Key Players in Economic Development*

Ashani Amarasinghe[†] Roland Hodler[‡]
Paul A. Raschky[§] Yves Zenou[¶]

December 2023

Abstract

This paper analyzes the role of networks in the spatial diffusion of local economic shocks in Africa. We show that geographic connectivity, along with road and ethnic connectivity, are important for diffusing economic spillovers over longer distances. We then determine the key players, that is, 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.

Keywords: Networks, spatial spillovers, key player centrality, natural resources, transportation, Africa.

JEL classification: O13, O55, R12.

*We thank the editor, two anonymous referees, Markus Brückner, John Gibson, 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.

†School of Economics, University of Sydney and SoDa Laboratories, Monash University; email: ashani.amarasinghe@sydney.edu.au.

‡Department of Economics, University of St.Gallen; CEPR, CESifo, and OxCarre; email: roland.hodler@unisg.ch.

§Department of Economics and SoDa Laboratories, Monash University; email: paul.raschky@monash.edu.

¶Department of Economics, Monash University, CEPR, and IZA; email: yves.zenou@monash.edu.

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 and Durlauf, 2001; Crespo Cuaresma, 2011).

This paper's main objective is to better understand the role of network's in propagating economic shocks. We highlight the roles of geographic, transport, and ethnic networks in the context of regional economic development and assume the three networks as given. To determine which districts are most central in propagating local economic shocks across space, we compute the key player centrality of each African district. The calculation of the key player centrality uses the position of each district within the networks and the estimated parameters of the spatially lagged indicator of economic activity. To identify these parameters we rely on plausibly exogenous variation in local mining wealth.[1]

For this purpose, we construct a balanced panel dataset of 5,944 African districts (ADM2, second subnational level) and yearly data from 1997–2020, in which our measure of local economic activity is nighttime light intensity. The basic econometric framework is a spatial Durbin model that allows for spatial autoregressive processes with the dependent and explanatory variables. 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. We interpret the estimated coefficient of the spatially lagged indicator of economic activity in this set up as the effect of a district's connectivity on its own economic activity.

[1]It is important to note that the focus of this paper is not on the spatial spillovers of mineral wealth shocks. See, e.g., Mamo et al. (2021) for a recent empirical analysis with this focus. Instead, we use local shocks to mineral wealth to identify the parameters in the Spatial Durbin Model that are required to calculate the key player centrality measure. We discuss the implications of using mining shocks as opposed to other local economic shocks in Section 3.2.

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 that builds on the work by Berman et al. (2017). In particular, we rely on cross-sectional variation in the neighboring districts' mining opportunities, and intertemporal fluctuations in the world price of the minerals extracted in these districts, as the sources of exogenous temporal variation in the districts' performance. Hence, we estimate the local average treatment effect (LATE) of economic shocks related to price-induced increases in natural resource rents. Our 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 geographic, ethnic, and road connectivity all foster the spatial diffusion of economic shocks and thereby increase local economic activity. However, they impact local economic activity in different ways.[2]

We then turn to measuring the network centrality of all the districts in Africa. The 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. While these districts are typically characterized by high local economic activity as well as good connectivity.

Finally, we conduct 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. These counterfactual policy exercises illustrate how our approach and its results could help policymakers to design more informed economic policies.[3]

[2]In Online Appendix A (titled "Theory"), we present a conceptual framework that derives the empirical specifications estimated in the paper. The main aim of the theoretical model is to explain how spatial spillovers between districts operate. They are mainly due to complementarities in interactions between districts. Many reasons could be behind such complementarities: for example, if a district has a high level of economic activity, then people from that district might migrate to the neighboring districts, thereby increasing the activities of neighboring districts. Unfortuntaley we are constrained from examining these in detail due to data limitations.

[3]Two important aspects need to be considered when interpreting the results from these policy analyses. First, we focus on the positive effects of these hypothetical interventions on overall economic activity, but are not able to account for local variation in costs. Conducting a more complete welfare analysis

Our paper relates to different strands of the literature. First, we contribute to the empirical literature on the effects of networks in economics. This literature has so far mostly focused on using micro data to test predictions from network theory.[4] A major challenge in these studies is the endogeneous sorting of agents in networks and the endogeneous formation of networks (e.g., Bifulco et al., 2011; Bramoullé et al., 2020). There are also some recent papers that study network effects from a macroeconomic perspective.[5] These papers study production or supply chain networks and document that the structure of the *production network* (input-output matrix) is key in determining whether and how microeconomic shocks – affecting only a particular firm or technology along the chain – propagate throughout the economy and shape macroeconomic outcomes (e.g., Acemoglu et al., 2016; Carvalho et al., 2021).

Our network analysis is not at the agent or firm level but at a more aggregate geographic location level, namely districts. We propose an explicit network analysis and determine the key players (districts) in each country, that is, the ones that need to be supported if the government wants to maximize total economic growth.[6] To the best of our knowledge, this is the first study that looks at the role of networks in explaining spatial economic spillovers across an entire continent.[7]

Second, we contribute to the literature on the economic effects of natural resource revenues.[8] To date, there has been little research on spatial spillovers of local resource extraction. A notable exception is Aragon's and Rud's (2013) study on the spillovers from a large gold mine in Peru. Closer to us, Mamo et al. (2019) look at spatial spillovers from

would go beyond the scope of this paper. Second, the usual caveats of using LATE parameters for a counterfactual policy analysis apply.

[4]For recent overviews, see Bramoullé et al. (2016) and Jackson et al. (2017).

[5]For a recent overview, see Carvalho and Tahbaz-Salehi (2019).

[6]There are very few papers that empirically determine key players in networks. Exceptions include König et al. (2017), Lindquist and Zenou (2014) and Lee et al. (2021). However, in these studies, the key players are determined at the individual level, which creates network formation issues that are difficult to address. In our paper, this is not the case because the links between districts are pre-determined by their locations.

[7]The key-player centrality assumes that the network does not change as a consequence of the removal. Indeed, the original definition of the key-player centrality in **?** is such that, when a node is removed, the network does not change. In our spatial context, it makes even more sense. If we hypothetically remove a district (of course this does not happen in reality), then it seems reasonable that the spatial network does not change, at least in the short run, that is, new roads are not constructed or the geography/topology of the districts does not change.

[8]For an overview, see van der Ploeg (2011).

mine discoveries in Africa. They find little evidence for spatial spillovers. Our analysis differs from theirs by focusing on resource price fluctuations rather than less frequent mine discoveries, by using an IV strategy, by using ethnic and road networks in addition to geographic networks, and by letting a network-theoretic framework guide our empirical approach, which allows us to determine key districts.

Third, we contribute to the literature on the effects of ethnic diversity on aggregate outcomes. Traditionally, this literature has investigated the relationship between ethnic diversity and aggregated economic outcomes (e.g., Alesina and La Ferrara, 2005) while more recent studies use subnational information to explain differences in economic outcomes across ethnic homelands (e.g., Michalopoulos and Papaioannou, 2013; De Luca et al., 2018). Our contribution to this literature is to show how economic shocks propagate through the ethnic network and thereby affect economic development outside subnational districts. The pattern of wider propagation within ethnic groups than across homelands may contribute to ethnic inequality (Alesina et al., 2016).

Fourth, like us, Storeygard (2016), Bonfatti and Poelhekke (2017), and Jedwab and Storeygard (2022) focus on road networks in Africa. In line with other recent papers on the effects of roads on local economic development (e.g., Asher and Novosad, 2020), these studies are mainly concerned with the effect road connectivity on the economic development in the spatial unit itself.[9] Our paper complements this literature 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) which is of importance, in particular, for the road and geographic networks. However, 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.

