

Modelagem de Dados Abertos e Busca Informada na Estrada Real: uma Perspectiva DCAI¹

Davi Machado da Rocha¹

*Universidade de São Paulo (USP)
São Carlos – SP – Brasil*

davimrocha@gmail.com

Abstract. Open digital sources describing historical routes are often too fragmented to yield a navigable graph directly. This paper investigates informed search on the Estrada Real as a data-intensive e-Science workflow under the Data-Centric AI (DCAI) lens. Four heterogeneous sources are unified into a weighted directed graph (108 nodes, 226 edges, 6 routes), making four modeling decisions explicit: homonym disambiguation, historical transversals as synthetic edges, lightweight elevation provenance, and an admissible topology-derived heuristic. Over this curated artifact, 45 controlled runs and 34,668 directed pairs per algorithm show that reverse Dijkstra makes admissibility auditable and shifts A* pruning by up to 32 percentage points; UCS and A* remain optimal in all pairs, while Greedy search diverges in 3,715 cases ($\approx 10.7\%$). The primary contribution is a reproducible data-curation workflow, not a novel search algorithm. Interactive explorer: <https://buscaestrada-real.netlify.app/>.

Resumo. Fontes digitais abertas sobre rotas históricas são frequentemente fragmentadas demais para produzir diretamente um grafo navegável. Este artigo investiga a busca informada na Estrada Real como um workflow de e-Science intensivo em dados, sob a ótica da Inteligência Artificial Centrada em Dados (DCAI). Quatro fontes heterogêneas são unificadas em um grafo dirigido ponderado (108 nós, 226 arestas, 6 rotas), explicitando quatro decisões de modelagem: desambiguação de homônimos, transversais históricas como arestas sintéticas, proveniência leve de elevação e heurística admissível derivada da topologia. Sobre esse artefato curado, 45 execuções controladas e 34.668 pares dirigidos por algoritmo mostram que o Dijkstra reverso torna a admissibilidade auditável e desloca a poda do A* em até 32 pontos percentuais; UCS e A* permanecem ótimos em todos os pares, enquanto a Busca Gulosa diverge em 3.715 casos ($\approx 10,7\%$). A contribuição principal é um workflow reproduzível de curadoria de dados, e não um novo algoritmo de busca. Explorador interativo: <https://buscaestrada-real.netlify.app/>.

1 Introdução

Reconstruir computacionalmente redes geográficas históricas — fundir acervos dispersos em um grafo navegável — é um problema central de e-Science aplicada às Humanidades Digitais. As fontes chegam fragmentadas — traços GPX, KML parciais, JSON reconstruídos, bases municipais —, e é preciso uni-las em uma rede coerente antes de qualquer análise. A Estrada Real, que ligou as regiões auríferas de Minas Gerais ao litoral entre os

séculos XVII e XIX, é uma instância representativa: documentação histórica abundante, várias fontes abertas e os defeitos típicos do gênero — toponímia inconsistente, georreferenciamento aproximado e homônimos —, o que torna a reconstrução, antes de tudo, um problema de curadoria de dados.

Sobre a rede reconstruída, a busca informada cumpre papel duplo: recuperar as rotas ótimas para diferentes viajantes históricos e verificar se o grafo curado sustenta a conectividade documentada. Não se busca descobrir caminhos desconhecidos, mas qualificá-los por custo, na linhagem dos modelos de *least-cost path* (Seção 2). Nesse cenário multifonte, decisões de modelagem dos dados podem deslocar o resultado tanto quanto, ou mais que, a escolha do algoritmo: é a leitura central da *Data-Centric AI* (DCAI), que itera sobre os dados mantendo o modelo fixo [Sambasivan et al. 2021, Whang et al. 2023, Zha et al. 2025], no espírito do quarto paradigma [Hey et al. 2009].

Este artigo aplica essa lente a um problema clássico de Inteligência Artificial — busca informada em grafos — para defender uma tese específica: **neste grafo curado da Estrada Real, decisões de modelagem dos dados movem o desempenho da busca — a poda do A* e a otimalidade da Gulosa — mais do que a escolha entre UCS, Gulosa e A***. Chamamos isso de *dominar*, em sentido mensurável (efeito sobre poda e otimalidade) e não como alegação de generalidade: a evidência provém de um único grafo curado (Seção 6). Quatro decisões foram explicitadas — homônimos, transversais sintéticas, proveniência de elevação e heurística admissível —, e cada uma alterou mensuravelmente o comportamento dos algoritmos, efeito confirmado pela varredura de 34.668 pares além do recorte experimental original.

