

Laboratório de Robôs - como seguir cientistas de dados ? Possíveis aplicações da Teoria do Ator-Rede para a pesquisa antropológica do campo algorítmico.¹

Oscar Arruda d'Alva

Universidade Federal do Ceará (UFC)

Instituto Brasileiro de Geografia e Estatística (IBGE)

Apresentação

Quando propusemos o presente artigo à 32^a. Reunião Brasileira de Antropologia (RBA), eram iniciais, como de resto ainda são, nossos estudos sobre a teoria do Ator-Rede, corrente de pesquisa em teoria social que se originou, nos anos 1980, a partir dos trabalhos em ciência, tecnologia e sociedade de Bruno Latour, Michel Callon e John Law. O artigo seria uma oportunidade para sondar “possíveis aplicações da teoria do Ator-Rede para a pesquisa antropológica do campo algorítmico”, esse o objetivo que cunhamos para a missão. Após alguns meses de leituras e pesquisas para a escrita deste relato, a única certeza que resiste é a da inadequação do subtítulo do artigo, que, apesar de tudo, deve ser mantido como testemunho de nosso aprendizado neste processo.

Como diria Latour, a teoria do Ator-Rede “não se aplica a coisa alguma”² (LATOURE, 2012, p.205) e se pudemos aprender alguma coisa seguindo os seus passos de formiga³, é que não devemos nos apressar em fechar objetos dentro de caixas-pretas⁴, como “campos algorítmicos”, por mais elegantes que sejam. Pelo contrário, nos cabe seguir os atores e observar os seus próprios esforços em fechar as tais caixas, bem como acompanhar as controvérsias que as reabrem. Do título, fiquemos apenas com a pesquisa (ANTropológica?) e com os possíveis caminhos oferecidos pela teoria do Ator-Rede para

¹ Trabalho apresentado na 32^a Reunião Brasileira de Antropologia, realizada entre os dias 30 de outubro e 06 de novembro de 2020.

² O sentido dessa afirmação é o de que: “ferramentas nunca são meras ferramentas a serem aplicadas: sempre modificam os objetos que se tem em mente.” (LATOURE, 2012, p. 208)

³ Analogia favorita de Bruno Latour à sigla inglesa ANT para *Actor-Network Theory*, no decorrer do texto utilizaremos esta sigla como abreviação para Teoria do Ator-Rede.

⁴ Termo recorrente na ANT: “A expressão caixa-preta é usada em cibernética sempre que uma máquina ou um conjunto de comandos se revela complexo demais. Em seu lugar é desenhada uma caixinha preta, a respeito da qual não é preciso saber nada, a não ser o que entra e o que dela sai”. (LATOURE, 2011, p.4)

seguirmos cientistas de dados e seus laboratórios. Talvez também possamos manter os robôs, pelo menos enquanto alegoria desengonçada deste relato.

As principais referências utilizadas como um guia de viagem para a teoria do Ator-Rede foram os livros de Bruno Latour: “Ciência em Ação – como seguir cientistas e engenheiros sociedade afora” (1987) e “Reagregando o Social – uma introdução à teoria do Ator-Rede” (2005). Se no primeiro encontramos um Latour pensando, por meio de uma série de estudos de casos, as possibilidades de um campo de pesquisa dedicado às relações entre ciência, tecnologia e sociedade – o qual futuramente desaguaria na teoria Ator-Rede – no segundo livro, vemos um autor preocupado em posicionar adequadamente uma teoria já estabelecida, criticada e, bem ou mal, utilizada e traduzida por uma comunidade de pesquisadores da tecnociência.

Nesse artigo fazemos uma breve apresentação da Teoria do Ator-Rede, na qual buscamos situar o seu caráter de contraposição à teoria social tradicional, nos apoiando principalmente em Bruno Latour, Michel Callon e John Law. Concluímos a revisão teórica apresentando as cinco fontes de incerteza da ANT, discutidas por Latour (2012). Em seguida utilizamos a primeira fonte de incertezas – a natureza dos grupos – como um experimento empírico das possibilidades oferecidas pela ANT para a pesquisa de nosso objeto, a ciência de dados.

A ideia inicial era a de desenvolver empiricamente cada uma das cinco fontes de incertezas. De fato conseguimos levantar uma quantidade considerável de dados, mas a narrativa ficaria demasiadamente extensa e extrapolaria os objetivos do presente artigo. Nos pareceu melhor limitar, neste momento, nossas pretensões e após colher as impressões e críticas de nossos colegas do Grupo de Trabalho “*Big data e thick data: pensando o lugar da antropologia digital*” da RBA, avaliar a possibilidade de dar continuidade a este percurso narrativo, abordando as demais fontes de incertezas como um roteiro para o estudo da ciência de dados, em futuros trabalhos.

Teoria do Ator-Rede: uma ANT(i)Teoria?

Talvez possamos melhor definir a Teoria do Ator-Rede⁵ como uma teoria negativa ou uma ANT(i)teoria. O tema é controverso mesmo entre os seus fundadores, que sempre resistiram em se apresentar como teóricos. Para Callon (1999, p. 194), por exemplo a “ANT não é uma teoria. Isso é o que lhe dá força e adaptabilidade”, enquanto para Latour “longe de ser uma teoria do social, sempre foi um método cru para aprender dos atores sem impor sobre eles uma visão a priori de suas capacidades construídas no mundo” (LAW e HASSARD, 1999, p.20) e finalmente para Law (1999, p.7), “o desejo de apontar e nomear, de transformar o que hoje chamamos de ANT em uma teoria, tem feito tanto mal quanto bem”. Seja como for, o acúmulo de pesquisas destes autores no campo dos estudos da ciência e tecnologia, convergiu para um diagnóstico de que havia algo de fundamentalmente errado na teoria social tradicional. Tal diagnóstico partiu de dois pontos fundamentais: o próprio conceito de “social” e a relação de causalidade subjacente à teoria social, ou seja, entre aquilo que deveria ser explicado e aquilo que fornece uma explicação.

