

APLICAÇÃO DE FONTES DE DADOS ALTERNATIVAS (BIG DATA) AO PLANEAMENTO E OPERAÇÃO DE SISTEMAS DE TRANSPORTES

Filipe Viegas

VTM Global, Linda-a-Velha, Portugal, filipe.viegas@vtm-global.com

Isabel Pimenta

VTM Global, Linda-a-Velha, Portugal, isabel.pimenta@vtm-global.com

RESUMO

A crescente digitalização do transporte gera dados que são armazenados em diferentes fontes, cujo processamento e análise permitem otimizar a gestão e operação das infraestruturas e veículos, promovendo a sua maior eficiência, resiliência e sustentabilidade nos sistemas de mobilidade. Esta publicação baseia-se na experiência dos autores em trabalhos realizados em diferentes sistemas de transporte situados em diversos países, e explora a aplicação de dados provenientes de Big Data, como sensores, veículos conectados, sistemas de GPS, telefonia móvel, Bluetooth e sistemas de bilhética, para a realização de análises no planeamento estratégico e operacional de transportes.

São apresentados casos de estudo em geografias como Portugal e México, que demonstram como estes dados suportam a gestão, otimizam operações oferecendo um apoio quantificado na tomada de decisão na priorização de investimento ou na otimização de operações. Estas fontes oferecem maior abrangência temporal quando comparados com métodos clássicos de recolha de dados convencionais no terreno, assegurando precisão na identificação de padrões de mobilidade, sustentando a priorização de intervenções estratégicas e investimentos. Este trabalho destaca o papel essencial das fontes alternativas de dados na conceção de sistemas de transporte adaptados aos desafios atuais e futuros da mobilidade.

PALAVRAS-CHAVE: Fontes Alternativas de Dados, Otimização de Planeamento e Operação de Transportes, Big Data em Mobilidade, Digitalização no Transporte, Decisões Estratégicas Baseadas em Dados.

ÍNDICE

1	Introdução	3
2	Visão Geral de Fontes de Dados Big Data em Transportes	4
2.1	Fontes baseadas em dados de GPS.....	4
2.2	Fontes baseadas em dados de rede móvel	5
2.3	Fontes baseadas em simulação de mobilidade	6
2.4	Fontes associadas a transportes públicos (dados de bilhética eletrónica, contadores automáticos de passageiros, GPS em veículos).....	8
3	Aplicações Práticas com Fontes de Dados Big Data	9
	REFERÊNCIAS BIBLIOGRÁFICAS	10

1 Introdução

A rápida evolução tecnológica que se tem verificado nas últimas décadas transformou profundamente a forma como recolhemos, armazenamos e analisamos dados. Na área dos transportes, esta tendência tem sido particularmente evidente, com a proliferação de dispositivos conectados — desde telemóveis a sensores GPS instalados em frotas automóveis — que produzem volumosos registos georreferenciados de deslocações. Esses dados massivos, frequentemente agrupados sob o termo “Big Data”, oferecem novas oportunidades para compreender de forma detalhada os padrões de mobilidade em áreas urbanas e regionais, ultrapassando as limitações dos métodos tradicionais de recolha de dados, como contagens manuais de tráfego, inquéritos pontuais e levantamentos exaustivos em campo.

Tradicionalmente, as análises de tráfego e a projeções de procura em sistemas de transportes baseavam-se em processos demorados e dispendiosos, como a realização de inquéritos origem-destino (O/D) junto dos utilizadores, a recolha de dados de contagens de tráfego em pontos específicos da rede ou ainda a observação direta do desempenho dos sistemas de transportes. Embora continuem a desempenhar um papel relevante — sobretudo em termos de validação e calibração de modelos ou de recolha de informação detalhada em zonas específicas — tais métodos apresentam várias limitações: a atualização não é contínua, os custos associados são geralmente elevados e a quantidade de informação recolhida é de reduzida amplitude espacial e/ou temporal para caracterizar padrões de mobilidade em larga escala.

Em contraste, fontes Big Data aplicadas aos transportes oferece a possibilidade de obter, de forma expedita, informação sobre deslocações em grande escala, com atualizações em intervalos frequentes (inclusive em tempo quase real), permitindo uma visão dinâmica do comportamento do sistema de transportes. As fontes variam significativamente: podem advir de dispositivos GPS instalados em veículos, de registos de telemóveis e/ou de aplicações móveis, de validações em sistemas de bilhética eletrónica (no caso dos transportes públicos), de contadores automáticos de veículos ou passageiros e de diversas outras plataformas que agregam e analisam estes dados. No entanto, como qualquer tecnologia emergente, o recurso a fontes Big Data em transportes também requer prudência na interpretação dos resultados: as diferenças de cobertura de amostra, a precisão dos dados e as metodologias de processamento empregues pelos fornecedores podem influenciar a qualidade e a aplicabilidade dos resultados obtidos.