Contribuições. (C1) Um processo de fusão para quatro fontes abertas da Estrada Real, com desambiguação de homônimos por contexto de rota e marcação leve de proveniência. (C2) Evidência empírica de que, neste cenário, decisões de dados (heurística topológica, desambiguação, transversais) movem o desempenho mais que a escolha entre UCS, Gulosa e A*. (C3) Caracterização topológica da subotimalidade da Gulosa via varredura completa, isolando a cauda em corredores específicos.

2 Trabalhos Relacionados

Reconstrução de redes geográficas históricas. A própria Estrada Real já foi reconstituída com cartografia histórica e SIG, cruzando um mapa de 1778 com remanescentes georreferenciados [Umbelino et al. 2009]. Em escala mais ampla, redes viárias antigas têm sido reconstruídas por modelos de custo-benefício [Stahlberg et al. 2023]. Esses trabalhos centram-se no traçado ou no modelo de custo sobre uma rede assumida; aqui, o foco são as decisões de curadoria que transformam fontes abertas em grafo navegável e seu efeito sobre a busca.

Fusão de fontes heterogêneas. A literatura sistematiza três etapas canônicas — *schema mapping*, *record linkage* e fusão [Bleiholder and Naumann 2009]. Em redes geográficas, elas envolvem toponímia inconsistente, georreferenciamento aproximado e a necessidade de preservar homônimos como entidades distintas.

Busca informada e DCAI. Busca em grafos sobre custos arbitrários é tratada em [Russell and Norvig 2020]; A* com heurística admissível preserva otimali-

dade [Hart et al. 1968]. A literatura costuma centrar a escolha entre UCS, Gulosa e A*; já a DCAI sustenta que, estabilizadas as arquiteturas, os ganhos vêm da iteração sobre qualidade, cobertura e proveniência dos dados [Sambasivan et al. 2021, Whang et al. 2023, Zha et al. 2025] — programa voltado sobretudo a aprendizado supervisionado, aqui estendido à busca clássica, onde a heurística (uma decisão de dados) pode pesar mais que esse contraste algorítmico.

3 Decisões de Modelagem dos Dados

As fontes abertas da Estrada Real divergem em formato, granularidade e qualidade. Esta seção resume as decisões que as transformam em um grafo adequado a busca informada; cada decisão reaparece nos achados experimentais (Seção 5).

3.1 Fusão de quatro fontes heterogêneas

Quatro fontes alimentam o grafo: traços GPX/KML dos quatro caminhos oficiais (Caminho Velho, Novo, Sabarabuçu e Diamantes) [Instituto Estrada Real 2026], quatorze etapas GPX dos Descaminhos do Ouro [Descaminhos do Ouro 2026], JSON do Novo Caminho da Piedade [Valente 2006] reconstruído pelo autor a partir da Rodovia dos Tropeiros e de nós históricos manuais, e coordenadas municipais [IBGE 2026] para corrigir *waypoints* (pontos de referência georreferenciados) deslocados. As seis rotas resultantes são os quatro caminhos oficiais, o conjunto Descaminhos e a Piedade.

O pipeline segue três etapas canônicas [Bleiholder and Naumann 2009].

Schema mapping: as fontes são normalizadas em uma representação única de *nó* (*node_id*, coordenadas, elevação, *semantic_labels*, *route_origins*) e *aresta dirigida* (distância, ganho/perda de elevação, superfície, geometria, modos permitidos, *status*).

Record linkage: cada *waypoint* rotulado é pareado a uma *localidade canônica* (a entrada única daquele lugar no grafo) comparando nomes após *normalização ASCII* (remoção de acentos e caixa). Os *tracks* (linhas GPS) são ancorados às localidades (*snap-ping*) sob dois limiares: 5 km quando a localidade é declarada no nome do track, 800 m por proximidade.

Fusão reutiliza nós entre rotas: quando *n_ouro_preto* já foi criado pelo Caminho Velho, outras rotas apenas anexam suas *route_origins*; o *entity linking* decorre da chave canônica (o *node_id* padronizado).

Tabela 1. Dicionário de dados: principais campos do grafo curado e do modelo de custo.

