

Social Networks and the Political Salience of Ethnicity

Nicholas Eubank^{1*}

¹*Center for the Study of Democratic Institutions, Vanderbilt University, Nashville, TN, USA; nick@nickeubank.com*

ABSTRACT

Ethnic politics scholars are increasingly convinced that (a) the political salience of ethnicity and (b) the correlation between ethno-linguistic fractionalization (ELF) and poor development are driven by the dense social networks shared by co-ethnics. By this argument, social networks allow ethnic parties to leverage inbuilt networks to share information and support collective action, while ethnically fragmented communities struggle to hold politicians accountable. This paper provides the first comprehensive empirical test of the assumption underlying this argument. Using seven months of telecommunications data from 9 million mobile subscribers in Zambia — which includes records of almost 2 billion calls and SMS messages — to measure social networks across an entire country, this paper finds that electoral constituencies with high ELF also have more fragmented social networks, especially in rural areas. It also finds other potential cleavages that have not achieved political salience (namely, religious identity and employment sector) are not correlated with network fragmentation, consistent with the idea that ethnicity achieves salience because it offers an organizational advantage not offered by other cleavages. Finally, it also finds that both voter knowledge and public goods are negatively correlated with network fragmentation, consistent with the network-proxy hypothesis.

*This project would not have been possible without the generosity of numerous parties, including Real Impact Analytics, Real Impact's Chief Data Scientists Gautier Krings, and its CEO Sébastien Deletaille. The author is also indebted to Ramya Parthasarathy, Ernest

Online Appendix available from:

http://dx.doi.org/10.1561/100.00017044_app

Supplementary Material available from:

http://dx.doi.org/10.1561/100.00017044_supp

MS submitted on 15 March 2017; final version received 30 July 2018

ISSN 1554-0626; DOI 10.1561/100.00017044

© 2019 N. Eubank

Keywords: Political networks; ethno-linguistic fractionalization; ethnic politics; collective action

Research on the politics of ethnicity is dominated by efforts to explain two prominent empirical regularities. First, across a broad range of regions and institutional contexts, ethnicity consistently achieves a high degree of political salience (e.g. Basedau *et al.*, 2011; Chandra, 2007; Horowitz, 1985; Keefer, 2010; Laitin, 1998; Posner, 2005). This is true not only in terms of political rhetoric, but also political party organization and voting behavior. And second, ethnically fragmented communities tend to experience worse development outcomes and low investment in public goods (e.g. Alesina *et al.*, 1999; Alesina and La Ferrara, 2005; Easterly and Levine, 1997; Habyarimana *et al.*, 2007; Khwaja, 2009; Miguel and Gugerty, 2005).

In search of explanations for these regularities, scholars increasingly suggest that the salience of ethnicity is not driven by ethnicity per se, but rather social network structures that are assumed to be closely related to ethnicity.¹ Miguel and Gugerty (2005, p. 2330), for example, argue “that social sanctions and coordination are possible within [ethnic] groups due to the dense networks of information and mutual reciprocity that exist in groups but are not possible across groups.” Similarly, Habyarimana *et al.* (2009) argue that if co-ethnics are more well connected in terms of their social networks, then co-ethnics may be more “findable” and thus subject to social sanctioning. And in the literature on ethnic violence, Fearon and Laitin (1996, p. 719) assert that ethnic groups are “often characterized by relatively dense social networks,” and that “across groups [...] social networks are less developed” so that “it is more difficult to get information on a potential trading or social partner from ‘across the tracks.’”

Mundia, Neil Malhotra, Katherine Casey, Jim Fearon, Kate Baldwin, the Zambian Central Statistics Organization, Adriane Fresh, Vincent Traag, Jeremy Weinstein, Jonathan Rodden, Viji Jagannathan and the Stanford GSB Center for Global Business and the Economy, Donovan Neale-May, Jure Leskovec, Kate Baldwin, Romain Wacziarg, Tamás Nepusz, James Moody, and participants in the Stanford GSB Organizational Behavior Seminar, the 2015 WGAPE Meeting at Brown University, and Duke Geospatial Data, Governance, and the Future of Development Aid conference, and numerous anonymous referees. Financial support provided by a Stanford Vice-Provost for Graduate Education Diversity Dissertation Research Opportunity grant, a Stanford GSB PhD Research grant, a Stanford Center for Philanthropy and Civil Society (PACS) Fellowship, and a Dolores Zohrab Liebmann Fellowship. Additional acknowledgments in Appendix V.

**Although the QJPS replicates every article and is committed to making replication material available, this article makes use of sensitive data that cannot be publicly shared. The results were replicated by our replication assistant, but the data is not included in the replication files available on the QJPS website. The associated replication file contains the code needed for scholars with access to the data to fully replicate the reported result.

¹Note that this literature does not tend to take a strong position on *why* ethnicity and network structure are closely related. These explanations are consistent with the possibility that values or other factors related to ethnicity shape network structure, which in turn drives political salience.

The political networks literature — which documents ways in which network structure shapes capacity for political action — provides additional support for this explanation. Well-connected social networks may help communities overcome collective action problems associated with getting constituents to political rallies or protests (Habyarimana *et al.*, 2009; Jackson *et al.*, 2012), and to turnout on election day (Cox, 2015; Rojo and Wibbels, 2014; Siegel, 2009). Well-connected networks may also support information diffusion (Larson, 2017b; Larson and Lewis, 2017),² making it easier for politicians to learn what citizens want and develop appealing policy platforms, and for citizens to monitor elected officials and hold them accountable. Well-connected networks may also help voters coordinate their support around a candidate, and avoid vote-splitting.

