Universidade de São Paulo
Instituto de Física
SBI-IFUSP

305M810T05759

Aplicações de Mecânica Estatística a Sistemas Sociais: Interação e Evolução Cultural

João Pedro Jericó

Dissertação de mestrado apresentada ao Instituto de Física para a obtenção do título de Mestre em Ciências

Orientador:
Prof. Dr. Renato Vicente

Comissão Examinadora:
Prof. Dr. Renato Vicente (IME - USP)
Prof. Dr. André de Pinho Vieira (IF - USP)
Prof. Dr. Leonardo Paulo Maia (IFSC - USP)

São Paulo
2012

530.13
J55a

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Jericó, João Pedro

Aplicações de mecânica estatística a sistemas sociais: interação e evolução cultural. – São Paulo, 2012.

Dissertação (Mestrado) – Universidade de São Paulo. Instituto de Física – Depto. de Física Geral

Orientador: Prof. Dr. Renato Vicente

Área de Concentração: Física Estatística

Unitermos: 1. Mecânica Estatística; 2. Evolução; 3. Psicologia Cognitiva.

USP/IF/SBI-038/2012

Agradecimentos

Meus profundos agradecimentos ao meu orientador, professor Renato Vicente, por sua considerável dedicação e paciência, pelos problemas sugeridos, pelas áreas de pesquisa apresentadas, pela motivação constante e pelos inúmeros conselhos.

Ao professor Nestor Caticha, por me apresentar à Mecânica Estatística e por demonstrar um interesse motivador a qualquer dúvida ou resultado que lhe era apresentado.

Ao professor Roberto Schonmann, pelo curso de pós-graduação "Tópicos em Modelagem Matemática em Ciências Biológicas e Sociais" dado no IME, cujo conteúdo constitui boa parte da introdução teórica desta dissertação.

Às agências de fomento CNPq e FAPESP, pelo apoio financeiro ao longo de todo o Mestrado.

Aos amigos do grupo de Teoria da Informação: Alex, Bruno, Diogo, Jonatas, Leonardo e Rafael, pelas envolventes discussões, tanto as produtivas quanto as absolutamente irrelevantes, e pelas idéias e contribuições diretas e indiretas a este trabalho.

Aos meus amigos da graduação: André, Bruno, Gabriel, Henrique, Lucas, Michel, Paulo, Petre, Ricardo e Thaenan, pela convivência e amizade ao longo dos últimos seis anos, pelas disciplinas enfrentadas, pelas muitas risadas e pelas tardes de completa improdutividade.

À minha mãe, Márcia, e ao meu pai, João Pedro, pelo apoio absolutamente incondicional, pela importância dada à minha educação e pelos incessantes estímulos à minha formação acadêmica.

À Marina, por tudo.

Muito obrigado.

Resumo

Nesta dissertação de mestrado aplicamos métodos de mecânica estatística a dois problemas relacionados à interação e evolução cultural. O primeiro diz respeito à evolução e manutenção da cooperação altruística. Apresentamos a dinâmica Fisher-Wright em dois níveis, desenvolvida em [49], descreve a evolução do comportamento altruísta através de uma estruturada em grupos com migração e, assim, contorna algumas das limitações de outros mecanismos propostos ao longo do último século. Neste trabalho, aplicamos este mecanismo ao modelo desenvolvido por Boyd, Gintis e Bowles [12], que explica a manutenção da cooperação através da punição altruística, mostrando que o processo Fisher-Wright em dois níveis permite uma abordagem mais robusta e natural, obtendo viabilidade para o altruísmo mesmo com taxas de migração significativas. O segundo problema descrito nesta dissertação consiste de um modelo, proposto em [16], para certos resultados quantitativos em suporte à Teoria de Fundamentos Morais obtidos por Jonathan Haidt [32]. Nestre trabalho estudamos a influência da rede topológica nos resultados deste modelo de agentes, em particular utilizando a rede do Facebook como exemplo de rede realista.

Abstract

In this dissertation we apply statistical mechanics methods to two problems involving social interaction and cultural evolution. The first problem concerns the evolution and maintenance of altruistic cooperation. We discuss the two level Fisher-Wright mecanism proposed in [49] which deals with the question of evolutionary viability of altruistic behavior in a group structured population with migration, and is capable of circumventing certain limitations of other mecanisms proposed in the last century. We use this framework to analyse Boyd, Gintis and Bowles' model [12] for the maintenance and proliferation of cooperation through altruistic punishment, showing that the two level Fisher-Wright process allows for a more robust and natural approach, showing altruism to be viable even at significantly high migration rates. The second problem deals with a model for explaining certain quantitative data obtained by Jonathan Haidt [32] in his Moral Foundation Theory. We study the agent based interaction model proposed in [16] in a scenario where the influcence graph consists of the Facebook social network to understand the network's influence in the results.

Sumário

1 Introdução **9**

1.1 O Problema do Altruísmo 10

1.2 Teoria de Fundamentos Morais 15

1.3 Organização Desta Dissertação 16

2 Evolução da Punição Altruística **17**

2.1 Punição Altruística 17

2.2 Modelo de Boyd - Gintis - Bowles 19

2.3 O Modelo Fisher-Wright 28

2.4 Análise do Processo 2lFW Para o Modelo BGB 30

2.5 Regime de Seleção Fraca 37

2.6 Simulação 41

2.7 Modelo de Limiar 43

3 Dinâmica de Classificações Morais **47**

3.1 Resultados Empíricos 47

3.2 Mecânica Estatística de Classificações Morais 49

3.3 Redes Barabási-Albert e Facebook 53

3.4 Introdução de Agentes Fixos 59

4 Conclusão **63**

4.1 Perspectivas 64

Capítulo 1

Introdução

O problema da extração de um comportamento agregado macroscópico a partir de regras microscópicas de partículas é central a diversas áreas da ciência. Não é surpreendente, portanto, que o corpo de técnicas e ferramentas desenvolvidas ao longo dos últimos dois séculos em mecânica estatística esteja cada vez mais sendo aplicado em áreas além da física, tradicionalmente em informação [24] e comunicações [40], mas também recentemente têm sido vistas diversas aplicações em economia [17] [5], biologia [7] [21] [41] e ciências sociais [43].

Esta última aplicação não é tão estranha quanto possa parecer. Laplace, no final do século XVIII, já havia demonstrado que a distribuição de sexos em nascimentos em Paris possuía formato idêntico à de moedas sendo jogadas ao acaso [8], o que era evidência bastante forte de que o número próximo de homens e mulheres era resultante de um processo aleatório, e não prova da sabedoria de Deus em prover esposas para todos.

Existem evidências [15] de que o próprio Maxwell, na metade do século XIX, deduziu sua distribuição de velocidades de partículas em um gás ideal em parte inspirado pelas obras do historiador Henry Buckle [8]. Maxwell estava convencido de que tentar entender o movimento individual de cada partícula em um gás era uma tarefa sem sentido e qualquer tentativa de descrever o comportamento do gás deveria ser feita de forma estatística. Em suas palavras:

> *Estas uniformidades observadas em experimentos com quantidades de matéria contendo milhões e milhões de moléculas são uniformidades da mesma categoria que aquelas explicadas por Laplace e ponderadas por Buckle, surgindo através da aglutinação de uma grande variedade de causas, nenhuma delas uniforme às outras.* [15]

Nesta dissertação, aplicamos os métodos da mecânica estatística a dois problemas relacionados. O primeiro problema trata da evolução da cooperação altruística. Pela descrição

do processo de seleção natural, o comportamento altruísta não poderia sobreviver, pois um indivíduo que se prejudica para ajudar os outros deveria deixar menos descendentes que partilha seus genes, e esta tendência genética para a cooperação deveria desparecer em algumas gerações. Existem no entanto diversas manifestações de comportamentos altruístas fortes na natureza e na sociedade humana. Explicar esta aparente inconsistência é um problema central em biologia evolutiva [22] e foram desenvolvidos ao longo do século passado diversos mecanismos diferentes que tentam solucionar este problema. Todos estes mecanismos, porém, apresentam limitações amplamente discutidas [11] [45] e não são capazes de justificar plenamente a existência do comportamento altruístico. Apresentaremos o mecanismo Fisher-Wright em dois níveis, introduzido em [49], que é um modelo microscópico para a reprodução e migração de alelos de um gene estruturados em grupos ou demos. Esta abordagem generaliza alguns mecanismos existentes mas não sofre de algumas das limitações. Neste trabalho, como um exemplo das vantagens deste novo mecanismo, será feita a sua implementação em um modelo desenvolvido por Boyd, Gintis e Bowles no qual a evolução e a manutentação do altruísmo ocorrem através da punição altruística.

O segundo problema abordado nesta dissertação descreve uma tentativa de explicar os resultados da Teoria de Fundações Morais de Jonathan Haidt [32]. Haidt utilizou questionários com perguntas sobre valores morais e concluiu que o julgamento moral é dividido em cerca de cinco (ou seis [33]) categorias e que a importância dada a cada uma destas cinco categorias varia de acordo com o indivíduo. Ele observou que existe uma correlação muito forte entre a posição política declarada de cada participante do questionário e suas classificações morais. Neste trabalho será apresentado um modelo de interação de agentes, introduzido em [16], onde tenta-se explicar os resultados obtidos por Haidt, com algum sucesso. A interação deste modelo ocorre em alguma rede topológica e será feito o estudo da influência do tipo de rede nos resultados, em particular utilizando a rede do Facebook como exemplo de rede realista.

1.1 O Problema do Altruísmo

A seleção natural é o mecanismo central por trás da evolução biológica [48] [18]. De modo geral, neste mecanismo cada indivíduo possui um *fitness*[1] [38], que pode ser definido [50] como o número médio de descendentes de um indivíduo que sobreviverão até a fase reprodutiva [20]. Supondo que os indivíduos em uma população infinita possuam um gene com dois alelos, A e N, com *fitness* f_A e f_N, considerados por ora fixos. Então, cada geração nova possui $N_A(t+1) = N_A(t)f_A$ indivíduos com o gene A e $N_N(t+1) = N_N(t)f_N$ indivíduos do tipo N. Se p for a fração de indivíduos A na população e não houver estrutura topológica, então a variação de p entre a geração t e $t+1$ é dada por

[1] A tradução do termo *fitness*, "aptidão", possui um sentido diferente em português e, portanto, optou-se por utilizar o termo original.

$$\Delta p(t) = p(t)(1-p(t))\frac{f_A - f_N}{p(t)f_A + (1-p(t))f_N}. \tag{1.1}$$

Esta equação é conhecida como equação do replicador [38] [44], e é utilizada para descrever seleção natural sem mutações (também dita seleção natural pura). Para dois alelos, só existe um equilíbrio caso haja fixação de um dos alelos (isto é, $p = 0$ ou $p = 1$) ou $f_A = f_N$.

Um problema não trivial para se entender seleção natural é justificar o comportamento altruísta [19] [11] [20]. Um comportamento altruísta forte será definido como aquele em que o indivíduo (ou animal) diminui o seu próprio *fitness* exclusivamente para aumentar o *fitness* de outros. Nesta discussão, um egoísta será o indivíduo que interage com um altruísta mas não age reciprocamente (e assim se beneficia da diminuição do *fitness* do outro). Suporemos que a tendência aos comportamentos altruísta ou egoísta é genética.

É possível encontrar diversas observações de comportamento altruístico na natureza e em seres humanos [11]. Certas espécies de aves, por exemplo, ao ver um predador emitem um aviso sonoro para alertar seu bando do perigo iminente [20]. Isso faz com que o bando inteiro possa fugir e se safar, mas coloca a ave que fez o alerta em um perigo maior, pois ela chama a atenção do predador. Em um exemplo ainda maior de altruísmo em animais, abelhas operárias de uma colméia ferroam aqueles que tentam roubar o mel produzido. No entanto, o ferrão está ligado ao abdômen e ao fugir após uma ferroada a abelha muito provavelmente terá o sistema digestivo arrancado e morrerá, não deixando mais descendentes.

Em seres humanos, existem evidências [39] de que grupos pré-históricos caçavam animais de até duas toneladas, o que certamente exigia uma grande cooperação entre os caçadores que individualmente arriscavam a sua vida para fornecer alimento para o resto do grupo. Em [10], Samuel Bowles mostra que a existência de conflitos primitivos gerou uma pressão evolutiva onde os grupos nos quais havia cooperação altruística, na forma de voluntários à guerra (que pagam o custo de arriscar a própria vida para beneficiar os outros membros de seu grupo na forma de segurança), eram mais bem sucedidos e obtinham territórios mais férteis e com fácil acesso à água.

Não é simples explicar este comportamento altruísta utilizando o mecanismo de seleção natural. A justificativa originalmente dada para a sua existência é de que o comportamento altruísta beneficia o grupo inteiro em que um indivíduo vive e, portanto, o grupo todo consegue prosperar se cada membro diminui o seu *fitness* para aumentar o dos outros [19] [22] [47]. A idéia deste mecanismo, chamado de seleção de grupo, é que a unidade de seleção natural seja não o indivíduo mas o grupo. Assim, grupos que possuem um maior *fitness* deixam mais descendentes com os alelos de seus integrantes e, por isso, grupos compostos de altruístas deixam mais descendentes com esse comportamento. O modelo talvez mais famoso

de seleção de grupo é conhecido como "modelo de *haystack*" (ou "modelo de palheiros"), desenvolvido por Maynard Smith, exemplificado por grupos de animais que vivem e se reproduzem em palheiros distintos por algumas gerações. Após esse período, saem todos ao mesmo tempo, se misturam e voltam aos palheiros de forma aleatória, gerando novos grupos. Aqueles grupos com um maior número de altruístas têm uma população maior e ao serem misturados o alelo altruísta se espalhará pela população global, crescendo em proporção ao alelo egoísta.

O mecanismo de seleção de grupo, no entanto, é problemático em alguns aspectos importantes. Primeiramente, a resolução da seleção de grupo não possui boa definição. Uma aldeia pode ser considerada um grupo, mas as diferentes famílias dentro desta aldeia podem ser consideradas grupos distintos competindo entre si, e um novo tipo de "cooperação super altruística" pode surgir entre elas. Como definir em qual nível de agregação a seleção de grupo se aplica?

Além desse problema de resolução, a principal crítica à seleção de grupo é que, embora um grupo inteiro se beneficie de seus componentes serem altruístas, a reprodução ocorre no nível do indivíduo. Por definição, um indivíduo egoísta em um grupo composto apenas por altruístas possui um *fitness* maior do que seus pares e terá mais descendentes, que partilham a sua genética egoísta e têm também maior *fitness* que o resto do grupo. Eles, portanto, se tornarão dominantes, até que os altruístas se extinguam. Mesmo se for possível eliminar o fenótipo egoísta (e forem desprezadas mutações), migrantes podem proliferar às custas dos altruístas. É fácil ver na equação do replicador (1.1) que se $f_A < f_N$ para qualquer valor de p, o único equilíbrio estável é $p = 0$. Mesmo que $p = 1$ seja equilíbrio, uma perturbação por mutação ou migrante leva o sistema a $p = 0$.

Em face a essas dificuldades, a ideia de seleção de grupo perdeu força, e a partir dos anos 60, com os trabalhos de John Maynard Smith e William Hamilton [34] foi sugerido o mecanismo chamado seleção de parentesco (*kin selection*). Nesta teoria, a unidade de seleção é o gene e, dado que altruísmo aumenta o *fitness* dos outros indivíduos, incluindo os que partilham desse genótipo, garante-se que o alelo altruísta seja copiado (através da reprodução dos indivíduos) a uma taxa maior do que o alelo egoísta. A representação mais comum e popular desta idéia é a chamada regra de Hamilton, onde dado que o B é aumento de *fitness* que o comportamento altruísta proporciona aos outros e C o custo dele para o invidíduo, então o altruísmo será favorecido pela seleção natural caso

$$rB > C, \tag{1.2}$$

onde r é o coeficiente de parentesco, definido como a probabilidade de que um outro indivíduo possua os mesmos alelos devido a descendência comum.

A regra de Hamilton simplifica uma idéia central da seleção de parentesco, de que o *fitness* relevante para a seleção natural não é o individual, mas sim o chamado *fitness* inclusivo, que é a soma do *fitness* do indivíduo focal somado ao de todos os outros indivíduos que interagem com ele, ponderados pelo coeficiente de parentesco. O *fitness* inclusivo pode ser considerado como o *fitness* do alelo na população. Ele também apresenta uma característica que é comum a todos os mecanismos de evolução do altruísmo: a ideia de que para que seja possível o alelo altruísta proliferar, indivíduos que o possuem devem interagir mais frequentemente do que o que seria esperado por simples acaso (também chamado de *positive assortment*, algo como sorteio favorável).

O grande mérito da seleção de parentesco é explicar como o altruísmo pode proliferar sem que esse comportamento seja dominado por um alelo egoísta com *fitness* maior. No entanto, observa-se em diversos exemplos [9] [36] cooperação entre pessoas que não são familiares próximas e, portanto, possuem um coeficiente de parentesco muito baixo. Até mesmo para sociedades caçadores-coletoras, o coeficiente médio de parentesco estimado é baixo, da ordem de 5% [27], sendo necessária uma proporção B/C extremamente alta para que o altruismo seja viável somente por seleção de parentesco.

Dada esta ineficiência da seleção de parentesco em explicar a evolução do altruísmo, outros mecanismos foram propostos [45]. O mecanismo de reciprocidade direta, por exemplo, pressupõe que uma interação entre indivíduos provavelmente occorerá diversas vezes e, portanto, altruístas interagindo entre si sustentam uma cooperação benéfica para ambos, enquanto egoístas não conseguem se aproveitar de outros indivíduos durante muitas interações e a longo prazo terão um *fitness* menor. Em seu livro, Robert Axelrod [6] descreve um torneio feito entre programas de computador jogando um jogo do dilema do prisioneiro iterado, e mostra que a estratégia campeã era a chamada Tit-for-Tat, que coopera na primeira rodada e nas rodadas subsequentes só o faz se o outro agente também cooperar. Axelrod mostra como essa estratégia é de fato superior e pode invadir grupos egoístas mesmo com probabilidade de interação entre agentes com Tit-for-Tat relativamente baixa (da ordem de 10%, esta probabilidade de interação essencialmente faz o papel do coeficiente de parentesco), dado que as interações tenham uma chance alta de durar mais que uma rodada. Talvez um dos exemplos mais interessantes de reciprocidade direta em humanos é o chamado sistema "Viva e deixe viver" das trincheiras da 1ª Guerra Mundial, onde era comum que dois exércitos se enfrentassem em trincheiras durante meses. Isso levou à uma situação onde nenhum dos dois lados iniciava agressões e qualquer agressão era prontamente retaliada, o que reforçava a cooperação mútua. A grande relevância deste exemplo é que ele demonstra como estes mecanismos permitem que o altruísmo evolua como melhor estratégia mesmo entre grupos antagônicos. É possível mostrar [38] que se B e C são como em (1.2) e w é a probabilidade da interação terminar a cada rodada (isto é, para interações únicas, $w = 1$), então o altruísmo é viável se

$$wB > C. \tag{1.3}$$

De fato, a solução que os oficiais ingleses e alemães encontraram para resolver o problema da cooperação em trincheiras (que, do ponto de vista de uma guerra, é ruim) foi mudar esquadrões inteiros de trincheiras com frequência, desta forma diminuindo sensivelmente w e tornando a cooperação muito menos viável.

As interações humanas, no entanto, não dependem só de reciprocidade direta. A maior parte das interações entre humanos é assimétrica e não se repete muitas vezes. Ainda assim, há cooperação altruística [9] entre seres humanos que não se conhecem e não esperam se encontrar novamente. Experimentos em laboratórios [14] [25] mostram que seres humanos são altruístas mesmo em completo anonimato e quando a cooperação envolve dividir uma quantia significativa de dinheiro. O mecanismo de reciprocidade indireta baseia-se na reputação adquirida através da cooperação. Ser altruísta aumenta a reputação do indivíduo, que será recompensado pela cooperação de outros que o reconhecerem como tal.

