

Reflexões sobre os percursos metodológicos em ciência social computacional

Reflections on methodological approaches in computational social science

Reflexiones sobre los caminos metodológicos en la ciencia social computacional

Recebido em 12-09-2021
Modificado em 12-11-2021
Aceito para publicação 25-11-2021

 <https://doi.org/10.47456/simbitica.v8i4.37345>

17

Ronaldo Baltar

Doutor em Sociologia (USP,1996), Professor do Departamento de Ciências Sociais da Universidade Estadual de Londrina, Coordenador do Programa Infosoc de Formação Complementar em Ciência Social Computacional – UEL. E-mail: baltar@uel.br

Cláudia Siqueira Baltar

Doutora em Demografia (Unicamp, 2008), Professora do Departamento de Ciências Sociais da Universidade Estadual de Londrina, Coordenadora do Observatório de Populações e Política Públicas (ObPPP) – UEL. E-mail: cbaltar@uel.br

Resumo

O objetivo do artigo é refletir sobre os desafios e a importância para as ciências sociais em se envolverem de forma mais próxima com os métodos da ciência social computacional. O principal argumento é que o diálogo interdisciplinar, propiciado pelos métodos computacionais aplicados aos estudos das questões sociais, pode trazer contribuições importantes para as ciências sociais, sobretudo em virtude da contínua e crescente transformação das fontes e dados de pesquisa qualitativa ou quantitativa em registros digitais. O artigo mostra uma visão sobre como se constitui o campo da ciência social computacional e aborda alguns obstáculos epistemológicos que dificultam a introdução dessa formação nas ciências sociais, sobretudo em língua portuguesa.

Palavras-chave: ciência social computacional; simulação baseada em agentes; análise de redes sociais; ciência dos dados; ciências sociais.

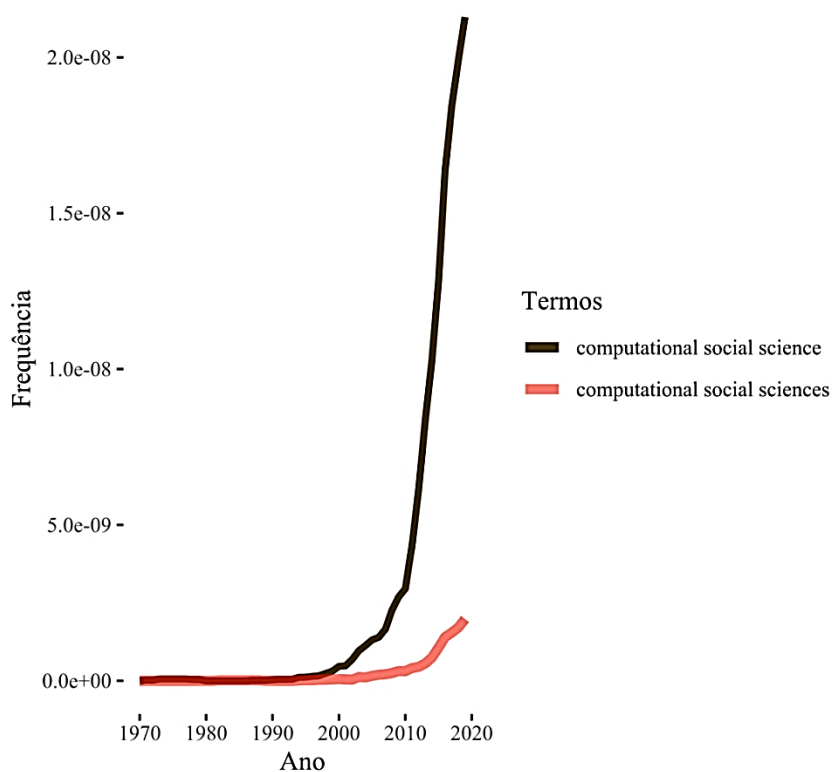


Introdução

Ciência social computacional é um novo campo do conhecimento e, ao mesmo tempo, um espaço interdisciplinar de convergência entre várias áreas do conhecimento científico. O número de publicações tem crescido rapidamente, desde 2003, em várias áreas do conhecimento, como sociologia, psicologia social, administração de empresas, educação, ciência política, antropologia entre outras (Edelmann *et al.*, 2020:64).

Conforme pode-se observar no Gráfico 1, a partir dos dados extraídos das publicações indexadas pelo Google Books, em língua inglesa (Google Books, 2021), há um crescimento acentuado no número de ocorrência do termo “ciência social computacional”, a partir do ano 2000, com um incremento ainda maior de 2010 em diante.

Gráfico 1 - Frequência comparativa de ocorrência dos termos "computational social science" e "computacional social sciences" no corpus de língua inglesa da base Google Books entre os anos de 1920 e 2019

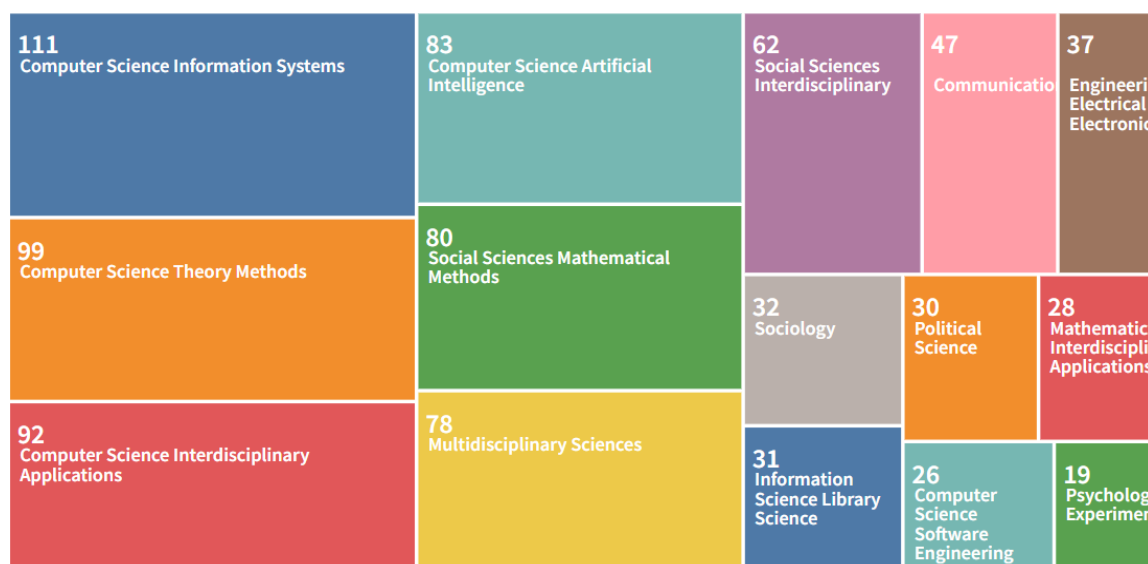


Fonte: Google Ngram, corpus “English 2019”. Elaboração própria. Nota: a busca foi feita a partir de 1920, mas as referências aos termos só aparecem a partir da década de 1970.

O termo utilizado de forma mais comum, em inglês, está no singular: ciência social computacional. Há também um crescimento do termo no plural – ciências sociais computacionais –, mas bem menos significativo. Permanece um debate subjacente ao uso do termo no singular ou no plural, sendo o primeiro a afirmação de um campo único, que emerge

como uma nova ciência, e o segundo, menos enfático nesse ponto, apontando para uma coleção de métodos que se somaria ao curso das disciplinas já estabelecidas (Chen; Yu, 2018). Essa não é uma discussão muito clara e não é o ponto central do debate sobre as fronteiras entre as ciências sociais e a ciência da computação. O ponto principal a ser destacado é que, utilizado no singular ou no plural, a essência da ciência social computacional é a interdisciplinaridade (Conte *et al.*, 2013). Nesse texto, utilizaremos o termo no singular para acompanhar a tradução das publicações em língua inglesa.

Gráfico 2 - Manifesto of Computational Science: número de citações pelas 15 áreas do conhecimento com mais referências entre 2012 e 2021



Fonte: Clarivate Web of Science.

O caráter interdisciplinar da ciência social computacional pode ser observado no Gráfico 2. Tomando-se como referência a publicação, em 2012, de pesquisadores europeus e dos Estados Unidos, liderados por Rosaria Conte e Nigel Gilbert, o *Manifesto of computational social science* (Conte *et al.*, 2012), de acordo com os dados da base de indexação de periódicos Web of Science (WoS); (Clarivate Analytics, 2021), entre as 15 áreas com maiores citações do artigo, estavam a ciência da computação (vários subcampos), ciências multidisciplinares, ciências sociais (sociologia, ciência política e outros subcampos), ciência da informação (o que inclui biblioteconomia), engenharia elétrica, matemática aplicada, comunicação e psicologia experimental. A área de antropologia não aparece no Gráfico 2, mas está entre as 20 áreas mais citadas, de acordo com os dados da base WoS.

Vale ressaltar que não há resultados expressivos em língua portuguesa. No entanto, uma consulta à base do Scielo (“SciELO - Brasil”, 2021) para “ciência social computacional”, em julho de 2021, retornou cinco referências e uma delas, fazia menção a “ciências sociais computacionais”. A consulta, no mesmo período, feita na base WoS, indicava apenas um

resultado em português, sendo a tradução do Manifesto de Rosaria Conte, Nigel Gilbert e demais coautores. No Portal de Periódicos da Capes (“Portal periódicos. CAPES”, 2021), para a mesma data, a consulta indicava sete publicações com o termo no singular e três, no plural, sendo que destes, dois apareciam também entre as sete referências da pesquisa “ciência social computacional”. Os resultados do Portal de Periódicos da Capes englobam os resultados da base Scielo.

