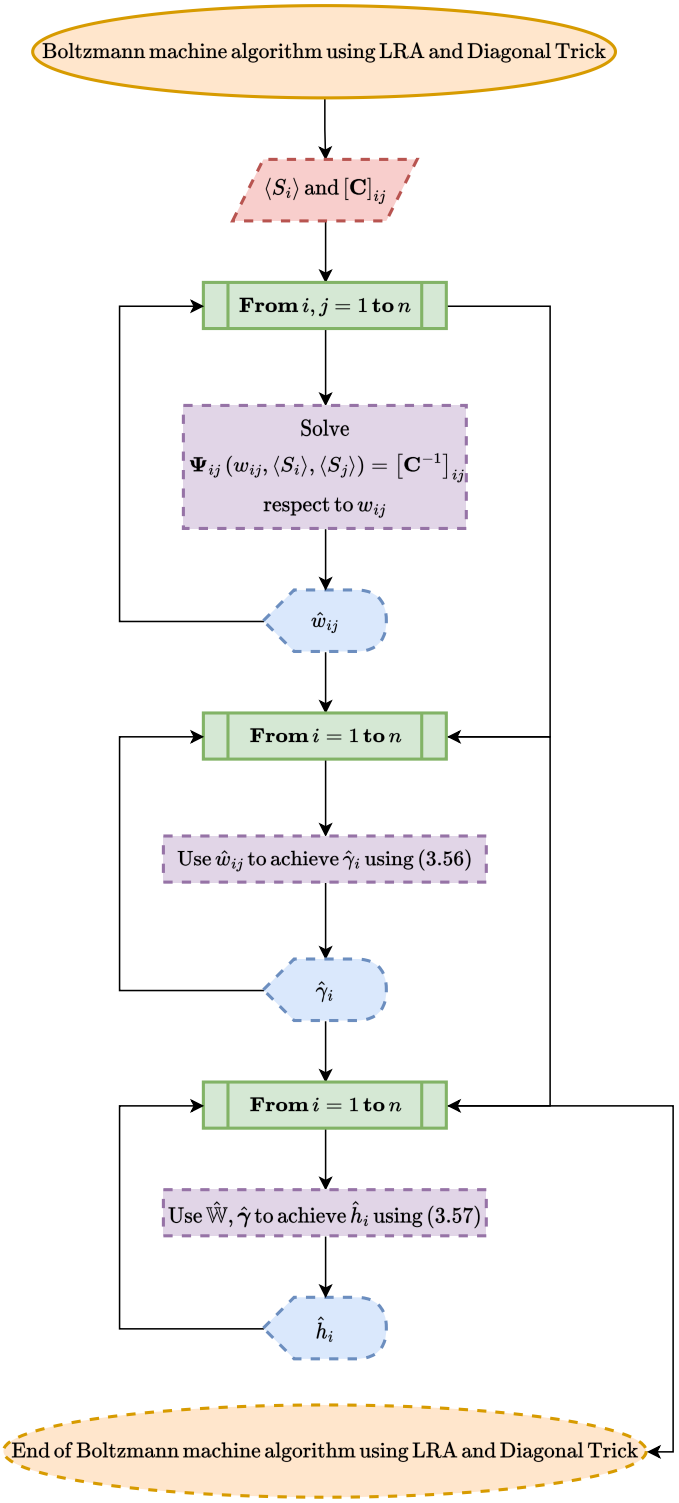
$$\hat{h}_i = (2-n)\,\mathrm{arctanh}\left(\langle \sigma_i \rangle\right) + \frac{1}{2}\sum_{j/i}\sum_{c_j=\pm 1} \mathrm{arctanh}\left(\frac{\langle \sigma_i \rangle + c_j \xi_{ij}\left(\langle \sigma_i \rangle, \langle \sigma_j \rangle, \hat{w}_{ij}\right)}{1 + c_j \langle \sigma_j \rangle}\right) + \hat{\gamma}_i \langle \sigma_i \rangle. \tag{3.57}$$

end for

return $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$

Computational cost of procedure 2 is equal to procedure 1. The extended learning algorithm improves mainly the inference of parameter $\hat{\mathbb{H}}$. In fact, extension of naive Bethe free energy along the diagonal problem is regarded as the higher - order approximation of the Bethe free energy [4, 6]. Figure 4 resumes the procedure aformentioned exposed.

Figure 4 – Flowchart for procedure 2

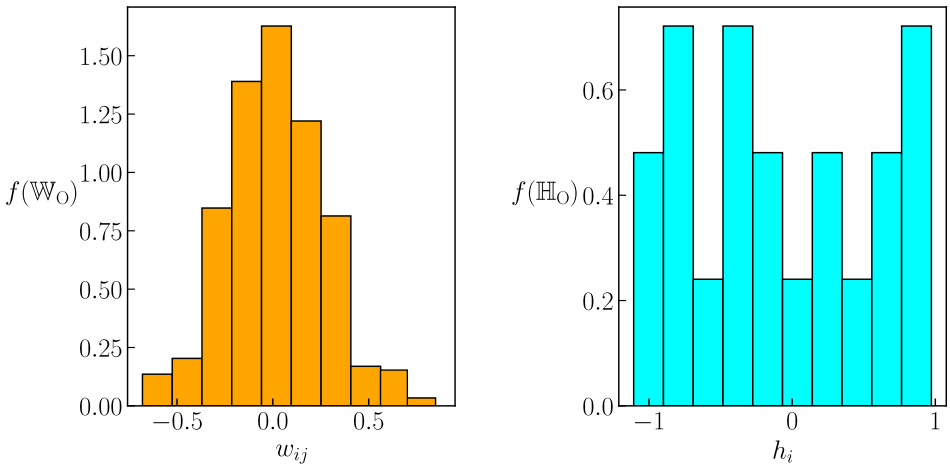


Source: Author.

3.5 Checking if the Learning Algorithm is Working Properly

In order to evaluate if the learning algorithm proposed in the last section is working properly, it is mandatory to prove that procedure (2) is capable of inferring desired parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ using a set of data points generated from a known set of observed parameters $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$ (subscript O stands for "observed"). With this aim, random data points were set up by sampling the Ising model [8] on a complete connected Boltzmann machine with n visible spins using Montecarlo simulation. First thing was to define the set of observed parameters $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$. To do that, random weights and local fields were sampled. For weights, normal distribution $w_{ij} \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 0.5$ were set up and for local fields, uniform distribution $h_i \sim \mathcal{U}(h_{\min}, h_{\max})$ where $h_{\min} = -1$ and $h_{\max} = 1$ were used. Reason for choosing these distributions resides in the fact that we want to resemble a spin glass system [9]. Observed weight parameter $\mathbb{W}_{\mathrm{O}}$ is represented as $n \times n$ symmetric matrix and $\mathbb{H}_{\mathrm{O}}$ arises as a n - one dimensional vector. For all simulations we fixed the number of visible spins to $n = 20$. Figure 5 shows the probability distribution function for a particular set $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$.

Figure 5 – Probability distribution function for a particular set of observed parameters $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$. For weights and local fields, it were used a normal distribution $w_{ij} \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 0.5$ and a uniform distribution $h_i \sim \mathcal{U}(h_{\min}, h_{\max})$ where $h_{\min} = -1$ and $h_{\max} = 1$, respectively. $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$ will be employed to test if procedure (2) is working properly.



Source: Author.

Subsequently, dynamics of Ising energy function

$$\mathcal{H}\left(\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\right) = -\sum_{i<j} w_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i,$$

was evaluated using the Metropolis algorithm [10]. As already established, $\sigma_i \in \{\pm 1\}$. For each sampling process, it were employed $t_{MC} = 10^4$ Montecarlo steps to reach thermal equilibrium and more $t_{MC} = 4 \times 10^3$ for storing the random data points. At the end, synthetic data consist in a $d_i^j \equiv n \times m$ binary matrix, that is, $d_i^j \in \{\pm 1\}$, where n is again associated to the number of visible spins and m stands as the Montecarlo steps employed to store the random data. It is important to highlight that given that the state probability configuration of Ising model depends on parameter β[3], synthetic data was generated for different values of β_{O} (again superscript O stands for "observed") ranging from $\beta_{\mathrm{O}} = 0.1$ to $\beta_{\mathrm{O}} = 0.9$, and for each sample and each β_{O} a new set $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$ was assigned. We use 10 samples for each β_{O}.

Two numerical techniques were employed to learn desired parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ using procedure 2. First, *Brent's method* [11] was implemented in a fixed range to find an initial hint for solutions and subsequently, that value was used as the initial guess of *Newton's method* [12]. First and second derivatives of the Hessian function (3.40) were computed to assure convergence of Newton method. Convergence tolerance was set up to be 2×10^{-12}.

After that, to verify how well procedure 2 is working, expected value $\langle \sigma_i \rangle$, covariance matrix $[\mathbf{C}]_{ij}$ and third order Ursell function $u(\sigma_i, \sigma_j, \sigma_k)$ [13] were computed in two cases. First case uses synthetic data produced using observed parameters $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$ and the second case employs inferred ones $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$. For the last case, synthetic data was generated by evaluating the Ising model at $\beta := K_B T = 1$[4]. This process was repeated for all samples in the range of β_{O} aforementioned. Expected value $\langle \sigma_i \rangle$ was defined in (3.1), covariance matrix $[\mathbf{C}]_{ij}$ is defined by

$$[\mathbf{C}]_{ij} = \langle \sigma_i \rangle \langle \sigma_j \rangle - \langle \sigma_i \sigma_j \rangle , \tag{3.58}$$

where $\langle \sigma_i \sigma_j \rangle$ was defined in (3.1) and the third order Ursell function $u(\sigma_i, \sigma_j, \sigma_k)$ is given by

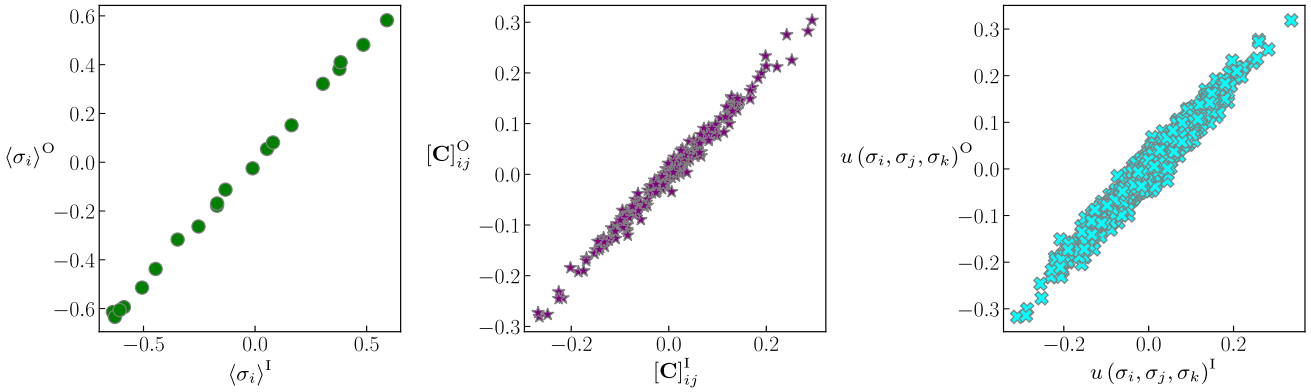
$$u(\sigma_i, \sigma_j, \sigma_k) = \langle \sigma_i \sigma_j \sigma_k \rangle - \langle \sigma_i \rangle \langle \sigma_j \sigma_k \rangle - \langle \sigma_j \rangle \langle \sigma_k \sigma_i \rangle - \langle \sigma_k \rangle \langle \sigma_i \sigma_j \rangle + 2 \langle \sigma_i \rangle \langle \sigma_j \rangle \langle \sigma_k \rangle . \tag{3.59}$$

[3] Known as inverse temperature. However, in this work we will refer to it just as the β parameter. In fact, this new parameter is hidden in Boltzmann machines, but as will be exposed below, it is possible to infer it too.

[4] Known as inverse operation temperature.

Figure 6 shows comparison of $\langle\sigma_i\rangle$, $[\mathbf{C}]_{ij}$ and $u\left(\sigma_i,\sigma_j,\sigma_k\right)$ when computed using observed - inferred parameters for a random sample of $\beta_{\mathrm{O}}=0.6$. Observed values are located in the vertical axis and inferred values are in the horizontal axis of each plot. Error bars indicate the discrepancy between inferred and observed values. O(I) superscripts are intended to differentiate between observed and inferred quantities.

Figure 6 – Comparison of quantities computed using observed and inferred parameters. Expected value $\langle\sigma_i\rangle$, covariance matrix $[\mathbf{C}]_{ij}$ and third order Ursell function $u\left(\sigma_i,\sigma_j,\sigma_k\right)$ were measured using data produced using observed parameters and data generated using inferred parameters at $\beta=1$. Observed parameters and data were sampled at $\beta_{\mathrm{O}}=0.6$. Superscript O(I) means for observed and inferred quantities. Error bars indicate the discrepancy between observed and inferred quantities.



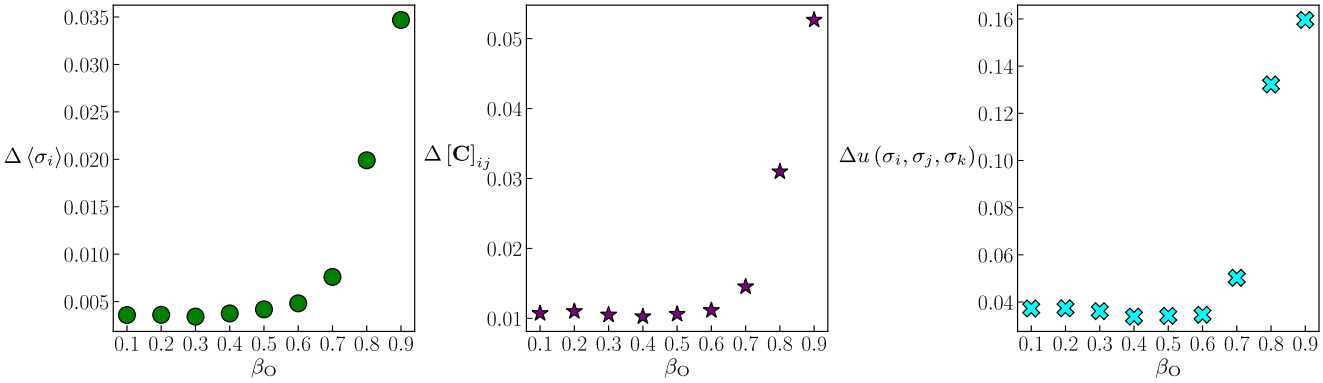
Source: Author.

As noted, error bars show that discrepancy between observed - inferred parameters are small. To better realize this, lets define the discrepancy over the three measured quantities as [6, 14]

$$\begin{aligned}
\Delta\langle\sigma_i\rangle &= \frac{1}{T}\sum_t \sqrt{\frac{1}{n^2}\sum_j\left(\langle\sigma_i\rangle_j^{\mathrm{O}}-\langle\sigma_i\rangle_j^{\mathrm{I}}\right)^2},\\
\Delta[\mathbf{C}]_{ij} &= \frac{1}{T}\sum_t \sqrt{\frac{1}{n^2}\sum_{i>j}\left([\mathbf{C}_{ij}]^{\mathrm{O}}-[\mathbf{C}_{ij}]^{\mathrm{I}}\right)^2},\\
\Delta u\left(\sigma_i,\sigma_j,\sigma_k\right) &= \frac{1}{T}\sum_t \sqrt{\frac{1}{n^2}\sum_l\left(u\left(\sigma_i,\sigma_j,\sigma_k\right)_l^{\mathrm{O}}-u\left(\sigma_i,\sigma_j,\sigma_k\right)_l^{\mathrm{I}}\right)^2},
\end{aligned} \tag{3.60}$$

where first summation is taken over all samples T. Discrepancy close to zero tell us that quantities were well reconstructed. Figure 7 shows the discrepancy in function of β_{O} parameter for each quantity already computed.

Figure 7 – Discrepancy among observed and inferred quantities in function of β_{O}. Discrepancies exhibit an increase in error in reconstruction of quantities along β_{O}. However, overall results for all samples proves that procedure (2) is working properly at the time to reconstruct $\langle\sigma_i\rangle$, $[\mathbf{C}]_{ij}$ and $u(\sigma_i,\sigma_j,\sigma_k)$. Maximum of discrepancy for $\langle\sigma_i\rangle$ was about 0.4%, for $[\mathbf{C}]_{ij}$ was below 6% and for $u(\sigma_i,\sigma_j,\sigma_k)$ is close to 16%. Maximum value of discrepancy in third order Ursell function is associated with the learning limit derives from the LRA approximation.



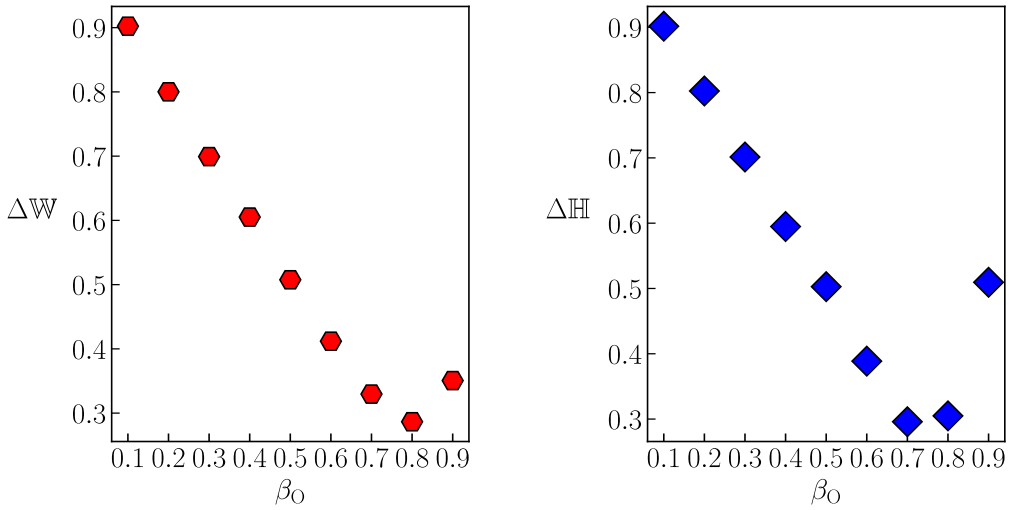
Source: Author.

In each plot, it is observed that discrepancy for all computed quantities increases with β_{O}. In other words, the learning algorithm works better in reconstructing $\langle\sigma_i\rangle$, $[\mathbf{C}]_{ij}$ and $u(\sigma_i,\sigma_j,\sigma_k)$ for low values of β_{O}. However, overall results for all samples show that in general, procedure (2) is working properly. Maximum of discrepancy for $\langle\sigma_i\rangle$ was about 0.4%, for $[\mathbf{C}]_{ij}$ below 6% and finally for $u(\sigma_i,\sigma_j,\sigma_k)$ was close to 16%. High value of discrepancy in third order Ursell function can be associated with an upper limit in learning derived from the fact that we are working with the LRA approximation. In the same way, lets define the discrepancy $\Delta\mathbb{W}$ and $\Delta\mathbb{H}$ as [6, 14]

$$\Delta\mathbb{W}=\frac{1}{T}\sum_{t}\sqrt{\frac{\sum_{i>j}\left(w_{ij}^{\mathrm{O}}-w_{ij}^{\mathrm{I}}\right)^{2}}{\sum_{i>j}\left(w_{ij}^{\mathrm{O}}\right)^{2}}},\qquad \Delta\mathbb{H}=\frac{1}{T}\sum_{t}\sqrt{\frac{\sum_{j}\left(h_{ij}^{\mathrm{O}}-h_{ij}^{\mathrm{I}}\right)^{2}}{\sum_{j}\left(h_{ij}^{\mathrm{O}}\right)^{2}}},\tag{3.61}$$

where again the average is taken over all samples and O and I stands as observed and inferred parameters, respectively. In this case, parameter discrepancy close to one express that inferred parameters were not well reconstructed. Figure 8 shows the parameters discrepancy computed in all ranges of β_{O}.

Figure 8 – Discrepancy of parameters in function of β_O. For both parameters, its discrepancy decreases as β_O grows. Note that, for $\beta_O \geq 0.7$ behaviour change and discrepancy increases again. Apparently, high values of discrepancy lead us to conclude that procedure (2) is not working properly in reconstructing parameters $\{\mathbb{W}_O, \mathbb{H}_O\}$. However, this apparent mistake indicates that a hidden feature in the learning algorithm has not been taken into account.



Source: Author.

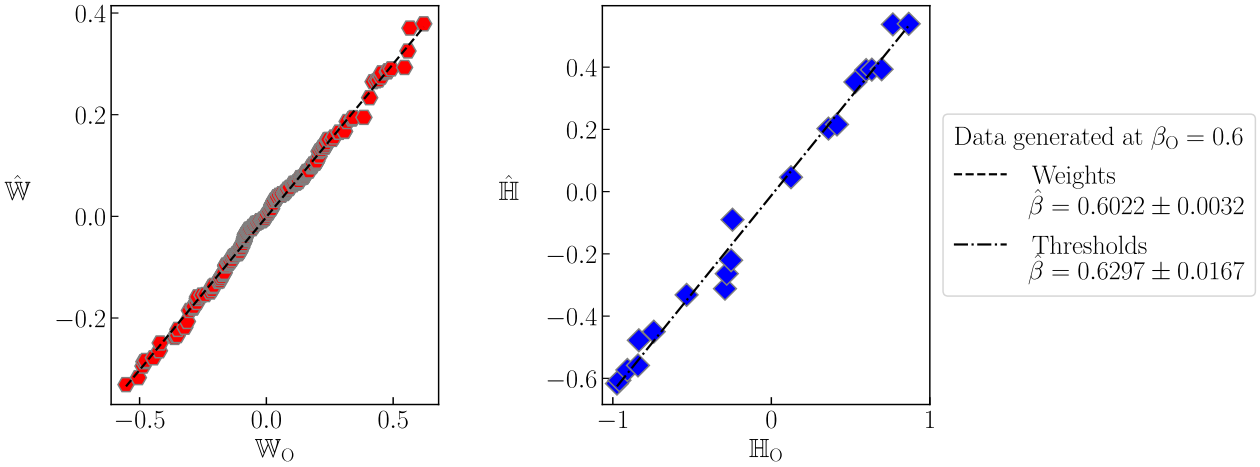
Discrepancy in parameters exposes a decreasing behaviour by increasing β_O. However, for $\beta_O \geq 0.7$, behaviour changes and discrepancy increases again. The most highlighted feature of the previous result is the highest discrepancy values for all β_O. This could lead us to conclude that the learning algorithm (2) does not work properly when reconstructing the observed parameters. However, this apparent mistake can be easily understood and overcome. High values of discrepancy in parameters must be due to a hidden feature that the learning algorithm does not cover. In other words, values of discrepancy can be decreased if that hidden feature is taken into account in the learning algorithm.

3.6 Learning β: A hidden feature in Boltzmann machines

In extremal conditions (3.24) and (3.25), it was assumed that $\beta = K_B T = 1$. However, this is just a trick to overcome the lack of equations needed to obatin analytically β. Main problem resides in the fact that β is a free parameter and therefore may take any real value. To figure out how to learn this new parameter, results of inferred parameters were checked by hand. This process led us to realize that high values in parameter discrepancy is due to the lack of scaling in learned parameters by β_O in which observed parameters were generated. In other words, to the best of our knowledge, there is a scale relation between observed and inferred parameters and seemingly this properly is hidden.

In order to make this property visible, first thing made was to compute the expected value of inferred - observed ratio for weights and local fields, i.e., $E\left[\hat{\mathbb{W}}/\mathbb{W}_{\mathrm{O}}\right]$ and $E\left[\hat{\mathbb{H}}/\mathbb{H}_{\mathrm{O}}\right]$. As an example, for a random sample of $\beta_{\mathrm{O}} = 0.3$ results were $E\left[\hat{\mathbb{W}}/\mathbb{W}_{\mathrm{O}}\right] = 0.3123 \pm 0.1216$ and $E\left[\hat{\mathbb{H}}/\mathbb{H}_{\mathrm{O}}\right] = 0.28698 \pm 0.0742$. Note that these two values are very close to the original β_{O}. This implies that if there is a scale relation between inferred and observed parameters, ratio between them must be equal to β_{O}. In order to prove that, a linear fitting process was carried out between $\hat{\mathbb{W}}$ vs. $\mathbb{W}_{\mathrm{O}}$ and $\hat{\mathbb{H}}$ vs. $\mathbb{H}_{\mathrm{O}}$ for each observed β_{O}. If the previous assertion is true, the angular coefficient obtained from the fitting process must correspond to the observed β_{O}. Figure 9 shows the linear fitting for a random sample of $\beta_{\mathrm{O}} = 0.6$. Inferred parameters are located in the vertical axis and observed parameters are in the horizontal axis of each plot.

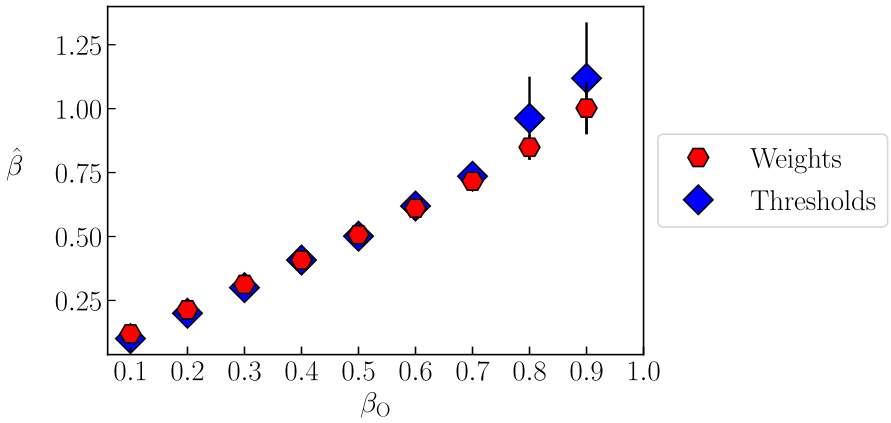
Figure 9 – Linear fitting between inferred and observed parameters. A random sample of $\beta_{\mathrm{O}} = 0.6$ was chosen. Inferred and observed parameters are located in the vertical and horizontal axis, respectively. Black dashed line represents linear fitting for weights and black dash-dotted line is associated with the linear fitting of local fields. Angular coefficient of linear fitting corresponds to the observed β_{O} in which synthetic data was generated.



Source: Author.

The linear fitting process (black dashed line for weights and black dash-dotted line for local fields) shows explicitly that angular coefficient (named as $\hat{\beta}$) corresponds to β_{O} in which the synthetic data was generated. In order to better conclude that, the linear fitting process was carried out for all samples and the average value of $\hat{\beta}$ was computed. Figure 10 shows angular coefficients obtained from linear fitting for each β_{O}.

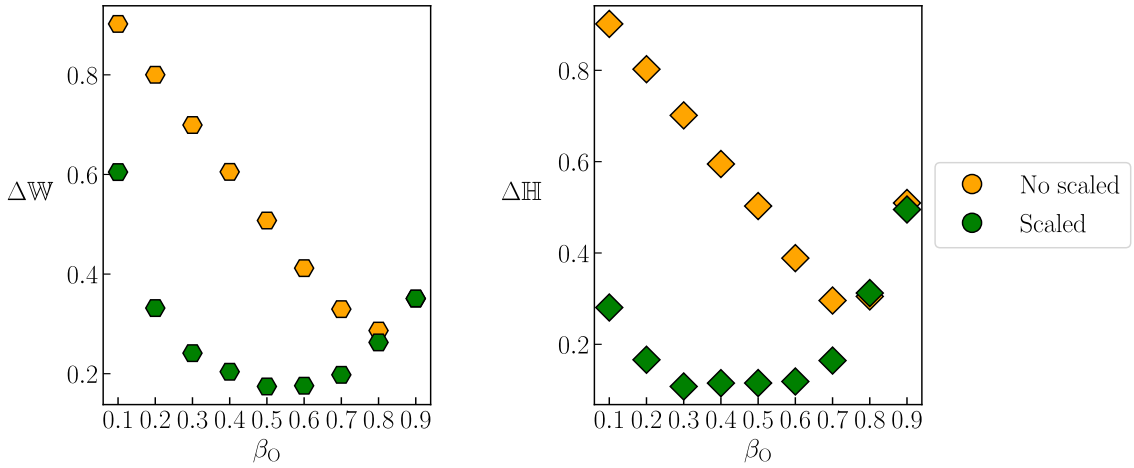
Figure 10 – Average angular coefficients in function of β_O. In order to prove that a scaling relation between inferred and observed parameters exists, a linear fitting between them for all samples were made. The average angular coefficient shows to be close to the original β_O in which synthetic data was generated. Error bars show the discrepancy between real and inferred values of β.



Source: Author.

From the Figure, we can note that in each case, the average value of angular coefficient obtained from linear regression is close to the original β_O in which synthetic data was generated. Previous results establish that despite inferred parameters are learned assuming $\hat{\beta} = 1$, the actual inferred parameters must be scaled by $\hat{\beta}$ in which its synthetic data was generated. If this is true, discrepancy of parameters must decrease. Figure 11 shows the parameters discrepancies (3.61) computed using both no scaling and scaling by $\hat{\beta}$.

Figure 11 – Scaled discrepancy of parameters in function of β_O. Green color recognizes scaled discrepancies and yellow color is for non-scaled discrepancies. A considerably decreasing in parameters discrepancies is observed when they were scaled by the average inferred angular coefficient $\hat{\beta}$. This proves that inferred parameters need to be scaled by $\hat{\beta}$ in order to reconstruct well the observed ones.

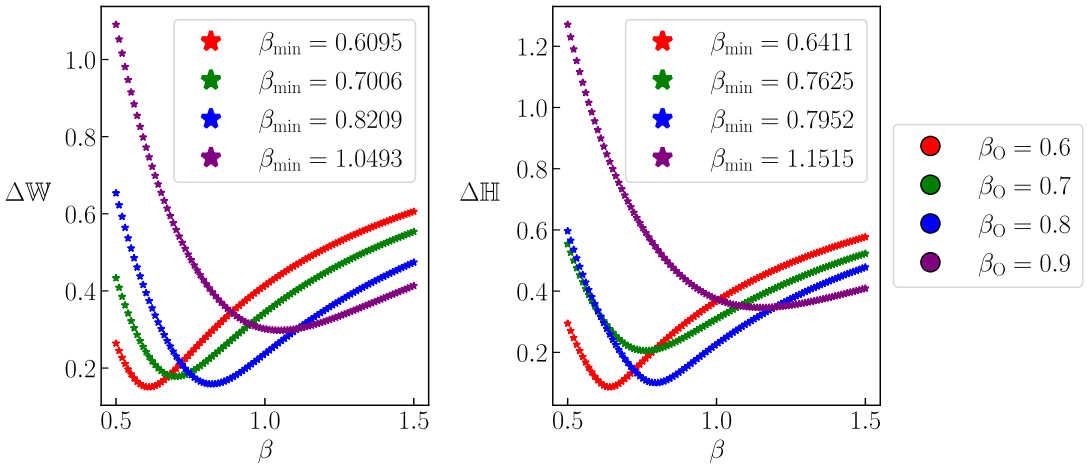


Source: Author.

As expected, the above Figure exposes the fact that discrepancy must decrease when learned parameters are scaled by $\hat{\beta}$. Note that values of green markers (scaled

discrepancy) are always below from the yellow ones (no scaled discrepancy), except for $\beta_{\mathrm{O}} = 0.8, 0.9$ where discrepancy remains almost equal. In this sense, to overcome the situation for $\beta_{\mathrm{O}} = 0.8, 0.9$, it is sufficient to show that discrepancy in parameters must be a minimum in $\hat{\beta}$. To do that, scaling of inferred parameters set was carried out in a range $\beta = 0.1$ to $\beta = 1.5$ and subsequently, the discrepancy (3.61) was again computed. Figure 12 shows the result from random samples of data generated among $\beta_{\mathrm{O}} = 0.6$ to $\beta_{\mathrm{O}} = 0.9$.

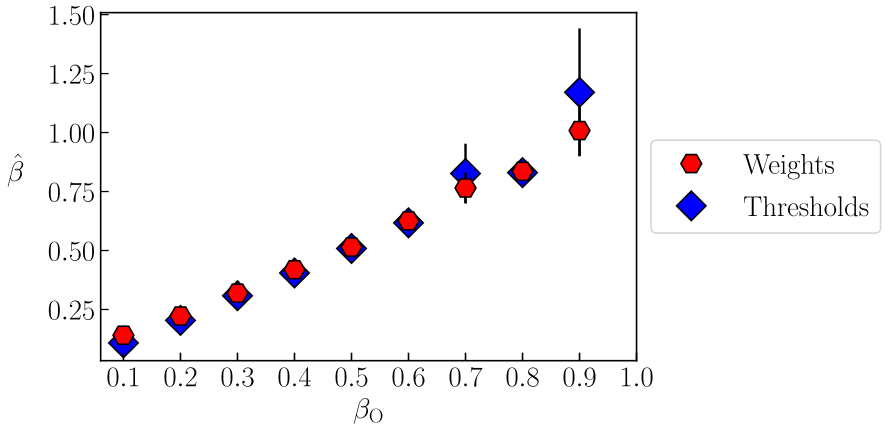
Figure 12 – Dependence of parameter discrepancy with respect to β. Inferred parameters were scaled in a fixed range and the discrepancy (3.61) was computed. If the scaling relation between observed and inferred parameters exists, then discrepancy must be a minimum in β_{O}. Each line in the plot belongs to a random sample of data generated from $\beta_{\mathrm{O}} = 0.6$ to $\beta_{\mathrm{O}} = 0.9$.



Source: Author.

As observed, in each sample, the minimum of discrepancy is close to β_{O} in which parameters were learned. For the remaining samples, minimum of discrepancy was obtained and the average by β_{O} was computed as exposed in Figure 13. As noted in the Figure, average minimum of $\hat{\beta}$ is close to the real observed β_{O}, besides, errors bars got evident as $\beta_{\mathrm{O}} \to 1$.

Figure 13 – Average minimum of scaled discrepancy in function of β_{O}. For each sample, discrepancy among observed and scaled inferred parameters was computed. Subsequently, the minimum was attained and the average by each β_{O} was computed. Error bars show how far the inferred $\hat{\beta}$ is from β_{O}. Average minimum of $\hat{\beta}$ is close to the real observed β_{O}.



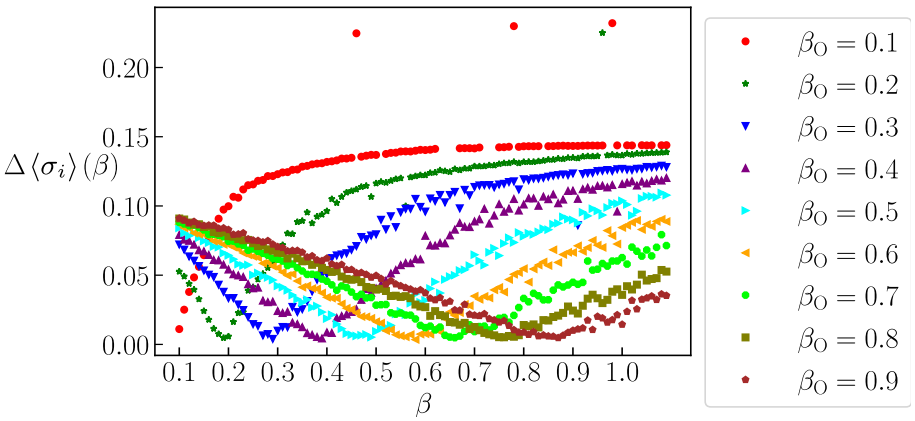
Source: Author.

Previous results lead us to conclude that learned parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ must be scaled by $\hat{\beta}$ in order to reconstruct well observed parameters $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$. Thing now is that in real applications, we just only having the empirical data and no observed parameters, and then the question raising is, **how β can be learned using only the empirical data?** In order to answer this question, random states were generated using the aforementioned Montecarlo process. Instead of using observed parameters for the sampling process, we used inferred parameters. In this case, inferred parameters were scaled by β. The sampling process were carried out in a range $\beta_{\min} = 0.1$ to $\beta_{\max} = 1.0$ and in each step the discrepancy $\Delta \langle \sigma_i \rangle (\beta)$ defined as

$$\Delta \langle \sigma_i \rangle (\beta) = \sqrt{\frac{1}{n^2} \sum_j \left(\langle \sigma_i \rangle_j^{\mathrm{O}} - \langle \sigma_i \rangle_j^{\mathrm{I}} (\beta) \right)^2}, \tag{3.62}$$

was computed. Here, $\langle \sigma_i \rangle^{\mathrm{O}}$ stands as the observed real expected value and $\langle \sigma_i \rangle^{\mathrm{I}} (\beta)$ is the expected value computed using random states generated at β. Figure 14 shows the discrepancy for each β_{O}.