O interessante deste mecanismo é que ele é praticamente exclusivo de seres humanos (e talvez dos primatas). Outros animais não possuem capacidade cognitiva suficiente para monitorar as interações entre outros indivíduos e julgar a reputação de cada um. É necessário desenvolver linguagem para poder obter informação sobre o comportamento dos outros na forma de fofoca. Isso oferece uma justificativa para o altruísmo em humanos ser consideravelmente mais abrangente que o de outros animais: nós temos acesso a um mecanismo extra de evolução da cooperação.

Mas nossa capacidade cognitiva é limitada. Assim, deve haver uma probabilidade q de se saber a reputação de um outro indivíduo qualquer da sociedade. Pode-se mostrar [46] que o altruísmo é viável no mecanismo de reciprocidade indireta se

$$qB > C. \tag{1.4}$$

No entanto, a probabilidade q de se saber a reputação de outro indivíduo em uma sociedade humana é muito baixa. Os próprios experimentos descritos anteriormente são feitos de forma anônima e, portanto, possuem $q = 0$. Logo a recriprocidade indireta não conta toda a história.

Observa-se, portanto, que a necessidade de que B/C seja irrazoável para que o altruísmo evolua é um problema pervasivo dos mecanismos propostos, pois deve-se explicar a cooperação entre pessoas sem correlação genética, em interações sem repetição e onde a reputação é desconhecida. Para solucionar este problema, é geralmente proposto como comportamento adicional a punição de agentes egoístas, que será um dos foco de estudo deste trabalho.

1.2 Teoria de Fundamentos Morais

Em seus artigos [32], Jonathan Haidt divide o julgamento moral em cinco categorias principais: questões relacionadas à integridade física (nomeada por ele de *Harm*), à justiça e igualdade (*Fairness*), à lealdade ao grupo (*Group*), ao respeito a autoridades (*Authority*) e a pureza e santidade (*Purity*). Ele fez uma série de entrevistas e aplicou questionários com afirmações de teor moral correspondentes a cada uma destas cinco categorias. Como exemplo, relacionado à integridade física o entrevistado deveria colocar em uma escala de 1 a 5 que iria de "discordo fortemente" a "concordo fortemente" a situação "Se eu visse uma mãe batendo em sua criança, ficaria revoltado.". Outros exemplos incluem "Se eu fosse um soldado e discordasse das ordens de meu superior, obedeceria de qualquer forma, pois esse é o meu dever." A lista completa de perguntas em mais detalhes pode ser vista em [32].

Ao descrever os resultados, Haidt conclui [31] que pessoas com filiações políticas diferentes (liberais e conservadores[2]) dão graus de importância diferentes para cada uma das cinco categorias quando fazem julgamento moral sobre um assunto. Em particular, liberais parecem dar mais relevância a questões morais ligadas à integridade física e justiça, e menos relevância às outras categorias. Indivíduos mais conservadores, por outro lado, consideram todas as categorias igualmente relevantes, o que significa que dão mais valor a lealdade ao grupo, respeito a autoridade e santidade do que liberais. Todas as filiações políticas descritas no trabalho de Haidt (e que aparecem nos dados deste trabalho) são *auto declaradas*, isto é, cada pessoa ao responder o seu questionário preenchia sua filiação política em uma escala de 1 a 7, onde 1 é considerado *fortemente liberal* e 7 *fortemente conservador*.

O modelo que deu origem a este trabalho [16] é uma tentativa de modelar a formação da opinião moral dos indivíduos que responderam o questionário descrito. Para a construção deste modelo baseado em fundações morais, foram utilizadas duas evidências empíricas importantes. A principal é o fato de que existem diferenças entre a forma como liberais e conservadores tratam informação novas e informações corroborativas [4]. No experimento "Go / No-Go" citado, os participantes devem prestar atenção a um monitor e apertar um botão toda vez que o sinal "Go" aparecer. Isto é feito de forma regular e frequente, de forma que a resposta se torne habitual e previsível. Em um número pequeno de testes, uma mensagem de "No-Go" aparece na tela, indicando que o botão não deve ser apertado. O estudo mostra que os liberais apresentam um aumento bastante significativo de atividades neurais relacionadas a conflitos quando era necessária a inibição da resposta em um cenário "No-Go". Isto resultou em melhor acurácia nos testes de "No-Go", porém não alterou os resultados dos testes com confirmação de padrões ("Go"). A conclusão dos autores é que

[2] A pesquisa foi feita com um público americano, portanto deve-se levar em consideração que as definições de liberal e conservador neste trabalho correspondem às posições políticas americanas. Estas definições possuem outros sentidos no Brasil. O modelo será essencialmente livre de semântica, portanto esta diferença não será relevante, mas é importante manter isto em mente.

uma filiação política conservadora está relacionada a uma maior persistência em manter um padrão de respostas habitual, enquanto uma filiação mais liberal está relacionada a uma adaptação mais rápida dadas novas informações. Esta diferença será a motivação de um dos principais parâmetros do modelo, como será visto adiante.

A segunda evidência empírica utilizada é um estudo mostrando que a rejeição gera reações cerebrais análogas à dor [23]. Os participantes do experimento citado têm imagens de fMRI feitas enquanto jogam um jogo virtual de arremesso de bolas, onde acreditavam estar jogando com outros participantes, enquanto na realidade estavam interagindo com um programa de computador.

Em um primeiro momento, eles foram avisados que, por dificuldades técnicas o scanner de fMRI ainda não estava ligado e eles não iriam participar desta rodada. Na segunda rodada, os participantes participavam do arremesso de bolas, mas eram excluídos pelo computador (que acreditavam ser outras pessoas) e assistiam de forma passiva ao jogo da mesma forma que na rodada anterior. A diferença foi que, ao serem ignorados por seus pares, o scan de fMRI mostrou bastante atividade no córtex cingular anterior, área do cérebro geralmente relacionada à dor. Isso significa que existe um custo social da rejeição, algo que será implementado no modelo a seguir como um aumento na energia entre dois agentes discordantes.

1.3 Organização Desta Dissertação

Esta dissertação está organizada da seguinte forma: no capítulo 2 serão apresentadas a punição altruística como justificativa para a evolução da cooperação e o modelo Boyd - Gintis - Bowles (BGB) que implementa a punição em uma dinâmica *haystack*. As limitações e aproximações do modelo BGB serão discutidas e o mesmo mecanismo de punição será implementando no processo Fisher-Wright em dois níveis, que também será apresentado. Os resultados obtidos serão comparados e discutidos.

No capítulo 3 serão introduzidos a Teoria de Fundamentos Morais de Jonathan Haidt e o modelo de agentes interagentes desenvolvido para justificar os resultados obtidos por Haidt. Após a comparação dos resultados, será feita uma discussão sobre a rede topológica utilizada e o mesmo modelo de interação será considerado na rede do Facebook e em redes de Barabási-Albert variadas para se estudar a influência de uma rede realista no equilíbrio de opiniões obtido. Por fim, um breve estudo sobre a influência de agentes fixos na rede é discutido.

O capítulo 4 é composto de considerações finais sobre ambos os trabalhos, bem como as perspectivas de futuros desenvolvimentos.

Capítulo 2

Evolução da Punição Altruística

Neste capítulo descreveremos o modelo de Boyd - Gintis - Bowles [12] que indica que o mecanismo de punição altruística pode resolver o problema da cooperação em um modelo com dinâmica de grupos similar ao modelo *haystack* e pode explicar a evolução se implementado um sorteio favorável. Discutiremos as vantagens e limitações do modelo BGB e será feita a implementação da mesma estratégia com a dinâmica Fisher-Wright em dois níveis para que se possa comparar os resultados obtidos e discutir as vantagens desta dinâmica.

2.1 Punição Altruística

Em face à ineficácia dos modelos discutidos na Introdução em justificar a evolução do alelo altruísta via seleção natural, a busca de outras evidências empíricas que estejam relacionadas à cooperação fez com que se empregasse a punição como mais uma explicação para o comportamento altruísta. Supõe-se que um novo comportamento possível ao indivíduo que coopera seja reduzir voluntariamente o seu *fitness* para reduzir o de outros agentes. Este tipo de punição é dita altruística pois um indivíduo que pune um egoísta essencialmente fica exposto ao custo de punição, que necessariamente implica em riscos. O benefício resulta da promoção de cooperação em escala mais ampla do que sem o mecanismo.

Existem fortes evidências empíricas para o mecanismo de punição. Em experimentos feitos em laboratório [29], participantes jogam um jogo de bens públicos que funciona da seguinte forma: em grupos de quatro pessoas, cada indivíduo começa com 20 unidades monetárias e a cada rodada podem pagar uma unidade monetária do seu montante para que todos os quatro participantes recebam 0,4 unidades. É claro então que se todos os participantes cooperarem, receberão 1,6 unidades monetárias pelo custo de uma. Mas cada participante,

individualmente, é incentivado a não cooperar, mantendo seu dinheiro, e receber o benefício dos outros. A interação é repetida ao longo de diversas rodadas, sempre com grupos novos, o que garante que não exista cooperação devido à reciprocidade direta. A interação é anônima, o que garante que não haja cooperação devido a efeitos de reputação. O que Fehr e Gächter observaram é que no experimento da forma como descrito, a taxa de cooperação começava baixa e caía ainda mais ao longo das rodadas. No entanto, quando foi introduzida a opção de qualquer participante pagar uma unidade monetária para retirar três de outro jogador, a cooperação crescia de forma significativa. Participantes puniam ativamente os que não cooperavam, e estes corrigiam suas ações e passavam a cooperar nas rodadas seguintes, mesmo sabendo que estavam jogando com outros indivíduos.

O sorteio de novos grupos anônimos nesse experimento faz com que a punição não possa ser explicada com base em ganhos materiais indiretos, isto é, os ganhos que um participante obtém no longo prazo devido à cooperação de um indivíduo que foi punido. O ato de punição é estritamente altruístico neste sentido porque um punidor paga os custos e não recebe os benefícios de um egoísta reformado. Também não há uma motivação de vingança por parte do punidor, por ter sofrido pessoalmente a injustiça. Em [26], os autores conduzem um experimento em que os participantes jogam um jogo do ditador, no qual um jogador recebe 20 euros e deve dividir da forma como quiser este dinheiro com um segundo indivíduo. No experimento, os participantes a serem testados ficam na posição de recipiente do dinheiro a ser dividido ou assistem passivamente à interação entre dois jogadores. Após a divisão do dinheiro cada participante tem a possibilidade de punir o ditador (que divide o dinheiro) a um custo para si. Em ambos os casos, quando recipientes do dinheiro ou observadores passivos, os participantes puniam os ditadores de forma proporcional à desigualdade da divisão (isto é, virtualmente nada para uma divisão 10:10, e significantemente para uma divisão 20:0). Isso exclui a possibilidade da punição ser executada devido à satisfação pela vingança pessoal. Os experimentadores também observaram que em ambos os casos o ato de punir participantes egoístas ativava a região do cérebro ligada à recompensa. Isso indica que existe um senso de moralidade por trás da punição, pois o ato de punir indivíduos que agem de forma injusta é prazeroso.

A punição também não está restrita à cooperação em situações de pequena escala em laboratório. Em conflitos primitivos a punição parece ser um mecanismo central para manter a cooperação custosa de combate [37] mesmo em grande escala (populações de 500 mil pessoas sem governo central), onde membros da população lutam ao lado de completos estranhos.

No entanto, por ser uma espécie de altruísmo forte, a punição apresenta os mesmos problemas que a cooperação. Supondo um grupo composto majoritariamente de altruístas punidores, que rapidamente castigam e eliminam em poucas gerações qualquer egoísta que apareça por migração ou mutação. Um indivíduo com um alelo "egoísta de segunda ordem" poderia invadir este grupo se cooperasse incondicionalmente mas deixasse a tarefa de punir os egoístas para seus colegas. Assim, este indivíduo recebe todos os benefícios da cooperação

e não recebe a penalidade no *fitness* por ser punido nem paga o custo de punir. Em algumas gerações, este grupo de cooperadores não punidores deslocaria os altruístas e a população ficaria novamente vulnerável aos egoístas. Um modelo que utiliza a punição para justificar a sobrevivência do altruísmo deve ser resistente a essa invasão de egoístas de segunda ordem.

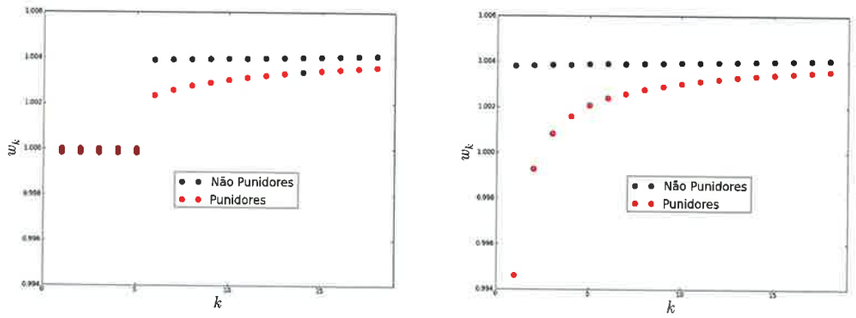
2.2 Modelo de Boyd - Gintis - Bowles

Como uma tentativa de resolver o problema do egoísmo de segunda ordem e apresentar a punição como solução para o problema na evolução do altruísmo, Boyd, Gintis e Bowles desenvolveram [12] um mecanismo de cooperação coordenada que funciona da seguinte forma: suponha uma população composta de um número muito grande de agentes organizados em grupos de tamanho $n = 18$, formados aleatoriamente. Existem dois alelos na população, o alelo P, de agentes punidores e o alelo N, de não-punidores. Os membros de um grupo participam de um jogo de bens públicos, como o empregado no experimento em [29] descrito na seção anterior, onde, a cada rodada, podem pagar c para contribuir uma quantidade b a um montante público, que será dividido entre todos os participantes. Assim, a cada uma das T rodadas, todos recebem bk/n, onde k é o número de indivíduos que pagaram o custo de contribuição c. A interação é repetida durante T rodadas, mas antes da interação há uma fase de sinalização onde os indivíduos de tipo P pagam um custo q para sinalizar sua intenção de punir. Na primeira rodada de interação, não-punidores nunca cooperam e punidores cooperam sempre que o número de agentes que sinalizaram na primeira fase for maior do que um limiar τ pré-definido. Toda vez que cooperam, punem aqueles que não o fizeram. O efeito da punição é fixo, u, e o custo por não-punidor é dado por c_P/k^a, de forma que existe um custo fixo por não-punidor, c_P, mas este é dividido entre todos os punidores a uma escala dada por a.

Nas $T-1$ rodadas seguintes, indivíduos do tipo P cooperam e punem se mais que τ agentes cooperaram e puniram nas rodadas anteriores e não cooperam e não punem se este limiar não foi atingido. Indivíduos do tipo N cooperam apenas se foram punidos na rodada anterior e nunca punem. Após as T rodadas, os grupos são desmanchados, cada indivíduo deixa um número de descendentes (com o mesmo alelo, não há mutação) igual ao seu *payoff*, que fará o papel do *fitness* neste modelo, e novos grupos são formados ao acaso a partir da nova geração, de forma semelhante ao modelo de *haystack*.

No modelo BGB, a cooperação ou não são estratégias adotadas pelos indivíduos em face à punição (ou falta de). Além disso, qualquer agente (do tipo P ou N), que em uma certa rodada irá cooperar, possui uma probabilidade e de erroneamente não o fazer. Isto é feito para modelar erros de percepção e transmissão de informação em uma situação de cooperação (por exemplo, agentes com intenção de cooperar não o fazem por erro) e o quanto isto afeta o equilíbrio. A estratégia Tit-for-Tat descrita na seção anterior é um exemplo de

estratégia dominante no dilema do prisioneiro que, apesar disso, é completamente invalidada se for introduzida uma taxa de erro como descrita. Daí a importância de se estudar o comportamento de um mecanismo em face à erros.



(a) Modelo BGB com fase de coordenação (b) Modelo BGB sem fase de coordenação

Figura 2.1: Comparação do payoff (em função do número de punidores) para ambos os tipos de agente no modelo BGB original e o *payoff* do modelo sem fase de coordenação. Observe que a diferença entre os *payoffs* antes do limiar ser atingido é significativa e faz com que altruístas raros sejam extintos rapidamente segundo a equação do replicador (1.1).

A fase de coordenação anterior à interação resolve os problemas comuns em mecanismos de punição de diversas formas. Caso os punidores não estejam em número suficiente (maior que τ), arcarão apenas com o custo de sinalização q, que é pequeno. Após passar o limiar, a punição (dependendo do valor de τ) se torna eficaz e o custo, dividido entre punidores, se torna eficiente. A figura 2.1 ilustra o efeito da coordenação nos *payoffs*. Em uma situação de indivíduos que punem indiscriminadamente, agentes punidores isolados sofreriam uma pressão evolutiva enorme e seriam extintos em poucas gerações. Com a fase de coordenação, os indivíduos do tipo P se reproduzem ligeiramente menos que os não punidores, e isso dura até que eventualmente um grupo, por sorte, atinja o limiar de punidores. A partir deste ponto, os punidores começam a punir em grupos suficientemente grandes e conseguem sobreviver.

O modelo também elimina o problema de cooperadores não punidores, pois estes não sinalizam e logo não são considerados na fase de coordenação. Como estes indivíduos são castigados se não cooperam da mesma forma que não cooperadores tradicionais, eles deslocam esta população, não a população de punidores.

Da mesma forma, um indivíduo com estratégia de "mentir" na fase de coordenação, que sinaliza intenção de punir mas não o faz, pagará o custo q mas após a primeira rodada os punidores perceberão que estão em número menor do que τ e pararão de cooperar e punir.

Assim, este agente "mentiroso" receberá os benefícios da cooperação na primeira rodada sem receber nenhuma punição, e nas rodadas seguintes receberá o mesmo *payoff* que os outros não-cooperadores. É possível mostrar [13] que este tipo de indivíduo não consegue proliferar quando $q > u$, condição que será pressuposta em diante.

Pode-se calcular analiticamente os pontos de equilíbrio deste modelo. Os *payoffs* de um agente punidor P e de um agente não-punidor N em um grupo contendo j outros indivíduos do tipo P são dados por:

$$W_P(j) = \begin{cases} W_0 - q, & \text{se } j < \tau \\ W_0 - q + (1-e)\left(\frac{b(j+1)}{n} - c\right) - \frac{e(j+1)+(n-j-1)c_P}{(j+1)^a} - eu & \\ +(T-1)\left((b-c)(1-e) - \frac{nec_P}{(j+1)^a} - eu\right), & \text{de outro modo} \end{cases} \tag{2.1}$$

$$W_N(j) = \begin{cases} W_0, & \text{se } j \leq \tau \\ W_0 - u + (1-e)\frac{bj}{n} + (T-1)\left((b-c)(1-e) - eu\right), & \text{de outro modo} \end{cases} \tag{2.2}$$

onde W_0 é o *payoff* base se não houver interação. Para determinar qual é a fração p de altruístas para a qual a população entra em equilíbrio, é necessário calcular os *payoffs* médios (sobre todos os grupos) da população. Assumindo grupos formados de forma aleatória, tem-se:

$$\overline{W}_P(p) = \sum_{j=0}^{n-1} W_P(j)\mathbb{P}(\text{Bin}(n-1,p) = j)$$

$$\overline{W}_N(p) = \sum_{j=0}^{n-1} W_N(j)\mathbb{P}(\text{Bin}(n-1,p) = j), \tag{2.3}$$

onde $\mathbb{P}(\text{Bin}(n,p) = j)$ é a probabilidade de uma variável aleatória com distribuição binomial de N tentativas e probabilidade p de sucesso obter j sucessos. Esta probabilidade é dada por

$$\mathbb{P}(\text{Bin}(n,p) = j) = \binom{n}{j} p^j (1-p)^{n-j}. \tag{2.4}$$

Escrevendo $\overline{W}_P(p) = W_P$ para não carregar a notação tem-se, segundo (1.1),

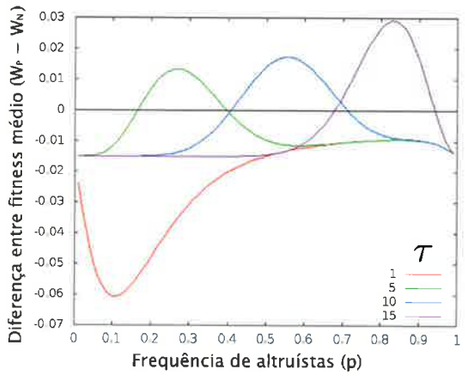


Figura 2.2: Diferença $W_P - W_N$ dos *payoffs* médios de cada estratégia em função de p, para diferentes valores de τ, calculados com os parâmetros padrão definidos no texto. Os equilíbrios não triviais da dinâmica se dão quando $W_P - W_N = 0$. $W_p - W_N$ define a alteração em p, portanto um zero da função com tangente positiva define um equilíbrio instável e um zero com tangente negativa define um equilíbrio estável. Não há equilíbrio para $\tau = 1$ com estes parâmetros, e portanto o altruísmo não é viável neste caso.

$$\Delta p = p(1-p)\frac{W_P - W_N}{pW_P + (1-p)W_N}. \tag{2.5}$$

Como mencionado anteriormente, esta equação possui os seguintes pontos fixos: $p = 0$, onde todos os agentes são não cooperadores, $p = 1$, onde todos os agentes são punidores, e p^* tal que $W_P(p^*) - W_N(p^*) = 0$.

