





*Felippe Alves Pereira*

# Quebra de Simetria Espontânea, Limites Cognitivos e Complexidade de Sociedades



## *Resumo*

Neste trabalho foi elaborado um modelo para a construção de estruturas sociais em relação a características cognitivas das pessoas. Com base em estudos nos campos da neurociência e da psicologia social, além do uso das técnicas de mecânica estatística e da teoria de aprendizado de máquina, fomos capazes de estabelecer um modelo de agentes que trocam opiniões, aprendem e escolhem seus parceiros com base nas informações que obtêm.

A dinâmica das relações sociais introduzida aqui é capaz de produzir complexidade na estrutura das relações sociais e na distribuição de opiniões de forma correlacionada. Essa dinâmica nos dá um parâmetro que mede o nível de organização social e nos permite interpretar o fenômeno de formação de comunidades como uma estrutura que requer uma "*massa crítica*" para se estabelecer.

Com base nesses resultados, propusemos um modelo capaz de reproduzir parte do comportamento observado no plenário brasileiro ao longo dos anos em relação a votações de projetos de lei sob diferentes mandatos presidenciais.



# *Abstract*

In this work we establish a model for social structure construction related with human cognitive character. From neuroscience and social psychology studies, and also relying on technique from statistical mechanics and machine learning theory, we were able to establish a model of agents who exchange opinions, learn together and choose their partnerships using the information they gather.

The social relationship dynamics here introduced is capable of generating complexity on the social structure level and a on opinion distribution in a correlated way. This dynamics gives a parameter representing the degree to which society is organized and allow us to see the phenomena of community formation as a structure requiring a "*critical mass*" to fixate.

These results guided us in proposing a model able to replicate a portion of the behavior observed at the Brazilian plenary through year of voting law projects under different president mandates.





# Sumário

1	<i>Um Panorama do Estudo de Fenômenos Sociais</i>	11
1.1	<i>Aprendizado por Reforço</i>	13
1.2	<i>Evidências do Aprendizado Social</i>	14
1.3	<i>Moral e a Primazia da Intuição</i>	15
2	<i>Aspectos Matemáticos dos Modelos de Complexidade Social</i>	17
2.1	<i>Aprendizado de Máquina</i>	17
2.1.1	<i>Aprendizado Supervisionado Online</i>	18
2.1.2	<i>Um Modelo mais Simples de Aprendizado</i>	24
2.2	<i>Modelos de Agentes e Mecânica Estatística</i>	26
2.2.1	<i>Estabelecendo a Linguagem do Modelo de Agentes</i>	26
2.2.2	<i>Mecânica Estatística de Sistemas Sociais</i>	29
3	<i>Experimentos Computacionais e Análise dos Resultados</i>	39
3.1	<i>Monte Carlo em Sociedades de Agentes e a Termodinâmica de Consenso</i>	39
3.1.1	<i>As Fases de Uma Sociedade</i>	40
3.1.2	<i>Um Retrato (Estatístico) da Semelhança Entre Agentes</i>	42
3.2	<i>Uma Dinâmica para as Relações Sociais</i>	43
3.2.1	<i>Complexidade no Nível das Opiniões</i>	45
3.2.2	<i>O Efeito da Dinâmica no Tecido Social</i>	47
3.2.3	<i>Outra Forma de Ordem</i>	48

3.3	<i>Jogos de Partidos</i>	52
4	<i>Conclusões e Futuros Estudos</i>	57
4.1	<i>Uma Breve Revisão dos Resultados</i>	57
4.2	<i>O que há pela frente?</i>	58

# 1

## *Um Panorama do Estudo de Fenômenos Sociais*

*“ ... I can see no other escape from this dilemma (lest our true aim be lost for ever) than that some of us should venture to embark on a synthesis of facts and theories, albeit with second-hand and incomplete knowledge of some of them, and at the risk of making fool of ourselves. So much for my apology. ”*

— Erwin Schrödinger, *What is Life?* (1944)

Os conceitos de fenômeno social e de natureza humana se entrelaçam. A complexidade e diversidade observadas na sociedade despertam o interesse de cientistas em diferentes campos, dedicados a questionar e entender os mecanismos que as produzem. É possível, por exemplo, que um sociólogo se interesse em entender como diferentes grupos lidam com resolução de conflitos, enquanto um psicólogo esteja preocupado com a abordagem de cada indivíduo envolvido numa situação de conflito, um neurocientista queira analisar a ativação cerebral em cada abordagem descrita pelo psicólogo, um estatístico queira entender as correlações entre estrutura social e comportamento, e a um físico talvez caiba tentar entender a dinâmica que rege esse tipo de fenômeno.

Uma característica, talvez a mais interessante, de sistemas sociais é a exibição simultânea de estruturas organizadas e ricas em diversidade. Embora as pessoas tendam a reproduzir o comportamento de seus pares, ainda observamos diversidade cultural, formação de estados e dialetos num mesmo idioma por exemplo. O estudo dessa característica das sociedades humanas tem sido abordada por diferentes estratégias na pesquisa contemporânea.

Nesse contexto, o psicólogo social J. Haidt e seus colegas na formulação da *Teoria dos Fundamentos Morais*<sup>1</sup> estabeleceram as bases para relacionar a resposta neural de indivíduos em questões de conflito moral com a participação deste na manutenção da sociedade. Através de resultados sobre o aprendizado de comportamento moral, da intuitividade dos

<sup>1</sup> Haidt 2007; Graham et al. 2011.

julgamentos e do mapeamento da moral <sup>2</sup>, a Teoria dos Fundamentos Morais, propõe que o domínio da moralidade humana não é composto apenas por questões relacionadas ao modo como dois indivíduos deviam se tratar, mas também por características que dizem respeito ao papel do indivíduo no grupo. Além disso, a diversidade encontrada no conceito de moral através das culturas pode ser explicada tomando a moral como um conjunto de mecanismos, biológicos ou culturais, que regulam o egoísmo e tornam a vida social possível. Essa linha de pensamento contribui para a compreensão da complexidade social na escala das interações humanas e será um ponto de partida útil na elaboração dos modelos estudados neste trabalho.

Numa perspectiva mais coletivista, R. Axelrod<sup>3</sup> propôs um modelo qualitativo para a diversidade cultural em termos de agentes interagentes, no qual ele foi capaz de mostrar a possibilidade da coexistência de diferentes culturas numa dada região como uma consequência da homofilia local. Nesse modelo, a capacidade de interagir com maior probabilidade com agentes mais parecidos é capaz de produzir fronteiras separando regiões de agentes que são compatíveis e incompatíveis. Esse resultado é interessante por si, mas o modelo deixa aberto generalizações em diversos sentidos como a topologia do grafo onde os agentes interagem e o efeito das particularidades no mecanismo de aprendizado.

O uso de modelos de agentes em grafos tem sido amplamente empregado no estudo de sistemas sociais. O presente trabalho se insere numa série de outros no esforço de entender melhor o papel das características humanas na constituição das estruturas sociais. Integram essa série, modelos relacionando o aprendizado moral com as estruturas cognitivas da interação de indivíduos<sup>4</sup> e modelos de surgimento de autoridade e desenvolvimento econômico<sup>5</sup>.

O ponto em comum nesses trabalhos recentes são algumas bases empíricas que proporcionam a abordagem de comportamentos sociais como consequência das capacidade humana de aprendizado, das respostas neurológicas a dilemas morais e conflitos sociais ou nas escolhas de estratégias, das estruturas de informação provenientes das interações sociais e de uma compreensão um pouco mais ampla dos processos de aprendizado.

A exploração dos fenômenos sociais como consequência das características psico ou neurológicas dos indivíduos distingue esses recentes esforços das linhas mais ‘cinemáticas’ de estudo<sup>6</sup>. Em outras palavras, além da descrição de como um certo fenômeno ocorre são interessantes também as causas desse fenômeno. Seguiremos, então, com uma breve apresentação da motivações empíricas sobre o comportamento humano e sua possível influência nas construções sociais.

<sup>2</sup> citar o que tiver

<sup>3</sup> Axelrod 1997.

<sup>4</sup> Nestor Caticha e Renato Vicente 2011; R. Vicente et al. 2014; César 2014.

<sup>5</sup> Papa 2014; Calsaverini 2013.

<sup>6</sup> Castellano, Fortunato e Loreto 2009.

## 1.1 Aprendizado por Reforço

Uma das características, não exclusivamente, humana é sua capacidade de aprender. Somos capazes de aprender comportamento, criar abstrações e fazer previsões com base em experiência prévia em diversas circunstâncias. Embora essa afirmação não seja chocante em si, talvez surpreenda o fato de que estamos aprendendo como funcionam nossos mecanismos de aprendizado, a nível neurológico, psicológico e social.

Um desses mecanismos é o de aprendizado por reforço, ou em outras palavras a habilidade de aprender com erros e acertos. Para isso é necessário que tenhamos aparatos neurais para planejar uma ação, prever o resultado e detectar erros na previsão para poder ajustar uma próxima ação.

Estudos realizados por (Holroyd e Coles 2002) mostram como o mecanismo de as regiões do Córtex Cingulado Anterior<sup>7</sup> em conjunto com os sistemas dopaminérgicos do mesencéfalo funcionam como um mecanismo de detecção de erros e recompensas durante o aprendizado de uma tarefa. Esse mecanismo é estudado através da análise de eletroencefalograma, mais especificamente na ocorrência de Negatividade Relacionada ao Erro<sup>8</sup>. A existência de tal mecanismo de aprendizado por reforço é uma das bases que fundamenta o uso da teoria de aprendizado de máquina para modelar o comportamento social.

A tarefa de detectar conflitos exercida pelo córtex cingulado anterior está diretamente relacionada à sensação de dor física. Em paralelo à capacidade de aprendizado, os estudos de (Somerville, Heatherton e Kelley 2006; Eisenberger, Lieberman e Williams 2003) apontam um aumento na atividade do córtex cingulado anterior quando um indivíduo vivencia exclusão social ou é avaliado negativamente por outras pessoas. O fato da região do cérebro responsável pela dor física também ser ativada durante a experiência de conflitos sociais sugere que há um custo, talvez análogo à dor física, a ser pago ao se envolver em tais conflitos. Essa relação de analogia entre dor física e dor por exclusão social nos permite sustentar a hipótese do *custo cognitivo* que regulará as interações no nosso modelo.

Outro resultado interessante sobre aprendizado, no estudo desenvolvido por (Amodio et al. 2007), é a correlação entre a amplitude da Negatividade Relacionada ao Erro e o posicionamento político auto declarado de indivíduos. Essa diferença na amplitude da ativação se reflete na sensibilidade do indivíduo para responder na resposta de conflitos. Em particular, foi observado que conservadores tendem a apresentar menores amplitudes de ERN quando comparados a liberais. Dada a relevância da detecção e resolução de conflitos, é de se esperar que liberais e conservadores<sup>9</sup> apresentem diferentes *estilos cognitivos* de aprendizado, o usaremos essa abordagem ao formular o custo cognitivo.

<sup>7</sup> tradução livre de *Anterior Cingulate Cortex*(ACC)

<sup>8</sup> tradução livre de *Error-Related Negativity* (ERN)

<sup>9</sup> essas denominações de *liberal* e *conservador* são relativas ao espectro político do Estados Unidos, embora resultado citado seja de aplicabilidade mais ampla

## 1.2 Evidências do Aprendizado Social

A formação da opinião de um indivíduo com relação à algum dilema é influenciada não apenas pela informação objetiva coletada <sup>10</sup> pelo seu aparato sensorial, como também pela *pressão social* de indivíduos envolvidos na discussão de tal dilema. O conteúdo dessa afirmação foi o tema de estudos de *M. Sherif* e *S. Asch* <sup>11</sup>.

<sup>10</sup> ou ao menos tão objetiva quanto possível, dadas as limitações biológicas

<sup>11</sup> Sherif 1937; Asch 1955.

O experimento de *Sherif* é estabelecido da seguinte forma: um grupo de participantes é colocado numa sala totalmente escura exceto por um ponto de luz projetado a uma distância fixa de um grupo de participantes. Dentre eles, dois estão de prévio acordo, sem o conhecimento do terceiro, em estimar um certo deslocamento para o ponto de luz dentro de um certo intervalo de valores. Após alguns segundos observando o ponto, que está parado, os participantes revelam suas estimativas para o deslocamento do ponto de luz, fenômeno conhecido como ilusão autocinética, e uma nova rodada de observações é feita. No dia seguinte, o experimento é repetido apenas com o terceiro participante. A hipótese é que o número de observações que o participante ‘*ingênuo*’ faz dentro do intervalo pré determinado pelos experimentadores é influência da interação entre os participantes.

Os resultados desse experimento mostram que as estimativas do participante ingênuo convergem para o mesmo valor das do grupo, caindo dentro do intervalo determinado, dependendo do quão grosseira é a percepção do indivíduo sobre essa estimativa do grupo. Além disso, a estimativa persiste durante o experimento individual.

O experimento realizado por *Asch* tem como objetivo determinar como esse tipo de influência varia conforme a variação do tamanho do grupo, ou da precisão da opinião do grupo com relação ao objetivo entre outras circunstâncias. Para isso, grupo entre seis e 9 indivíduos, sendo todos com exceção de um confederados do estudo, aos quais eram apresentadas duas cartas. Uma carta continha uma barra preta e a outra continha três barras pretas de diferentes comprimentos, das quais os participantes deveriam escolher aquela com o mesmo comprimento da barra da outra carta. Sempre uma das três barra era a correta, e o experimento foi repetido com uma variedade de diferenças relativas entre as três barras, de modo a regular a dificuldade da tarefa.

Para verificar a influência do grupo, que era instruído a escolher uma das opções erradas um certo número de vezes, foi medido número de vezes que o indivíduo ingênuo errava na escolha da barra correta quando em grupo e de forma individual. Observou-se que a taxa de erro, tipicamente da ordem de 1% ao realizar a tarefa individualmente, aumentava para cerca de 30% quando realizada sob na presença do grupo.

Num estudo mais recente, (Campbell-Meiklejohn et al. 2010) mostrou através da análise de imagens de ressonância magnética funcional

o efeito da influência social na atribuição de valor a algo. A atribuição de valor está associada à uma ativação na região do estriado ventral no cérebro, e a influência social gera uma ativação nessa região quando do indivíduo sob o qual atua.

Esses estudos mostram o efeito direto da influência do grupo na constituição da opinião de indivíduos. Aliada ao conhecimento sobre os mecanismos de aprendizado por reforço, de dor por exclusão social e aos diferentes estilos cognitivos, a influência social formará o mecanismo de aprendizado de comportamento social. O ingrediente final para a construção do modelo diz respeito sobre o espaço no qual existe um certo comportamento pode ser '*codificado*'. Mais especificamente, tomaremos como referência o mapeamento da moral humana e a intuitividade dos julgamentos morais, apresentados a seguir.

### 1.3 Moral e a Primazia da Intuição

A Teoria do Fundamentos morais desenvolvida por *J. Haidt* e *J. Graham*<sup>12</sup> contribui para a compreensão da moral humana através de quatro princípios, dos quais dois são essenciais na elaboração deste trabalho.

<sup>12</sup> Haidt 2007; Graham et al. 2011.

O primeiro diz respeito à primazia da intuição, ou seja, ao modo automático e inconsciente de tomar decisões morais. O uso da palavra *primazia* tem como principal objetivo lembrar que, embora a intuição seja a primeira fonte de conclusões, ela não necessariamente domina o julgamento do indivíduo, podendo ser corrigida posteriormente por mecanismos a nível da consciência e do raciocínio.

Essa característica da intuitividade nos permite fazer a associação entre a evolução da cognição moral e a resposta a um estímulo adequado. Essa associação proporciona um dos vínculos interpretativos entre o aprendizado de comportamento social e a teoria do aprendizado de máquina.

O segundo princípio importante para nós é o mapeamento das características morais num conjunto de *fundamentos universais* da moral humana. Isso significa que, embora os valores morais variem entre culturas e épocas da civilização, há características em comum entre todas elas. Essas características são, de certa forma, independentes entre si e formam uma base sobre a qual a moral é construída dentro da população. Em particular, *Haidt* e seus colegas determinaram através da análise de questionários aplicados sobre uma vasta extensão do planeta que são apenas os fundamentos da moral humana, a saber Violência/Cuidado, Justiça/Reciprocidade, Inclusão em Grupo/Lealdade, Autoridade/Respeito e Pureza/Santidade.

A contribuição desse princípio para o nosso modelo é a justificativa na representação de características cognitivas dos indivíduos através de um vetor em algum espaço adequado. Note que, embora a Teoria dos

Fundamentos Morais verse apenas sobre a moralidade humana, faremos uma extrapolação e aplicaremos esse conceito sobre alguma característica cognitiva qualquer, na esperança de prover algum modelo que seja capaz de fazer previsões.

As considerações feitas aqui não ilustram a completeza de cada referência feita, que trazem resultados interessantes por si. Entretanto, as motivações empíricas apresentadas são suficientes para o desenvolvimento do trabalho, com exceção de alguns resultados específicos da teoria do aprendizado de máquinas, que serão apresentados no capítulo a seguir. Daremos, portanto, seguimento ao desenvolvimento da teoria, sem nos aprofundarmos mais nesses assuntos.



## 2

# *Aspectos Matemáticos dos Modelos de Complexidade Social*

Este capítulo se compromete com a introdução dos conceitos matemáticos básicos na elaboração de modelos para explicar o comportamento social. Espera-se, contudo, que o leitor esteja familiarizado com a linguagem do cálculo diferencial e integral em várias variáveis e tenha boa noção dos conceitos de probabilidade e estatística. Dito isso, o tratamento dos tópicos neste capítulo está longe de ser considerado completo ou rigoroso, tendo como principal função uma de referência para a interpretação e análise dos resultados apresentados no capítulo 3. Um tratamento mais abrangente do estudo de aprendizado de máquinas pode ser encontrado em (Engel e Den Broeck 2001), enquanto tópicos específicos de teoria de grafos podem ser estudados em (M.E.J. Newman 2010).

Começaremos com uma breve revisão da teoria de *Aprendizado de Máquina*, introduzimos as linguagens da *Teoria de Grafos* e da *Mecânica Estatística*. Em seguida, elaboramos o conceito de *Modelos de Agentes* e estabelecemos o contexto de sua aplicação em sistemas sociais através da identificação das características observadas na realidade com os elementos do modelo. Finalmente, uma análise de campo médio é feita para obter uma previsão teórica do que se espera do modelo.

### 2.1 Aprendizado de Máquina

A investigação de fenômenos sociais demanda uma certa compreensão de *indivíduo* e suas relações com seu arredores e com os demais indivíduos. Uma pessoa é, por si, um sistema altamente complexo, e a riqueza do comportamento humano torna impraticável a elaboração de uma representação matemática perfeita para o indivíduo. Todavia, é possível focar nas características que pareçam, em uma primeira aproximação, mais relevantes para a compreensão do fenômeno em estudo. Por exemplo, para compreender o comportamento de multidões em trânsito, a

posição e a estratégia para se mover são mais relevantes do sua opinião sobre política, entretanto esta se torna mais importante num contexto de eleições. Para abordar fenômenos sociais relacionados à diversidade de opiniões ou ideologias políticas, parece razoável dar foco no modo cada indivíduo constrói e representa esses conceitos.

### 2.1.1 Aprendizado Supervisionado Online

Como visto no capítulo 1, comportamento social é aprendido e que a imitação de comportamento é uma das formas pela qual aprendizado se dá<sup>1</sup>. O estudo do Aprendizado de Máquina, em particular o de *aprendizado supervisionado*, será adequado para a descrever o comportamento de indivíduos em contato num contexto de troca de *opiniões*.

<sup>1</sup> veja o capítulo 1 para mais detalhes.