A ANT apropriou-se da noção de Gabriel Tarde⁶ que considerava o social como um fluido circulante e não um tipo específico de organismo. Nesta visão alternativa, social “não designa um domínio da realidade ou um item especial; é antes o nome de um movimento, um deslocamento, uma transformação, uma translação, um registro. É uma associação entre entidades de modo algum reconhecíveis como sociais” (LATOURE, 2012, p.99). O social, deixa, portanto, de ser tratado como uma substância homogênea e passa a ser compreendido como uma série de associações entre elementos heterogêneos. Seguindo esta definição a sociedade deixa de representar um contexto em que ocorrem atividades não sociais e que pode ser utilizado para explicar aspectos que escapam a outros domínios. Para a ANT, pelo contrário, e aqui reside uma primeira inversão de causalidade, o social é, justamente, o que precisa ser explicado por associações

⁵ A Teoria do Ator-Rede se origina no final dos anos 1980 a partir dos trabalhos na área dos estudos das ciências, tecnologia e sociedade desenvolvidos principalmente por Bruno Latour, John Law e Michel Callon. No web site <http://wp.lancs.ac.uk/sciencestudies/the-actor-network-resource-thematic-list/> mantido por John Law é possível encontrar um registro cronológico da principais obras ANT, bem como de seus críticos.

⁶ Sobre o esquecimento ativo da obra de Jean-Gabriel de Tarde (1843-1904) pela sociologia ver VARGAS(2000)

específicas de entidades não-sociais. Segundo Law (2003), essa inversão decorreu do principal objeto de estudo da ANT, a mecânica do poder:

Se quisermos entender a mecânica do poder e da organização, é importante não começar assumindo tudo o que desejamos explicar. Por exemplo, é uma boa ideia não tomar como certo que existe um sistema macrossocial, por um lado, e pedaços de detalhes microssociais derivados do outro. Se fizermos isso, fechamos a maior parte das questões interessantes sobre as origens do poder e da organização. Em vez disso, devemos começar com uma lousa em branco. Por exemplo, podemos começar com interação e assumir que a interação é tudo que existe. Então, podemos perguntar como alguns tipos de interação mais ou menos bem-sucedidos em se estabilizando e se reproduzindo: como é que superam as resistências e parecem tornar-se "macrossocial"; como é que parecem gerar os efeitos de tal poder, fama, tamanho, escopo ou organização com a qual todos estamos familiarizados (LAW, 2003, p.2. Tradução nossa.)

A regra do social passa a ser a fluidez, enquanto a inércia e a solidez tornam-se as exceções que precisam ser explicadas. Tal explicação, por sua vez, deve recorrer não aos vínculos sociais, mas aos materiais que proporcionam estabilidade aos agregados sociais. Daí decorre a importância dos objetos para esta sociologia. Os humanos e não humanos são tratados simetricamente como atores, desde que façam diferença no curso da ação, modifiquem uma situação e ajudem a explicar a durabilidade e a extensão de uma interação. (LATOUR, 2012, pp. 97-122)

De acordo com Callon (1999), a ANT foi desenvolvida para analisar situações em que é difícil separar humanos e não humanos, e é justamente essa abertura para os não humanos que representa sua ruptura com correntes mais ortodoxas das ciências sociais:

Não há atores modelo. A identidade do ator e da ação depende precisamente das configurações, e cada uma delas pode ser compreendida apenas se concordarmos em dar aos humanos todos os não-humanos que estendem suas ações. É precisamente porque a ação humana não é apenas humana mas também é desdobrada, delegada e formatada em redes com múltiplas configurações, que a diversidade da ação e dos atores é possível. (CALLON, 1999, p. 194. Tradução nossa.)

Na ANT o recurso às noções de força social, explicação social e estrutura social são, portanto, combatidos e substituídos pelo método de estudar as associações, seguir os atores e suas próprias teorias sobre a constituição do social (LATOUR, 2012, p.31). Da compreensão de que o social só se deixaria observar por traços disseminados na constituição de uma nova associação, decorre a opção metodológica tomada por Latour desde o início dos estudos sobre a tecnociência: “nossa entrada no mundo da ciência e

tecnologia será pela porta de trás, a da ciência em construção, e não pela entrada mais grandiosa da ciência acabada” (LATOURE, 2011, p.6)

Para Callon, Law e Rip (1998) os caminhos que levaram ao estudo dos laboratórios como locais estratégicos de mudança social pressupõem o abandono da dicotomia que entre ciência e política:

A ideia de que há um método científico especial, um reino onde a verdade prospera na ausência do poder, é um mito. De fato, é particularmente importante seguir os atores de perto quando eles entram em lugares estratégicos, pois com frequência é do interesse das forças em ação esconder a forma com que eles agem (CALLON, LAW, RIP, 1998, p. 4, Tradução nossa.)

O foco na ciência em construção e nas controvérsias, que envolvem os processos coletivos de produção de fatos e máquinas, é coerente com uma segunda inversão de causalidade da ANT, o abandono simétrico do uso da “sociedade” e da “natureza” como recursos explicativos para os fenômenos. A ANT assume que é na resolução das controvérsias que reside a causa da representação da natureza e a causa da estabilização da sociedade e não o contrário. (LATOURE, 2011)

O estudo das controvérsias é, portanto, fundamental para a ANT. O analista deve acompanhar como estas se formam e como se encerram, desdobrando-as e rastreando suas conexões, ao invés se precipitar em estabilizá-las ou decidir como resolvê-las. (LATOURE, 2012, p.44) Utilizando este mesmo princípio para a explicação da própria ANT, Latour (2012) procurou organizar as principais fontes de controvérsias da teoria social examinando cinco grandes incertezas: a natureza dos grupos; a natureza das ações, a natureza dos objetos, a natureza dos fatos e finalmente a natureza das ciências sociais.