Finally, we relate to the vast literature on quantitative spatial or regional economics, which develops general equilibrium models that analyze the distribution of economic devel-

[9]This is not to say that these studies have ignored spatial spillovers. For example, Asher and Novosad (2020) look at local employment spillover effects between Indian villages within a 5km radius.

opment across space.[10] Typically, this literature explains why economic activity is higher in one district compared to another one due to exogenous local factors (e.g., geography) and endogenous economic mechanisms (e.g., trade). In our study, we are relatively agnostic about the particular type of spatial interdependence (e.g., migration, trade, knowledge spillovers) between districts. However, by using tools from network theory, we are able to show how a district's level of connectivity influences its ability to diffuse economic shocks through the network and to determine the key districts in a given country.

2 Data

Our units of observation are administrative units at the second subnational level (ADM2), which we call districts.[11] The final dataset consists of yearly observations for 5,944 districts from 53 African countries over the period 1997–2020.[12] The average (median) size of a district is approximately 39km^2 (6km^2) and the average (median) population is around 105,000 (38,000).

For our purpose, there are a number of advantages of using administrative units rather than grid cells (e.g., Berman et al., 2017, Henderson et al., 2018) or ethnic homelands (Michalopoulos and Papaioannou, 2013, 2014). First and foremost, policymakers operating within a country's administrative framework typically take their project allocation decisions based on administrative units. Second, changes in economic activity in the African hinterland, e.g., agricultural production or mining, may often be reflected by changes in nighttime lights in a district's urban hub. We thus want to ensure that a district's urban hub belongs to the same spatial units as its hinterland.

[10] For a recent overview, see Redding and Rossi-Hansberg (2017).

[11] 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. Figure B1 in the Online Appendix shows the boundaries of each district in Africa.

[12] Table B1 in the Online Appendix lists all the African countries in our sample and provides the number of districts per country.

2.1 Dependent variable: Nighttime lights

Like many contributions inspired by Chen and Nordhaus (2011) and Henderson et al. (2012), we use satellite data on the intensity of nighttime lights as a proxy for economic activity. More specifically, we use the harmonized global nighttime light dataset by Li et al. (2020), who combine the nighttime light data from the Defense Meteorological Satellite Program (DMSP)/Operational Linescan System (OLS), which is available for the years 1992–2013, and the more precise data from the Visible Infrared Imaging Radiometer Suite (VIIRS), which is available since 2012. Like the DMSP/OLS data, the harmonized global nighttime light data are also available for pixels of less than one square kilometer and on a scale from 0 to 63, with higher values implying more intense nighttime lights.

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.

Henderson et al. (2012) and Hodler and Raschky (2014) document 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 in the Online Appendix). 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.[13]

A number of recent studies highlight that nighttime lights are an imperfect proxy for economic activity and local development. For example, Gibson et al. (2021) show that the correlation between DMSP nighttime light data and official income data from Indonesia is the weaker the smaller the geographical unit is. This is particularly a concern if one

[13]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.

uses nighttime light data in a cross-sectional setting to explain differences between small subnational units. Moreover, there is an ongoing debate about what economic variable is ultimately captured by nighttime lights. For example, Asher et al. (2021) find that granular nighttime light data are strongly correlated with population, employment or per capita consumption at a very local level in India. In addition, nighttime lights have also been used to proxy for electrification (e.g., Baskaran et al., 2015; Asher et al., 2021).

Despite these weaknesses, there does not exist any other proxy for regional economic activity that is measured in a consistent and comparable way across the entire African continent and available annually for a period of 20 or more years. We acknowledge it is hard to know whether the economic spillovers we measure are more closely proxying for population, employment or per capita consumption. However, importantly, we can rule out that they are simply reflecting electrification of larger areas due to mining booms. We conduct a robustness check where we use geo-referenced data about the extent of the electricity grid and show that our results are not driven by access to the electricity grid (see Table E12 in the Online Appendix). Finally, the use of district fixed effects means that small or sparsely populated districts with zero nighttime lights throughout the sample period cannot drive our results.

2.2 Connectivity matrices

We construct three connectivity matrices to measure spatial spillovers.

2.2.1 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 $\tilde{d}_{i,j} = 1/d_{i,j}(1 + e_{i,j})$.

As a simpler alternative we also construct a weighting matrix for geographic connectivity based on contiguity.[14] Defining geographic connectivity using the inverse altitude-adjusted distance as opposed to contiguity proves advantageous on three accounts. 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 distance cutoff of 100 km.That is, we set the spatial weight as $\omega_{i,j} = 1/\tilde{d}_{i,j}$ if the geodesic distance $d_{i,j}$ is less than 100 km, and $\omega_{i,j}$=0 otherwise. The focus on a distance cutoff of 100 km is based on the previous literature. In particular, Aragon and Rud (2013) provide a careful study of the economic spillovers of mining. They find that the spillovers decline with distance and become insignificant beyond 100 km. Others find that spatial spillovers become close to zero already a lower distances (e.g., Kotsadam and Tolonen, 2016, at 50 km), but we prefer to err on the side of a distance cutoff that is too high rather than too low in our main specification.

2.2.2 Ethnic connectivity

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 (1959) documents the spatial distribution of ethnic homelands in Africa and subdivides the continent into over 800 ethnic homelands.[15]

[14]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.

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

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. Comparing the maps of district boundaries in Figure B1 with the boundaries of ethnic homelands in Figure B2 shows that, for many areas, multiple, adjacent districts share the same ethnicity. Therefore, any spillover effects between contiguous ethnic regions are captured by the geographic connectivity matrix and the estimated coefficient of the spatial lag using the ethnic connectivity matrix only captures the effects of more distant districts that share the same ethnicity.

2.2.3 Major Road connectivity

Roads are, arguably, a key form of connectivity between districts. Roads enable non-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).[16] We accessed the OSM data in early 2016 and extracted information about *major roads* for the African continent.[17] There is no consistent definition of major roads across African countries within the OSM data. As such, we include any road that is defined as “highway”, “motorway”, or “primary road” in OSM. It is important to note that the

[16]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.

[17]Figure B3 in the Online Appendix shows the road network.

type and quality of these major road types can largely differ between and within African countries. Using aerial photography of some of these major roads that connect the district of Ambaca in Angola to other districts (see Figure B4. in the Online Appendix). we show that a "highway" can mean the road to be paved but also an unpaved "dirty road". Unfortunately, geo-referenced data about road quality or road material are not available for the entire African continent. However, for our application, information about the road material or quality is not required as we rely on OSM's definition of major roads, which ultimately reflects the local definition of major roads as the data is sourced from local, publicly available data. The aerial photos also show connectivity of other, minor roads or paths, which are not included in this connectivity matrix but are captured by the geographic connectivity matrix.

For the purpose of this study, we use the OSM major road polylines and intersect them with the district boundary polygons and generate a network graph of the road network.[18] 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).[19] 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 and again use 100 km in the main specification.

Table 1 shows the correlation structure of the three connectivity matrices. We observe

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

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

that the correlation between the geographic and the major road network measure is 0.61 which is, relatively high. It is also important to note that our road connecting matrix is based on *major* road connections and not *all* road connections. There is a high likelihood that each district has some form of basic road connection with other districts. However, the road connection matrix used in our analysis uses major roads and no t every district is connected through a major (i.e. highway) road connection and therefore the correlation between the geographic and the road connection is not (close to) one.

Table 1: Correlation Between Connectivity Matrices

	(1) $Eth\ W\ Light_{jt}$	(2) $Inv\ Dist\ W\ Light_{jt}$	(3) $Inv\ Road\ W\ Light_{jt}$
$Eth\ W\ Light_{jt}$	1.000		
	(1.000)		
$Inv\ Dist\ W\ Light_{jt}$	0.299	1.000	
	(0.219)	(1.000)	
$Inv\ Road\ W\ Light_{jt}$	0.212	0.614	1.000
	(0.184)	(0.512)	(1.000)

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

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 test, we construct new connectivity matrices that take into account national borders (see Section E.10 in the Online Appendix).