A figura 2.2 mostra a diferença $W_P(p) - W_N(p)$ para alguns valores de τ. Quando a diferença é negativa, o *fitness* médio dos egoístas é maior do que a dos punidores e p diminui. Caso contrário, p aumenta. É claro então que alguns valores de τ apresentam dois equilíbrios diferentes (além dos triviais). Caso p^* seja um equilíbrio e $W_P - W_N$ seja positivo para p um pouco menor que p^*, então p^* é um equilíbrio estável, pois caso p aumente Δp será negativo e fará com que p diminua, e caso p fique menor que p^*, aumentará. Assim, p^* é um equilíbrio estável. Da mesma forma, se a diferença é negativa para valores menores que p^*, este é um equilíbrio instável.

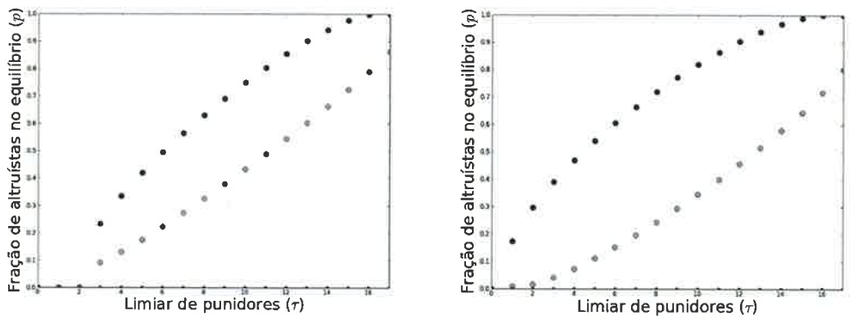
O cálculo dos pontos de equilíbrio foi feito numericamente por Boyd, Gintis e Bowles e os resultados principais encontram-se nas figuras 2.3 a 2.6. Os parâmetros padrões escolhidos pelos autores e sua justificativa encontram-se a seguir. Salvo menção contrária, todo cálculo numérico do texto é feito com estes parâmetros.

- $W_0 = 1$, o que significa que na ausência de interação, cada indivíduo tem um descendente que sobrevive até a fase reprodutiva.
- $c = 0,01$. Todos os outros parâmetros serão dados em função de c, e ele é escolhido desta forma para que a força da seleção seja da ordem de 1% da reprodução natural, definida por W_0.
- $b = 2c$ ou $b = 4c$, para que o altruísmo, pela regra de Hamilton, seja viável apenas para irmãos ou meio irmãos. Isso implica que se o modelo for bem sucedido com coeficientes de parentesco menores que $r = 0,25$, já é superior aos modelos de seleção de parentesco padrões.
- $u = 1,5c$. Para garantir que os indivíduais cooperem quando punidos, u deve ser maior do que o custo de se cooperar, $c - b/n$.
- $c_P = u$, para que o custo de punição por não-punidor seja o mesmo que a penalidade recebida ao ser punido, o que coloca punidores raros em desvantagem.
- $q = u$, para que não sejam possíveis estratégias mentirosas em que um agente sinaliza sem intenção de punir.
- $a = 2$, para que o custo de punição caia rapidamente com o número de punidores.
- $e = 0,1$, para simular uma taxa de erro não desprezível.

Os pontos cheios dos gráficos são os equilíbrios estáveis e os pontos vazios são equilíbrios instáveis. Como pode-se ver, nem sempre existem equilíbrios polimórficos instáveis ou estáveis (o que é mais acentuado para certas escolhas de parâmetros, como mostram as figuras 2.4, onde se remove a dependência quadrádica do custo de punição), mas a maior parte dos valores de τ possui equilíbrios estáveis com $p > 0$. Isto é muito interessante pois mostra que o mecanismo de fato é eficaz em garantir a manutenção da cooperação. Uma população que está em um equilíbrio estável não obtém fixação de egoístas.

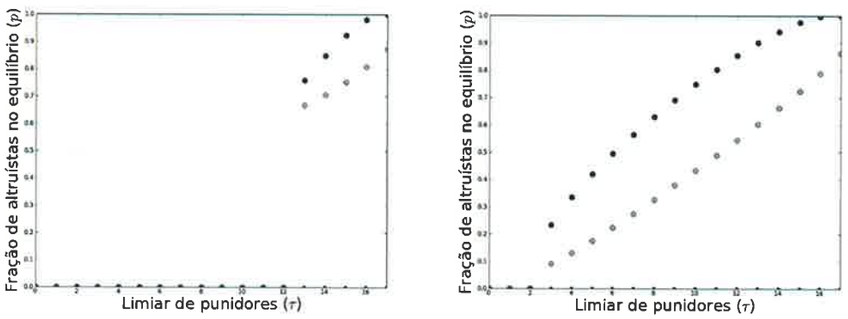
Mas enquanto o modelo descrito acima garante a manutenção da cooperação, não garante, sobre qualquer combinação de parâmetros, a sua seleção, isto é, a possibilidade de agentes altruístas invadirem uma sociedade de não cooperadores. É possível observar que em todos os casos onde há um equilíbrio polimórfico estável p_e^*, existe também um $p_i^* < p_e^*$ que é instável. Logo, uma fração $p \approx 0$ de punidores estaria abaixo do equilíbrio instável e seria extinta.

Como já mencionado na seção anterior, um aspecto básico que possibilita a evolução da cooperação é o *assortment* positivo, no qual agentes com o mesmo alelo (no nosso caso, P ou N) possuem uma maior probabilidade de interagirem do que seria esperado ao acaso. Boyd, Gintis e Bowles inseriram um coeficiente de parentesco r positivo em seu modelo, definido da



(a) b=2c (b) b=4c

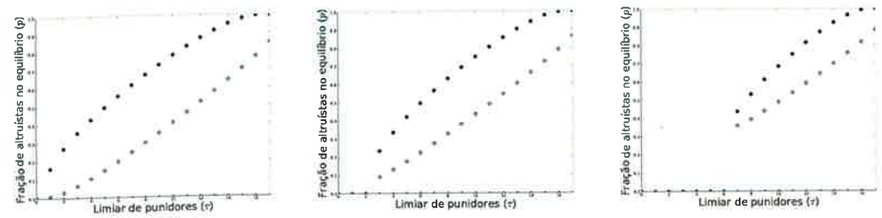
Figura 2.3: Pontos de equilíbrio do modelo BGB em função de τ, o limiar no qual altruístas começam a punir os egoístas, para dois valores de b. Os pontos representam os equilíbrios não triviais (isto é, $0 < p < 1$) da equação (2.5), calculados através dos zeros da diferença $W_P - W_N$ como mostra a figura 2.2. Os pontos cheios representam equilíbrios estáveis e os pontos vazios equilíbrios instáveis. Para os valores de τ onde não há equilíbrios não triviais (por exemplo $\tau < 3$ para $b = 2c$), o altruísmo não é viável. Observe que em ambas as figuras, a manutenção do altruísmo é garantida, mas sua evolução é impossível pois qualquer valor de p menor que o valor do equilíbrio instável implica na extinção do altruísmo.



(a) $a = 1$ (b) $a = 2$

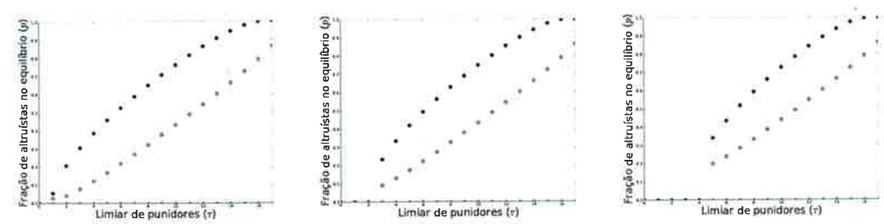
Figura 2.4: Pontos de equilíbrio do modelo BGB em função de τ para dois valores de a, o expoente do termo $1/(j+1)^a$ em (2.1) que divide o custo de punição entre todos os altruístas. Os pontos cheios representam equilíbrios estáveis e os pontos vazios equilíbrios instáveis. Observa-se que no modelo em que o custo de punição diminui linearmente com o número de punidores, altruístas só conseguem sobreviver quando em grande maioria.

seguinte forma: suponha que em uma dada geração, um grupo é sorteado aleatoriamente e os seus integrantes são ordenados de forma aleatória. O grupo escolhido é dito o grupo focal e o



(a) $e = 0.01$ (b) $e = 0.1$ (c) $e = 0.2$

Figura 2.5: Pontos de equilíbrio do modelo BGB em função de τ para três frequências de erro (e) diferentes. Os pontos cheios representam equilíbrios estáveis e os pontos vazios equilíbrios instáveis. Observa-se que uma taxa alta de erros (que simula tanto erros de cooperação de fato, isto é, o agente teve intenção de cooperar mas não o fez, quanto erros de percepção, onde um agente de fato coopera mas outros indivíduos acreditam no contrário) faz com que o altruísmo se torne menos viável. Ainda assim, ao contrário do mecanismo *Tit-for-Tat* de Axelrod [6] mencionado na introdução, o altruísmo no modelo BGB não colapsa imediatamente quando o erro é introduzido nas interações.



(a) $c_P = u/2$ (b) $c_P = u$ (c) $c_P = 2u$

Figura 2.6: Pontos de equilíbrio do modelo BGB em função de τ para três valores de c_P, o custo fixo de punição por agente egoísta a ser castigado. Os pontos cheios representam equilíbrios estáveis e os pontos vazios equilíbrios instáveis. Conforme o custo fixo de punição aumenta, o altruísmo se torna menos viável, sendo completamente inviável para limiares (τ) baixos.

primeiro indivíduo da ordenação do grupo é dito o indivíduo focal. Então definimos o evento A_1 caso o focal seja punidor e N_1 caso não o seja. Observe que esta ordenação aleatória é equivalente a pegar um indivíduo do grupo ao acaso, porém permite uma notação mais consistente. A probabilidade do segundo agente nesta ordenação (dito cofocal) ser punidor (evento A_2) ou não punidor (N_2) é dada por:

$$\begin{aligned} \mathbb{P}(A_2|A_1) &= r + (1-r)p, & \mathbb{P}(N_2|A_1) &= (1-r)(1-p) \\ \mathbb{P}(A_2|N_1) &= (1-r)p & \mathbb{P}(N_2|N_1) &= r + (1-r)(1-p) \end{aligned} \tag{2.6}$$

Colocando r em evidência chega-se à seguinte definição alternativa:

$$r = \mathbb{P}(A_2|A_1) - \mathbb{P}(A_2|N_1) \tag{2.7}$$

Os autores então implementaram um coeficiente r positivo definindo a probabilidade de um grupo com um altruísta possuir k outros punidores como

$$\mathbb{P}(\mathrm{Bin}(n, r + (1-r)p) = k) \tag{2.8}$$

ao invés de $\mathbb{P}(\mathrm{Bin}(n,p) = k)$. Isto constitui uma aproximação não trivial, que será discutida mais adiante. A figura 2.7 mostra os equilíbrios obtidos numericamente utilizando os mesmos parâmetros da figura 2.3, mas com um valor de $r = 0,07$.

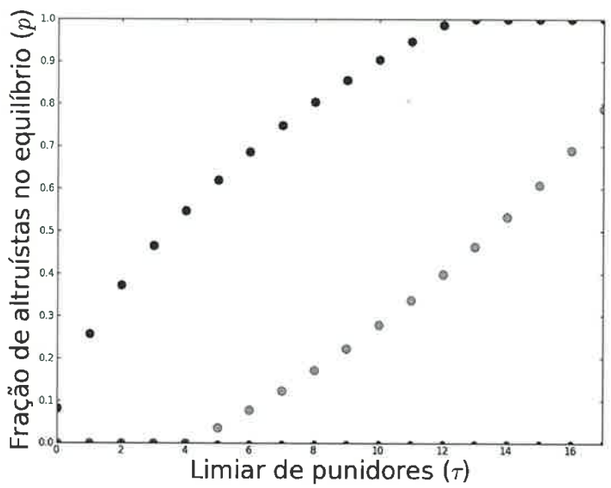


Figura 2.7: Pontos de equilíbrio do modelo BGB em função de τ para $b = 4c$ e $r = 0,07$. Os pontos cheios representam equilíbrios estáveis e os pontos vazios equilíbrios instáveis. Com a introdução do *assortment* positivo no sorteio dos grupos através de (2.8), obtém-se equilíbrios estáveis sem a presença de equilíbrios instáveis em frequências menores. Isto significa que é possível que o altruísmo evolua a partir de qualquer valor não nulo de p.

Como é possível ver na figura 2.7, para valores de $\tau < 5$ não existe um equilíbrio instável, apenas estáveis. Isso significa que para qualquer fração $0 < p < 1$, o sistema entrará em um equilíbrio polimórfico com sobrevivência de altruísmo. Isso é importante pois se suposermos

que os primatas começam em um estado ancestral dominado por não-cooperadores, o surgimento do altruísmo através de mutações genéticas é possível segundo este modelo. Mais interessante ainda: o coeficiente de parentesco estimado para grupos caçadores coletores é de aproximadamente $0,054$ [27], um número muito baixo para ser explicado apenas por seleção de parentesco, mas compatível com os resultados do modelo BGB.

Calculamos o coeficiente de parentesco mínimo necessário segundo o modelo BGB para, dado um valor de τ, ser possível o altruísmo se propagar a partir de altruístas raros. Isto é feito utilizando os parâmetros padrão escolhidos por Boyd, Gintis e Bowles e para cada τ aumenta-se o valor de r até que não haja equilíbrio instável e sobre apenas o equlíbrio estável não trivial. A figura 2.8 mostra os valores de r_s para cada τ. Como pode-se ver, ele começa com valores compatíveis com $0,054$ para τ baixo, mas rapidamente cresce para valores não razoáveis.

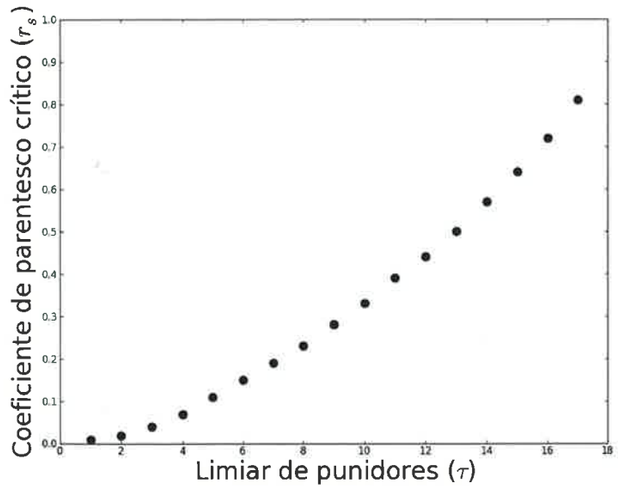


Figura 2.8: Valores de r_s em função de τ para $b = 4c$. Estes valores representam o r mínimo necessário para que o alelo altruísta possa se propagar para qualquer valor de $p > 0$. Estes valores foram encontrados numericamente para cada valor de τ tomando $r = 0$ no cálculo de $W_P - W_N$ (figura 2.2), aumentando gradativamente em incrementos de $0,005$ e tomando o último falor de r para o qual a equação do replicador possui apenas um equilíbrio não trivial, sendo este estável.

A análise de Boyd, Gintis e Bowles, no entanto, é essencialmente macroscópica. Ao fazer a análise através dos valores médios de W_P e W_N, desprezam-se flutuações em p. Isso é equivalente a supor que o número de altuístas iniciais é pequeno, mas macroscópico (isto é, da ordem de 1%, não um único indivíduo). Não é exatamente o que desejamos modelar. Se o comportamento cooperativo surge através de mutações, deve ser viável que ele sobreviva a

partir de um único indivíduo.

A aproximação feita na introdução do *assortment* positivo também deve ser avaliada. Ao substituir $\mathbb{P}(\text{Bin}(n-1,p)=k)$ por $\mathbb{P}(\text{Bin}(n-1,r+(1-r)p)=k)$ e $\mathbb{P}(\text{Bin}(n-1,(1-r)p)=k)$ nas equações (2.3), supõe-se que a probabilidade dos $n-1$ outros indivíduos serem altruístas ou não é independente. Explicitamente, os autores fazem a suposição de que, se A_3 for evento em que o terceiro indivíduo na ordenação aleatória do grupo focal seja do tipo A e A_2 e A_1 são os eventos em que o focal e o cofocal são altruístas, então

$$\mathbb{P}(A_3A_2|A_1) = \mathbb{P}(A_3|A_1)\mathbb{P}(A_2|A_1) = (p+r(1-p))^2. \tag{2.9}$$

Isto é, pressupõem que após o primeiro altruísta em um grupo, os seguintes aparecem de forma independente e com probabilidade que depende apenas da existência de ao menos um indivíduo do tipo A no grupo. Observe que não é possível calcular o valor de $\mathbb{P}(A_3|A_2A_1)$ analiticamente pois não se tem a distrbuição conjunta $\mathbb{P}(A_3,A_2,A_2)$. De fato, este é um problema diretamente proveniente de se implementar o coeficiente de parentesco no modelo a partir das definições (2.6). Como elas tratam da conjunta entre A_1 e A_2, é necessário fazer uma aproximação para se obter a conjunta de n indivíduos.

Esses dois problemas principais da dinâmica do BGB são inerentes a diversos modelos de biologia evolutiva. Como uma tentativa de contorná-los, foi desenvolvido o processo Fisher-Wright em dois níveis [49].

2.3 O Modelo Fisher-Wright

O modelo Fisher-Wright (também chamado de Wright-Fisher) [50] foi a primeira dinâmica probabilística de difusão gênica em uma população, desenvolvido por Fisher [30] e Wright [53] no início da década de 30, e será a base para o processo Fisher-Wright em dois níveis que será apresentado na próxima seção. Neste modelo, o processo de difusão de alelos em uma população ocorre de forma equivalente a um sorteio com reposição de bolas em uma caixa. Supondo que existam n indivíduos haplóides na população, sendo n fixo para todas as gerações, e cada indivíduo possui a variante A ou N de um mesmo gene. Então, no processo Fisher-Wright, cada indivíduo de uma nova geração possui o alelo A com probabilidade n_A/n, onde n_A é o número de indivíduos da geração anterior que possuem este alelo. Da mesma forma, um indivíduo da nova geração possui a variante N com probabilidade $1-n_A/n$.