Taken together, the link between ethnicity and network structure, on the one hand, and network structure and capacity for political activity on the other have the potential to explain both empirical regularities noted above — the political salience of ethnicity and the tendency for ethnically fragmented communities to experience worse development outcomes. If ethnic groups share well-connected networks that support political action, for example, then political parties that organize along ethnic lines will inherit a free organizational advantage over non-ethnic parties. This would make ethnic parties more competitive than non-ethnic parties. And if ethnically fragmented communities lack networks that support collective political behaviors, they would also have difficulty holding politicians accountable, leading to shirking, underinvestment in public goods, and poor development outcomes.

But while this *network-proxy hypothesis* has generated substantial excitement in the ethnic politics literature in recent years (e.g. Dionne, 2015; Habyarimana *et al.*, 2007, 2009; Larson, 2017a; Miguel and Gugerty, 2005; Rojo and Wibbels, 2014), the motivating assumption of this argument — that ethnically fragmented groups have more fragmented social networks — has never been systematically tested. Instead, most existing work (like Miguel and Gugerty (2005) and Fearon and Laitin (1996) quoted above) simply assumes this relationship to exist. And while there are reasons to believe this may be the case, much of the evidence cited in support of this assumption — like evidence of ethnic homophily — is not sufficient to substantiate it (as discussed in Section “Measuring Social Network Fragmentation”). As a result, this growing literature rests on a dangerously untested foundation.

This paper fills that gap by providing the first comprehensive test of this literature’s underlying assumption: that ethnically fragmented communities have more fragmented social networks. I use seven months of detailed mobile telecommunications data to measure the structure of social networks in each of Zambia’s 150 National Assembly electoral districts. In particular, I use

²Though Larson and Lewis (2017) offer a theoretical nuance to this argument, along with preliminary empirical evidence suggested where links differ qualitatively, density of ties may not be monotonically associated with faster information diffusion.

this data — which includes records of almost 2 billion calls and SMS (text) messages — to construct a measure of network fragmentation for each district. By pairing this measure with geo-coded census data on the ethnic composition of districts, this paper systematically examines the relationship between ethnic fractionalization and network fragmentation across the universe of electoral constituencies in Zambia.

In doing so, this analysis joins three other studies aiming to test this assumption. First, Larson and Lewis (2017) map social networks in two villages in Uganda and find that networks are denser in an ethnically heterogeneous village, though they also find that information diffuses more effectively in a homogeneous village. Dionne (2015) maps networks in four villages of Malawi and finds cross-ethnic ties to be as common as intra-ethnic ties. And finally Habyarimana *et al.* (2009) find that lab subjects find random strangers more quickly if they are co-ethnics, suggesting that co-ethnics have more well-connected networks. Unlike past studies, however, this analysis is not limited by the logistical constraint of studying a relatively small number of geographically-confined networks, or to measuring small-scale network properties. Instead, using cell-phone meta-data this analysis is able to measure the social networks of *an entire country*, and study variation in network structure at the most *politically-relevant* scale for theories of party organization and political accountability — at the scale of electoral constituencies.

This analysis finds that ethno-linguistic fractionalization (ELF) and network fragmentation are generally positively correlated, as predicted by the network-proxy hypothesis. This result is especially strong in rural constituencies, suggesting that urbanization may diminish the salience of ethnicity. It also finds that network fragmentation is *not* correlated with other social divisions that have failed achieved political salience in Zambia, like religious and economic divisions, consistent with the idea that ethnicity offers organizational advantages to voters and party leaders that other organizing cleavages do not. And finally, this analysis also examines whether network fragmentation is associated with lower voter knowledge and public goods provision, as the network-proxy hypothesis suggests. It is, albeit not as strongly as network fragmentation is associated with ELF. Finally, as detailed in Section “Public Goods and Voter Knowledge”, the interrelationship of public goods, voter knowledge, ELF, and network fragmentation is also consistent with the network-proxy hypothesis.

Throughout this analysis, every effort is made to provide an authentic test of the network-proxy hypothesis as put forth by past scholars. The measure of network fragmentation used in this analysis, for example, is constructed to measure the properties of networks past scholars have suggested should vary with ethnic composition as closely as possible. Similarly, ethnic fractionalization

is measured using the canonical measure of ethnic fractionalization used to establish the empirical regularities that the network-proxy hypothesis aims to rationalize — ELF. And finally, ELF is calculated with respect to a dimension of identity that has the key features of an ethnic cleavage that the network-proxy hypothesis aims to explain: it is based on ascriptive features that are not easily chosen or changed by individuals (thus meeting a standard definition of ethnicity (Fearon, 1999)) and it is a politically salient.

These results constitute the first direct validation of the assumption that ethnic fragmentation is associated with network fragmentation. This both affirms an untested yet critical assumption underlying numerous past studies — improving our confidence in those results — and also provides a firm foundation for future researchers that the motivating assumption of this literature is well founded. Moreover, in showing that voter knowledge and public goods tend to be negatively correlated with network fragmentation, it provides further support for the network-proxy hypothesis. And finally, in finding that this relationship is particularly strong in rural communities, this research suggests a potential avenue for future research into the causes of the attenuation of the network–ethnicity relationship in urban communities, causes with may also have implications for efforts to diminish the political salience of ethnicity more broadly.