Figure 14 – Discrepancy of expected value $\langle \sigma_i \rangle$ as a function of β. Each set of inferred parameters were scaled in β which synthetic data was generated. Subsequently, Metropolis sampling using the scaled parameters was carried out in order to generate random states in a range $\beta_{\text{min}} = 0.1$ to $\beta_{\text{max}} = 1.0$. In each step the discrepancy between the real expected value and the expected value obtained from the sampling process was computed. Minimum of discrepancy is shown to be at the observed β_{O} in whom observed parameters were originated.



Source: Author.

Figure visually exposes that minimum discrepancy is close to β_{O} in all cases. This ensures that a based expected value procedure can be established in order to learn $\hat{\beta}$. In fact, a complete procedure must include fluctuations of expected values $\langle \sigma_i \rangle$ as a function of β. To this aim, let's introduce the first and second order fluctuation values $\chi^1(\mathbb{W}, \mathbb{H}, \beta)$, $\chi^2(\mathbb{W}, \mathbb{H}, \beta)$ as[5]

$$\begin{aligned} \chi^1(\mathbb{W}, \mathbb{H}, \beta) &= \frac{\partial^2}{\partial h_i^2} \left(-\beta \ln Z_B \left(\mathbb{W}, \mathbb{H}, \beta \right) \right), \\ \chi^2(\mathbb{W}, \mathbb{H}, \beta) &= \frac{\partial^3}{\partial h_i^3} \left(-\beta \ln Z_B \left(\mathbb{W}, \mathbb{H}, \beta \right) \right), \end{aligned} \tag{3.63}$$

where $Z_B \left(\mathbb{W}, \mathbb{H}, \beta \right)$ is the partition function. Taking into account these quantities, we propose the following procedure to learn $\hat{\beta}$.

[5] Known in physics as first and second order susceptibility, respectively.

Procedure 3 Boltzmann machine algorithm for learning $\hat{\beta}$

Require: $\langle\sigma_i\rangle$, $[\mathbf{C}]_{ij}$ and $u(\sigma_i, \sigma_j, \sigma_k)$;

Require: $\hat{\mathbb{W}}, \hat{\mathbb{H}}$.

stop $\leftarrow$ *false*

scale $\leftarrow 1$

$\beta_{\text{aux}} \leftarrow 0$

$\tau \leftarrow 1 \times 10^{-9}$ ▷ Convergence parameter

while stop = *false* **do**

$Z_B\left(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right) \leftarrow Z_B\left(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right)$ /scale

Solve

$$\langle\langle\sigma_i\rangle\rangle = \frac{\partial}{\partial h_i}\left(-\hat{\beta}\ln Z_B\left(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right)\right), \tag{3.64}$$

respect to $\hat{\beta}$ to get β_1.

if $|\beta_1 - \beta_{\text{aux}}| \leq \tau$ **or** $\beta_1 \leq 0$ **then**

stop $\leftarrow$ *true*

else if $|\beta_1 - \beta_{\text{aux}}| \geq \tau$ **then**

stop $\leftarrow$ *false*

scale $\leftarrow 1 - \beta_1$

$\beta_{\text{aux}} \leftarrow \beta_1$

end if

end while

Use β_1 to compute χ^1 and χ^2

$$\chi^1 = \frac{\beta_1}{n}\sum_{i<j}[\mathbf{C}]_{ij},$$

$$\chi^2 = \frac{\beta_1^2}{n}\sum_{i<j<k}\langle\sigma_i\sigma_j\sigma_k\rangle,$$

where n stands by number of visible spins.

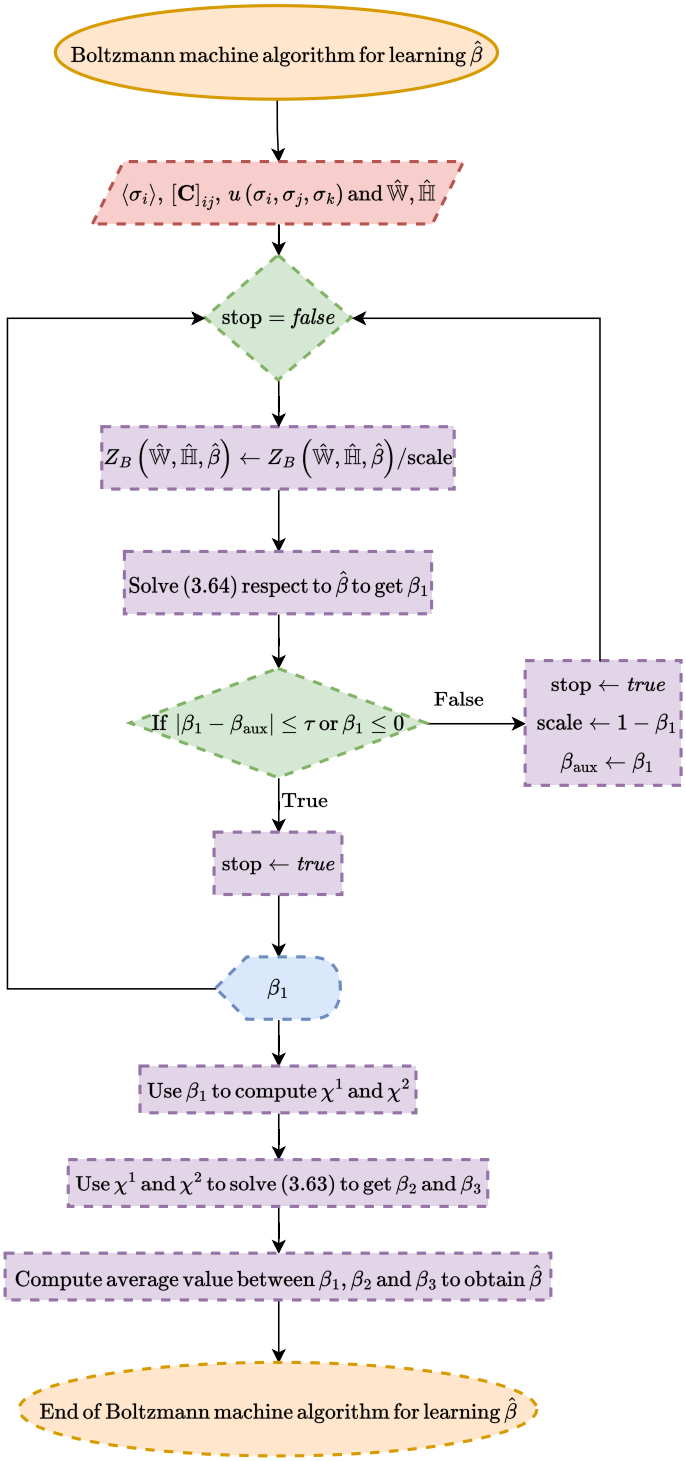
Solve equations (3.63) respect to $\hat{\beta}$ in order to get β_2 and β_3.

return $\hat{\beta} = \text{mean}(\beta_1, \beta_2, \beta_3)$ ▷ mean operator computes average value

Proposed procedure uses mainly the partition function to learn $\hat{\beta}$. Note that in each iteration, the partition function is scaled by the solution satisfying that the expected value of $\langle\sigma_i\rangle$ when computed using $Z_B\left(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right)$ is equal to the real observed. This loop ends when convergence is achieved. Convergence is measured as the difference between

actual and previous solution of β_1. After this loop, β_1 is used to compute first and second fluctuations values and equations (3.63) are solved separately to find β_2 and β_3. Finally, the learned parameter $\hat{\beta}$ is attained by computing the average value among all solutions. Figure 15 resumes the aformentioned procedure.

Figure 15 – Flowchart for procedure 3

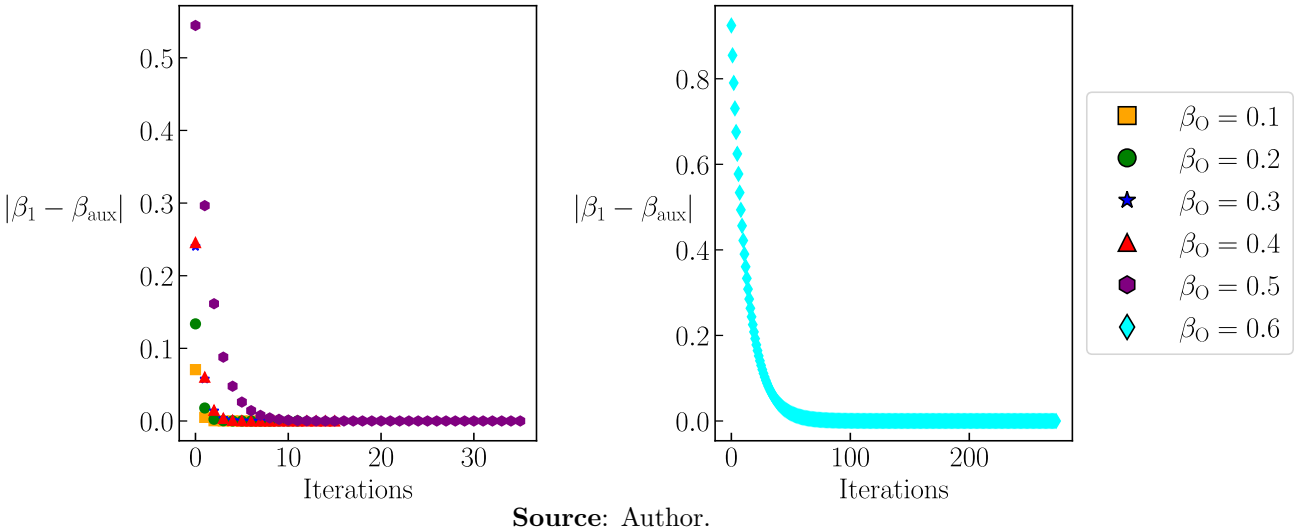


Source: Author.

In order to test if the algorithm is working properly, synthetic data was generated using the Montecarlo approach as already exposed. In each β_O and each set of synthetic data, $\left\{\hat{\mathbb{W}},\hat{\mathbb{H}}\right\}$ were inferred using procedure 2. Subsequently, covariance matrix (3.58) and third Ursell function (3.59) were computed to estimate the value of $\hat{\beta}$. In order to solve equations established in procedure 3, we use a root-finding method. Specifically, we use

Powell's method [15] with a convergence tolerance of 2×10^{-12}. First thing mandatory to prove is the convergence condition to find β_1. For all simulations, we set the convergence parameter to be $\tau = 1 \times 10^{-9}$. Figure 16 shows the evaluation of convergence parameter in each iteration for a random sample in different values of β_{O}.

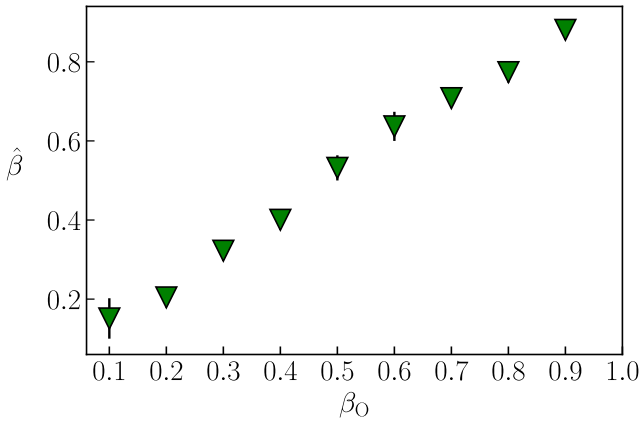
Figure 16 – Convergence condition of procedure 3. In order to prove that the procedure is working properly, synthetic data was generated, weights and local fields were inferred using procedure 2 and statistical quantities were measured. For $\beta_{\mathrm{O}} = 0.1$ to $\beta_{\mathrm{O}} = 0.6$ the quantity of iterations increase with β_{O}, meaning that the convergence condition is mediated by the difference $|\beta_1 - \beta_{\mathrm{aux}}|$. For the remaining range $\beta_{\mathrm{O}} = 0.7$ to $\beta_{\mathrm{O}} = 0.9$, it was realized that convergence condition works taking into account the second condition of the if statement.



Source: Author.

As observed, for $\beta_{\mathrm{O}} = 0.1$ to $\beta_{\mathrm{O}} = 0.6$, convergence condition is mainly mediated by the evaluation of difference $|\beta_1 - \beta_{\mathrm{aux}}|$. In fact, the quantity of iterations grows as β_{O} approaches to $\beta_{\mathrm{O}} = 0.6$. For the remaining range $\beta_{\mathrm{O}} = 0.7$ to $\beta_{\mathrm{O}} = 0.9$ we realize that the convergence condition was mediated by the second condition in the if statement, that means, convergence condition stops to scale partition function because solutions of β_1 are forbidden. Verified the convergence condition of procedure 3, the next step carried out was to implement it by complete. To do that, for each β_{O}, we implement a framework in which random weights, local fields and synthetic data are generated, inferred parameters are computed using procedures 2 and 3 and if the difference $\left|\beta_{\mathrm{O}} - \hat{\beta}\right|$ is less or equal to 1×10^{-2} results are saved. This process is iteratively repeated until a total amount of ten samples are attained by each β_{O}. Figure 17 shows the comparison between $\hat{\beta}$ and β_{O}.

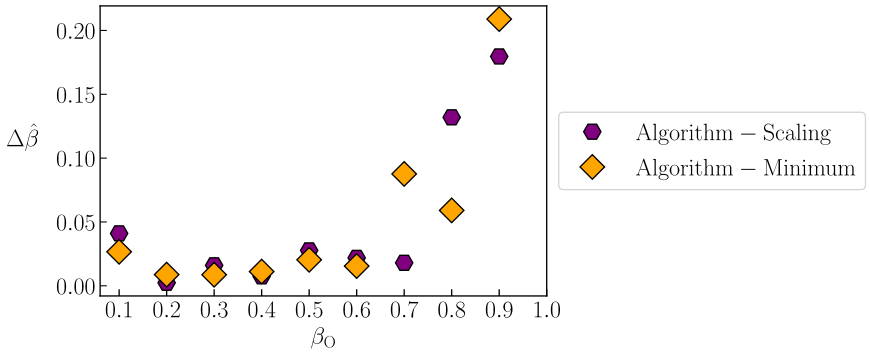
Figure 17 – Comparison between $\hat{\beta}$ and β_{O} obtained using proposed procedure. Synthetic data was generated for different values of β_{O}. Parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ were inferred using procedure 2 and $\hat{\beta}$ was learned using procedure 3. Each point is the average of almost ten samples by each β_{O}. Linear relation among observed β_{O} and learned $\hat{\beta}$ shows that procedure 3 can be implemented to effectively infer the last parameter needed for a complete Boltzmann machine description.



Source: Author.

In the plot, each point corresponds to the average of almost ten samples. Error bar shows the discrepancy between the $\hat{\beta}$ learned and β_{O}. Previous Figure shows that procedure 3 can be implemented as a methodology for learning the β_{O} parameter in which data was generated. Note the linear relation between learned and observed quantities. A last characterization of the effectiveness of procedure 3 can be carried out by computing the difference of $\hat{\beta}$ with the values obtained via scaling hypothesis (see Figure 10) and minimum scaled discrepancy (see Figure 13) as shown in Figure 18.

Figure 18 – Final comparison for $\hat{\beta}$. Purple hexagon represents the difference between $\hat{\beta}$ and the average angular coefficients obtained from the observed-inferred scaling. Orange diamonds are related to the difference between $\hat{\beta}$ and the average minimum value obtained from the minimization of the observed-inferred discrepancy. Difference remains small in the range, for extreme values the difference turns greater.



Source: Author.

In Figure, purple hexagon represents the difference between $\hat{\beta}$ and the average angular coefficient obtained by scaling observed and inferred parameters. In the same

way, orange diamonds are associated with the difference between $\hat{\beta}$ and the average value obtained from the minimization with respect to β of the discrepancy between observed-inferred parameters. Note from $\beta = 0.2$ to $\beta = 0.7$ difference remains small, almost zero. In contrast, for extreme values ($\beta \longrightarrow 0$ or $\beta \longrightarrow 1$) difference turns greater, implying that the proposed algorithm is more sensible and therefore, learning turns complex, therefore, an improvement of the proposed algorithm must be made in order to overcome the extreme values sensibility.

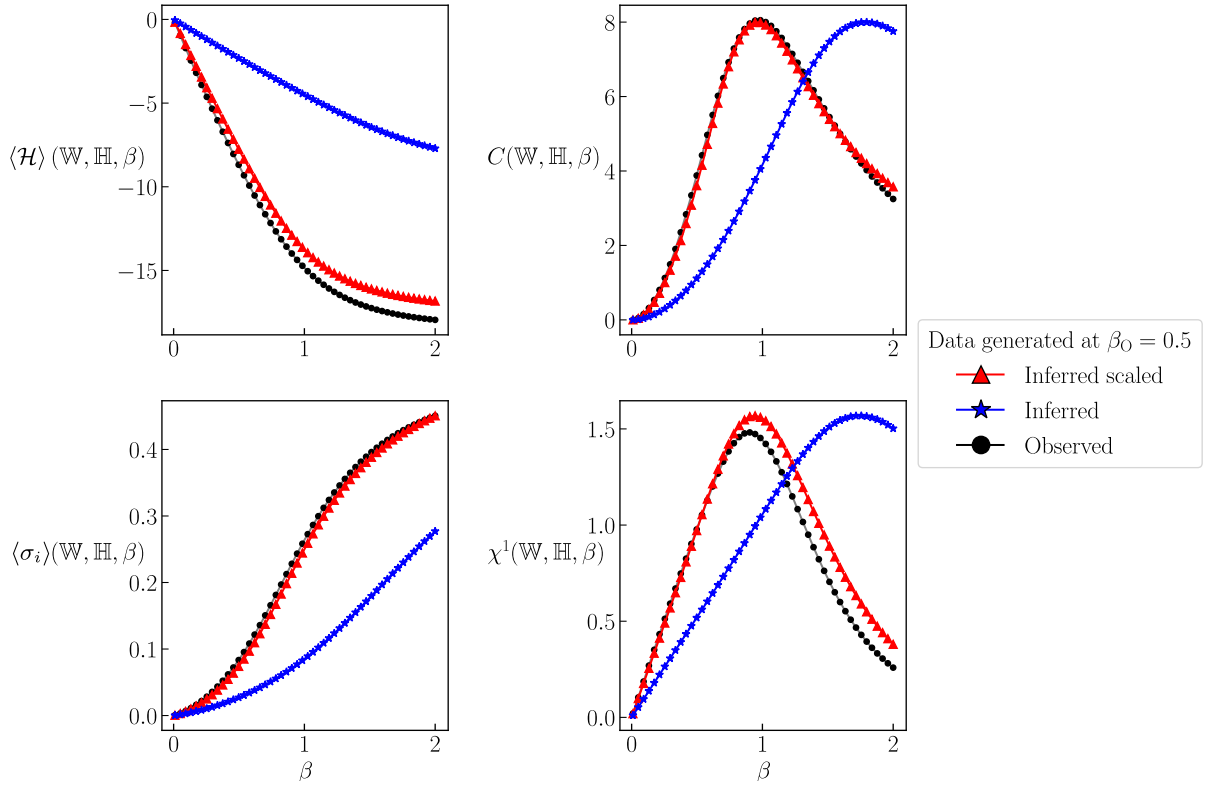
Now, in order to prove that a set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ of inferred parameters are successfully describing all properties of the real observed data, the next step was to characterize properties derived from the partition function. Explicitly, it were computed the dependence in β of energy function average value $\langle \mathcal{H} \rangle (\mathbb{W}, \mathbb{H}, \beta)$, its fluctuation value $C(\mathbb{W}, \mathbb{H}, \beta)$[6], average value $\langle \sigma_i \rangle (\mathbb{W}, \mathbb{H}, \beta)$ and its first order fluctuation $\chi^1(\mathbb{W}, \mathbb{H}, \beta)$ defined by

$$\begin{aligned} \langle \mathcal{H} \rangle (\mathbb{W}, \mathbb{H}, \beta) &= -\frac{\partial}{\partial \beta} \left(\ln Z_B (\mathbb{W}, \mathbb{H}, \beta) \right), \\ C(\mathbb{W}, \mathbb{H}, \beta) &= \frac{\partial^2}{\partial \beta^2} \left(\ln Z_B (\mathbb{W}, \mathbb{H}, \beta) \right), \\ \langle \sigma_i \rangle (\mathbb{W}, \mathbb{H}, \beta) &= \frac{\partial}{\partial h_i} \left(-\beta \ln Z_B (\mathbb{W}, \mathbb{H}, \beta) \right), \\ \chi^1(\mathbb{W}, \mathbb{H}, \beta) &= \frac{\partial^2}{\partial h_i^2} \left(-\beta \ln Z_B (\mathbb{W}, \mathbb{H}, \beta) \right). \end{aligned} \tag{3.65}$$

To do this task, we compute these quantities in two cases: using observed parameter set $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}\}$ and inferred parameter set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$. However, in the second case, we compute the characterization in another two cases. The first one uses the inferred parameters and the second scales $\hat{\mathbb{W}}$ and $\hat{\mathbb{H}}$ by $\hat{\beta}$ learned. The characterization for all quantities in all cases was made in a fixed range $\beta_{\min} = 0.01$ to $\beta_{\max} = 2$. To compute the partition function, we generate all 2^n possible states of $\{\vec{\sigma}\}$ to evaluate each quantity in (3.65). Figure 19 shows the result of the characterization from a random sample of synthetic data generated in $\beta_{\mathrm{O}} = 0.5$.

[6] Known in physics as specific heat.

Figure 19 – Partition function characterization of observed and inferred parameters. Quantities (3.65) were computed using three cases. The first one uses the observed set $\{\mathbb{W}_O, \mathbb{H}_O\}$ (black-dotted line), the second uses the inferred set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ (blue-starred line) and the last case employs the inferred parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ scaled by $\hat{\beta}$ (red-triangled line). Note that inferred scaled values reconstruct better the characterization when compared with the observed case.



Source: Author.

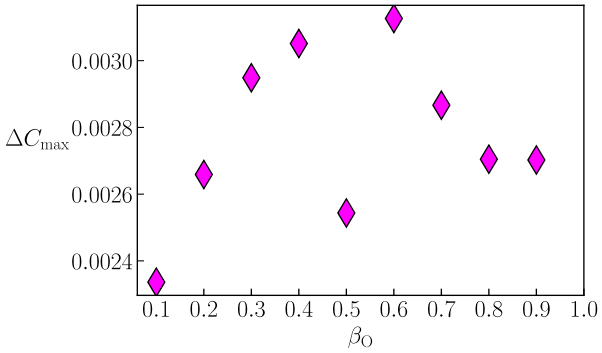
In each panel, black-dotted line corresponds to the characterization computed using observed set $\{\mathbb{W}_O, \mathbb{H}_O\}$, while the blue-stared line represents the characterization using inferred set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ and finally, the blue-triangled line is associated with the characterization obtained using $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ but scaled by $\hat{\beta}$. Note that scaled inferred values seem to reconstruct better the characterization of observed data in comparison with the non-scaled case. In order to assure this conclusion, we computed the error of maximum value of $C(\mathbb{W}, \mathbb{H}, \beta)$ between observed and inferred-scaled characterization defined by

$$\Delta C_{\max} = \frac{1}{T}\sum_{T}\left|\frac{C_{\max}\left(\mathbb{W}_O, \mathbb{H}_O, \beta_O\right) - C_{\max}\left(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right)}{C_{\max}\left(\mathbb{W}_O, \mathbb{H}_O, \beta_O\right)}\right|, \tag{3.66}$$

where $C_{\max}(\mathbb{W}_O, \mathbb{H}_O, \beta_O)$, $C_{\max}\left(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right)$ are the maximum values of $C(\mathbb{W}, \mathbb{H}, \beta)$ in the observed, inferred-scaled case, respectively and T is the number of samples by each β_O. We did this task in a similar way with the learning of β_O. However, in this case,

we add a new step in the computation related to finding the maximum value and again results were saved if the difference $\left|\beta_{\mathrm{O}} - \hat{\beta}\right|$ is less or equal to 1×10^{-2}. Figure 20 shows the average value of $\Delta C_{\max}$ for different values of β_{O}.

Figure 20 – Average error value of maximum values of $C(\mathbb{W}, \mathbb{H}, \beta)$ between observed and inferred-scaled parameters. The maximum value of Average error value of maximum values of $C(\mathbb{W}, \mathbb{H}, \beta)$ given in (3.65) was computed using $\{\mathbb{W}_{\mathrm{O}}, \mathbb{H}_{\mathrm{O}}, \beta_{\mathrm{O}}\}$ and $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$. Subsequently, the average of $\Delta C_{\max}$ was estimated using ten samples in which the difference $\left|\beta_{\mathrm{O}} - \hat{\beta}\right|$ is less or equal to 1×10^{-2}. The maximum value of discrepancy in all range of β_{O} is below 3%, meaning that inferred-scaled parameters reconstruct well the partition function characterization.

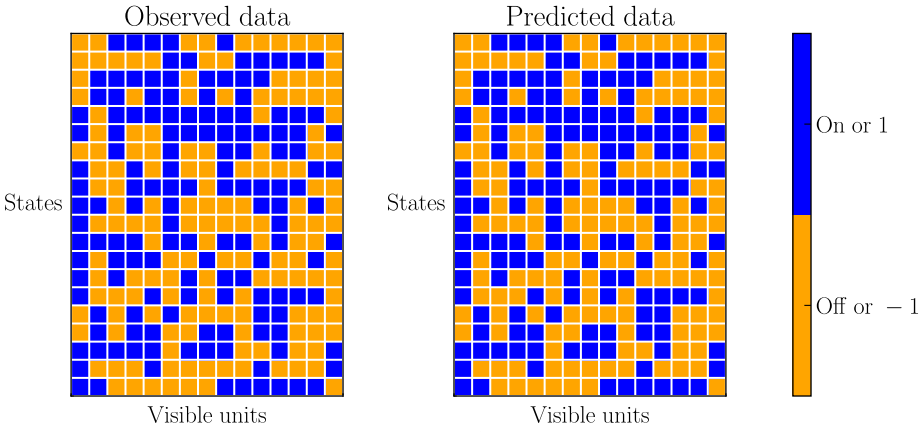


Source: Author.

From Figure 20 can be conclude that the maximum value of discrepancy $\Delta C_{\max}$ is below 3%, meaning that $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ is capable of reconstructing all the four possible characterizations of the partition function given by (3.65). This ensures that parameter $\hat{\beta}$ is mandatory in order to the Boltzmann machine reconstructs in a better way, either the statistical properties or partition function characterization contained in the original data. Finally, the last step to prove that procedures 2 and 3 are working properly is to test if they are capable of reconstructing the original data, that is to say, if the Boltzmann machine can recover the input-output states of the original dataset. To do that, we compute the value of the energy function in (3.65) for all sequences of the original synthetic data generated for each sample in each β_{O}. To this computation it were used the inferred parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ scaled by $\hat{\beta}$. Each sequence is a one-dimensional vector with n elements[7]. After that, a simple algorithm for returning the combination from all 2^n possibilities of $\vec{\sigma}$ who has the energy value was implemented. We iterated the algorithm over all m sequences[8]. Figure 21 shows the comparison between original and predicted data for a random sample of data generated at $\beta_{\mathrm{O}} = 0.10$.

[7] Remember, n is the quantity of visible spin that was set up to $n = 20$.
[8] m was set up to be $m = 4 \times 10^3$

Figure 21 – Comparison of observed and predicted data. Left (right) matrix is associated with observed (predicted) data. In each matrix, columns represent visible spins and rows correspond to a sequence. For better visualization, matrices were plotted with only $n = 15$ visible spins and $m = 20$ random sequences. Visually it is possible to include that procedure 2 and 3 reconstruct with precision all input-output states of the original data.



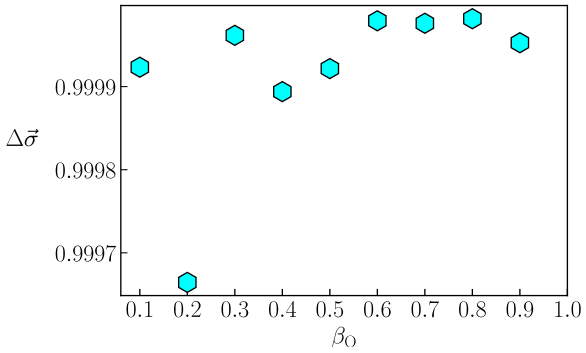
Source: Author.

In each matrix, each column represents a visible spin and each row is associated with a sequence of input-output states. For the sake of better visualization, matrices were plotted with only $n = 15$ visible spins and $m = 20$ random sequences. Besides, two colors were assigned for each possible input-output. Blue color was assigned to on values, i.e., $d_i^j = 1$ and yellow express off values or $d_i^j = -1$. By comparing visually observed and inferred data (left and right matrices), it is possible to attain that procedures 2 and 3 reconstruct with high precision the input-output states of the original data. To better describe this assertion, the input-output success prediction was computed for each sample in all values of β_{O}. Let's define the input-output success prediction $\Delta\vec{\sigma}$ as

$$\Delta\vec{\sigma} = \frac{1}{T}\sum_{T}\left(1 - \frac{1}{m}\sum_{ij}\frac{1}{2n}\left|d_i^j - \hat{d}_i^j\right|\right), \tag{3.67}$$

where d_i^j ($\hat{d}_i^j$) are the observed (inferred) input-output states, respectively, n is the number of visible spins, m is the quantity of sequences and T is the number of samples for each β_{O}. Denominator in third fraction containing $2n$ was set up to normalize the summation. As close the value of (3.67) to 1 better is the prediction success. Figure 22 shows the success prediction for all values of β_{O}.

Figure 22 – Input-output success prediction of Boltzmann machine. The success prediction of data predicted by the Boltzmann machine was computed using (3.67). As exposed in the Figure, prediction success is above of 99% for all values of β_{O}



Source: Author.

Figure shows that for all values of β_{O}, prediction success is above 99%, meaning that procedures 2 and 3 can surely recover the original data and therefore they can be applied to real data.

4 DATA MINING AND MANIPULATION

It is discussed the data mining process, a procedure for generating synthetic data and a new algorithm for detecting clusters based on the adjacency matrix. In detail, the first part is dedicated to describe how data from three different Lower Houses in the world was obtained, processed and transformed to a numerical fashion. Subsequently, a procedure to generate synthetic data constrained to have the same statistical properties of the Lower Houses processed data is exposed. Finally, based on a simple like-percolation process an algorithm for detecting clusters in a network is presented.

4.1 Setting up Legislative data

In order to apply procedures 2 - 3 already discussed in the last chapter for learning desired parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ in Legislative Systems, the first thing to do is set up the binary data points necessary to implement that process. With this in mind, the political party majority opinion will be used to construct the binary data array $d_i^j \in \{\pm 1\}$. Here i represents a political party and j is associated to a bill. The construction of political party majority opinion is based on the roll - call vote[1] database for three different Legislative Lower Houses. Specifically, we will study the United States House of Representatives [16], the House of Commons of the United Kingdom [17] and the Chamber of Deputies of Brazil [18]. The United States House of Representatives is composed of 435 voting Members to legislate for a two year term. The House of Commons of the United Kingdom is made up by an elected body consisting of 650 Members of the Parliament to legislate for a four year term and the Chamber of Deputies of Brazil is organized by 513 elected Deputies to legislate for a four year term.

Lower Houses of these three Legislative systems were choose mainly because they have come to wield more power and exert significant political influence due to its Members are more numerous. The reason for working with the political majority opinion instead of roll - call vote of Lower Houses Members is because we need to compute the partition function for a system of n interacting units and computationally this computation, for a small quantity of visible units, can be made exactly. In fact, computations will be

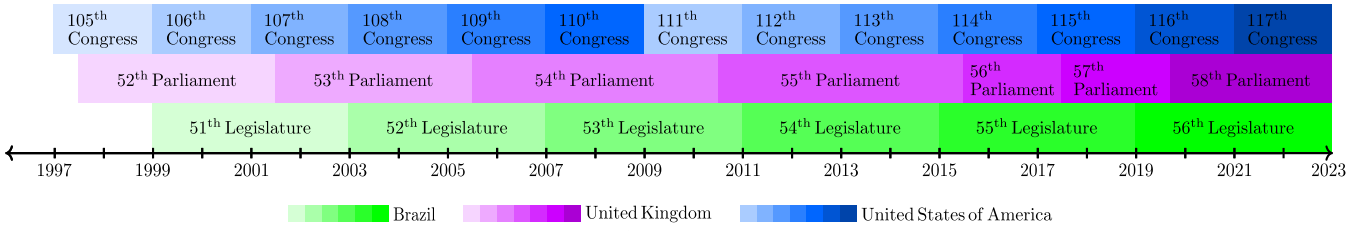
[1] Roll - call vote is a voting system in which it is possible to identify voters and their votes. Generally, roll - call vote process is made either by calling individually each voter when they vote on commissions or by electronic vote when they vote in plenary sessions.

restricted to a maximum quantity of $n = 23$ political parties. Construction of political party majority matrix for each Lower House is explained as follows.

4.2 Legislative Lower House Databases

Results of roll - call vote for each aforementioned Lower Houses were collected using the public and free available data found at their websites[2]. That process was carried out by implementing a simple *API* with the aim to search, read and store in a *CSV* file the result of roll - call vote for each bill discussed. For the United States House of Representatives data was collected from 1997 to 2022. This period encloses from 105$^{\text{th}}$ to 117$^{\text{th}}$ Congress meeting. For the House of Commons of the United Kingdom data ranges from 1997 to 2022 compassing from 52$^{\text{th}}$ to 58$^{\text{th}}$ Parliament. For the Chamber of Deputies of Brazil data extends from 1999 to 2022 spanning from 51$^{\text{th}}$ to 56$^{\text{th}}$ Legislature. Figure 23 summarizes legislative periods to be studied.

Figure 23 – Legislative periods to be studied. Data from the Legislative Lower House of the United States of America, United Kingdom and Brazil were collected using the public and free available data found at their websites. The period of time studied ranges from the 105$^{\text{th}}$ to 117$^{\text{th}}$ Congress, 52$^{\text{th}}$ to 58$^{\text{th}}$ Parliament and 51$^{\text{th}}$ to 56$^{\text{th}}$ Legislature for the United States of America, United Kingdom and Brazil, respectively. Bluish color, purplish colors and greenish colors are used to illustrate Legislative periods of the United States of America, United Kingdom and Brazil.

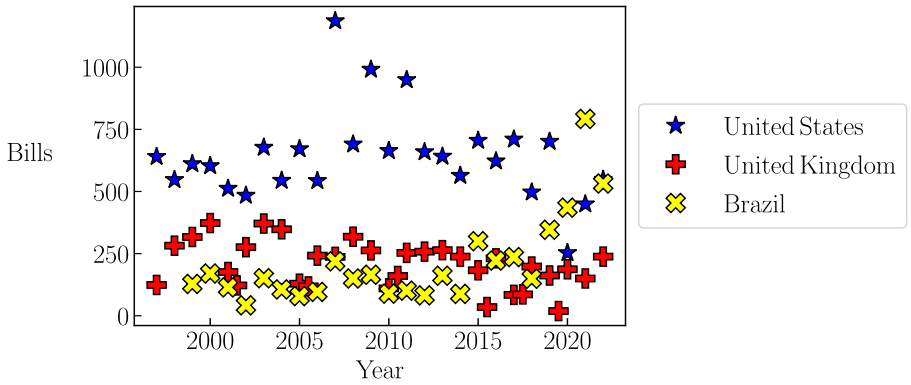


Source: Author.

In Figure 23, bluish colors, purplish colors and greenish colors are used to illustrate Legislative periods of the United States of America, United Kingdom and Brazil, respectively. At the end of the collecting process, files containing roll - call vote results were grouped by year so that the quantity of files corresponds to the quantity of bills voted in each Lower House. Figure 24 shows the quantity of bills discussed in each Lower House.

[2] Websites are https://clerk.house.gov/Votes/, https://www.publicwhip.org.uk/ and http://www2.camara.leg.br/ for the United States of America, United Kingdom and Brazil, respectively.

Figure 24 – Quantity of bills voted in Lower Houses databases. A simple *API* was implemented to search, read and store in *CSV* files the result of roll - call vote for each Lower House under study. Files were grouped by year so that quantity of files corresponds to the quantity of bills discussed in each Lower House.



Source: Author.

Subsequently, all possible vote options for each Legislative Lower House database were identified. Vote options allow voting Members to express their stance against every bill discussed[3]. Table 2 shows all vote options identified for each Lower House database.

Table 2 – Identified vote options in Lower House databases. For each Lower House it was identified all possible options that a voting Member has in order to express its stance against a bill discussed.

