Spatial Diffusion of Economic Shocks in Networks*

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

September, 2018

Abstract

The aggregate economic impact of any project depends on its effects within the chosen administrative region as well as its economic spillovers into other regions. However, little is known about how these spillovers propagate through geographic, ethnic and road networks. In this paper, we analyze both theoretically and empirically the role of these networks in the spatial diffusion of local economic shocks. We develop a network model that shows how a district's level of prosperity is related to its position in the network. The network model's first-order conditions are used to derive an econometric model of spatial spillovers that we estimate using a panel of 5,944 districts from 53 African countries over the period 1997–2013. To identify the causal effect of spatial diffusion, we exploit cross-sectional variation in the location of mineral mines and exogenous time variation in world mineral prices. Our results show that road and ethnic connectivity are particularly important factors for diffusing economic spillovers over longer distances. We then use the estimated parameters from the econometric model to calculate the key player centralities, which determine which districts are key in propagating local economic shocks across Africa. We further show how counterfactual exercises based on these estimates and the underlying network structure can inform us about the potential gains from policies that increase economic activity in specific districts or improve road connectivity between districts.

Keywords: Economic development, networks, spatial spillovers, key player centrality, natural resources, transportation, Africa.

JEL classification: O13, O55, R12.

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

[†]Department of Economics, Monash University; email: ashani.amarasinghe@monash.edu.

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

[§]Department of Economics, Monash University; email: paul.raschky@monash.edu.

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

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 reasons could be that in many African countries, economic activity is concentrated in a few geographic areas,[1] and that geography, poor transport infrastructure and ethnic heterogeneity may limit the extent of spatial economic spillovers (e.g., Easterly and Levine, 1997, Bloom et al., 1998, Brock et al., 2001, Masanjala and Papageorgiou, 2008, Crespo Cuaresma, 2011).

The aim of this paper is to estimate the extent of spatial economic spillovers between African districts, highlight the role of geographic, transport and ethnic networks in the context of regional economic development in Africa and to determine which districts are key in propagating local economic shocks across Africa. By doing so, we not only present a novel link between network theory and macro-level measures of economic activity, but (to the best of our knowledge) we are also the first to estimate a well-identified econometric model of spatial economic spillovers across an entire continent.

We first develop a simple theoretical model that describes how one district's prosperity depends on its economic activity and its connectivity with other districts in a multi-district network framework. The first-order conditions are used to estimate an econometric model of spatial spillovers in which the economic activity of one district depends on the economic activity of neighboring districts, which are defined in terms of geographic, ethnic and road connectivity networks.

We estimate this econometric model using a balanced panel dataset of 5,944 African districts (ADM2, second subnational level) and yearly data from 1997–2013. 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 in the dependent

[1]Masanjala and Papageorgiou (2008) show that African countries' GDPs heavily rely on mineral exports, which are highly localized. Ades and Glaeser (1995) and Henderson (2002) point out the importance of primate cities in general, while Storeygard (2016) empirically underpins this point for primate cities in Africa.

and explanatory variables. We interpret the estimated coefficient of the spatial lag of the dependent variable in that model as the effect of a district's connectivity on its own economic activity. Our preferred specifications include time-varying controls as well as district and country-year fixed effects to account for all time-invariant differences across districts and country-year specific shocks that affect all districts in a country and year, respectively.

The major empirical challenge is that the estimated parameter is likely to be biased due to reverse causality and time-varying omitted variables. We address this problem by applying an instrumental variables (IV) strategy similar to that of Berman et al. (2017). This strategy relies on cross-sectional variation in the neighboring districts' mining opportunities and fluctuations in the world price of the minerals extracted in these districts as the source of exogenous temporal variation in these districts' performance.

One potential threat to our identification strategy is that mineral resources in Africa are often clustered in a number of neighboring districts. Hence, a positive price increase in a particular mineral may not only increase economic activity and prosperity in a district through spatial spillovers but could also directly impact prosperity in the district if this district happens to have an endowment of that particular mineral itself. Therefore, in our specifications we control for the mineral wealth of the district itself. A second concern is that changes in the neighboring districts' mineral wealth may systematically trigger violent conflict in these districts and these conflicts might spillover to other districts (Berman et al., 2017). We therefore control for conflict in a district and the neighboring districts. A third concern is that spatial spillovers may be fiscal rather than economic. Suppose the central government channels resource-based government revenues back to mining provinces (ADM1, first subnational level). In this case, non-mining districts belonging to a province in which there are some mining districts may well benefit from these government transfers, but we may not want to think of these transfers as economic spillovers. We therefore present specifications in which we compare spillovers to neighboring districts belonging to the same or different provinces.[2]

[2] In addition to these concerns, notice that given this identification strategy, we estimate the local average treatment effect (LATE) of a particular type of economic shock for a particular subset of districts.

Our estimation results are as follows: Individually, geographic, ethnic and road connectivity all increase local economic activity, but they impact local economic activity in different ways. With respect to geographic connectivity, positive economic shocks only seem to affect other districts within very close geographic proximity. A positive income shock in one district only systematically increases economic activity in districts that are within a 70km radius. The spillover effect is no longer statistically significant beyond this distance. Ethnic connectivity transmits positive economic shocks to other districts that share the same ethnicity but are not necessarily contiguous. Connectivity via major roads is the most important determinant of spatial spillovers and positive economic shocks diffuse to districts well beyond 100km if they are connected by better road infrastructure.

We then turn to measuring the network centrality of all the districts in Africa. Our simple model and the estimated coefficients on the spatial lag variable allow us to calculate Katz-Bonacich and key-player network centralities.[3] Based on the key-player centrality, we determine the "key" districts in African countries, i.e., the districts that contribute most to economic activity across Africa. These districts are typically characterized by high local economic activity as well as good connectivity.

We think that our approach has important implications for policymakers in Africa as well as international donors and development agencies. A planner who decides where to locate a particular developmental project or where to build a new or better road may need to consider many aspects, but one of them should be the potential of this project to generate spatial economic spillovers. Therefore, we 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.

Among other things, these specific economic shocks could generate both positive and negative spatial spillovers and, as such, we ultimately estimate the total net effect of the economic spillovers. We discuss these issues in more detail below.

[3]We also calculate betweenness and eigenvector centrality. These two centrality measures are parameter free and only depend on the topology of the network (see e.g., Jackson, 2008).

The rest of the paper is structured as follows. In the next section, we highlight our contribution to the literature. In Section 3, we develop the formal model. We then describe the data in Section 4, discuss the empirical strategy in Section 5, and present the results from our estimates in Section 6. Using these results, we identify the districts that are the most central in Section 7 and discuss counterfactual policy exercises in Section 8. In Section 9, we briefly conclude.

2 Related Literature

Our paper contributes to four different strands of the literature. First, we contribute to the empirical literature on the effects of networks in economics.[4] This literature has so far mostly focused on using micro data to test predictions from network theory (see e.g., Calvó-Armengol et al., 2009, for education or Cawley et al., 2017, for obesity). A major challenge in these studies is the endogeneous sorting of agents in networks and the endogeneous formation of networks (see, in particular, Bramoullé et al., 2009, 2016, and Blume et al., 2011). There are also some recent papers that study network effects from a macroeconomic perspective (see, in particular, Acemoglu et al., 2012, 2015, Carvalho, 2015). 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. For example, localized disturbances such as the 2011 earthquake in Japan, the financial crisis, the 2007–2009 recession (Carvalho, 2015) or natural disasters (Barrot and Sauvagnat, 2016) affect decisions by many firms. These micro decisions then propagate through the production network and the resulting synchronized behavior affects business cycles.

Our contribution to the network literature is twofold. First, compared to the microeconomic strand of the literature, we do not have the problem of endogenous network formation of agents since our unit of analysis is the geographical location of a district

[4]For overviews on the economics of networks, see Ioannides (2012), Jackson (2008, 2014), Jackson and Zenou (2015), Bramoullé et al. (2016), and Jackson et al. (2017).

in a country, which is clearly exogenous.[5] Also, we propose a credible IV strategy that deals with the reflection problem and the endogeneity of the spatial lag variable. Second, compared to the macroeconomic literature, we focus on the geographical location of districts rather than the location of firms in a production network. Moreover, we propose an explicit network analysis and determine the key districts in each country, i.e., the ones that need to be supported if the government wants to maximize total economic growth. In that sense, our paper may be closer to that of König et al. (2017), who present a key player analysis for the conflict in the Democratic Republic of the Congo. We, however, believe that ours is the first study that looks at the role of networks in explaining spatial economic spillovers across an entire continent.

A second literature to which we contribute is the one on the economic effects of natural resource revenues. This literature has traditionally focused on whether and when natural resources are a curse for a country's economic development (e.g., Sachs and Warner, 2001, van der Ploeg, 2011). More recently, the focus of this literature has turned to the local effects of resource extraction. The local economic effects may be positive if the extractive industries hire local workers or if their presence leads to an increase in the demand for locally produced goods and services (e.g., Aragon and Rud, 2013). The effects may also be positive if the central government channels a disproportioned share of the taxes and fees paid by the extractive industry back to the governments of the resource-rich provinces or districts (e.g., Caselli and Michaels, 2013). Resource extraction may have negative effects on the local economy if it causes environmental pollution (e.g., Aragon and Rud, 2016) or conflict (Berman et al., 2017).

