

FUNDAÇÃO JOÃO PINHEIRO  
Escola de Governo Professor Paulo Neves de Carvalho

Luiz Henrique Valle Xavier

DELIBERAÇÃO E SISTEMAS DELIBERATIVOS:  
contribuições da Ciência de Dados

Belo Horizonte

2020

Luiz Henrique Valle Xavier

**DELIBERAÇÃO E SISTEMAS DELIBERATIVOS:**  
contribuições da Ciência de Dados

Trabalho apresentado ao Curso de Especialização em Administração Pública, Planejamento e Gestão Governamental da Escola de Governo Professor Paulo Neves de Carvalho da Fundação João Pinheiro, como requisito parcial para obtenção do título de especialista.

Orientadora: Flávia de Paula Duque Brasil

Belo Horizonte  
2020

X3d Xavier, Luiz Henrique Valle.  
Deliberação e sistemas deliberativos [manuscrito] : contribuições da Ciência de Dados / Luiz Henrique Valle Xavier. – 2020. [8], 53 f. : il.

Monografia de conclusão de Curso (Especialização em Administração Pública Planejamento e Gestão Governamental) – Fundação João Pinheiro, Escola de Governo Professor Paulo Neves de Carvalho, 2020.

Orientadora: Flávia de Paula Duque Brasil

Bibliografia: f. 58-61

1. Ciências sociais – Computador. 2. Democracia deliberativa. 3. Computational Social Science. I. Brasil, Flávia de Paula Duque. II. Título.

CDU 3:692.31

Luiz Henrique Valle Xavier

**DELIBERAÇÃO E SISTEMAS DELIBERATIVOS:**

contribuições da Ciência de Dados

Trabalho apresentado ao Curso de Especialização em Administração Pública, Planejamento e Gestão Governamental da Escola de Governo Professor Paulo Neves de Carvalho da Fundação João Pinheiro, como requisito parcial para obtenção do título de especialista.

Aprovado na Banca Examinadora

---

Prof<sup>a</sup>. Dra. Flávia de Paula Duque Brasil, Fundação João Pinheiro

---

Prof. Dr. Ágnez de Lélis Saraiva, Fundação João Pinheiro

Belo Horizonte, 24 de novembro de 2020

## **AGRADECIMENTOS**

Em primeiro lugar, agradeço à Adriana. Sem seu apoio, incentivo e parceria nada seria possível. Somos um em dois.

À minha orientadora, Professora Flávia Brasil, por aceitar o desafio e me guiar mesmo à distância.

À Prodemge, pela oportunidade.

À Fundação João Pinheiro, por abraçar a sua missão.

Aos colegas de turma, pela convivência.

E em especial ao Hugo, ao Gustavo e ao Augusto, por fazerem tudo valer a pena.

## RESUMO

A democratização do exercício do poder constitui um fundamento essencial da democracia. A partir dos debates das últimas décadas, relativos à qualidade democrática, ganham relevo as concepções de democracia participativa e deliberativa como condutores dos processos de tomada de decisão pelas instituições políticas, destacando-se a abordagem sistêmica, de natureza complexa, implicando diversos desafios metodológicos para descrição e interpretação de seus resultados. Considera-se, ainda, o crescimento das práticas de democracia digital envolvendo processos de participação em deliberações *on-line* e as formas de comunicação e interação política por meio de redes sociais digitais, nas quais se observa uma pulverização e descentralização dos agentes formadores de opinião. Esse cenário demanda a aplicação de tecnologias avançadas de análise de dados em contexto de larga escala para a realização de estudos diversos. Essa monografia propõe uma aproximação entre as áreas da Ciência de Dados e o campo da Democracia Deliberativa com o objetivo de potencializar o aprimoramento desta última, ao trazer à tona atores, debates e ideias muitas vezes excluídos dos processos deliberativos. Para tanto, realizou-se uma pesquisa exploratória com objetivo de definir um recorte no campo da democracia deliberativa com afinidade para a aplicação de métodos e ferramentas de Ciência de Dados. Nessa direção, a pesquisa identificou a *Computational Social Science* como uma metodologia de pesquisa aderente às demandas das Ciências Sociais. Em seguida realizou-se uma pesquisa de natureza descritiva e exploratória baseada em revisão bibliográfica para descrever adequadamente os processos deliberativos como um objeto de estudo de Ciência de Dados. A etapa seguinte relaciona os métodos e técnicas da Ciência de Dados à abordagem de sistemas deliberativos sob a perspectiva metodológica de pesquisa da *Computational Social Science*, resultando na identificação de possibilidades de colaboração entre as áreas.

Palavras-Chave: Democracia Deliberativa, Sistemas Deliberativos, Ciência de Dados, Pegadas Digitais, *Computational Social Science*

## **ABSTRACT**

The democratization of the exercise of power is an essential foundation of democracy. From the debates of the last decades regarding democratic quality, the conceptions of participatory and deliberative democracy as drivers of decision-making processes by political institutions gain prominence, highlighting the systemic approach, of a complex nature, implying several methodological challenges for description and interpretation of its results. It is also considered the growth of digital democracy practices involving processes of participation in online deliberations and the forms of communication and political interaction through digital social networks, where there is a dispersion and decentralization of opinion-forming agents. This scenario demands the application of advanced data analysis technologies in a large-scale context for carrying out studies. This monograph proposes an approximation between the areas of Data Science and Deliberative Democracy in order to enhance the improvement of the latter by bringing up actors, debates and ideas that are often excluded from deliberative processes. For that, exploratory research was carried out with the objective of defining a cut in the field of deliberative democracy with affinity for the application of methods and tools of Data Science. In this sense, the research identified Computational Social Science as a research methodology that adheres to the demands of the social sciences. Then, a research of a descriptive and exploratory nature was carried out based on a bibliographic review to adequately describe the deliberative processes as an object of study of Data Science. The next step tried to relate the methods and techniques of Data Science to the approach of deliberative systems under the methodological research perspective of Computational Social Science, resulting in the identification of possibilities for collaboration between the areas.

**Keywords:** Deliberative Democracy, Deliberative Systems, Data Science, Digital Footprints, Computational Social Science

## SUMÁRIO

1	INTRODUÇÃO	9
2	DEMOCRACIA DELIBERATIVA: DEFINIÇÕES E ABORDAGENS	14
2.1	Sistemas deliberativos: definições e características	18
3	COMPUTATIONAL SOCIAL SCIENCE: MODELOS E MÉTODOS	26
3.1	Teoria da informação	31
3.1.1	Teoria da Decisão	32
3.2	Métodos de Ciência de Dados	33
3.2.1	<i>Big data</i>	35
1.1.1.1	Características de <i>big data</i> nas análises em Ciências Sociais	36
1.1.1.2	Limitações de <i>big data</i> na análise em Ciências Sociais	37
3.2.2	Mineração de dados	39
1.1.1.3	Análise exploratória de dados	41
1.1.1.4	Mineração de padrões frequentes	42
1.1.1.5	Agrupamento	42
1.1.1.6	Classificação	43
3.2.3	Aprendizagem de máquina	44
3.3	Simulações computadorizadas	45
3.3.1	Sistemas Complexos	45
3.3.2	Modelos baseados em agentes	45
3.4	Métodos de análise	47
3.4.1	Análise de redes sociais	47
3.4.2	Processamento de Linguagem Natural	48
4	COMPUTATIONAL SOCIAL SCIENCE: APLICAÇÕES EM DEMOCRACIA DELIBERATIVA	49
4.1	Técnicas para abordagem operacional	51
4.2	Técnicas para abordagem estrutural	52
5	CONSIDERAÇÕES FINAIS	54
	REFERÊNCIAS	58

## 1 INTRODUÇÃO

A democratização do exercício do poder constitui um fundamento essencial da democracia. Mas o que signif Democracia Deliberativa, Sistemas Deliberativos, Ciência de Dados, Pegadas Digitais, *Computational Social Science* ica esse poder e como ele se mantém? Habermas (HABERMAS, 1997, v. II) discute que o poder do Estado se manifesta na estabilidade da ordem por ele mantida. Nessa perspectiva, a estabilidade vale como medida para a legitimidade, ao se considerar o modelo elitista de democracia que se afirmou especialmente na primeira metade do século XX, apontando estritamente para a participação via voto como fundamento democrático (SANTOS; AVRITZER, 2002).

A partir dos debates das últimas décadas relativos à qualidade democrática, ganham relevo as concepções de democracia participativa e deliberativa, que requerem, complementarmente aos arranjos de representação eleitoral, o alargamento das formas de participação e deliberação para tomadas de decisões políticas.

Nessa perspectiva, a legitimidade passa por processos deliberativos que têm fundamento nas interações comunicativas - condutores dos processos de tomada de decisão pelas instituições políticas. Nas últimas décadas, é possível observar “um crescente consenso entre teóricos políticos e cientistas políticos empíricos de que a legitimidade de uma democracia depende em parte da qualidade da deliberação que informa os cidadãos e seus representantes.” (MANSBRIDGE et al., 2012, p. 1)<sup>1</sup>.

A partir do entendimento de que a legitimidade de uma democracia depende das práticas deliberativas e da qualidade da deliberação realizada pelos cidadãos, dentre as abordagens deliberativas mais recentes, destaca-se a perspectiva dos sistemas deliberativos, que busca abranger e conectar os diversos contextos e momentos nos quais ocorrem as deliberações públicas. Assim, há entre pesquisadores no campo deliberativo esforços orientados à compreensão, caracterização, definição dos limites, desenho e formatos dos sistemas deliberativos.

No âmbito internacional, a análise dos diferentes fóruns deliberativos e suas relações com o sistema político tem ganhado a atenção recente de autores que se dedicam ao estudo da teoria deliberativa, reunidos em torno da discussão sobre o que denominaram de um modelo

---

<sup>1</sup> Tradução nossa do original: “[...] have seen growing agreement among political theorists and empirical political scientists that the legitimacy of a democracy depends in part on the quality of deliberation that informs citizens and their representatives.” (MANSBRIDGE et al., 2012, p. 1)

sistêmico de democracia deliberativa. Embora a abordagem sistêmica não seja inteiramente nova, inclusive no âmbito da teoria deliberativa na qual já tinha sido objeto da atenção de Habermas (2003), o enfoque dado à questão pode ser considerado promissor, na medida em que busca analisar as diversas instâncias deliberativas do sistema político e suas formas de interação mútua. (SILVA, 2014, p. 3)

Nessa linha, no campo deliberativo destaca-se a abordagem sistêmica, implicando diversos desafios metodológicos para descrição e interpretação de seus objetos de pesquisa. Ou seja, tem se tornado mais urgente olhar para os processos deliberativos, em si complexos, especialmente quando se propõe considerar as práticas deliberativas de larga escala (MANSBRIDGE et al., 2012) e interconectadas.

Mais além, a perspectiva dos sistemas deliberativos incorpora os processos de formação informal de opinião pública e sua possível influência nas tomadas de decisão nos âmbitos dos sistema político-administrativo. E há que se considerar, ainda, o crescimento das práticas de democracia digital que envolve processos de participação em deliberações *on-line* e as formas de comunicação e interação política por meio de redes sociais.

No contexto de uma “sociedade em rede”, na qual se destaca o papel extenso das tecnologias informais nos diversos âmbitos da vida social (CASTELLS, 2013), o que se observa neste contexto é uma pulverização e descentralização dos agentes formadores de opinião. Neste contexto, requer-se que pesquisadores dos sistemas deliberativos considerem uma série de condições específicas do desenho dos processos por deliberação em meios digitais, como descreve Macintosh (2008):

Uma variedade de métodos de coleta e análise de dados deve ser usada, dependendo das ferramentas de *eParticipation* e do contexto de uso. Estes incluem:

- observação de campo de atores relevantes usando a ferramenta em um ambiente do mundo real
- entrevista e discussão em grupo com atores relevantes
- análise de questões on-line e discussão
- análise da documentação do projeto
- estatísticas de uso das ferramentas e análise do arquivo de log do servidor

Incluindo métodos, essa interação de análise garante que a avaliação forneça evidências do que as pessoas realmente fazem com as ferramentas de *eParticipation*, bem como o que dizem que fazem. É importante ressaltar que o uso de métodos mistos permite a

triangulação de métodos e resultados e, portanto, ajuda a maximizar sua validade. (MACINTOSH et al. 2008, p. 6)<sup>2</sup>

Em paralelo à tendência de se pensar os processos sociais por meio de abordagens sistêmicas e ao entendimento de que se deve considerar a inter-relação entre as partes como relevantes nesse tipo de abordagem, a Ciência da Computação - aliada à Estatística e à Matemática - vem propondo o desenvolvimento e aplicação de tecnologias de análise de dados num contexto de larga escala. O conjunto dessas tecnologias tem se consolidado como uma nova área do conhecimento denominada Ciência de Dados, notadamente em seu subcampo *Computational Social Science*.

Assim, considerando os desenvolvimentos teóricos no campo da teoria democrática deliberativa, o problema que esse trabalho se propõe a responder é: como a Ciência de Dados pode contribuir com as abordagens voltadas para os sistemas deliberativos no contexto da democracia digital?

A presente monografia tem por objetivo, portanto, explorar os métodos e ferramentas de Ciência de Dados com potencial para contribuir para as abordagens sistêmicas de democracia deliberativa. Tem-se em vista identificar, sistematizar e analisar as várias aplicações das mais importantes ferramentas de computação que têm aderência com métodos das Ciências Sociais.

Em relação à metodologia, na primeira fase dessa pesquisa, realizou-se uma pesquisa exploratória com objetivo de definir um recorte no campo aplicado da democracia deliberativa em que seria empiricamente possível propor a intervenção e/ou contribuição de métodos e ferramentas de Ciência de Dados. Nessa direção, a pesquisa apontou o campo específico de *Computational Social Science* como uma área que já tem grande aderência com a demanda das Ciências Sociais. No contexto da Administração Pública em suas interfaces com a Ciência Política, a opção foi considerar as abordagens de democracia deliberativa, especificamente no que se refere aos sistemas deliberativos.

---

<sup>2</sup> Tradução nossa do original "A range of data gathering and analysis methods should be used depending, on eParticipation tools and context of use. These include: field observation of relevant actors using the tool in a real-world setting; interviewing and group discussion with relevant actors; analysis of online questions and discussion; analysis of project documentation; usage statistics from the tools and server logfile analysis.

Including methods that analyse interaction ensures that the evaluation provides evidence of what people actually do with eParticipation tools, as well as what they say they do. Importantly, using mixed methods allows triangulation of methods and results and, therefore, helps to maximise their validity." (MACINTOSH et al. 2008, p. 6)

No que se refere ao percurso metodológico e às técnicas empregadas, a pesquisa – de natureza descritiva e exploratória – recorreu basicamente à revisão bibliográfica. Como método inicial, utilizou-se o sistema de indexação de publicações científicas Google Scholar. A primeira busca foi por “*deliberative system*” associada a “*deliberative democracy*” cujo resultado indexa, dentre outros trabalhos revisados, o livro “*Deliberative systems: Deliberative democracy at the large scale*” (PARKINSON; MANSBRIDGE, 2012), que se constituiu como uma referência basilar para o trabalho, especialmente no que se refere aos sistemas deliberativos. Destaca-se que a referida obra reúne trabalhos precursores e centrais para a abordagem dos sistemas deliberativos.

