I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13071

# Key Players in Economic Development

Ashani Amarasinghe
Roland Hodler
Paul A. Raschky
Yves Zenou

MARCH 2020

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13071

# Key Players in Economic Development

**Ashani Amarasinghe**
*Monash University*

**Roland Hodler**
*University of St.Gallen, CEPR, CESifo and OxCarre*

**Paul A. Raschky**
*Monash University*

**Yves Zenou**
*Monash University, CEPR and IZA*

MARCH 2020



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## Key Players in Economic Development*

This paper analyzes the role of networks in the spatial diffusion of local economic shocks in Africa. We show that road and ethnic connectivity are particularly important factors for diffusing economic spillovers over longer distances. We then determine the key players, i.e., which districts are key in propagating local economic shocks across Africa. Using these results, we conduct counterfactual policy exercises to evaluate the potential gains from policies that increase economic activity in specific districts or improve road connectivity between districts.

**JEL Classification:** O13, O55, R12

**Keywords:** networks, spatial spillovers, key player centrality, natural resources, transportation, Africa

**Corresponding author:**
Yves Zenou
Department of Economics
Monash University
Caulfield East VIC 3145
Australia
E-mail: yves.zenou@monash.edu

* We thank Markus Brückner, Matthew Leister, George Marbuah, Martin Ravallion and Mathias Thoenig as well as conference and seminar participants at the Australasian Development Economics Workshop, Australasian Public Choice Conference, CSAE Conference, Development Economics Network Switzerland, Econometric Society European Meeting, German Development Economics Conference, IDEAs-CDES Summer School in Development Economics, Swiss Development Economics Workshop, Monash University and University of Melbourne for their helpful comments. We also thank Shaun Astbury, Thu Phan, and Georgie Stokol for excellent research assistance. Ashani Amarasinghe gratefully acknowledges support of the Australian Government through its Research Training Program Scholarship. Roland Hodler and Paul Raschky gratefully acknowledge financial support from the Australian Research Council (ARC Discovery Grant DP150100061). This research was supported in part by the Monash eResearch Centre and eSolutions-Research Support Services through the use of the Monash Campus HPC Cluster.

# 1 Introduction

In recent decades, the majority of African countries have experienced an unprecedented period of aggregate economic growth. However, the gains from this rise in aggregate income have been unequally distributed between individuals and regions within those countries (Beegle et al., 2016). The reason for this could be that in many African countries, economic activity is concentrated in a few geographic areas, and the geography, poor transport infrastructure, and ethnic heterogeneity may limit the extent of spatial economic spillovers (e.g., Brock et al., 2001; Crespo Cuaresma, 2011).

The aim of this paper is to estimate the extent of spatial economic spillovers between African districts, to highlight the roles of geographic, transport, and ethnic networks in the context of regional economic development, and to determine which districts are key in propagating local economic shocks across Africa.

We constructed a balanced panel dataset of 5,944 African districts (ADM2, second subnational level) and yearly data from 1997–2013, in which our measure of local economic activity was nighttime light intensity. The basic econometric framework is a spatial Durbin model that allows for spatial autoregressive processes with the dependent and explanatory variables. We interpret the estimated coefficient of the spatial lag of the dependent variable in this model as the effect of a district's connectivity on its own economic activity. Our preferred specifications include time-varying controls as well as district and country-year fixed effects to account for all time-invariant differences across districts and country-year specific shocks that affect every district in a country and year, respectively.

The major empirical challenge is that the estimated parameter is likely to be biased due to reverse causality and time-varying omitted variables. We address this problem by applying an instrumental variables (IV) strategy, which relies on cross-sectional variation in the neighboring districts' mining opportunities and fluctuations in the world price of the minerals extracted in these districts as the sources of exogenous temporal variation in the examined districts' performance. Note, however, that the primary goal is not to obtain estimates to evaluate the importance of the spatial lags but to have a well-identified spatial econometric model to show how economic shocks propagate though a network. We show that, individually, geographic, ethnic, and road connectivity all increase local economic activity, however, they impact local economic activity in different ways.

We then turn to measuring the network centrality of all the districts in Africa. Our estimated coefficients on the spatial lag variable allow us to calculate Katz-Bonacich and key-player network centralities. Based on the key-player centrality, we determine the "key" districts in African countries, i.e., the districts that contribute most to economic activity across Africa. These districts are typically characterized by high local economic activity as well as good connectivity.

Our paper contributes to two main strands of the literature. First, we contribute to the empirical literature on the effects of networks in economics, particularly key players.[1] There are very few papers that empirically calculate key players in networks (exceptions include König et al. (2017), Lindquist and Zenou (2014); Liu et al. (2020); for an overview, see Zenou (2016)) but they do not determine them for a whole continent. Moreover, in these studies, the key players are determined at the individual level, which creates network formation issues that are difficult to deal with. In our paper, this is not the case because the links between districts are pre-determined by their locations.

Second, we contribute to the literature that studies the importance of transport networks and, more broadly, market access for subnational economic development in Africa. Studies in this area often focus on the construction of new highways, e.g., in China or India (Banerjee et al. (2012), Faber (2014), Alder (2015)). Our counterfactual policy exercise of improving road connectivity between districts is related to Alder (2015), who has studied the consequences of a counterfactual Indian highway network that mimics the design of the Chinese highway network. Like us, Storeygard (2016), Bonfatti and Poelhekke (2017), and Jedwab and Storeygard (2018) have focused on road networks in Africa. However, we differ from these papers by focusing on the importance of roads in shaping the spatial diffusion of economic shocks across Africa. Those spillover effects could be driven by increased market access (e.g., Donaldson (2018), Donaldson and Hornbeck (2016)), which is of importance for the road and geographic networks. However,

[1] For overviews of the economics of networks, see Jackson (2008) and Jackson et al. (2017).

economic spillovers also occur because of numerous other channels, such as technology diffusion or transfer payments, and our aim is to estimate the *aggregate spillover effects* of those various channels.

# 2 Estimating Spatial Spillover Effects for African Districts

The first step of our analysis consisted of estimating the paramaters of a standard Spatial Durbin model. We constructed a balanced panel dataset in which the units of observation were the administrative regions in Africa at the second subnational level (ADM2), which we labelled *districts*.[2] The final dataset consisted of yearly observations for 5,944 districts from 53 African countries over the period 1997–2013.[3] The average (median) size of a district was 39km$^2$ (6km$^2$), and the average (median) population was around 150,000 (55,000). In the Online Appendix B, we provide a detailed description of all variables. Here, we provide a quick description of our main variables.

## 2.1 Data description

Our outcome variable was nighttime light intensity based on satellite data from the the National Oceanic and Atmospheric Administration (NOAA). From the raw satellite data, NOAA removes readings affected by cloud coverage, northern or southern lights, readings that are likely to reflect fires, other ephemeral lights, and background noise. The objective is that the reported nighttime lights are primarily man-made. The NOAA provides annual data for the time period from 1992 onward in output pixels that correspond to less than one square kilometer. The data is presented on a scale from 0 to 63, with higher values implying more intense nighttime lights. Work by Henderson et al. (2012) and Hodler and Raschky (2014) has already shown that nighttime light is a good proxy for economic

[2]In a robustness test presented in Table H1 in the Online Appendix, we used $0.5 \times 0.5$ degree grid cells instead of ADM2 regions and show that all our results are qualitatively the same.

[3]Table B3 in the Online Appendix lists all the African countries in our sample and provides the number of districts per country.

development at the national and subnational level, respectively, which makes this suitable for the present work's purpose.

To construct our dependent variable, $Light_{it}$, we took the logarithm of the average nighttime light pixel value in district $i$ and year $t$. To avoid losing observations with a reported nighttime light intensity of zero, we followed Michalopoulos and Papaioannou (2013, 2014) and Hodler and Raschky (2014) in adding 0.01 before taking the logarithm. We construct three connectivity matrices to measure spatial spillovers.

*Ethnic connectivity* - To measure ethnic connectivity between districts, we first overlayed the district (ADM2) boundaries with the boundaries of the ethnic homelands from Murdock (1959). Each district was assigned the ethnicity of the ethnic homeland in which it is located. For districts that fell into more than one ethnic homeland, we assigned the ethnicity of the ethnic homeland that covered the largest part of the district. We then constructed our ethnic connectivity matrix, $\omega_{i,j}$, where an element was 1 if the ethnicity in district $i$ was the same as the ethnicity in district $j$ and 0 otherwise.

*Geographic connectivity* - We based the weighting matrix for geographic connectivity on geographic distance. We constructed this weighting matrix as follows. First, we calculated the centroid of each district. Second, we calculated the geodesic distance $d_{i,j}$ connecting the centroids of districts $i$ and $j$. Third, following Acemoglu et al. (2015), we measured the variability of the altitude $e_{i,j}$ along the geodesic connecting the centroids of districts $i$ and $j$ and used elevation data from GTOPO30. Finally, we calculated the inverse of the altitude-adjusted geodesic distance as $\tilde{d}_{i,j} = 1/d_{i,j}(1 + e_{i,j})$. The main specification used a cutoff of 70 km (for reasons made explicit below). In this case, we set the spatial weight as $\omega_{i,j} = 1/\tilde{d}_{i,j}$ if the geodesic distance $d_{i,j}$ was less than 70 km and $\omega_{i,j} = 0$ otherwise.

*Road connectivity* - To construct connectivity via the road network, we obtained data from OpenStreetMap (OSM). We accessed the OSM data in early 2016 and extracted information about major roads (e.g., highways and motorways) for the African continent. We intersected these roads with the district boundary polygons and generated a network graph of the road network. The resulting road connectivity matrix assigns a value equal to the inverse of the road distance in km between districts $i$ and $j$ if they are connected via a major road and 0 if they are not connected. We, again, constructed different weighting matrices by truncating at different distance cutoffs.

Our identification strategy made use of cross-sectional information concerning the location of mining projects and temporal variations in the world prices of the corresponding minerals. Information on mining activity came from the SNL Minings & Metals database. We used the point locations to assign the mining projects to districts and to identify all districts where a mine was active for at least one year during our sample period. Across Africa, 4% of all districts are mining districts. The indicator variable $Mine_i^r$ is equal to one if district $i$ had a mining project that extracted resource $r$ and was active for at least one year during our sample period. Data on the world prices of minerals was sourced from the World Bank, IMF, USGS, and SNL.[4] $Price_t^r$ is the logarithm of the yearly nominal average price of resource $r$ in USD.

## 2.2 Empirical strategy

The aim of the empirical analysis was to estimate the following equation:[5]

$$Light_{it} = \sum_{k=1}^{3}\sum_{j=1}^{J} \rho_k \omega_{k,i,j} Light_{jt} + \mathbf{X}_{it}\boldsymbol{\beta} + \sum_{k=1}^{3}\sum_{j=1}^{J}\sum_{m=1}^{M} \rho_k^m \omega_{k,i,j} X_{jt}^m + \alpha_i + CT_{ct} + \epsilon_{it}, \quad (1)$$

where $\omega_{1,i,j}$ is the $(i,j)$ cell of the adjacency matrix based on geographic connectivity; $\omega_{2,i,j}$ is the $(i,j)$ cell of the adjacency matrix based on road connectivity; $\omega_{3,i,j}$ is the $(i,j)$ cell of the adjacency matrix based on ethnic connectivity; $\mathbf{X}_{it} = \left(X_{it}^1, ..., X_{it}^M\right)$, the $(1 \times M)$ vector of time-variant, district-level characteristics, and $\boldsymbol{\beta} = \left(\beta^1, ..., \beta^M\right)^T$, a $(M \times 1)$ vector of parameters; $\rho_k^m$ are the coefficients of the spatial lags; $\alpha_i$ and $CT_{ct}$ are district and country-year fixed effects, respectively; and $\epsilon_{it}$ is an error term that is assumed to be $\epsilon_{it} \sim N(0, \sigma^2 I_n)$.[6] We row-normalized the adjacency matrices, i.e., we normalized them so that the sum of each row became equal to 1.

[4]Refer to Table B4 in the Online Appendix for more information on the data sources.
[5]In Appendix A, we propose a simple model that microfounds equation (1).
[6]Table D2 in the Online Appendix presents estimation results without control variables and with year rather than country-year fixed effects.

In this specification, local spillovers to other districts not only operate through the spatial lag of the dependent variable but also occur due to spatial autoregressive processes in the explanatory variables as well (Spatial Durbin Model). In the spatial context, spillovers might not only run from district $j$ to $i$ but also from $i$ to $j$. In addition, economic activity (and, therefore, $Light_{jt}$) might also be simultaneously determined by other unobserved shocks. Thus, estimating equation (1) using OLS can yield biased and inconsistent estimates.

To address this problem, we estimated a 2SLS model that exploits exogenous variation in the economic value of mineral resources in the mining districts.[7] The idea was that more valuable mining districts increase spillover effects such that the level of economic activity in neighboring districts will be positively affected. In particular, in the first stage, we used interaction terms between the time-invariant indicators of mining activity and the time-variant exogenous world prices for minerals as instrumental variables:

$$Light_{jt} = \gamma MP_{jt} + \mathbf{X}_{jt}\boldsymbol{\beta} + \sum_{k=1}^{3}\sum_{i=1}^{J}\sum_{m=1}^{M} \rho_k^m \omega_{k,j,i} X_{it}^m + \alpha_j + CT_{ct} + u_{jt}, \tag{2}$$

where

$$MP_{jt} \equiv \frac{1}{R_j}\sum_{r=1}^{R_j}(Mine_j^r \times Price_t^r), \tag{3}$$

with $R_j = \sum_r Mine_j^r$ being the number of different minerals extracted in district $j$. Hence, for each mining region, this instrumental variable captures the average of the world prices (in logs) for all minerals that are extracted in the relevant district at some time during the sample period. For all other districts, this instrumental variable is zero.

In order for this instrument to be relevant, it is key that fluctuations in world mineral prices have a first-order effect on the mining districts' economies. We know from Brückner et al. (2012) that oil price shocks lead to an increase in per capita GDP growth at the country-level (also see Brunnschweiler and Bulte (2008)) while works by Aragon and Rud (2013) and Caselli and Michaels (2013) have shown that resource extraction can have positive effects on local economic development.

[7]The results are very similar if we estimate our model using a quasi-maximum likelihood approach as suggested by Anselin (1988).

In our sample, we found that only 4% of all districts were mining districts. Minerals, however, account for a large proportion of export earnings in many African countries, especially strategically important minerals such as diamonds, gold, uranium, and bauxite. Given the small number of mining districts and the importance of minerals at the country level, it seems plausible to assume that fluctuations in world mineral prices are relevant for mining districts.

With respect to the instrument's validity, our identification strategy rests upon the assumption that price shocks in the mining sector in district $j$ affect $Light_{it}$ in district $i$ only through $Light_{jt}$. We took a number of measures to mitigate the risk of the exclusion restriction being violated. First, we included district-fixed effects in equation (2), which absorb all time-invariant characteristics at the district level, including suitability for mining activity. The vector of country-year fixed effects accounts for any time-variant factors that might simultaneously drive mineral prices and aggregated economic development. Second, the work by Berman et al. (2017) showed that mining activity could lead to increased conflict as parties dispute over ownership of lucrative mines. This, in turn, could adversely affect economic activity. Therefore, we controlled for district-level conflict events in our specifications. Third, the exclusion restriction also relies on the assumption of exogeneity of world prices, i.e., no single district can affect the world price of a commodity. For this reason, we conducted a robustness check of our specifications by excluding countries in the top ten list of producers for any mineral.

The statistical inference in our setting is further complicated by the clustering structure of the error terms in our econometric model. The traditional spatial clustering approach proposed by Conley (1999) imposes the same spatial kernel (geographic distance) to all units in the sample. However, our empirical model does not only assume dependence based on geographic distance but also relatedness through ethnic and road networks. Therefore, we applied a novel estimator developed by Colella et al. (2018) that allowed us to account for dependence across our observations' error terms in a more flexible form.[8] In practice, we corrected for clustering at the network level where observations

[8]This procedure was implemented in Stata 14 using the "acreg" command by Colella et al. (2018).

were assumed to belong to the same cluster once they are linked through at least one of the three networks (geography, ethnicity, roads).[9]

Our setting implies that we needed to estimate the local average treatment effect (LATE) of economic shocks related to windfalls in natural resource rents. These shocks and the resulting spatial spillovers may have very particular effects on consumption, investment, and government expenditure. Moreover, we estimated the LATE for districts with a certain network proximity to mining districts. In our sample, 31% of all districts were within a geodesic or road distance of less than 70 km to a mining district or shared an ethnic connection with a mining district. These districts might systematically differ from the average African district. For these reasons, one needs to be careful when drawing more general policy conclusions based on the estimated spatial spillover effects.

In addition, it is *a priori* unclear whether mining-related income shocks only generate positive spillover effects for other districts. Windfalls in natural resource rents in one district could lead to migration of labor and capital from other connected districts into the mining district. A mining boom could also lead the government to shift public expenditure and infrastructure projects away from nearby districts into the mining district. As such, the estimated parameters of the spatial lags represent the net effect from mining-related economic shocks in connected districts.

## 2.3 Estimation results

Table 1 presents our estimates from equations (1) and (2).[10] We start with specifications that include each weighting matrix individually. Column (1) provides the results of the OLS estimates for *ethnic connectivity* while column (2) provides the comparable IV estimates. The OLS estimates show that coefficient $\rho_1$ (coefficient on $Ethnicity\,W\,Light_{jt}$) was positive and statistically significant. In column (2), we present comparable IV estimates. The coefficient of interest in the *first stage* of the IV estimate, $\gamma$, was positive and statistically significant. This, along with the high $F$ statistics of the first stage, indicates

[9]Table E3 in the Online Appendix presents the results using the standard Conley-type spatial clustering approach (Conley 1999). Our results remain the same.

[10]Table 1 only reports the coefficients on the variables of main interest to improve the readability of the table. Table D1 in the Online Appendix reports the coefficients on the control variables.

that our instrumental variable was a strong predictor of economic activity in district $j$. The coefficient of interest of the *second stage*, $\rho_1$, was positive but not statistically significant.

Second, we consider the *inverse-distance connectivity.* In Figure D5 in the Online Appendix, we show that the coefficient of interest from the IV estimates declined in magnitude in the cutoff distance and became statistically insignificant for cutoff distances beyond 70 km. As a result, in all our estimations, we truncated the adjacency matrix at 70 km. In columns (3) and (4) of Table 1, we focus on geographic connectivity and, therefore, use the weighting matrix based on the inverse of the altitude-adjusted geodesic distance between districts $i$ and $j$, truncating the matrix at 70 km. The coefficient of interest, $\rho_2$, was positive and statistically significant at the 1% level in both the OLS and the IV estimations.

