

Estado da publicação: O preprint não foi publicado em outro meio.

CIÊNCIAS SOCIAIS COMPUTACIONAIS: CARTOGRAFIA DE UM CAMPO EMERGENTE E SEUS SENTIDOS PARADIGMÁTICOS

Nilton Sainz

<https://doi.org/10.1590/SciELOPreprints.13736>

Submetido em: 2025-10-12

Postado em: 2025-10-13 (versão 1)

(AAAA-MM-DD)

A moderação deste preprint recebeu o(s) endosso(s) de:

- Adriano Codato (ORCID: <https://orcid.org/0000-0002-5015-4273>)

CIÊNCIAS SOCIAIS COMPUTACIONAIS: CARTOGRAFIA DE UM CAMPO EMERGENTE E SEUS SENTIDOS PARADIGMÁTICOS

NILTON SAINZ¹

RESUMO: A digitalização da vida social transformou profundamente a produção e a circulação da informação, gerando uma quantidade inédita de rastros digitais. Nesse contexto, as *Ciências Sociais Computacionais* (CSS) emergem como um campo interdisciplinar dedicado a explicar fenômenos sociais por meio de métodos computacionais e da análise de dados em larga escala. Este estudo investiga se as CSS configuram uma transformação teórico-metodológica nas ciências sociais ou se permanecem como um conjunto instrumental de técnicas analíticas. A pesquisa adota um desenho misto que combina uma análise bibliométrica de 100 documentos publicados entre 1997 e 2025 nas bases Scopus e Web of Science, com uma revisão de escopo de 14 trabalhos fundacionais altamente citados. Os dados foram processados no *R* com o pacote *bibliometrix* e visualizados no *Gephi* para mapear a estrutura temática e intelectual do campo. Os resultados indicam rápida consolidação institucional do campo após o artigo de Lazer e colaboradores em 2009, concentrada em instituições de elite dos Estados Unidos, China e Europa. A rede temática é dominada por *big data* e *machine learning*, enquanto a teoria social clássica ocupa posição periférica. A literatura mais influente é predominantemente programática, voltada à definição e legitimação do campo. As CSS encontram-se num ponto de inflexão entre a inovação metodológica orientada por dados e a explicação social informada por teoria. Seu amadurecimento depende da integração entre rigor teórico e sofisticação computacional, bem como da expansão institucional e formação profissional.

Palavras-chave: ciências sociais computacionais; análise bibliométrica; revisão de sistemática de escopo; big data; métodos de pesquisa interdisciplinares.

MAPPING THE EMERGING FIELD OF COMPUTATIONAL SOCIAL SCIENCE: PARADIGMATIC DIMENSIONS AND INTELLECTUAL STRUCTURE

ABSTRACT: The digitalization of social life has profoundly reshaped the production and circulation of information, giving rise to an unprecedented abundance of digital traces. In this context, Computational Social Science (CSS) has emerged as an interdisciplinary field seeking to explain social phenomena through computational methods and large-scale data analysis. This study investigates whether CSS represents a genuine theoretical-methodological transformation within the social sciences or remains primarily an instrumental set of analytical techniques. A mixed-method design combined a bibliometric analysis of 100 documents (1997–2025) indexed in Scopus and Web of Science with a scoping review of 14 highly cited foundational works. Data were processed in R using the *bibliometrix* package and visualized in *Gephi* to map thematic and intellectual structures. Findings reveal rapid institutional consolidation of the field after Lazer (2009), concentrated in elite institutions in the United States, China, and Europe. The thematic network is dominated by big data and machine learning, while classical social theory occupies peripheral positions. The most influential literature remains largely meta-scientific and programmatic, focusing on defining the field rather than empirically applying its methods. CSS currently stands at a crossroads between data-driven

¹ ORCID: <https://orcid.org/0000-0002-3957-2714>, sainznilton@gmail.com, Universidade Federal do Paraná, Programa de Pós-Graduação em Ciência Política, INCT-ReDem. Curitiba, PR, Brasil.

methodology and theoretically informed social explanation. Its maturity depends on integrating computational innovation with robust social theory and expanding beyond Global North infrastructures. This synthesis may redefine how social sciences generate and validate knowledge in the digital age.

Keywords: computational social science; bibliometric analysis; scoping review; big data; interdisciplinarity.

1 INTRODUÇÃO

A sociedade contemporânea vive uma transformação profunda na forma como produz, circula e processa informações sociais. A digitalização das interações humanas gerou uma quantidade inédita de rastros digitais (Lazer *et al.*, 2009; Lazer *et al.*, 2020), registros automáticos de ações, comunicações e comportamentos cotidianos. Essa nova ecologia informacional não apenas amplia o volume de dados disponíveis, mas altera qualitativamente a natureza do conhecimento possível sobre os fenômenos sociais. Se antes as ciências sociais dependiam de entrevistas e levantamentos retrospectivos, hoje é possível observar, em tempo real, dinâmicas de mobilidade, comunicação e comportamento coletivo em escala global. Essa revolução informacional redefine as condições de observação e explicação social, inaugurando uma nova fase na história das ciências sociais.

É nesse contexto que emergem as Ciências Sociais Computacionais (CSS), um campo interdisciplinar dedicado a investigar fenômenos sociais por meio de métodos computacionais aplicados a grandes volumes de dados. Suas origens remontam à segunda metade do século XX, quando pesquisadores como Thomas Schelling, Hebert Simon e James Coleman começaram a empregar simulações e modelos formais para compreender sistemas sociais complexos. No início do século XXI, com o crescimento exponencial dos dados digitais e da capacidade computacional, o campo assume forma própria, marcada pelo artigo seminal de Lazer *et al.* (2009), que reconhece o potencial transformador do Big Data nas ciências sociais.

A presente pesquisa tem como objetivo compreender como o campo das Ciências Sociais Computacionais tem se estruturado e consolidado na literatura científica internacional. Pergunta-se se as CSS configuram uma transformação teórico-metodológica nas ciências sociais ou se permanecem como um conjunto instrumental de técnicas analíticas. Para responder a essa questão, foi adotada uma abordagem mista: 1) uma análise bibliométrica da literatura, a fim de mapear o atual cenário do campo e sua estrutura intelectual; 2) segundo uma revisão de escopo qualitativa, identificando os principais objetos e objetivos dos estudos, as técnicas aplicadas, a mobilização de teoria social e a definição para CSS. Essa abordagem permite mapear tanto a estrutura de produção e circulação do conhecimento quanto os fundamentos epistemológicos que orientam o desenvolvimento das CSS.

As CSS ocupam posição estratégica na renovação das ciências sociais contemporâneas. Ao integrar métodos computacionais, estatística e teoria social, o campo redefine as fronteiras disciplinares e amplia as possibilidades explicativas sobre fenômenos sociais complexos. Contudo, no Brasil, o debate ainda é incipiente: há poucos grupos de pesquisa consolidados, escassa formação metodológica específica e quase nenhuma presença institucional em programas de pós-graduação. Mapear a consolidação internacional do campo, portanto, oferece uma referência crucial para o fortalecimento da pesquisa nacional e para o desenvolvimento de capacidades científicas capazes de acompanhar as transformações globais em curso.

A compreensão desse processo de consolidação insere-se em debates clássicos sobre a dinâmica do conhecimento científico. Seguindo a interpretação de Kuhn, os novos campos do conhecimento emergem quando anomalias acumuladas em paradigmas anteriores criam espaço para novos modos de observação e explicação (Kuhn, 1962), o que se adapta perfeitamente para o surgimento das CSS. Sob essa perspectiva, as CSS podem ser interpretadas como um caso contemporâneo de formação paradigmática, no qual novos instrumentos técnicos geram transformações teóricas, metodológicas e institucionais no interior das ciências sociais.