2.3 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.[20]

[20] An interesting alternative would be to use georeferenced information on geological data about mineral deposits instead of actual mining operations. However, to our knowledge there is no complete, publicly available dataset that contains georeferenced information about geological suitability or deposits covering all minerals for ADM2 regions in Africa available. The only data that we are aware of is for Gold, which is referenced in Girard et al.(2022). However, this is only for one mineral and would further decrease the number of potentially treated districts and further amplify concerns regarding LATE. In addition, there is a concern that using suitability instead of actual operations might create a weak instrument problem. Without more detailed information about the status of the exploration (which can be endogenous) it is unclear how a positive price shock for a mineral deposit (with existing extraction capacity) might increase economic activity in the district in the, relative, short run.

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 over 500 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.[21] 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 time-invariant 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.[22]

Data on world prices of minerals are sourced from the World Bank and USGS (see Table B2 in the Online Appendix for more information on the data sources). $Price_t^r$ is the logarithm of the yearly nominal price of resource r in USD.

2.4 Control variables

Our main time-varying control variable at the district level is $Population_{it}$. It measures a district's total population (in logs) and is derived based on the Gridded Population of the World (GPW) v4 from the Center for International Earth Science Information Network (CIESIN, 2016).

We further control for conflict events using geo-referenced data from the PRIO/Uppsala Armed Conflict Location and Event Database (ACLED). 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

[21]Figure B5 in the Online Appendix shows the spatial distribution of mining projects across Africa.

[22]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. Therefore, our time-invariant indicator of mining assumes a value of 1 for the whole sample period, if an active mine was located in the district in at least one year of the sample period, and 0 otherwise.

value of one if any conflict-related event occurred in district i in year t, and zero otherwise.

Table 2 provides the descriptive statistics for our key variables.

Table 2: Descriptive statistics

Variable	Observations	Mean	Std. Dev.	Min.	Max.
$Light_{it}$	142,656	-0.430	2.719	-4.605	4.143
$Conflict_{it}$	142,656	0.185	0.388	0	1
$Population_{it}$	142,656	10.788	2.148	-4.605	16.286
MP_{it}	142,656	0.213	1.102	0	10.525

3 Empirical Analysis

The aim of the empirical analysis is to estimate the following equation:[23]

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

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; ρ_k^m are the coefficients of the spatial lags; α_i and CT_{ct} are district and country-year fixed effects, respectively; and ϵ_{it} is an error term that is assumed to be $\epsilon_{it} \sim N(0, \sigma^2 I_n)$.[24] We row-normalize the adjacency matrices such that the sum of each row is equal to 1.

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

[23] In Online Appendix A, we propose a simple model that microfounds equation (1).

[24] Table D2 in the Online Appendix presents estimation results with step-wise addition of control variables.

other unobserved shocks. Thus, estimating equation (1) using OLS can yield biased and inconsistent estimates.

To address this problem, we estimate a 2SLS model that exploits exogenous variation in the economic value of mineral resources in the mining districts.The idea is 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 use 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. Note that the component $Mine_j$ is time-invariant, $Mine_j \in \{0, 1\}$, and is one if the mine is active at least for one year during the sample period. In combination with district fixed effects, using a time-invariant mining activity variable, addresses the potential problem of endogenous opening and closing of mines due to economic activity or other confounders, as well as the concern that more developed or better connected districts might be more or less suitable for mineral extraction.[25]

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

[25]In addition, one can also consider this variable as a proxy for a district's suitability for mining and occurrence of mineral deposits.

networks. Therefore, we apply a novel estimator by Colella et al. (2018) that allows us to account for dependence across our observations' error terms in a more flexible form. In practice, we correct for clustering at the network level where observations were assumed to belong to the same cluster once they are linked through at least one of the three networks (geography, ethnicity, roads).[26]

3.1 Validity and Relevance of the Instrumental Variable

In order for our instrumental variable 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 (see, also, Brunnschweiler and Bulte, 2008) while works by Aragon and Rud (2013), Caselli and Michaels (2013) or von der Goltz and Prabhat (2019) have shown that resource extraction can have positive effects on local economic development.

Minerals 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 take a number of measures to mitigate the risk of the exclusion restriction being violated. First, we include 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, Berman et al. (2017) show that mining activity could increase conflict. This, in turn, could adversely affect economic activity. Therefore, we control for district-level

[26] We implement this procedure in Stata 18 using the "acreg" command by Colella et al. (2018). Moreover, Table E3 in the Online Appendix presents the results using the standard Conley-type spatial clustering approach (Conley, 1999).

conflict events in our specifications. Third, the exclusion restriction also relies on the assumption of exogeneity of world mineral prices, requiring that no single district can affect the world market price of a commodity. For this reason, we conduct a robustness check of our specifications by excluding countries in the top ten list of producers for any mineral (see Table E5 in the Online Appendix). Fourth, it is possible that mining operations are spatially correlated and that the neighbors of a mining district are also more likely to be mining districts. While any time-invariant unobservables are captured by the district fixed effects, we explicitly control for any direct, time-variant, mining shocks by including MP_{it} in our vector of control variables $\mathbf{X}_{it}$ in equation (2).

Fifth, it is possible that an increase in the resource wealth in one district can have a direct effect across districts through regional or national public expenditures. Therefore, our instrumental variable, MP_{jt}, can have a direct effect on the outcome variable and thereby violating the exclusion restriction. We address this potential concern in two ways: in Table E1, we present results that include province-year fixed effects. Province refers to ADM1 regions and this large set of fixed effects, capture all shocks in a province and year, including, increased public expenditure due to a mining boom in some ADM2 districts of the ADM1 province. In addition, given that a lot of mining operations are geographically clustered, we explicitly control for any direct, time-variant, mining shocks by including MP_{it}, which capture for example increased investments in exploration activity in neighbouring districts suitable for mining.

Sixth, recent contributions have found spatial spillover effects of mines on labor-market outcomes (Kotsadam and Tolonen, 2016) and child mortality (Benshaul-Tolonen, 2019), which could violate the exclusion restriction. However, the spatial reach of these effects is very narrow and unlikely to cross the district borders. For example, Benshaul-Tolonen (2019) does not find a statistically significant effect of mining activity on infant mortality beyond 10km while Kotsadam and Tolonen (2016) find that the labor-market effect of mine openings are close to zero after 50km from the mine. In addition, any mining related labor-market effects on neighboring districts are more likely to capture spatial spillover effects as opposed to direct effects. Indeed, it first requires a mine to increase activity in its own districts (both by increasing own production as well as a growth in service facilities in the immediate proximity of the mine). The spillover effect is then through workers who travel from neighbouring districts, work at the mine, and either send part of the salary back home or spend it at a later point in their home district.

3.2 Mining Shocks and Spatial Diffusion of Economic Activity

Our setting implies that we estimate a local average treatment effect (LATE) of economic shocks related to price-induced increases in natural resource rents on the performance in districts with a certain network proximity to mining districts. These shocks and the resulting spatial spillovers may have different effects on consumption, investment, and government expenditure than other economic shocks, and districts close to mining districts may differ from the average African district. We thus need to be careful when drawing more general policy conclusions based on the estimated spatial spillover effects (similar to all the other contributions estimating a LATE). However, it is worth emphasizing that the LATE that we estimate is of first-order importance in itself. The mining sector has long been a large and important sector in many African countries. Moreover, it will become even more important as a result of the need to rapidly decarbonize the global economy and to transition towards renewable sources of energy. This green energy transition raises the global demand for certain minerals (Hund et al., 2020, Herrington, 2021), and this demand is increasingly met by mining activities in the Global South, including many African countries (Reichl and Schatz, 2022).

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.

It is well documented that the economic benefits from mining activity often come with large negative externalities, by increasing the likelihood of violent conflict (e.g. Berman et al., 2017), polluting the natural environment (e.g., Aragon and Rudd, 2016), or negative health impacts (e.g., von der Goltz and Prabhat, 2019). While we directly control for potential mining related spatial diffusion of conflict in our vector of control variables X, we observe that mining related pollution and health effects are strongest in the close proximity (within the district) of a mine.