Third, in columns (5) and (6), we focus on *road connectivity* based on our matrix of inverse road distances, again, truncating the matrix at 70 km. The coefficient of interest, $\rho_3$, was positive and statistically significant at the 1% level in both the OLS and the IV estimations. As in Figure D5 in the Online Appendix, the coefficient of interest from the IV estimates would remain positive and statistically significant for cutoff distances of 100 km and beyond, indicating that the extent of positive spillovers spreads farther when focusing on the actual transport infrastructure rather than just geography.

Finally, the last two columns of Table 1 include spatial lags with weights based on ethnic, geographic, *and* road connectivity. That is, they report our estimates of the whole model as described in equations (1) and (2). We observed that three coefficients of interest, i.e., $\rho_1$, $\rho_2$, and $\rho_3$, were all positive and statistically significant at the 1% level in the OLS estimates. The same holds true for the IV estimates, except that the spatial spillovers via purely geographic connectivity were only statistically significant at the 10% level. The coefficient estimates further suggest that geographic connectivity tends to be less important than ethnic and road connectivity.[11]

[11]In the Online Appendix E (Sections E1–E11), we conduct a series of robustness checks and show that the estimation of the main coefficients does not change.

Table 1: Connectivity based on ethnicity, geography and roads

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| | Dependent variable: $Light_{it}$ | | | | | | | |
| $Ethnicity\ W\ Light_{jt}$ | 0.552*** | 0.271 | | | | | 0.160*** | 0.342*** |
| | (0.015) | (0.176) | | | | | (0.013) | (0.122) |
| $Inv\ Dist\ W\ Light_{jt}$ | | | 0.550*** | 0.639*** | | | 0.246*** | 0.305** |
| | | | (0.012) | (0.131) | | | (0.011) | (0.124) |
| $Inv\ Road\ W\ Light_{jt}$ | | | | | 0.556*** | 0.280** | 0.393*** | 0.361*** |
| | | | | | (0.010) | (0.113) | (0.015) | (0.116) |
| First stage: | Dependent variable: $Light_{jt}$ | | | | | | | |
| $MP_{jt}$ | | 0.121*** | | 0.124*** | | 0.119*** | | 0.123*** |
| | | (0.015) | | (0.013) | | (0.014) | | (0.015) |
| First-stage F-stat | | 67.83 | | 85.00 | | 71.90 | | 63.64 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* The even columns report standard fixed effects regressions with district and country-year fixed effects, and the odd columns report IV estimates. $Ethnicity\,W\,Light_{jt}$ is weighted $Light_{jt}$, with weights based on the row-normalized ethnicity matrix. $Inv\,Dist\,W\,Light_{jt}$ ($Inv\,Road\,W\,Light_{jt}$) is weighted $Light_{jt}$, with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70 km. Additional control variables include population, conflict, and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information concerning the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3)). The first stage further includes the control variables indicated in equation (2). Standard errors, clustered at the network level, are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10% levels, respectively.

# 3 Most central districts

We now turn to the main aim of our paper: determining the districts that play a key role in African economies due to their connectivity and their propagation of positive spillovers. In Online Appendix A, we formally define different well-known centrality measures in network theory, that is, betweenness centrality, eigenvector centrality, Katz-Bonacich centrality, and key-player centrality. In particular, the key-player centrality (Ballester et al. (2006); Zenou (2016)) is the district which, once removed, will reduce total nighttime lights the most. We computed these four different centrality measures for all 5,444 districts from the 53 African countries in our sample. The underlying computation used the coefficient estimates, particularly the estimated $\rho$'s, reported in column (8) of Table 1. In our discussion here, we will mainly focus on the key-player ranking based on the geographic network, the road network, *and* the ethnicity network for two large countries that feature prominently in the literature: Nigeria and Kenya.[12] Figure 1 compares the key-player centrality of the districts (top row) with the districts' average nighttime light intensity (middle row) and population density (bottom row).

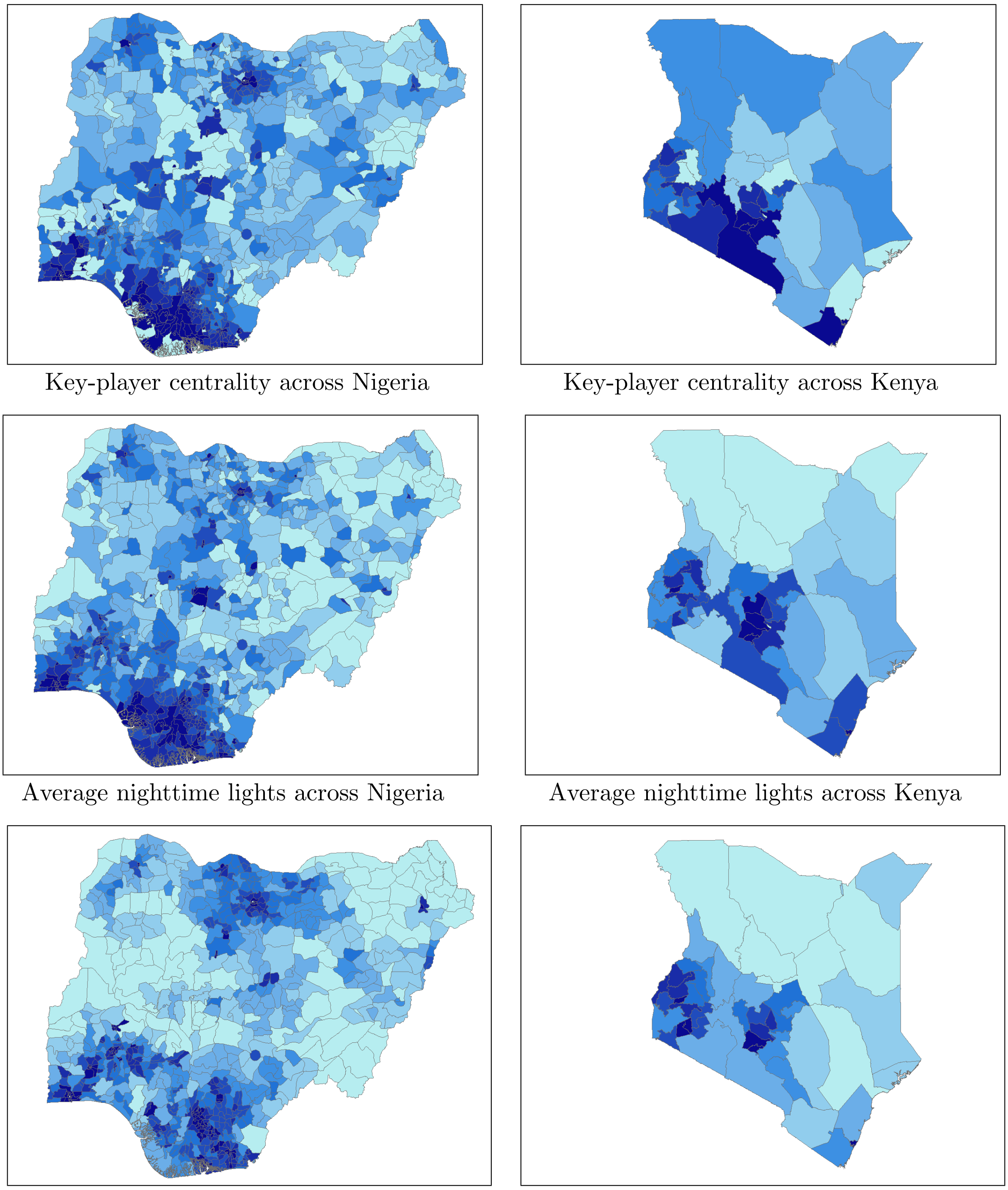
The top seven districts with the highest key-player centralities were part of the Lagos metropolitan area, which is the primate city of Nigeria and its economic hub. Seven other districts belonging to the top-ten key districts of Nigeria were also part of Lagos State. The two remaining districts in the key-player ranking belonged to the Kano metropolitan area, which is the second largest metropolitan area in Nigeria, and the economic hub of the country's north.

The key district in Kenya is Nairobi, which is the capital and the primate city. It is followed by Mombassa, which is Kenya's second largest city and home to Kenya's largest seaport (see the right column of Figure 1). The key districts encompassed or were part of the primate city in many other African countries as well, including Ethiopia (Addis Ababa) and South Africa (Johannesburg). The overall pattern suggests that *primate cities* tend to be the key districts' development in Africa, which resonates with the findings of

[12] Table F1 in the Online Appendix presents the ten most central districts (according to the key-player centrality) for Kenya and Nigeria, while Table F2 presents the same information for the five most populous African countries aside from Nigeria.

Ades and Glaeser (1995), Henderson (2002), and Storeygard (2016) among others.

Figure 1: Key-player Centrality, Nighttime Light Intensity, and Population Density in Nigeria and Kenya



Key-player centrality across Nigeria

Key-player centrality across Kenya

Average nighttime lights across Nigeria

Average nighttime lights across Kenya

Population density across Nigeria

Population density across Kenya

*Notes:* Darker colors indicate higher values.

# 4 Policy experiments

The key-player rankings are valuable in showing which districts are most economically important. However, relying on key-player rankings for policymaking has two disadvantages. First, the key districts are typically economically active and well-connected while policymakers may be interested in the benefits from either promoting local economic activity or improving the network structure, e.g., by building roads. Second, key-player rankings capture the total effect of having a particular district while policymakers are generally better advised to focus on the "marginal" effects of increasing local economic activity or improving the network structure. In this section, therefore we illustrate how our approach allows for counterfactual exercises that can inform policymakers.

The first policy experiment consists of increasing economic activity, i.e., nighttime lights, in each district, one at a time. This experiment may mimic large public investments within the given districts. We proceeded as follows. First, we added the value of 10 to the average nighttime light pixel value in the treatment district, which corresponded to an increase of one standard deviation.[13] Second, we took the logarithm of the now higher average nighttime light pixel value and recalculated the spatially lagged dependent variables with the new values, while keeping the estimated $\rho$'s from Table 1 (column (8)). Third, we recalculated the predicted nighttime lights (in logs) for each African district and computed the sum across all districts. Fourth, we compared this sum, which included the increase in nighttime lights in one district and the subsequent spatial spillovers, to the sum of the district-level nighttime lights (in logs) across Africa from the baseline, i.e., in the absence of any policy intervention. We repeated this exercise for each of the 5,944 districts.[14]

The maps in Panel (a) of Figure 2 show the districts in Nigeria (left panel) and Kenya (right panel), where this counterfactual increase in economic activity would create a stronger (darker colors) and lesser (lighter colors) overall impact.[15]

[13]The nighttime light pixel values were top-coded at 63, and 0.01% of the districts in our sample had average nighttime light pixel values above 53. Nevertheless, we increased the average nighttime light pixel value of these districts by 10 as well.

[14]In all steps and for all districts, we averaged the variables used over the sample period.

[15]Table G1 in the Online Appendix lists the Top 10 districts with the largest overall impact from this

There were various types of districts where the overall impact was particularly high. First, for both Nigeria and Kenya, the districts with the highest impact also had high key-player centrality because they are economically active and well-connected, such as the districts from Lagos State. Second, in Nigeria, the top districts includes some districts in Bayelsa and Delta, which are both oil-producing states in the Niger Delta. These districts are economically quite active and well-connected, but they have a low key-player centrality because they are conflict-ridden. An increase in economic activity, however, has a positive impact exactly because of the dense network in the Niger Delta. Third, in Kenya, the districts with the biggest impact included three poor districts that ranked at the bottom in terms of key-player centrality because of their low nighttime light values. In these districts, an increase in absolute nighttime lights leads to a large overall impact, mainly because we measure economic benefits using the logarithm of nighttime lights. Our use of logged values implies that an increase in economic activity is more valuable in poorer districts.

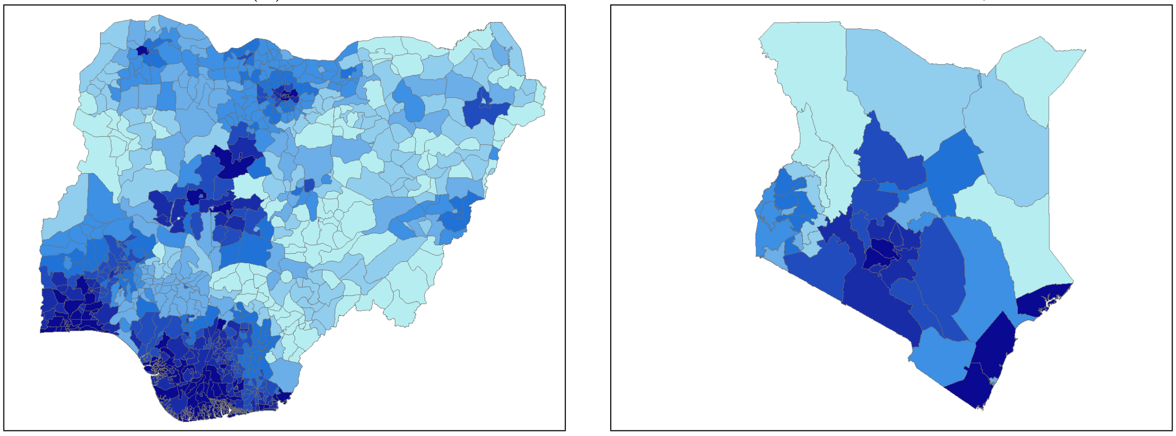
The second policy experiment consisted of increasing the road connectivity of each district, again one at a time. This experiment mimics improvements in the road infrastructure. We proceeded as follows. First, for any given district, we determined the set of contiguous districts with which the given district was not yet linked via a major road, and we chose the district with the highest average nighttime light value from this set of districts. Second, we added a link between these two districts (with a value of 1) in the non-normalized road connectivity matrix. Third, we re-normalized the road connectivity matrix and then recalculated the spatially lagged dependent and independent variables using this new matrix. Fourth, we recalculated the predicted nighttime lights (in logs) for each African district and computed the sum across all African districts. Finally, we, again, compared this sum with the sum of nighttime lights (in logs) across Africa from the baseline. We repeated this exercise for each of the 5,944 districts to identify the districts for each country that have the largest overall impact when improving their road connectivity.

counterfactual increase while Table H1 presents the same ranking for the five other most populous African countries.

Panel (b) in Figure 2 maps the results of this policy experiment for districts in Nigeria (left panel) and Kenya (right panel). Again, districts where a counterfactual improvement of road connectivity would create a stronger overall increase in nighttime light are presented by darker colors.[16]

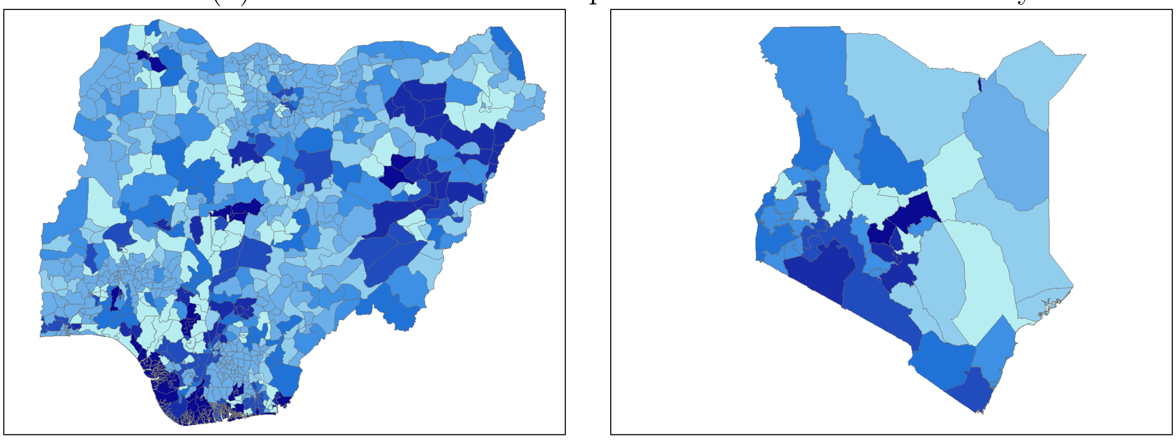
Figure 2: Results of the Counterfactual Policy Experiments - Nigeria and Kenya

Panel (a): The Counterfactual Increase in Economic Activity



Nigeria Kenya

Panel (b): The Counterfactual Improvement in Road Connectivity



Nigeria Kenya

*Notes:* Darker colors indicate a higher overall impact.

The top districts in Nigeria are all in the Niger Delta. The top two, Boony and Orika, are both islands with intense nighttime lights but poor road connectivity. Improving their road connectivity would lead to positive economic spillovers from these two districts into

[16]Table G2 in the Online Appendix lists the top 10 districts with the largest overall impact from this counterfactual increase while Table H2 presents the same ranking for the five other most populous African countries.

other districts in the Niger Delta and beyond. The districts with the strongest overall impact in Kenya included many districts with high key-player centrality. In addition, the list included some very dark/poor districts (Machakos, Wajir, and Meru), where an increase in economic activity from better road connectivity would be particularly valuable. Along similar lines, better road connectivity would also be of value in many dark/poor districts in Northeastern Nigeria.

# 5 Concluding remarks

In this paper, we studied the role of geographical, ethnic, and road networks for the spatial diffusion of local economic shocks using a panel dataset of 5,944 districts from 53 African countries over the period 1997–2013. Our main aim was to calculate the key-player centralities by performing counterfactual exercises, which consisted of removing a district and all its direct "links" (in the adjacency matrices representing the geographical, ethnic, and road networks) and computing the economic loss to an average African district. We found that primate cities are key for a country's economic development due to their high economic activity and good connectivity, suggesting that policies focusing on the major cities are justified from a growth perspective. We further conducted two counterfactual policy exercises; the first increased economic activity in each district, one at a time, and the second each district's road connectivity.

These counterfactual exercises illustrate the potential of our approach for informing policymakers in Africa as well as international donors and development agencies. A planner who decides where to locate a particular developmental project or where to build a new or better road may need to consider many aspects, but one of them should be the potential of this project to generate spatial economic spillovers. Therefore, we conducted counterfactual exercises to show how the estimated coefficients and the underlying network structure can inform us about the aggregate economic effects of policies that increase economic activity in particular districts or improve road connectivity between districts.

## References


Acemoglu, D., Garcia-Gimeno, C. and J.A. Robinson (2015), "State capacity and economic development: a network approach," *American Economic Review* 105, 2364–2409.

Ades, A. and E. Glaeser (1995), "Trade and circuses: Explaining urban giants," *Quarterly Journal of Economics* 90, 195–228.

Alder, S. (2015), "Chinese roads in India: the effect of transport infrastructure on economic development," unpublished manuscript, University of North Carolina at Chapel Hill.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Aragon, F.M. and J.P. Rud (2013), "Natural resources and local communities: evidence from a Peruvian gold mine," *American Economic Journal: Economic Policy* 5, 1–25.