In our paper, we focus on spatial spillovers of local economic shocks originating from the mining sector. Positive spatial spillovers may also result from an increased demand or increased fiscal transfers to the province in which the mining activity takes place, while conflict could lead to negative spatial spillovers. To date, there has been little research on spatial spillovers of local resource extraction. A notable exception is Aragon and

[5]The ethnic network that we use is exogenous as well since we rely on maps of traditional ethnic homelands. The road network is exogenous to the extent that the contemporary network of major roads in Africa largely originates from roads which were built by colonial powers based on their own needs (Rodney, 1982).

Rud's (2013) study on the spillovers from a large gold mine in Peru. Closer to us, Mamo et al. (2017) look at spatial spillovers from 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 that studies the importance of transport networks and, more broadly, market access for subnational economic development in Africa. Studies in this area often focus on the construction of new highways in, e.g., China or India (Banerjee et al., 2012, Faber, 2014, Alder, 2015). Like us, Storeygard (2016), Bonfatti and Poelhekke (2017), and Jedwab and Storeygard (2018) focus on road networks in Africa. Storeygard (2016) and Jedwab and Storeygard (2018) investigates the effect of transportation costs on the income of African cities. Bonfatti and Poelhekke (2017) document that African roads typically connect mines directly to the coast and study how this pattern affects bilateral trade with neighboring versus overseas countries. We differ from these papers by focusing on the importance of roads in shaping the spatial diffusion of economic shocks across Africa. Those spillover effects could be driven by increased market access (e.g. Donaldson 2018, Donaldson and Hornbeck 2016, Jedwab and Storeygard 2018) which is of importance in particular for the road and geographic networks. However, economic spillovers also occur because of a numerous other channels such as technology diffusion or transfer payments and our aim is to estimate the aggregate spill-over effects of those various channels.

Fourth, we contribute to the literature on the effects of ethnic diversity on aggregate outcomes (e.g., Easterly and Levine, 1997, Alesina et al., 2003, Alesina and La Ferrara, 2005). Traditionally, this literature has focused on the ethnic composition at the country level. More recently, two strands have developed that take the information about the spatial distribution of different ethnic groups into account. A first strand uses this information to explain differences in economic outcomes across ethnic homelands (e.g., Michalopoulos and Papaioannou, 2013, Burgess et al., 2015, Dimico, 2017, Hodler and Raschky, 2017, De Luca et al., 2018). A second strand is more macro in nature. It uses information about the spatial distribution of ethnic groups to develop country-level measures of ethnic segregation (Alesina and Zhuravskaya, 2011, Hodler et al., 2017) or ethnic inequality (Alesina et al., 2016) and shows how these measures correlate with aggregate economic outcomes. Like these two strands, we also use information on the spatial distribution of economic groups. However, we link the more micro literature to the macro one using subnational regions as units of analysis, but with a focus on how economic shocks propagate through the ethnic network and thereby affect economic development outside these small units.

At a more general level, we relate to the growing number of empirical studies analyzing the differences in economic development between and within subnational units rather than countries (e.g., Gennaioli et al., 2014, Henderson et al., 2018, Michalopoulos and Papaioannou, 2017, and some of the studies cited above). Within this strand of literature, a few studies investigate the role of spatial spillovers on regoinal development using spatial econometric techniques (e.g., Crespo-Cuaresma et al., 2014, Higgins et al., 2006, Mamo et al., 2017). A common problem in this context is the potential endogeneity of the spatial lag variable. To the best of our knowledge, we are the first to address this issue by employing an IV approach to estimate the spatial spillover effects in the context of regional economic development.[6]

3 A simple model

3.1 The case of a single network

Consider a network linking different districts. A network (graph) $\boldsymbol{\omega}$ is the pair (N, E) consisting of a set of nodes (here districts) $N = \{1, \ldots, n\}$ and a set of edges (links) $E \subset N \times N$ between them. The neighborhood of a node $i \in N$ is the set $N_i = \{j \in N : (i, j) \in E\}$. The adjacency matrix $\boldsymbol{\Omega} = (\omega_{ij})$ keeps track of direct links so that $\omega_{ij} \in]0, 1]$

[6] Harari and La Ferrara (2018) employ a similar approach in a robustness check. However, they focus on spillovers in conflict rather than economic development.

if a link exists between districts i and j, and $\omega_{ij} = 0$ otherwise.[7] We assume that the adjacency matrix $\mathbf{\Omega}$ is row-normalized so that the sum of each of its rows is equal to 1, i.e., $\sum_j \omega_{ij} = 1$ for all i.[8] In the data, $\mathbf{\Omega} = (\omega_{ij})$ will capture connectivity based on geography, the road network or ethnicity (see Section 4.2).

The level of prosperity p_i of a district i is given by:

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

where l_i is the economic activity in district i, X_i captures the characteristics of district i that determine the marginal effect of its own economic activity on its own prosperity, and $\rho > 0$ captures the cross (spillover) effects between own prosperity and the neighbors' activities since

$$\frac{\partial^2 p_i}{\partial l_i \partial l_j} = \rho \omega_{ij}.$$

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

Each district i or, more exactly, the local politicians in charge of the district choose the district's own economic activity level l_i, taking as given the choices of all the other districts. The utility function of district i is given by:

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

The first-order condition yields:

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

or in matrix form:

$$\mathbf{l} = (\mathbf{I} - \rho \mathbf{\Omega})^{-1} (\mathbf{X} + \boldsymbol{\varepsilon}) =: \mathbf{C}^{BO}_{\mathbf{X}+\boldsymbol{\varepsilon}}(\rho, \boldsymbol{\omega}) \tag{3}$$

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

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

where $\mathbf{l}$ is a column-vector of l_is, $\mathbf{I}$ is the identity matrix, and $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are the vectors corresponding to the X_is and ε_is, respectively. In (3), $\mathbf{C}^{BO}_{\mathbf{X}+\boldsymbol{\varepsilon}}(\rho, \boldsymbol{\omega})$, whose ith row is $C^{BO}_{i,X_i+\varepsilon_i}(\rho, \boldsymbol{\omega})$, is the weighted *Katz-Bonacich centrality* (due to Bonacich, 1987, and Katz, 1953), where the weights are determined by the sum of X_i and ε_i for each district i. Denote by $\mu_1(\mathbf{\Omega})$ the spectral radius of $\mathbf{\Omega}$. Then, if $\rho\mu_1(\mathbf{\Omega}) < 1$, there exists a unique interior equilibrium given by (2) or (3). Since the adjacency matrix $\mathbf{\Omega}$ is assumed to be row-normalized, it holds that $\mu_1(\mathbf{\Omega}) = 1$. Thus, the condition for existence and uniqueness can be written as $\rho < 1$.

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

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

Consider now a network where the two districts are linked to each other (i.e., $\omega_{12} = \omega_{21} = 1$). Then, if $\rho < 1$, the unique interior Nash equilibrium is given by:

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

In other words, because of complementarities, in the dyad, the level of activity of each district is much higher than when the districts are not connected. The factor $1/(1-\rho) > 1$ is the *network multiplier.*[9]

Observe that the way we modeled spillover effects (see (2)) is similar to the way urban economists have been modeling agglomeration effects. For example, in Ahlfeldt et

[9]Observe that if we keep the ex ante heterogeneity, we have that, if $\rho < 1$, then the unique interior Nash equilibrium is equal to:

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

al. (2015), agglomeration effects are modeled as production externalities. In our case, spillover effects might capture those effects but could also be driven by other effects as well.[10]

So far, we have assumed that $\rho > 0$, which implies strategic complementarities. However, it may be the case, especially in Africa, that there are negative spillovers on neighboring districts so that $\rho < 0$. For example, if a district i has a high activity level, then it may be that the neighboring districts are negatively affected by this because their residents may move to district i, which is more prosperous. The equilibrium activity levels are still given by (2) or (3). In that case, the condition for the existence and uniqueness of equilibrium is no longer given by $\rho\mu_1(\mathbf{\Omega}) < 1$ but by $\rho\mu_{\min}(\mathbf{\Omega}) > -1$, where $\mu_{\min}(\mathbf{\Omega})$ is the *lowest* eigenvalue of $\mathbf{\Omega}$.

3.2 The case of multiple spillover effects

In the real world, there is more than one type of spillovers between districts. For example, in our main specifications below, we use different adjacency matrices $\mathbf{\Omega} = (\omega_{ij})$ that keep track of the (inverse) spatial distance between districts, the road network and the proximity in terms of ethnicity. As a result, we now assume that:[11]

$$p_i = X_i l_i + \rho_1 l_i \sum_{j=1}^{J} \omega_{1,ij} l_j + \rho_2 l_i \sum_{j=1}^{J} \omega_{2,ij} l_j + \rho_3 l_i \sum_{j=1}^{J} \omega_{3,ij} l_j + l_i \varepsilon_i \tag{4}$$

where $\rho_1 > 0$, $\rho_2 > 0$ and $\rho_3 > 0$. We now have three adjacency matrices $\mathbf{\Omega}_1 = (\omega_{1,ij})$, $\mathbf{\Omega}_2 = (\omega_{2,ij})$ and $\mathbf{\Omega}_3 = (\omega_{3,ij})$, which are all assumed to be row-normalized. This means that the prosperity of a district is affected differently by the different ways we measure the "proximity" between neighborhoods. Remember that $\mu_1(\mathbf{A})$ denotes the spectral radius of a matrix $\mathbf{A}$. Thus, we have the following result:

Proposition 1. *Consider a model where the prosperity of a district is given by (4). Then,*

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

[11] Our analysis extends straightforwardly to the case of n adjacency matrices.