We purposefully exclude oil and gas extraction from our analysis because of the problem of gas flaring. Gas flaring refers to the combustion of associated gas that is a byproduct of oil and gas production and associated refining processes. Gas flares are active throughout the year and are thereby recorded as a stable nightlight source. Due to their brightness and spatial extent, one could misinterpret gas flares as small cities. Elvidge et al. (2009) have documented this phenomena and provide shapefile with the extent of the areas impacted by gas flares. It is common practice in the literature to remove these gas-flare affected areas from the spatial boundary file prior to the calculation of the local nighttime light values. As such, it is not possible to measure any world price related expansion or contraction in oil and gas production, and local economic activity in those districts.

Besides mining, agriculture is another important sector in many African countries. A number of recent empirical papers have exploited spatial and temporal variations in climate (e.g., Harari and La Ferrara, 2018; Berman et al., 2021) or soil productivity (e.g., Couttenier and Soubeyran, 2014) as shocks to the local agricultural sector. We prefer to exploit mining shocks rather than agricultural shocks because crop locations and soil composition are more spread out in space. This would make it relatively more difficult to justify the exclusion restriction in a spatial model that aims at identifying spillover effects between relatively small spatial units.

3.3 Estimation results

Table 3 presents our estimates from equations (1) and (2).[27] 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 ρ_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, γ, is positive and statistically significant, indicating that our instrumental variable is a strong predictor of economic activity in district j. The coefficient of interest of the *second stage*, ρ_1, is also positive and statistically significant.

Second, in columns (3) and (4), we focus on geographic connectivity using the weighting matrix based on the inverse of the altitude-adjusted geodesic distance between districts i and j, truncating the matrix at 100 km. The coefficient of interest, ρ_2, is positive and statistically significant at the 1% level in both the OLS and the IV estimations. In Figure D1 in the Online Appendix, we show that the coefficient of interest from the IV estimates is positive, fairly constant in the cutoff distance and statistically significant for up to 140 km, but increases for higher cutoff distances. Hence, our results do not strongly depend on the exact distance cutoff used in the analysis.

Third, in columns (5) and (6) of Table 3, we focus on *road connectivity* based on our matrix of inverse road distances, again, truncating the matrix at 100 km. The coefficient of interest, ρ_3, is positive and statistically significant at the 1% level in both the OLS and the IV estimations. In Figure D1 in the Online Appendix, we show that the coefficient of interest from the IV estimates is positive, fairly stable, and statistically significant for cutoff distances from 60–500 km.

Finally, the last two columns of Table 3 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).

[27] Table 3 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.

Table 3: 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: $Light_{it}$							
$Ethnicity\ W\ Light_{jt}$	0.547***	0.737***					0.160***	0.310**
	(0.009)	(0.142)					(0.009)	(0.139)
$Inv\ Dist\ W\ Light_{jt}$			0.672***	1.287***			0.432***	0.958***
			(0.010)	(0.179)			(0.013)	(0.153)
$Inv\ Road\ W\ Light_{jt}$					0.502***	0.935***	0.269***	0.327**
					(0.009)	(0.206)	(0.011)	(0.137)
First stage:	Dependent variable: $Light_{jt}$							
MP_{jt}		0.057***		0.065***		0.058***		0.078***
		(0.017)		(0.016)		(0.017)		(0.017)
First-stage F-stat		11.17		16.30		12.36		20.44
Observations	142,656	142,656	142,656	142,656	142,656	142,656	142,656	142,656
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 100 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.

We observe that the three coefficients of interest, i.e., ρ_1, ρ_2, and ρ_3, are 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 ethnic and road connectivity are only statistically significant at the 5% level in column (8). The coefficient estimates further suggest that geographic connectivity tends to be more important than ethnic and road connectivity.

To better interpret the magnitude of the coefficients in this specification, we provide a simple example with three districts, labeled 1, 2 and 3. First, let us consider a model with only one adjacency matrix $\mathbf{\Omega}$. Denote the initial matrix by $\widetilde{\mathbf{\Omega}}$ and the row-normalized one by $\mathbf{\Omega}$. Assume that the network is *complete*. We have:

$$\widetilde{\mathbf{\Omega}} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \text{ so that } \mathbf{\Omega} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

Assume $\rho = 0.737$ (which corresponds to the estimated ρ_1 for the ethnic network in column (2) of Table 3) so that $\rho = 737 < \mu_1(\mathbf{\Omega}) = 1$. Then, given the districts' observable and unobservable characteristics, a 10% increase of the nighttime lights in district $i = 1, 2, 3$ increases the nighttime lights in each of the other two districts by 3.685%.[28]

Consider, now, the case of three adjacency matrices $\mathbf{\Omega}_1$, $\mathbf{\Omega}_2$ and $\mathbf{\Omega}_3$ and assume that the networks are different so that each row-normalized network is equal to:

$$\mathbf{\Omega}_1 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \mathbf{\Omega}_2 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{\Omega}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{4}$$

$\mathbf{\Omega}_1$ represents an *ethnicity network* in which all districts share the same ethnicity; $\mathbf{\Omega}_2$ is a *geographic network* in which the geodesic distance between district 1 and each of the other districts is within 100km, whereas the distance between districts 2 and 3 exceeds

[28]Indeed, we have $l_1 = (0.737)\left(\frac{l_2+l_3}{2}\right) + X_1 + \varepsilon_1 = 0.3685\, l_2 + 0.3685\, l_3 + X_1 + \varepsilon_1$. Thus an increase of 10% of the nighttime light of district 2 (or district 3) will increase the nighttime light of district 1 by 3.685%. The same is true for all the other districts since the network is complete.

100km; $\mathbf{\Omega}_3$ is a *road network* with a single road between districts 1 and 2.[29] Using the estimates from column (7) of Table 3, assume $\rho_1 = 0.160$, $\rho_2 = 0.432$ and $\rho_3 = 0.269$. It is easily verified that

$$\rho_1\mathbf{\Omega}_1 + \rho_2\mathbf{\Omega}_2 + \rho_3\mathbf{\Omega}_3 = \begin{pmatrix} 0 & 0.565 & 0.296 \\ 0.781 & 0 & 0.08 \\ 0.512 & 0.08 & 0 \end{pmatrix}$$

so that

$$\mu_1\left(\rho_1\mathbf{\Omega}_1 + \rho_2\mathbf{\Omega}_2 + \rho_3\mathbf{\Omega}_3\right) = 0.807 < 1.$$

Let us now interpret equation (5) in Section A of the Online Appendix. For district 1, we have:

$$l_1 = (0.160 + 0.432)\left(\frac{l_2 + l_3}{2}\right) + 0.269\, l_2 + X_1 + \varepsilon_1 = 0.565\, l_2 + 0.296\, l_3 + X_1 + \varepsilon_1$$

Similarly, for the two other districts, we obtain:

$$\begin{aligned} l_2 &= 0.160\left(\frac{l_1 + l_3}{2}\right) + (0.432 + 0.269)\, l_1 + X_2 + \varepsilon_2 = 0.781\, l_1 + 0.08\, l_3 + X_2 + \varepsilon_2 \\ l_3 &= 0.160\left(\frac{l_1 + l_2}{2}\right) + 0.432\, l_1 + X_3 + \varepsilon_3 = 0.512\, l_1 + 0.08\, l_2 + X_3 + \varepsilon_3 \end{aligned}$$

We can now interpret the different coefficients as follows. Given the observable and unobservable characteristics of each district, a 10% increase of the nighttime lights of the other districts increases the nighttime lights of districts 1, 2 and 3 by 8.61%, 8.61% and 5.92%, respectively. This is because district 3 is *less connected* than districts 1 and 2 and the spillover effects are relatively similar between the different matrices.

[29] This is a simple theoretical example. In the data, the adjacency matrices $\mathbf{\Omega}_2$ and $\mathbf{\Omega}_3$ are weighted by the inverse distance between the two districts.

3.4 Robustness checks

We now discuss a number of robustness checks where the Online Appendix contains the corresponding tables.