The validation of this assumption was by no means a forgone conclusion. Indeed, many other promising hypotheses aimed at explaining the salience of ethnicity that were premised on seemingly self-evident assumptions have struggled in the face of subsequent empirical interrogation. For example, the theory that ethnic coalitions reflect shared policy preferences (and therefore that bad outcomes in ethnically fragmented communities stem from policy disagreements (Alesina *et al.*, 1999)) has been met with recent studies that show policy preferences do *not* appear to vary dramatically across ethnic groups (e.g. Desmet *et al.*, 2015; Habyarimana *et al.*, 2009. See Lieberman and McClendon (2012) for a dissenting view), and ethnic parties do not appear to be structured to maximize policy influence (Keefer, 2010). The idea that people simply care more about co-ethnics than non-co-ethnics (Charness and Rabin, 2002; Chen and Li, 2009) and so under-invest in public goods that also benefit non-co-ethnics (Ejdemyr *et al.*, 2015) also does not square with laboratory results that show that participants in anonymous altruism games seem just as generous to co-ethnics as non-coethnics (Berge *et al.*, 2015; Dionne, 2015; Habyarimana *et al.*, 2009). And finally, the idea that ethnicity is strategically valuable because the recognizability and immutability of ethnicity allow minimal winning coalitions to ensure new-comers do not sneak in to claim benefits (Fearon, 1999) has been challenged by evidence that in many African contexts ethnicity is not that easy to identify in strangers (Casey, 2016; Habyarimana *et al.*, 2009; Harris and Findley, 2012).

Ethnicity in Zambian Politics

Ethnicity has always been a core organizing concept of political debate in Zambia. Today, the political salience of ethnicity is evident not only in aggregate voting patterns (Erdmann, 2007; Posner, 2005) but also in individual-level survey data. Using data from a 1996 post-election survey, Posner and Simon (2002) conclude that while economic conditions do matter, “ethnicity and urban/rural location explain the lion’s share of the variance in patterns of support for the incumbent regime” (Posner and Simon, 2002, p. 329). Similarly, using data from the 2004 wave of Afrobarometer, Basedau *et al.* (2011) examine the relationship between ethnicity and vote choice. They conclude that while ethnicity is not a strong predictor of political preferences in all African countries, Zambia is subject to “comparatively strong or medium to strong ethnicization of party politics,” (Basedau *et al.*, 2011, p. 467) based on their ability to predict vote choice using ethnicity. Indeed, they note that ethnicity was a significant predictor of voting preferences for *all* political parties they examined (MMD, UPND, UNIP, and PF). And finally, Erdmann (2007) reaches similar conclusions based on data from focus groups and field surveys: “ethnicity can still be viewed as the major factor explaining party affiliation in Zambia, and to a lesser degree voter alignment.” (p. 30). The same author reports that in focus groups, “most of the participants could not detect any programme or policy difference between the parties. And the few who said they could detect differences were, when directly challenged, almost all unable [sic] to name any difference” (Erdmann, 2007, p. 23).

Despite agreement on the political salience of ethnicity in Zambia, the question of *why* ethnicity is politically salient remains largely unanswered. As Laitin (1998, p. 248) has argued, “ethnic entrepreneurs cannot create ethnic solidarities from nothing. They must, if they are to succeed, be attuned to the micro-incentives that real people face.” So what micro-incentives are at play in Zambia that make ethnicity so preferable to a different cleavage?³ Class differences, for example, are widely noted in Zambia and have clear policy implications, and yet have never achieved political salience (Chikulo, 1988; Posner, 2005).

³Posner (2005) provides a partial answer to this question in the Zambian context, arguing that strategic political considerations drive the specific dimensions of identity that become politically salient. Yet, this explanation begins from the assumption that citizens will mobilize along *some* dimension of identity. The one exception to this oversight comes in a discussion of why Zambians do not organize along class lines. Here, Posner (2005, p. 86) argues “[t]he answer is that, by and large, class identities are not sufficiently deeply felt for them to play this role.” This implies that it is the emotional salience of ethnicity that leads to its viability as a political cleavage, implicitly supporting the social-psychological mechanism discussed below.

Explaining the Salience of Ethnicity

The network-proxy hypothesis provides one potential explanation. If ethnicity is closely tied to network structure, then ethnically aligned constituencies may be better able to overcome the collective action problem, coordinate activities, and share information.

Consider, first, the collective action problem. All supporters (or opponents) of a candidate or cause benefit from rallies and protests whether they participate or not, so many individuals may prefer to free-ride on the attendance of others, even if they support the cause. Social pressure on non-contributing group members can solve this problem (Ostrom, 1990). Jackson *et al.* (2012) and Wolitzky (2012) have shown that the ability to apply social pressure is a function of network connectedness.⁴

Social networks may also help to overcome coordination problems, like rallying voters around a specific candidate (Siegel, 2009). When more than two candidates from a given community stand for election, avoiding vote-splitting is critical to political effectiveness. Moreover, well-connected social networks may also help overcome mundane logistical challenges like learning about voter preferences, coordinating on objectives, or scheduling and publicizing political events.

In addition, social networks may help facilitate information diffusion, which has implications for electoral accountability. Theoretical and empirical research has consistently found that citizens are only able to hold politicians accountable when they know what those politicians are doing (e.g. Besley, 2007; Ferraz and Finan, 2008; Reinikka and Svensson, 2004, among many others). Citizens cannot reward good behavior and punish poor behavior if they cannot observe politician behavior in the first place. Moreover, poor information diffusion also makes the application of social pressure more difficult (Fearon and Laitin, 1996; Larson, 2017b).