Na próxima seção nos debruçaremos sobre a primeira fonte de incertezas, a natureza dos grupos, tendo como objeto o estudo do agrupamento “cientista de dados”. Seguiremos os passos oferecidos pela ANT para o estudo de grupos, na medida em que levantamos e analisamos dados empíricos e seguimos os atores e suas controvérsias, como um experimento das possibilidades desta teoria para a pesquisa de nosso objeto: a ciência de dados.

Primeira Fonte de Incertezas: A formação do grupo “Cientistas de Dados”

Profissão do futuro, modismo, panaceia, nova arma do capitalismo de vigilância, novo paradigma nas ciências... estas são algumas das visões paradoxais que uma busca pelo termo “ciência de dados” na internet pode nos oferecer. Mas afinal o que é a ciência de dados, e quem são os cientistas de dados? Seguindo os passos de formiga da Teoria Ator-Rede, vamos deixar um pouco de lado os macro conceitos e tentar nos aproximar dos fatos quentes da tecnociência em construção, dirigindo nosso olhar para os atores, suas associações e controvérsias, rastreando as pistas deixadas pelas suas atividades na formação desse novo grupo.

Para Latour (2012, pp. 51-52) “não há grupos, apenas formação de grupos” e a natureza desses agrupamentos é a primeira fonte de incertezas com que lidamos em um estudo da tecnociência. Se uma característica fundamental do mundo social é o empenho de algumas pessoas em desenhar fronteiras que as separam das outras, o papel do analista não consiste em reconhecer a existência inquestionável dessas fronteiras, mas em rastrear as pistas e traços deixados pelos atores na formação dos grupos, tendo como ponto de partida justamente as controvérsias acerca dos agrupamentos.

A ciência de dados parece nos oferecer uma excelente oportunidade de pesquisa ANT, uma vez que se trata de um campo bastante recente e em plena construção, e, como tal, objeto de muitas controvérsias. A literatura técnica (AGGARWAL, 2015; BRAMMER, 2016; SKIENA, 2017), bem como as ementas de novos cursos universitários⁷, concebem a ciência de dados como um campo interdisciplinar emergente, que gradualmente vem delimitando suas fronteiras na intersecção entre a estatística e a ciência da computação e abrindo-se para um vasto campo de aplicações. A análise das pistas, traços e vestígios, deixados pelos atores para a formação deste agrupamento, nos levaram aos Estados Unidos da América, país onde verificamos consideráveis investimentos e controvérsias na formação do novo grupo.

Uma das primeiras citações ao termo “cientista de dados” em um documento oficial nos Estados Unidos ocorreu em 2005, quando o *National Science Board (NSB)*⁸

⁷ A novidade e interdisciplinaridade do campo é destacada no programa dos novos cursos de Ciência de Dados recentemente inaugurados na Universidade de Harvard (<https://gsas.harvard.edu/programs-of-study/all/data-science>), MIT (<https://stat.mit.edu/>) e UC Berkeley (<https://data.berkeley.edu/>)

⁸ O *National Science Board* estabelece as políticas da *National Science Foundation* e atua como assessor do Congresso e do Presidente dos Estados Unidos da América. (<https://www.nsf.gov/nsb/>)

em relatório da *National Science Foundation (NSF)*⁹, propôs um conjunto de políticas e estratégias para o aprimoramento de acervos de dados digitais, e identificou entre as prioridades, o apoio a um grupo emergente de pesquisadores:

“A NSF, trabalhando em parceria com gestores e a comunidade em geral, deve agir para desenvolver e amadurecer a carreira de cientista de dados e para garantir que as iniciativas de pesquisa incluam um número suficiente de cientistas de dados de alta qualidade.” (NSB, 2005, pg.12. Tradução nossa.)

Para o NSB, conselho que define as estratégias de um dos maiores orçamentos públicos para pesquisas científicas do mundo, os acervos de dados digitais haviam se tornado essenciais para a pesquisa e educação, e os cientistas de dados seriam atores-chaves para a manutenção, acumulação e ampliação da nova rede digital. Segundo o NSB (2005) vários fatores já conhecidos contribuíram para esse fenômeno, como: a proliferação de sensores e instrumentos automatizados geradores de dados em formato digital; a redução nos custos de armazenamento de grandes bancos de dados e a facilidade de compartilhamento de dados via internet.

A questão da gestação de uma nova geração de cientistas para lidar com os desafios da gestão e análise de quantidades crescentes de dados também foi objeto de considerações do *National Research Council da National Academy of Sciences*¹⁰ por meio do *Comitê on the Analysis of Massive Data*, constituído em 2010. O relatório do comitê identificou a necessidade de novas abordagens para a era do Big Data e apontou desafios que vão além das dificuldades de armazenamento, indexação e gestão de grandes bases de dados, mas que envolvem a questão da inferência, ou de como gerar conhecimento a partir dos dados, levando em consideração os possíveis vieses na mineração de dados e o necessário controle de erros. Neste sentido o relatório reforça o caráter interdisciplinar da nova geração de “cientista de dados”, destacando a “combinação íntima” da estatística e da ciência da computação (em aspas no texto original):

⁹ A National Science Foundation (NSF) é uma agência federal independente criada pelo Congresso dos Estados Unidos da América em 1950 para promover o progresso da ciência. Com um orçamento anual de US\$ 8,3 bilhões (ano fiscal de 2020), é a fonte de financiamento de aproximadamente 27% do orçamento federal para pesquisa básica conduzida em faculdades e universidades dos EUA. (<https://www.nsf.gov/nsb/>)

¹⁰ A *National Academy of Sciences* (NAS) é uma sociedade privada sem fins lucrativos norte americana de estudiosos ilustres. Estabelecido por um Ato do Congresso, assinado pelo presidente Abraham Lincoln em 1863, a NAS é encarregado de fornecer consultoria independente e objetiva à nação em questões relacionadas à ciência e tecnologia.