A first set of checks show that our results are robust to small changes in the empirical specification. Table D2 provides estimates when incorporating variables gradually. Table E1 replaces the country-year fixed effects with province-year fixed effects, thereby controlling for province-specific economic and political variation over time. 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 0.5×0.5 degrees (i.e., around 55 $\times$ 55km at the equator) instead of ADM2 regions. Results suggest a less important role of road connectivity for the spatial economic spillovers, maybe because roads are more likely to connect district capitals than grid cells.

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 coefficient on ethnic connectivity becomes statistically insignificant. 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 geographic and road networks remain quantitatively similar when controlling for fiscal spillovers. The importance of the ethnic network declines within this set of estimates, potentially because, given the clustered nature of ethnic homelands, the ADM1 network already accounts for ethnic connectivity.

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. 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 an increase in the importance of the ethnic network.

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 and the other only connected districts in other countries. The results in Tables E10 and E11 reveal that geographic and road connectivity is the primary source of within country spillovers, while ethnic and geographic connectivity is more important for between country spillovers.

Lastly, in Table E12, we consider whether the spillovers of economic activity are driven by connectivity to the electricity grid. Although electricity is a key source of nighttime light, its importance is not so apparent in Africa where large parts of the continent emit nighttime light despite being excluded from the electricity grid. Nevertheless, to identify the role played by the electricity network in generating spatial spillovers of economic activity, we construct an "electricity" network that identifies all districts connected to the electricity grid. In columns (1) and (2), we consider the effect of the spatial lag of this new connectivity network in isolation. In columns (3) and (4), we include the spatial lag of the electricity network together with the connectivity networks in our baseline specification, and we find that our IV estimates remain strong. Finally, in columns (5) and (6), we exclude all districts connected to the electricity grid from our sample and re-estimate the baseline specification. These results therefore provide evidence that economic activity in Africa, as measured by nighttime light, are not merely driven by the electricity network, and that alternative sources of nighttime light play an important role in generating economic activity.

4 Most central districts

We now use our theoretical model, our connectivity matrices and our estimates of the spillover effects to calculate various centrality measures and determine the districts that play a key role in African economies due to their connectivity.

4.1 Theory: Different definitions of node centralities

There are different centrality measures (see Jackson (2008) 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.

4.1.1 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:

$$C_i^E(\boldsymbol{\omega}) = \frac{1}{\mu_1(\boldsymbol{\Omega})} \sum_{j=1}^{n} g_{ij} C_j^E(\boldsymbol{\omega}) \tag{5}$$

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:

$$\mu_1(\boldsymbol{\Omega})\, \mathbf{C}^E(\boldsymbol{\omega}) = \boldsymbol{\Omega}\mathbf{C}^{\mathbf{E}}(\boldsymbol{\omega}) \tag{6}$$

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.

4.1.2 Katz-Bonacich centrality

In our theoretical model (Section A on the Online Appendix), 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 equation (4) in Section A of the Online Appendix, 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 ρ. We will use the estimated value of ρ (IV estimates). We also need to check that the condition $\rho\mu_1(\mathbf{\Omega}) < 1$ is satisfied.

4.1.3 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.[30] For that, consider the game with strategic complements developed in the theory section (Section A of the Online Appendix) for which the utility is given by equation (3) in Section A of the Online Appendix, 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 equation (1) or (4) in Section A of the Online Appendix. 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:

$$\max\{L^*(\boldsymbol{\omega}) - L^*(\boldsymbol{\omega}^{[-i]}) \mid i = 1, ..., n\} \tag{7}$$

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

$$C_{i,u_i}^{KP}(\rho, \boldsymbol{\omega}) = \frac{C_{i,u_i}^{BO}(\rho, \boldsymbol{\omega}) \sum_j m_{ji}(\rho, \boldsymbol{\omega})}{m_{ii}(\rho, \boldsymbol{\omega})} \tag{8}$$

where $C_{i,u_i}^{BO}(\rho, \boldsymbol{\omega})$ is the weighted Katz-Bonacich centrality of district i (see equation (4) in Section A of the Online Appendix) and $m_{ij}(\rho, \boldsymbol{\omega})$ is the (i, j) cell of the matrix $\mathbf{M}(\rho, \boldsymbol{\omega})= (\mathbf{I} - \rho\mathbf{\Omega})^{-1}$. Ballester et al. (2006, 2010) have shown that the district i^* that solves (7) 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 (8) 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

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

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 $i^* = \arg\max_i C^{KP}_{i,u_i}(\rho, \boldsymbol{\omega})$, where[31]

$$C^{KP}_{i^*,u_i}(\rho, \boldsymbol{\omega}) = L^*(\boldsymbol{\omega}) - L^*\left(\boldsymbol{\omega}^{[-i]}\right). \tag{9}$$

4.2 Empirical Results

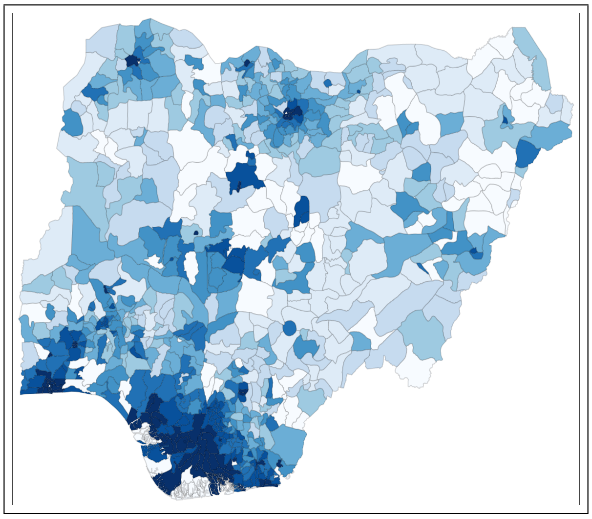
We compute these four different centrality measures for all 5,944 districts from the 53 African countries in our sample. In our discussion here, however, we will mainly focus on the key player ranking based on the geographic network, the road network, *and* the ethnicity network for two large countries that feature prominently in the literature: Nigeria and Kenya. 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).[32]

The ten Nigerian districts with the highest key player centralities are all parts of metropolitan areas: seven are part of the Lagos metropolitan area, which is the primate city of Nigeria and its economic hub; two are part of the Kano metropolitan area, which is the second largest city in Nigeria and the economic hub of the country's north; and one is part of the Port Harcourt metropolitan area, which is the fifth largest city and the main hub in the oil-producing Niger Delta. The average intercentrality value in our calculations is 4.4 units, while the intercentrality value for the top district – Ikeja from the Lagos metropolitan area – is 80.4, suggesting the loss from removing Ikeja would be around 18 times higher than the loss from removing an average district.

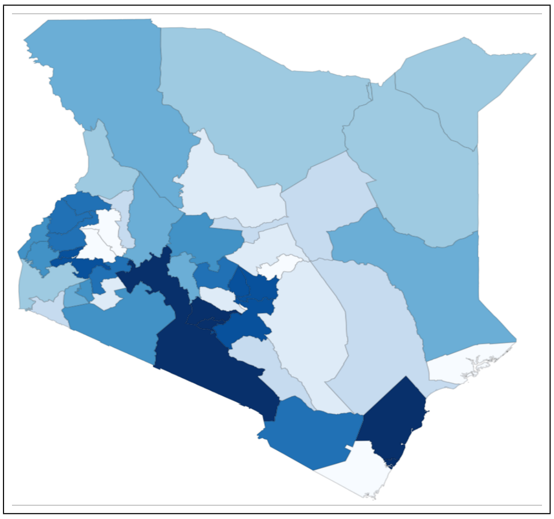
[31]Ballester et al. (2006) define the key player in (8) only when the adjacency matrix $\boldsymbol{\Omega}$ is not row-normalized. Since we use row-normalized adjacency matrices when estimating the ρs, we will determine the key player numerically based on its definition in (9).

[32]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 and Kenya.