Nesse contexto, a VTM Global tem vindo a integrar de forma progressiva diversas fontes de dados Big Data na sua prática de consultoria de transportes e mobilidade, tanto em Portugal como noutros países. A experiência acumulada permitiu identificar as potencialidades e desafios de cada abordagem, ajudando a orientar a seleção de dados consoante o tipo de estudo em causa — sejam diagnósticos de tráfego em redes rodoviárias ou análises de procura em sistemas de transporte público.

Neste artigo, a partir da experiência obtida, apresenta-se uma visão geral de diferentes fontes de Big Data aplicáveis ao planeamento e à operação dos sistemas de transportes, descrevendo para cada uma delas as principais características e casos de uso. Além disso, é dado enfoque específico a Big Data em sistemas de transportes públicos, através de dados de bilhética eletrónica, destacando o seu papel fundamental no planeamento e na gestão e otimização de sistemas de transporte coletivo. Numa secção subsequente, discutem-se as metodologias de processamento envolvidas e as diferenças-chave entre estas fontes, com o objetivo de orientar profissionais do setor na sua correta aplicação. Finalmente, são apresentados exemplos práticos de projetos em Portugal e no México, recorrendo à plataforma TomTom MOVE, evidenciando as vantagens do uso de Big Data na tomada de decisões.

Conclui-se, pelos casos apresentados, que a adoção de fontes de dados alternativas proporciona benefícios evidentes aos decisores públicos e privados que planeiam e gerem infraestruturas e sistemas de transportes, sendo, no entanto, crucial considerar as especificidades de cada fonte, assim como assegurar a compatibilização com métodos tradicionais de recolha de dados. As reflexões do presente artigo contribuem assim para a divulgação e ampliação de boas práticas no planeamento e operação de sistemas de transportes.

2 Visão Geral de Fontes de Dados Big Data em Transportes

Nos últimos anos, o setor dos transportes passou a dispor de vastas fontes de dados Big Data, possibilitando análises e decisões mais informadas e quantificadas no planeamento e operação de sistemas de mobilidade. Estas fontes podem ser agrupadas em diferentes categorias conforme a origem dos dados. Nos sub-capítulos seguintes apresenta-se uma visão geral das principais fontes – desde *dados GPS* recolhidos em veículos automóveis, passando por *dados de redes móveis*, por *modelos de simulação de mobilidade*, até *dados gerados pelos sistemas de transporte público*. Cada grupo oferece perspetivas complementares, e a integração de todas permite uma compreensão completa dos padrões de deslocação e do desempenho dos sistemas de transporte.

2.1 Fontes baseadas em dados de GPS

Nesta tipologia de fonte de dados proveniente de localização GPS provenientes de veículos automóveis e dispositivos móveis é possível inferir condições de circulação tráfego e padrões de mobilidade. Empresas como a **TomTom**, **INRIX** ou **StreetLight Data** recolhem massivamente *trajetos GPS* de frotas automóveis, aplicações de navegação e outros dispositivos. Esses dados – muitas vezes designados “*floating car data*” – representam posições e velocidades de veículos ao longo do tempo, captadas por satélite. Estima-se que os serviços da TomTom recolhem dados de cerca de 1 em cada 5 veículos em circulação globalmente, cobrindo mais de 80 países, com um fluxo diário na ordem dos 61 mil milhões de pontos GPS registados¹. De modo semelhante, a INRIX agrega diariamente aproximadamente 1,5 mil milhões de registos GPS de mais de 300 milhões de veículos conectados, graças a parcerias com construtores automóveis e apps móveis². Esta recolha contínua e ubíqua de posições permite monitorizar fluxos de tráfego na malha rodoviária praticamente em tempo real.

As principais características e vantagens destas fontes incluem a elevada precisão espacial e temporal e a cobertura alargada da rede viária. As localizações GPS têm precisão de poucos metros, possibilitando identificar em que via ou segmento o veículo circula, e são obtidas com frequência de segundos, permitindo reconstruir *trajetos completos e velocidades instantâneas*. Outra vantagem significativa é o facto de dispensarem infraestrutura fixa: os veículos e smartphones funcionam como sensores móveis, evitando a instalação de estações de contagem ou câmaras ao longo das vias. Isto torna a solução de rápida implementação, como sublinha a TomTom ao promover plataformas “*infrastructure-less*”, já disponíveis em dezenas de países³. Adicionalmente, muitas destas fontes contam com extensos arquivos históricos – por exemplo, a TomTom e a INRIX acumulam mais de uma década de dados – o que permite análises de tendências de longo prazo (como evolução do congestionamento) com grande detalhe espacial⁴. Em resumo, os dados GPS fornecem uma imagem pormenorizada do desempenho viário (velocidades, tempos de viagem, atrasos) e dos caminhos efetivamente percorridos pelos utilizadores.