Vote Options		
United States	**United Kingdom**	**Brazil**
Aye	Aye	Aye
Present	Tellaye	Abstention
	Both	Art. 17
Absence	Absence	Absence
	Tellnay	Obstruction
Nay	Nay	Nay

Source: Author.

Aye and *Nay* vote options express either supporting or opposition to the bill that is being voted. *Abstention* is the possibility to refuse to be part of a voting session in order to avoid expressing a stance but counts as part of the quorum. *Obstruction* is a political device used to create interference on legislative procedures. Generally, obstruction means abandonment of plenary by voting Members to avoid quorum count. *Absence* is the physical non-appearance at the time to vote by a voting Member. *Tellaye* and *Tellnay* are appointed to verify the count when there is a division in the Commons in order to report the result back to the House. *Both*, *Present* and *Art. 17* votes does not count towards or against the passage of a bill, but it contributes towards the quorum. With all of this,

[3] It worth of comment that voting Members may use one and only vote option by bill voted.

a small section of a particular file for each Lower House database and a random year is exposed in Figure 25.

Figure 25 – Pictorial representation of Lower House datasets. For each Lower House and year, a database containing name, political party, state and vote for each Voting Member was created.

Name	Party	State	Vote	Name	Party	State	Vote
Abercrombie	Democratic	Hawaii	Nay	Angela Portela	Pt	Roraima	Aye
Ackerman	Democratic	New York	Nay	Edio Lopes	Pmdb	Roraima	Aye
Aderholt	Republican	Alabama	Nay	Francisco Rodrigues	Dem	Roraima	Nay
Allen	Democratic	Maine	Nay	Arlindo Chinaglia	Pt	São Paulo	Art. 17
Andrews	Democratic	New Jersey	Nay	Maria Helena	Psb	Roraima	Aye
Archer	Republican	Texas	Nay	Sebastião Bala Rocha	Pdt	Amapá	Abstention
Armey	Republican	Texas	Aye	Beto Faro	Pt	Pará	Aye
Bachus	Republican	Alabama	Nay	Elcione Barbalho	Pmdb	Pará	Aye
Baesler	Democratic	Kentucky	Nay	Armando Monteiro	Ptb	Pernambuco	Obstruction
Baker	Republican	Louisiana	Nay	Eduardo Valverde	Pt	Rondonia	Absence

United States - 1997

% Name	Party	State	Vote
% Stephen Farry	Alliance (front bench)	North Down	Aye
% Nigel Adams	Con (front bench)	Selby and Ainsty	Nay
% Bim Afolami	Con	Hitchin and Harpenden	Nay
% Adam Afriyie	Con	Windsor	Nay
% Nickie Aiken	Con	Cities of London and Westminster	Nay
% Peter Aldous	Con	Waveney	Nay
% Lucy Allan	Con	Telford	Nay
% Sir David Amess	Con (front bench)	Southend West	Nay
% Lee Anderson	Con	Ashfield	Nay
% Stuart Anderson	Con (front bench)	Wolverhampton South West	Nay

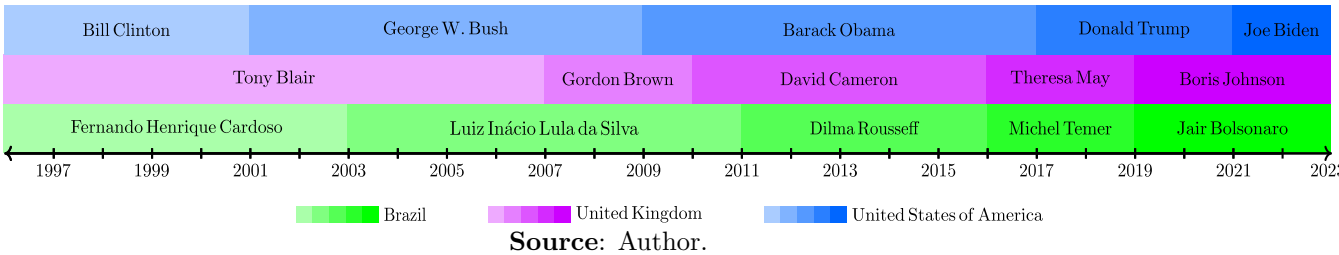
Brazil - 2007

United Kingdom - 1997

Source: Author.

As observed, for each Lower House and year, a database containing name, political party, state and vote for each Lower House Member was created. Databases will help us to construct the political party majority opinion array for each year. Another benefit information obtained from the Legislative databases is the name of Chief Executive power in each Legislative period as illustrated in Figure 26.

Figure 26 – Names of Chief Executive power for each Legislative database. For each term, the name of Executive Chief is shown. Bluish, purplish and greenish colors were set up to represent the United States of America, United Kingdom and Brazil Executive terms.



Source: Author.

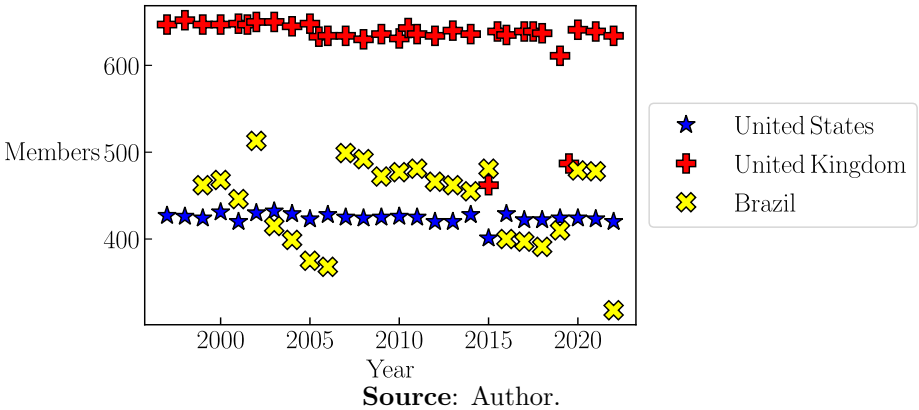
In Figure 26, bluish, purplish and greenish colours indicated each Executive term for the United States of America, United Kingdom and Brazil, respectively. This

information will be useful at the time to analyse results derived from Boltzmann machine computations.

4.3 Lower House Databases Manipulation

In order to create the political party majority opinion matrix, a manipulation over each Lower House database was carried out. In detail, a script was designed for taking all files containing the results of roll - call vote by year and transform them into an individual file. Specifically, the script appends the result of roll - call vote for each Lower House Member throughout year. If a Lower House Member does not have a vote in a particular bill a missing vote is appended. After that, voting Members that either not vote all year around or possess missing votes were removed and remaining ones were grouped by political party. Figure 27 shows the quantity of voting Members in each Lower House database by year after this process.

Figure 27 – Quantity of voting Members in Lower House databases. Roll - call vote was appended by voting Member throughout year and stored in an individual file for each year and Lower House database. Voting Members that either do not vote all year around or possess missing votes were removed from the final file.



Source: Author.

For all Lower Houses databases, the quantity of voting Members is ever less than the real elected parliamentary body. Afterwards, a random variable to map vote options to integer real numbers was created for each Lower House defined like

$$V_{\text{US}} = \begin{cases} 2 & \text{if Aye,} \\ 1 & \text{if Present,} \\ -1 & \text{if Absence,} \\ -2 & \text{if Nay.} \end{cases} \quad V_{\text{UK}} = \begin{cases} 2 & \text{if Aye,} \\ 1 & \text{if Tellaye,} \\ 0 & \text{if Both,} \\ -1 & \text{if Absence,} \\ -2 & \text{if Tellno,} \\ -3 & \text{if Nay.} \end{cases} \quad V_{\text{BR}} = \begin{cases} 2 & \text{if Aye,} \\ 1 & \text{if Abstention,} \\ 0 & \text{if Art. 17,} \\ -1 & \text{if Absence,} \\ -2 & \text{if Obstruction,} \\ -3 & \text{if Nay,} \end{cases} \tag{4.1}$$

where V_{US}, V_{UK} and V_{BR} stands as the random variable for the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively. Values of each random variable were designated in order to differentiate vote options used to express approval stance (positive values), neutral stance (zero value) and disapproval stance (negative values). We also wanted to highlight Aye/Nay at the highest values. At the end of this process a roll - call vote matrix $\mathbf{V}_{nk} = (V_{1k}, V_{2k}, \cdots, V_{nk})'$ is obtained for each year and Lower House. $\mathbf{V}_{nk}$ is a sequence of k values of random variable (4.1) spawned in n different realizations. Here, n counts for the number of voting Members (see Figure 27) and k represents the quantity of bills voted by year (see Figure 24). Each element of $\mathbf{V}_{nk}$ represents the numerical value of the vote of a voting Member n at bill k. A pictorial representation of the roll - call vote matrix is

$$\mathbf{V}_{nk} = \begin{array}{c} \\ P_1 \\ \\ \\ P_2 \\ \\ \\ P_m \\ \\ \end{array} \overbrace{\begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \\ M_4 \\ M_5 \\ M_6 \\ \vdots \\ M_n \end{array}}^{\text{Bills by year}} \begin{pmatrix} -3 & 2 & 1 & -1 & \cdots & 2 \\ 0 & -1 & 0 & 2 & \cdots & 1 \\ -2 & 2 & -3 & -3 & \cdots & -3 \\ 2 & 3 & -1 & 0 & \cdots & 0 \\ -2 & 1 & 2 & -1 & \cdots & -3 \\ 2 & 0 & 1 & -3 & \cdots & -2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -2 & -1 & 0 & -3 & \cdots & 1 \end{pmatrix}, \tag{4.2}$$

where M_i represents i - voting Member and P_j depicts that Members were grouped by m-political party. Observed that a color box was drawn to better notice the political party ordering as well as to highlight that there are m political parties. Roll - call vote matrix (4.2) will be used to create the political party majority opinion matrix for each year and Lower House.

4.4 Creating political party majority opinion vote matrix

Finally, using the fact that voting Members are grouped by political party in (4.2) is straightforward to obtain the political party majority opinion vote matrix. It is worth of comment that due to the restriction of binary arrays $d_i^j \in \{\pm 1\}$, this matrix must have only two possible values, meaning that the political party is being expressing either disapproving (-1) or approving (1) stance against a bill. Specifically, for each Lower House in each year the following procedure is proposed.

Procedure 4 Political parties' majority opinion algorithm.

Require: $\mathbf{V}_{nk}$ for each year and Lower House (Equation (4.2)).

$m \longrightarrow$ Quantity of political parties

$k \longrightarrow$ Number of bill discussed by year

$\tilde{\mathbf{V}}_{mk} \longrightarrow$ Political parties' majority matrix

for $i = 1$ **to** m **do**

Compute sum over Members from i-party for all vote options in each bill.

for $j = 1$ **to** k **do**

if Nay$[j] \neq$ Aye$[j]$ **then**

if Nay$[j] >$ Aye$[j]$ **then**

$\tilde{\mathbf{V}}_{ij} = 1$

else

$\tilde{\mathbf{V}}_{ij} = -1$

end if

else

if Sum of positive vote options$[j] >$ Sum of negative vote options$[j]$ **then**

$\tilde{\mathbf{V}}_{ij} = 1$

else

if Sum of positive options$[j] <$ Sum negative options$[j]$ **then**

$\tilde{\mathbf{V}}_{ij} = -1$

else

$\tilde{\mathbf{V}}_{ij} = \text{rand}(-1, 1)$ ▷ rand means binary random choice.

end if

end if

end if

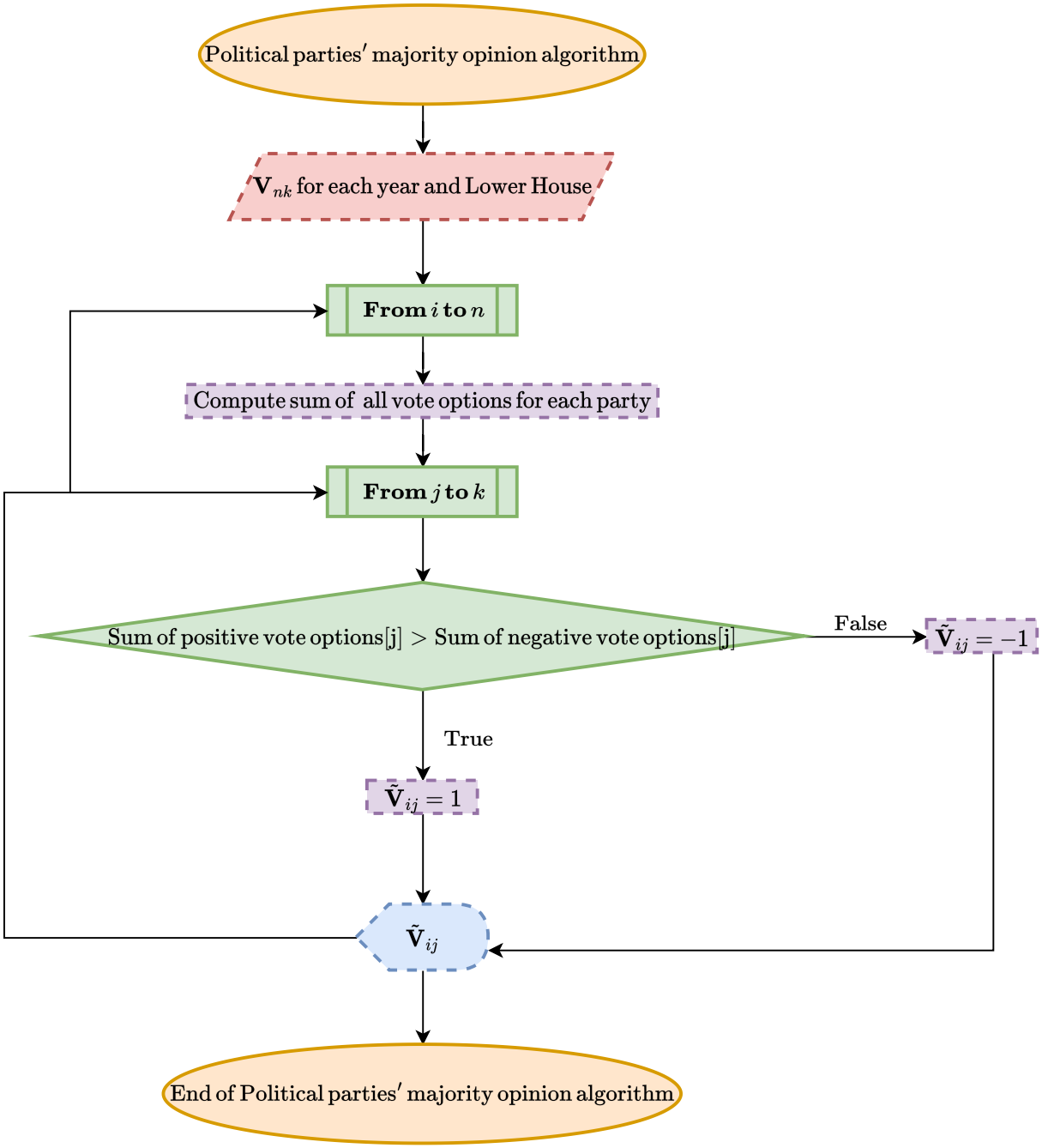
end for

end for

return $\tilde{\mathbf{V}}_{ij}$

Note that procedure 4 points out to compute majority opinion first by comparing *Aye* and *Nay* majorities, in case of draw, it compares now between approving and disapproving majorities and if no condition is met until there, a random binary choice is set up. Figure 28 resumes the aformentioned procedure.

Figure 28 – Flowchart for procedure 4



Source: Author.

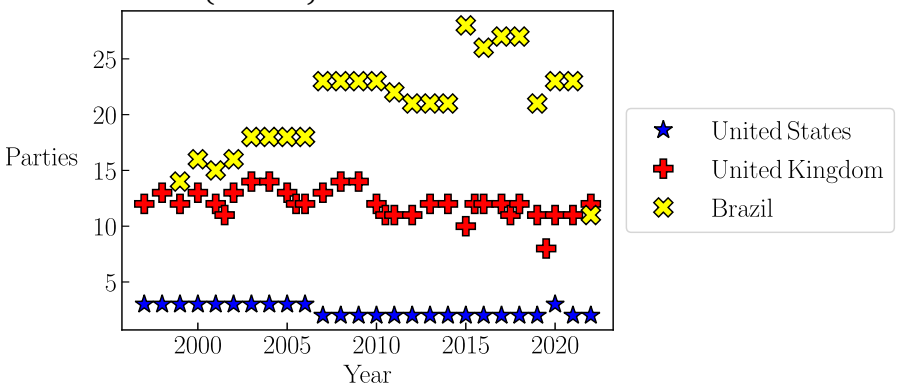
This procedure returns a binary array containing the majority opinion for each party in each bill by year. A pictorial representation of the political party majority opinion

matrix is given by

$$\tilde{\mathbf{V}}_{mk} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ \vdots \\ P_m \end{array} \overbrace{\begin{pmatrix} -1 & 1 & 1 & -1 & \cdots & 1 \\ 1 & -1 & -1 & 1 & \cdots & 1 \\ -1 & 1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & 1 & -1 & \cdots & 1 \end{pmatrix}}^{\text{Bills by year}}, \tag{4.3}$$

where again P_m represents the m - political party. Matrix (4.3) will be employed to learn parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ for each Lower House through all years. Figure 29 shows the quantity of political parties for each Lower House by year after the computation of the majority opinion binary array using the procedure 4 aforementioned.

Figure 29 – Quantity of political parties found in Lower Houses databases. For each year and Lower House database, the political party majority opinion array was computed using procedure 4. At the end, a like binary array (4.3) is obtained containing the majority opinion for political party. This array will be employed to learn parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ employing procedures exposed in Chapter 1.



Source: Author.

4.5 Generating political party majority random data

In order to prove that procedures proposed in Chapter 1 work well when applied to the majority opinion array (4.3) and therefore, that our conclusions cannot be obtained by chance, we decided to generate simulated data encompassing all original statistical properties for each year in each Lower House. Specifically, it was implemented a procedure that generates a binary array constrained to have the same correlation structure and

cumulative distribution function when compared to each political party, year and Lower House. Technically, we implement a procedure that uses a set of functions for generating correlated data already offered in *R* language and our procedure seeks for the binary array which minimizes the difference of the correlation matrix and marginal distribution between the original binary political party array (4.3) and the simulated one. Functions used were *corrcheck*, and *ordsample* from the *GenOrd* package [19, 20]. Proposed procedure is established as follows.

Procedure 5 Political party majority random data

Require: $\mathbf{F}\left[\mathbf{V}_{mk}\right]$ (Cumulative distribution function for each party)

Require: $\boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]$ (Correlation matrix).

Use corrcheck $\left(\boldsymbol{\rho}\left[\mathbf{V}_{mk}\right], \mathbf{F}\left[\mathbf{V}_{nk}\right]\right)$ to verify if all correlation coefficients are feasible to simulate.

In case, change not feasible coefficients using lower and upper bounds limits returned by corrcheck $\left(\boldsymbol{\rho}\left[\mathbf{V}_{nk}\right], \mathbf{F}\left[\mathbf{V}_{nk}\right]\right)$.

$d \longrightarrow$ Number of binary values for each political party;

$\tau \longrightarrow$ Minimum tolerance value of the difference between simulated and real $\mathbf{F}\left[\mathbf{V}_{nk}\right]$, $\boldsymbol{\rho}\left[\mathbf{V}_{nk}\right]$, respectively;

$\mathbf{V}'_{nd} \longrightarrow$ Simulated political party majority opinion array.

while $\gamma_1 \geq \tau$ and $\gamma_2 \geq \tau$ **do**

$\mathbf{V}'_{nd} \longrightarrow$ ordsample $\left(\boldsymbol{\rho}\left[\mathbf{V}_{nk}\right], \mathbf{F}\left[\mathbf{V}_{nk}\right]\right)$

Compute cumulative distribution function for each party to obtain $\mathbf{F}\left[\mathbf{V}'_{md}\right]$

$\gamma_1 = \text{mean}\left(\left|\mathbf{F}\left[\mathbf{V}'_{md}\right] - \mathbf{F}\left[\mathbf{V}_{mk}\right]\right|\right)$ ▷ mean is the average value.

$\gamma_2 = \text{mean}\left(\left|\boldsymbol{\rho}\left[\mathbf{V}'_{md}\right] - \boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]\right)\right.$ ▷ Only in upper-triangle elements.

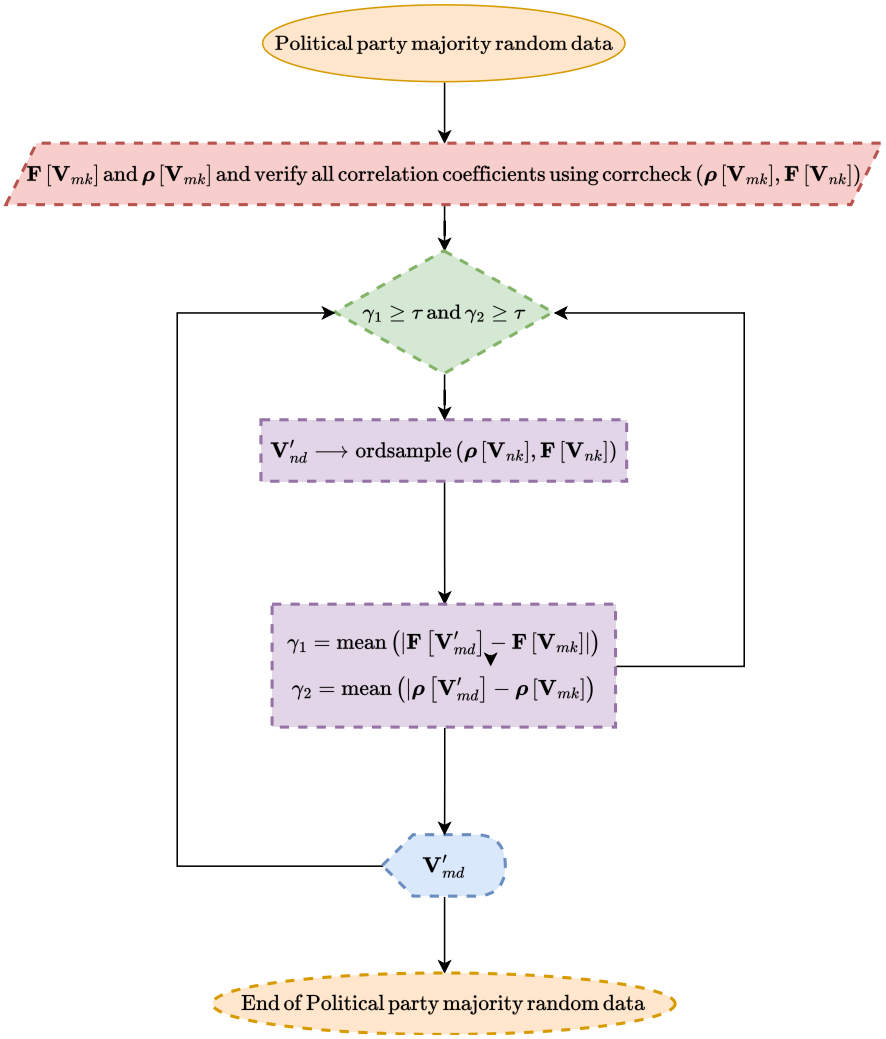
end while

return $\mathbf{V}'_{md}$

Procedure works as follows. It is required to compute the cumulative distribution function for each party and the correlation matrix for (4.3). Prior to generate the data, the procedure checks if all correlation coefficients are feasible to construct. This process is carried out by computing the upper and lower bounds for each discrete variable using its marginal distribution. Subsequently, correlated ordinal data is generated by the *ordsample* function. Binary data is generated first by computing the correlation matrix of a standard normal variable yielding the desired one and after that, a discretization is made over the resulting variable set. Afterwards, the cumulative distribution function of

the simulated binary array is computed by each party along its correlation matrix. Finally, the difference between real and simulated cumulative and upper-triangle correlation values is obtained. Note that simulated binary array is generated iteratively until the minimum tolerance of difference for cumulative and correlation values is met. It is worth of comment that proposed procedure does not work if correlation matrix has negative eigenvalues. In our simulation for all years and Lower Houses, it was established a minimum tolerance of $\tau = 10^{-3}$ and the quantity of data generated for each party was set up to be $d = 10^5$. Figure 30 resumes the aformentioned procedure.

Figure 30 – Flowchart for procedure 5



Source: Author.

4.6 Nearest correlated cluster algorithm (NECO)

In order to better analyse the political parties' intrinsic inner relationship with the results generated by the Boltzmann machine procedures already exposed in Chapter 1, a last procedure is presented. This procedure aims to detect clusters based on the correlation of the sequences of the Lower Houses Members votes. This procedure is based in a "like-percolation" process computed over the correlation distance matrix of the roll-call vote matrix (4.2). The aim is then to find the minimum value of the correlation distance in which it is possible to construct a complete connected network. In other words, we seek for the value of correlation distance in which unconnected - connected phase transition happens. Actually, this procedure is an extension of the Minimal Spanning Tree (MST) based procedure for reordering correlation matrices [1]. In simple words, the method

computes the MST of the correlation distance matrix obtained from the roll-call vote matrix (4.2). The MST is arisen as a subnetwork that connects all Lower Houses members constrained to the sum of their correlation distance values is minimum. Reordering of the correlation matrix is made then, by taking into account the position in which each pair of Lower Houses members appears as the spanning tree grows [1]. In this sense, NECO procedure aims for better visualization of reordered matrices produced by this method. Procedure is established as follows.

Procedure 6 Nearest correlated cluster algorithm (NECO)

Require: $\mathbf{V}_{nk} \longrightarrow$ Roll - call vote matrix (4.2).

Compute correlation distance matrix $\boldsymbol{d}[\mathbf{V}_{nk}] = \sqrt{2\,(1 - \boldsymbol{\rho}[\mathbf{V}_{nk}])}$

Sort $\boldsymbol{\rho}[\mathbf{V}_{nk}]$ using MST-based procedure described in [1].

Consider $\boldsymbol{d}[\mathbf{V}_{nk}]$ as a complete undirected connected network.

In the weighted adjacency list $[v_i, v_j, d(i,j)]$, sort weights in descending order.

head $\longrightarrow [1, 2, 3, \ldots, n]$; ▷ Head vector for computing giants components.

mass $\longrightarrow [1, 1, 1, \ldots, 1]$; ▷ Mass vector array of size equal to n.

$\tau \longrightarrow []$; ▷ Threshold array.

$n_1 \longrightarrow []$; ▷ First giant component array.

$n_2 \longrightarrow []$; ▷ Second giant component array.

$m_1 \longrightarrow 0$;

$m_2 \longrightarrow 0$;

$i, j \longrightarrow 0$;

while $m_1 \neq 1$ and $m_2 \neq 0$ **do**

  while $v_i \neq \text{head}[v_i]$ **do**

    $a_i = \text{head}[v_i]$

    $v_i = \text{head}[a_i]$

  end while

  while $v_j \neq \text{head}[v_j]$ **do**

    $a_j = \text{head}[v_j]$

    $v_j = \text{head}[a_j]$

  end while

  if $v_i \neq v_j$ **then**

    $\text{head}[v_j] = v_i$

    $\text{mass}[v_i] += \text{mass}[v_j]$

    $\text{mass}[v_j] = 0$

  end if

  Sort in descending order mass array

  $m_1 = \text{mass}[0]/n$ ▷ mass[0] means first element of mass array.

  $m_2 = \text{mass}[1]/n$ ▷ mass[1] means second element of mass array.

  Append m_1 to n_1

  Append m_2 to n_2

  Append $d(i,j)$ to τ

end while

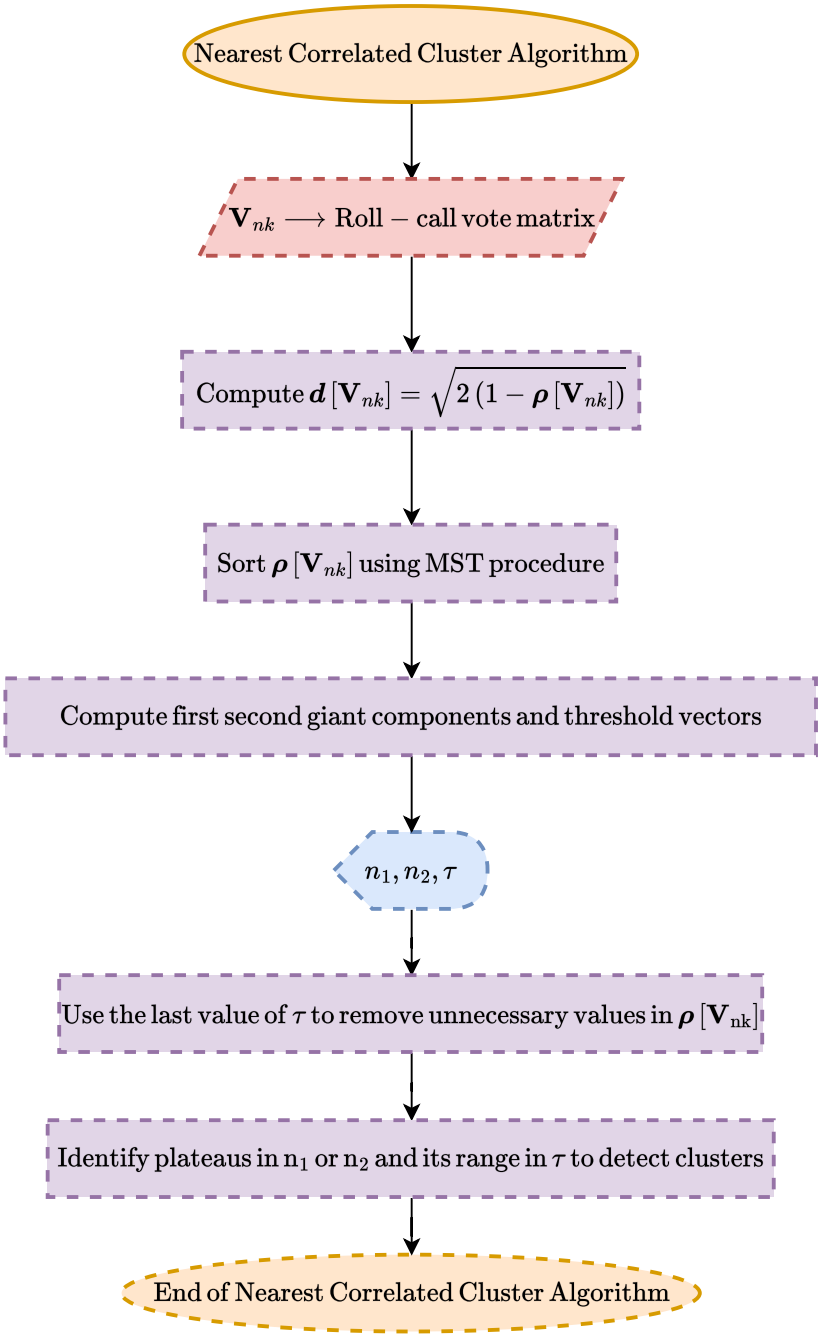
return n_1, n_2, τ

Use the last value of τ to remove unnecessary values in $\boldsymbol{\rho}[\mathbf{V}_{nk}]$

Identify plateaus of n_1 or n_2 and its range in τ to detect clusters in $\boldsymbol{\rho}[\mathbf{V}_{nk}]$

Procedure 6 is based on the computation of the first and second giant components of a complete connected network. Complete procedure works as follows: First, a residual network is created by adding one edge step by step. Rule for adding edges begins from the smallest to the largest weights (in this case, values of correlation distance). Subsequently, in each step it is computed the first and second giant component of the residual network. Giant component is a connected sub-network containing a proportion of the entire nodes of the original vertex set [21]. Finally, the procedure runs until phase transition is found. Phase transition is the state in which, first giant component must be equal to one and the second giant component is zero. In the end, procedure returns three arrays: the first and second giants components arrays and another one containing the weights in which plateaus for the previous giants components are observed. These arrays help us to identify clusters on the adjacency matrix (in this case, correlation matrix). Figure 31 resumes the aformentioned procedure.

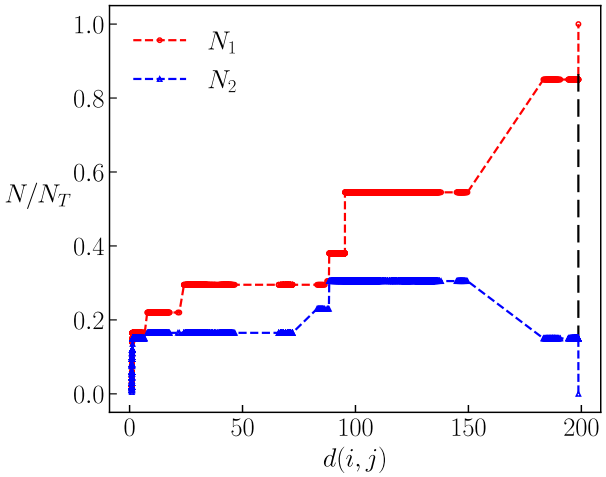
Figure 31 – Flowchart for procedure 6



Source: Author.

To exemplify how cluster identification arise, we apply procedure 6 over a toy network. Figure 32 shows the first (red - dashed) and second (blue - dashed) normalized giant component plateaus for the toy network.

Figure 32 – First (red-dashed) and second (blue-dashed) normalized giant components of a toy network. N is the number of connected nodes in each weight $d(i,j)$ and N_T stands as the size of the network. Vertical black dashed line indicates the value in which unconnected-connected phase transition occurs.

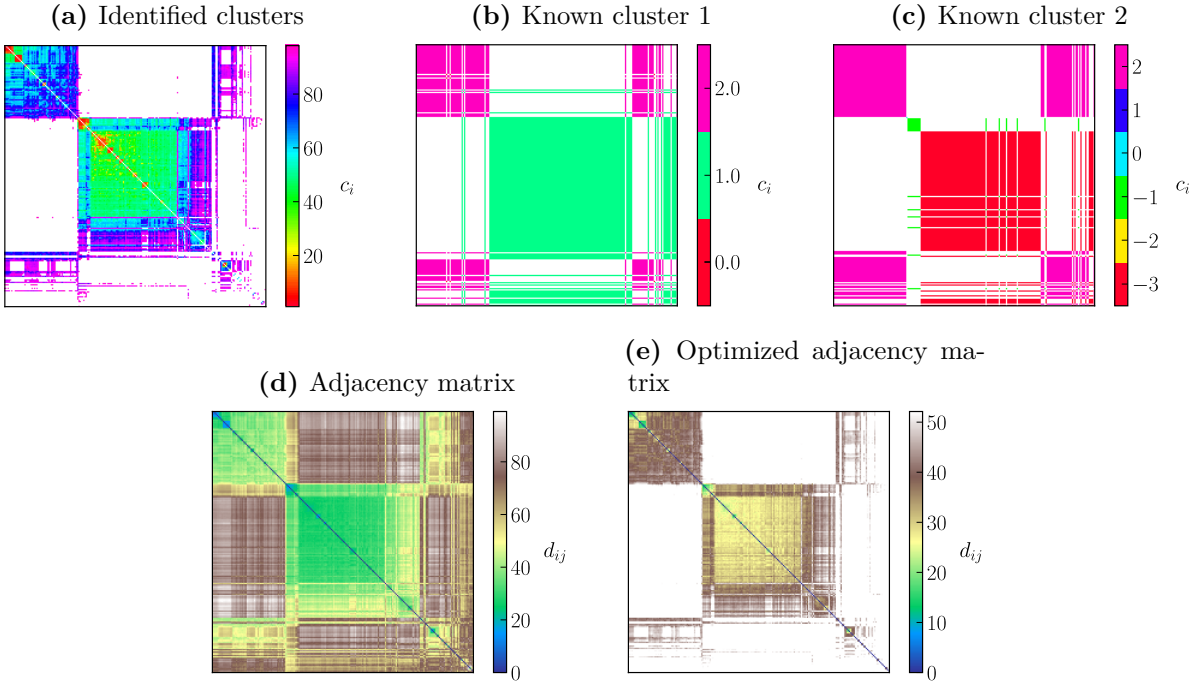


Source: Author.

Note that value in which the unconnected-connected phase transition happens is depicted by a vertical black dashed line. Identification of clusters is as follows: First use the last value of τ to remove unnecessary values in the correlation matrix (value highlighted by the vertical black dashed line). Subsequently, by counting plateaus from the first component array n_1 it is possible to identify the number of clusters in the correlation matrix[4]. To display each cluster counted, use threshold array τ to recognize correlation ranges for clusters. Figure 33 shows the final result of cluster identification over our toy network. To better visualize, displayed matrices were re-ordered using the MST-based procedure detailed explained in [1].

[4] Same result is attained by doing the counting process over the second giant component.