“A análise massiva de dados não é província de um campo só, mas é antes um empreendimento totalmente interdisciplinar. Soluções para grandes problemas de dados exigirão uma combinação íntima de ideias da ciência da computação e estatística, com contribuições essenciais também necessárias das matemáticas aplicada e pura, da teoria de otimização e de várias áreas de engenharia.” (NRC, 2013, p.5. Grifo nosso. Tradução nossa.)

Em 2012, o *Big Data Research and Development Initiative*¹¹ lançado pela administração Obama, financiou diversos programas federais para o desenvolvimento da pesquisa e infraestrutura para Big Data, principalmente nas áreas de defesa e saúde¹². Como bem observou Latour (2011), a “tecnociência é um assunto militar [...] Faz parte de uma máquina de guerra e deve ser estudada como tal”. Já sobre a grande fatia de recursos reservadas à saúde na pesquisa civil, o sociólogo chama a atenção para o fato de que: “assim como a sobrevivência do organismo político, a do organismo físico é um assunto em que todos estão direta e virtualmente interessados” (LATOUR, 2011, pp.269-268). Aqui nos interessa registrar que a iniciativa aportou considerável apoio financeiro à constituição de centros de pesquisa em Big Data nas universidades norte americanas.

A Universidade da Califórnia em Berkeley (UC Berkeley) nos oferece um estudo de caso interessante. Em 2018 os Departamentos de Ciências da Computação e Ciência Estatística passaram a integrar uma nova divisão, intitulada “Divisão de Computação, Ciência de Dados e Sociedade”, onde são oferecidos cursos em ciência de dados e mantidos laboratórios de pesquisa na área. Dois fatos relevantes da pré-história da Divisão estão ligados a um mesmo ator, o Professor Michael Jordan. Jordan atuou como coordenador do *Comitte on the Analysis of Massive Data da National Academy of Sciences* e, com recursos da iniciativa Obama, fundou o AMP Lab (*Algorithms, Machines and People*)¹³, laboratório que tem desenvolvido as mais relevantes plataformas e ferramentas para processamento de Big Data¹⁴ da atualidade, utilizadas por corporações do porte da Amazon, Google e Microsoft. Recentemente, em fevereiro de 2020 a Divisão comemorou o recebimento anônimo de US \$252 mi para implantação de um Data Hub

¹¹ <https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal>

¹² O *National Institute of Health* (NIH) e a *Defense Advanced Research Projects Agency* (DARPA) responderam pelos principais programas e investimentos em Big Data. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf

¹³ Sobre o financiamento obtido pelo Data Lab: <https://data.ber> <https://news.berkeley.edu/2012/03/29/nsf-big-data-grant/>

¹⁴ Foram concebidas no AMP LAB entre outras plataformas: Apache Spark (<https://spark.apache.org/>), Apache Mesos (<http://mesos.apache.org/>), e Alluxio (<https://www.alluxio.io/>).

no campus, o valor consta como a maior doação individual da história da UC Berkeley¹⁵, sem dúvida uma demonstração de força do novo agrupamento.

Seguindo os passos do Professor Jordan, rastreamos a história de seu deslocamento do Departamento de Ciência da Computação para o Departamento de Estatística. A pesquisa em ciência estatística na UC Berkeley iniciou com a chegada do matemático e estatístico polonês Jerzy Neyman (1894-1981) em 1938, responsável pela criação do Stat Lab¹⁶. Neyman tem seu lugar de honra entre os pais da disciplina¹⁷, suas notáveis contribuições incluem a introdução do cálculo do intervalo de confiança e de procedimentos de estratificação, que conduziram aos modernos procedimentos de pesquisa por amostragem (DESROSIÈRES, 1998, p. 204). Após décadas de glória e importantes contribuições para o desenvolvimento da ciência estatística e suas aplicações, o departamento passou a enfrentar um esvaziamento de quadros a partir dos anos 1990. De acordo com o relato de professores: “Os fundadores envelheceram e começaram a se aposentar neste período[...] nós falhamos em perceber o envelhecimento desta geração até quando era quase muito tarde” (SPEED, PITMAN, RICE, 2012, p.14, tradução nossa). É nesse contexto que se dá a chegada do Professor Jordan:

Em meados da década de 1990, Breiman exortou o departamento a reconhecer a explosão de um tipo de estatística teórica e aplicada, crescendo dentro dos departamentos de Ciência da Computação, que contrastava dramaticamente com a visão do assunto mantida na (maioria) departamentos de Estatística. Ele argumentou veementemente que deveríamos reduzir a lacuna entre essas duas abordagens, fazendo nomeações de pessoas com esse *background*, geralmente denominado de *Machine Learning*. Após uma discussão considerável entre Estatística e Ciência da Computação e uma pesquisa bastante extensa, Michael Jordan foi nomeado em 1998 para fazer exatamente isso, e sua entrada em cena foi rapidamente percebida como um sucesso. (SPEED, PITMAN e RICE, 2012, p.16, tradução nossa)

Diferentemente da UC Berkeley, onde a ciência de dados estabeleceu um novo e poderoso agrupamento reunindo os departamentos de estatística e ciência da computação, um caminho diverso foi trilhado pela Universidade da Califórnia em Los Angeles (UCLA). Aqui, após intensos debates sobre a criação de um novo curso de ciência de

¹⁵ <https://data.berkeley.edu/news/largest-gift-berkeley%E2%80%99s-history-will-lead-hub-data-science>

¹⁶ Consta que “a contribuição do trabalho estatístico para o esforço dos Estados Unidos na Segunda Guerra Mundial, logo tornou-se o principal foco do laboratório, em particular pesquisa extensiva em padrões de bombardeios” (SPEED, PITMAN, RICE, 2012, p.1, tradução nossa).