These factors help to explain why political entrepreneurs may choose to organize along ethnic lines as opposed to along other cleavages: ethnic coalitions come with inbuilt networks that enhance political effectiveness. Moreover, this advantage is likely to be especially acute in developing countries, where political parties tend to be poorly institutionalized and underfunded. Parties organized along other dimensions can develop their own machinery, of course, but doing so diverts scarce resources from other political activities, undermining competitiveness, and making this option less desirable.

These factors also help to explain why *citizens* may be willing to mobilize along ethnic lines. By joining ethnic coalitions, voters access effective political organizations that can better share information and sanction politicians (through protest or coordination around alternative candidates), likely leading to better governance.

⁴Studies of voting also suggest social pressure increases turnout (Gerber *et al.*, 2008).

And finally, these factors may also explain why ethnically fragmented communities have difficulty holding politicians accountable, resulting in lower investment in public goods and poor development outcomes.

These insights help to explain why, if ethnicity is closely tied to network structure, it makes sense for ethnicity to achieve such political salience in various settings. Despite the promise of this explanation, however, the network-proxy hypothesis is arguably the least-tested explanation for ethnic salience — a shortcoming remedied here.

Measuring Social Network Fragmentation

Existing work has argued that co-ethnics tend to be well connected to one another, and that they tend to have more connections with one another than with non-co-ethnics (Fearon and Laitin, 1996; Habyarimana *et al.*, 2009; Miguel and Gugerty, 2005). With that in mind, this section develops a measure of network fragmentation that reflects this characterization as closely as possible in three steps.

First, I construct a social network from cell-phone meta-data. Each cell-phone subscriber forms a node or vertex in this network, and connections are placed between individuals who call or text one another. This process is discussed in more detail in Section “Data”.

Second, this social network of all Zambian cell-phone subscribers is partitioned into groups so that (a) the members of each group are as well connected to one another as possible, and (b) there are as few connections running *between* groups as possible. If current characterizations of ethnic networks are accurate, then these groups — called *network communities* — should roughly reflect ethnic divisions.

This partitioning is based only on patterns of interaction. The network will be partitioned along whatever lines are most clearly defined in the network. If, as network-proxy theorists have argued, ethnicity is the dominant determinant of network divisions (and that it is for this reason that ethnicity is more likely to achieve political salience than other potential cleavages), then these network communities should mirror ethnic divisions. However, if other divisions (like class) play a larger role in shaping patterns of interaction, then these partitions will not reflect ethnic divisions.

In the third and final step, I compute a measure of network fragmentation at the level of electoral constituencies. More specifically, I calculate network fragmentation as one minus the Herfindahl index of network community assignments for residents of an electoral constituency. This measure is perfectly analogous to ELF, which is computed as one minus the Herfindahl index of individuals’ ethnicities. Indeed, if the network has been partitioned into

network communities that parallel ethnic divisions, then these two measures should be roughly equivalent and thus highly correlated in subsequent analyses.

Note that in the second step, the partitioning algorithm is applied to the global network of Zambia; that is, it does *not* take into account information about the spatial distribution of users. Only after users are partitioned is information about user residency used. I employ this *global partition strategy* rather than the alternative — a *constituency partition strategy*, in which sub-networks consisting only of the residents of each electoral constituency are created and partitioned — for several reasons. First, it best utilizes all of the available data. As discussed in more detail below, while this analysis makes use of network meta-data for all Zambian cell-phone subscribers, only subscribers of the Partner Telecom (PT) who have provided the cellphone meta-data used in this analysis can be geo-referenced. Thus as non-PT subscribers could not be assigned to a constituency, a constituency partition strategy that first subsets on residents of a constituency would require dropping a significant portion of the available network data. This would not only reduce the statistical power of estimates of network structure, but also likely lead to biased estimates, as most aspects of network topology are not preserved under sampling, even when that sampling is i.i.d. (Kolaczyk, 2005; Leskovec and Faloutsos, 2006). A global partition, by contrast, comes much closer (as close as possible in this context) to capturing the full network.

Second, the global partition strategy allows me to capture cross-constituency and cross-village ties which can provide important information about the strength of social ties. For example, suppose i and j live in Constituency 1, and are both friends with a , b , and c , who live in Constituency 2. A global partitioning strategy allows for these out-of-constituency ties to increase the likelihood that i and j will be assigned to the same network community. A constituency partitioning strategy, by contrast, would ignore mutual connections to a , b , and c .

And finally, the global partition strategy allows me to examine the spatial distribution of members of each network community as a test of measure validity, as detailed in Section “Measure Validation”.

It is important to emphasize that the measure of network fragmentation created through these three steps is qualitatively different from measures used in most studies of ethnicity and networks. Due primarily to data constraints, most studies rely on *local* network measures (like *homophily*) that only consider individuals and their immediate connections. The network fragmentation measure used in this analysis, by contrast, takes into account not only local network features, but also *meso-scale* network structure: the structure of the network at the level of medium to large groups (Jackson *et al.*, 2017). As most network-proxy theories are about group-level dynamics, this is critical to properly testing the network-proxy hypothesis.

Assignment to Network Communities

Individual subscribers are partitioned into *network communities* using a Constant Potts Model (CPM). As described in detail in Appendix D, the CPM assigns each subscriber to one and only one network community so as to maximize the density of links within each network community and minimize the number of links running between communities.