Figure 33 – Nearest correlated cluster algorithm result. Procedure 6 was applied over a toy network in order to identify its clusters. Figures 33a - 33c show the identified cluster and the comparison with two known clusters. Figures 33d - 33e show the original and optimized adjacency matrix. Optimized adjacency matrix is compute by eliminating all values greater or equal to the value in which unconnected-connected transition occurs. Note that clusters obtained using proposed procedure resemble real clusters.



Source: Author.

Identified cluster are located in Figure 33a which is compared with two known cluster exposed in Figures 33b - 33c. Note how identified clusters resembles real known clusters. In fact, procedure 6 reveals inner sub clusters. Note that quantity of clusters (Right vertical color bar) in the upper - left matrix is greater in comparison to the number of known clusters for upper - center and upper - left matrices. Finally, Figures 33d - 33e show the original and optimized adjacency matrices. Optimized adjacency matrix is computed by eliminating all values greater or equal to the value in which unconnected-connected transition occurs. In general, proposed procedure works in any adjacency network without regarding the metric used to measure weight between two vertices. However, at last, that measure function must satisfy the axioms of a distance metric.

5 RESULTS AND ANALYSIS

It is presented the analysis of roll-call vote data using the interactionism and statistical approach. We begin describing results obatine using interactionism procedures and subsequently, results of statistical approach are discussed. Finally, Legislative Consensual Regimes are proposed by joining results of both approaches.

5.1 Input quantities for extended Boltzmann machine learning process

Previously, in Chapter 2 was described how roll-call vote data was transformed from raw data to political majority opinion matrix for each Lower House defined as

$$\tilde{\mathbf{V}}_{mk} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ \vdots \\ P_m \end{array} \overbrace{\begin{pmatrix} -1 & 1 & 1 & -1 & \cdots & 1 \\ 1 & -1 & -1 & 1 & \cdots & 1 \\ -1 & 1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & 1 & -1 & \cdots & 1 \end{pmatrix}}^{\text{Bills by year}} . \tag{5.1}$$

This matrix is the main key to compute the input quantities needed to do the learning process detailed exposed in Chapter 1. Remember that $m-$political parties are located in rows and $k-$bills voted by year are spanned in columns as shown in (5.1). Input quantities are listed as follows:

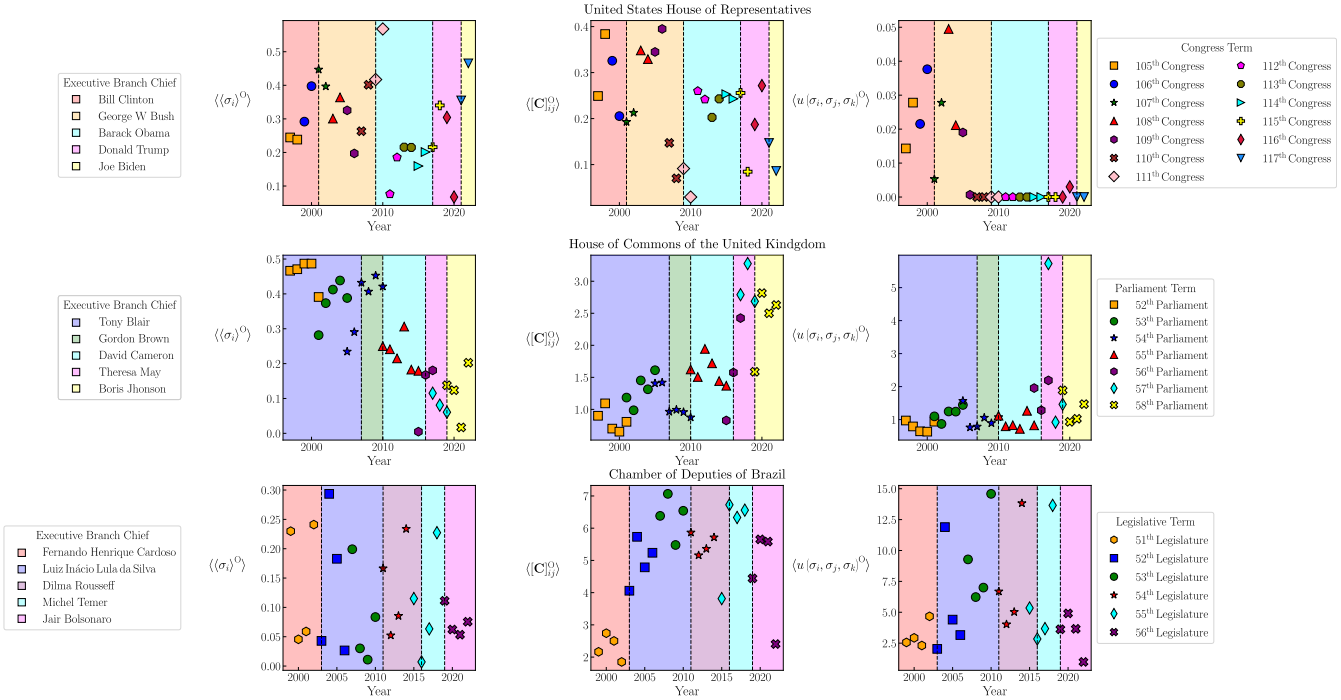
Political party opinion mean value: This value is computed by averaging the sequence of values spanned over all bills by year and do this process for each political party (green, cyan and purple rectangles in (5.1)). At the end, a $m-$vector $\langle \sigma_i \rangle^{\mathrm{O}}$ is obtained representing the opinion mean value for each party. $\langle \sigma_i \rangle^{\mathrm{O}}$ is a value between zero and one;

Political party covariance matrix: Covariance matrix is computed by evaluating the expected value (3.58) among all pairwise sequences of values spanned over columns, i.e. all possible combinations of green, cyan and oranges rectangles. Final result must be a $m-$squared $[\mathbf{C}]^{\mathrm{O}}_{ij}$ matrix;

Third order political party degree of interaction: Third order Ursell function (3.59) is computed using the two previous quantities. This value measures how the interaction among triplets of parties is. At the end, we obtain a $s-$vector $u\left(\sigma_i,\sigma_j,\sigma_k\right)^{\mathrm{O}}$ where $s=m(m-1)(m-2)/3!$ counts all possible combinations of triplets among parties.

Note that superscript O means observed data, that is, quantities computed using Lower Houses data. Listed quantities were obtained for each Lower House and year. Figure 34 shows the expected value of these three quantities for each Lower House in a period of time spanning from 1997 to 2022.

Figure 34 – Expected value of input quantities for learning process. Expectation value of $\langle\sigma_i\rangle^{\mathrm{O}}$ (left-panels), $[\mathbf{C}]_{ij}^{\mathrm{O}}$ (center-panels), $u\left(\sigma_i,\sigma_j,\sigma_k\right)^{\mathrm{O}}$ (right-panels) was computed for each Lower House and year. In all panels, black dashed lines and coloured zones were used to separate executive terms and a marker was associated to a legislative term. Executive and Legislative term are displayed in the legends boxs located at right and left of each group of panels. Lower values of $\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle$ indicates dissensus among political parties. Values of $\left\langle[\mathbf{C}]_{ij}^{\mathrm{O}}\right\rangle$ and $\left\langle u\left(\sigma_i,\sigma_j,\sigma_k\right)^{\mathrm{O}}\right\rangle$ from Chamber of Deputies of Brazil are greater than values for the United States House of Representatives and the House of Commons of the United Kingdom mainly because the number of political parties is greater and therefore, the number of possible combinations is greater in consequence. 10-base log scale was set up in $\left\langle[\mathbf{C}]_{ij}^{\mathrm{O}}\right\rangle$ and $\left\langle u\left(\sigma_i,\sigma_j,\sigma_k\right)^{\mathrm{O}}\right\rangle$ to better observe values for all three Lower Houses.



Source: Author.

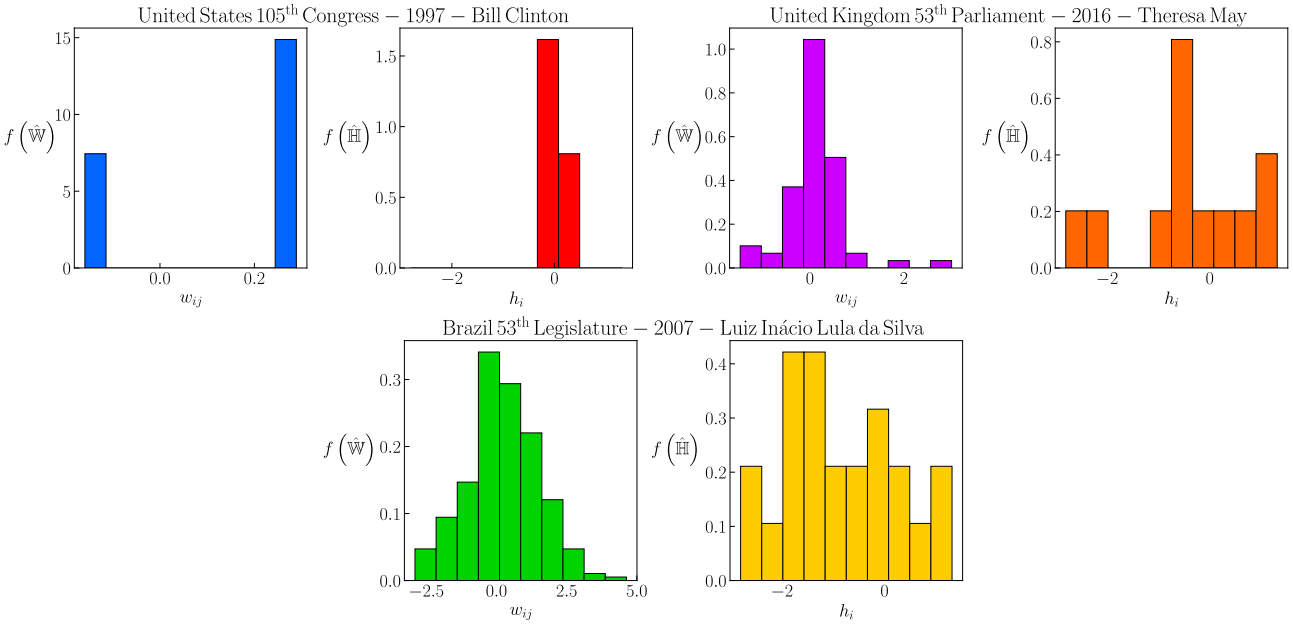
In Figure 34 lefts panel show expected value of $\langle\sigma_i\rangle^{\mathrm{O}}$, center panels are for expected values of $[\mathbf{C}]_{ij}^{\mathrm{O}}$ and right panel are intended for expected value of $u\left(\sigma_i,\sigma_j,\sigma_k\right)^{\mathrm{O}}$. Black dashed lines and coloured zones are intended to highlight each executive term and

a marker was assigned for each Legislative term. Note that if $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ is close to zero, this is an indication of a high degree of dissensus among political parties. Conversely, greater values is associated with a more consensual sense among them. Dissensual or polarized state is defined as the situation in which groups of political parties strongly disagree. That is to say, is the configuration in which a subset of political parties agree to approve the bill whereas the remaining subset enforces for rejecting it, for example. Conversely, a consensual state is the case in which all political parties possess the same opinion in relation to approve or disprove the bill. On the other hand, values of $\left\langle [\mathbf{C}]_{ij}^{\mathrm{O}} \right\rangle$ and $\left\langle u\left(\sigma_i, \sigma_j, \sigma_k\right)^{\mathrm{O}} \right\rangle$ from Brazil are greater than values for the United States and United Kingdom mainly because the number of political parties in this Lower House is greater and therefore, possible combinations of three political parties are greater in consequence. In fact, for some years of the United States Lower House, $\left\langle u\left(\sigma_i, \sigma_j, \sigma_k\right)^{\mathrm{O}} \right\rangle$ is zero because there are only two political parties. Note that 10-base log scale was set up in $\left\langle [\mathbf{C}]_{ij}^{\mathrm{O}} \right\rangle$ and $\left\langle u\left(\sigma_i, \sigma_j, \sigma_k\right)^{\mathrm{O}} \right\rangle$ to better observe values for all three Lower Houses.

5.2 Learned parameters $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$: From statistics to interactions

By using input quantities $\left\{\langle \sigma_i \rangle^{\mathrm{O}}, [\mathbf{C}]_{ij}^{\mathrm{O}}\right\}$ as input information in procedure 2 it is possible to learn parameter set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ for each year and Lower House. As a result, $\hat{\mathbb{W}}$ must be a $m-$squared symmetrical matrix and $\hat{\mathbb{H}}$ must be a $m-$vector. The importance of learning set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ resides in the fact that analyses of Lower Houses change from a statistical (opinion mean value and covariance matrix) to an interaction (weights and thresholds) point of view. This change promotes a deeper understanding of the intrinsic inner political relationships because it is possible to carry out a profound characterization of Lower Houses collective behaviour. Figure 35 shows the probability distribution function for the set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ in a random year for each Lower House.

Figure 35 – Probability distribution function of learned parameter set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ for one random year in each Lower House. Learned parameters were obtained using procedure 2 using input quantities $\left\{\langle\sigma_i\rangle^{\mathrm{O}}, [\mathbf{C}]^{\mathrm{O}}_{ij}\right\}$. For each pair $\left(f\left(\hat{\mathbb{W}}\right), f\left(\hat{\mathbb{H}}\right)\right)$ a label indicating the legislative term number, year and name of Chief Executive power was set up. Distributions for the United States Lower House are sparser in comparison with the others. For the United Kingdom and Brazil Lower Houses distribution for weights and thresholds set seems to be normally and uniformly distributed, respectively.



Source: Author.

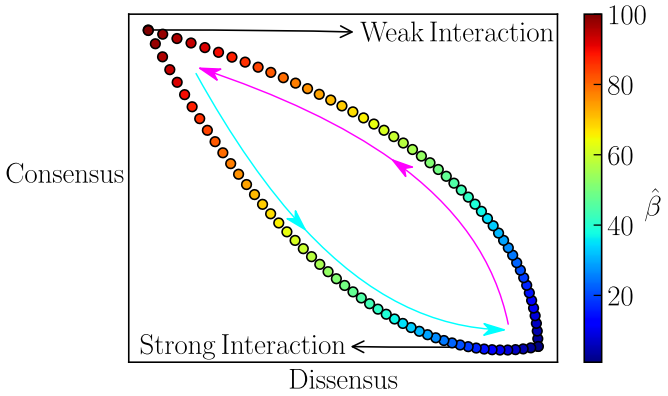
In Figure, for each pair $\left(f\left(\hat{\mathbb{W}}\right), f\left(\hat{\mathbb{H}}\right)\right)$, a label indicating the legislative term number (see Figure 23), year and name of Chief Executive power was set up (see Figure 26). Note that distributions for the United States Lower House is sparser than distributions for the United Kingdom and Brazil Lower Houses because the number of political parties oscillates between two and three. In comparison, the distribution function for weight sets in the United Kingdom and Brazil Lower Houses seems to be normally distributed around $w_{ij} = 0$. In the same way, distribution functions for threshold sets in both Lower Houses show to be uniformly distributed.

5.3 $\hat{\beta}$ learned: A new way for measuring the degree of political interaction

First thing mandatory to establish is the significance of $\hat{\beta}$ in Legislative Houses analysis. This parameter can be considered as an inverse measure of the degree of legislative interaction among political parties. Legislative interaction can be understood as a naive way to quantify how much political parties debate, discuss or share ideas when they come to present, argue and subsequently vote bills in legislative sessions. Numerically speaking, large $\hat{\beta}$ is related to weak degree of legislative interaction whereas small $\hat{\beta}$ is a sign of

strong legislative interaction among political parties. Figure 36 illustrates a simple way to understand the significance of $\hat{\beta}$.

Figure 36 – Significance of $\hat{\beta}$ in Legislative process. Beta can be understood as an inverse measure of legislative interaction among political parties. For large $\hat{\beta}$, consensual states are found, meaning that political parties have the same opinion to approve-disapprove the bill by using small quantities of debates. Contrary, small $\hat{\beta}$ is associated with a dissensual state in which political parties interact in large quantities of debates. $\hat{\beta}$ value was distributed as shown in the right color bar. The fuchsia-aqua cycle makes reference to the irreversible transition from consensus to dissensus for an individual bill.



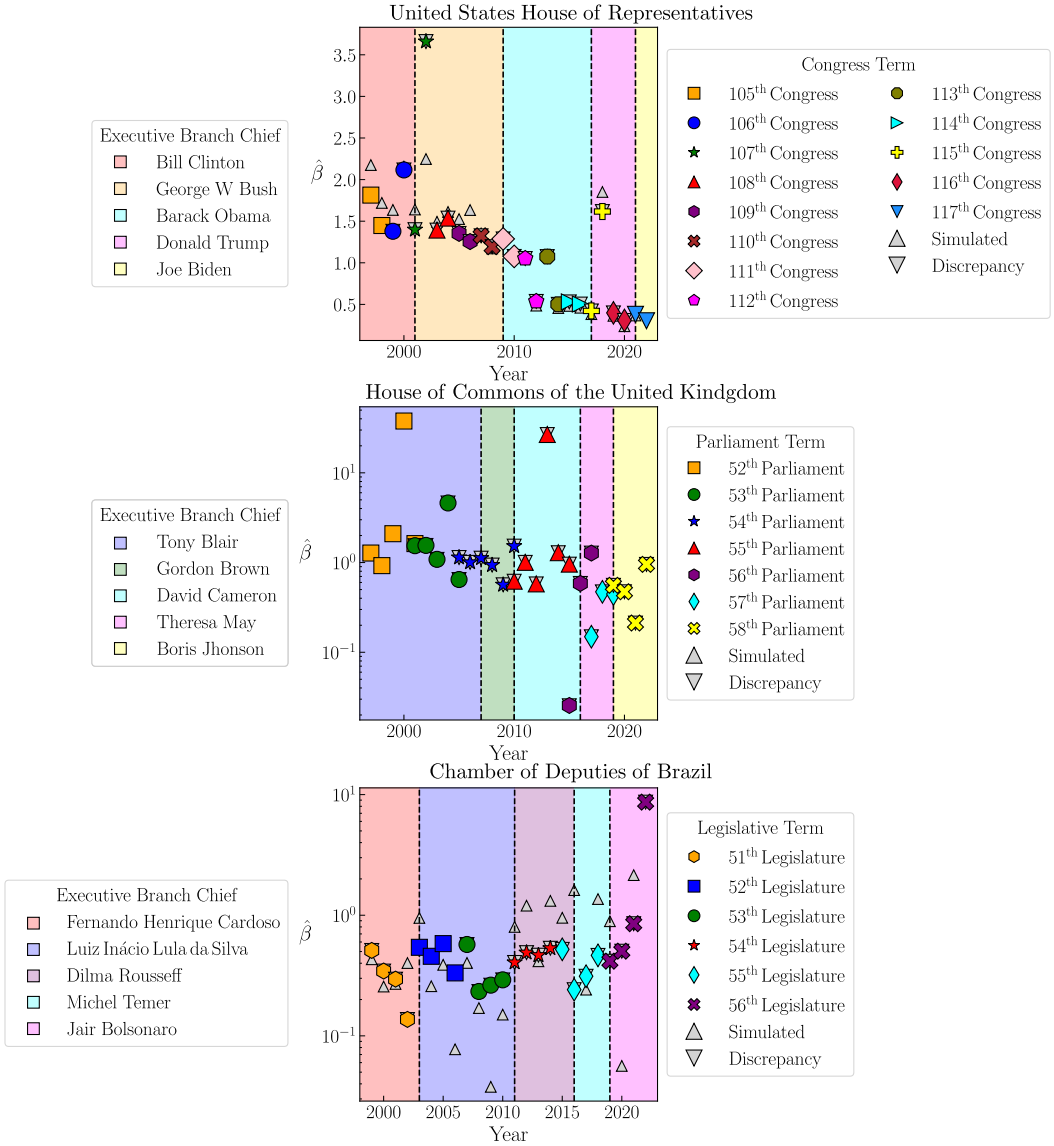
Source: Author.

Figure shows a set of coloured distributed points according to the color bar located at the right. Note that strong and weak degree of interaction were labelled for small $\hat{\beta}$ (most bluish point located at right-bottom) and large $\hat{\beta}$ (most reddish point located at the left-top), illustrating the premise aforementioned established. Besides that, Figure 36 shows another interesting feature that can be ingrained between $\hat{\beta}$ and consensual - dissensual states (vertical and horizontal axes). Small values of $\hat{\beta}$ shall be connected to the dissensual state because groups of political parties strongly disagree to approve-disapprove, meaning in more debates and discussions. In the opposite, a consensual state shall be associated with large values of $\hat{\beta}$ due to the fact that political parties have the same opinion for approving-disapproving bills, implying that either none or small quantity of debates or discussion are needed. As a major conclusion, $\hat{\beta}$ can be used to characterize transition from consensual to dissensual state of political parties. The fuchsia-aqua cycle in Figure 36 illustrates how the consensual-dissensual transition must happen. Note that direction and color of cycle defines an irreversible transition between consensus and dissensus. The fuchsia cycle means that for bill k_i, consensus is reached only from a dissensus state. Reverse cycle (aqua cycle) happens when political parties begin to discuss a new bill k_{i+1}, for example.

Legislative interaction parameter $\hat{\beta}$ and the association with consensus-dissensus

transition can be easily observed in Lower Houses Legislative procedures. Lets analyze the following example. It is known that bills with high importance over future executive branch power actions are sent to the Legislative branch in the first year term. This is because of the high level of controversy generated among political parties that are against and in favour of executive power proposals, resulting in a highly polarized state [1]. These bills are discussed, reviewed, voted on and in the majority of cases, amendments to specific bill issues are made in order to be discussed again, meaning in large quantities of debates and legislative sessions. As the time goes and amendments are made, more political parties begin to agree with the bill, and therefore the number of legislative sessions decreases, that is, the consensual state is reached. Establishing the significance of $\hat{\beta}$ in Legislative analysis, it is time to look at the result obtained from the Legislative Lower Houses databases. Set $\left\{\langle\sigma_i\rangle^{\mathrm{O}}, [\mathbf{C}]^{\mathrm{O}}_{ij}, u\left(\sigma_i, \sigma_j, \sigma_k\right)^{\mathrm{O}}\right\}$ for each Lower House was used as the input of Procedure 3 in order to learn $\hat{\beta}$. Figure 37 shows $\hat{\beta}$ learned from procedure 3 for the three different Lower Houses.

Figure 37 – $\hat{\beta}$ learned in Lower Houses Databases. Set $\left\{\langle\sigma_i\rangle^{\mathrm{O}}, [\mathbf{C}]^{\mathrm{O}}_{ij}, u\left(\sigma_i,\sigma_j,\sigma_k\right)^{\mathrm{O}}\right\}$ was used as the input of Procedure 3 in order to learn $\hat{\beta}$. Upper, center and lower panel are related to the United States of America House of Representatives, United Kingdom House of Commons and Brazil Chamber of Deputies. In all panels, black dashed lines and coloured zones were used to separate executive terms. Executive and Legislative term are displayed in the legends boxs located at right and left of each panel. In each panel, a marker, representing a specific Legislative term shows $\hat{\beta}$ learned for each year and Lower House. Gray down triangles are associated with $\hat{\beta}$ obtained from the minimization of the discrepancy between learned parameters and learned scaled parameters and gray up triangles are values of $\hat{\beta}$ learned obtained procedure 3 using simulated data gathered from Procedure 6. For some years in the United States of America and Brazil panels, $\hat{\beta}$ obtained from simulated data are away from the value obtained using the real data. This is mainly because some correlation values were not feasible to be reconstructed and therefore, these values were recomputed. In the case of the United Kingdom it was not possible to generate simulated data because the majority of correlation values were not feasible to be reconstructed.



Source: Author.

Upper, center and lower panels of Figure 37 are related to the United States House of Representatives, the House of Commons of the United Kingdom and Chamber of Deputies of Brazil. Executive terms (see Figure 26) were coloured and black dashed

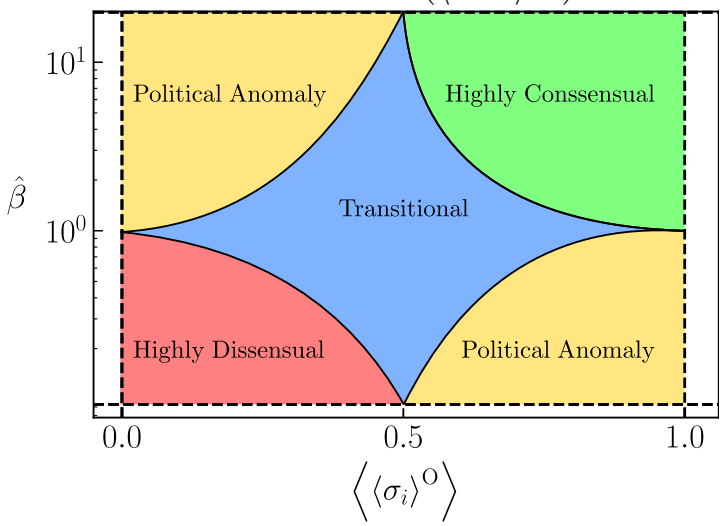
lines were used to separate them and they can be recognize by seeing the legends boxs located at right and left of each panel. Each panel shows $\hat{\beta}$ learned using procedure 3 for the United States of America, United Kingdom and Brazil, respectively. Besides, there are two more markers for each panel. Gray down triangles are associated with $\hat{\beta}$ obtained from the minimization of the discrepancy between learned parameters and learned scaled parameters, that is, by dividing $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ by $\hat{\beta}$ as detailed in Chapter 1. On the other hand, gray up triangles are values of $\hat{\beta}$ learned obtained procedure 3, but in this case, simulated data gathered from Procedure 6 was used as the input parameters. Note that for some years in the United States House of Representatives and Chamber of Deputies of Brazil panels, $\hat{\beta}$ obtained from simulated data are away from the value obtained using the real data. This issue comes from the fact that in order to generate simulated data, some correlation values were not feasible to be reconstructed and therefore, these values were recomputed. In the case of the House of Commons of the United Kingdom it was not possible to generate simulated data because the majority of correlation values between political parties were not feasible to be reconstructed.

Note that there are three large values of $\hat{\beta}$ learned that are worth of commenting on. Specifically, these points are located in 2001 for the United States House of Representatives and in 2000 and 2013 for the House of Commons of the United Kingdom. It was proposed that large value of $\hat{\beta}$ learned implies a high degree of consensus among political parties and these points prove that this argument is true. Explicitly, in 2001, the United States of America experience terrorist attacks, for that reason, the House of Representatives have to deal with bills related to the interception and obstruction of terrorism [22]. Given the situation, these bills were approved almost by unanimous consent. On the other hand, the large value of $\hat{\beta}$ in 2000 is associated with the introduction of the House of Lords Act 1999 [23]. An Act to restrict membership of the House of Lords by virtue of a hereditary peerage. The aim of this act was to make the House of Lords more democratic and representative. This Act decreased the membership of the House from 1,330 in October 1999 to 669 in March 2000. At that time, Act 1999 was well received by Commons. Finally, in 2003, the House of Commons had to discuss the government motion related to the use of chemical weapons in Syria [24, 25]. The motion was defeated by the House of Commons, meaning that at some point, Commons partially or completely agreed and therefore, large values of $\hat{\beta}$ does not make sense.

5.4 Consensual and Dissensual zones: Analysing $\hat{\beta}$ and $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$

Inconsistency in the value of $\hat{\beta}$ for the United kingdom in 2003 can be understood if a comparison between $\hat{\beta}$ learned and the expected value of the political party opinion mean value $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ is made. It was established that $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ close to zero (one) is associated with a high degree of dissensus (consensus). In this sense, by analysing all combinations of small and large values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ and plotting them, we can define regions of dissensus, consensus and for the best of our knowledge, regions of political anomalies. Our proposal is depicted in Figure 38.

Figure 38 – Consensual and dissensual states regions. In the plot, the horizontal axis is related to the expected value of the political party opinion mean value and the vertical axis it is related to $\hat{\beta}$ learned which is plotted in log scale. Four coloured regions were depicted according to all possible combinations of coordinates $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$. Red and green coloured zones located at the bottom-left and upper-right are associated to dissensual and consensual states, given that $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ are both either small or large. Yellow coloured regions located at upper-left and bottom-right of the plot are associated with political anomalies given the inconsistency between values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$. Finally, the blue region located at the centre of the plot is related to all intermediary values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$.

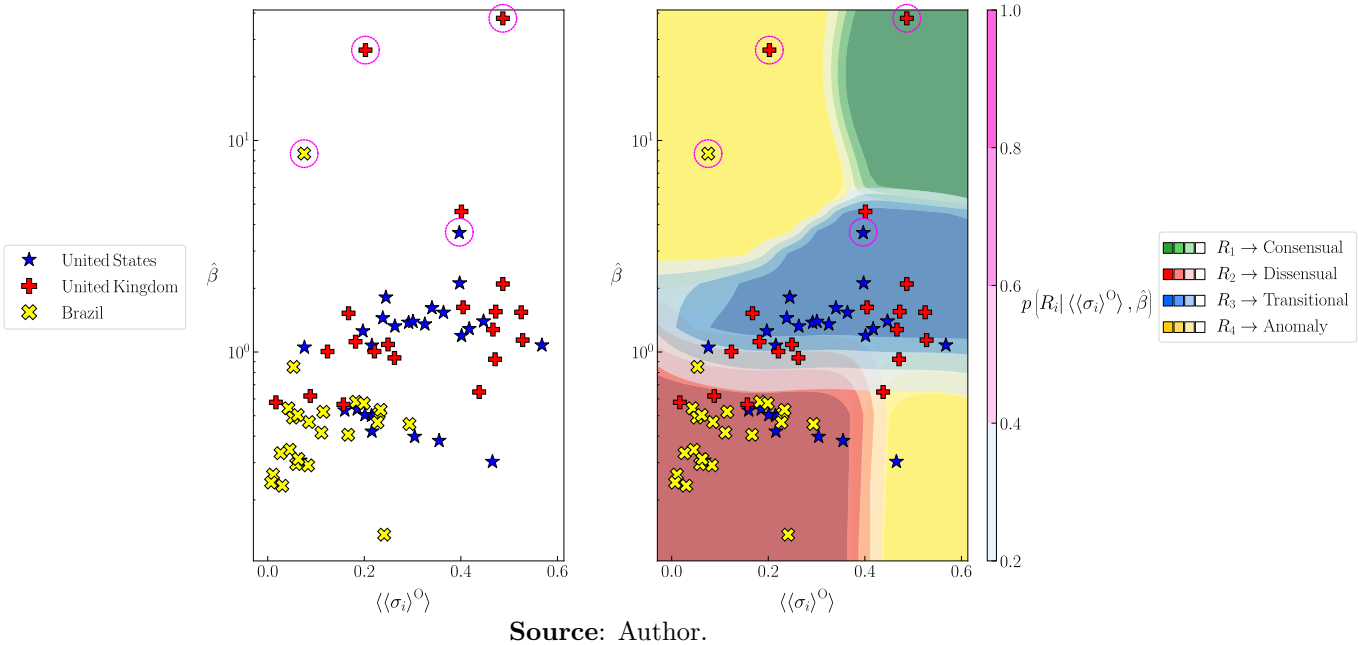


Source: Author.

Figure 38 exhibits in the horizontal axis the expected value of the political party opinion mean value with bounds fixed between zero and one and in the vertical axis it is found in a log scale, $\hat{\beta}$ learned. Four coloured regions were depicted according to all possible combinations of coordinates $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$. When both values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ are either small or large a highly dissensual or consensual state among political parties can be associated. Dissensual and consensual regions were red and green coloured and these zones are located at the bottom-left and upper-right of the plot. On the other hand, when values

of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ are either small-large or large-small this is a sign of political anomaly due to the fact that it is not possible to comprehend the fact that for example, political parties globally disagree in opinion but they are not interacting to decide that. Anomaly regions were yellow coloured and they are located at upper-left and bottom-right of the plot. Finally, there is a blue zone named as transitional region. This region encompasses all intermediate values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ that are between dissensual and consensual zones. Figure 39 shows the plot of $\hat{\beta}$ vs $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ using the result obtained for the three lower Houses.

Figure 39 – Consensual and dissensual zones for the three Lower Houses. In both panels, is exposed the values of $\hat{\beta}$ learned in the horizontal and $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ in the vertical axis. Blue stars, red plus and yellow rotated plus are associated with the values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ for the United States of America, the United Kingdom and Brazil Lower Houses. In the right panel, it is shown the same plot but this time with coloured zones. Coloured zones were obtained by doing a classification of all possible regions described on the right center legend. The classification of these zones is based on a Multilayer Perceptron algorithm. Colormaps on the center right represent what is the probability for a point $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ to belong to a region R_i. The possible values for probabilities are depicted in the fuchsia colorbar next to the right panel. Note that probability value is related to the contrast of color. Fuchsia circle was used to highlight points worth of commenting.



Source: Author.

In both panels of Figure 39 is exposed the values of $\hat{\beta}$ learned and $\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle$ in the horizontal and vertical axis, respectively. Again, blue stars, red plus and yellow rotated plus are associated with the values of $\left(\left\langle \langle \sigma_i \rangle^{\mathrm{O}} \right\rangle, \hat{\beta}\right)$ for the United States of America, the United Kingdom and Brazil Lower Houses. The right panel of Figure 39 shows the same plot but this time with coloured zones. Each coloured zone is associated with each possible

region described above. Legend on the right center of the Figure shows the colormap chosen for each region. In each coloured zone, the colormap reveals what is the probability for belonging to a region R_i given a point $\left(\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle, \hat{\beta}\right)$. Observe the levels of the colorbar next to the right panel. A fuchsia colormap was selected only to exhibit how the probability value is associated with the contrast of the color. These zones were obtained by employing a Multilayer Perceptron Classifier Algorithm [26] over the set of points obtained for the three Lower Houses. In the machine learning context, highly dissensual, highly consensual, anomaly and transitional zones are depicted as the decision zones learned for the Multilayer Perceptron over a multiclass classification problem. Observe how the regions better improve the way that points $\left(\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle, \hat{\beta}\right)$ can be interpreted. Specifically, all points $\left(\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle, \hat{\beta}\right)$ for Brazil Lower House are located in the dissensual zone; For the United States of America points $\left(\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle, \hat{\beta}\right)$ are distributed among dissensual and transitional zones and finally, points for the United Kingdom are distributed among all zones. A highlight of three points of $\hat{\beta}$ learned, two for the United States House of Representatives, two for the House of Commons of the United Kingdom and one more for the Chamber of Deputies of Brazil were made in Figure 39. These points were outstanded using a fuchsia circle. The two points mentioned of the House of Commons of the United kingdom are located in the consensual and anomaly zone. Point located in the consensual zones is associated with the implementation of Act 1999 and the point located in the anomaly zone is for the discussion about use of chemical weapons in Syria. Points related to the United States located at the edge of the transitional zone towards the consensual zone is related with bills discussed after the 2001 terrorist attacks and the other one located in the anomaly zone is related to the COVID-19 epidemic state of emergency. The point highlighted in Brazil is related to the first election of Luis Inácio Lula da Silva as the President of Brazil in 2002.