Especialmente a partir de Mansbridge et al. (2012) foi possível uma aproximação entre a abordagem sistêmica da deliberação com o conceito geral de sistemas complexos, central na Ciência de Dados. Nessa direção, a partir de uma busca no repositório do Instituto de Pesquisa Econômica Aplicada (IPEA), especificamente no Centro de Pesquisa Ciência e Sociedade,<sup>3</sup> dentre outros trabalhos, destaca-se a contribuição da obra de Furtado, Sakowski e Tóvolli (2015) que analisa os fundamentos da teoria dos Sistemas Complexos aplicados em diversas pesquisas sobre decisões em Políticas Públicas.

Apesar de não tratar especificamente sobre Sistemas Deliberativos, o livro constitui uma das chaves para este trabalho ao descrever pesquisas em que os métodos utilizados pela Ciência de Dados podem contribuir para as tomadas de decisões governamentais. Tendo em vista o objetivo da monografia, a revisão bibliográfica voltou-se, ainda, para a *Computational Social Science* e seus recursos aplicáveis à análise proposta.

O presente trabalho estrutura-se em cinco capítulos, considerando-se esta introdução e as considerações finais. O capítulo 2 desenvolve-se em torno do tema da democracia deliberativa, especificamente da abordagem sistêmica das práticas deliberativas. Tal abordagem aponta para uma demanda por métodos e ferramentas com potencial para lidar com redes complexas de dados e informações. Portanto, o referido capítulo, em que se descreve o contexto complexo das práticas deliberativas, tem a função de criar os fundamentos para a exploração e descrição das ferramentas com potencial de uso no contexto da abordagem sistêmica.

---

<sup>3</sup>< <https://www.ipea.gov.br/cts/pt/>>

No campo específico da Ciência da Computação, no terceiro capítulo, realizou-se uma abordagem descritiva para compreensão dos métodos e ferramentas da *Computational Social Science*. O capítulo 4 detém-se na análise relativa ao potencial das tecnologias, ferramentas e métodos que poderiam ser empregados para as pesquisas sistêmicas. Finalmente, no último capítulo são apresentadas as considerações finais abordando os limites desse trabalho. Em resumo, o trabalho destaca a modelagem e as simulações computacionais que podem contribuir significativamente para as abordagens dos sistemas deliberativos.

## 2 DEMOCRACIA DELIBERATIVA: DEFINIÇÕES E ABORDAGENS

A democracia deliberativa é antes uma perspectiva de prática democrática que adota os procedimentos deliberativos como ideais para justificação das relações políticas, tendo a razão e liberdade para a deliberação entre cidadãos como legitimadora do poder do governo. Nos termos de Maldonado (2007), a democracia deliberativa:

É uma concepção de democracia que adota o procedimento deliberativo como ideal de justificação política. Reivindicando a razão e não o poder como fundamento, não faz depender a legitimidade da democracia do princípio da maioria ou da soberania popular, mas do raciocínio público livre entre cidadãos iguais. Nem o voto, nem a agregação de interesses, nem o autogoverno constituem a essência da democracia, mas a própria deliberação como procedimento de decisão. (MALDONADO, 2007, p. 38)<sup>4</sup>

A democracia deliberativa pode ser pensada como um modelo geral e ideal de democracia, como resultado a “generalização de uma política deliberativa deve resultar em uma ordem política mais racional e um aumento na justiça das leis e instituições sociais.” (MALDONADO, 2007, p.40)<sup>5</sup> Na democracia deliberativa considera-se um pressuposto ético a participação de todos os afetados ou seus representantes nas decisões políticas, por meio de procedimento deliberativo que se baseia na argumentação racional e imparcial como critério.

A democracia deliberativa é também entendida como uma abordagem teórica que tem um lastro na prática democrática, tratando-se de considerar a deliberação como um fenômeno social complexo que envolve a interação entre atores sociais em instâncias formais e informais da arena política.

Nesse sentido, a democracia é entendida na vertente deliberativa a partir dos fluxos comunicacionais na esfera pública e sua possibilidade de influenciar os sistemas político e administrativo, gerando novas práticas deliberativas e participativas. Conforme Avritzer (1996, p.15), a democracia associa-se às práticas

---

<sup>4</sup> Tradução nossa do original: "Se trata de una concepción de la democracia que adopta el procedimiento deliberativo como ideal de justificación política. Reclamando como fundamento la razón antes que el poder, no hace depender la legitimidad de la democracia del principio de mayoría ni de la soberanía popular, sino del libre razonamiento público entre ciudadanos iguales. Ni el voto, ni la agregación de intereses, ni el autogobierno constituyen la esencia de la democracia, sino la deliberación misma como procedimiento de decisión." (MALDONADO, 2007, P. 38)

<sup>5</sup> Tradução nossa do original: "[...] generalización de una política deliberativa debe resultar un orden político más racional, y un incremento de la justicia de las leyes e instituciones sociales." (MALDONADO, 2007, p.40)

discursivas “que têm seu lugar nas redes públicas de comunicação com as quais os processos de institucionalização legal e utilização administrativa do poder estão indissolavelmente ligados” (AVRITZER, 1996, p.15). Assim, nos termos de Santos e Avritzer (2003), a democracia constitui uma gramática de organização da sociedade e das relações entre Estado e sociedade.

A democracia deliberativa, portanto, “é claramente dependente de um processo público de comunicação, através do qual a sociedade debate, argumenta e toma decisões sobre questões políticas relevantes [...] (LEAL, 2011, p.44) Nessa instância discursiva, atuam as forças de conformação da opinião pública que definem um campo muito específico em relação à consideração das práticas democráticas.

[...]um dos fundamentos da política deliberativa não é simplesmente a oportunidade de manifestação da opinião de cada um, mas a formatação de posições sociais – coletivas – sobre informações e temas focalizados, consensualmente. Assim, o que gera a esfera pública não são agregados estatísticos de opiniões individuais – por mais críticas que sejam –, mas deliberações forjadas racionalmente entre todos os envolvidos; (Leal, 2011, p. 64)

Pode-se trazer como síntese a perspectiva de que a democracia deliberativa deve ser considerada “a partir da ideia de cogestão pública dos interesses comunitários, em todas as suas fases existenciais (constitutivas, aplicativas e avaliativas), chamando à responsabilidade a cidadania como sujeito histórico que é.” (LEAL, 2011, p. 43). E mais, um “problema central ainda é o de se construir metodologias de interlocução política capazes de viabilizar deliberações públicas compartilhadas qualitativa e quantitativamente” (LEAL, 2011, p. 44).

A democracia deliberativa pode ser entendida a partir do que Habermas chama de “implementação aproximativa do processo”. Nesse sentido, considera-se que não se tem um processo perfeito de deliberação, mas se têm critérios procedimentais como uma referência tanto para nortear a prática como para observação analítica dos processos deliberativos:

a) a inclusão de todas as pessoas envolvidas; b) chances reais de participação no processo político, repartidas equitativamente; c) igual direito a voto nas decisões; d) o mesmo direito para a escolha dos temas e para o controle da agenda; e) uma situação na qual todos os participantes, tendo à mão informações suficientes e bons argumentos, possam formar uma compreensão articulada acerca das matérias a serem regulamentadas e dos interesses controversos. (HABERMAS *apud* LEAL, 2011, p.44)

Esse modelo racional que orienta a prática formal e institucionalizada das deliberações democráticas orienta também as práticas de deliberação *on-line* em experiências de democracia digital. Entretanto, Maia (2001) considera que o resultado é mais próximo de uma representação ampliada ao cidadão do que realmente a promoção do debate político. A facilitação promovida pela tecnologia para os processos de deliberação não promove a inclusão dos cidadãos no debate político.

A prática da deliberação não é algo que ocorre espontaneamente em sociedades complexas. É dependente de um contexto institucional e de procedimentos que promovam a prática dialógica e a valorização da participação das pessoas nos debates desde a esfera pública. A esfera pública, bem como coloca Habermas (2001):

Em sociedades complexas, a esfera pública forma uma estrutura intermediária entre o sistema político, de um lado, e os setores privados do mundo da vida e sistemas de ação especializados em termos de funções, de outro lado. Ela representa uma rede supercomplexa que se ramifica espacialmente num sem número de arenas internacionais, nacionais, regionais, comunais e subculturais, que se sobrepõem umas às outras; essa rede se articula objetivamente de acordo com os pontos de vista funcionais, temas, círculos políticos, assumindo a forma de esferas públicas mais ou menos especializadas, porém ainda assim acessível a leigos (...) (HABERMAS *apud* MAIA, 2001, p. 4)

Na instância informal da esfera pública que passa pelas conversações, debates e diálogos que ocorrem na vida cotidiana e em espaços informais da vida pública, tem-se a questão:

(...) como transformar os fluxos comunicativos surgidos nas esferas públicas periféricas em poder comunicativo, a fim de informar e subsidiar os processos de deliberação nos corpos políticos de tomada de decisão. Como transportar tais demandas para as instâncias políticas formais, de maneira a contribuir na formulação de programas administrativos? Por certo, muitas das práticas comunicativas que acontecem nos contextos socioculturais da vida diária e esferas públicas periféricas não alcançam as instâncias formais do sistema político. Permanecem sem expressão política e, conseqüentemente, sem eficácia política. (MAIA, 2001, p. 11-12)

Com Cohen (2007) compreende-se o papel das instituições em promover uma esfera pública formal e resguardar um procedimento ideal de deliberação que garanta os princípios éticos do modelo da democracia deliberativa:

As instituições em uma democracia deliberativa não servem simplesmente para implementar os resultados da deliberação, como se a livre deliberação pudesse ocorrer na ausência de instituições adequadas. Nem o compromisso com as decisões deliberativas, nem a capacidade de as alcançar, são algo que podemos assumir independentemente da ordem institucional. As instituições devem oferecer a estrutura para a formação da vontade; determinam se há igualdade, se a deliberação é livre e fundamentada, se há autonomia, etc. (COHEN, 2007, p.137)<sup>6</sup>

Nessa linha, para além das deliberações na sociedade civil e esfera pública, tem-se também as práticas deliberativas promovidas por instâncias de governo e instituições participativas, ou seja, as interfaces entre Estado e sociedade voltadas para a participação. O conceito de sistemas deliberativos visa a encampar esse amplo conjunto de práticas deliberativas e de suas conexões.

Outro elemento importante para descrição da deliberação neste contexto é a ideia - mais ampla e inclusiva - da deliberação como uma forma de diálogo. “Bohman considera a deliberação mais como uma forma de diálogo do que uma forma de discurso.” e ainda completa que para esse autor “(...) o diálogo público é possível mesmo quando não existe acordo unânime entre os partícipes ou quando os interlocutores não se encontram face a face”. (BOHMAN apud FARIA, 2007, p.103).

Ainda nessa direção, muito além da prática discursiva argumentativa, a deliberação inclui outras instâncias de comunicação que ampliam em muito a visão das possibilidades de perscrutar a esfera pública informal e nos leva a uma ampliação do universo possível de análise dessa mesma esfera. Assim quando

(...), em contraposição à Habermas, Cohen e Bohman e junto com Young e os democratas da diferença, defende não só o discurso racional como meio da deliberação autêntica, mas também outras formas de expressão como a retórica, a emoção, o humor, a fofoca e o testemunho. Dryzek acredita que tais formas podem ser indutoras de maior inclusão política. (FARIA, 2007, p.104)

A visão alarga- se mais ao se considerar a esfera pública de forma mais ampliada, como indica Faria (2007):

---

<sup>6</sup> Tradução nossa do original: “Las instituciones en una democracia deliberativa no sirven simplemente para implementar los resultados de la deliberación, como si la libre deliberación pudiera realizarse en ausencia de las instituciones adecuadas. Ni el compromiso con las decisiones deliberativas, ni la capacidad para llegar a ellas, son algo que podamos asumir de manera independiente al orden institucional. Las instituciones deben ofrecer el marco para la formación de la voluntad; ellas determinan si hay igualdad, si la deliberación es libre y razonada, si hay autonomía, etc.” (COHEN, 2007, p.137)

Dryzek sugere ainda que a esfera pública seja pensada como “uma arena de contestação de discursos sobrepostos” cujo resultado no sistema político nem sempre é decisivo, mas que, em determinados momentos, faz diferença e induz mudanças no conteúdo da política pública (FARIA, 2007, p. 104).

Em síntese, não uma única conceituação e entendimento sobre o que se considera deliberação, sobre o funcionamento da esfera pública e sobre os diferentes lugares nos quais ocorrem as práticas deliberativas. Nesse sentido, no debate contemporâneo, diversos teóricos deliberativos conduzem à ideia central de que as práticas de deliberação compõem um sistema complexo e exige uma abordagem que vislumbra e busca compreender tal complexidade.

Os sistemas complexos, conforme Furtado, Sakowski e Tóvolli (2015) discorre

“podem ser definidos de forma ampla e abraçar conceitos de diferentes campos da ciência, da física à biologia, à computação e às Ciências Sociais. O conceito central de sistemas complexos pressupõe sistemas dinâmicos, não lineares, que contêm grande número de interações entre as partes. Esses sistemas se modificam, de modo a aprenderem, evoluírem e adaptarem-se e geram comportamentos emergentes e não determinísticos. (FURTADO; SAKOWSKI; TÓVOLI, 2015, p.23)

Nessa direção, tem-se desenvolvimentos teóricos no campo da democracia deliberativa nucleados em torno da perspectiva sistêmica (MANSBRIDGE et al., 2012). A democracia deliberativa é vislumbrada como sistema que merece uma abordagem capaz de dar conta da complexidade de tal fenômeno social.

## **2.1 Sistemas deliberativos: definições e características**

Ao propor a abordagem sistêmica para o campo da deliberação, Mansbridge et al. (2012) – que constitui uma referência central dessa perspectiva - convoca a entender as diversas instâncias de deliberação além do desenho institucional e das experiências locais mais facilmente observáveis:

Para entender o objetivo maior da deliberação, sugerimos que é necessário ir além do estudo de instituições e processos individuais para examinar sua interação no sistema como um todo. Reconhecemos que a maioria das democracias são entidades complexas nas quais uma ampla variedade de instituições, associações e locais de contestação realizam trabalho político -

incluindo redes informais, a mídia, grupos organizados de defesa, escolas, fundações, instituições privadas e sem fins lucrativos, legislaturas, agências executivas e os tribunais. Assim, defendemos o que pode ser chamado de abordagem sistêmica da democracia deliberativa. [...] Um sistema aqui significa um conjunto de partes distinguíveis, diferenciadas, mas em algum grau interdependentes, frequentemente com funções distribuídas e uma divisão de trabalho, conectadas de forma a formar um todo complexo. Requer diferenciação e integração entre as partes. Requer alguma divisão funcional do trabalho, de modo que algumas partes realizam trabalhos que outras não podem fazer tão bem. E requer alguma interdependência relacional, de modo que uma mudança em um componente trará mudanças em alguns outros. (MANSBRIDGE ET AL., 2012 p. 1 - 4)<sup>7</sup>

Mansbridge et al. (2012) aponta características dos sistemas deliberativos que são comuns a todo sistema. Ou seja, há um conjunto de fundamentos que explicam a natureza e características dos sistemas. O primeiro fundamento é que há uma interação muito importante e observável entre as partes. A consideração dessa característica do sistema, permite considerar que a democracia deliberativa seja analisada em termos sociais de grande escala.