Como se pode observar pelos dados bibliométricos, a ciência social computacional tem crescido a um ritmo bastante acelerado em publicações de língua inglesa. Segundo os dados da base WoS, até julho de 2021, havia 645 artigos com o termo “computational social science”, em algum dos campos (tópico, título ou resumo). Esses artigos foram citados por outros 9.410 artigos, em um total de 10.919 citações (sem contar as autocitações). Em 1999, há uma publicação e zero citações. Em 2010, foram 6 publicações e 70 citações, que aumentaram para 101 publicações e 1.104 citações, em 2017. No ano de 2021, ainda em curso, foram 86 publicações de artigos e 1.685 citações. As publicações e citações crescem em uma curva semelhante à frequência de termos extraídos da base Google Books, mostrada no Gráfico 1.

A expansão da ciência social computacional, em língua inglesa, apresenta desafios epistemológicos e oportunidades metodológicas que as ciências sociais, em língua portuguesa, não deveriam ignorar. O presente artigo tem por objetivo apresentar alguns argumentos que possam clarear a compreensão sobre o que é a ciência social computacional e qual a sua importância para as ciências sociais, em língua portuguesa, de se aproximarem mais desse movimento, que tem crescido em centros de pesquisa e universidades norte-americanas, europeias e asiáticas. Não se trata apenas de uso de *software* para pesquisa. Também não se trata de substituir o conhecimento teórico e metodológico das ciências sociais por dados ou procedimentos computacionais. O argumento principal é que o diálogo interdisciplinar, propiciado pelos métodos computacionais aplicados aos estudos das questões sociais, pode trazer contribuições importantes para as ciências sociais, sobretudo diante da crescente transformação das fontes e dados de pesquisa qualitativa ou quantitativa em registros digitais.

O artigo está dividido em quatro partes. Primeiro, será apresentada e discutida a concepção de que a ciência social computacional não é um substituto para campos já consolidados das ciências sociais. Trata-se de uma área necessariamente interdisciplinar, que tem um histórico de formação em distintos campos do conhecimento. Para se compreender melhor o entrelaçamento dessas diferentes áreas, a ciência social computacional será apresentada como um conjunto de possíveis caminhos que tanto podem ser integrados, como podem ser percorridos de forma autônoma.

Assim, áreas de simulação social, análise de redes sociais, estatística computacional, ciência dos dados, humanidades digitais, inteligência artificial, entre outras, formam itinerários ou percursos formativos para a ciência social computacional. Dessa forma, elucida-se que a ciência social computacional não é uma área a mais, nem uma unificação de todas essas áreas em uma nova ciência, mas, pode ser entendida como a combinação de qualquer uma dessas trilhas para a análise de fenômenos sociais a partir de dados digitais.

Em seguida, serão apresentados alguns obstáculos epistemológicos que precisam ser desfeitos para que as ciências sociais, principalmente nos centros de pesquisa e formação, em língua portuguesa, possam, efetivamente, se aproximar e contribuir com o aprimoramento desses percursos metodológicos computacionais aplicados às questões sociais. O principal obstáculo epistemológico é o espectro do positivismo que ronda as ciências sociais, espalhando desconfiança com quaisquer aproximações com métodos computacionais ou quantitativos. A parte seguinte irá discutir as possibilidades epistemológicas que a ciência social computacional pode trazer para as diferentes áreas de estudos das ciências sociais. Assim, desfazendo essas dissonâncias epistemológicas, a ciência social computacional é apresentada, aqui, como um conjunto de caminhos metodológicos possíveis de serem combinados, para que, quem se interessar por percorrer essas trilhas, possa fazer uma formação básica na área especificamente escolhida.

Itinerários para se compreender a ciência social computacional

Apesar de um campo de crescente interesse de pesquisa, a ciência social computacional não pode ser definida como um campo de conhecimento com fronteiras bem delimitadas. Para quem se aproxima desse novo campo, há várias combinações e entradas possíveis, começando pelos trabalhos em simulação baseada em agentes (Payne *et al.*, 2019) e estudos de sociedades artificiais (Epstein; Axtell, 1996), passando pelas análises de redes sociais (Leon; Rodríguez-Rodríguez, 2017), web semântica e os impactos da comunicação digital na vida moderna (Dasgupta, 2010), até a ciência dos dados aplicada aos fenômenos sociais (Chen; Yu, 2018). A ciência social computacional faz ainda intersecção com outros campos que estudam os fenômenos sociais com recursos computacionais aplicados à análise qualitativa assistida por computador (Caliandro; Gandini, 2016), processamento de linguagem natural (Priante *et al.*, 2016), os métodos computacionais aplicados à estatística, incluindo recursos tradicionais (Freedman, 2008) e novos métodos bayesianos (Jackman, 2009), além das humanidades digitais (Schreibman; Siemens; Unsworth, 2008).

A área de humanidades digitais possui vários pontos interligados ao caminho traçado pela ciência social computacional, mas formou-se por outro itinerário. As motivações de pesquisa de onde partem os estudos que formam as humanidades digitais, situam-se no campo de conhecimento que se poderia chamar, apenas para demarcar um espaço de atuação acadêmico, como socioculturais, históricos, bibliográficos, literários e linguísticos. A associação internacional Alliance of Digital Humanities Organizations – ADHO, fundada em 2005, tem como uma de suas atividades a concessão do prêmio de pesquisa em homenagem à Roberto Busa, considerado o pioneiro no campo das Humanidades Digitais. Essa homenagem, The Busa Prize, concedida trienalmente para pessoas que tenham uma carreira de pesquisa na área, ilustra o trajeto de constituição das Humanidades Digitais.

Roberto Busa, filósofo, teólogo e jesuíta, iniciou, nos anos de 1940, um grande projeto em linguística computacional, em parceria com Universidades e com a empresa de computadores International Business Machines – IBM. O projeto foi chamado de Index Thomisticus (Breve introdução al Corpus Thomisticum), o qual fez a indexação de toda a obra de Tomás de Aquino, para análise textual computacional. Um projeto de grandes proporções para a época, para o qual foram necessários 11 milhões de cartões perfurados e 30 anos de trabalho, para que o *corpus* viesse a público em formato de CD-ROM, nos anos 1980 (IBM 100, 2012).

Essas várias possibilidades metodológicas estão articuladas, na ciência social computacional, em torno do objetivo de compreender a organização, a estrutura e as transformações sociais a partir dos recursos computacionais. O que a ciência social computacional pode acrescentar às ciências sociais, de modo geral, refere-se à possibilidade de se analisar a complexidade das relações sociais a partir de modelos, simulações e dados, articulando microprocessos sociais a macroestruturas organizacionais e vice-versa.

Historicamente, a ciência social computacional apresentou alguns demarcadores que foram se formando ao longo das últimas cinco décadas, juntamente com a popularização dos computadores. A ideia de um campo de conhecimento chamado de ciência social computacional surge com o advento da computação eletrônica, na década de 1970.

Naquele momento, ainda não se tratava, especificamente, de lidar com a questão dos grandes volumes de dados. Antes, pretendia-se compreender a aplicação da lógica de computação para a resolução de questões sociais. Trabalhos como os modelos de segregação, de Thomas Schelling (2006), e o programa de análise de troca de votos, de James Coleman (2006), publicado originalmente em 1976, são exemplos de aplicação da lógica computacional às questões sociais, com pouco ou nenhum uso de equipamentos. Mesmo no desenvolvimento da própria ciência da computação, a proposição da máquina universal de Turing (1950) e o

autômato celular de Von Neumann (1966) também são exemplos de referências importantes para lógica de programação, que foram desenvolvidos, basicamente, sem o uso direto de computadores, embora ambos tenham contribuído de forma decisiva para a construção e os avanços dos primeiros computadores em seus respectivos países.

O adjetivo “computacional” significou também, por algumas décadas, o uso de determinado *software* aplicado à pesquisa. O tratamento de coleções de dados, incluindo os dados sociais, por programas de computadores, era um recurso metodológico já em prática nas ciências sociais, tanto para estudos quantitativos (Berquó; Souza; Gotlieb, 1981) como qualitativos (Bardin, 2011:28), desde meados da década de 1960. O *software* Statistical Package for the Social Sciences (SPSS), originalmente, criado em Fortran pelo cientista político Norman H. Nie, teve sua primeira versão disponível para o público em 1968 (Nie, 1975). Os pesquisadores Karl Jöreskog e Dag Sörbom criaram a primeira versão do *software* LISREL (linear structural relations) em 1970 (Jöreskog; Thiilo, 1972). O programa para análise de dados qualitativos, Etnograph, criado pelo cientista social John Seidel, foi lançado ao público em 1984 (“Etnograph Software”, 2008), sendo reconhecido como o pioneiro que abriu um campo próprio de investigação: a análise de dados qualitativos assistida por computador – CAQDAS – Computer Assisted Qualitative Data Analysis Software – (Fielding, 1993). Em 1985, a empresa Ventana Systems, em Boston, nos Estados Unidos, lançou o programa Vensim, que se tornaria padrão para a simulação de sistemas dinâmicos. Esses são exemplos de programas que tiveram grande impacto na forma como a pesquisa em ciências sociais passou a ser realizada, tanto com métodos quantitativos, quanto com métodos qualitativos. O uso de *software* aplicado à pesquisa foi um dos principais fatores motivadores de inovações metodológicas em todos os campos das ciências sociais, desde os anos 1970 até a virada do século XXI. A partir de 1990, começaram a ser desenvolvidas as diferentes linguagens e ambientes de programação, que estão entre as mais utilizadas em análise de dados atualmente, como Java, R e Python.

Nos anos 1980, o termo “ciência social computacional”, como um novo campo de pesquisa, que amalgamaria a ciência da computação e as ciências sociais, começou a ser utilizado com mais frequência, tanto nas áreas tradicionais das ciências sociais – sociologia, ciência política, antropologia –, como em áreas como a informática, matemática, estatística e física (Edelmann *et al.*, 2020). Essas eram áreas que, por muito tempo, estiveram afastadas do diálogo interdisciplinar sobre os processos sociais. A ciência social computacional foi apontada como uma oportunidade de convergência, que poderia somar conhecimentos em busca de resultados mais profícuos no estudo de fenômenos sociais.