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

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

or, in matrix form,

$$\mathbf{l} = (\mathbf{I} - \rho_1\mathbf{\Omega}_1 - \rho_2\mathbf{\Omega}_2 - \rho_3\mathbf{\Omega}_3)^{-1}(\mathbf{X} + \boldsymbol{\varepsilon})$$

Proof: We need to show that $\mathbf{I}-\mathbf{A}$ is non-singular (i.e., invertible), where $\mathbf{A} \equiv \rho_1\mathbf{\Omega}_1 + \rho_2\mathbf{\Omega}_2 + \rho_3\mathbf{\Omega}_3$. We know that $\mathbf{I}-\mathbf{A}$ is non-singular if $\mu_1(\mathbf{A}) := \mu_1(\rho_1\mathbf{\Omega}_1 + \rho_2\mathbf{\Omega}_2 + \rho_3\mathbf{\Omega}_3) < 1$ (see, e.g., Meyer, 2000, p. 618). From the first-order condition (5), we see that the solution is always interior. ∎

In terms of empirical implications, the models with a single and multiple spillovers are very different, as we discuss in detail when interpreting the results of our econometric model of spatial spillovers (see Section 6.1).

4 Data

Our units of observation are administrative units at the second subnational level (ADM2), which we call districts.[12] The final dataset consists of yearly observations for 5,944 districts from 53 African countries over the period 1997–2013.[13] The average (median) size of a district is 39km^2 (6km^2) and the average (median) population is around 150,000 (55,000).

For our purpose, there are a number of advantages of using administrative units rather than other 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

[12]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 A1 in the Online Appendix shows the boundaries of each district in Africa.

[13]Table A1 in the Online Appendix lists all the African countries in our sample and provides the number of districts per country.

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.[14]

4.1 Dependent variable: Nighttime lights

Satellite data on the intensity of nighttime lights comes from the the National Oceanic and Atmospheric Administration (NOAA). Weather satellites from the US Air Force circle the earth 14 times per day and measure light intensity. The NOAA uses evening observations during the dark half of the lunar cycle in seasons when the sun sets early, but removes observations affected by cloud coverage, or northern or southern lights. It further processes the data by setting readings that are likely to reflect fires, other ephemeral lights or background noise to zero.[15] The objective is that the reported nighttime lights are primarily man-made. The NOAA then provides annual data for the time period from 1992 onwards for output pixels that correspond to less than one square kilometer. The data come on a scale from 0 to 63, with higher values implying more intense nighttime lights.

Nighttime lights are a proxy for economic activity, as most forms of consumption and production in the evening require light. Moreover, public infrastructure is often lit at night. It is, therefore, not surprising that Henderson et al. (2012) and Hodler and Raschky (2014) find a high correlation between changes in nighttime light intensity and GDP at the level of countries and subnational administrative regions, respectively. Using data from Gennaioli et al. (2014), we also find a high correlation between nighttime lights and subnational GDP for 82 subnational administrative regions from nine African countries (see Table B1 in the Online Appendix).

To construct our dependent variable, $Light_{ict}$, we take the logarithm of the average nighttime light pixel value in district i of country c in year t. To avoid losing observations with a reported nighttime light intensity of zero, we follow Michalopoulos and

[14]In a robustness test presented in Table G1 in the Online Appendix, we use 0.5×0.5 degree grid cells instead of ADM2 regions and show that all our results are qualitatively the same.

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

Papaioannou (2013, 2014) and Hodler and Raschky (2014) in adding 0.01 before taking the logarithm.

4.2 Connectivity matrices

We construct three connectivity matrices to measure spatial spillovers.

4.2.1 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 (1958) documents the spatial distribution of ethnic homelands in Africa and subdivides the continent into over 800 ethnic homelands.[16]

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, $\omega_{ic,jc}$, where elements are 1 if the ethnicity in district i is the same as the ethnicity in district j, and 0 otherwise.

4.2.2 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 $d_{ic,jc}$ connecting the centroids of districts i and j. Third, following Acemoglu et al. (2015), we measure the variability of altitude, $e_{ic,jc}$, 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 $\tilde{d}_{ic,jc} = 1/d_{ic,jc}(1 + e_{ic,jc})$.

[16] Figure A2 in the Online Appendix shows the digitized version of Murdock's original map.

Defining geographic connectivity using the inverse altitude-adjusted distance as opposed to contiguity proves advantageous on three accounts.[17] 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, $e_{ic,jc}$, we account for the topology of the landscape. Districts separated by a mountainous terrain, for example, receive a lower connectivity weight, as opposed to districts connected via a flat surface. Third, measuring geographic connectivity based on geodesic distance allows truncation at different distances, enabling the determination of the extent of spillovers. Leveraging this advantage, we construct different weighting matrices by varying the distance considered in defining a district's neighbors. The main specification will use a cutoff of 70km (for reasons made explicit below). In this case, we set the spatial weight as $\omega_{ic,jc} = 1/\tilde{d}_{ic,jc}$ if the geodesic distance $d_{ic,jc}$ is less than 70km, and $\omega_{ic,jc}$=0 otherwise.

4.2.3 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).[18] We accessed the OSM data in early 2016 and extracted information about major

[17] As an alternative we construct a weighting matrix for geographic connectivity based on contiguity. The contiguity matrix indicates whether districts i and j share a common border or, at least, a common point along their borders. We report estimates based on the contiguity matrix in Table J1 in the Online Appendix. Further, we report estimates based on geodesic distances but without adjustment for the variability in altitude in Table K1 in the Online Appendix.

[18] 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.

roads (e.g., highways and motorways) for the African continent.[19] We intersect these roads with the district boundary polygons and generate a network graph of the road network.[20] 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).[21] In the second step, we identify the shortest path on the road segments between each district and calculate the distance on that path. If districts A and B are adjacent and connected via a major road, we assign a distance value of 1km. If districts A and B are not adjacent, but connected via the road network, they are assigned the road distance between the closest road and district boundary node of A and the closest road and district boundary node of B (i.e., the road travel distance through all the district that one has to cross to get from district A to district B).

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

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

[Table 1 about here]

Note that these connectivity matrices capture spillover effects within and between countries, because in our main analysis we are interested in capturing the overall economic spillovers between African districts. However, in a robustness analysis[22] we construct new connectivity matrices that take into account national borders.

4.3 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.

[19] Figure A3 in the Online Appendix shows the road network.

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

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

[22] See Section M in the Online Appendix

We describe the construction of the respective variables in turn.

Our information on mining activity comes from the SNL Minings & Metals database. This database covers 3,487 mining projects across Africa that were active during our sample period. For each project, it contains information about the point location, i.e., the geographic coordinates, and the (potentially multiple) resources extracted at this location.[23]

We use the point locations to assign the mining projects to districts and identify all districts where a mine was active for at least one year during our sample period. Across Africa, 4% of all districts are mining districts. The indicator variable $Mine^{r}_{ic}$ is equal to one if district i of country c 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.[24]

Data on world prices of minerals are sourced from the World Bank, IMF, USGS and SNL (see Table A2 in the Online Appendix for more information on the data sources). $Price^{r}_{t}$ is the logarithm of the yearly nominal average price of resource r in USD.

4.4 Control variables

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

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

[23]Figure A4 in the Online Appendix shows the spatial distribution of mining projects across Africa.

[24]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.

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 about here]

5 Empirical strategy

The aim of the empirical analysis is to estimate equation (5) from the theoretical model. By inserting county subscript c and time subscript t, the equation can be written as:

$$l_{ict} = \rho_1 \sum_{j=1}^{J} \omega_{1,ic,jc} l_{jct} + \rho_2 \sum_{j=1}^{J} \omega_{2,ic,jc} l_{jct} + \rho_3 \sum_{j=1}^{J} \omega_{3,ic,jc} l_{jct} + X_{ict} + \varepsilon_{ict} \tag{6}$$

where $\omega_{1,ic,jc}$ is the (ic, jc) cell of the adjacency matrix based on geographic connectivity, $\omega_{2,ic,jc}$ is the (ic, jc) cell of the adjacency matrix based on road connectivity, and $\omega_{3,ic,jc}$ is the (ic, jc) cell of the adjacency matrix based on ethnic connectivity. For the sake of the exposition, we rewrite equation (6) in a more compact form:

$$l_{ict} = \sum_{k=1}^{3} \sum_{j=1}^{J} \rho_k \omega_{k,,ic,jc} l_{jct} + X_{ict} + \varepsilon_{ict}, \tag{7}$$

Denote by $\mathbf{X}_{ict} = \left(X_{ict}^1, ..., X_{ict}^M\right)$ the $(1 \times M)$ vector of time-variant, district-level characteristics and by $\boldsymbol{\beta} = \left(\beta^1, ..., \beta^M\right)^T$ a $(M \times 1)$ vector of parameters. Then, using $Light_{ict}$ to measure the level of economic activity l_{ict} and adding district and country-year fixed effects to equation (7), we get the following econometric specification:

$$Light_{ict} = \sum_{k=1}^{3} \sum_{j=1}^{J} \rho_k \omega_{k,ic,jc} Light_{jct} + \mathbf{X}_{ict} \boldsymbol{\beta} + \alpha_i + CT_{ct} + \epsilon_{ict} \tag{8}$$

where α_i and CT_{ct} are district and country-year fixed effects, respectively, and ϵ_{ict} is an error term that is assumed to be $\epsilon_{ict} \sim N(0, \sigma^2 I_n)$. As is standard in spatial econometrics, we row-normalize the adjacency matrices, i.e., we normalize them so that the sum of each row becomes equal to 1.