Os dados de GPS são amplamente utilizados tanto no planeamento estratégico como na gestão operacional do tráfego rodoviário. Ao nível de planeamento, permitem identificar os principais pontos de congestionamento que permitem o desenvolvimento de cenários de melhoria: por exemplo, simulando o efeito de uma nova via ou de uma alteração de semáforos, usando históricos de fluxos e velocidades. Também suportam a calibração de modelos de tráfego, fornecendo matrizes origem-destino inferidas e distribuição de percursos escolhidos.

¹<https://www.itsinternational.com/daily-news/its-world-congress-2024/tomtom-moves-traffic-data-new-heights#:~:text=TomTom%20traffic%20data%20also%20enables,cost%2C%20and%20environmental%20impact%20data>

²<https://canvasbusinessmodel.com/pt/products/inrix-swt-analysis?srsid=AfmBOop180HMBG4p6eAaF7u9E2eXmDrZS29Hv7xKBsgQ5dkC-frVmYes#:~:text=A%20Inrix%20construiu%20uma%20base,de%20300%20milh%C3%B5es%20ve%C3%A9culos%20conectados>

³<https://www.itsinternational.com/daily-news/its-world-congress-2024/tomtom-moves-traffic-data-new-heights#:~:text=While%20many%20hardware,instantaneously%20in%20over%2080%20countries>

⁴<https://www.itsinternational.com/daily-news/its-world-congress-2024/tomtom-moves-traffic-data-new-heights#:~:text=TomTom%20traffic%20data%20also%20enables,cost%2C%20and%20environmental%20impact%20data>

Na operação diária, são usados em *centros de gestão de tráfego* para detetar incidentes ou filas em tempo real, acionando painéis de mensagem variável ou informações aos condutores. Ferramentas comerciais como TomTom MOVE disponibilizam relatórios prontos-a-utilizar baseados nestes dados; por exemplo, é possível avaliar como o trânsito em torno de um estádio varia em dias de jogo face a dias normais⁵, ou quantificar o impacto de medidas excepcionais (como restrições durante a pandemia) nos volumes de viagens. A TomTom publica um índice anual de congestionamento cobrindo mais de 400 cidades no mundo, calculado a partir destes grandes repositórios de dados de velocidades⁶. Já a INRIX fornece *scorecards* de tráfego comparando cidades e identificando os corredores com maiores atrasos. Em Portugal, concessionárias de autoestradas e entidades responsáveis pela gestão de infraestruturas rodoviárias, tais como a IP ou os municípios, têm recorrido a dados GPS para monitorizar tempos de viagem e níveis de serviço nas vias, detetando em que períodos se excedem os tempos de percurso contratuais ou onde há necessidade de expansão ou revisão de tempos semaforicos, entre outras alterações necessárias. Em suma, as fontes baseadas em GPS tornaram-se essenciais para um gestão pró-ativa do tráfego rodoviário, combinando abrangência espacial com detalhe temporal.

Embora os dados GPS ofereçam excelente precisão e informação de tráfego automóvel, dependem de uma frota equipada ou de utilizadores de apps específicas. Para obter uma visão mais abrangente da mobilidade – incluindo viagens que não se realizam em transporte individual ou áreas fora da rede viária principal – recorre-se a fontes baseadas em dados de redes móveis, descritas a seguir.

2.2 Fontes baseadas em dados de rede móvel

Este segundo grupo de fontes de dados explora a informação produzida pelas redes de telemóvel para inferir padrões de deslocação. Aqui, empresas como a **Nommon**, **Kido Dynamics** ou **Teralytics** estabelecem parcerias com operadores de telecomunicações para tirar partido de registos gerados pelos telemóveis dos utilizadores. Sempre que um telefone se conecta a uma antena, inicia ou termina uma chamada, ou mesmo periodicamente em segundo plano, são produzidos *logs* de rede contendo identificadores da torre e carimbos de tempo (*timestamps*). A partir destes registos – devidamente anonimizados e agregados para proteger a privacidade – é possível reconstruir de forma aproximada os movimentos dos aparelhos no território. Em essência, cada telemóvel funciona como um *tracer* da pessoa que o transporta, indicando sucessivas localizações ao longo do dia.

Com técnicas de inteligência artificial, estes fornecedores extrapolam os dados amostrais que recolhem para representar a população inteira de uma região. No caso da Alemanha, a Teralytics baseia as suas análises nos movimentos de mais de 30 milhões de dispositivos móveis, extrapolados de forma a representar toda a população do país⁷. Essa abordagem assegura que mesmo quem não utiliza aplicativos de navegação ou dispositivos GPS dedicados é incluído na *fotografia* da mobilidade, assumindo que a maioria das deslocações são realizadas por indivíduos portadores de um telemóvel.