No ano de 1983, George Garson, professor da Universidade da Carolina do Norte, nos Estados Unidos, junto com Ronald Anderson, da Universidade de Minnesota, fundaram a *Social Science Computer Review*. Foi o primeiro periódico de caráter interdisciplinar com o objetivo específico de analisar e disseminar estudos sobre os usos metodológicos da computação em ciências sociais (sociologia, antropologia, ciência política, economia, psicologia) e os impactos da informática na sociedade.

Em 1995, foi criada em Nova York, Estados Unidos, a revista *Computational and Mathematical Organization Theory (CMOT)*, por iniciativa de William A. Wallace, engenheiro de sistemas e pesquisador do Rensselaer Polytechnic Institute, e Kathleen Carley, socióloga da Carnegie Mellon University, criadora do *software* para análise dinâmica de redes, ORA, e umas das fundadoras do Center for Computational Analysis of Social and Organizational Systems (CASOS), na mesma Universidade. Na Inglaterra, no ano de 1998, iniciou-se a publicação do *Journal of Artificial Societies and Social Simulation (JASSS)*, por iniciativa de Nigel Gilbert, da Universidade de Surrey, na Inglaterra, e diretor do Centre for Research in Social Simulation (CRESS).

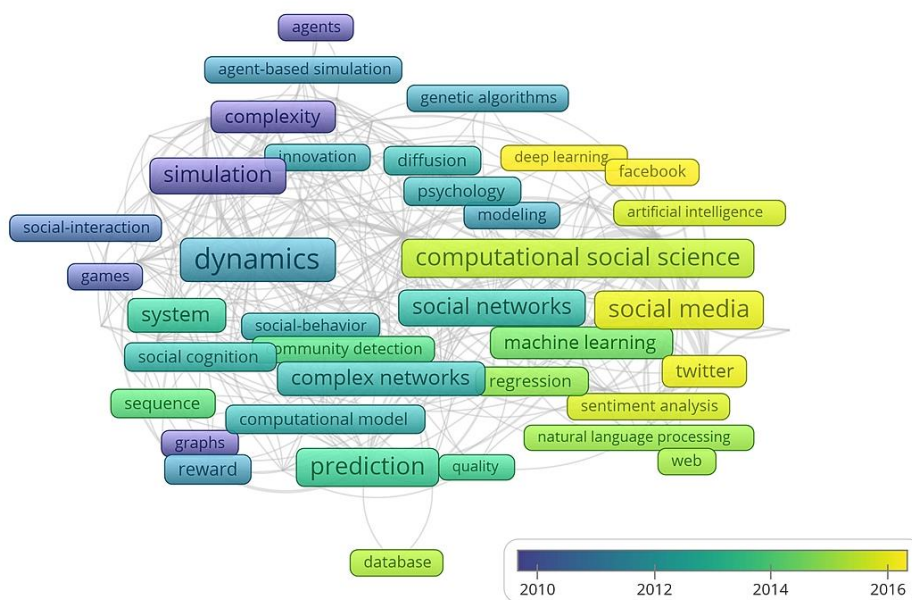
Essa publicação europeia trazia uma forte associação entre ciência social computacional e simulação baseada em agentes – um campo de estudos que começava a se estabelecer como campo interdisciplinar nesse período – principalmente através dos Simpósios de Simulação Social (SimSoc), realizados a partir do ano de 1992 (Gilbert e Troitzsch, 2005: ix). Em 2003, pesquisadores envolvidos com a Revista JASSS, articulada em torno dos estudos sobre simulação social baseada em agentes, fundaram a European Social Simulation Association (Moss *et al.*, 2002).

Tais iniciativas europeias disseminaram os métodos de simulação e se consolidaram como espaço interdisciplinar de pesquisa, associando o termo ciência social computacional com a simulação social baseada em agentes. Em 1999, Uri Wilensky, cientista da computação no Center for Connected Learning and Computer-Based Modeling (CCL), na Northwestern University, nos Estados Unidos, criou a linguagem e o ambiente de programação Netlogo (Wilensky; Rand, 2015), utilizado para construção de simulações com diagramas de fluxos, como também para modelos com agentes. Outros programas dedicados à simulação de agentes eram usados por profissionais e pesquisadores, como Swarm e Mason (Gilbert e Troitzsch, 2005: 256), mas a criação do Netlogo facilitou e ampliou o acesso a iniciantes, estudantes e professores, que demonstravam interesse em se aproximar do tema e, rapidamente, desenvolveram vários modelos de simulação baseados em agentes, em várias áreas do conhecimento, da química à filosofia.

O Netlogo foi concebido com base na proposta de aplicação do pensamento computacional para educação infantil, com a linguagem de programação Logo (Abelson; Goodman; Rudolph, 1974). Porém, mais do que um programa para o ensino, o Netlogo tornou-se uma ferramenta para pesquisa avançada, possibilitando a linguagem comum para os estudos em ciência social computacional, em várias instituições de ensino e pesquisa.

Além dos modelos baseados em agentes, os métodos de simulação social computacional, iniciados por diferentes caminhos, desde os anos 1950, compreendiam, entre outros, os sistemas dinâmicos, microssimulação, autômatos celulares e a análise de redes sociais. Cada uma dessas abordagens possui pontos de origens distintos. Os sistemas dinâmicos têm como fundamento os modelos físicos, baseados em equações diferenciais (Duggan, 2016). Modelos baseados em agentes individuais têm uma contribuição a partir da teoria dos jogos e da sociologia matemática (Coleman, 2006), enquanto modelos baseados em multiagentes têm como origem a ciência da computação, sobretudo a área de inteligência artificial (Wilensky; Rand, 2015). Os autômatos celulares se originam na ideia de criação de mecanismos computacionais para autorreplicação de Von Neuman, que funcionam a partir de regras simples de contagem das mudanças de estados das células adjacentes (Cood, 1968). A análise de redes sociais teve dois caminhos distintos que se encontraram posteriormente. Um caminho pela psicologia (Burt; Kilduff; Tasselli, 2013) e as ciências sociais (Scott, 1988), outro pela ciência da computação e da matemática dos grafos (Brandes, 2005).

Gráfico 3 - Coocorrência de palavras-chave, por ano, entre os mil artigos mais citados, contendo a palavra-chave Computational Social Sciences



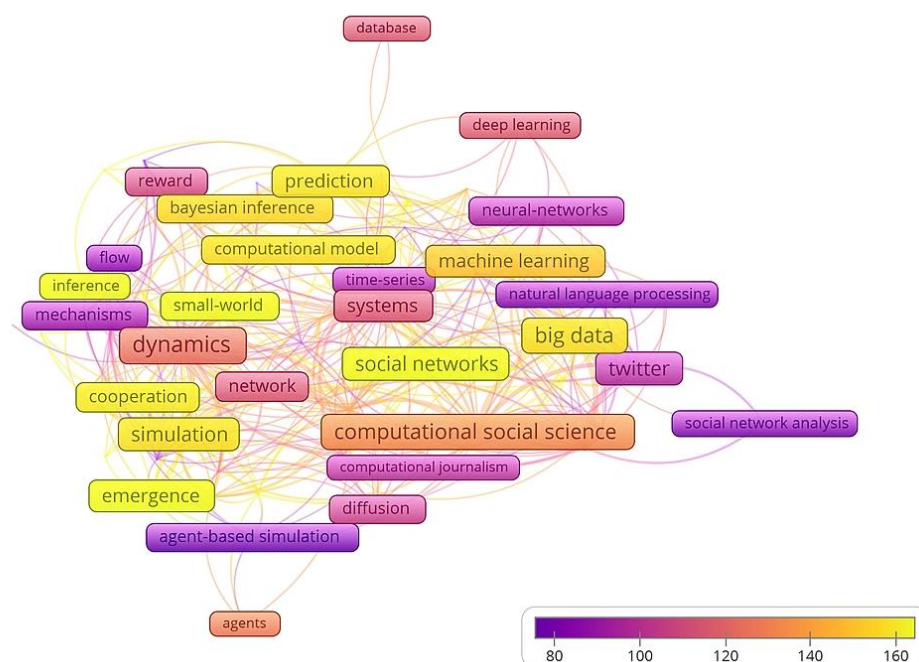
Fonte: Clarivate Web of Science. Elaboração própria.

O Gráfico 3 mostra os termos-chaves mais citados ao longo do tempo, entre os mil artigos mais citados, de um total de 17.956 artigos na base WoS, que contém as palavras “computational social 26gente26” no título ou no resumo. No gráfico, as cores mais escuras à esquerda indicam maior quantidade de citações em anos anteriores, enquanto a cor mais clara, aponta as citações mais recentes. O período coberto para a consulta foi de 1944 a 2021. Os mil artigos com maior número de citações estão entre os anos de 2010 e 2016. Pode-se observar que, em 2010, a ocorrência principal se concentrava em termos correlacionados com simulação, complexidade e teoria dos jogos. A associação da ciência social computacional com simulação baseada em agentes foi passando para a análise de redes sociais entre 2010 e 2012. Nos anos seguintes, de 2013 a 2015, passaram a ter mais citações trabalhos em ciência social computacional correlacionados com bases de dados, aprendizagem de máquina, processamento natural de linguagem, análise de regressão, todas vinculadas à área de inteligência artificial e ciência dos dados. Após 2016, começam a aparecer os termos “deep learning” e análise de sentimentos, vinculados aos estudos de mídia social, como Facebook e Twitter. É importante destacar, no Gráfico 3, que a ocorrência, entre 2012 e 2013, do termo associado à pesquisa qualitativa (quality) já se destacava entre os mil artigos mais citados em ciência social computacional, na base WoS. O gráfico mostra como a denominação “ciência social computacional” vem se adaptando ao surgimento ou revigoramento de novos campos de pesquisa.