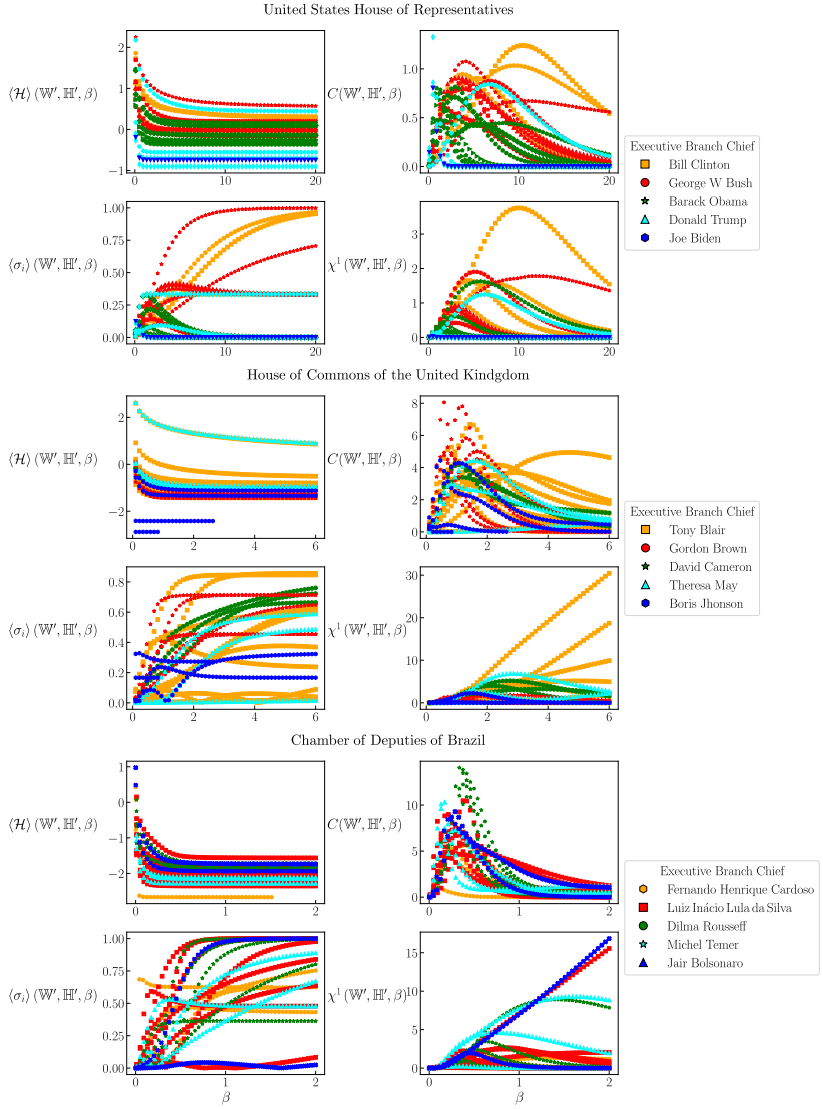
5.5 Finding $\hat{\beta}_{\max}$: An estimator to know the transition from dissensual to consensual state

Above result shows that it is possible to characterize dissensual and consensual states just by measuring $\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle$, learning $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ to obtain $\hat{\beta}$ and classifying $\left(\left\langle\langle\sigma_i\rangle^{\mathrm{O}}\right\rangle, \hat{\beta}\right)$ to assign it a zone. However, the assignment of zones (Consensual, dissensual, transitional and anomaly) are static in the sense that the chosen zone is extracted from the political parties' majority opinion array. In a more dynamic point of view one can ask if is it possible

to know the value in which the transition from dissensual to consensual states happens for each year in each Lower House? In order to gather this value a characterization of the partition function $Z_B\left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right) = Z_B\left(\hat{\mathbb{W}}/\hat{\beta}, \hat{\mathbb{H}}/\hat{\beta}, \beta\right)$ defined in (3.3) respect to β must be made[1]. Specifically, it were computed the dependence in β of energy function average value $\langle \mathcal{H} \rangle \left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$, its fluctuation value $C(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$, average value $\langle \sigma_i \rangle \left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$ and its first order fluctuation $\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ defined in (3.65) using set $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ learned for each year in each Lower House. The characterization was carried out by evaluating the aformentioned functions in a defined range of Legislative interaction. Result of this characterization is exposed in Figure 40.

[1] Note that $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}\right\}$ were scaled by $\hat{\beta}$ in accordance with results found in Chapter 1.

Figure 40 – Characterization of $Z_B\left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$ using $\left\{\hat{\mathbb{W}}, \hat{\mathbb{H}}, \hat{\beta}\right\}$ learned for each year in each Lower House. It was computed the dependence in β of energy function average value $\langle \mathcal{H} \rangle \left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$, its fluctuation value $C(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$, average value $\langle \sigma_i \rangle \left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$ and its first order fluctuation $\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ defined in (3.65) using learned parameters. Results for each Lower House are composed of four panels. Four upper panels are associated with the United States House of Representatives, Four center panels are intended for the House of Commons of the United Kingdom and the four lower panels are for the Chamber of Deputies of Brazil. A marker and color was assigned to identify each Executive term in each Lower House as depicted in the right boxes.

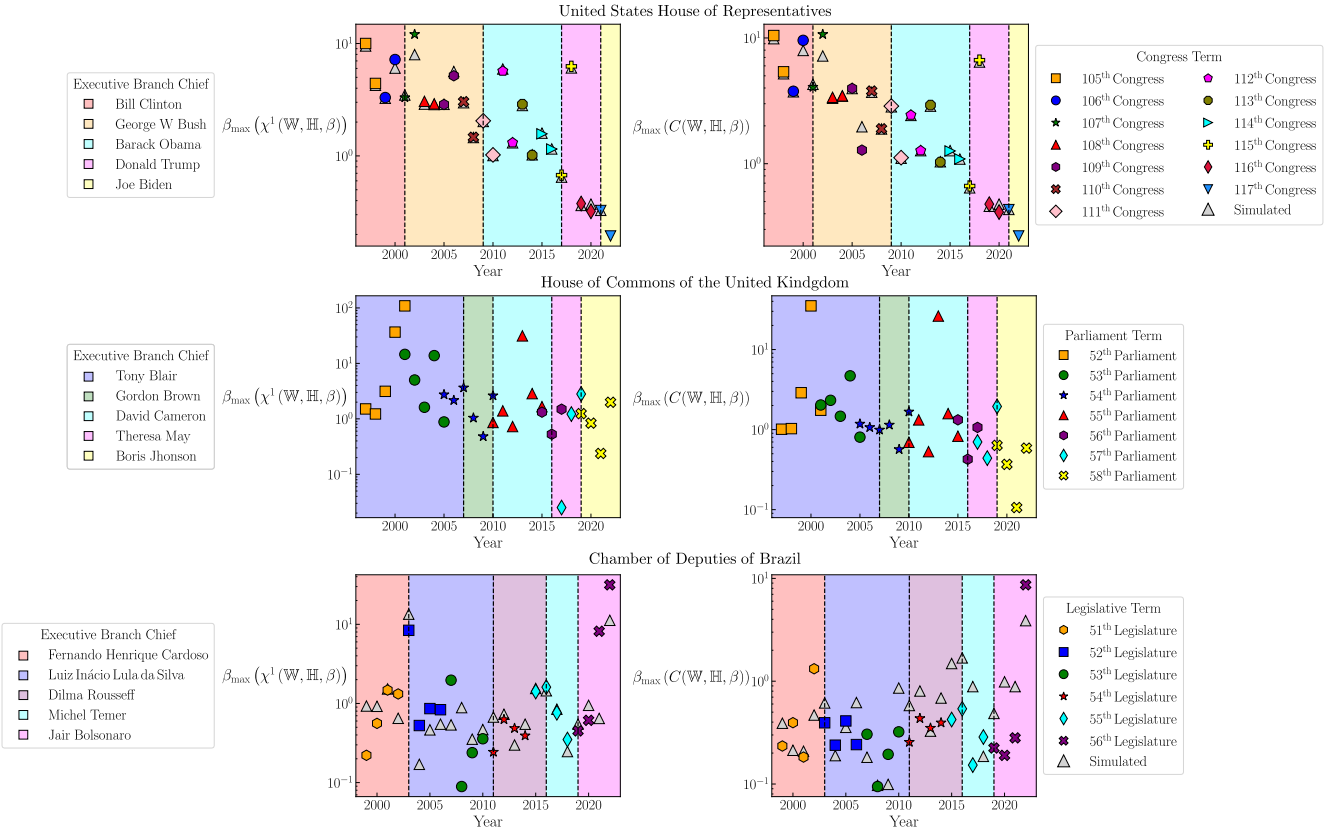


Source: Author.

Figure 40 shows the result of the partition function characterization for the three Lower Houses. Results for each lower House consist of four panel groups. The upper group is associated with the United States House of Representatives, center group is for the House of Commons of the United Kingdom and the lower group is related to the Chamber of Deputies of Brazil. Note that the right legend boxes show that for each Executive term a marker and color was associated. In order to find values in which transition from dissensual to consensual states happens it is necessary to compute the value of β in

which the fluctuation of $\langle \sigma_i \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ is maximum. This is because $\langle S_i \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ is an estimator of how much Political parties are in agreement or disagreement given the interaction value β. $\langle \sigma_i \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ close to zero implies in dissensus and close to one are associated to consensus. In fact, $\langle S_i \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ has the same role with $\langle \sigma_i \rangle^{\mathrm{O}}$ due to it is easy to realize that $\langle S_i \rangle^{\mathrm{O}}$ is the measure of the average political opinion among parties when $\beta = \hat{\beta}$, that is $\langle \sigma_i \rangle^{\mathrm{O}} = \langle S_i \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \hat{\beta})$. Figure 41 shows maximum values computed over the fluctuation of $\langle \sigma_i \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ and $\langle \mathcal{H} \rangle (\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)$ namely as $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ and $\beta_{\max}\left(C(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$, respectively, for each year in each Lower House. Maximum values were computed averaging results obtained using *Brent* [11] and *Golden* [27] numerical methods.

Figure 41 – Maximum values of $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ and $\beta_{\max}\left(C(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$. Brent and Golden methods were used to find maximum values. Final values arise as the average between these two results. Set of two panels corresponds to the result for each Lower House. The upper group is dedicated to the United States House of Representatives, the center group is for the House of Commons of the United Kingdom and the lower group is related to the Chamber of Deputies of Brazil. The value in which the transition from dissensual to consensual state happens is associated with $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$. Large values of $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ can be realized by how well organized a Lower House is in order to discuss bills.



Source: Author.

Figure 41 shows three groups of two panels. The upper group is for the United

States House of Representatives, the center group is for the House of Commons of the United Kingdom and the lower group is for the Chamber of Deputies of Brazil. Left panels are associated to the maximum value of β obtained from $\chi^1\left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$ whereas right panels are related to the maximum values computed through $C\left(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta\right)$. Again, colored zones and black dashed lines are used to delimitate executive terms and a colored marker was assigned for each Legislative term. The value in which the transition from dissensual to consensual state happens is associated with $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$. Transition in this scenario must be associated with the capacity of Political Parties to maximize the degree of Legislative interaction. In other words, $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ can be understood as an estimator of the optimal number of discussions, sessions and debates to be held by Lower Houses members in order to maximize Legislative effectiveness. Remember that β is an inverse measure of Legislative interaction among political parties. In this sense, large values of $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ implies small Legislative interaction, meaning that transition is set up with a small number of discussions or debates. Contrary, if $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ is small, a high level of Legislative interaction is needed for triggering the transition, suggesting more discussions.

Note that in the United States House of Representatives from 1997 to 2015, transition must take place in $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right) > 1$. This result makes sense because Legislative term in the United States is only a 2-year term, therefore a restricted number of legislative sessions were fixed. Take attention now to the period of time from 2019 to 2022 ($\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right) < 1$). If Legislative term continues to be the same, then this change it is explained by the fact that number of Legislative sessions have been increased along time in the United States Lower House. On the other hand, values of $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ for the House of Commons of the United Kingdom follow a linear function like fashion. Note that values alternate around $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right) = 1$ but when $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right) > 1$ values are smaller when compared to the values obtained from the United States House of Representatives. This comparison makes sense because legislative term in the United Kingdom is greater than in the United States, therefore, more time to discuss bills is provided and consequently $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ for each year in this Lower House must be smaller than values in the United States Lower House. Finally, take attention to the fact that values of $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ for the Chamber of Deputies of Brazil are the smallest of the three Lower Houses. This is because the Legislative term in Brazil is

four year-term, therefore, Deputies have a long range available for discussing bills. As a final conclusion, $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ measures capacity for maximizing legislative work, when $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ is large, less discussions are needed to give a decision on a bill, in contrast, $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}', \hat{\mathbb{H}}', \beta)\right)$ small is an indicator of large quantity of discussion are mandatory.

5.6 Lower Houses Members analysis via nearest correlated cluster algorithm

Results discussed so far have been extracted just by employing the political parties' majority opinion matrix to run the extended Bolztmman machine discussed in Chapter 1. In other words, we have used an interaction point of view to analyze Lower Houses. Now, it is time for dealing with the statistical way to analyze Lower Houses. The aim is to better understand Legislative collective behavior features. This information can be extracted from the analysis of the roll-call vote matrix based on the Nearest Correlated Cluster Algorithm (NECO) exposed in Chapter 2. The roll-call vote matrix for a particular Lower House in a defined year is given by

$$\mathbf{V}_{nk} = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \\ M_4 \\ M_5 \\ \vdots \\ M_n \end{array} \overbrace{\begin{pmatrix} -3 & 2 & 1 & -1 & \cdots & 2 \\ 0 & -1 & 0 & 2 & \cdots & 1 \\ -2 & 2 & -3 & -3 & \cdots & -3 \\ 2 & 3 & -1 & 0 & \cdots & 0 \\ -2 & 1 & 2 & -1 & \cdots & -3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -2 & -1 & 0 & -3 & \cdots & 1 \end{pmatrix}}^{\text{Bills by year}}, \tag{5.2}$$

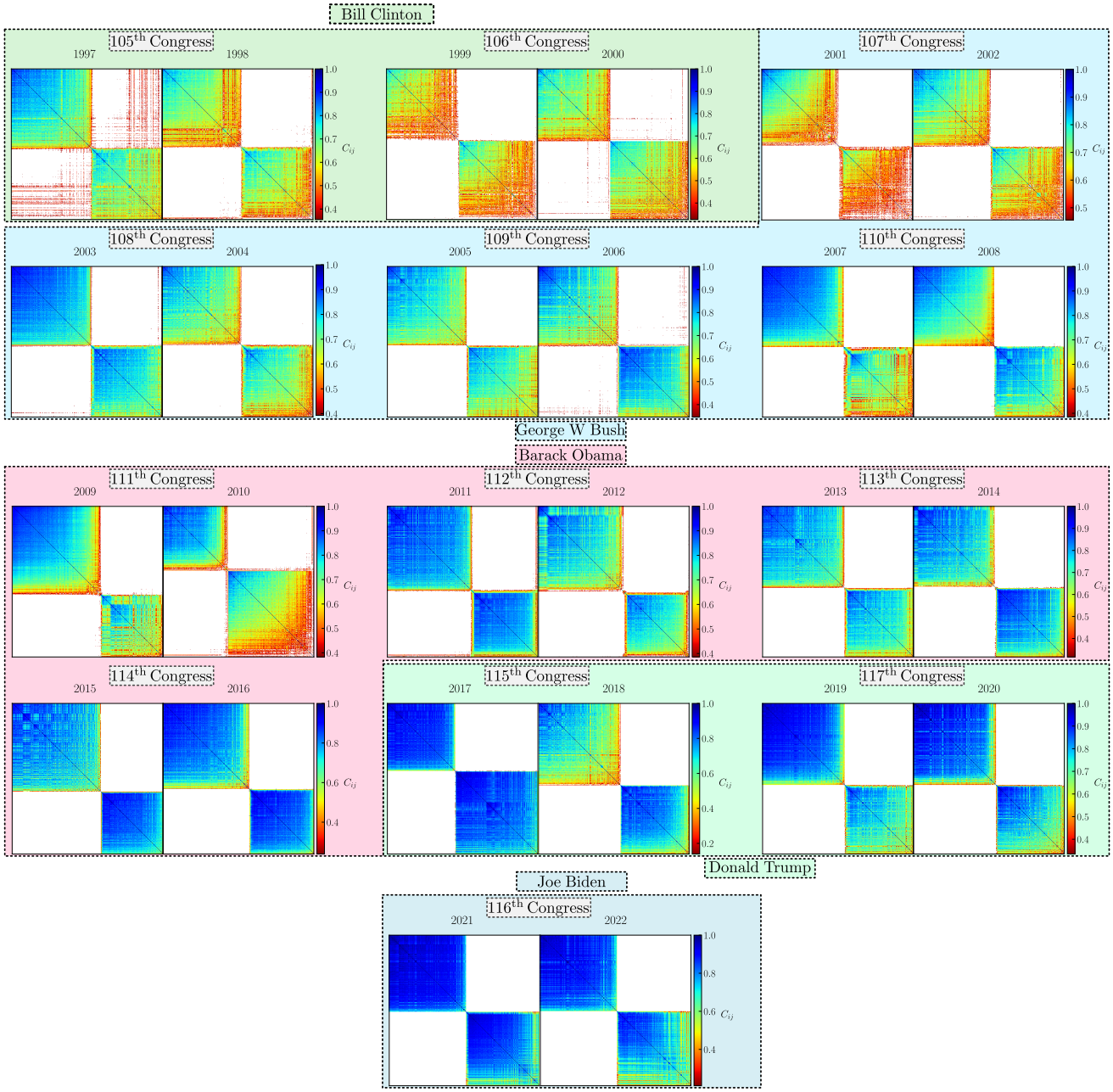
where Lower Houses members are located in the rows. This matrix is the input quantity in procedure 6. Note first step in procedure is to computes the correlation distance matrix

$$\boldsymbol{d}\left[\mathbf{V}_{nk}\right] = \sqrt{2\left(1 - \boldsymbol{\rho}\left[\mathbf{V}_{nk}\right]\right)}, \tag{5.3}$$

where $\boldsymbol{\rho}[\mathbf{V}_{nk}]$ is the correlation matrix. This matrix is obtained by computing the correlation distance coefficient among all pairwise sequences of votes among Lower Houses members (green, cyan and purple rectangles in (5.2)). In other words, correlation distance values are gathered from the computation on the correlation distance between pairs of votes sequences spanned over all bills voted in a year. Correlation distance varies in the fixed range $[0, 2]$. If $\boldsymbol{d}[\mathbf{V}_{nk}] = 0$ then $\boldsymbol{\rho}[\mathbf{V}_{nk}] = +1$ meaning that sequences of votes of Lower Houses members are the same. Contrary, if $\boldsymbol{d}[\mathbf{V}_{nk}] = 2$ then $\boldsymbol{\rho}[\mathbf{V}_{nk}] = -1$ and sequences of votes of Lower Houses members are opposed. Closer correlation distance means highly correlated Lower Houses Members. Finally, if $\boldsymbol{d}[\mathbf{V}_{nk}] = \sqrt{2}$ then $\boldsymbol{\rho}[\mathbf{V}_{nk}] = 0$ and therefore sequences of votes between Lower Houses members are independent.

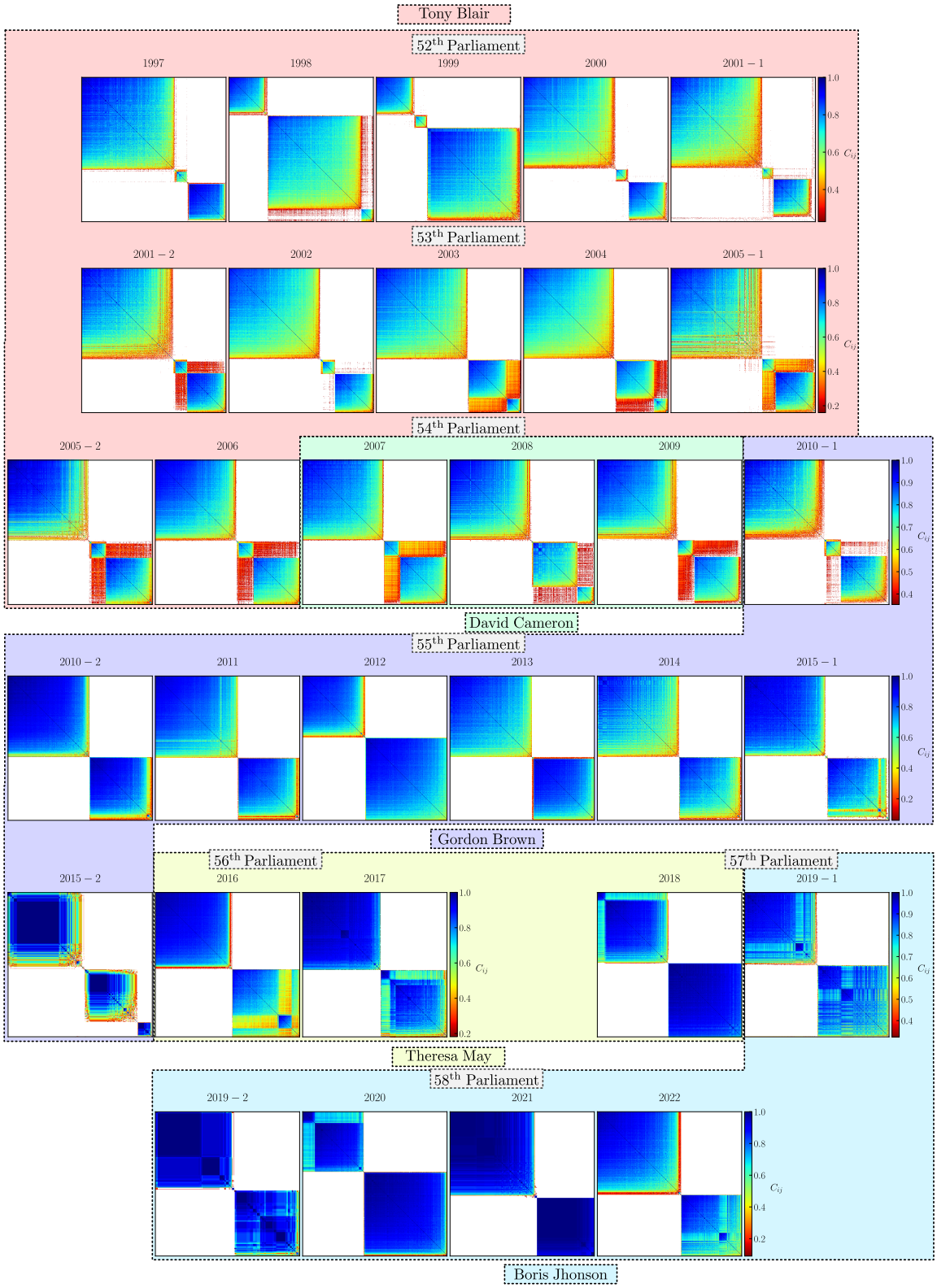
The roll-call vote matrices for the three Lower Houses in each year were computed and subsequently, procedure 6 was used. The first result that this procedure dispatches are the ordered correlation matrices with unnecessary values removed. Figures 42, 43, 44 show the resulting matrices for the United States House of Representatives, House of Commons of the United Kingdom and Chamber of Deputies of Brazil, respectively.

Figure 42 – Correlation matrices for the United States House of Representatives. Roll-call vote matrices for this Lower House were used as the input of procedure 6. Ordered matrices with unnecessary values removed were grouped by Congress term and a box surrounding groups of matrices indicates the name of Executive Chief Branch. A color box was designed for each Executive term. Right colorbar indicates the value of correlation. Blueish color is intended for consensus and reddish colors are related to dissensus. C_{ij} label is intended for denominating correlation coefficient between Representative i and j. As a general feature, ordered matrices present two main clusters.



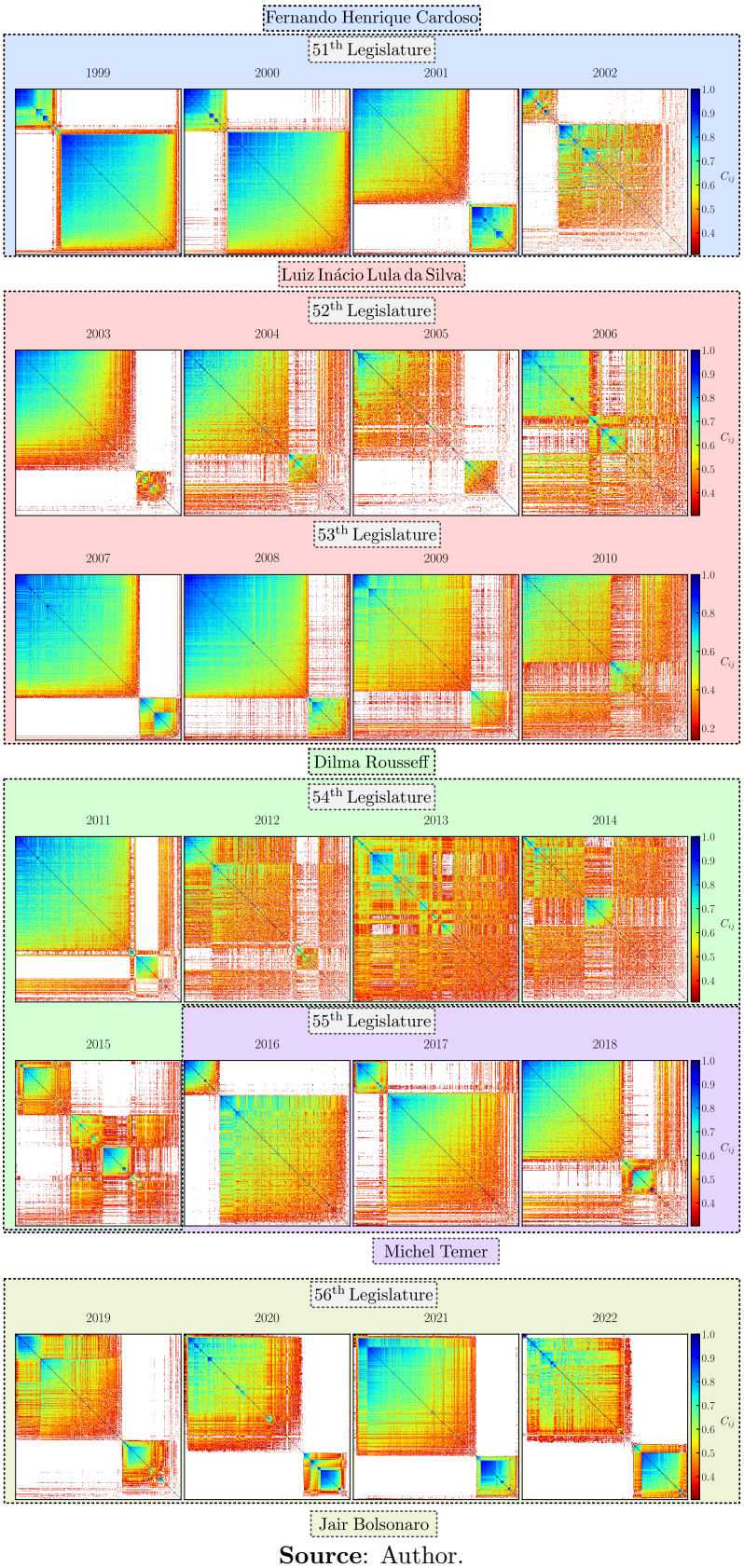
Source: Author.

Figure 43 – Correlation matrices for the House of Commons of the United Kingdom. Roll-call vote matrices for this Lower House were used as the input of procedure 6. Ordered matrices with unnecessary values removed were grouped by Parliament term and a box surrounding groups of matrices indicates the name of Executive Chief Branch. A color box was designed for each Executive term. Right colorbar indicates the value of correlation. Blueish color is intended for consensus and reddish colors are related to dissensus. C_{ij} label is intended for denominating correlation coefficient between Common i and j. A transition from three main clusters to two is observed from 2010-I. Besides, note that an anomalous symmetry is observed in 2015-II, precisely a year before the Brexit referendum.



Source: Author.

Figure 44 – Correlation matrices for the Chamber of Deputies of Brazil. Roll-call vote matrices for this Lower House were used as the input of procedure 6. Ordered matrices with unnecessary values removed were grouped by Legislative term and a box surrounding groups of matrices indicates the name of Executive Chief Branch. A color box was designed for each Executive term. Right colorbar indicates the value of correlation. Blueish color is intended for consensus and reddish colors are related to dissensus. C_{ij} label is intended for denominating correlation coefficient between Deputy i and j. An anomalous behavior is observed from 2012 to 2015. This period of time is associated with the pre-impeachment period of Dilma Rousseff.

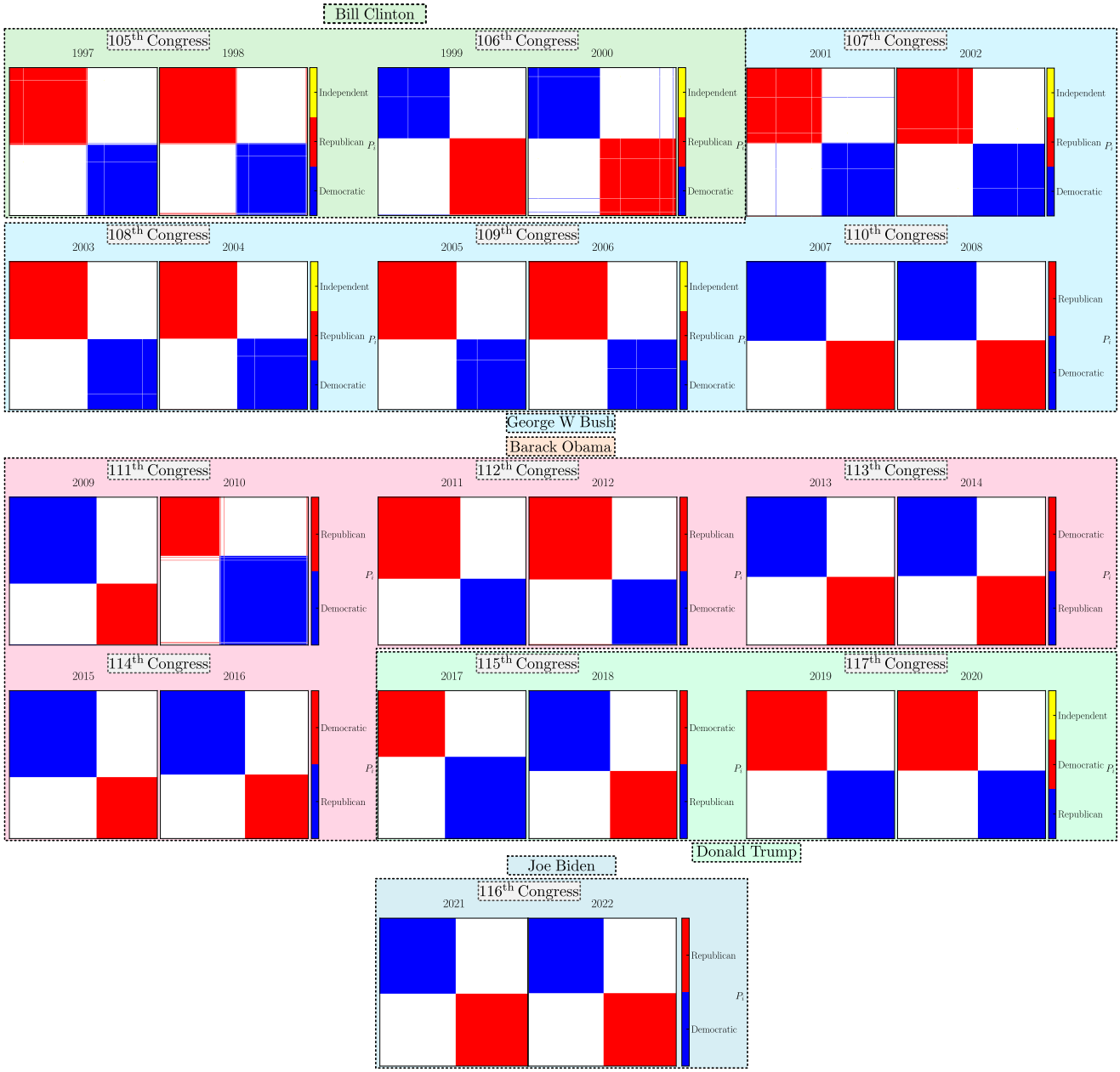


Source: Author.

Figures 42, 43 and 44 show ordered matrices with unnecessary values removed. Note that matrices were grouped by Legislative term and a box surrounding groups of matrices indicates an Executive term. Each colored box represents a specific Executive term. Color bar found in the rightest part of each Figure indicates the values correlation. Blueish color is intended for consensus degree (correlation values greater than zero and smaller or equal to one) and reddish colors are related to dissensus (correlation values smaller than zero and smaller or equal to minus one). In the United States House of Representatives Figure is observed two main clusters and this feature remains for all executive terms. For the House of Commons of the United Kingdom three mains clusters are noted from 1997 to 2010-1. From that time, a transition for two main clusters appears. Besides that, note that in 2015-2 exhibits an anomalous symmetry. This year is associated with the pre-Brexit referendum. For the Chamber of Deputies of Brazil a different behavior is observed. Note that in each Legislature, a transition from three to one cluster happens. In fact, the transition is really fast for the first Executive Term of Dillma Rousseff ended in 2015, the pre-impeachment year.

Taking advantage of the ordered matrices produced by procedure 6, it is possible to change color features in matrices. Instead of coloring by correlation, we can assign a color by political party. Figures 45, 46 and 47 show the result for the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil.

Figure 45 – Correlation matrices colored by political parties for the United States House of Representatives. If a pair (R_i, R_j) of Representatives belong to the same political party a specific color is assigned. Possibilities for colors are exposed in the colorbar found in the rightest part of Figure. P_i label in the colorbar is intended for i-Political party. Grouping by Legislative Congress and Executive term was maintained.

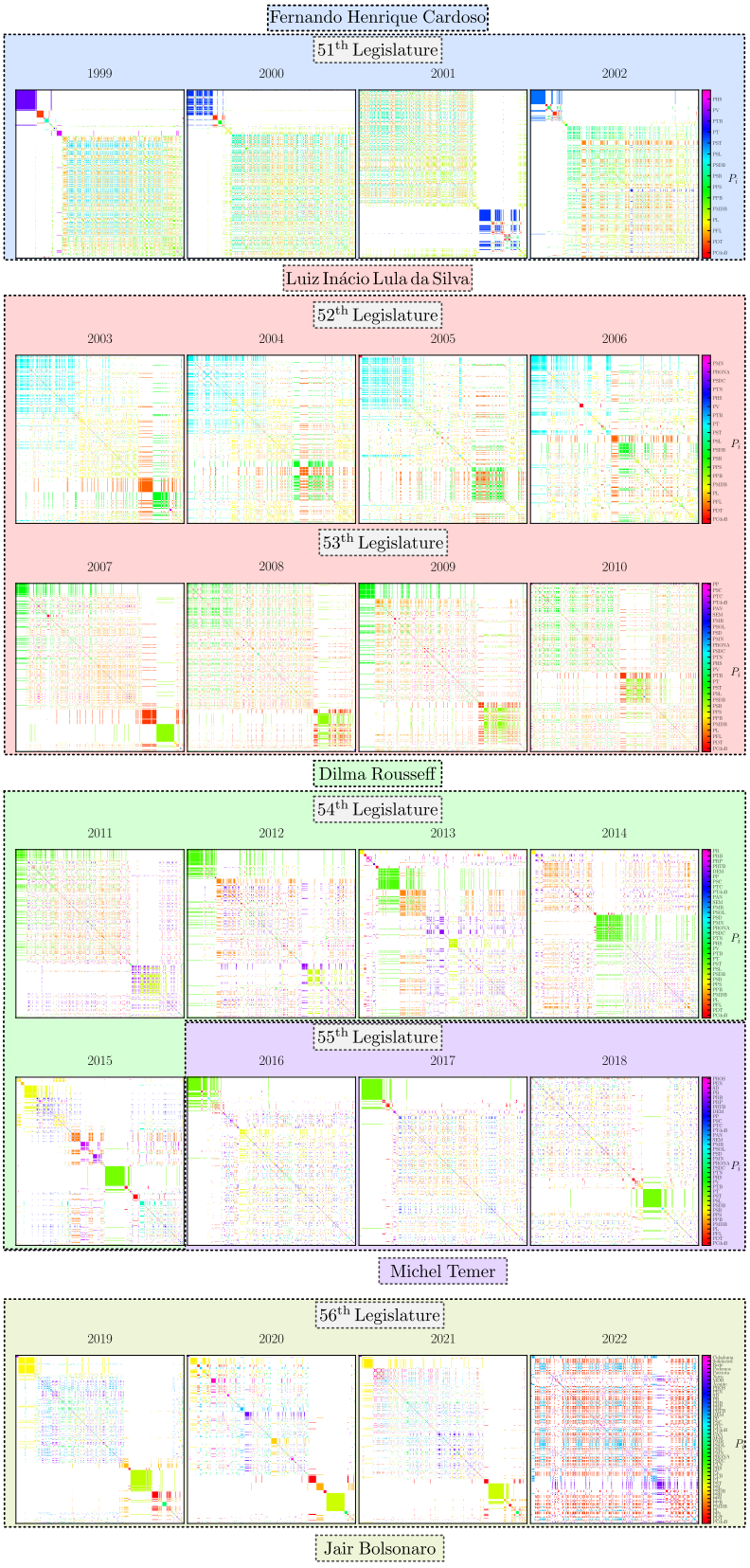


Source: Author.

Figure 46 – Correlation matrices colored by political parties for the House of Commons of the United kingdom. If a pair (C_i, C_j) of Commons belong to the same political party a specific color is assigned. Possibilities for colors are exposed in the colorbar found in the rightest part of Figure. P_i label in the colorbar is intended for i-Political party. Grouping by Legislative Congress and Executive term was maintained.

Source: Author.

Figure 47 – Correlation matrices colored by political parties for the Chamber of deputies of Brazil. If a pair (D_i, D_j) of Deputies belong to the same political party a specific color is assigned. Possibilities for colors are exposed in the colorbar found in the rightest part of Figure. P_i label in the colorbar is intended for i-Political party. Grouping by Legislative Congress and Executive term was maintained.