Campo	Significado
<i>node_id</i>	identificador canônico da localidade; chave de fusão entre rotas
<i>route_origins</i>	rotas históricas a que o nó pertence (proveniência)
<i>surface</i>	superfície da aresta: asfalto, calçamento, terra, misto, trilha ou marítimo
<i>elevation_gain/loss</i>	ganho/perda de elevação ao percorrer a aresta (m)
<i>modos</i>	meios de deslocamento admitidos na aresta
<i>status</i>	<i>measured</i> (traço GPX) ou <i>synthetic</i> (transversal/marítima documentada, sem GPX)
<i>elevation_status</i>	<i>measured_gpx</i> ou <i>manual_no_altimetry</i> (Piedade sem altimetria)
$\sigma_p(s)$	fator de superfície do agente <i>p</i> no piso <i>s</i> (Tabela 2)
α_p	peso da elevação (custo de subir) do agente <i>p</i>

3.2 Desambiguação de homônimos

“Pequeri” existe no Caminho Velho (Congonhas/Alto Maranhão) e na Zona da Mata, a mais de 200 km; sem desambiguação, as duas localidades virariam um atalho artificial. O contexto de rota atua como chave secundária: `pequeri` gera IDs distintos em cada corredor, e uma auditoria automatizada verifica que ambos persistem após a fusão.

3.3 Conhecimento histórico injetado: transversais como arestas sintéticas

A historiografia documenta transversais entre o Caminho Velho e o Novo: o **Caminho de Cima** (São João del Rei $\leftrightarrow$ Carandaí, ~ 66 km) e o **Borda do Campo / Rio das Mortes** (Barbacena $\leftrightarrow$ São João del Rei, ~ 67 km) [Calaes and Ferreira 2009, Costa 2015]. Sem GPX, entram como **arestas sintéticas** (`status = synthetic`): distância haversine $\times$ tortuosidade (1,3) e ganho/perda inferidos das cotas (altitudes) dos nós extremos. Uma aresta marítima análoga liga Porto Estrela ao Rio de Janeiro. Essas arestas testam apenas a ligação historicamente documentada entre dois pontos, sem reproduzir o traçado real (Achado A4).

3.4 Proveniência leve e heurística topológica

Proveniência leve. Por *proveniência leve* entende-se anotar origem e qualidade de cada dado com marcas simples, sem modelo formal. A Piedade introduz arestas sem altimetria medida; o campo `elevation_status` (Tabela 1) torna explícita a subestimação dos custos de elevação nesse corredor.

Heurística topológica. Uma versão preliminar do A^* usava $h(n) = \sigma_{\min}(p) \cdot \text{haversine}(n, t)$, em que n é o nó corrente, t o destino, p o agente e haversine a distância geográfica. Embora intuitiva, ela ignorava o grafo curado (conectividade real, arestas sintéticas, lacunas de altimetria), reduzindo a poda e a auditabilidade da otimalidade.

A substituição foi por um **Dijkstra reverso no grafo transposto** (Dijkstra a partir do destino t sobre o grafo de arestas invertidas, dando o custo mínimo de cada nó até t), com custo relaxado $c_p^-(e) = d(e) \cdot \sigma_{\min}(p)$, em que $\sigma_{\min}(p) = \min_s \sigma_p(s)$, e $h(n) = \text{dist}_{c_p^-}(n, t)$. A relaxação ignora elevação e usa o menor fator de superfície, garantindo $h(n) \leq C_p^*(n, t)$ (admissibilidade); a consistência decorre de a heurística derivar de caminhos reais do grafo: uma estatística pré-computada sobre o grafo curado, não fórmula fechada sobre coordenadas (Achado A2).

4 Definição Formal e Algoritmos

Problema de busca. Dado o grafo dirigido ponderado $G = (V, E)$ resultante da Seção 3, uma origem $s \in V$, um destino $t \in V$ e um agente p , busca-se um caminho $\pi = (s, \dots, t)$ que minimize:

$$C_p(\pi) = \sum_{e \in \pi} c_p(e), \quad c_p(e) = d(e) \cdot \sigma_p(s(e)) + g(e) \cdot \alpha_p \quad (1)$$

em que $d(e)$ é a distância da aresta, $s(e)$ a superfície atribuída à aresta, $\sigma_p(\cdot)$ o fator de

penalização da superfície para o agente p , $g(e)$ o ganho de elevação e α_p o peso atribuído à subida. A perda de elevação fica registrada, mas não entra no custo: para os três agentes, o esforço relevante é subir.