Ballester, C., Calvó-Armengol, A. and Y. Zenou (2006), "Who's who in networks. Wanted: the key player," *Econometrica* 74, 1403–1417.

Banerjee, A., Duflo, E. and N. Qian (2012), "Transportation infrastructure and economic growth in China," *NBER Working Paper* 17897.

Beegle, K., Christiaensen, L., Dabalen, A. and I. Gaddis (2016), *Poverty in a Rising Africa*, Washington, DC: World Bank Publications.

Berman, N., Couttenier, M., Rohner, D. and M. Thoenig (2017), "This mine is mine! How minerals fuel conflicts in Africa," *American Economic Review* 107, 1564–1610.

Bonfatti, R., and S. Poelhekke (2017), "From mine to coast: transport infrastructure and the direction of trade in developing countries," *Journal of Development Economics* 127, 91–108.

Brock, W.A. and S.N. Durlauf (2001), "What have we learned from a decade of empirical research on growth? Growth empirics and reality," *World Bank Economic Review* 15, 229–272.

Brückner, M., Ciccone, A. and A. Tesei (2012), "Oil price shocks, income, and democracy," *Review of Economics and Statistics* 94, 389–399.

Brunnschweiler, C.N., and E.H. Bulte (2008), "The resource curse revisited and revised: A tale of paradoxes and red herrings," *Journal of Environmental Economics and Man-*

*agement* 55, 248–264.

Caselli, F. and G. Michaels (2013), "Do oil windfalls improve living standards? Evidence from Brazil," *American Economic Journal: Applied Economics* 5, 208-238.

Colella, F., Lalive, R., Sakalli, S. and M. Thoening (2018), "Inference with arbitrary clustering," unpublished manuscript, Department of Economics, HEC University of Lausanne.

Conley, T.G. (1999), "GMM estimation with cross sectional dependence," *Journal of Econometrics* 92, 1-45.

Crespo Cuaresma, J. (2011), "How different is Africa? A comment on Masanjala and Papageorgiou," *Journal of Applied Econometrics* 26, 1041–1047.

Donaldson, D. (2018), "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review* 108, 899–934.

Donaldson, D. and R. Hornbeck (2016), "Railroads and American economic growth: A market access approach," *Quarterly Journal of Economics* 131, 799–858.

Faber, B. (2014), "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies* 81, 1046–1070.

Henderson, V.J. (2002), "Urban primacy, external costs, and quality of life," *Resource and Energy Economics* 24, 95–106.

Henderson, V.J., Storeygard A. and D.N. Weil (2012), "Measuring economic growth from outer space," *American Economic Review* 102, 994–1028.

Hodler, R. and P.A. Raschky (2014), "Regional favoritism," *Quarterly Journal of Economics* 129, 995–1033.

Jackson, M.O. (2008), *Social and Economic Networks*, Princeton, NJ: Princeton University Press.

Jackson, M.O., Rogers, B.W. and Y. Zenou (2017), "The economic consequences of social network structure," *Journal of Economic Literature* 55, 49–95.

Jedwab, R., and A. Storeygard (2018), "The heterogeneous effects of transportation investments: Evidence from sub-Saharan Africa 1960–2010," Unpublished manuscript.

König, M.D., Rohner, D., Thoenig, M. and F. Zilibotti (2017), "Networks in conflict:

theory and evidence from the Great War of Africa," *Econometrica* 85, 1093–1132.

Lindquist, M.J. and Y. Zenou (2014), "Key players in co-offending networks," CEPR Discussion Paper No. 9889.

Liu, X., Patacchini, E., Zenou, Y. and L-F. Lee (2020), "Who is the key player? A network analysis of juvenile delinquency," *Journal of Business and Economic Statistics*, forthcoming.

Michalopoulos, S. and E. Papaioannou (2013), "Pre-colonial ethnic institutions and contemporary African development," *Econometrica* 81, 113–152.

Michalopoulos, S. and E. Papaioannou (2014), "National institutions and subnational development in Africa," *Quarterly Journal of Economics* 129, 151–213.

Murdock, G.P. (1959), *Africa: Its Peoples and their Culture History*, New York: McGraw-Hill Book Company.

Storeygard, A. (2016), "Farther on down the road: transport costs, trade and urban growth," *Review of Economic Studies*, 83, 1263–1295.

Zenou, Y. (2016), "Key players," In: Y. Bramoullé, B.W. Rogers and A. Galeotti (Eds.), *The Oxford Handbook of the Economics of Networks*, Oxford: Oxford University Press, pp. 244–274.

# Online Appendix: Key Players in Economic Development

Ashani Amarasinghe* Roland Hodler†
Paul A. Raschky‡ Yves Zenou§

Not for Publication

*Department of Economics, Monash University; email: ashani.amarasinghe@monash.edu.
†Department of Economics, University of St.Gallen; CEPR, CESifo, and OxCarre; email: roland.hodler@unisg.ch.
‡Department of Economics, Monash University; email: paul.raschky@monash.edu.
§Department of Economics, Monash University; CEPR and IZA; email: yves.zenou@monash.edu.

# Contents

# A Theory

## A.1 A simple theoretical model

### A.1.1 The case of a single network

Consider a network linking different districts. A network (graph) $\boldsymbol{\omega}$ is the pair $(N, E)$ consisting of a set of nodes (here districts) $N = \{1, \ldots, n\}$ and a set of edges (links) $E \subset N \times N$ between them. The neighborhood of a node $i \in N$ is the set $N_i = \{j \in N : (i, j) \in E\}$. The adjacency matrix $\boldsymbol{\Omega} = (\omega_{ij})$ keeps track of direct links so that $\omega_{ij} \in ]0, 1]$ if a link exists between districts $i$ and $j$, and $\omega_{ij} = 0$ otherwise.[1] We assume that the adjacency matrix $\boldsymbol{\Omega}$ is row-normalized so that the sum of each of its rows is equal to 1, i.e., $\sum_j \omega_{ij} = 1$ for all $i$.[2] In the data, $\boldsymbol{\Omega} = (\omega_{ij})$ will capture connectivity based on geography, the road network or ethnicity.

We assume that the level of economic activity $l_i$ of a district $i$ is given by:

$$l_i = \rho \sum_{j=1}^{J} \omega_{ij} l_j + X_i + \varepsilon_i \tag{1}$$

Indeed, we assume that the economic activity of district $i$ is simply a function of the economic activity of neighboring districts, of the observable characteristics $X_i$ (such as its population) and unobservable characteristics $\varepsilon_i$ of this district. In this equation, $\rho$ captures the spillover effects of economic activities between neighboring districts.

Observe that the total level of activity $l_i$ of district $i$ is given by (1) because we would like to describe activities at the *district* level and, more importantly, the transmission of economic shocks between districts. Our theoretical framework, and the empirical analysis, implicitly acknowledge that there are a plethora of possible transmission channels (e.g., prices, wages, trade, or migration). However, the main focus of this paper is analyzing the effect of a district's position on the diffusion of economic shocks within a network. For that purpose, applying a simple, network theoretical model to a more aggregate setting is sufficient.

There are different ways one can microfound equation (1). Let us propose a simple way of doing so.

Assume that the level of prosperity $p_i$ of a district $i$ is given by:

$$p_i = X_i l_i + \rho l_i \sum_{j=1}^{J} \omega_{ij} l_j + l_i \varepsilon_i \tag{2}$$

where, as above, $l_i$ is the economic activity in district $i$, $X_i$ captures the characteristics of

[1] In spatial econometrics, the adjacency matrix is called the "connectivity matrix." Throughout the paper, we will use these terms interchangeably.

[2] All our theoretical results hold if the adjacency matrix is not row-normalized.

district $i$ and $\varepsilon_i$ is the error term. In sum, the prosperity level of a district is determined by the district's *observable* and *unobservable* characteristics, the economic activity of the district, and the spillover effects of the economic activity of neighboring districts.

Assume that the entity in charge of district $i$ (this could be an institution or a local government or a local politician) chooses the district's own economic activity level $l_i$, taking as given the choices of all the other districts. The payoff function of the entity in charge of district $i$ is then given by:

$$U_i = p_i - \frac{1}{2}l_i^2 = X_i l_i + \rho l_i \sum_{j=1}^{J} \omega_{ij} l_j + l_i \varepsilon_i - \frac{1}{2}l_i^2 \tag{3}$$

Indeed, the payoff function consists of the prosperity level of district $i$ minus the cost of maintaining this prosperity level, which is, quite naturally, increasing in economic activity.

Then, taking the first-order condition of (3) leads to (1), which can be written in matrix form as follows:

$$\mathbf{l} = (\mathbf{I} - \rho\boldsymbol{\Omega})^{-1}(\mathbf{X} + \boldsymbol{\varepsilon}) =: \mathbf{C}^{BO}_{\mathbf{X}+\boldsymbol{\varepsilon}}(\rho, \boldsymbol{\omega}) \tag{4}$$

where $\mathbf{l}$ is a column-vector of $l_i$s, $\mathbf{I}$ is the identity matrix, and $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are the vectors corresponding to the $X_i$s and $\varepsilon_i$s, respectively. In (4), $\mathbf{C}^{BO}_{\mathbf{X}+\boldsymbol{\varepsilon}}(\rho, \boldsymbol{\omega})$, whose $i$th row is $C^{BO}_{i,X_i+\varepsilon_i}(\rho, \boldsymbol{\omega})$, is the weighted *Katz-Bonacich centrality* (due to Bonacich, 1987, and Katz, 1953), where the weights are determined by the sum of $X_i$ and $\varepsilon_i$ for each district $i$. Denote by $\mu_1(\boldsymbol{\Omega})$ the spectral radius of $\boldsymbol{\Omega}$. Then, if $\rho\mu_1(\boldsymbol{\Omega}) < 1$, there exists a unique interior equilibrium given by (1) or (4). Since the adjacency matrix $\boldsymbol{\Omega}$ is assumed to be row-normalized, it holds that $\mu_1(\boldsymbol{\Omega}) = 1$. Thus, the condition for existence and uniqueness can be written as $\rho < 1$.

Consider again (1). Then, $\rho$ has an easy interpretation. In social networks, it is called the social or network multiplier. Here, it is the strength of spillovers in terms of nighttime lights between neighboring districts. To illustrate this, consider the case of a dyad (two districts, i.e., $N = 2$). For simplicity, assume that the two districts are ex ante identical so that $X_1 + \varepsilon_1 = X_2 + \varepsilon_2 = X + \varepsilon$. In that case, if there were no network (empty network) so that the two districts were not linked, then (1) will be given by:

$$l_1^{empty} = l_2^{empty} = X + \varepsilon$$

Consider now a network where the two districts are linked to each other (i.e., $\omega_{12} = \omega_{21} = 1$). Then, if $\rho < 1$, we obtain:

$$l_1^{dyad} = l_2^{dyad} = \frac{X + \varepsilon}{1 - \rho}$$

In other words, because of complementarities, in the dyad, the level of activity of each district is much higher than when the districts are not connected. The factor $1/(1-\rho) > 1$ is the *network multiplier*.[3]

Observe that the way we modeled spillover effects (see (1)) is similar to the way urban economists have been modeling agglomeration effects. For example, in Ahlfeldt et al. (2015), agglomeration effects are modeled as production externalities. In our case, spillover effects might capture those effects but could also be driven by other effects as well.[4]

### A.1.2 The case of multiple networks

In the real world, there is more than one type of spillovers between districts. For example, in our main specifications below, we use different adjacency matrices $\mathbf{\Omega} = (\omega_{ij})$ that keep track of the (inverse) spatial distance between districts, the road network and the proximity in terms of ethnicity. In that case, (1) would be written as:

$$l_i = \rho_1 \sum_{j=1}^{J} \omega_{1,ij} l_j + \rho_2 \sum_{j=1}^{J} \omega_{2,ij} l_j + \rho_3 \sum_{j=1}^{J} \omega_{3,ij} l_j + X_i + \varepsilon_i \tag{5}$$

where $\rho_1 > 0$, $\rho_2 > 0$ and $\rho_3 > 0$. We now have three adjacency matrices $\mathbf{\Omega}_1 = (\omega_{1,ij})$, $\mathbf{\Omega}_2 = (\omega_{2,ij})$ and $\mathbf{\Omega}_3 = (\omega_{3,ij})$, which are all assumed to be row-normalized.

This equation says that the spillover effects in terms of economic activities between districts are affected differently by the ways we measure the "proximity" between neighboring districts.[5]

## A.2 Theory: Different definitions of node centralities

There are different centrality measures (see Jackson, 2008, for an overview). We first introduce two non micro-founded, purely topological centrality measures and then two micro-founded measures that are strongly linked to our simple model.

[3] Observe that if we keep the ex ante heterogeneity, if $\rho < 1$, we obtain:

$$\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \frac{1}{(1-\rho^2)} \begin{pmatrix} X_1 + \varepsilon_1 + \rho (X_2 + \varepsilon_2) \\ X_2 + \varepsilon_2 + \rho (X_1 + \varepsilon_1) \end{pmatrix}$$

[4] See the overviews by Duranton and Puga (2004) and Fujita and Thisse (2013) who provide different micro-foundations of spillover effects in the context of urban agglomeration.

[5] As above, we can provide a microfoundation of this equation by assuming that the prosperity level of a district is given by:

$$p_i = X_i l_i + \rho_1 l_i \sum_{j=1}^{J} \omega_{1,ij} l_j + \rho_2 l_i \sum_{j=1}^{J} \omega_{2,ij} l_j + \rho_3 l_i \sum_{j=1}^{J} \omega_{3,ij} l_j + l_i \varepsilon_i$$

Then, if $\mu_1 (\rho_1 \mathbf{\Omega}_1 + \rho_2 \mathbf{\Omega}_2 + \rho_3 \mathbf{\Omega}_3) < 1$, there exists a unique interior equilibrium given by (5).

### A.2.1 Non micro-founded centrality measures

The two most commonly used individual-level measures of network centrality are betweenness centrality and eigenvector centrality.

The *betweenness centrality*, $C_i^{BE}(\boldsymbol{\omega})$, describes how well located an individual district in the network in terms of the number of shortest paths between other districts that run through it. Denote the number of shortest paths between districts $j$ and $k$ that district $i$ lies on as $P_i(jk)$, and let $P(jk)$ denote the total number of shortest paths between districts $j$ and $k$. The ratio $P_i(jk)/P(jk)$ tells us how important district $i$ is for connecting districts $j$ and $k$ to each other. Averaging across all possible $jk$ pairs gives us the betweenness centrality measure of district $i$:

$$C_i^{BE}(\boldsymbol{\omega}) = \sum_{j \neq k: i \notin \{j,k\}} \frac{P_i(jk)/P(jk)}{(n-1)(n-2)/2}$$

It has values in $[0, 1]$.

The *eigenvector centrality*, $C_i^{E}(\boldsymbol{\omega})$, is defined using the following recursive formula:

$$C_i^{E}(\boldsymbol{\omega}) = \frac{1}{\mu_1(\boldsymbol{\Omega})} \sum_{j=1}^{n} g_{ij} C_j^{E}(\boldsymbol{\omega}) \tag{6}$$

where $\mu_1(\boldsymbol{\Omega})$ is the largest eigenvalue of $\boldsymbol{\Omega}$. According to the Perron-Frobenius theorem, using the largest eigenvalue guarantees that $C_i^{E}(\boldsymbol{\omega})$ is always positive. In matrix form, we have:

$$\mu_1(\boldsymbol{\Omega})\, \mathbf{C}^{E}(\boldsymbol{\omega}) = \boldsymbol{\Omega}\mathbf{C}^{\mathbf{E}}(\boldsymbol{\omega}) \tag{7}$$

The eigenvector centrality of a district assigns relative scores to all districts in the network based on the concept that connections to high-scoring districts contribute more to the score of the district in question than equal connections to low-scoring agents.

### A.2.2 Katz-Bonacich centrality

In our theoretical model (Section A), we have shown that the unique Nash equilibrium of our game in terms of nighttime lights is equal to the *Katz-Bonacich centrality* of the district. As a result, the level of nighttime lights in district $i$ is given by its weighted Katz-Bonacich centrality, defined in (4), i.e.

$$\mathbf{C}^{BO}_{\mathbf{X}+\boldsymbol{\varepsilon}}(\rho, \boldsymbol{\omega}) =: (\mathbf{I} - \rho\boldsymbol{\Omega})^{-1} (\mathbf{X} + \boldsymbol{\varepsilon})$$

Importantly, in order to calculate the Katz-Bonacich centrality of each district $i$, we need to know the value of $\rho$. We will use the estimated value of $\rho$ (IV estimates). We also need to check that the condition $\rho\mu_1(\boldsymbol{\Omega}) < 1$ is satisfied.

### A.2.3 Key-player centrality

The Katz-Bonacich centrality was based on the outcome of a Nash equilibrium. Let us now focus on the planner's problem. The key question is as follows: Which district, once removed, will reduce total nighttime lights the most? In other words, which district is the key player? Ballester et al. (2006) have proposed a measure, *key-player centrality*, that answers this question.[6] For that, consider the game with strategic complements developed in the theory section (Section A) for which the utility is given by (3), and denote $L^*(\boldsymbol{\omega}) = \sum_{i=1}^{n} l_i^*$ the total equilibrium level of activity in network $\boldsymbol{\omega}$, where, assuming $\phi\mu_1(\boldsymbol{\omega}) < 1$, $l_i^*$ is the Nash equilibrium effort given by (1) or (4). Also, denote by $\boldsymbol{\omega}^{[-i]}$ the network $\boldsymbol{\omega}$ without district $i$. Then, in order to determine the *key player*, the planner will solve the following problem:

$$\max\{L^*(\boldsymbol{\omega}) - L^*(\boldsymbol{\omega}^{[-i]}) \mid i = 1, ..., n\} \tag{8}$$

Then, the *intercentrality* or the *key-player centrality* $C_i^{KP}(\rho, \boldsymbol{\omega})$ of district $i$ is defined as follows:

$$C_{i,u_i}^{KP}(\rho, \boldsymbol{\omega}) = \frac{C_{i,u_i}^{BO}(\rho, \boldsymbol{\omega}) \sum_j m_{ji}(\rho, \boldsymbol{\omega})}{m_{ii}(\rho, \boldsymbol{\omega})} \tag{9}$$

where $C_{i,u_i}^{BO}(\rho, \boldsymbol{\omega})$ is the weighted Katz-Bonacich centrality of district $i$ (see equation (4)) and $m_{ij}(\rho, \boldsymbol{\omega})$ is the $(i, j)$ cell of the matrix $\mathbf{M}(\rho, \boldsymbol{\omega})= (\mathbf{I} - \rho\boldsymbol{\Omega})^{-1}$. Ballester et al. (2006, 2010) have shown that the district $i^*$ that solves (8) is the key player if and only if $i^*$ is the district with the highest *intercentrality* in $\boldsymbol{\omega}$, that is, $C_{i^*,u_i}^{KP}(\rho, \boldsymbol{\omega}) \geq C_{i,u_i}^{KP}(\rho, \boldsymbol{\omega})$, for all $i = 1, ..., n$. The intercentrality measure (9) of district $i$ is the sum of $i$'s centrality measures in $\boldsymbol{\omega}$, and its contribution to the centrality measure of every other district $j \neq i$ also in $\boldsymbol{\omega}$. It accounts both for one's exposure to the rest of the group and for one's contribution to every other exposure. This means that the key player $i^*$ in network $\boldsymbol{\omega}$ is given by $i^* = \arg\max_i C_{i,u_i}^{KP}(\rho, \boldsymbol{\omega})$, where[7]

$$C_{i^*,u_i}^{KP}(\rho, \boldsymbol{\omega}) = L^*(\boldsymbol{\omega}) - L^*\left(\boldsymbol{\omega}^{[-i]}\right). \tag{10}$$

## References


Ahfeldt, G.M., Redding, S.J., Sturm, D.M. and N. Wolf (2015), "The economics of density: Evidence from the Berlin wall," *Econometrica* 83, 2127–2189.

