Spatial Inefficiencies in Africa's Trade Network

Tilman Graff[*]

December 6, 2022

Abstract

I assess the efficiency of transport networks for every country in Africa. Using spatial data from various sources, I simulate trade flows over more than 70,000 links covering the entire continent. I maximise over the space of networks and find the optimal road system for every African state. My simulations predict that Africa would gain 1% of total welfare from better organising its national road systems. I then construct a novel dataset of local network inefficiency and find that colonial infrastructure projects significantly skew trade networks towards a sub-optimal equilibrium. I also find evidence for regional favouritism and inefficient aid provision.

JEL-codes: R42, R12, O18, O11

1 Introduction

Trade costs in Africa are the highest in the world, severely inhibiting interregional trade (Limao and Venables, 2001; The Economist, 2015; Nugent and Lamarque, 2022). Sub-Saharan Africa's coverage with paved roads is by far the lowest of any world region, with only 31 total paved road kilometres per 100 square kilometres of land, compared to 134 in other low-income countries (Foster and Briceño-Garmendia, 2010). The World Bank has identified an annual infrastructure gap amounting to 93 billion US dollars and urges countries in Sub-Saharan Africa to spend almost one per cent of GDP on building new roads (Foster and Briceño-Garmendia, 2010; Nugent and Lamarque, 2022). This reasoning is also reflected in the composition of development aid – in 2017, by far the largest share of World Bank lending to African countries was allocated to transport infrastructure projects (The World Bank, 2017). There appears to be a clear consensus that Africa needs more roads.

I investigate a neglected, yet powerful second source of spatial inefficiency in Africa's transport system. I don't ask if the continent has too *few* roads, but rather analyse whether the current infrastructure is *in the wrong place*. Do Africa's roads connect the right areas to promote trade? How would a social planner design a perfect transport network which optimises welfare in a given country? Which African country is closest to its hypothetical optimum? And why are some locations systematically cut off from the national trade system?

In this paper, I derive the unique optimal trade network for every country in Africa. Using data from satellites and online routing services, I first construct an interconnected economic topography of more than 10,000 square grid cells covering the entire continent. I then employ a simple network

[*]Harvard University. tgraff@g.harvard.edu. I thank my supervisor Ferdinand Rauch for invaluable support, and Pol Antras, Antonia Delius, Dave Donaldson, Lukas Freund, Doug Gollin, Johannes Haushofer, Gabriel Kreindler, Ameet Morjaria, Felix Samy Soliman, Fabio Schmidt-Fischbach, Evan Soltas, Tony Venables, and seminar participants at Oxford, LSE, NBER, and the World Bank for helpful feedback and discussion.

trade model to simulate trade flows through more than 70,000 links spanning all of Africa. In a second step, I use a variant of a recently established framework by Fajgelbaum and Schaal (2020) to optimise over the space of networks and find the optimally redesigned transport system given the underlying economic fundamentals for every African country. An intuitive thought experiment demonstrates this process: suppose the social planner were to observe the spatial distribution of roads, people, and economic activity in a given country before being allowed to lift all roads from their current location, freely shuffle them around, and then reorganise them in the most efficient way for mutual trade. The planner is not allowed to build completely new roads, but is only allowed to move infrastructure from one part of the country to another. In this exercise, she takes into account local incentives for trade between all sets of neighbours on a complex network graph, regional differences in trade costs caused by geographical and network characteristics, and heterogeneous costs to constructing new roads depending on the underlying terrain.

I then compare these optimal networks to the current system. I argue that the degree to which the optimum differs from the status quo can be interpreted as an intuitive measure for the inefficiency of a country's current road network. I show that potential welfare gains from reshuffling roads would improve overall welfare on the continent by about 1%. I also identify South Sudan as the country with the most inefficient transport network in Africa.

On the regional level, this scenario creates winners and losers. The model identifies some areas as having *too* many roads and decides to put them to better use somewhere else. These areas were inefficiently overendowed with transportation infrastructure before the reshuffling exercise. Other regions, however, did not have enough infrastructure given their relative position in the network and are now awarded additional roads by the social planner. I identify these areas as discriminated against by the current transportation network design. By comparing welfare levels before and after the hypothetical intervention, I create a novel dataset of local infrastructure discrimination for more than 10,000 cells covering the entire African continent.

Why are some regions systematically cut off from the benefits of efficient trade? I use a variety of empirical designs to analyse the substantial spatial variation present in my dataset. Firstly, I investigate the long-run effects of large infrastructure investments from the colonial area. Similarly to Jedwab and Moradi (2016), I find a persistent impact of railway lines constructed by the colonial powers over a century ago. Plausibly exogeneous variation in the number of kilometres crossing a given area significantly skews the current trade network towards a suboptimal state today. Even though many of the railway lines have fallen into disarray since independence, regions close to colonial railroads still have too much road infrastructure given their relative position in the network. In contrast, railway lines that were planned, but by historical accident never built, do not predict any significant departure from the optimal spatial distribution.

Secondly, I analyse how spatial inefficiencies in Africa's trade system are related to ethnic power dynamics. I find no conclusive evidence that ethnic regions that are politically discriminated against have more or less than optimal infrastructure stocks. However, I do find evidence for regional favouritism – home regions of national leaders have significantly more infrastructure than is nationally efficient. Finally, I investigate the extent to which foreign aid projects have succeeded in alleviating the imbalances in Africa's transport networks. I present descriptive evidence demonstrating that World Bank funds have not gone towards the regions most in need of additional infrastructure. Instead, areas that are identified as having too many roads are associated with more Bank lending. The same patterns hold for development aid from China.

My study contributes to several strands of literature. In analysing the impact of transport revolutions, I add to the large body of work devoted to identifying the economic returns to improving infrastructure systems. A series of rigorous studies have gauged the welfare effects of the expansion of the US railway network in the 19th century (Donaldson and Hornbeck, 2016; Swisher, 2017), colonial railway and rural road systems in India (Donaldson, 2018; Burgess and Donaldson, 2012; Asher and Novosad, 2020), or highway systems in China (Faber, 2014; Baum-Snow et al., 2017). In contrast to these studies, I do not analyse the impact of existing transport revolutions, but rather measure how much a hypothetical first-best transport system would improve welfare (see also Santamaria, 2022). Methodologically, I harness recent advances bringing insights from the optimal transport literature into the economics discourse. Most directly, I apply the framework by Fajgelbaum and Schaal (2020) to construct the optimal trade network for every African country. They are the first to optimise over the space of networks in order to find the globally efficient transport system in an economics context, though the problem has long featured prominently in the mathematics literature (for a textbook treatment, see Bernot et al., 2009; Galichon, 2016). Previous studies in economics relied on stepwise heuristics to eliminate suboptimal counterfactual networks like Alder (2022) in India or Burgess et al. (2015) in Kenya, but did not include a derivation of the globally optimal network design. Once constructed, my network features trade on a two-dimensional lattice geometry and is hence related to the theoretical work of Allen and Arkolakis (2014, 2022). I also contribute to the literature employing regional trade models to explain subnational welfare disparities caused by internal transport geography in a development context (Atkin and Donaldson, 2015; Storeygard, 2016; Coşar and Fajgelbaum, 2016; Fiorini et al., 2021; Gorton and Ianchovichina, 2022). In constructing the efficient network to mitigate these dispersions over space, I also contribute to new explorations into conditions and characteristics of optimal spatial policies (Fajgelbaum and Gaubert, 2020). In my three strands of empirical inquiry, I first add to the literature examining long-run persistence of colonial transportation revolutions in Africa (Jedwab and Moradi, 2016; Jedwab et al., 2017). I also contribute to the literature examining how ethnic relations relate to comparative development in Africa (Michalopoulos and Papaioannou, 2013, 2014, 2016) and add to our understanding of how ethnic (De Luca et al., 2018) and regional favouritism (Hodler and Raschky, 2014; Burgess et al., 2015) skew public goods spending towards an inefficient allocation. Lastly, I contribute to the literature on the distribution and effects of foreign aid (Clemens et al., 2012; Nunn and Qian, 2014; Dreher et al., 2019; Dreher and Langlotz, 2020).

2 A model of optimal transport networks

In this paper, I derive the unique optimal goods trade network for every country in Africa. To be able to maximise over the space of networks, I harnesses an altered version of a framework by Fajgelbaum and Schaal (2020). The model is introduced in the following paragraphs.

2.1 Geography

Following the set-up and notation of Fajgelbaum and Schaal (2020), I consider a set of locations $\mathcal{I} = \{1, ..., I\}$. Each location $i \in \mathcal{I}$ inhabits a number of homogeneous consumers L_i. This number is treated as given and fixed for every location, such that consumers are not allowed to move between locations. Each consumer has an identical set of preferences characterised by

$$u = c^{\alpha}$$

where c denotes per capita consumption. Every consumer in location i consumes c_i and $C_i = L_i c_i$ denotes total consumption in location i.

There is a set of goods $\mathcal{N}$ denoted by $n = \{1, ..., N\}$. Total consumption in each location is defined as the CES aggregation of these goods

$$C_i = \left(\sum_{n=1}^{N} (C_i^n)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

where σ denotes the standard elasticity of substitution and C_i^n denotes the consumption of good n in location i. Locations specialise in the production of goods such that each location only supplies one variety $n \in \mathcal{N}$. Let $Y_i = Y_i^n$ denote total production in location i.

2.2 Network topography

Locations $\mathcal{I}$ represent nodes of an undirected network graph. Each location i is directly connected to a set of neighbours $N(i) \in \mathcal{I} \setminus \{i\}$. I consider locations to be arranged on a two-dimensional square lattice where each node is connected to its eight surrounding nodes to the north, north-east, east, and so on.

All goods can be traded within the network. Let $Q_{i,k}^n$ denote the total flow of good n travelling between nodes i and $k \in N(i)$. While goods can only be traded between neighbouring nodes, nothing prevents them from travelling long distances through the network by passing multiple locations after each other. Sending goods from location i to location $k \in N(i)$ incurs trade costs, which are modelled in the canonical iceberg form. I follow Fajgelbaum and Schaal and model iceberg trade costs for trading good n between neighbouring locations i and k as

$$\tau_{i,k}^n(Q_{i,k}^n, I_{i,k}) = \delta_{i,k}^{\tau} \frac{(Q_{i,k}^n)^{\beta}}{I_{i,k}^{\gamma}} \tag{1}$$

where $I_{i,k}$ is defined as the level of *infrastructure* on the edge between nodes i and k. More infrastructure on a given link decreases the cost of trading between them. $\delta_{i,k}^{\tau}$ is a scaling parameter, which allows trade costs to be flexibly adjusted for any given origin-destination pair. Trade costs also depend on $Q_{i,k}^n$, the total flow of goods on the link. Higher existing trade volumes on a given edge make sending an additional good more costly, a dynamic Fajgelbaum and Schaal refer to as *congestion externality*. The social planner realises this and takes congestion into account when determining optimal trade flows.

In equilibrium, each location cannot consume and export more than it produced and imported. More formally

$$C_i^n + \sum_{k \in N(i)} Q_{i,k}^n (1 + \tau_{i,k}^n(Q_{i,k}^n, I_{i,k})) \leq Y_i^n + \sum_{j \in N(i)} Q_{j,i}^n \tag{2}$$

must hold for every n and i.