A construção dessa associação é feita estabelecendo o cenário de aprendizado *aluno-professor*. Tanto professor quanto aluno são representados por vetores  $K$ -dimensionais,  $\mathbf{w}, \bar{\mathbf{w}} \in R^K$ , envolvidos na discussão de questões  $\mathbf{x}_a \in \mathbb{R}^K$ , também representadas por vetores no mesmo espaço. Cada exemplo do conjunto  $D_t = \{y_i, \dots, y_t\}$  é composto por uma ‘questão’  $\mathbf{x}_a$  e uma resposta  $\tau_a$  a ela dada pelo professor. Os exemplos em  $D_t$  são independentes entre si e identicamente distribuídos pela  $P(y_a|\mathbf{w}) = P(\tau_a|\mathbf{x}_a, \mathbf{w})P(\mathbf{x}_a)$ . Através do conjunto de exemplos  $D_t$ , o aluno  $\bar{\mathbf{w}}$  tentará aprender o comportamento do professor  $\mathbf{w}$ .

Usando o teorema de *Bayes* temos

$$P(\mathbf{w}|D_t) = \frac{Q(\mathbf{w})P(D_t|\mathbf{w})}{\int d\mathbf{w}' Q(\mathbf{w}')P(D_t|\mathbf{w}')} \quad (2.1)$$

onde  $Q(\mathbf{w})$  é uma distribuição à priori para  $\mathbf{w}$ . Considere a apresentação de um novo exemplo  $y_{t+1} = y = (\tau, \mathbf{x})$  ao conjunto  $D_t$ , independente dos demais. A postura no aprendizado Bayesiano é assumir a distribuição à posteriori  $P(\mathbf{w}|D_t)$  obtida com os exemplos anteriores como a priori  $Q(\mathbf{w}|D_t)$  sobre a apresentação do novo exemplo. Com isso, a atualização da distribuição fica

$$\begin{aligned} P(\mathbf{w}|y, D_t) &= \frac{Q(\mathbf{w}|D_t)P(y|\mathbf{w})}{\int d\mathbf{w}' Q(\mathbf{w}'|D_t)P(y'|\mathbf{w}')} \\ &= \frac{Q(\mathbf{w}|D_t)P(\tau|\mathbf{x}, \mathbf{w})}{\int d\mathbf{w}' Q(\mathbf{w}'|D_t)P(\tau|\mathbf{x}, \mathbf{w}')} \end{aligned} \quad (2.2)$$

onde foi usado o fato do novo exemplo também ser independente dos demais, ou seja,

$$P(y, D_t|\mathbf{w}) = P(y|D_t, \mathbf{w})P(D_t|\mathbf{w}) = P(y|\mathbf{w})P(D_t|\mathbf{w})$$

e o fato de  $P(\mathbf{x})$  ser independente de  $\mathbf{w}$ .

O custo computacional de armazenar o conjunto de exemplos e computar a atualização da distribuição posterior crescem a cada apresentação de um novo exemplo. Seria interessante, do ponto de vista da eficiência computacional<sup>2</sup>, ser capaz de fazer boa inferência com base nos

<sup>2</sup> Não apenas do ponto de vista tecnológico, mas também como uma questão evolutiva. Imagine que os indivíduos de uma espécie tenham evoluído sua capacidade de armazenar memória e inferir sobre fontes de perigo com base nas suas experiências de vida. Ambas tarefas tem um custo energético, demandam tamanho cerebral e têm influência direta na sobrevivência dos indivíduos, então é de se esperar que espécies mais eficientes tenham vantagem, num contexto de seleção natural, quando em competição com outras menos eficientes. Para uma abordagem interessante da evolução de programas veja (Juan Pablo Neirotti e Caticha 2003; Juan P Neirotti e Caticha 2006)

exemplos sem a inconveniência advinda do tamanho do seu conjunto. Para isso, faremos uma aproximação gaussiana<sup>3</sup>  $Q(\mathbf{w}|D_t) \rightarrow G(\mathbf{w}|\theta)$ , com  $\theta = (\bar{\mathbf{w}}, \mathbf{C})$  representando o ponto no espaço dos parâmetros da família gaussiana, a saber sua média  $\bar{\mathbf{w}}$  e a correlação  $\mathbf{C}$ , de modo que

$$G(\mathbf{w}|\theta) = \det(2\pi\mathbf{C})^{-\frac{1}{2}} \exp \left[ (\mathbf{w} - \bar{\mathbf{w}})^T \mathbf{C}^{-1} (\mathbf{w} - \bar{\mathbf{w}}) \right]$$

e a equação (2.2) fica

$$P(\mathbf{w}|y\theta) = \frac{Q(\mathbf{w}|\theta)P(\tau|\mathbf{x}\mathbf{w})}{\int d\mathbf{w}' Q(\mathbf{w}'|\theta)P(\tau|\mathbf{x}\mathbf{w}')} \quad (2.3)$$

Com a adição de um novo exemplo, a distribuição posterior  $P(\mathbf{w}|y\theta)$  não necessariamente pertence à família das distribuições gaussianas. Essa situação é remediada projetando a posterior no espaço dos parâmetros adequados com o mínimo possível de descarte da informação provida pelo novo exemplo  $y$ , que é possível através da maximização da entropia relativa<sup>4</sup>:

$$S[P(\mathbf{w}|y\theta) : G(\mathbf{w}|\theta_+)] = - \int d\mathbf{w} P(\mathbf{w}|y\theta) \ln \frac{P(\mathbf{w}|y\theta)}{G(\mathbf{w}|\theta_+)} \quad (2.4)$$

onde  $\theta_+ = (\bar{\mathbf{w}}_+, \mathbf{C}_+)$  indica o ponto no espaço paramétrico após as apresentação do novo exemplo.

Para minimizar a equação em (2.4) em relação à  $\theta_+$ , basta substituir  $P(\mathbf{w}|y\theta)$  pela identidade em (2.3) e obter os extremos em relação a  $\bar{\mathbf{w}}_+$  e  $\mathbf{C}_+$ . Chamando  $Z = \int d\mathbf{w}' G(\mathbf{w}'|\theta)P(\tau|\mathbf{x}\mathbf{w}')$ , e denotando  $\partial_{\bar{\mathbf{w}}_+} = \frac{\partial}{\partial \bar{\mathbf{w}}_+}$ , a maximização em relação a  $\bar{\mathbf{w}}_+$  é feita<sup>5</sup> da seguinte forma

$$\begin{aligned} 0 = \partial_{\bar{\mathbf{w}}_+} S &= -\partial_{\bar{\mathbf{w}}_+} \int d\mathbf{w} \frac{1}{Z} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) \ln \frac{P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta)}{Z G(\mathbf{w}|\theta_+)} \\ &= \frac{1}{Z} \int d\mathbf{w} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) \partial_{\bar{\mathbf{w}}_+} \ln G(\mathbf{w}|\theta_+) \\ &= -\frac{1}{Z} \int d\mathbf{w} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) (\mathbf{w} - \bar{\mathbf{w}}_+)^T \mathbf{C}_+^{-1} \\ &= \bar{\mathbf{w}}_+^T \mathbf{C}_+^{-1} - \frac{1}{Z} \int d\mathbf{w} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) \mathbf{w}^T \mathbf{C}_+^{-1} \end{aligned}$$

de modo a obter

$$\Rightarrow \bar{\mathbf{w}}_+^T = \frac{1}{Z} \int d\mathbf{w} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) \mathbf{w}^T \quad (2.5)$$

Para calcular a integral em (2.5), faça a mudança de variáveis  $\mathbf{u} = \mathbf{w} - \bar{\mathbf{w}}$ , de modo que<sup>6</sup>

<sup>3</sup> Solla e Winther 1999; Oppen 1996.

<sup>4</sup> ou de forma equivalente, através da minimização da divergência de Kullback-Liebler

<sup>5</sup> usando as identidades

$$\partial_v (\mathbf{v}^T A \mathbf{v}) = 2\mathbf{v}^T A$$

para um vetor  $\mathbf{v}$  e uma matriz  $A$ , e

$$\begin{aligned} \ln G(\mathbf{w}|\theta_+) &= -\frac{K}{2} \ln 2\pi \\ &\quad -\frac{1}{2} \ln \det \mathbf{C}_+ \\ &\quad -\frac{1}{2} (\mathbf{w} - \bar{\mathbf{w}}_+)^T \mathbf{C}_+^{-1} (\mathbf{w} - \bar{\mathbf{w}}_+) \end{aligned}$$

<sup>6</sup> usando as identidades

$$\begin{aligned} \partial_a f(a+b) &= \partial_b f(a+b) \\ \partial_{\mathbf{u}} G(\mathbf{u}|0, \mathbf{C}) &= -\mathbf{u}^T \mathbf{C}^{-1} G(\mathbf{u}|0, \mathbf{C}) \end{aligned}$$

$$\begin{aligned}
\bar{\mathbf{w}}_+^T &= \frac{1}{Z} \int d\mathbf{w} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) \mathbf{w}^T \\
&= \frac{1}{Z} \int d\mathbf{u} P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}}) G(\mathbf{u}|0, \mathbf{C}) (\mathbf{u} + \bar{\mathbf{w}})^T \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} \int d\mathbf{u} P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}}) G(\mathbf{u}|0, \mathbf{C}) \mathbf{u}^T \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} \int d\mathbf{u} P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}}) [-\partial_{\mathbf{u}} G(\mathbf{u}|0, \mathbf{C})] \mathbf{C} \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} \int d\mathbf{u} G(\mathbf{u}|0, \mathbf{C}) [\partial_{\mathbf{u}} P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}})] \mathbf{C} \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} \int d\mathbf{u} G(\mathbf{u}|0, \mathbf{C}) [\partial_{\bar{\mathbf{w}}} P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}})] \mathbf{C} \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} \partial_{\bar{\mathbf{w}}} \left[ \int d\mathbf{u} G(\mathbf{u}|0, \mathbf{C}) P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}}) \right] \mathbf{C} \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} (\partial_{\bar{\mathbf{w}}} \langle P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}}) \rangle_{\mathbf{u}}) \mathbf{C} \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} (\partial_{\bar{\mathbf{w}}} \langle P(\tau|\mathbf{x}\mathbf{w}) \rangle_{\mathbf{w}}) \mathbf{C} \\
&= \bar{\mathbf{w}}^T + \frac{1}{Z} (\partial_{\bar{\mathbf{w}}} Z) \mathbf{C} = \bar{\mathbf{w}}^T + (\partial_{\bar{\mathbf{w}}} \ln Z) \mathbf{C}
\end{aligned}$$

e concluímos, portanto

$$\Rightarrow \bar{\mathbf{w}}_+ = \bar{\mathbf{w}} + \mathbf{C} (\partial_{\bar{\mathbf{w}}} \ln Z)^T = \bar{\mathbf{w}} + \mathbf{C} (\partial_{\bar{\mathbf{w}}} \ln \langle P(\tau|\mathbf{x}\mathbf{w}) \rangle_{\mathbf{w}})^T \quad (2.6)$$

com a notação  $\langle \dots \rangle_{\mathbf{z}} = \int d\mathbf{z} P(\mathbf{z}) [\dots]$ .

Seguindo passos análogos a (2.5) e (2.6), a minimização da equação (2.4) relativa a  $\mathbf{C}_+$  leva a

$$\begin{aligned}
\mathbf{C}_+ &= \mathbf{C} + \mathbf{C} \partial_{\bar{\mathbf{w}}} (\partial_{\bar{\mathbf{w}}} \ln Z)^T \mathbf{C} \\
&= \mathbf{C} + \mathbf{C} \partial_{\bar{\mathbf{w}}} (\partial_{\bar{\mathbf{w}}} \ln \langle P(\tau|\mathbf{x}, \mathbf{u} + \bar{\mathbf{w}}, \mathbf{C}) \rangle_{\mathbf{u}})^T \mathbf{C}
\end{aligned} \quad (2.7)$$

e com isso, estabelecemos a dinâmica de aprendizado Bayesiano *online*, dada pelas equações (2.6) e (2.7), a menos da determinação da verossimilhança  $P(\tau|\mathbf{x}\mathbf{w})$ .

A escolha da forma da verossimilhança está diretamente associada com a *arquitetura* da máquina que produz os exemplos, neste caso, o professor  $\mathbf{w}$ . Sejam  $h = \frac{1}{\sqrt{K}} \mathbf{x}^T \mathbf{w}$  e  $f(h) = \text{sgn}(h)$  a classificação gerada por  $\mathbf{w}$  para o assunto  $\mathbf{x}$ , a resposta  $\tau$  do professor para a questão será dada pela distribuição

$$\begin{aligned}
P(\tau|h) &= \varepsilon \Theta(\tau f(h)) + (1 - \varepsilon) \Theta(-\tau f(h)) \\
&= \varepsilon + (1 - 2\varepsilon) \Theta(\tau f(h)) \\
&= \varepsilon + (1 - 2\varepsilon) \Theta(\tau h)
\end{aligned} \quad (2.8)$$

onde  $\varepsilon$  é a probabilidade de trocar o sinal da classificação  $f(h)$ . Neste caso, temos a verossimilhança dependente apenas de  $h$ , de modo que

$P(\tau|\mathbf{x}\mathbf{w}) = \int dh \delta\left(h - \frac{1}{\sqrt{K}}\mathbf{x}^T\mathbf{w}\right) P(\tau|h)$ , e podemos escrever

$$\begin{aligned}\langle P(\tau|\mathbf{x}\mathbf{w}) \rangle_{\mathbf{w}} &= \int d\mathbf{w} P(\tau|\mathbf{x}\mathbf{w}) G(\mathbf{w}|\theta) \\ &= \int d\mathbf{w} \int dh \delta\left(h - \frac{1}{\sqrt{K}}\mathbf{x}^T\mathbf{w}\right) P(\tau|h) G(\mathbf{w}|\theta) \\ &= \int dh \int d\mathbf{w} G(\mathbf{w}|\theta) \delta\left(h - \frac{1}{\sqrt{K}}\mathbf{x}^T\mathbf{w}\right) P(\tau|h) \\ &= \int dh P(h|\mathbf{x}\theta) P(\tau|h)\end{aligned}\quad (2.9)$$

Para determinar a distribuição  $P(h|\mathbf{x}\theta)$ , basta escrever

$$\delta\left(h - \frac{1}{\sqrt{K}}\mathbf{x}^T\mathbf{w}\right) = \int_{-\infty}^{\infty} dt \frac{1}{2\pi} e^{it\left(h - \frac{1}{\sqrt{K}}\mathbf{x}^T\mathbf{w}\right)}$$

e calcular integral completando a forma quadrática de modo a obter uma integral gaussiana, levando ao resultado

$$P(h|\mathbf{x}\theta) = \frac{1}{\sqrt{2\pi\lambda^2}} e^{-\left(\frac{h-\bar{h}}{\lambda}\right)^2} \quad (2.10)$$

com  $\lambda^2 = \frac{1}{K}\mathbf{x}^T\mathbf{C}\mathbf{x}$  e  $\bar{h} = \frac{1}{\sqrt{K}}\mathbf{x}^T\bar{\mathbf{w}}$ .

Substituindo (2.10) em (2.9), obtemos a *classificação Bayesiana* estimada

$$\langle P(\tau|\mathbf{x}\mathbf{w}) \rangle_{\mathbf{w}} = \varepsilon + (1 - 2\varepsilon) H\left(-\frac{\tau\bar{h}}{\lambda}\right) \quad (2.11)$$

onde usamos a definição

$$H(z) = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} dt e^{-\frac{1}{2}t^2}$$

Para finalizar, façamos duas hipóteses relacionadas ao conteúdo de  $\mathbf{w}$  e de  $\mathbf{x}_a$  que ajudarão interpretar os resultados obtidos. Considere que as características representadas por  $\mathbf{w}$  sejam independentes entre si, de modo que  $\mathbf{C} = \gamma^2\mathbb{1}$  e que todos os vetores exemplo satisfazem  $\mathbf{x}_a^T\mathbf{x}_a = K$ . Com isso, temos  $\lambda^2 = \gamma^2$  e usando o resultado (2.11), podemos reescrever as equações (2.6) e (2.7) da seguinte forma <sup>7</sup>:

$$\bar{\mathbf{w}}_+ = \bar{\mathbf{w}} - \frac{\mathbf{x}}{\sqrt{K}} \partial_{\bar{h}} V(\bar{h}|\gamma, \tau, \varepsilon) \quad (2.12)$$

$$\gamma_+^2 = \gamma^2 - \gamma^2 \partial_{\bar{h}} \partial_{\bar{h}} V(\bar{h}|\gamma, \tau, \varepsilon) \quad (2.13)$$

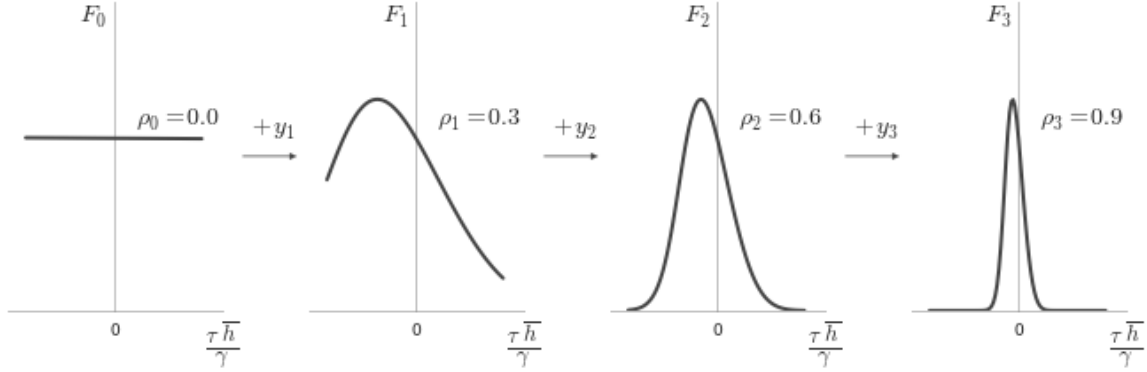
que definem as equações de aprendizado Bayesiano online. Nas equações acima, definimos implicitamente a função

$$\begin{aligned}V(\bar{h}|\gamma, \tau, \varepsilon) &= -\gamma^2 \ln \langle P(\tau|\mathbf{x}\mathbf{w}) \rangle_{\mathbf{w}} \\ &= -\gamma^2 \ln \left[ \varepsilon + (1 - 2\varepsilon) H\left(-\frac{\tau\bar{h}}{\gamma}\right) \right]\end{aligned}\quad (2.14)$$

<sup>7</sup> usando também a regra da cadeia

$$\partial_{\bar{\mathbf{w}}} f(\bar{h}) = \frac{1}{\sqrt{K}} \mathbf{x}^T \partial_{\bar{h}} f(\bar{h})$$

para enfatizar a solução das equações como uma descida pelo gradiente da função energia  $V(\bar{h}|\gamma, \tau, \varepsilon)$ . No contexto de aprendizado, a energia é a taxa de erro de classificação atingida pelo aluno. No contexto de aprendizado social, esse custo estará vinculado à discordância entre dois agentes através da opinião sobre um dado assunto, e a minimização dessa energia estará associada a diminuição das discordâncias.