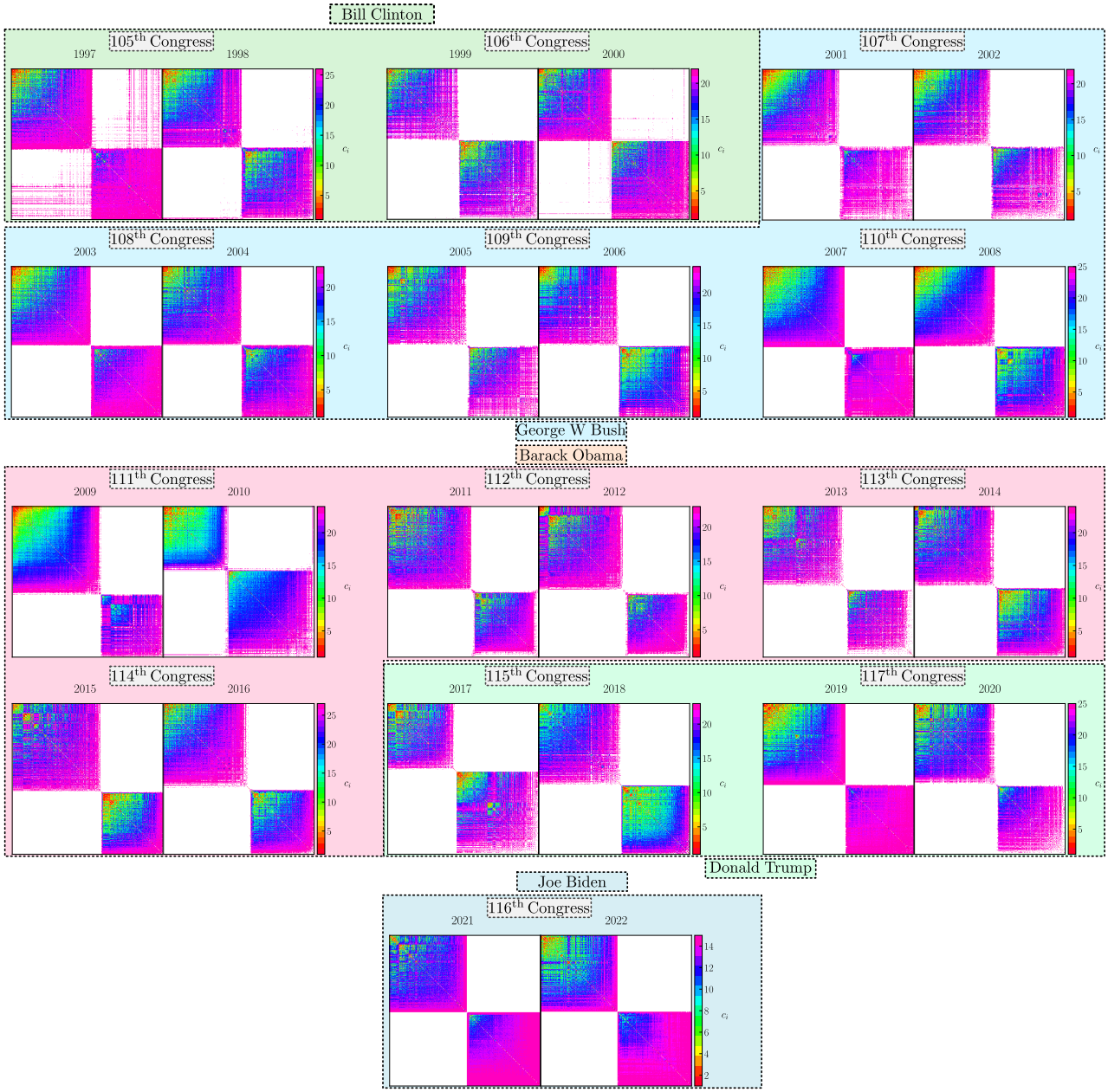


Source: Author.

Ordered political parties matrices were plotted in the following way. If a pair of Lower Houses members (M_i, M_j) belong to the same political party, a predefined color is set up. Possibilities for colors are exhibited in the colorbar found at the rightest of each Figure. Note that Legislative and executive term grouping features were maintained. An interesting fact that can be highlighted from ordered matrices colored by political parties is how it is possible to realize how strong the political parties cohesive degree is. Note that color clusters for Figures concerning the United States House of Representatives and the House of Commons of the United Kingdom are well defined in comparison with color clusters from the Chamber of Deputies of Brazil. This is a sign that in the first two Lower Houses mentioned, decisions and opinions are strongly related to the political parties collective orientation.

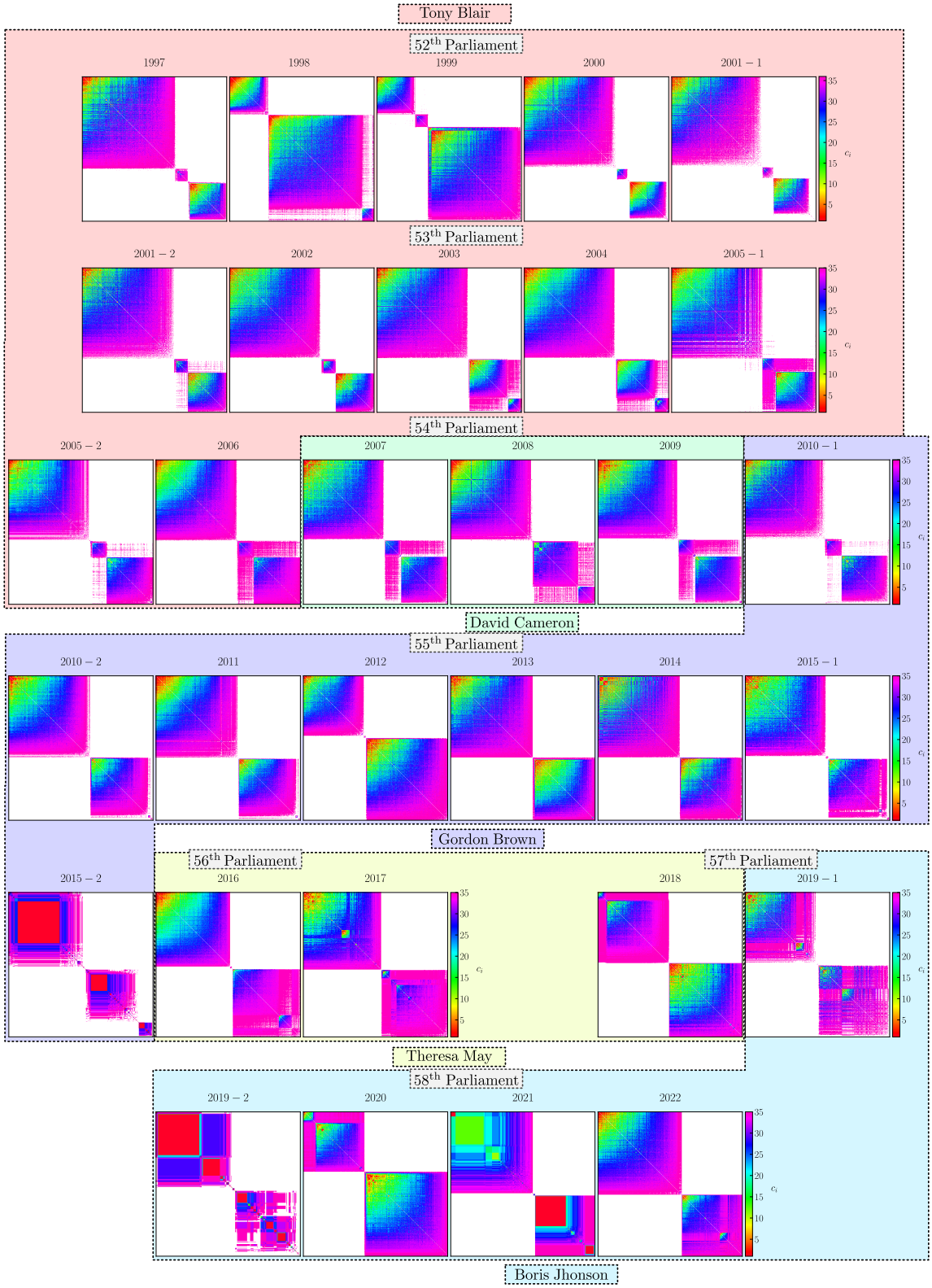
Last result obtained from the NECO algorithm is the cluster detection based on values of correlation distance. Figures 48, 49 and 50 show ordered matrices along with the detected cluster found by procedure 6 for the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively.

Figure 48 – Ordered correlation matrices colored by learned clusters for the United States House of Representatives. Using procedure 6 it is possible to detect clusters based on the values of the correlation distance coefficient. For each cluster learned a color was assigned. Possibilities for color are shown in the colorbar located at the righest of Figure. c_i label in the colorbar is intended for i-detected cluster.



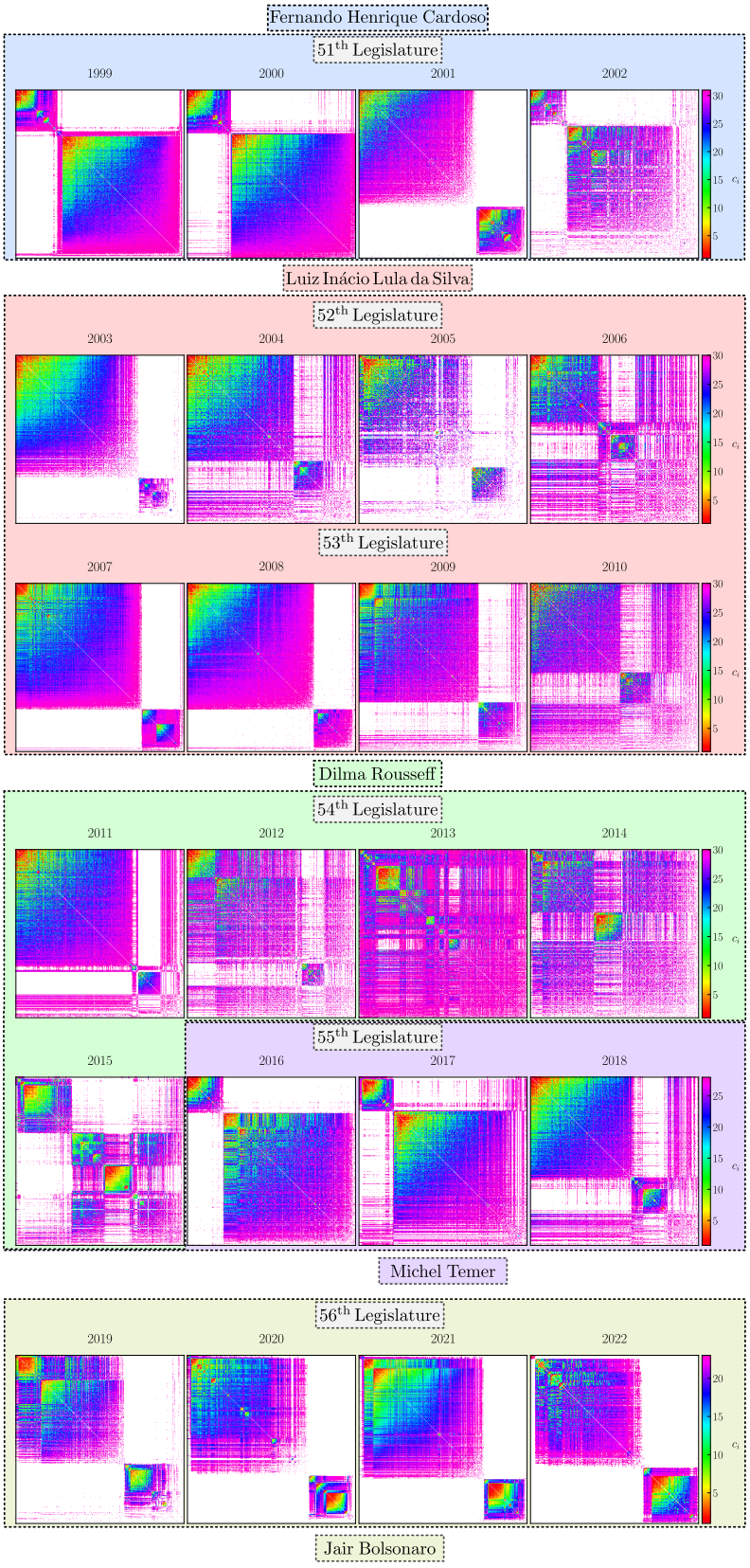
Source: Author.

Figure 49 – Ordered correlation matrices colored by learned clusters for the House of Commons of the United Kingdom. Using procedure 6 it is possible to detect clusters based on the values of the correlation distance coefficient. For each cluster learned a color was assigned. Possibilities for color are shown in the colorbar located at the righest of Figure. c_i label in the colorbar is intended for i-detected cluster.



Source: Author.

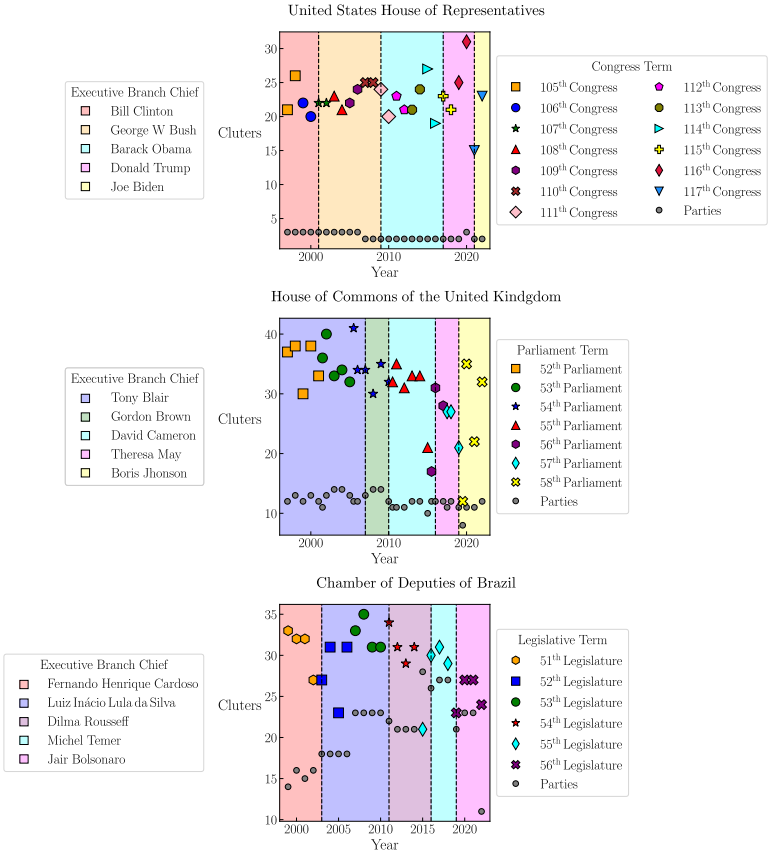
Figure 50 – Ordered correlation matrices colored by learned clusters for the Chamber of Deputies of Brazil. Using procedure 6 it is possible to detect clusters based on the values of the correlation distance coefficient. For each cluster learned a color was assigned. Possibilities for color are shown in the colorbar located at the righest of Figure. c_i label in the colorbar is intended for i-detected cluster.



Source: Author.

As shown in Figures, procedure 6 is able to detect clusters only by having the information of the correlation distance matrix. It is worth commenting that this technique of cluster detection does not have any prior knowledge in order to learn and classify each cluster. The procedure is based on the computation of first and second giant components when a complete connected network is diluting, that is, when edges are being removed from the network. This argument comes to the play in the sense that detected clusters are unbiased to any political information. In fact, the cluster detection feature is capable of recognizing inner undetected political groups in Lower Houses. Note for example in 2015-2, 2019-2 and 2021 parliament years for the House of Commons of the United kingdom, the clusters detected exhibit a different pattern. These years are precisely connected with the pre and post Brexit period. Figure 51 shows the quantity of clusters detected in each year by each Lower House. Note that actual number of political parties was set up in each panel of Figure for comparison. A profound study of these clusters could bring new insights about partisanship and legislative workflow.

Figure 51 – Cluster learned by NECO algorithm. Procedure 6 was applied over the roll-call vote matrix for the three Lower Houses in order to detect inner undetected clusters. Upper panel is for the United States House of Representatives, the central panel is for the House of Commons of the United Kingdom and the lower panel is for the Chamber of Deputies of Brazil. A label for legislative terms (right legend of each panel) was imposed and colored boxes represent Executive terms (left legend of each panel). Point label in the right legend is intended for recognizing the actual number of political parties for each Lower House.

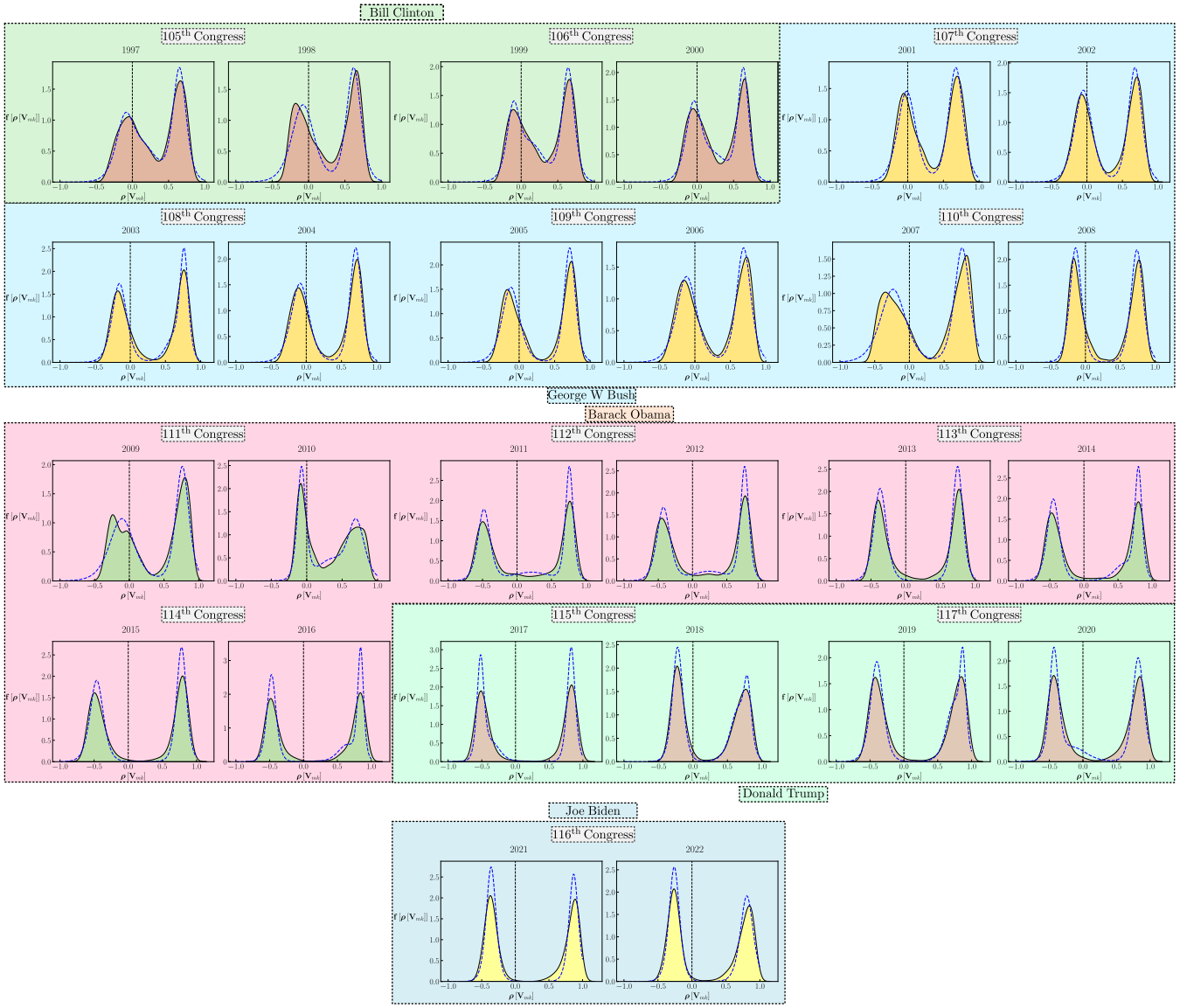


Source: Author.

5.7 Lower Houses members correlation distributions

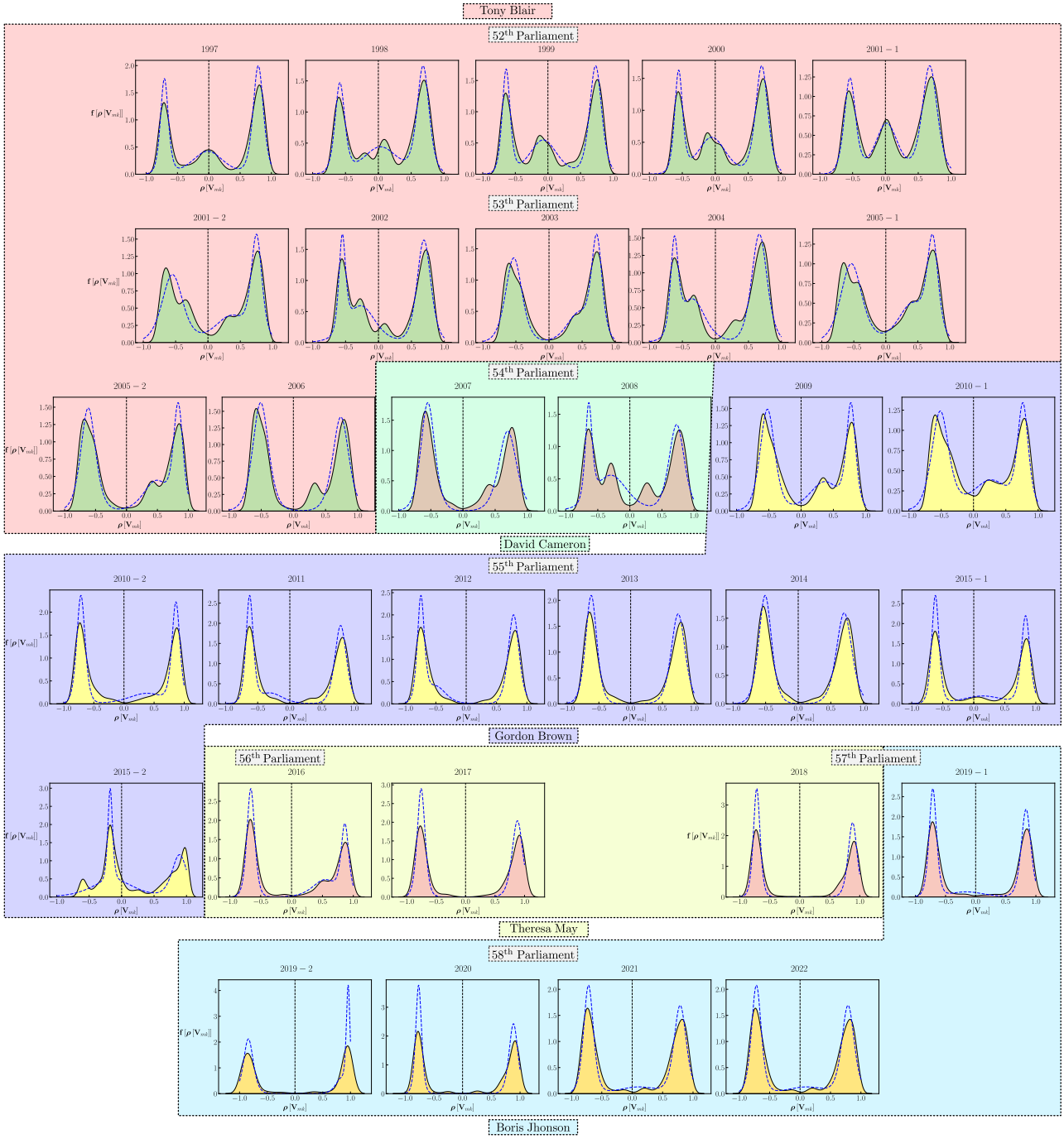
By observing the behavior of ordered correlation matrices (see Figures 42, 43 and 44) based on NECO algorithm, one can realize that correlation coefficients must possess a defined functional form. So, probability distribution function of the correlation matrix $\mathbf{f}\left[\boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]\right]$ was computed for each year in each Lower Houses. Figures 52, 53 and 54 show the distributions obtained. Executive and Legislative terms were colored labeled in each Figure.

Figure 52 – Probability distribution function for correlation matrices for the United States House of Representatives. Executive and Legislative terms were colored labeled. Blue dashed line representing the best fit function for each year. A well defined bimodal state is observed all years. Correlation values are distributed both in positive and negative axis, meaning that a coexistence of dissensual and consensual state is present in this Lower House.



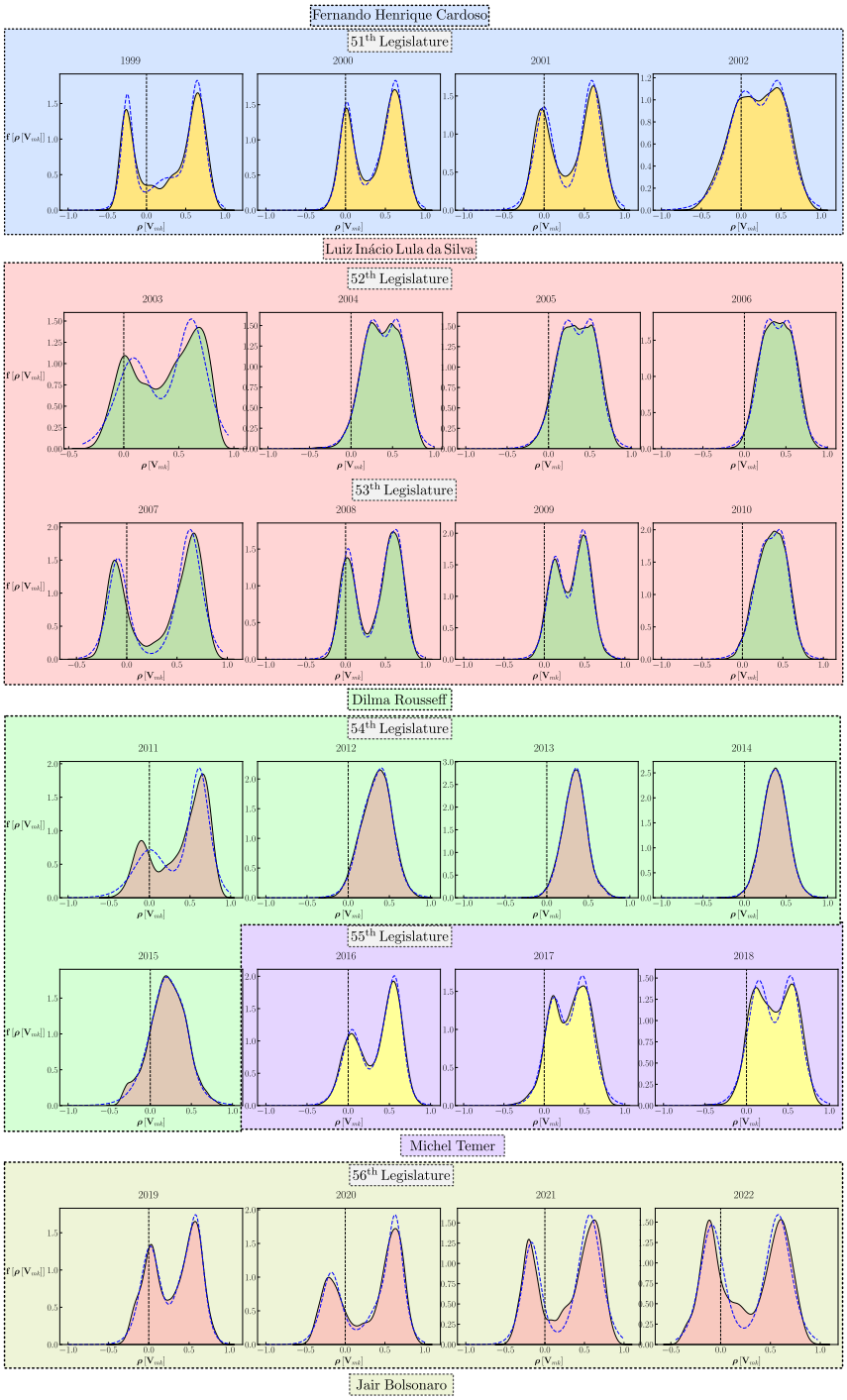
Source: Author.

Figure 53 – Probability distribution function for correlation matrices for the House of Commons of the United Kingdom. Executive and Legislative terms were colored labeled. Blue dashed line representing the best fit function for each year. A trimodal state is observed in some years. Trimodal state is related to the coexistence of dissensual, consensual and random political states.



Source: Author.

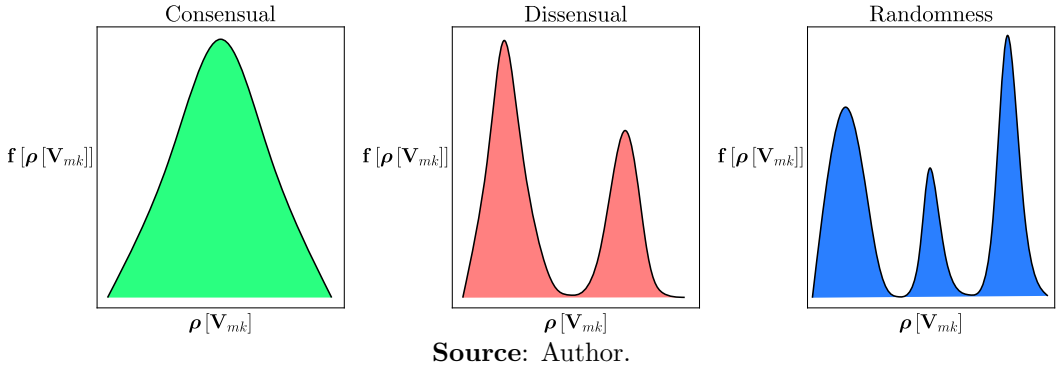
Figure 54 – Probability distribution function for correlation matrices for the Chamber of Deputies of Brazil. Executive and Legislative terms were colored labeled. Blue dashed line representing the best fit function for each year. A transition from bimodal to unimodal states is recognized in each Legislature.



Source: Author.

Presence of modes in the probability distribution function for correlation values can be associated with the existence of three possible main political states. Name these political states as **Consensual**, **Dissensual** and **Political randomness**. Dissensual political state means that correlation values are distributed over positive and negative axis and therefore, probability distribution arise as bimodal-like distribution. Dissensual state means that a portion of Lower House Members are agreeing among themselves, but they are in disagreement with the remaining portion. Consensual, the contrary political state, is associated with the fact that correlation values are distributed either over positive or negative axis, a unimodal-like distribution. Consensual state means that all Lower House members are in agreement either to approve or disprove. Finally, Political randomness is the state in which correlation values are distributed all around axis, including the particular case of $\boldsymbol{\rho} = 0$, implying in a trimodal-like distribution. Political randomness is likely to dissensual state, except for the presence of a portion of Lower Houses members apparently voting freely. Political randomness could mean either that Lower Houses Members are voting without taking into account political party orientation or that bills discussed need a lot of amendments, meaning that some parts are being approved and other ones are rejected. Figure 55 shows an example of the three possible political states along a skecth of how the distribution look like.

Figure 55 – Political states for correlation probability distributions. According to modes of probability distribution, three political states can be defined: Consensual, dissensual and political randomness. Left central and right panels show how correlation probability distributions must look like taking into account each possible political state.

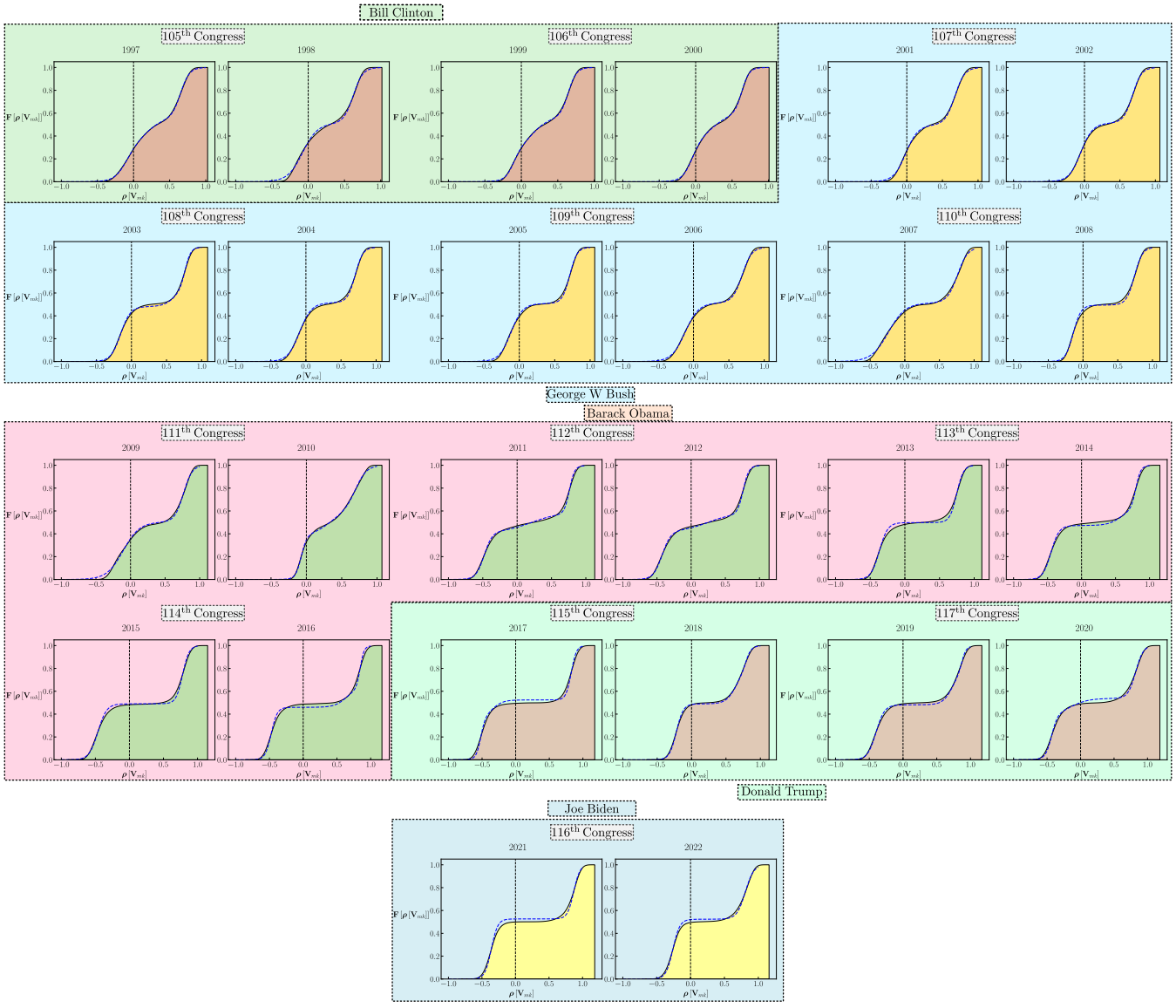


Source: Author.