Bonacich P. (1987), "Power and centrality: a family of measures," *American Journal of Sociology* 92, 1170–1182.


[6]For an overview of the way the key player is determined in different areas, see Zenou (2016).

[7]Ballester et al. (2006) define the key player in (9) only when the adjacency matrix $\boldsymbol{\Omega}$ is not row-normalized. Since we use row-normalized adjacency matrices when estimating the $\rho$s, we will determine the key player numerically based on its definition in (10).

Duranton, G. and D. Puga (2004), "Micro-foundations of urban agglomeration economies," In: J.V. Henderson and J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier, pp. 2063–2117.

Fujita, M. and J.-F. Thisse (2013), *Economics of Agglomeration. Cities, Industrial Location and Globalization*, Second edition, Cambridge: Cambridge University Press.

Jackson, M.O. (2008), *Social and economic networks*, Princeton, NJ: Princeton University Press.

Katz, L. (1953), "A new status index derived from sociometric analysis," *Psychometrika* 18, 39–43.

# B Additional Data Description

## B.1 Description of our variables

### B.1.1 Dependent variable: Nighttime lights

Satellite data on the intensity of nighttime lights comes from the the National Oceanic and Atmospheric Administration (NOAA). Weather satellites from the US Air Force circle the earth 14 times per day and measure light intensity. The NOAA uses evening observations during the dark half of the lunar cycle in seasons when the sun sets early, but removes observations affected by cloud coverage, or northern or southern lights. It further processes the data by setting readings that are likely to reflect fires, other ephemeral lights or background noise to zero.[8] The objective is that the reported nighttime lights are primarily man-made. The NOAA then provides annual data for the time period from 1992 onwards for output pixels that correspond to less than one square kilometer. The data come on a scale from 0 to 63, with higher values implying more intense nighttime lights.

Nighttime lights are a proxy for economic activity, as most forms of consumption and production in the evening require light. Moreover, public infrastructure is often lit at night. It is, therefore, not surprising that Henderson et al. (2012) and Hodler and Raschky (2014) find a high correlation between changes in nighttime light intensity and GDP at the level of countries and subnational administrative regions, respectively. Using data from Gennaioli et al. (2014), we also find a high correlation between nighttime lights and subnational GDP for 82 subnational administrative regions from nine African countries (see Table C1).[9]

To construct our dependent variable, $Light_{it}$, we take the logarithm of the average nighttime light pixel value in district $i$ and year $t$. To avoid losing observations with a reported nighttime light intensity of zero, we follow Michalopoulos and Papaioannou (2013, 2014) and Hodler and Raschky (2014) in adding 0.01 before taking the logarithm.

### B.1.2 Connectivity matrices

We construct three connectivity matrices to measure spatial spillovers.

**Ethnic connectivity**

[8]Readings due to fires and other ephemeral lights are identified by their high brightness and infrequent occurrence. Background noise is identified by setting light intensity thresholds based on areas expected to be free of detectable lights (Baugh et al., 2010).

[9]Furthermore, Bruederle and Hodler (2018) show that nighttime lights are correlated with wealth indicators based on the Development and Health Surveys (DHS) in a large cross-section of African localities. Survey-based wealth indicators are, however, not suitable for our empirical analysis, which exploits within-district variation. Reasons are threefold: First, DHS and other surveys take only place every few years. Second, some districts are not even surveyed in each survey round. Third, DHS and other surveys are typically not representative at the district level, such that the composition of the respondents may change within districts that are surveyed in every round.

Africa is known for its ethnic diversity. Members of the same ethnic group share similar cultural traits and behavioral norms, which may influence their ability to cooperate and their willingness to maintain economic relations. The work by Murdock (1958) documents the spatial distribution of ethnic homelands in Africa and subdivides the continent into over 800 ethnic homelands.[10]

To measure ethnic connectivity between districts, we first overlay the district (ADM2) boundaries with the boundaries of the ethnic homelands from Murdock. Each district is assigned the ethnicity of the ethnic homeland in which it is located. For districts that fall into more than one ethnic homeland, we assign the ethnicity of the ethnic homeland that covers the largest part of the district. We then construct our ethnic connectivity matrix, $\omega_{i,j}$, where elements are 1 if the ethnicity in district $i$ is the same as the ethnicity in district $j$, and 0 otherwise.

**Geographic connectivity**

We base the weighting matrix for geographic connectivity on geographic distance. We construct this weighting matrix as follows: First, we calculate the centroid of each district. Second, we calculate the geodesic distance $d_{i,j}$ connecting the centroids of districts $i$ and $j$. Third, following Acemoglu et al. (2015), we measure the variability of altitude, $e_{i,j}$, along the geodesic connecting the centroids of districts $i$ and $j$. We use elevation data from GTOPO30. Finally, we calculate the inverse of the altitude-adjusted geodesic distance as $\tilde{d}_{i,j} = 1/d_{i,j}(1 + e_{i,j})$.

Defining geographic connectivity using the inverse altitude-adjusted distance as opposed to contiguity proves advantageous on three accounts.[11] First, by incorporating all districts within a given radius, connectivity is extended to districts beyond those merely sharing a common border or a point. Second, by incorporating variability in altitude, $e_{i,j}$, we account for the topology of the landscape. Districts separated by a mountainous terrain, for example, receive a lower connectivity weight, as opposed to districts connected via a flat surface. Third, measuring geographic connectivity based on geodesic distance allows truncation at different distances, enabling the determination of the extent of spillovers. Leveraging this advantage, we construct different weighting matrices by varying the distance considered in defining a district's neighbors. The main specification will use a cutoff of 70km (for reasons made explicit below). In this case, we set the spatial weight as $\omega_{i,j} = 1/\tilde{d}_{i,j}$ if the geodesic distance $d_{i,j}$ is less than 70km, and $\omega_{i,j}$=0 otherwise.

**Road connectivity**

Roads are, arguably, a key form of connectivity between districts. Roads enable non-

[10] Figure B2 in the Online Appendix shows the digitized version of Murdock's original map.

[11] As an alternative we construct a weighting matrix for geographic connectivity based on contiguity. The contiguity matrix indicates whether districts $i$ and $j$ share a common border or, at least, a common point along their borders. We report estimates based on the contiguity matrix in Table E7 in the Online Appendix. Further, we report estimates based on geodesic distances but without adjustment for the variability in altitude in Table E8 in the Online Appendix.

contiguous districts to connect with one another and allow connectivity to extend to greater distance. Moreover, while the inverse distance matrix assumes that all districts within a given (altitude-adjusted) distance are by default connected, the road network presents an actual mechanism of connectivity, which can lead to a more realistic quantification of spillovers.

To construct connectivity via the road network we obtained data from OpenStreetMap (OSM).[12] We accessed the OSM data in early 2016 and extracted information about major roads (e.g., highways and motorways) for the African continent.[13] We intersect these roads with the district boundary polygons and generate a network graph of the road network.[14] In a first step, the road polylines are split into segments whenever they intersect with a district boundary. For each segment (edge), we then calculate the road travel distance in km between each intersection (node).[15] In the second step, we identify the shortest path on the road segments between each district and calculate the distance on that path. If districts A and B are adjacent and connected via a major road, we assign a distance value of 1km. If districts A and B are not adjacent, but connected via the road network, they are assigned the road distance between the closest road and district boundary node of A and the closest road and district boundary node of B (i.e., the road travel distance through all the district that one has to cross to get from district A to district B).

The road connectivity matrix assigns a value equal to the inverse of the road distance in km between districts $i$ and $j$ if they are connected via a major road, and 0 if they are not connected. We again construct different weighting matrices by truncating at different distance cutoffs.

Table B1 in the Online Appendix shows the correlation structure of the three connectivity matrices. Note that these connectivity matrices capture spillover effects within and between countries, because in our main analysis we are interested in capturing the overall economic spillovers between African districts. However, in a robustness analysis[16] we construct new connectivity matrices that take into account national borders.

[12] OSM is an open-source mapping project where information about roads (and other objects) is crowd-sourced by over two million volunteers worldwide, who can collect data using manual surveys, handheld GPS devices, aerial photography, and other commercial and government sources. (See https://openstreetmap.org for more information and https://geofabrik.de for the shapefiles.) We opted for the OSM instead of the World Bank's African Infrastructure Country Diagnostic (AICD) database because the AICD data does not contain information for countries with Mediterranean coastline as well as Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia.

[13] Figure B3 shows the road network.

[14] The road connectivity analysis between ADM2 polygons was conducted in ArcMap 10.2 using arcpy. The python scripts are available upon request.

[15] If the road starts/ends in a district, we calculate the distance between the start/end point and the intersection.

[16] See Section E.

### B.1.3 Mining data and instrumental variables

Our identification strategy makes use of cross-sectional information on the location of mining projects and temporal variation in the world prices of the corresponding minerals. We describe the construction of the respective variables in turn.

Our information on mining activity comes from the SNL Minings & Metals database. This database covers 3,487 mining projects across Africa that were active during our sample period. For each project, it contains information about the point location, i.e., the geographic coordinates, and the (potentially multiple) resources extracted at this location.[17]

We use the point locations to assign the mining projects to districts and identify all districts where a mine was active for at least one year during our sample period. Across Africa, 4% of all districts are mining districts. The indicator variable $Mine_i^r$ is equal to one if district $i$ has a mining project that extracts resource $r$ and is active for at least one year during our sample period. Following Berman et al. (2017), the underlying idea is that this time-invariant variable should capture a district's suitability for mining, in particular its geology, rather than endogenous decisions on production or the opening and closing of mines.[18]

Data on world prices of minerals are sourced from the World Bank, IMF, USGS and SNL (see Table B4 in the Online Appendix for more information on the data sources). $Price_t^r$ is the logarithm of the yearly nominal average price of resource $r$ in USD.

### B.1.4 Control variables

Our main time-varying control variable at the district level is $Population_i$. It measures a district's total population (in logs) and is derived based on the population data from the Center for International Earth Science Information Network (CIESIN).

In most specifications, we further control for conflicts using data extracted from the PRIO/Uppsala Armed Conflict Location and Event Database (ACLED). This is a geo-referenced database on dyadic conflict from 1997 to 2015. It includes nine different types of conflict-related events, including battles and violence against civilians as well as some non-violent events. We use the indicator variable $Conflict_{it}$, which takes a value of one if any conflict-related event occurred in district $i$ in year $t$, and zero otherwise.

Table B2 in the Online Appendix provides the descriptive statistics for our key variables.

[17] Figure B4 shows the spatial distribution of mining projects across Africa.

[18] Berman et al. (2017) restrict their sample to grid cells where a mine operates in all years or no year. This methodology significantly reduces the number of mining districts in our case and thus weakens the relevance of the instrumental variable.

## B.2 Correlation structure of the three connectivity matrices

The following table provides the correlations between our three connectivity matrices

Table B1: Correlation Between Connectivity Matrices

| | (1)<br>$Eth\ W\ Light_{jt}$ | (2)<br>$Inv\ Dist\ W\ Light_{jt}$ | (3)<br>$Inv\ Road\ W\ Light_{jt}$ |
|---|---|---|---|
| $Eth\ W\ Light_{jt}$ | 1.000<br>(1.000) | | |
| $Inv\ Dist\ W\ Light_{jt}$ | 0.341<br>(0.379) | 1.000<br>(1.000) | |
| $Inv\ Road\ W\ Light_{jt}$ | 0.382<br>(0.412) | 0.410<br>(0.429) | 1.000<br>(1.000) |

*Notes:* Correlation between demeaned variables (demeaned with respect to country-year fixed effects) presented in parenthesis.

## B.3 Descriptive statistics of our main variables

Table B2: Descriptive statistics

| Variable | Observations | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| $Light_{it}$ | 101,048 | -1.257 | 2.703 | -4.605 | 4.143 |
| $Conflict_{it}$ | 101,048 | 0.126 | 0.332 | 0 | 1 |
| $Population_{it}$ | 101,048 | 10.733 | 2.129 | -4.605 | 16.204 |
| $MP_{it}$ | 101,048 | 0.218 | 1.233 | -2.161 | 11.133 |

*Notes:* See Section B1 for the definitions of all variables. Note that $Light_{it}$, $Population_{it}$ and $MP_{it}$ are in logs.

## B.4 Subnational Districts and Countries

Our analysis focuses on 5944 subnational districts, i.e., ADM2 regions, in 53 countries across the African continent. The shapefile containing the ADM boundary polygons comes from the GADM database of Global Administrative Areas, version 1, available at http://gadm.org. Boundary polygons at the ADM2 level are available for all African countries, except Egypt and Libya, for which they are only available at the ADM1 level. The list of countries and the number of subnational regions belonging to each country appear in Table B3. The geographic dispersion of subnational regions is graphically represented in Figure B1.

Figure B1: Districts in Africa

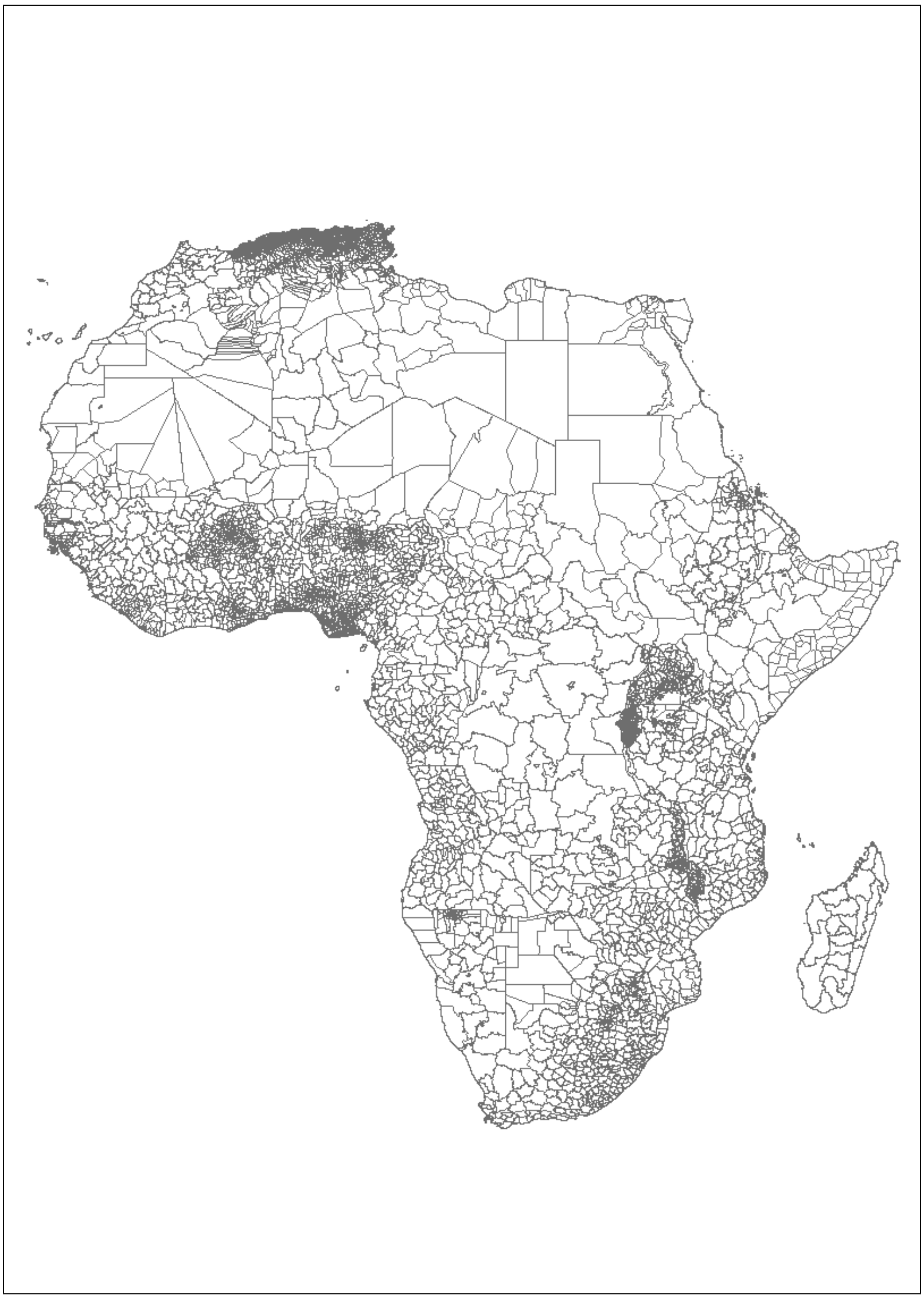


Table B3: List of Countries

| | Country | No. of districts |
|---|---|---|
| 1 | Algeria | 1,504 |
| 2 | Angola | 163 |
| 3 | Benin | 76 |
| 4 | Botswana | 25 |
| 5 | Burkina Faso | 301 |
| 6 | Burundi | 133 |
| 7 | Cameroon | 58 |
| 8 | Cape Verde | 16 |
| 9 | Central African Republic | 51 |
| 10 | Chad | 53 |
| 11 | Comoros | 3 |
| 12 | Ivory Coast | 50 |
| 13 | Democratic Republic of the Congo | 38 |
| 14 | Djibouti | 11 |
| 15 | Egypt | 26 |
| 16 | Equatorial Guinea | 6 |
| 17 | Eritrea | 50 |
| 18 | Ethiopia | 72 |
| 19 | Gabon | 37 |
| 20 | Gambia | 13 |
| 21 | Ghana | 137 |
| 22 | Guinea | 34 |
| 23 | Guinea-Bissau | 37 |
| 24 | Kenya | 48 |
| 25 | Lesotho | 10 |
| 26 | Liberia | 66 |
| 27 | Libya | 32 |
| 28 | Madagascar | 22 |
| 29 | Malawi | 253 |
| 30 | Mali | 51 |
| 31 | Mauritania | 44 |
| 32 | Mauritius | 10 |
| 33 | Morocco | 54 |
| 34 | Mozambique | 128 |
| 35 | Namibia | 107 |
| 36 | Niger | 36 |
| 37 | Nigeria | 775 |
| 38 | Republic of Congo | 46 |
| 39 | Rwanda | 142 |
| 40 | Sao Tome and Principe | 2 |
| 41 | Senegal | 30 |
| 42 | Sierra Leone | 14 |
| 43 | Somalia | 74 |
| 44 | South Africa | 354 |
| 45 | Sudan | 26 |
| 46 | Swaziland | 4 |
| 47 | Tanzania | 136 |
| 48 | Togo | 21 |
| 49 | Tunisia | 267 |
| 50 | Uganda | 162 |
| 51 | Western Sahara | 4 |
| 52 | Zambia | 72 |
| 53 | Zimbabwe | 60 |