I follow the contribution of the Fajgelbaum and Schaal (2020) framework and proceed to endogenize infrastructure provision $I_{i,k}$ in order to facilitate optimal trade flows. Analytically, this problem nests the static trade flow exercise outlined above. The social planner chooses an infrastructure network, and given the network proceeds to compute the optimal trade flows subject to (2). To make the problem more interesting, I introduce a constraint on infrastructure. This is specified in fairly straightforward manner as the *Network Building Constraint*

$$\sum_i \sum_{k \in N(i)} \delta^I_{i,k} I_{i,k} \leq K \tag{3}$$

where $\delta^i_{i,k}$ denotes the cost of building infrastructure on the edge between nodes i and k. Total spending on infrastructure is constrained by K, the sum originally spent on building the *existing* road network of a country. I observe the current road network of the economy, infer how much it must have cost to build it, and set K equal to this amount. The social planner's task of choosing an optimal network hence amounts to a reallocation exercise. She gathers all road building material available in the economy and gets to redistribute it in a way that is welfare optimising according to the model. Improving infrastructure between two nodes in order to foster local trade hence comes at the cost of having to take away infrastructure elsewhere. I argue that the extent to which the social planner has to rearrange existing edges serves as a sensible measure of spatial inefficiency in the existing network.[1]

2.3 Planner's problem and equilibrium

In the nested problem, the social planner observes localities, endowments, population, and preferences and solves for trade flows between nodes that maximise overall welfare. She also solves for the optimal transport network which induces welfare-maximising trade flows in the nested problem while respecting the *Network Building Constraint* (3). The full planner's problem can hence be stated as

$$\max_{\left\{\left\{C_i^n, \{Q_{i,k}^n\}_{k \in N(i)}\right\}_n, \; c_i, \{I_{i,k}\}_{k \in N(i)}\right\}_i} \quad \sum_i L_i u(c_i)$$

subject to

$$L_i c_i \leq \left(\sum_{n=1}^{N} (C_i^n)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \forall i \in \mathcal{I} \qquad \textsc{CES Consumption}$$

$$C_i^n + \sum_{k \in N(i)} Q_{i,k}^n (1 + \tau_{i,k}^n(Q_{i,k}^n, I_{i,k})) \leq Y_i^n + \sum_{j \in N(i)} Q_{j,i}^n, \forall i \in \mathcal{I}, n \in \mathcal{N} \qquad \textsc{Balanced Flows Constraint}$$

$$\sum_i \sum_{k \in N(i)} \delta^I_{i,k} I_{i,k} \leq K \qquad \textsc{Network Building Constraint}$$

$$I_{i,k} = I_{k,i}, \forall i \in \mathcal{I}, k \in N(i) \qquad \textsc{Infrastructure Symmetry}$$

$$C_i^n, c_i, Q_{i,k}^n, I_{i,k} \geq 0, \forall i \in \mathcal{I}, n \in \mathcal{N}, k \in N(i). \qquad \textsc{Non-Negativity}^2$$

My version of the planner's problem follows the baseline Fajgelbaum and Schaal (2020) model. However, my model has four important differences. First, in my model all goods are tradeable and no local amenities exist. Second, I do not allow workers to migrate between places and hence differences in marginal utility might still exist between nodes. Third, my model remains agnostic about the production function of each location and no analysis of the optimal use of input factors is

[1] I also impose infrastructure symmetry and restrict $I_{i,k} = I_{k,i} \; \forall \; i, k \in N(i)$.

[2] As will be discussed in chapter 3, I calibrate my version of the model with an even stronger lower bound to infrastructure $I_{i,k}$ than mere non-negativity. For reasons discussed below, I simulate the model while binding $I_{i,k} \geq 4$. For all other variables, merely non-negativity is required.

undertaken. Fourth, I impose infrastructure symmetry. All these changes are undertaken with the later calibration and reshuffling exercise in mind.

While optimising over the space of networks might appear daunting, Fajgelbaum and Schaal (2020) provide conditions under which deriving the unique spatial optimum is both ensured and feasible. Instead of solving for every single infrastructure link, I follow the authors and recast the problem in its dual representation as a set of first-order conditions from the subproblems, which only depend on Lagrange multipliers of each constraint. There are considerably fewer multipliers than primal control variables, namely one for every good in every node (interpretable as local prices). I am hence left to only find a price field from which under the convexity assumptions, all other properties follow.[3] As spelled out more formally in the technical appendix (see section A), I obtain the optimal network by constructing the Lagrangian corresponding to the planner's problem, deriving its first-order conditions, and recasting them as functions of the Lagrange parameters. Numerical optimisation now yields the solution to the dual problem and inserting the parameters back into the derived first-order conditions, I can immediately derive the optimal infrastructure network $I_{i,k}$, optimal trade flows $Q^n_{i,k}$ over this network, and ensuing consumption patterns C^n_i in each location.

3 Calibration of current and optimal trade network designs

To calibrate a topography of economic activity and trade in all African countries, I construct a novel network representation covering the entire continent and enrich its nodes and edges with data from a variety of sources.

3.1 Network nodes

I first divide the entire continent into grid cells of 0.5 degrees latitude by 0.5 degrees longitude (roughly 55 by 55 kilometres at the equator). For all of Africa, this amounts to 10,167 cells. I then aggregate spatial data on economic and geographic characteristics onto this grid cell level. I calibrate L_i with data on 2015 population totals from the *Gridded Population of the World* dataset (2016).[4] On average, a cell is home to 110,000 people (median 25,000). The most populous cell contains Cairo and inhabits almost 18 million people. 212 cells are uninhabited. To proxy for heterogeneities in economic activity over space, I rely on the established practise of using satellite imagery of light intensity at night (Henderson et al., 2012). Data on 2010 night luminosity come from Henderson et al. (2018) and are also aggregated onto my study's 0.5×0.5 degree grid resolution to form Y_i.

3.2 Network edges

To quantify the degree to which network nodes are connected to each other, I make use of the open source routing service OPEN STREET MAP (OSM). The OSM routing algorithm is specified for cars and takes into account differential speeds attainable on different types of roads. For every centroid location, I scan OSM for the optimal route to each of their respective eight surrounding neighbours (or less for coastal grid cells). For all of the resulting almost 75,000 routes, I gather distance travelled,

[3] It is still a quite demanding task to solve the ensuing dual problem, even numerically. Invoking duality reduces the scale of the problem, but I am still left with optimising over $I \times N$ variables.

[4] This NASA-funded project gathers data from hundreds of local census bureaus and statistical agencies in order to construct a consistent high-resolution spatial dataset of the world's population. When a datasource only reports population totals for large, higher-level administrative districts, the dataset smoothes population uniformly over the entire area. GPW does not employ any auxiliary data sources – like satellite data – to weight-adjust population totals over space (Doxsey-Whitfield et al., 2015). Africa is the continent with the coarsest resolution of administrative input data. However, the average coverage of $(57\text{KM})^2$ neatly matches the grid cell resolution of my study.

average speed, and step-by-step coordinates of the travel path.[5][6] For some particularly remote areas, the nearest street is very far away, such that the car routing provided by OSM is not sensible. To counter these cases, I also calculate for all 70,000+ connections the outside option of walking the entire link in a straight line at 4 km/h. I then identify cases in which walking directly is actually faster than using OSM's proposed route (plus the travel to and from roads). In these cases, I replace OSM's route with the walking distance and constant 4 km/h speed.

My study is concerned with the optimal domestic road network for each country in Africa. I hence divide the entire grid along national borders. Reoptimising roads solely within a country's borders comes at the risk of undervaluing roads built primarily for international trade. A highway connecting to an important trading post or port just beyond a country's borders might look inefficient to a social planner who is only given national data. I hence create a buffer of 120km around each country and allow the social planner to take these border regions into account when computing the optimal network.

Figure 1 presents the resulting road networks for four countries. I plot all connections within each country and their border buffer (in grey). Figure 1a displays every OSM connection for Burkina Faso, which appears overall fairly well connected. Connections in which walking were the preferred alternative are displayed in thin straight lines and fairly rare. Figure 1b presents the case of Mali, which paints a different picture: for many connections through the Sahara in the north of the country, walking straight lines is actually the fastest way to get from A to B. The DRC in Figure 1c displays a clear lack of infrastructure in the middle of the country. Small Rwanda in Figure 1d zooms in on the actual roads taken and displays the intricacies of the optimal routing provided by OSM.

Relying on the open source community of OSM does come with some drawbacks. The most pressing concern is that data on the position and quality of roads are user-generated and hence subject to reporting bias. Richer areas may appear to be equipped with more roads if local residents have the time and necessary access to a computer to enter their neighbourhoods into the database. While this is certainly troubling, I believe this bias to be much more important on finer resolutions than the operating one in this study. Start and destination of the elicited routes are on average more than 55 kilometres apart and travel will hence take place mostly on larger roads and national highways. It is unlikely that these major streets are systematically underreported in OSM, the primary open source routing platform on the internet.[7]

I use the average attainable speed between locations according to the OSM algorithm as a proxy for the quality of current infrastructure on the edge between them. If two locations are linked by a faster connection, I assume this to be the result of higher infrastructure $I_{i,k}$ on this edge. I hence set

$$I_{i,k} = \text{Average Speed}_{i,k} \tag{5}$$

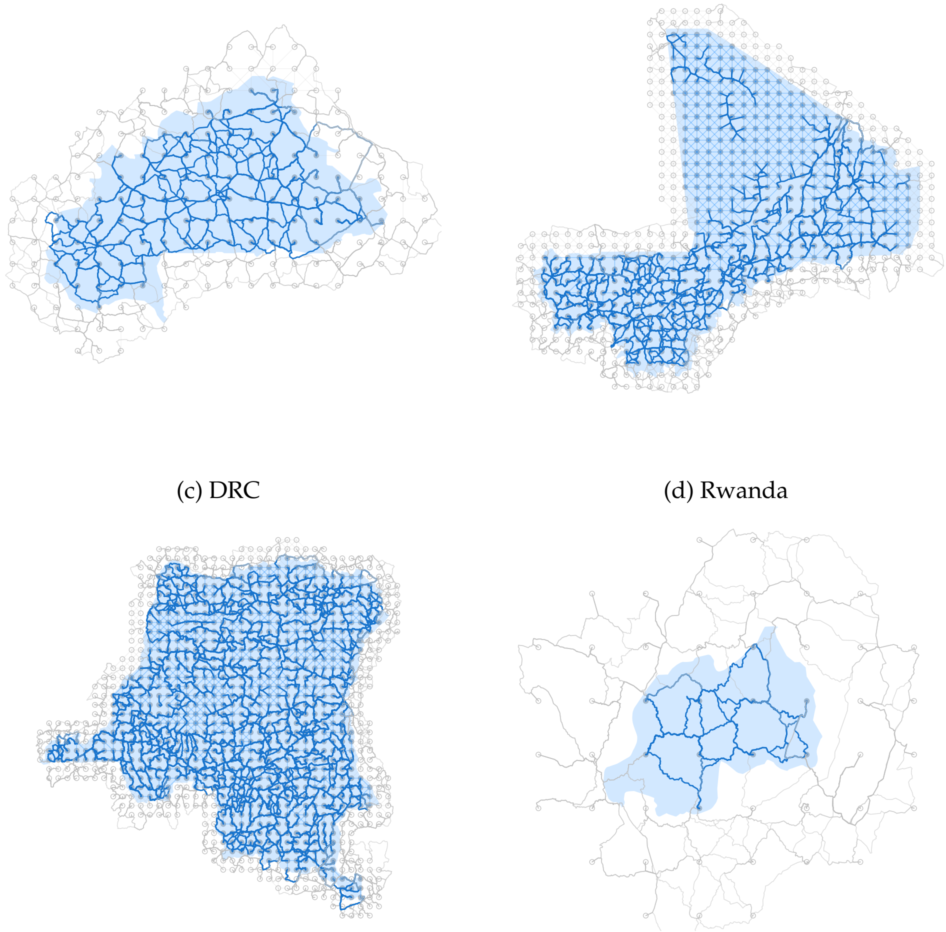
This measure is naturally bound from below at 4 km/h, as walking the air-line distance is always

[5]Scans of OSM were conducted in July 2019. The service does not allow a retrospective scan over past road databases, so a time difference between lights (2010), population (2015), and roads (2019) can not be overcome.

[6]If either start or destination location do not directly fall onto a street, the optimal route jumps to the nearest road and goes from there. To take this into account, I add a walking distance to the travel path. Agents are assumed to walk in straight lines to the nearest street at a fixed speed of 4 km/h. They then take the car and drive the route with average speed as specified by OSM, before they potentially have to walk the last stretch again to their exact centroid destination.

[7]In some rare cases (less than 0.1 per cent of all connections), the OSM algorithm cannot find any route between two neighbouring centroid locations. This is mostly due to an obvious geographic impossibility to connect two nodes. In Guinea-Bissau, for instance one location lies on the Bolama Islands just off the shore of mainland Guinea-Bissau. Its neighbouring locations are all on the mainland and hence unreachable by car. In other cases, both locations to be connected are in deep jungle or swampy regions. In all these cases, I treat the link as if the two locations were not neighbours in the first place. That implies I even forgo the backup possibility of walking the entire distance, assuming that agents cannot walk between islands or through the densest jungle.

Figure 1: Road networks for different countries as scanned off OSM



(a) Burkina Faso

(b) Mali

(c) DRC

(d) Rwanda

Road networks as scanned off Open Street Map (OSM). Blue lines represent routes from each grid cell centroid to each of its eight surrounding neighbours. These routes may include a portion walked by foot in order to get to the nearest street. Grey lines are connections abroad. Connections in which walking the entire distance is faster are printed as thin lines. Data scanned in July 2019.

available as a backup. Empirically, average speeds range between 6 km/h (Mauritania, where most routes go through the desert and have to be covered by walking) and 33 km/h (Swaziland).