By inspecting Figures 52, 53 and 54, the probability distribution function for correlation values for the United States Houses of Representatives possesses a bimodal mode meaning in dissensual state. For the case of the House of Commons of the United Kingdom the dissensual state is present except for some years in which it appears to

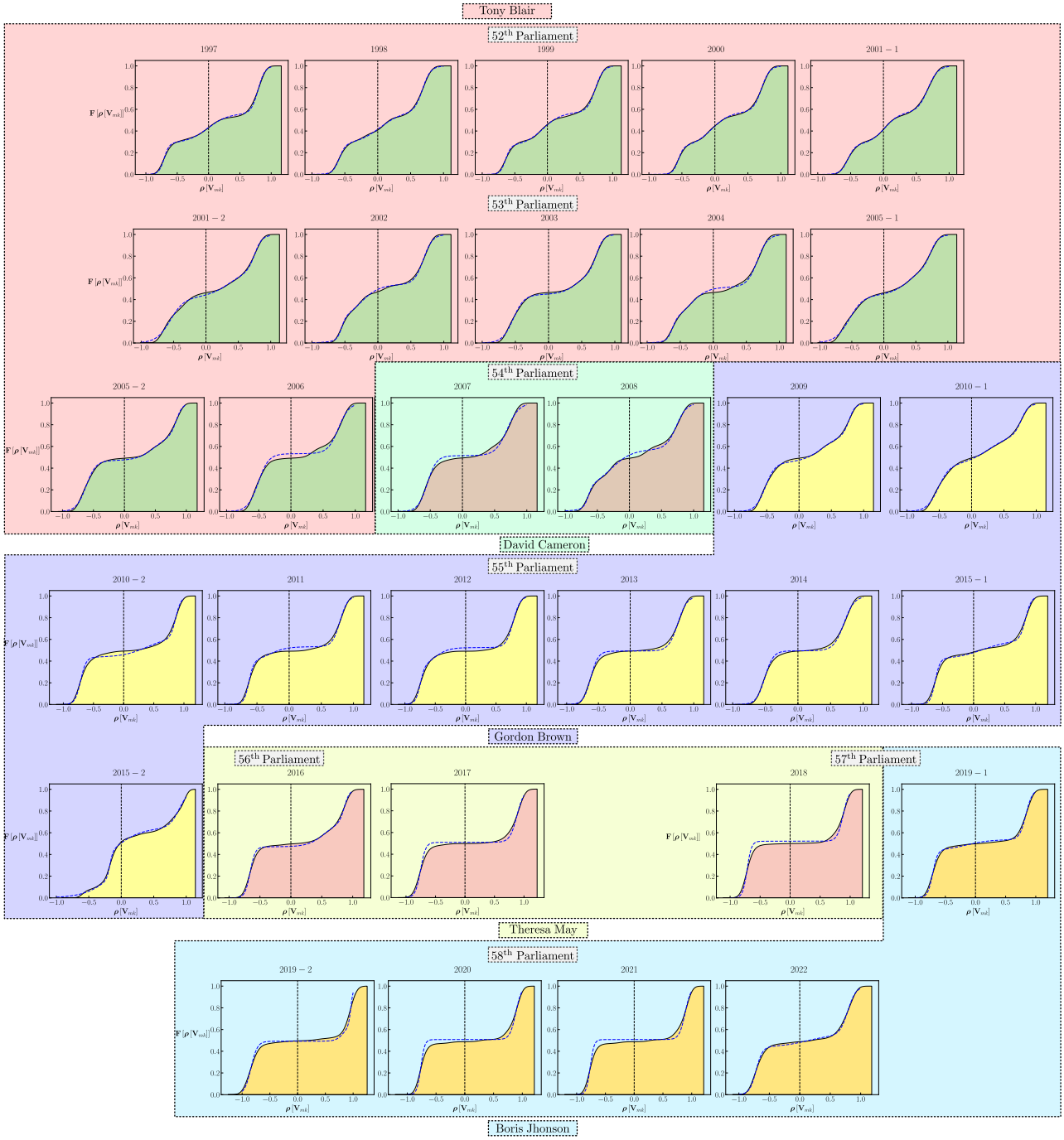
have a trimodal mode. Take attention to distributions for 1998, 1999, 2000, 2000-1 and 2015-2. In these years, political randomness is present. Finally, the probability distribution function for the Chamber of Deputies of Brazil shows the transition from bimodal mode (dissensual state) to unimodal. Note that the final state at the end of Legislatures would be either consensual or political randomness. The functional form of these distributions can be acquired by finding the best function that describes all Lower Houses distributions. Blue dashed lines in distributions show the result of a nonlinear least squares regression technique used for finding a unique function for all three Lower Houses. The regression process was implemented over the cumulative distribution functions of the correlation values, mainly because $\mathbf{F}\left[\boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]\right]$ is independent of bin range. Cumulative distribution for each year was obtained after sorting correlation values from the smallest to the largest in a Zipf Plot [28]. Figures 56, 57 and 58 show the cumulative distribution functions along its best fit regression curve for the three Lower Houses.

Figure 56 – Cumulative distribution function for correlation matrices for the United States House of Representatives. Executive and Legislative terms were colored labeled. Blue dashed line represents the best fit function for each year.



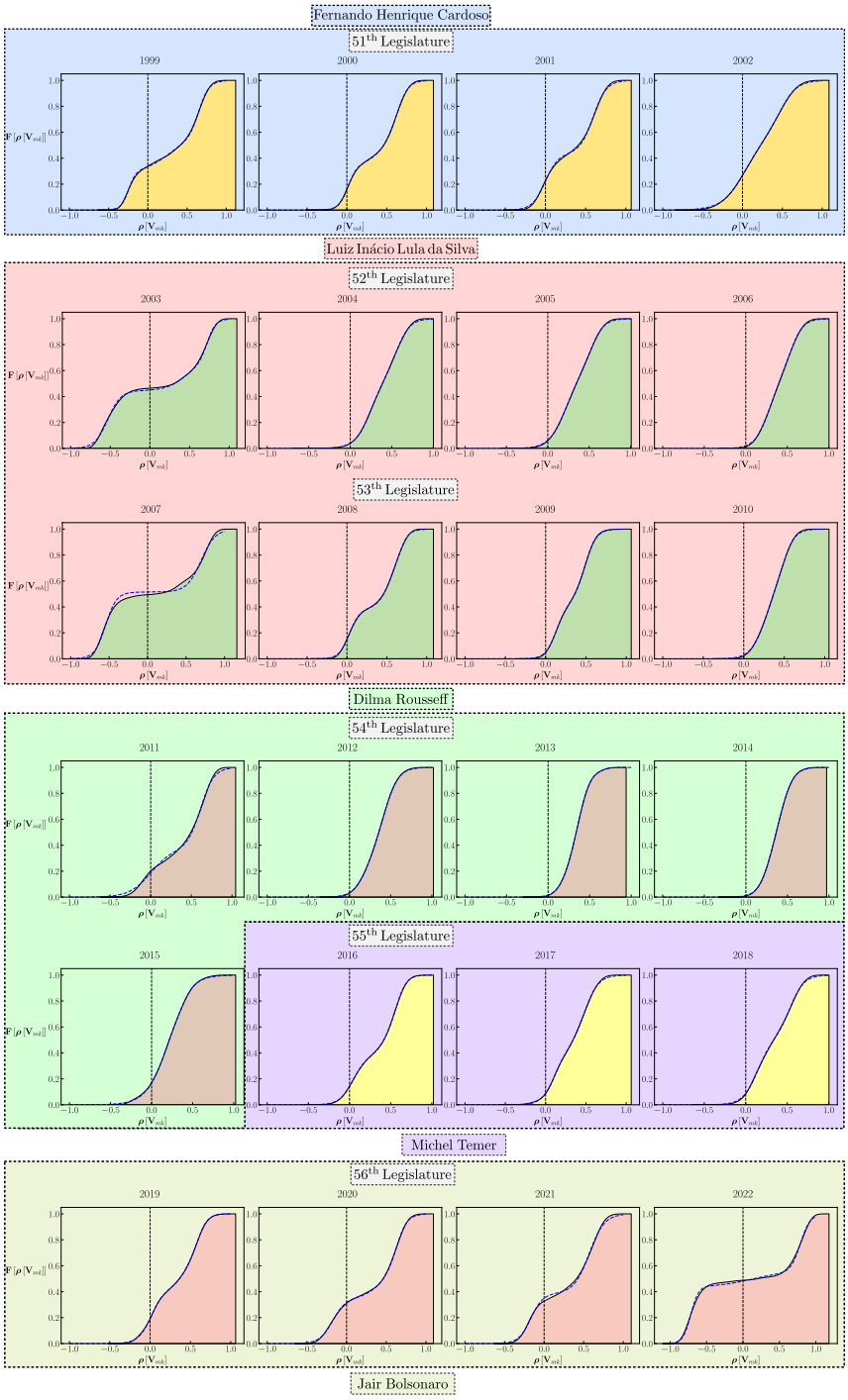
Source: Author.

Figure 57 – Cumulative distribution function for correlation matrices for the House of Commons of the United Kingdom. Executive and Legislative terms were colored labeled. Blue dashed line representing the best fit function for each year.



Source: Author.

Figure 58 – Cumulative distribution function for correlation matrices for the Chamber of Deputies of Brazil. Executive and Legislative terms were colored labeled. Blue dashed line representing the best fit function for each year.



Source: Author.

The best fit function was found by minimizing the sum of residual squares between cumulative distribution of correlation matrices and target function, constrained to the function must have basic properties of a cumulative distribution, that is $\mathbf{F}\left[\boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]\right] \geq 0 \; \forall \, \boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]$. Function describing the cumulative distribution for Lower Houses is

$$\begin{aligned}\mathbf{F}\left[\boldsymbol{\rho}\left[\mathbf{V}_{mk}\right]\right] = & \frac{c_1}{1+\exp\left(\frac{-(\boldsymbol{\rho}[\mathbf{V}_{mk}]-\mu_1)}{s_1}\right)} + \frac{c_2}{1+\exp\left(\frac{-(\boldsymbol{\rho}[\mathbf{V}_{mk}]-\mu_2)}{s_2}\right)} \\ & + \frac{c_3}{1+\exp\left(\frac{-(\boldsymbol{\rho}[\mathbf{V}_{mk}]-\mu_3)}{s_3}\right)},\end{aligned} \tag{5.4}$$

where μ_1, μ_2, μ_3 and s_1, s_2, s_3 are the means and scale values. c_1, c_2 and c_3 are constants that must satisfy $c_1 + c_2 + c_3 = 1$. Constrains for mean values are $\mu_i \in \mathbb{R}$ and for scale values are $s_i > 0$. Cumulative distribution function (5.4) is a combination of three logistic cumulative distributions to encapsulate all possible political states, it all depends on the configuration of μ_i and c_i. For example, if $\mu_1 < 0$, $\mu_2 > 0$, $c_2 = 1 - c_1$ and $c_3 = 0$, probability distribution function (5.4) would be associated to dissensual state. In order to find parameters for proposed distribution, a Sequential Least Squares Programming method [29] was implemented with an accuracy fixed to 1×10^{-8}. For the sake of simplicity, for the United States House of Representatives and House of Commons of the United Kingdom a combination of just two logistic distribution was used. Tables 3, 4 and 5 show parameters found for the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively.

Table 3 – Fitted parameters for cumulative distribution function for the United States House of Representatives. A Sequential least squares method was used for finding the best fit parameters. The method is based on the minimization of the least squares between empirical cumulative distribution and the functional value.

United States House of Representatives						
Year	c_1	c_2	μ_1	μ_2	s_1	s_2
1997	$0.47882 \pm 3.7 \times 10^{-4}$	$0.52118 \pm 1.3 \times 10^{-4}$	$0.64067 \pm 9.0 \times 10^{-5}$	$-0.01924 \pm 8.0 \times 10^{-5}$	$0.0673 \pm 8.0 \times 10^{-5}$	$0.11782 \pm 6.0 \times 10^{-5}$
1998	$0.5103 \pm 2.1 \times 10^{-4}$	$0.4897 \pm 9.8 \times 10^{-4}$	$-0.07691 \pm 1.2 \times 10^{-4}$	$0.6207 \pm 2.6 \times 10^{-4}$	$0.102 \pm 9.0 \times 10^{-5}$	$0.06622 \pm 1.7 \times 10^{-4}$
1999	$0.51857 \pm 1.9 \times 10^{-4}$	$0.48143 \pm 7.2 \times 10^{-4}$	$-0.02222 \pm 1.1 \times 10^{-4}$	$0.63015 \pm 1.8 \times 10^{-4}$	$0.10312 \pm 8.0 \times 10^{-5}$	$0.06411 \pm 1.4 \times 10^{-4}$
2000	$0.46727 \pm 5.3 \times 10^{-5}$	$0.53273 \pm 1.6 \times 10^{-4}$	$0.61838 \pm 1.3 \times 10^{-4}$	$0.00343 \pm 8.0 \times 10^{-5}$	$0.05976 \pm 1.0 \times 10^{-4}$	$0.09537 \pm 6.0 \times 10^{-5}$
2001	$0.5044 \pm 3.6 \times 10^{-4}$	$0.4956 \pm 1.1 \times 10^{-4}$	$0.67368 \pm 1.0 \times 10^{-4}$	$-0.01421 \pm 6.0 \times 10^{-5}$	$0.0687 \pm 8.0 \times 10^{-5}$	$0.08499 \pm 5.0 \times 10^{-5}$
2002	$0.4862 \pm 2.2 \times 10^{-4}$	$0.5138 \pm 6.0 \times 10^{-5}$	$0.67736 \pm 6.0 \times 10^{-5}$	$-0.05042 \pm 3.0 \times 10^{-5}$	$0.06325 \pm 4.0 \times 10^{-4}$	$0.0833 \pm 3.0 \times 10^{-5}$
2003	$0.49311 \pm 4.6 \times 10^{-4}$	$0.50689 \pm 1.0 \times 10^{-4}$	$0.74235 \pm 1.1 \times 10^{-4}$	$-0.14415 \pm 5.0 \times 10^{-5}$	$0.05304 \pm 7.0 \times 10^{-5}$	$0.07611 \pm 4.0 \times 10^{-5}$
2004	$0.48456 \pm 3.2 \times 10^{-4}$	$0.51544 \pm 9.0 \times 10^{-5}$	$0.70223 \pm 7.0 \times 10^{-5}$	$-0.08781 \pm 4.0 \times 10^{-5}$	$0.05472 \pm 5.0 \times 10^{-5}$	$0.08426 \pm 4.0 \times 10^{-5}$
2005	$0.49189 \pm 5.1 \times 10^{-4}$	$0.50811 \pm 1.2 \times 10^{-4}$	$0.71176 \pm 1.1 \times 10^{-4}$	$-0.11725 \pm 6.0 \times 10^{-5}$	$0.05234 \pm 8.0 \times 10^{-5}$	$0.08239 \pm 5.0 \times 10^{-5}$
2006	$0.51331 \pm 9.0 \times 10^{-5}$	$0.48669 \pm 4.2 \times 10^{-4}$	$-0.11731 \pm 5.0 \times 10^{-5}$	$0.68761 \pm 1.2 \times 10^{-4}$	$0.09489 \pm 4.0 \times 10^{-5}$	$0.0675 \pm 8.0 \times 10^{-5}$
2007	$0.5135 \pm 1.6 \times 10^{-4}$	$0.4865 \pm 2.2 \times 10^{-3}$	$-0.23594 \pm 1.1 \times 10^{-4}$	$0.75961 \pm 6.0 \times 10^{-4}$	$0.12107 \pm 8.0 \times 10^{-5}$	$0.07325 \pm 2.2 \times 10^{-4}$
2008	$0.49373 \pm 1.3 \times 10^{-4}$	$0.50627 \pm 7.4 \times 10^{-4}$	$-0.14072 \pm 5.0 \times 10^{-5}$	$0.73955 \pm 1.8 \times 10^{-4}$	$0.05545 \pm 4.8 \times 10^{-5}$	$0.05786 \pm 1.1 \times 10^{-4}$
2009	$0.50394 \pm 1.8 \times 10^{-4}$	$0.49606 \pm 1.0 \times 10^{-3}$	$-0.1063 \pm 1.3 \times 10^{-4}$	$0.7643 \pm 2.7 \times 10^{-4}$	$0.11758 \pm 9.0 \times 10^{-5}$	$0.06302 \pm 1.5 \times 10^{-4}$
2010	$0.44573 \pm 2.1 \times 10^{-4}$	$0.55427 \pm 1.4 \times 10^{-3}$	$-0.0596 \pm 3.0 \times 10^{-5}$	$0.63123 \pm 5.1 \times 10^{-4}$	$0.04937 \pm 4.0 \times 10^{-5}$	$0.11169 \pm 2.9 \times 10^{-4}$
2011	$0.49946 \pm 1.8 \times 10^{-3}$	$0.50054 \pm 1.7 \times 10^{-4}$	$0.7445 \pm 3.9 \times 10^{-4}$	$-0.44905 \pm 1.0 \times 10^{-4}$	$0.0497 \pm 1.8 \times 10^{-4}$	$0.0799 \pm 9.0 \times 10^{-5}$
2012	$0.49924 \pm 1.3 \times 10^{-3}$	$0.50076 \pm 1.6 \times 10^{-4}$	$0.74417 \pm 3.1 \times 10^{-4}$	$-0.39388 \pm 9.0 \times 10^{-5}$	$0.05354 \pm 1.6 \times 10^{-4}$	$0.08378 \pm 1.0 \times 10^{-5}$
2013	$0.50158 \pm 9.8 \times 10^{-4}$	$0.49842 \pm 1.5 \times 10^{-4}$	$0.74735 \pm 2.1 \times 10^{-4}$	$-0.36824 \pm 6.0 \times 10^{-5}$	$0.04898 \pm 1.2 \times 10^{-4}$	$0.06028 \pm 6.0 \times 10^{-5}$
2014	$0.49211 \pm 1.4 \times 10^{-3}$	$0.50789 \pm 1.5 \times 10^{-4}$	$0.77679 \pm 3.1 \times 10^{-4}$	$-0.4376 \pm 7.0 \times 10^{-5}$	$0.05073 \pm 1.4 \times 10^{-4}$	$0.06752 \pm 6.0 \times 10^{-5}$
2015	$0.50951 \pm 3.3 \times 10^{-4}$	$0.49049 \pm 4.2 \times 10^{-3}$	$0.78081 \pm 7.2 \times 10^{-5}$	$-0.46252 \pm 5.3 \times 10^{-4}$	$0.04749 \pm 5.0 \times 10^{-5}$	$0.06439 \pm 8.2 \times 10^{-5}$
2016	$0.49682 \pm 2.9 \times 10^{-3}$	$0.50318 \pm 1.5 \times 10^{-4}$	$0.82239 \pm 4.9 \times 10^{-4}$	$-0.46121 \pm 5.0 \times 10^{-5}$	$0.04366 \pm 1.7 \times 10^{-4}$	$0.05277 \pm 5.0 \times 10^{-5}$
2017	$0.50374 \pm 3.0 \times 10^{-4}$	$0.49626 \pm 2.0 \times 10^{-4}$	$0.81961 \pm 1.5 \times 10^{-5}$	$-0.49579 \pm 7.9 \times 10^{-5}$	$0.04293 \pm 4.8 \times 10^{-5}$	$0.0489 \pm 1.2 \times 10^{-5}$
2018	$0.49785 \pm 3.3 \times 10^{-4}$	$0.50215 \pm 6.0 \times 10^{-5}$	$0.71895 \pm 1.0 \times 10^{-4}$	$-0.21795 \pm 2.0 \times 10^{-5}$	$0.07306 \pm 6.0 \times 10^{-5}$	$0.0522 \pm 2.0 \times 10^{-5}$
2019	$0.50115 \pm 1.5 \times 10^{-3}$	$0.49885 \pm 1.1 \times 10^{-4}$	$0.80731 \pm 4.0 \times 10^{-4}$	$-0.39392 \pm 5.0 \times 10^{-5}$	$0.06678 \pm 1.4 \times 10^{-4}$	$0.06671 \pm 4.0 \times 10^{-5}$
2020	$0.49918 \pm 2.7 \times 10^{-3}$	$0.50082 \pm 1.3 \times 10^{-4}$	$0.80943 \pm 5.1 \times 10^{-4}$	$-0.40757 \pm 5.0 \times 10^{-5}$	$0.06498 \pm 1.8 \times 10^{-4}$	$0.06264 \pm 5.0 \times 10^{-5}$
2021	$0.47394 \pm 2.6 \times 10^{-3}$	$0.52606 \pm 1.0 \times 10^{-4}$	$0.86806 \pm 4.1 \times 10^{-4}$	$-0.36004 \pm 3.0 \times 10^{-5}$	$0.04613 \pm 1.1 \times 10^{-4}$	$0.04798 \pm 2.0 \times 10^{-5}$
2022	$0.52365 \pm 6.0 \times 10^{-5}$	$0.47635 \pm 5.2 \times 10^{-4}$	$-0.2598 \pm 2.0 \times 10^{-5}$	$0.80808 \pm 1.3 \times 10^{-4}$	$0.05105 \pm 1.0 \times 10^{-5}$	$0.06217 \pm 6.0 \times 10^{-5}$

Source: Author.

Table 4 – Fitted parameters for cumulative distribution function for the House of Commons of the United Kingdom. A Sequential least squares method was used for finding the best fit parameters. The method is based on the minimization of the least squares between empirical cumulative distribution and the functional value.

House of Commons of the United Kingdom										
Year	c_1	c_2	c_3	μ_1	μ_2	μ_3	s_1	s_2	s_3
1997	$0.43662 \pm 4.7 \times 10^{-3}$	$0.27707 \pm 4.1 \times 10^{-3}$	$0.2863 \pm 1.8 \times 10^{-4}$	$0.78999 \pm 1.4 \times 10^{-3}$	$-0.01575 \pm 8.3 \times 10^{-3}$	$-0.70895 \pm 1.7 \times 10^{-4}$	$0.055 \pm 4.8 \times 10^{-4}$	$0.16352 \pm 7.9 \times 10^{-3}$	$0.04122 \pm 1.8 \times 10^{-4}$
1998	$0.27623 \pm 2.4 \times 10^{-4}$	$0.41431 \pm 8.0 \times 10^{-5}$	$0.30946 \pm 1.2 \times 10^{-4}$	$-0.58532 \pm 6.0 \times 10^{-5}$	$0.68155 \pm 3.0 \times 10^{-5}$	$0.03349 \pm 9.0 \times 10^{-5}$	$0.04853 \pm 5.0 \times 10^{-5}$	$0.06089 \pm 2.0 \times 10^{-5}$	$0.17551 \pm 1.0 \times 10^{-4}$
1999	$0.4261 \pm 2.4 \times 10^{-4}$	$0.26455 \pm 8.0 \times 10^{-5}$	$0.30935 \pm 1.2 \times 10^{-4}$	$0.73242 \pm 6.0 \times 10^{-5}$	$-0.64087 \pm 3.0 \times 10^{-5}$	$-0.0708 \pm 9.0 \times 10^{-5}$	$0.06165 \pm 5.0 \times 10^{-5}$	$0.04017 \pm 2.0 \times 10^{-5}$	$0.1408 \pm 1.0 \times 10^{-4}$
2000	$0.24957 \pm 9.0 \times 10^{-5}$	$0.41949 \pm 1.7 \times 10^{-4}$	$0.33094 \pm 1.3 \times 10^{-4}$	$-0.55556 \pm 3.0 \times 10^{-5}$	$0.70517 \pm 4.0 \times 10^{-5}$	$-0.05287 \pm 9.0 \times 10^{-5}$	$0.03986 \pm 2.0 \times 10^{-5}$	$0.0622 \pm 4.0 \times 10^{-5}$	$0.14303 \pm 9.0 \times 10^{-5}$
2001-1	$0.42341 \pm 3.5 \times 10^{-3}$	$0.30026 \pm 6.3 \times 10^{-3}$	$0.27633 \pm 7.2 \times 10^{-4}$	$0.6753 \pm 3.2 \times 10^{-3}$	$0.01746 \pm 9.7 \times 10^{-2}$	$-0.54103 \pm 7.0 \times 10^{-3}$	$0.07639 \pm 4.6 \times 10^{-3}$	$0.11438 \pm 3.5 \times 10^{-3}$	$0.0567 \pm 1.1 \times 10^{-4}$
2001-2	$0.4174 \pm 2.2 \times 10^{-4}$	$0.31295 \pm 3.4 \times 10^{-3}$	$0.26964 \pm 1.5 \times 10^{-3}$	$-0.5505 \pm 1.7 \times 10^{-4}$	$0.75208 \pm 1.2 \times 10^{-3}$	$0.36826 \pm 3.1 \times 10^{-3}$	$0.1066 \pm 1.2 \times 10^{-4}$	$0.05459 \pm 4.3 \times 10^{-4}$	$0.17156 \pm 3.4 \times 10^{-3}$
2002	$0.34178 \pm 2.8 \times 10^{-3}$	$0.45957 \pm 2.1 \times 10^{-3}$	$0.19865 \pm 2.1 \times 10^{-3}$	$-0.2675 \pm 4.5 \times 10^{-4}$	$0.68977 \pm 6.8 \times 10^{-3}$	$-0.55019 \pm 7.6 \times 10^{-3}$	$0.14401 \pm 8.3 \times 10^{-3}$	$0.06972 \pm 2.5 \times 10^{-3}$	$0.03339 \pm 1.5 \times 10^{-3}$
2003	$0.44667 \pm 1.7 \times 10^{-4}$	$0.33352 \pm 5.4 \times 10^{-3}$	$0.21981 \pm 8.3 \times 10^{-3}$	$-0.53146 \pm 7.0 \times 10^{-5}$	$0.72087 \pm 2.1 \times 10^{-4}$	$0.44493 \pm 7.5 \times 10^{-3}$	$0.08249 \pm 5.0 \times 10^{-5}$	$0.05274 \pm 4.8 \times 10^{-4}$	$0.11859 \pm 1.6 \times 10^{-3}$
2004	$0.19593 \pm 1.6 \times 10^{-3}$	$0.48429 \pm 3.6 \times 10^{-3}$	$0.31978 \pm 3.7 \times 10^{-3}$	$-0.6206 \pm 6.3 \times 10^{-3}$	$0.66487 \pm 2.2 \times 10^{-3}$	$-0.34983 \pm 4.4 \times 10^{-3}$	$0.03804 \pm 1.0 \times 10^{-3}$	$0.078 \pm 3.3 \times 10^{-3}$	$0.12794 \pm 3.4 \times 10^{-3}$
2005-1	$0.44106 \pm 3.4 \times 10^{-3}$	$0.27283 \pm 1.5 \times 10^{-3}$	$0.28611 \pm 1.6 \times 10^{-3}$	$-0.5387 \pm 4.2 \times 10^{-3}$	$0.73244 \pm 3.0 \times 10^{-3}$	$0.41171 \pm 1.3 \times 10^{-4}$	$0.11036 \pm 3.3 \times 10^{-3}$	$0.057 \pm 5.5 \times 10^{-3}$	$0.14566 \pm 4.0 \times 10^{-4}$
2005-2	$0.47002 \pm 1.4 \times 10^{-4}$	$0.23855 \pm 3.4 \times 10^{-3}$	$0.29143 \pm 1.9 \times 10^{-3}$	$-0.61863 \pm 7.0 \times 10^{-5}$	$0.48265 \pm 1.1 \times 10^{-3}$	$0.83682 \pm 3.8 \times 10^{-3}$	$0.07895 \pm 6.0 \times 10^{-5}$	$0.13554 \pm 3.5 \times 10^{-4}$	$0.05012 \pm 5.8 \times 10^{-3}$
2006	$0.5333 \pm 1.6 \times 10^{-4}$	$0.4667 \pm 3.5 \times 10^{-3}$	$--$	$-0.50852 \pm 7.0 \times 10^{-5}$	$0.7377 \pm 1.1 \times 10^{-3}$	$--$	$0.08176 \pm 5.0 \times 10^{-5}$	$0.08253 \pm 3.5 \times 10^{-4}$	$--$
2007	$0.51494 \pm 1.2 \times 10^{-4}$	$0.48506 \pm 2.4 \times 10^{-3}$	$--$	$-0.54657 \pm 5.0 \times 10^{-5}$	$0.69471 \pm 8.3 \times 10^{-4}$	$--$	$0.0719 \pm 4.0 \times 10^{-5}$	$0.09186 \pm 2.6 \times 10^{-4}$	$--$
2008	$0.37321 \pm 5.8 \times 10^{-4}$	$0.20158 \pm 4.9 \times 10^{-4}$	$0.42521 \pm 1.5 \times 10^{-3}$	$-0.29642 \pm 4.0 \times 10^{-4}$	$-0.64077 \pm 9.0 \times 10^{-5}$	$0.7302 \pm 5.7 \times 10^{-4}$	$0.16786 \pm 3.4 \times 10^{-4}$	$0.03464 \pm 9.0 \times 10^{-5}$	$0.07959 \pm 2.6 \times 10^{-5}$
2009	$0.44992 \pm 3.3 \times 10^{-4}$	$0.27782 \pm 1.5 \times 10^{-3}$	$0.27226 \pm 2.2 \times 10^{-3}$	$-0.52151 \pm 7.0 \times 10^{-5}$	$0.77314 \pm 1.4 \times 10^{-4}$	$0.3682 \pm 2.3 \times 10^{-3}$	$0.07586 \pm 7.0 \times 10^{-5}$	$0.04725 \pm 2.2 \times 10^{-4}$	$0.15826 \pm 1.4 \times 10^{-3}$
2010-1	$0.27428 \pm 1.3 \times 10^{-3}$	$0.4255 \pm 7.1 \times 10^{-4}$	$0.30021 \pm 2.0 \times 10^{-3}$	$0.76532 \pm 1.7 \times 10^{-4}$	$-0.51509 \pm 9.0 \times 10^{-5}$	$0.27252 \pm 1.3 \times 10^{-3}$	$0.05541 \pm 2.4 \times 10^{-4}$	$0.08766 \pm 1.2 \times 10^{-4}$	$0.19475 \pm 1.7 \times 10^{-3}$
2010-2	$0.4355 \pm 1.1 \times 10^{-3}$	$0.37847 \pm 5.5 \times 10^{-3}$	$0.18602 \pm 5.4 \times 10^{-3}$	$-0.71416 \pm 1.3 \times 10^{-4}$	$0.86289 \pm 1.1 \times 10^{-3}$	$0.38243 \pm 1.2 \times 10^{-3}$	$0.04609 \pm 7.6 \times 10^{-3}$	$0.04395 \pm 2.9 \times 10^{-3}$	$0.20177 \pm 3.7 \times 10^{-3}$
2011	$0.46733 \pm 4.6 \times 10^{-4}$	$0.39294 \pm 8.7 \times 10^{-4}$	$0.13973 \pm 8.9 \times 10^{-4}$	$0.79028 \pm 1.2 \times 10^{-4}$	$-0.62558 \pm 1.2 \times 10^{-3}$	$-0.30793 \pm 7.0 \times 10^{-5}$	$0.06007 \pm 6.0 \times 10^{-5}$	$0.03756 \pm 4.9 \times 10^{-4}$	$0.12923 \pm 6.0 \times 10^{-5}$
2012	$0.34896 \pm 1.3 \times 10^{-3}$	$0.17557 \pm 1.1 \times 10^{-3}$	$0.47547 \pm 4.1 \times 10^{-4}$	$-0.74709 \pm 5.0 \times 10^{-5}$	$-0.50425 \pm 8.9 \times 10^{-4}$	$0.78144 \pm 1.0 \times 10^{-4}$	$0.03799 \pm 8.0 \times 10^{-5}$	$0.11017 \pm 3.0 \times 10^{-4}$	$0.05934 \pm 6.0 \times 10^{-4}$
2013	$0.50744 \pm 8.1 \times 10^{-4}$	$0.49256 \pm 8.0 \times 10^{-5}$	$--$	$0.74836 \pm 2.4 \times 10^{-4}$	$-0.62073 \pm 3.0 \times 10^{-5}$	$--$	$0.07298 \pm 1.1 \times 10^{-4}$	$0.05875 \pm 3.0 \times 10^{-5}$	$--$
2014	$0.50521 \pm 5.5 \times 10^{-4}$	$0.49479 \pm 8.0 \times 10^{-5}$	$--$	$0.70524 \pm 1.9 \times 10^{-4}$	$-0.50707 \pm 3.0 \times 10^{-5}$	$--$	$0.07928 \pm 1.0 \times 10^{-4}$	$0.0651 \pm 3.0 \times 10^{-5}$	$--$
2015-1	$0.17373 \pm 3.0 \times 10^{-3}$	$0.40733 \pm 4.6 \times 10^{-3}$	$0.41894 \pm 1.5 \times 10^{-4}$	$0.14794 \pm 6.9 \times 10^{-3}$	$0.8363 \pm 1.0 \times 10^{-3}$	$-0.6193 \pm 5.0 \times 10^{-5}$	$0.22044 \pm 1.0 \times 10^{-3}$	$0.04702 \pm 3.1 \times 10^{-4}$	$0.03909 \pm 5.0 \times 10^{-5}$
2015-2	$0.35441 \pm 4.4 \times 10^{-3}$	$0.40713 \pm 5.1 \times 10^{-4}$	$0.23846 \pm 1.5 \times 10^{-3}$	$0.89298 \pm 4.3 \times 10^{-4}$	$-0.13498 \pm 6.0 \times 10^{-5}$	$-0.17565 \pm 5.2 \times 10^{-4}$	$0.07683 \pm 4.2 \times 10^{-4}$	$0.21026 \pm 7.0 \times 10^{-5}$	$0.02373 \pm 4.2 \times 10^{-4}$
2016	$0.47042 \pm 1.0 \times 10^{-4}$	$0.28805 \pm 2.4 \times 10^{-3}$	$0.24153 \pm 9.2 \times 10^{-3}$	$-0.65963 \pm 2.0 \times 10^{-5}$	$0.86448 \pm 2.1 \times 10^{-4}$	$0.54406 \pm 7.8 \times 10^{-3}$	$0.04158 \pm 2.0 \times 10^{-5}$	$0.04042 \pm 2.4 \times 10^{-4}$	$0.13327 \pm 1.3 \times 10^{-3}$
2017	$0.50933 \pm 1.2 \times 10^{-4}$	$0.49067 \pm 4.0 \times 10^{-3}$	$-0.75388 \pm 3.0 \times 10^{-5}$	$--$	$0.88008 \pm 7.4 \times 10^{-4}$	$0.04389 \pm 3.0 \times 10^{-5}$	$--$	$0.06014 \pm 2.0 \times 10^{-4}$	$--$
2018	$0.47794 \pm 3.2 \times 10^{-3}$	$0.52206 \pm 1.1 \times 10^{-4}$	$0.88181 \pm 5.5 \times 10^{-4}$	$--$	$-0.71724 \pm 2.0 \times 10^{-5}$	$0.04929 \pm 1.5 \times 10^{-4}$	$--$	$0.03692 \pm 2.0 \times 10^{-5}$	$--$
2019-1	$0.43572 \pm 1.3 \times 10^{-3}$	$0.46293 \pm 3.7 \times 10^{-4}$	$0.10136 \pm 1.3 \times 10^{-3}$	$-0.71107 \pm 5.0 \times 10^{-5}$	$0.83708 \pm 9.0 \times 10^{-5}$	$-0.15633 \pm 1.3 \times 10^{-3}$	$0.04074 \pm 8.0 \times 10^{-5}$	$0.05302 \pm 5.0 \times 10^{-5}$	$0.19569 \pm 4.5 \times 10^{-4}$
2019-2	$0.49228 \pm 2.2 \times 10^{-3}$	$0.34701 \pm 1.1 \times 10^{-3}$	$0.16071 \pm 6.0 \times 10^{-4}$	$-0.84067 \pm 7.0 \times 10^{-4}$	$0.95908 \pm 1.7 \times 10^{-4}$	$0.83402 \pm 9.8 \times 10^{-3}$	$0.058 \pm 6.1 \times 10^{-3}$	$0.02209 \pm 2.2 \times 10^{-4}$	$0.0684 \pm 4.7 \times 10^{-3}$
2020	$0.49214 \pm 4.1 \times 10^{-3}$	$0.50786 \pm 1.3 \times 10^{-4}$	$--$	$0.88565 \pm 7.0 \times 10^{-4}$	$-0.76461 \pm 3.0 \times 10^{-5}$	$--$	$0.05068 \pm 1.9 \times 10^{-4}$	$0.03372 \pm 3.0 \times 10^{-5}$	$--$
2021	$0.45601 \pm 1.9 \times 10^{-4}$	$0.25686 \pm 9.1 \times 10^{-3}$	$0.28713 \pm 1.6 \times 10^{-2}$	$-0.81757 \pm 3.0 \times 10^{-5}$	$0.98728 \pm 3.9 \times 10^{-4}$	$0.8235 \pm 7.2 \times 10^{-3}$	$0.02182 \pm 3.0 \times 10^{-5}$	$0.00773 \pm 1.4 \times 10^{-4}$	$0.08357 \pm 1.5 \times 10^{-3}$
2022	$0.44859 \pm 3.7 \times 10^{-4}$	$0.43892 \pm 1.5 \times 10^{-3}$	$0.11249 \pm 1.5 \times 10^{-3}$	$0.78855 \pm 1.5 \times 10^{-3}$	$-0.71825 \pm 1.2 \times 10^{-4}$	$0.10789 \pm 9.0 \times 10^{-5}$	$0.06706 \pm 7.0 \times 10^{-4}$	$0.05308 \pm 6.0 \times 10^{-5}$	$0.21742 \pm 1.1 \times 10^{-4}$

Source: Author.

Table 5 – Fitted parameters for cumulative distribution function for the Chamber of Deputies of Brazil. A Sequential least squares method was used for finding the best fit parameters. The method is based on the minimization of the least squares between empirical cumulative distribution and the functional value.