A partir dessa introdução, o artigo está organizado da seguinte forma: a Seção 2 apresenta o percurso metodológico da pesquisa, incluindo as etapas de coleta, tratamento e análise dos dados bibliométricos. A Seção 3 divide-se em duas partes. Na primeira, descreve os principais resultados do levantamento quantitativo, abrangendo o mapeamento temático e as redes de cocitação. Na segunda parte, por meio da revisão de escopo, analisa qualitativamente a literatura de referência e fundacional do campo. A Seção 5 apresenta as considerações finais dessa pesquisa e propõe caminhos para o desenvolvimento futuro da área, com ênfase na integração entre teoria social, inovação metodológica e estruturação de uma CSS no Brasil.

2 MATERIAIS E MÉTODOS

O desenho metodológico desta investigação combina análise bibliométrica com revisão sistemática de escopo, permitindo mapear simultaneamente a estrutura quantitativa do campo e aprofundar a análise qualitativa de sua literatura fundacional. Esta abordagem dual possibilita compreender tanto os padrões amplos de produção e circulação do conhecimento quanto as características conceituais, teóricas e epistemológicas que definem as Ciências Sociais Computacionais.

Para construir o corpus de análise bibliométrica, foram realizadas buscas sistemáticas em três bases de dados acadêmicas: Scopus, Web of Science (WoS) e Scientific Electronic Library Online

(SciELO). A seleção destas bases visou garantir cobertura abrangente da literatura científica de alto impacto indexada internacionalmente (Scopus e WoS) e, simultaneamente, verificar a presença do campo em contextos regionais específicos, particularmente na produção científica latino-americana (SciELO).

As estratégias de busca foram desenhadas para capturar documentos que não apenas empregam o termo “computational social Science” em seus metadados, mas que se dedicam fundamentalmente à reflexão conceitual, teórica ou metodológica sobre o campo. Buscou-se identificar trabalhos que definem, discutem ou problematizam a própria identidade das CSS, em contraste com estudos que meramente aplicam métodos computacionais sem reflexão epistemológica explícita. No Quadro 1, estão apresentadas as strings de busca empregadas nas bases:

Quadro 1 – Strings de buscas aplicadas nas bases indexadoras Scopus, WoS e SciELO

Base	String de busca	Resultado
Scopus	(TITLE(computational social science) AND TITLE-ABS-KEY(“concept*” OR “paradigm” OR “disciplin*” OR “definition” OR “research”)) AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “re”))	94
Web Of Science	“computational social science” (Title) and “concept*” OR “paradigm” OR “disciplin*” OR “definition” OR “research” (Topic) and Article or Review Article (Document Types)	57
SciELO	“computational social science” (Title) and “concept*” OR “paradigm” OR “disciplin*” OR “definition” OR “research” (Topic) and Article or Review Article (Document Types)	0

Fonte: Elaboração própria com base em Scopus, WoS e SciELO (Setembro de 2025).

As buscas foram realizadas em setembro de 2025, sem restrições temporais, contemplando todos artigos e revisões publicados. A busca inicial resultou em 151 documentos: 94 registros na Scopus e 57 na Web of Science. A busca na base SciELO não retornou resultados, achado que constitui, por si só, evidência relevante sobre a concentração geográfica da produção científica no campo. Após a unificação das bases e exclusão de duplicatas, o corpus bibliométrico consolidado totalizou 100 documentos únicos.

O processamento e análise dos dados bibliométricos foram realizados utilizando a linguagem de programação R no ambiente *RStudio*. O principal pacote empregado foi o *bibliometrix*, ferramenta especializada para análise científica de dados bibliométricos que permitiu importar registros em diferentes formatos, processar metadados, remover duplicatas e consolidar uma base de dados unificada. O pacote *dplyr* foi utilizado para manipulação e ajustes nos dados. As análises de cocorrência de palavras-chave, redes de citação, colaboração institucional e mapas temáticos foram

geradas através das funcionalidades nativas do *bibliometrix*. Visualizações de redes foram posteriormente refinadas no *software Gephi* para aprimoramento estético e analítico.

O corpus bibliométrico final é composto por documentos publicados entre 1997 e 2025, provenientes de 84 fontes distintas. O conjunto apresenta taxa de crescimento anual de 8,16% no número de publicações e média de 34,21 citações por documento, indicando um campo de pesquisa em expansão constante com impacto acadêmico crescente. A Tabela 1 sintetiza as principais características descritivas do corpus bibliométrico consolidado.

Tabela 1 – Descrição do perfil das publicações presentes no corpus da análise bibliométrica

Métrica	Valor
Período Temporal	1997-2025
Fontes (Periódicos, etc.)	84
Documentos	100
Taxa de Crescimento Anual (%)	8,16
Idade Média do Documento	5,8 anos
Média de Citações por Doc.	34,21
Autores	293
Palavras-Chave (Autor)	540
Palavras-Chave (Indexadas)	412

Fonte: Elaboração própria com base em Scopus e WoS (2025).

A partir deste corpus bibliométrico, procedeu-se à seleção de um subconjunto de documentos para análise aprofundada através de revisão sistemática de escopo. O critério de seleção fundamentou-se na relevância e influência dos trabalhos, operacionalizada através de indicadores de citação. Foram identificados os documentos mais citados segundo duas métricas complementares geradas pelo *bibliometrix*: “*Most Global Cited Documents*” (documentos com maior número absoluto de citações globais segundo as bases de dados) e “*Most Local Cited References*” (documentos mais citados dentro do próprio corpus analisado) revelando referências centrais para a comunidade científica do campo. Selecionaram-se os 10 documentos com maior pontuação em cada uma destas métricas, totalizando inicialmente 20 trabalhos potenciais.

Após remoção dos documentos que apareciam em ambas as listas, o corpus final da revisão de escopo consolidou-se em 14 documentos seminais publicados entre 1999 e 2024. Este conjunto representa a literatura fundacional de maior impacto e influência no campo das Ciências Sociais

Computacionais, constituindo amostra intencional estratégica para análise qualitativa aprofundada. Os 14 trabalhos foram analisados integralmente e sistematicamente segundo quatro questões estruturantes: (1) Qual o objeto de estudo destes trabalhos? (2) Quais técnicas metodológicas empregam? (3) Mobilizam teoria social clássica? (4) Propõem definições para as Ciências Sociais Computacionais e, em caso afirmativo, como as definem? As respostas a estas questões foram sistematizadas em matriz analítica comparativa, permitindo identificar padrões, tensões e lacunas estruturais na literatura fundacional do campo.

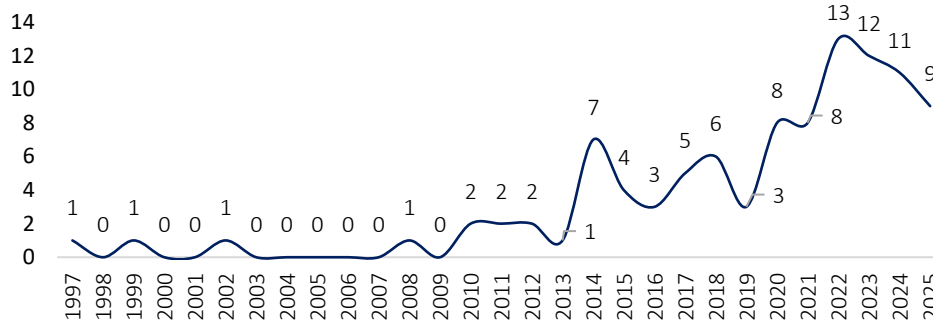
3 RESULTADOS

3.1 *Bibliometric analysis* para um mapeamento quantitativo da literatura

A análise dos dados bibliométricos permite traçar um perfil detalhado do campo das CSS, revelando a sua trajetória de crescimento, os seus centros de produção intelectual e os seus canais de circulação científica. Esta análise é complementada por uma interpretação das redes de coocorrência de palavras-chave e de cocitação de autores, que revelam a estrutura temática e a base intelectual do campo.