To parameterise iceberg trade costs defined in equation (1), I follow Fajgelbaum and Schaal and set $\beta = 1.245$ and $\gamma = 0.6225$. I calibrate $\delta^{\tau}_{i,k}$ following Atkin and Donaldson (2015)'s estimate of the distance coefficient of the gravity equation. Directly taking the average of the authors' two point estimates for Ethiopia and Nigeria, I calculate

$$\delta^{\tau}_{i,k} = 0.0466 \times \ln(\text{Distance}_{i,k}) \tag{6}$$

as the trade cost elasticity to distance travelled.[8]

$\delta^{I}_{i,k}$ from equation (3) denotes the relative constant cost of increasing the average speed on a given link by one. I follow Fajgelbaum and Schaal who in turn make use of a recent study by Collier et al. (2015), which estimates infrastructure building costs in developing countries. Readily applying their

[8] Atkin and Donaldson (Table 2, page 44) estimate the coefficient as 0.0374 for Ethiopia and 0.0558 for Nigeria. My parameter is the simple average of these two point estimates.

specification, I calculate

$$\ln(\delta^I_{i,k,c}) = \delta^I_c + 0.12 \times \ln(\text{Ruggedness}_{i,k,c}) + \ln(\text{Distance}_{i,k,c}) \tag{7}$$

as the constant cost of increasing infrastructure $I_{i,k}$ on the link between i and k in country c. Distance$_{i,k,c}$ denotes the road distance travelled between nodes and enters positively, implying that longer roads are costlier to develop as every single road kilometre will have to be improved. The term Ruggedness$_{i,k,c}$ denotes the average ruggedness between grid cells i and k and enters positively, highlighting the additional expenses accompanied with building on uneven terrain.[9] δ^I_c is a country-specific scaling parameter. Its main purpose it to ensure that equation (3) is satisfied with $K = 1$. I first appraise the infrastructure network $I_{i,k}$ of all countries and then flexibly alter δ^I_c for each nation individually in order to comply with equation (3).

To build incentives for trade, I introduce $N = 6$ different varieties. First, the four most populous grid cells of a given country are assumed to be producing their own variety. This creates incentives for trade between major cities. Another "international" variety is supplied by the three most populous grid cells within each country's border buffer. This ensures incentives for international trade with big localities beyond the border. I also collect data from Lloyd's List on grid cells which are home to a major international port.[10] Every port location (within a country's borders or within its international buffer) not covered by the previous varieties is assumed to also produce the "rest of the world" variety. Lastly, every other location is assumed to produce a sixth, "agricultural" variety.

After these steps, a discretised network representation exists for every African country. Figures 2a,2c,2e present such networks for three countries. Nodes are printed larger proportional to their population. Edges are drawn thicker proportional to the initial infrastructure investment. Grid cells producing a variety other than the agricultural variety are highlighted with an additional circle around them.

3.3 Trade network optimisation

For each country, I conduct two simulations. In both exercises, I calibrate the curvature parameter of the utility function at $\alpha = 0.4$ and the elasticity of substitution parameter at $\sigma = 4$. In the first simulation exercise, infrastructure $I_{i,k}$ is treated as fixed. This is to obtain a baseline estimate of the spatial variation of welfare in each country. The social planner would like to overcome these differences, but is confronted with trade costs which might leave certain remote areas much worse off than well-connected ones.

Following this static exercise, I proceed to the main task of endogenizing the infrastructure matrix $I_{i,k}$. With the network building constraint binding total infrastructure investment at the level of the *current* road network, the social planner is now free to reshuffle roads within the country in order to improve connections as she chooses. If she wants to improve the connection between two given locations, she will have to take away infrastructure from somewhere else in the country. This reallocation exercise does not seek to identify where to place the optimal next investment, but rather represents a hypothetical scenario in which every road can be lifted from the ground, reshuffled, and eventually located someplace else.[11] The procedure does not measure how many roads a country

[9] Data on local ruggedness come from Henderson et al. (2018) and is described in more detail with other geographical covariates below.

[10] I use the open-access portal at https://directories.lloydslist.com/port and hand-code the locations of the 90 biggest ports in Africa. Figure A.4a prints the resulting locations.

[11] Note that equation (3) only fixes $\sum_i \sum_{k \in N(i)} \delta^I_{i,k} I_{i,k} = K$. Hence, not the overall sum of infrastructure is fixed, but more precisely the overall cost of infrastructure. This still allows the social planner to take away one unit of infrastructure on a

has, but rather how well they are placed.

When conducting these optimisations, I fix infrastructure in the international buffer around each country at their current level. While the planner takes into account the economic geography of a country and its surroundings, she can only reallocate infrastructure within the country in question.[12]

I conduct the reallocation scenario for every African country. Six small countries (Cape Verde, Comoros, The Gambia, Mauritius, São Tomé and Príncipe, and Reunion) are too small to form a sensible network as they only show up as a single location in the dataset and are henceforth no longer considered. Optimisations are performed with a version of the optimisation toolkit provided by Fajgelbaum and Schaal (2020). When conducting the simulations, I bind the social planner's set of permissible roads from below at 4 km/h. This is motivated by the assumption at the beginning that walking straight lines at this speed is an outside option and always available to any traveler. I also impose a maximum speed of 120 km/h on any given link (such that $4 \leq I_{i,k} \leq 120 \, \forall \, i \in \mathcal{I}, k \in N(i)$).

Figure 2 visualises this reallocation exercise for several countries. Subfigure 2a displays the discretised network representation of Burkina Faso. The edges to this network are printed almost evenly thick, implying that infrastructure is fairly evenly distributed across the country. Subfigure 2b then displays the country after the network reshuffling exercise. Three patterns stand out. First, the social planner sees a clear need to connect the populous areas in the center of the country with each other. For that, the social planner is willing to salvage some of the apparently less important infrastructure in the north of the country. Second, there still is a benefit to having a few trails connecting the center with a regional hub in the south-west producing it's own variety. Third, nodes are printed in a colour scale corresponding to individual welfare gains and losses for each location. As can be seen from first-glance, most southern regions (brighter colors) stand to gain from this scenario, while the big cities on average seem to lose (darker colors).

The Democratic Republic of Congo in Figures 2c – 2d displays a more decentralised optimal network solution. The social planner sees need to better connect the center of the country to its surroundings and the populous border regions. This is in line with the common perception of DRC's periphery being notoriously poorly connected to the centers of power and commerce. As a result, large parts in the southern center of the country gain welfare in this scenario, while many border regions (especially in the west) lose out. Small Rwanda in Figures 2e – 2f helps to illustrate some of the forces at hand in a less crowded graph. Starting from a fairly evenly distributed transport network, the reallocation dynamics lead to much more variation in infrastructure provision. Some links are deemed superfluous and hence reduced to the smallest admissible level, while others are scaled to multiple times their starting infrastructure stock.

4 A measure of spatial transport network inefficiency

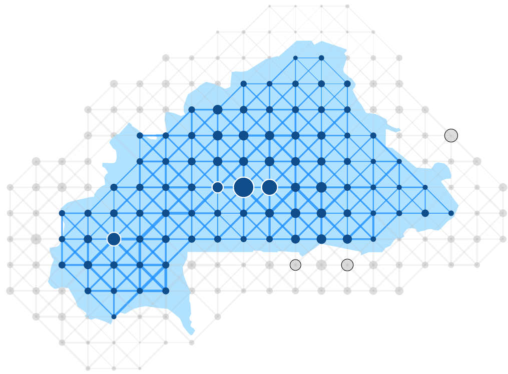
After successfully reshuffling a country's transport network, national welfare will by construction (weakly) increase. While overall production (light output) will be unaffected by the exercise, any welfare gains are solely caused by enabling mutual benefits from trade through connecting the right

very expensive (high $\delta^I_{i,k}$) link and exchange it for much more than one unit on a cheaper (low $\delta^I_{i,k}$) link.

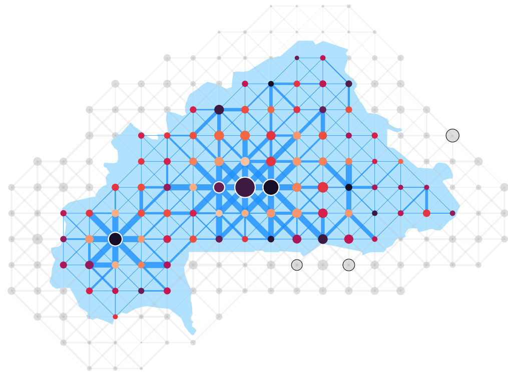
[12] There are two reasons why I conduct the simulation procedure within countries and not over the entire African continent. One is computational; the requirements for numerically solving the model increase quadratically in the number of locations $\mathcal{I}$. The largest country in Africa (Algeria) is made up of almost 900 locations and already strains computing power quite heavily. Simulating all of Africa's 10,000+ locations at once is then almost unattainable with available technology. The second reason is interpretational; while lifting a country's roads from the ground and flexibly reshuffling them across the nation is already a fictitious scenario, it still operates within a government transport authority's locus of control. Regions disadvantaged by their own government can reasonably be considered discriminated against. This is less the case if one were to optimise over the entire continent. Without a central planning body for all of Africa, it is hard to interpret why a road in e.g. Tunisia should rather be moved into Namibia.