A abordagem sistêmica não exige que tomemos uma nação ou grande governo como nosso objeto de estudo. Escolas e universidades, hospitais, mídia e outras organizações podem ser entendidos ao longo das linhas oferecidas por uma abordagem de sistema deliberativo. Mas, ao permitir a possibilidade de aumentar a escala e a complexidade das inter-relações entre as partes, essa abordagem nos permite pensar sobre as decisões democráticas sendo tomadas no contexto de uma variedade de locais e instituições deliberativas, interagindo entre si para produzir um sistema deliberativo saudável. (MANSBRIDGE et al., 2012, p. 2)<sup>8</sup>

---

<sup>7</sup> Tradução nossa do original: "To understand the larger goal of deliberation, we suggest that it is necessary to go beyond the study of individual institutions and processes to examine their interaction in the system as a whole. We recognize that most democracies are complex entities in which a wide variety of institutions, associations, and sites of contestation accomplish political work – including informal networks, the media, organized advocacy groups, schools, foundations, private and non-profit institutions, legislatures, executive agencies, and the courts. We thus advocate what may be called a systemic approach to deliberative democracy.[...] A system here means a set of distinguishable, differentiated, but to some degree interdependent parts, often with distributed functions and a Division of labour, connected in such a way as to form a complex whole. It requires both differentiation and integration among the parts. It requires some functional division of labour, so that some parts do work that others cannot do as well. And it requires some relational interdependence, so that a change in one component will bring about changes in some others." (MANSBRIDGE et al., 2012 p. 1 - 4)

<sup>8</sup> Tradução nossa do original: "The systemic approach does not dictate that we take a nation or large polity as our object of study. Schools and universities, hospitals, media, and other organizations can be understood along the lines offered by a deliberative system approach. But in allowing for the possibility of ratcheting up the scale and complexity of interrelations among the parts, this approach enables us to think about democratic decisions being taken in the context of a variety of deliberative venues and

O segundo fundamento é que “a interação entre as partes pode levar à auto-organização do sistema, sem a necessidade de controle central” (FURTADO; SAKOWSKI; TÓVOLI, 2015, p.24). E o terceiro atributo refere-se às interações em sistemas complexos e seus efeitos temporais: “ações, em um determinado momento, refletem-se sobre as possibilidades e limitações nos momentos seguintes. É por isso que os sistemas complexos são caracterizados como sistemas que aprendem, se adaptam e evoluem.” (FURTADO; SAKOWSKI; TÓVOLI, 2015, p. 24).

Essas características dos sistemas complexos podem ser associadas à base para a abordagem sistêmica das deliberações, devendo-se considerar que há especificidades dos sistemas deliberativos, visto que são sistemas sociais. Conforme as descrições de Mansbridge et al. (2012) os sistemas deliberativos também aprendem e são altamente dinâmicos:

Esperamos que um sistema deliberativo altamente funcional seja redundante ou potencialmente redundante na interação, de modo que quando uma parte deixa de desempenhar um papel importante, outra pode preencher ou evoluir ao longo do tempo para preencher. Tal sistema incluirá verificações e equilíbrios de várias formas, de maneira que os excessos em uma parte sejam controlados pela ativação de outras partes do sistema. (MANSBRIDGE et al., 2012, p. 5)<sup>9</sup>

Os sistemas deliberativos devem seguir princípios básicos da democracia deliberativa que constituem seus fundamentos:

1) a participação na deliberação é regulada por normas de igualdade e simetria; todos têm as mesmas chances de iniciar atos de fala, questionar, interrogar e abrir o debate; 2) todos têm o direito de questionar os tópicos fixados no diálogo; e 3) todos têm o direito de introduzir argumentos reflexivos sobre as regras do procedimento discursivo e o modo pelo qual elas são aplicadas ou conduzidas. Não há *prima facie* regras que limitem a agenda da conversação, ou a identidade dos participantes, contanto que cada pessoa ou grupo excluído possa mostrar justificadamente que são atingidos de modo relevante pela norma proposta em questão. (BENHABIB *apud* LEAL 2011, p. 66-67)

---

institutions, interacting together to produce a healthy deliberative system.” (MANSBRIDGE et al., 2012, p. 2)

<sup>9</sup>Tradução nossa do original: “We expect that a highly functional deliberative system will be redundant or potentially redundant in interaction, so that when one part fails to play an important role another can fill in or evolve over time to fill in. Such a system will include checks and balances of various forms so that excesses in one part are checked by the activation of other parts of the system.” (MANSBRIDGE et al., 2012, p. 5)

Do que se trata quando se toma a deliberação democrática como objeto específico da abordagem sistêmica? Trata-se de lidar com um objeto com os limites muito ampliados em relação à pesquisa, muito além dos fóruns e instituições de participação locais. É de se esperar, portanto, que o pesquisador, passe a lidar com um universo de informação especializada, gerada pelos inúmeros órgãos e instituições nacionais e supranacionais que atuam como agentes na condição de larga escala das práticas democráticas:

Como um primeiro corte, adotamos aqui uma abordagem institucional em que o sistema deliberativo é conceituado e avaliado no funcionamento dentro das fronteiras dos estados-nação, estados supranacionais, órgãos internacionais de tomada de decisão e as instituições internacionais com as quais os estados-nação e estados supranacionais são vinculados. Nossa análise se aplica a todas as instituições governamentais e não governamentais, incluindo redes de governança e as redes informais de amizade que conectam indivíduos e grupos discursivamente em questões de interesse comum. (MANSBRIDGE et al., 2012, p. 5)<sup>10</sup>

Dessa forma, os sistemas deliberativos incluem as diversas práticas deliberativas, incluindo os espaços institucionais nos quais se formulam políticas públicas e se tomam decisões de governo. Os sistemas deliberativos entrecruzam-se com os processos de formação de políticas públicas, desde a construção de suas agendas, passando formação (informal) da opinião pública até às tomadas de decisões, nos casos nos quais se prevê formas de participação institucionalizada.

Aceita-se, portanto, que diversas interações discursivas não formais atuam como forças que formulam as decisões. É fato que essas interações discursivas se dão através das redes sociais presenciais, bem como das redes de interação digital. É possível visualizar essas interações na sua estrutura de dados, ora como dados informacionais organizados ou ainda como pegadas digitais. Todas essas transações discursivas que circulam por diversos meios e formas são parte dos sistemas deliberativos mesmo que o observador não consiga perceber a relação direta com as decisões e com o processo de tomada de construção de políticas públicas.

---

<sup>10</sup> Tradução nossa do original: "As a first cut, we adopt here an institutional approach in which the deliberative system is conceptualized and evaluated as it functions within the boundaries of nation states, supranational states, international decisionmaking bodies, and the international institutions with which the nation states and supranational states are linked. Our analysis applies to all governmental and non-governmental institutions, including governance networks and the informal friendship networks that link individuals and groups discursively on matters of common concern." (MANSBRIDGE et al., 2012, p. 5)

Então, propõem-se considerar a metáfora gráfica de uma rede de nós e não mais o modelo em que um círculo central é seguido de círculos periféricos:

Um mapa de nós no sistema deliberativo revelaria muitos nós, com múltiplas formas de comunicação entre eles. Esses nós incluiriam órgãos do Estado-nação em diferentes níveis de governo e com suas diferentes casas legislativas, agências administrativas, militares e as equipes de todos eles; organismos internacionais em diferentes níveis e suas equipes; corporações multinacionais e negócios locais; comunidades epistêmicas; fundações; partidos políticos e facções dentro desses partidos; campanhas partidárias e outros fóruns partidários; corpos religiosos; escolas; universidades com seus departamentos, campos e associações disciplinares; sindicatos, grupos de interesse, associações voluntárias e organizações não governamentais (ONGs) tanto ad hoc quanto de longa data; movimentos sociais com seus enclaves e sua participação mais ampla; a mídia, incluindo a internet, blogs, mídia social, mídia interativa, livros, revistas, jornais, filmes e televisão; conversa informal entre indivíduos politicamente ativos ou menos ativos, poderosos ou marginalizados; e formas de conhecimento subjugado e local que raramente surgem para serem acessadas por outros sem alguma abertura no sistema deliberativo. (MANSBRIDGE et al., 2012, p. 10)<sup>11</sup>

O que se tem é uma rede de conexões que geram trânsito de comunicação e interação dentro de um sistema. O desafio, então, para o pesquisador é identificar, analisar e estabelecer padrões dentro desse sistema. Mas há um grande desafio na análise e descrição de todo sistema:

Por definição, sistemas complexos são difíceis de se descrever, explicar e prever, por isso mesmo não podemos esperar políticas ideais. Mas podemos ter esperança de melhorar, de fazer melhor. Dispondo-se de mais ferramentas, especialmente as ferramentas da ciência da complexidade, que se encontram em constante maturação e evolução, melhores cientistas e gestores de políticas públicas. (FURTADO; SAKOWSKI; TÓVOLI, 2015, p.19)

---

<sup>11</sup> Tradução nossa do original: "A map of nodes in the deliberative system would reveal many nodes, with multiple forms of communication among them. Those nodes would include nation state bodies at different levels of government and with their different legislative houses, administrative agencies, the military, and the staffs of all of these; international bodies at different levels and their staffs; multinational corporations and local businesses; epistemic communities; foundations; political parties and factions within those parties; party campaigns and other partisan forums; religious bodies; schools; universities with their departments, fields, and disciplinary associations; unions, interest groups, voluntary associations and non-governmental organizations (NGOs) both ad hoc and long-standing; social movements with both their enclaves and their broader participation; the media including the internet, blogs, social media, interactive media, books, magazines, newspapers, film, and television; informal talk among politically active or less active individuals whether powerful or marginalized; and forms of subjugated and local knowledge that rarely surface for access by others without some opening in the deliberative system." (MANSBRIDGE et al., 2012, p. 10)

Tomando sistemas deliberativos como sistemas complexos, tem-se o complicador de que uma deliberação considerada de baixa qualidade ou não democrática em uma instância individual pode contribuir para uma deliberação geral saudável, sob uma perspectiva de sistemas.

Como já posto, na vertente da democracia deliberativa, o debate público e as deliberações legitimam as ações dos governos e afetam a agenda das políticas públicas. Por isso é tão importante fortalecer os sistemas deliberativos, suas redes de debates midiáticos que, somados ao processo eleitoral, teriam o maior potencial de dar expressão à vontade dos cidadãos. (MENDONÇA, 2016).

O diálogo, as conversações, o debate público em locais informais ou mesmo dentro das instituições públicas, os discursos midiáticos, as interações comunicativas nas redes sociais presenciais e virtuais, os debates nos fóruns, as apresentações em audiências públicas, as defesas técnicas dos *experts*, as manifestações de grupos ativistas são o que formam a opinião pública. Portanto, compreender a opinião pública é tomar como desafio extrair dessas instâncias informação.

As TIC complexificam ainda mais esse cenário, na medida que permitem que uma parcela muito ampliada da população possa comunicar e mediar sua vontade política. Há um crescimento do número de atores deliberando fora das instituições formais de participação devido a atuação nas redes sociais digitais. E mesmo como instância formal, as tecnologias permitem as deliberações *on-line* e outros instrumentos de participação. Toda essa conversação política nas redes sociais é bastante fragmentada e descentralizada.

A consequência é um novo cenário das interações políticas que se consolida nas redes sociais virtuais e nas comunicações em rede, capaz de afetar as deliberações democráticas. Esse novo lugar de deliberação informal tem um caráter transnacional e avança além das fronteiras dos Estados e permite a formação de redes de interesses internacionais. Tais redes formam comunidades imaginadas, surgem novas dinâmicas espaciais e promovem o redescobrimto de identidades (BLANCO; DE LA FUENTE, 2010, p.4).

Toda essa interação digital cria uma nuvem de dados informacionais que são base para descrição do próprio fenômeno de formação da opinião pública nestes espaços virtuais - é a instância da democracia digital. É certo que há uma maior complexidade do fenômeno da formação da opinião pública em que se deve

considerar o fluxo intenso da informação, a interação em redes sociais e o exponencial crescimento dos meios de comunicação e ainda o papel cada vez mais significativo dos ativistas por meios digitais (MENDONÇA, 2016).

Para o modelo deliberativo de democracia digital a diferença e multiplicidade dos diálogos são considerados um fator positivo, pois

[...] assume que sempre existe diferença entre os sujeitos, diferença que requer um processo de discurso racional-crítico para que indivíduos de orientação privada se tornem cidadãos de orientação pública e para o desenvolvimento da opinião pública que possa orientar racionalmente a tomada de decisão democrática.” (DAHLBERG, 2001, p.116)<sup>12</sup>.

Nessa linha, Andrés Serbin (2005) atenta ainda para outra característica da deliberação contemporânea. Em contrapartida a essa interação política internacional, há o fortalecimento e crescimento de redes regionais com poder de influência e mobilização dos atores locais e que atuam fortemente nas instâncias digitais.

Os vínculos formados entre as iniciativas da sociedade civil nos níveis regional e global criam novas sinergias que contribuem para a democratização e o aumento da influência dos cidadãos nos processos internacionais de tomada de decisão. As organizações e redes da sociedade civil estão gradualmente desenvolvendo novas formas de “diplomacia cidadã” tanto em nível global quanto regional. (SERBIN, 2005, p.2)<sup>13</sup>.

Essa nova dimensão do embate político pode ser denominada de esfera pública digital que atua no contexto da democracia digital. O termo democracia digital abarca processos de deliberação digitais impulsionados por atores não governamentais - esfera informal - associada às tecnologias de redes sociais, bem como as iniciativas de participação e deliberação digital promovidas por instituições formais de participação.

---

<sup>12</sup> Tradução nossa do original: “[...] assumes that difference always exists between subjects, difference which necessitates a process of rational-critical discourse in order for privately-oriented individuals to become publicly-oriented citizens and for public opinion to develop that can rationally guide democratic decision-making.” (DAHLBERG, 2001, p.116)

<sup>13</sup> Tradução nossa do original: “The linkages formed between civil society initiatives at the regional and global levels create new synergies that contribute to the democratization of, and increased influence of citizens in international decision-making processes. Civil society organizations and networks are gradually developing new forms of “citizen diplomacy” on both the global and the regional levels.” (SERBIN, 2005, p.2)

Em síntese, a democracia digital pode ser definida a partir do modelo de democracia deliberativa que envolve a promoção da participação e da deliberação a fim de construir discursos mais racionais em torno das questões políticas e sobre as decisões dos governos.

Configuram-se, assim, diversos objetos de pesquisa a partir da abordagem sistêmica. A partir da discussão efetuada por Furtado, Sakowski e Tóvolli (2015, p. 23) considera-se que a compreensão desses objetos pode se beneficiar do uso de metodologias associadas a sistemas complexos. São exemplos dessas metodologias: análise de redes, modelagem baseada em agentes, simulação numérica, teoria dos jogos, formação de padrões e outras abordagens no âmbito de sistemas complexos.

Furtado, Sakowski e Tóvolli (2015) defendem que essas metodologias - discutidas no capítulo seguinte - podem contribuir com as análises de sistemas sociais. Nessa direção, argumenta-se que as metodologias e métodos da Ciência de Dados também podem ser relevantes para a compreensão dos sistemas deliberativos.