A trajetória de publicação em CSS, conforme ilustrado no Gráfico 1, revela um campo em claro processo de consolidação. Após um período de publicações esparsas entre 1997 e 2008, o campo experimenta um ponto de viragem em 2009, coincidindo com a publicação do influente artigo “Computational Social Science” na revista *Science* por Lazer *et al.* A década de 2010 marca a verdadeira ascensão do campo, com um crescimento exponencial que atinge o seu pico em 2022, com 13 publicações. A tendência geral aponta para uma estabilização da produção em níveis elevados, característica de um campo que transita da fase de “emergência” para a de “consolidação”.

Gráfico 1 – Evolução Temporal das Publicações em CSS (1997-2025)



Fonte: Elaboração própria (setembro de 2025).

A análise das afiliações institucionais, apresentadas no Gráfico 2, evidencia que a consolidação das Ciências Sociais Computacionais está fortemente associada à presença de infraestruturas institucionais especializadas. As cinco instituições mais produtivas no corpus: University of Pennsylvania, Wuhan University, Tsinghua University, Leibniz Institute for the Social Sciences (GESIS) e George Mason University exemplificam modelos distintos de institucionalização do campo.

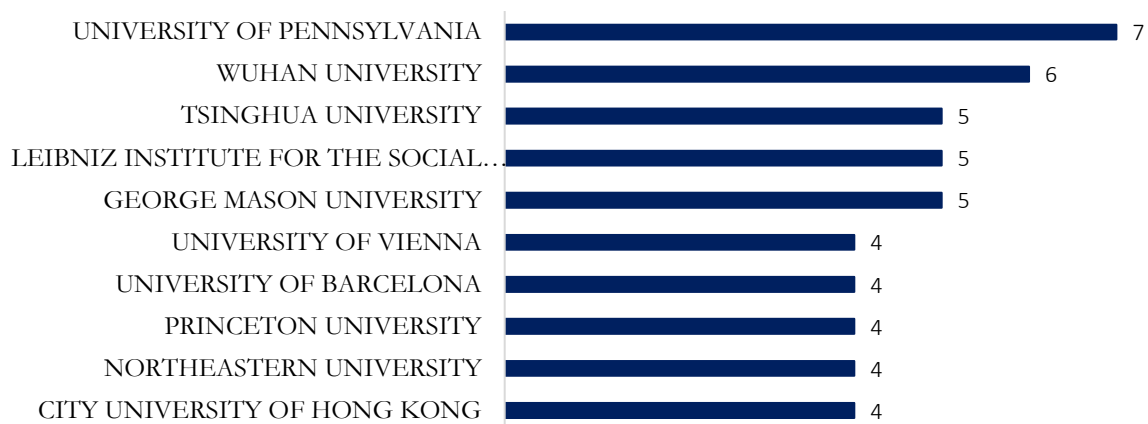
A University of Pennsylvania destaca-se pela integração entre ciências sociais, comunicação e engenharia em seu *Computational Social Science Lab*, voltado à pesquisa aberta com aplicação de métodos computacionais em problemas socialmente relevantes e dados em larga escala, assumindo parcerias com a indústria. A George Mason University representa um dos casos mais precoces de institucionalização, com programa formal de pós-graduação e foco em modelagem e simulação social. A instituição possui um programa de doutorado dedicado às CSS com objetivo formar cientistas sociais computacionais profissionais por meio de uma abordagem interdisciplinar, fornecendo ferramentas e métodos computacionalmente intensivos (como agent-based modeling e análise de redes) para resolver problemas complexos em diversas áreas das ciências sociais.

Já a Wuhan University reflete a expansão asiática do campo, com crescente investimento em ciência de dados aplicada às ciências sociais. Um exemplo é o *Geocomputation Center for Social Science*, que promove a pesquisa em ciências sociais espaciais e a cooperação internacional, alavancando a *expertise* em ciência geoespacial para desenvolver e aplicar a teoria e os métodos de dados espaciais em pesquisa, formação e prática social. Outra universidade chinesa com protagonismo na área é Tsinghua University, que possui o *Future Intelligence Lab*, dedicado ao desenvolvimento de modelos de inteligência artificial e ciência de dados, aplicando machine learning para modelar e controlar sistemas complexos com foco em inteligência incorporada, ciência urbana e computação social. Além disso, a universidade chinesa recebeu o *Annual International Conference on Computational Social Sciences* em 2021, na escola de Ciências Sociais e parece também possuir um instituto focado na área, chamado de Computational Social Science Institute of Tsinghua University.

O *GESIS - Leibniz Institute for the Social Sciences*, na Alemanha, atua como centro europeu de referência em dados e métodos de pesquisa social digital, oferecendo infraestrutura e padrões abertos para o uso ético e transparente de dados comportamentais. O departamento de CSS é o setor de expertise metodológica do *GESIS* que foca em dados digitais e métodos computacionais, sendo crucial para a inovação dos serviços fundamentais que a instituição oferece para as ciências sociais.

De modo geral, essas instituições demonstram que a produção científica em CSS é sustentada por ecossistemas institucionais complexos que combinam formação interdisciplinar, capacidade computacional, colaboração internacional e parcerias com a indústria. Essa concentração reflete não apenas o investimento que vem sendo realizado, mas a consolidação de um modelo de pesquisa ancorado em infraestrutura, expertise técnica e integração entre domínios disciplinares, um modelo ainda incipiente em muitas regiões do mundo, onde a ausência de centros comparáveis limita a participação equitativa no avanço do campo.

Gráfico 2 – Top 10 Universidades e Institutos de Pesquisa por Número de Publicações



Fonte: Elaboração própria (setembro de 2025).

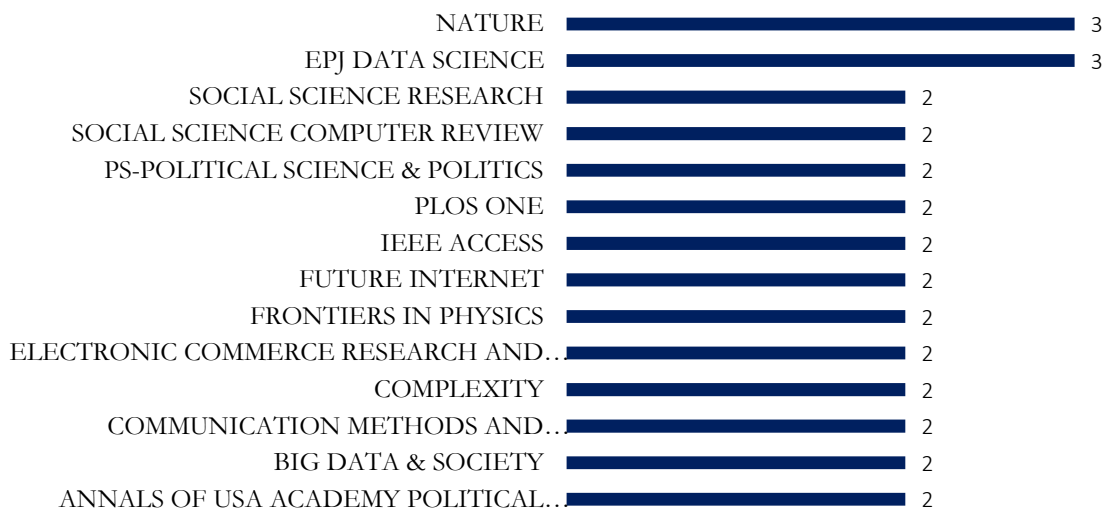
Finalmente, a análise dos principais periódicos onde a produção em CSS é publicada (Gráfico 3) revela estratégia dual de legitimação que opera simultaneamente através da busca por visibilidade em periódicos generalistas de elite científica e a construção de canais especializados próprios. A presença da Nature no topo da lista com três publicações, juntamente com Science (identificada na revisão de escopo com artigos fundacionais incluindo Lazer et al. (2009) e Centola (2010)), constitui evidência de que o campo conseguiu capturar atenção da comunidade científica ampla, transcendendo fronteiras disciplinares tradicionais.