## B.5 Ethnic Homelands

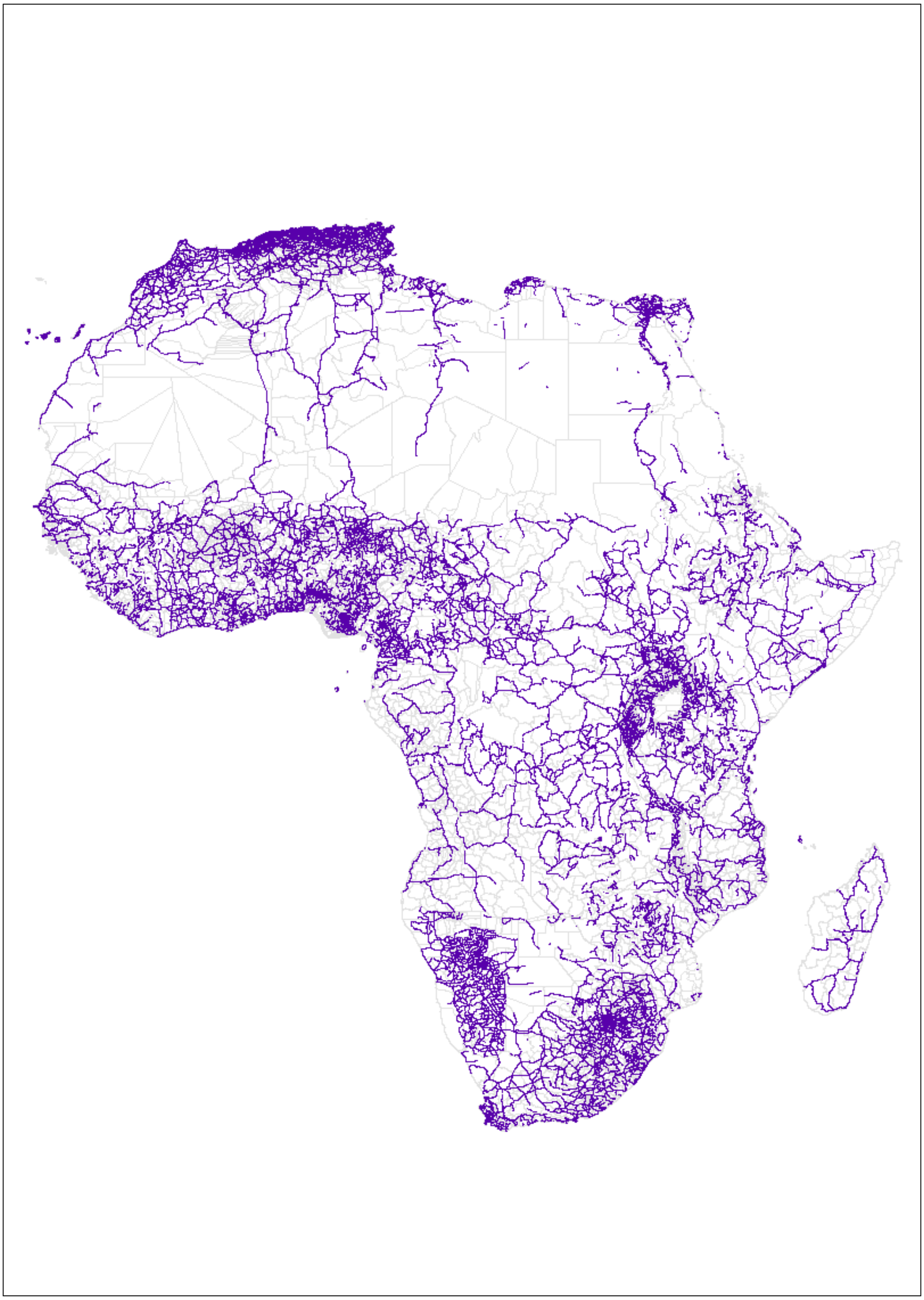
The ethnic connectivity matrix is based on the digitized map version of the Murdock (1958) map of the boundaries of ethnic homelands in Africa shown in Figure B2. Using the spatial overlay tool in ArcMap 10.2, we combined the ADM2 polylines from Figure B1 with the ethnic homeland polylines from Figure B2 to assign each district the ethnicity of the ethnic homeland that covers the largest area of this district.

Figure B2: Ethnic Homelands in Africa

## B.6 Road Network

Figure B3 shows the network of primary and secondary roads (in purple) from the OpenStreetMap data, with the district boundaries in the background (in light-gray).

Figure B3: Primary and Secondary Roads in Africa



## B.7 Mines & Minerals

Figure B4 shows the location of mines from the SNL Minings & Metals database, with the district boundaries in the background. Table B4 lists the different types of minerals covered in this database as well as the source for the information on the world market price of these minerals.

Figure B4: Distribution of Mines in Africa

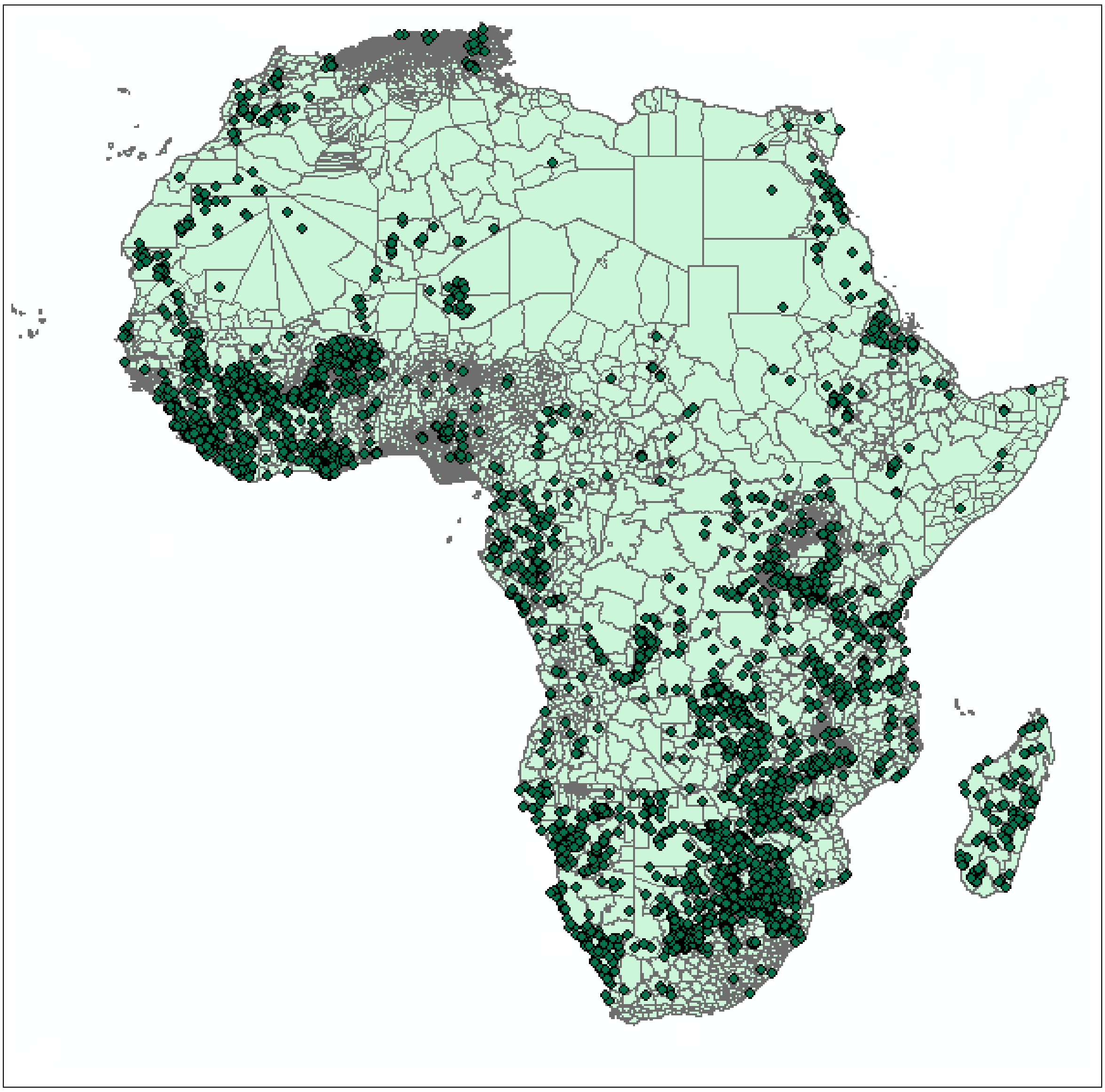


Table B4: List of Minerals

| | Name | Measure | Source |
|---|---|---|---|
| 1 | Antimony | Tonnes | USGS Commodity Prices |
| 2 | Bauxite | Tonnes | USGS Commodity Prices |
| 3 | Chromite | Tonnes | USGS Commodity Prices |
| 4 | Coal | Tonnes | World Bank |
| 5 | Cobalt | Tonnes | USGS Commodity Prices |
| 6 | Copper | Tonnes | SNL-Thomas Reuters |
| 7 | Diamond | Carats | USGS Commodity Prices |
| 8 | Gold | Ounces | SNL-Thomas Reuters |
| 9 | Graphite | Tonnes | USGS Commodity Prices |
| 10 | Ilmenite | Tonnes | USGS Commodity Prices |
| 11 | Iron | Tonnes | World Bank |
| 12 | Lanthanide | Tonnes | USGS Commodity Prices |
| 13 | Lead | Tonnes | World Bank |
| 14 | Lithium | Tonnes | USGS Commodity Prices |
| 15 | Managanese | Tonnes | USGS Commodity Prices |
| 16 | Nickel | Tonnes | World Bank |
| 17 | Niobium | Tonnes | USGS Commodity Prices |
| 18 | Palladium | Ounces | SNL-Thomas Reuters |
| 19 | Phosphate | Tonnes | World Bank |
| 20 | Platinum | Ounces | SNL-Thomas Reuters |
| 21 | Potash | Tonnes | World Bank |
| 22 | Rutile | Tonnes | USGS Commodity Prices |
| 23 | Silver | Ounces | World Bank |
| 24 | Tin | Tonnes | SNL-Thomas Reuters |
| 25 | Tantalum | Tonnes | USGS Commodity Prices |
| 26 | Tungsten | Tonnes | USGS Commodity Prices |
| 27 | Uranium Oxide | Pounds | International Monetary Fund |
| 28 | Vanadium | Tonnes | USGS Commodity Prices |
| 29 | Yttrium | Tonnes | SNL-Thomas Reuters |
| 30 | Zinc | Tonnes | SNL-Thomas Reuters |
| 31 | Zircon | Tonnes | USGS Commodity Prices |

# C Correlation between Subnational GDP and Subnational Nighttime Lights in Africa

Hodler and Raschky (2014, Appendix B) document a strong correlation between GDP per capita and nighttime lights in subnational administrative regions using the subnational GDP data by Gennaioli et al. (2014). In Table C1, we replicate their analysis using their data, but restricting the sample to the 82 subnational regions from the nine African countries for which Gennaioli et al. (2014) provide subnational GDP data. These countries are: Benin, Egypt, Kenya, Lesotho, Morocco, Mozambique, Nigeria, South Africa, and Tanzania. Comparing the results reported in Table C1 with those in Hodler and Raschky (2014) suggests that the relation between subnational GDP per capita and subnational nighttime lights is very similar in Africa as elsewhere.

Table C1: Subnational GDP and Nighttime Lights in Africa

| | (1) | (2) |
|---|---|---|
| $Light_{it}$ | 0.291*** | 0.354*** |
| | (0.005) | (0.047) |
| R-squared | 0.688 | 0.688 |
| Observations | 1,200 | 1,200 |
| Region FE | NO | YES |

*Notes:* Dependent variable is the logarithm of regional GDP per capita. OLS regressions. $Light_{it}$ is the logarithm of average nighttime lights. Robust standard errors in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

# D Main Result (complete table) and coefficients of the spatial lags

Table D1: Connectivity based on ethnicity, geography and roads

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | Dependent variable: $Light_{it}$ | | | | |
| $Ethnicity\ W\ Light_{jt}$ | 0.552*** | 0.271 | | | | | 0.160*** | 0.342*** |
| | (0.015) | (0.176) | | | | | (0.013) | (0.122) |
| $Inv\ Dist\ W\ Light_{jt}$ | | | 0.550*** | 0.639*** | | | 0.246*** | 0.305** |
| | | | (0.012) | (0.131) | | | (0.011) | (0.124) |
| $Inv\ Road\ W\ Light_{jt}$ | | | | | 0.556*** | 0.280** | 0.393*** | 0.361*** |
| | | | | | (0.010) | (0.113) | (0.015) | (0.116) |
| $MP_{it}$ | 0.125*** | 0.118*** | 0.120*** | 0.116*** | 0.111*** | 0.112*** | 0.115*** | 0.107*** |
| | (0.014) | (0.015) | (0.013) | (0.014) | (0.013) | (0.014) | (0.014) | (0.016) |
| $Population_{it}$ | 0.243*** | 0.179*** | 0.178*** | -0.023 | 0.162*** | -0.098*** | 0.095*** | -0.220*** |
| | (0.033) | (0.035) | (0.027) | (0.031) | (0.026) | (0.033) | (0.026) | (0.036) |
| $Ethnicity\ W\ Population_{jt}$ | -0.377*** | 0.096 | | | | | -0.171*** | -0.256*** |
| | (0.041) | (0.060) | | | | | (0.035) | (0.057) |
| $Inv\ Dist\ W\ Population_{jt}$ | | | -0.247*** | 0.363*** | | | -0.021 | 0.240*** |
| | | | (0.039) | (0.042) | | | (0.036) | (0.046) |
| $Inv\ Road\ W\ Population_{jt}$ | | | | | -0.128*** | 0.582*** | 0.048 | 0.475*** |
| | | | | | (0.035) | (0.043) | (0.041) | (0.050) |
| $Conflict_{it}$ | -0.011* | -0.012** | -0.011* | -0.011* | -0.010** | -0.013** | -0.008 | -0.006 |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.005) | (0.006) | (0.005) | (0.006) |
| $Ethnicity\ W\ Conflit_{jt}$ | -0.031*** | -0.043** | | | | | -0.019* | -0.020 |
| | (0.011) | (0.018) | | | | | (0.011) | (0.017) |
| $Inv\ Dist\ W\ Conflit_{jt}$ | | | -0.018* | -0.018 | | | -0.009 | -0.012 |
| | | | (0.011) | (0.015) | | | (0.011) | (0.015) |
| $Inv\ Road\ W\ Conflict_{jt}$ | | | | | -0.004 | -0.018 | 0.004 | 0.001 |
| | | | | | (0.009) | (0.012) | (0.010) | (0.013) |
| First stage: | | | | Dependent variable: $Light_{jt}$ | | | | |
| $MP_{jt}$ | | 0.121*** | | 0.124*** | | 0.119*** | | 0.123*** |
| | | (0.015) | | (0.014) | | (0.014) | | (0.015) |
| First-stage F-stat | | 63.83 | | 85.00 | | 71.90 | | 63.64 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |

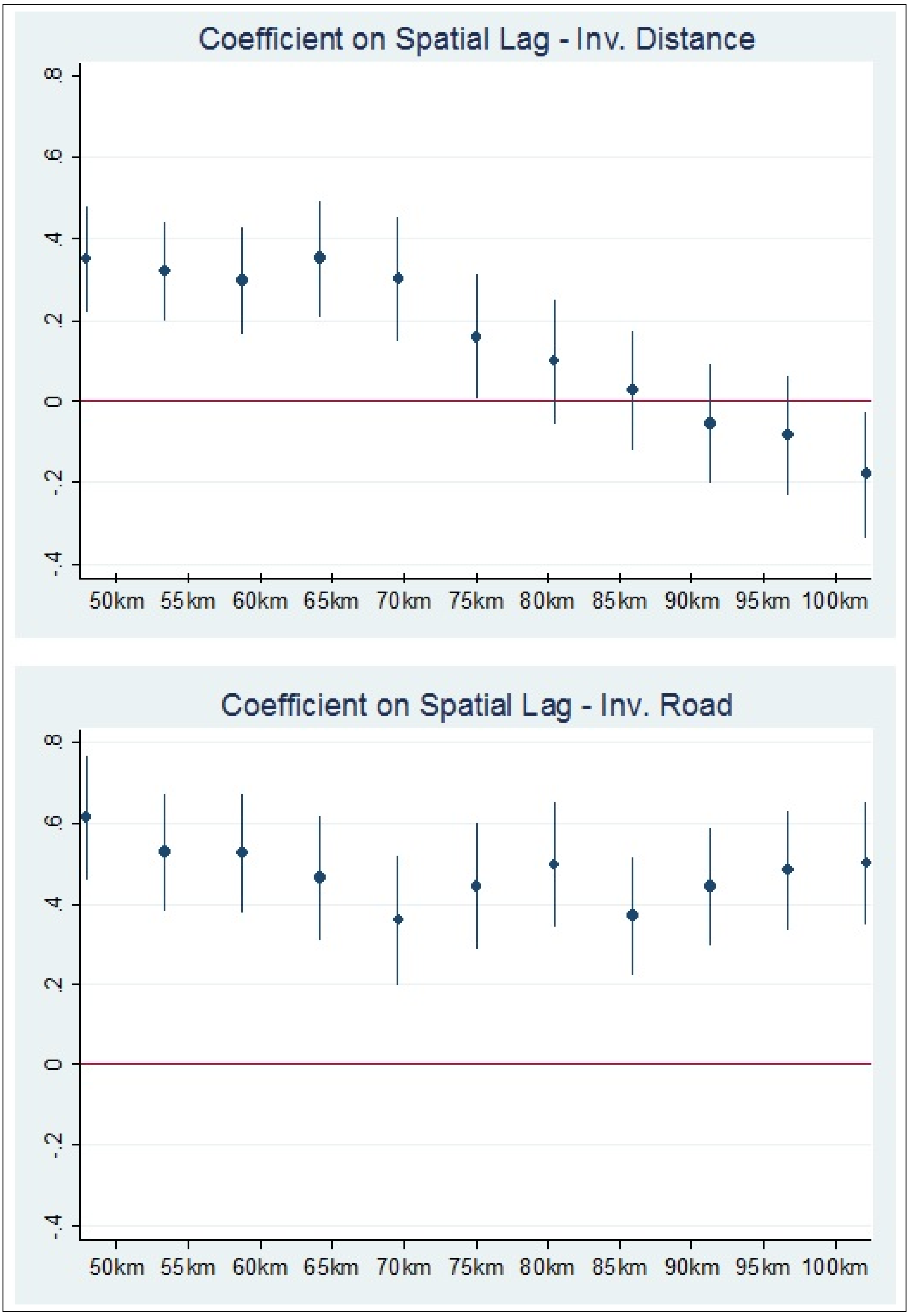
*Notes:* This table corresponds to Table 1 in the main paper, but reports the coefficient estimates on all (second-stage) control variables. Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B1 for the definitions of all variables. The first stage further includes the control variables indicated in Section B1. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10% level, respectively.