O gradiente da função  $V(\bar{h}|\gamma, \tau, \varepsilon)$  pode ser visto como uma *força* que atua sobre  $\bar{w}$  na direção que aproxima o aluno do professor, composta pela direção  $\tau \mathbf{x}$  e uma amplitude  $F$ . Chamaremos essa amplitude de *Função de Modulação Bayesiana*, e para futuras referências

$$\begin{aligned} F(\bar{h}|\gamma, \tau, \varepsilon) &= -\tau \partial_{\bar{h}} V(\bar{h}|\gamma, \tau, \varepsilon) = \tau \gamma^2 \partial_{\bar{h}} \ln \left[ \varepsilon - (1 - 2\varepsilon) H \left( -\frac{\tau \bar{h}}{\gamma} \right) \right] \\ &= \frac{\gamma}{\sqrt{2\pi}} (1 - 2\varepsilon) \frac{\exp \left[ -\frac{1}{2} \left( \frac{\bar{h}}{\gamma} \right)^2 \right]}{\varepsilon + (1 - 2\varepsilon) H \left( -\frac{\tau \bar{h}}{\gamma} \right)} \end{aligned} \quad (2.15)$$

É possível mostrar<sup>8</sup> que a grandeza  $\gamma$  está relacionada com semelhança entre aluno e professor, dada por  $\rho = \frac{\mathbf{w}^T \bar{\mathbf{w}}}{\|\mathbf{w}\| \|\bar{\mathbf{w}}\|}$ , da seguinte forma

$$\gamma^2 = \frac{1 - \rho^2}{\rho^2} \quad (2.16)$$

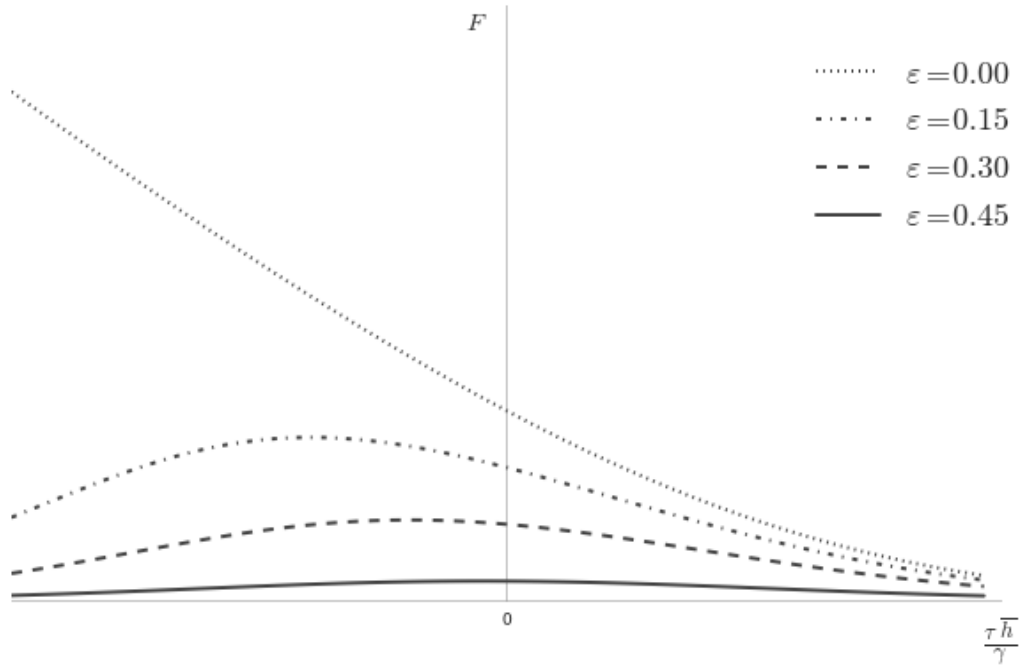
Essa relação facilita a compreensão da evolução de  $C$  ao longo do aprendizado, associando a adaptação na função de modulação diretamente ao desempenho do aluno condizente com os acertos relativos ao professor. Tal adaptação da função de modulação pode ser descrita como um ajuste da relevância dada pelo aluno ao conteúdo dos exemplos de acordo com a experiência obtida até então.

Figura 2.1: Evolução da função de modulação  $F_t$  ao longo da apresentação de exemplos  $y_t$ , para  $t = 1, 2, \dots$ , paralelamente à evolução da semelhança  $\rho_t$  entre o aluno e o professor, para  $\varepsilon$  fixo. Valores de  $\frac{\tau \bar{h}}{\gamma}$  positivos ou negativos ocorrem quando o aluno classifica correta ou incorretamente o exemplo apresentado, o seu valor absoluto está associado com o grau de surpresa trazido pelo exemplo.

<sup>8</sup> Kinouchi 1996; Renato Vicente, Kinouchi e Caticha 1998.

O sinal de  $\frac{\tau \bar{h}}{\gamma}$  indica que o aluno classificou correta ou incorretamente o exemplo apresentado, enquanto seu valor absoluto indica o grau de *surpresa/tédio* que o exemplo causa no aluno. Para dar compreender o que se passa, olhemos para a grandeza  $\frac{\tau \bar{h}}{\gamma}$ . Essa interpretação vem do fato que  $\tau = \pm 1$  e, portanto,  $|\bar{h}|$  indica quanta certeza o aluno tem em sua resposta, e portanto, um erro quando há muita certeza gera um mais surpresa do que um acerto numa situação semelhante. Ao longo do aprendizado,  $\rho$  tende a aumentar, fazendo com que a função de modulação se adapte para dar mais relevância a exemplos que causem maiores surpresas.

A desconfiança do aluno sobre a classificação do professor, representada pela estimativa  $\varepsilon$  da probabilidade do professor estar errado, faz com que a função de modulação deduz a relevância de surpresas muito grandes, dando à função de modulação um caráter adaptativo algoritmos que aprendem correção de erros e aqueles que o fazem por busca de correlações. Uma ilustração do comportamento da função de modulação em relação parâmetro  $\rho$  e  $\varepsilon$  é dada pelas figuras 2.1 e 2.2



Com base no cenário estabelecido, fica evidente que ao fixarmos um valor de  $\rho$ , fazendo com que a equação (2.13) fique  $C_+ = C$ , a forma da função de modulação fica congelada. Isso é equivalente a fixar um al-

Figura 2.2: Comportamento da função de modulação  $F$  com respeito ao aumento da desconfiança do aluno sobre possíveis erros do professor, para  $\rho$  fixo. Valores de  $\frac{\tau \bar{h}}{\gamma}$  positivos ou negativos ocorrem quando o aluno classifica correta ou incorretamente o exemplo apresentado, o seu valor absoluto está associado com o grau de surpresa trazido pelo exemplo.

goritmo de aprendizado para o aluno que, na abordagem apresentada, é completamente determinado por  $\rho$  e  $\varepsilon$ . A escolha de congelar a dinâmica para  $\rho$ , ao menos no que diz respeito à modelagem de comportamento humano, é motivada estudos apontando que o desenvolvimento das áreas sociais do cérebro se desenvolvem no período da adolescência. Nos estudos de (Choudhury, Blakemore e Charman 2006; Blakemore 2008; Moriguchi et al. 2007) é estabelecida, através de análises *fMRI*, mudanças na atividade e na estrutura de áreas como o córtex pré-frontal, o córtex parietal e o córtex temporal superior, associados com o desenvolvimento da cognição social. Neste trabalho, teremos sempre fixado um valor de  $\rho$ , sinalizando que a estratégia de aprendizado dos agentes foi previamente estabelecida, numa analogia com agentes “adultos” do ponto de vista de cognição social.

A dedução acima foi feita com base na teoria de aprendizado de máquina e deixou de fora vários aspectos interessantes desse campo de estudo. Leitores interessados no assunto podem consultar (Hastie 1993; Engel e Den Broeck 2001) para abordagem completa e (Kinouchi 1996; Solla e Winther 1999; Oppen 1996) para aprofundar os conceitos apresentados nessa seção.

### 2.1.2 Um Modelo mais Simples de Aprendizado

A dedução acima nos proporciona um algoritmo de aprendizado supervisionado que maximiza o aprendizado por exemplo apresentado, mas essa vantagem é obtida pelo preço da interpretação do algoritmo, dificultada pela forma não muito amigável da função de modulação (2.15). Para tentar sanar esse problema, vamos abstrair as características da função de modulação ótima e tentar criar uma forma mais simples que as mantenha.

Primeiramente, note que a equação (2.15) como função da variável que mede  $\frac{-\tau h}{\gamma}$  a concordância entre as respostas do aluno e do professor pode ser vista como uma atribuição de pesos para cada grau de concordância. Em outras palavras, a função de modulação é responsável pela diferença na relevância dada a exemplos que corroboram a classificação do aluno e aqueles que o surpreendem, com  $\frac{-\tau h}{\gamma}$  respectivamente positivos e negativos. Essa é a principal característica dos algoritmos de aprendizado e, portanto, a principal forma de caracterizá-los. Por exemplo, os algoritmos tradicionais como o de *Hebb* ou o *Perceptron* tem como única distinção o fato do primeiro dar mesmo valor a todos os exemplos e o segundo dar valor apenas aos que trazem novidade, que surpreendem.

No presente contexto, a forma de ajustar a relevância dada à exemplos com carga corroborativa ou surpreendente é fixando os valores do estilo cognitivo  $\rho$  e da desconfiança  $\varepsilon$ . Podemos pensar uma forma sim-



plificada,  $F_{MF}$ , para a função de modulação (2.15) que tenha essas características de regular o peso entre corroboração e novidade através dos parâmetros  $\rho$  e  $\varepsilon$ . Será útil também definir um custo cognitivo simplificado,  $V_{MF}$ , associado ao erro na classificação.

$$F_{MF}(z|\rho, \varepsilon, z_0) = \frac{1-\rho}{2} - \frac{\rho}{2}\text{sgn}(z) + \frac{1}{2}\text{sgn}(z+z_0) \quad (2.17)$$

$$V_{MF}(z|\rho, \varepsilon, z_0) = -\frac{1-\rho}{2}z + \frac{\rho}{2}|z| - \frac{1}{2}|z+z_0| \quad (2.18)$$

onde as grandezas  $z$  e  $z_0$  estão relacionadas com o parâmetro de concordância  $\frac{-\tau h}{\gamma}$ .

Para entender as grandezas  $z$  e  $z_0$ , vamos olhar para o gráfico das funções (2.17), ilustrado na figura 2.3, e associá-las às características descritas acima.

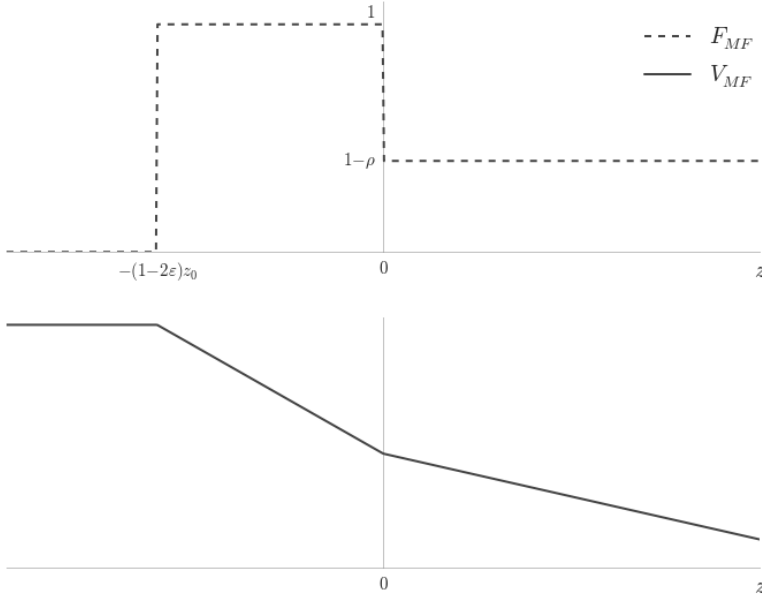


Figura 2.3: Ilustração das função aproximadas  $F_{MF}$  e  $V_{MF}$

Como discutido acima,  $F_{MF}$  regula a relevância entre óbvio e novidade, logo  $z$  deve ser uma função de  $\frac{-\tau h}{\gamma}$ . O parâmetro  $z_0$  é um valor de discordância que indica um conflito tão grande com o conhecimento do aluno que faz com que ele a ignore, ou seja é o ponto a partir do qual o aluno atribui a informação daquele exemplo como 'besteira' e não o absorve. É evidente, olhando para a função de modulação Bayesiana (2.15), que  $z_0$  deve ser uma função de  $\varepsilon$  e  $\rho$ .

Para determinar o comportamento de  $z_0$ , basta otimizar o erro quadrático entre  $F_{MF}$  e  $F$  em relação ao valor de  $z$ , que levará ao valor  $z$

para o qual  $F$  tem a metade da altura máxima, que será o valor de  $z_0$ . Alternativamente, é possível considerar  $z_0$  como um grau de liberdade a mais no modelo e estudar sua influência no comportamento do campo médio.

## 2.2 Modelos de Agentes e Mecânica Estatística

### 2.2.1 Estabelecendo a Linguagem do Modelo de Agentes

A hipótese central deste estudo de fenômenos sociais é que as propriedades de uma sociedade surgem da interação entre os indivíduos que a integram. Essa forma de abordar o problema tem a vantagem de requerer apenas a compreensão do que é um indivíduo e como ele atua com outros semelhantes a ele, ou seja, é necessária apenas uma compreensão local do sistema para inferir algumas características globais. É claro que essa estratégia pode ser criticada por um excesso de simplificação ao desconsiderar características particulares, tanto no que diz respeito ao indivíduo quanto a suas relações. Entretanto, é necessário ter em mente que o objetivo não é detalhar o resultado de cada interação possível entre indivíduos, mas sim tentar reproduzir o comportamento macroscópico que observamos em diversas culturas e, se possível, sua relação com aspectos intrínsecos da natureza humana.

Para elaborar um modelo de agentes capaz de descrever fenômenos sociais atribuídos a cultura, moral ou estratégias políticas, entre outros fenômenos relacionados a comportamento aprendido, precisamos entender ou, no mínimo emular, como se dá o processo de aprendizado social. Considere um conjunto de vértices  $\mathcal{V} = \{1, \dots, N\}$  e um conjunto de vetores  $\mathcal{W} = \{\mathbf{w}_i | i \in \mathcal{V}\}$ , representando *agentes* que interagem através da troca de informações e aprendem uns com os outros, de acordo com a estrutura das *relações sociais* estabelecidas em  $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$  denotando  $(ij) \in \mathcal{E}$  quando os agentes  $i$  e  $j$  são parceiros sociais<sup>9</sup>. Cada agente é representado por vetor  $\mathbf{w}_i \in \mathbb{R}^K$  que representa sua experiência<sup>10</sup> e um índice  $i \in \mathcal{V}$ , e pode receber a *opinião* de um agente com quem se relaciona sobre um *assunto* ou *questão*  $\mathbf{x} \in \mathbb{R}^K$  que tem alguma relação semântica com os vetores  $\mathbf{w}$ .

A interação entre dois agentes é regulada pelo custo cognitivo  $V$ , atribuído ao processo de aprendizado da *opinião*  $h = \frac{1}{\sqrt{K}} \mathbf{w} \cdot \mathbf{x}$  de um agente pelo outro. Considerando a soma do custo cognitivo sobre todos os pares de agentes no grafo temos o custo social total

$$\mathcal{H} = \sum_{(i,j) \in \mathcal{E}} J_{ij} V_{ij} \quad (2.19)$$

Chamaremos de *Sociedade de Agentes*, ou apenas *sociedade*, ao conjunto  $\mathcal{S} = (\mathcal{G}, \mathcal{W}, \mathcal{H})$ , onde  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  é o grafo das interações sociais.

<sup>9</sup> Note que as relações estabelecidas em  $\mathcal{E}$  não são, necessariamente, simétricas. É possível que um agente considere outro um parceiro sem reciprocidade

<sup>10</sup> num contexto determinado pelo fenômeno social em estudo

A natureza dos vetores  $\mathcal{W}$  depende do fenômenos em estudo. No caso de aprendizado moral, como em (César 2014; R. Vicente et al. 2014), o vetor  $\mathbf{w}_i$  representa a matriz moral do agente  $i$ , e o processo de aprendizado entre os agentes pode levar ou não a um consenso sobre como é o comportamento *ético* daquela sociedade. Por outro lado, numa escala mais ampla, como no modelo em (Axelrod 1997), os vetores  $\mathcal{W}$  podem representar características culturais de agrupamentos de pessoas, cenário no qual cada agente representa, digamos, uma vila e a troca de características culturais entre agentes representa a dinâmica de estruturas culturais, possibilitando o surgimento de fronteiras isolando diferentes culturas que, de alguma forma, se tornaram *incompatíveis*.

A topologia do grafo social, em particular o número de parceiros sociais, tem uma grande influência na possibilidade da sociedade experimentar uma transição de fases. Somado a isso, o fato das estruturas sociais dependerem da dinâmica entre os indivíduos cria uma dinâmica para as relações entre agentes. Esse efeito mútuo da influência do grafo nas características do indivíduo e vice-versa deve ser contemplado pelo modelo. Para isso, introduzimos a matriz de adjacência das relações sociais, com elementos  $R_{ij}$ . Estabeleceremos uma dinâmica para as relações sociais de modo que os valores de  $R_{ij}$  estejam entre 0 e 1, se anulando quando os agentes  $i$  e  $j$  deixam de interagir, ou seja quando  $(ij) \notin \mathcal{S}$ .

A determinação da função custo cognitivo está relacionada com a dinâmica de aprendizado supervisionado desenvolvido na seção 2.1 através de uma analogia. Cada agente na sociedade pode desempenhar tanto o papel de *aluno* quanto o de *professor*. Essa alternância de papéis é interpretada como uma sequência de diálogos entre agentes, nos quais ora um agente expressa sua opinião sobre uma questão, ora recebe a opinião de algum parceiro social. Na situação em que a opinião de um colega é recebida, o agente se comporta como um aluno, tentando aprender a reproduzir o comportamento do outro agente.<sup>11</sup>

Com base nas equações (2.12), (2.13) e (2.14), podemos estabelecer um *evento* envolvendo os parceiros sociais  $(ij) \in \mathcal{E}$  como uma relação *aluno/professor* na qual  $i$  aprende a opinião  $h_j = \frac{1}{\sqrt{K}} \mathbf{w}_j \cdot \mathbf{x}$  do agente  $j$  a respeito de um assunto  $\mathbf{x}$ . O desempenho de  $i$  com relação a essa tarefa é estimada pelo custo cognitivo  $V_{ij}(\mathbf{x}) = V(\mathbf{w}_i, \mathbf{w}_j, \mathbf{x})$ .

Como argumentado em (César 2014), pessoas passam por diferentes estágios de aprendizado ao longo vida. Quando mais jovens, as estratégias de aprendizado social estão se formando, e podemos associar a resposta de uma criança ao se defrontar com diversos exemplos com a evolução da função de modulação ilustrada na figura 2.1. Ao passo que novidade é encontrada, a criança altera, de forma inconsciente, a relevância relativa dada a exemplos surpreendentes ou corroborativos. Entretanto, após uma certa idade o modo como as pessoas aprendem

<sup>11</sup> É possível impor um comportamento antagonista, no qual um agente ativamente caminha no sentido oposto ao colega locutor. A analogia se mantém considerando que o professor seria um agente com a orientação oposta à do agente que expressa a opinião.

fica ‘congelado’, sendo associado a um algoritmo de aprendizado fixo no contexto de aprendizado de máquinas. Em termos das equações da dinâmica de aprendizado (2.12) e (2.13), isso equivale à fazer  $\gamma_i$  constante, e por consequência  $\rho_i$  constante. Desse modo, fixando  $\rho_i$ , nossa analogia trata de agentes ‘adultos’ no aspecto de aprendizado social. Com base nos argumentos dados em (Nestor Caticha e Renato Vicente 2011), a tendência de associação entre indivíduos mais parecidos nos estimula a fixar um único  $\rho$  para grupos de agentes parceiros ou mesmo para toda uma sociedade. Com essas considerações, podemos escrever um custo cognitivo e a função de modulação para um evento entre os parceiros  $i$  e  $j$  da seguinte maneira

$$V_{ij}(\mathbf{x}) = -\gamma^2 \ln \left[ \varepsilon + (1 - 2\varepsilon) H \left( -\frac{\tau_j h_i}{\gamma} \right) \right] \quad (2.20)$$

$$F_{ij}(\mathbf{x}) = \frac{\gamma(1 - 2\varepsilon)}{\sqrt{2\pi}} \frac{\exp \left[ -\frac{1}{2} \left( \frac{h_i}{\gamma} \right)^2 \right]}{\varepsilon + (1 - 2\varepsilon) H \left( -\frac{\tau_j h_i}{\gamma} \right)} \quad (2.21)$$

onde  $\tau_j = \text{sgn}(h_j)$  e  $\gamma^2 = \frac{1-\rho^2}{\rho^2}$ .