### 3 COMPUTATIONAL SOCIAL SCIENCE: MODELOS E MÉTODOS

Warren Weaver, no clássico artigo *Science and Complexity* (WEAVER, 1948) afirma que existem três tipos de problemas que vêm sendo trabalhados ao longo da história da ciência. Anteriormente a 1900, a ciência se preocupava largamente com problemas de poucas variáveis. A partir do século 20, cientistas desenvolveram técnicas poderosas de probabilidade para lidar com problemas que Weaver chamou de problemas de complexidade desorganizada, cujo número de variáveis é muito grande e em que cada uma das várias variáveis pode ter um comportamento totalmente desconhecido. Por fim existem os problemas de complexidade organizada, que lidam simultaneamente com um considerável número de fatores que estão inter-relacionados em um todo orgânico, problemas das ciências econômicas, biológicas e políticas não podem ser manipulados com técnicas estatísticas tão efetivas em descrever comportamentos de médias. Esses problemas necessitam de métodos e ferramentas mais poderosas para avançar a um nível de conhecimento ainda maior que as conquistas do século dezenove dos problemas de simplicidade ou as vitórias do século 20 sobre os problemas de complexidade desorganizada. Nas palavras de Weaver (1948):

Alguns cientistas buscarão e desenvolverão para si próprios novos tipos de arranjos colaborativos; que esses grupos terão membros provenientes essencialmente de todos os campos da ciência; e que essas novas formas de trabalhar, efetivamente instrumentadas por enormes computadores, contribuirão muito para o avanço que o próximo meio século certamente alcançará no tratamento dos complexos, mas essencialmente orgânicos, problemas das ciências biológicas e sociais. (WEAVER, 1948, p. 542)<sup>14</sup>

De fato, essa previsão vem se confirmando com a abordagem da Ciência de Dados aplicada nessas e em outras áreas do conhecimento. A Ciência da Computação vem evoluindo no desenvolvimento e aplicação de tecnologias de análise de dados. O conjunto dessas tecnologias tem se consolidado como uma nova área do conhecimento, multidisciplinar e com foco no estudo de dados visando a extração de

---

<sup>14</sup> Tradução nossa do original: “[...] some scientists will seek and develop for themselves new kinds of collaborative arrangements; that these groups will have members drawn from essentially all fields of science; and that these new ways of working, effectively instrumented by huge computers, will contribute greatly to the advantage which the next half century will surely achieve in handling the complex, but essentially organic, problems of the biological and social sciences.” (WEAVER, 1948, p. 542)

conhecimento, detecção de padrões e/ou obtenção de *insights*<sup>15</sup> para orientar tomadas de decisão em áreas muito diversas.

Particularmente, no contexto das Ciências Sociais, a computação tem sido utilizada para modelar, simular e analisar os fenômenos sociais a partir das técnicas desenvolvidas na Ciência de Dados. Uma abordagem que tem sido proposta é a da *Computational Social Science*, que, de forma sucinta, pode ser entendida como parte das Ciências Sociais que se apropria de abordagens computacionais para estudar o fenômeno social, trazendo para as Ciências Sociais uma capacidade de predição dos fenômenos sociais com altos índices de precisão.

Lazer et al. (2009) discute como a “digitalização” do comportamento humano tem gerado rastros ou pegadas digitais que formam uma fonte de dados abundante e preciosa para ajudar a compreender as interações humanas, alavancando a “capacidade de coletar e analisar dados com amplitude, profundidade e escala sem precedentes”. Isso tem se tornado possível com a consolidação dos métodos de “*big data*”. Segundo Hilbert (2015):

Seu [big data] impacto nas Ciências Sociais foi comparado com o impacto da invenção do telescópio para astronomia e da invenção do microscópio para biologia (fornecendo um nível de detalhe sem precedentes sobre o sistema de interesse). Confrontados com tal aumento no nível de granularidade perceptível na dinâmica social, os cientistas sociais têm a obrigação inevitável de fazer uso dele para informar análises, políticas e elaboração de projetos. (HILBERT, 2015, p. 2)<sup>16</sup>

As palavras de Hilbert mostram que o termo *big data* não deve ser entendido como apenas “mais dados”, antes disso, é uma mudança de paradigma científico. O estudo das correlações entre os dados é o cerne da Ciência de Dados, assim como o estudo das relações biológicas são o cerne das Ciências Biológicas e o estudo das relações humanas são o cerne das Ciências Sociais.

No contexto da *Computational Social Science*, as técnicas de *big data* se concentram mais especificamente na fonte de dados gerada pelas pegadas digitais

---

<sup>15</sup> O significado desse termo, obtido do site *léxico.com*, é “*an accurate and deep understanding*”. Optou-se por utilizar o termo em inglês para preservar o significado original.

<sup>16</sup> Tradução nossa do original: “Its impact on the social sciences has been compared with the impact of the invention of the telescope for astronomy and the invention of the microscope for biology (providing an unprecedented level of detail about the system of interest). Confronted with such an increase in the level of perceivable granularity in social dynamics, social scientists have an inevitable obligation to make use of it to inform analysis, policy and project design.” (HILBERT, 2015, p. 2)

(dados registrados a cada interação digital) que vão gerar indicadores substitutos na correlação com a variável de interesse. A grande vantagem dessa abordagem é a possibilidade de obter dados correlacionados em grande quantidade, em tempo real e a baixo custo.

Com as ferramentas estatísticas e computacionais adequadas, a análise dos dados ganha novas proporções: agora é possível realizar sistematicamente a descoberta do conhecimento diretamente dos dados. Essa abordagem permite que os cientistas se concentrem na interpretação e avaliação dos padrões descobertos, poupando anos de pesquisa que seriam gastos na análise desses dados, se utilizassem o método tradicional.

A descoberta do conhecimento a partir da explicitação dos padrões de correlação entre variáveis que ocorrem em grandes bases de dados é chamada de mineração de dados. A partir disso, é possível encontrar modelos matemáticos que governam relações entre essas variáveis. Esses modelos matemáticos possibilitam prever qual é o comportamento esperado para uma parte do mesmo conjunto de dados, mas que não foram analisados na geração do modelo. Essa é a essência da aprendizagem de máquina. Porém, a própria característica dessa abordagem impõe uma limitação fundamental: como os dados estão todos no passado, os modelos de aprendizagem de máquina são incapazes de prever mudanças de comportamento dos dados no futuro.

Lazer et al. (2014) discute essa limitação a partir de um estudo de Ginsberg, et al. (2009) que obteve grande sucesso ao gerar um modelo de predição capaz de detectar surtos de gripes em áreas de grande população a partir apenas da análise de consultas de pesquisa no Google. O mesmo modelo que se mostrou muito efetivo nas predições na época do estudo, no ano de 2009, se mostrou cada vez mais ineficaz com o passar do tempo. A resposta para essa limitação passa por uma modelagem, baseada na teoria, que permita as variáveis do modelo serem ajustadas com valores nunca antes observados. Aqui ocorre a virada para outro método, a simulação computacional.

Hilbert (2015) propõem uma abordagem metodológica para a *Computational Social Science* baseada em três eixos: empírico, analítico e teórico. O eixo empírico se apropria dos conceitos e técnicas de *big data* que extrapolam as tecnologias tradicionais de bancos de dados para lidar de forma mais apropriada com enormes volumes de dados; dados oriundos de variadas fontes; disponibilidade da

informação em velocidade adequada, apesar da quantidade dos dados; variedade das fontes de dados; veracidade da informação; utilidade da informação gerada. O eixo analítico aborda tecnologias tais como inteligência artificial, aprendizagem de máquina, análise de redes sociais e processamento de linguagem natural. Por fim, o eixo teórico lida com a simulação computacional dos fenômenos sociais analisados.

A abordagem de três eixos será adotada nesta monografia como guia para descrever métodos de Ciência de Dados aplicáveis ao estudo da democracia deliberativa como fenômeno social.

Conforme visto nos capítulos anteriores, considera-se neste trabalho que os sistemas deliberativos podem, para fins de modelagem e análise, ser classificados como sistemas complexos. Ao descrever os elementos dos sistemas complexos, Furtado, Sakowski e Tóvolli (2015) apresentam três conjuntos de atributos:

Em primeiro lugar, está a ideia forte de interação entre as partes, considerando as escalas e dando relevância espacial e temporal. Essas interações, por sua vez, levam a um sistema que não é redutível a suas partes. Ou seja, um sistema que não pode ser descrito apenas pelos atributos de suas partes componentes. [...] Em segundo lugar, a interação entre as partes pode levar à auto-organização do sistema, sem a necessidade de controle central. Isto implica que as interações locais podem gerar comportamentos que emergem de baixo para cima. Esse conceito poderoso pode ser ilustrado para o leitor iniciante com o exemplo de voo de pássaros em formação. Nenhum pássaro específico controla a direção e posição de todas as aves em voo. Cada pássaro só observa aqueles mais próximos e sincroniza seu voo com seus vizinhos imediatos. Como resultado, a partir das aves individuais (de baixo para cima), o voo coordenado emerge. Um terceiro atributo a ressaltar é que os sistemas complexos podem experimentar feedback. Em sistemas complexos, as interações têm efeitos no tempo: ações, em um determinado momento, refletem-se sobre as possibilidades e limitações nos momentos seguintes. É por isso que os sistemas complexos são caracterizados como sistemas que aprendem, se adaptam e evoluem. (FURTADO, SAKOWSKI E TÓVOLI, 2015, p.22)

Tais atributos estão presentes na abordagem de sistêmica proposta por Mansbridge et al (2012) para a democracia deliberativa:

Um sistema aqui significa um conjunto de partes distinguíveis, diferenciadas, mas em algum grau interdependentes, frequentemente com funções distribuídas e uma divisão de trabalho, conectadas de modo a formar um todo complexo.

[...] Uma abordagem sistêmica, em nossa visão, não requer que cada componente tenha uma função ou que cada componente seja interdependente entre si, de modo que uma mudança em um trará

automaticamente uma mudança em todos os outros. Se um componente contribui para uma função, não é necessário que a função seja cumprida de forma otimizada em um local, uma vez que em um sistema deliberativo a mesma função pode ser distribuída por vários subsistemas. [...] Nosso ponto é que normativamente, na abordagem sistêmica, todo o ônus da tomada de decisão e da legitimidade não recai sobre um fórum ou instituição, mas é distribuído entre diferentes componentes em diferentes casos.

Esperamos que um sistema deliberativo altamente funcional seja redundante ou potencialmente redundante na interação, de modo que quando uma parte deixa de desempenhar um papel importante, outra pode preencher ou evoluir ao longo do tempo para preencher. Tal sistema incluirá verificações e equilíbrios de várias formas, de tal maneira que os excessos em uma parte sejam controlados pela ativação de outras partes do sistema. Também imaginamos sistemas que são dinâmicos em vez de estáticos. (MANSBRIDGE et al, 2012, p.4-5)<sup>17</sup>

Espera-se que, com esse mapeamento das características de sistemas deliberativos para os elementos dos sistemas complexos, seja reforçada a hipótese de que, assim como Furtado, Sakowski e Tóvolli (2015) propõem passa as políticas públicas, no caso dos sistemas deliberativos as "metodologias disponíveis para o estudo de tais sistemas tornam-se disponíveis [...] contribuindo para melhorar a sua análise". Dessa forma, conclui-se que a abordagem de três eixos da *Computational Social Science* pode ser considerada válida no contexto da democracia deliberativa.

Nas próximas seções deste capítulo serão descritos modelos e métodos aplicáveis em *Computational Social Science* e que podem ser apropriados pelos pesquisadores de democracia deliberativa com potencial de melhorar a efetividade dos estudos nessa área.

---

<sup>17</sup> Tradução nossa do original: "A system here means a set of distinguishable, differentiated, but to some degree interdependent parts, often with distributed functions and a division of labour, connected in such a way as to form a complex whole.

[...] A systemic approach, in our view, does not require that every component have a function or that every component be interdependent with every other such that a change in one will automatically bring about a change in all others. If a component does contribute to a function, it is not necessary that the function be fulfilled optimally in one location, since in a deliberative system the same function may be distributed across various subsystems. [...] Our point is that normatively, in the systemic approach the entire burden of decision-making and legitimacy does not fall on one forum or institution but is distributed among different components in different cases.

We expect that a highly functional deliberative system will be redundant or potentially redundant in interaction, so that when one part fails to play an important role another can fill in or evolve over time to fill in. Such a system will include checks and balances of various forms so that excesses in one part are checked by the activation of other parts of the system. We also envision systems that are dynamic rather than static." (MANSBRIDGE et al, 2012, p.4-5)

### 3.1 Teoria da informação

A teoria da informação é a área do conhecimento que estuda a definição, quantificação, armazenamento dos elementos envolvidos numa troca de informações por um meio ou canal de comunicação. Surge a partir do artigo “*A Mathematical Theory of Communication*” de Claude E. Shannon (SHANNON, 1948) que investiga o problema fundamental da comunicação: reproduzir exatamente ou aproximadamente uma mensagem de um local de origem para outro local. Esse artigo introduziu os conceitos de entropia (medida da quantidade de informação contida em uma variável aleatória) e de informação mútua (medida da quantidade de informação que uma variável aleatória contém em relação a outra) como expressões probabilísticas que formam a base do modelo matemático da teoria da informação. Dado que a entropia é a soma de todos os possíveis valores que uma variável pode assumir, ela pode ser vista como a medida da incerteza de que um desses valores ocorra.

O objetivo de Shannon (1948) era resolver o problema da transmissão de dados por um canal de comunicação ruidoso e para isso demonstrou como esses dados podem ser codificados de forma a permitir comunicação eficiente entre dois pontos. O legado desse trabalho foi a formalização matemática do conceito de informação aplicável em diversas áreas.

O fundamento por trás da teoria da informação é: a ocorrência de um evento improvável é mais informativa do que a ocorrência de um evento provável. Esse fundamento pode ser formalizado por três propriedades:

- a) Os eventos prováveis devem ter baixo conteúdo de informação e, em casos extremos, os eventos com garantia de ocorrência não devem ter nenhum conteúdo de informação;
- b) Eventos menos prováveis devem ter maior conteúdo de informação;
- c) Os eventos independentes devem ter informações adicionais. Por exemplo, descobrir que uma moeda lançada deu cara duas vezes deve transmitir o dobro de informações do que descobrir que uma moeda lançada deu cara.

O conhecimento de teoria da informação é base para os métodos de mineração de dados (discutido na seção 3.2.2) e de aprendizagem de máquina (discutido na seção 3.2.3). Esta monografia não descreve as fórmulas ou os teoremas estatísticos que fundamentam essas técnicas. Para um aprofundamento nessas

questões, indica-se os roteiros de estudo que Mackay (2003) sugere: formação em reconhecimento de padrões e redes neurais (MACKAY, 2003, p. vii) e formação em inferência bayesiana e aprendizado de máquina (MACKAY, 2003, p. x).

### 3.1.1 Teoria da Decisão

O princípio da teoria da decisão é a obtenção de boas escolhas sob um cenário de incertezas.