A maior atração de pessoas para o estudo na área de ciência social computacional pós 2018 tem ocorrido mais pelo interesse em abordar conflitos e polarizações em redes sociais, com dados do Facebook e Twitter. Uma década antes, a simulação social baseada em agentes era o foco atrativo para a formação em ciência social computacional. Essa mudança ao longo do tempo não é um problema. Ao contrário, mostra que se trata de uma área dinâmica, aberta às rápidas transformações, tanto nas técnicas computacionais de pesquisa, quanto na atenção aos fenômenos sociais estudados.

Apesar das mudanças recentes na ciência social computacional, alguns conceitos se mantêm em destaque nos estudos sobre os fenômenos sociais com os recursos de informática. De acordo com o Gráfico 4, selecionando-se os mil artigos mais citados com a ocorrência do termo “ciência social computacional”, com a base WoS, a mesma do Gráfico anterior, foram destacadas as 25 palavras-chave de maior ocorrência em todos os períodos, de 1944 a 2021. A gradação de cores segue da mais escura, com número de 80 citações aproximadamente, para a cor mais clara, com 180 citações.

Gráfico 4 – Coocorrência de palavras-chave por frequência de citação, entre os mil artigos mais citados, contendo a palavra-chave Computational Social Science



Fonte: Clarivate Web of Science.
Elaboração própria.

É possível observar que alguns conceitos e técnicas de pesquisa se destacam em todos os períodos, como as palavras relacionadas à simulação baseada em agentes (*agent-based simulation*): recompensa (*reward*), cooperação (*cooperation*) e difusão (*diffusion*). Também aparecem as palavras que indicam a abordagem das redes sociais (*social networks*, *social network analysis*, *twitter*, *small world*). Estão entre os mais citados os termos ligados à inteligência artificial e ciência dos dados: *big data*, base de dados (*database*), processamento de linguagem natural (*natural language processing*), redes neurais (*neural networks*), aprendizagem de máquina (*machine learning*) e *deep learning*. Aparecem também os termos relacionados à análise da dinâmica de sistemas, um dos pilares de fundação da teoria dos sistemas: dinâmica (*dynamics*), fluxo (*flow*), simulação (*simulation*), sistemas (*systems*), mecanismos (*mechanism*), séries temporais (*time-series*). Desse conjunto surgem também alguns conceitos que são importantes para a ciência social computacional: inferência (*inference*), previsão (*prediction*) e inferência bayseana (*baysean inference*).

Figura 1 – Itinerários metodológicos da ciência social computacional



Notas: CADQAS – *Computer assisted qualitative data analysis software* (Análise de Dados Qualitativos Assistida por Programa de Computador). PCA – *Principal Component Analysis* (Análise de Componentes Principais).
Elaboração própria.

Com base no histórico e na frequência dos termos mais destacados que se relacionam às citações mais frequentes, é possível apresentar a ciência social computacional como itinerários metodológicos que se entrecruzam. A Figura 1 é uma representação esquemática que não guarda proporcionalidade exata com as citações de palavras apresentadas nos Gráficos anteriores. O objetivo é propor uma visão dinâmica, a partir do conhecimento da diversidade formativa que compõe a ciência social computacional e suas mudanças ao longo dos anos. Pensar a ciência social computacional como uma área de conhecimento fechada, por exemplo, vinculada, exclusivamente, à simulação social, à análise de redes sociais ou ao tratamento de *bigdata* seria incorreto e empobrecedor do potencial que o campo oferece para o estudo dos fenômenos sociais.

Ao invés de uma nova área de conhecimento sobreposta ou concorrente a áreas já existentes, pode-se enxergar a ciência social computacional como possíveis entrecruzamentos de trilhas metodológicas. A coocorrência de palavras-chave demonstra que as pesquisas de maior repercussão na área não estão confinadas e transitam sem problemas por várias estações do conhecimento.

A Figura 1 é uma analogia com um mapa de transportes. Pode-se imaginar a constituição dessa região do conhecimento perpassada por oito linhas ou itinerários formativos: Estatística, Sistema Complexos, Teoria dos sistemas, Humanidades digitais, Pesquisa Qualitativa, Análise de redes, Inteligência Artificial e Ciência dos dados. Cada uma dessas linhas possui sua própria história, seus próprios métodos e seus praticantes, necessariamente, estão fazendo pesquisa em ciência social computacional. Mas quem entra na área de ciência social computacional irá percorrer um ou mais dentre esses itinerários. Para cada itinerário há uma sequência de estações ou pontos de parada.

São os conceitos ou métodos que fundamentam o caminho analítico percorrido. Não há pretensão, com essa figura, em apresentar um mapa detalhado e exaustivo para cada área. Na verdade, são pontos mínimos para destacar o trajeto. As estações maiores indicam pontos de interligação com as demais áreas. A ciência social computacional seria o ponto de partida ou de chegada nesse percurso. Pode-se percorrer vários desses caminhos sem, necessariamente, passar pela ciência social computacional. Mas não há como chegar ou partir da ciência social computacional sem passar por algum desses itinerários. Alguns conceitos foram acrescentados porque, na literatura, se apresentam como caminhos relevantes para a discussão na ciência social computacional, como, por exemplo, a reflexão sobre os modelos estruturais causais de Judea Pearl (Pearl; Mackenzie, 2018).

Como esses caminhos se cruzam de fato? Por exemplo, pesquisas qualitativas assistidas por computador (CADQDAS) têm utilizado recursos de aprendizagem de máquina para auxiliar a codificação dos documentos, comparando a notação manual de diferentes pesquisadores que trabalham sobre um tema. No caminho inverso, os métodos de aprendizagem de máquina se beneficiam da pesquisa qualitativa, que alimentam e treinam modelos preditivos com mais acuidade (Chen *et al.*, 2018:39-2).

A aplicabilidade desse mapa, parte da concepção de que fenômenos sociais resultam em padrões coletivos de regulação das ações individuais. Esses padrões coletivos não existem *a priori*, não são imposição de vontade de um único indivíduo sobre os demais, nem residem na imaginação das pessoas. Emergem da interação manifesta entre os indivíduos, cotidianamente, e se cristalizam em estruturas sociais ao longo do tempo. As estruturas são mutáveis, o que se

percebe como um processo social por meio de modelos analíticos quantitativos ou em combinação com modelos qualitativos.

Um processo de mudança pode ser observado comparando-se as diferentes etapas ou estados dos fenômenos, em seguidos intervalos de tempo. Aquilo que muda de maneira mais vagarosa, e pode até parecer imutável em um primeiro momento, aparece como ordem ou estrutura, e serve de referência para a compreensão do sentido das transformações em curso. Assim, identificar a estruturação ou ordenação de algum fenômeno é essencial para a compreensão de qualquer processo de mudança.

Para se identificar a mudança, é necessário encontrar os padrões que formam a ordem ou a estruturação do fenômeno em análise. Mas não há constância indefinida naquilo que se percebe, num primeiro momento, como ordenado ou estruturado. Com o tempo, tudo muda. Mas, é possível compreender e estimar previsões probabilísticas de determinados tipos de processos sociais por meio dos padrões ou modelos de análise. Essa é a essência e o objetivo da análise computacional aplicada às ciências sociais.

Em um conjunto de relações complexas, não se pode antever a resultante de um único trajeto, mas se pode prever, com certa margem de erro estimada, quais as trajetórias possíveis para determinados eventos, se conhecidos os seus estados e as interrelações que estabelecem entre si. Olhando-se em sentido inverso, com essas mesmas informações, pode-se também explicar, com certa margem de erro, como os eventos se combinaram em padrões para resultar no fenômeno observado.

Obstáculos epistemológicos

Percorrer qualquer um dos percursos indicados na Figura 1 não é uma tarefa fácil. Pode-se dizer que há quatro visões, não necessariamente consolidadas como tal, mas que influenciam o que, potencialmente, pode se desenvolver como ciência social computacional: 1) o *mainstream* das ciências sociais, que rejeita ou desconfia dos recursos de informática; 2) grupos de pesquisa em ciências sociais que vêm adotando recursos de informática para análise e visualização de dados (principalmente, sobre redes sociais), mas desconhecem outros recursos de ciência social computacional; 3) o *mainstream* da ciência da computação, ciência dos dados e estatística em geral, que faz pesquisas e estudos aplicados às questões sociais (mobilidade urbana, redes sociais, etc.), mas ignoram, deliberadamente ou não, a contribuição das ciências sociais; 4) por fim, grupos, em ambos os lados da trincheira acadêmica, que tentam construir a ciência social computacional como um campo interdisciplinar, com a participação de todas as áreas do

conhecimento envolvidas e – respeitadas as diferenças metodológica e epistemológicas de cada campo do conhecimento – com princípios básicos comuns de comunicação de resultados de pesquisa e compartilhamento de recursos analíticos, com vistas ao crescimento colaborativo do entendimento dos fenômenos sociais.

Foi com base nessa quarta concepção que, no ano de 2012, um grupo de pesquisadores de diversas instituições europeias (Itália, França, Reino Unido, Espanha, Holanda, Hungria, Polônia e Suíça) e dos Estados Unidos lançou o Manifesto de Ciência Social Computacional, publicado originalmente, em inglês, no número 214 da revista *The European Physical Journal – Special Topics*. O artigo recebeu uma tradução para o português em 2013 (Conte *et al.*, 2013).

O Manifesto tinha como propósito lançar as bases e estimular a pesquisa interdisciplinar no campo de ciências sociais e das ciências da computação. O cenário que motivou o lançamento do Manifesto era a conturbada situação de crise das hipotecas, que resultou na quebra de vários bancos nos Estados Unidos, de 2008 a 2010, com efeitos bastante graves na economia europeia e em vários outros países do mundo.