As características e vantagens dos dados de rede móvel diferem em alguns aspetos das fontes GPS. A maior vantagem é sem dúvida a elevada taxa de amostragem populacional: como referido, abrange-se um espectro muito amplo de utilizadores – condutores, passageiros de transporte público, peões, ciclistas – qualquer indivíduo com um telemóvel emitindo sinal. Isto permite obter uma visão macroscópica dos fluxos de viagens em área urbana ou de âmbito regional/nacional, incluindo origens e destinos de deslocações com representatividade estatística adequada. Releva-se que, os dados são recolhidos de forma contínua e passiva, sem depender da participação ativa do utilizador (não requerendo a instalação de aplicativos específicos nem equipamento novo). Em termos de precisão, a granularidade espacial é inferior àquela obtida através de GPS

⁵ <https://www.tomtom.com/products/tomtom-move/>

⁶ <https://www.itsinternational.com/daily-news/its-world-congress-2024/tomtom-moves-traffic-data-new-heights#:~:text=TomTom%20traffic%20data%20also%20enables,cost%2C%20and%20environmental%20impact%20data>

⁷ <https://teralytics.net/matrix/mobility-trends/#:~:text=means%20of%20transport%20we%20move,represent%20the%20entire%20German%20population>

– tipicamente da ordem de dezenas a algumas centenas de metros, condicionada pela densidade da rede de antenas – mas suficiente para análises zonais (entre bairros ou freguesias/concelhos) e identificação das principais linhas de desejo das deslocações. A dimensão temporal varia conforme o tipo de registo utilizado: *CDRs* (*Call Detail Records*) tradicionais registam eventos de chamada ou SMS e têm intervalos irregulares, enquanto dados de sinalização ou de aplicativos podem fornecer *pings* frequentes (minuto a minuto). Em qualquer dos casos, é possível obter séries temporais extensas, permitindo analisar tendências sazonais ou variações horárias dos padrões de mobilidade. Uma vantagem importante é a possibilidade de analisar períodos prolongados do passado: operadoras móveis preservam registos históricos, possibilitando análises *ex post* de eventos. A título de exemplo, é possível analisar padrões de mobilidade antes, durante e após uma intervenção física numa infraestrutura ou serviço ou de política de alteração de tarifário de estacionamento ou de transportes públicos, alteração de sentidos de circulação, zonas de restrição à circulação automóvel, entre outras, que se revela muito útil na avaliação de impactos. Outra característica que se destaca de forma muito positiva é que, apesar do volume massivo de dados brutos, o resultado é apresentado de forma anonimizada e agregada (fluxos entre zonas, contagens em eixos, etc.), cumprindo as normas de privacidade europeias atualmente em vigor.

As fontes baseadas em redes móveis demonstram assim grande utilidade no planeamento de transportes e ordenamento do território. A das utilizações mais comuns é a construção de matrizes Origem-Destino (O/D) em larga escala, alimentando modelos de previsões de procura em infraestruturas e sistemas de transportes. Em regiões onde não estão disponíveis dados de inquéritos à mobilidade, os dados de telemóveis permitem obter estimativas robustas de volumes de deslocações entre zonas previamente definidas, como seja uma deslocação de uma zona de residência para uma zona de emprego, por exemplo, em diferentes períodos do dia. Autoridades de transportes têm utilizado estes dados para redesenhar redes de transporte público ou identificar necessidades de novos serviços – detetando, por exemplo, corredores de elevada procura entre duas áreas mal servidas pela oferta existente. Também no planeamento rodoviário, concessionárias utilizam dados de mobilidade de operadores móveis para identificar origens/destinos do tráfego em autoestradas, ajudando a definir estratégias comerciais (descontos regionais, novos acessos) e prever a procura futura em eixos interurbanos. Adicionalmente, estes dados demonstram ser relevantes na monitorização de eventos excepcionais: durante a pandemia foi possível quantificar a queda de mobilidade quase em tempo real e orientar ações (por ex., ajustar oferta de transporte público em função da redução de passageiros). Destacam-se casos de utilização em algumas cidades europeias que avaliaram o impacto de eventos de grande escala (festivais, maratonas, etc.) analisando a afluência e os percursos dos participantes através de dados agregados de telemóveis. Em termos operacionais, empresas como a Nommon e Kido Dynamics desenvolveram *dashboards* para entidades e de infraestruturas ou serviços de transportes, elaborando *mapas de calor* de deslocações e comparando padrões dia-a-dia – permitindo, por exemplo, a gestão de tráfego em tempo real consoante a procura observada (p. ex. gestão de semáforos adaptativa a volumes de tráfego). Em síntese, as fontes baseadas em dados de rede móvel oferecem uma perspetiva macroscópica e multimodal da mobilidade, complementando as fontes de dados GPS.