Observe que neste modelo não há pressão evolutiva, cada indivíduo da nova geração escolhe aleatoriamente um pai da geração anterior e seu alelo é idêntico ao do seu pai, o que significa que também não há mutação. Dado que em uma geração t existam $n_A(t)$ indivíduos

com o alelo A, a probabilidade de que k indivíduos sejam do tipo A na geração seguinte é dada, assim como no sorteio de bolas em uma caixa com reposição, pela distribuição binomial de n tentativas com k sucessos e probabilidade $n_A(t)/n$:

$$\mathbb{P}\left(n_A(t+1)=k|n_A(t)\right) = \binom{n}{k}\left(\frac{n_A(t)}{n}\right)^k\left(\frac{1-n_A(t)}{n}\right)^{n-k} \tag{2.10}$$

É fácil calcular o valor esperado de $n_A(t+1)$, dado por $\mathbb{E}(n_A(t+1)|n_A(t)) = n_A(t)$, o que é esperado de um processo de difusão de alelos sem seleção. Observe que existe uma probabilidade não nula de que $n_A(t+1)=0$, e, por não haver mutação, este é um ponto absorvente da dinâmica. O mesmo ocorre quando $n_A=n$. Para n finito, o sistema percorre todos os pontos do espaço de fase e, portanto, certamente cai em um destes dois pontos, ditos pontos de fixação, em tempo finito. É simples mostrar [28] que por dispersão gênica (isto é, sem seleção ou mutação) a probabilidade do alelo do tipo A fixar a partir de uma população com k indivíduos com este alelo é dada por k/n. Como o alelo ou fixa ou se extingue para n finito, a probabilidade de extinção (e, consequentemente, a probabilidade do alelo N se fixar) é dada por $1-k/n$.

Para incluir seleção e mutação na dinâmica, é necessário alterar a probabilidade com a qual um indivíduo da nova geração escolhe seu pai na geração anterior. Ao invés de escolher de forma aleatória uniforme (e portanto, com probabilidade k/n de escolher um tipo A), as probabilidades são proporcionais ao *fitness* normalizado dos indivíduos. Se v_A for a diferença de *fitness* entre indivíduos do tipo A e do tipo N (A podendo ser um alelo deletério, caso $v_A<0$), m_1 a fração de alelos do tipo A que mutam para a variante N ao serem transmitidos de pai para filho e m_2 a taxa de mutação de tipos N para alelos do tipo A, então a probabilidade de cada indivíduo da nova geração ser do tipo A é dada por:

$$p_k = \frac{(1+v_A)k}{(1+v_A)k+N-k}(1-m_1) + \frac{N-k}{(1+v_A)k+N-k}m_2, \tag{2.11}$$

que deve ser substituído na equação (2.10) no lugar de n_A/n, obtendo como distribuição do número de alelos do tipo A na geração seguinte uma binomial com n tentativas e probabilidade p_k de sucesso.

O modelo Fisher-Wright com seleção porém sem mutação é a base da dinâmica do processo Fisher-Wright em dois níveis, que será descrito na próxima seção.

2.4 Análise do Processo 2lFW Para o Modelo BGB

O processo Fisher - Wright de dois níveis, proposto em [49], é um mecanismo interessante por formular uma descrição microscópica analiticamente tratável para uma população estruturada em grupos com migração, na qual o *assortment* positivo surge como consequência desta estrutura.

A dinâmica consiste em uma população divida em g grupos contendo n indivíduos cada, onde n é um inteiro qualquer e supõe-se $g \to \infty$. Os indivíduos são haplóides e existem duas variantes do gene na população: o alelo comum N que leva ao comportamento não-cooperador, e o alelo mutante altruísta A, inicialmente raro. A reprodução é, como no BGB, assexuada e sem mutações, e os *fitness* de um altruísta e de um não-altruísta em um grupo com k indivíduos do tipo A[1] são dados, respectivamente, por

$$w_k^A = 1 + \delta v_k^A, \quad w_k^N = 1 + \delta v_k^N, \tag{2.12}$$

onde a dependência do número de altruístas no *fitness* individual está em v_k. A taxa fixa nestas equações possui o mesmo papel de W_0 no modelo BGB, é o *fitness* individual base, caso não houvesse interação. As funções v_k^A e v_k^N representam o benefício (ou custo) no *fitness* de ser altruísta ou não-altruísta ao interagir com outros indivíduos e o parâmetro δ regula a força da seleção, em um papel semelhante ao parâmetro c no modelo BGB. Se $\delta = 0$, não há seleção e os grupos sofrem disperção gênica similar ao modelo Fisher-Wright original.

Além do *fitness* individual, cada grupo possui um *fitness* igual a média de seus indivíduos. Definimos que um grupo é do tipo k se este possui k indivíduos do tipo A. Um grupo do tipo k então possui *fitness*

$$\overline{w}(k) = \frac{k w_k^A + (n-k) w_k^N}{n}. \tag{2.13}$$

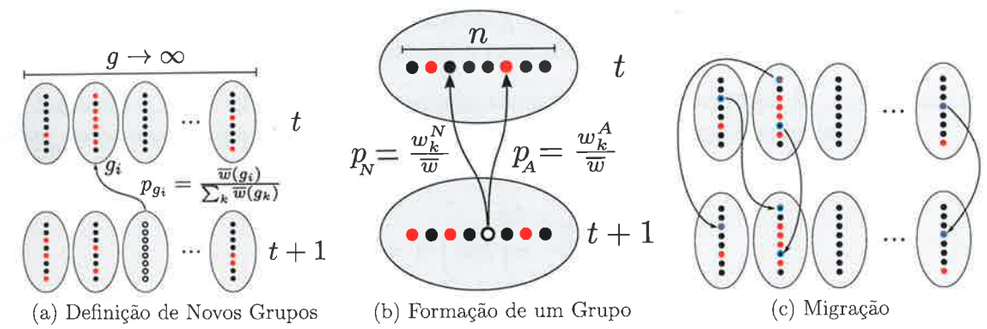
Dado o *fitness*, a seleção é feita através de dois processos do tipo Fisher-Wright distintos, como descritos na seção anterior, com seleção porém sem mutação. Cada grupo na nova geração possui um grupo paterno da geração anterior escolhido independentemente com probabilidade proporcional ao *fitness* dos grupos da geração anterior. Isto é, se w_l é o *fitness* do grupo l, então um grupo l' da geração seguinte o escolhe como paterno com probabilidade

[1] Note que enquanto no modelo BGB, j era o número de outros altruístas, isto é, $j = 1, \ldots, n-1$, neste *framework* k é o número total de altruístas no grupo, incluindo o focal.

$$\mathbb{P}(l' = l) = \frac{w_l}{\sum_j w_j}. \tag{2.14}$$

Definido o seu grupo paterno, cada um dos n indivíduos deste novo grupo define o seu pai escolhendo de forma independente com probabilidade proporcional ao *fitness* dos indivíduos do grupo paterno. Como não há mutação, um novo indivíduo sempre possui a mesma variante que seu pai.

O parâmetro m representa a taxa de migração. Após os novos grupos serem formados neste processo, seleciona-se uma fração m da população inteira (isto é, dos ng indivíduos) e faz-se uma permutação aleatória deste subconjunto, de forma a se manter constante o tamanho n de cada grupo. Isto é uma forma de modelar a migração dos agentes, e o parâmetro m controla o quão forte ou fraca é a migração. A taxa de migração separa as regiões de viabilidade e inviabilidade da variante A.



(a) Definição de Novos Grupos (b) Formação de um Grupo (c) Migração

Figura 2.9: Esquema da dinâmica do mecanismo Fisher-Wright em dois níveis. Na figura a) os grupos paternos são definidos através de um processo Fisher-Wright em que cada grupo escolhe seu paterno com probabilidade proporcional ao *fitness* (médio) de cada grupo da geração anterior. Na figura b), uma vez definido o paterno, o grupo da nova geração define seus integrantes através de outro processo Fisher-Wright e cada indivíduo escolhe seu pai com probabilidade proporcional ao *fitness* dos agentes da geração anterior. Na figura c), após os dois processos Fisher-Wright, uma fração m dos indivíduos da população é definida como migrante e sofre uma permutação aleatória.

w_k^A e w_k^N definem o modelo que será analisado com o *framework*. No caso do BGB, é o jogo dos bens públicos com a fase de coordenação, economia de escalas na punição e introdução de erros na cooperação. Para demonstrar a versatilidade do mecanismo, a análise será inicialmente feita para funções de *fitness* arbitrárias. Pressupõe-se, sem perda de generalidade, que o *fitness* em um grupo sem altruístas é dado por $w_0^N = 1$ (o que implica $v_0^N = 0$).

Até agora as funções v_k^A e v_k^N são gerais. Apesar de estarmos discutindo o altruísmo

e nomeando os indivíduos como tal, o processo Fisher-Wright de dois níveis como descrito acima não faz absolutamente nenhuma suposição sobre as funções e pode ser utilizado para estudar a difusão de alelos que descrevem qualquer tipo de conmportamento que seja descrito por funções de fitness adequadas. Para obter resultados específicos relacionados ao altruísmo, serão feitas duas restrições sobre v_k^A e v_k^N.

- $v_1^A < 0$, o que implica $w_1^A < 1 = w_0^N$. Isso significa que um altruísta isolado em um grupo sempre tem menos descendentes que os não altruístas.
- $v_n^A > 0$, o que implica $w_n^A > 1 = w_0^N$. Isso significa que os membros de um grupo composto apenas por altruístas sempre deixam mais descendentes do que um grupo composto apenas por não altruístas.

Apenas como um exemplo, o *fitness* do jogo de bens públicos utilizado no modelo BGB onde cada altruísta gera um benefício b para os outros a um custo c para si pode ser representado neste *framework* da seguinte forma:

$$v_k^A = -c + \frac{kb}{n}, \qquad (2.15)$$

$$v_k^N = \frac{kb}{n}.$$

É fácil ver que as duas condições definidas acima são satisfeitas com este modelo. Um estudo mais detalhado do jogo de bens públicos implementado no mecanismo Fisher Wright em dois níveis pode ser encontrado em [49].

Estamos interessados em saber para quais condições de migração (isto é, para quais valores de m) o altruísmo é viável dados v_k^A e v_k^N. A análise do BGB será feita transformando os *payoffs* (2.1) e (2.2) em funções de *fitness* adequadas. Supondo um estado ancestral dominado pela variante N, deseja-se definir viabilidade a partir de um único mutante A. É claro que este alelo pode sumir nas primeiras iterações por puro acaso, mesmo com uma grande pressão evolutiva a favor de altruístas (isto é, mesmo com $v_2^A \gg v_2^N$). Qualquer afirmação sobre sobrevivência neste modelo deve então ser probabilistica.

O altruísmo é viável quando a probabilidade de que o altruísmo se espalhe não vá a zero na aproximação $g \to \infty$, que será suposta durante todo este trabalho salvo menção contrária. Quando $m = 1$, a variante A se extingue com probabilidade 1 pois, segundo a primeira condição sobre v_k^A, $v_1^A < 0$ e logo, $w_1^A < 1$. Como g é muito grande, a taxa

de migração garante que os descendentes do mutante original acabam isolados em grupos distintos. Se $N^A(t)$ for o número de altruístas total na geração t, então $m = 1$ garante que $\mathbb{E}(N^A(t)) = (w_1^A)^t \to 0$. No caso $m = 0$, o tamanho finito de n possibilita que um grupo acabe com n altruístas em poucas gerações, e grupos gerados através deste possuem aptidão média $w_n^A > 1 = w_0^N$, logo o mutante A certamente viável. Estas deduções serão feitas com maior cuidado, mas fica claro que o importante é entender o comportamento quando $0 < m < 1$.

Para entender o que ocorre nas primeiras gerações, deve-se notar o seguinte: como é pressuposto que $g \to \infty$, e apenas um grupo possui um altruísta, nas primeiras gerações o *fitness* médio de todos os grupos é dado em excelente aproximação por $\overline{w}_0 = w_0^N = 1$. Nas primeiras gerações, portanto, a probabilidade de um grupo escolher um paterno do tipo k é dada por $\overline{w}_k/g$. Se $N(t) = (N_1(t), \ldots, N_n(t))$ for um vetor com entradas $N_k(t)$ representando o número de grupos de tipo k na geração t, então a aproximação de $\overline{w}_0 = 1$ implica que os grupos com $k \geq 1$ geram sucessores de forma independente, sem interferirem uns com os outros e é possível definir um processo de ramificação [35] no vetor de tipos $N(t)$. Neste formalismo, é possível, enquanto o número de grupos com altruístas for muito menor do que g e a aproximação $\overline{w} \approx 1$ for boa, calcular o crescimento do altruísmo como se segue.

Suponha que um grupo da nova geração tenha um parental do tipo k. A probabilidade deste grupo ser do tipo k' nesta fase inicial é dada pela distribuição binomial:

$$p(k, k') = \mathbb{P}\left(\mathrm{Bin}\left(n, \frac{kw_k^A}{n\overline{w}_k}\right) = k'\right), \qquad (2.16)$$

pois cada indivíduo do novo grupo escolhe de forma independente seu pai no grupo paterno, logo cada indivíduo possui probabilidade $\frac{kw_k^A}{kw_k^A + (n-k)w_k^N}$ de ser do tipo A. Como cada grupo do tipo k é escolhido como parental com probabilidade $\overline{w}_k/g$, o número de grupos gerados é descrito por uma distribuição $\mathrm{Bin}(g, \overline{w}_k/g)$, que possui média $\overline{w}_k$. Levando em conta (2.16), um grupo do tipo k cria em média $\overline{w}_k p(k, k')$ grupos do tipo k'.

Podemos estudar a dinâmica do sistema definindo um vetor de tipos $N(t) = (N_1(t), \ldots, N_n(t)$, onde $N_k(t)$ é o número de grupos de tipo k na geração t. Assim, definindo a matriz promotora do processo (*driving matrix*) M

$$M_{k,k'} = \overline{w}_k p(k, k'), \qquad (2.17)$$

então para $m = 0$, $N(t)$ pode-se calcular o valor esperado de $N(t+1)$ da forma

$$\mathbb{E}(N(t+1)|N(t)) = N(t)M. \tag{2.18}$$

Para introduzir uma taxa de migração não nula, observa-se que, para $g \to \infty$, na fase inicial a probabilidade da permutação aleatória remover um altruísta de um grupo e colocá-lo em outro grupo com tipo maior que zero é extremamente baixa, pois estes grupos são extremamente raros. Assim, os altruístas removidos geram todos grupos do tipo 1. A matriz que define a remoção destes agentes é dada por

$$A_{k',k''} = \mathbb{P}(\mathrm{Bin}(k', 1-m)) = k''). \tag{2.19}$$

A saída de altruístas devido à migração se dá através da equação $E(N(t+1)|N(t)) = N(t)A$. Cada altruísta em um grupo do tipo k' possui uma probabilidade $(1-m)$ de permanecer no grupo e não sofrer migração. Por isso, a probabilidade de que após a migração permanecam k'' altruístas é dada por uma distribuição binomial de k' tentativas e probabilidade de sucesso $(1-m)$.

A criação de grupos do tipo 1 por altruístas migrantes pode ser representada pela matriz

$$B_{k',k''} = \begin{cases} mk', & \text{se k"=1,} \\ 0, & \text{de outro modo.} \end{cases} \tag{2.20}$$

Assim, um grupo do tipo k' gera em média mk' grupos do tipo 1, pois cada altruísta migrante cai em um grupo de tipo 0. O valor esperado de $N(t+1)$ com migração é dado então por

$$\mathbb{E}(N(t+1)|N(t)) = N(t)M(A+B). \tag{2.21}$$

Como $N^A(t) = N_1(t) + 2N_2(t) + \ldots nN_n(t)$, é claro que a sobrevivência do altruísmo depende do vetor $N(t)$ neste processo dinâmico. Como a matriz $M(A+B)$ possui apenas entradas estritamente positivas, o Teorema de Perron-Frobenius [35] garante que existe um autovalor ρ real, positivo único, com módulo maior do que os outros autovalores do espectro de $M(A+B)$ e cujo autovetor correspondente possui todas as componentes estritamente positivas. Seja ν este autovetor, normalizado de forma a representar uma distribuição de probabilidades (isto é, $\nu_1 + \ldots + \nu_n = 1$). Uma consequência conhecida do Teorema de Perron-Frobenius é que pode-se descrever a evolução do sistema apenas por ρ e ν, da forma:

$$\mathbb{E}(N(t)) = N(0)(M(A+B))^t = C\rho^t\nu, \tag{2.22}$$

para algum número positivo C. Isso significa que a distribuição de altruístas entra no equilíbrio ν após um certo tempo e cresce à uma taxa ρ a cada geração, mantendo a proporção entre os grupos segundo ν. É importante observar que este equilíbrio é válido para a fase inicial, mas após uma quantidade grande de gerações, suficiente para que $N(t)$ convirja para $\nu(t)$. Esta fase onde o número de gerações é alto o suficiente para a distribuição dos tipos de grupo ser ν mas baixo o suficiente para que $N_A \ll ng$ é dita fase inicial estacionária.

É simples, dada a equação (2.22), definir um critério para a evolução do altruísmo. Lembrando que a matriz $M(A+B)$ depende de m e, consequentemente, ρ também é uma função de m, o alelo A possui probabilidade positiva de sobreviver se

$$\rho(m) > 1. \tag{2.23}$$

A taxa crítica de migração m_s é encontrada então resolvendo a equação

$$\rho(m_s) = 1. \tag{2.24}$$

Agora é possível justificar os argumentos sobre a migração nos extremos. Quando $m = 0$ a matriz $M(A+B)$ se reduz à matriz M, cujo autovalor é $w_n^A > 1$, e o autovetor correspondente é $(0, \ldots, 1)$. Logo, $m_s > 0$, não importando o modelo.

Quando $m = 1$, a matriz $A = 0$ e $M(A+B) = MB$, que possui apenas as entradas da primeira coluna não nulas. Isso significa que o seu autovetor deve ser do tipo $(a, 0, 0, \ldots, 0)$ e, consequentemente, $\nu = (1, 0, \ldots, 0)$. Mas então $(1, 0, \ldots, 0)MB = (w_1^A, 0, \ldots, 0)$ e como por definição $w_1^A < 1$, isso significa que o altruísmo não sobrevive. Logo $m_s < 1$ não importa o modelo. Isso implica que para cada par de funções de *fitness* existe uma taxa de migração crítica $0 < m_s < 1$ para a qual o altruísmo é viável se $m < m_s$.

Uma forma alternativa desta condição de viabilidade pode ser deduzida da seguinte forma: na fase inicial estacionária, se $N(t) = C\nu$ para alguma constante C, então $\mathbb{E}(N(t+1)) = C\rho\nu$. Logo, $\mathbb{E}(N^A(t+1)) = C\rho\sum_k k\nu_k$. Mas como descrito acima, na fase inicial um indivíduo do tipo A em um grupo com k altruístas gera em média w_k^A descendentes, portanto a identidade $N^A(t) = C\sum_k k\nu_k$ leva a identidade $\mathbb{E}(N^A(t+1)) = C\sum_k w_k^A k\nu_k$. Igualando estas duas equações, chega-se a uma condição alternativa de viabilidade:

$$\frac{\sum_k w_k^A k\nu_k}{\sum_k k\nu_k} = \rho > 1. \tag{2.25}$$

A igualdade acima mostra que o autovalor de Perron-Frobenius da matriz $M(A+B)$ é a média de w_k^A na população quando atingido o equilíbrio da fase inicial estacionária ν e para o altruísmo ser viável esta média deve ser maior do que *fitness* médio dos agentes de tipo N (que é igual a w_0^N devido aos agentes do tipo N serem a grande maioria na fase inicial). Esta condição é equivalente, para qualquer valor de $\delta > 0$, a

$$\frac{\sum_k v_k^A k\nu_k}{\sum_k k\nu_k} > 0 \tag{2.26}$$

O cálculo dos autovalores de $M(A+B)$ pode ser feito numericamente. Para isso, é necessário definir os modelos v_k^A e v_k^N a serem estudados. O objetivo deste trabalho de Mestrado foi utilizar o modelo BGB descrito na seção anterior como um exemplo de aplicação do processo Fisher-Wright em dois níveis. Para isso, os *payoffs* do BGB (eqs. (2.1) e (2.2)) podem ser escritos como funções v_k^A e v_k^N da seguinte forma:

$$v_k^A = \begin{cases} -q, & \text{se } k \leq \tau, \\ -q - \frac{ek+(n-k)c_P}{k^a} + (1-e)\left(\frac{bk}{n} - c\right) - eu + & \\ (T-1)\left((b-c)(1-e) - \frac{nec_P}{k^a} - eu\right), & \text{de outro modo.} \end{cases} \tag{2.27}$$

$$v_k^N = \begin{cases} 0, & \text{se } k \leq \tau, \\ -u + \frac{bk}{n} + (T-1)\left((b-c)(1-e) - eu\right), & \text{de outro modo,} \end{cases} \tag{2.28}$$

onde $n = 18$, como no trabalho de Boyd, Gintis e Bowles. Observe que $\delta = 1$ e estas funções são numericamente idênticas as funções do BGB para qualquer composição de grupos.