While most technical aspects of the CPM can be relegated to appendices, there is one parameter in the model that it is critical to discuss. When assigning individuals to network communities, the CPM attempts to meet two conflicting criteria: (a) maximize the density of links inside each community, and (b) minimize the number of links between groups. The more the model attempts to maximize (a), the more it will assign individuals to small, internally-densely connected groups; the more the model attempts to minimize (b), the more it will assign individuals to large, internally-loosely connected groups. As such, the model requires specification of a *resolution parameter* (commonly denoted γ) that determines the trade-off between these two goals.

Traditionally, specification of the resolution parameter is treated as a nuisance and an estimation challenge (referred to as the *problem of scale*). It is addressed by choosing the value of γ that gives rise to the most (myopically) statistically significant community structure, meaning the community structure which is least likely to emerge in a network where edges are randomly distributed between nodes.^{5,6}

While convenient, this approach is deeply problematic for social scientists. The resolution parameter reflects a deep and substantively-important truth about communities: communities can exist on many different scales, and the question of which scale is most relevant depends on the question being asked. In the study of intra-village politics, for example, we may be interested in the structure of the small communities that form within villages; in the study of national politics, by contrast, we are likely to be more interested in community structure at the level of large regions.⁷

⁵All community detection algorithms address the problem of scale. Modularity-optimizing community detection algorithms, for example, implicitly follow the strategy described above. Others, like InfoMap, use an analogous information-theoretic criterion.

⁶It is worth noting that this kind of problem of scale is not unique to network analysis, and these two approaches employed here are portable to other contexts. An analogous problem exists in many inductive algorithms, including any clustering algorithm in which individual observations are inductively grouped into buckets based on similarity. In text-as-data topic models, for example, the number of topics allowed is a near-perfect analogue. Similarly, the results of spatial clustering analyses depend on the scale at which clustering is measured — clustering of voters is very different within cities than at the level of states, for example. As such, discussion of how best to address the problem of scale is of broad and increasing relevance as political science adopts more and more tools from statistics and computer science.

⁷To illustrate, consider a network of academics. Network communities that consist of members of a sub-discipline (like Americanists, Comparativists, and Normative Theorists

This paper departs from reliance on this myopic statistical criterion (hereafter referred to as the *atheoretic criterion*) in two ways. First, network communities are calculated for values of the resolution parameter γ across the parameter space (from small, internally dense communities to large, looser communities), and results are presented for all values. This not only provides a sense of robustness of results, but also a novel opportunity to gain insights into the scale at which network structure appears to matter for political outcomes, turning a nuisance parameter into an opportunity for learning.

Second, as discussed in more detail in Section “Public Goods and Voter Knowledge”, this analysis also uses variation in scale to further test the network-proxy hypothesis. If ethnicity achieves political salience because ethnic fractionalization is correlated with network fragmentation, and network fragmentation impedes the ability of communities to effectively engage in political activity, then not only should ELF and network fragmentation be correlated, but the scale (value of γ) at which they are correlated should also be the scale at which network fragmentation is negatively correlated with political outcomes like public goods and voter knowledge.

Geo-Referencing Subscribers

The output of the CPM is an assignment of each individual subscriber to a network community. In order to convert these individual-level assignments into information about network fragmentation in each electoral districts, it is first necessary to identify the electoral district in which each subscriber resides.

Geo-referencing of subscribers takes advantage of the fact that most calls are routed through the nearest antenna tower. The cell-phone meta-data used for this analysis identifies the antenna tower that handles every call placed by a subscriber of the PT. Using this information, combined with data on the GPS location of these antenna towers, it is possible to geo-reference cell-phone subscribers to specific zones based on the assumption that a user’s home is closer to his or her most-used cell-tower than any other tower. This process is detailed in Appendix F, along with information on several additional refinements — such as restricting attention to calls made before 8 am and after 6 pm (when subscribers are most likely at home and not at work) and using information on the second-most-used cell-phone tower.

in a political science department) may be the most clearly delimited groups in a myopic statistical sense (since we talk most to colleagues whose work is most related to our own). However, if we wish to use this network to predict voting behavior in a University-level debate over how funds should be allocated across departments, we know *intra*-departmental divisions are less likely to shape voting behavior than super-department network communities (social science departments, natural science departments).

Measuring District Fragmentation

The previous two sections detail how each individual cell-phone subscriber is inductively assigned to a network community, and how each individual cell-phone user is geo-referenced. This section describes how these results are combined to create a measure of whether residents of each electoral district are divided into many small network communities (high fragmentation) or into a small number of large network communities (low fragmentation). As previously noted, this is computed in the same manner as ELF. So, if network communities are coincident with ethnic groups, the network fragmentation measure computed here will be equivalent to ELF.

This aggregation allows subsequent analyses at the level of the electoral constituency. Most forms of political action — like rallying around an opposition candidate, or applying pressure to an elected official — require participation of a substantial share of an elected official’s constituents to be effective. As such, constituency-level measures of network structure constitute the most appropriate unit of analysis for the study of ethnicity, network structure, public goods, and voter knowledge.

As detailed in Appendix G, most public goods are under the control of Zambia’s national government, and so the focus of this analysis is on the electoral districts (*constituencies*) for members of Zambia’s national legislature, the National Assembly. Zambia’s national government has primary *de facto* authority over public goods in Zambia, including water, electricity, education, police, and more. As a result, the ability of citizens to influence their elected representatives in the National Assembly is likely to affect the quality of service they experience across nearly all government departments.