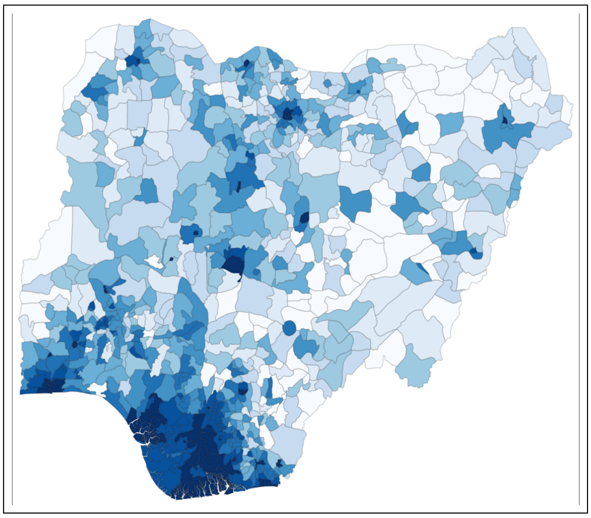
Figure 1: key player Centrality, Nighttime Light Intensity, and Population Density in Nigeria and Kenya



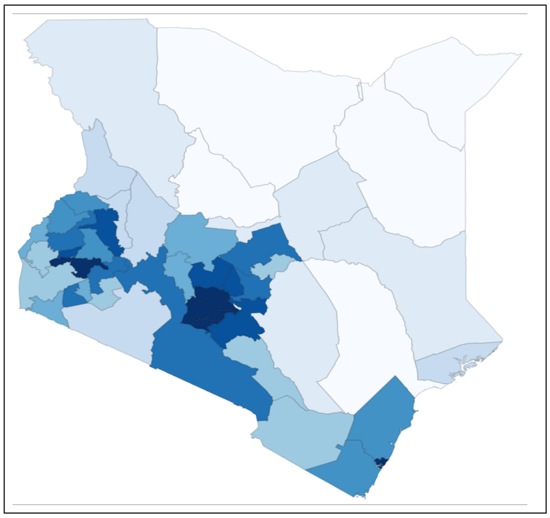
Key player centrality across Nigeria



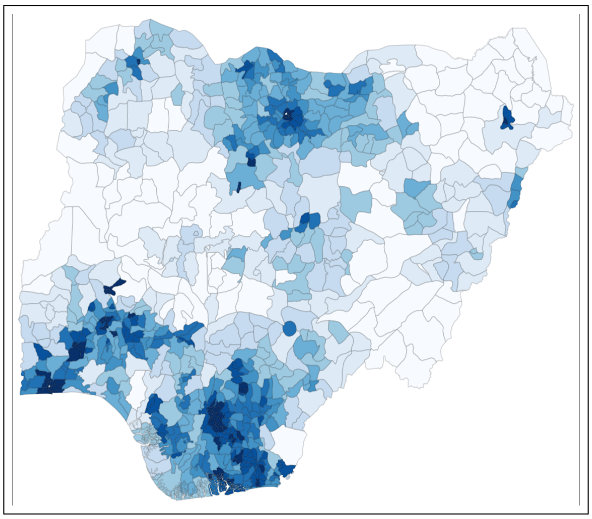
Key player centrality across Kenya



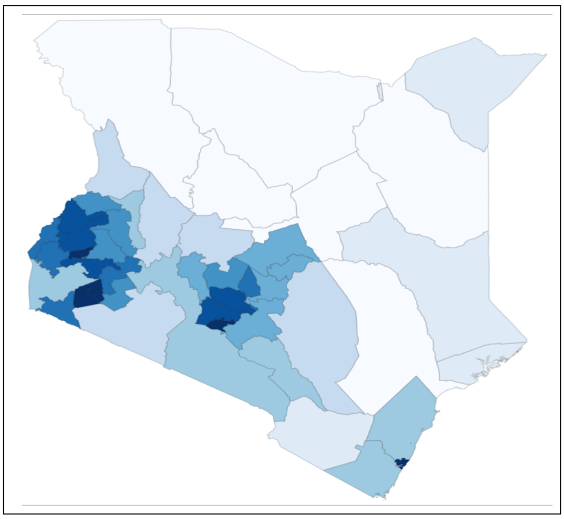
Average nighttime lights across Nigeria



Average nighttime lights across Kenya



Population density across Nigeria



Population density across Kenya

Notes: Darker colors indicate higher values.

The key district in Kenya is Nairobi (with an intercentrality value of 34.7), which is the capital and the primate city. Mombassa, which is Kenya's second largest city and home to Kenya's largest seaport, has the third highest key player centrality. These and other Kenyan districts with high key player centrality all tend to have relatively intense nighttime lights. The key districts encompassed or is part of the primate city in many other African countries as well, including Ethiopia (Addis Ababa) and South Africa (Johannesburg), but not in Egypt (Suhaj) and the Democratic Republic of the Congo (Kivu). 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), and Storeygard (2016) among others.

However, our findings go beyond these well-established results. Out of the 5,944 African districts, only very few are home to cities that qualify as primate cities (mainly capital and port cities). It is also not the case that the most developed districts are automatically the most central ones. For example, Killifi (Kenya) is ranked 2nd in key player centrality, but is ranked 20 out of 48 districts in terms on nighttime light. West Gojan (Ethiopia) is ranked 7 in key player centrality but is 24 out of 72 districts in terms of nighttime light. In the Democratic Republic of Congo, Sud-Kivu is ranked 1 in key player ranking but 18 out of 38 districts in terms of nighttime light. Similarly, in the same country, Tshilenge is ranked 9 in key player but is ranked 28 out of 38 in terms of nighttime light. As such, our analysis can also shed light on a large number of other districts in Africa, outside of the relatively small group of primate cities, which are important localities for diffusing the benefits of positive economic shocks to other districts. In other words, our findings can provide important information to policy-makers about areas that have been so far neglected but where development projects and public investments could potentially generate large economic spillover effects.

5 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, policymakers may be interested in the benefits from either promoting local economic activity or improving the network structure beyond those already well connected, e.g., by building new 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.

A few comments are in order. 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.[33] Fourth, these (and most other) policy experiments are based on estimates derived using the existing connectivity networks. If these policies were implemented in reality, there is the possibility that the network dynamics would change over time. Therefore, our counterfactual policy experiments are most informative about short- to medium-run effects rather than long-run effects.

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 proceed as follows. First, we add the value of 10 to the average nighttime light pixel value in the treatment district, which corresponds to an increase of one standard deviation.[34] Second, we take the logarithm of the now higher

[33]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.

[34]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 increase the average nighttime light pixel

average nighttime light pixel value and recalculate the spatially lagged dependent variables with the new values, while keeping the estimated ρ's from Table 3 (column (8)). Third, we recalculate the predicted nighttime lights (in logs) for each African district and compute the sum across all districts. Fourth, we compare this sum, which includes 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 repeat this exercise for each of the 5,944 districts.[35] The maps in Panel (a) of Figure 2 show the results for Nigeria (left panel) and Kenya (right panel), with darker colors implying a stronger overall impact of the counterfactual increase in economic activity.[36]

There are various types of districts where the overall impact of this investment is particularly high. First, for both Nigeria and Kenya, the overall impact is high in some districts with high key player centrality, which are economically active and well-connected. Second, in Nigeria, the top districts include districts in Bayelsa and Rivers, which are both oil-producing states in the Niger Delta. These districts are economically quite active and well-connected, but 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 top districts include poor districts that rank 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 differences in the overall impact of investing in the different districts are large. The expected overall impact from increasing the average nighttime light pixel value of a randomly chosen Nigerian (Kenyan) district by 10 is 23.2 (15.5). In contrast, the overall

value of these districts by 10 as well. Approximately 9% of the districts in our sample experienced an increase in nighttime lights of more than 10 units over the sample period.