Apesar de todo o valor dos dados observacionais (GPS e redes móveis), existem situações em que se pretende estimar cenários hipotéticos ou variáveis não diretamente medidas (como o motivo da viagem, ou a mobilidade futura num plano estratégico). Nesses casos, recorre-se a ferramentas de simulação de mobilidade baseadas em Big Data, as quais são abordadas no sub-capítulo seguinte.

2.3 Fontes baseadas em simulação de mobilidade

As fontes desta categoria caracterizam-se por gerarem dados sintéticos de mobilidade através de modelos calibrados com Big Data, oferecendo uma visão *virtual* do comportamento de viagens. Ao contrário das fontes puramente empíricas, aqui combina-se uma variedade de dados de entrada e técnicas de modelação para simular as deslocações de uma população inteira. Um exemplo destacado é a plataforma Replica: desenvolvida originalmente pela Sidewalk Labs, que integra *dados anonimizados de localização de telemóvel* (provenientes de aplicativos móveis), dados demográficos de fontes públicas (p. ex. censos) e informações de transportes

públicos, entre outras, para construir uma “população sintética” de viajantes⁸. Esta população virtual corresponde estatisticamente à real (em termos de distribuição de residências, empregos, perfis socioeconômicos, etc.), mas não representa indivíduos específicos – garantindo privacidade. Com base no conjunto de informação referida, o Replica modela as viagens de cada indivíduo sintético ao longo de um dia típico, atribuindo origens, destinos, modos de transporte e motivos às deslocamentos. O resultado é uma simulação detalhada que indica, a título de exemplo, quantos indivíduos estão numa dada via a determinada hora, como se deslocam (a pé, de bicicleta, automóvel, transporte público, etc.) e com que motivo (trabalho, escola, lazer, etc.). Estes modelos aproximam-se de um “mundo virtual” do transporte, onde se pode avaliar qualquer métrica de mobilidade de forma consistente. A StreetLight Data, embora conhecida sobretudo pelo uso de dados reais, também se enquadra nesta categoria uma vez que utiliza modelos de aprendizagem automática para *extrapolar e sintetizar* informações de mobilidade a partir de dados massivos dispersos. Esta plataforma traduz *trilhões de pontos de dados* de veículos conectados e dispositivos IoT em indicadores úteis sobre como e por onde os indivíduos se deslocam⁹. Na prática, o StreetLight desenvolve uma simulação contínua do tráfego e deslocamentos, inferindo matrizes O/D, rotas preferenciais e quota modal a partir dos vestígios digitais, mesmo onde não há observações diretas suficientes. Em ambos os casos – Replica e StreetLight – têm, como objetivo disponibilizar dados prontamente utilizáveis, de forma flexível, para qualquer segmento espacial ou temporal de interesse, que representam “contagens virtuais” ou “inquéritos sintéticos” gerados a partir do Big Data.

Entre as características e vantagens destas abordagens de simulação, destaca-se a capacidade de preencher lacunas e responder a questões de “e se?” no planeamento. Por exemplo, se numa certa zona faltam dados empíricos (poucos veículos com GPS ou poucas antenas móveis), com um modelo sintético é possível estimar as trocas de viagens naquela área baseando-se nos comportamentos gerais conhecidos e em variáveis socioeconômicas. Também permite imputar atributos às viagens – como o motivo ou a cadeia de atividades – algo que não é possível obter a partir de dados brutos de GPS ou telemóvel. Outra vantagem é a possibilidade de testar cenários hipotéticos: é possível simular, por exemplo, o efeito de se introduzir uma nova linha de metro ou de se pedonalizar uma avenida, antecipando como a “população sintética” reajustaria as suas rotinas de viagem. Desta forma, é possível disponibilizar informação relevante aos decisores através de uma ferramenta para *avaliação prospectiva* de políticas de mobilidade em momento anterior à sua implementação. Importa realçar que, apesar de serem modelos, estes processos estão ancorados em dados reais, o que lhe confere credibilidade e precisão. A calibração regular com dados atualizados garante que a simulação reflita mudanças entretanto ocorridas nos padrões de mobilidade (por exemplo, alterações pós-pandemia ou adoção de novos modos como trotinetes). Por fim, há um benefício de privacidade: ao trabalhar com agentes sintéticos agregados, sendo eliminado o rastreio de indivíduos reais, mitigando riscos de divulgação de informação pessoal enquanto se obtêm resultados fiáveis a nível agregado¹⁰.