Table D2: Main Results - Stepwise addition of control variables

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| | Dependent variable: $Light_{it}$ | | | | | | | | | | | |
| $Ethnicity\ W\ Light_{jt}$ | 0.154*** | 0.291 | 0.154*** | -0.155 | 0.160*** | 0.380*** | 0.154*** | 0.038 | 0.160*** | 0.342*** | 0.221*** | 0.424*** |
| | (0.013) | (0.234) | (0.013) | (0.218) | (0.013) | (0.126) | (0.013) | (0.187) | (0.013) | (0.122) | (0.013) | (0.126) |
| $Inv\ Dist\ W\ Light_{jt}$ | 0.246*** | -0.266* | 0.246*** | -0.460*** | 0.246*** | 0.309** | 0.246*** | -0.374** | 0.246*** | 0.305** | 0.255*** | 0.663*** |
| | (0.011) | (0.148) | (0.011) | (0.151) | (0.011) | (0.128) | (0.011) | (0.147) | (0.011) | (0.124) | (0.011) | (0.114) |
| $Inv\ Road\ W\ Light_{jt}$ | 0.396*** | 0.678*** | 0.396*** | 0.523*** | 0.393*** | 0.377*** | 0.396*** | 0.491*** | 0.393*** | 0.361*** | 0.414*** | 0.540*** |
| | (0.015) | (0.161) | (0.015) | (0.160) | (0.015) | (0.119) | (0.015) | (0.152) | (0.015) | (0.116) | (0.015) | (0.102) |
| $MP_{it}$ | | | 0.115*** | 0.115*** | 0.114*** | 0.106*** | 0.115*** | 0.117*** | 0.115*** | 0.107*** | 0.078*** | 0.142*** |
| | | | (0.014) | (0.015) | (0.014) | (0.016) | (0.014) | (0.015) | (0.014) | (0.016) | (0.078) | (0.019) |
| $Population_{it}$ | | | | | 0.096*** | -0.224*** | | | 0.095*** | -0.220*** | 0.202*** | -0.274*** |
| | | | | | (0.026) | (0.037) | | | (0.026) | (0.036) | (0.028) | (0.038) |
| $Ethnicity\ W\ Population_{jt}$ | | | | | -0.171*** | -0.261*** | | | -0.171*** | -0.256*** | -0.174*** | -0.348*** |
| | | | | | (0.035) | (0.057) | | | (0.035) | (0.057) | (0.031) | (0.064) |
| $Inv\ Dist\ W\ Population_{jt}$ | | | | | -0.018 | 0.237*** | | | -0.021 | 0.240*** | -0.162*** | 0.103** |
| | | | | | (0.036) | (0.045) | | | (0.036) | (0.046) | (0.032) | (0.049) |
| $Inv\ Road\ W\ Population_{jt}$ | | | | | 0.046 | 0.467*** | | | 0.048 | 0.475*** | 0.046 | 0.456*** |
| | | | | | (0.041) | (0.051) | | | (0.041) | (0.050) | (0.041) | (0.053) |
| $Conflict_{it}$ | | | | | | | -0.008 | -0.012** | -0.008 | -0.006 | -0.012** | -0.002 |
| | | | | | | | (0.005) | (0.006) | (0.005) | (0.006) | (0.046) | (0.007) |
| $Ethnicity\ W\ Conflict_{jt}$ | | | | | | | -0.019* | -0.043** | -0.019* | -0.020 | -0.039*** | -0.021 |
| | | | | | | | (0.011) | (0.017) | (0.011) | (0.017) | (0.010) | (0.024) |
| $Inv\ Dist\ W\ Conflict_{jt}$ | | | | | | | -0.008 | -0.014 | -0.009 | -0.012 | -0.012 | 0.043** |
| | | | | | | | (0.011) | (0.015) | (0.011) | (0.015) | (0.010) | (0.019) |
| $Inv\ Road\ W\ Conflict_{jt}$ | | | | | | | 0.004 | -0.004 | 0.004 | 0.001 | -0.001 | 0.002 |
| | | | | | | | (0.010) | (0.013) | (0.010) | (0.013) | (0.010) | (0.014) |
| First stage: | Dependent variable: $Light_{jt}$ | | | | | | | | | | | |
| $MP_{jt}$ | | 0.126*** | | 0.123*** | | 0.122*** | | 0.124*** | | 0.123*** | | 0.081*** |
| | | (0.016) | | (0.015) | | (0.015) | | (0.015) | | (0.015) | | (0.015) |
| First-stage F-stat | | 66.53 | | 65.32 | | 63.07 | | 65.95 | | 63.64 | | 27.84 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Year FE | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | NO | NO |

*Notes:* Even columns report standard fixed effects regressions, and odd columns IV estimates. See Section B1 for the definitions of all variables. The first stage further includes the control variables indicated in Section B1. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10% level, respectively.

Figure D5: Coefficients on the Spatial Lag of the Dependent Variable




*Notes:* Dots show the coefficients on $Inv\,Dist\,W\,Light_{jt}$ and $Inv\,Road\,W\,Light_{jt}$ from the second-stage regression reported in Table 1, column (8), when applying different distance cutoffs for the weighting matrices for geographic and road connectivity. The vertical lines show the 90% confidence interval based on standard errors clustered along the relevant network.

Figure D5 presents the coefficient estimates for $\rho_2$ and $\rho_3$, i.e., the coefficients on $Inv\ Dist\ W\ Light_{jt}$ and $Inv\ Road\ W\ Light_{jt}$, and the corresponding 90% confidence intervals from re-running our main IV specification (corresponding to column (8) of Table 1) for various cutoff distances. Spatial spillovers via purely geographic connectivity are decreasing in the cutoff distance and become statistically insignificant for cutoffs above 70km, while spatial spillovers via the road network remain large in magnitude and statistically significant even for considerably larger cutoff distances. For the subsequent analysis, we use IV estimates that are based on the largest cutoff distance at which the three coefficients of interest, i.e., $\rho_1$, $\rho_2$ and $\rho_3$, are all positive and statistically significant at the 10% level. That is, we use the estimates from column (8) in Table 1, where the cutoff distance is at 70km.

# E Robustness checks

We now discuss a number of robustness checks for which E1–E10 present the corresponding tables.

A first set of checks show that our results are robust to small changes in the empirical specification. Table E1 replaces the country-year fixed effects with province-year fixed effects, thereby controlling for province-specific economic and political variation over time. If anything, the coefficients become larger in our IV estimates. Table E2 adds a temporal lag to the spatial lag of the explanatory variables as spatial spillovers may occur in the future period. Results remain quantitatively similar. Table E3 shows that standard errors become smaller when using the traditional Conley-type spatial clustering approach (Conley, 1999). Table E4 is based on the exactly same specification as our main results, but the unit of observations are rectangular grid cells of 0.5 $\times$ 0.5 degrees (i.e., around 55 $\times$ 55km at the equator) instead of ADM2 regions. Results remain similar, but suggest a slightly more (less) important role of road (geographic) connectivity for the spatial economic spillovers.

A second set of robustness checks tackles potential threats to our identification strategy. In our IV estimates, we exploit the variation of world mineral prices as a source of exogenous shocks, which is then propagated amongst neighbors based on different levels of connectivity. Our identification relies on the assumption that mining activity in a single unit does not influence world mineral prices. Given that our units are subnational districts, this assumption appears reasonable. Nevertheless, Table E5 excludes the districts which belong to countries that are among the top ten producers for any mineral under consideration. Results remain qualitatively similar, but the impact of geographic connectivity decreases. Table E6 shows that our results are not driven by fiscal spillovers. Fiscal spillovers would occur if non-resource-extractive districts benefit from economic activity in resource-extractive districts belonging to the same province purely because government revenues get channeled to resource-rich provinces. To control for fiscal spillovers, we add an additional connectivity matrix that captures whether two districts belong to the same province. The results suggest that the spatial lag related to this new connectivity matrix matters as well and, therefore, that fiscal spillovers may be present. More importantly for our purpose, we see that the spatial lags of our three main connectivity matrices remain quantitatively similar when controlling for fiscal spillovers.

A third set of robustness checks is based on different definitions of the three connectivity matrices. Tables E7 and E8 present results when geographic connectivity is proxied by contiguity and by inverse geodesic distance without adjustment for variability in altitude along the geodesic, respectively. Results remain similar. Table E9 replaces our binary ethnic connectivity matrix with a matrix that suggests an intermediate level of connectivity between districts of related ethnic groups. In particular, a pair of districts is still assigned a value of 1 if they share the same ethnicity, but a value of 0.5 if they do not share the same ethnicity, but belong to the same culture group according to Murdock (1969). A pair of districts that belong to different culture groups still get a value of 0. The coefficient estimates suggest a drop in the importance of the ethnic network, which is consistent with the idea that it is primarily co-ethnicity that matters for spatial economic spillovers.

Lastly, so far, we have made no difference between the spillovers from connected districts within the same country and spillovers from connected districts located in other countries. National borders are likely to affect the magnitude of the spatial economic spillovers, and the impact of borders might be different for each connectivity type. Therefore, we construct two new sets of connectivity matrices: one includes only connected districts in the same country (Table E10), and the other only connected districts in other countries (Table E11). The results in Tables M1 and M2 reveal that road connectivity is the primary source of within country spillovers, while ethnic and geographic connectivity is more important for between country spillovers.

## E.1 Robustness: Province-Year Fixed Effects

Table E1: Province-Year Fixed Effects

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | Dependent variable: $Light_{it}$ | | | | |
| $Ethnicity\ W\ Light_{jt}$ | 0.044** | 0.301 | | | | | -0.029 | 0.469*** |
| | (0.019) | (0.187) | | | | | (0.019) | (0.151) |
| $Inv\ Dist\ W\ Light_{jt}$ | | | 0.118*** | 0.748*** | | | 0.037** | 0.473*** |
| | | | (0.015) | (0.119) | | | (0.014) | (0.101) |
| $Inv\ Road\ W\ Light_{jt}$ | | | | | 0.220*** | 0.779*** | 0.214*** | 0.484*** |
| | | | | | (0.013) | (0.108) | (0.017) | (0.119) |
| First stage: | | | | Dependent variable: $Light_{jt}$ | | | | |
| $MP_{jt}$ | | 0.132*** | | 0.131*** | | 0.133*** | | 0.136*** |
| | | (0.015) | | (0.014) | | (0.015) | | (0.016) |
| First-stage F-stat | | 74.15 | | 89.56 | | 84.44 | | 69.15 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Province-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and province(ADM1)-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\, W\, Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized ethnicity matrix. $Inv\, Dist\, W\, Light_{jt}$ ($Inv\, Road\, W\, Light_{jt}$) is weighted $Light_{jt}$ with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.2 Robustness: Temporal and Spatial Lags

Table E2: Temporal and Spatial Lags

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | Dependent variable: $Light_{it}$ | | | | |
| *Ethnicity W* $Light_{jt-1}$ | 0.350*** | 0.457** | | | | | 0.114*** | 0.356*** |
| | (0.019) | (0.200) | | | | | (0.017) | (0.130) |
| *Inv Dist W* $Light_{jt-1}$ | | | 0.334*** | 0.695*** | | | 0.147*** | 0.366*** |
| | | | (0.014) | (0.138) | | | (0.013) | (0.132) |
| *Inv Road W* $Light_{jt-1}$ | | | | | 0.333*** | 0.279** | 0.229*** | 0.317*** |
| | | | | | (0.011) | (0.117) | (0.014) | (0.122) |
| First stage: | | | | Dependent variable: $Light_{jt-1}$ | | | | |
| $MP_{jt-1}$ | | 0.111*** | | 0.115*** | | 0.111*** | | 0.114*** |
| | | (0.015) | | (0.014) | | (0.015) | | (0.016) |
| First-stage F-stat | | 51.56 | | 65.42 | | 56.89 | | 49.56 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. *Ethnicity W* $Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized ethnicity matrix. *Inv Dist W* $Light_{jt}$ (*Inv Road W* $Light_{jt}$) is weighted $Light_{jt}$ with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.3 Robustness: Spatial Clustering of Standard Errors

Table E3: Spatial Clustering of Standard Errors

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | Dependent variable: $Light_{it}$ | | | | | | | |
| *Ethnicity W* $Light_{jt}$ | 0.552*** | 0.271 | | | | | 0.160*** | 0.342*** |
| | (0.015) | (0.165) | | | | | (0.012) | (0.107) |
| *Inv Dist W* $Light_{jt}$ | | | 0.550*** | 0.639*** | | | 0.246*** | 0.305*** |
| | | | (0.012) | (0.132) | | | (0.011) | (0.112) |
| *Inv Road W* $Light_{jt}$ | | | | | 0.556*** | 0.280** | 0.393*** | 0.361*** |
| | | | | | (0.012) | (0.126) | (0.013) | (0.108) |
| First stage: | Dependent variable: $Light_{jt}$ | | | | | | | |
| $MP_{jt}$ | | 0.121*** | | 0.124*** | | 0.119*** | | 0.123*** |
| | | (0.013) | | (0.014) | | (0.013) | | (0.013) |
| First-stage F-stat | | 82.84 | | 84.89 | | 79.70 | | 84.66 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns report IV estimates. See Section B for the definitions of all variables. *Ethnicity W* $Light_{jt}$ is weighted $Light_{jt}$, with weights based on the row-normalized ethnicity matrix. *Inv Dist W* $Light_{jt}$ (*Inv Road W* $Light_{jt}$) is weighted $Light_{jt}$, with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Spatially clustered standard errors are in parentheses, allowing for spatial correlation up to 70km and for infinite serial correlation. ***, **, * indicate significance at the 1, 5 and 10% level, respectively.

## E.4 Robustness: Grid-Cells as Unit of Observation

Table E4: Grid-Cells as Unit of Observation

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | Dependent variable: $Light_{it}$ | | | | |
| *Eth W* $Light_{jt}$ | 0.517*** | 0.356*** | | | | | 0.168*** | 0.392*** |
| | (0.010) | (0.094) | | | | | (0.010) | (0.055) |
| *Inv Dist W* $Light_{jt}$ | | | 0.518*** | 0.422*** | | | 0.338*** | 0.250*** |
| | | | (0.007) | (0.074) | | | (0.012) | (0.037) |
| *Inv Road W* $Light_{jt}$ | | | | | 0.420*** | 0.604*** | 0.173*** | 0.413*** |
| | | | | | (0.007) | (0.053) | (0.010) | (0.067) |
| First stage: | | | | Dependent variable: $Light_{jt}$ | | | | |
| $MP_{jt}$ | | 0.258*** | | 0.247*** | | 0.231*** | | 0.230*** |
| | | (0.020) | | (0.018) | | (0.018) | | (0.020) |
| First-stage F-stat | | 165.30 | | 197.93 | | 172.06 | | 128.58 |
| Observations | 175,695 | 175,695 | 175,695 | 175,695 | 175,695 | 175,695 | 175,695 | 175,695 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* The units of observation are rectangular grid cells of 0.5 × 0.5 degrees (i.e., around 55 × 55km at the equator). Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns report IV estimates. See Section B for the definitions of all variables. *Ethnicity W* $Light_{jt}$ is weighted $Light_{jt}$, with weights based on the row-normalized ethnicity matrix. *Inv Dist W* $Light_{jt}$ (*Inv Road W* $Light_{jt}$) is weighted $Light_{jt}$, with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.5 Robustness: Excluding Top Mineral Producers

Table E5: Dropping Large Players

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | Dependent variable: $Light_{it}$ | | | | | | | |
| *Ethnicity W* $Light_{jt}$ | 0.546*** | 0.394*** | | | | | 0.120*** | 0.477*** |
| | (0.015) | (0.125) | | | | | (0.014) | (0.102) |
| *Inv Dist W* $Light_{jt}$ | | | 0.630*** | 0.398*** | | | 0.362*** | -0.001 |
| | | | (0.012) | (0.128) | | | (0.015) | (0.107) |
| *Inv Road W* $Light_{jt}$ | | | | | 0.552*** | 0.420*** | 0.317*** | 0.497*** |
| | | | | | (0.011) | (0.092) | (0.014) | (0.098) |
| First stage: | Dependent variable: $Light_{jt}$ | | | | | | | |
| $MP_{jt}$ | | 0.211*** | | 0.214*** | | 0.208*** | | 0.211*** |
| | | (0.019) | | (0.018) | | (0.018) | | (0.019) |
| First-stage F-stat | | 127.84 | | 149.95 | | 132.30 | | 118.38 |
| Observations | 95,030 | 95,030 | 95,030 | 95,030 | 95,030 | 95,030 | 95,030 | 95,030 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Sample is restricted to districts of countries that do not belong to the top ten producers for any mineral under consideration over the period 1997–2013 (see Table A2). Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\, W\, Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized ethnicity matrix. $Inv\, Dist\, W\, Light_{jt}$ ($Inv\, Road\, W\, Light_{jt}$) is weighted $Light_{jt}$ with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.6 Robustness: Fiscal Channel