Espaço de estados. Cada nó é um estado, totalizando 108 estados; as ações disponíveis são as arestas dirigidas que dele partem. A transição é determinística e o teste de objetivo verifica se o nó atual coincide com o destino.

Funções de custo por agente. Os agentes são perfis de custo que representam viajantes históricos com prioridades distintas sobre o mesmo grafo: variar o agente é variar a função de custo. A Tabela 2 apresenta os fatores de superfície e o peso de elevação: o tropeiro penaliza trilhas (mulas carregadas), o contrabandista as prefere pela cobertura e a tropa oficial desloca-se melhor em estradas formais. Os rótulos de superfície descrevem o substrato atual de cada trecho (rodovias que recobrem o traçado histórico, por exemplo), não o piso colonial, servindo de proxy do esforço de deslocamento.

Tabela 2. Fatores de custo por agente e superfície.

Superfície	Tropeiro	Contrabandista	Tropa
Asfalto	0,8	2,5	0,7
Calçamento	0,9	1,8	0,8
Terra	1,0	1,0	1,2
Misto	1,3	1,2	1,5
Trilha	2,5	0,6	3,0
Marítimo	1,0	1,0	1,0
α	0,008	0,001	0,005

Algoritmos. UCS expande o nó de menor $g(n)$ e equivale a Dijkstra com destino fixo [Dijkstra 1959]; serve como baseline ótimo. Gulosa expande o nó de menor $h(n)$, sem garantia de otimalidade. A* expande o nó de menor $g(n) + h(n)$ [Hart et al. 1968]; com a heurística admissível da Seção 3.4, preserva otimalidade.

5 Avaliação Experimental

5.1 Configuração

Cinco cenários cobrem travessia total, Caminho Velho integral, transversais como atalhos e um caso representativo de *divergência da Gulosa* (Bananal → Alto Maranhão) — par origem–destino no qual a rota gulosa custa mais que a ótima. Três agentes e três algoritmos completam $5 \times 3 \times 3 = 45$ execuções; a varredura cobre todos os $108 \times 107 = 11.556$ pares dirigidos por agente, totalizando 34.668 execuções por algoritmo. Métricas: custo $C_p(\pi)$ e nós expandidos. O explorador público em <https://buscaestradareal.netlify.app/> reproduz as 45 execuções; grafo curado, tabela de desambiguação e scripts da varredura estão disponíveis em [Rocha 2026].

5.2 Resultados

A Tabela 3 reúne as 45 execuções controladas. UCS e A* coincidem em custo nas 15 combinações (cenário, agente); a Gulosa coincide em 13/15, divergindo apenas em Bananal → Alto Maranhão para tropeiro e tropa oficial.

Tabela 3. Resultados das 45 execuções (custo / nós expandidos). *Transversal*: arestas sintéticas (Seção 3.3); em **negrito, divergências da Gulosa.**

Cenário	Agente	UCS		Gulosa		A*	Tipo de rota
		custo	nós	custo	nós	custo / nós	
Diam.→Rio	Tropeiro	1082,2	94	1082,2	48	1082,2 / 73	Marítima
	Contrab.	951,6	99	951,6	48	951,6 / 78	Marítima
	Tropa Of.	1186,3	94	1186,3	48	1186,3 / 74	Marítima
OP→Paraty	Tropeiro	803,0	105	803,0	26	803,0 / 63	Transversal
	Contrab.	879,3	108	879,3	26	879,3 / 97	Transversal
	Tropa Of.	854,4	105	854,4	26	854,4 / 67	Transversal
Tirad.→Pet.	Tropeiro	401,3	74	401,3	16	401,3 / 23	Transversal
	Contrab.	366,9	80	366,9	16	366,9 / 30	Transversal
	Tropa Of.	449,4	75	449,4	16	449,4 / 28	Transversal
Barb.→Guar.	Tropeiro	515,4	94	515,4	17	515,4 / 28	Transversal
	Contrab.	484,5	97	484,5	17	484,5 / 46	Transversal
	Tropa Of.	565,5	94	565,5	17	565,5 / 38	Transversal
Banan.→AM	Tropeiro	837,3	73	888,4	28	837,3 / 57	Marítima + transversal
	Contrab.	750,1	74	750,1	28	750,1 / 54	Piedade
	Tropa Of.	927,4	73	960,1	28	927,4 / 66	Marítima + transversal

5.3 Cinco achados ancorados em decisões de dados

A1 — Poda do A* acompanha $\sigma_{\min}$ (ver Seção 3.4). Em quatro dos cinco cenários a poda do A* cresce com a amplitude dos fatores de superfície ($\sigma_{\min}$); a exceção é Bananal → Alto Maranhão, onde a penalidade da tropa a trilhas inverte o padrão.