W_0 não é utilizado pois entra como o termo unitário fixo na definição de w_k^A. Caso seja necessário alterar o valor da aptidão base, isto é feito alterando o valor de δ.

Utilizando os parâmetros padrão novamente, calcula-se m_s para diferentes valores de τ (figura 2.10). O valor de $\delta = 1$ foi preservado para manter os *payoffs* individuais idênticos ao modelo BGB.

É interessante ressaltar as diferenças entre esta análise e a análise original feita em [12], descrita na Seção 2.2. Embora o modelo estudado seja o mesmo, a saber, a punição coordenada com economia de escala e erros de cooperação descrita por Boyd, Gintis e Bowles, a dinâmica da evolução é diferente. Algumas das vantagens do mecanismo Fisher-Wright em dois níveis já foram ressaltadas, mas é claro agora que a dinâmica do BGB é essencialmente a

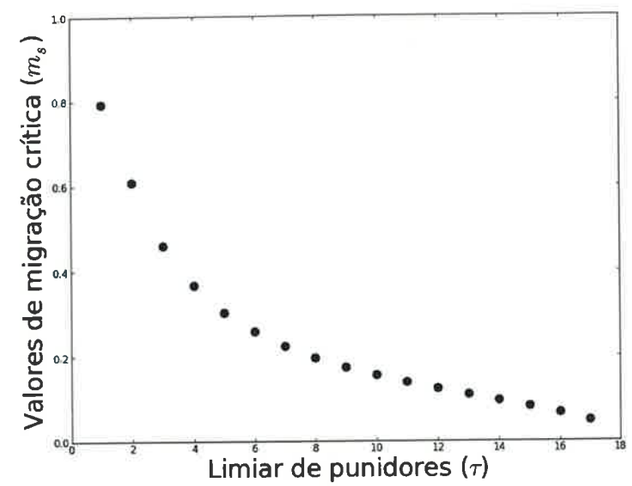


Figura 2.10: Taxas de migração críticas em função de τ para o processo 2lFW, calculadas encontrando m tal que o autovalor de Perron-Frobenius da matriz $M(A+B)$ seja unitário (equação (2.24)), com os parâmetros idênticos ao modelo BGB com $b = 4c$. Para $\tau = 1$, por exemplo, o altruísmo possui probabilidade não nula de sobreviver até mesmo com uma taxa de migração de 80%.

descrita aqui, porém com $m = 1$. Como visto, essa configuração não pode suportar evolução da cooperação, independentemente do modelo utilizado, sendo necessária a introdução, à mão, do *assortment* positivo para contornar este problema. Mas esta introdução é feita de forma aproximada e gera problemas. Daí um dos interesses neste mecanismo, que permite que o grau de parentesco emerja da dinâmica. O parâmetro utilizado para regular a difusão do alelo altruísta e o *assortment* positivo é a taxa de migração m.

2.5 Regime de Seleção Fraca

A relação entre o coeficiente de parentesco e m, no entanto, não é simples de ser definida em um caso de seleção genérico, mas é fácil de calcular em um regime de seleção fraca ($\delta \to 0$). Esta aproximação, além de permitir que se defina r em função de m, dá origem a uma condição de viabilidade mais simples. Na equação (2.26), se tomarmos $\delta \to 0$, ela se torna

$$\frac{\sum_k v_k^A k \nu_k^0}{\sum_k k \nu_k^0} > 0, \tag{2.29}$$

onde ν^0 é o autovetor (de Perron-Frobenius) da matriz $M^0(A+B)$, onde a matriz M^0 é a matriz M com $\delta \to 0$, que não depende das funções w_k^N e w_k^A, apenas de n e m. Esta simplificação é importante pois ν_k^0 pode ser calculado para qualquer modelo, e a condição de viabilidade consiste apenas em calcular o valor médio de v_k^A em uma população com distribuição demográfica ν^0. Observe que esta condição independe de v_k^N.

Pode-se definir uma distribuição π_k dada por

$$\pi_k = \mathbb{P}_t^0(K^D = k), \text{ para } t \gg 1, \tag{2.30}$$

onde K^D é o número de indivíduos do tipo A que são idênticos por descendência ao focal. Idêntico por descendência (*identical by descent* ou IBD) significa que, ao seguir a linhagem de ambos nas gerações anteriores, encontra-se um ancestral comum antes de um evento de migração (ver figura 2.11).

É claro que $k^D \leq k$ e na fase inicial $k^D = k$. Assim, π_k é a probabilidade de, dado que o focal é altruísta, o seu grupo ser do tipo k. Para isso, deve-se levar em conta o viés de tamanho. π_k é dado então por:

$$\pi_k = \frac{k\nu_k^0}{\sum_{k'} k' \nu_{k'}^0} \tag{2.31}$$

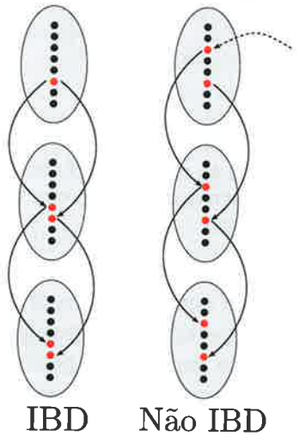


Figura 2.11: Esquema demonstrando a definição de indivíduos X e Y idênticos por descendentes. Dois agentes são idênticos por descendentes se é possível seguir a linhagem de ambos e encontrar um ancestral comum antes de um evento de migração.

Com esta definição, pode-se escrever a condição de viabilidade da seguinte forma alternativa

$$\sum_k v_k^A \pi_k > 0. \tag{2.32}$$

Em nossa aplicação do modelo BGB, basta calcular π_k numericamente e utilizar v_k^A para calcular a condição de viabilidade nesta aproximação. Na figura 2.12, pode-se ver a comparação dos valores de m_s encontrados originalmente segundo (2.26) e encontrados com a aproximação de seleção fraca. A proximidade dos valores de m_s

para todos os valores de τ permite concluir que é a aproximação de seleção fraca é bastante razoável. Isto já era previsto pois Boyd, Gintis e Bowles escolheram o parâmetro $c = 0,01$ com a intenção de que o efeito do altruísmo na sobrevivência de descendentes fosse da ordem de 1%.

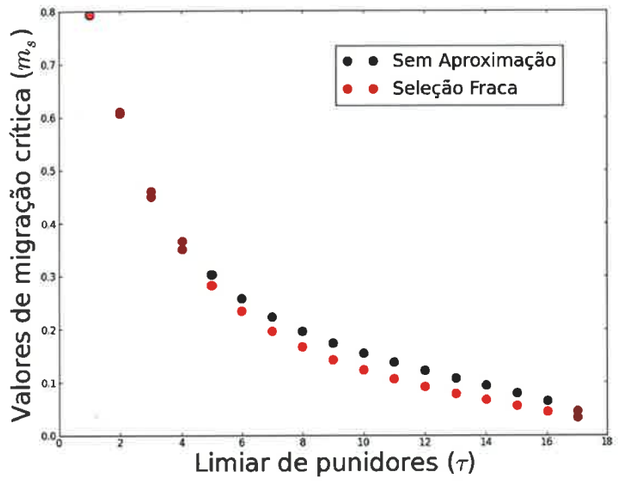


Figura 2.12: m_s em função de τ calculado pela condição (2.32) no regime de seleção fraca, em comparação com o obtido na seção anterior demonstrado na figura 2.10. Observa-se que a aproximação $\delta \to 0$ é bastante razoável, o que justifica o cálculo de r_s para o processo 2IFW que é feito em (2.35).

A relação entre o coeficiente de parentesco e a taxa de migração no regime de seleção fraca é feita da seguinte forma: define-se uma grandeza R_t dita coeficiente de parentesco genético definida por:

$$R_t = \mathbb{P}_t(D|A_1), \tag{2.33}$$

onde D representa o evento do cofocal ser idêntico por descendentes ao focal.

Na fase inicial, é simples mostrar que $r_t = R_t$. Basta ver que $r_t = \mathbb{P}_t(A_2|A_1) - \mathbb{P}_t(A_2|N_1) = \mathbb{P}_t(A_2|A_1)$, pois $\mathbb{P}_t(A_2|N_1)$ é da ordem de $p(t)$, a fração de altruístas na sociedade, que é muito pequeno na fase inicial.

Mas $r_t = \mathbb{P}_t(A_2|A_1) = \mathbb{P}_t(A_2 D|A_1) + \mathbb{P}_t(A_2 \overline{D}|A_1)$, onde $\overline{D}$ é a negação de D, isto é, o evento onde o focal e cofocal não são idênticos por descendência. Mas como descrito ao justificar o uso do formalismo de processos de ramificação, na fase inicial a probabilidade

de um evento de migração colocar um altruísta em um grupo de tipo $k \geq 1$ é nula. Isso significa que $\mathbb{P}_t(A_2\overline{D}|A_1) = 0$ na fase inicial e $r_t = \mathbb{P}_t(A_2D|A_1) = \mathbb{P}_t(D|A_1) = R_t$. A segunda igualdade é evidente pelo fato de que não existe mutação no modelo e os indivíduos são haplóides. Logo, se o focal é do tipo A e é idêntico por descendência ao co-focal, este deve ser também altruísta.

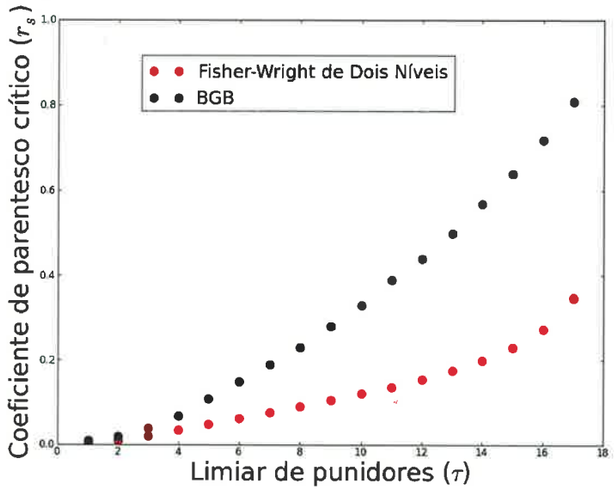


Figura 2.13: Comparação de r_s obtido pela dinâmica do BGB e descrito na figura 2.8 e pelo modelo Fisher-Wright em dois níveis com a aproximação $\delta \to 0$, definido pela equação (2.35). É possível ver que a introdução *ad-hoc* do *assortment* positivo no modelo BGB faz com que o valor necessário de r para que o altruísmo seja viável seja muito maior do que a implementação via migração do processo 2lFW.

O problema de calcular r_t se reduz ao problema de calcular R_t. Mas na aproximação de $\delta \to 0$, isto é fácil. Se o focal ou co-focal forem migrantes, eles não são idênticos por descendência. Se são ambos não migrantes, eles escolheram o mesmo parente com probabilidade $1/n$ (pois $w_k^A \approx w_k^N$) e são IBD. Com probabilidade $1 - 1/n$ eles escolheram parentes diferentes e são IBD se seus parentes forem IBD. Logo,

$$R_{t+1}^0 = (1-m)^2 \left(\frac{1}{n} + \left(1 - \frac{1}{n}\right) R_t^0 \right) \tag{2.34}$$

Na fase inicial estacionária, é possível mostrar [49] que R_t^0 independe de t. Logo, para esta fase (onde a distribuição de tipos é dada por ν^0) pode-se resolver (2.34) em função de R^0 e, usando que $R^0 = r^0$, tem-se

$$r^0 = \frac{(1-m)^2}{n - (n-1)(1-m)^2} \tag{2.35}$$

Com esta relação, é possível calcular os coeficientes mínimos de parentesco necessários para a sobrevivência do altruísmo, e compará-los com os do BGB, como mostra a figura 2.13

Pode-se ver que a dinâmica descrita pelo modelo Fisher-Wright em dois níveis não só possui um coeficiente de parentesco que surge de forma natural e evolui com o tempo, mas o coeficiente mínimo necessário para que o altruísmo seja viável é menor do que no modelo BGB, diferença que aumenta bastante conforme o valor de τ cresce. Além disso, mesmo no caso ideal $\tau = 1$, o r_s do Fisher Wright é menor do que o do BGB. Isto é um resultado extremamente interessante.

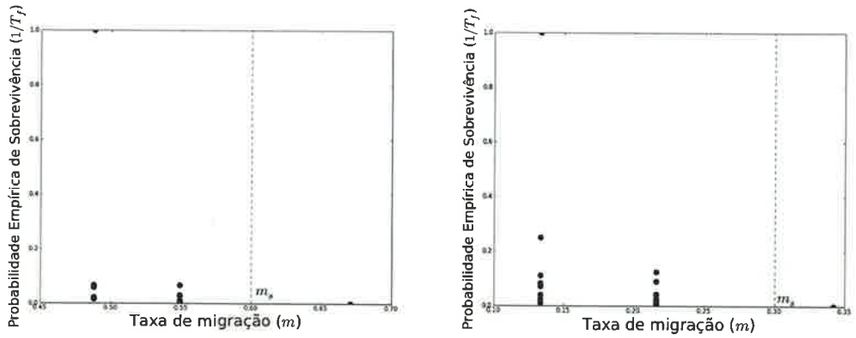
2.6 Simulação

Para corroborar os resultados obtidos analiticamente, foi feita a simulação da dinâmica como descrita no início da seção. No entanto, a simulação apresenta diversas dificuldades a serem levadas em conta.

Primeiramente, o número de grupos não é infinito. Isto cria um efeito de escala finita onde a transição ao redor do m_s não é bem definida. Isto é complicado pelo fato de que o altruísmo é considerado viável no modelo se existe uma probabilidade não nula de que ele não se extingua na fase inicial. É de se esperar então que esta probabilidade seja muito pequena para $m \to m_s$, e a consequência é que se observa uma grande quantidade de simulações onde o altruísmo morre mesmo quando viável.

Além disso, há uma complicação adicional onde, com m perto de m_s, mesmo quando os agentes do tipo A não morrem antes da fase inicial estacionária o equilíbrio polimórfico resultante possui uma fração p^* de altruístas muito baixa. Quando $g \to \infty$ isto não é um problema pois a partir do momento em que $N_A(t) \gg 1$ as flutuações são desprezíveis quando comparadas a N_A e este equilíbrio polimórfico, embora baixo, permanece constante. No entanto, como g é finito, se p^* é muito baixo o sistema certamente atingirá o estado $N_A(t) = 0$ em pouco tempo devido às flutuações, e este estado é absorvente. Essa situação gera a seguinte complicação: se o número de gerações na simulação for muito alto, corre-se o risco do altruísmo morrer devido a flutuações causadas por g finito. Se o número de gerações for baixo, corre-se o risco de não atingir o equilíbrio e se obter falsos positivos.

O que optou-se em fazer na simulação foi, ao invés de tentar obter exatamente o com-



(a) $\tau = 2$ (b) $\tau = 5$

Figura 2.14: Probabilidades Empíricas de Sobrevivência para alguns valores da taxa de migração m para dois valores do limiar de punidores τ. Estas probabilidades foram estimadas rodando uma simulação da dinâmica com $n = 18$ agentes por grupo e $g = 10000$ grupos, até o primeiro evento de sobrevivência, representado por um ponto no gráfico. Isto foi repetido dez vezes para cada m, e a probabilidade empírica é dada por $1/T_f$, onde T_f é o número de simulações necessárias para que, ao final de 30000 gerações, exista ao menos um alelo A na população. As simulações são compatíveis com os valores teóricos para m_s.

portamento do sistema na vizinhança de m_s, buscar evidências que corroborem os resultados teóricos, rodando a simulação com valores um pouco abaixo e um pouco acima de m_s e se observou a probabilidade empírica de sobrevivência, dada por $1/T_f$, onde T_f é o tempo até o primeiro evento de sobrevivência. Este tempo foi medido da da seguinte forma: rodou-se a simulação partindo de um apenas um altruísta e com $n = 18$ e $g = 10000$. O tempo até o primeiro evento de sobrevivência é o número de instâncias do programa rodadas até que houvesse sobrevivência do altruísmo, e o critério de sobrevivência foi se após 30000 gerações do processo Fisher-Wright em dois níveis ainda havia altruístas na sociedade.

Para cada valor de m foram medidos dez valores de T_f e com a dez estimativas de $1/T_f$ construiu-se a figura 2.14. A migração na simulação foi implementada sorteando-se aleatoriamente mng agentes da população e contando o número de altruístas neste grupo. Cada um destes mng indivíduos recebe um alelo A ou N seguindo o procedimento de retiradas de bolas de uma caixa sem reposição. Isto significa que se havia K altruístas entre os mng indivíduos (supondo N_m como o maior inteiro menor que mng), então o primeiro indivíduo escolhido recebe o alelo A com probabilidade K/N_m e recebe o alelo N com probabilidade $1 - K/N_m$. Se este indivíduo se tornar um tipo A, o segundo escolhido recebe o alelo A com probabilidade $(K-1)/(N_m-1)$, mas se se tornar um tipo N, o segundo indivíduo recebe o alelo A com probabilidade $K/(N_m-1)$. Esse procedimento segue até os N_m indivíduos tenham sido sorteados, e os K alelos do tipo A tenham sido distribuidos aleatoriamente entre os indivíduos migrantes.

Como pode-se ver, as simulações corroboram fortemente os resultados teóricos. Embora seja difícil obter valores precisos na vizinhança de m_s, os resultados obtidos mostram que de fato existe uma transição em torno de m_s. Uma investigação mais cuidadosa exigiria que se aumentasse tanto o número de grupos, para diminuir o efeito de tamanho finito e consequentemente diminuir o efeito de flutuações que atingem o estado absorvente $p = 0$, e o número de repetições, para que se evitem falsos negativos quando a probabilidade de sobrevivência do altruísmo é muito baixa.

2.7 Modelo de Limiar

Uma questão interessante ao estudar o modelo BGB é entender o quão crucial é a fase de coordenação para a evolução do altruísmo em comparação com as outras características do modelo, como economia de escala e sensibilidade a erros. Isto é feito aproximando as funções de fitness para um modelo limiar da forma

$$v_k^A = \begin{cases} -c, & \text{se } k \leq \tau, \\ -c + a, & \text{de outro modo.} \end{cases} \tag{2.36}$$

$$v_k^N = \begin{cases} 0, & \text{se } k \leq \tau, \\ a', & \text{de outro modo,} \end{cases} \tag{2.37}$$

que tenta abstrair apenas a fase de sinalização onde punidores decidem cooperar ou se abster. Comparamos as taxas de migração críticas desta aproximação com os resultados das funções de *fitness* (2.1) e (2.2). O valor de c a ser utilizado é bem evidente, o custo obrigatório aos altruístas no BGB é o custo de sinalização q. Para a e a', é necessário definir qual valor de k será utilizado para definir o benefício fixo. Optou-se por fazer a aproximação para os dois casos limite, $k \to \tau + 1$ e $k \to n$.[2] Desta forma, as aproximações são dadas por:

[2] A notação $k \to n$ foi escolhida para não causar confusão com o número real de altruístas no grupo.