Chamber of Deputies of Brazil						
Year	c_1	c_2	μ_1	μ_2	s_1	s_2
1999	$0.40374 \pm 1.9 \times 10^{-4}$	$0.59625 \pm 9.0 \times 10^{-4}$	$-0.20237 \pm 1.0 \times 10^{-4}$	$0.60674 \pm 2.7 \times 10^{-4}$	$0.07926 \pm 9.1 \times 10^{-5}$	$0.09200 \pm 1.9 \times 10^{-4}$
2000	$0.60495 \pm 1.4 \times 10^{-4}$	$0.39504 \pm 6.5 \times 10^{-5}$	$0.60121 \pm 3.3 \times 10^{-5}$	$0.02896 \pm 3.0 \times 10^{-5}$	$0.08306 \pm 3.3 \times 10^{-5}$	$0.06741 \pm 2.4 \times 10^{-5}$
2001	$0.56503 \pm 2.3 \times 10^{-4}$	$0.43496 \pm 1.0 \times 10^{-4}$	$0.59309 \pm 5.6 \times 10^{-5}$	$-0.00310 \pm 5.0 \times 10^{-5}$	$0.08304 \pm 5.8 \times 10^{-5}$	$0.08032 \pm 3.9 \times 10^{-5}$
2002	$0.42223 \pm 2.7 \times 10^{-4}$	$0.57776 \pm 2.2 \times 10^{-4}$	$0.47703 \pm 7.1 \times 10^{-5}$	$0.02627 \pm 9.2 \times 10^{-5}$	$0.10479 \pm 5.7 \times 10^{-5}$	$0.14137 \pm 3.3 \times 10^{-5}$
2003	$0.44638 \pm 3.5 \times 10^{-4}$	$0.55361 \pm 9.3 \times 10^{-4}$	$0.08371 \pm 1.1 \times 10^{-4}$	$0.61728 \pm 1.6 \times 10^{-4}$	$0.10649 \pm 9.1 \times 10^{-5}$	$0.09249 \pm 2.0 \times 10^{-4}$
2004	$0.56346 \pm 2.5 \times 10^{-4}$	$0.43653 \pm 3.0 \times 10^{-4}$	$0.24576 \pm 6.2 \times 10^{-5}$	$0.56489 \pm 6.1 \times 10^{-5}$	$0.09618 \pm 2.3 \times 10^{-5}$	$0.08016 \pm 4.7 \times 10^{-5}$
2005	$0.40984 \pm 4.3 \times 10^{-4}$	$0.59015 \pm 3.7 \times 10^{-4}$	$0.52649 \pm 8.9 \times 10^{-5}$	$0.21476 \pm 9.9 \times 10^{-5}$	$0.07756 \pm 6.5 \times 10^{-5}$	$0.10027 \pm 3.4 \times 10^{-5}$
2006	$0.46285 \pm 7.7 \times 10^{-4}$	$0.53714 \pm 6.8 \times 10^{-4}$	$0.54098 \pm 1.5 \times 10^{-4}$	$0.27296 \pm 1.3 \times 10^{-4}$	$0.077231 \pm 8.8 \times 10^{-5}$	$0.08434 \pm 5.2 \times 10^{-5}$
2007	$0.59970 \pm 1.3 \times 10^{-2}$	$0.40029 \pm 7.0 \times 10^{-3}$	$0.63019 \pm 2.0 \times 10^{-3}$	$-0.08874 \pm 3.8 \times 10^{-2}$	$0.07641 \pm 3.8 \times 10^{-2}$	$0.06558 \pm 4.1 \times 10^{-2}$
2008	$0.60496 \pm 1.1 \times 10^{-4}$	$0.39503 \pm 5.1 \times 10^{-5}$	$0.59107 \pm 2.4 \times 10^{-5}$	$0.03532 \pm 2.3 \times 10^{-5}$	$0.08066 \pm 2.5 \times 10^{-5}$	$0.06696 \pm 1.8 \times 10^{-5}$
2009	$0.57405 \pm 1.2 \times 10^{-4}$	$0.42594 \pm 8.8 \times 10^{-5}$	$0.48456 \pm 2.0 \times 10^{-5}$	$0.13377 \pm 2.8 \times 10^{-5}$	$0.07087 \pm 2.0 \times 10^{-5}$	$0.06756 \pm 1.7 \times 10^{-5}$
2010	$0.43414 \pm 6.5 \times 10^{-4}$	$0.56585 \pm 5.9 \times 10^{-5}$	$0.49372 \pm 1.1 \times 10^{-4}$	$0.25969 \pm 1.3 \times 10^{-4}$	$0.07068 \pm 5.5 \times 10^{-5}$	$0.08987 \pm 4.1 \times 10^{-5}$
2011	$0.37569 \pm 1.8 \times 10^{-4}$	$0.62430 \pm 4.9 \times 10^{-4}$	$0.00289 \pm 1.5 \times 10^{-4}$	$0.61385 \pm 9.4 \times 10^{-5}$	$0.13159 \pm 1.1 \times 10^{-4}$	$0.08169 \pm 9.5 \times 10^{-5}$
2012	$0.62503 \pm 4.1 \times 10^{-4}$	$0.37496 \pm 3.9 \times 10^{-4}$	$0.43027 \pm 6.6 \times 10^{-5}$	$0.20852 \pm 1.2 \times 10^{-4}$	$0.08408 \pm 2.6 \times 10^{-5}$	$0.08675 \pm 4.3 \times 10^{-5}$
2013	$0.16115 \pm 5.1 \times 10^{-4}$	$0.83884 \pm 5.2 \times 10^{-4}$	$0.19339 \pm 1.6 \times 10^{-4}$	$0.36274 \pm 5.8 \times 10^{-5}$	$0.06454 \pm 4.7 \times 10^{-5}$	$0.07758 \pm 1.6 \times 10^{-5}$
2014	$0.49540 \pm 6.5 \times 10^{-4}$	$0.50459 \pm 6.6 \times 10^{-4}$	$0.28440 \pm 1.0 \times 10^{-4}$	$0.44291 \pm 9.1 \times 10^{-5}$	$0.07539 \pm 2.5 \times 10^{-5}$	$0.07344 \pm 2.5 \times 10^{-5}$
2015	$0.08747 \pm 1.7 \times 10^{-4}$	$0.91252 \pm 2.0 \times 10^{-4}$	$0.41516 \pm 7.7 \times 10^{-5}$	$0.19209 \pm 5.5 \times 10^{-5}$	$0.05494 \pm 1.3 \times 10^{-4}$	$0.12932 \pm 1.6 \times 10^{-5}$
2016	$0.59022 \pm 1.4 \times 10^{-4}$	$0.40977 \pm 8.7 \times 10^{-5}$	$0.54125 \pm 2.5 \times 10^{-5}$	$0.05982 \pm 4.7 \times 10^{-5}$	$0.07302 \pm 2.7 \times 10^{-5}$	$0.08946 \pm 3.2 \times 10^{-5}$
2017	$0.57356 \pm 2.5 \times 10^{-4}$	$0.42643 \pm 1.9 \times 10^{-4}$	$0.48073 \pm 4.6 \times 10^{-5}$	$0.11212 \pm 5.6 \times 10^{-5}$	$0.08693 \pm 4.6 \times 10^{-5}$	$0.08049 \pm 3.1 \times 10^{-5}$
2018	$0.48761 \pm 4.6 \times 10^{-4}$	$0.51238 \pm 3.2 \times 10^{-4}$	$0.53753 \pm 8.7 \times 10^{-5}$	$0.14308 \pm 9.2 \times 10^{-5}$	$0.08384 \pm 9.6 \times 10^{-5}$	$0.09020 \pm 4.7 \times 10^{-5}$
2019	$0.54127 \pm 2.1 \times 10^{-4}$	$0.45872 \pm 1.1 \times 10^{-4}$	$0.54849 \pm 4.4 \times 10^{-5}$	$0.03068 \pm 5.5 \times 10^{-5}$	$0.07850 \pm 4.9 \times 10^{-5}$	$0.08933 \pm 3.9 \times 10^{-5}$
2020	$0.60705 \pm 2.4 \times 10^{-4}$	$0.39294 \pm 9.1 \times 10^{-5}$	$0.59969 \pm 5.8 \times 10^{-5}$	$-0.14860 \pm 6.9 \times 10^{-5}$	$0.08179 \pm 5.2 \times 10^{-5}$	$0.09820 \pm 5.7 \times 10^{-5}$
2021	$0.38611 \pm 1.0 \times 10^{-4}$	$0.61388 \pm 3.4 \times 10^{-4}$	$-0.16833 \pm 6.0 \times 10^{-5}$	$0.56899 \pm 9.4 \times 10^{-5}$	$0.07673 \pm 5.1 \times 10^{-5}$	$0.09587 \pm 8.2 \times 10^{-5}$
2022	$0.53797 \pm 5.6 \times 10^{-4}$	$0.46202 \pm 1.9 \times 10^{-4}$	$0.57828397 \pm 1.5 \times 10^{-4}$	$-0.08757 \pm 9.7 \times 10^{-5}$	$0.08461 \pm 1.4 \times 10^{-4}$	$0.07882 \pm 7.7 \times 10^{-5}$

Source: Author.

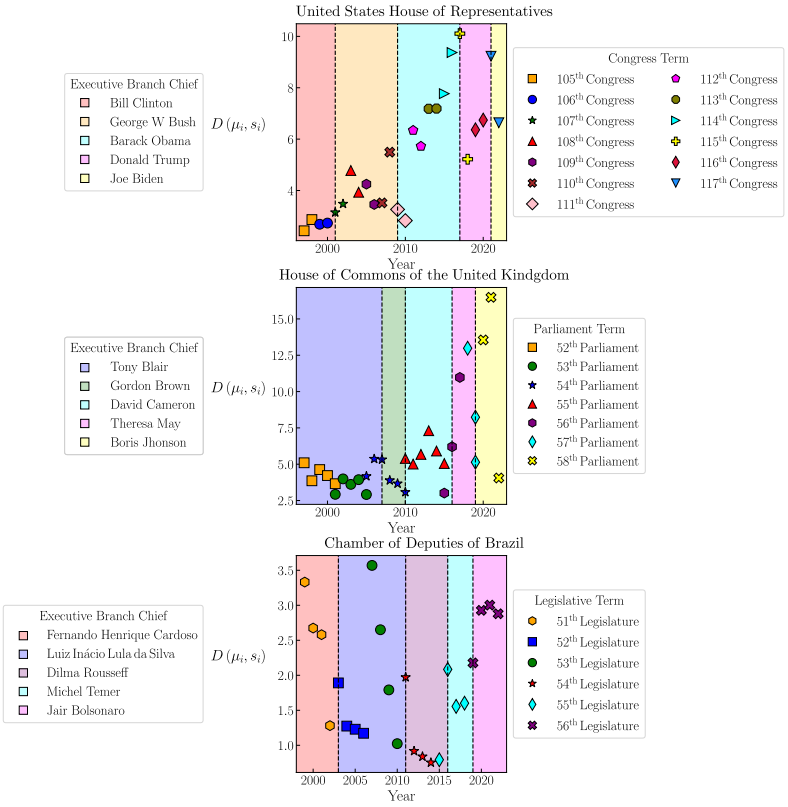
5.8 Bimodality index: Another way to characterize dissensual and consensual states

In the last section, it was mentioned that political states can be recognized by inspecting modes of the correlation probability distribution function. In this sense, fitted parameters for Lower Houses correlation probability distributions can be used to characterize these political states. Classification is done by computing the bimodal index [30]

$$D\left(\mu_i, s_i\right) = \frac{\left|\mu_j - \mu_i\right|}{2\sqrt{s_i^2 + s_j^2}}, \tag{5.5}$$

where μ_i and s_i are the mean and scale values obtained from the non-linear regression. By establishing $D\left(\mu_i, s_i\right)$ as a function of mean and scale values, it is possible to classify Political states. Numerically, if $D\left(\mu_i, s_i\right) > 1$ means that either dissensual or political randomness exists in the probability distribution function for correlation values; contrary, if $D\left(\mu_i, s_i\right) \leq 1$ means that only consensual state is present. Definition of bimodal index is based only on pairs of mean and scales values, for a trimodal-like distribution bimodal index is computed in he following way. First bimodal index for all combinations of pairs (μ_i, s_i) are computed and if resulting values are greater than one, the final value arise as the average, contrary the classification is discarded and probability distribution gets unclassified. This was the case for some years in the House of Commons of the United Kingdom. Figure 59 shows the bimodal index measured in each year for each Lower House under study.

Figure 59 – Bimodal index for Lower Houses. Upper, central and lower panel shows the bimodal index for the United States House of Representatives, The House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively. Legislative terms were marker coloured labeled and each executive term was boxed colored. Given that $D(\mu_i, s_i) > 1$ for all years in the United States House of Representatives and the House of Commons of the United Kingdom, it confirms the coexistence of either consensual or political randoness states. For the Chamber of Deputies of Brazil, $D(\mu_i, s_i) > 1$ at the beginning of Legislatures, as it goes, $D(\mu_i, s_i)$ decreases to become $D(\mu_i, s_i) \leq 1$, meaning that a transition from dissensual to consensual state happened.



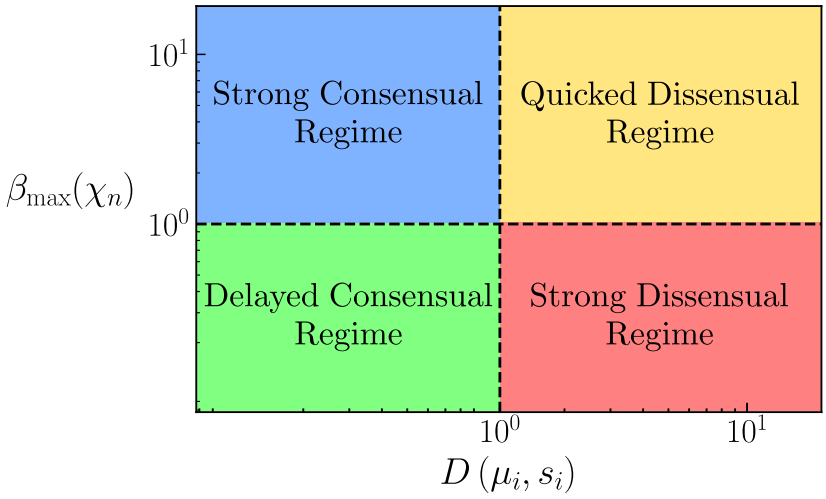
Source: Author.

Upper, central and lower panel of Figure 59 shows the bimodal index for the United States House of Representatives, The House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively. Legislative terms were marker coloured labeled and each executive term was boxed colored. Panel for the United States House of Representatives confirm the existence of dissensual state given that $D(\mu_i, s_i) > 3$. The same argument holds for the House of Commons of the United Kingdom. Note that $D(\mu_i, s_i) > 2.5$, meaning that there is coexistence of either consensual or political randoness states. Finally, the bimodal index for the Chamber of Deputies of Brazil shows the transition from bimodal to unimodal already visualized in Figure 54. Observe that $D(\mu_i, s_i) > 1$ at the beginning of each Legislature, except for 2015, the first year of 54^{th} Legislature, and as the Legislature goes, bimodal index falls to become $D(\mu_i, s_i) \leq 1$, meaning that a transition from dissensual to consensual state happened.

5.9 Legislative consensual-dissensual regimes: Joining interaction and statistical analysis

Last step in order to completely characterize Lower Houses is to join results obtained from the interaction point of view, gathered using the extended Boltzmann machine and the statistical point of view, by means of the Nearest Correlated Cluster Algorithm (NECO). By joining these two approaches, Legislative consensual regimes appear. A legislative consensual regime can be defined as the configuration of parameters $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$ in which a Lower House will be in either consensus or dissensus. Remember that $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)$ estimates the optimal degree of Legislative interaction in which the transition from dissensus to consensus would happened. That is to say, $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)$ measures the capacity for Political parties to maximize Legislative effectiveness by setting an optimal number of discussions. Take attention that $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)$ is extracted from Political parties data. If $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)$ is large it means that a low level of Political interaction among political parties is needed in order to reach consensus. On the other hand, $D\left(\mu_i, s_i\right)$ classifies political states in Lower Houses, just by analyzing the correlation probability distribution among Lower House members. If $D\left(\mu_i, s_i\right) > 1$ means dissensual state, a portion of Lower House Members agreeing among themselves, but in disagreement with the remaining portion. Contrary, if $D\left(\mu_i, s_i\right) \leq 1$ is associated with consensual state, Lower House members are either in agreement or disagrement. Legislative regimes arise then by plotting values $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$. Our proposal for these regimes must look like as shown in Figure 60.

Figure 60 – Legislative consensual-dissensual regimes. Based on the values of $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$, consensual and dissensual regimes can be proposed. Regimes are classified in terms of the political state in which Lower Houses are and the optimal value of Legislative interaction in which transition from dissensus to consensus happened. Four regimes are proposed: Strong - delayed consensual regimes and strong - quicked dissensual regimes. Black dashed lines are used to indicate boundaries for each regime.



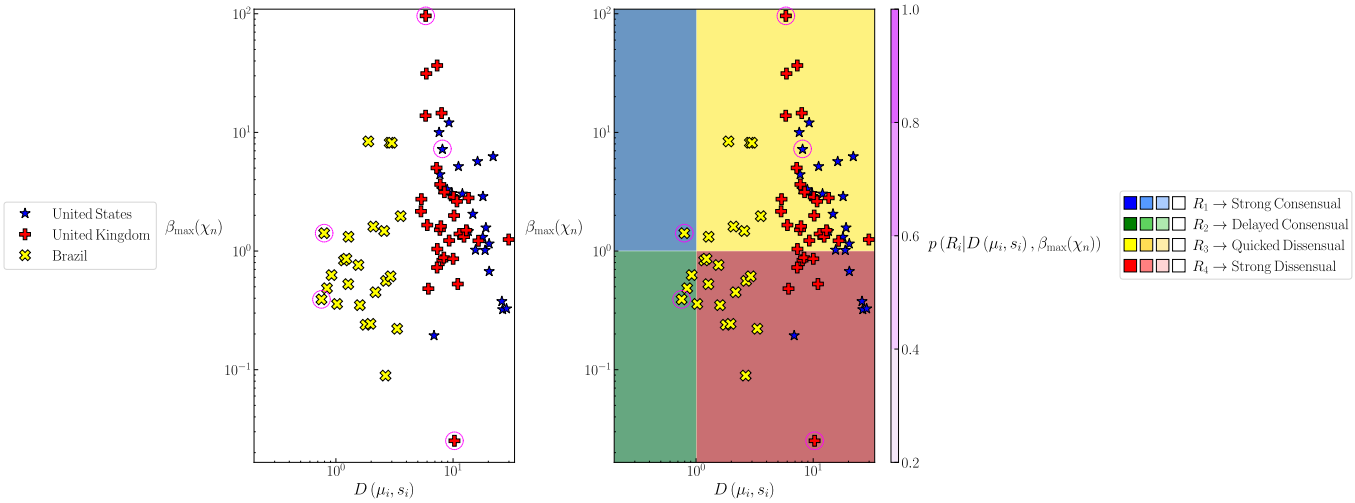
Source: Author.

Vertical and horizontal axis of Figure 60 shows $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)$ and $D\left(\mu_i, s_i\right)$. Vertical and horizontal dashed lines were set up as boundaries for four depicted Legislative regimes: Strong - Delayed Consensual regimes (blue and green zones) and Strong - Quicked dissensual regimes (red and yellow zones).

Strong consensual regime is the zone in which $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right) > 0$ and $D\left(\mu_i, s_i\right) \leq 1$. In this zone are found all Lower Houses in which a low degree of Political interaction is sufficient to set the consensual state. On the counter part, Delayed consensual regime, the region in which $1 \leq \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right) < 0$ and $D\left(\mu_i, s_i\right) \leq 1$, is the zone in which predominates Lower Houses whose all members are agreeing and the degree of Legislative political interaction needed to set up this configuration is large. Delayed means for a big number of discussions. On the other hand, Strong dissensual regime is the zone in which $1 \leq \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right) < 0$ and $D\left(\mu_i, s_i\right) > 1$. This regime is characterized by the fact that despite the large degree of political interaction, political polarization among Lower Houses members is observed. Finally, Quicked dissensual regime, containing values $\beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right) > 1$ and $D\left(\mu_i, s_i\right) > 1$, is a regime in which a small number of discussions are necessary for political parties decide to disagree. Quicked dissensual state can be related to the configuration in which political parties globally decide for polarized opinions. Figure 61 exposes consensual regimes detected using all values of

$\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$ from the three Lower Houses under study.

Figure 61 – Legislative consensual regimes for Lower Houses. Left panel shows values $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$ for the United States House of Representatives (Blue stars), the House of Commons of the United Kingdom (red plus) and the Chamber of Deputies of Brazil (yellow rotated plus). Right panel exposes the same points but this time four colored zones were added. These zones correspond to the dissensual-consensual regimes already proposed. Names for regimes are depicted in the right legend. For each zone, color contrast is associated with the probability for a point $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$ to be in a region R_i, note the fuchsia colorbar. Regimes were obtained from a classification technique applied over random data using Decision Tree Algorithm.



Source: Author.

Left panel shows values of $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$. Blue stars, red plus and yellow rotated plus are associated with the values of the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil, respectively (see left legend box). Right panel shows the same data but this time with four colored zones. Each colored zone is associated with dissensual-consensual regimes aforementioned proposed, take attention to the right legend box. In order to obtain these regimes, a classification technique based Decision Tree algorithm [31] was implemented. Regimes were gathered by generating random data taking into account boundaries for each regime. Regimes arise then as decision zones. Decision zone for each regime indicates which is the probability to be in a regime R_i given a configuration $\left(D\left(\mu_i, s_i\right), \beta_{\max}\left(\chi^1(\hat{\mathbb{W}}, \hat{\mathbb{H}}, \beta)\right)\right)$. Note that probability values are assigned according to the fuchsia colorbar next to the right panel. Fuchsia circles around a marker are used to highlight some years.

Majority of values for the United States House of Representatives and the House of Commons of the United kingdom are found at Quicked Dissensual Regime. Meaning that in these two Lower Houses encourage polarization among main political parties. The

fact that a small quantity of optimal number of discussions are needed to set polarization is held in the Legislative term, smaller in comparison to the Chamber of Deputies of Brazil. The remaining portion of points for these two Lower Houses are located in the Strong Dissensual Regime, specifically, years are 1997, 2018, 2021, 2022 for the United States House of Representatives and 2003, 2005, 2010, 2017, 2019 for the House of Commons of the United Kingdom. Respect to the Chamber of Deputies of Brazil, points are distributed at all Regimes. Take attention to the three points in the Delayed Consensual Regime, years for these are 2012, 2013 and 2014. This coincides with the pre-impeachment period, culminating in 2015, point precisely located at the Strong Consensual Regime. In fact, two of these years were fuchsia circled highlighted (2014 and 2015). Highlighted points for the House of Commons of the United Kingdom are 2001 (located in the Quicked Regime) and 2019 (Strong Regime) which coincides with the terrorist attack in the United States and the Covid-19 pandemic, respectively.

6 CONCLUSIONS.

A complete method for analysing political collective behaviour in Legislative system was presented. It were gathered information about roll-call vote data for the United States House of Representatives, the House of Commons of the United Kingdom and the Chamber of Deputies of Brazil as detailed exposed in Chapter 2. Exposed method aims for analysing transformed and processed roll-call vote data using two approaches: the interactionism and the statistical. Main objective of this method is to characterize legislative consensual-dissensual regimes.

Interactionism approach aims for extract information of about legislative behaviour by studying the dynamics of like spin glass system in which sites of lattice represents political parties. Parameters mandatory are extracted from the extended Boltzmann machine learning algorithms presented in Chapter 1. Algorithm exposed were carefully validated by testing them on synthetic data. Analysing values the political party mean opinion value against the degree of political interaction dissensual and consensual zones arise. These zones helps to characterize what kind of position (conssensual- dissensual) political parties took in each year in each Lower House. Finding the value in which transition from dissensus to consensus happened in each year and Lower House, an estimator of the optimal degree of Legislative interaction needed to reach Legislative consensus is achieved.

Purpose of statistical approach is gather Legislative political information by analysing the correlation matrix computed from Lower Houses members' roll-call vote data for each year. Minimal Spanning Tree ordered correlation matrices shows time evolution of collective consensual and dissensual features for Lower Houses studied. Parties coloured matrices and clusters detected matrices gives us information of how cohesive is the collaboration among Lower Houses members. Probability distribution function offers a quantitative way to characterize the different possible political states (Dissenssus, political randomness and conssensus). By computing bimodal index, a fast way to recognize these kind of political states is achieved.

By join the value in which transition from consensus to dissensual happened in the interactionism approach with the bimodal index measured in the statistical approach, Legislative consensual regimes arise. Four regimes were defined in function of the values coming from the interactionism and statistical approach. Meaning of these regimes were

detailed exposed and the final results gives a powerful way to characterize collective legislative political behaviour.

These method could be used to foresee political storms, to identify political anomalies and as unbiased tool to audit Legislative activity.

BIBLIOGRAPHY

1 MARENCO, L.; CARMONA, H. A.; CARDOSO, F. M.; JR, J. S. A.; CESAR, C. L. Time evolution of the behaviour of brazilian legislative representatives using a complex network approach. **Plos one**, Public Library of Science San Francisco, CA USA, v. 15, n. 2, p. e0226504, 2020.

2 COLLIRI, T. Evaluating presidential support in the brazilian house of representatives through a network-based approach. **arXiv preprint arXiv:2109.03638**, 2021.

3 FETTER, F.; GAMERMANN, D.; BRITO, C. On the stability of the brazilian presidential regime: A statistical analysis. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 571, p. 125832, 2021.

4 YASUDA, M.; TANAKA, K. Approximate learning algorithm in boltzmann machines. **Neural computation**, MIT Press, v. 21, n. 11, p. 3130–3178, 2009.

5 KAPPEN, H. J.; RODRÍGUEZ, F. d. B. Efficient learning in boltzmann machines using linear response theory. **Neural Computation**, MIT Press, v. 10, n. 5, p. 1137–1156, 1998.

6 RICCI-TERSENGHI, F. The bethe approximation for solving the inverse ising problem: a comparison with other inference methods. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2012, n. 08, p. P08015, 2012.

7 KUBO, R. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. **Journal of the Physical Society of Japan**, The Physical Society of Japan, v. 12, n. 6, p. 570–586, 1957.

8 BRUSH, S. G. History of the lenz-ising model. **Reviews of modern physics**, APS, v. 39, n. 4, p. 883, 1967.

9 MÉZARD, M.; PARISI, G.; VIRASORO, M. A. **Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications**. [S.l.]: World Scientific Publishing Company, 1987. v. 9.

10 METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **Journal of Chemical Physics**, v. 21, p. 1087–1092, 1953.

11 BRENT, R. P. **Algorithms for minimization without derivatives**. [S.l.]: Courier Corporation, 2013.

12 ATKINSON, K. E. **An introduction to numerical analysis**. [S.l.]: John wiley & sons, 2008.

13 URSELL, H. The evaluation of gibbs' phase-integral for imperfect gases. In: CAMBRIDGE UNIVERSITY PRESS. **Mathematical Proceedings of the Cambridge Philosophical Society**. [S.l.], 1927. v. 23, n. 6, p. 685–697.

14 NGUYEN, H. C.; ZECCHINA, R.; BERG, J. Inverse statistical problems: from the inverse ising problem to data science. **Advances in Physics**, Taylor & Francis, v. 66, n. 3, p. 197–261, 2017.

15 POWELL, M. J. A hybrid method for nonlinear equations. **Numerical methods for nonlinear algebraic equations**, Gordon and Breach, 1970.

16 UNITED States House of Representatives. Disponível em: https://www.house.gov/. Acesso em: 14 fev. 2023.

17 HOUSE of Commons of the United Kingdom. Disponível em: https://www.parliament.uk/business/commons/. Acesso em: 14 fev. 2023.

18 CâMARA dos Deputados do Brasil. Disponível em: https://www.camara.leg.br/. Acesso em: 14 fev. 2023.

19 FERRARI, P. A.; BARBIERO, A. Simulating ordinal data. **Multivariate Behavioral Research**, Taylor & Francis, v. 47, n. 4, p. 566–589, 2012.

20 BARBIERO, A.; FERRARI, P. A. Simulation of correlated poisson variables. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 31, n. 5, p. 669–680, 2015.

21 ESSAM, J. W. Percolation theory. **Reports on progress in physics**, IOP Publishing, v. 43, n. 7, p. 833, 1980.

22 REPRESENTATIVES, U. of States House of. **H.R.3162 - Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism**. 2009. Disponível em: https://www.congress.gov/bill/107th-congress/house-bill/3162. Acesso em: 14 fev. 2023.

23 KINGDOM, H. of Commons of the U. **House of Lords Act 1999**. 1999. Disponível em: https://www.parliament.uk/about/living-heritage/evolutionofparliament/houseoflords/house-of-lords-reform/from-the-collections/from-the-parliamentary-collections-lords-reform/lords-reform-1963-1999/houseoflordsact1999/. Acesso em: 14 fev. 2023.

24 SIPRI. **Syria vote in British Parliament**. 2013. Disponível em: https://www.sipri.org/commentary/2013/30-aug-2013-syria-vote-british-parliament. Acesso em: 14 fev. 2023.

25 KINGDOM, H. of Commons of the U. **Recall Parliament**. 2013. Disponível em: https://www.parliament.uk/about/how/occasions/recallparliament/. Acesso em: 14 fev. 2023.

26 RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986.

27 KIEFER, J. Sequential minimax search for a maximum. **Proceedings of the American mathematical society**, v. 4, n. 3, p. 502–506, 1953.

28 CRISTELLI, M.; BATTY, M.; PIETRONERO, L. There is more than a power law in zipf. **Scientific reports**, Springer, v. 2, n. 1, p. 1–7, 2012.

29 MA, Y.; ZHANG, N.; LI, J. Improved sequential least squares programming–driven feasible path algorithm for process optimisation. In: MONTASTRUC, L.; NEGNY, S. (Ed.). **32nd European Symposium on Computer Aided Process Engineering**. Elsevier, 2022, (Computer Aided Chemical Engineering, v. 51). p. 1279–1284. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780323958790502149.

30 ASHMAN, K. M.; BIRD, C. M.; ZEPF, S. E. Detecting bimodality in astronomical datasets. **arXiv preprint astro-ph/9408030**, 1994.

31 BREIMAN, L. **Classification and regression trees**. [S.l.]: Routledge, 2017.

APPENDIX A – AXIOMS OF THE INFORMATION THEORY

The four axioms of the information theory are:

1. Let $f(n) = S\left(\frac{\vec{1}}{n}\right) = S\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$. The lower the probability of an event the larger the information obtained by knowing that such event happened. First axiom establishes that$f(n)$ is a monotonically increasing function of n;
2. For two independent Random variables $x, y \rightarrow \vec{p}_x = \frac{\vec{1}}{n} \rightarrow \vec{p}_y = \frac{\vec{1}}{m}$ the joint probability to find $p[x \,\text{and}\, y] = \frac{1}{n} \times \frac{1}{m}$. The second axiom establishes that $f(n \times m) = f(n) + f(m)$;
3. Breaking the complete set $C = \{1, 2, \cdots, n\}$ with a probability vector $\vec{p}_n = (p_1, p_2, \cdots, p_n)$ of a random variable into two subsets $A = \{1, 2, \cdots, k\}$ and $B = \{k+1, k+2, \cdots, n\}$, with the probability vectors $\vec{p}_k = (p_1, p_2, \cdots, p_k) \rightarrow \vec{p}_{n-k} = (p_{k+1}, p_{k+2}, \cdots, p_n)$, the chance that an element belongs to set A is $p_A = \sum_{i=1}^{k} p_i$ and to belong to set B is $p_B = \sum_{i=k+1}^{n} p_i$. The axiom affirms that the process to first choose the subset and then the element of the subset must be equal to choose the element directly from the whole set at once. The 3$^{\text{th}}$ axiom, therefore, establishes that: $S(\vec{p}_n) = S(p_A, p_B) + p_A S\left(\frac{\vec{p}_k}{p_A}\right) + p_B S\left(\frac{\vec{p}_{n-k}}{p_B}\right)$;
4. The 4$^{\text{th}}$ and last axiom simply requires that $S(p, 1-p)$ be continous in p.

Given the four axioms we can verify that the function $S(\vec{p}) = -c \sum_{i=1}^{n} p_i \log_a p_i \rightarrow a > 1 \rightarrow c > 0$ satisfies all four axioms. Shannon also proved that this the unique function compatible with the axioms.

1. $f(n) = c \log n$ and $f(m) = c \log m$ then $f(n) + f(m) = c \log n + c \log m = c \log(n \times m) = f(n \times m)$;

2. This straight forward with the algebra:

$$
\begin{aligned}
& S\left(p_A, p_B\right)+p_A S\left(\frac{\vec{p}_k}{p_A}\right)+p_B S\left(\frac{\vec{p}_{n-k}}{p_B}\right) \\
& =-c\left[p_A \log_a p_A+p_B \log_a p_B+p_A \sum_{i=1}^{k}\left(\frac{p_i}{p_A} \log_a \frac{p_i}{p_A}\right)+p_B \sum_{i=k+1}^{n}\left(\frac{p_i}{p_B} \log_a \frac{p_i}{p_B}\right)\right] \\
& =-c\left[p_A \log_a p_A+p_B \log_a p_B+\sum_{i=1}^{k}\left(p_i \log_a p_i\right)-\left(\sum_{i=1}^{k} p_i\right) \log_a p_A\right] \\
& -c\left[\sum_{i=k+1}^{n}\left(p_i \log_a p_i\right)-\left(\sum_{i=k+1}^{n} p_i\right) \log_a p_B\right] \\
& =-c\left[p_A \log_a p_A+p_B \log_a p_B+\sum_{i=1}^{n}\left(p_i \log_a p_i\right)-p_A \log_a p_A-p_B \log_a p_B\right] \\
& =-c \sum_{i=1}^{n}\left(p_i \log_a p_i\right) \\
& =S\left(\vec{p}_n\right)
\end{aligned}
\tag{A.1}
$$

3. $S\left(p, 1-p\right)=-c\left[p \log p+(1-p) \log \left(1-p\right)\right]=-c \log \left[p^p(1-p)^{(1-p)}\right]$ is a continuous function of p

APPENDIX B – MAXIMIZING ENTROPY

$$\text{Maximze}\left\{S=-c\sum_{i=1}^{n}p_i\ln p_i \quad \text{subjected to the constraint } \sum_{i=1}^{n}p_i=1\right\}$$

$$\text{Lagrangian}: \quad L(\lambda,\vec{x})=-c\sum_{i=1}^{n}p_i\ln p_i-\lambda\left[\sum_{i=1}^{n}p_i-1\right]$$

$$\frac{\partial L}{\partial\lambda}=0 \quad \rightarrow \quad \sum_{i=1}^{n}p_i=1$$

$$\frac{\partial L}{\partial p_j}=0 \quad \rightarrow \quad -c\sum_{i=1}^{n}\ln p_i\frac{\partial p_i}{\partial p_j}-c\sum_{i=1}^{n}p_i\frac{\partial\ln p_i}{\partial p_j}-\lambda\sum_{i=1}^{n}\frac{\partial p_i}{\partial p_j} \tag{B.1}$$

$$=-c\sum_{i=1}^{n}\delta_{ij}\ln p_i-c\sum_{i=1}^{n}\delta_{ij}\frac{p_i}{p_i}-\lambda\sum_{i=1}^{n}\delta_{ij}=$$

$$=-c\ln p_j-c-\lambda=0$$

$$\ln p_j=-1-\frac{\lambda}{c} \quad \rightarrow \quad p_j=e^{-1-\frac{\lambda}{c}}=p_o \quad \rightarrow \quad \sum_{i=1}^{n}p_o=np_o=1 \quad \rightarrow \quad p_o=\frac{1}{n}$$

APPENDIX C – INEQUALITY RELATION FOR ENTROPY

The logarithm is convex function meaning it is Always below the tangent line in any point, in particular at the $x = 1$ point, when the tangent line is $y_{\tan} = x - 1$. Therefore $\ln x \leq x - 1 \quad \forall x$, with the equality valid only at $x = 1$.

This means that:

$$\begin{aligned}
&\ln \frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1 \\
&\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1 \right) = \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i = 1 - 1 = 0 \\
&\sum_{i=1}^{n} p_i \ln \frac{q_i}{p_i} \leq 0 \\
&\sum_{i=1}^{n} p_i \ln q_i \leq \sum_{i=1}^{n} p_i \ln p_i \\
&- c \sum_{i=1}^{n} p_i \ln p_i \geq -c \sum_{i=1}^{n} p_i \ln q_i
\end{aligned} \tag{C.1}$$

Now, we can define a new probability $q_{ij} = p(x_i)\, p(y_j)$ because:

$$1 = \left[\sum_{i=1}^{n} p_i \right] \left[\sum_{j=1}^{m} p_j \right] = \sum_{i=1}^{n} \sum_{j=1}^{m} p_i p_j \quad \rightarrow \quad \sum_{i=1}^{n} \sum_{j=1}^{m} q_{ij} = 1 \tag{C.2}$$

Therefore:

$$S(x) + S(y) = -c \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_a q_{ij} \geq -c \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_a p_{ij} = S(x, y) \tag{C.3}$$

The equality is valid only when x and y are independent because $p(x_i, y_j) = p(x_i)\, p(y_j)$.