Publicações em periódicos como Nature e Science conferem legitimidade científica, são lidas por audiências multidisciplinares, possuem fatores de impacto elevados, e sua seleção editorial rigorosa funciona como selo de qualidade que valida campos emergentes. Contudo, esta estratégia de legitimação externa coexiste com processo de institucionalização interna através da criação e consolidação de canais de comunicação especializados. A presença significativa de EPJ Data Science, Social Science Computer Review e Big Data & Society indica emergência de periódicos que funcionam

como “canais oficiais” do campo, estabelecendo padrões de qualidade, definindo agendas de pesquisa e construindo identidade disciplinar compartilhada.

A diversidade temática dos periódicos identificados demonstra a interdisciplinaridade das CSS. Periódicos especializados em física (Frontiers in Physics, Complexity) refletem a tradição da ciência das redes e modelagem de sistemas complexos que constitui uma das raízes históricas do campo. Periódicos de ciência política (PS-Political Science & Politics), comunicação (Communication Methods and Measures), e interdisciplinares das ciências sociais (Social Science Research) e áreas do desenvolvimento tecnológico, como engenharia e computação (IEEE Access) demonstram a incorporação não apenas dos métodos computacionais em suas pesquisas, mas também dos objetos e problemas que as CSS estão dedicadas.

Gráfico 3 – Principais periódicos por número de publicações (mínimo 2 publicações)



Fonte: Elaboração própria (setembro de 2025).

A análise da rede de coocorrência de palavras-chave oferece visualização cartográfica da estrutura conceitual das Ciências Sociais Computacionais, revelando não apenas quais temas são investigados, mas como estes temas se relacionam, quais conceitos funcionam como pontes integradoras e quais permanecem periféricos ou isolados. A configuração espacial dos termos, suas conexões e suas centralidades relativas constituem evidências sobre a estrutura de poder epistêmico do campo: quais conceitos definem a identidade compartilhada, quais abordagens gozam de legitimidade central e quais perspectivas ocupam posições marginais.

A característica mais imediatamente evidente da rede é a dominância absoluta de “big data” como nó central massivo que funciona como hub gravitacional conectando praticamente todos os

sub-campos e linhas de pesquisa. Esta centralidade não é meramente topológica mas epistemológica: *big data* constitui o conceito organizador que confere identidade compartilhada a um campo fragmentado em múltiplas disciplinas, métodos e objetos substantivos. A magnitude do nó *big data* em relação a todos os outros termos sugere que o campo define-se primariamente não por questões teóricas compartilhadas, objetos substantivos comuns ou compromissos epistemológicos unificados, mas pela disponibilidade e análise de um tipo específico de recurso empírico: dados em escala sem precedentes.

Esta centralidade hegemônica apresenta implicações ambivalentes para o desenvolvimento do campo. Por um lado, fornece identidade operacional clara e critério de pertencimento relativamente objetivo: pesquisas que empregam ou analisam *big data* podem reivindicar legitimamente o rótulo de CSS. Esta definição instrumental pode facilitar a agregação rápida de pesquisadores de backgrounds diversos, acelerando o crescimento exponencial observado na análise temporal. Por outro lado, identidade fundada primariamente em características dos dados como sua escala, origem digital, disponibilidade massiva, privilegia dimensões técnicas e metodológicas sobre considerações teóricas ou substantivas. A rede revela, portanto, campo estruturado em torno de meios (dados e métodos para analisá-los) mais do que fins (questões teóricas sobre sociedade).

O grafo de rede revela a estrutura multifacetada do campo das Ciências Sociais Computacionais (CSS), organizado em quatro grandes núcleos temáticos. O primeiro, metodológico-sociológico, articula “sociology”, “text analysis”, “computer simulation” e “content analysis”, refletindo o predomínio da análise textual e da simulação computacional como principais pontes entre métodos computacionais e questões clássicas das ciências sociais. O segundo, machine learning e comportamento, conecta “big data”, “machine learning”, “behavioral research” e “social theory”, indicando uma integração ainda incipiente entre aprendizagem de máquina, modelagem e explicação teórica, o que pode estar associado ao movimento de “modelagem integrativa”.

A modularidade da rede ficou em 0,475, o que pode ser considerada uma estrutura modular moderada-forte e que corresponde a clusters bem delineados, mas com zonas de interseção. Contudo, a distribuição dos clusters confirma a centralidade de *big data*, *machine learning*, *behavioral research* e *text analysis* como eixos estruturantes do campo, enquanto os termos teoricamente orientados permanecem em posição com baixa centralidade na rede temática. Além dessa análise concentrada nos agrupamentos, também podemos extrair os termos que não possuem um alto grau na rede, mas que podem estar presentes como motores dos estudos em ascensão no campo, como é o caso de “artificial intelligence”, “natural language processing” e “climate change”.

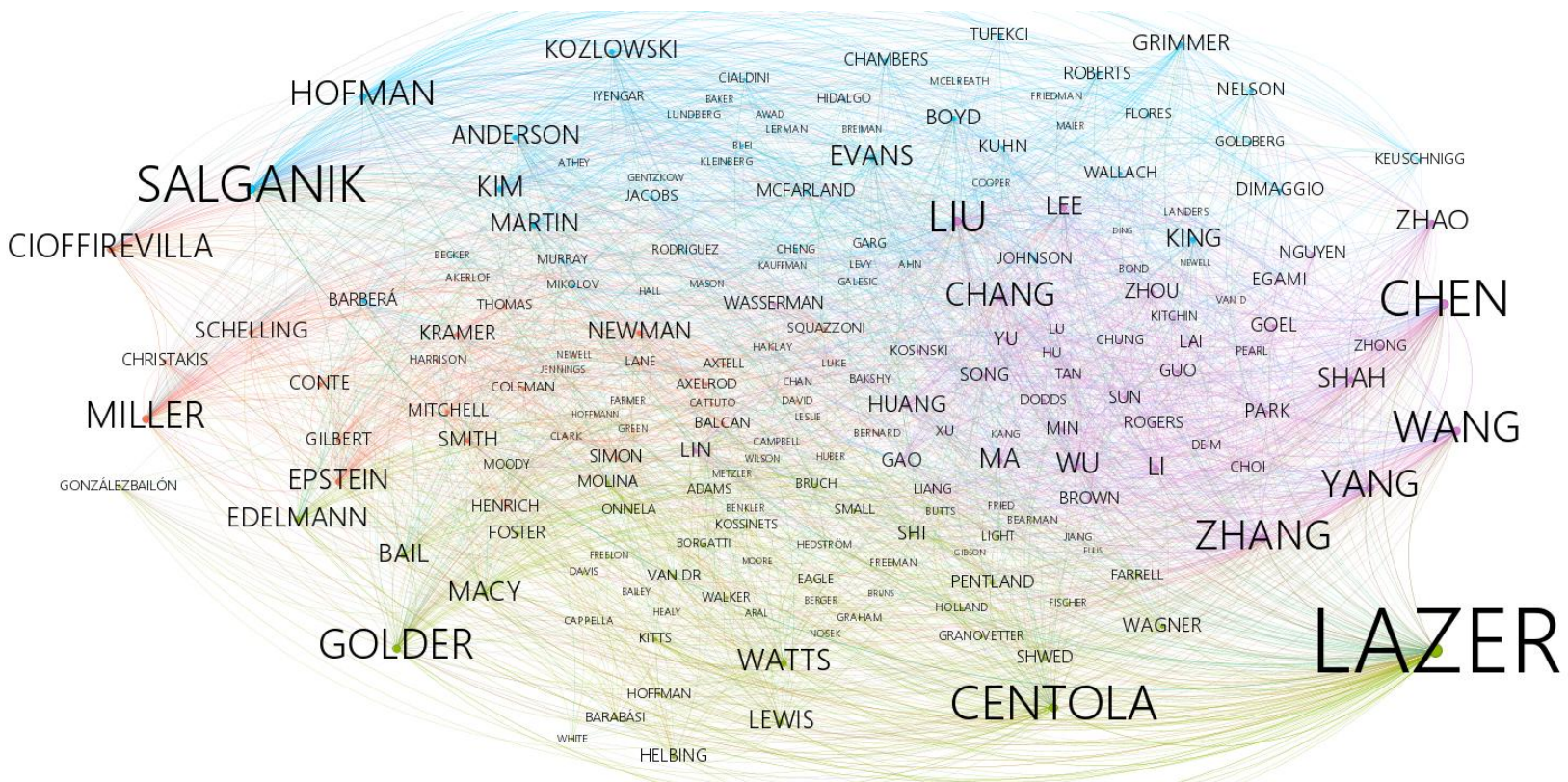
e da análise de redes, com destaque para a presença de Granovetter e Barabási. Este núcleo integra abordagens matemáticas e sociológicas sobre estrutura, difusão e dinâmica de interação online, constituindo a base conceitual e analítica predominante do campo na atualidade.