¹⁷ Neyman estudou em Londres com Karl Pearson e foi amigo e colaborador de seu filho Egon Pearson. (Desrosières, 1998). Recebeu a Medalha Guy da *Royal Statistical Society* e a Medalha Nacional de Ciências dos Estados Unidos. <https://royalsocietypublishing.org/doi/10.1098/rsbm.1982.0015>

dados, a administração optou por criar um Mestrado de Estatística Aplicada subordinado ao Departamento de Estatística¹⁸. Outras iniciativas no campo da ciência de dados e big data podem ser encontradas como ramos de pesquisa em outros departamentos, mas não alcançaram a força de um novo agrupamento como verificou-se na UC Berkeley. A tardia constituição do departamento de estatística da UCLA em 1998, até então subordinado ao departamento de matemática¹⁹, talvez ajude a explicar a resistência e não adesão ao novo rótulo.

Um elemento importante para nossa investigação é precisamente a associação entre os campos da estatística e ciência da computação na gestação do novo agrupamento da ciência de dados. É interessante observar como a *American Statistical Association* (ASA)²⁰ parece acolher a interface com a ciência da computação, compreendendo-a como uma oportunidade para alavancar a ciência estatística. “*Leveraging Statistics with computer Science*” é o título de um artigo de 2014 do Grupo de Trabalho em Big Data da ASA. O artigo reforça a importância da associação entre estatística e ciência da computação para o tratamento e análise de grandes quantidades de dados; identifica a estatística como “a mais madura das ciências de dados” e especula que “a idade do Big Data será uma era de ouro para a estatística” (ASA, 2014, tradução nossa). O artigo lista uma série de áreas em transformação pelos *big datas* e para cada área discute avanços metodológicos e ferramentas da ciência estatística como: *Data visualization, Clustering, Variance decomposition, Multiple hypothesis testing, Causal inference, Pattern mining, Anomaly detection, Markov Chain Monte Carlo* (MCMC), etc.

A *Computing Research Association* (CRA)²¹, por outro lado, parece conferir menor ênfase, tanto para a ciência de dados, quanto para a ciência estatística, no conjunto de artigos produzidos sobre o tema “*Big Data and National Priorities*”²². De fato não encontramos, entre esses artigos, nenhuma menção ao termo “*data science*” ou “*data scientist*”, bem como são restritas e contraditórias as menções à ciência estatística. A associação de ciências em um ambiente de competição por recursos, certamente não se

¹⁸ <https://master.stat.ucla.edu/>

¹⁹ Sobre a história de décadas de tentativas frustradas de lograr autonomia pelo departamento: <http://statistics.ucla.edu/about/our-history/statistics-at-ucla/>

²⁰ A ASA é a segunda associação profissional mais antiga dos EUA em operação contínua, fundada em Boston em 1839.

²¹ A CRA foi constituída em 1972 e conta entre seus membros mais de 200 organizações norte-americanas ativas na pesquisa em computação: departamentos acadêmicos de ciência da computação e engenharia da computação; laboratórios e centros na indústria, governo e academia; e sociedades profissionais afiliada

²² Analisamos o conteúdo dos oito artigos publicados pelo *Computing Community Consortium* da CRA sob o tema *Big Data and National Priorities* <https://cra.org/ccc/resources/ccc-led-whitepapers/>

dá sem controvérsias, as quais como já observamos, são valiosas para o estudo da tecnociência, vamos desdobrá-las a seguir.

Em um dos artigos da CRA sobre Big Data, Eric Joel Horvitz, diretor do Laboratório de pesquisas da Microsoft, apresenta o *Machine Learning* e o *Data mining* como sub disciplinas da ciência da computação que cresceram sobre métodos estatísticos “mais antigos”:

Métodos de *Machine Learning*, construindo sobre e ampliando procedimentos estatísticos mais antigos para a inferência de dados, floresceram no final dos anos 1980 e anos 1990. Os métodos foram estimulados, e por sua vez influenciaram, o acúmulo de grandes quantidades de dados em uma variedade de campos. O *Machine Learning* e o *Data Mining* evoluíram para uma subdisciplina ativa e vigorosa da ciência da computação com uma comunidade apaixonada de pesquisadores. (HORVITZ, 2010, p.3, Tradução nossa.)

Em um outro artigo para o mesmo debate da CRA, Horvitz se refere ao emprego de métodos estatísticos por órgãos de Estado na área de segurança como algo referente ao passado, quando comparados às pesquisas de computação mais recentes, como o *active learning*, *learning from network features* e *machine learning* reunidos sob o título de “*data centric computational analysis*”:

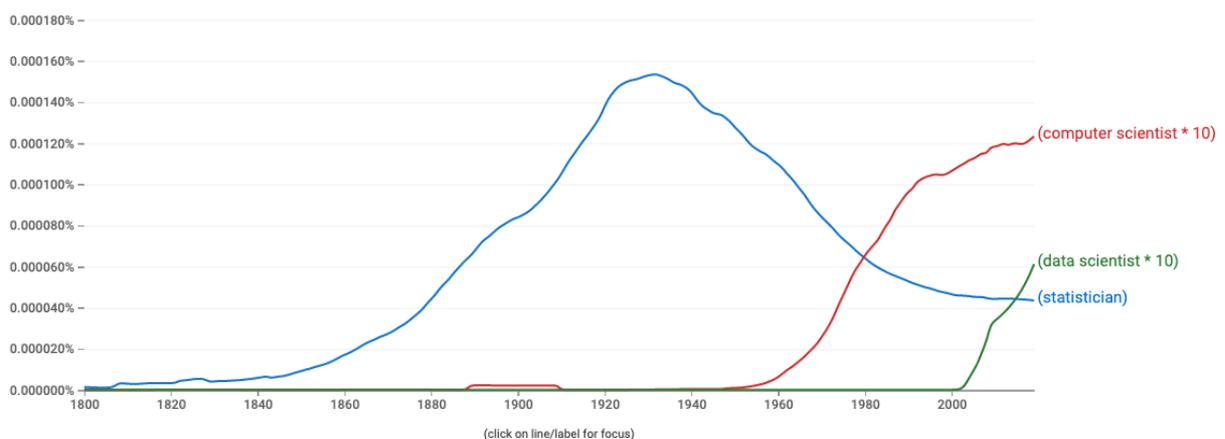
Inteligência, vigilância e reconhecimento colocam desafios que são adequados para análises computacionais centradas em dados. Métodos estatísticos para fundir informações foram, sem dúvida, usados para análise de evidências por agências governamentais engajadas em segurança. No entanto, as pesquisas de computação mais recentes sobre tópicos como o *active learning*, o *learning from network features* e o uso de indicadores de confiabilidade podem aprimorar nossa capacidade de reunir dados de fontes e sensores heterogêneos.” (Horvitz e Mitchell, 2010, pg.5. Grifo nosso. Tradução nossa.)