Importa destacar que estas soluções atualmente apenas se encontram disponíveis no mercado norte-americano onde várias cidades e regiões já adotaram estas fontes para apoio em planeamento de transportes. Áreas metropolitanas como Kansas City, Chicago ou Portland colaboraram com o Replica para obter um retrato integral e abrangente das viagens urbanas e testar virtualmente intervenções no sistema de transportes. A título de exemplo, refira-se a cidade de Portland que utilizou o modelo para compreender os padrões de deslocamento dos residentes e avaliar as possíveis mudanças com a entrada de um novo sistema BRT (Bus Rapid Transit) em momento anterior à sua implementação, ajustando o projeto de acordo com os resultados simulados. Permitiu ainda quantificar quantas viagens em transporte individual poderiam ser substituídas por deslocamentos a pé ou de bicicleta se fossem implantadas infraestruturas adequadas, produzindo argumentos objetivos para investimentos em mobilidade ativa. Já a plataforma StreetLight tem sido amplamente empregue por consultores de transportes e agências públicas para estudos de procura mais céleres. Em projetos de planeamento de redes viárias e de transportes, o StreetLight permite obter matrizes O/D e indicadores de

⁸<https://venturebeat.com/business/sidewalk-labs-spins-out-replica-to-help-city-planners-create-virtual-populations-with-big-data/#:~:text=In%20a%20nutshell%2C%20Replica%20leverages,along%20the%20lines%20of%20SimCity>

⁹<https://www.streetlightdata.com/travel-data-analytics/#:~:text=The%20travel%20analytics%20provided%20by,air%2C%20bus%2C%20and%20rail%20travel>

¹⁰<https://venturebeat.com/business/sidewalk-labs-spins-out-replica-to-help-city-planners-create-virtual-populations-with-big-data/#:~:text=In%20a%20nutshell%2C%20Replica%20leverages,along%20the%20lines%20of%20SimCity>

mobilidade praticamente instantâneos, sem a necessidade de realizar inquéritos de tráfego, processo mais demorado. Uma aplicação que se destaca é o caso do planeamento do serviço **Uber Air** (transporte aéreo urbano): através dos dados e modelos do StreetLight, foi possível estimar a procura para viagens aéreas urbanas com base em padrões terrestres existentes, identificando que corredores poderiam captar mais utilizadores¹¹. Do ponto de vista de eficiência, estas ferramentas proporcionam uma forma custo-efetiva de avaliar padrões O/D em grande escala, com enviesamento reduzido e num período de tempo significativamente inferior ao tempo necessário recorrendo a levantamentos e subsequente modelação tradicionais. Ou seja, com simulações baseadas em Big Data é possível obter em semanas ou dias, informação que pelos métodos clássicos exigiria meses de trabalho de campo. Em conclusão, as fontes baseadas em simulação de mobilidade permitem às equipas de planeamento antecipar e simular virtualmente alterações no sistema de transportes, reduzindo a incerteza na tomada de decisão complementando os dados observados com um exercício preditivo.

2.4 Fontes associadas a transportes públicos (dados de bilhética eletrónica, contadores automáticos de passageiros, GPS em veículos)

Os sistemas de transporte público produzem uma grande quantidade de dados operacionais diariamente. Entre as principais fontes de Big Data neste domínio estão: os dados de bilhética eletrónica (registos das validações de títulos de transporte em torniquetes ou validadores dentro dos veículos ou nos cais de embarque), os contadores automáticos de passageiros (APC, do inglês *Automatic Passenger Counters*) instalados nos veículos, e os sistemas de localização GPS/AVL presentes na frota (Autocarros, comboios, metros, etc.). Cada validação de bilhete eletrónico realizada – seja com um cartão *smartcard* (ex.: passe Navegante, Andante) ou bilhete via app/QR code – gera um registo contendo o identificador do cartão (que pode ser anonimizado), a linha/estação, e a hora em que o passageiro entrou no sistema. Em sistemas de transporte fechados (como metro ou comboio urbano com barreiras de entrada/saída), obtém-se tanto a estação de entrada como a de saída, permitindo identificar o trajeto completo do passageiro. Em sistemas abertos (como autocarros urbanos ou suburbanos), o registo é realizado apenas no embarque. Esses registos alimentam bases de dados extensas: por exemplo, milhões de validações mensais numa grande cidade, armazenando informação detalhada sobre a utilização de cada linha e percurso¹². Com recurso a programação é possível construir a cadeia de viagem, com as sucessivas validações num período temporal pré-definido. É ainda possível construir matrizes O/D considerando os diversos modos de transporte em presença, desde que com integração do sistema de bilhética, inferindo o destino da viagem tomando com origem da validação seguinte o destino da viagem anterior, num período temporal pré-estabelecido.

Por outro lado, os APCs consistem em sensores (tipicamente infravermelhos nas portas dos veículos, ou câmaras digitais com algoritmos de contagem) que contabilizam entradas e saídas de passageiros em cada paragem. Cada veículo equipado com APC regista, para cada paragem número de passageiros entrados e saídos, permitindo aferir a ocupação do veículo em cada momento.