Os autores do Manifesto procuraram estabelecer as bases de uma cooperação acima das fronteiras disciplinares, para estimular um engajamento coletivo em projetos de pesquisa de ciência social computacional, compreendendo a ciência da complexidade, ciência dos dados, ciência da simulação baseada em modelos de Agentes, estatística computacional, entre outras.

As dificuldades para se analisar fenômenos sociais com dados são várias, tanto do ponto de vista metodológico quanto epistemológico (Lazer *et al.*, 2020). Autores de diferentes correntes de pensamento, em momentos distintos, postularam algo que se pode vislumbrar como um consenso: em ciências sociais, não deveria haver teoria que não se ancorasse em evidências, nem análise de evidências que não estivesse fundamentada por formulações teóricas. A explicação científica para as questões sociais requer a articulação entre teoria e dados, sejam estes quantitativos, qualitativos, documentais ou de qualquer outra fonte válida para o entendimento dos fenômenos sociais.

Embora a aplicação de métodos quantitativos em sociologia tenha uma fundamentação epistemológica bem clara (Raftery, 2002), ainda persiste uma certa desconfiança em relação à análise de dados quantitativa em ciências sociais, sobretudo quando se trata de métodos computacionais avançados (Conte *et al.*, 2013:48). Como consequência, permanece uma rejeição para a inclusão de modelos analíticos baseado em dados como elemento para a explicação sociológica.

Um dos problemas básicos que alimenta esta desconfiança encontra-se na associação equivocada entre métodos quantitativos e computacionais e “positivismo”. Esta barreira se

solidifica como um obstáculo epistemológico na pesquisa em ciências sociais. Bachelard define um obstáculo epistemológico como “um fato mal interpretado por uma época” que permanece para o historiador, um fato, mas “para o epistemólogo, é um obstáculo, um contra-pensamento” (Bachelard, 1996:22).

Para que as ciências sociais possam avançar metodologicamente com a incorporação de recursos computacionais de análise de dados quantitativos, e mesmo qualitativos, tratados com recursos computacionais, na busca por explicações demonstráveis para os fenômenos sociais, é necessário desfazer-se deste obstáculo. A associação entre positivismo, métodos quantitativos e recursos computacionais é um fato mal interpretado que se tornou um “contra-pensamento”. O efeito é a inibição de estudos que busquem a formação e a aplicação de métodos computacionais nas ciências sociais. De modo geral, as buscas por explicações fundamentadas em dados, sejam quantitativos ou qualitativos, são descredenciadas por serem positivistas. Como afirma Judea Pearl, quando se proíbe a manifestação de uma ideia, proíbe-se o pensamento, o que sufoca o desenvolvimento de métodos e ferramentas associados (Pearl; Mackenzie, 2018:9). Análises fundamentadas em recursos computacionais não são o mesmo que métodos positivistas. São recursos metodológicos à disposição das ciências sociais, já incorporados por outras áreas do conhecimento, que contam com um amadurecimento de décadas de aplicação e reflexão sobre suas possibilidades e limites (Raftery, 2002).

O positivismo, que alimenta o imaginário acadêmico, possui muitas interpretações diferentes (Lacerda, 2009). Para evitar este obstáculo, é importante lembrar as bases do sistema filosófico proposto por Augusto Comte, no século XIX, que não é nem quantitativo, nem científico.

O positivismo apresenta-se como uma revelação de verdade lógica e espiritual. A obra síntese de Comte, “O Catecismo Positivista”, é um texto narrado como um diálogo no qual um sacerdote positivista apresenta seus postulados filosóficos-teológicos para uma discípula (Comte, 1978). Como documento de formação para a catequese da religião positiva, as ideias são apresentadas somente como argumento de autoridade, não havendo demonstrações, nem evidências.

Distante da obra de Comte, o positivismo foi transformado em um espantalho argumentativo (Saussure, 2018:178). É uma construção falaciosa, utilizada geralmente como refutação *a priori*, quando uma pesquisa busca contrastar argumentos de autoridade com evidências empíricas.

Apesar de Comte ser visto como defensor do “cientificismo” e do “empirismo”, na verdade, o desenvolvimento da ciência, nos séculos XIX e XX, deu-se ao largo e em sentido

oposto ao seu positivismo. Nenhum pesquisador, de nenhuma área, desenvolveu métodos científicos inspirados em Comte, simplesmente porque não há método científico no positivismo. A origem desta associação do positivismo com a idealização da ciência advém de uma citação que Comte faz, inspirado em Condorcet, no início do Curso de Filosofia Positiva de Comte.

Condorcet, já no século XVIII, propunha que a matemática e o método de observação e experimentação da física poderiam ser aplicados para fazer avançar o que, na época, se chamava de ciências morais e políticas. O Marquês de Condorcet propunha a “matemática social” em sua defesa da razão contra o que ele e outros iluministas entendiam por “tirania dos preconceitos” e a “hipocrisia das superstições”, que favoreciam as credices e perpetuavam as desigualdades. Como iluminista, acreditava que o método científico, aplicado ao entendimento das questões da sociedade, poderia identificar e aproximar os interesses individuais e os interesses coletivos. Esse entendimento racional do social resultaria em leis, instituições e serviços públicos voltados ao bem comum, preservando as diferenças e aproximando as pessoas (Feldman, 2005: 33). A aposta revolucionária era que a nova “matemática social” seria a promotora do progresso e questionadora dos governos e autoridades que se sustentavam no misticismo religioso, na idolatria personalista e no fanatismo político.

Comte e seus seguidores, na França, foram influenciados por Condorcet e incorporaram a defesa da ciência como ruptura contra a ignorância e promotora do progresso social. Comte usou a ideia do método científico como base para a sua crítica à metafísica e ao escolasticismo. Mas Comte acreditava que, para a realização de uma “síntese” de seu projeto filosófico, era necessário a criação da chamada “religião da humanidade”, assemelhada ao catolicismo.

Como resultado, o positivismo como sistema filosófico conflita com quase todas as formas de pesquisa científica, principalmente com o uso de dados quantitativos para questões humanas. A começar pela discordância de Comte em relação a Adolphe Quételet sobre o uso de probabilidade para o estudo de questões sociais. Dessa polêmica, Comte abandonou seu projeto original de denominar sua ciência aglutinadora do positivismo como “física social”, termo adotado pelo autor belga. Comte mudou sua designação original e adotou o termo “sociologia”.

Na virada do século XIX para o século XX, o positivismo tornou-se alvo de grandes críticas na Europa, sobretudo França, Alemanha, Inglaterra e Itália, com reflexos nos Estados Unidos. Mas não diretamente o trabalho de Comte, como havia sido nas décadas anteriores. Exceto alguns autores franceses, como Émile Durkheim, que dirigiram seus comentários críticos diretamente aos escritos de Comte, os demais, em geral, criticavam um “espírito positivista” genérico (Giddens, 1978:237) que, segundo esses autores, havia dominado o pensamento

europeu, no século XIX. Era uma crítica, na verdade, à ciência e aos ideais do iluminismo – que passaram a ser chamados de “espírito positivista” (Hughes, 2002:34).

O “espírito positivista” tornou-se assim sinônimo de “cientificismo” e passou a ser o alvo principal da crítica tanto do conservadorismo, quanto dos partidários da “mudança social”. Autores marxistas, sobretudo alemães, já no século XX, criticavam-no porque entendiam o positivismo como o método científico da burguesia (Bottomore, 1978:132). Conservadores, sobretudo religiosos, criticavam o positivismo porque defendia a separação entre Estado e Igreja, promovia o agnosticismo e defendia que a moral humana deveria ser submetida à racionalização e às leis naturais que governam o mundo. Esse personagem positivista imaginário alimenta ainda hoje o negacionismo.

Pela via da disputa política nas universidades dos Estados Unidos, o espantinho do positivismo foi usado para criticar os métodos chamados, pejorativamente, de tecnicistas, empiristas e quantitativistas, a partir do final da Segunda Guerra Mundial, geralmente associados ao uso de estatística em ciências sociais (Giddens, 1978:239). A partir da intensificação dessa disputa acadêmica, reflexo da radicalização política nos anos 1960, pesquisar com dados em sociologia recebeu o rótulo definitivo de “positivismo”, sinônimo de alienação, falta de empatia com as demandas sociais, racionalização cartesiana das relações humanas, até chegar ao ponto de métodos quantitativos passarem a ser considerados como não-sociológicos.

Independente do método, compreender um fenômeno social é diferente de posicionar-se a favor ou contra algo. Até porque, para haver adesão a uma ideia ou causa não é necessário haver a compreensão dos processos sociais envolvidos. Pelo contrário, o século XXI tem mostrado que é mais comum o inverso: a firmeza retórica que sustenta uma posição é tanto maior quanto o desdém que se demonstra pelo esforço científico de compreensão dos fenômenos como um todo. Desfazer-se do obstáculo que representa o uso falacioso do positivismo contra a pesquisa com dados é necessário para que as ciências sociais possam ampliar o uso de recursos computacionais avançados. Romper com esse obstáculo significa, ainda, abrir caminho para se desfazer as divergências entre teoria e pesquisa empírica, métodos quantitativos e métodos qualitativos, entre outras disputas desnecessárias (Teddlie; Tashakkori, 2009).

Um outro obstáculo a ser superado está na distinção entre ciências *hard* e *soft*, tendo como base uma *pseudo* separação entre conhecimento objetivo das ciências exatas contra o conhecimento subjetivo das ciências humanas. O debate sobre objetividade e subjetividade não é novo e tem dividido pensadores há vários séculos. São conceitos que variam no tempo e, desde o século XIX, a polêmica tem sido usada para demarcar disputas acadêmicas (Daston; Galison, 2007). A disputa em torno dessas duas palavras demarca trincheiras em campos que se veem

como opostos. De fato, se colocam em lados opostos, mas não pela fundamentação do método científico em si. A oposição ocorre pela necessidade mútua de desqualificação dos trabalhos que não pertencem ao círculo restrito dos que são legitimados como cânones em cada grupo.