Formalmente, o conceito da teoria da decisão é bem simples: dado uma ação  $\mathbf{a}$  que pode influenciar um estado  $\mathbf{x}$ , cuja distribuição de probabilidade é  $p(\mathbf{x} | \mathbf{a})$ , há um ganho que se recebe quando ocorre o estado  $\mathbf{x}$  ao ser escolhida a ação  $\mathbf{a}$ . Esse ganho é governado por uma função  $U(\mathbf{x}, \mathbf{a})$ , chamada função de utilidade. A tarefa da teoria da decisão é selecionar a ação que maximize a utilidade esperada, uma vez que a utilidade esperada,  $E[u | \mathbf{a}]$ , é definida pelo somatório de  $p(\mathbf{x} | \mathbf{a}) U(\mathbf{x}, \mathbf{a})$  para cada estado  $\mathbf{x}$ , ou seja:

$$E[u] = \sum_{\mathbf{x}} p(\mathbf{x} | \mathbf{a}) U(\mathbf{x}, \mathbf{a})$$

O problema computacional se resume a calcular o maior valor de  $E[U | \mathbf{a}]$  sobre  $\mathbf{a}$ , de forma que esse valor determina a melhor ação a ser adotada de forma a alcançar o estado desejado.

Em problemas reais, a escolha de uma função de utilidade apropriada pode ser uma tarefa difícil. Outro dificultador é determinar a melhor decisão quando se deve tomar não apenas uma ação, mas uma sequência de ações, cada uma influenciando um estado futuro, que poderá causar algum efeito na ação subsequente. Isso faz com que o conjunto de possibilidades de ação seja enorme e determinar a melhor decisão pode se tornar impossível, porque a utilidade esperada de uma ação não pode ser calculada exatamente. É o que ocorre, por exemplo, em jogos de tabuleiro como o jogo de xadrez e explica porque é tão complicado construir a máquina perfeita para a tomada de decisão.

A teoria da decisão formaliza um método para tomada racional de decisão a partir da maximização do valor esperado de utilidade sobre todos os estados possíveis do mundo. Mas, em muitos casos, o tomador de decisão pode utilizar outros critérios para obter a melhor escolha.

O critério otimista identifica o melhor que poderia acontecer em cada ação e então escolhe a ação com o maior valor. Ao presumir que poderá obter o máximo possível, então decide de acordo com o melhor cenário possível. O exemplo típico é o jogador de loteria, que procura grandes recompensas mesmo ignorando as probabilidades de efetivamente obter algum ganho.

Outro critério é o pessimista, que busca identificar o pior que poderia acontecer em cada ação e então escolhe a ação com o maior retorno. Aqui se assume que o pior caso ocorrerá, e, portanto, se busca o maior ganho para esse caso. Um exemplo é o investidor em caderneta de poupança, que aceita receber uma remuneração menor para evitar investimentos de maior risco.

Um terceiro critério é o oportunista, onde, a partir de uma tabela de perdas possíveis, escolhe (antes que o evento ocorra) aquela opção que trará o menor prejuízo. É uma estratégia de tomada de decisão que visa minimizar as perdas ao invés de maximizar os ganhos.

### **3.2 Métodos de Ciência de Dados**

O termo Ciência de Dados remete à uma reflexão intrigante: o que poderia significar uma ciência que se dedica ao estudo dos dados? Inicialmente é necessário entender o que é um dado. No contexto da computação, entende-se o dado como uma informação dissociada do seu significado, ou seja, fora de um contexto. Logo, informação é o dado associado a um contexto que lhe fornece sentido. Por exemplo, o número 31 é um dado, não expressa nenhum significado por si, mas no contexto de tempo decorrido, representa uma idade; enquanto no contexto de código de área de telefonia, representa uma localidade.

Porto e Ziviani (2014) lembram que o estudo metódico para a extração de conhecimento, de forma generalizada e em escala e a partir de uma imensa massa de dados é um desafio técnico-científico em computação. Os autores relacionam a resposta a esse desafio à emergência de uma nova espécie de ciência que incorpora elementos, técnicas e teorias de ciências básicas.

Provost e Fawcett (2013) definem a Ciência de Dados de forma sucinta como um conjunto de princípios fundamentais que apoiam e orientam a extração de informações e conhecimento dos dados. Van der Aalst (2016) enfatiza o caráter

interdisciplinar e indica caminhos para a elaboração de atividades investigativas, propondo uma definição detalhada para a Ciência de Dados:

A Ciência de Dados é um campo interdisciplinar com o objetivo de transformar dados em valor real. Os dados podem ser estruturados ou não estruturados, grandes ou pequenos, estáticos ou streaming. O valor pode ser fornecido na forma de previsões, decisões automatizadas, modelos aprendidos com os dados ou qualquer tipo de visualização de dados que forneça *insights*. A Ciência de Dados inclui extração de dados, preparação de dados, exploração de dados, transformação de dados, armazenamento e recuperação, infraestruturas de computação, vários tipos de mineração e aprendizagem, apresentação de explicações e previsões e a exploração de resultados levando em consideração questões éticas, sociais, legais e aspectos de negócios. (VAN DER AALST, 2016, p10)<sup>18</sup>

O autor apresenta a ideia de transformar dados em valor para indivíduos, organizações e sociedade. Como valor, entende-se conhecimento - uma síntese de informações enriquecidas com um significado mais abrangente. Van der Aalst (2014, p.2) afirma que a Ciência de Dados visa a responder perguntas baseadas nos dados, classificando-as em quatro categorias:

- a) Notificação: o que aconteceu?
- b) Diagnóstico: por que isso aconteceu?
- c) Predição: o que vai acontecer?
- d) Recomendação: qual é o melhor que pode acontecer?

As subseções a seguir descrevem algumas técnicas e métodos de Ciência de Dados bastante utilizados em *Computational Social Science*.

---

<sup>18</sup> Tradução nossa do original: "Data science is an interdisciplinary field aiming to turn data into real value. Data may be structured or unstructured, big or small, static or streaming. Value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing infrastructures, various types of mining and learning, presentation of explanations and predictions, and the exploitation of results taking into account ethical, social, legal, and business aspects." (VAN DER AALST, 2016, p10)

### 3.2.1 Big data

Os métodos de big data foram desenvolvidos para lidar com conjuntos de dados muito grandes ou complexos para serem tratados pelos softwares de processamento de dados tradicionais. Laney (2001) definiu três dimensões de desafios que as organizações de gerenciamento de dados deveriam lidar: volume, velocidade e variedade. Mais tarde uma nova dimensão, a veracidade, se juntou a esses desafios para formar o que é conhecido atualmente como os quatro “V’s do *big data*”:

- a) **Volume:** refere-se ao tamanho de alguns conjuntos de dados. As abordagens para lidar com o crescimento de dados localmente resolvem bem esse problema, afinal, dispositivos de armazenamento estão cada vez mais desenvolvidos, mas como lidar com dados distribuídos globalmente?
- b) **Velocidade:** refere-se à capacidade de processar os dados num intervalo de tempo conveniente. Num banco de dados local, os dados estão disponíveis tão logo sejam registrados, mas em bancos de dados distribuídos o registro de dados é uma tarefa mais complexa, pois precisa levar em consideração outros fatores além da capacidade de processamento e de armazenamento local. Existem análises de dados cujos resultados somente são úteis se forem obtidos num tempo adequado. Por exemplo, de que adianta um modelo de previsão do tempo que só gera resultados depois do dia da previsão?
- c) **Variedade:** os bancos de dados tradicionais são muito bons para lidar com dados estruturados e provenientes de uma única fonte de dados, como os dados de uma ficha de cadastro, por exemplo. Mas como lidar com dados não estruturados e originados de diversas fontes, tais como imagens sons e vídeos?
- d) **Veracidade:** refere-se ao quanto pode-se confiar nos dados, ou seja, o quanto o dado é verdadeiro. Esse aspecto se mostra ainda mais importante ao manipular os dados, juntando dados de várias fontes diferentes para construir uma informação mais completa. Ao fazer isso é preciso garantir que a informação gerada seja verdadeira.

### 1.1.1.1 Características de *big data* nas análises em Ciências Sociais

Uma forma de tentar compreender a sociedade é analisando as pegadas digitais. Essas pegadas digitais são os dados que as pessoas geram ao interagir com os meios digitais, ou seja, ao realizar uma chamada no telefone celular, ao acessar um site, ao interagir numa rede social. Até mesmo ao se deslocar de um lugar para o outro com um celular no bolso são gerados dados de localização dessa pessoa. Todos esses dados estão armazenados em algum lugar e podem ser acessados e analisados. Em suma, a pegada digital é um efeito colateral da interação das pessoas com o ambiente digital, portanto é produzida automaticamente como um dado de baixo custo e quase inevitável.

A pegada digital nos dá a evidência empírica, num paradigma científico computacional, para estudar a sociedade. Não é necessário coletar esses dados, eles já foram coletados, mas eles precisam ser armazenados e essa armazenagem pode ser extremamente grande. Geralmente essa pegada digital é produzida em tempo real e pode-se analisá-la também em tempo real.

Hilbert (2015) cita os atributos que caracterizam *big data* para análises em Ciências Sociais:

- a) Permite substituir a amostragem aleatória pela captura de todo o universo de dados disponível. Mocanu et al. (2013) usou postagens no Twitter para mostrar que esses dados permitem o estudo da geografia da língua em escalas que vão desde a agregação em nível de país até bairros específicos da cidade.
- b) Geralmente os dados estão disponíveis em tempo real. Um exemplo notável dessas fontes de dados são as redes sociais on-line como Twitter, Facebook e Instagram, dentre outras. Chunara et al. (2012) também utilizando postagens no Twitter conseguiu estimar padrões epidemiológicos do surto de cólera do Haiti em 2010 com até duas semanas de antecedência em relação aos dados oficiais.
- c) São dados confusos e incompletos, o que pode ser compensado pela redundância de dados de diferentes fontes, podendo ser recombinados num processo chamado de fusão de dados.

- d) A detecção de padrões por meio de ferramentas de mineração de dados e de aprendizado de máquina, leva a uma forma automatizada de informar o processo de tomada de decisão.

#### 1.1.1.2 Limitações de *big data* na análise em Ciências Sociais

Existem limitações do uso da pegada digital para fins de pesquisa em Ciências Sociais. Algumas limitações mais comuns são:

- a) Representatividade: a pegada digital não é necessariamente representativa, trabalhar com toda a população de usuários do Facebook, significa lidar com todos os usuários do Facebook. Mas isso não significa que é toda a população mundial. Portanto, a pegada digital é tendenciosa porque é influenciada pela exclusão digital, ou seja, existe uma parte da população sobre representada. Em contrapartida há outra parte sub representada no mundo digital.
- b) Realidade: trata-se de dados e não da realidade, assim como uma pegada não é o pé. Essa confusão, pode levar a tomada de decisões errôneas quando não absurdas. Pode-se citar, como exemplo, o tema do filme *Minority Report*, em que, num futuro ficcional, existe um departamento de polícia pré-crime, que tem por objetivo prender futuros assassinos antes que cometam os seus crimes, baseados em previsões feitas por pessoas que possuem poderes paranormais preditivos. No filme, dados preditivos determinavam quais pessoas seriam julgadas e condenadas, automaticamente, por crimes que não aconteceram e nunca virão a ocorrer. Ainda que seja um exemplo fictício, ilustra o cuidado necessário com a automatização das tomadas de decisão baseadas nas análises de dados.
- c) Significância, discriminação e personalização: há um significado que pode ser detectado nos dados por meio de mineração de dados e aprendizado de máquina. Mas esse significado sempre faz sentido? Deve-se ter sempre em mente que ao descobrir padrões ocultos nos dados, esses padrões descobertos refletem tudo o que está nos dados. Refletem inclusive discriminações, preconceitos e até mesmo racismo.

Por exemplo, imagine-se que exista uma ferramenta hipotética capaz de agilizar a análise de currículos e convidar automaticamente os melhores candidatos para entrevista. Essa ferramenta utilizaria apenas os dados de contratações de funcionários por empresas brasileiras para tomar como base para uma decisão. É muito provável que essa ferramenta irá convidar mais homens brancos para entrevistas, visando cargos de alto escalão e mais mulheres pretas para cargos de serviços gerais. Isso ocorreria porque essa discriminação está presente nos dados que deram origem aos padrões. Um problema parecido ocorre quando se trabalha com algoritmos baseados na classificação dos dados. Classificar dados significa encontrar similaridades e diferenças entre eles, separar os dados em grupos a partir dessas similaridades e, às vezes, dar um rótulo para cada um desses grupos. Essa é a ideia por trás dos algoritmos de personalização nos sistemas de recomendação (MACEDO, 2015). Por outro lado, isso é a mesma coisa que discriminar. Por exemplo, ao classificar pessoas pela cor dos olhos, provavelmente o grupo dos olhos azuis será composto quase que exclusivamente por pessoas brancas, enquanto ao classificar pelo tipo de cabelo, o grupo do cabelo crespo terá uma predominância de pessoas pretas. Ainda que não intencionalmente, inteligência artificial pode ser racista e será, se os dados que a alimentam tiverem viés racista. Por outro lado, os algoritmos podem ser controlados de forma a desconsiderar variáveis que possam gerar classificações preconceituosas, discriminatórias e racistas ou mesmo garantir que essas variáveis tenham distribuição uniforme em cada um dos agrupamentos. Os resultados podem não ser tão precisos quanto seriam sem essa intervenção, mas essa perda de precisão pode ser medida de forma a se decidir fazer ou não a intervenção.

- d) Correlação versus causalidade: essa é talvez uma das limitações mais importantes da Ciência de Dados. Correlação não significa causalidade. Correlação significa basicamente que duas coisas caminham juntas, ou, mais formalmente, é qualquer relação estatística (causal ou não causal) entre duas variáveis. Causalidade significa que um evento causa outro. Quando não existe uma relação de causalidade em uma correlação,

trata-se de uma correlação espúria. Correlações espúrias podem ocorrer por coincidência ou por influência de uma terceira variável, chamada de variável de confusão. Métodos de mineração de dados e de aprendizagem de máquina não são orientados por teoria, eles apenas descobrem padrões e correlações estatísticas, por isso não são capazes de interpretar os resultados, portanto não revelam automaticamente relações de causalidade. Por outro lado, com mais e melhores dados é mais fácil detectar variáveis espúrias e com isso gerar apenas correlações com maior potencial de causalidade.

- e) Passado não é igual ao futuro: por mais óbvio que isso pareça, é importante porque significa que os dados dizem algo apenas sobre o passado. Predições baseadas nos dados são efetivas quando o futuro é uma repetição do passado. De fato, muitas vezes é assim que acontece. Um exemplo simples é que depois de um dia sempre vem uma noite e depois de uma noite sempre vem um dia. Entretanto, existem fenômenos em que os eventos futuros são independentes dos eventos passados. Por exemplo, após uma sequência de dias ensolarados, não há nenhuma garantia que o dia seguinte também será ensolarado. Mas uma teoria que relaciona as variáveis como umidade, temperatura, velocidade e direção do vento e pressão atmosférica numa fórmula matemática complexa poderia prever, com boa precisão, uma ocorrência inesperada de chuvas após um dia ensolarado. É por isso que métodos de mineração de dados e aprendizagem de máquina não são suficientes para substituir uma teoria.

### 3.2.2 Mineração de dados

Zaki e Meira (2014) definem mineração de dados como o processo de descoberta de padrões originais, interessantes e perspicazes, e de modelos preditivos, compreensíveis e descritivos a partir de conjuntos de dados em larga escala. As principais tarefas de mineração de dados abrangem a análise exploratória de dados, mineração de padrões frequentes, agrupamento e classificação.