The specification in equation (8) assumes that local spillovers to other districts only operate through the spatial lag of the dependent variable. However, it is possible that spillover effects occur due to spatial autoregressive processes in the explanatory variables as well. To account for this, we extend the model in equation (8) to a Spatial Durbin Model by including spatial lags of the explanatory variables, $\omega_{k,ic,jc}X^m_{jct}$:

$$Light_{ict} = \sum_{k=1}^{3}\sum_{j=1}^{J}\rho_k\omega_{k,ic,jc}Light_{jct}+\mathbf{X}_{ict}\boldsymbol{\beta}+\sum_{k=1}^{3}\sum_{j=1}^{J}\sum_{m=1}^{M}\rho_k^m\omega_{k,ic,jc}X^m_{jct}+\alpha_i+CT_{ct}+\epsilon_{ict} \quad (9)$$

where ρ_k^m are the coefficients of the spatial lags.

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 $Light_{jct}$) might also be simultaneously determined by other unobserved shocks. Therefore, estimating equation (9) using OLS can yield biased and inconsistent estimates.

Traditionally, scholars have either used a quasi-maximum likelihood estimation procedure (e.g., Anselin, 1988, Lee, 2004) or a GMM instrumental variable (IV) approach that uses internal instruments (including higher order spatial lags of contiguity matrices) and estimates the model using a standard 2SLS approach (e.g., Kelejian and Prucha, 1998, Kelejian and Robinson, 1993). However, Gibbons and Overman (2012) criticize both approaches because the validity assumption is often unlikely to hold. They propose the use of exogenous instruments (as opposed to internal instruments) in a standard spatial IV specification.

We follow their suggestion and 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 time-invariant indicators of mining activity and time-variant exogenous world prices for minerals as instrumental variables:

$$Light_{jct} = \gamma MP_{jct} + \mathbf{X}_{jct}\boldsymbol{\beta} + \sum_{k=1}^{3}\sum_{i=1}^{J}\sum_{m=1}^{M}\rho_k^m\omega_{k,jc,ic}X^m_{ict} + \alpha_j + CT_{ct} + u_{jct} \quad (10)$$

where

$$MP_{jct} \equiv \frac{1}{R_{jc}} \sum_{r=1}^{R_{jc}} (Mine^r_{jc} \times Price^r_t) \tag{11}$$

with $R_{jc} = \sum_r Mine^r_{jc}$ being the number of different minerals extracted in district j of country c. Hence, for each mining region, this instrumental variable captures the average of the world prices (in logs) of all the minerals that are extracted in this district at some time during the sample period. For all other districts, this instrumental variable is zero.

This IV strategy was used by Berman et al. (2017), who focus on the effect of exogenous variation in the economic value of mines on the likelihood of local conflict. The use of the time-invariant indicator $Mine^r_{jc}$ addresses the potential endogeneity between economic activity and mining activity. Hence, our identification strategy operates as a differences-in-differences estimator, which exploits between-variation in a district's suitability for extracting specific minerals, combined with the exogenous time variation in the world price of these minerals.

For this instrument to be relevant, it is key that fluctuations in world mineral prices have a first-order effect on the mining districts' economies. Even though only 4% of all districts are mining districts, minerals account for a large proportion of export earnings in many African countries, especially strategically important minerals such as diamonds, gold, uranium and bauxite. For example, in countries such as Botswana and Congo, minerals account for over 80% of export income (USGS, 2014). 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 affects $Light_{ict}$ in district i only through $Light_{jct}$. We take a number of measures to mitigate the risk that the exclusion restriction is violated. First, we include district fixed effects in equation (10), which absorbs all time-invariant characteristics at the district level, including suitability for mining activity. The vector of country-year fixed effects accounts for any time-variant factors that might simultaneously drive mineral prices and aggregated economic development. Second, the work by Berman et al. (2017) shows that mining activity could lead to increased conflict as parties dispute over ownership of lucrative mines. This, in turn, could adversely affect economic activity. Therefore, we control for district-level conflict events in our specifications. Third, the exclusion restriction also relies on the assumption of exogeneity of world prices, i.e., no single district can affect the world price of a commodity. For this reason, we conduct a robustness check of our specification by excluding countries in the top ten list of producers for any mineral.

The statistical inference in our setting is further complicated by the clustering structure of the error terms in our econometric model. The traditional spatial clustering approach proposed by Conley (1999) allows for both cross-sectional spatial clustering and location-specific serial correlation in the error terms. This approach imposes the same spatial kernel (geographic distance) to all units in the sample. However, our empirical model does not only assume dependence based on geographic distance but also relatedness through ethnic and road networks. As such, Conley-type spatial clustering might not adequately capture unobserved correlation due to common shocks to districts which are linked by ethnicity or major roads. To address this problem we apply a novel estimator developed by Colella et al. (2018) that allows us to account for dependence across our observations' error terms in a more flexible form.[25] In practice, we correct for clustering at the network level where observations are assumed to belong to the same cluster once they are linked through at least one of the three networks (geography, ethnicity, roads).[26]

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

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

[26]Table F1 in the Online Appendix shows that applying the traditional Conley-type spatial clustering approach (Conley 1999) yields smaller standard errors.

general policy conclusions based on the estimated spatial spillover effects.

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

6 Estimation results

6.1 Main results

Table 3 presents our estimates of equations (9) and (10).[27]

[Table 3 about here]

We start with specifications that include each weighting matrix individually. First, we analyze the spillover effects using the spatial weighting matrix based on ethnic connectivity. Column (1) provides the results of the OLS estimates, while column (2) provides the comparable IV estimates. Both columns include district fixed effects and country-year fixed effects to filter out time-invariant district-specific features and country-specific time-varying characteristics. The OLS estimates show that coefficient ρ_1, i.e., the coefficient on $Ethnicity\,W\,Light_{jct}$, is positive and statistically significant at the 1% level, suggesting that economic activity in districts inhabited by the same ethnic group indeed has a positive impact on economic activity in district i. However, endogeneity concerns restrict us from drawing causal inference based on these OLS estimates. In column (2), we thus present comparable IV estimates. The coefficient of interest in the *first stage* of the IV estimate, γ, is positive and statistically significant at the 1% level. This, along with the high F statistics of the first stage, indicates that our instrumental variable is a

[27]Table 3 only reports the coefficients on the variables of main interest to improve the readability of the table. Table C1 in the Online Appendix reports the coefficients on the control variables.

strong predictor of economic activity in district j. The coefficient of interest of the *second stage*, ρ_1, is positive but not statistically significant at conventional levels.

Second, in Figure 1, we show that the coefficient of interest from the IV estimates declines in magnitude in the cutoff distance and become statistically insignificant for cutoff distances beyond 70km. As a result, in all our estimations, we will truncate the adjacency matrix at 70km. In columns (3) and (4) of Table 3, we focus on geographic connectivity and therefore use the weighting matrix based on the inverse of the altitude-adjusted geodesic distance between districts i and j, truncating the matrix at 70km. The coefficient of interest, ρ_2, is positive and statistically significant at the 1% level in both the OLS and the IV estimates.

[**Figure 1 about here**]

Third, in columns (5) and (6), we focus on road connectivity based on our matrix of inverse road distances, again truncating the matrix at 70km. The coefficient of interest, ρ_3, is positive and statistically significant at the 1% level in both the OLS and the IV estimates. As in Figure 1, the coefficient of interest from the IV estimates would remain positive and statistically significant for cutoff distances of 100km and beyond, indicating that the extent of positive spillovers spreads farther when focusing on actual transport infrastructure rather than just geography.

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 (9) and (10). We observe that 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 geographic connectivity are only statistically significant at the 10% level. The coefficient estimates further suggest that geographic connectivity tends to be less important than ethnic and road connectivity.

Figure 1 presents the coefficient estimates for ρ_2 and ρ_3, i.e., the coefficients on $Inv\ Dist\ W\ Light_{jct}$ and $Inv\ Road\ W\ Light_{jct}$, and the corresponding 90% confidence intervals from re-running our main IV specification (corresponding to column (8) of Table

3) for various cutoff distances. Spatial spillovers via purely geographic connectivity are decreasing in the cutoff distance and become statistically insignificant for cutoffs above 70km, while spatial spillovers via the road network remain large in magnitude and statistically significant even for considerably larger cutoff distances. For the subsequent analysis, we use IV estimates that are based on the largest cutoff distance at which the three coefficients of interest, i.e., ρ_1, ρ_2 and ρ_3, are all positive and statistically significant at the 10% level. That is, we use the estimates from column (8) in Table 3, where the cutoff distance is at 70km.