$k \to n$

$$v_k^A = \begin{cases} -q, & \text{se } k \leq \tau, \\ -q + T\left((b-c)(1-e) - \frac{ec_P}{n^{a-1}} - eu\right), & \text{de outro modo.} \end{cases} \tag{2.38}$$

$$v_k^N = \begin{cases} 0, & \text{se } k \leq \tau, \\ -u + (1-e)b + (T-1)\left((b-c)(1-e) - eu\right), & \text{de outro modo.} \end{cases} \tag{2.39}$$

$k \to \tau + 1$

$$v_k^A = \begin{cases} -q, & \text{se } k \leq \tau, \\ -q - \frac{e(\tau+1)+(n-\tau-1)c_P}{(\tau+1)^a} + (1-e)\left(\frac{b(\tau+1)}{n} - c\right) - eu + \\ (T-1)\left((b-c)(1-e) - \frac{nec_P}{(\tau+1)^a} - eu\right), & \text{de outro modo.} \end{cases} \tag{2.40}$$

$$v_k^N = \begin{cases} 0, & \text{se } k \leq \tau, \\ -u + \frac{bk}{n} + (T-1)\left((b-c)(1-e) - eu\right), & \text{de outro modo.} \end{cases} \tag{2.41}$$

Calculamos os valores de m_s através da condição de viabilidade (2.26), conforme mostrados na figura 2.15. É possível observar que os valores de m_s (e, consequentemente, r_s) da aproximação limiar e do modelo BGB original são bastante parecidos, o que indica que estrutura fundamental do modelo que permite a evolução da cooperação é mesmo a coordenação de altruístas, o que é uma observação bastante interessante.

A distribuição no limite de seleção fraca π_k, embora comum a todos os modelos, ainda precisa ser encontrada numericamente para cada par de parâmetros m e n. Deseja-se encontrar uma forma analítica para que a condição de viabilidade possa ser calculada para qualquer fitness v_k^A sem a necessidade de cálculos numéricos.

Se for assumido que $n \to \infty$, os cumulantes da distribuição π_k descrevem uma distribuição Beta com parâmetros $\alpha = 1$ e $\beta = 2mn$ [49], cuja função densidade de probabilidade é dada por $f_{m,n} = 2mn(1-x)^{2mn}$, $x = k/n$. Assim,

$$\mathbb{P}\left(\frac{K^D}{n} > x\right) \approx \int_x^1 f_{m,n}(x')dx' = (1-x)^{2mn} \tag{2.42}$$

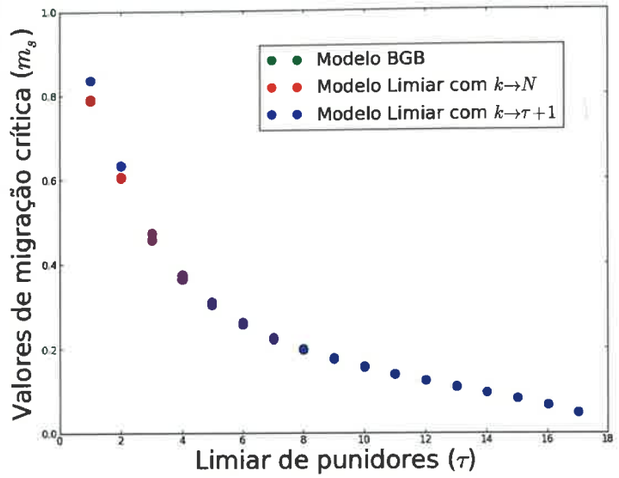


Figura 2.15: Comparação da taxa de migração crítica m_s da aproximação de limiar com os resultados da figura 2.10. Ambas as aproximações são muito boas, o que indica que a característica crucial do modelo BGB que permite a evolução e manutenção do altruísmo é a fase de coordenação.

Infelizmente, não é possível ainda assim encontrar uma forma analítica para π_k. Para a aproximação limiar é possível calcular a condição de viabilidade da seguinte forma:

$$\sum_k \pi_k v_k^A > 0 \Rightarrow \sum_{k>\tau} \pi_k a > c \tag{2.43}$$

Se for feita a aproximação de $\sum_{k>\tau+1} \pi_k \to \int_{\tau+1}^1 f_{m,n}(x')dx' = (1 - \frac{\tau+1}{n})^{2nm}$, válida para $n \to \infty$, então a condição de viabilidade para a aproximação limiar pode ser escrita como

$$a(1 - \frac{\tau+1}{n})^{2nm_s} > c, \tag{2.44}$$

de onde obtemos que

$$m_s = \frac{\log(c/a)}{2n\log(1-(\tau+1)/n}, \tag{2.45}$$

que é uma equação fechada para m_s em função de a e c. A figura 2.16 mostra a comparação de m_s para a aproximação limiar no limite $n \to \infty$ e os resultados originais obtidos em 2.10. É possível ver que apesar da equação (2.45) envolver três aproximações distintas, a saber, $\delta \to 0$, $n \to \infty$ e a aproximação ao modelo limiar, o resultado ainda é bastante próximo enquanto o cálculo de m_s se tornou significativamente mais simples.

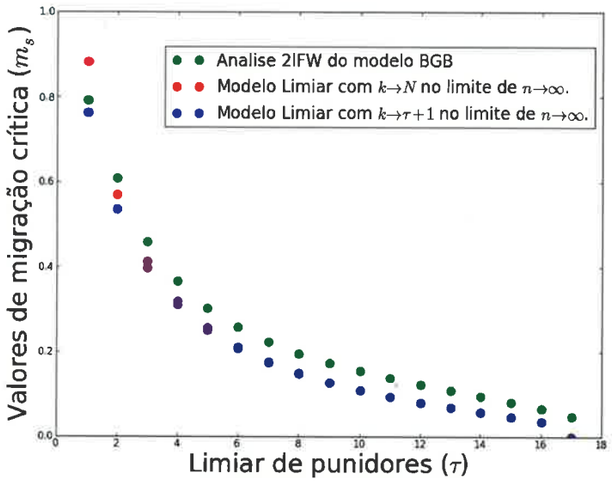


Figura 2.16: Comparação da taxa de migração crítica m_s da aproximação de limiar no limite de $n \to \infty$ com os resultados da figura 2.10. O cálculo da curva da aproximação limiar é exato, dado pela equação (2.45), enquanto o valor de m_s do modelo BGB é calculado numericamente através dos autovalores da matriz $M(A+B)$.

Capítulo 3

Dinâmica de Classificações Morais

Neste capítulo, apresentamos o modelo de interação de agentes, desenvolvido em [16] para explicar os resultados empíricos obtidos por Haidt em seus questionários, e seus resultados. A última seção detalha um estudo da dependência do modelo em relação às propriedades da rede topológica onde os agentes interagem, utilizando em particular a rede do Facebook, que é interessante por representar exatamente uma rede de interações sociais realista.

O modelo descrito a seguir foi definido através de regras microscópicas de interação entre os agentes, baseadas nas duas observações empíricas descritas na introdução. O comportamento macroscópico do sistema foi medido e comparado com os resultados dos questionários do Haidt, cujo tratamento é descrito a seguir.

3.1 Resultados Empíricos

O questionário feito por Haidt, que pode ser encontrado em [2], é dividido em duas partes. Na primeira, o participante deveria classificar de 0 a 5 a relevância de certas considerações ao decidir se algo é certo ou errado. Exemplos incluem "Se alguém mostrou ou não falta de lealdade", "Se alguém agiu ou não de acordo com as tradições da sociedade" e "Se alguém agiu ou não de uma forma que Deus aprovaria". Uma classificação 0 indicaria que a consideração é absolutamente irrelevante ao decidir se algo é certo ou errado, e uma classificação 5 indicaria que a consideração é extremamente relevante.

Na segunda parte, participantes deveriam classificar de 0 a 5 o quanto concordavam com certas afirmações. Exemplos incluem "Compaixão com aqueles que estão sofrendo é uma das virtudes mais cruciais.", "Eu tenho orgulho da história de meu país" e "Nunca é correto matar um ser humano". Uma classificação 0 indicaria total desacordo e uma classificação 5

total acordo com a afirmação.

Cada uma das considerações e afirmações estava relacionada a um dos fundamentos morais definidos por Haidt. Assim, para cada um dos $n = 14246$ participantes, calculou-se um vetor Z_i de cinco dimensões cujas entradas são as médias das respostas relacionadas a cada fundamento moral. No questionário, havia também um campo onde o participante deveria preencher sua filiação política (*political affiliation* ou p.a.) de 1 a 7, 1 sendo totalmente liberal e 7 sendo totalmente conservador.

Os vetores de classificação dos conservadores e ultraconservadores (p.a. $= 6$ e 7) apresentavam a menor variação e por isso foram escolhidos para definir a "classificação média" Z com a qual todos os outros participantes seriam comparados. Isto foi feito calculando o vetor J médio normalizado entre todos os participantes com p.a. 6 e 7. Para cada p.a., foi construído um histograma de $J_i \cdot Z$ com ambos os vetores normalizados. Estes histogramas podem ser vistos na figura 3.1 e serão comparados com os resultados obtidos pelo modelo que será apresentado na próxima seção.

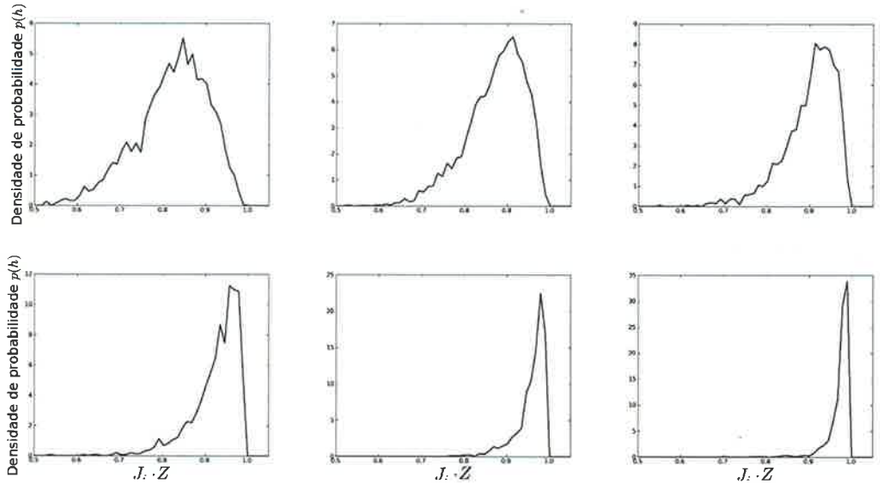


Figura 3.1: Histogramas do produto interno $J_i \cdot Z$ entre o vetor de classificação moral de cada indivíduo e o vetor médio Z dos conservadores, para filiações políticas de 1 a 6. Observa-se que os ultraliberais (p.a. $= 1$) são os que possuem maior variação de classificações, enquanto os conservadores (p.a.$= 6$) possuem uma menor variação de classificações. Por essa razão, seu vetor médio foi escolhido como orientação para o histograma.

3.2 Mecânica Estatística de Classificações Morais

Define-se uma rede (cuja importância da estrutura topológica é um dos objetos de estudo) com N agentes, cada agente i possui um vetor de classificação moral J_i $M-$dimensional normalizado, M sendo o número de fundações morais dos indivíduos na sociedade. Em alinhamento com o trabalho de Haidt, escolhemos utilizar $M = 5$, mas os resultados permanecem qualitativamente inalterados para valores de M próximos.

Dado um vetor de assunto a ser debatido x_μ, também normalizado e $M-$dimensional, o julgamento do agente i sobre o assunto μ é dada por $h_{i\mu} = J_i \cdot x_\mu \in [-1, 1]$. O vetor x_μ representa quais julgamentos morais o assunto envolve, podendo tratar apenas de uma classificação moral caso seja um vetor da base de $\mathbb{R}^M$, ou ser relevante para todas as classificações. Novamente em alinhamento com o trabalho de Haidt, a classificação do assunto não depende do indivíduo, ou seja, um assunto com componente positiva apenas em *Harm*, por exemplo, será visto desta forma por todos os agentes que irão discutí-lo.

Com estas definições, pode-se introduzir um custo[1] social de discordância entre dois agentes interagentes i e j dado por

$$V_\delta(h_i^\mu, h_j^\mu) = -\frac{1+\delta}{2} h_{i\mu} h_{j\mu} + \frac{1-\delta}{2} |h_{i\mu} h_{j\mu}|, \tag{3.1}$$

onde $\delta \in [0, 1]$ é dito o parâmetro de corroboração. Este custo modela a evidência empírica de que a rejeição, neste caso representada por agentes que classificam de forma diferente, causa uma sensação próxima da dor e, portanto, classificar de forma diferente de seus vizinhos incorre em um custo social. Esse custo pode ser entendido em um contexto de mecânica estatística do equilíbrio como a energia da interação entre os agentes i e j, e pode-se definir a energia total desta sociedade de agentes como

$$H = \sum_{(i,j)} \sum_{\mu} V_\delta(h_i^\mu, h_j^\mu), \tag{3.2}$$

onde a soma em (i, j) é a soma entre todos os pares de vizinhos na rede.

A intuição por trás da energia de interação definida é a seguinte: o campo $h_i^\mu = J_i \cdot x^\mu$ possui dois aspectos importantes ao modelo. O seu sinal representa a posição do agente i

[1] "Custo", neste contexto, possui um sentido diferente da Parte I. Por exemplo, o custo descrito na equação (3.1) pode ser positivo ou negativo, ao invés do custo c na Parte I, sempre negativo.

sobre o assunto μ, se ele concorda ($h_i^\mu > 0$[2]) ou discorda ($h_i^\mu < 0$). A magnitude do campo h_i^μ representa o quão importante o assunto é para o agente.

Um par de agentes que possuem a mesma posição sobre o assunto ($h_i^\mu h_j^\mu > 0$) terá uma energia de interação (sobre o assunto μ) $V_\delta = -\delta h_i^\mu h_j^\mu$, e agentes com opiniões contrárias ($h_i^\mu h_j^\mu < 0$) terá uma energia de interação $V_\delta = h_i^\mu h_j^\mu$. A interpretação desta energia é a seguinte: dois agentes com opiniões fortes ($|h_i^\mu| = 1$) e contrárias terão uma energia máxima, $+1$. Dois agentes com opiniões fortes e concordantes terão energia mínima, $-\delta$. Um agente indiferente ($h_i^\mu = 0$) ao assunto μ não contribuirá na Hamiltoniana, não importando a posição de seus vizinhos.

Com isto em mente, a interpretação de δ como um parâmetro de corroboração faz sentido. Agentes com $\delta = 0$ são indiferentes à interação com vizinhos que concordam com suas opiniões, apenas levam em conta a discordância. Por isso, são chamados no modelo de agentes *novelty seeking*, ou seja, que buscam novidades. Agentes com $\delta = 1$ consideram igualmente opiniões contrárias e opiniões corroborantes, portanto são ditos agentes *corroboration seeking*, que buscam corroboração. Fica claro, portanto, que o parâmetro δ serve para implementar no modelo a evidência empírica descrita de que há diferentes estilos cognitivos para se tratar informações novas e informações corroborantes. Pretende-se estabelecer uma relação numérica entre δ e a filiação política autodeclarada, que permitirá atribuir diferentes estratégias cognitivas a diferentes posições políticas, corroborando a evidência descrita.

O estado de equilíbrio deste modelo é encontrando através de simulação numérica utilizando o método de Monte Carlo Metropolis-Hastings [42], simulado em redes com valores de δ homogêneos. Essa aproximação é feita pois não se obtêm resultados qualitativamente diferentes ao rodar a simulação em redes heterogêneas e esta simplificação facilita a análise e comparação com os dados.

Além disso, análises qualitativas mostraram que quando existem vários assuntos sendo discutidos, o resultado é equivalente a se discutir um assunto (normalizado) $x = \sum_i x_i$, dito *Zeitgeist*. Portanto, a simulação numérica será feita com apenas um assunto discutido, o que facilita a análise e apresentação dos dados e não altera qualitativamente os resultados.

Feitas estas aproximações, fixa-se uma temperatura (α, neste modelo interpretada como a pressão social dos pares) e faz-se a simulação para cem valores de δ entre 0 e 1, espaçados em intervalos de $0,01$. A rede utilizada nas simulações originais em [16] foi a rede Barabási-Albert com $m = 8$. A rede Barabási-Albert com parâmetro m é construida da seguinte forma: gera-se uma rede inicial com $m_0 \geq m$ nós, com grau de cada nó no mínimo 1 (isto é, cada nó inicial tem ao menos um vizinho). Adiciona-se então um nó por vez na rede, e este nó é conectado a m nós anteriores, com probabilidade $p_i = \frac{k_i}{\sum_j k_j}$ para que o nó i seja incluso

[2]A definição de concordância na verdade possui simetria de paridade, podendo ser definida com sinais opostos. A definição presente, é, no entanto, a mais intuitiva.

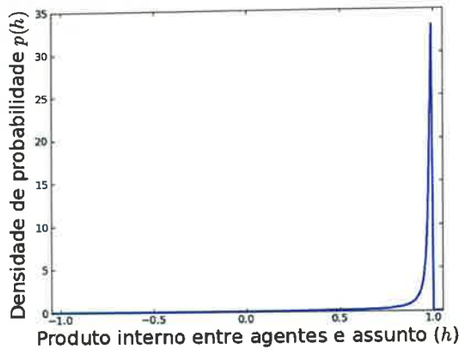


Figura 3.2: Exemplo de resultado de uma simulação do modelo. Fixado uma temperatura α e um valor do parâmetro de corroboração δ, inicia-se com agentes com vetores de opinião J_i aleatórios em uma rede (originalmente a rede Barabási-Albert com $m = 8$) e encontra-se o equilíbrio deste sistema através do método de Monte Carlo Metropolis-Hastings, com energia dada por (3.2). O resultado é visualizado através de um histograma normalizado dos campos $h_i = J_i \cdot x$, onde x é o assunto sendo discutido. Quanto menor a variância do histograma, maior o consenso entre os agentes.

na conexão com o novo nó. Itera-se este procedimento até que o número de nós na rede atinja o desejado. A rede resultante sempre possui distribuição de graus do tipo $P(k) \propto k^{-3}$, independente do valor de m utilizado. O valor de m define, entre outras coisas, o número médio de vizinhos por agentes, cuja importância será discutida adiante.

A rede Barabási-Albert é relevante pois exibe uma distribuição de graus do tipo lei de potência[3], que é uma distribuição comumente encontrada em redes que descrevem algum tipo de comportamento humano [16] [3]. Exemplos de rede do tipo lei de potência são: sites de internet (um link entre dois sites configura uma aresta direcionada), autoria de artigos em comunidades científicas, atores de Hollywood (trabalhar em um mesmo filme gera uma conexão entre dois atores), entre outros diversos exemplos [3].

Os resultados da simulação podem ser visualizados através de um histograma normalizado dos valores da opinião $h_i = J_i \cdot x$, que possui todas as informações relevantes sobre os agentes e permite a visualização rápida do consenso ou dissenso na sociedade, como mostra a figura 3.2. Observa-se nos resultados (quer serão discutidos em mais detalhes adiante) que quanto maior o δ, maior o consenso dos agentes sobre o assunto discutido.