Computation of District Fragmentation

The district-level measure of network fragmentation is calculated in a manner analogous to ELF: for each electoral district d and community partition of the network into network communities $c \in C$, network fragmentation (NF) is computed as one minus the Herfindahl index of communities:

$$NF_d = 1 - \sum_{c \in C} \left(\frac{n_{c,d}}{n_d} \right)^2 \quad (1)$$

where n_d is the total number of cell-phone subscribers who reside in district d , and $n_{c,d}$ is the number of cell-phone subscribers in district d assigned to community c . This generates a single measure of network fragmentation for each electoral district that can be interpreted as the probability that any two randomly selected residents of a district are in different network communities.

One nuance to this calculation is that available data can only geo-reference cell-phone subscribers to specific zones, not points,⁸ so adjustments must be made for situations where subscribers are assigned to zones that are not completely contained within an electoral district. As discussed in detail in Appendix M, this is accomplished by assuming that users are distributed in proportion to the area of the assigned zone that falls within each district and in proportion to each district's population density.

Data

Cell-Phone Meta-data

Network data for this analysis consists of records of all SMS (text) and voice transactions passing through the network of a major Zambian Telecommunications company from December 2011 to June 2012. This data — commonly referred to as *Call Detail Records* or CDRs — includes approximately 2 billion transactions. Each transaction includes type (voice call or SMS), anonymized identifiers for both caller and receiver, date and time of transaction, duration, and GPS coordinates of the antenna tower through which the call was placed. The anonymized identifiers for callers and receivers are stable codes that allow usage patterns to be tracked over time but cannot be used to identify individuals or their phone numbers.

This dataset comes from one of the three dominant cell-phone providers in Zambia. Because of the sensitivity of the data, the telecom providing the data has asked to be referred to only as the Partner Telecom (PT), and certain commercially-sensitive statistics (like the exact location or number of antenna towers) have been omitted or are presented in an intentionally imprecise manner. As of early 2012, it can be said that PT had a market share of 25–40%,⁹ had between 500 and 700 antenna towers, and provided service in 148 of Zambia's 150 constituencies.

Two types of cell-phone users appear in the network data: PT subscribers and non-PT subscribers. Because the dataset is constructed by aggregating PT subscriber data, it includes (a) all transactions by PT subscribers and (b) transactions between PT subscribers and non-PT subscribers. As a result, nearly all cell-phone users in Zambia appear in the data, including non-PT most subscribers.¹⁰ Not included in the data are (a) non-PT

⁸For example, a crude geo-referencing would assign a user to the area in which the user's most-used tower is the closest tower. Full details of the more detailed but analogous geo-referencing strategy used in this paper can be found in Appendix F.

⁹The range of market shares for the top three telecoms in 2012.

¹⁰The 2010 census puts Zambia's population at 12.5 million individuals; and there are approximately 9 million users in the cell-phone meta-data.

subscribers who never call PT subscribers, (b) calls that occur between pairs of non-PT subscribers, and (c) geo-spatial information on non-PT subscribers.¹¹

These data make it possible to measure social networks in a fundamentally different way from past studies that collected data by asking individuals for the names of their closest friends in a geographically-bounded area (e.g. Banerjee *et al.*, 2014; Dionne, 2015; Fafchamps and Vicente, 2013; Rojo and Wibbels, 2014) and provide several advantages. First, they provide exceptional *breadth* — they cover an entire country — making it possible to estimate the properties of networks in 150 national electoral districts (population $\sim 100,000$). This makes cross-sectional analyses of the relationship between network structure and outcomes of interest possible at the most politically-relevant level. Second, the data have exceptional *depth*, capturing even the weak social ties sociologists have found to be exceptionally important, but which are often censored or forgotten when individuals are asked to list their five or ten closest friends (Granovetter, 1973). Third, the data include network ties between villages and between villages and cities, providing a more complete picture of networks than geographically-bounded surveys. And finally, unlike self-reported data, these data measure *actual communication patterns*, reducing concern of under-reporting of inter-class or inter-ethnic relationships. This is not to say meta-data perfectly capture social networks — like all other methods of network mapping, it has limitations — but these shortcomings are balanced by the advantages meta-data provides, making it an important compliment to traditional survey mapping.

Because of these advantages, meta-data has become an increasingly popular tool for analyzing networks in recent years, especially among computer scientists and economists. It has been used to measure internal migration and mobility (Blumenstock, 2012; Wesolowski *et al.*, 2012), estimate wealth and socio-economic status in hard-to-surveil contexts (Blumenstock *et al.*, 2015), gain insights into the topology of real-world networks (Onnela *et al.*, 2007), examine the spatial distribution of network communities (Barthélemy, 2011; Blondel *et al.*, 2010), study social processes like the spread of knowledge through networks (Björkegren, 2015), and estimate the credit worthiness of unbanked populations (Björkegren and Grissen, 2015). To the best of this author’s knowledge, however, this is the first use of cell-phone meta-data for the study of political phenomena.

¹¹Users are geo-referenced using information on the antenna towers to which they connect. Telecoms manage their own towers, so PT CDRs only include information on tower routing for PT subscribers.

Cell Phone Use in Zambia

The rapid penetration of cell-phones in Zambia makes them a useful tool for measuring social networks. According to a 2012 nationally-representative Afrobarometer survey of 1,200 adult citizens (Afrobarometer, 2012), households report an average of 0.79 phones per adult and 79.6% of households have at least one cell-phone. Even in rural communities, 67.0% of households report owning a mobile phone. Phones are also well used. Among respondents in households with a phone, fully 83.0% report using the phone at least once a day. Moreover, data from Research ICT Africa (2008) also suggests that these phones are used extensively for social purposes, not just business, as shown in Table 1.