Por fim, os sistemas AVL (Automatic Vehicle Location) com GPS disponibilizam a posição em tempo real de cada veículo, registando coordenadas em curtos intervalos temporáveis, tipicamente cinco segundos. Estes dados são utilizados para aferição de cumprimento contratual s comparando-os com os percursos e horários planeados. Assim, geram-se dados sobre atrasos, adiantamentos, velocidade média por troço, tempos de percurso efetivos, etc., para cada viagem realizada. Todos estes dados – bilhética, contagens e GPS – são em geral integrados nos centros de controlo das empresas de transporte e armazenados para análise histórica, sendo disponibilizados às Autoridades de Transportes, o que lhes permite um conhecimento detalhado da mobilidade nestes sistemas.

¹¹ <https://learn.streetlightdata.com/uber-air>

¹² <https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997258221/Dissertacao%20Andre%20Caseiro.pdf#:~:text=Mas%20a%20grande%20vantagem%20da,monitoriza%C3%A7%C3%A3o%20e%20gest%C3%A3o%20da%20opera%C3%A7%C3%A3o>

As características e vantagens destas fontes residem na fidelidade com que retratam a utilização real do sistema de transporte e a sua performance. Diferem das fontes anteriores por serem *exclusivas dos operadores* (ou autoridades de transporte) e não estarem disponíveis de forma aberta em bruto, sendo, no entanto, o seu valor analítico muito relevante. A bilhética eletrónica, por exemplo, tem o potencial de proporcionar informação valiosa aos operadores sobre a forma como os passageiros utilizam o serviço. Estes sistemas registam quantos passageiros foram transportados em cada viagem, onde embarcaram (e eventualmente desembarcaram) e a que horas, entre outros dados úteis para a *monitorização e gestão da operação*. Através do tratamento destes registos, é possível construir matrizes origem-destino dos passageiros, elaborar mapas de carga dos veículos e conhecer em detalhe os padrões de mobilidade dos utilizadores. Esta informação estruturada e detalhada permite ajustar a oferta à procura – por exemplo, identificando linhas ou horários de pico onde é necessário reforço de veículos, ou, inversamente, períodos de baixa utilização onde se podem otimizar recursos. Permite ainda reestruturar serviços proporcionando ligações diretas em pares O/D relevantes em que as deslocações impõem transbordo, melhorando a experiência de utilização. Os contadores automáticos de passageiros cumprem um papel similar, inclusive captando dados de utilizadores que não validam bilhete (por ex., em linhas gratuitas ou em fraudes), garantindo que a contagem de passageiros é o mais completa possível. Já os dados de GPS dos veículos permitem quantificar indicadores de desempenho em tempo real ou quase, como a *pontualidade* (percentagem de viagens cumpridas no horário previsto) ou o *intervalo efetivo entre autocarros*. Com esta informação, os operadores e/ou as Autoridades de Transportes podem identificar e quantificar atrasos e irregularidades intervindo atempadamente (acionando autocarros de reserva, reencaminhando veículos, informando passageiros de previsões de espera, etc.). No pós-operação, os históricos de GPS permitem calcular, por exemplo, a *velocidade média por corredor* e identificar pontos sistemáticos de atraso, fundamentando melhorias em infraestruturais ou na de tráfego rodoviário. Em síntese, a vantagem destes dados reside em permitirem a medição direta do desempenho do serviço e do comportamento do passageiro, com elevado nível de detalhe e abrangendo virtualmente 100% das viagens realizadas no sistema (dado que quase todos os passageiros validam título ou são contados e todos os veículos dispõem de GPS).

Para os engenheiros de transportes e decisores públicos, estas fontes de informação representam uma evolução substancial: substitui-se a dependência exclusiva de inquéritos pontuais e modelos teóricos por uma gestão orientada por dados objetivos e em tempo real, levando a um planeamento mais responsivo e a uma operação mais eficiente bem como a um planeamento mais fundamentado e eficaz.

3 Aplicações Práticas com Fontes de Dados Big Data

[WIP]