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 $\rho = 0.271$ (which corresponds to the estimated ρ_1 for the ethnic network in column (2) of Table 3) so that $\rho = 0.271 < \mu_1(\mathbf{\Omega}) = 1$. 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 1.355%.[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:

$$\mathbf{\Omega}_1 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \mathbf{\Omega}_2 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{\Omega}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad (12)$$

$\mathbf{\Omega}_1$ represents an ethnicity network in which all districts share the same ethnicity; $\mathbf{\Omega}_2$ is

[28]Indeed, we have $l_1 = (0.271)\left(\frac{l_2+l_3}{2}\right) + X_1 + \varepsilon_1 = 0.1355\, l_2 + 0.1355\, l_3 + X_1 + \varepsilon_1$. Thus an increase of 10% of the nighttime light of district 2 (or district 3) will increase the nighttime light of district 1 by 1.355%. The same is true for all the other districts since the network is complete.

a geographic network in which the geodesic distance between district 1 and each of the other districts is within 70km, whereas the distance between districts 2 and 3 exceeds 70km; and $\mathbf{\Omega}_3$ is a road network with a single road between districts 1 and 2.[29] Using the estimates from column (8) of Table 3, assume $\rho_1 = 0.342$, $\rho_2 = 0.305$ and $\rho_3 = 0.361$. It is easily verified that

$$\rho_1\mathbf{\Omega}_1 + \rho_2\mathbf{\Omega}_2 + \rho_3\mathbf{\Omega}_3 = \begin{pmatrix} 0 & 0.685 & 0.333 \\ 0.837 & 0 & 0.171 \\ 0.476 & 0.171 & 0 \end{pmatrix}$$

so that

$$\mu_1\left(\rho_1\mathbf{\Omega}_1 + \rho_2\mathbf{\Omega}_2 + \rho_3\mathbf{\Omega}_3\right) = 0.934 < 1$$

Let us now interpret equation (5). For district 1, we have:

$$l_1 = (0.342 + 0.305)\left(\frac{l_2 + l_3}{2}\right) + 0.361\, l_2 + X_1 + \varepsilon_1 = 0.685\, l_2 + 0.333\, l_3 + X_1 + \varepsilon_1$$

Similarly, for the two other districts, we obtain:

$$\begin{aligned} l_2 &= 0.342\left(\frac{l_1 + l_3}{2}\right) + (0.305 + 0.361)\, l_1 + X_2 + \varepsilon_2 = 0.837\, l_1 + 0.171\, l_3 + X_2 + \varepsilon_2 \\ l_3 &= 0.342\left(\frac{l_1 + l_2}{2}\right) + 0.305\, l_1 + X_3 + \varepsilon_3 = 0.476\, l_1 + 0.171\, l_2 + X_3 + \varepsilon_3 \end{aligned}$$

We can now interpret the different coefficients as follows. Given the observable and unobservable characteristics of districts 2 and 3, a 10% increase of the nighttime lights in district 1 increases the nighttime lights by 8.37% and 4.76% in districts 2 and 3, respectively. The increase in nighttime lights is larger in district 2 because it is better connected to district 1 due to the road connecting these two districts.

[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.

6.2 Robustness checks

We now discuss a number of robustness checks. The Online Appendix (Sections D–M) presents the corresponding tables.

A first set of checks show that our results are robust to small changes in the empirical specification. Table D1 replaces the country-year fixed effects with province-year fixed effects, thereby controlling for province-specific economic and political variation over time. If anything, the coefficients become larger in our IV estimates. Table E1 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 F1 shows that standard errors become smaller when using the traditional Conley-type spatial clustering approach (Conley, 1999). Table G1 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 55×55km at the equator) instead of ADM2 regions. Results remain similar, but suggest a slightly more (less) important role of road (geographic) connectivity for the spatial economic spillovers.

A second set of robustness checks tackles potential threats to our identification strategy. In our IV estimates, we exploit the variation of world mineral prices as a source of exogenous shocks, which is then propagated amongst neighbors based on different levels of connectivity. Our identification relies on the assumption that mining activity in a single unit does not influence world mineral prices. Given that our units are subnational districts, this assumption appears reasonable. Nevertheless, Table H1 excludes the districts which belong to countries that are among the top ten producers for any mineral under consideration. Results remain qualitatively similar, but the impact of geographic connectivity decreases. Table I1 shows that our results are not driven by fiscal spillovers. Fiscal spillovers would occur if non-resource-extractive districts benefit from economic activity in resource-extractive districts belonging to the same province purely because government revenues get channeled to resource-rich provinces. To control for fiscal spillovers, we add an additional connectivity matrix that captures whether two districts belong to the same province. The results suggest that the spatial lag related to this new connectivity matrix matters as well and, therefore, that fiscal spillovers may be present. More importantly for our purpose, we see that the spatial lags of our three main connectivity matrices remain quantitatively similar when controlling for fiscal spillovers.

A third set of robustness checks is based on different definitions of the three connectivity matrices. Tables J1 and K1 present results when geographic connectivity is proxied by contiguity and by inverse geodesic distance without adjustment for variability in altitude along the geodesic, respectively. Results remain similar. Table L1 replaces our binary ethnic connectivity matrix with a matrix that suggests an intermediate level of connectivity between districts of related ethnic groups. In particular, a pair of districts is still assigned a value of 1 if they share the same ethnicity, but a value of 0.5 if they do not share the same ethnicity, but belong to the same culture group according to Murdock (1969). A pair of districts that belong to different culture groups still get a value of 0. The coefficient estimates suggest a drop in the importance of the ethnic network, which is consistent with the idea that it is primarily co-ethnicity that matters for spatial economic spillovers.

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

7 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.

7.1 Theory: Different definitions of node centralities

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

7.1.1 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:

$$C_i^E(\boldsymbol{\omega}) = \frac{1}{\mu_1(\boldsymbol{\Omega})} \sum_{j=1}^{n} g_{ij} C_j^E(\boldsymbol{\omega}) \tag{13}$$

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:

$$\mu_1(\boldsymbol{\Omega}) \mathbf{C}^E(\boldsymbol{\omega}) = \boldsymbol{\Omega}\mathbf{C}^{\mathbf{E}}(\boldsymbol{\omega}) \tag{14}$$

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.

7.1.2 Katz-Bonacich centrality

In our theoretical model (Section 3), 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 (3), 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 $\rho\mu_1(\boldsymbol{\Omega}) < 1$ is satisfied.

7.1.3 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 3) for which the utility is given by (1), 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 (2) or (3). 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:

$$\max\{L^*(\boldsymbol{\omega}) - L^*(\boldsymbol{\omega}^{[-i]}) \mid i = 1, ..., n\} \tag{15}$$

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

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

$$C_{i,u_i}^{KP}(\rho, \boldsymbol{\omega}) = \frac{C_{i,u_i}^{BO}(\rho, \boldsymbol{\omega}) \sum_j m_{ji}(\rho, \boldsymbol{\omega})}{m_{ii}(\rho, \boldsymbol{\omega})} \tag{16}$$

where $C_{i,u_i}^{BO}(\rho, \boldsymbol{\omega})$ is the weighted Katz-Bonacich centrality of district i (see equation (3)) and $m_{ij}(\rho, \boldsymbol{\omega})$ is the (i, j) cell of the matrix $\mathbf{M}(\rho, \boldsymbol{\omega})= (\mathbf{I} - \rho\boldsymbol{\Omega})^{-1}$. Ballester et al. (2006, 2010) have shown that the district i^* that solves (15) 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 (16) of district i is the sum of i's centrality measures in $\boldsymbol{\omega}$, and its contribution to the centrality measure of every other district $j \neq i$ also in $\boldsymbol{\omega}$. It accounts both for one's exposure to the rest of the group and for one's contribution to every other exposure. This means that the key player i^* in network $\boldsymbol{\omega}$ is given by $i^* = \arg\max_i C_{i,u_i}^{KP}(\rho, \boldsymbol{\omega})$, where[31]

$$C_{i^*,u_i}^{KP}(\rho, \boldsymbol{\omega}) = L^*(\boldsymbol{\omega}) - L^*\left(\boldsymbol{\omega}^{[-i]}\right). \tag{17}$$

7.2 Empirical results

We compute these four different centrality measures for all 5,444 districts from the 53 African countries in our sample. In our discussion, we focus on two large countries that feature prominently in the literature: Nigeria and Kenya. Table 4 presents information on the ten most central districts (according to the key-player centrality) of each of these two countries; and Figure 2 maps with information on these two countries' districts.[32]

[Table 4 and Figure 2 about here]

Column (4) of Table 4 presents the main ranking of interest, i.e., the key-player ranking based on the geographic network, the road network *and* the ethnicity network. The

[31] Ballester et al. (2006) define the key player in (16) 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 (17).

[32] Table N1 in the Online Appendix presents information on the ten most central districts for the five most populous African countries besides Nigeria. These countries are (in decreasing order of their population): Ethiopia, Egypt, the Democratic Republic of the Congo, South Africa, and Tanzania. The rankings for all other African countries are available upon request.

underlying computation thus uses the coefficient estimates, in particular the estimated ρ's, reported in column (8) of Table 3. The district with the highest key-player centrality is Ikeja, which is part of the Lagos metropolitan area (often simply called Lagos) and the capital of Lagos State (province). Lagos is the primate city of Nigeria and its economic hub. Seven other districts belonging to the top-ten key districts of Nigeria are also part of Lagos State. The two remaining districts in the key-player ranking belong to the Kano metropolitan area (often simply Kano) in Kano State. Kano is the second largest metropolitan area in Nigeria and the economic hub of the country's northern part. The left column of Figure 2 compares key-player centrality of Nigerian districts (top row) with the districts' average nighttime light intensity (middle row) and population density (bottom row). In general, the most central districts are also districts that are more developed and more densely populated districts themselves but we see that there are still important differences between the three different indicators.

Nigeria is no exception in this respect. The key district in Kenya is Nairobi, which is the capital and the primate city. It is followed by Mombassa, which is Kenya's second largest city and home to Kenya's largest seaport (see the right column of Figure 2). The key districts encompass or are part of the primate city in many other African countries as well, including Ethiopia (Addis Ababa) and South Africa (Johannesburg). The overall pattern suggests that *primate cities* tend to be the key districts for economic development in Africa. This finding resonates with findings on the importance of primate cities for development. Ades and Glaeser (1995) and Henderson (2002) point out the importance of primate cities in general, while Storeygard (2016) empirically highlights the importance of primate cities in Africa. In addition, Henderson et al. (2017) document high spatial inequality in late-developing countries, with one or a few urban areas being much more economically developed than the rest of these countries. The finding that primate cities are the key districts suggests that they are not only characterized by higher local economic activity but that they also tend to cause more spatial economic spillovers than other districts.