A mineração de dados é um campo interdisciplinar que mescla conceitos de áreas afins, como sistemas de banco de dados, estatísticas, aprendizado de

máquina e reconhecimento de padrões por meio de algoritmos que permitem obter *insights* de uma grande massa de dados. Por fim, é parte de um processo maior de descoberta de conhecimento, que inclui tarefas de:

- a) Pré-processamento: extração de dados, limpeza de dados, fusão de dados e redução de dados;
- b) Processamento: obtenção de *insights*;
- c) Pós-processamento: interpretação de padrões e de modelos, geração e confirmação de hipóteses.

Um conjunto de dados podem ser representados como uma matriz de dados  $n \times d$ , com  $n$  linhas e  $d$  colunas, sendo que as linhas correspondem às entidades do conjunto de dados e as colunas representam os atributos ou propriedades dessas entidades. Dependendo do domínio do aplicativo, as linhas também podem ser referidas como entidades, instâncias, exemplos, registros, transações, objetos, pontos, tuplas e assim por diante. Da mesma forma, as colunas também podem ser chamadas de atributos, propriedades, recursos, dimensões, variáveis, campos e assim por diante. O número de instâncias  $n$  é conhecido como o tamanho dos dados, enquanto o número de atributos  $d$  é chamado de dimensionalidade dos dados. A análise de um único atributo é denominada análise univariada, enquanto a análise simultânea de dois atributos é chamada de análise bivariada e a análise simultânea de mais de dois atributos é chamada de análise multivariada.

Uma questão importante é que mesmo num conjunto massivo de dados, pode-se encontrar apenas um pequeno número efetivo de dados interessantes. Na verdade, os dados em uma variedade de domínios exibem a propriedade de que algumas coisas são muito comuns, mas a maioria das coisas é bastante rara.

Os atributos podem ser classificados em dois tipos principais, numéricos ou categóricos, dependendo dos tipos de valores que assumem. Um atributo numérico é aquele cujo valor é um número real ou inteiro. Um atributo categórico é aquele em que o valor é composto por um conjunto de símbolos. Por exemplo, **Sexo** e **Educação** podem ser atributos categóricos cujos valores seriam **(M, F)** e **(Fundamental, Médio, Superior)**, respectivamente. Os atributos categóricos podem ser de dois tipos:

- a) Nominal: em que os valores do atributo não são ordenados e, portanto, apenas as comparações de igualdade são significativas.

- b) Ordinal: em que os valores dos atributos são ordenados e, portanto, tanto as comparações de igualdade quanto às comparações de desigualdades são permitidas, embora não seja possível quantificar a diferença entre os valores.

As subseções a seguir descrevem brevemente as principais tarefas de mineração de dados. Informações mais conceituais e abrangentes sobre mineração de dados podem ser obtidas em Han, Pei e Kamber (2011) enquanto um aprofundamento dos algoritmos e das descrições matemáticas das principais técnicas podem ser obtidas em Zaki e Meira (2014).

### 1.1.1.3 Análise exploratória de dados

A análise exploratória de dados tem como objetivo explorar os atributos numéricos e categóricos dos dados individualmente ou em conjunto para extrair características-chave da amostra de dados por meio de estatísticas que fornecem informações sobre a centralidade, dispersão e dependência. É importante descrever os dados como um grafo, no qual os nós denotam os pontos e as arestas ponderadas denotam as conexões entre os pontos. Isso permite extrair atributos topológicos importantes que fornecem *insights* na forma de estruturas e modelos de redes e de grafos.

Há métodos que fornecem uma conexão fundamental entre a visão local independente dos dados e a visão abrangente que lida com semelhanças entre os pontos. Esses métodos permitem realizar análises não lineares usando métodos algébricos e estatísticos em espaços de alta dimensão, compreendendo dimensões "não lineares". Permitem extrair dados complexos, desde que haja uma maneira de medir a semelhança entre dois pares de objetos abstratos. Uma vez que a mineração de dados lida com conjuntos de dados massivos, com milhares de atributos e milhões de pontos, outro objetivo da análise exploratória é reduzir a quantidade de dados a serem minerados.

#### 1.1.1.4 Mineração de padrões frequentes

Se refere à tarefa de extrair padrões úteis e informativos a partir de conjuntos de dados massivos e complexos. Os padrões podem compreender conjuntos de valores de atributos correlacionados ou padrões complexos, como sequências (que consideram relacionamentos explícitos de precedência, ou posicional ou temporal, entre pontos), ou grafos (que consideram relações arbitrárias entre pontos). O objetivo principal é descobrir tendências e comportamentos ocultos nos dados para entender melhor as interações entre os pontos e atributos.

#### 1.1.1.5 Agrupamento

É a tarefa de particionar os pontos em grupos naturais chamados *clusters*<sup>19</sup>, de forma que os pontos dentro de um grupo sejam muito semelhantes, enquanto os pontos entre os clusters sejam tão diferentes quanto possível. Existem diferentes tipos de *clusters*, dependendo das características dos dados e dos agrupamentos desejados. Os tipos de agrupamentos podem ser:

- a) agrupamento baseado em representatividade: seu objetivo é particionar o conjunto de dados em *clusters*. Para cada *cluster* existe um ponto representativo que o resume, uma escolha comum desse ponto é a média. São adequados para encontrar clusters em formas convexas, mas têm dificuldade em encontrar os *clusters* não convexas, pois dois pontos de *clusters* diferentes podem estar mais próximos do que dois pontos no mesmo cluster;
- b) agrupamento baseado em densidade: os métodos baseados em densidade são adequados para encontrar os clusters não convexas;
- c) agrupamento hierárquico: o objetivo do agrupamento hierárquico é criar uma sequência de partições aninhadas, que podem ser convenientemente visualizadas por meio de uma árvore ou hierarquia. Os *clusters* na hierarquia variam do granulado fino ao granulado grosso - o nível mais baixo da árvore (as folhas) consiste em cada ponto em

---

<sup>19</sup> Entende-se que a melhor tradução para o termo seria "agrupamento", no sentido de grupo ou aglomeração. Para não confundir com o termo adotado para *clustering*, que também é agrupamento, mas no sentido do nome da ação de agrupar, optou-se por adotar o termo em língua inglesa.

seu próprio cluster, enquanto o nível mais alto (a raiz) consiste em todos os pontos em um grupo;

- d) agrupamento baseado em grafo: dado um grafo, o objetivo é agrupar os nós usando as arestas e seus pesos, que representam a similaridade entre os nós incidentes. O agrupamento de grafos está relacionado ao agrupamento hierárquico divisivo, pois muitos métodos particionam o conjunto de nós para obter os agrupamentos finais usando a matriz de similaridade pareada entre os nós.

### 1.1.1.6 Classificação

A tarefa de classificação é prever o rótulo ou classe para um determinado ponto não rotulado. Formalmente, um classificador é um modelo ou função  $M$  que prevê o rótulo da classe  $y$  para um dado exemplo de entrada  $x$ , ou seja,  $y = M(x)$ , onde  $y$  pertence ao conjunto de rótulos de classe (um valor de atributo categórico). Para construir o modelo, é necessário um conjunto de pontos com seus rótulos de classe corretos, esse é chamado de conjunto de treinamento. Depois de aprender o modelo  $M$ , pode-se prever automaticamente a classe para qualquer ponto novo. Existem vários modelos diferentes de classificação, cada um com outros tantos algoritmos de classificação. Alguns modelos de classificação mais utilizados são:

- a) Árvores de decisão: uma árvore de decisão particiona recursivamente o espaço de dados em regiões “puras” que contêm pontos de dados de apenas uma classe, com relativamente poucas exceções. Um de seus pontos fortes é que ele produz modelos que são mais fáceis de entender em comparação com outros métodos;
- b) Classificadores probabilísticos: um classificador muito utilizado e poderoso é o classificador de Bayes que se baseia no teorema de Bayes para prever classe como aquela que maximiza probabilidade  $P(c_i | x)$ , onde  $c_i$  é um dos rótulos do conjunto. A tarefa principal é estimar a função de densidade de probabilidade conjunta para cada classe. Essa função é modelada por meio de uma distribuição normal;
- c) Máquinas de vetores de suporte (SVM): é um dos classificadores mais eficazes para muitos domínios de problemas diferentes. Seu objetivo é encontrar o hiperplano ideal que maximize a margem entre as classes.

Máquinas de vetores de suporte podem ser usadas para encontrar limites não lineares, que, no entanto, correspondem a algum hiperplano linear em algum espaço “não linear” de alta dimensão.

### 3.2.3 Aprendizagem de máquina

Assim como na mineração de dados, aprendizado de máquina oferece métodos automatizados de análise de dados. De fato, as duas abordagens são muito parecidas e compartilham diversos métodos. Murphy (2012) define aprendizado de máquina como um conjunto de métodos para detectar automaticamente padrões nos dados e, em seguida, usar os padrões descobertos para prever dados futuros ou para realizar outros tipos de tomada de decisão sob incerteza. No aprendizado de máquina, a incerteza vem de várias formas: qual é a melhor previsão sobre o futuro com alguns dados anteriores? Qual é o melhor modelo para explicar alguns dados? Qual medição deve-se realizar a seguir?

O aprendizado de máquina geralmente é dividido em dois tipos principais: aprendizagem preditiva ou supervisionada e aprendizagem descritiva ou não-supervisionada.

Na abordagem de aprendizagem preditiva ou supervisionada, o objetivo é aprender um mapeamento das entradas às saídas, dado um conjunto rotulado de pares de entradas-saída. Quando os dados de saída são categóricos, o problema é conhecido como classificação ou reconhecimento de padrão, e quando são valores numéricos, o problema é conhecido como regressão.

Na abordagem de aprendizagem descritiva ou não supervisionada, o objetivo é encontrar “padrões interessantes” a partir apenas dos dados de entrada. Neste caso, o problema não é tão bem definido, uma vez que não há informação sobre quais tipos de padrões procurar. Também não há uma forma óbvia para avaliar o resultado como se dá no aprendizado supervisionado, pois neste último é possível comparar uma saída prevista para uma determinada entrada com o valor observado.

Informações mais detalhadas sobre fundamentos de aprendizagem de máquina podem ser obtidas em Murphy (2012) e em Bishop (2006). Goodfellow et al. (2016) aborda um tipo especial de aprendizagem de máquina conhecido por *deep learning* que trata de técnicas mais avançadas de aprendizagem.

### **3.3 Simulações computadorizadas**

Muitas vezes estudar um problema do mundo real pode se tornar uma tarefa extremamente difícil de se realizar devido à complexa relação entre um grande número de variáveis. Uma estratégia muito útil nesses casos, é simplificar o problema inicial de forma que a investigação se concentre nos aspectos mais relevantes ao objeto de estudo. Essa é a essência da modelagem: a definição de modelos na forma de uma representação simplificada da realidade.

A simulação computacional consiste de um programa de computador que executa regras de interação definidas por uma modelagem matemática projetada com a finalidade de prever o comportamento de algum aspecto de fenômenos do mundo real. Um tipo especial de simulação discreta que não depende de um modelo com uma equação subjacente, mas pode ser representada formalmente, é a simulação baseada em agente.

Na simulação baseada em agente, as entidades individuais (como moléculas, células, árvores ou consumidores) no modelo são representadas diretamente (ao invés de sua densidade ou concentração) e possuem um estado interno e um conjunto de comportamentos ou regras que determinam como o estado do agente é atualizado de uma etapa de tempo para a próxima.

#### **3.3.1 Sistemas Complexos**

Segundo Bar-Yam (2015), sistemas complexos são sistemas em que o comportamento coletivo não satisfaz o teorema central do limite, isto é, os componentes não são independentes nem totalmente dependentes. Sistemas complexos são compostos por um grande número de agentes heterogêneos que agem localmente, seguindo regras simples, sob racionalidade limitada, sem controle central e em constante adaptação e aprendizagem.

#### **3.3.2 Modelos baseados em agentes**

Nesta seção, descreve-se a modelagem baseada em agentes devido à sua adequação para simulação de sistemas complexos, facilidade de implementação e maturidade experimental.

A modelagem baseada em agentes (ABM)<sup>20</sup> é uma técnica utilizada para sistemas de simulação computacional e que vem se mostrando muito poderosa para simular o comportamento humano em sistemas complexos. Essa abordagem, passa pela definição de uma entidade autônoma de tomada de decisão, o agente, que é capaz de avaliar individualmente a sua situação para tomar decisões baseadas em um conjunto de regras, de forma a apresentar comportamentos apropriados para a função que representam. Um modelo baseado em agentes consiste de um sistema de agentes e dos relacionamentos entre eles. Para Bonabeau (2002), "ABM é mais uma mentalidade do que uma tecnologia [pois] consiste em descrever um sistema a partir da perspectiva de suas unidades constituintes."

Bonabeau afirma que ABM apresenta três características: captura de fenômenos emergentes, descrição natural de um sistema e flexibilidade. Fenômenos emergentes são os resultados das interações de entidades individuais, que, por definição, não podem ser reduzidos às partes do sistema. Um fenômeno emergente pode ser contra intuitivo, pois pode ter propriedades que são desacopladas das propriedades individuais. Por exemplo, um congestionamento, que resulta das interações entre os motoristas de veículos individuais, pode estar se movendo na direção oposta à dos carros que o causam. Logo, ABM é especialmente útil quando se deseja simular comportamentos com potencial para fenômenos emergentes, tais como:

- a) Comportamento individual não linear.
- b) Comportamento individual que exhibe memória, dependência do caminho e histerese<sup>21</sup>, comportamento não markoviano<sup>22</sup> ou correlações temporais, incluindo aprendizagem e adaptação.
- c) Interações heterogêneas e que podem gerar efeitos de rede.
- d) Sistema linearmente estável, mas instável a perturbações maiores.

---

<sup>20</sup> Muitas vezes conhecido pela abreviação ABM do termo em inglês *agent-based modeling*.

<sup>21</sup> Por definição: fenômeno apresentado por determinados sistemas físicos cujas propriedades dependem de sua história precedente. Um exemplo bem conhecido é o fenômeno de magnetização que ocorre após expor um objeto ferromagnético a um campo magnético.

<sup>22</sup> Propriedade de uma cadeia de Markov, em que a distribuição de probabilidade do próximo estado depende apenas do estado atual e não na sequência de eventos que precederam.

Por fornecer uma descrição natural de um sistema, ABM faz o modelo parecer mais próximo da realidade. Por exemplo, é mais natural descrever como os compradores se movem em um supermercado do que propor equações que governam a dinâmica da densidade de compradores. Ao se basear no comportamento individual das partes, ABM se contrapõe à modelagem por processos de negócio.

### **3.4 Métodos de análise**

#### **3.4.1 Análise de redes sociais**

Pode-se descrever uma rede como um sistema de nodos e elos; uma estrutura sem fronteiras; uma comunidade não geográfica; um sistema de apoio ou um sistema físico que se pareça com uma árvore ou uma rede. A rede social, derivando deste conceito, passa a representar um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados. O conceito de rede tem uma dupla aplicação: a utilização estática, que explora a rede estrutura para melhor compreender a sociedade ou um grupo social por sua estrutura, seus nós e suas ramificações; e a utilização dinâmica, que explicita a rede sistema, trabalhando as redes como uma estratégia de ação no nível pessoal ou grupal, para gerar instrumentos de mobilização de recursos (MARTELETO, 2001).