Um olhar panorâmico²³ sobre as ciências em questão pode, juntamente com as pistas que vimos colhendo dos atores, nos ajudar a compreender de um ponto de vista histórico, as controvérsias na intersecção destes grupos. Tanto a estatística quanto a computação são ciências que provocaram importantes rupturas e construíram novos objetos e agregados que afetaram profundamente o que hoje concebemos como social ou sociedade. Tanto do ponto de vista político e administrativo, quanto cognitivo, a emergência do pensamento estatístico e probabilístico da segunda metade do século XIX

²³ Latour (2012, p. 271) se refere a “panoramas” enquanto dispositivos que encenam a totalidade e que podem nos ser úteis desde que estejamos conscientes de que: “esses relatos coerentes e completos podem tornar-se pontos de vistas mais cegos, mais locais e mais parciais, esses panoramas devem ser estudados cuidadosamente porque propiciam a única ocasião para ver a história total como um todo. Suas visões totalizadoras não devem ser descartadas como um ato de megalomania profissional, mas sim acrescentadas, como tudo o mais, à multiplicidade de locais que queremos desdobrar. (LATOURE, 2012, p. 273)

ao início do século XX, gerou profundas transformações no Estado, nas formas de governo e nas ciências, ao introduzir novas formas de quantificação (DESROSIÈRES, 1998). A ciência da computação, que se beneficiou do impacto do pensamento estatístico e probabilístico nas ciências, e em especial na física, a partir do desenvolvimento da cibernética (WIENER, 2000), vem desde meados do século XX ampliando as possibilidades e capacidades de quantificação e acumulação de informações em níveis inimagináveis, e em escala e velocidades crescentes.

Gráfico 1 – Panorama histórico: *Computer Scientist, Data Scientist e Statistician*



Nota: Gráfico gerado com a ferramenta *Google Ngram*. As curvas representam a frequência de citações em fontes impressas dos termos “*statistician*” (estatístico), “*computer scientist*” (cientista da computação) e “*data scientist*” (cientista de dados) obtidos entre 1800 e 2019. Os termos “*data scientist*” e “*computer scientist*” foram multiplicados por uma constante (10) para melhor visualização junto ao termo “*statistician*” que apresenta maior escala de frequência de citações. Sobre o uso do *Google Ngrams* para o estudo de narrativas ver o artigo de Shiller (2017), em que o autor apresenta o conceito de economia narrativa e aplica um modelo clássico da epidemiologia (SIR) para analisar “epidemias narrativas” que seguem padrão semelhante à distribuição normal.

O gráfico 1 nos ajuda a ilustrar visualmente o panorama histórico destas ciências²⁴. Se a era dos estatísticos que iniciou no século XIX, teve seu pico por volta dos anos 1930, a era dos cientistas da computação que iniciou por volta dos anos 1950 parece viver hoje o seu apogeu. Neste sentido, os cientistas de dados que começam a despontar no século XXI, podem se apresentar como uma oportunidade para novas agregações e translações de interesses e recursos – como parece ser o caso do Professor Jordan, do AMP LAB, da UC Berkeley e da ASA - ou como um inconveniente para grupos já consolidados –como nos indicam os *papers* da CRA, o professor Horvitz no laboratório de pesquisas da Microsoft ou os estatísticos da UCLA que lutaram por tantos anos por

²⁴ Segundo Latour (2012, p.273), também permite ao leitor “equipar-se com um desejo de totalidade e centralidade”

seu status de departamento. Nos interessa principalmente saber quais são as associações mais fortes e mais fracas na construção deste híbrido.