O parâmetro  $\varepsilon$ , introduzido no contexto do aprendizado de máquina como uma estimativa do aluno para a probabilidade de receber uma informações equivocadas do professor, desempenha o papel de uma *desconfiança* no contexto de aprendizado social, possibilitando a rejeição de opiniões muito dissonantes. A interpretação da informação trazida por exemplos através de surpresa ou corroboração para o aluno no cenário do aprendizado de máquina é traduzido no contexto de aprendizado social para uma dicotomia entre *concordância* e *discordância* das opiniões. Dessa forma, fica clara uma interpretação do custo cognitivo como um preço a ser pago pela discordância entre dois agentes e o custo social,  $\mathcal{H}$ , como uma energia necessária para sustentar uma sociedade com um certo grau de diversidade nas opiniões dos agentes.

É evidente, pela forma do potencial  $V_{ij}$ , que embora opiniões ‘absurdas’ do agente  $j$  sejam ignoradas pelo agente  $i$ , o mínimo global de  $\mathcal{H}$  estará mais próximo do estados de  $\mathcal{W}$  que desempenham o maior valor possível de concordância entre todos os agentes, caracterizando um consenso global quando o agentes trocam informação sobre uma questão apenas, como verificado em (Nestor Caticha e Renato Vicente 2011).

Para sustentar a coexistência de diversidade entre agentes com as mesmas características cognitivas, a saber  $\rho$  e  $\varepsilon$ , é necessário criar um mecanismo que reduz o custo pago em presença de discordância, aliviando a pressão social. Isso é feito dando a cada agente o poder de construir ou destruir suas relações sociais por meio das opiniões recebidas, através do controle das relações sociais  $R_{ij}$ . A dinâmica das relações

sociais é definida da seguinte forma

$$R_{ij}^+ = (1 - \varphi)R_{ij} + \lambda R_{ij}(1 - R_{ij})\text{sgn}(h_i h_j) \quad (2.22)$$

onde  $R_{ij}^+$  é a relação entre  $i$  e  $j$  após um evento de interação,  $\lambda$  uma parâmetro que controla a ‘euforia’ no ajuste da relação e  $\varphi$  um parâmetro que controla o ‘esquecimento’ dessa relação. A grandeza  $J_{ij} = J(R_{ij})$  que aparece na equação (2.19) está relacionada, de alguma forma a ser escolhida de acordo com o contexto específico do fenômeno estudado, com as relações sociais  $R_{ij}$  e tornará possível a regulação do comportamento de aproximação ou rejeição relativo a uma opinião. Note que as relações sociais se intensificam caso os agentes concordem num evento e são reduzidas caso contrário, fazendo o papel de uma espécie de registro da taxa de concordância do agente  $i$  com as opiniões do agente  $j$ .

### 2.2.2 Mecânica Estatística de Sistemas Sociais

Do que foi discutido na elaboração da linguagem que usaremos ao lidar com fenômenos sociais, ficam claros dois pontos importantes do modelo: sua natureza estatística e a importância delegada ao custo social,  $\mathcal{H}$ . Para avançar, faremos uso de métodos típicos da mecânica estatística, possibilitando a compreensão de alguns termos usado de forma vaga anteriormente, como ‘consenso’ ou ‘pressão’ e providenciando algumas previsões de comportamento do modelo.

Primeiramente, vamos determinar a probabilidade  $P(\mathcal{W})$  de termos uma sociedade  $\mathcal{S} = (\mathcal{W}, \mathcal{G}, \mathcal{H})$  num determinado estado  $\mathcal{W}$  com uma configuração social  $\mathcal{G}$  fixa. Dos argumentos apresentados, somos levados a esperar um algum valor para a função custo, ou seja devemos ter  $P(\mathcal{W})$  tal que  $\langle \mathcal{H} \rangle = E \in \mathbb{R}$ , onde  $\langle \dots \rangle$  denota o valor esperado sobre  $P(\mathcal{W})$ . Embora não tenhamos o valor  $E$  para cada estado, essa expectativa indica nosso interesse na avaliação da função  $\mathcal{H}$ .

A determinação de  $P(\mathcal{W})$  é feita através da maximização da entropia relativa dela com alguma à priori  $Q(\mathcal{W})$ , que tomamos como uma distribuição uniforme em  $\mathcal{W}$ <sup>12</sup>, sujeita ao vínculo de valor esperado do custo social. A entropia relativa entre  $P$  e  $Q$  é dada por

$$\begin{aligned} S[P(\mathcal{W}) : Q(\mathcal{W})] &= - \int d\mathcal{W} P(\mathcal{W}) \ln \frac{P(\mathcal{W})}{Q(\mathcal{W})} \\ &\quad + \beta \left[ E - \int d\mathcal{W} P(\mathcal{W}) \mathcal{H} \right] \end{aligned} \quad (2.23)$$

Igualando a zero derivada funcional de  $S$  relativa a  $P$ , notando que  $Q$  é uma constante, obtemos a forma a distribuição  $P$  que maximiza a entropia, a saber a distribuição de Boltzmann, dada por

$$P(\mathcal{W}) = \frac{1}{Z} \exp(-\beta \mathcal{H}[\mathcal{W}]) \quad (2.24)$$

<sup>12</sup> Essa escolha é guiada pelo princípio da máxima ignorância e pelo fato da distribuição à priori carregar o mínimo de informação a respeito de um conjunto de variáveis, a saber apenas os possíveis valores que elas podem tomar.

onde  $Z$  é a função de partição

$$Z = \int dW \exp(-\beta \mathcal{H}[W]) \quad (2.25)$$

O parâmetro  $\beta$ , introduzido como um multiplicador de *Lagrange* associado ao vínculo do valor esperado do custo social, pode ser interpretado analisando a distribuição  $P$  obtida. Note que, supondo um valor fixo de  $\beta \neq 0$ , estados com custos sociais mais elevados se tornam menos prováveis de acordo com  $P(W)$ . Isso implica que a evolução da sociedade sobre a dinâmica definida por  $\mathcal{H}$  deve seguir na direção de reduzir os custos sociais. Neste caso,  $\beta$  é uma espécie de *pressão*, determinando a escala de flutuações do custo social relativo ao valor esperado  $E$ .

A natureza da pressão  $\beta$  depende em parte do contexto, embora seja claro que ela está relacionada com a pressão sobre cada agente perante a exibição de um comportamento ‘*transgressor*’ relativo a seus parceiros, e por esse motivo a chamaremos de *pressão social*<sup>13</sup>. Essa interpretação nos leva a crer que, fixadas as relações sociais, ou seja fazendo  $R_{ij}$  constantes, e dada suficiente pressão encontraremos a sociedade apenas em estados de *consenso* de opinião a respeito da questão colocada. Para testar essa intuição, seguiremos com um estudo de campo médio visando estabelecer as condições em que tal situação de consenso pode ocorrer.

Considere uma sociedade  $\mathcal{S}_0 = (W, \mathcal{G}, \mathcal{H}_0)$ , com as relações sociais  $\mathcal{G}$  fixadas a função custo social  $\mathcal{H}_0 = \sum J_{ij} V_{ij}^0$  com

$$V_{ij}^0 = -\frac{1-\rho}{2} h_i h_j + \frac{\rho}{2} |h_i h_j| - \frac{1}{2} |h_i h_j + \eta| \quad (2.26)$$

uma aproximação do custo cognitivo, como destacado na seção 2.1 com a equação (2.18). Os agentes são descritos por vetores  $\mathbf{w}_i \in \mathbb{R}^K$  e trocam opiniões sobre um único assunto  $\mathbf{x} \in \mathbb{R}^K$ . Vamos supor que  $|\mathbf{w}_i|^2 = |\mathbf{x}|^2 = K$  para todo  $i$  e chamar  $h_i = \frac{1}{\sqrt{K}} \mathbf{w}_i \cdot \mathbf{x}$ , de modo que  $-\sqrt{K} \leq h_i \leq \sqrt{K}$  e  $\eta = (1 - 2\varepsilon)K$ . Essa escolha para o valor de  $\eta$  está relacionada com o parâmetro  $z_0$  comentado na seção 2.1.2, tratando-o como um grau de liberdade extra, e pode ser encarada como um modelo para o comportamento de  $\eta$  em termos de  $\varepsilon$ . Como veremos, a influência de  $\eta$  para um valor linear em  $\varepsilon$  não afeta as fases observadas no campo médio, dissonando do resultado obtido para o potencial Bayesiano.

Para possibilitar o tratamento da função de partição (2.25), além da introdução do potencial  $V^0$ , vamos supor que os  $J$  é homogêneo em  $\mathcal{G}$ , de modo que

$$J_{ij} = \begin{cases} J_0 > 0 & \text{se } i \text{ e } j \text{ são parceiros sociais} \\ 0 & \text{caso contrário} \end{cases}$$

e que  $\mathcal{G}$  é um grafo regular não direcionado, ou seja, cada agente tem o mesmo número  $n_0$  de parceiros e todas as relações são reciprocadas.<sup>14</sup>

<sup>13</sup> No contexto de concorrência partidária, daremos um nome mais sugestivo para a pressão, mas por enquanto, pressão social é suficiente para o entendimento do que segue.

<sup>14</sup> Essa hipótese é bem restritiva, mas não é artificial, dado que grafos sociais frequentemente apresentam uma estrutura quase regular como grafos de *mundo pequeno*, onde o grau médio de cada vértice é representativo globalmente. Esse pode não ser o caso em estruturas mais assimétricas como grafos de *Barabasi-Albert*. Mas em se tratando do campo médio, a estrutura do grafo seria descartada de todo modo, então essas consideração servem apenas para indicar as situações em

Com essas hipóteses adicionais, podemos nos questionar qual é a distribuição  $P_* = \prod_i^N P_i$ , sujeita ao vínculo do valor esperado do custo social, que melhor aproxima a distribuição  $P(\mathcal{W})$  por um sistema de agentes independentes? Para responder essa pergunta, usamos mais uma vez a maximização da entropia em relação às distribuições  $P_i$ . A entropia relativa entre  $P$  e  $P_*$  é

$$\begin{aligned} S[P_* : P] &= - \int d\mathcal{W} P_* \ln \frac{P_*}{P} + \beta \left[ E - \int d\mathcal{W} P_* \mathcal{H} \right] \\ &= - \sum_j^N \int d\mathbf{w}_j P_j \ln \frac{P_j}{P} \\ &\quad + \beta \left[ E - \sum_{(kj) \in \mathcal{G}} \int d\mathbf{w}_k d\mathbf{w}_j P_k P_j J_0 V_{kj}^0 \right] \end{aligned} \quad (2.27)$$

onde usamos implicitamente a normalização de cada  $P_j$ . Tomando a derivada funcional de  $S$  em relação a  $P_i$  e igualando a zero teremos, usando os fatos de  $P$  e  $E$  serem independentes de  $P_i$

$$\begin{aligned} 0 &= \frac{\delta S}{\delta P_i} = 1 - \ln P_i - \beta J_0 \sum_{j \in n(i)} \int d\mathbf{w}_j P_j V_{ij}^0 \\ \Rightarrow P_i &= \frac{1}{Z_i} \exp \left[ -\beta J_0 \sum_{j \in n(i)} \int d\mathbf{w}_j P_j V_{ij}^0 \right] \end{aligned} \quad (2.28)$$

denotando por  $n(i) = \{j | (ij) \in \mathcal{G}\}$  o conjunto dos parceiros sociais do agente  $i$ .

Para prosseguir, será necessário trabalhar com a integral

$$\begin{aligned} I_i &= \int d\mathbf{w}_j P_j V_{ij}^0 \\ &= \int d\mathbf{w}_j P_j \left[ -h_i h_j + \frac{\rho}{2} |h_i h_j| - \frac{1}{2} |h_i h_j + \eta| \right] \end{aligned} \quad (2.29)$$

que aparece no lado direito da equação (2.28). Somente nesse ponto a escolha do custo cognitivo (2.26) se justifica, tendo em vista que mesmo a definição dos parâmetros de ordem, no que seguirá, seria mais difícil, senão impossível, com o potencial (2.20). Note também que, o potencial (2.26) é uma forma um pouco mais geral daquele usado em (Nestor Caticha e Renato Vicente 2011), se reduzindo àquele no caso em que a desconfiança é nula, ou seja  $\varepsilon = 0$ . De fato, a análise de campo médio apresentada aqui segue os moldes da estabelecida em (Nestor Caticha e Renato Vicente 2011; César 2014).

Para determinar a distribuição  $P_i$ , definimos os parâmetros de ordem

$$m = \int d\mathbf{w}_j P_j \frac{1}{\sqrt{K}} \mathbf{w}_j \cdot \mathbf{x} \quad (2.30)$$

$$r = \int d\mathbf{w}_j P_j \left| \frac{1}{\sqrt{K}} \mathbf{w}_j \cdot \mathbf{x} \right| \quad (2.31)$$

e substitua em (2.29) para obter

$$I_i = -m \frac{1-\rho}{2\sqrt{K}} \mathbf{w}_i \cdot \mathbf{x} + r \frac{\rho}{2\sqrt{K}} |\mathbf{w}_i \cdot \mathbf{x}| - \frac{1}{2\sqrt{K}} |m \mathbf{w}_i \cdot \mathbf{x} + \eta| \quad (2.32)$$

Note que assumimos a homogeneidade dos parâmetros de ordem, ou seja, fizemos  $m_j = m$  e  $r_j = r$  para todo  $j$ . Isso não significa que os são idênticos, mas que são descritos de forma idêntica. Essa hipótese está relacionada com a homogeneidade das relações, explicitada pela constante  $J_0$ . Fizemos também a escolha de comutar a integral com o valor absoluto no termo que envolve a desconfiança para facilitar as contas e por manter a coerência, mas sem uma justificativa formal para isso. Substituindo (2.32) em (2.28) e lembrando que todos agentes têm o mesmo número de parceiros  $n_0$ , chegamos à conclusão

$$P_i = \frac{1}{Z_i} e^{-\beta J_0 n_0 I_i} \quad (2.33)$$

O resultado obtido para as probabilidade  $P_i$  nos leva traz um conjunto de equações que precisam ser resolvidas de forma autoconsistente, a saber

$$m = \frac{1}{Z} \int d\mathbf{w}_i \left( \frac{1}{\sqrt{K}} \mathbf{w}_i \cdot \mathbf{x} \right) e^{-\beta J_0 n_0 I_i(m,r)} \quad (2.34)$$

$$r = \frac{1}{Z} \int d\mathbf{w}_i \left| \frac{1}{\sqrt{K}} \mathbf{w}_i \cdot \mathbf{x} \right| e^{-\beta J_0 n_0 I_i(m,r)} \quad (2.35)$$

$$Z = \int d\mathbf{w}_i e^{-\beta J_0 n_0 I_i(m,r)} \quad (2.36)$$

A solução desse sistema de equações resulta na determinação das distribuições  $P_i$  para todo os agentes, e por consequência determina nossa aproximação de campo médio. Entretanto, a equação (2.33) nos dá a probabilidade  $P_i = P(\mathbf{w}_i)$  de encontrar um agente com um estado interno  $\mathbf{w}_i$  do agente. Essa situação é inconveniente por dois motivos, a saber, primeiramente, não temos acesso direto ao estado cognitivo de pessoas nos processos formadores de opinião e, também, porque estamos de fato interessado nas opiniões. Para resolver esse conflito entre o resultado (2.33) e sua praticidade, façamos

$$\begin{aligned} P(h) &= \int d\mathbf{w} \delta \left( h - \frac{1}{\sqrt{K}} \mathbf{w} \cdot \mathbf{x} \right) P(\mathbf{w}) \\ &= \frac{1}{C} (1 - h^2) e^{\beta J_0 n_0 \left[ \frac{1-\rho}{2} h m - \frac{\rho}{2} |h| r + \frac{1}{2} |h r + \eta| \right]} \end{aligned} \quad (2.37)$$



com a nova função de partição  $C$  dada por

$$C = \int_{-\sqrt{K}}^{\sqrt{K}} dt (1 - t^2) e^{\beta J_0 n_0 \left[ \frac{1-\rho}{2} mt - \frac{\rho}{2} r|t| + \frac{1}{2} |rt + \eta| \right]} \quad (2.38)$$

Os parâmetros ordem  $m$  e  $r$  são os valores esperados das opiniões e das suas ‘convicções’ sobre  $x$  em toda a sociedade. Com isso podemos estudar as condições definidas pela pressão social  $\beta$ , pelo estilo cognitivo  $\rho$  e pela desconfiança  $\varepsilon$  na formação de consenso através dos valores dos parâmetros de ordem. O comportamento típico do parâmetro  $m$ , que mede a média das opiniões em relação à questão dada, está ilustrado na figura 2.4.

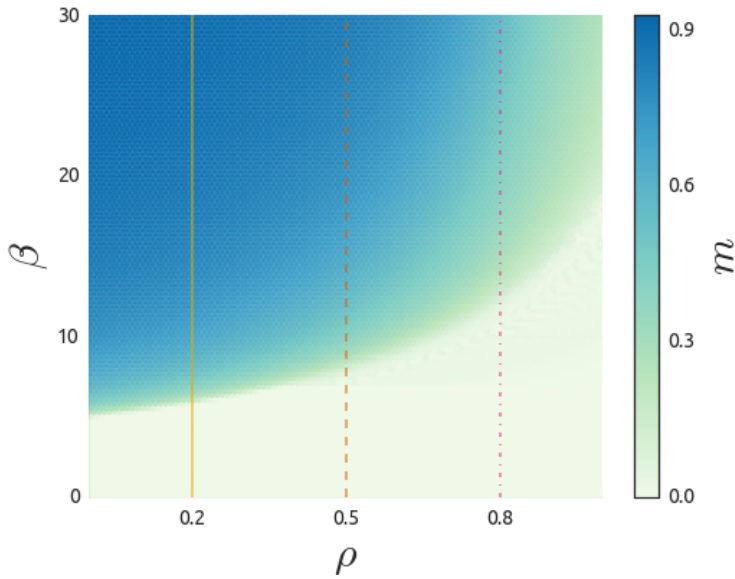


Figura 2.4: Diagrama de fases da solução  $m$  das equações de campo médio, no plano  $\beta \times \rho$ . Esse comportamento não se altera para baixos valores da desconfiança  $\varepsilon$  neste cenário de agentes similares em estratégia de aprendizado e motivados a cooperar.

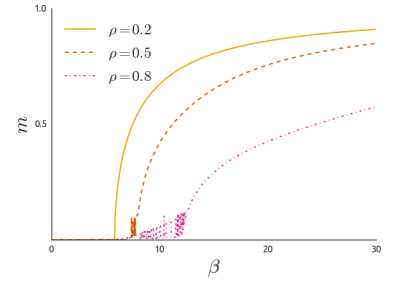


Figura 2.5: Curvas de consenso correspondentes às retas verticais na figura 2.4.

A figura 2.4 mostra que, seguindo uma das linhas verticais que fixa um valor de  $\rho$ , há uma transição de fase entre os estados de dissenso e consenso com relação às opiniões dos agentes a respeito da questão  $x$ . Sem se desprender do contexto, no cenário em que a concordância é estimulada, dado o parâmetro  $J_0 > 0$ , existirá uma pressão social crítica que levará essa sociedade ao consenso. Esse tipo de comportamento é que esperamos dentro de um grupo com alguma ideologia ou de agentes que compartilhem um certo conjunto de características não muito disperso, codificado nos vetores  $\mathbf{w}_i$ . Analogamente, situações onde a concordância seja muito relevante, talvez em condições de perigo à sociedade, é de se esperar que a sociedade atinja o consenso em detrimento da manutenção de diferenças.