O segundo agrupamento (laranja), centrado em Epstein, Coleman e Schelling, articula teoria social — especialmente a escolha racional e a modelagem baseada em agentes — com métodos de simulação e modelagem matemática. Este cluster funciona como elo entre a tradição teórica da sociologia e os instrumentos computacionais contemporâneos, oferecendo uma ponte entre explicação causal e experimentação generativa. Também destaque no cluster laranja, a contribuição de Claudio Cioffi-Revilla aborda o objetivo das CSS em torno da investigação da dinâmica e organização humana e social em todos os níveis de análise, incluindo cognição, tomada de decisão, comportamento, grupos, sociedades e o sistema mundial, isso habilitado por instrumentos da computação avançada para impulsionar a investigação e a teoria social.

O cluster azul tem como principal expoente Matthew Salganik, autor de *Bit by Bit: Social Research in the Digital Age* (2018), obra que sistematiza o uso de dados digitais e propõe a modernização dos desenhos de pesquisa à luz das transformações tecnológicas. Nessa mesma direção metodológica, destaca-se também Hofman, cujos trabalhos propõem a integração das prioridades epistêmicas da ciência social voltada à explicação e da ciência da computação centrada na predição. Este agrupamento caracteriza-se, assim, pela ênfase na inovação metodológica e na reflexão sobre os próprios métodos computacionais aplicados às ciências sociais. Soma-se ao cluster metodológico, as contribuições de Grimmer (análise textual computacional), Evans (interações entre métodos computacionais e ferramentas sociológicas) e Gary King (inferência causal e métodos quantitativos em ciência política).

Por fim, o cluster chinês, composto por Chen, Wang, Zhang, Liu e Yang, constitui um polo autônomo e denso de citação, sinalizando a consolidação de uma comunidade científica nacional com crescente autonomia intelectual e capacidade produtiva. Este agrupamento representa a emergência de um novo centro de produção em CSS fora do eixo anglo-americano, ainda que mantenha forte diálogo metodológico com as tradições dominantes do campo.

Imagem 2 – Grafo de rede de cocitação de autores



Fonte: Elaboração própria (setembro de 2025).

A análise bibliométrica revela um campo em rápida consolidação institucional, mas marcado por desequilíbrios estruturais significativos. Após crescimento exponencial desencadeado pelo artigo fundacional de Lazer et al. (2009), as CSS estabeleceram-se através de infraestruturas robustas em instituições de elite – particularmente Estados Unidos, China e Europa Ocidental – com laboratórios dedicados, programas de formação e publicação em periódicos de alto impacto. A estrutura temática organiza-se hegemonicamente em torno de “big data” como conceito integrador, com clusters metodológicos mais centrais que engajamentos teóricos substantivos. A base intelectual concentra-se em cânone estreito dominado por Lazer, Centola e Salganik, com presença periférica de teoria social e emergência de polo chinês que fragmenta geograficamente o conhecimento.

A revisão de escopo que se segue examina qualitativamente os trabalhos mais citados para compreender se o campo articula uma genuína transformação teórico-metodológica nas ciências sociais ou consolida-se primariamente como conjunto instrumental de técnicas analíticas.

3.2 *Scoping review* para uma análise qualitativa da literatura

O Quadro 2 apresenta a literatura selecionada para esta etapa qualitativa, organizada cronologicamente pelo ano de publicação. Cada artigo foi analisado com o objetivo de compreender os objetos e as definições propostas pelos artigos, e como os estudos mobilizam a tensão epistemológica entre teoria e método.

Para fins analíticos, considerou-se como “teoria social clássica” a menção ou diálogo explícito com autores de reconhecida tradição nas Ciências Sociais, como Bourdieu, Latour, Giddens, Elias, entre outros. A classificação da abordagem foi realizada a partir da identificação de seu propósito central (objetivo do artigo). O rótulo “Transformação teórico-metodológica” foi atribuído a artigos de caráter programático, reflexivo ou de revisão, que discutem a transformação epistemológica ou paradigmática da pesquisa social a partir de técnicas computacionais. Já a categoria “Conjunto de técnicas” foi reservada a trabalhos empíricos ou aplicados, que empregam ou divulgam métodos computacionais específicos como análise de redes, modelagem baseada em agentes, aprendizado de máquina ou mineração de texto sem necessariamente propor um enquadramento teórico mais amplo.

Quadro 2 – Literatura selecionada para a revisão de escopo ordenada por ano de publicação

Ano	Autores	Título	Periódico /editora	Objeto	Teoria Social Clássica	Abordagem
1999	Epstein, J. M.	Agent-Based Computational Models And Generative...	Complexity	Modelos baseados em agentes	Não	Transformação teórico-metodológica
2009	Lazer, D. et al.	Computational Social Science	Science	Fundação do campo CSS	Não	Transformação teórico-metodológica
2010	Cioffi-Revilla, C.	Computational social science...	WIREs Computational Statistics	Revisão do campo CSS	Sim	Transformação teórico-metodológica
2010	Centola, D.	The Spread of Behavior in an Online Social Network Experiment	Science	Difusão de comportamentos em redes	Não	Conjunto de técnicas

Ano	Autores	Título	Periódico /editora	Objeto	Teoria Social Clássica	Abordagem
2012	Giles, J.	Making the links: from e-mails to social networks, the digital traces...	Nature	Dados digitais transformando ciências sociais	Sim	Transformação teórico-metodológica
2012	Conte, R. et al.	Manifesto of computational social science	The European Physical Journal Special Topics	Manifesto da CSS	Sim	Transformação teórico-metodológica
2013	Grimmer, J. & Stewart, B. M.	Text as Data: The Promise and Pitfalls of Automatic...	Political Analysis	Análise automática de textos políticos	Não	Conjunto de técnicas
2014	Chang, R. M. et al.	Understanding the paradigm shift to computational...	Decision Support Systems	Mudança de paradigma científico	Sim	Transformação teórico-metodológica
2018	Salganik, M. J.	Bit by Bit: Social Research in the Digital Age	Princeton University Press	Pesquisa social na era digital	Sim	Transformação teórico-metodológica
2020	Lazer, D. et al.	Computational social science: Obstacles and opportunities	Science	Obstáculos e oportunidades da CSS	Não	Transformação teórico-metodológica
2020	Edelmann, A. et al.	Computational Social Science and Sociology.	Annual Review of Sociology	CSS e Sociologia	Sim	Transformação teórico-metodológica
2021	Hofman, J. M. et al.	Integrating explanation and prediction in computational...	Nature	Explicação e predição	Não	Transformação teórico-metodológica
2022	Baden, C. et al.	Three Gaps in Computational Text Analysis Methods...	Communication Methods and Measures	Lacunas em análise textual	Não	Conjunto de técnicas
2024	Ziems, C. et al.	Can Large Language Models Transform Computational...	Computational Linguistics	LLMs e CSS	Sim	Transformação teórico-metodológica

Fonte: Elaboração própria (2025).

A análise dos 14 textos mais citados em Ciências Sociais Computacionais revela um padrão estruturante que caracteriza a literatura fundacional do campo: uma predominância marcante de trabalhos dedicados à auto-reflexão, à delimitação de fronteiras disciplinares e à legitimação epistemológica, em detrimento da aplicação empírica de métodos computacionais a fenômenos sociais concretos.