Figure 2: Reallocation scenario for different countries

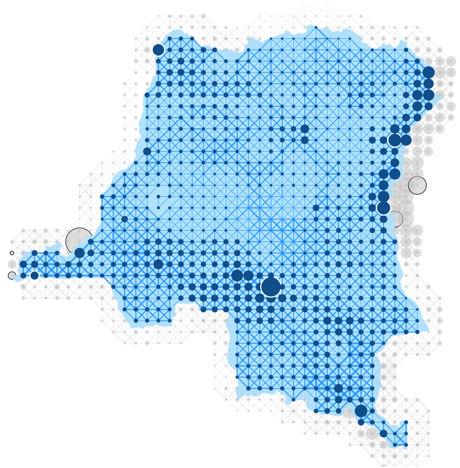
(a) Burkina Faso before reallocation



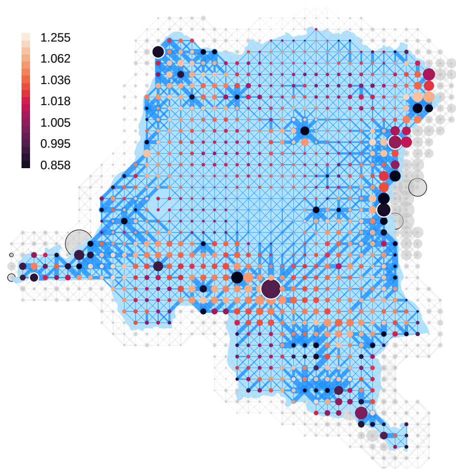
(b) Burkina Faso after reallocation



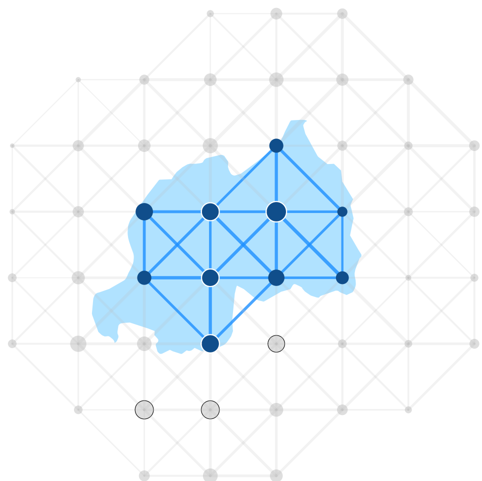
(c) DRC before reallocation



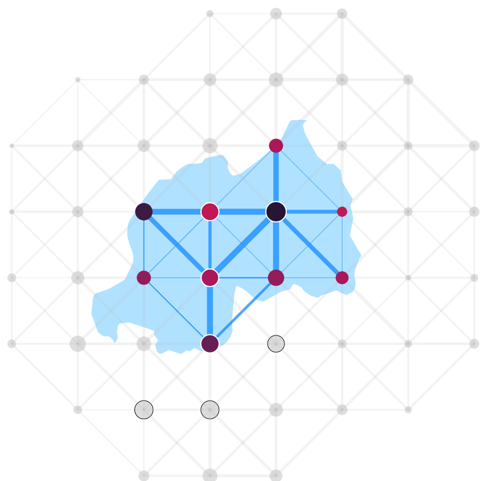
(d) DRC after reallocation



(e) Rwanda before reallocation

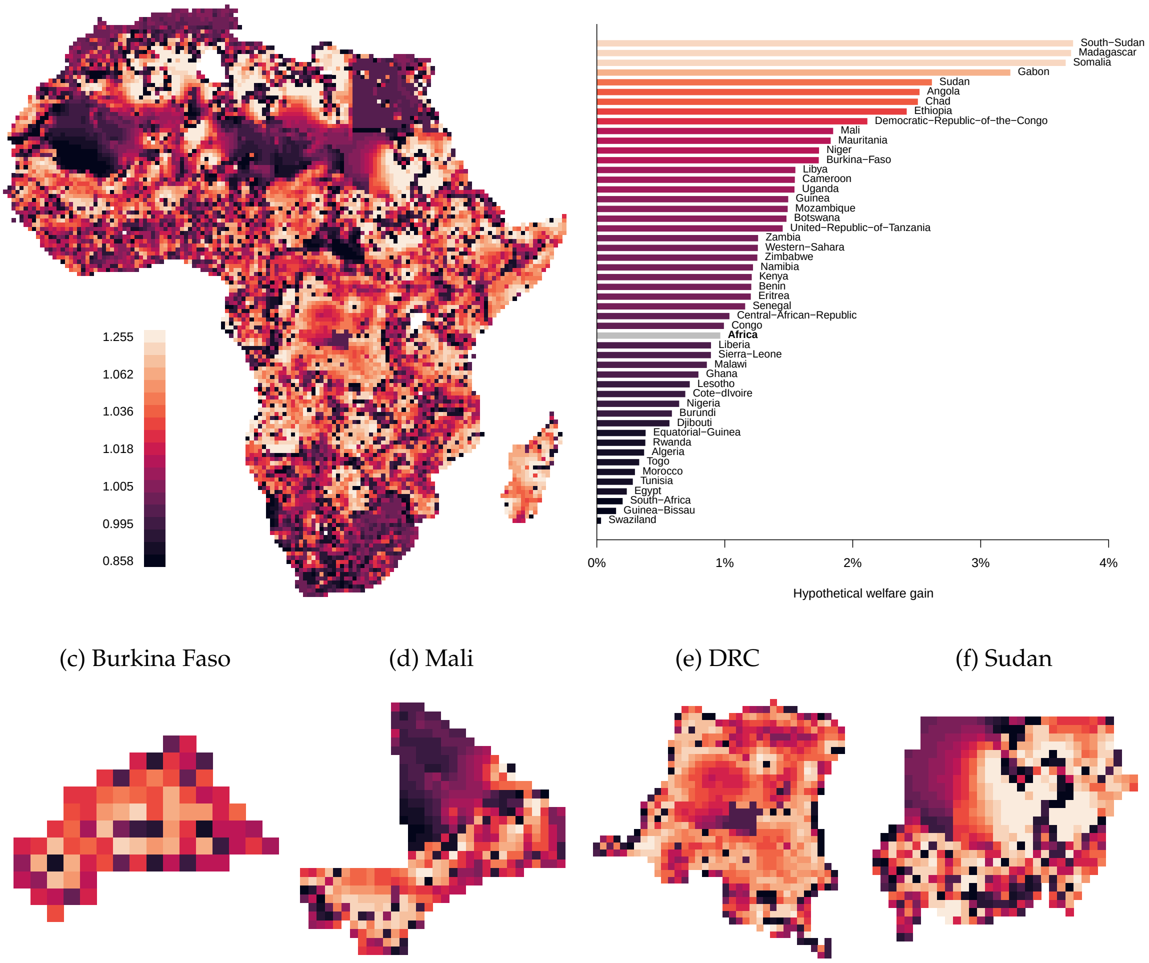


(f) Rwanda after reallocation



Results from optimally reshuffling roads in three African countries. In each network graph, every node represents a grid cell centroid location with radius proportional to the size of its local population. Edges are drawn thicker depending on their allotted infrastructure $I_{i,k}$ *(i.e. average attainable speed). In the optimal networks on the right, nodes are coloured based on their relative welfare gains and losses, with more light areas gaining more. Color scheme is the same as in Figure 3a.*

Figure 3: Africa by network inefficiency



(a) Λ_i across the continent

(b) Welfare gains by country

(c) Burkina Faso

(d) Mali

(e) DRC

(f) Sudan

Discrimination index at the grid cell level (a) and welfare gain under the optimal reallocation counterfactual for each country (b). Country gains are computed by comparing the population-weighted mean of the discrimination index of all cells in a country. (c-f) zoom in on four countries and their discrimination index. Maps show each country as the 0.5 × 0.5 degree grid used for the network optimisation. For each map, darker shaded cells correspond to lower Λ_i levels and hence less infrastructure discrimination compared to the optimal network. All maps share the same color scale, reported in (a).

locations. Nevertheless, they are not negligible. The DRC of Figure 2d, for instance, stands to gain 2.1% of overall welfare just by reshuffling its roads. Burkina Faso (1.7%) and Rwanda (0.4%) are closer to their hypothetical optimum.

Figure 3b reports hypothetical welfare gains for all African countries. Some nations like South Africa (0.2% welfare gains) or Tunisia (0.3%) perform better than the three countries from above. Many countries are leaving much more on the table, like Somalia (3.7%) or Madagascar (also 3.7%). No African country has a more inefficiently placed road network than South Sudan. Its citizens stand to gain 3.8% of overall welfare if just their roads were better placed. This maybe comes as no surprise, as the world's newest country has largely inherited a road network that was not conceived to sustain an independent nation, but rather connect it to its former capital up north. For the entire continent, optimal reallocation of national road systems would improve overall welfare by 1.0%.

Forgone welfare gains can be conceived as an intuitive measure for overall network inefficiency. The closer hypothetical gains to zero, the more efficient the current allocation of roads. Vice-versa, if a country stands to gain a lot from reshuffling, then the current network is deemed more inefficient. On a simple cross-section, countries with less efficient networks are weakly correlated with less property rights ($p = 0.1$), less 2010 log GDP per capita ($p = 0.39$), yet also less corruption ($p = 0.2$).[13]

While each country only stands to gain overall welfare from the reallocation procedure, individual locations might very well lose in the process. Intuitively, some regions might be equipped with far too many good roads such that the social planner takes these roads away to use someplace else. Comparing each grid cell's welfare before and after the major reshuffling can help to identify regions which are currently over or underprovided for. More formally, I define

$$\Lambda_i = \frac{\text{Welfare under the optimal infrastructure}_i}{\text{Welfare under the current infrastructure}_i} \tag{8}$$

as the *Local Infrastructure Discrimination Index* Λ_i for grid cell i. Areas with high Λ_i scores ($\Lambda_i > 1$) would be gaining under the optimal reallocation scenario and are hence under provided for in the network's current state. A score of $\Lambda_i < 1$ on the other hand, implies that a region is *too* well off given its position in the network today and hence should be stripped off some of its infrastructure to increase overall welfare.

Figure 3a displays the spatial distribution of Λ_i over all 10,000+ grid cells of the entire continent. The darker a grid cell's shade, the more it is advantaged by the inefficiencies of the current network. When interpreting this map, note that grid cells are undergoing the reshuffling scenario solely within their respective country (while taking into account respective international buffer regions). National borders hence play a role and can at times even clearly be inferred from the printed map. Keeping this in mind, the map reveals substantial spatial variation in the index across the African continent. The luckiest region stands to loose almost 25% of total welfare if the fictitious social planner intervened and reshuffled roads away from it. On the other extreme of the spectrum, the residents of the most discriminated grid cell are missing out on a welfare hike of almost 40%. On average, a grid cell gains 2.7 per cent of welfare, the median cell gains 1.1%.[14] Figures 3c – 3f zoom in on a few countries. In Mali, the Saharan north west has roads reallocated away from it, which benefits the more populous south. In the DRC, border regions are printed in dark shades and hence clearly lose out in favor of the center. Sudan shows a reallocation of roads away from the Nile, which can clearly be inferred from the map.

In interpreting Λ_i, keep in mind that this is a measure of differences in *welfare*. In this, it need not be a direct mapping of changes in actual infrastructure provision. Indeed, the highly non-linear nature of the optimal reallocation scenario can lead to situations in which a certain region substantially profits from the optimal policy, even though it is not directly granted additional roads. Local changes in welfare can instead be caused also by fortuitous peculiarities of geography – maybe a neighbouring region emerges as a local trade hub, or the optimal network leads to broader availability in the variety of goods, all without directly targeting each individual grid cell with additional roads. In my full dataset, changes in welfare Λ_i are hence not significantly correlated with local changes in infrastructure.

In the final section, I analyse patterns behind the heterogeneity of infrastructure discrimination over space.

[13]Data from The World Bank (2017). For corruption and property rights, data are only available for 35 countries and correlations are hence performed on this truncated sample.

[14]Note that this number is higher than the 1% aggregate welfare gain for the continent because of variation in population across space.

5 The determinants of Africa's spatial trade network imbalances

Why are some African roads not in the right place to promote beneficial trade? To investigate which areas have too much or too little infrastructure, I employ the *local infrastructure discrimination index* Λ_i as dependent variable in a standard OLS regression setting. In the base specification, I estimate

$$\tilde{\Lambda}_{i,c} = \beta v_{i,c} + \mathbf{X}_{i,c}\gamma + \delta_c + \epsilon_{i,c} \tag{9}$$

for a variety of different independent variables $v_{i,c}$ of grid cell i in country c. The dependent variable $\tilde{\Lambda}_{i,c}$ corresponds to the z-scored transformation of Λ_{ic}. δ_c denotes country fixed effects, $\mathbf{X}'_{i,c}$ is a vector of controls, and β is the coefficient of interest. To account for spatial autocorrelation, I follow Bester et al. (2011) and construct a higher-level spatial grid of 3 degrees latitude by 3 degrees longitude and cluster standard errors within each of these higher-level grid cells.

Country fixed effects make observations comparable internationally by accounting for the fact that reallocation is performed for each country individually. The vector of controls $\mathbf{X}'_{i,c}$ captures observable characteristics of each grid cell that plausibly account for some of the variation in Λ_i. Henderson et al. (2018) show that a surprisingly parsimonious set of geographical and agricultural covariates explains a substantial part of the global variation in economic activity. Making use of their data, I include in $\mathbf{X}$ each grid cell's average altitude, average temperature and precipitation, land suitability for agriculture, length of the annual growing period, and an index for the stability of malaria transmission. I also include mutually exclusive (and collectively exhaustive) dummy variables classifying each grid cell into one of twelve predominant vegetation regions (or *biomes*, see Henderson et al., 2018).[15] To flexibly control for any broad geographic trend over the entire continent, I additionally add fourth-order polynomials of both latitude and longitude for each grid cell. To take into account that some regions have a natural advantage in conducting trade, I also include indicators for whether a grid cell's centroid is within 25 kilometres of a natural harbour, big lake, or navigable river respectively (again using data from Henderson et al., 2018), and whether a grid cell is at the border of a country.

I call the set of controls outlined so far *"Geographic Controls"*. They are in principle unaffected by human decisions about the design of trade networks. Another set of covariates, however, poses more difficulties. These are the variables that were already used to calibrate the optimal reallocation simulation from above, namely a cell's population, light output, ruggedness, and classification into urban and rural.[16] I call these *"Simulation Controls"*. It is crucial to be aware that Λ_i is, among others, already a product of the intricate interplay between these factors. There is hence a danger for plain OLS to detect a spurious, mechanical relationship between them, potentially biasing results. In the main analysis, I include the *"Simulation Controls"*; yet results are largely robust to excluding them.[17]

In this paper, I investigate three potential sources of network inefficiency in Africa: colonial era infrastructure investments, ethnic power relations, and foreign aid.

5.1 Colonial infrastructure investments

Between 1890 and 1960, British, French, Belgian, German, Italian, and Portuguese administrations all undertook efforts to permeate their colonies with more or less sophisticated railway networks

[15] Only eight of the twelve vegetation patterns are actually present on the African continent, the other indicators (biomes 4, 6, 8, and 11) are dropped from consideration.

[16] Recall that population and output were components of the planner's problem, ruggedness went into the cost of building new infrastructure $\delta^I_{i,k}$, and the urban/rural classification determined which good a cell produced.

[17] Tables A.1,A.2,A.3 in the appendix replicate the tables of the main text while dropping the simulation controls.

(Jedwab and Moradi, 2016). There were two main motivations for this: supporting the extractive economies and ensuring military domination (Jedwab et al., 2017). These colonial railroads have been found to have a persistent impact on the spatial organisation of economic activity today. Jedwab and Moradi (2016) show how urbanisation started to centre around railway tracks in the decades following their construction. Even as most railway lines have fallen into disarray and road traffic has replaced trains as the most important means of transportation, economic activity today still clusters in places close to the former rail lines.

Did the transport revolution coordinate the economy on an efficient spatial equilibrium? To investigate whether railways from the colonial period still have an impact on trade network inefficiency today, I overlay the 10,000+ grid cells of my data set with every railway line built by the colonial powers in Sub-Saharan Africa. Figure A.1 prints in red 237 lines built between 1890 and the various independence dates. Data on railroad positioning comes from Jedwab and Moradi (2016), with the exception of South Africa, for which I manually digitise a map from Herranz-Loncán and Fourie (2017).[18]

For every grid cell, I compute the total number of colonial railway kilometres crossing the cell. This serves as a measure for the stock of physical transport capital invested into each region. The majority of cells (91%) are not crossed by a colonial railway and hence have zero railroad kilometres. Those that are intersected by a line usually have between 20 and 60 railroad kilometres, while some important rail crossings or transport hubs have up to 100 kilometres of railroads. Every railroad line comes with a classification of being constructed primarily for military purposes or mining purposes (or neither, or both), allowing for more nuanced further analysis. Lastly, to account for potential endogeneity concerns, I also compute the same statistics for a set of railway lines the colonisers planned, but never built. As Jedwab and Moradi (2016) explain, these *"placebo"* projects were not realised only for a series of arguably random historical events like unforeseeable cuts to financing, the outbreak of wars, or sudden retirements of administration officials. These tracks are printed in blue in Figure A.1.[19]

Table 1 displays results from OLS estimation of equation (9) with $\tilde{\Lambda}_i$ on the left hand side and total rail kilometres as explanatory variable v_i. Column (1) reports the baseline association between colonial railways and present-day infrastructure discrimination: every 50 kilometres of colonial railway construction are associated with grid cells losing 0.24 standard deviations of welfare at the hand of areas without any investment. Column (2) repeats the exercise but using "placebo" railroads. Reassuringly, we find no association between infrastructure discrimination and infrastructure that was never built. Columns (3) and (4) distinguish between railroads built for different purposes: while both military and mining purpose railroads are associated with too much infrastructure today, it is in particular areas with military railroads that lose out under the reallocation scenario.