Table E6: Controlling for Connectivity based on ADM1 Networks

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | | $Light_{it}$ | | | |
| $Ethnicity\ W\ Light_{jt}$ | 0.315*** | 0.274 | | | | | 0.126*** | 0.295** |
| | (0.014) | (0.169) | | | | | (0.013) | (0.123) |
| $Inv\ Dist\ W\ Light_{jt}$ | | | 0.386*** | 0.522*** | | | 0.209*** | 0.288** |
| | | | (0.013) | (0.127) | | | (0.012) | (0.114) |
| $Inv\ Road\ W\ Light_{jt}$ | | | | | 0.441*** | 0.396*** | 0.360*** | 0.382*** |
| | | | | | (0.012) | (0.115) | (0.015) | (0.108) |
| $ADM1\ W\ Light_{jt}$ | 0.434*** | 0.256 | 0.329*** | 0.834*** | 0.270*** | 0.236 | 0.142*** | 0.292** |
| | (0.018) | (0.266) | (0.014) | (0.152) | (0.013) | (0.148) | (0.014) | (0.130) |
| First stage: | | | | | $Light_{jt}$ | | | |
| $MP_{jt}$ | | 0.123*** | | 0.126*** | | 0.122*** | | 0.124*** |
| | | (0.015) | | (0.015) | | (0.015) | | (0.016) |
| First-stage F-stat | | 63.59 | | 74.76 | | 66.71 | | 62.21 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country Year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional Controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\,W\,Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized ethnicity matrix. $Inv\,Dist\,W\,Light_{jt}$ ($Inv\,Road\,W\,Light_{jt}$) is weighted $Light_{jt}$ with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. $ADM1\,W\,Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized ADM1 matrix, which identifies whether districts belong to the same ADM1 unit. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.7 Robustness: Contiguity Network

Table E7: Connectivity based on contiguity

| | (1) | (2) |
|---|---|---|
| | OLS | IV |
| | Dependent variable: $Light_{it}$ | |
| *Contiguity W* $Light_{jt}$ | 0.746*** | 0.607*** |
| | (0.008) | (0.158) |
| First stage: | Dependent variable: $Light_{jt}$ | |
| $MP_{jt}$ | | 0.127*** |
| | | (0.014) |
| First-stage F-Stat | | 83.17 |
| Observations | 101,048 | 101,048 |
| District FE | YES | YES |
| Country-year FE | YES | YES |
| Additional controls | YES | YES |

*Notes:* Column (1) reports the standard fixed effects regression with district and country-year fixed effects, and column (2) the IV estimates. See Section B for the definitions of all variables. *Contiguity W* $Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized contiguity matrix. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.8 Robustness: Geodesic Network without Adjustment for Variability in Altitude

Table E8: Geodesic Network without Adjustment for Variability in Altitude

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| | Dependent variable: $Light_{it}$ | | | |
| $Ethnicity\ W\ Light_{jt}$ | | | 0.110*** | 0.396*** |
| | | | (0.013) | (0.124) |
| $Inv\ Dist\ W\ Light_{jt}$ | 0.664*** | 0.750*** | 0.374*** | 0.439*** |
| | (0.012) | (0.156) | (0.013) | (0.132) |
| $Inv\ Road\ W\ Light_{jt}$ | | | 0.330*** | 0.396*** |
| | | | (0.014) | (0.113) |
| First stage: | Dependent variable: $Light_{jt}$ | | | |
| $MP_{jct}$ | | 0.127*** | | 0.124*** |
| | | (0.014) | | (0.015) |
| First-stage F-stat | | 88.61 | | 64.99 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\, W\, Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized ethnicity matrix. $Inv\, Dist\, W\, Light_{jt}$ ($Inv\, Road\, W\, Light_{jt}$) is weighted $Light_{jt}$ with weights based on the row-normalized matrix of the inverse geodesic distances without adjustment for the variability in altitude (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (3) in the main text). The first stage further includes the control variables indicated in equation (2) in the main text. Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.9 Robustness: Alternative Ethnicity Network

Table E9: Alternative Ethnicity Network

| | (1) OLS | (2) IV | (3) OLS | (4) IV |
|---|---|---|---|---|
| | Dependent variable: $Light_{it}$ | | | |
| $Ethnicity\ W\ Light_{jt}$ | 0.750*** | 1.146 | 0.189*** | 0.717 |
| | (0.032) | (0.710) | (0.017) | (0.467) |
| $Inv\ Dist\ W\ Light_{jt}$ | | | 0.283*** | 0.271** |
| | | | (0.013) | (0.109) |
| $Inv\ Road\ W\ Light_{jt}$ | | | 0.410*** | 0.426*** |
| | | | (0.017) | (0.112) |
| First stage: | Dependent variable: $Light_{jt}$ | | | |
| $MP_{jt}$ | | 0.121*** | | 0.123*** |
| | | (0.024) | | (0.023) |
| First-stage F-stat | | 26.72 | | 29.07 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\,W\,Light_{jt}$ is weighted $Light_{jt}$ with weights based on the row-normalized alternative ethnicity matrix, as discussed in Section 6.2. $Inv\,Dist\,W\,Light_{jct}$ ($Inv\,Road\,W\,Light_{jt}$) is weighted $Light_{jt}$ with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{it}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jt}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (13)). The first stage further includes the control variables indicated in equation (12). Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## E.10 Networks within and across Countries

We create two new sets of connectivity matrices: The first one only includes connected districts $j$ which are in the same country as district $i$ (*Within Country*), while the second one only includes connected districts $j$ that are in other countries than district $i$ (*Outside Country*).

Table E10 below looks at only within country connectivity. Accordingly, the ethnicity matrix here only captures districts which are of the same ethnicity and belong to the same country. The inverse distance (road) matrix captures districts where the geodesic (road) distance is less than 70km and which belong to the same country. Compared to the main specification that allows for spill-overs within and between countries, the estimated $\rho$'s for ethnic and inverse distance connectivity are also positive but smaller in magnitude and no longer statically significant. In contrast, the $\rho$ for road connectivity is larger in magnitude and highly statistically significant.

Table E11 isolates outside-country spillover effects. Accordingly, the ethnicity matrix here only captures district which are of the same ethnicity and which *do not* belong to the same country. The inverse distance (road) matrix captures districts where the geodesic (road) distance is less than 70km, and *do not* belong to the same country. Here a different pattern emerges. Ethnic and geographic connectivity are more important for between country spill-overs, while the effect of road-connectivity is smaller and not statistically significant.

Table E10: Within-Country Networks

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | | $Light_{ict}$ | | | |
| $Ethnicity\ W\ Light_{jct}$ | 0.484*** | -0.035 | | | | | 0.144*** | 0.147 |
| $(Within\ Country)$ | (0.014) | (0.140) | | | | | (0.013) | (0.101) |
| $Inv\ Dist\ W\ Light_{jct}$ | | | 0.504*** | 0.393*** | | | 0.236*** | 0.115 |
| $(Within\ Country)$ | | | (0.011) | (0.110) | | | (0.012) | (0.113) |
| $Inv\ Road\ W\ Light_{jct}$ | | | | | 0.479*** | 0.278*** | 0.318*** | 0.482*** |
| $(Within\ Country)$ | | | | | (0.011) | (0.103) | (0.015) | (0.106) |
| First stage: | | | | | $Light_{jct}$ | | | |
| $MP_{jct}$ | | 0.121*** | | 0.128*** | | 0.117*** | | 0.122*** |
| | | (0.014) | | (0.014) | | (0.014) | | (0.015) |
| First-stage F-stat | | 75.79 | | 90.17 | | 70.16 | | 68.83 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\,W\,Light_{jct}\,(Within\,Country)$ is weighted $Light_{jct}$ for districts belonging to the same country, with weights based on the row-normalized ethnicity matrix. $Inv\,Dist\,W\,Light_{jt}\,(Within\,Country)$ $(Inv\,Road\,W\,Light_{jct}\,(Within\,Country))$ is weighted $Light_{jct}$ for districts belonging to the same country, with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{ict}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jct}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (13)). The first stage further includes the control variables indicated in equation (12). Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

## Table E11: Outside-Country Networks

| | (1) OLS | (2) IV | (3) OLS | (4) IV | (5) OLS | (6) IV | (7) OLS | (8) IV |
|---|---|---|---|---|---|---|---|---|
| | | | | | $Light_{ict}$ | | | |
| *Ethnicity W* $Light_{jct}$ | 0.262*** | 0.291** | | | | | 0.177*** | 0.382*** |
| (*Outside Country*) | (0.010) | (0.146) | | | | | (0.010) | (0.096) |
| *Inv Dist W* $Light_{jct}$ | | | 0.221*** | 0.504*** | | | 0.102*** | 0.678*** |
| (*Outside Country*) | | | (0.010) | (0.140) | | | (0.011) | (0.123) |
| *Inv Road W* $Light_{jct}$ | | | | | 0.228*** | 0.380*** | 0.151*** | 0.144 |
| (*Outside Country*) | | | | | (0.009) | (0.111) | (0.010) | (0.105) |
| First stage: | | | | | $Light_{jct}$ | | | |
| $MP_{jct}$ | | 0.127*** | | 0.126*** | | 0.116*** | | 0.126*** |
| | | (0.013) | | (0.012) | | (0.013) | | (0.013) |
| First-stage F-stat | | 91.73 | | 105.79 | | 86.46 | | 90.49 |
| Observations | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 | 101,048 |
| District FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Country-year FE | YES | YES | YES | YES | YES | YES | YES | YES |
| Additional controls | YES | YES | YES | YES | YES | YES | YES | YES |

*Notes:* Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns IV estimates. See Section B for the definitions of all variables. $Ethnicity\,W\,Light_{jct}\,(Outside\,Country)$ is weighted $Light_{jct}$ for districts not belonging to the same country, with weights based on the row-normalized ethnicity matrix. $Inv\,Dist\,W\,Light_{jct}\,(Outside\,Country)$ ($Inv\,Road\,W\,Light_{jct}\,(Outside\,Country)$) is weighted $Light_{jct}$ for districts not belonging to the same country, with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and $MP_{ict}$ as well as weighted population and conflict in districts $j \neq i$. $MP_{jct}$ is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (13)). The first stage further includes the control variables indicated in equation (12). Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10%-level, respectively.

# F Key Player Rankings

Table F1 presents information on the ten most central districts (according to the key-player centrality) of each of these two countries; and Figure 1 compares key-player centrality of districts (top row) with the districts' average nighttime light intensity (middle row) and population density (bottom row).

Column (4) of Table F1 presents the main ranking of interest, i.e., the key-player ranking based on the geographic network, the road network *and* the ethnicity network. The underlying computation thus uses the coefficient estimates, in particular the estimated $\rho$'s, reported in column (8) of Table 1. The top 7 districts with the highest key-player centrality are part of the Lagos metropolitan area which is the primate city of Nigeria and its economic hub. Seven other districts belonging to the top-ten key districts of Nigeria are also part of Lagos State. The two remaining districts in the key-player ranking belong to the Kano metropolitan area which is the second largest metropolitan area in Nigeria and the economic hub of the country's north.

The key district in Kenya is Nairobi, which is the capital and the primate city. It is followed by Mombassa, which is Kenya's second largest city and home to Kenya's largest seaport (see the right column of Figure 1). The key districts encompass or are part of the primate city in many other African countries as well, including Ethiopia (Addis Ababa) and South Africa (Johannesburg). The overall pattern suggests that *primate cities* tend to be the key districts development in Africa which resonates with the findings of Ades and Glaeser (1995), Henderson (2002), or Storeygard (2016), among others.

Column (5) in Table F1 shows the ranking for the Katz-Bonacich centrality, again based on the estimates taking the geographic network, the road network and the ethnicity network into account. We see that the districts that rank high in terms of key-player centrality also tend to rank high in terms of Katz-Bonacich centrality in Nigeria, but not in Kenya. This is because Katz-Bonacich and key-player centralities capture different aspects of centrality. The former is a recursive measure highlighting the importance of being connected to central districts while the latter is a welfare measure that also takes into account the negative impact of cutting links on neighboring districts.

Columns (6) and (7) show the rankings for the two other centrality measures: betweenness and eigenvector centrality. Looking at Nigeria, we see that the districts from Lagos State that are top ranked in terms of key-player centrality tend to rank poorly in terms of these alternative centrality measures. This is not surprising given that the betweenness and eigenvector centralities are pure topological measures, which capture either the number of paths that go through a district (betweenness centrality) or the links to other central districts (eigenvector centrality), and Lagos State is situated at the coast in the country's south-east bordering Benin.

Columns (8)-(10) also give rankings of key-player centrality, but in each of these columns we compute the ranking based on the coefficient estimates from regressions including one network only. We see that most districts that rank high in overall key-player centrality also rank high in any type of single-network key-player centrality. This suggests that most key districts are important due to their geographic, ethnic *and* road connectivity. For many countries, the overall key-player centrality is most highly correlated with the key-player centrality based on the road network, which indicates that road connectivity may be of particular importance.

Table F1: Top-Ten Key Player Rankings

| (1) Country | (2) Province | (3) District Rank | (4) Overall KP Rank | (5) Overall Katz-Bon Rank | (6) Overall Betw. Rank | (7) Overall Eig. Rank | (8) Ethnicity KP Rank | (9) Road KP Rank | (10) Inv.Dist KP Rank |
|---|---|---|---|---|---|---|---|---|---|
| | | | **Nigeria** | | | | | | |
| Nigeria | Lagos | Ikeja | 1 | 16 | 564 | 428 | 4 | 1 | 2 |
| Nigeria | Lagos | LagosIsland | 2 | 24 | 707 | 424 | 32 | 2 | 18 |
| Nigeria | Kano | Fagge | 3 | 237 | 479 | 475 | 6 | 5 | 6 |
| Nigeria | Lagos | Agege | 4 | 14 | 641 | 433 | 8 | 14 | 4 |
| Nigeria | Lagos | Ajeromi/Ifelodun | 5 | 17 | 578 | 438 | 5 | 13 | 9 |
| Nigeria | Lagos | Apapa | 6 | 21 | 713 | 439 | 10 | 17 | 14 |
| Nigeria | Kano | Tarauni | 7 | 239 | 579 | 475 | 14 | 9 | 11 |
| Nigeria | Lagos | Mainland | 8 | 23 | 712 | 433 | 1 | 3 | 1 |
| Nigeria | Lagos | Surulere | 9 | 19 | 569 | 433 | 3 | 4 | 7 |
| Nigeria | Lagos | Amuwo Odofin | 10 | 12 | 432 | 440 | 26 | 7 | 12 |
| | | | **Kenya** | | | | | | |
| Kenya | Nairobi | Nairobi* | 1 | 23 | 41 | 4 | 1 | 1 | 1 |
| Kenya | Coast | Mombasa | 2 | 41 | 45 | 8 | 2 | 2 | 2 |
| Kenya | Coast | Kwale | 3 | 37 | 9 | 8 | 17 | 48 | 3 |
| Kenya | Rift Valley | Nakuru | 4 | 20 | 3 | 26 | 8 | 4 | 5 |
| Kenya | Central | Kiambu | 5 | 24 | 40 | 5 | 4 | 3 | 4 |
| Kenya | Eastern | Machakos | 6 | 30 | 17 | 5 | 9 | 46 | 6 |
| Kenya | Central | Murang'a | 7 | 22 | 29 | 3 | 5 | 6 | 7 |
| Kenya | Central | Nyeri | 8 | 25 | 18 | 25 | 7 | 7 | 8 |
| Kenya | Rift Valley | Narok | 9 | 19 | 12 | 30 | 23 | 42 | 35 |
| Kenya | Central | Kirinyaga | 10 | 21 | 31 | 7 | 19 | 9 | 11 |

*Notes:* Overall KP Rank is based on the $\rho$s estimated in column (8) of Table 1. Overall Katz-Bonacich Rank is based on the $\rho$s estimated in column (8) of Table 1 and a weighting vector of 1. Ethnicity KP Rank is based on $\rho_1$ estimated in column (2) of Table 1. Inv.Dist KP Rank is based on $\rho_2$ estimated in column (4) of Table 1. Road KP Rank is based on $\rho_3$ estimated in column (6) Table 1. * indicate districts that are (part of) capital cities. Nigeria has 775 districts, and Kenya has 48 districts.