Wetherell, Plakans e Wellman (1994) descrevem uma rede social como uma estrutura social formada por uma rede de laços que conectam membros e canalizam recursos. A rede social enfoca as características dos laços e não as características dos membros individuais e vê as comunidades como redes de relações individuais que as pessoas fomentam, mantêm e usam em sua vida diária.

A análise de rede social (SNA) é uma estratégia para investigar estruturas sociais. Enquanto a teoria social individualista considera que os atores individuais fazem escolhas sem levar em consideração o comportamento dos outros, na SNA os relacionamentos entre os atores se tornam a primeira prioridade e as propriedades individuais são apenas secundárias. Os dados relacionais são o foco das investigações. Ressalta-se, entretanto, que as características individuais, bem como os vínculos relacionais, são necessárias para a compreensão plena dos fenômenos sociais.

Distinguem-se duas formas principais de SNA: a análise de rede do ego e a análise de rede global. Em estudos de 'ego', a rede de uma pessoa é analisada, enquanto nas análises de redes globais, tenta-se encontrar todas as relações entre os participantes da rede.

Otte e Rousseau (2002) mostram que a análise de redes sociais é um exemplo típico de uma ideia que pode ser aplicada em muitos campos. Com a teoria matemática dos grafos como base, ela se tornou uma abordagem multidisciplinar com aplicações em sociologia, ciências da informação, ciências da computação, geografia etc.

Malini (2016) propõe uma abordagem de análise de rede sociais digitais baseada em perspectivas, onde é realizada uma modelagem mais abrangente de perfis de usuários de forma a diminuir a influência dos perfis com grande popularidade, e com isso identificar pontos de vista como rastros de tempos com posição topológica estrutural num grafo. Essa abordagem permite a geração de dados mais interessantes para análise de fenômenos sociais no meio digital.

### **3.4.2 Processamento de Linguagem Natural**

Processamento de Linguagem Natural (NLP) é um conjunto de técnicas de Inteligência Artificial que tem por objetivo entender a linguagem dos seres humanos, ou seja, permitir que os computadores tenham a capacidade de entender e compor textos. Isso significa reconhecer o contexto, fazer análises sintáticas, semânticas, lexicais e morfológicas, criar resumos, extrair informações, interpretar os sentidos, analisar sentimentos e aprender conceitos com os textos processados.

Ainda que os computadores se baseiam no processamento de linguagens, as linguagens computacionais possuem uma característica fundamental que é a ausência de ambiguidade, o que não se observa nas linguagens humanas. Muitas vezes, uma simples mudança de entonação ou na expressão facial é suficiente para alterar drasticamente a interpretação de um mesmo texto.

#### **4 COMPUTATIONAL SOCIAL SCIENCE: APLICAÇÕES EM DEMOCRACIA DELIBERATIVA**

No capítulo 2, estabeleceu-se uma delimitação descritiva dos fundamentos da democracia deliberativa que apontam para a compreensão da prática democrática como um sistema em que atuam uma rede complexa de atores, com interconexões de influência entre as redes formais e informais de deliberação política e tomada de decisão pública. No capítulo 3, apresentou-se uma abordagem computacional de pesquisa científica para as Ciências Sociais e uma breve descrição de algumas técnicas de Ciência de Dados comumente aplicadas nessa abordagem.

Este capítulo tem como proposta pensar de forma crítica o uso das ferramentas de análise de dados que podem ser aplicadas aos diferentes contextos dos sistemas deliberativos. Mais além, os fundamentos da democracia deliberativa são tomados como um norte ético, podendo orientar o desenho de projetos aplicados à Ciência de Dados.

O objetivo principal dessa análise é ampliar a compreensão das práticas de Ciência de Dados como capazes de reconfigurar métodos de pesquisa e promoção da democracia deliberativa. Como já posto, a Ciência de Dados apresenta soluções técnicas para questões sociais em contextos sociais complexos. Portanto, é preciso uma aproximação com o campo da democracia deliberativa especificamente o debate ampliado e sistêmico sobre a formação da opinião pública e entendimento de que forma ocorrem as deliberações políticas.

A democracia deliberativa parte da premissa de que as deliberações no âmbito da esfera pública e espaços públicos afetam a agenda das políticas públicas, orientam e legitimam as ações dos governos. Porém, mesmo toda a rede estruturada das instituições formais e informais de participação não são capazes de fazer emergir a vontade pública por completo, pois ainda excluem uma parcela significativa dos cidadãos com interesses nas deliberações.

Interações discursivas que se dão através das redes sociais e das redes de interação digital agem como forças atuando na formulação de decisões, ainda que não se perceba uma relação direta com o processo de construção de políticas públicas e que essa relação seja contingencial, como problematiza Habermas (1997, v I) e é mais presente em democracias consolidadas e avançadas.

O crescimento do número de atores deliberando fora das instituições formais de participação combinado a um fluxo intenso, fragmentado e descentralizado

da informação e interação nas redes sociais geram uma maior complexidade do fenômeno da formação da opinião pública. No contexto da democracia digital, por meio de processos de deliberação digitais espontâneas, impulsionados por atores não governamentais e também por iniciativas de participação e deliberação digital promovidas por instituições formais de participação, esse crescimento é ainda maior.

Toda essa interação digital cria uma nuvem de dados informacionais que são base para descrição do próprio fenômeno de formação da opinião pública nestes espaços virtuais. É possível visualizar essas interações na sua estrutura de dados, ora como dados informacionais organizados e ora como pegadas digitais.

Uma compreensão aprofundada dos fenômenos sociais em larga escala é tarefa extremamente complexa e exige métodos científicos capazes de lidar com esse nível de complexidade.

Dessa forma, é possível uma aproximação da computação com a abordagem sistêmica da democracia deliberativa. Não se trata de mera instrumentalização de recursos computacionais aplicados aos sistemas deliberativos. Dada a complexidade desses sistemas, o meio digital se mostra como um novo ambiente para a sua realização, pois a ausência de barreiras físicas e espaciais torna possível a interação entre atores independentemente de estarem inseridos nos níveis micro ou macro de instâncias deliberativas.

Mas a simples digitalização das deliberações pode levar a um efeito colateral adverso - uma inundação de dados. Para lidar com esse enorme volume de dados complexos, propõe-se lançar mão das técnicas de *Computational Social Science*. Para tanto, identifica-se duas abordagens complementares:

- a) Abordagem operacional, que visa o aperfeiçoamento da análise discursiva dos processos deliberativos que permita lidar com a complexidade de um debate ampliado a níveis impossíveis de serem analisados pelos métodos tradicionais e dessa forma promover a melhoria na eficiência das tomadas de decisões;
- b) Abordagem estrutural, que visa modelar os sistemas deliberativos como um sistema complexo o que permitiria estudar o processo deliberativo como um todo, possibilitando realizar previsões e simulações que permitiriam identificar, propor e testar pontos de melhorias nos processos deliberativos.

#### 4.1 Técnicas para abordagem operacional

Segundo Howard (2001), os governos em todo o mundo estão cada vez mais utilizando plataformas *on-line* e mídias sociais para se envolver e determinar as opiniões de seus cidadãos. Lawrence et al (2017) analisaram um debate, promovido pelo Departamento de Transporte dos Estados Unidos numa plataforma de deliberação *on-line*. O *corpus* do estudo era composto de comentários dos usuários, nos quais foram identificadas 23.682 palavras, 1.657 segmentos (ou seja, unidades de discurso que constituem componentes de redes de argumento), 209 turnos (ou seja, comentários de usuários trocados durante o diálogo) e 70 mapas, ou seja, visualizações de redes de argumento, cada um correspondendo a uma discussão de troca de usuários. Em debates deliberativos *on-line* como esse, os tomadores de decisão geralmente não têm tempo para processar os dados resultantes desse engajamento e acabam contando apenas com um resumo superficial dos pontos levantados e a força da opinião de cada lado de uma questão. Com isso informações preciosas podem acabar se perdendo.

O desafio é identificar questões fundamentais para o debate em um contexto deliberativo: questões controversas, questões que foram resolvidas com sucesso e questões que devem ser tratadas posteriormente devido à falta de consenso e de compreensão mútua. Encontrar e classificar essas questões são tarefas árduas e demoradas em discussões da vida real. Com a utilização de técnicas de mineração de dados é possível automatizar esse trabalho, cuja execução manual e em larga escala se torna virtualmente impossível na prática.

Lawrence et al (2017) discutem duas abordagens para automatizar a análise de debates: a mineração de controvérsias e a mineração de argumentos.

A mineração de controvérsia examina os processos de como um problema ou um evento atrai opiniões conflitantes em redes dinâmicas e dialógicas. Embora a mineração de controvérsias forneça uma indicação de onde ocorrem as questões controversas, ela não oferece o entendimento mais rico obtido pelo estudo da estrutura argumentativa.

A mineração de argumentos visa desenvolver métodos e técnicas de extração automática de argumentos de textos em linguagem natural. Um argumento é uma unidade de discurso complexa com limites facilmente reconhecíveis por humanos, mas difíceis de determinar por um computador. Por essa razão, a

mineração de argumentos é frequentemente apoiada por estruturas retóricas de documentos, esquemas de argumentos ou relações dialógicas.

Para realizar essas tarefas pode-se utilizar técnicas de processamento de linguagem natural com o objetivo de extrair a estrutura léxica e sintática de um diálogo. Técnicas de análise de sentimento são aplicadas para identificar e dimensionar o grau de concordância, discordância ou neutralidade em relação às questões controversas. Tarefas de mineração de dados são realizadas para descobrir padrões nos diálogos conhecidos como indicadores de discurso, que são palavras que servem de pista para a função argumentativa da proposição. Por fim, é realizada a etapa de análise computacional do discurso por meio da aprendizagem de máquina. Nesta etapa, uma parte dos dados é rotulada manualmente para treinar os algoritmos de classificação. O objetivo é obter um modelo de classificação automático capaz de inferir rótulos para uma base de dados não treinados e com esse modelo classificar os relacionamentos entre as proposições.

Uma abordagem como essa permite automatizar o entendimento aprofundado das questões e com isso oferecer valor real aos tomadores de decisão e aos responsáveis pelas políticas públicas na interpretação e obtenção de *insights* em debates complexos em grande escala. Essa análise automatizada do discurso pode ajudar a lidar com a complexidade dos julgamentos políticos em um processo deliberativo. Ao mesmo tempo permite a inclusão de um número maior de atores excluídos destes processos.

## 4.2 Técnicas para abordagem estrutural

A metodologia de análise de redes sociais pode ser utilizada para identificar os fluxos de informação e as construções sociais e simbólicas em redes de movimentos sociais (MARTELETO, 2001). De forma análoga, essa metodologia pode ser útil para identificar fluxos de informações nas redes sociais digitais. No lugar de agrupamentos como os movimentos sociais, no ambiente digital, formam-se comunidades virtuais deliberativas cujos agentes interagem internamente entre si e externamente uns com os outros.

No contexto da análise de redes sociais digitais, Malini (2016) se apropria das pegadas digitais para propor um método perspectivista de análise de rede como uma abordagem que "se dedica a estudar como as relações de compartilhamentos,

respostas, inscrições, comentários, favoritadas, curtidas, indicações [...] formam rastros sociais que expressam, conjuntamente, pontos de vistas coletivos” (MALINI, 2016, p. 2), A partir dessa abordagem é possível identificar “pistas para a construção de objetos de pesquisas a partir da descoberta do caráter relacional dos dados digitais” (MALINI, 2016, p. 3) que não sejam meramente dominadas pela simples popularidade dos influenciadores digitais mas valorizadas pela “densidade de relações altamente conectadas entre os perfis que fazem emergir ricos pontos de vistas coletivos capazes de influir nos sentidos dos acontecimentos sociais” (MALINI, 2016, p. 5). Dessa forma, a análise perspectivista de rede permite acompanhar a evolução da construção da opinião nas redes sociais digitais.

Uma abordagem que visa a automação sistematizada desse método pode permitir ampliar o horizonte da pesquisa. Enquanto na análise perspectivista o foco está no estudo da construção de pontos de vista, a automação sistemática desse método permitiria ampliar o horizonte da pesquisa para o estudo da interação entre pontos de vista. Essa pode ser a base metodológica para estabelecer o processo comunicativo numa modelagem computacional de sistemas deliberativos.

Da análise de redes tradicionais é possível o estudo dos sistemas deliberativos em diversos níveis de granularidade, ou seja, possibilitando que sejam estudadas as interações desde os níveis das comunidades locais, passando pelos níveis regionais, nacionais e alcançando níveis transnacionais. Modelos baseados em agentes podem ser utilizados para compreender os comportamentos individuais e o efeito deste em relação ao grupo.

A cada nível de interação, grandes volumes de dados podem ser gerados. Analisar esses dados manualmente é uma tarefa extenuante e demorada, podendo levar a erros que diminuem a confiabilidade dos resultados. Para lidar com esse problema, aplica-se técnicas de *big data* e mineração de dados. Os processos podem ser automatizados a partir de predições obtidas pelas técnicas de aprendizagem de máquina.

## 5 CONSIDERAÇÕES FINAIS

Nesta monografia buscou-se mostrar que a Ciência da Computação, por meio da obtenção e análise dos rastros que as pessoas produzem no meio digital, pode se tornar uma poderosa aliada na geração de conhecimento relativos à sociedade, às práticas sociais e aos sistemas deliberativos. Esse conhecimento, aplicado às pesquisas dos cientistas sociais, pode elevar a compreensão das relações sociais a níveis nunca antes vistos. Por outro lado, tende também a ser utilizada para fins alheios aos interesses públicos. Esse dualismo remete à reflexão sobre como usar essas informações: devem ser disponibilizadas ou ocultadas? Preservadas ou descartadas? Transformadas ou preservadas na forma original? Devem ser considerados nas tomadas de decisões dos governos? Cada uma dessas perguntas e outras mais estão associadas a dilemas éticos e resultam em algum compromisso.

Pode-se considerar que o melhor, talvez, seja ignorar essas questões, mas pelo menos no nível estatal, essa não é uma opção aceitável. Pois os dados, as pegadas digitais já existem e continuam a ser produzidos a cada instante. Esse tipo de informação pode ser obtido por qualquer pessoa com relativa facilidade e, com o uso de técnicas e ferramentas adequadas, podem dizer muito sobre cada indivíduo e também sobre a própria sociedade.

Para que o Estado cumpra o seu objetivo perante a sociedade é preciso compreender essa sociedade da melhor forma possível. Se o Estado não o fizer, outra instituição ou mesmo outro Estado poderá fazê-lo e as consequências podem não ser as mais desejáveis.

Aqui, defende-se que a partir de critérios estabelecidos, o Estado pode apropriar-se das informações geradas pela sociedade e usá-las em benefício dessa sociedade, guardados os limites e os devidos direitos e garantias relativas à proteção da privacidade e de dados pessoais, neutralidade da rede e outros pontos que inclusive constituem objeto de regulamentação da legislação federal.

Dentre esses possíveis benefícios, sugere-se a aplicação no aprimoramento da democracia por meio do desenvolvimento de mecanismos de aproximação entre o Estado e o cidadão.