Note, porém que maiores valores de  $\rho$  tendem a apresentar pressões

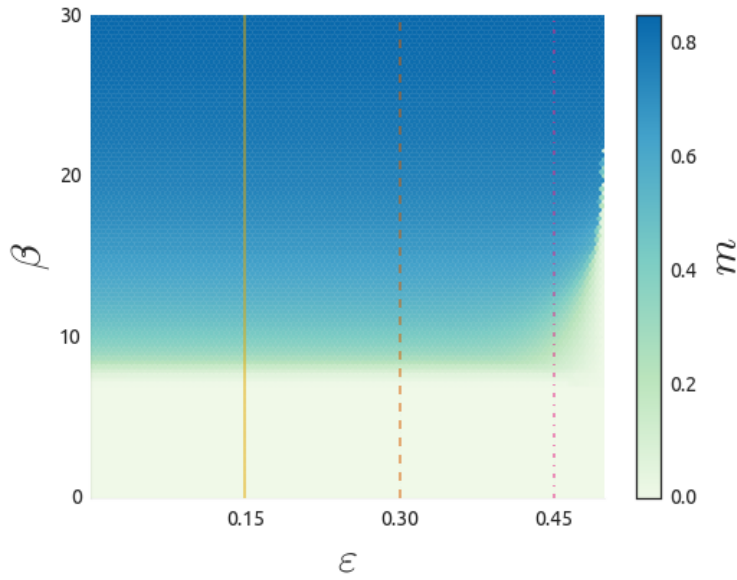


Figura 2.6: Diagrama de fases da solução  $m$  das equações de campo médio, no plano  $\beta \times \epsilon$ . Note como apenas valores de  $\epsilon$  próximos de 0.5 introduzem dificuldades ao surgimento de consenso. Isso significa que, em campo médio, apenas sociedade em que quase nenhuma confiança pode ser atribuída aos parceiros é possível impedir o consenso.

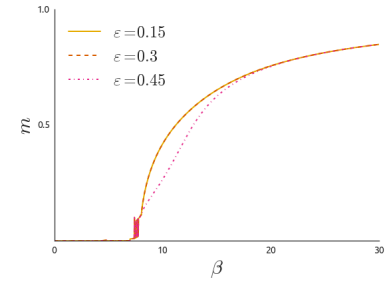


Figura 2.7: Curvas de consenso correspondentes às retas verticais na figura 2.6.

críticas mais elevadas. Podemos imaginar numa sociedade com diferentes valores de  $\rho$  a possibilidade de parte dos agentes estarem sob suficiente pressão social para obter consenso, mas outra parte não. Logo, a analogia com ameaças à sociedade nos levaria à associar essa variação a diferenças no que os agentes consideram ameaça ou não. Essa aparente dicotomia sugere uma relação entre os comportamentos liberal e conservador exibido pelas pessoas e uma análise mais detalhada desse relação pode ser encontrada em (Nestor Caticha e Renato Vicente 2011; R. Vicente et al. 2014; César 2014).

Uma propriedade interessante desse resultado é sua robustez perante mudanças na desconfiança  $\epsilon$ , como pode ser visto na figura 2.6. Essa condição nos diz relações internas de algum segmento social tendem a atingir consenso, mesmo com pouca confiança nas opiniões dos parceiros. Tendo em vista a diversidade de opiniões relativas a um dado assunto, essa conclusão é de fato irreal. O caso é que fixadas as relações sociais, a pressão dos pares eventualmente conduz ao consenso numa ‘vitória’ por resistência.

Vejamos um cenário em que uma sociedade tem dois grupos ‘rivals’ coexistindo. Nessa situação, cada agente  $i$  pode integrar um grupo  $A = \{1, \dots, N_A\}$  ou um grupo  $B = \{1, \dots, N_B\}$ . O custo cognitivo da interação entre os quaisquer dois agentes ( $ij$ ) é dado pela equação (2.26), porem as grandezas  $J_{ij}$  que regulam as interações sociais dependem dos

grupo de  $i$  e  $j$ . Considere

$$J_{ij} = \begin{cases} 1 & \text{se } i \text{ e } j \text{ estão no mesmo grupo} \\ -J & \text{com } J > 0 \text{ caso contrário} \end{cases} \quad (2.39)$$

de modo que o custo social pode ser escrito

$$\mathcal{H}_{AB} = \sum_{\substack{i \in A \\ j \in A}} V_{ij}^0 + \sum_{\substack{i \in B \\ j \in B}} V_{ij}^0 - J \sum_{\substack{i \in A \\ j \in B}} V_{ij}^0 \quad (2.40)$$

e podemos aplicar uma aproximação de campo médio análoga à feita para o cenário anterior.

Vamos aproximar a distribuição de *Boltzmann* para a sociedade com custo social  $\mathcal{H}_{AB}$  por uma distribuição  $P_{AB} = \prod_{i \in A} P_A(\mathbf{w}_i) \prod_{j \in B} P_B(\mathbf{w}_j)$ , novamente através da maximização da entropia. Seguindo essa estratégia, obteremos as seguintes formas para  $P_A(\mathbf{w}_i)$  e  $P_B(\mathbf{w}_j)$

$$P_A(\mathbf{w}_i) \propto e^{-\beta I_A(\mathbf{w}_i)} \quad (2.41)$$

$$P_B(\mathbf{w}_j) \propto e^{-\beta I_B(\mathbf{w}_j)} \quad (2.42)$$

com os termos  $I_A$  e  $I_B$  dados por

$$I_A(h_i) = \sum_{k \in A} \int d\mathbf{w}_k P_A(\mathbf{w}_k) V_{ik}^0 - J \sum_{l \in B} \int d\mathbf{w}_l P_B(\mathbf{w}_l) V_{il}^0 \quad (2.43)$$

$$I_B(h_j) = \sum_{l \in B} \int d\mathbf{w}_l P_B(\mathbf{w}_l) V_{jl}^0 - J \sum_{k \in A} \int d\mathbf{w}_k P_A(\mathbf{w}_k) V_{jk}^0 \quad (2.44)$$

e para lidar com as integrais que aparecem em  $I_A$  e  $I_B$ , definimos as seguintes grandezas

$$m_A = \frac{1}{Z_A} \int d\mathbf{w}_k e^{-\beta I_A(\mathbf{w}_k)} \frac{1}{\sqrt{K}} \mathbf{w}_k \cdot \mathbf{x} \quad (2.45)$$

$$r_A = \frac{1}{Z_A} \int d\mathbf{w}_k e^{-\beta I_A(\mathbf{w}_k)} \left| \frac{1}{\sqrt{K}} \mathbf{w}_k \cdot \mathbf{x} \right| \quad (2.46)$$

$$Z_A = \int d\mathbf{w}_k e^{-\beta I_A(\mathbf{w}_k)} \quad (2.47)$$

para o grupo  $A$  e analogamente para o grupo  $B$

$$m_B = \frac{1}{Z_B} \int d\mathbf{w}_l e^{-\beta I_B(\mathbf{w}_l)} \frac{1}{\sqrt{K}} \mathbf{w}_l \cdot \mathbf{x} \quad (2.48)$$

$$r_B = \frac{1}{Z_B} \int d\mathbf{w}_l e^{-\beta I_B(\mathbf{w}_l)} \left| \frac{1}{\sqrt{K}} \mathbf{w}_l \cdot \mathbf{x} \right| \quad (2.49)$$

$$Z_B = \int d\mathbf{w}_l e^{-\beta I_B(\mathbf{w}_l)} \quad (2.50)$$

que definem um conjunto de equações auto consistentes que determinam as probabilidades  $P_A$  e  $P_B$ .

Usando o mesmo truque da equação (2.37) obtemos as distribuições de opinião dos grupos  $A$  e  $B$ , a menos da determinação das equações

acima, a saber

$$P_A(h) = \frac{1}{Z_A} (1 - h^2) e^{-\beta U_A(h)} \quad (2.51)$$

$$P_B(h) = \frac{1}{Z_B} (1 - h^2) e^{-\beta U_B(h)} \quad (2.52)$$

com as funções de partição

$$Z_A = \int_{-\sqrt{K}}^{\sqrt{K}} dt (1 - t^2) e^{-\beta U_A(t)} \quad (2.53)$$

$$Z_B = \int_{-\sqrt{K}}^{\sqrt{K}} dt (1 - t^2) e^{-\beta U_B(t)} \quad (2.54)$$

e os logaritmos  $U_A$  e  $U_B$  dados por

$$\begin{aligned} U_A(h) = & -\frac{1-\rho}{2} (N_A m_A - J N_B m_B) h \\ & + \frac{\rho}{2} (N_A r_A - J N_B r_B) |h| \\ & - \frac{1}{2} (N_A |m_A h + \eta| - J N_B |m_B h + \eta|) \end{aligned} \quad (2.55)$$

$$\begin{aligned} U_B(h) = & -\frac{1-\rho}{2} (N_B m_B - J N_A m_A) h \\ & + \frac{\rho}{2} (N_B r_B - J N_A r_A) |h| \\ & - \frac{1}{2} (N_B |m_B h + \eta| - J N_A |m_A h + \eta|) \end{aligned} \quad (2.56)$$

A solução desse conjunto de equações auto consistentes nos leva à distribuição de opiniões no caso em que dois grupos antagônicos coexistem sem uma relação de dominância ou estrutura hierárquica entre eles, em outras palavras numa situação de pluralidade de opiniões. Para medir essa pluralidade, olhamos para o parâmetro de ordem ‘*antiferromagnético*’

$$m_s = \frac{m_A - m_B}{2},$$

que se anula quando os grupos  $A$  e  $B$  compartilham da mesma opinião e tem valores mais próximos de  $\pm 1$  de acordo com o quão opostas suas opiniões são. O comportamento dessa sociedade simples é ilustrado na figura 2.8

Note que o aumento da desconfiança influencia a convicção de cada grupo em sua própria opinião, de modo que  $m_s$  apresenta valores menores conforme a desconfiança  $\varepsilon$  aumenta. Isso nos indica que a desconfiança associada às relações sociais pode ser um mecanismo que possibilita a diversidade de opiniões numa sociedade.

Para explorar essa nova possibilidade, investigaremos no capítulo 3 dinâmicas para a formação de grupos dentro de uma sociedade e outros cenários em que a coexistência de grupos leva a comportamentos de interesse em eventos reais.

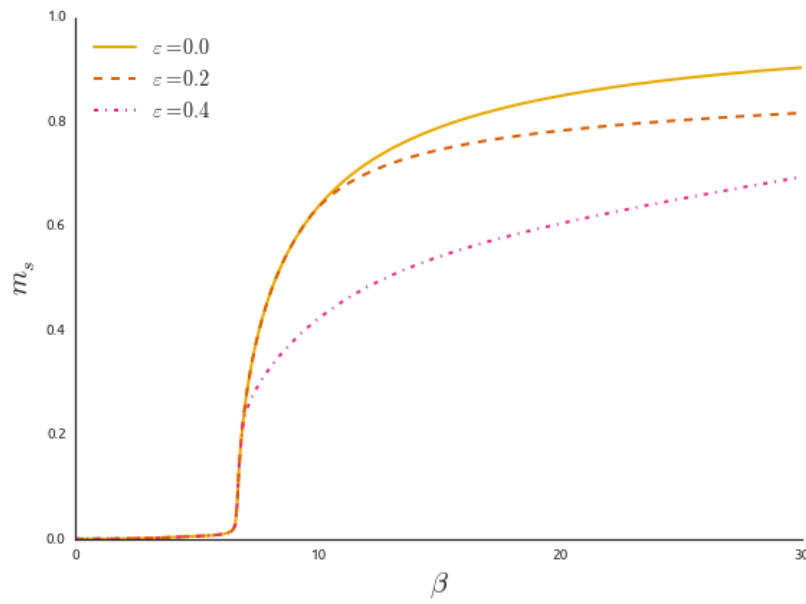


Figura 2.8: Curvas da pluralidade de opiniões  $m_s$  em função da pressão  $\beta$  para diferentes valores da desconfiança  $\varepsilon$  e valor fixo do algoritmo de aprendizado  $\rho = 0.5$ . Note que conforme a desconfiança aumenta, o valor absoluto de  $m_s$  cai, comportamento que se repete em  $m_A$  e  $m_B$ , indicando a dificuldade na formação das opiniões intragrupo em situações de desconfiança.



## Experimentos Computacionais e Análise dos Resultados

Apesar do esforço no sentido de simplificar ao máximo durante a elaboração do modelo para abordar relações sociais, a complexidade das equações da dinâmica não permitem um tratamento analítico além do desenvolvido no 2 com as soluções de campo médio em condições bem restritivas. Para explorar melhor as capacidades do modelo passamos agora a uma análise computacional da termodinâmica resultante da evolução de uma sociedade sob certas condições.

Para isso, começaremos descrevendo brevemente os métodos utilizados, sem detalhes de implementação, e seguiremos com a construção de cenários e o resultados neles obtidos. O tratamento apresentado neste capítulo supõe familiaridade com o algoritmo de *Metropolis* em Monte Carlo e com a termodinâmica de modelos magnéticos, como o de Ising.

### 3.1 Monte Carlo em Sociedades de Agentes e a Termodinâmica de Consenso

Para elaborar uma dinâmica de Monte Carlo para um sistema de agentes, vamos nos basear na forma como as relações ocorrem no cotidiano. Podemos imaginar um cenário em que duas pessoas se encontram, expõem suas opiniões, possivelmente aprendem algo e seguem suas vidas. Um passo ‘microscópico’ do algoritmo deve seguir esse roteiro, visando ser minimamente realista.

Para tentar formalizar essa sequência de eventos, consideremos  $\mathcal{S}$  uma sociedade com agentes  $\mathcal{W}$  em  $\mathbb{R}^K$ ,  $\mathcal{G}$  um grafo completo e custo cognitivo

$$V_{ij} = -\gamma^2 \ln \left[ \varepsilon + (1 - 2\varepsilon) H \left( -\frac{\tau_j h_i}{\gamma} \right) \right] \quad (2.20)$$

O encontro de dois agentes, num passo microscópico  $t$ , é emulado sorteando um agente  $i$  de maneira uniforme em  $\mathcal{S}$  e, em seguida, sorteando um de seus parceiros  $j$  com probabilidade  $p(j|i)$ . A conversa entre os agentes  $i$  e  $j$  pode resultar numa mudança no vetor cognitivo  $\mathbf{w}_i(t)$ , que ocorre através da proposta de um novo vetor  $\mathbf{v} \in \mathbb{R}^K$ . A proposta é

aceita com base na diferença do custo social entre o vetor proposto e o estado atual do agente  $i$ . De maneira mais precisa, a probabilidade da proposta ser aceita é dada por

$$p(\mathbf{w}_i(t) \rightarrow \mathbf{v}) = \min \left( 1, e^{-\beta J_{ij} \Delta V} \right), \quad (3.1)$$

com  $\Delta V = V(\mathbf{v}, \mathbf{w}_j) - V(\mathbf{w}_i(t), \mathbf{w}_j)$ , de modo que  $\mathbf{w}_i(t+1) = \mathbf{v}$  com probabilidade  $p(\mathbf{w}_i(t) \rightarrow \mathbf{v})$ . O termo  $J_{ij}$  pode depender da relação  $R_{ij}$  entre os agentes, mas vamos mantê-lo constante  $J_{ij} = \frac{1}{K}$ , por enquanto, para estudar o cenário de formação de consenso. Ao longo da simulação foi mantido o vínculo de normalização sobre os agentes, de forma que  $|\mathbf{w}_i(t)|^2 = K$  para todo  $t \in \{1, \dots, T\}$  e todo  $i \in \mathcal{S}$ , assim como sobre a questão  $|\mathbf{x}|^2 = K$ . Além disso, a cada parceiro  $j$  do agente  $i$  foi dada a mesma probabilidade de ser escolhido, ou seja

$$p(j|i) = \frac{R_{ij}}{\sum_k R_{ik}} \quad (3.2)$$

com  $R_{ij}$  constantes, neste cenário, e dados por

$$R_{ij} = \begin{cases} 1, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases} \quad (3.3)$$

Essa dinâmica de interação entre os agentes é repetida um número  $T = MN$  de vezes, onde  $N$  é o número de agentes em  $\mathcal{S}$  e  $M$  é o número médio de vezes que cada agente é escolhido para aprender com um parceiro. Em outras palavras,  $M$  é o número de passos ‘macroscópicos’, ou número de passos de Monte Carlo. Em cada passo macroscópico, são medidas a média das opiniões e seus valores absolutos e suas respectivas correlações

$$m = \frac{1}{N\sqrt{K}} \sum_{i=1}^N h_i \quad (3.4)$$

$$r = \frac{1}{N\sqrt{K}} \sum_{i=1}^N |h_i| \quad (3.5)$$

$$q = \frac{1}{N^2 K} \sum_{(ij)} R_{ij} h_i h_j \quad (3.6)$$

além de outras grandezas relacionadas ao desempenho do Monte Carlo, como as taxas de aceitação. Para mais detalhes sobre o método de Monte Carlo, recomendamos ao leitor o texto de (MEJ Newman e Barkema 1999)

### 3.1.1 As Fases de Uma Sociedade

Os resultados das simulações dessa sociedade nos dizem quais as condições de pressão social  $\beta$ , estilo de aprendizado  $\rho$  e desconfiança  $\varepsilon$  nas



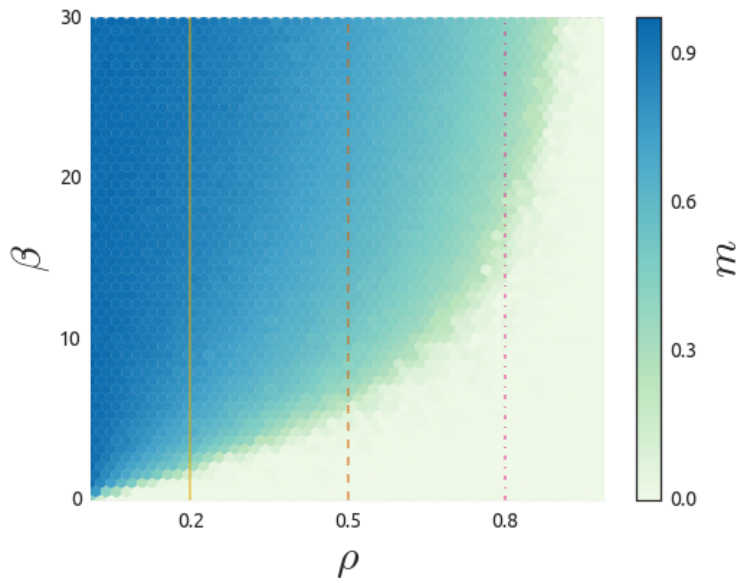


Figura 3.1: Diagrama de fases do consenso  $m$  no plano  $\beta \times \rho$ , para desconfiança  $\varepsilon = 0$ .

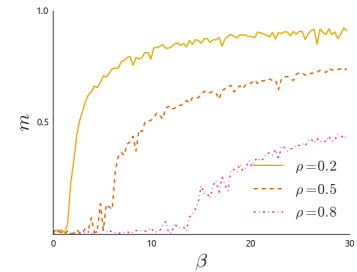


Figura 3.2: Curvas de consenso correspondentes às retas verticais na figura 3.1.

quais consenso podem surgir. O diagrama de fases na figura 3.1 mostra as fases de consenso numa sociedade com desconfiança  $\varepsilon = 0$ . Como visto na discussão do capítulo 2, existe uma linha de transição entre as fases de desordem e consenso na sociedade e, ao menos qualitativamente, os resultados nas figuras 2.4 e 3.1 são equivalentes.

A diferença que surge na forma da fronteira que delimita as regiões vem do fato que o algoritmo de aprendizado definido por (2.20), a saber a função de modulação Bayesiana, é muito mais eficiente do que aquele oriundo da aproximação de campo médio (2.26). Por diferença na eficiência de um algoritmo de aprendizado entende-se que um número menor de exemplos precisa ser apresentado ao aluno para que ele aprenda a regra do professor. No contexto de aprendizado social, isso significa que dois parceiros precisam interagir um número menor de vezes para que cheguem a uma opinião comum, e isso se reflete numa pressão crítica menor para que a sociedade atinja o consenso. Essa é a exata diferença entre as figuras 2.4 e 3.1.