Columns (5) and (6) trace out the spatial gradient of infrastructure discrimination close to colonial railroad lines. I cluster grid cells in bins depending on the distance of their centroid to the nearest realised and placebo line. Grids with centroids further than 40KM away from a rail serve as the omitted category. We observe a substantial gradient, with grid cells within 10KM of a realised railroad being 0.33 standard deviations too well off. Magnitudes are similar for cells withing 20KM of a railroad, after which the association tapers off. Cells between 20 and 30KM out are still favoured by the current network, yet the planner reallocated much fewer infrastructure away from them. The association becomes indistinguishable from zero even further out. No comparable associations can

[18]No comparable data are available for Madagascar, Egypt, and the Maghreb countries, which also saw some colonial railway construction. Findings are robust to excluding grid cells from these countries.

[19]Data for placebo lines also come from Jedwab and Moradi (2016) and Herranz-Loncán and Fourie (2017)

Table 1: Colonial railroads and local infrastructure discrimination index

	Infrastructure discrimination $\tilde{\Lambda}$ (z-scored)				Real	Placebo
	(1)	(2)	(3)	(4)	(5)	(6)
50 KM of Colonial Railroads	-0.238***					
	(0.0507)					
50 KM of Colonial Placebo Railroads		-0.198				
		(0.154)				
50 KM of Colonial Railroads for Military Purposes			-0.289***			
			(0.0742)			
50 KM of Colonial Railroads for Mining Purposes				-0.161**		
				(0.0575)		
<10KM to railroad					-0.332***	-0.0973
					(0.0674)	(0.0511)
10-20KM to railroad					-0.377***	-0.0736
					(0.0721)	(0.0635)
20-30KM to railroad					-0.141*	0.00604
					(0.0669)	(0.0656)
30-40KM to railroad					0.0295	0.0621
					(0.0542)	(0.0499)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	10158	10158	10158	10158	10158	10158
R2	0.165	0.162	0.164	0.162	0.168	0.162

Results of estimation of equation (9) *on the sample of 0.5 $\times$ 0.5 degree grid cells for the entire African continent (excluding six small countries, see text). Dependent variable is the z-scored Local Infrastructure Discrimination Index $\tilde{\Lambda}_i$ for each grid cell. Columns (1)-(4) estimate the effect of colonial infrastructure investments as measured by the total number of colonial railroad kilometres of different types crossing a cell. Geography controls, consisting of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, indicators for the 12 predominant agricultural biomes, indicators for whether a cell is a capital, within 25 KM of a natural harbour, navigable river, or lake, the fourth-order polynomial of latitude and longitude, and an indicator of whether the grid cell lies on the border of a country's network. Simulation controls are comprised of population, night lights, and ruggedness. These indicators went into the original infrastructure reallocation simulation and are hence not orthogonal to Λ. Standard errors are clustered on the 3 $\times$ 3 degree level and are shown in parentheses.*

be found for placebo railroads in column (6). This is suggestive evidence that the confounding effect of colonial infrastructure policies is locally contained. Areas blessed with a close-by railway line are still too well off today, which comes at the expense of their neighbouring regions just a few kilometres away.

The effects described in Table 1 are small, yet remarkable. Across the African continent, areas that received large infrastructure investments a century ago are still *too* well off given their position in the national trade network. In contrast, areas that were not crossed by tracks are inefficiently short on infrastructure today. To see that this is a non-trivial finding, note that, firstly, most of the colonial railway lines have been in disrepair for decades and thus do not immediately dictate trade flows today. Secondly, recall that the optimal network reallocation and construction of Λ_i was based on roads and cars, not rails and trains. The implication is hence *not* that colonial railway systems themselves are inadequate to efficiently sustain inter-regional trade today. Rather the transport revolution a century ago coordinated the entire economy into a certain spatial equilibrium, which persists even though it has become inefficient. African nations would benefit from moving to a better equilibrium, but are locked in the current state. The social planner identifies this, seeks to overcome these misallocations, and moves infrastructure away from regions once considered important by the colonisers. The even stronger findings for military railroads reinforces that point: while mining railroads arguably still dictate trade flows today, military railroads have become completely obsolete and are much more clearly associated with imbalanced infrastructure stocks across space. In short, Jedwab and Moradi show that colonial investments helped the economy to coordinate on one of many spatial equilibria – my findings suggest that this is not the optimal one.[20]

5.2 Regional favouritism

The political economy of public good provision along ethnic lines presents an interesting case for the spatial imbalances in my dataset. Are regions with less political clout systematically discriminated against in the provision of trade infrastructure? One avenue to investigate differential infrastructure provision to privileged places is to look at the birthplaces of political leaders. Existing studies have shown that the rise to power of a new national leader leads to temporarily more consumption and output in the leader's birth region (Hodler and Raschky, 2014) and ethnic homeland (De Luca et al., 2018). During the leader's time in office, birth region and ethnic homelands also benefit from more foreign aid and infrastructure investment being channeled their way (Dreher et al., 2019; Burgess et al., 2015). To investigate whether ethnic favouritism accounts for imbalances in trade infrastructure provision over space, I make use of a dataset of African national leaders provided by Dreher et al. (2019). The data entail information about the birth region and time in office of 117 heads of state holding power in 44 African countries dating back to 1969.[21] Using Open Street Map, I obtain coordinates for birthplaces and spatially merge them with my grid dataset. I then use this information to calculate for each cell the total number of years someone born there has held high office. This allows me to analyse whether a region's over provision with transport infrastructure covaries with personal ties to national power.

Table 2 columns (1-4) investigate effects of such regional favouritism. Column (1) estimates equation (10) on the full sample of grid cells with the explanatory variable being the total number of years someone born in that cell was holding national power. The covariate enters with a small and insignificant coefficient, implying that for each year one of their members was in power, an ethnic homeland is about 0.01 standard deviations too well off given their relative position in the country's trade network. Column (2) looks at the extensive margin by correlating Λ_i with a dummy for whether a grid cell ever sent someone to lead the country. We find a marginally significant negative relationship: grid cells that ever were the birthplace for a national leader are 0.25 standard deviations too well off today. If the social planner were to intervene, she would strip cells with regional ties to power from some infrastructure and reallocate it towards areas with no such ties. Columns (3-4) exclude capital grid cells from the dataset in an attempt to counter biases stemming from some regions just being close to power generally. The negative associations from columns (1-2) prevail.

Lastly, I turn to ethnic homelands, which are arguably the more relevant unit of observation when investigating regional favouritism on the African continent. I follow Michalopoulos and Papaioannou (2013, 2014, 2016) and intersect an ethnolinguistic map of pre-colonial homelands from Murdock (1959) with current national borders.[22] I spatially aggregate my grid cell measure of network inefficiency Λ_i onto the ethnicity-country level by assigning each grid cell an ethnicity based on its centroid location and weighing grid cells by their respective population. Figure A.3 presents the spatial variation of the local infrastructure discrimination index Λ_h for each ethnicity-country

[20]Note that the spatial equilibrium induced by colonial railroads could have still been optimal *at the time*. My argument solely concerns the persistent effects of investments a century ago on network efficiency *today*.

[21]No data on national leaders are reported for Algeria, Western Sahara, South Sudan, Somalia, and Djibouti.

[22]Ethnicity data are available for every country but Western Sahara. Ethnicities present in more than one country count as multiple observations. Of the 835 inhabited homelands identified by Murdock, 314 are split in two or more parts by the current national borders, creating 1,212 ethnicity-country observations.

pair h.[23]

I investigate patterns of infrastructure discrimination along ethnic lines by way of the negative treatment of groups excluded from the government (*ethnic discrimination*). To measure ethnic discrimination, I rely on three measures of political relationships between ethnicities. I make use of the Ethnic Power Relations (EPR) database by Vogt et al. (2015), which globally identifies "politically relevant ethnic groups and their access to state power" (Vogt et al., 2015, p. 1328) over the past seven decades. Not every ethnic homeland inhabits a group that is "politically relevant", significantly truncating the to 432 observations.[24] Firstly, EPR reports a yearly time series of political discrimination for every group in the sample. In particular, a group is coded as discriminated against by the central government if there is "active, intentional, and targeted discrimination by the state against group members in the domain of public politics" (Vogt et al., 2015, p. 1331). I follow Michalopoulos and Papaioannou (2016) and analyse effects of a dummy variable taking on the value one if a group has experienced discrimination in at least one year between 1960 and 2010 and use this measure to investigate whether infrastructure discrimination covaries with political discrimination. Secondly, I broaden the definition of ethnic discrimination and more generally look at groups which are excluded from the central government. As defined by EPR, this classification entails all groups that are discriminated against (from above), plus groups that are defined as either powerless or self-excluded (Vogt et al., 2015, p. 1331).[25] Thirdly, I analyse the effects of an EPR indicator denoting whether an ethnicity was part of a civil war with an explicitly ethnic dimension at some point between 1960 and 2010. The construction of this indicator is identical to the dummies described above and is obtained from Michalopoulos and Papaioannou (2016).

To analyse patterns of infrastructure discrimination on the ethnic homeland level, I estimate a slightly different version of equation (9)

$$\tilde{\Lambda}_{h,c} = \beta v_{h,c} + \mathbf{X}_{h,c}\gamma + \delta_c + \epsilon_{h,c} \tag{10}$$

where $\tilde{\Lambda}_{h,c}$ is the z-scored local infrastructure discrimination index for homeland h in country c, $\mathbf{X}'_{h,c}$ and δ_c again denote controls and country fixed effects respectively, $v_{h,c}$ are the explanatory covariates discussed above, and β is the coefficient of interest. The number of ethnicity observations (about 430) is significantly smaller than the number of grid cells. In order to avoid overfitting, I slightly truncate the set of controls $\mathbf{X}'_{h,c}$. I replace the latitude and longitude polynomials as well as dummies indicating proximity to a natural harbour, river, lake, and national border with two continuous measures of distance to the nearest border and distance to the coast. As homelands are much more irregularly shaped than grid cells, I also include the natural logarithm of each homelands' area (as in Michalopoulos and Papaioannou, 2016). Apart from these adjustments, $\mathbf{X}'_{h,c}$ entails all geographical and simulation controls of the models on the grid cell level. I follow Michalopoulos and Papaioannou (2016) and double-cluster standard errors at both the country level as well as the ethnic family level using the mechanism proposed by Cameron et al. (2011).

Table 2 columns (5-7) report results. The estimates of being discriminated against or involved in an ethnic war are not significantly different from zero, implying that the social planner would on average not reallocate infrastructure away or towards those areas. Slightly puzzling is the result in

[23]The homeland which would benefit most from national reshuffling of roads is the Kababish in Sudan (who would gain 18%). The most disproportionally advantaged ethnic homelands in Africa are those of the Kreish (11%) of the Central African Republic (who stand to lose 10% of welfare if optimal networks were imposed).

[24]Merging EPR observations with ethnic homelands is non-trivial. Thankfully, I am able to rely on the conversion established by Michalopoulos and Papaioannou (2016)

[25]I again rely on the transformation by Michalopoulos and Papaioannou (2016) who code an indicator as one if the ethnic group has experienced exclusion from the government at any point between 1960 and 2010.