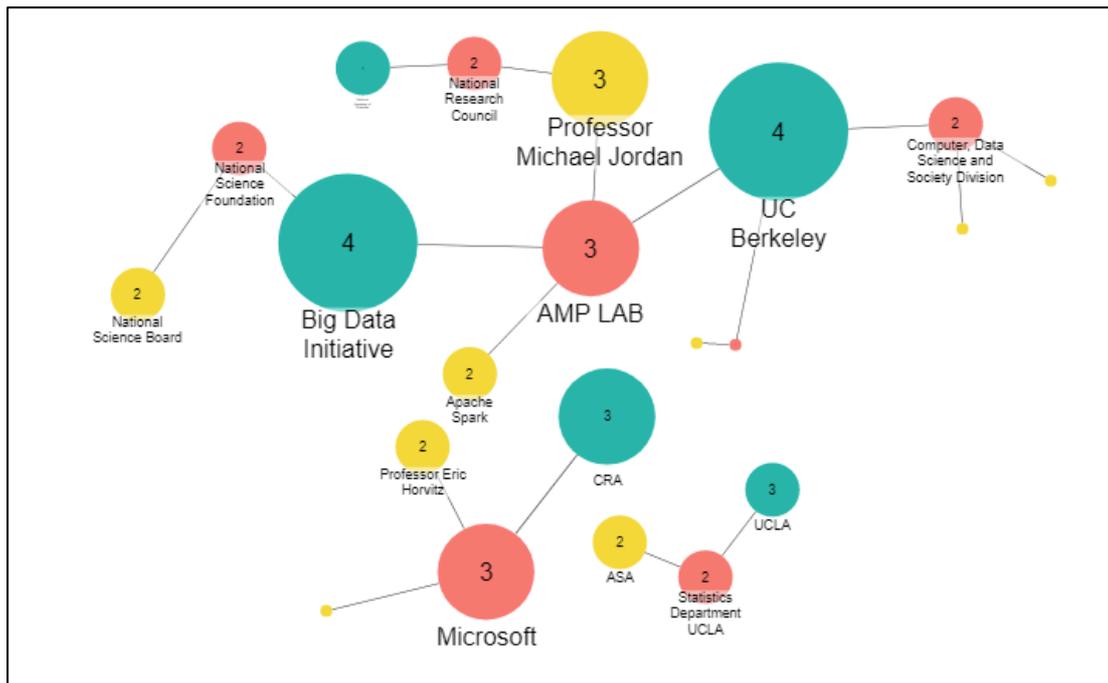
Conclusão: reunir mais mediadores

Após levantarmos esse conjunto inicial de pistas, atores e controvérsias, podemos agora retornar para nosso guia de viagem e refletir sobre alguns meios práticos apontados por Latour (2012, pp.55-58) para delinear grupos. Segundo o autor, em qualquer controvérsia a respeito da formação de grupos alguns itens sempre estarão presentes: **(1) Porta vozes**: falam pela existência de grupos e funcionam como oficiais de recrutamento; **(2) Antigrupos**: outros agrupamentos são classificados negativamente como recurso para traçar as fronteiras do grupo; **(3) Consolidação de Fronteiras**: busca de um amplo leque de recursos contra as pressões dos grupos antagônicos; **(4) Mobilização**: profissionais especializados são mobilizados para estudar e definir o grupo.

Em nosso relato guiado pela primeira fonte de incertezas – a natureza dos grupos - pudemos identificar alguns dos elementos apontados acima: Vimos o professor Michael Jordan, a Divisão de Ciência de Dados da UC Berkeley e a ASA atuando como porta-vozes de um novo híbrido; observamos o Departamento de Estatística da UCLA e o chefe de pesquisas da Microsoft, Eric Horvitz, resistindo ao movimento de agrupamento e comportando-se como potenciais antigrupos; assistimos a busca de uma parafernália de objetos científicos pelos atores para defender suas posições e consolidar as fronteiras de seus grupos, e, finalmente; identificamos uma série de profissionais, estudos, objetos e recursos sendo mobilizados para defender ou se contrapor ao novo agrupamento.

Os traços colhidos até aqui são ainda insuficientes para a descrição da rede da ciência de dados, e tampouco, é nossa intenção interromper o fluxo de controvérsias com explicações ou conclusões neste momento. Por hora, acreditamos que o experimento atendeu à nossa tarefa de exercitar as possibilidades da ANT e ao mesmo tempo indica possíveis caminhos futuros para a pesquisa.

Gráfico 2 – Representação do relato em forma de rede



Nota: Gráfico de rede utilizando a ferramentas Zoom Network Chart e Power BI.

Podemos agora retomar a frase inicial da primeira incerteza, “não há grupos, apenas formação de grupos”. Subjacente a esta afirmação está a definição fluida do social de Latour (2012), para quem os agregados sociais não são objeto de uma definição ostensiva, mas apenas de uma definição performativa: “o objeto de uma definição ostensiva permanece [...] mas o objeto de uma definição performativa desaparece quando não é mais representado – ou, caso permaneça, isso significa que outros atores entraram em cena”. Daí a importância dos atores que trazemos para a narrativa das associações e as ações que desempenham.

Entre as entidades que participam da composição de associações e agregados sociais, Latour (2012) chama a atenção para uma incerteza fundamental quanto a sua natureza íntima: funcionam como intermediários ou mediadores? Intermediários transportam significados ou forças sem transformá-los, já mediadores transformam, traduzem, distorcem e modificam o significado ou os elementos que supostamente veiculam. Do conjunto, ainda limitado, de entidades que conseguimos reunir, suspeitamos ter nos aproximado de alguns mediadores muito interessantes como os professores Jerzy Newman e Michael Jordan da UC Berkeley, que em seus laboratórios arregimentaram vultosos recursos, articularam interesses e operaram inovações e

translações nos campos da estatística e da ciência da computação, com impacto considerável sobre o mundo social.

O aprofundamento destas suspeitas demandará o desdobramento de mais controvérsias e o levantamento de mais mediadores, como, por exemplo, os atores humanos e não humanos reunidos para fortalecer posições e operar novos deslocamentos. Aqui precisaremos adentrar em outras fontes de incerteza que abrangem as ações e os objetos, as quais, como mencionamos inicialmente, serão objeto de novas pesquisas e futuros relatos. Por ora cumpre finalizar esta narrativa, apesar de sentirmos ainda ser necessário, após termos concluído o experimento, refletir sobre seus efeitos em nossa trajetória de pesquisa.

Pós-escrito: relato-rede e assimetrias

Escrever é uma tarefa arriscada, uma travessia incerta no meio de um redemoinho de dados e teorias, em que importa menos o destino e mais as aventuras no caminho. Cumprimos nossa tarefa? Logramos “aplicar” o que não se aplica? uma coisa é certa, a ferramenta que não era ferramenta, modificou o artífice. Nossa pesquisa sai transformada deste relato. Seguindo os passos da ANT, descobrimos novas perspectivas, novas formas de olhar o nosso objeto, fomos literalmente deslocados. Neste sentido a ANT se revelou um mediador poderoso.