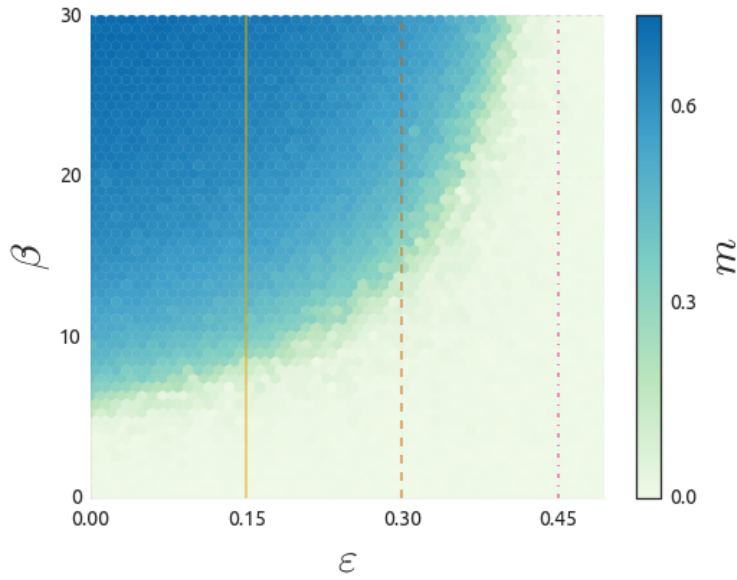


Figura 3.3: Diagrama de fases do consenso  $m$  no plano  $\beta \times \epsilon$ , para desconfiança  $\rho = 0.5$ .

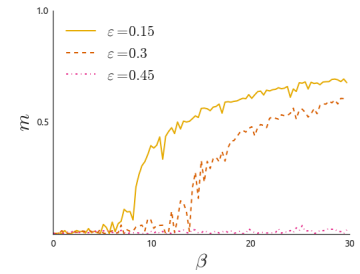


Figura 3.4: Curvas de consenso correspondentes às retas verticais na figura 3.3.

Outra diferença que observamos entre os resultados da simulação de Monte Carlo e o resultado do campo médio é a dependência da pressão crítica com relação ao parâmetro de desconfiança  $\epsilon$ . De fato, na simulação de Monte Carlo, o papel da desconfiança no surgimento de consenso é similar ao estilo cognitivo  $\rho$ , dificultando o consenso conforme seu valor aumenta. Essa característica não foi capturada pela aproximação de campo médio, o que explica as diferenças entre as figuras 2.6 e 3.3. Entretanto, a previsão de que apenas a desconfiança não é capaz de produzir grupos de opiniões distintas é corroborada, devido tamanha similaridade entre  $\rho$  e  $\epsilon$  no aparecimento de consenso, o que justifica nossa investida em mecanismos alternativos para tentar entender esse fenômeno.

### 3.1.2 Um Retrato (Estatístico) da Semelhança Entre Agentes

Outra forma interessante de ver o fenômeno de consenso é através dos histogramas de similaridade  $\psi_{ij}$  definida por

$$\psi_{ij} = \frac{1}{K} \mathbf{w}_i \cdot \mathbf{w}_j \quad (3.7)$$

que nada mais é do que o cosseno do ângulo formado pelos vetores cognitivos dos agentes  $i$  e  $j$ . De acordo com os diagrama nas figuras 3.1 e 3.3, a sociedade tem duas fases, uma desordenada e uma de consenso. Essas fases são reflexo da distribuição das opiniões a respeito de  $\mathbf{x}$ , que

estão diretamente relacionadas com  $\psi_{ij}$ , e portanto devem ser visíveis nos histogramas de similaridade.

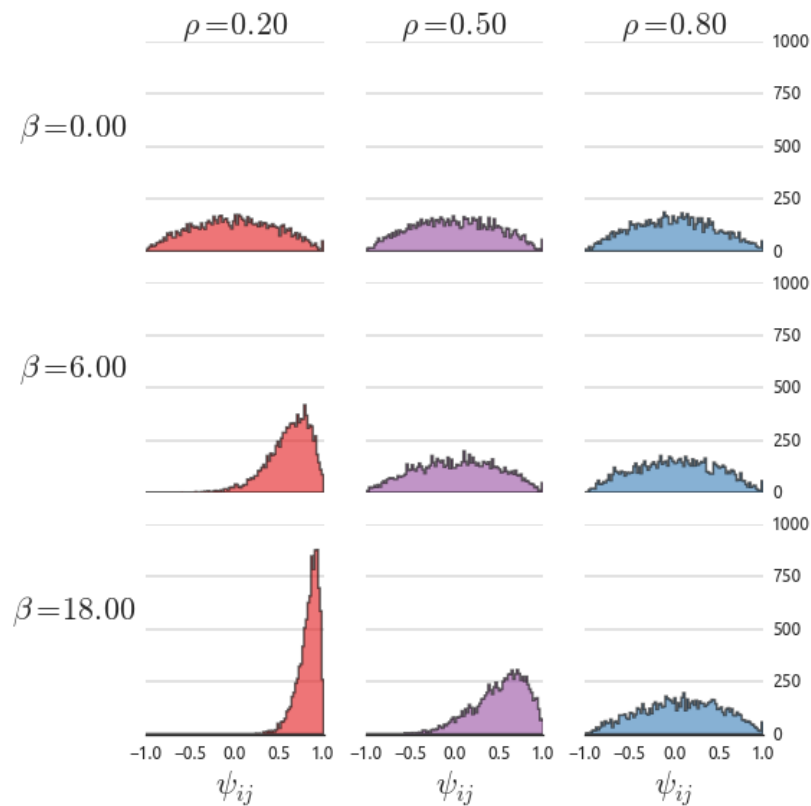


Figura 3.5: Histogramas da similaridade  $\psi_{ij}$  para alguns valores do estilo cognitivo  $\rho$  e da pressão social  $\beta$  com desconfiança  $\varepsilon = 0$  fixa no modelo para o surgimento de consenso.

As figuras 3.5 e 3.6 mostram que conforme a pressão social aumenta, e a sociedade caminha para um consenso, a média dos histogramas de semelhança caminha em direção ao valor 1, e a variância em torno da média vai diminuindo. A intensidade desse fenômeno está relacionada com os valores do estilo cognitivo  $\rho$  e da desconfiança  $\varepsilon$ , sendo mais perceptíveis conforme menores seus valores. Essa característica de baixos valores de  $\rho$  estarem associados com uma menor flutuação tolerada nas opiniões é associada com um comportamento conservador, como mostrado em (César 2014; R. Vicente et al. 2014; Nestor Caticha e Renato Vicente 2011).

### 3.2 Uma Dinâmica para as Relações Sociais

Um efeito colateral da interação dos agentes, não contemplado pelo modelo acima, é a evolução das relações sociais. As próximas perguntas que nos fazemos são como a escolha de parceiros sociais pode ser feita pelos agentes e como essa escolha afeta as fases de desordem e consenso

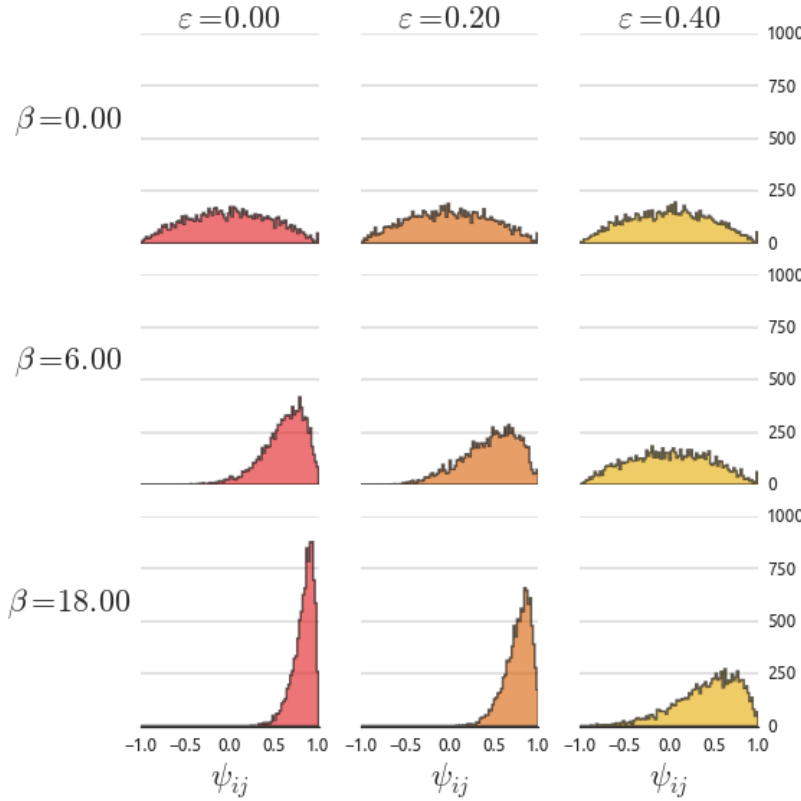


Figura 3.6: Histogramas da similaridade  $\psi_{ij}$  para alguns valores da desconfiança  $\varepsilon$  e da pressão social  $\beta$  com estilo cognitivo  $\rho = 0.2$  fixo no modelo para o surgimento de consenso.

na sociedade? Para respondê-las, implementaremos um mecanismo de controle das relações baseados nos choques de opinião. Isso é feito definido equações para  $R_{ij}$  em função das opiniões dos agentes  $i$  e  $j$ .

Restringindo entre 0 e 1 os valores de  $R_{ij}$ , e interpretando o valor nulo como indicativo do agente  $i$  não considerar o agente  $j$  como parceiro.<sup>1</sup> A evolução de  $R_{ij}$  é dada pela equação (2.22),

$$R_{ij}(t+1) = R_{ij} + \lambda R_{ij}(1 - R_{ij})\text{sgn}(h_i h_j), \quad (3.8)$$

com fazendo nulo o termo de esquecimento, ou seja  $\varphi = 0$ .

As relações sociais  $R_{ij}$  determinam a probabilidade  $p(j|i)$  de cada parceiro  $j$  ser escolhido pelo parceiro  $i$  nessa rodada. A escolha mais simples que podemos fazer para relacionar  $R_{ij}$  e  $p(j|i)$  é a seguinte

$$p(j|i) = \frac{R_{ij}}{\sum_k R_{ik}}, \quad (3.9)$$

que pode ser interpretada diretamente como uma medida de distância<sup>2</sup> entre os agentes  $i$  e  $j$ . A simulação de Monte Carlo é feita de modo análogo ao caso anterior, no qual as relações sociais estavam fixas, com a exceção da escolha do par  $(ij)$  ser feita usando a probabilidade  $p(j|i)$ .

<sup>1</sup> Estamos lidando com um grafo direcionado, representando possíveis não reciprocidades em relações entre agentes.

<sup>2</sup> na verdade seria o inverso da distância, dado que agentes interagem tem probabilidade nula de interação quando sua relação é também nula.

Apenas ajustar a probabilidade de interação entre os agentes não é capaz de produzir polarização de forma consistente. Com a finalidade de produzir esse efeito, vamos definir a constante de  $J_{ij}$ , em analogia com o modelo ‘antiferromagnético’ usado para o campo médio no capítulo 2, mas desta vez como uma função explícita das relações sociais

$$J_{ij} = \frac{1}{K} \operatorname{sgn} \left( R_{ij} - \frac{1}{2} \right) \quad (3.10)$$

Definida dessa maneira, a grandeza  $J_{ij}$  direciona o agente  $i$  no sentido de aprender a opinião de  $j$ , caso sua relação esteja ‘acima da média’, ou no sentido de se opor a  $j$ , caso esteja abaixo. Por ‘acima da média’ entenda que o agente  $i$  prefere interagir com  $j$  mais do que com outros agentes.

### 3.2.1 Complexidade no Nível das Opiniões

O primeiro resultado dessa dinâmica pode ser observado através de histogramas da *semelhança*,  $\psi_{ij}$ <sup>3</sup>. Para interpretar os histogramas, consideremos três possíveis situações de equilíbrio da sociedade.

<sup>3</sup> definida pela equação (3.7)

Considere, primeiramente, a situação com baixa ou nula pressão social,  $\beta$ , indicando que o custo cognitivo da discordância é baixo. Neste caso, não esperamos que os agentes aprendam as opiniões de seus parceiros e, portanto, nenhum consenso pode ser atingido. Essa situação é a que ocorre na região clara dos diagramas de fases nas figuras 3.1 e 3.3. Neste caso, os vetores cognitivos dos agentes estarão distribuídos de maneira uniforme na  $K$ -esfera, e a *semelhança* entre eles terá média e moda nulas.

Na situação representada pela região escura dos diagramas de fases acima, representando as regiões de consenso, a situação é oposta. O custo de discordar de um parceiro é elevado, dada a alta pressão social, de modo que os agentes são levados a aprender a opinião uns dos outros. Isso conduz a uma distribuição de vetores bem mais localizada ao redor do vetor  $x$  ou sua oposta  $-x$ . A *semelhança* entre os agentes terá, portanto, média e moda próximos de 1, de acordo com os valores do estilo cognitivo  $\rho$  e da desconfiança  $\varepsilon$ . Esse é o caso em que representa os histogramas nas figuras 3.5 e 3.6 da seção anterior.

A terceira situação, que é a novidade desse modelo, é uma em que opiniões opostas em relação a  $x$  podem coexistir. É claro que para isso, é necessário uma pressão crítica grande o bastante para que o aprendizado seja necessário, como na situação de consenso. Entretanto, dada a possibilidade de eliminar relações com agentes que discordam permite que opiniões opostas se mantenham, até que o consenso seja atingido em cada uma delas separadamente de modo que há uma rivalidade global. Nesse caso, a *semelhança* entre os agentes também terá média nula, mas apresentará duas modas, cada uma tendendo a  $\pm 1$ , que indica que uma fração dos agentes compartilha opiniões semelhantes mas opostas

a uma outra fração da sociedade.

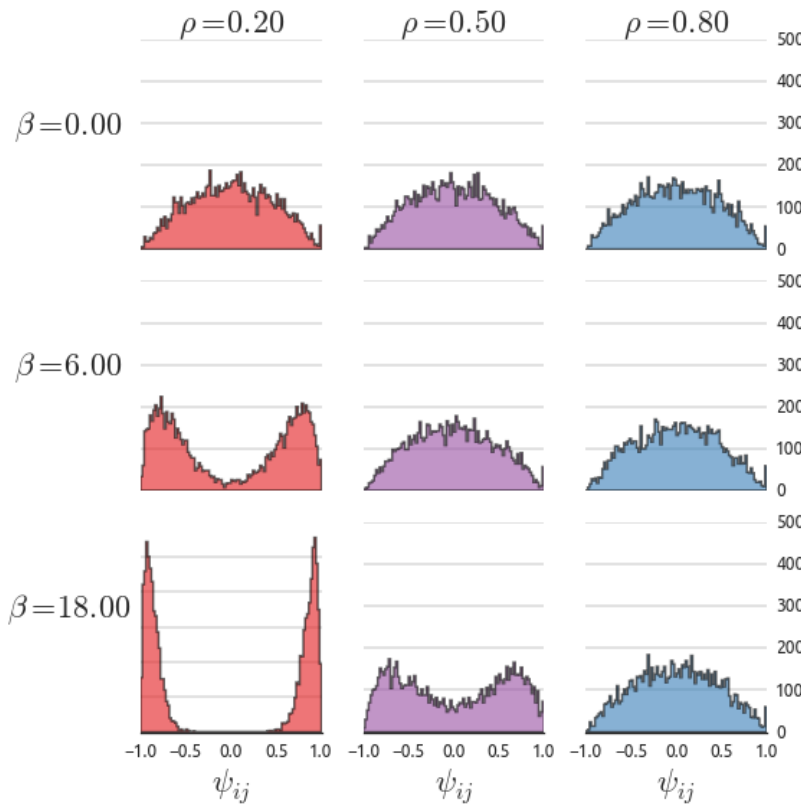


Figura 3.7: Histogramas da similaridade  $\psi_{ij}$  para alguns valores do estilo cognitivo  $\rho$  e da pressão social  $\beta$  com desconfiança  $\varepsilon = 0$  fixa no modelo com a dinâmica das relações sociais.

As três situações acima são ilustrativas e diversos estados intermediários podem surgir, dependendo dos parâmetros que controlam o sistema. As assinaturas estatísticas dessas situações ajudam, porém, a interpretar as fases da sociedade com relação à dinâmica das relações sociais.

A figura 3.7 mostra os histogramas de similaridade em diferentes condições de estilo cognitivo  $\rho$  e pressão social  $\beta$ , mantendo fixado o valor da desconfiança  $\varepsilon = 0$ . Note que estilos cognitivos mais conservadores, dados por valores mais baixos de  $\rho$ , experimentam polarização em condições mais brandas de pressão social e desconfiança.

A figura 3.8 mostra histogramas análogos, desta vez em função da desconfiança  $\varepsilon$  e da pressão social  $\beta$ . Note que a influência da desconfiança sobre o comportamento dos regimes de dissenso, polarização e consenso é similar à do estilo cognitivo, no sentido de dificultar o consenso conforme seu valor aumenta.

É evidente que a distribuição de opiniões com relação a  $x$  está diretamente relacionada com os histogramas das figuras 3.7 e 3.8. Com isso podemos concluir que a sociedade tem duas 'regiões de opinião', em

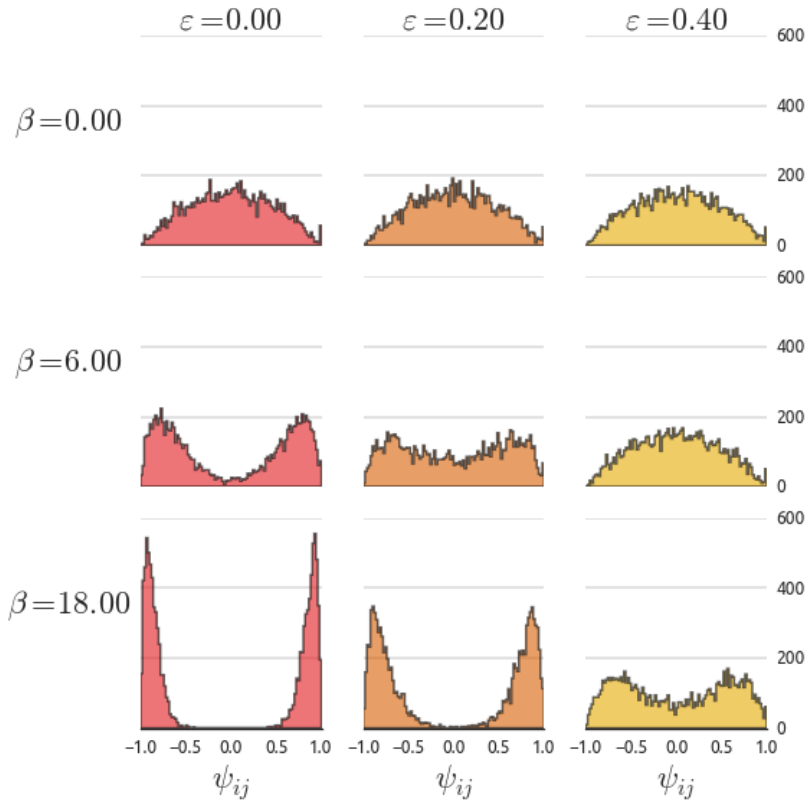


Figura 3.8: Histogramas da similaridade  $\psi_{ij}$  para alguns valores da desconfiança  $\varepsilon$  e da pressão social  $\beta$  com estilo cognitivo  $\rho = 0.2$  fixo no modelo com a dinâmica das relações sociais.

contraposição à situação de consenso. Do ponto de vista da teoria de aprendizado de máquina, podemos dizer que uma sociedade polarizada apresenta uma maior complexidade do que no estado de consenso.

### 3.2.2 O Efeito da Dinâmica no Tecido Social

Outro aspecto dessa complexidade pode ser observada nas matrizes das relações sociais. Como apontado na seção anterior, as relações sociais formam, inicialmente, um grafo completo, ou seja todos os agentes são parceiros. Conforme a sociedade evolui, relações podem ser fortalecidas ou destruídas de acordo com a concordância ou discordância dos agentes durante o encontro.