A predominância de trabalhos meta-científicos manifesta-se de múltiplas formas ao longo do corpus analisado. Artigos como o seminal “*Computational Social Science*” de Lazer e colaboradores (2009), publicado na *Science*, dedicam-se fundamentalmente a apresentar o campo emergente, seus potenciais e desafios institucionais, sem aplicar efetivamente métodos computacionais a questões empíricas. O “*Manifesto of computational social Science*” de Conte e colaboradores (2012) representa paradigmaticamente esta orientação programática, estabelecendo uma agenda de pesquisa e definindo objetivos para o campo, mas permanecendo no plano da elaboração conceitual. Mesmo a atualização de Lazer e colaboradores (2020), publicada uma década após o artigo fundacional, mantém este caráter reflexivo, focando-se nos obstáculos institucionais que impedem o avanço da CSS. Esta literatura tem como objeto central o próprio campo das Ciências Sociais Computacionais: suas definições, seus métodos potenciais, suas promessas epistemológicas e seus desafios estruturais. Trata-se, portanto, de um corpo de conhecimento que está mais preocupado em definir o que a CSS é do que em demonstrar empiricamente o que a CSS pode fazer.

Esta característica auto-reflexiva torna-se ainda mais evidente quando se analisa a evolução das definições propostas para as Ciências Sociais Computacionais ao longo do período estudado. A trajetória conceitual do campo pode ser periodizada em três fases distintas, cada uma marcada por uma ênfase particular que reflete tanto os desenvolvimentos tecnológicos quanto as preocupações epistemológicas de cada momento. A primeira fase, representada pelo trabalho pioneiro de Epstein (1999), caracteriza-se pela centralidade da modelagem baseada em agentes e pela concepção da CSS como ciência social generativa. A segunda fase, inaugurada pelo artigo fundacional de Lazer e colaboradores (2009), opera um deslocamento radical no centro de gravidade conceitual do campo: da simulação para os dados massivos. A definição proposta passa a enfatizar a capacidade de coleta e análise de dados numa escala que pode revelar padrões de comportamento individual e de grupo por meio dos “rastros digitais” (Lazer *et al.*, 2009).

A terceira fase, correspondente aos trabalhos mais recentes do corpus, reflete uma maturidade crescente do campo através de uma reflexão crítica sobre seus próprios fundamentos epistemológicos e sobre o impacto de novas tecnologias. Hofman e colaboradores (2021) abordam diretamente a

tensão fundamental entre explicação e predição na CSS, argumentando que o campo, ao focar excessivamente no desempenho preditivo de modelos de machine learning, corre o risco de se afastar da busca por explicações causais que constitui o objetivo central das ciências sociais tradicionais. Esta terceira fase caracteriza-se, portanto, por definições mais abrangentes e epistemologicamente sofisticadas, que integram análise de dados massivos, modelagem e uma reflexão explícita sobre os objetivos últimos da investigação social computacional.

A literatura de maior relevância revela que as Ciências Sociais Computacionais têm sido majoritariamente definidas a partir de sua atenção ao comportamento humano e aos fenômenos sociais, distinguindo-se de outras vertentes da computação por manter o foco nas dinâmicas e estruturas próprias da vida social. A centralidade dos dados em larga escala, associados ao Big Data, aparece como eixo organizador das abordagens mais recentes, marcando uma segunda fase de consolidação do campo. Nesse contexto, a interdisciplinaridade emerge como um princípio estruturante, situando as CSS na confluência entre as ciências sociais, a ciência da computação e as ciências naturais. Ainda que em menor medida, a modelagem e a simulação permanecem como componentes relevantes, indicando uma tentativa de integração entre tradições metodológicas distintas e de continuidade em relação às origens mais formais e experimentais do campo.

Esta preocupação obsessiva com a definição e redefinição do campo, manifesta na evolução conceitual ao longo das três fases identificadas, configura um paradoxo fundamental que a revisão de escopo captura. As Ciências Sociais Computacionais justificam sua existência e relevância através do potencial de responder a questões sociais complexas utilizando novos métodos que superariam as limitações das abordagens tradicionais. Os manifestos e artigos programáticos enfatizam repetidamente a capacidade do campo de abordar “grandes problemas” sociais, desde instabilidade financeira e desigualdades estruturais até a disseminação de desinformação e a degradação da qualidade democrática. Contudo, a literatura mais influente e mais citada do campo ainda não demonstra esse potencial de forma empírica.

A escassez de estudos empíricos no corpus de trabalhos fundacionais sugere que o campo tem atravessado uma fase prolongada de construção identitária, característica de disciplinas emergentes que necessitam estabelecer legitimidade acadêmica e demarcar fronteiras disciplinares antes de consolidar um programa de pesquisa empírica robusto (Hofman et al., 2021). Os poucos exemplos empíricos presentes no corpus (Baden et al., 2022; Centola, 2010; Grimmer et al., 2021), revelam que o campo encontra-se em um momento de transição necessária: após uma fase inicial, inevitável e produtiva, de auto-definição e legitimação, as Ciências Sociais Computacionais necessitam

urgentemente traduzir seu imenso potencial metodológico em investigações empíricas substantivas que demonstrem sua capacidade de transformar efetivamente a compreensão dos fenômenos sociais contemporâneos.

A análise da mobilização de teoria social clássica e das técnicas metodológicas empregadas revela uma clivagem epistemológica que estrutura o campo das Ciências Sociais Computacionais. Exatamente metade dos textos analisados estabelece um diálogo explícito com a tradição teórica das ciências sociais, enquanto a outra metade não o faz. Esta divisão não constitui uma mera diferença de ênfase, mas expõe uma tensão fundamental sobre a natureza, os objetivos e a própria identidade do campo. De um lado, encontra-se uma visão que concebe a CSS como continuação das ciências sociais por outros meios, utilizando ferramentas computacionais para responder a debates teóricos de longa data sobre estrutura e agência, processos de micro a macro, emergência de fenômenos coletivos e mecanismos causais. De outro, manifesta-se uma perspectiva que trata a CSS como disciplina essencialmente nova, onde o valor reside primariamente na capacidade de extrair padrões, fazer previsões e revelar regularidades em dados massivos, independentemente de enquadramentos teóricos prévios derivados da tradição sociológica, antropológica ou da ciência política.

Os artigos que mobilizam teoria social clássica demonstram uma preocupação explícita em situar a CSS dentro de tradições intelectuais estabelecidas, conectando as inovações metodológicas a questões teóricas fundamentais das ciências sociais.

O trabalho de Chang e colaboradores (2014) exemplifica paradigmaticamente esta abordagem ao empregar a teoria das revoluções científicas de Thomas Kuhn para enquadrar a ascensão da CSS como uma mudança de paradigma. Os autores não se limitam a celebrar as novas possibilidades técnicas, mas as situam explicitamente dentro de uma narrativa histórico-epistemológica sobre transformações na estrutura da ciência. Adicionalmente, Chang e colaboradores dialogam com Runkel e McGrath sobre o “dilema das três pontas (chifres)” da pesquisa social sobre a impossibilidade de maximizar simultaneamente generalização, controle e realismo, argumentando que a abundância de dados digitais oferece uma resolução parcial deste dilema metodológico clássico.

De forma similar, o Manifesto de Conte e colaboradores (2012) aborda diretamente o problema sociológico clássico da ligação micro-macro, um debate central em teóricos como James Coleman, argumentando que a modelagem baseada em agentes constitui a ferramenta ideal para estudar a emergência de fenômenos coletivos a partir de ações e interações individuais. O manifesto também mobiliza conceitos de teoria dos jogos, modelos de aprendizagem social, racionalidade

limitada e emergência, estabelecendo pontes explícitas entre a tradição das ciências sociais e as novas capacidades computacionais.