Com os resultados da simulação, pode-se estabelecer a relação entre o parâmetro δ e a filiação política auto declarada nas respostas do questionário do Haidt. Para isso, roda-se a simulação para centenas de valores de δ entre 0 e 1, obtendo diversos histogramas de forma

[3]Isto é, uma distribuição de graus da forma $P(k) \propto k^{-\gamma}$

similar ao mostrado na figura 3.2. Os dados provenientes de [32] são transformados em histogramas conforme descrito na Seção 3.1 (ambos os histogramas normalizados com área unitária), e para cada valor de afiliação política (p.a.), mede-se a distância entre o histograma dos dados e cada um dos histogramas gerados pela simulação com valores de δ diferentes. A distância usada é a distância euclidiana, isto é, dados dois histogramas h e g, a distância entre eles é definida por

$$d(h,g) = \sum_i (h_i \Delta h_i - g_i \Delta g_i)^2 \tag{3.3}$$

Encontra-se então o valor δ que corresponde à distância mínima ao histograma de cada filiação política. Com isso se obtém, para cada valor de α, um gráfico da relação entre as filiações políticas e o parâmetro de corroboração (figura 3.3).

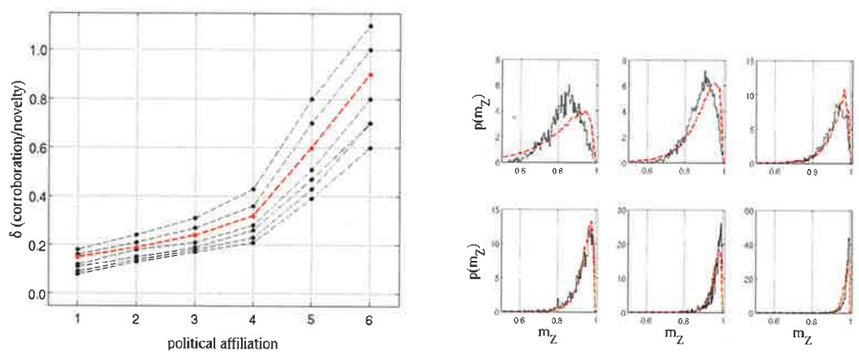


Figura 3.3: A primeira figura mostra a relação entre as relações políticas e o parâmetro de corroboração δ para diferentes valores de α, para a rede Barabási-Albert com M=8 e a segunda figura mostra a comparação dos histogramas simulados com melhor aproximação e os dados de cada filiação. A comparação entre p.a. e δ foi feita, para cada p.a., minimizando a distância (definida por (3.3) entre os histogramas dos dados com os histogramas gerados para diversos valores de δ. Para $\alpha = 8$, a figura mostra a comparação dos histogramas representando os dados dos questionários e o histograma simulado que miniza a distância. Os histogramas são muito parecidos, o que corrobora a idéia de que as classificações morais segundo o modelo de Haidt são formadas a partir de interação entre indivíduos. Figura retirada de [16].

Por fim, faz-se a comparação entre os histogramas obtidos através dos dados e os histogramas obtidos na simulação com a melhor aproximação, como mostra a figura 3.3.

Esta comparação é bastante interessante, em alguns casos com uma concordância quase perfeita com os dados. Isso favorece a idéia de que as classificações morais feitas pelos entrevistados nos questionários de Haidt [32] se devem parcialmente à interação dos indivíduos

com os seus pares, e que é possível associar a estratégia cognitiva (isto é, o parâmetro de corroboração δ) de cada indivíduo com a sua filiação política, corroborando a evidência empírica citada.

3.3 Redes Barabási-Albert e Facebook

Os resultados descritos acima, como já mencionado, foram obtidos utilizando uma rede Barabási-Albert com $m = 8$. Esta rede foi utilizada por possuir um algoritmo computacional extremamente eficiente de se gerar redes com distribuição de grau do tipo lei de potência que, como justificado, é o tipo de rede mais comum observada em interações humanas.

Neste trabalho estudamos uma topologia social mais realista, mais especificamente a do Facebook. Para isso, foi utilizado o material suplementar ao trabalho de Traud et al. [51], que inclui matrizes de conexão do Facebook[4] para cinco universidades americanas [5]. Utilizando o pacote de C igraph [1], converteram-se as matrizes em grafos não direcionados onde dois nós são vizinhos caso sejam amigos no Facebook.

Uma das questões mais importantes que surge ao se trabalhar com estes dados é se os resultados são robustos às diferentes redes de universidades, isto é, se redes distintas do Facebook são topologicamente semelhantes, uma condição importante para que faça sentido se estudar "a rede" do Facebook. Para descobrir isso, rodamos as simulações descritas na seção anterior para as redes transformadas em grafos da forma descrita. A figura 3.4 mostra o histograma do produto interno entre as classificações e o assunto discutido nestas redes, estudadas em uma simulação com $\alpha = 8$, $\delta = 0.4$.

As diferenças nas simulações entre as redes são muito pequenas, como é possível ver na figura. O único resultado razoavelmente distoante das demais é o histograma com a rede do Caltech, cuja diferença se deve ao fato de possuir uma ordem de grandeza a menos de agentes e metade do número de vizinhos médios para cada agente (a influência disto nos resultados será vista adiante). Ainda assim, quando comparada à rede Barabási-Albert utilizada nas simulações originais, percebe-se que mesmo com resultados diferentes das demais, há muito pouca diferença entre a rede do Caltech e as outras quatro. Esta consistência reafirma a importância de se estudar a rede do Facebook, pois mostra que a diferença entre ela e a rede Barabási-Albert com $m = 8$ não é mero acidente.

Estes resultados também implicam algo importante: pode-se analisar a topologia do Face-

[4] A é uma matriz quadrada de ordem n onde $A_{ij} = 1$ caso i e j sejam amigos no Facebook e $A_{ij} = 0$ caso contrário

[5] A saber, Caltech ($n = 769$), Georgetown University ($n = 9414$), University of Oklahoma ($n = 17425$), Princeton ($n = 6595$) e University of North Carolina at Chapel Hill ($n = 18163$)

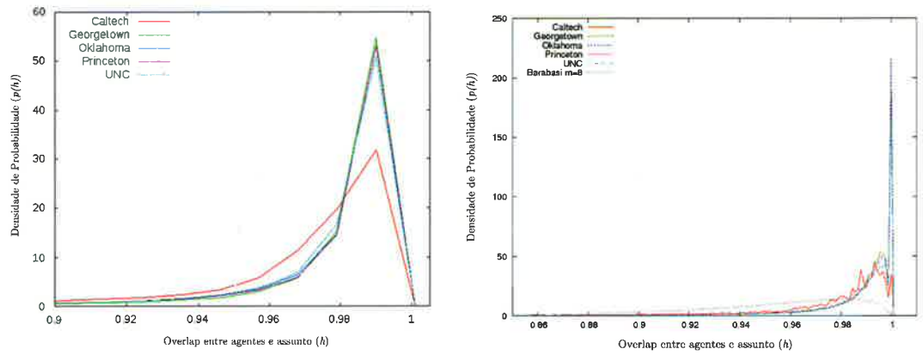


Figura 3.4: A primeira figura mostra o histograma do produto interno entre a classificação dos agentes e o assunto (h_i) das cinco redes de Facebook estudadas, onde os resultados obtidos com as redes do Facebook são essencialmente os mesmos, com exceção da rede do Caltech, pois esta possui uma ordem a menos de grandeza de agentes e menos da metade do número médio de vizinhos por agente. Na segunda figura, observa-se que quando comparado aos resultados da rede utilizada originalmente, a rede Barabási-Albert com $m = 8$, a diferença entre a rede do Caltech e as demais fica desprezível. Isso significa que as redes do Facebook possuem estrutura topológica bastante similar, o que é um bom sinal, e é possível estudar a simulação em apenas uma delas, pois os resultados serão essencialmente os mesmos.

book rodando as simulações apenas em uma das cinco redes usadas. Para isso, foi utilizada a rede de Princeton, por ser a menor das quatro redes com a mesma ordem de grandeza ($n \sim 10^4$). De agora em diante, quando for citada "a rede do Facebook", será subentendido que foi utilizada a rede de Princeton, salvo menção contrária.

O trabalho feito com a rede de Princeton foi o mesmo que no artigo original: roda-se o algoritmo Metropolis com um determinado valor de α, fazendo a simulação para uma grande quantidade de valores de δ. Assim, obtêm-se histogramas do campo $h_i = J_i \cdot x$ da forma da figura 3.2. Ao se obter esta coleção de histogramas, deseja-se descobrir qual é a relação entre o parâmetro δ e a filiação política autodeclarada do trabalho de Haidt, isto é, gerar uma figura parecida com a figura (3.3). Como dito anteriormente, isto é feito minimizando a distância euclidiana entre os dados das sete filiações políticas e os histogramas gerados. A figura equivalente à figura 3.3 gerada para a rede do Facebook pode ser vista na figura 3.5.

Esta análise mostrou um resultado curioso: embora o gráfico seja qualitativamente semelhante ao da rede original Barabási-Albert com $m = 8$, os valores correspondentes de δ cairam bastante, com a filiação política muito liberal (p.a. $= 1$) sendo correspondente a $\delta = 0.01$. Além disso, existem valores de p.a. distintos com valores de δ idênticos para o mesmo α, o que significa que, segundo o modelo, alguma das filiações políticas diferentes possuem a mesma estratégia cognitiva.

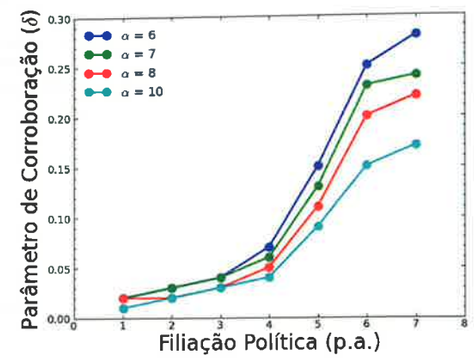


Figura 3.5: Relação entre as filiações políticas e o parâmetro de corroboração δ para diferentes valores de α com a topologia do Facebook, calculados da mesma forma que a figura 3.3. Observa-se que os valores de δ correspondendes às filiações políticas são significativamente mais baixos que os da rede Barabási-Albert de $m = 8$ mostrados na figura 3.3 e, para alguns valores de α, duas filiações políticas correspondem ao mesmo δ.

Para entender a razão destes resultados não esperados, foram feitas as comparações gráficas dos histogramas, similares à figura 3.3, como mostra a figura 3.6.

A comparação dos histogramas mostrou que mesmo usando os valores de δ que minimizam a distância entre os dados da simulação e os dados reais, os histogramas mais próximos ainda são uma aproximação pobre para o resultado dos questionários, muito piores do que os da figura 3.3. A figura nos mostra que na rede do Facebook o consenso é extremamente sensível a δ, fazendo que consenso maior do que os resultados experimentais ocorra ainda para valores de δ bastante pequenos, uma diferença significativa das simulações originais. Para entender por que isso acontece, fez-se a análise das características da topologia do Facebook quando comparada à rede Barabási-Albert com $m = 8$ e a rede quadrada. Descobriu-se que o maior grau de consenso ocorre pois nas redes do Facebook os agentes possuem um número médio de vizinhos da ordem de 100, bastante elevado comparado à rede BAM8, que possui número médio de vizinhos igual a 16. Esse número certamente altera os resultados, pois a energia de interação de um indivíduo é da ordem do seu número de vizinhos, portanto o aumento da energia causado por um agente discordante é quase dez vezes maior na rede do Facebook.

Para uma melhor investigação da influência do número de vizinhos no equilíbrio da sociedade de agentes, foram feitas comparações utilizando duas redes com número médio de vizinhos (representado por $\overline{k}$) similar porém com distribuições de graus $P(k)$ diferentes. Um destes testes foi feito utilizando uma rede cúbida e uma rede Barabási-Albert com $m = 3$. Ambas as redes possuem $\overline{k} = 6$, porém a rede cúbica possui distribuição de vizinhos com $P(k) = 1$ para $k = 6$ e $P(k) = 0$ para $k \neq 6$, e a rede Barabási-Albert com $m = 3$ possui

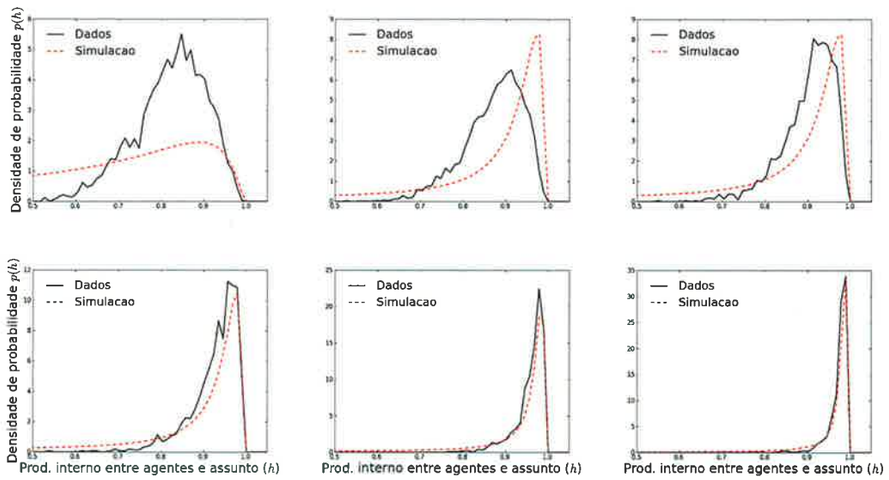


Figura 3.6: Comparação do histograma dos dados reais e a simulação com a rede do Facebook (Caltech), para filiações políticas de 1 a 6. Observe que, para todas as filiações políticas, os histogramas da simulação são menos compatíveis com os dados do que a simulação original, e de forma perceptível para filiações políticas mais liberais. Isto mostra que a rede artificial Barabási-Albert com $m = 8$ é uma melhor representação para a rede de interações que gerou os dados do que a rede (real) do Facebook, o que requer atenção especial.

distribuição de graus $P(k) \propto \frac{1}{k^3}$, sendo portanto redes bastante distintas. A figura 3.7 mostra os resultados da simulação com as duas redes.

É possível ver que os resultados são bastante parecidos para redes com $\overline{k}$ similar, embora com distribuições diferentes. Isso fica mais nítido quando comparado às outras redes distintas com $\overline{k}$ próximos, em particular uma rede quadrada e uma rede Barabási-Albert com $m = 2$ (ambas com $\overline{k} = 4$) e uma rede Barabási-Albert com $m = 4$ ($\overline{k} = 8$). Observa-se que apesar das distribuições diferentes, o parâmetro da rede que parece ditar semelhança entre os resultados é dado pelo número médio de vizinhos.

Isso nos fez pensar o quão realística a rede do Facebook é de fato. O fenômeno que estamos tentando modelar diz respeito à existência de interação social que força conformidade de classificação moral entre agentes, que possuem diferentes estratégias cognitivas em relação a opiniões corroborativas. Essencialmente, estamos modelando agentes que influenciam os julgamentos morais uns dos outros. Como descrito acima, nas redes do Facebook utilizadas cada agente possui em média da ordem 100 vizinhos, isto é, 100 amigos. Mas o critério de amizade em redes sociais como o Facebook não é "pessoas que influenciam seu julgamento moral", como estamos pressupondo no modelo, e sim um critério significamentemente mais frágil, com muitos usuários possuindo amigos que nunca conheceram na vida real ou que mal

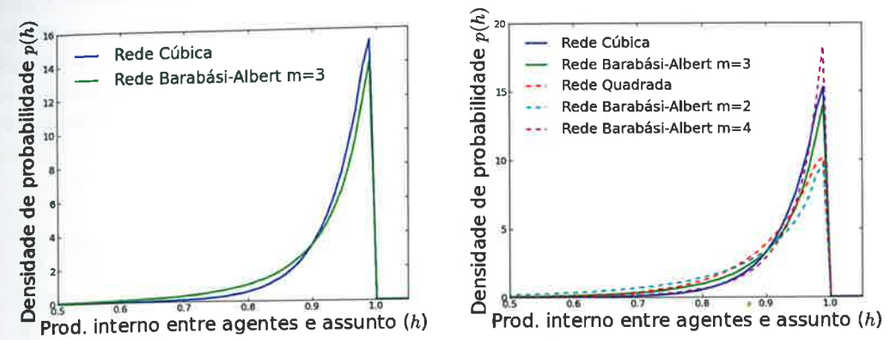


Figura 3.7: A primeira figura mostra uma comparação entre o histograma da simulação na rede cúbica e na rede Barabási-Albert com $m = 3$, o que indica que, apesar da distribuição de graus $p(k)$ completamente diferente nas duas redes, o fato do número médio de vizinhos $\overline{k} = 6$ em ambas faz com que os resultados sejam muito parecidos. Isto é mais acentuado na segunda figura, quando se compara as mesmas redes com redes com $\overline{k} = 4$ e $\overline{k} = 8$. Percebe-se que o fator principal a determinar o equilíbrio das classificações J_i é o número de vizinhos médios por agente.

conhecem. Armados com os resultados da simulação e esta análise, chegamos a conclusão que a rede do Facebook é de fato um substrato pobre para nosso modelo, apesar de ser uma rede realista de interações sociais.

Após considerarmos o fato de que o modelo pode ser considerado invariante por topologia da rede, importando apenas o número médio de vizinhos, e que a rede do Facebook é de fato pobre para os nosso modelos, a pergunta que surgiu é se é possível inverter o processo: dada uma série de simulações feitas em redes Barabási-Albert com diferentes valores de m (e portanto com diferentes valores de $\overline{k}$), tenta-se encontrar qual das redes produz resultados mais próximos dos dados reais. Como este método, deve ser possível obter uma melhor estimativa para o número médio de pessoas que influenciaram a opinião dos usuários que responderam o questionário em [32].

Foi feita então a minimização de histogramas para redes Barabási-Albert com m variando de 1 a 30, da mesma forma como descrita para os resultados originais, e descobriu-se que a rede cujos histogramas são os mais próximos (segundo a distânica euclidiana definida anteriormente) dos histogramas empíricos é a rede Barabási-Albert com $m = 11$, que possui $\overline{k} = 22$. Como pode-se ver pela figura 3.9, a aproximação é bastante superior aos resultados da rede do Facebook e os do artigo original, obtendo concordância significativamente melhor para os dados de Haidt com p.a. ≥ 3. A comparação dos resultados nas redes do Facebook e Barabási-Albert com $m = 11$ pode ser vista na figura 3.10.

Após ester valor ter sido descoberto, procurou-se obter alguma corroboração para este

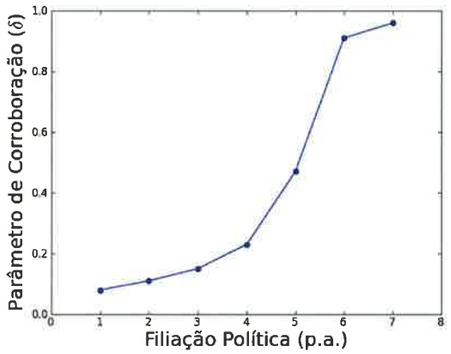


Figura 3.8: Relação entre a filiação política e o parâmetro de corroboração δ para a rede Barabási-Albert com $m = 11$. Observe que δ varia todo intervalo de 0 a 1.