Estamos conscientes de que apenas começamos a trilhar um caminho e que muito trabalho ainda será necessário para que nosso relato desdobre mesmo que uma mínima parte da rede da ciência de dados e se torne ele também um mediador. Para Latour (2012), os relatos textuais são o laboratório do cientista social e um bom relato ANT é aquele que tece uma rede, que desdobra os atores como redes de mediações, daí o hífen na palavra ator-rede. Rede é entendida como “uma expressão para avaliar quanta energia, movimento e especificidade nossos próprios relatos conseguem incluir. Rede é conceito, não coisa. É uma ferramenta que nos ajuda a descrever algo, não algo que esteja sendo descrito.” (LATOURE, 2012, p. 192)

Em nossa trajetória (vide o resumo inicialmente apresentado para este artigo), saímos de uma posição um tanto imobilizada pela teoria social tradicional, explicamos. Ao tratar o tema da ciência de dados queremos sobretudo abordar a questão do poder e da dominação, que intuímos estarem passando por rápidas transformações, em face do

desenvolvimento tecnológico desigual e da assimetria de recursos, que tendem a ampliar desigualdades. Também identificamos uma série de riscos à autonomia dos indivíduos frente a uma intensificação de novas formas de controle, proporcionadas pela governança por algoritmos. Essa intuição básica que herdamos da sociologia crítica, irremediavelmente nos levou a um beco sem saída: a reificação.

Se o desenvolvimento tecnológico leva a uma inevitável autonomização do social e a uma reificação dos sujeitos e de todos os âmbitos da vida social, não resta muito o que fazer. A sensação de que tudo o que tínhamos a dizer ou explicar neste campo, já estava de certa forma dito e explicado causou uma certa paralisia intelectual. Neste sentido a subversão apresentada pela teoria Ator-Rede nos apresentou um caminho para tentar compreender a mecânica concreta destas formas de dominação e de como se instauram desigualdades e assimetrias. Em uma palavra final, deixando um pouco de lado as explicações sociais e a distância crítica, para nos aproximar o máximo possível dos atores, e assim compreender os elos mais fortes e mais fracos desta rede, quem sabe possamos construir uma pesquisa com maiores chances de contribuir para modificar esse estado de coisas.

Referências Bibliográficas

AGGARWAL, C.C. **Data Mining: The Textbook**. Cham, Switzerland: Springer, 2015.

ASA. Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society. **American Statistical Association**. Jul. 2, 2014. Disponível em: <<https://www.amstat.org/asa/files/pdfs/POL-BigDataStatisticsJune2014.pdf>>. Acesso em: 10 set. 2020.

BRAMER, M. **Principles of Data Mining**. London, United Kingdom: Springer, 2007.

BRYANT, R. E. et al. From Data to Knowledge to Action: Enabling the Smart Grid. Sep. 21, 2010. **Computing Community Consortium - Computer Research Association**. Disponível em: <<https://cra.org/ccc/resources/ccc-led-whitepapers/#big-data-and-national-priorities>>. Acesso em: 22 set. 2020.

CALLON, M. **Actor-network theory: the market test**. The Sociological Review 1999. Vol. 47, Issue S1, pp. 181 – 195. May, 1999.

CALLON, M.; LAW, J.; RIP, A. **Mapping the Dynamics of Science and Technology**. London, United Kingdom: Springer, 1998

DESROSIÈRES, A. **The Politics of Large Numbers: a history of statistical reasoning**. Cambridge, Massachusetts, and London, England : Harvard University Press, 1998.

HORVITZ, T.M. From Data to Predictions and Decisions: Enabling Evidence-Based Healthcare. Sep. 16, 2010. **Computing Community Consortium - Computer Research Association**. Disponível em: <<https://cra.org/ccc/resources/ccc-led-whitepapers/#big-data-and-national-priorities>>. Acesso em: 22 set. 2020.

HORVITZ, T.M.; MICHELL, T. From Data to Knowledge to Action: A Global Enabler for the 21st Century. **Computing Community Consortium - Computer Research Association**. Sep. 11, 2010. Disponível em: <<https://cra.org/ccc/resources/ccc-led-whitepapers/#big-data-and-national-priorities>>. Acesso em: 22 set. 2020.

KUBAT, M. **An Introduction to Machine Learning**. Cham, Switzerland: Springer, 2017.

LATOUR, B. **Ciência em Ação**: como seguir cientistas e engenheiros sociedade afora. São Paulo: Ed. Unesp, 2011.

LATOUR, B. **Reagregando o Social**: uma introdução à teoria do Ator-Rede. Salvador: Edufba, Bauru: Edusc, 2012.

LAW, J. Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity; **Centre for Science Studies, Lancaster University**. Nov. 30, 2003. Disponível em <<http://www.comp.lancs.ac.uk/sociology/papers/Law-Notes-on-ANT.pdf>>. Acesso em 10 set. 2020.

LAW, J.; HASSARD, J. **Actor network theory and after**. Contemporary Sociology 30:96. Oxford: Blackwell, 1999.

NRC, National Reserach Council of the National Academies. **Frontiers in Massive Data Analysis**. Washington, D.C.: The National Academies Press, 2013.

NSB, National Science Board. **Long-Lived digital data Collections: enabling Research and education in the 21st Century**. National Science Foundation, 2005.

SHILLER,R.J. **Narrative Economics**. Cowles Foundation Discussion Paper no. 2069. Presidential address delivered at the 129th annual meeting of the American Economic Association, January 7, 2017, Chicago, IL.

SKIENA, S.S. **The Data Science Design Manual**. Cham, Switzerland: Sringer, 2017.

SPEED,T.;PITMAN J.; RICE J. A Brief History of the Statistics Department of the University of California at Berkeley. **Department of Statistics - UC Berkeley**. Dez. 2012. Disponível em: <<https://arxiv.org/pdf/1201.6450.pdf>>. Acesso em: 15 set. 2020.

VARGAS, E.V. **Antes Tarde do que Nunca**: Gabriel Tarde e a Emergência das Ciências Sociais. Rio de Janeiro: Contra Capa Livraria, 2000.

WIENER, N. **Cibernética e Sociedade**:O Uso Humano de Seres Humanos. 7a. Ed. São Paulo: Editora Cultrix, 2000.