Para a ciência social computacional essa postura conflitiva torna-se um impeditivo, dado que sua constituição só pode ocorrer quando há a colaboração entre diferentes campos do conhecimento, indo além das fronteiras entre “Ciências Exatas” e “Humanidades”, “Hard Science” e “Soft Science”, STEM (Science, Technology, Engineering and Mathematics – Ciência, Tecnologia, Engenharia e Matemática) e o restante do conhecimento humano, “métodos positivistas” e “métodos críticos”.

Para quem está do lado “*hard*” da fronteira, o conhecimento é verdadeiro quando produzido por uma ciência fundamentada na objetividade, isto é, neutra e não influenciada pelos desejos de quem julga a realidade. Para um cientista “*hard*”, a contraparte “*soft*” do *campus* acadêmico seria pseudociência.

Uma consequência desse equívoco é a afirmação oposta de que a objetividade é equivalente à abordagem matemática dos fenômenos. O uso de estatística ou qualquer recurso quantitativo não torna nenhum estudo, por si só, mais ou menos verossímil ou objetivo em comparação a outros, baseados em narrativas orais, verbais, textuais, imagéticas ou quaisquer outras fontes de evidências, desde que sejam demonstráveis. Vários autores, em várias áreas do conhecimento, já demonstraram a inutilidade dessa divisão (Bourdieu; Chamboredon; Passeron, 2007).

A matemática é uma ferramenta importante para a pesquisa na medida em que permite dimensionar, construir modelos, classificar, prever e identificar os fatores de mudança nos padrões sociais. Mas é curioso a defesa de que a ciência “*hard*” tem como fundamento justamente a probabilidade. A pesquisa com desenho experimental controlado randômico, ou o padrão ouro das áreas biomédicas, não é a única forma de se fazer ciência em nenhuma área do conhecimento (Shipley, 2002:8). Também não é garantia de isenção, de rigor, nem de verdade, como vem sendo demonstrado por atitudes públicas, na disputa pelos falsos métodos de cura da Covid-19.

Para quem se posiciona do lado “*soft*”, o conhecimento formulado com objetividade seria um mito ingênuo, quando muito aplicável a poucos ramos dos estudos da natureza. A neutralidade esconderia uma postura insensível da pesquisa, fria, sem empatia com as dificuldades humanas. Os interesses humanos estariam sempre presentes e não permitiriam a neutralidade, nem objetividade em nenhum campo do conhecimento. A escolha em se fazer método de pesquisa, sobretudo métodos quantitativos, já carregaria a defesa antecipada de um

ponto de vista, mesmo que não consciente. Assumir a subjetividade, isto é, a defesa das ideias a partir de um ponto de vista explícito, seria então a única forma de se buscar algum conhecimento válido, não necessariamente verdadeiro, mas justo.

Para evitar perder-se em círculos discursivos e falsas divergências nominais, deve-se resgatar alguns entendimentos metodológicos básicos, que são comuns para a realização de estudos científicos em qualquer campo do conhecimento.

Objeto é tudo o que se constitui em foco da indagação do sujeito. Aquilo que se quer conhecer é denominado de objeto do conhecimento (ou objeto de estudos). O objeto do conhecimento é um fenômeno que possui características que lhe são imanentes e independem do observador. Logo, são características objetivas aquelas que são próprias do objeto. São características subjetivas aquelas que são atribuídas ao objeto pelo sujeito que quer conhecê-lo. Por exemplo, quando um pesquisador de arte, analisando uma escultura, a descreve como feita de bronze, ressalta uma característica objetiva, inerente e inseparável do objeto. Quando o mesmo crítico afirma que esta peça é bela, está atribuindo uma característica subjetiva ao objeto, isto é, uma característica que ele, sujeito, percebe e imputa ao objeto. Ser belo é uma característica atribuída pelo sujeito ao objeto. Não é uma característica que está no objeto, mas no sujeito que o observa.

As características subjetivas, uma vez que são atribuições do sujeito ao objeto, variam com cada observador. Diferem das características objetivas, que são próprias do objeto do conhecimento e não variam de observador para observador. A neutralidade não se aplica ao objeto ou ao método, mas ao sujeito que escolhe estudar um objeto a partir de um método. Como todo estudo, de fato, parte de uma posição prévia, dado que o sujeito escolhe o que quer estudar e como estudar, mesmo em relação aos temas da natureza, a não neutralidade não é uma opção, pois qualquer reflexão ou estudo parte, sim, de um ponto de partida de livre escolha de quem faz a pesquisa.

Assim, a questão não é ser neutro ou não, mas ser capaz de olhar o problema por todos os seus ângulos e ir além das concepções prévias. Por isso, explicitar seus pressupostos é realmente uma atitude necessária para qualquer estudo. Porém, há uma diferença enorme entre se afirmar que toda pesquisa parte de pressupostos definidos pelo sujeito dos estudos, logo pressupostos subjetivos, e assumir que se pode distorcer a realidade para caber dentro dessas preconcepções formuladas já desde o início da pesquisa.

Para isso há o método, para garantir que o ponto de chegada da pesquisa seja o mais próximo da realidade possível, ou seja das características objetivas que compõem o objeto de estudos. A subjetividade é o ponto de partida, a objetividade é o ponto de chegada e o método

conduz a razão nesse percurso em busca da validação do conhecimento, isto é, procura verificar se a subjetividade inicial corresponde à realidade objetiva. Se houver correspondência, a proposição estará demonstrada. Se a concepção subjetiva, que motiva a busca do conhecimento, não encontrar correspondência na manifestação objetiva do fenômeno observado, o que fazer? A atitude científica aponta como caminho refazer a proposição à luz das evidências e buscar nova validação nas manifestações objetivas do fenômeno. A postura não científica, ou negacionista, percorre o caminho inverso. Nega a realidade em nome da defesa de suas proposições iniciais, sempre corretas por princípio, na visão do seu proponente.

Essa distinção é simples, mas é vital para a construção da atitude científica. Se o fundamento das ciências sociais consiste em explicar os fenômenos sociais a partir de suas manifestações históricas concretas, portanto imanentes ao processo social que as constitui, então os métodos computacionais podem ser de grande serventia.

Se o propósito de quem faz ciência social é reafirmar pontos de vista, tanto a informática quanto qualquer outro método serão dispensáveis. Não inúteis, porque os recursos empíricos sempre podem ser usados como reforço a um ou outro argumento retórico, mas desnecessários. Pois, quem anuncia como verdades suas impressões sobre o mundo, não precisa de dados ou evidências. Afinal, se os dados vierem a desmentir as preconcepções que sustentam suas visões de mundo, então esses dados não serão confiáveis. Se confirmam o que já se sabia, então não serão necessários. Quando cientistas sociais negam evidências que não condizem com o que sua teoria espera da realidade, acabam por fazer uma espécie de ciência exata baseada em argumentações, um efeito semelhante à reificação criticada por Norbert Elias (Elias, 2008:12).

Ninguém há que renunciar seus ideais para se fazer ciência, muito menos em ciência social computacional. Mas há que se decidir o que se está a fazer: ou o propósito é reafirmar crenças, ou descobrir explicações para os problemas concretamente propostos. O que não se pode é confundir a defesa de um ideal com a explicação da realidade. Se um ideal precisa se sustentar em distorções enviesadas dos fatos, então essa não é uma boa causa a ser defendida, por mais que seus postulados sejam altivos. E não há recurso computacional que possa auxiliar nessa tarefa.

A resistência em se fazer pesquisa com dados, muitas vezes, encobre o receio de se descobrir que a compreensão dos fenômenos sociais não pode ser confinada em preconcepções. Não que o uso de dados não possa ser manipulado para justificar preconceitos. A eugenia, frenologia e o *lysenkoismo* são provas de que quando o preconceito e a ideologia se impõem sobre a razão, a ciência faz muitas vítimas (Ben-Rafael, 2003:123). Mas, a pesquisa baseada em dados e evidências, ainda que não sem percalços, certamente, é o caminho mais seguro para

fazer avançar o conhecimento sobre os problemas com que se defrontam as ciências sociais. Nesse sentido, os itinerários da ciência social computacional têm um potencial muito grande a ser explorado, somando-se ao acúmulo teórico e metodológico das várias áreas de estudo das ciências sociais dos fenômenos sociais.

As possibilidades abertas pelo universo digital

Superados os obstáculos epistemológicos, os recursos metodológicos computacionais abrem grandes possibilidades para as ciências sociais. O século XXI está testemunhando a capacidade de combinação de métodos, ou, ao menos, de intenso diálogo, que permite, de fato, a compreensão da dinâmica social como jamais seria possível nos séculos anteriores. Essa possibilidade está dada pela transformação de todos os dados da vida cotidiana em formato digital. A distinção entre qualitativo e quantitativo tornou-se sem sentido. Toda informação utilizada para uma pesquisa social pode ser analisada do ponto de vista computacional.

A grande mudança não está no *bigdata*, mas na digitalização da vida social. Certamente, a constatação desse fenômeno levanta o problema da privacidade e liberdade dos indivíduos. Esse é mais um dos pontos a serem analisados pela ciência social computacional, na defesa da privacidade e da liberdade individual. Mas, a questão aqui, não diz respeito aos dados pessoais, mas ao fato de que qualquer dado, seja a obra de Tomás de Aquino, seja o Censo Demográfico, toda e qualquer informação, para a análise em ciências sociais, ou está em formato digital ou será convertida em tal formato antes de ser analisada.