There is also little evidence of systematic differences in usage across ethnic groups. As shown in Table 2, phone ownership rates are relatively similar across ethnic groups. And while a limited sample size makes strong inferences difficult,¹² as shown in Table 3, ethnic groups do not appear to sort into different carriers.

Table 1: Who do people call most?

	Share of respondents
Family members	51.2%
Friends	44.0%
Business clients	4.1%
Business suppliers	0.6%
Information services	0.2%

Source: Research ICT Africa (2008).

Table 2: Phone ownership by ethnicity.

Ethnic group	Avg. num. phones per adult
Barotse	0.80
Bemba	0.80
NW	0.74
Nyanja	0.85
Tonga	0.72
Total	0.79

Source: Afrobarometer (2012).

¹²Most enumerators appear to have skipped the open-ended “Who is your primary carrier” question.

Table 3: Subscribers by ethnicity.

	CelTel	ZamTel	MTN
Bemba	42.2%	25.0%	76.5%
Lozi	3.5%	0.0%	0.0%
NW	3.5%	0.0%	0.0%
Nyanja	38.7%	71.4%	17.6%
Tonga	12.1%	3.6%	5.9%
No. Obs.	256	28	17

Source: Research ICT Africa (2008).

This high penetration, high level of use, and the fact phones are being used for social as well as business purposes suggest that cell-phone meta-data provide a good source of data for studying social networks.

Network Generation

Cell-phone meta-data is used to generate four distinct empirical networks. In each of these networks, the vertices of the network are individual cell-phone subscribers and edges are added between subscribers who communicate with one another. However, as summarized in Table 4, these networks differ in (a) the amount of communication required for two vertices to be considered connected and (b) the types of calls considered.

First, networks are differentiated by the threshold for connection. In the inclusive network specifications, nodes that exchange at least three SMSs or calls over the 7-month period of the data are treated as connected. This cutoff is chosen to exclude missed calls and single back-and-forth exchanges. In the more restrictive specifications, edges are only placed between individuals who have communicated at least once per month on average (seven times), thus capturing only strong relationships. See Appendix C for details on the decisions to specify unweighted and undirected networks.

Second, networks are differentiated by the type of communications considered. The *All-Days* networks consider calls placed on any day, while *Weekend-Only* networks consider only calls placed on the weekend, restricting attention to calls likely to be non-commercial in nature.

Table 4: Different network filters.

	>2 Calls or >2 SMS	>6 Calls or >6 SMS
All-Days	Everyone	Strong Contacts
Weekend-Only	All Social	Strong Social

Table 5: Network summary statistics.

Network	Callers	Connections	Average degree
All-Days, >2 Calls	7,982,865	68,282,737	8.6
All-Days, >6 Calls	6,294,867	37,105,650	5.9
Weekend-Only, >2 Calls	6,285,375	32,626,556	5.2
Weekend-Only, >6 Calls	4,395,327	15,915,679	3.6

For each of these networks, the 1% of users with the most connections are dropped from all networks to exclude large firms. This corresponds to dropping users with more than 170 and 91 distinct contacts for the All-Day networks, and 109 and 55 distinct users for the Weekend-Only networks.¹³

Summary statistics for these different networks are presented in Table 5. It shows that changes in the communication threshold have a substantial effect on the number of users included, as many users are light users. When one moves from the most to the least inclusive criteria, the total number of users drops by almost half. Similarly, the restriction to weekend connections also causes a substantial reduction in the number of users, suggesting that many phones are used primarily for business.

Theoretical guidance on which of these network filters is best does not exist. Theories of how network fragmentation may impede the ability of citizens to hold politicians accountable — like information diffusion models — are agnostic about the nature of connections. Similarly, models that suggest fragmented networks impede social sanctioning sometimes focus on sanctioning among friends and family (like Jackson *et al.*, 2012), but sometimes suggest that sanctioning may also occur in commercial relationships (e.g. Fearon and Laitin, 1996; Larson, 2017b). Rather than make a theoretically unfounded decision about which network to use based on which generates the best-looking results — the type of ex-post decision that has been implicated as a potential threat to the integrity of empirical social science research (Ioannidis, 2005; Open Science Collaboration, 2015) — this analysis presents results for all four of these networks in parallel.

Measure Validation

This section provides a number of descriptive statistics about CPM-generated network communities. It shows that these communities have numerous

¹³Note that the restriction of attention to calls made in the mornings and evenings noted in Section “Measuring Social Network Fragmentation” for ge-ocoding users is not used in network generation. That restriction is only meant to ensure so inferences about one’s place of residence are not influenced by calls made at work.

Table 6: Population-weighted community statistics.

	Mean size	Median size	Largest size
All-Days, >2	8,058	34	88,724
All-Days, >6	50	42	507
Weekend-Only, >2	45	38	401
Weekend-Only, >6	26	24	195

properties that comport closely with intuitive concepts of politically and socially salient networks, providing assurances that later analyses rest on a solid foundation.

Community Sizes

Table 6 presents the subscriber-weighted distribution of network community sizes.¹⁴ It shows that the median network community consists of 24–42 individuals for the four network specifications, though all networks show significant right-skew. This is most pronounced for the All-Days > 2 network where the largest communities are more than 100 times larger than in any other network. This may suggest that the network is insufficiently filtered, although as previously noted, there is limited theoretical guidance to determine what network filter is most appropriate. Community size histograms can be found in Appendix H.