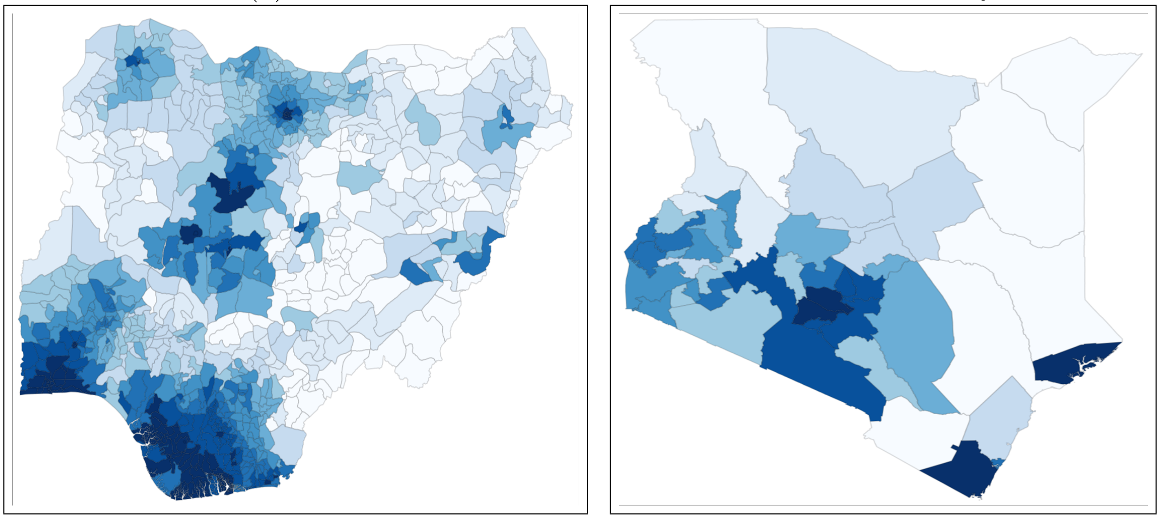
[35]In all steps and for all districts, we average the variables used over the sample period.

[36]Table G1 in the Online Appendix lists the ten districts with the largest overall impact from this counterfactual increase, while Table H1 presents the same ranking for the five other most populous African countries.

impact from the same increase in nighttime lights are 306.9 (73.4) for Brass (Lamu), which is the top district in Nigeria (Kenya). Hence, compared to a random district selection, policy makers can increase the overall impact of their investment by a factor of around 13 (5) when targeting the best possible district.

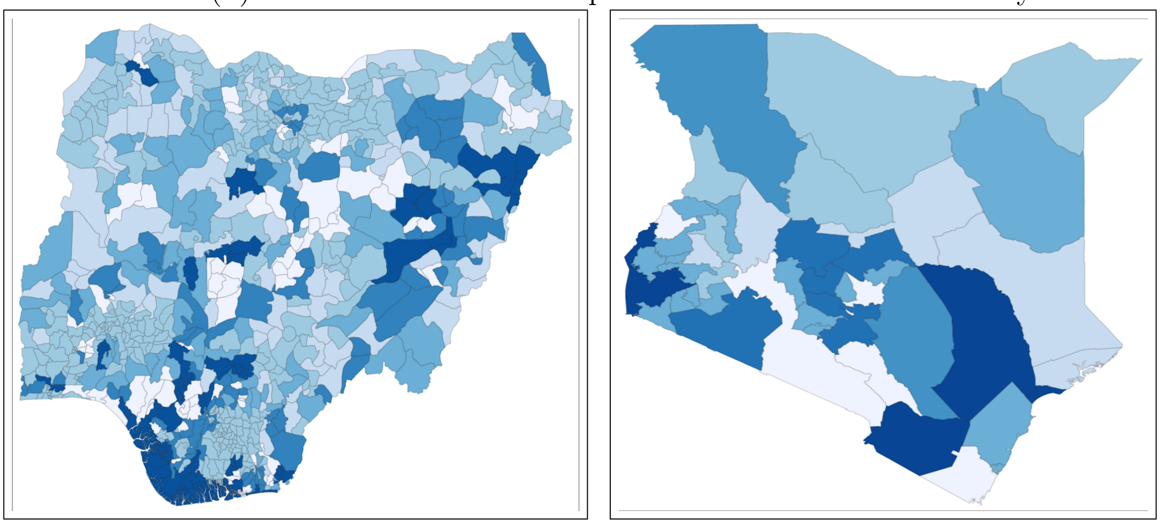
Figure 2: Results of the Counterfactual Policy Experiments - Nigeria and Kenya

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 second policy experiment consists of increasing the road connectivity of each district, again one at a time. This experiment mimics improvements in the road infrastructure. We proceed as follows. First, for any given district, we determine the set of geographically close districts with which the given district is not yet linked via a major road, and we choose the district with the highest average nighttime light value from this set of districts. Second, we add a link between these two districts (with a value of 1) in the non-normalized road connectivity matrix. Third, we re-normalize the road connectivity matrix and then recalculate the spatially lagged dependent and independent variables using this new matrix. Fourth, we recalculate the predicted nighttime lights (in logs) for each African district and compute the sum across all African districts. Finally, we, again, compare this sum with the sum of nighttime lights (in logs) across Africa from the baseline. We repeat 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. Panel (b) in Figure 2 maps the results for Nigeria (left panel) and Kenya (right panel), with darker colors implying a stronger overall impact of the counterfactual improvement of road connectivity.[37]

The top ten districts in Nigeria are all in the Niger Delta. The first, Khana, is on the mainland, while the next two, Bonny and Okrika, are islands with intense nighttime lights but poor road connectivity. Improving their road connectivity would lead to positive economic spillovers from these two districts into other districts in the Niger Delta and beyond. Taita Taveta, which borders Mombasa, is the district for which better road connectivity generates the strongest overall impact in Kenya. The top ten Kenyian districts further include some districts with high key player centrality as well as some dark/poor districts, where an increase in economic activity from better road connectivity would be particularly valuable. Similarly, better road connectivity would also be of value in many dark/poor districts in Northeastern Nigeria.

The differences in the overall impact of improved road connectivity are again large. For example, the overall impact is around 35 (28) times larger for Khana (Taita Taveta) than a randomly chosen Nigerian (Kenyan) district. Our network-based approach, which allows identifying the districts where improvements in the transport infrastructure reap

[37]Table G2 in the Online Appendix lists the ten districts with the largest overall impact from this counterfactual increase for Nigeria and Kenya, while Table H2 presents the same ranking for the five other most populous African countries.

the highest return, may thus become a very valuable tool for policymakers.

6 Concluding remarks

In this paper, we study the role of geographic, 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 is to calculate the key player centralities by performing counterfactual exercises, which consist 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 find that primate cities are important for a country's economic development due to their high economic activity and good connectivity. However, outside of the relatively small group of primate cities, we show that a large number of other districts in Africa are also important localities for diffusing the benefits of positive economic shocks to other districts. We further conduct two counterfactual policy exercises; the first increased economic activity in each district, one at a time, and the second improved each district's road connectivity. These counterfactual exercises 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.

More generally, we believe that using network theory to regional growth can lead to interesting and surprising policy implications because it provides new tools based not only on direct but also on indirect spillovers, which explicitly take into account the complex topology of the network connecting different regions. Our counterfactual exercises illustrate the potential of this approach for informing policymakers in Africa as well as international donors and development agencies.

Even though our analysis has some limitations, we believe that identifying key districts is very important for policy purposes, especially for helping targeting the districts that spread economic activity the most.[38] This is a first stab at a complex problem and we

[38]Targeting in networks in developing countries is a very active research area (e.g., Banerjee et al.,

hope that more research will be conducted on this topic in the future. This paper has used exogenous shocks to local mining wealth to identify the key districts for diffusing economic spillovers. Future work could apply our framework and use exogenous variation of other local economic shocks, such as short-term variation in agricultural yields. Exploration of the mechanisms that drive the propagation of shocks, and how they differ across different types of networks, would also be an interesting extension of our analysis.

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.