Table 2: Ethnic favoritism

	$\tilde{\Lambda}$: entire sample		$\tilde{\Lambda}$: excluding capitals		$\tilde{\Lambda}$: ethnic homeland level		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total years in power	-0.00927		-0.0121*				
	(0.00474)		(0.00538)				
Ever in power dummy		-0.253*		-0.296*			
		(0.110)		(0.122)			
Discriminated against					-0.194		
					(0.158)		
Excluded from government						-0.365**	
						(0.121)	
Involved in ethnic war							-0.235
							(0.180)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	10158	10158	10109	10109	494	494	494
R2	0.162	0.162	0.164	0.164	0.255	0.264	0.257

Persistent impacts of holding power on local infrastructure discrimination in leaders' birthplaces and across ethnic homelands. Columns 1-4 estimates equation (10) *on the grid cell level. Independent variable in columns (1) and (3) is the number of years since 1969 someone born in the homeland was the country's leader. In (2) and (4), the independent variable is a dummy indicating whether the cell ever held power. Columns (3)–(4) replicate the regressions while excluding observations containing a country's capital. Geography controls, consisting of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, indicators for the 12 predominant agricultural biomes, indicators for whether a cell is a capital, within 25 KM of a natural harbour, navigable river, or lake, the fourth-order polynomial of latitude and longitude, and an indicator of whether the grid cell lies on the border of a country's network. Simulation controls are comprised of population, night lights, and ruggedness. These indicators went into the original infrastructure reallocation simulation and are hence not orthogonal to* Λ. *Columns (5–7) estimate equation 10 on the ethnic homeland by country level. Geography controls there are slightly truncated (see text). Standard errors are (double-)clustered on the country level (and the ethnic-family level for homelands) and are reported in parentheses.*

column (6): homelands excluded from government on average have too *high* infrastructure stocks (yielding lower Λ_i discrimination scores and hence a negative estimate). One could interpret this as preliminary evidence of access to power and access to infrastructure being substitutes, wherein elites tilt the trade network in the country towards regions they want to keep quiet and away from government. Of course, the estimations in this section are not well identified and as a result could also just be spurious.

5.3 Foreign aid

Africa is the primary target of international aid. In 2017, African countries received more World Bank aid than Europe, Central Asia, Latin America, and the Caribbean combined (The World Bank, 2017). Of almost 12 billion US dollars worth of lending commitments, the biggest share was awarded to projects aimed at improving transportation infrastructure.[26] The World Bank is not alone – in the past decade, non-traditional players have entered and disrupted the international development aid system (Dreher and Lohmann, 2015). Most notably, China has emerged as a significant donor nation, funding development projects in at least 50 African countries since the turn of the millennium (Strange et al., 2017). Yet despite the vast amount of resources involved, foreign aid has not yet been unequivocally proven to be linked with positive economic outcomes in recipient countries (Burnside

[26]The transport sector made up 18% of total IBRD and IDA lending to African nations, followed by water and sanitation (14%), energy and extractives (14%), and public administration (12%) (The World Bank, 2017).

and Dollar, 2000; Easterly et al., 2004; Rajan and Subramanian, 2008; Clemens et al., 2012; Clemens and Kremer, 2016; Nunn and Qian, 2014).

To investigate whether international development aid is quantitatively associated to my measure of trade network inefficiency, I make use of two datasets of geo-referenced aid flows to Africa. Firstly, AidData (2017) in cooperation with the World Bank, tracks over 5,600 lending lines from the World Bank to African nations and reports precise coordinates of over 60,000 projects financed through these funds, totalling more than 300 billion US dollars. The sample comprises all projects approved between 1996–2014. As Strandow et al. (2011) describe, attributing projects to locations relies on a double-blind coding procedure of various World Bank documents. Secondly, I explore patterns from a similar database on Chinese aid projects by Strange et al. (2017). They resort to reports from numerous local and international media outlets to track official and unofficial financing lines to over 1,500 projects worth 73 billion US dollars in the period 2000–2011.[27]

For the purpose of this study, I exclude aid projects with no clear-cut geographical target like unconditional lending lines to the central government or assistance for political parties. I also exclude flows with unknown or only vague information on eventual project location.[28] I also ignore projects which were still under construction or otherwise not fully completed by the end of 2017. Together, these steps truncate the World Bank sample by 35% and the China sample by 52%. In Figure A.2, I map the spatial distribution of aid projects from both remaining samples. I aggregate the total value of aid disbursements from the remaining 10,786 World Bank projects and 1,420 Chinese projects onto the grid cell level. Of the 10,158 grid cells of my sample, more than 21% have received some form of assistance from either source.[29]

Do donor institutions identify places most in need of additional infrastructure? I employ various indicators of aid provision in the standard grid cell level framework based on equation (9). I rely on two measures to quantify the prevalence of foreign aid: the total value of aid disbursements to a grid cell in 2011 US dollars and the number of distinct project sites within a given cell. I also put additional emphasis on infrastructure by separately analysing variation in funds going only to infrastructure projects in the transportation sector.

Table 3 reports results. Columns (1–4) investigate the spatial distribution of World Bank assistance. The estimates reveal seemingly opposing objectives between the Bank and the social planner. Negative estimates in columns (1) through (4) imply that grid cells receiving more World Bank assistance score lower on the discrimination index Λ_i. Every additional million US dollar flowing into an area is associated with the grid cell being about 0.004 standard deviations too well off. Focusing on transport sector projects only, results are qualitatively similar, yet much stronger. The average transport infrastructure project size of around 3 million US dollars goes to grid cells which stand to lose 0.02 standard deviations of welfare under the reallocation exercise. Similar effects hold on the extensive margin reported in columns (3) and (4). Columns (5–8) present very similar results for

[27] As Strange et al. point out, media reports are often based on initial press releases and do not necessarily follow up on the eventual disbursement of every promised dollar. In that, the dataset is likely to capture Chinese funding *commitments* rather than actual *disbursements*. Insofar as donors usually commit to more than they eventually deliver, these figures present an upper bound of realised development assistance. Furthermore, while AidData (2017) claim their dataset on World Bank projects to be exhaustive, the dataset on Chinese aid will naturally miss some unofficial flows, as significant parts of Chinese involvement remain untracked.

[28] Specifically, I exclude all projects with a precision code of more than 3 – this corresponds to projects only identified at province-level or above. The remaining entries are geo-coded either exactly (61%), within a 25 kilometre radius (4%), or with municipality-level precision (35%) (Strandow et al., 2011)

[29] All disbursements are adjusted to 2011 US dollars. For projects with multiple sites, I assume total disbursement value to be split evenly between sites. On average, these cells receive aid volumes of more than 30 million US dollars. The area receiving the most total World Bank funding is the grid cell containing Uganda's capital Kampala. The biggest beneficiary of Chinese development assistance is a grid cell in the south of Congo-Kinshasa, where Chinese funds of almost 5 billion US dollars helped construct a vast copper mining infrastructure.

Table 3: International aid and local infrastructure discrimination

	$\tilde{\Lambda}$: World Bank				$\tilde{\Lambda}$: China			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total disbursements (mil $)	-0.00480***				-0.000634***			
	(0.00115)				(0.000180)			
Total transport sector disbursements (mil $)		-0.00718*				-0.000726**		
		(0.00278)				(0.000276)		
Number of projects			-0.0287***				-0.0379***	
			(0.00500)				(0.00994)	
Number of transport sector disbursements				-0.0422***				-0.0744*
				(0.00819)				(0.0306)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	10158	10158	10158	10158	10158	10158	10158	10158
R2	0.164	0.163	0.166	0.165	0.163	0.162	0.163	0.163

Grid cell level estimations of equation (9) with z-scored local infrastructure discrimination $\tilde{\Lambda}_i$ as dependent variable and different measures of foreign aid flows into grid cells as explanatory covariates. Columns (1–4) investigate World Bank assistance. Column (1) analyses total disbursement value from World Bank projects approved from 1996–2014 in 2011 US dollars, which were completed by 2017. (2) only uses a subset of projects in the transport sector. (3)–(4) use the same data but focus on the number of distinct project sites within each grid cell. Columns (5–8) repeat the same estimations, but with data on Chinese aid projects between 2000–2011. Geography controls, consisting of altitude, temperature, average land suitability, malaria prevalence, yearly growing days, average precipitation, indicators for the 12 predominant agricultural biomes, indicators for whether a cell is a capital, within 25 KM of a natural harbour, navigable river, or lake, the fourth-order polynomial of latitude and longitude, and an indicator of whether the grid cell lies on the border of a country's network. Simulation controls are comprised of population, night lights, and ruggedness. Chinese aid data are more likely to reflect commitments rather than actual disbursements. Standard errors are clustered on the 3×3 degree level and are shown in parentheses.

Chinese aid. Chinese assistance also systematically flows into privileged cells, with intensive margin point estimates of the association ranging between a quarter and a tenth of the World Bank results. On the extensive margin, more Chinese projects are similarly associated with higher trade network imbalances. For each new development site financed by China in a certain cell, the social planner intervenes and allocates 0.04 to 0.07 standard deviations of welfare away from the cell (columns 7–8).

These relationships should by not interpreted as causal effects. Since the placement of aid projects is not random, numerous other channels could account for the patterns depicted in Table 3. The donor's investment strategies might for example be motivated by increasing returns to scale. If the World Bank believes in an environment with multiple equilibria, where small initial investments set in motion a dynamic of spillover externalities, labour migration, and follow-up investments, it is often the right decision to fund projects in places that will not immediately harness their full capabilities (Krugman, 1991; Duranton and Venables, 2017). These investments will necessarily appear inefficient in promoting optimal trade *today*, yet spur transformative development *tomorrow* (see Michaels et al., 2021). Embedding the reallocation exercise in a New Economic Geography framework of increasing returns and labour mobility might be a valuable extension to better evaluate specific place-based policies.

6 Conclusion

In this study, I have identified spatial inefficiencies in Africa's trade network. I first constructed a comprehensive economic topography of the entire continent, bringing together data from a variety of sources like satellites, census bureaus, and open source online routing services. I then presented a simple six-sector endowment network trade model and simulated the flow of goods through the internal geography formed by 10,000 African regions and almost 75,000 network connections. Harnessing the recent theoretical contribution by Fajgelbaum and Schaal (2020), I proceeded to endogenize the transport network in order to derive the unique optimally reorganised road network for every country in Africa.

In the second part of this study, I compared each country's current network to its hypothetically optimal one. I ranked countries by overall network efficiency and presented a fine-resolution spatial dataset quantifying which sub-national areas are disadvantaged by the status quo. I empirically investigated patterns of trade network imbalances over space and linked inefficiencies to persistent lock-in effects caused by colonial infrastructure investments and differential treatment on the basis of regional favouritism. I found no consistent association with ethnic discrimination. I also documented how development assistance by the World Bank and China has not targeted the regions identified as most in need of additional investment.

In contributing a comprehensive spatial measure on the differential provision of a primary public good covering an entire continent, my study provides the quantitative foundation for many more research questions pertaining to inequality over space. Future research designs could employ my dataset to analyse regional roots of conflict, political activism, social mobility, or subjective overall wellbeing. Another interesting avenue for inquiry could be to investigate whether infrastructure inefficiency spatially covaries with the provision of other public goods like education, health, or security.

Identifying spatial inefficiencies and understanding their historical, cultural, and political roots can be the first step in outlining effective place-based policies. Equipped with an unparalleled availability of spatial data and computing power, policymakers in Africa and around the world should feel empowered to combat local imbalances and design powerful interventions to better connect millions.

References

AidData (**2017**). World Bank Geocoded Research Release, Version 1.4.2. `http://aiddata.org/data/world-bank-geocoded-research-release-level-1-v1-4-2` (Accessed: 2018-04-06). Williamsburg, VA and Washington, DC

Alder, Simon (**2022**). Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development. Working Paper

Allen, Treb and Costas Arkolakis (**2014**). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics* 129(3), pp. 1085–1140

——— (**2022**). The Welfare Effects of Transportation Infrastructure Improvements. *The Review of Economic Studies* forthcoming

Asher, Sam and Paul Novosad (**2020**). Rural Roads and Local Economic Development. *American Economic Review*

Atkin, David and Dave Donaldson (**2015**). Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs. Working Paper, National Bureau of Economic Research

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang (**2017**). Roads, Railroads, and Decentralization of Chinese Cities. *The Review of Economics and Statistics* 99(3), pp. 435–448

Bernot, Marc, Vicent Caselles, and Jean-Michel Morel (**2009**). *Optimal Transportation Networks*, volume 1955 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg

Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen (**2011**). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics* 165(2), pp. 137–151

Burgess, Robin and Dave Donaldson (**2012**). Railroads and the Demise of Famine in Colonial India. Working Paper, LSE and NBER

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel (**2015**). The Value of Democracy: Evidence from Road Building in Kenya. *The American Economic Review* 105(6), pp. 1817–1851

Burnside, Craig and David Dollar (**2000**). Aid, Policies, and Growth. *The American Economic Review* 90(4), pp. 847–868

Cameron, A. Colin, Jonah Gelbach, and Douglas Miller (**2011**). Robust Inference with Multi-way Clustering. *Journal of Business & Economic Statistics* 29(2), pp. 238–249

Center, Socioeconomic Data and Applications (**2016**). Gridded Population of the World (GPW), v4 | SEDAC. `http://sedac.ciesin.columbia.edu/data/collection/gpw-v4` (Accessed: 2017-11-26)

Clemens, Michael A. and Michael Kremer (**2016**). The New Role for the World Bank. *Journal of Economic Perspectives* 30(1), pp. 53–76

Clemens, Michael A., Steven Radelet, Rikhil R. Bhavnani, and Samuel Bazzi (**2012**). Counting Chickens when they Hatch: Timing and the Effects of Aid on Growth. *The Economic Journal* 122(561), pp. 590–617