A2 — Heurística geométrica versus topológica (ver Seção 3.4). *Achado central para a leitura DCAI.* A heurística geométrica não incorporava a conectividade curada, reduzindo a poda e dificultando a auditoria; o Dijkstra reverso torna a admissibilidade demonstrável. Para o tropeiro, Ouro Preto → Paraty salta de 8% para 40% de poda; Barbacena → Guaratinguetá, de 39% para 70%. O ganho de até 32 pp supera a diferença de trocar UCS por A* com a mesma heurística.

A3 — Invariância da Gulosa em exploração, mas não em otimalidade (ver Seção 3.4). A Gulosa expandiu o mesmo número de nós para os três agentes em cada cenário (48, 26, 16, 17, 28), pois a heurística só difere por $\sigma_{\min}$; em custo, coincidiu com o ótimo em 13/15 combinações — em Bananal → Alto Maranhão, só o contrabandista foi ótimo, pois a rota gulosa privilegia trilhas na Piedade.

A4 — Transversais nas rotas ótimas (ver Seção 3.3). As transversais entram nas rotas ótimas de Ouro Preto → Paraty, Tiradentes → Petrópolis e Barbacena → Guaratinguetá (todos os agentes) e de Bananal → Alto Maranhão (tropeiro e tropa): conectividade documentada muda rotas sem novas regras de busca.

A5 — Varredura completa do grafo (ver Seções 3.4 e 3.2). Nos $108 \times 107 = 11.556$ pares dirigidos por agente, UCS e A* foram ótimos em 100% dos casos; a Gulosa diverge em **3.715** ($\approx 10,7\%$) (Tabela 4). A cauda (os casos de maior divergência) combina Piedade com transversais e depende da desambiguação de homônimos: sem ela, o efeito *buraco de minhoca* (atalho espúrio entre localidades homônimas distantes) contaminaria a comparação entre pares.

Tabela 4. Resumo da subotimalidade da Gulosa na varredura completa.

Agente	Div.	Taxa	Razão méd.	Maior razão (caso)
Contrabandista	1155	10,0%	1,067	1,404 (Petrópolis → Baependi)
Tropa Oficial	1279	11,1%	1,054	1,294 (SJDR → Sto. Ant. Leite)
Tropeiro	1281	11,1%	1,050	1,286 (Ouro Preto → Caeté)

6 Discussão

No escopo deste artefato, os achados sustentam a tese do artigo: a heurística topológica (Achado A2) desloca a poda do A* além do que a troca de algoritmo conseguiria; a varredura mostra que ela não cura tudo (10,7% dos pares divergem), mas a cauda é caracterizável, indicando que novos ganhos dependem mais de dados adicionais que de outro algoritmo. A desambiguação de homônimos é o caso extremo: sem ela, custos artificiais e divergências espúrias contaminariam as caudas — cascata de dados típica da DCAI [Sambasivan et al. 2021].

Limitações. O grafo cobre paradas relevantes, não um inventário exaustivo; trechos manuais da Piedade não têm altimetria GPX; transversais testam conectividade documentada, não traçado real; e a monotonicidade da poda em $\sigma_{\min}$ é tendência empírica, não garantia formal. A heurística por Dijkstra reverso deriva do próprio grafo, de modo que sua vantagem de poda é em parte endógena à modelagem. Os pesos de custo por agente (Tabela 2) são ilustrativos, não validados empiricamente; ablações de transversais, aresta marítima, tortuosidade e desses pesos ficam como trabalho futuro.

7 Conclusão

Este artigo aplicou DCAI à busca informada em redes geográficas históricas. A evidência das 45 execuções controladas e das 34.668 da varredura aponta na mesma direção: neste artefato, decisões de modelagem dos dados moveram o desempenho da busca mais do que a escolha entre UCS, Gulosa e A*; generalizar exige replicar a curadoria em novos grafos (trabalho futuro). O resultado também contribui para e-Science em Humanidades Digitais ao transformar fontes dispersas em artefato computacional auditável e explorável: as rotas e os custos por agente funcionam como hipóteses quantitativas sobre percursos plausíveis sob diferentes prioridades de viagem.