Column (5) in Table 4 shows the ranking for the Katz-Bonacich centrality, again based on the estimates taking the geographic network, the road network and the ethnicity network into account. We see that the districts that rank high in terms of key-player centrality also tend to rank high in terms of Katz-Bonacich centrality in Nigeria, but not in Kenya. This is because Katz-Bonacich and key-player centralities capture different aspects of centrality. The former is a recursive measure highlighting the importance of being connected to central districts while the latter is a welfare measure that also takes into account the negative impact of cutting links on neighboring districts.

Columns (6) and (7) show the rankings for the two other centrality measures: betweenness and eigenvector centrality. Looking at Nigeria, we see that the districts from Lagos State that are top ranked in terms of key-player centrality tend to rank poorly in terms of these alternative centrality measures. This is not surprising given that the betweenness and eigenvector centralities are pure topological measures, which capture either the number of paths that go through a district (betweenness centrality) or the links to other central districts (eigenvector centrality), and Lagos State is situated at the coast in the country's south-east bordering Benin. For other countries, including Kenya, the districts that rank high in terms of key-player centrality also tend to rank high in terms of eigenvector centrality, but less so in terms of betweenness centrality. These findings suggest that it is not easy to determine the key districts with simple topological measures since the key-player centrality depends on the districts' local economic activity, i.e., their nighttime lights, *and* the spatial spillovers they generate to connected districts.

Columns (8)-(10) also give rankings of key-player centrality, but in each of these columns we compute the ranking based on the coefficient estimates from regressions including one network only. Ikeja, which is top ranked in Nigeria when looking at all three networks jointly, is also top ranked when focusing just on the road network (column (9)). It is also highly ranked when focusing just on the ethnicity network (column (8)) or the geographic network (column (10)). More generally, we see that all Nigerian districts that rank high in overall key-player centrality also rank high in any type of single-network key-player centrality. This pattern holds true for many other countries, including Kenya. This suggests that most key districts are important due to their geographic, ethnic *and* road connectivity. For many countries, including Nigeria but not Kenya, the overall key-player centrality is most highly correlated with the key-player centrality based on the road network, which indicates that road connectivity may be of particular importance.

8 Policy experiments

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

At a general level, there are two types of counterfactual exercises that we can pursue. First, we can study how changes within one or more districts propagate through the geographic, ethnic and road networks and, thereby, how these changes affect all other districts. Such an exercise provides insights that go beyond the sometimes more narrow policy evaluations that only focus on the local benefits of local policies. In Section 8.1, we perform such an analysis by computing the aggregate effect of increasing economic activity, i.e., nighttime lights, in each single district, one at a time. The second type of counterfactual exercise is based on changing the network structure. In Section 8.2, we study the aggregate economic effect of linking neighboring districts by a major road.

A few comments are in order before presenting these two policy experiments: First, the socially optimal location of a development project depends on costs and benefits, and our approach does not take into account the fact that the costs of implementing a certain project or building a certain road may differ across districts. Second, it is impossible to compare the benefits of different development projects or different project locations without an underlying social welfare function. Here, as in the previous section, we (implicitly) measure social welfare in a district by the logarithm of the average nighttime light pixel value, and we give equal weight to all districts when computing aggregate social welfare. Needless to say, one could apply our approach using alternative social welfare functions. Third, these policy experiments do not explicitly take into account the congestion effects that may occur in urban districts when new people move in .[33] Therefore, our counterfactual policy experiments are most informative about short- to medium-run effects rather than long-run effects.

8.1 Policy experiment 1: Increasing local economic activity

The first policy experiment consists in 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 average nighttime light pixel value and recalculate the spatially lagged dependent variables with the new values, but keep 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 African districts. Fourth, we compare this sum, which includes the increase in nighttime lights in one district and the subsequent spatial spillovers, with 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 Figure 3 show the districts in Nigeria (left panel) and Kenya (right panel) where this counterfactual increase in economic activity would create a stronger (darker colors) and lower (lighter colors) overall impact.[36]

[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 are top-coded at 63; and 0.01% of the districts in our sample have an average nighttime light pixel values above 53. We nevertheless increase the average nighttime light pixel value of these districts by 10 as well.

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

[36]Table O1 in the Online Appendix lists the Top 10 districts with the largest overall impact from this counterfactual increase in economic activity and compares the ranks from this policy exercise to the

[Figure 3 about here]

There are various types of districts where the overall impact is particularly high. First, for both Nigeria and Kenya, the districts with the highest impact also have high key-player centrality because they are economically active and well connected, such as the districts from Lagos State. Second, in Nigeria, the top districts includes some districts in Bayelsa and Delta, which are both oil-producing states in the Niger Delta. These districts are economically quite active and well connected but have a low key-player centrality because they are conflict-ridden. An increase in economic activity, however, has a positive impact exactly because of the dense network in the Niger Delta. Third, in Kenya, the districts with the biggest impact includes three 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.

8.2 Policy experiment 2: Improving road connectivity

The second policy experiment consists in 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 contiguous districts with which the given district is not yet linked via a major road, and we then 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 base-

district's overall key-player rank. Table P1 in the Online Appendix presents the same ranking for the five most populous African countries besides Nigeria.

line. We repeat this exercise for each of the 5,944 districts to identify, for each country, the districts that have the largest overall impact when improving their road connectivity.

Figure 4 maps the results of this policy experiment for districts in Nigeria (left panel) and Kenya (right panel). Again, districts where a counterfactual improvement of road connectivity would create a stronger overall increase in nighttime light are presented by darker colours.[37]

[**Figure 4 about here**]

The top districts in Nigeria are all in the Niger Delta. The top two, Boony and Orika, are both islands with intense nighttime lights but poor road connectivity. Improving their road connectivity would lead to positive economic spillovers from these two districts to other districts in the Niger Delta and beyond. The districts with the strongest overall impact in Kenya again include many districts with high key-player centrality. In addition, the list includes some very dark/poor districts (Machakos, Wajir and Meru), where an increase in economic activity from better road connectivity would be particularly valuable. Along similar lines, better road connectivity would also be valuable in may dark/poor districts in North Eastern Nigeria.

9 Concluding remarks

In this paper, we study the role of geographical, ethnic and road networks for the spatial diffusion of local economic shocks. We first develop a simple network model that describes how a district's prosperity is determined by its own economic activity, its observable and unobservable characteristics, and the spillover effects of the economic activity of neighboring districts. Using a panel dataset of 5,944 districts from 53 African countries over the period 1997–2013, we estimate the model's first-order conditions using an econometric model of spatial spillovers. To identify the causal effect of spatial diffusion, we exploit

[37]Table O2 in the Online Appendix lists the Top 10 districts with the largest overall impact from this counterfactual increase in economic activity and compares the ranks from this policy exercise to the district's overall key-player rank. Table P2 in the Online Appendix presents the same ranking for the five most populous African countries besides Nigeria.

cross-sectional variation in the location of mineral mines and time variation in world mineral prices.

We then use our simple model and the estimated parameters to calculate various network centrality measures for each district. In particular, we calculate the key-player centralities by performing a counterfactual exercise, which consists 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 indeed key for a country's economic development due to their high economic activity and good connectivity, suggesting that policies focusing on the major cities are justified from a growth perspective.

We go one step further by conducting two counterfactual policy exercises based on our estimates and the underlying network structure. The first looks at the aggregate effects of increasing economic activity in each district, one at a time, and the second at the aggregate effects from improving each district's road connectivity. These counterfactual exercises illustrate the potential of our approach for informing policymakers.

References

Acemoglu, D., Carvalho, V.M., Ozdaglar, A. and A. Tahbaz-Salehi (2012), "The network origins of aggregate fluctuations," *Econometrica* 80, 1977–2016.

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 (2015), "Networks and the macroeconomy: an empirical exploration," *NBER Macroeconomics Annual* 30, 273–335.

Alder, S. (2015), "Chinese roads in India: the effect of transport infrastructure on economic development," *Mimeo*, University of North Carolina.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and R. Wacziarg (2003), "Fractionalization," *Journal of Economic Growth* 8, 155–194.

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.

Alesina, A. and E. Zhuravskaya (2011), "Segregation and the quality of government in a cross section of countries," *American Economic Review* 101, 1872–1911.

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

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

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

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

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

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

Barrot, J.-N. and J. Sauvagnat (2016), "Input specificity and the propagation of idiosyncratic shocks in production networks," *Quarterly Journal of Economics* 131, 1543-–1592.

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*, 114–130.

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

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

Bloom, D.E., Sachs, J.D., Collier, P., and C. Udry, (1998), "Geography, demography, and economic growth in Africa," *Brookings Papers on Economic Activity* 2, 207–295.

Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), "Identification of social interactions," In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Science, pp. 853–964.

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

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., Djebbari, H. and B. Fortin (2009), "Identification of peer effects through social networks," *Journal of Econometrics* 150, 41–55.

Bramoullé, Y., Rogers, B.W. and A. Galeotti (2016), *The Oxford handbook of the economics of networks*, Oxford: Oxford University Press.

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.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A. and G. Padro i Miquel (2015), "The value of democracy: Evidence from road building in Kenya," *American Economic Review* 105, 1817–1851.

Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009), "Peer effects and social networks in education," *Review of Economic Studies* 76, 1239–1267.

Carvalho, V.M. (2015), "From micro to macro via production networks," *Journal of Economic Perspectives* 28, 23–48.

Cavalli-Sforza, L.L., Menozzi, P. and A. Piazza (1994), *The History and Geography of Human Genes*, Princeton: Princeton University Press.

Cawley, J., Han, E., Kim, J. and E.C. Norton (2017), "Testing for peer effects using genetic data," NBER Working Paper No. 23719.