REFERÊNCIAS BIBLIOGRÁFICAS

- AirSage. (2020). AirSage Data for Transportation Planners. Disponível em: <https://cdn2.hubspot.net/hubfs/4944247/eBooks/AirSage%20Data%20for%20Transportation%20Planners.pdf>. Acedido em: 11 de abril de 2025.
- AirSage. (2023). How Mobile Location Data Is Revolutionizing the Transportation Industry. Disponível em: <https://airsage.com/mobile-location-data-is-revolutionizing-the-transportation-industry/>. Acedido em: 10 de abril de 2025.
- Bessa, P. F. M. (2022). *Transportation management in an era of big data: from data to knowledge*. Dissertação de Mestrado, Faculdade de Engenharia da Universidade do Porto, Porto.
- Canvas Business Model. (s.d.). INRIX SWOT Analysis. Disponível em: <https://canvasbusinessmodel.com/pt/products/inrix-swot-analysis>. Acedido em: 11 de abril de 2025.
- Caseiro, A. (2019). *Ferramentas de Big Data para Monitorização e Gestão da Operação*. Dissertação de Mestrado, Instituto Superior Técnico, Lisboa. Disponível em: <https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997258221/Dissertacao%20Andre%20Caseiro.pdf>. Acedido em: 11 de abril de 2025.
- Crist, P. (2015). *Big Data and Transport: Understanding and Assessing Options*. Discussion Paper 2015/06. International Transport Forum, OCDE, Paris.
- INRIX. (s.d.). INRIX Roadway Analytics – plataforma de análise de tráfego. Disponível em: <https://www.inrix.com/products/roadway-analytics/>. Acedido em: 11 de abril de 2025.
- INRIX. (2023). INRIX Roadway Analytics – Improving transportation planning and mobility across cities. Disponível em: <https://inrix.com/products/roadway-analytics/>. Acedido em: 10 de abril de 2025.
- ITS International. (s.d.). TomTom moves traffic data to new heights. Disponível em: <https://www.itsinternational.com/daily-news/its-world-congress-2024/tomtom-moves-traffic-data-new-heights>. Acedido em: 11 de abril de 2025.
- Kido Dynamics. (s.d.). Making Smarter Decisions with Data. Disponível em: <https://www.kidodynamics.com/>. Acedido em: 11 de abril de 2025.
- Kido Dynamics. (2023). *Kido: Big Data Solutions for Mobility* (Documento técnico interno).
- Nommon. (s.d.). Mobility Insights – Best-in-class origin-destination matrices from mobile network data. Disponível em: <https://www.nommon.es/products/mobility-insights/>. Acedido em: 11 de abril de 2025.
- O Estado de S. Paulo. (2023). Por que os dados do seu celular podem revolucionar o transporte? Disponível em: <https://www.estadao.com.br/brasil/por-que-os-dados-do-seu-celular-podem-revolucionar-o-transporte/>. Acedido em: 10 de abril de 2025.
- Parente, R. G. A. (2021). *Caracterização dos padrões de mobilidade dos usuários de transporte público por ônibus em Fortaleza através do big data de bilhetagem eletrônica*. Dissertação de Mestrado em Engenharia de Transporte, Universidade Federal do Ceará, Fortaleza.
- Replica. (s.d.). The Replica Platform: A holistic picture of mobility. Disponível em: <https://www.replicahq.com/platform>. Acedido em: 11 de abril de 2025.
- Replica. (2023). Sidewalk Labs spins out Replica to help city planners create virtual populations with big data. Disponível em: <https://venturebeat.com/business/sidewalk-labs-spins-out-replica-to-help-city-planners-create-virtual-populations-with-big-data/>. Acedido em: 11 de abril de 2025.
- StreetLight Data. (2019). *StreetLight Insight Metrics: Our Methodology and Data Sources*. Disponível em: https://learn.streetlightdata.com/hubfs/White%20Papers/Methodology%20and%20Data%20Sources/StreetLight%20Data_Methodology%20and%20Data%20Sources.pdf. Acedido em: 11 de abril de 2025.

- StreetLight Data. (2023a). Travel Data Analytics – How we capture mobility across air, bus, and rail travel. Disponível em: <https://www.streetlightdata.com/travel-data-analytics/>. Acedido em: 11 de abril de 2025.
- StreetLight Data. (2023b). Uber Air Case Study. Disponível em: <https://learn.streetlightdata.com/uber-air>. Acedido em: 11 de abril de 2025.
- Teralytics. (s.d.). Teralytics – plataforma para visualizar e compreender o movimento de pessoas e veículos. Disponível em: <https://teralytics.net/>. Acedido em: 11 de abril de 2025.
- Teralytics. (2023). Mobility Trends: Representing the entire German population. Disponível em: <https://teralytics.net/matrix/mobility-trends/#:~:text=means%20of%20transport%20we%20move,represent%20the%20entire%20German%20population>. Acedido em: 11 de abril de 2025.
- TomTom. (s.d.). TomTom MOVE – Traffic Analytics. Disponível em: <https://www.tomtom.com/products/tomtom-move/>. Acedido em: 10 de abril de 2025.
- TomTom. (2023). *TomTom Traffic Data Solutions* (Documento técnico).
- Willumsen, L. (2021). *Use of Big Data in Transport Modelling*. Discussion Paper 2021/05. International Transport Forum, OCDE, Paris. Disponível em: <https://www.itf-oecd.org/sites/default/files/docs/big-data-transport-modelling.pdf>. Acedido em: 11 de abril de 2025.