O *bigdata* tem a ver com a capacidade de processamento e não com a capacidade de análise dos dados, que se baseia em modelos teóricos. Os grandes volumes de dados ganharam visibilidade como eixo da ciência social computacional com a disseminação do termo *bigdata*, cunhado em 2008 (Bryant *et al.*, 2008) e popularizado por um artigo da Revista Wired, no mesmo ano (Anderson, 2008). Com o uso cada vez mais cotidiano de redes sociais e da telefonia celular, a grande massa de informações digitais chamou a atenção para novos recursos computacionais e até novos campos do conhecimento científico, que se propunham a desvendar o “dilúvio de dados”, desafiador para as teorias e métodos estabelecidos até a primeira década do século XXI (Lazer; Radford, 2017:25).

De acordo com Kitchin (2014:3), a nova era do *bigdata* representa a quarta mudança de paradigma na ciência, desde a Renascença. O autor define a mudança de paradigma. O primeiro paradigma seria a ciência experimental, que tinha como objeto de estudos e preocupação central a descrição dos fenômenos da natureza. O segundo paradigma, que cobre o período após a

Renascença e até antes do advento dos computadores, fundamentou-se na construção de modelos e na busca de generalização das formulações teóricas sobre os fenômenos. O terceiro paradigma, a era da ciência computacional, começa com a popularização da informática e tem como eixo a simulação e a análise dos sistemas complexos. O quarto paradigma se caracteriza pelo *bigdata*, momento no qual as ciências são orientadas pelos dados.

Sob esse novo paradigma, os procedimentos estatísticos veem sendo reformulados para atender à necessidade de exploração de grandes quantidades de variáveis e dados, gerados por fontes diversas e não experimentais. Nas fases anteriores, a estatística era utilizada, prioritariamente, para a confirmação de hipóteses em experimentos controlados com pequenas quantidades de dados e variáveis. A era da ciência orientada para os dados levanta, para as ciências sociais, alguns questionamentos metodológicos, como o acesso e a manipulação de dados, que se atualizam quase que em tempo real, e epistemológicos, como o lugar da teoria na análise de dados (Chen; Yu, 2018:18).

Uma vez que todos os dados se transformam em informações digitais, não há justificativa epistemológica para a persistência da dicotomia entre quantitativo e qualitativo, *hard* ou *soft science*, como se fossem campos éticos distintos. A virtude não está no método, mas no rigor e transparência das ações de quem trilha o caminho da pesquisa, sabendo que fazer perguntas e procurar evidências é mais importante do que ter respostas prontas sobre tudo.

Os processos sociais ocorrem por meio de relações concretas, estabelecidas por pessoas reais. Os fenômenos sociais, em ciência social computacional, são descritos e validados por modelos (Lave; March, 1993). Um modelo computacional é abstraído da realidade, não fruto da imaginação destituída de amparo empírico. Construir um modelo computacional requer base empírica e fundamentação teórica. Com os dados digitais, é possível não só a confrontação do modelo com as evidências sobre a realidade, como também a replicação do resultado, facilitando o compartilhamento e a colaboração científica em ciências sociais.

Os métodos computacionais permitem o entendimento de um fenômeno a partir da explicação demonstrável sobre sua manifestação. Manifestação esta que se torna passível de ser analisada a partir de sua natureza digital por diferentes métodos e técnicas de classificação, inferência e predição dos resultados. A análise computacional procura tornar aparentes as interconexões entre os fatos ou eventos, que tornam possível a manifestação do fenômeno. Dessa maneira, as ciências sociais ganham uma dimensão a mais para a compreensão dos fenômenos sociais, na medida em que se torna possível conhecer as suas implicações de forma preditiva.

A expansão da ciência social computacional em direção a se tornar um campo de conhecimento mais amplo nas ciências sociais em língua portuguesa, com ganhos cumulativos

de compreensão teórica sobre os fenômenos sociais, baseada em evidências e resultados reproduzíveis, requer colaboração no compartilhamento de dados e métodos, formação colaborativa com trocas de experiências constantes e infraestrutura compartilhada (Lazer *et al.*, 2020:1061).

Formar pessoas capacitadas para a pesquisa computacional seguindo um ou mais dos percursos apresentados na Figura 1, com base rigorosa, toma tempo. A pesquisa computacional requer pensamento criativo. Algumas vezes, será necessário olhar além do campo específico de formação disciplinar para ser capaz de trilhar novos caminhos metodológicos. Não adianta aprender apenas a utilizar um *software* e reproduzir esquemas de análises, como receitas. É necessário entender metodologicamente, de fato, o que se está analisando e como deve ser feita a análise de forma adequada. É necessário formar pesquisadoras e pesquisadores em ciências sociais com *computational thinking* (Wing, 2008).

Considerações finais

Novas questões estão propostas para as ciências sociais que podem ser respondidas com auxílio da ciência social computacional (Watts, 2016). Não se trata mais de voltar às indagações dos séculos XIX e XX, mas agora com roupagem de informática. Os novos desafios resultam exatamente das transformações resultantes das relações sociais que se moldaram a partir da informatização da sociedade. Por exemplo, uma nova questão trata da emergência da cultura digital, na qual o mundo virtual é tão ou mais impregnado de sentidos e significados definidores dos papéis sociais do que o mundo não virtual. Ainda há fronteiras que delimitam o mundo virtual do mundo não virtual?

Não há como responder a questões como estas sem os recursos metodológicos, teóricos e conceituais que perpassam alguma das trilhas da ciência social computacional. Do mesmo modo, não há como pensar as novas dimensões da sociedade civil digital, que aparece com perspectiva definidora dos parâmetros de sociabilidade política no século XXI, sem esses recursos computacionais. São vários temas e desafios, como por exemplo, o conflito entre a vida social e redes e a exposição controladora da privacidade ou as implicações das relações humanas e máquinas cada vez mais humanizadas. Em todos esses novos desafios, que vão muito além do exposto aqui, permanece a necessidade de entendimento dos processos que perpetuam as desigualdades. A ciência social computacional pode auxiliar a compreensão mais profunda desses processos.

Assim, são inúmeras as questões levantadas com o objetivo de findar as discussões aqui

apresentadas. Uma vez que, deve-se levar em consideração reflexões tais quais: como os indivíduos se organizam coletivamente; como a organização emerge em diferentes níveis de relações sociais; quais os processos sociais de mudança existentes; quais os tensionamentos resultantes, dado que as pessoas afetam e são afetadas de forma diferente e desigual por estas interações, e, por fim, qual o sentido da mudança. As respostas a estas perguntas podem ser buscadas pela ciência social computacional de forma clara, fundamentada em evidências, para que as pessoas possam utilizá-las para entender e planejar a mudança social, seja através de políticas públicas, ações coletivas, controle social, ensino ou qualquer outra forma que precise da compreensão de como os fenômenos sociais emergem e como se transformam.

O conhecimento, para avançar em sua capacidade de compreensão e explicação do mundo, incluindo os fenômenos sociais, precisa de dados e evidências, teorias e modelos, métodos e meios de pesquisa, formas de disseminação e averiguação colaborativa sobre os resultados. Desse modo, os recursos computacionais podem se somar aos estudos sobre as questões sociais em todas essas fases, contribuindo sobremaneira para os resultados da pesquisa científica.

Referências

- ABELSON, Hal; GOODMAN, Nat; RUDOLPH, Lee (1974), *LOGO Manual*. Artificial Intelligence Lab, Massachusetts Institute of Technology [Consult. 06-06-2021]. Disponível em <https://dspace.mit.edu/handle/1721.1/6226>
- ANDERSON, Chris (2008), “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”. *Wired*, 23 jun. 2008 [Consult. 28-07-2021]. Disponível em <https://www.wired.com/2008/06/pb-theory/>.
- BACHELARD, Gaston (1996), *A formação do espírito científico*. Rio de Janeiro, Contra Ponto.
- BARDIN, Laurence (2011), *Análise de Conteúdo*. São Paulo, Edições 70.
- BEN-RAFAEL, Eliezer (Org.) (2003), *Sociology and ideology*. Boston, Brill.
- BERQUÓ, Elza Salvatori; SOUZA, José Maria Pacheco de; GOTLIEB, Sabina Léa Davidson (1981), *Bioestatística*. São Paulo, EPU.
- BOTTOMORE, Tom (1978), “Marxism and Sociology”, in T. Bottomore e R. Nisbet (Orgs.). *A History of Sociological Analysis*. New York, Basic Books.
- BOURDIEU, Pierre; CHAMBOREDON, J. C; PASSERON, Jean-Claude (2007), *Ofício de sociólogo: metodologia da pesquisa na sociologia*. Petrópolis, Vozes.
- BRANDES, Ulrik (2005), *Network Analysis: Methodological Foundations*. Nova Iorque, Springer Science & Business Media.
- BREVE INTRODUÇÃO AL CORPUS THOMISTICUM [s. d.]. [Consult. 28-07-2021]. Disponível em: <https://www.corpusthomicum.org/wintrop.html>.