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

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

Crespo Cuaresma, J., Doppelhofer, G. and M. Feldkircher (2014), “The determinants of economic growth in European regions,” *Regional Studies* 48, 44–67.

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.

Dimico, A. (2017), “Size matters: The effect of the size of ethnic groups on development,” *Oxford Bulletin of Economics and Statistics*, 79, 291–318.

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

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

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

Easterly, W. and R. Levine (1997), “Africa’s growth tragedy: Policies and ethnic divisions,” *Quarterly Journal of Economics* 122, 1203–1250.

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

Franck, R. and I. Rainer (2012), “Does the leader’s ethnicity matter? Ethnic favoritism, education and health in Sub-Saharan Africa,” *American Political Science Review* 106, 294–325.

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

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

Gibbons, S. and H.G. Overman (2012), “Mostly pointless spatial econometrics,” *Journal* *of Regional Science* 52, 172–191.

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

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.

Higgins, M.J., Levy, D. and A.T. Young (2006) "Growth and convergence across the United States: evidence from county-level data," *Review of Economics and Statistics* 88, 671–681

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

Hodler, R. and P.A. Raschky (2017), "Ethnic politics and the diffusion of mobile technology in Africa," *Economics Letters* 159, 78–81.

Hodler, R., Valsecchi, M. and A. Vesperoni (2017), "Ethnic geography: Measurement and evidence," CEPR Discussion Paper 12378.

Hsiang, S. (2010), "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America," *Proceedings of the National Academy of Sciences of the United States of America* 107, 15367–15372.

Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.

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

Jackson, M.O. (2014), "Networks in the understanding of economic behaviors," *Journal of Economic Perspectives* 28, 3–22.

Jackson, M.O. and Y. Zenou (2015), "Games on networks," In: Young, P. and S. Zamir (Eds.), *Handbook of game theory Vol. 4*, Amsterdam: Elsevier Publisher, pp. 91–157.

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

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

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

Kelejian, H.H. and D.P. Robinson. (1993), "A suggested method of estimation of spatial interdependent models with autocorrelated errors, and an application to a county expenditure model," *Papers in Regional Science* 72, 297–312.

Kelejian, H.H. and I.R. Prucha. (1998), "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *Journal of Real Estate Finance and Economics* 17, 99–121.

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.

Lee, L.-F. (2004), "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica* 72, 1899–1925

Mamo, N., Bhattacharyya, S., Moradi, A. and R. Arezki (2017), "Intensive and extensive margins of mining and development: evidence from Sub-Saharan Africa," CSAE Working Paper 2017-05.

Masanjala, W. H. and C. Papageorgiou (2008), "Rough and lonely road to prosperity: a reexamination of the sources of growth in Africa using Bayesian model averaging," *Journal of Applied Econometrics* 23, 671–682.

Meyer, C.D. (2000), *Matrix Analysis and Applied Linear Algebra*, Philadelphia, PA: American Mathematical Society.

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.

Michalopoulos, S. and E. Papaioannou (2017), "Spatial patterns of development: A meso approach," NBER Working Paper 24088.

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

Rodney, W. (1982), *How Europe underdeveloped Africa*, Washington DC: Howard University Press.

Sachs, J. and A. Warner (2001), "The curse of natural resources," *European Economic Review* 45, 827–838.

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

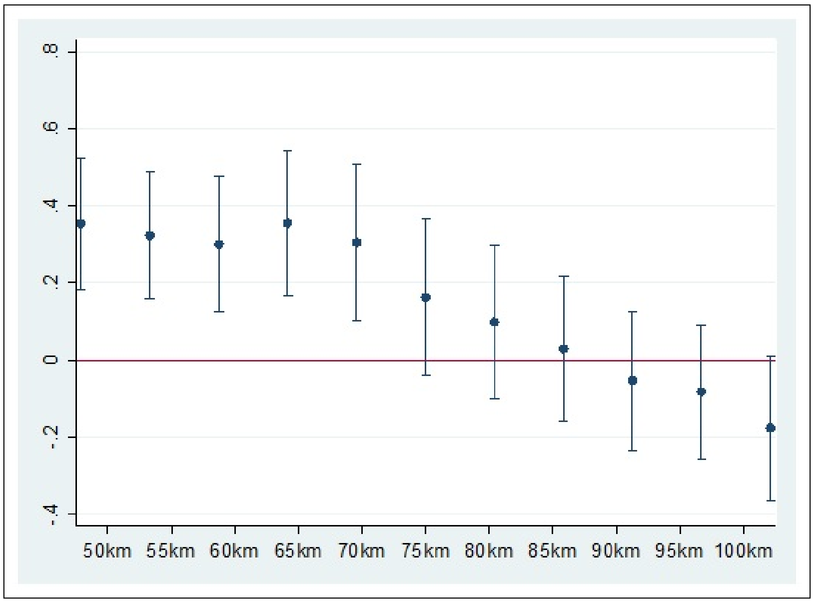
USGS (2014), "*Minerals yearbook*," Retrieved from https://minerals.usgs.gov/minerals/pubs/myb.html.

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

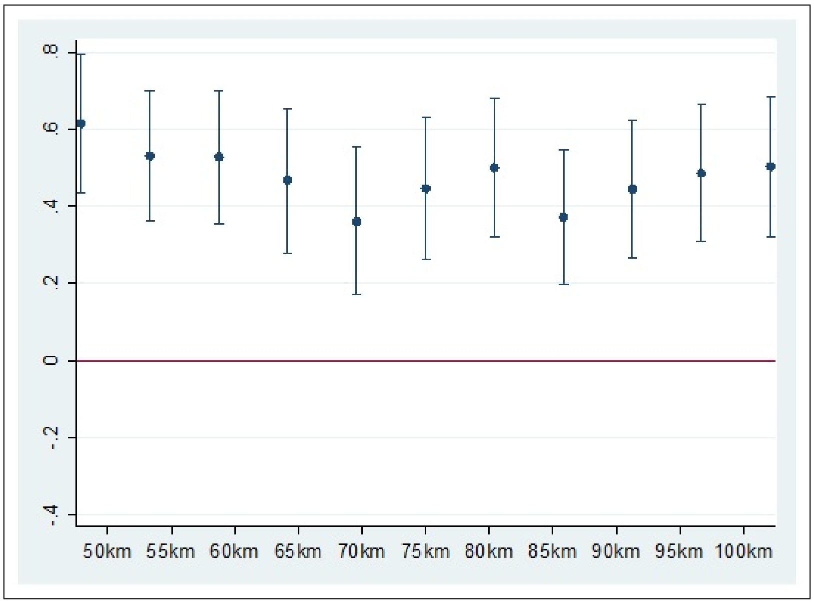
Zenou, Y. (2016), "Key players," In: Y. Bramoullé, B.W. Rogers and A. Galeotti (Eds.), *The Oxford Handbook of the Economics of Networks*, Oxford: Oxford University Press, pp. 244–274.

Figures and Tables

Figure 1: Coefficients on the Spatial Lag of the Dependent Variable



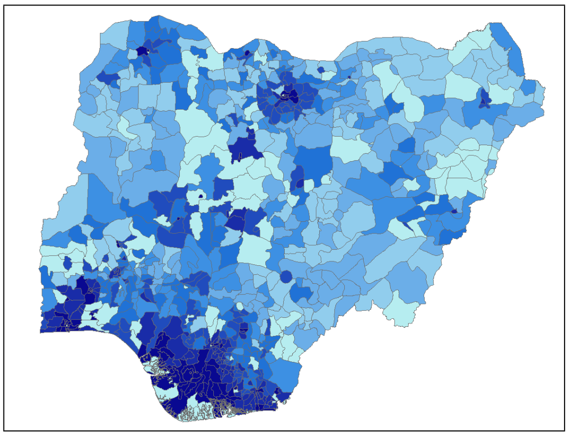
(a) Coefficient on $Inv\,Dist\,W\,Light_{jct}$



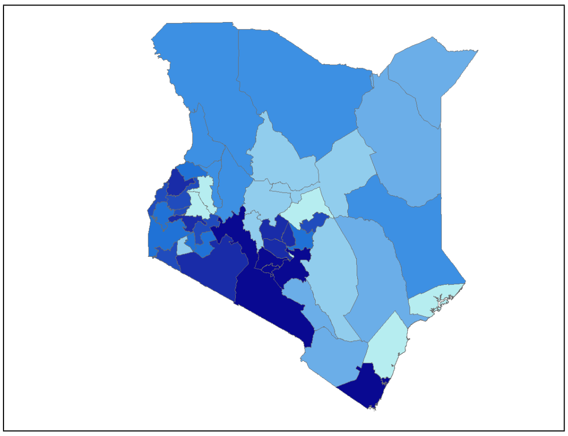
(b) Coefficient on $Inv\,Road\,W\,Light_{jct}$

Notes: Dots show the coefficients on $Inv\,Dist\,W\,Light_{jct}$ and $Inv\,Road\,W\,Light_{jct}$ from the second-stage regression reported in Table 3, column (8), when applying different distance cutoffs for the weighting matrices for geographic and road connectivity. The vertical lines show the 90% confidence interval based on standard errors clustered along the relevant network.