Ao se apresentar como uma concepção ética de manifestação da vontade popular, as abordagens da democracia deliberativa podem, potencialmente, beneficiar-se da leitura que a computação permite realizar por meio da subárea da

Ciência de Dados aplicada sobre as pegadas digitais. Esse potencial é enorme e pode contribuir tanto para melhorar os processos deliberativos, ao disponibilizar ferramentas poderosas aos pesquisadores da área, quanto para melhorar a qualidade das deliberações, ao propor mecanismos para a análise profunda de debates em larga escala. E, por fim, contribuir para uma melhoria na capacidade de tomada de decisão fundamentada na real necessidade ou vontade da sociedade afetada por essa decisão.

Esta monografia buscou demonstrar que os processos deliberativos podem ser analisados sob a perspectiva dos sistemas complexos, que se mostram aderentes com os métodos da *Computational Social Science*. Dessa forma, apresenta-se uma abordagem possível para pesquisa empírica de larga escala no campo da democracia deliberativa.

Para tanto, é preciso identificar na prática empírica da deliberação possibilidades de automação, utilizando os métodos e as ferramentas de Ciência de Dados de forma a garantir a inclusão e os princípios éticos que a democracia deliberativa apregoa. Isso é fundamental, pois como visto no capítulo 3, a Ciência de Dados pode ser excludente e atender a práticas antiéticas.

Esta monografia limitou-se a realizar uma aproximação entre duas áreas muito distintas: a democracia deliberativa e a Ciência da Computação. Pela democracia deliberativa assumiu-se a abordagem sistêmica pelo seu potencial de gerar objetos de estudo na forma de sistemas complexos. Pelo lado da Ciência da Computação, assumiu-se que a Ciência de Dados apresenta um conjunto de métodos e ferramentas adequadas para a realização das análises desses objetos de estudo.

A evolução prática deste trabalho passaria pela formação de equipes interdisciplinares provenientes das áreas da Ciência Política, Ciência da Computação e estatística que podem adotar uma abordagem de pesquisa *botton-up* e devem focar no aprimoramento de técnicas para abordagem operacional (seção 4.1) do estudo dos sistemas deliberativos. Por outro lado, equipes que adotarem uma abordagem de pesquisa *top-down* devem focar no aprimoramento de técnicas para abordagem estrutural (seção 4.2) de pesquisa dos sistemas deliberativos.

Por fim, esta monografia cumpre o objetivo de apresentar o campo empírico de prática democrática aos cientistas de dados, fazendo uma aproximação entre duas áreas que têm estado muito distantes. É importante perceber que a automação deverá ocorrer num futuro próximo, à medida que mais e melhores dados abertos se tornem

disponíveis. Portanto, é importante que essa automação se direcione a partir dos princípios éticos da democracia deliberativa, tendo em vista a complexidade dos sistemas deliberativos e as dificuldades e desafios que os pesquisadores, cidadãos e tomadores de decisão têm quando consideram esse sistema como complexo.

Por outro lado, não foi possível demonstrar, a partir da pesquisa realizada, que já existem resultados publicados mostrando uma contribuição efetiva da Ciência de Dados com os sistemas deliberativos. O que se encontrou foram estudos que utilizam essas ferramentas e métodos para análise dos dados, porém sem apresentar conexões dessas análises com o processo decisório.

Outra limitação foi a ausência de discussão relativa à regulamentação do uso de dados, o que se justifica, de um lado, pelo recorte adotado no presente trabalho e, de outro lado, pela abrangência, complexidade e implicações do tema da regulação, que requer uma abordagem específica e aprofundada como objeto principal de pesquisa. Nesse sentido, a limitação referida aponta para questões necessárias e importantes a serem contempladas em outras frentes de estudos.

A legislação básica pertinente é definida principalmente pela Lei nº 12.965, de 23 de abril de 2014, conhecida como Marco Civil da Internet e pela Lei nº 13.709, de 14 de agosto de 2018 que é a Lei Geral de Proteção de Dados Pessoais (LGPD). O Marco Civil da Internet estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. O texto da lei regulamenta pontos essenciais como garantia da liberdade de expressão, proteção da privacidade, proteção dos dados pessoais, neutralidade da rede, responsabilização dos agentes, dentre outros. A LGPD dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, tendo como fundamentos o respeito à privacidade, à autodeterminação informativa, à liberdade de expressão, à inviolabilidade da intimidade, ao desenvolvimento econômico e tecnológico e a inovação, à livre iniciativa, à livre concorrência e a defesa do consumidor e os direitos humanos, ao livre desenvolvimento da personalidade, à dignidade e ao exercício da cidadania. Para além dos marcos legais mencionados, pode-se ainda supor que o próprio avanço tecnológico pode implicar a ampliação dos debates e demandar novos instrumentos regulatórios ou reformulações do aparato legal existente.

A partir das discussões efetuadas na presente monografia, mostra-se necessário formar equipes multidisciplinares que promovam a inter colaboração entre cientistas de dados e cientistas políticos. Essa demanda poderá levar, inclusive, ao

aprimoramento dos currículos acadêmicos dessas áreas, em que tanto os cientistas políticos aprendam a aplicar o potencial da Ciência de Dados em suas pesquisas, quanto os cientistas de dados sejam sensibilizados quanto às idiossincrasias das Ciências Sociais e em especial a Ciência Política de forma a promover o desenvolvimento de novas tecnologias que considerem essas características tão específicas e tão diferentes daquelas pautadas pelos resultados determinísticos aos quais estão acostumados enquanto pesquisadores de Ciências Exatas.

Durante a realização desta monografia, constatou-se que já existe uma grande quantidade de dados que são utilizados pelos pesquisadores na linha da democracia deliberativa, seja na área de Ciência Política, seja em suas interfaces com a administração pública e outras áreas. Mostrou-se que os desenvolvimentos teóricos recentes no campo dos sistemas deliberativos apontam para a complexidade, incluindo fóruns e mini públicos, dentre outros espaços deliberativos, de modo que as pesquisas, podem estender seu foco de abordagem inclusive no que se refere à escala, além dos próprios governos nacionais. Essa ampla graduação de escalas passou a ter uma grande relevância nas decisões de políticas públicas em virtude das mídias digitais que geram um grande volume de dados e que, por sua vez, não podem ser analisados com métodos de análise tradicionais.

Outra percepção possível ao longo desse trabalho é que existe um desconhecimento do potencial das ferramentas e métodos da Ciência de Dados. Os pesquisadores consultados têm tentado uma aproximação, mas ainda não se concebe um uso das ferramentas de Ciência de Dados para atender a demanda por informação e análise das deliberações capaz de percorrer todo o processo deliberativo, desde a formação da opinião nas esferas públicas até a tomada de decisão nas esferas institucionais.

Dessa feita, pode-se considerar que qualquer governo ou entidade política que realmente quiser melhorar o sistema deliberativo no qual está inserido e obter melhores resultados nas suas deliberações, precisa realizar análises complexas das informações de forma a obter um entendimento mais aprofundado dos processos por meio das ferramentas de dados.

## REFERÊNCIAS

- ALVES, Priscila Mello. Inteligência Artificial e Redes Neurais  
<https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/106-inteligencia-artificial-e-redes-neurais>
- AVRITZER, Leonardo. A moralidade da democracia: ensaios em teoria habermasiana e teoria democrática. São Paulo: Perspectiva, 1996.
- BAR-YAM, Yaneer. Aplicações de sistemas complexos em políticas públicas no mundo. In: FURTADO, Bernardo Alves; SAKOWSKI, Patrícia AM; TÓVOLLI, Marina H. Modelagem de sistemas complexos para políticas públicas. 2015. p. 267 – 290.
- BISHOP, Christopher M. Pattern recognition and machine learning. Springer, 2006.
- BLANCO, Víctor Sampedro; DE LA FUENTE, Jorge Resina. Opinión pública y democracia deliberativa en la Sociedad Red. Ayer, p. 139-162, 2010.
- BRASIL. LEI Nº 12.965, DE 23 DE ABRIL DE 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil, Brasília,DF, abr 2014. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/l12965.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm)>. Acesso em: 27 out. 2020.
- BRASIL. LEI Nº 13.709, DE 14 DE AGOSTO DE 2018. Lei Geral de Proteção de Dados Pessoais (LGPD), Brasília,DF, abr 2014. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm)>. Acesso em: 27 out. 2020.
- CASTELLS, Manuel. A Sociedade em Rede. Rio de Janeiro: Paz e Terra, 2013.
- CHUNARA, Rumi; ANDREWS, Jason R.; BROWNSTEIN, John S. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. The American journal of tropical medicine and hygiene, v. 86, n. 1, p. 39-45, 2012.
- COHEN, Joshua. Deliberación y legitimidad democrática. Cuaderno gris, 2007, 9: 127-146.
- DAHLBERG, Lincoln. The Internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. Information, communication & Society, v. 4, n. 4, p. 615-633, 2001.
- FARIA, Cláudia Feres. O que há de radical na teoria democrática contemporânea: análise do debate entre ativistas e deliberativos. Revista Brasileira de Ciências Sociais, 2010, 25.73: 101-111.
- FURTADO, Bernardo Alves; SAKOWSKI, Patrícia AM; TÓVOLLI, Marina H. Modelagem de sistemas complexos para políticas públicas. 2015.
- GINSBERG, Jeremy et al. Detecting influenza epidemics using search engine query data. Nature, v. 457, n. 7232, p. 1012-1014, 2009.

GOODFELLOW, Ian et al. Deep learning. Cambridge: MIT press, 2016.

HABERMAS, Jürgen. DIREITO E DEMOCRACIA: Entre facticidade e validade. Tradução: Flávio Beno Siebeneichler. Rio de Janeiro: Tempo Brasileiro, 1997. V. 1

HABERMAS, Jürgen. DIREITO E DEMOCRACIA: Entre facticidade e validade. Tradução: Flávio Beno Siebeneichler. Rio de Janeiro: Tempo Brasileiro, 1997. V. 2

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. Data mining: concepts and techniques. Elsevier, 2011.

HILBERT, Martin. e-Science for Digital Development: ICT4ICT4D. Development Informatics Working Paper Series [http://www.seed.manchester.ac.uk/medialibrary/IDPM/working\\_papers/di/di-wp60.pdf](http://www.seed.manchester.ac.uk/medialibrary/IDPM/working_papers/di/di-wp60.pdf), v. 60, 2015.

HILBERT, Martin. Big data for development: A review of promises and challenges. Development Policy Review, v. 34, n. 1, p. 135-174, 2016.

HOWARD, Mark et al. E-government across the globe: how will e'change government. e-Government, v. 90, p. 80, 2001.

LANEY, Doug. 3D data management: Controlling data volume, velocity and variety. META group research note, v. 6, n. 70, p. 1, 2001.

LAWRENCE, John et al. Using argumentative structure to interpret debates in online deliberative democracy and erulemaking. ACM Transactions on Internet Technology (TOIT), v. 17, n. 3, p. 1-22, 2017.

LAZER, David et al. Social science. Computational social science. Science (New York, NY), v. 323, n. 5915, p. 721-723, 2009.

LAZER, David et al. The parable of Google Flu: traps in big data analysis. Science, v. 343, n. 6176, p. 1203-1205, 2014.

LEAL, Rogério Gesta, et al. A democracia deliberativa como nova matriz de gestão pública: alguns estudos de casos. 2011.

MACEDO, Augusto Q.; MARINHO, Leandro B.; SANTOS, Rodrygo LT. Context-aware event recommendation in event-based social networks. In: Proceedings of the 9th ACM Conference on Recommender Systems. 2015. p. 123-130.

MACINTOSH, A; WHYTE, A. Towards an evaluation framework for eParticipation. In: Transforming Government: People, Process & Policy. 2008. p. 16-30.

MAIA, Rousiley. Democracia e a internet como esfera pública virtual: aproximando as condições do discurso e da deliberação. Consultado a, 2001.

MALDONADO, Manuel Arias. La política en la teoría deliberativa. Notas sobre deliberación, decisión y conflicto. Revista Española de Ciencia Política, 2007.

- MALINI, Fábio. Um método perspectivista de análise de redes sociais: cartografando topologias e temporalidades em rede. XXV Encontro Anual da Compós, p. 1-30, 2016.
- MANSBRIDGE, Jane et al. A systemic approach to deliberative democracy. In: PARKINSON, John; MANSBRIDGE, Jane (Org.). *Deliberative systems: Deliberative democracy at the large scale*. Cambridge University Press, 2012. p. 1 – 26
- MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Fundamentos de Metodologia Científica*. 5. ed. São Paulo: Atlas, 2003.
- MARTELETO, Regina Maria. Análise de redes sociais-aplicação nos estudos de transferência da informação. *Ciência da informação*, v. 30, n. 1, p. 71-81, 2001.
- MENDONÇA, Ricardo Fabrino. Vídeo (45 min). Mitigating systemic dangers: the role of connectivity inducers in a deliberative system. Publicado pelo canal IESP-UERJ: [https://www.youtube.com/watch?v=OL\\_Y86rH-Xs](https://www.youtube.com/watch?v=OL_Y86rH-Xs) . 2016.
- MORGADO, Priscila Zanandrez Martins. *Sistemas Participativos: O que dizem essas experiências? III Encontro Internacional Participação, Democracia e Políticas Públicas*. Vitória, 2017
- MOCANU, Delia et al. The twitter of babel: Mapping world languages through microblogging platforms. *PloS one*, v. 8, n. 4, p. e61981, 2013.
- MURPHY, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- OTTE, Evelien; ROUSSEAU, Ronald. Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, v. 28, n. 6, p. 441-453, 2002.
- PORTO, Fábio; ZIVIANI, Artur. *Ciência de dados. III Seminário de Grandes Desafios da Computação no Brasil*, Rio de Janeiro, RJ, 2014.
- PROVOST, Foster; FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. *Big data*, v. 1, n. 1, p. 51-59, 2013.
- SALTZ, Jeffrey S.; STANTON, Jeffrey M. *An introduction to data science*. Sage Publications, 2017.
- SANTOS, Boaventura de Sousa; AVRITZER, Leonardo. Para ampliar o cânone democrático. In: SANTOS, Boaventura de Sousa (Org.). *Democratizar a democracia: os caminhos da democracia participativa*. Rio de Janeiro: Civilização Brasileira, 2003.
- SERBIN, Andrés. Effective regional networks and partnerships. *People Building Peace*, v. 2, p. 45-58, 2005.
- SILVA, Eduardo Moreira da. Sistemas deliberativos e análise de redes: similitudes e diferenças nas pesquisas sobre instituições participativas. In: IX Encontro da Associação Brasileira de Ciência Política. Brasília, 2014.

VAN DER AALST, Wil MP. Data Scientist: The Engineer of the Future. In K. Mertins, F. Benaben, R. Poler, and J. Bourrieres, editors, Proceedings of the I-ESA Conference, volume 7 of Enterprise Interoperability, pages 13–28. Springer, Berlin, 2014.

VAN DER AALST, Wil MP. Data science in action. In: Process mining. Springer, Berlin, Heidelberg, 2016. p. 3-23.

VIEIRA, Mónica Brito; SILVA, Filipe Carreira da. Democracia deliberativa hoje: desafios e perspectivas. Revista Brasileira de Ciência Política, 2013, 10: 151-194.

WEAVER, Warren. Science and Complexity. American Scientist, [s.l.], n. 36, p. 536–544, 1948.

WETHERELL, Charles; PLAKANS, Andrejs; WELLMAN, Barry. Social networks, kinship, and community in Eastern Europe. The Journal of Interdisciplinary History, v. 24, n. 4, p. 639-663, 1994.