Essa orientação teórica também se expressa em trabalhos que, mesmo sem necessariamente recorrer a autores clássicos, articulam suas contribuições a debates conceituais centrais das ciências sociais. Cioffi-Revilla (2010) interpreta as Ciências Sociais Computacionais como uma resposta à necessidade de compreender as dinâmicas sociais em múltiplos níveis de análise. Giles (2012) evidencia como o uso de rastros digitais e dados massivos permite testar e reformular teorias sociais clássicas, consolidando empiricamente o potencial transformador do campo. Edelman *et al.* (2020) discutem a integração entre métodos computacionais e a agenda tradicional da sociologia, destacando suas tensões e potencialidades. Salganik (2018) propõe uma combinação entre técnicas digitais e desenho de pesquisa orientado por teoria social, defendendo o rigor conceitual na aplicação dos métodos computacionais. Por sua vez, Ziems *et al.* (2024) exploram o uso de modelos de linguagem para operacionalizar constructos teóricos, reforçando o diálogo entre inovação técnica e fundamentos conceituais das ciências sociais.

Em contraste marcante, a outra metade dos artigos fundacionais foca-se primariamente em dimensões metodológicas, computacionais e institucionais, sem estabelecer diálogos explícitos com a teoria social clássica. Os trabalhos seminais de Lazer e colaboradores (2009, 2020), apesar de sua influência fundacional no campo, dedicam-se essencialmente a descrever as novas possibilidades oferecidas pelos dados digitais massivos e a identificar obstáculos técnicos e institucionais ao desenvolvimento da CSS. O artigo de 2009 menciona brevemente teoria de redes sociais, mas seu foco predominante recai sobre as capacidades metodológicas, uso de técnicas e sobre desafios práticos como acesso a dados, privacidade, ética e risco de monopolização do conhecimento por empresas privadas. A atualização de 2020 aprofunda esta orientação, concentrando-se em obstáculos institucionais: estruturas universitárias inadequadas, modelos de acesso a dados, questões éticas operacionais e formação de pesquisadores. Em ambos os casos, a justificativa para o campo não reside em sua capacidade de resolver debates teóricos de longa data, mas na novidade dos dados e na potência dos métodos computacionais *per se*.

O trabalho pioneiro de Epstein (1999) estabelece as bases epistemológicas da modelagem baseada em agentes, mas o faz a partir de referenciais da filosofia da ciência e da teoria econômica, sem recorrer à tradição sociológica clássica. Grimmer e Stewart (2021), assim como Baden *et al.* (2022), concentram-se na avaliação técnica de métodos de análise automatizada de texto, discutindo desempenho e validade sem articulação com debates teóricos das ciências sociais. Centola (2010), por

sua vez, exemplifica o potencial experimental da pesquisa digital ao investigar a difusão de comportamentos em redes sociais, mas permanece ancorado em uma abordagem metodológica, sem diálogo explícito com teorias sociológicas.

A análise das técnicas metodológicas presentes no corpus reforça essa clivagem entre tradição teórica e inovação instrumental. A análise de redes sociais desponta como método transversal e estruturante do campo, consolidando a compreensão dos fenômenos sociais por meio de suas propriedades relacionais. Os experimentos online ganham destaque como estratégia de investigação em larga escala, enquanto a modelagem baseada em agentes, presente desde os trabalhos fundadores, mantém relevância na explicação de processos emergentes. Já as técnicas associadas ao Big Data, machine learning e processamento de linguagem natural expandem-se nos estudos mais recentes, acompanhando a crescente disponibilidade de dados e recursos computacionais.

A clivagem entre artigos orientados pela teoria e artigos orientados pelos dados manifesta visões fundamentalmente diferentes sobre o que constitui uma contribuição científica válida no campo das Ciências Sociais Computacionais. Para aqueles que mobilizam teoria social clássica, o valor da CSS reside em sua capacidade de operacionalizar, testar e refinar teorias existentes sobre fenômenos sociais, utilizando dados e métodos que antes eram inacessíveis. A contribuição distintiva da CSS, nesta perspectiva, não provém meramente da aplicação de técnicas sofisticadas, mas da capacidade de iluminar debates teóricos de longa data com evidências empíricas em escala e granularidade sem precedentes.

Em contraste, para aqueles que priorizam a orientação pelos dados e métodos, o valor da CSS reside em sua capacidade preditiva, na revelação de padrões antes invisíveis e na geração de resultados “nascidos dos dados” que emergem da análise exploratória de grandes volumes de informação, independentemente de hipóteses teóricas prévias. Esta visão enfatiza a ciência orientada por dados (*data-driven science*) que permite a transição de estudos “escassos em dados” para estudos “ricos em dados”, promovendo descobertas indutivas que podem eventualmente gerar novas teorias, mas não dependem necessariamente de teorias existentes para orientar a investigação inicial (Chang et al., 2014; Edelman et al., 2020). A contribuição principal, nesta perspectiva, é a capacidade de fazer previsões acuradas, identificar padrões robustos ou desenvolver aplicações práticas eficazes, mesmo que os mecanismos causais subjacentes permaneçam incompletamente compreendidos.

A tensão epistemológica que atravessa as CSS evidencia tanto seus riscos quanto suas oportunidades. A ausência de um diálogo sistemático com a teoria social pode levar à fragmentação conceitual e à redescoberta de ideias já consolidadas sob novos rótulos computacionais. Por outro

lado, a rigidez teórica e o conservadorismo metodológico podem inibir a inovação e impedir que o campo reconheça fenômenos emergentes próprios da vida social digitalizada e que métodos computacionais podem e devem servir simultaneamente objetivos explanatórios e preditivos, especialmente em contextos de intervenção ou mudança onde compreender mecanismos causais é essencial para generalizar além dos padrões observados nos dados de treinamento.

A literatura recente sugere que o futuro das CSS depende menos da escolha entre teoria ou técnica e mais da capacidade de articular ambas em abordagens verdadeiramente integrativas capazes de explicar e prever, interpretar e modelar. Essa integração demanda formação interdisciplinar, colaboração entre áreas e o cultivo de um pluralismo epistêmico que preserve a criatividade científica sem renunciar ao rigor conceitual. Assim, a tensão que hoje pode parecer dividir a consolidação do campo, se bem administrada, constituir o motor da consolidação e da maturidade teórica do campo.

4 CONSIDERAÇÕES FINAIS

Esta pesquisa investigou como as Ciências Sociais Computacionais têm sido conceituadas, estruturadas e consolidadas na literatura científica internacional, buscando compreender se o campo configura uma transformação teórico-metodológica nas ciências sociais ou se está sendo tratado predominantemente como conjunto instrumental de técnicas analíticas. A análise bibliométrica de 100 documentos publicados entre 1997 e 2025 revelou a rápida consolidação institucional, com crescimento exponencial após o artigo fundacional de Lazer et al. (2009), mas marcado por concentração nos Estados Unidos, China e Europa. A estrutura temática organiza-se hegemonicamente em torno de “big data” como conceito integrador, com clusters metodológicos ocupando posições mais centrais que engajamentos teóricos substantivos. A base intelectual concentra-se em cânone estreito dominado por Lazer, Watts e Salganik, com presença periférica de teoria social clássica.

A centralidade hegemônica de “big data” apresenta implicações ambivalentes: por um lado, fornece identidade operacional clara e critério de pertencimento objetivo que facilitou agregação rápida de pesquisadores de backgrounds diversos, acelerando crescimento exponencial; por outro, identidade fundada primariamente em características dos dados privilegia dimensões técnicas sobre considerações teóricas, estruturando campo em torno de meios (dados e métodos para analisá-los) mais do que fins, que são as questões substantivas sobre sociedade.

A evolução conceitual do campo pode ser descrita através de três fases distintas. A ênfase em simulação generativa (Epstein, 1999), o deslocamento para dados massivos e rastros digitais (Lazer et

al., 2009), até reflexão epistemológica contemporânea sobre integração de explicação e predição (Hofman *et al.*, 2021), o que demonstra amadurecimento progressivo, mas também persistência de lacunas estruturais. A literatura fundacional, exemplificada paradigmaticamente pela centralidade de Lazer cujos trabalhos de 2009 e 2020 tornaram-se referências obrigatórias, dedicou-se intensamente a definir identidade do campo, mas produziu relativamente pouco conhecimento substantivo sobre fenômenos sociais concretos através de aplicações empíricas sistemáticas.