Como é possível constatar olhando as figuras 3.9 e 3.7, conforme a distribuição de similaridade se tornam mais estritas em relação às modas  $\pm 1$ , estruturas mais evidentes são encontradas nas matrizes de relação social. Isto é, na situação de polarização é possível observar dois blocos de relações sociais bem definidos, representando agente que interagem mais frequentemente com outros no mesmo bloco do que com agentes do outro bloco. Essa é noção de *comunidade* no estudo de redes

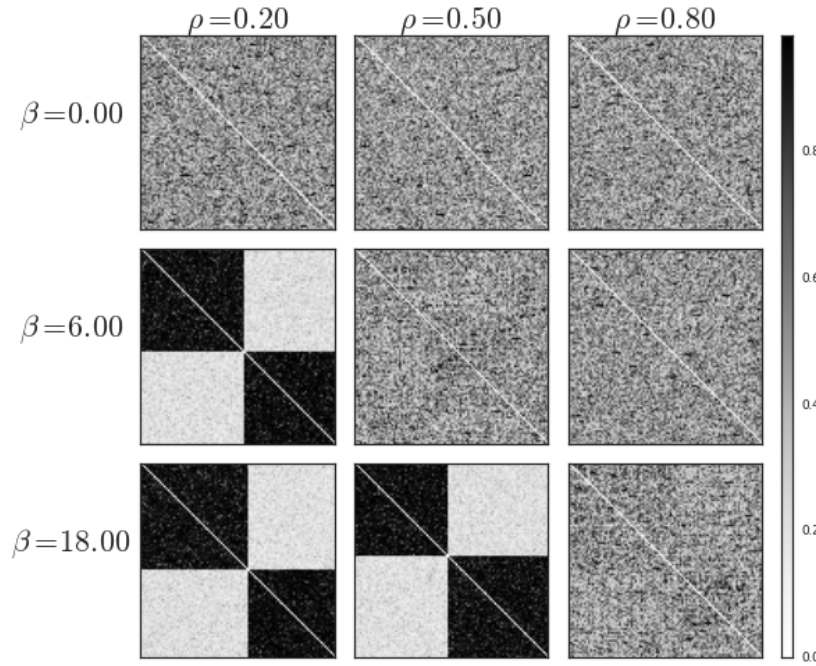


Figura 3.9: Matrizes das relações sociais  $R_{ij}$  para alguns valores do estilo cognitivo  $\rho$  e da pressão social  $\beta$  com desconfiança  $\varepsilon = 0.0$  fixo.

sociais<sup>4, 5</sup>

Em contraposição à estrutura de comunidade do estado polarizado, podemos observar um grafo completo, dado por uma matriz com um único bloco ou comunidade e relativo ao estado de consenso, ou um grafo aleatório, em que surge nenhuma comunidade e que representa a situação de dissenso nas figuras 3.9 e 3.10.

Essa complexidade a nível de estrutura social é de uma natureza diferente da complexidade de opiniões. Embora, por construção, as duas devam estar relacionadas, precisamos de uma forma de garantir que a influência do aprendizado social na estrutura das relações não é coincidência.

### 3.2.3 Outra Forma de Ordem

Determinar a influência da opinião dos agentes na estrutura social, vamos introduzir um novo parâmetro de ordem

$$q = \sum_{(ij)} \text{sgn} \left[ \left( R_{ij} - \frac{1}{2} \right) h_i h_j \right] \quad (3.11)$$

Para entender como esse parâmetro nos permite associar o estado de polarização às estruturas de comunidade, considere apenas o termo da soma referente aos parceiros  $i$  e  $j$ . Se a relação  $R_{ij}$  entre eles é boa, ou seja  $R_{ij} > \frac{1}{2}$ <sup>6</sup>, então eles devem estar na mesma comunidade, se te-

<sup>4</sup> M.E.J. Newman 2010.

<sup>5</sup> As matrizes apresentadas nas figuras foram analisadas usando o algoritmo de SPIN de (Tsafrir et al. 2005), que reorganiza as linhas e colunas em uma matriz de distâncias de modo que linhas e colunas vizinha sejam as representem menores distâncias.

<sup>6</sup> que faz com que  $J_{ij} = \frac{1}{K}$



mos polarização na sociedade, a opinião deles deve ter o mesmo sinal, ou seja,  $h_i h_j > 0$ . Por outro lado, se eles não estão na mesma comunidade,  $R_{ij} < 1/2^7$ , quando há polarização eles devem discordar, ou seja  $h_i h_j < 0$ . Nesses dois casos, esse termo de  $q$  será positivo, pois  $\text{sgn} \left[ \left( R_{ij} - \frac{1}{2} \right) h_i h_j \right] > 0$ . Se, em oposição, os agente  $i$  e  $j$  estão na mesma comunidade mas discordam em opinião, ou concordam em opinião mas estão em comunidades distintas, esse termo será negativo.

O parâmetro  $q$ , que chamaremos de *medida de estrutura*, é a média sobre todas as relações sociais dessa adequação entre opinião e comunidade, e pode ser vista justamente como um valor de quão bem definidas são as estruturas de comunidade. Para chegar a essa conclusão basta perceber que o valor de  $q$  será mais próximo de 1 quão melhor for descrever os agentes com mesma opinião como uma comunidade e mais próximo de zero quando essa descrição não for boa.

As figuras 3.11 e 3.12 mostram uma transição entre as fases de desordem e polarização, semelhante à transição de fases na sociedade que atinge consenso. De modo análogo aos resultados da seção 3.1, para valores de pressão social  $\beta$  maiores que uma dada pressão crítica, que depende do estilo cognitivo  $\rho$  e da desconfiança  $\varepsilon$ , a sociedade experimenta uma fase em que comunidade de opinião bem formada coexistem. Abaixo da pressão crítica, porém, não é possível atribuir a multitude de opiniões a comunidades bem estruturadas, e a sociedade vive uma fase de desordem.

<sup>7</sup> fazendo com que  $J_{ij} = -\frac{1}{K}$

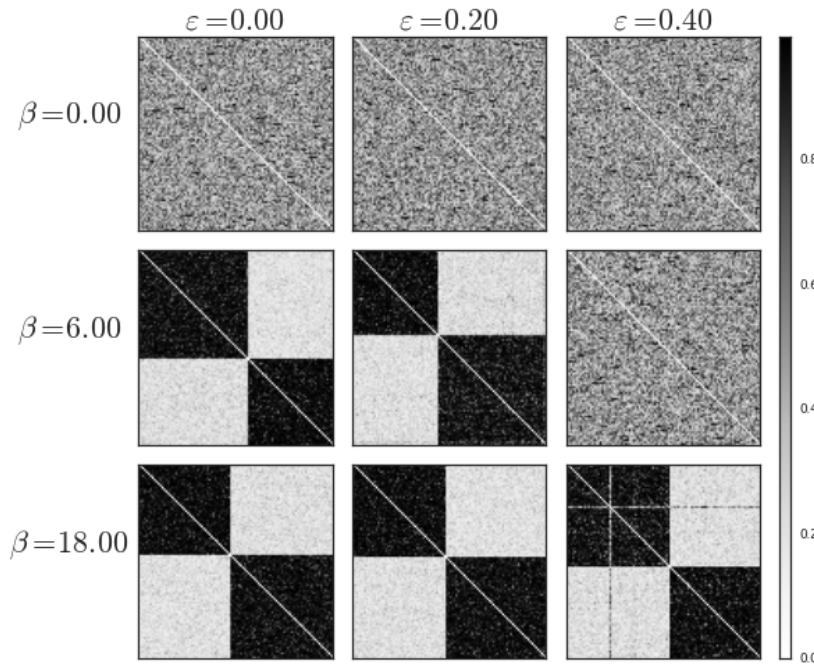


Figura 3.10: Matrizes das relações sociais  $R_{ij}$  para alguns valores da desconfiança  $\varepsilon$  e da pressão social  $\beta$  com estilo cognitivo  $\rho = 0.2$  fixo.

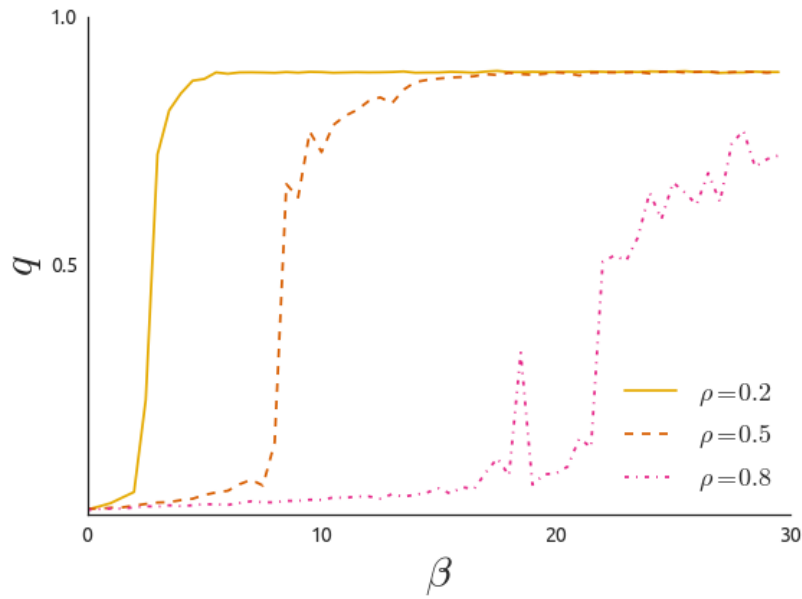


Figura 3.11: Curvas da medida de estrutura  $q$  em função da pressão social  $\beta$  para diferentes valores do estilo cognitivo  $\rho$  e com desconfiança  $\varepsilon = 0$  fixa.

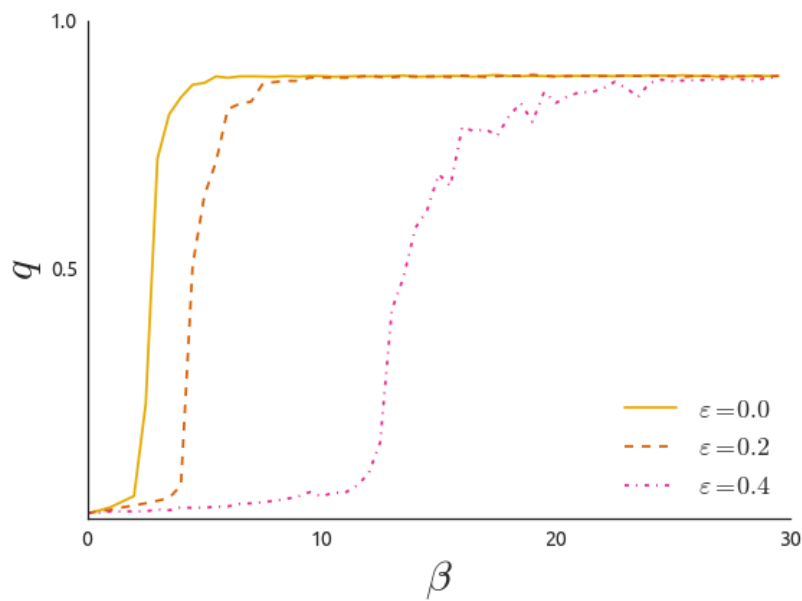


Figura 3.12: Curvas da medida de estrutura  $q$  em função da pressão social  $\beta$  para diferentes valores da desconfiança  $\varepsilon$  e com estilo cognitivo  $\rho = 0.2$  fixo.

O efeito do estilo cognitivo e da desconfiança na pressão crítica que possibilita a estruturação de comunidades é análogo ao caso de consenso. Para valores de  $\rho$  maiores, associados a agentes mais liberais, a formação de comunidades exige maior pressão social. O mesmo pode ser dito sobre os valores de  $\varepsilon$ .

Podemos ver como essa transição se dá ao longo da evolução da sociedade tomando como exemplo a figura 3.13, que mostra o valor de  $q$  ao longo das interações entre agentes para alguns valores da pressão social  $\beta$  e com  $\rho$  e  $\varepsilon$  fixos.

É possível notar que, para baixos valores de pressão social, os valores de  $q$  aumentam por um curto período de tempo, e convergem para zero em seguida. Para valores mais altos de pressão, a medida de estrutura cresce até convergir para o valor 1. Com base na concepção de  $q$  como uma medida da estruturação da sociedade, podemos interpretar esse resultado como uma espécie fenômeno que requer uma ‘*massa crítica*’. Isso é, as comunidades começam a se formar concomitantes com o aprendizado dos agentes, mas só se estruturam bem se uma porção suficiente de agentes compartilha opiniões mais concentradas.

Com a presente análise, concluímos que as habilidades de escolher o indivíduos com que um agente se relaciona e de escolher a natureza dessa relação são suficientes para criar complexidade social. Essa complexidade diz respeito tanto à estrutura das relações sociais quanto à distribuição das opiniões e, de fato, é estabelecida como consequência do acoplamento entre as dinâmicas de aprendizado social e da relações entre agentes. Os resultados apresentados nessa seção são novidade e proporcionam um modelo simples para a formação de estruturas sociais através de um vínculo direto entre a estrutura cognitiva dos agentes e sua interação social.

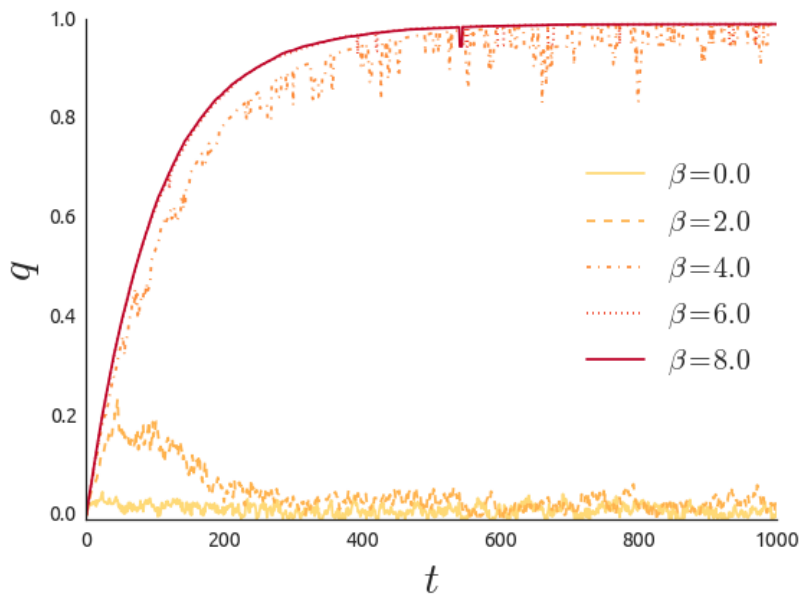


Figura 3.13: Curvas da medida de estrutura  $q$  em função do tempo  $t$ , em unidades de número médio de interações por agente, para diferentes valores da pressão social  $\beta$  e com desconfiança  $\varepsilon = 0$  e com estilo cognitivo  $\rho = 0.2$  fixos.

### 3.3 Jogos de Partidos

Uma vez estabelecida uma dinâmica para as relações sociais que possibilitam o surgimento de grupos com opiniões distintas, podemos nos perguntar como a coexistência desses grupos é afetada por mudanças na pressão social e nas relações entre grupos. Uma nicho social interessante para estudar a influência da estrutura e das relações de poder é o das votações parlamentares de um governo democrático.

Com base numa análise, que embora lúdica é bastante interessante, apresentada no *blog Todas as Configurações Possíveis*<sup>8</sup>, podemos nos perguntar como a mudança de um presidente pode afetar as votações no plenário. A análise lá apresentada consiste da projeção nas componentes principais da matriz de correlação entre os votos de parlamentares, em várias votações de projetos de lei, ao longo de vários mandatos presidenciais. O resultado mostra a forte tendência de partidos como o PMDB em manter seu apoio a projetos da posição, enquanto partidos rivais como PT e PSDB se opõem sistemática e independentemente de quem está no governo.

<sup>8</sup> Marino s.d.(a); Marino s.d.(b).

Nossa ideia é usar a plataforma de aprendizado social e estrutura de comunidade para tentar reproduzir esse comportamento. Para isso, ao invés de considerar a multitude de partidos políticos no Brasil, lidaremos apenas com três partidos, que vivenciam a troca de poder ao e o tempo de dois mandatos presidenciais.

O cenário é estabelecido com dois *partidos ideológicos* formados por agentes concentrados ao redor de suas respectivas *agendas*, representando os parlamentares de partidos rivais e os quais chamaremos *A* e *B*. Um terceiro partido, ao qual damos o rótulo *C*, sem uma ideologia fixada integra o plenário e participa das votações. As agendas políticas dos partidos serão denotadas por  $x_A$  e  $x_B$ .

Sendo rivais, os integrantes dos partidos *A* e *B* não procuram uns aos outros para discutir projetos. Essa situação é bem descrita pela fase de polarização com duas comunidades estudada na seção anterior. Além disso, esses dois partidos têm ideologias fortes, de modo que eles não aceitam as opiniões dos integrantes do partido *C* e com isso os integrantes dos partidos *A* e *B* discutem apenas entre si. O partido *C*, por outro lado, não tem ideologia e se permite interagir com qualquer parlamentar. Para representar essa estrutura, criamos uma matriz de adjacência, a nível ultra métrico, que representa as possíveis interações entre parlamentares de acordo com seus partidos, ilustrada da figura 3.14

Não estamos, neste momento, preocupados com a formação dos partidos e vamos supor que no período considerado nenhum parlamentar troca de partido, de modo que a estrutura da figura ?? é fixa.

A dinâmica de aprendizado é similar à usada nos cenários de sociedade com consenso ou polarização, com a exceção de não existir apenas

	A	B	C
A			
B			
C			

Figura 3.14: Matriz das relações partidárias no modelo com três partidos.

uma questão. De fato, as agendas política dos partidos *A* e *B* representam os projetos propostos por esses partidos e serão, portanto, as questões a serem discutidas pelos agentes. Vamos convencionar que os parlamentares dos partidos *A* e *B* discutem apenas as agendas de seus partidos, enquanto os parlamentares de *C* escolhem uniformemente entre as duas agendas quando procuram a opinião de outro agente, sendo que quem decide a agenda discutida é o agente que faz o papel de “professor”.

Com relação ao estilo cognitivo dos parlamentares, temos que escolher considerar as possíveis diferenças com respeito a políticas liberais ou conservadoras ou considerar os parlamentares com uma mesma orientação nesse sentido. Esse ponto é delicado e difícil de justificar sem um estudo mais profundo sobre esse aspecto da política brasileira, sendo nenhum de conhecimento do autor. Por esse motivo, vamos escolher o mais simples, todos os parlamentares são semelhantes em estilo cognitivo.

Em se tratando de política, é certamente um disparate afirmar que não há desconfiança com relação às posições dos parlamentares com relação aos projetos em tramitação. Entretanto, os resultados que apre-

sentamos até então mostram que o efeito da desconfiança nos regimes da sociedade é similar ao do estilo cognitivo, a saber dificultar o aparecimento de estruturas de ordem. É com isso em mente que vamos fixar  $\varepsilon = 0$  nesse modelo, numa forma talvez inadequada de incluir os efeitos da desconfiança no estilo cognitivo<sup>9</sup>.

O papel da pressão social nesse contexto precisa ser revisto. Lembrando das fases de uma sociedade onde consenso pode surgir, parece natural que dentro de cada partido o valor de  $\beta$  deva estar acima da pressão crítica necessária para atingir o consenso. Porém, não é possível definir de forma heurística o regime no plenário como um todo porque, embora os partidos *A* e *B* esteja 'isolados' e em consenso, o partido *C* pode mudar sua opinião. A capacidade de um partido influenciar o outro deve estar relacionada com seu '*poder político*'. Vamos então associar o poder político do partido com relação à presidência, ou seja, o partido que estiver no governo deve ter mais poder para pressionar os parlamentares do que aquele que os demais. Digamos que os poderes políticos do governo e da oposição são dados por  $J_G$  e  $J_O$  com  $J_G > J_O$ . A pressão política será a pressão social efetiva  $\beta J_{ij}$ , onde

$$J_{ij} = \begin{cases} J_O & \text{se } j \text{ não faz parte do partido do governo} \\ J_G & \text{se } j \text{ faz parte do partido do governo} \end{cases} \quad (3.12)$$

Os partidos viverão dois mandatos de mesma duração, sendo o primeiro sob o governo do partido *A* e o segundo sob o governo do partido *B*. A transição de governo é representada pela troca dos valores de  $J_{ij}$ . Para tornar as coisas interessantes, e de certa forma próximas da situação real, vamos assumir que o número de parlamentares do partido *C* é maior que os dos partidos *A* e *B*. A pergunta que queremos responder é: como o partido *C* se comporta com relação às agendas de *A* e *B*? Ou, em outras palavras, qual dos partidos rivais recebe o apoio do partido *C*?