- BRYANT, Randal E; MELLON, Carnegie; KATZ, Randy H; LAZOWSKA, Edward D. (2008), "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society". *Computing Community Consortium committee of the Computing Research Association*. Washington DC, Academic Press.
- BURT, Ronald S.; KILDUFF, Martin; TASSELLI, Stefano (2013), "Social Network Analysis: Foundations and Frontiers on Advantage". *Annual Review of Psychology*, v. 64, n. 1, pp. 527-547.
- CALIANDRO, Alessandro; GANDINI, Alessandro (2016), *Qualitative Research in Digital Environments: A Research Toolkit*. Londres, Routledge.
- CHEN, Nan-Chen; DROUHARD, Margaret; KOCIELNIK, Rafal; SUH, Jina; ARAGON, Cecilia R (2018), "Using Machine Learning to Support Qualitative Coding in Social Science: Shifting The Focus to Ambiguity". *ACM Transactions on Interactive Intelligent Systems*, v. 9, n. 4, pp. 21.
- CHEN, Shu-Heng; YU, Tina (2018), "Big Data in Computational Social Sciences and Humanities: An Introduction". In: CHEN, Shu-Heng (org.). *Big Data in Computational Social Science and Humanities. Computational Social Sciences*. Nova Iorque, Springer International Publishing, pp. 1-25.
- CLARIVATE ANALYTICS. *Coleção principal da Web of Science* (2021), [Consult. 07-07-2021]. Disponível em <https://www-webofscience.ez78.periodicos.capes.gov.br/wos/woscc/basic-search>
- COLEMAN, James S (2006), *The mathematics of collective action*. New Brunswick, N.J, Aldine Transaction.
- COMTE, Auguste (1978), *Catecismo positivista: coleção os pensadores*. São Paulo, Abril Cultural.
- CONTE, Rosaria.; GILBERT, Nigel.; et al. (2012), "Manifesto of computational social science". *The European Physical Journal Special Topics*, v. 214, n. 1, pp. 325-346.
- CONTE, Rosaria; GILBERT, Nigel, et al. (2013), "Manifesto da Ciência Social Computacional". *Mediações - Revista de Ciências Sociais*, v. 18, n. 1, pp. 20-54.
- COOD, E. F. (1968), *Cellular automata*. Londres, Academic Press.
- DASGUPTA, Subhasish (Org.) (2010), *Social computing: concepts, methodologies, tools and applications*. Hershey, PA, Information Science Reference.
- DASTON, Lorraine; GALISON, Peter (2007), *Objectivity*. Nova York, Zone Books.
- DE SAUSSURE, Louis (2018), "The straw man fallacy as a prestige-gaining device", in S. Oswald; T. Herman e J. Jacquin (orgs.). *Argumentation and language-linguistic, cognitive and discursive explorations*. Argumentation Library, Berlin, Springer, v. 32, pp. 171-190.
- DUGGAN, Jim (2016), *System Dynamics Modeling with R*. Cham, Springer International Publishing (Lecture Notes in Social Networks).
- EDELMANN, Achim; WOLFF, Tom; MONTAGNE, Danielle; BAIL, Christopher A. (2020), "Computational Social Science and Sociology". *Annual Review of Sociology*, v. 46, n. 1, pp. 61-81.
- ELIAS, Norbert (2008), *Introdução à Sociologia*. Lisboa, Edições 70.
- EPSTEIN, Joshua M.; AXTELL, Robert (1996), *Growing artificial societies: social science from the bottom up*. Washington, D.C, Brookings Institution Press.
- ETHNOGRAPH (software) in L. Given (2008), *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, Estados Unidos, SAGE Publications.

- FELDMAN, Jacqueline (2005), "Condorcet et la mathématique sociale. Enthousiasmes et bémols". *Mathématiques et sciences humaines*, n. 172.
- FIELDING, Nigel (Org.) (1993), *Using computers in qualitative research*. Londres, Sage.
- FREEDMAN, David A (2008), *Statistical models and causal inference: a dialogue with social sciences*. Cambridge, Cambridge University Press.
- GIDDENS, Anthony (1978), "Positivism and Its Critics", in T. Bottomore e R. Nisbet (Orgs.). *A History of Sociological Analysis*. New York, Basic Books.
- GILBERT, G. Nigel; TROITZSCH, Klaus G. (2005), *Simulation for the social scientist*. Nova York, Open University Press.
- GOOGLE BOOKS. *Ngram Viewer* (2021). [Consult. 07-07-2021]. Disponível em <https://books.google.com/ngrams/info>
- HUGHES, H. Stuart (2002), *Consciousness and Society*. Revised edition. New Brunswick, N.J, Routledge.
- JACKMAN, Simon (2009), *Bayesian analysis for the social sciences*. Chichester, U.K, Wiley.
- JÖRESKOG, Karl G.; THIILO, Marielle van (1972), "Lisrel a General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables". *ETS Research Bulletin Series*, n. 2, pp. i-71.
- KITCHIN, Rob (2014), "Big Data, new epistemologies and paradigm shifts". *Big Data & Society*, v. 1, n. 1, pp. 205-229.
- LACERDA, Gustavo Biscaia de (2009), Augusto Comte e o "positivismo" redescobertos. *Revista de Sociologia e Política*, v. 17, n. 34, pp. 319-343.
- LAVE, Charles A.; MARCH, James G (1993), *An introduction to models in the social sciences*. Lanham, Md, University Press of America.
- LAZER, David M. J.; PENTLAND, Alex; WATTS, Duncan J.; ARAL, Sinan; ATHEY, Susan; CONTRACTOR, Noshir; FREELON, Deen; GONZALEZ-BAILON, Sandra; KING, Gary; MARGETTS, Helen; NELSON, Alondra; SALGANIK, Matthew J.; STROHMAIER, Markus; VESPIGNANI, Alessandro; WAGNER, Claudia (2020), "Computational social science: Obstacles and opportunities". *Science*, v. 369, n. 6507, pp. 1060-1062.
- LAZER, David; RADFORD, Jason (2017), "Data ex Machina: Introduction to Big Data". *Annual Review of Sociology*, v. 43, n. 1, pp. 19-39.
- LEON, R D; RODRÍGUEZ-RODRÍGUEZ, R (2017), "Social network analysis: a tool for evaluating and predicting future knowledge flows from an insurance organization". *Science direct*. v. 114, pp. 103-118.
- MOSS, Scott *et al.* (2002), A European Social Simulation Association. *Journal of Artificial Societies and Social Simulation*. v. 5, n. 3 [Consult. 06-07-2021]. Disponível em <https://www.jasss.org/5/3/9.html>
- NIE, Norman H. (Org.) (1975), *Statistical package for the social sciences: SPSS*. Nova York/Düsseldorf, McGraw-Hill.

- PAYNE, Diane; ELKINK, Johan A.; FRIEL, Nial; GRUND, Thomas U.; HOCHSTRASSER, Tamara; LUCAS, Pablo; OTTEWILL, Adrian (Orgs.) (2019), *Social Simulation for a Digital Society: Applications and Innovations in Computational Social Science*. Cham, Springer International Publishing.
- PEARL, Judea; MACKENZIE, Dana (2018), *The book of why: the new science of cause and effect*. NoVA York, Basic Books.
- PIONEERING THE COMPUTATIONAL LINGUISTICS AND THE LARGEST PUBLISHED WORK OF ALL TIME (2012). *IBM 100* [Consult. 28-07-2021]. Disponível em: https://web.archive.org/web/20120327122219/http://www.ibm.com/ibm100/it/en/stories/linguistica_computazionale.html
- PORTAL .PERIÓDICOS. CAPES (2021) [Consult. 07-07-2021]. Disponível em <http://www-periodicos-capes-gov-br>
- PRIANTE, Anna; HIEMSTRA, Djoerd; BROEK, van den, Tijs; SAEED, Aaqib; EHRENHARD, Michel; NEED, Ariana (2016), " #WhoAmI in 160 characters? Classifying social identities based on Twitter profile descriptions". *Workshop on Natural Language Processing and Computational Social Science*. Austin, EMNLP.
- RAFTERY, Adrian E. "Statistics in Sociology", 1950–2000 (2002), in M. Tanner e M. Wells (orgs.). *Statistics in the 21st century. Monographs on statistics and applied probability*. Boca Raton, Chapman & Hall/CRC.
- SCHELLING, Thomas C. (2006), *Micromotives and Macrobehavior*. Nova York, W. W. Norton & Company.
- SCHREIBMAN, Susan; SIEMENS, Ray; UNSWORTH, John (2008), *A Companion to Digital Humanities*. Nova York, John Wiley & Sons.
- SCIELO – BRASIL (2021) [Consult. 07-07-2021]. Disponível em <https://www.scielo.br/>
- SCOTT, John (1988), "Social Network Analysis". *Sociology*, v. 22, n. 1, pp. 109-127.
- SHIPLEY, Bill (2002), *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference*. Cambridge, Cambridge University Press.
- TEDDLIE, Charles; TASHAKKORI, Abbas (2009), *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, SAGE.
- TURING, Alan Mathison (1950), Computing Machinery and Intelligence. *Mind*, v. LIX, n. 236, pp. 433-460.
- VON NEUMANN, John (1966), *Theory of self-reproducing automata*. Urbana, University of Illinois Press [Consult. 02-07-2021]. Disponível em http://archive.org/details/theoryofselfrepr00vonn_0
- WATTS, Duncan (2016), "Computational Social Science: Exciting Progress and Future Challenges" *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Nova York: Association for Computing Machinery.
- WILENSKY, Uri; RAND, William (2015), *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, MIT Press.
- WING, Jeannette M (2008), "Computational thinking and thinking about computing". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 366, n. 1881, pp. 3717-3725.

Abstract

This paper discusses some challenges and the importance of methodological approximation between social sciences and computational social science. The main argument is that interdisciplinary dialogue, provided by computational methods applied to the studies of social phenomena, can bring important contributions to the social sciences, especially due to the continuous and growing transformation of qualitative or quantitative records into digital data. The study provides some insights into how the field of computational social science is constituted and addresses some epistemological obstacles that hinder that approximation, especially in the Portuguese language social sciences.

Keywords: computational social science; agent based social simulation; social network analysis; social sciences.

Resumen

Este estudio discute algunos desafíos y la importancia de la aproximación metodológica entre las ciencias sociales y las ciencias sociales computacionales. El argumento principal es que el diálogo interdisciplinario, proporcionado por métodos computacionales aplicados a los estudios de fenómenos sociales, puede aportar importantes contribuciones a las ciencias sociales, especialmente debido a la continua y creciente transformación de registros cualitativos o cuantitativos en datos digitales. El estudio proporciona algunas ideas sobre cómo se constituye el campo de las ciencias sociales computacionales y aborda algunos obstáculos epistemológicos que dificultan esa aproximación, especialmente en las ciencias sociales de la lengua portuguesa.

Palabras clave: ciencias sociales computacionales; simulación social basada en agentes; análisis de redes sociales; ciencias sociales.