Table F2: Top-Ten Key Player Rankings for Populous Countries

| (1) Country | (2) Province | (3) District Rank | (4) Overall KP Rank | (5) Overall Katz-Bon Rank | (6) Overall Betw. Rank | (7) Overall Eig. Rank | (8) Ethnicity KP Rank | (9) Road KP Rank | (10) Inv.Dist KP Rank |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **Ethiopia** | | | | | |
| Ethiopia | Addis Ababa | Zone 4* | 1 | 6 | 47 | 14 | 2 | 2 | 1 |
| Ethiopia | Addis Ababa | Zone 3* | 2 | 7 | 38 | 16 | 6 | 4 | 3 |
| Ethiopia | Addis Ababa | Zone 2* | 3 | 8 | 43 | 16 | 3 | 3 | 5 |
| Ethiopia | Addis Ababa | Zone 6* | 4 | 10 | 59 | 16 | 10 | 11 | 7 |
| Ethiopia | Oromia | North Shewa (K4) | 5 | 12 | 48 | 67 | 27 | 8 | 8 |
| Ethiopia | Addis Ababa | Zone 5* | 6 | 9 | 44 | 14 | 4 | 6 | 2 |
| Ethiopia | Tigray | Mekele | 7 | 2 | 49 | 70 | 5 | 7 | 6 |
| Ethiopia | Addis Ababa | Addis Ababa* | 8 | 4 | 38 | 16 | 1 | 1 | 4 |
| Ethiopia | Amhara | West Gojam | 9 | 17 | 10 | 9 | 18 | 9 | 9 |
| Ethiopia | Amhara | Bar Dar Sp. Zone | 10 | 14 | 59 | 57 | 7 | 10 | 10 |
| | | | | **Egypt** | | | | | |
| Egypt | Al Qalyubiyah | | 1 | 9 | 17 | 3 | 1 | 1 | 1 |
| Egypt | Al Gharbiyah | | 2 | 5 | 11 | 1 | 3 | 4 | 3 |
| Egypt | Al Minufiyah | | 3 | 10 | 16 | 6 | 2 | 3 | 4 |
| Egypt | Ash Sharqiyah | | 4 | 6 | 6 | 2 | 6 | 5 | 5 |
| Egypt | Ad Daqahliyah | | 5 | 2 | 8 | 5 | 9 | 11 | 2 |
| Egypt | Dumyat | | 6 | 1 | 22 | 20 | 8 | 9 | 11 |
| Egypt | Al Buhayrah | | 7 | 11 | 7 | 7 | 13 | 10 | 7 |
| Egypt | Al Qahirah* | | 8 | 8 | 15 | 8 | 10 | 6 | 6 |
| Egypt | Bani Suwayf | | 9 | 14 | 13 | 10 | 15 | 16 | 12 |
| Egypt | Kafr ash Shaykh | | 10 | 4 | 20 | 14 | 12 | 13 | 15 |
| | | | | **Democratic Republic of the Congo (DRC)** | | | | | |
| DRC | Kivu | Sud-Kivu | 1 | 3 | 9 | 2 | 20 | 1 | 20 |
| DRC | Bas-Congo | Matadi | 2 | 15 | 30 | 1 | 2 | 2 | 1 |
| DRC | Kasaï-Oriental | Mbuji-Mayi | 3 | 5 | 30 | 2 | 3 | 5 | 4 |
| DRC | Kasaï-Oriental | Tshilenge | 4 | 7 | 29 | 2 | 17 | 37 | 4 |
| DRC | Katanga | Lubumbashi | 5 | 11 | 30 | 2 | 1 | 3 | 2 |
| DRC | Kivu | Bukavu | 6 | 1 | 28 | 37 | 6 | 8 | 6 |
| DRC | Bas-Congo | Boma | 7 | 2 | 26 | 38 | 4 | 7 | 3 |
| DRC | Kinshasa City | Kinshasa* | 8 | 18 | 18 | 2 | 5 | 6 | 7 |
| DRC | Kivu | Nord-Kivu | 9 | 9 | 8 | 2 | 19 | 9 | 19 |
| DRC | Bas-Congo | Bas-Fleuve | 10 | 4 | 25 | 36 | 18 | 10 | 8 |
| | | | | **South Africa (SA)** | | | | | |
| SA | Gauteng | Johannesburg* | 1 | 82 | 228 | 4 | 1 | 1 | 1 |
| SA | Western Cape | Kuils River* | 2 | 263 | 339 | 228 | 10 | 3 | 5 |
| SA | Gauteng | Roodepoort* | 3 | 86 | 266 | 10 | 4 | 11 | 3 |
| SA | Western Cape | Mitchells Plain* | 4 | 261 | 344 | 228 | 8 | 2 | 8 |
| SA | Gauteng | Benoni* | 5 | 73 | 257 | 12 | 7 | 10 | 6 |
| SA | KwaZulu-Natal | Durban | 6 | 13 | 129 | 199 | 9 | 4 | 7 |
| SA | Gauteng | Boksburg* | 7 | 72 | 108 | 3 | 2 | 8 | 2 |
| SA | Gauteng | Kempton Park* | 8 | 77 | 200 | 5 | 17 | 6 | 9 |
| SA | Gauteng | Soweto* | 9 | 85 | 245 | 6 | 3 | 7 | 4 |
| SA | Gauteng | Soshanguve* | 10 | 104 | 313 | 26 | 14 | 12 | 13 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | **Tanzania** | | | | | |
| Tanzania | Dar-Es-Salaam | Ilala | 1 | 124 | 39 | 82 | 5 | 3 | 1 |
| Tanzania | Dar-Es-Salaam | Kinondoni | 2 | 126 | 37 | 7 | 4 | 1 | 2 |
| Tanzania | Arusha | Arusha | 3 | 40 | 110 | 83 | 3 | 2 | 4 |
| Tanzania | Dar-Es-Salaam | Temeke | 4 | 125 | 119 | 86 | 7 | 4 | 5 |
| Tanzania | Arusha | Arumeru | 5 | 39 | 99 | 28 | 24 | 7 | 18 |
| Tanzania | Kilimanjaro | Moshi Urban | 6 | 31 | 111 | 14 | 6 | 5 | 6 |
| Tanzania | Kilimanjaro | Moshi Rural | 7 | 36 | 26 | 17 | 21 | 15 | 30 |
| Tanzania | Kaskazini-Unguja | Kaskazini 'B' | 8 | 131 | 103 | 4 | 25 | 11 | 9 |
| Tanzania | Kilimanjaro | Rombo | 9 | 33 | 81 | 17 | 63 | 34 | 132 |
| Tanzania | Mwanza | Nyamagana | 10 | 28 | 114 | 111 | 2 | 6 | 3 |

*Notes:* Overall key-player (KP) rank is based on the $\rho$s estimated in column (8) of Table 1. Overall Katz-Bonacich rank is based on the $\rho$s estimated in column (8) of Table 1 and a weighting vector of 1. Ethnicity KP rank is based on $\rho_1$ estimated in column (2) of Table 1. Inverse distance KP rank is based on $\rho_2$ estimated in column (4) of Table 1. Road KP rank is based on $\rho_3$ estimated in column (6) Table 1. * indicate districts that are (part of) capital cities. South Africa has three capital cities i.e. Pretoria (Gauteng), Bloemfontein (Free State) and Cape Town (Western Cape). The number of districts per country are: Ethiopia 72, Egypt 26 (ADM1 level), DRC 38, South Africa 354, and Tanzania 136.

# G Top-Ten Rankings from Policy Experiments for Kenya and Nigeria

A few comments are in order before presenting these two policy experiments: First, the socially optimal location of a development project depends on costs and benefits, and our approach does not take into account the fact that the costs of implementing a certain project or building a certain road may differ across districts. Second, it is impossible to compare the benefits of different development projects or different project locations without an underlying social welfare function. Here, as in the previous section, we (implicitly) measure social welfare in a district by the logarithm of the average nighttime light pixel value, and we give equal weight to all districts when computing aggregate social welfare. Needless to say, one could apply our approach using alternative social welfare functions. Third, these policy experiments do not explicitly take into account the congestion effects that may occur in urban districts when new people move in.[19] Therefore, our counterfactual policy experiments are most informative about short- to medium-run effects rather than long-run effects.

Table G1 presents the ten districts in Nigeria and Kenya where a counterfactual increase in economic activity (see Section 8.1) would have the largest overall impact. Table G2 presents the ten districts in Nigeria and Kenya where a counterfactual improvement of the road connectivity (see Section 8.2) would have the largest overall impact.

[19]Observe that the vast majority of our districts in the network are rural districts, and thus the congestion effect might be less of a concern.

Table G1: Top-Ten Rankings from Policy Experiment 1 – Nighttime Lights

| (1) Country | (2) Province | (3) District | (4) Overall Rank | (5) Overall Key-Player Rank |
|---|---|---|---|---|
| | | **Nigeria** | | |
| Nigeria | Lagos | Surulere | 1 | 9 |
| Nigeria | Lagos | Mainland | 2 | 8 |
| Nigeria | Lagos | Shomolu | 3 | 16 |
| Nigeria | Lagos | Oshodi/Isolo | 4 | 18 |
| Nigeria | Bayelsa | Nembe | 5 | 771 |
| Nigeria | Lagos | Amuwo Odofin | 6 | 10 |
| Nigeria | Lagos | Alimosho | 7 | 14 |
| Nigeria | Lagos | Kosofe | 8 | 25 |
| Nigeria | Bayelsa | Brass | 9 | 769 |
| Nigeria | Delta | Warri South-West | 10 | 772 |
| | | **Kenya** | | |
| Kenya | Coast | Lamu | 1 | 48 |
| Kenya | Central | Machakos | 2 | 47 |
| Kenya | Coast | Kilifi | 3 | 46 |
| Kenya | Coast | Kwale | 4 | 3 |
| Kenya | Central | Kiambu | 5 | 5 |
| Kenya | Central | Murang'a | 6 | 7 |
| Kenya | Eastern | Machakos | 7 | 6 |
| Kenya | Eastern | Embu | 8 | 19 |
| Kenya | Rift Valley | Nakuru | 9 | 4 |
| Kenya | Central | Kirinyaga | 10 | 10 |

*Notes:* Overall rank reflects the district's overall impact from increasing its average nighttime light pixel value by 10 on average nighttime lights across African districts (see Section B for a more detailed explanation). This counterfactual exercise is based on the $\rho$s estimated in column (8) of Table 1. Nigeria has 775 districts, and Kenya has 48 districts.

Table G2: Top-Ten Rankings from Policy Experiment 2 – Roads

| (1) Country | (2) Province | (3) District | (4) Overall Rank | (5) Overall Key-Player Rank |
|---|---|---|---|---|
| **Nigeria** | | | | |
| Nigeria | Rivers | Bonny | 1 | 763 |
| Nigeria | Rivers | Okrika | 1 | 21 |
| Nigeria | Rivers | Khana | 3 | 99 |
| Nigeria | Delta | Burutu | 4 | 75 |
| Nigeria | Delta | Warri North | 5 | 79 |
| Nigeria | Rivers | Andoni/O | 6 | 127 |
| Nigeria | Delta | Ughelli South | 7 | 45 |
| Nigeria | Rivers | Abua/Odu | 8 | 42 |
| Nigeria | Rivers | Akukutor | 9 | 766 |
| Nigeria | Bayelsa | Yenegoa | 10 | 31 |
| **Kenya** | | | | |
| Kenya | Central | Machakos | 1 | 47 |
| Kenya | Eastern | Wajir | 2 | 31 |
| Kenya | Eastern | Meru | 3 | 43 |
| Kenya | Central | Nyeri | 3 | 8 |
| Kenya | Central | Kirinyaga | 5 | 10 |
| Kenya | Central | Murang'a | 6 | 7 |
| Kenya | Eastern | Machakos | 6 | 6 |
| Kenya | Rift Valley | Narok | 8 | 9 |
| Kenya | Coast | Kwale | 9 | 3 |
| Kenya | Rift Valley | Bomet | 10 | 24 |
| Kenya | Rift Valley | Nakuru | 10 | 4 |

*Notes:* Overall rank reflects the district's overall impact from adding a road link to the contiguous district with the highest average nighttime light pixel value to which there exists no road link (see Section 4 in the main text for a more detailed explanation). This counterfactual exercise is based on the $\rho$s estimated in column (8) of Table 1. Nigeria has 775 districts, and Kenya has 48 districts.

# H Top-Ten Rankings from Policy Experiments for Other Populous Countries

Table H1 presents the ten districts in Ethiopia, Egypt, DRC, South Africa, and Tanzania where a counterfactual increase in economic activity (see Section B) would have the largest overall impact. Table H2 presents the ten districts in the same country where a counterfactual improvement of the road connectivity (see Section B) would have the largest overall impact.

Table H1: Top-Ten Rankings from Policy Experiment 1 (Nighttime Lights) for Populous Countries

| (1) Country | (2) Province | (3) District | (4) Overall Rank | (5) Overall KP Rank |
|---|---|---|---|---|
| | | **Ethiopia** | | |
| Ethiopia | Amhara | West Gojam | 1 | 9 |
| Ethiopia | Addis Ababa | Addis Ababa | 2 | 8 |
| Ethiopia | Addis Ababa | Zone 5 | 3 | 6 |
| Ethiopia | Addis Ababa | Zone 4 | 4 | 1 |
| Ethiopia | Afar | Zone 5 | 5 | 70 |
| Ethiopia | Oromia | East Shewa | 6 | 69 |
| Ethiopia | Oromia | West Shewa | 7 | 13 |
| Ethiopia | Amhara | North Shewa (K3) | 8 | 68 |
| Ethiopia | Addis Ababa | Zone 2 | 9 | 3 |
| Ethiopia | Addis Ababa | Zone 3 | 10 | 2 |
| | | **Egypt** | | |
| Egypt | Suhaj | | 1 | 18 |
| Egypt | Al Jizah | | 2 | 24 |
| Egypt | Asyut | | 3 | 19 |
| Egypt | Qina | | 4 | 20 |
| Egypt | Al Minya | | 5 | 22 |
| Egypt | Matruh | | 6 | 26 |
| Egypt | As Suways | | 7 | 23 |
| Egypt | Al Bahr al Ahmar | | 8 | 25 |
| Egypt | Ad Daqahliyah | | 9 | 5 |
| Egypt | Al Wadi al Jadid | | 10 | 13 |

| | | | | |
|---|---|---|---|---|
| **Democratic Republic of the Congo (DRC)** | | | | |
| DRC | Katanga | Haut-Shaba | 1 | 37 |
| DRC | Équateur | Sud-Ubangi | 2 | 38 |
| DRC | Kivu | Sud-Kivu | 3 | 1 |
| DRC | Katanga | Lubumbashi | 4 | 5 |
| DRC | Kivu | Nord-Kivu | 5 | 9 |
| DRC | Bas-Congo | Boma | 6 | 7 |
| DRC | Kinshasa City | Kinshasa | 7 | 8 |
| DRC | Bandundu | Mai-Ndombe | 8 | 32 |
| DRC | Kasaï-Oriental | Tshilenge | 9 | 4 |
| DRC | Bas-Congo | Cataractes | 10 | 36 |
| **South Africa (SA)** | | | | |
| SA | Western Cape | Wynberg | 1 | 11 |
| SA | Gauteng | Pretoria | 2 | 22 |
| SA | Gauteng | Kempton Park | 3 | 8 |
| SA | Gauteng | Randburg | 4 | 16 |
| SA | Gauteng | Wonderboom | 5 | 37 |
| SA | Gauteng | Germiston | 6 | 13 |
| SA | Western Cape | Goodwood | 7 | 26 |
| SA | Gauteng | Alberton | 8 | 21 |
| SA | Gauteng | Bronkhorstspruit | 9 | 339 |
| SA | Gauteng | Johannesburg | 10 | 1 |
| **Tanzania** | | | | |
| Tanzania | Zanzibar West | Magharibi | 1 | 135 |
| Tanzania | Morogoro | Morogoro Rural | 2 | 136 |
| Tanzania | Iringa | Iringa Rural | 3 | 16 |
| Tanzania | Mwanza | Ilemela | 4 | 134 |
| Tanzania | Pwani | Mafia | 5 | 130 |
| Tanzania | Zanzibar South and Central | Zansibar Central | 6 | 11 |
| Tanzania | Arusha | Arumeru | 7 | 5 |
| Tanzania | Arusha | Simanjiro | 8 | 127 |
| Tanzania | Kaskazini-Unguja | Kaskazini 'B' | 9 | 8 |
| Tanzania | Kilimanjaro | Moshi Rural | 10 | 7 |

*Notes:* Overall rank is based on counterfactual analysis described in Section 4 in the main text and the $\rho$s estimated in column (8) of Table 1. Overall key-player (KP) rank is based on the $\rho$s estimated in column (8) of Table 1 as well. The number of districts per country are: Ethiopia 72, Egypt 26 (ADM1 level), DRC 38, SA 354, and Tanzania 136.

Table H2: Top-Ten Rankings from Policy Experiment 2 (Roads) for Populous Countries

| (1) Country | (2) Province | (3) District | (4) Overall Rank | (5) Overall KP Rank |
|---|---|---|---|---|
| | | **Ethiopia** | | |
| Ethiopia | Tigray | Central Tigray | 1 | 17 |
| Ethiopia | Tigray | Easetern Tigray | 2 | 16 |
| Ethiopia | Afar | Zone 5 | 3 | 70 |
| Ethiopia | Tigray | Southern Tigray | 4 | 65 |
| Ethiopia | Afar | Zone 4 | 4 | 71 |
| Ethiopia | Amhara | South Gonder | 6 | 66 |
| Ethiopia | Amhara | West Gojam | 6 | 9 |
| Ethiopia | Amhara | North Wollo | 8 | 54 |
| Ethiopia | Addis Ababa | Zone 5 | 9 | 6 |
| Ethiopia | SNNP* | Konso Special Woreda | 10 | 29 |
| **Southern Nations, Nationalities and Peoples* | | | | |
| | | **Egypt** | | |
| Egypt | Al Qalyubiyah | | 1 | 1 |
| Egypt | Ad Daqahliyah | | 1 | 5 |
| Egypt | Asyut | | 3 | 19 |
| Egypt | Al Minufiyah | | 3 | 3 |
| Egypt | Suhaj | | 3 | 18 |
| Egypt | Al Fayyum | | 3 | 11 |
| Egypt | Al Qahirah | | 3 | 8 |
| Egypt | Al Iskandariyah | | 3 | 12 |
| Egypt | Al Isma‘iliyah | | 3 | 15 |
| Egypt | Al Buhayrah | | 3 | 7 |
| Egypt | Al Gharbiyah | | 3 | 2 |
| Egypt | Kafr ash Shaykh | | 3 | 10 |
| Egypt | Dumyat | | 3 | 6 |
| Egypt | Bani Suwayf | | 3 | 9 |
| | | **Democratic Republic of the Congo (DRC)** | | |
| DRC | Bas-Congo | Bas-Fleuve | 1 | 10 |
| DRC | Bas-Congo | Boma | 2 | 7 |
| DRC | Bas-Congo | Matadi | 2 | 2 |
| DRC | Kivu | Sud-Kivu | 4 | 1 |
| DRC | Bas-Congo | Cataractes | 5 | 36 |
| DRC | Kasaï-Occidental | Lulua | 6 | 34 |
| DRC | Kivu | Nord-Kivu | 7 | 9 |
| DRC | Kasaï-Occidental | Kasaï | 8 | 21 |
| DRC | Équateur | Équateur | 9 | 16 |
| DRC | Orientale | Ituri | 10 | 15 |

| | | | | |
|---|---|---|---|---|
| **South Africa (SA)** | | | | |
| SA | Orange Free State | Bloemfontein | 1 | 115 |
| SA | Orange Free State | Botshabelo | 1 | 27 |
| SA | Orange Free State | Thaba'Nchu | 3 | 99 |
| SA | Orange Free State | Dewetsdorp | 4 | 283 |
| SA | Mpumalanga | Moutse | 5 | 65 |
| SA | Gauteng | Cullinan | 6 | 190 |
| SA | Mpumalanga | Moretele | 7 | 352 |
| SA | Mpumalanga | Mdutjana | 7 | 41 |
| SA | Mpumalanga | Mbibana | 9 | 114 |
| SA | Limpopo | Bochum | 10 | 349 |
| SA | Limpopo | Seshego | 10 | 77 |
| **Tanzania** | | | | |
| Tanzania | Kilimanjaro | Mwanga | 1 | 116 |
| Tanzania | Kilimanjaro | Same | 2 | 115 |
| Tanzania | Kagera | Bukoba Rural | 3 | 22 |
| Tanzania | Manyara | Simanjiro | 4 | 123 |
| Tanzania | Manyara | Karatu | 5 | 103 |
| Tanzania | Mtwara | Masasi | 6 | 72 |
| Tanzania | Manyara | Mbulu | 7 | 64 |
| Tanzania | Mwanza | Nyamagana | 8 | 10 |
| Tanzania | Mwanza | Lake Victoria | 8 | 40 |
| Tanzania | Mara | Lake Victoria | 10 | 94 |

*Notes:* Overall rank is based on counterfactual analysis described in Section 4 in the main text and the $\rho$s estimated in column (8) of Table 1. Overall key-player (KP) rank is based on the $\rho$s estimated in column (8) of Table 1 as well. The number of districts per country are: Ethiopia 72, Egypt 26 (ADM1 level), DRC 38, SA 354, and Tanzania 136.