Para respondê-la, consideramos duas situações distintas com relação à pressão social  $\beta$  no plenário. Numa delas, a oposição não tem coesão mas o governo tem, ou seja o valor de  $\beta$  está abaixo da pressão crítica de modo que a pressão política da oposição não esteja acima da crítica, mas a do governo sim. Na outra,  $\beta$  está acima da pressão crítica, e tanto governo quanto oposição são coerente, embora o governo seja exerça uma pressão política muito maior.

A resposta para essas perguntas pode ser obtida olhando para a evolução da adesão  $m_{rs}$ , com  $r, s \in \{A, B, C\}$ , definida como a média das opiniões dos agentes do partido *r* sobre a agenda do partido *s*. Pela construção do modelo, as adesões interessantes serão  $m_{CA}$ ,  $m_{CB}$ ,  $m_{AA}$  e  $m_{BB}$ , pois queremos saber o posicionamento do partido *C* e devemos monitorar os outros partidos para saber como se comportam com rela-

<sup>9</sup> Outro ponto delicado, mas não completamente sem fundamento. Para compreender melhor a motivação para isso, basta comparar os gráficos da função de modulação (2.15) para diferentes valores de  $\rho$  e  $\varepsilon$  e constatar que, com algumas restrições, aumentar  $\rho$  é qualitativamente equivalente a reduzir  $\varepsilon$  e vice versa, de modo que determinar cada um independentemente não é trivial

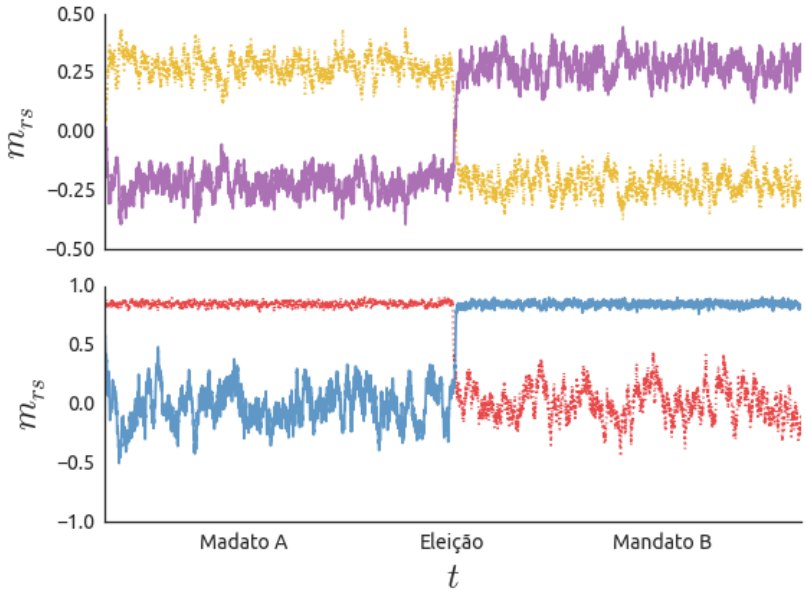


Figura 3.15: Evolução da adesão  $m_{rs}$  às agendas  $s$  pelo partido  $r$  ao longo de dois mandatos com pressão social  $\beta = 1$  fixo. Os parâmetros  $J_G = 10$  e  $J_O = 1$  são atribuídos de acordo com o mandato. Os partidos tem tamanho  $n_A = n_B = 25$  e  $n_C = 100$  e a eleição ocorre em  $t = 5000$ .

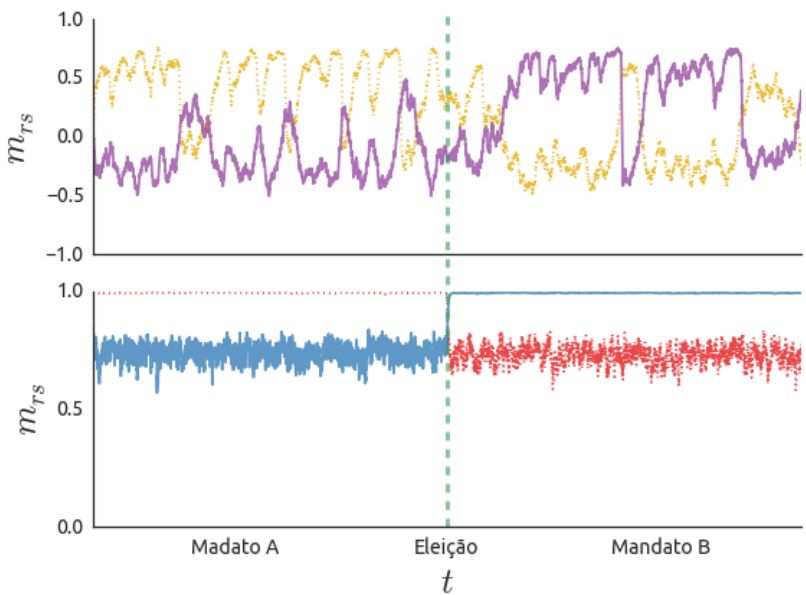


Figura 3.16: Evolução da adesão  $m_{rs}$  às agendas  $s$  pelo partido  $r$  ao longo de dois mandatos com pressão social  $\beta = 5$  fixo. Os parâmetros  $J_G = 50$  e  $J_O = 1$  são atribuídos de acordo com o mandato. Os partidos tem tamanho  $n_A = n_B = 25$  e  $n_C = 100$  e a eleição ocorre em  $t = 5000$ .

ção às próprias agendas.

A 3.15, quando a pressão social está abaixo da pressão crítica, mostra que, embora o partido  $B$  comece concentrado, a falta de coesão faz com que se desprenda da sua agenda e o mesmo ocorre quando  $A$  deixa de ser governo e passe a ser oposição. Nesse caso, não temos partidos

ideológico ao longo dois mandatos. Por conta disso, o partido *C* acaba apoiando o governo, seja do partido *A* ou do partido *B*.

No caso de pressão acima da pressão crítica, ilustrada na figura 3.16, é possível perceber que os partidos rivais *A* e *B* continuam focados em suas agendas, embora sofram um pequeno abalo no valor absoluto do consenso quando não estão no governo. Por outro lado, a adesão do partido *C* oscila entre governo e oposição, com amplitude bem definida, mas com frequência irregular. Entretanto, é possível notar que o partido *C* apoia o governo na maior parte do tempo. Esse resultado é sensível do tamanho relativos dos partidos e da relação de poder entre governo e oposição, podendo não ser observado caso a diferença de poder não seja muito grande ou caso os partidos sejam distribuídos de forma mais justa ou com a maioria em um dos partidos ideológicos. Todavia, esse resultado é bastante interessante por ser capaz de reproduzir parte do comportamento exibido no resultados de votações no plenário ao longo do anos, embora a confirmação através da medida das grandezas introduzidas seja decisivo nessa questão.

Neste capítulo concluímos que mecanismos de formação de estrutura social são capazes de introduzir complexidade nas relações sociais e na distribuição de opiniões. Vimos um exemplo de como fenômenos sociais, no caso as disputas políticas na câmara, podem ser possivelmente descritos através de um modelo preocupado em relacionar as características cognitivas aos fenômenos sociais<sup>10</sup>.

<sup>10</sup> note que embora tenha sido apresentado de forma um tanto lúdica, o modelo para as votações em plenário acabou se mostrando impressionante em vários aspectos.



## 4

# *Conclusões e Futuros Estudos*

### 4.1 Uma Breve Revisão dos Resultados

Neste trabalho desenvolvemos modelos para o estudo de fenômenos sociais com base em características cognitivas dos humanos. Como foi visto no capítulo 1, o crescente corpo de resultados a respeito do comportamento humano, tanto a nível neural quanto psico-social ajudam nos guiam pelo na captura dos elementos básico do comportamento social. Estabelecida a base empírica e fazendo uso das técnicas da mecânica estatística e da teoria de aprendizado de máquina, fomos capazes de estabelecer uma dinâmica para o aprendizado social, que exerce papel fundamental na interpretação dos fenômenos exibidos pelos modelos.

No capítulo 2 desenvolvemos a teoria elementar para a abordagem estatística de fenômenos de interação social através da troca de opiniões. O uso de ferramentas, como o método de máxima entropia, nos permitiu compreender através da análise campo médio qual seria o comportamento de agentes que aprendem em conjunto, bem como o de agentes que se antagonizam, em analogia com sistemas ferromagnéticos e anti-ferromagnéticos. Note que, embora este trabalho tenta tomado como modelo base outros trabalhos recentes na mesma área<sup>1</sup>, está é a primeira vez que uma análise um pouco mais profunda é feita em relação à desconfiança e ao desenvolver a sociedade baseada no custo cognitivo (2.20).

No capítulo 3 conseguimos comparar os resultados de campo médio com simulações usando o custo cognitivo associado à função de modulação Bayesiana, reproduzindo os resultados encontrado nas referências. Com base nesses resultados, elaboramos um modelo de construção (ou aniquilação) das relações sociais com base na troca de opiniões. Esse modelo se mostrou capaz de produzir a estrutura de comunidade com distintas opiniões, proporcionando inclusive uma medida do grau de organização dessa estrutura.

Por fim, um pequeno modelo de comportamento do plenário brasi-

<sup>1</sup> César 2014; R. Vicente et al. 2014; Nestor Caticha e Renato Vicente 2011.

leiro foi elaborado para ilustrar as algumas das possibilidades ao utilizar a nossa abordagem no estudo de fenômenos reais. Embora as bases sob as quais esse modelo foi contruído demandem mais estudo, seus resultados são interessantes pela capacidade de reproduzir o comportamento observado nas votação em plenário. Entretanto, esses resultados tratados como uma guia qualitativa e qualquer esperança de usar esse modelo para explicar quantitativamente fenômenos reais necessita de base experimental.

Além dos resultados apresentados, diversos pontos podem ser destacados com relação ao modelo utilizado. A escolha de usar apenas uma questão é uma limitação que pode ser facilmente removida, embora o custo computacional envolvido é um preço alto a pagar. A escolha da topologia da rede social como um grafo completo é irreal na maior parte das redes sociais. Porém, como um dos focos do trabalho era a construção das estruturas sociais, a escolha de alguma topologia mais realista não é facilmente justificável. Por fim, os resultados apresentados no capítulo 3 dependem de parâmetros usado nas simulações de Monte Carlo para os quais há pouca margem para interpretação dentro do modelo.

Por exemplo, o raio do cone dentro do qual o vetor cognitivo de um agente pode ser proposto está associado com a velocidade de convergência do algoritmo e pode também ser associado a uma velocidade de aprendizado. Porém, essa taxa não surge na dedução apresentada no capítulo 2 e precisaria ser introduzida artificialmente tendo como justificativa seu uso no algoritmo. Os resultados apresentados neste trabalho são aqueles mais robustos com relação a esses parâmetros, e diversas variações dos modelos que poderiam gerar outras interpretações para os fenômenos não foram apresentados por não estarem sob controle dos parâmetros fornecidos pela teoria.

## 4.2 O que há pela frente?

Embora o fenômeno de estruturação social possa ser um pouco melhor compreendido através dos resultados deste trabalho, diversas críticas podem ser feitas com relação à elaboração do modelo. Esforço deve ser desempenhado no sentido de justificar a forma da dinâmica das relações sociais que, diferente da dinâmica de aprendizado, foi estabelecida de forma heurística. Tal justificativa demanda não apenas uma análise matemática mais delicada das interações sociais como também de resultados experimentais diretamente relacionado a escolha de amigos ou de colegas de trabalho e outros comportamentos relacionados.

A falta do respaldo experimental é outra crítica que pode ser feita a este trabalho. A busca de resultados em áreas bem estudadas do comportamento social é prioridade para a sequência do trabalho apresentado aqui.

## Bibliografia

- Amodio, David M et al. (2007). "Neurocognitive correlates of liberalism and conservatism." Em: *Nature neuroscience* 10, pp. 1246–1247. ISSN: 1097-6256. DOI: [10.1038/nn1979](https://doi.org/10.1038/nn1979).
- Asch, Solomon E. (1955). *Opinions and Social Pressure*. DOI: [10.1038/scientificamerican1155-31](https://doi.org/10.1038/scientificamerican1155-31).
- Axelrod, R. (1997). *The Dissemination of Culture: A Model with Local Convergence and Global Polarization*. DOI: [10.1177/0022002797041002001](https://doi.org/10.1177/0022002797041002001).
- Blakemore, Sarah-Jayne (2008). "The social brain in adolescence." Em: *Nature reviews. Neuroscience* 9, pp. 267–277. ISSN: 1471-003X. DOI: [10.1038/nrn2353](https://doi.org/10.1038/nrn2353).
- Calsaverini, Rafael S. (2013). "Tópicos em Mecânica Estatística de Sistemas Complexos". Tese de doutorado.
- Campbell-Meiklejohn, Daniel K. et al. (2010). "How the opinion of others affects our valuation of objects". Em: *Current Biology* 20, pp. 1165–1170. ISSN: 09609822. DOI: [10.1016/j.cub.2010.04.055](https://doi.org/10.1016/j.cub.2010.04.055).
- Castellano, Claudio, Santo Fortunato e Vittorio Loreto (2009). "Statistical physics of social dynamics". Em: *Reviews of Modern Physics* 81, pp. 591–646. ISSN: 0034-6861. DOI: [10.1103/RevModPhys.81.591](https://doi.org/10.1103/RevModPhys.81.591). URL: <http://link.aps.org/doi/10.1103/RevModPhys.81.591>.
- César, Jônatas Eduardo da Silva (2014). "Mecânica Estatística de Sistemas de Agentes Bayesianos". Tese de doutorado.
- Choudhury, Suparna, Sarah-Jayne Blakemore e Tony Charman (2006). "Social cognitive development during adolescence." Em: *Social cognitive and affective neuroscience* 1, pp. 165–174. ISSN: 1749-5024. DOI: [10.1093/scan/nsi024](https://doi.org/10.1093/scan/nsi024).
- Eisenberger, Naomi I, Matthew D Lieberman e Kipling D Williams (2003). "Does rejection hurt? An fMRI study of social exclusion." Em: *Science (New York, N.Y.)* 302, pp. 290–292. ISSN: 0036-8075. DOI: [10.1126/science.1089134](https://doi.org/10.1126/science.1089134).
- Engel, A e C Van Den Broeck (2001). *Statistical Mechanics of Learning*. Vol. 5400, pp. 1–22. ISBN: 051102021X. DOI: [10.1007/978-3-642-01805-3](https://doi.org/10.1007/978-3-642-01805-3). URL: <http://books.google.com/books?id=qVo4IT9ByfQC>.

- Graham, Jesse et al. (2011). "Mapping the moral domain." Em: *Journal of personality and social psychology* 101, pp. 366–385. ISSN: 0022-3018. DOI: [10.1097/00005053-199107000-00016](https://doi.org/10.1097/00005053-199107000-00016).
- Haidt, J (2007). "The New Synthesis in Moral Psychology". Em: *Science* (New York, NY) 316, pp. 998–1002. ISSN: 1095-9203. DOI: [10.1126/science.1137651](https://doi.org/10.1126/science.1137651). URL: <http://www.sciencemag.org/cgi/doi/10.1126/science.1137651>.
- Holroyd, Clay B e Michael G H Coles (2002). *The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity*. Rel. téc., pp. 679–709. DOI: [10.1037/0033-295X.109.4.679](https://doi.org/10.1037/0033-295X.109.4.679).
- Kinouchi, Osame (1996). "Aprendizagem Ótima em Percéptrons a Partir de Exemplos com Ruído". Tese de Doutorado. Universidade de São Paulo.
- Marino, Ricardo. *A Valsa dos Partidos*. URL: <http://www.todasasconfiguracoes.com/2014/04/24/a-valsa-dos-partidos/>.
- *Há Partidos Politicos no Brasil*. URL: <http://www.todasasconfiguracoes.com/2013/09/14/ha-partidos-politicos-no-brasil/>.
- Moriguchi, Yoshiya et al. (2007). "Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence." Em: *Psychiatry and clinical neurosciences* 61, pp. 355–363. ISSN: 1323-1316. DOI: [10.1111/j.1440-1819.2007.01687.x](https://doi.org/10.1111/j.1440-1819.2007.01687.x). URL: <http://www.ncbi.nlm.nih.gov/pubmed/17610659>.
- Neirotti, Juan P e Nestor Caticha (2006). *Statistical mechanics of program systems*. DOI: [10.1088/0305-4470/39/33/006](https://doi.org/10.1088/0305-4470/39/33/006).
- Neirotti, Juan Pablo e Nestor Caticha (2003). "Dynamics of the evolution of learning algorithms by selection." Em: *Physical review. E, Statistical, nonlinear, and soft matter physics* 67, p. 041912. ISSN: 1539-3755. URL: <http://www.ncbi.nlm.nih.gov/pubmed/12786401>.
- Nestor Caticha e Renato Vicente (2011). "Agent-Based Social Psychology: From Neurocognitive Processes To Social Data". en. Em: URL: <http://www.worldscientific.com/doi/abs/10.1142/S0219525911003190?queryID=18/1262402>.
- Newman, M.E.J. (2010). *Networks. An introduction*, p. 772. ISBN: 978-0-19-920665-0 (Hbk).
- Newman, MEJ e GT Barkema (1999). *Monte Carlo methods in statistical physics*, p. 475. ISBN: 0198517971. URL: <http://adsabs.harvard.edu/abs/1999mcms.book.....N>.
- Opper, Manfred (1996). "On-line versus Off-line Learning from Random Examples: General Results". Em: *Physical Review Letters* 77.22, pp. 4671–4674. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.77.4671](https://doi.org/10.1103/PhysRevLett.77.4671). URL: <http://link.aps.org/doi/10.1103/PhysRevLett.77.4671>.

- Papa, Bruno Del (2014). "A study of social and economic evolution of human societies using methods of Statistical Mechanics and Information Theory". Tese de doutorado.
- Sherif, Muzafer (1937). "An Experimental Approach to the Study of Attitudes". Em: *Sociometry* 1, pp. 90–98. ISSN: 00380431. URL: <http://search.ebscohost.com/login.aspx?direct=true%5C&db=sih%5C&AN=18041645%5C&site=ehost-live>.
- Solla, Sara e Ole Winther (1999). "Optimal perceptron learning: as online Bayesian approach". Em: pp. 379–398. URL: <http://dl.acm.org/citation.cfm?id=304710.304758>.
- Somerville, Leah H, Todd F Heatherton e William M Kelley (2006). "Anterior cingulate cortex responds differentially to expectancy violation and social rejection." Em: *Nature neuroscience* 9, pp. 1007–1008. ISSN: 1097-6256. DOI: [10.1038/nn1728](https://doi.org/10.1038/nn1728).
- Tsafrir, D. et al. (2005). "Sorting points into neighborhoods (SPIN): Data analysis and visualization by ordering distance matrices". Em: *Bioinformatics* 21, pp. 2301–2308. ISSN: 13674803. DOI: [10.1093/bioinformatics/bti329](https://doi.org/10.1093/bioinformatics/bti329).
- Vicente, Renato, Osame Kinouchi e Nestor Caticha (1998). "Statistical mechanics of online learning of drifting concepts: A variational approach". Em: *Machine Learning*. Vol. 32, pp. 179–201. DOI: [10.1023/A:1007428731714](https://doi.org/10.1023/A:1007428731714). arXiv: [9801297](https://arxiv.org/abs/9801297) [cond-mat].
- Vicente, R. et al. (2014). "Moral foundations in an interacting neural networks society: A statistical mechanics analysis". Em: *Physica A: Statistical Mechanics and its Applications* 400, pp. 124–138. ISSN: 03784371. DOI: [10.1016/j.physa.2014.01.013](https://doi.org/10.1016/j.physa.2014.01.013). URL: <http://www.sciencedirect.com/science/article/pii/S037843711400017X>.