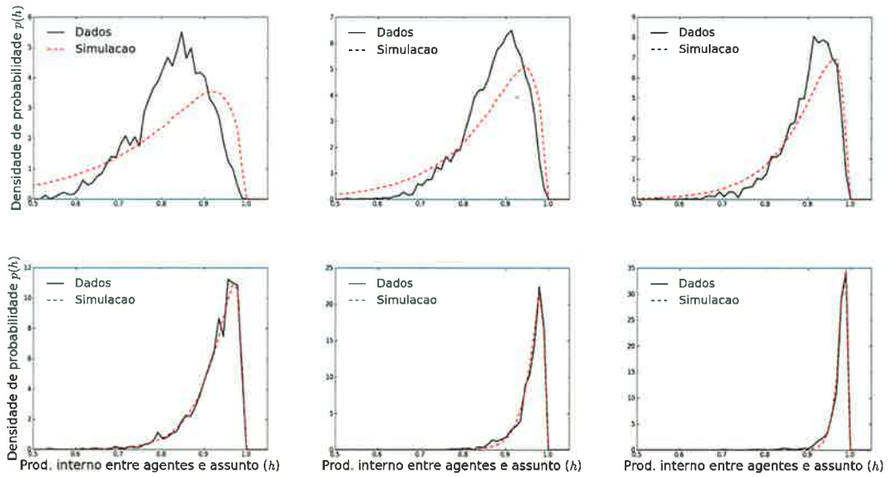


Figura 3.9: Comparação dos histogramas do produto interno entre a classificação dos agentes e o assunto discutido (h_i) simulados para a rede Barabási-Albert com $m = 11$, para filiações políticas de 1 a 6. A aproximação está bastante superior aos resultados originais com a rede de $m = 8$, sendo praticamente idêntica para filiações políticas mais conservadoras.

número. Em um artigo por Trusov et al. [52], os autores fizeram uma pergunta similar à posta neste trabalho: dada uma rede social, quantos amigos de cada usuário de fato influenciam o comportamento deste usuário? Os autores definiram influência como a correlação entre acessos da rede social, isto é, um usuário j influencia i caso o número de acessos de i suba

quando j acessa mais frequentemente a rede, e caso j pare de acessar, i também diminua seus acessos. O resultado obtido no trabalho foi que em uma rede social em que cada usuário possui em média 90 amigos, 22% destes amigos são considerados influentes. Isso significa que a subrede de influências desta rede social possui número médio de vizinhos $\overline{k} = 20$, uma impressionante corroboração para os resultados obtidos pelo modelo de agentes descrito neste trabalho.

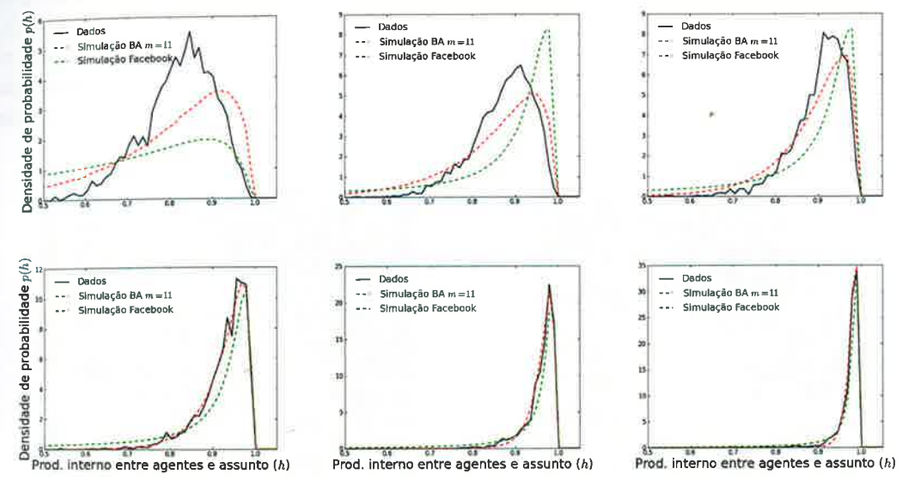


Figura 3.10: Comparação dos histogramas do produto interno entre a classificação dos agentes e o assunto discutido (h_i) simulados para a rede Barabási-Albert com $m = 11$ e para a rede do Facebook, para filiações políticas de 1 a 6. Observa-se que as simulações com a rede Barabási-Albert com $m = 11$ obtém resultados mais próximos dos empíricos.

3.4 Introdução de Agentes Fixos

Nesta seção investigaremos o quão sensíveis são os agentes à introdução de agentes fixos de opinião contrária. A definição de agentes fixos usada neste contexto é de agentes que nunca mudam de opinião, independente de interações. Estes agentes terão um vetor moral fixo $I = -x$, ou seja, maximizarão a energia com os agentes que atingiram o consenso.

A implementação dos agentes fixos é feita da seguinte forma: roda-se a simulação como descrita antes, através do algoritmo de Monte Carlo Metropolis-Hastings, e após atingir o equilíbrio (e, para a maioria dos valores de δ, o consenso), sorteia-se uma fração p de agentes na rede (para um total de pN fixos, N sendo o número total de agentes). Roda-se então

novamente a simulação, pulando os passos de Monte Carlo onde um agente fixo é sorteado para ser modificado.

Define-se um parâmetro de ordem q_i como a fração de agentes com $h < 0$ (isto é, discordando do assunto original x porém concordando com a posição I) ao final da simulação original e q_f como o mesmo valor, porém ao final da simulação com os fixos. O parâmetro $q = q_f - q_i$ representa então a fração de agentes na rede que passou de $h > 0$ (concordando com o assunto x) para $h < 0$ (concordando com a posição $I = -x$) após a interação com os fixos. Isto foi feito para diferentes valores de p e δ, com pressão social $\alpha = 8$ na rede Barabási-Albert com $m = 11$, descrita anteriormente como melhor aproximação para uma rede de influências real, e também em uma rede quadrada para melhor comparação e compreensão. A figura 3.11 ilustra os diagramas de fase obtidos através destas simulações.

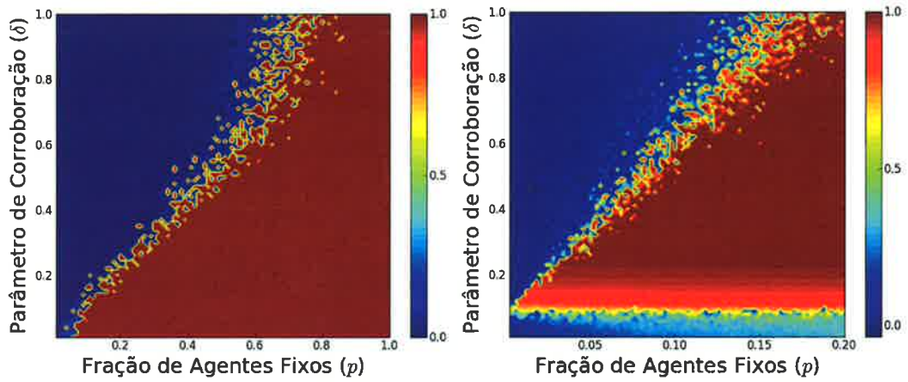


Figura 3.11: Diagrama de fases da simulação com agentes fanáticos para a rede Barabási-Albert com $m = 11$ e para a rede quadrada. Os eixos representam a fração de agentes fixos p e o parâmetro de corroboração δ e a cor do gráfico representa a fração de indivíduos convertidos q. Observe que o p na rede quadrada varia de 0 a 0.2, mostrando que ela é bem mais sensível a agentes fixos do que a rede Barabási-Albert com $m = 11$.

Pelos diagramas de fase, podemos ver que o número médio de vizinhos altera bastante a influência de agentes fixos na opinião média da rede, bastanto apenas que 20% dos agentes sejam fixos na rede quadrada para que estes convertam totalmente uma rede com $\delta = 1$. Para a rede Barabási-Albert com $m = 11$, precisa-se de quase 80% de agentes fixos para que se converta uma rede totalmente conservadora. É importante notar que a região com $\delta < 0.1$ no diagrama da rede quadrada possui $q < 1$ por um artefato do parâmetro de ordem. Estas redes não entram em consenso após o equilíbrio (isto é, não possuem $q_i = 0$ no equilíbrio), então não é possível haver conversão total quando não há um consenso.

Observa-se por estes resultados que indivíduos com uma estratégia cognitiva δ maior, ou seja, que consideram posições corroborantes igualmente relevantes, são mais resistentes a

mudança de classificação devido à interação com agentes fixos. Levando as conexões sobre filiações políticas e estratégias cognitivas estabelecidas na seção anterior a sério, conclui-se que liberais são mais influenciáveis quando discutindo com pessoas que não mudam de opinião, enquanto conservadores são mais resistentes a essa influência. Este resultado pode parecer um tanto óbvio, mas deve-se manter em mente que os diagramas de fase da figura 3.11 foram obtidos através da simulação de um modelo de agentes que interagem e atualizam suas opiniões através de estratégias cognitivas determinadas pelo parâmetro δ. Que este modelo reproduza o comportamento que se espera de cada filiação política é uma corroboração importante.

Capítulo 4

Conclusão

Um dos focos principais deste trabalho foi aplicar o modelo de cooperação coordenada de Boyd, Gintis e Bowles com a dinâmica Fisher-Wright em dois níveis. Com isto, esperamos ter demonstrado a flexibilidade e robustez de um tratamento de sistemas complexos ao problema da evolução. A vantagem do mecanismo Fisher-Wright em dois níveis deve-se principalmente ao tratamento estocástico feito e ao uso de processos de ramificação, que permite que se extraia um resultado macroscópico sem desprezar a dinâmica microscópica exemplificada na figura 2.9, que permite que se defina a viabilidade como uma probabilidade não nula de sobrevivência.

A introdução da taxa de migração também resolve o problema de se introduzir um *assortment* positivo. A definição tradicional de coeficiente de parentesco dada apenas para dois indivíduos, e é necessário usar aproximações para calcular o *assortment* ao se introduzir mais de dois altruístas em um grupo. Utilizando a dinâmica de reprodução dentro do grupo com migração, se obtém um *assortment* positivo natural, porém não extremo (como no modelo de infinitas ilhas de Wright) e que varia com o tempo. Além disso, pode-se definir o coeficiente de parentesco tradicional (quando no regime de seleção fraca) em função da taxa de migração, o que permite que se compare os resultados com modelos já existentes, que geralmente adotam seleção fraca.

Outro grande mérito deste framework é não segregar as idéias de seleção de parentesco e seleção de grupo. A competição entre grupos faz com que um grupo com a maior fitness média se prolifere, mas a competição entre indivíduos de um grupo faz com que seja importante que agentes tenham interações principalmente com aqueles idênticos por descendentes. Desta forma, é possível pensar nestes mecanismos como complementares ao invés de excludentes.

Com o exemplo do modelo de punição coordenada, demonstra-se que a implementação natural de *assortment* positivo do modelo somado com a estrutura multi nível de competição

permite que o altruísmo seja viável em condições mais favoráveis do que se obtém normalmente. Em particular, o coeficiente de parentesco crítico obtido com nosso framework é absolutamente compatível com aquele observado em sociedades caçadoras coletoras [27] e menor, consistentemente, do que o modelo introduzido por Boyd, Gintis e Bowles prevê.

A aplicação de mecânica estatística no problema da Teoria de Fundamentos Morais mostrou que existem evidências de que as classificações morais feitas em [32] foram formadas através da interação entre indivíduos e que é possível relacionar o tratamento de informação corroborativa e informação nova com a posição política auto declarada destas pessoas. O que é interessante deste modelo é que a implementação das regras da interação foi feita no nível microscópico, definindo apenas uma energia de interação dos agentes segundo as evidências empíricas descritas e a Teoria de Fundamentos Morais de Haidt, e o sistema resultante obtido possui uma distribuição de classificações no equilíbrio muito próxima ao dos resultados reais. Essa corroboração se tornou ainda maior quando otimizamos os resultados alterando a rede topológica que define as interações e descobriu-se que a rede cujo equilíbrio é o mais próximo ao dos dados de Haidt é muito parecida com as redes de influência reais em redes sociais.

4.1 Perspectivas

Algumas das futuras investigações planejadas para este modelo podem ser encontradas em [49]. O foco mais imediato de interesse é entender a fase avançada da dinâmica, quando o número de altruístas já é macroscópico e não vale mais a teoria de ramificação. Já se sabe que uma vez atingido este ponto o altruísmo é viável (caso contrário haveria sido extinguido ao chegar na fase inicial estacionária), mas é possível ainda que, dependendo de m, o equilíbrio de sobrevivência seja polimórfico e com uma fração baixa de indivíduos do tipo A. Outra extensão natural é a reprodução sexuada com organismos diplóides, onde indivíduos possuem dois alelos e cada agente descendente recebe um alelo de cada pai aleatoriamente. O altruísmo neste caso pode ser dominante ou recessivo. No fim de uma geração, apenas as mulheres migrariam. Esta análise é mais realista e seria interessante observar como os resultados se alteram. Em particular no modelo BGB, uma terceira extensão seria observar o que ocorre com uma sociedade com agentes de τ diferentes, especialmente se τ puder evoluir com o tempo.

Na dinâmica de classificação moral, uma das extensões mais naturais é rodar as simulações utilizando agentes com valores de δ diferentes. Pode-se definir diversas densidades de probabilidade $p(\delta)$ e estudar como a dinâmica, e o consenso no equilíbrio, dependem destas densidades.

Este trabalho de mestrado teve como tema geral a aplicação de métodos de mecânica estatística a problemas de sistemas sociais, e acreditamos ter colaborado no entendimento

de dois problemas interessantes e bastante relevantes em suas áreas. No entanto, espera-se que a contribuição mais importante deste trabalho seja mostrar ao leitor que a mecânica estatística possui ferramentas extremamente robustas e que podem ser aplicadas com sucesso a problemas em diversas áreas da ciência.

Referências Bibliográficas

[1] The igraph library for complex network research. Disponível em: http://igraph.sourceforge.net/.

[2] The Moral Foundations Questionnaire. Disponível em: http://faculty.virginia.edu/haidtlab/mft/index.php?t=questionnaires.

[3] Réka Albert and Albert-László Barabási. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[4] D.M. Amodio, J.T. Jost, S.L. Master, and C.M. Yee. Neurocognitive Correlates of Liberalism and Conservatism. *Nature Neuroscience*, 10:1246–1247, 2007.

[5] Masanao Aoki. *New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria and Externalities as Field Effects.* Cambridge University Press, 1998.

[6] Robert Axelrod. *The Evolution of Cooperation.* Basic Books, 1984.

[7] Ellen Baake and Wilfried Gabriel. Biological Evolution Through Mutation, Selection, and Drift: An Introductory Review, 1999. arXiv:99077372.

[8] Philip Ball. The Physical Modelling of Society: A Historical Perspective. *Physica A*, 314:1–14, 2002.

[9] Samuel Bowles. *Microeconomics: Behavior, Institutions, and Evolution.* Princeton University Press, 2006.

[10] Samuel Bowles. Did Warfare Among Ancestral Hunter-Gatherers Affect the Evolution of Human Social Behaviors? *Science*, 324:1293–1298, 2009.

[11] Samuel Bowles and Herbert Gintis. *A Cooperative Species: Human Reciprocity and Its Evolution.* Princeton University Press, 2011.

[12] Robert Boyd, Herbert Gintis, and Samuel Bowles. Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science*, 328:617–620, 2010.

[13] Robert Boyd, Herbert Gintis, and Samuel Bowles. Supplementar Online Material: Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare, 2010.

[14] Lisa Cameron. Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia. *Economic Inquiry*, 37:47–59, 1998.

[15] L. Campbell and W. Garnett. *The Life of James Clerk Maxwell.* Adamant Media Corporation, 2001.

[16] Nestor Caticha and Renato Vicente. Agent-based Social Psychology: from Neurocognitive Processes to Social Data. *Advances in Complex Systems*, 14:711–731, 2011.

[17] D. Challet, M. Marsili, and Y. C. Zhang. *Minority Games: Interacting Agents in Financial Markets.* Oxford University Press, 2005.

[18] Charles Darwin. *On the Origin of Species.* John Murray, 1859.

[19] Charles Darwin. *The Descent of Man.* John Murray, 1871.

[20] Richard Dawkins. *The Selfish Gene.* Oxford University Press, 1976.

[21] Barbara Drossel. Biological Evolution and Statistical Physics. *Advances in Physics*, 50:209–295, 2001.

[22] Lee Alan Dugatkin. *The Altruism Equation.* Princeton University Press, 2006.

[23] N. Eisenberger, M. Lieberman, and K. Williams. Does Rejection Hurt? An fMRI Study of Social Exclusion. *Science*, 302:290–292, 2003.

[24] A. Engel and C. Van den Broeck. *Statistical Mechanics of Learning.* Cambridge University Press, 2001.

[25] Urs Fischbacher Ernst Fehr and Elena Tougareva. Do High Stakes and Competition Undermine Fairness? Evidence from Russia, 2002. Disponível em: http://ssrn.com/abstract=320384 ou doi:10.2139/ssrn.320384.

[26] Alexander Strobel et al. Beyond Revenge: Neural and Genetic Bases of Altruistic Punishment. *NeuroImage*, 54:671–680, 2010.

[27] Kim R. Hill et al. Co-Residence Patterns in Hunter-Gatherer Societies Show Unique Human Social Structure. *Science*, 331:1286–1289, 2011.

[28] Alison Etheridge. Notas de Aula: Stochastic Models in Mathematical Genetics. http://www.stats.ox.ac.uk/ etheridg/popgen/notes.pdf.

[29] Ernst Fehr and Simon Gächter. Altruistic Punishment in Humans. *Nature*, 415:137–140, 2002.

[30] Ronald Fisher. *The Genetical Theory of Natural Selection.* 1930.

[31] Jesse Graham, Jonathan Haidt, and Brian A. Nosek. Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology*, 96:1029–1046, 2009.

[32] Jonathan Haidt. The New Synthesis in Moral Psychology. *Science*, 316:998–1002, 2007.

[33] Jonathan Haidt. The Moral Foundations of Occupy Wall Street, 2012. http://reason.com/archives/2011/12/30/the-moral-foundations-of-occupy-wall-str.

[34] W. D. Hamilton. The Genetical Evolution of Social Behaviour. *Journal of Theoretical Biology*, 7:1–16, 1964.

[35] Theodore E. Harris. *The Theory of Branching Processes.* Dover Publications, 2002.

[36] John A. List. Friend or Foe? A Natural Experiment of the Prisoner's Dilemma. *Review of Economics and Statistics*, 88:463–471, 2006.

[37] Sarah Matthew and Robert Boyd. Punishment Sustains Large-Scale Cooperation in Prestate Warfare. *Proceedings of the National Academy of Sciences*, 108:1375–1380, 2011.

[38] Richard McElreath and Robert Boyd. *Mathematical Models of Social Evolution.* University of Chicago Press, 2007.

[39] Richard Milo. Evidence for Hominid Predatation at Klasies River Mouth, South Africa and its Implications for the Behaviour of Modern Humans. *Journal of Archaeological Science*, 25:99–133, 1998.

[40] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation.* Oxford University Press, 2009.

[41] Richard A. Neher and Boris I. Shraiman. Statistical Genetics and Evolution of Quantitative Traits. *Reviews of Modern Physics*, 83:1283–1300, 2011.

[42] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics.* Oxford University Press, 1999.

[43] Mark Newman, Albert-László Barabási, and David Watts. *The Structure and Dynamics of Networks.* Princeton University Press, 2006.

[44] Martin A. Nowak. *Evolutionary Dynamics.* Belknap Press of Harvard University Press, 2006.

[45] Martin A. Nowak. Five Rules for the Evolution of Cooperation. *Science*, 314:1560–1563, 2006.

[46] Martin A. Nowak and Karl Sigmund. Evolution of Indirect Reciprocity By Image Scoring. *Nature*, 393:573–577, 1998.

[47] Samir Okasha. *Evolution and the Levels of Selection.* Oxford University Press, 2009.

[48] Mark Ridley. *Evolution.* Oxford University Press, 1993.

[49] Roberto H. Schonmann, Renato Vicente, and Nestor Caticha. Two-level Fisher-Wright framework with selection and migration: An approach to studying evolution in group structured populations, 2011. arXiv:1106.4783.

[50] John Maynard Smith. *Evolutionary Genetics.* Oxford University Press, 1998.

[51] Amanda L. Traud, Eric D. Kelsic, Peter J. Mucha, and Mason A. Porter. Community Structure in Online Collegiate Social Networks. arXiv:0809.0960, 2008.

[52] M. Trusov, A. V. Bodapati, and R. E. Bucklin. Determining Influential Users in Internet Social Networks. *Journal of Marketing Research*, XLVII:643–658, 2010.

[53] Sewall Wright. Evolution in Mendelian Populations. *Genetics*, 16:97–159, 1932.