Acemoglu, D., Akcigit, U. and W.R. Kerr (2016), "Networks and the macroeconomy: an empirical exploration," *NBER Macroeconomics Annual* 30, 273–335.

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

Alesina, A., Michalopoulos, S. and E. Papaioannou (2016), "Ethnic inequality," *Journal of Political Economy* 124, 428–488.

Alesina, A. and E. La Ferrara (2005), "Ethnic diversity and economic performance," *Journal of Economic Literature* 43, 762–800.

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.

Aragon, F.M. and J.P. Rud (2016). "Polluting industries and agricultural productivity: evidence from mining in Ghana," *Economic Journal* 126, 1980–2011.

Asher, S., Lunt, T., Matsuura, R. and P. Novosad (2021), "Development research at high geographic resolution: an analysis of night lights, firms, and poverty in India using the SHRUG open data platform," *World Bank Economic Review* 35, 845–871.

Asher, S. and P. Novosad (2020), "Rural roads and local economic development," *American Economic Review* 110, 797–823.

2013, 2019; Beaman et al., 2021a, 2021b).

Ballester, C., Calvó-Armengol, A. and Y. Zenou (2010), "Delinquent networks," *Journal of the European Economic Association* 8, 34–61.

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

Banerjee, A., Chandrasekhar, A.G., Duflo, E., and M.O. Jackson (2013), "The diffusion of microfinance," *Science* 341, 1236498.

Banerjee, A., Chandrasekhar, A.G., Duflo, E., and M.O. Jackson (2019), "Using gossips to spread information: Theory and evidence from two randomized controlled trials," *Review of Economic Studies* 86, 2453–2490.

Baskaran, T., Min, B. and Y. Uppal, (2015) "Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections," *Journal of Public Economics*, 126, 64–73.

Baugh, K., Elvidge, C., Ghosh, T. and D. Ziskin (2010), "Development of a 2009 stable lights product using DMSP-OLS data," *Proceedings of the 30th Asia-Pacific Advanced Network Meeting* 30, 114–130.

Beaman, L., BenYishay, A., Magruder, J. and A.M. Mobarak (2021a), "Can network theory-based targeting increase technology adoption?," *American Economic Review* 111, 1918–1943.

Beaman, L., Keleher, N., Magruder, J. and C. Trachtman (2021b), "Urban networks and targeting: Evidence from Liberia," AEA Papers and Proceedings 111, 572–576.

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

Benshaul-Tolonen, A. (2019), "Local industrial shocks and infant mortality," *Economic Journal* 129, 1561–1592.

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.

Berman, N., Couttenier, M., and Soubeyran, R. (2021), "Fertile ground for conflict," *Journal of the European Economic Association* 19, 82–127.

Bifulco, R., Fletcher, J.M. and S.L. Ross (2011), "The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health," *American Economic Journal: Economic Policy* 3, 25–53.

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.

Bramoullé, Y., Rogers, B.W. and A. Galeotti (2016), *The Oxford Handbook of the Economics of Networks*, Oxford: Oxford University Press.

Bramoullé, Y., Djebbari, H. and B. Fortin (2020), "Peer effects in networks: A survey," *Annual Review of Economics* 12, 603–629.

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.

Bruederle, A. and R. Hodler (2018), "Nighttime lights as a proxy for human development at the local level," *PloS one* 13.

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 Management* 55, 248–264.

Carvalho, V.M., Nirei, M., Saito, Y. and A. Tahbaz-Salehi (2021), "Supply chain disruptions: Evidence from the great east Japan earthquake," *Quarterly Journal of Economics* forthcoming.

Carvalho, V.M. and A. Tahbaz-Salehi (2019), "Production networks: A primer," *Annual Review of Economics* 11, 635–663.

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

Chen, X. and W.D. Nordhaus (2011), "Using luminosity data as a proxy for economic statistics," *Proceedings of the National Academy of Sciences* 108, 8589–8594.

Chen, X. and W.D. Nordhaus (2015), "A test of the new VIIRS lights data set: Population and economic output in Africa, " *Remote Sensing* 7, 4937–4947.

Center for International Earth Science Information Network (CIESIN) (2016), "Gridded Population of the World, Version 4 (GPWv4): Administrative Unit Center Points with Population Estimates." Columbia University and NASA Socioeconomic Data and Applications Center (SEDAC).

Colella, F., Lalive, R., Sakalli, S. and M. Thoening (2018), "Inference with arbitrary clustering," *IZA Discussion Paper No. 12584*.

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

Couttenier, M. and R. Soubeyran (2014), "Drought and civil war In Sub-Saharan Africa," *Economic Journal* 124, 201–244.

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

De Luca, G., Hodler, R., Raschky, P.A., and M. Valsecchi (2018), "Ethnic favoritism: An axiom of politics?," *Journal of Development Economics* 132, 115–129.

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

Elvidge, C.D., Ziskin, D., Baugh, K.E., Tuttle, B.T., Ghosh, T., Pack, D.W., Erwin, E.H., and M. A Zhizhin (2009), "Fifteen year record of global natural gas flaring derived from satellite data," *Energies* 2, 595–622.

Gennaioli, N., LaPorta, R., Lopez-de-Silanes, F. and A. Shleifer (2014), "Growth in regions," *Journal of Economic Growth* 19, 259–309.

Gibson, J., Olivia, S., Boe-Gibson, G., and C. Li (2021), "Which night lights data should we use in economics, and where?, " *Journal of Development Economics*, 149.

Girard, V., Molina-Millán, T. and G. Vic (2022), "Artisanal mining in Africa, " Nova Africa Working Paper 2201.

Harari, M. and E. La Ferrara (2018), "Conflict, climate and cells: A disaggregated analysis," *Review of Economics and Statistics* 100(4), 594–608.

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.

Henderson, V.J., Squires, T., Storeygard, A. and D.N. Weil (2018), "The global spatial distribution of economic activity: nature, history, and the role of trade," *Quarterly Journal of Economics* 133, 357—406.

Herrington, R. (2021), "Mining our green future," *Nature Reviews Materials* 6, 456–458.

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

Hund, K., La Porta, D., Fabregas, T.P, Laing, T., and J. Drexhage (2020), *Minerals for climate action: The mineral intensity of the clean energy transition*, Washington, DC: World Bank.

Jackson, M.O. (2008), "Network formation," In: Durlauf, S. and L. Blume (Eds.), *The New Palgrave Dictionary of Economics and the Law*, London: MacMillan Press, pp. 1–9.

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 (2022), "The average and heterogeneous effects of transportation investments: Evidence from sub-Saharan Africa 1960–2010," *Journal of the European Economic Association* 20, 1–38.

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.

Kotsadam, A. and A. Tolonen (2016), "African mining, gender, and local employment," *World Development*, 83, 325–339.

Li, X., Zhou, Y., Zhao, M., and X. Zhao (2020), "A harmonized global nighttime light dataset 1992–2018, " *Scientific Data* 7, 168.

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

Lee, L-F., Liu, X., Patacchini, E. and Y. Zenou (2021), "Who is the key player? A network analysis of juvenile delinquency," *Journal of Business and Economic Statistics* 39(3), 849–857.

Mamo, N., Bhattacharyya, S., Moradi, A. and R. Arezki (2019), "Intensive and extensive margins of mining and development: evidence from Sub-Saharan Africa," *Journal of Development Economics* 139, 28–49.

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.

Redding S. and E. Rossi-Hansberg (2017), "Quantitative spatial economics," *Annual Review of Economics* 9, 21–58.

Reichl, C. and M. Schatz (2022). *World Mining Data*, Federal Ministry of Agriculture, Regions and Tourism, Austria, and International Organizing Committee for the World Mining Congresses.

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

van der Ploeg, F. (2011), "Natural resources: curse or blessing?," *Journal of Economic Literature* 49, 366–420.

von der Goltz, J. and B. Prabhat (2019), "Mines: The local wealth and health effects of mineral mining in developing countries," *Journal of Development Economics*, 139, 1–16.

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.