Collier, Paul, Martina Kirchberger, and Mans Söderbom (**2015**). The cost of road infrastructure in low-and middle-income countries. *The World Bank Economic Review* 30(3), pp. 522–548

Coşar, A. Kerem and Pablo D. Fajgelbaum (**2016**). Internal Geography, International Trade, and Regional Specialization. *American Economic Journal: Microeconomics* 8(1), pp. 24–56

De Luca, Giacomo, Roland Hodler, Paul A. Raschky, and Michele Valsecchi (**2018**). Ethnic favoritism: An axiom of politics? *Journal of Development Economics* 132, pp. 115–129

Donaldson, Dave (**2018**). Railroads of the Raj: Estimating the impact of transportation infrastructure. *The American Economic Review* 108(4), pp. 899–934

Donaldson, Dave and Richard Hornbeck (**2016**). Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal of Economics* 131(2), pp. 799–858

Doxsey-Whitfield, Erin, Kytt MacManus, Susana B. Adamo, Linda Pistolesi, John Squires, Olena Borkovska, and Sandra R. Baptista (**2015**). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography* 1(3), pp. 226–234

Dreher, Axel, Andreas Fuchs, Roland Hodler, Bradley Parks, Paul Raschky, and Michael Tierney (**2019**). Aid on demand: African leaders and the geography of China's foreign assistance. *Journal of Development Economics* 140, pp. 44–71

Dreher, Axel and Sarah Langlotz (**2020**). Aid and growth. New evidence using an excludable instrument. *Canadian Journal of Economics/Revue canadienne d'économique* 53(3), pp. 162–1198

Dreher, Axel and Steffen Lohmann (**2015**). Aid and growth at the regional level. *Oxford Review of Economic Policy* 31(3-4), pp. 420–446

Duranton, Gilles and Anthony J Venables (**2017**). Place-Based Policies for Development. Working Paper, The World Bank

Easterly, William, Ross Levine, and David Roodman (**2004**). Aid, Policies, and Growth: Comment. *The American Economic Review* 94(3), pp. 774–780

Faber, Benjamin (**2014**). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies* 81(3), pp. 1046–1070

Fajgelbaum, Pablo D and Cecile Gaubert (**2020**). Optimal Spatial Policies, Geography, and Sorting*. *The Quarterly Journal of Economics* 135(2), pp. 959–1036

Fajgelbaum, Pablo D. and Edouard Schaal (**2020**). Optimal Transport Networks in Spatial Equilibrium. *Econometrica* 88(4), pp. 1411–1452

Fiorini, Matteo, Marco Sanfilippo, and Asha Sundaram (**2021**). Trade liberalization, roads and firm productivity. *Journal of Development Economics* 153, p. 102712

Foster, Vivien and Cecilia Briceño-Garmendia (**2010**). *Africa's Infrastructure: A Time for Transformation*. Africa Development Forum. World Bank Publications, Washington DC

Galichon, Alfred (**2016**). *Optimal Transport Methods in Economics*. Princeton University Press

Gorton, Nicole and Elena Ianchovichina (**2022**). Trade Networks in Latin America: Spatial Inefficiencies and Optimal Expansions. Technical report, The World Bank

Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David N Weil (**2018**). The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade. *The Quarterly Journal of Economics* 133(1), pp. 357–406

Henderson, J. Vernon, Adam Storeygard, and David N Weil (**2012**). Measuring Economic Growth from Outer Space. *The American Economic Review* 102(2), pp. 994–1028

Herranz-Loncán, Alfonso and Johan Fourie (**2017**). "For the public benefit"? Railways in the British Cape Colony. *European Review of Economic History* pp. 1–28

Hodler, Roland and Paul A. Raschky (**2014**). Regional Favoritism. *The Quarterly Journal of Economics* 129(2), pp. 995–1033

Jedwab, Remi, Edward Kerby, and Alexander Moradi (**2017**). History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya. *The Economic Journal* 127(603), pp. 1467–1494

Jedwab, Remi and Alexander Moradi (**2016**). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *The Review of Economics and Statistics* 98(2), pp. 268–284

Krugman, Paul (1991). History Versus Expectations. *The Quarterly Journal of Economics* 106(2), pp. 651–667

Limao, Nuno and Anthony J. Venables (2001). Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review* 15(3), pp. 451–479

Michaels, Guy, Dzhamilya Nigmatulina, Ferdinand Rauch, Tanner Regan, Neeraj Baruah, and Amanda Dahlstrand (2021). Planning Ahead for Better Neighborhoods: Long-Run Evidence from Tanzania. *Journal of Political Economy* 129(7), pp. 2112–2156

Michalopoulos, Stelios and Elias Papaioannou (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica* 81(1), pp. 113–152

——— **(2014)**. National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics* 129(1), pp. 151–213

——— **(2016)**. The Long-Run Effects of the Scramble for Africa. *The American Economic Review* 106(7), pp. 1802–1848

Murdock, George Peter (1959). *Africa: its peoples and their culture history*. McGraw-Hill, New York; London

Nugent, Paul and Hugh Lamarque (2022). *Transport Corridors in Africa*. James Currey, Rochester, NY. ISBN: 1847012949 Publisher: Boydell and Brewer

Nunn, Nathan and Nancy Qian (2014). US Food Aid and Civil Conflict. *The American Economic Review* 104(6), pp. 1630–1666

Rajan, Raghuram G and Arvind Subramanian (2008). Aid and growth: what does the cross-country evidence really show? *The Review of Economics and Statistics* 90(4), pp. 643–665

Santamaria, Marta (2022). Reshaping Infrastructure: Evidence from the division of Germany. Working Paper

Storeygard, Adam (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies* 83(3), pp. 1263–1295

Strandow, Daniel, Michael Findley, Daniel Nielson, and Josh Powell (2011). The UCDP and AidData codebook on georeferencing aid Version 1.1. Technical report, Department of Peace and Conflict Research, Uppsala University, Uppsala

Strange, Austin M., Axel Dreher, Andreas Fuchs, Bradley Parks, and Michael J. Tierney (2017). Tracking Underreported Financial Flows: China's Development Finance and the Aid–Conflict Nexus Revisited. *Journal of Conflict Resolution* 61(5), pp. 935–963

Swisher, I. V. (2017). Reassessing railroads and growth: Accounting for transport network endogeneity. Working Paper, Faculty of Economics, University of Cambridge

The Economist (2015). All aboard. *The Economist* February 26, 2015

The World Bank (2017). The World Bank Annual Report 2017. Technical report, The World Bank, Washington DC

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker, and Luc Girardin (2015). Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family. *Journal of Conflict Resolution* 59(7), pp. 1327–1342

Appendix

A Numerically solving the planner's problem

The full planner's problem on page 5 consists of a very large number of choice variables and hence requires vast computation efforts when solved directly. Fortunately, Fajgelbaum and Schaal (2020) provide guidance on how to transform this *primal* problem into its much simpler *dual* representation. The following section illustrates how to use their derivation to numerically solve my version of the model.

To show how a unique global optimum exists, first note that every constraint of the social planner's problem is convex but potentially for the *Balanced Flows Constraint*. However, the introduction of congestion causes even the *Balanced Flows Constraint* to be convex if $\beta > \gamma$. To see this, note that every part of the lengthy constraint is linear, but for the interaction term $Q^n_{i,k}\tau^n_{i,k}(Q^n_{i,k}, I_{i,k})$ representing total trade costs. Since $\tau^n_{i,k}$ was parameterised as in (1), this expands to

$$Q^n_{i,k}\tau^n_{i,k}(Q^n_{i,k}, I_{i,k}) = \delta^\tau_{i,k}\frac{(Q^n_{i,k})^{1+\beta}}{I^\gamma_{i,k}} \tag{A.1}$$

which is convex if $\beta > \gamma$. Under this condition, the social planner's problem is to maximise a concave objective over a convex set of constraints, guaranteeing that any local optimum is indeed a global maximum.[A.1] $\beta > \gamma$ describes a notion of congestion dominance: increased infrastructure expenditure might alleviate the powers of congestion, but it can never overpower it. It precludes corner solutions in which all available concrete is spent on one link, all but washing away trade costs and leading to overwhelming transport flows on this one edge. If $\beta > \gamma$, geography always wins.

Consider first the full Lagrangian of the primal planner's problem

$$\begin{aligned}\mathcal{L} = &\sum_i L_i u(c_i) - \sum_i \lambda^C_i \left[L_i c_i - \left(\sum_{n=1}^N (C^n_i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \right] \\ &- \sum_i \sum_n \lambda^P_{i,n} \left[C^n_i + \sum_{k\in N(i)} Q^n_{i,k}(1 + \tau^n_{i,k}(Q^n_{i,k}, I_{i,k})) - Y^n_i - \sum_{j\in N(i)} Q^n_{j,i} \right] \\ &- \lambda^I \left[\sum_i \sum_{k\in N(i)} \delta^i_{i,k} I_{i,k} - K \right] - \sum_i \sum_{k\in N(i)} \zeta^S_{i,k} \left[I_{i,k} - I_{k,i} \right] \\ &+ \sum_i \sum_{k\in N(i)} \sum_n \zeta^Q_{i,k,n} Q^n_{i,k} + \sum_i \sum_n \zeta^C_{i,n} C^n_i + \sum_i \sum_n \zeta^c_i c_i - \sum_i \sum_{k\in N(i)} \zeta^I_{i,k} \left[4 - I_{i,k} \right]\end{aligned} \tag{A.2}$$

This is a function of the choice variables $(C^n_i, Q^n_{i,k}, c_i, I_{i,k})$ in all dimensions $\langle i,k,n\rangle$ and the Lagrange multipliers $(\lambda^C, \lambda^P, \lambda^I, \zeta^Q, \zeta^C, \zeta^c, \zeta^I)$ also in $\langle i,k,n\rangle$. Standard optimisation yields first-order

[A.1] This is Fajgelbaum and Schaal Proposition 1.

conditions which can be collapsed to the following set of equations

$$\begin{aligned}
c_i &= \left(\frac{1}{\alpha}\Big(\sum_{n'}(\lambda^P_{i,n'})^{1-\sigma}\Big)^{\frac{1}{1-\sigma}}\right)^{\frac{1}{\alpha-1}} \\
C^n_i &= \left[\frac{\lambda^P_{i,n}}{(\sum_{n'}(\lambda^P_{i,n'})^{1-\sigma})^{\frac{1}{1-\sigma}}}\right]^{-\sigma} L_i c_i \\
Q^n_{i,k} &= \left[\frac{1}{1+\beta}\frac{I^\gamma_{i,k}}{\delta^\tau_{i,k}}\max\Big\{\frac{\lambda^P_{k,n}}{\lambda^P_{i,n}}-1,0\Big\}\right]^{\frac{1}{\beta}} \\
I_{i,k} &= \max\Bigg\{\Bigg[\frac{\kappa}{\lambda^I(\delta^I_{i,k}+\delta^I_{k,i})}\Bigg(\sum_n \max\Big\{(\delta^\tau_{i,k})^{-\frac{1}{\beta}}\lambda^P_{i,n}\Big(\frac{\lambda^P_{k,n}}{\lambda^P_{i,n}}-1\Big)^{\frac{1+\beta}{\beta}},0\Big\} \\
&\qquad + \sum_n \max\Big\{(\delta^\tau_{k,i})^{-\frac{1}{\beta}}\lambda^P_{k,n}\Big(\frac{\lambda^P_{i,n}}{\lambda^P_{k,n}}-1\Big)^{\frac{1+\beta}{\beta}},0\Big\}\Bigg)\Bigg]^{\frac{\beta}{\beta-\gamma}},4\Bigg\}
\end{aligned} \tag{A.3}$$

These directly follow the more general framework outlined in the technical appendix of Fajgelbaum and Schaal applied to my version of the model. In the final equation denoting optimal infrastructure supply, $\kappa = \gamma(1+\beta)^{-\frac{1+\beta}{\beta}}$, and the multiplier λ^I is such that adherence to the *Network Building Constraint* is ensured. Note that there is a typo in the original authors' paper which prints one of the exponents as $(\delta^\tau_{i,k})^{\frac{1}{\beta}}$ when it should be $(\delta^\tau_{i,k})^{-\frac{1}{\beta}}$. Through these algebraic manipulations, I have expressed all choice variables as functions of merely the Lagrange parameters λ^P over dimensions $\langle i,k,n\rangle$. I can hence recast the entire Lagrangian in much simpler form as

$$\begin{aligned}
\mathcal{L}(\lambda, x(\lambda)) &= \sum_i L_i u(c_i(\lambda)) \\
&- \sum_i\sum_n \lambda^P_{i,n}\Big[C^n_i(\lambda) + \sum_{k\in N(i)} Q^n_{i,k}(\lambda)(1+\tau^n_{i,k}(Q_{i,k}(\lambda)^n, I_{i,k}(\lambda))) - Y^n_i - \sum_{j\in N(i)} Q^n_{j,i}(\lambda)\Big]
\end{aligned} \tag{A.4}$$

where $x(\lambda)$ denote the choice variables as functions of the Lagrange parameters as derived above. Fajgelbaum and Schaal note that thanks to complementary slackness, all other constraints can be readily dropped from consideration and only the *Balanced Flows Constraint* remains part of the problem.