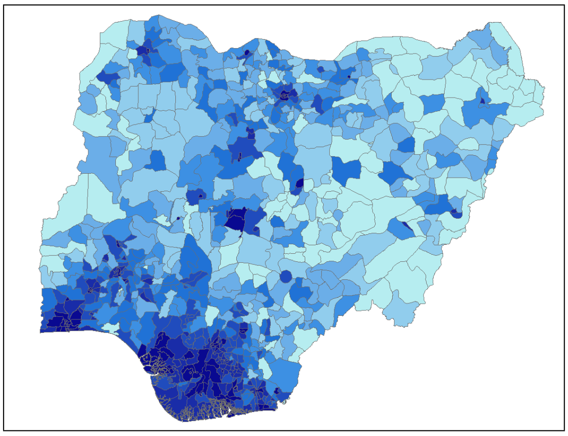
Figure 2: Key-player Centrality, Nighttime Light Intensity, and Population Density in Nigeria and Kenya



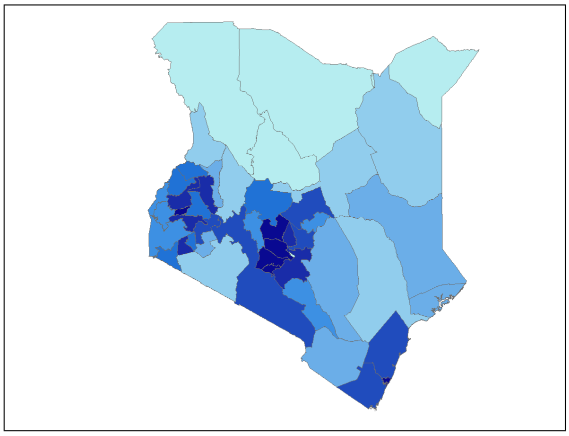
Key-player centrality across Nigeria



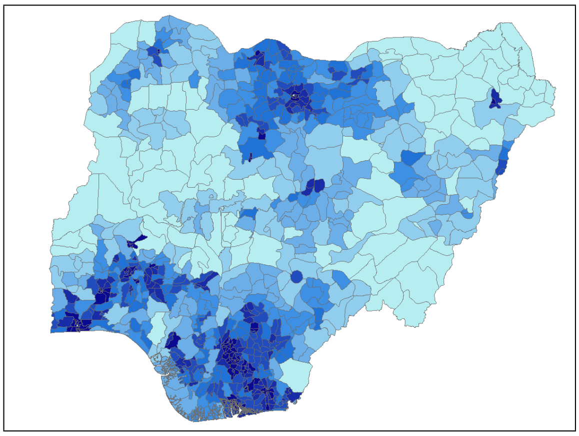
Key-player centrality across Kenya



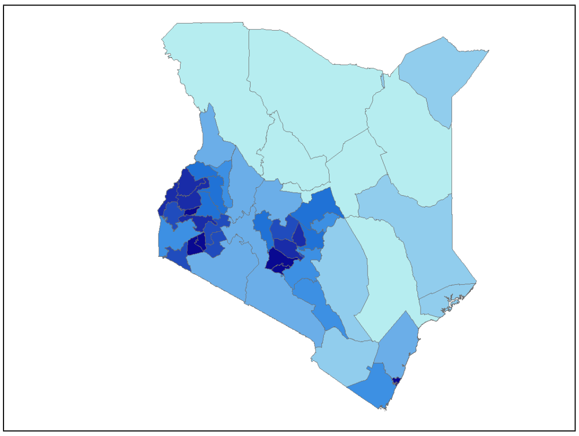
Average nighttime lights across Nigeria



Average nighttime lights across Kenya



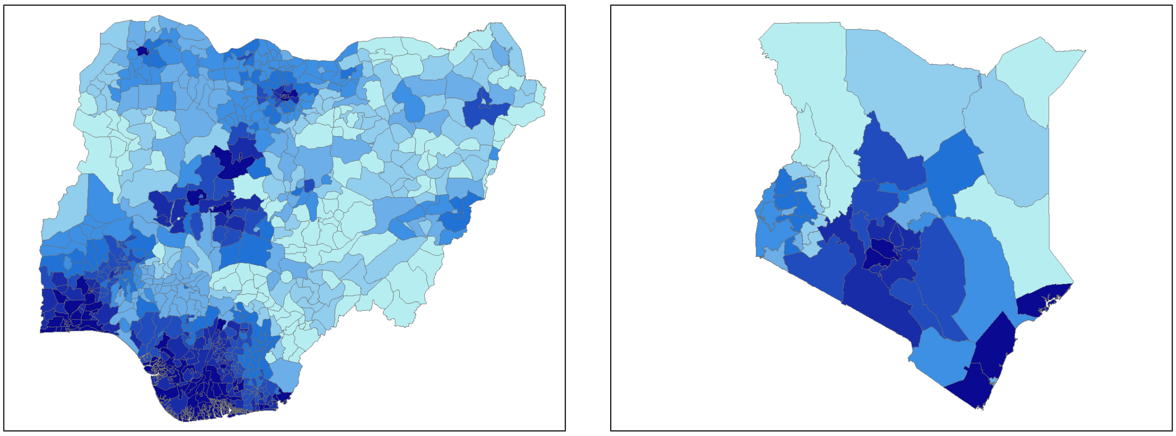
Population density across Nigeria



Population density across Kenya

Notes: This figure compares districts' key-player centrality (top row) with the average nighttime light intensity (middle row) and population density (bottom row) for Nigeria (left column) and Kenya (right column). Darker colors indicate higher values.

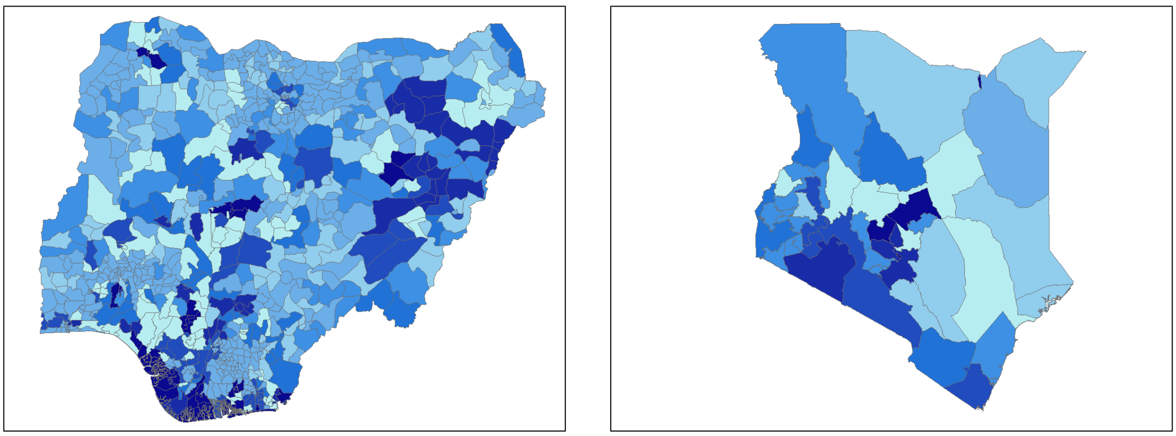
Figure 3: Policy Experiment 1: Counterfactual Increase in Economic Activity



Nigeria Kenya

Notes: This figure shows the overall increase in nighttime light intensity from a counterfactual increase in economic activity in the respective Nigerian (left) and Kenyan (right) districts. Darker colors indicate higher overall impact.

Figure 4: Policy Experiment 2: Counterfactual Improvement in Road Connectivity



Nigeria Kenya

Notes: This figure shows the overall increase in nighttime light intensity from a counterfactual improvement in the respective district's road connectivity for Nigerian (left) and Kenyan (right) districts. Darker colors indicate higher overall impact.

Table 1: Correlation Between Connectivity Matrices

	(1) *Eth W* $Light_{jct}$	(2) *Inv Dist W* $Light_{jct}$	(3) *Inv Road W* $Light_{jct}$
Eth W $Light_{jct}$	1.000		
	(1.000)		
Inv Dist W $Light_{jct}$	0.341	1.000	
	(0.379)	(1.000)	
Inv Road W $Light_{jct}$	0.382	0.410	1.000
	(0.412)	(0.429)	(1.000)

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

Table 2: Descriptive statistics

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

Notes: See Section 4 for the definitions of all variables. Note that $Light_{ict}$, $Population_{ict}$ and MP_{ict} are in logs.

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

Notes: Even columns report standard fixed effects regressions with district and country-year fixed effects, and odd columns report IV estimates. See Section 4 for the definitions of all variables. $Ethnicity\, W\, Light_{jct}$ is weighted $Light_{jct}$, with weights based on the row-normalized ethnicity matrix. $Inv\, Dist\, W\, Light_{jct}$ ($Inv\, Road\, W\, Light_{jct}$) is weighted $Light_{jct}$, with weights based on the row-normalized matrix of the inverse altitude-adjusted geodesic distances (inverse road distances) truncated at 70km. Additional control variables are population, conflict and MP_{ict} as well as weighted population and conflict in districts $j \neq i$. MP_{jct} is an interaction term based on cross-sectional information on the location of mines and time-varying world prices of the commodities produced in these mines (see equation (11)). The first stage further includes the control variables indicated in equation (10). Standard errors, clustered at the network level, are in parentheses. ***, **, * indicate significance at the 1, 5 and 10% level, respectively.

Table 4: Top-Ten Key Player Rankings

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

Notes: Overall KP Rank is based on the ρs estimated in column (8) of Table 3. Overall Katz-Bonacich Rank is based on the ρs estimated in column (8) of Table 3 and a weighting vector of 1. Ethnicity KP Rank is based on ρ_1 estimated in column (2) of Table 3. Inv.Dist KP Rank is based on ρ_2 estimated in column (4) of Table 3. Road KP Rank is based on ρ_3 estimated in column (6) Table 3. * indicate districts that are (part of) capital cities. Nigeria has 775 districts, and Kenya has 48 districts.