A tensão epistemológica entre abordagens orientadas pela teoria e orientadas pelos dados não representa falha do campo, mas característica de sua fase de consolidação. Contudo, os riscos desta tensão não resolvida são substanciais: ausência de diálogo sistemático com teoria social pode conduzir à fragmentação conceitual e redescoberta de ideias estabelecidas sob novos rótulos computacionais; inversamente, conservadorismo teórico e metodológico pode inibir inovação e impedir reconhecimento de fenômenos emergentes próprios da vida social digitalizada e na era da produção massiva de dados.

A literatura recente, particularmente trabalhos de Salganik sobre modernização de desenhos de pesquisa e Hofman sobre modelagem integrativa, sugere que o futuro das CSS depende menos da escolha binária entre teoria ou técnica e mais da capacidade de articular ambas em abordagens genuinamente integrativas capazes simultaneamente de explicar e prever, interpretar e modelar. Métodos computacionais devem servir objetivos explanatórios e preditivos, especialmente em contextos de intervenção onde compreender mecanismos causais é essencial para generalizar além de padrões observados em dados de treinamento.

A consolidação futura das CSS como campo maduro, rigoroso e substantivamente relevante demanda urgentemente investimentos institucionais deliberados em regiões atualmente marginalizadas. A concentração atual em infraestruturas de elite do Norte Global demonstra que produção científica robusta requer ecossistemas institucionais complexos incluindo programas de formação em múltiplos níveis, infraestruturas de dados e computação, e redes de colaboração interdisciplinar. Universidades e agências de fomento fora do eixo EUA-China-Europa devem priorizar criação de centros e institutos que capacitem pesquisadores locais, desenvolvam infraestruturas próprias e promovam agendas de pesquisa contextualizadas refletindo realidades e necessidades regionais. Cientistas sociais tradicionais enfrentam risco real de marginalização neste ponto de inflexão histórica se mantiverem posturas conservadoras que rejeitam métodos computacionais por princípio ou resistem a colaborações interdisciplinares necessárias. A transformação digital da sociedade torna dados massivos e capacidades computacionais recursos

empíricos incontornáveis; a questão não é se ciências sociais devem engajar-se com estes recursos, mas como fazê-lo de forma teoricamente informada, eticamente responsável e substantivamente relevante.

O caminho mais promissor para o amadurecimento e a democratização das Ciências Sociais Computacionais reside na adoção de uma abordagem integrativa, capaz de articular métodos computacionais avançados a teorias sociais robustas. Essa integração não apenas enriquece o repertório teórico das abordagens computacionais, mas também oferece às ciências sociais tradicionais novos instrumentos para revisitar questões clássicas com escala, precisão e granularidade inéditas. O fortalecimento desse campo tem potencial de impacto direto sobre desafios contemporâneos de política pública e governança, contribuindo para o aprimoramento de políticas, processos e instituições, bem como para o diagnóstico de fenômenos que afetam a qualidade democrática no século XXI. As Ciências Sociais Computacionais configuram, assim, um projeto científico transformador, cuja consolidação depende da superação das lacunas identificadas ao longo deste estudo: ampliação geográfica e institucional, integração teoria–método, diversificação epistemológica e compromisso com aplicações substantivas que demonstrem, de forma concreta, seu valor para a compreensão e intervenção sobre os fenômenos sociais contemporâneos.

REFERÊNCIAS

- Baden, C., Pipal, C., Schoonvelde, M., & Van Der Velden, M. A. C. G. (2022). Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda. Em *Communication Methods and Measures* (V. 16, Número 1, p. 1–18). Routledge. <https://doi.org/10.1080/19312458.2021.2015574>
- Centola, D. (2010). The Spread of Behavior in an Online Social Network Experiment. *Science*, 329(5996), 1194–1197. <https://doi.org/10.1126/science.1185231>
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. Em *Decision Support Systems* (V. 63, p. 67–80). Elsevier. <https://doi.org/10.1016/j.dss.2013.08.008>
- Cioffi-Revilla, C. A. (2010). Computational social science. Em *Wiley Interdisciplinary Reviews: Computational Statistics* (V. 2, Número 3, p. 259–271). <https://doi.org/10.1002/wics.95>
- Conte, R., Gilbert, G. N., Bonelli, G., Cioffi-Revilla, C. A., Deffuant, G., Kertesz, J., Loreto, V., Moat, H. S., Nadal, J. P., & Martínez-Sánchez, Á. E. (2012). Manifesto of computational social science. Em *European Physical Journal: Special Topics* (V. 214, Número 1, p. 325–346). <https://doi.org/10.1140/epjst/e2012-01697-8>

- Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, 46(1), 61–81. <https://doi.org/10.1146/annurev-soc-121919-054621>
- Epstein, J. M. (1999). Agent-based computational models and generative social science. Em *Complexity* (V. 4, Número 5, p. 41–60). [https://doi.org/10.1002/\(SICI\)1099-0526\(199905/06\)4:5%253C41::AID-CPLX9%253E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5%253C41::AID-CPLX9%253E3.0.CO;2-F)
- Giles, J. (2012). Computational social science: Making the links. Em *Nature* (V. 488, Número 7412, p. 448–450). <https://doi.org/10.1038/488448a>
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine Learning for Social Science: An Agnostic Approach. *Annual Review of Political Science*, 24(1), 395–419. <https://doi.org/10.1146/annurev-polisci-053119-015921>
- Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J. M., Margetts, H. Z., Mullainathan, S., Salganik, M. J., & Vazire, S. (2021). Integrating explanation and prediction in computational social science. Em *Nature* (V. 595, Número 7866, p. 181–188). *Nature Research*. <https://doi.org/10.1038/s41586-021-03659-0>
- Kuhn, T. S. (1962). *The structure of scientific revolutions* (2. ed., enlarged, 21. print). Univ. of Chicago Press.
- Lazer, D. M. J., Pentland, A. S., Watts, D. J., Aral, S., Athey, S., Contractor, N. S., Freelon, D., González-Bailón, S., King, G., & Margetts, H. Z. (2020). Computational social science: Obstacles and opportunities. Em *Science* (V. 369, Número 6507, p. 1060–1062). American Association for the Advancement of Science. <https://doi.org/10.1126/science.aaz8170>
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- Salganik, M. J. (2018). *Bit by bit: Social research in the digital age* (First paperback printing). Princeton University Press.
- Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can Large Language Models Transform Computational Social Science? *Computational Linguistics*, 50(1), 237–291. https://doi.org/10.1162/coli_a_00502

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS DA PESQUISA:

O conjunto de dados de apoio aos resultados deste estudo ainda não está disponível ao público. A base de dados utilizada e scripts bibliométricos serão publicados assim que a pesquisa for publicada em periódico.

FINANCIAMENTO:

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES). O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq) por meio do Instituto de Ciência, Tecnologia e Inovação Representação e Legitimidade Democrática (INCT – ReDem).

CONTRIBUIÇÃO DAS/DOS AUTORES/AS:

Nilton Sainz: Conceptualization, Methodology, Software, Data curation, Writing-Original draft preparation, Visualization, Investigation, Writing- Reviewing and Editing.

DECLARAÇÃO DE CONFLITO DE INTERESSE:

O autor declara não haver conflitos de interesse.

MINIBIOGRAFIAS DOS/DAS AUTORAS DO PAPER

Doutor em Ciência Política pela UFPR, Professor colaborador e Pós-Doc do Programa de Pós-Graduação em Ciência Política (UFPR), especialista em *Data Science e Big Data* pela UFPR e pesquisador associado ao INCT ReDem. Coordenador do Portal da Classe Política e VoxDem.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.