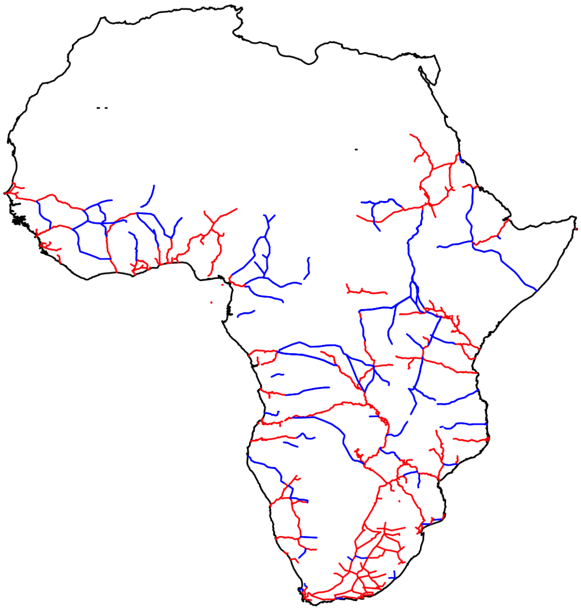
As Fajgelbaum and Schaal further explain, the dual of this problem can now be conceived as the minimisation of

$$\min_{\lambda\geq 0} \mathcal{L}(\lambda, x(\lambda))$$

which is an optimisation problem over merely $\|\lambda^P\| = I \times N$ variables. Fajgelbaum and Schaal interpret λ^P as a field of prices varying over goods and locations. I am left only to minimise equation (A.4) to obtain the price-field λ^P. I implement constrained optimisations within the `fmincon` environment in MATLAB and achieve fairly fast convergence. Solving for smaller networks (like Rwanda or Djibouti) is a matter of seconds, yet the largest countries (Algeria, Angola, DRC, and Sudan) each take about a day of computation time (on a five-year old device, nonetheless). Plugging the derived λ^P parameters into the various FOCs in (A.3) yields the optimal transport network $I_{i,k}$, trade flows between locations $Q^n_{i,k}$, and consumption patterns C^n_i and c_i.

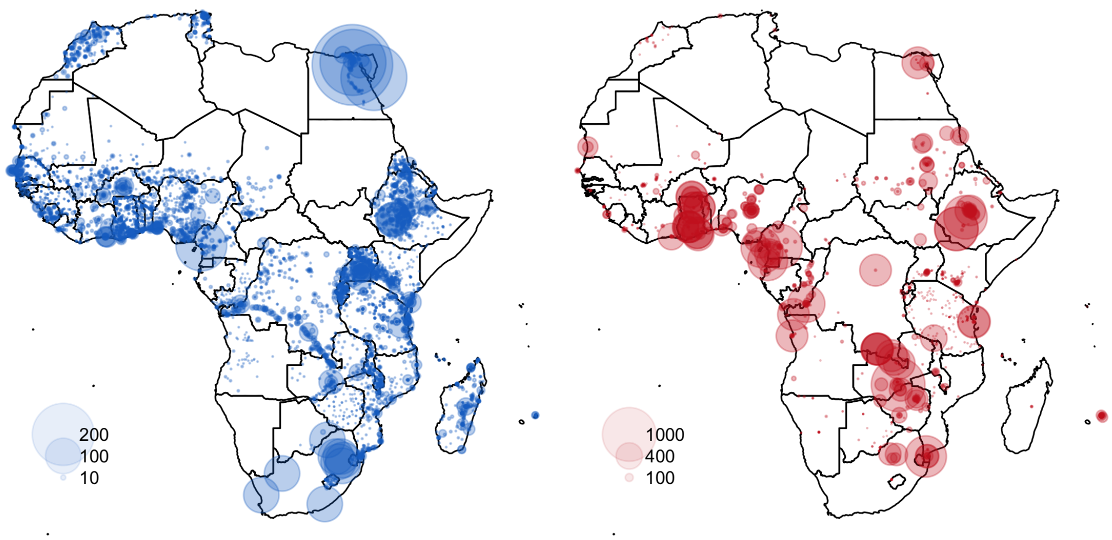
B Additional figures and tables

Figure A.1: Colonial railway network



Maps displaying the network of railway lines (red) and placebo railroads (blue). Data from Jedwab and Moradi (2016) and Herranz-Loncán and Fourie (2017). Railroads built by the colonial powers between 1890 and 1960 are printed in red. Lines that were initially planned but never actually built are printed in blue.

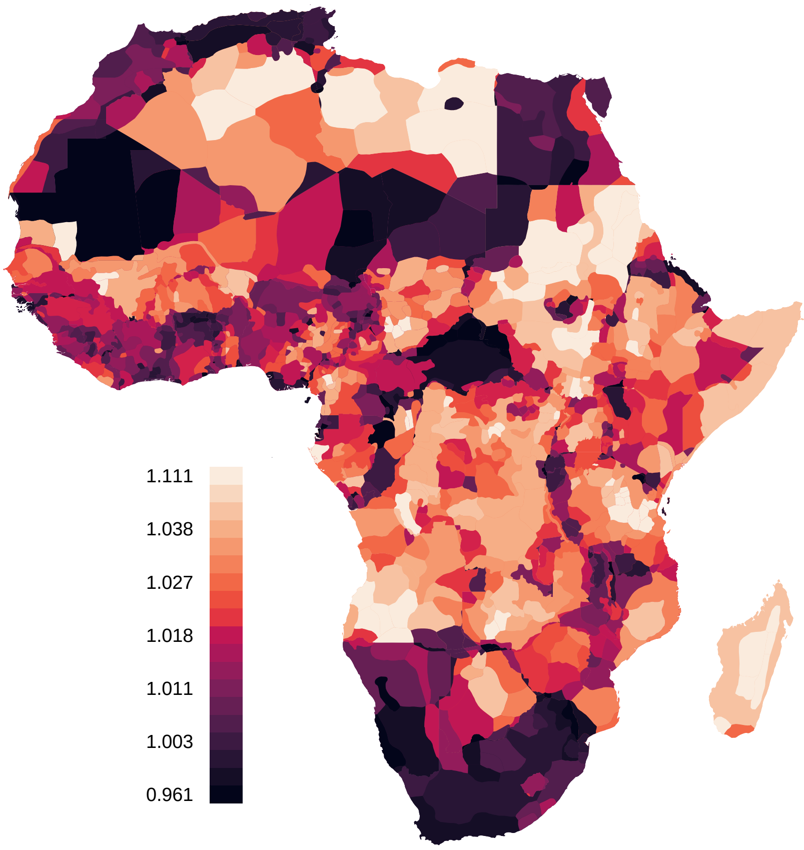
Figure A.2: Spatial distribution of development aid projects to African nations



(a) World Bank aid (b) Chinese aid

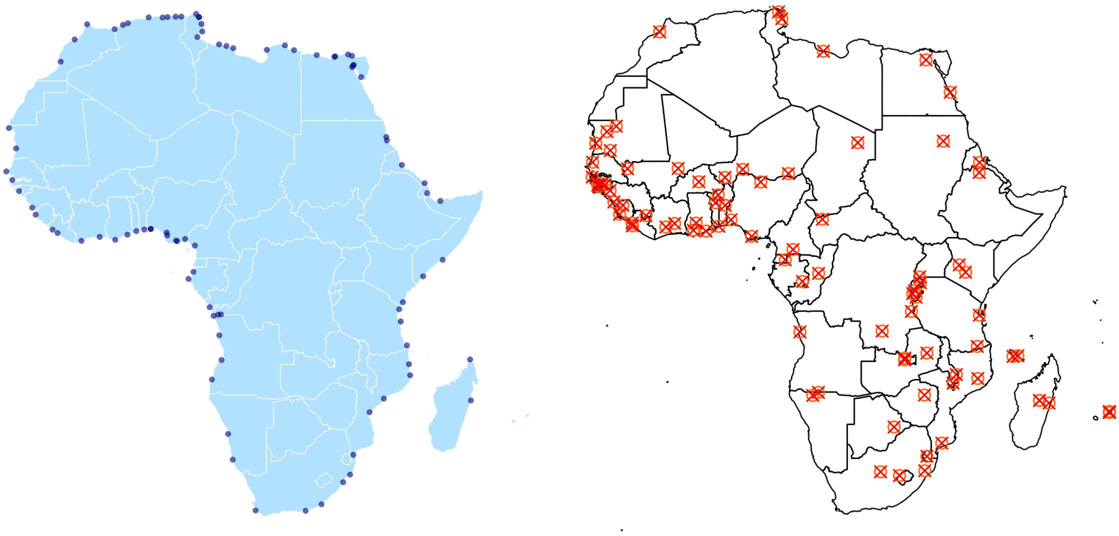
Foreign aid projects funded by the World Bank (A.2a) *and China* (A.2b)*. Each dot represents one project site with radius proportional to the logarithm of total disbursements flowing to each site. World Bank data comprise all projects approved between 1996–2014. Chinese data include tracked projects between 2000–2011. Map only depicts projects coded with sufficient precision to not be excluded (see text). If a project has multiple sites, total disbursements are assumed evenly distributed between locations. Data from AidData (2017) and Strange et al. (2017). Legend denotes disbursement values in million 2011 US dollars. Note that the legends have different scales.*

Figure A.3: Λ_h over ethnic homelands



Spatial distribution of local infrastructure discrimination index Λ_h, aggregated over ethnic homelands. Unit of observation is pre-colonial homelands as initially defined in an ethnolinguistic map by Murdock (1959) intersected by current political borders (following Michalopoulos and Papaioannou, 2016). Λ_h is from grid cell level.

Figure A.4: Further spatial data used



(a) Locations of ports

(b) Leaders' birthplaces

Spatial distribution of ports and leaders' birthplaces across the continent. Ports data is hand-coded from Lloyd's list at https://directories.lloydslist.com/port and corresponds to the 90 biggest ports in Africa. Birthplace data from Dreher et al. (2019).

Table A.1: Colonial railroads: No simulation controls

	Infrastructure discrimination $\tilde{\Lambda}$ (z-scored)				Real	Placebo
	(1)	(2)	(3)	(4)	(5)	(6)
50 KM of Colonial Railroads	-0.241***					
	(0.0500)					
50 KM of Colonial Placebo Railroads		-0.183				
		(0.154)				
50 KM of Colonial Railroads for Military Purposes			-0.285***			
			(0.0739)			
50 KM of Colonial Railroads for Mining Purposes				-0.159**		
				(0.0585)		
<10KM to railroad					-0.327***	-0.0960
					(0.0652)	(0.0511)
10-20KM to railroad					-0.374***	-0.0670
					(0.0716)	(0.0636)
20-30KM to railroad					-0.134*	0.0110
					(0.0666)	(0.0654)
30-40KM to railroad					0.0355	0.0630
					(0.0544)	(0.0500)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Controls	No	No	No	No	No	No
N	10158	10158	10158	10158	10158	10158
R2	0.163	0.160	0.162	0.160	0.166	0.160

Table A.2: Ethnic favoritism: No simulation controls

	$\tilde{\Lambda}$: entire sample		$\tilde{\Lambda}$: excluding capitals		$\tilde{\Lambda}$: ethnic homeland level		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total years in power	-0.0111*		-0.0135**				
	(0.00442)		(0.00517)				
Ever in power dummy		-0.275*		-0.331**			
		(0.106)		(0.115)			
Discriminated against					-0.194		
					(0.157)		
Excluded from government						-0.363**	
						(0.122)	
Involved in ethnic war							-0.227
							(0.178)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Controls	No	No	No	No	No	No	No
N	10158	10158	10109	10109	494	494	494
R2	0.160	0.160	0.160	0.161	0.253	0.262	0.255

Table A.3: International aid: No simulation controls

	$\tilde{\Lambda}$: World Bank				$\tilde{\Lambda}$: China			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total disbursements (mil $)	-0.00408***				-0.000594***			
	(0.000876)				(0.000176)			
Total transport sector disbursements (mil $)		-0.00654**				-0.000661*		
		(0.00201)				(0.000256)		
Number of projects			-0.0207***				-0.0311***	
			(0.00395)				(0.00903)	
Number of transport sector disbursements				-0.0350***				-0.0625*
				(0.00733)				(0.0287)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simulation Controls	No	No	No	No	No	No	No	No
N	10158	10158	10158	10158	10158	10158	10158	10158
R2	0.162	0.161	0.163	0.162	0.161	0.160	0.161	0.161