Trabalhos futuros. A* Ponderado para o *trade-off* otimalidade–eficiência; expansão de transversais e paradas; altimetria GPX para a Piedade; proveniência completa via PROV-O.

Agradecimentos

O autor agradece ao Prof. Ricardo Marcondes Marcacini (USP/ICMC) pela orientação na disciplina em que este trabalho se originou, e aos mantenedores do Instituto Estrada Real, do Descaminhos do Ouro e do Paisagem Cultural da Estrada Real [Paisagem Cultural 2026] pelos dados abertos que alimentam o grafo e o explorador.

Uso de IA Generativa

Este artigo utilizou ferramenta de IA generativa (LLM) para estruturação do texto e revisão linguística, conforme a chamada do XX BreSci. Todos os resultados experimentais provêm do pipeline implementado e auditado pelo autor (NetworkX 3.x, Python 3.11), a quem cabe a verificação metodológica e bibliográfica.

Referências

- Bleiholder, J. and Naumann, F. (2009). Data fusion. *ACM Computing Surveys*, 41(1):1–41. DOI: 10.1145/1456650.1456651.
- Calaes, G. D. and Ferreira, G. E., editors (2009). *A Estrada Real e a transferência da corte portuguesa: Programa Rumys – Projeto Estrada Real*. CETEM/MCT/CNPq/CYTED, Rio de Janeiro.
- Costa, A. G. (2015). Registros do Caminho Novo para as minas de ouro nos mapas antigos. In *Atas do VI Simpósio Luso-Brasileiro de Cartografia Histórica*, Braga.
- Descaminhos do Ouro (2026). Cicloturismo. Disponível em <https://cicloturismo.descaminhosdoouro.com.br/>. Acesso em abr. 2026 (cópia no Wayback Machine).
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271. DOI: 10.1007/BF01386390.
- Hart, P. E., Nilsson, N. J., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107. DOI: 10.1109/TSSC.1968.300136.
- Hey, T., Tansley, S., and Tolle, K., editors (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, WA.
- IBGE (2026). Coordenadas das sedes municipais. Disponível em <https://www.ibge.gov.br/>. Acesso em mar. 2026.
- Instituto Estrada Real (2026). Roteiros planilhados. Disponível em <https://institutoestradaareal.com.br/>. Acesso em mar. 2026.
- Paisagem Cultural da Estrada Real (2026). Inventário de bens culturais, sítios arqueológicos e grupos sociais. Disponível em <https://www.paisagemestradaareal.com.br/>. Acesso em abr. 2026.
- Rocha, D. M. (2026). *Estrada Real: grafo curado, desambiguação de homônimos e scripts da varredura* [dados]. Zenodo. DOI: 10.5281/zenodo.20867698.
- Russell, S. and Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition.
- Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., and Aroyo, L. M. (2021). “Everyone wants to do the model work, not the data work”: data cascades in high-stakes AI. In *Proc. of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15. DOI: 10.1145/3411764.3445518.
- Stahlberg, M. J., Sagnol, G., Ducke, B., and Klimm, M. (2023). Spatiotemporal reconstruction of ancient road networks through sequential cost–benefit analysis. *PNAS Nexus*, 2(2):pgac313. DOI: 10.1093/pnasnexus/pgac313.
- Umbelino, G., Carvalho, R., and Antunes, A. (2009). Uso da cartografia histórica e do SIG para a reconstituição dos caminhos da Estrada Real. *Revista Brasileira de Cartografia*, 61(1). DOI: 10.14393/rbcv61n1-44855.
- Valente, R. V. (2006). Café e escravidão no Caminho Novo da Piedade. In *Anais... USP*.
- Whang, S. E., Roh, Y., Song, H., and Lee, J.-G. (2023). Data collection and quality challenges in deep learning: a data-centric AI perspective. *The VLDB Journal*, 32(4):791–813. DOI: 10.1007/s00778-022-00775-9.
- Zha, D., Bhat, Z. P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., and Hu, X. (2025). Data-centric Artificial Intelligence: a survey. *ACM Computing Surveys*, 57(5):1–42. DOI: 10.1145/3711118.