Spatial Distributions of Communities

The CPM assigns subscribers to network communities based only on calling patterns and does not take into account physical locations, so the spatial distribution of network communities is epiphenomenal and can be used as a test of measure validity.

Figure 1 plots the spatial distribution of eight network communities selected at random from the 100 largest communities from the Weekend-Only > 2 network. The figure shows a heat map where redness indicates the share of network community members living in a location.¹⁵ As the figure shows, most communities consist either of (a) a group in a single city or (b) a relatively concentrated group in a rural locality with a small diaspora located in a nearby urban center.

¹⁴For the duration of Section “Measure Validation”, results are presented at the resolution selected by the atheoretic criterion for simplicity of presentation; results in formal analyses in later sections are presented at all resolutions.

¹⁵For simplicity, users are assumed to live at the centroid of the Second-Order Thiessen Polygon. For more on geocoding, see Appendix F.

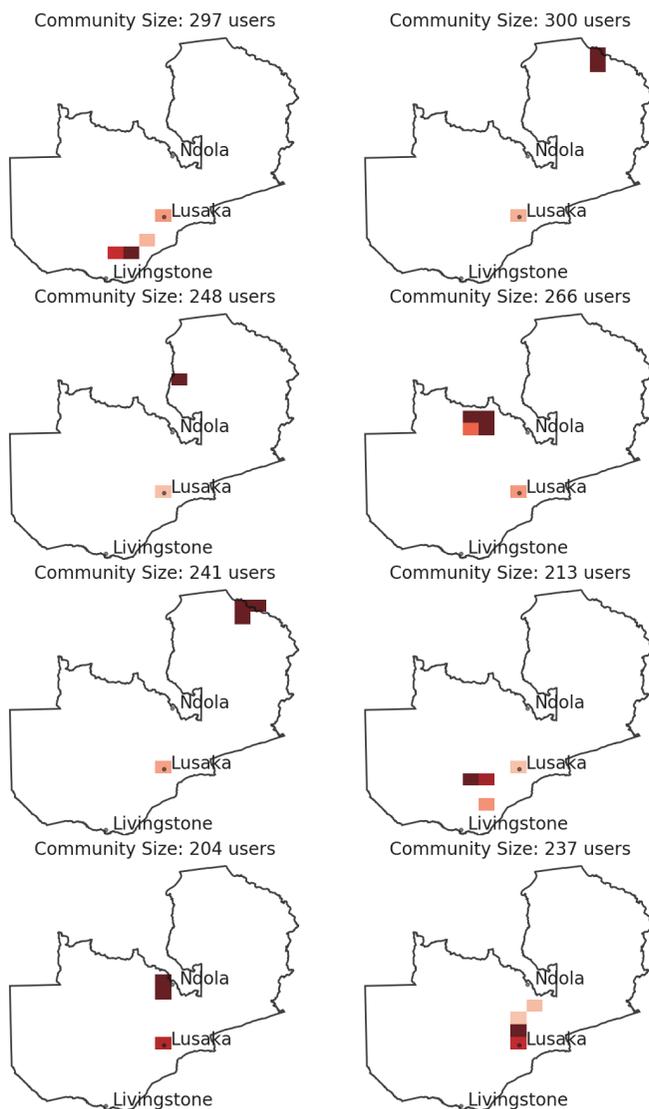


Figure 1: Spatial distribution of eight random communities from the *Weekend-Only > 2 Connections* network. Figures show heatmaps of the geo-referenced locations of members of eight network communities selected at random from the 100 largest communities from the weekend > 2 call network. The darkness of shading is proportional to the share of community members geo-located within a given area. The plots are relatively representative of the two main types of spatial distributions in the data: either (a) a large, densely clustered communities in a rural area and smaller diaspora communities located in one or more of Zambia's urban centers (e.g. top right), or (b) hyper-urban communities located almost entirely within major cities (e.g. bottom left).

This pattern of rural communities linked to an urban diaspora — especially diaspora in Lusaka or the Copperbelt — is consistent with known patterns of rural-to-urban migration in Zambia. As of 2010, rural-to-urban migrants made up just over 5% of Zambia’s population, and of these 624,000 individuals, more than half (53%) lived in Lusaka and 30% lived in the Copperbelt (Zambian Central Statistics Office, 2013, pp. 8–12). Similar patterns can also be seen in other networks, as shown in Appendix I.

This pattern is also corroborated by spatial calling patterns. Sixty-one percent of the average user’s contacts are located within 20 km of their location.¹⁶ This estimate likely underestimates the number of contacts in close proximity,¹⁷ but even so, it does suggest that urban migrants are a significant part of communication networks.

The overall distribution of subscribers also suggests accurate geo-referencing. For both constituencies and wards (a sub-constituency administrative unit, $N = \sim 1,420$), the correlation between number of subscribers and population is quite high — 0.78 at the constituency level and 0.81 for wards — and population shares are approximately in line with PT market share (see Appendix J for scatter plots).

Additional evidence of measure validity can be found in the structure of network communities in Zambia’s Western Province. As detailed in Appendix K, Western Province has stood apart from the rest of Zambia since the British South African Company first turned it into a semi-autonomous protectorate. Today, it has self-governance rights, and is home to multiple secessionist political parties. In light of this history, one might expect network communities in Western Province to be unusually internally oriented.¹⁸ This can be tested using a measure of *national integration* for each of Zambia’s nine provinces.

National Integration is defined as the degree to which residents of a province belong to network communities that include people from other provinces. A province where everyone is assigned to network communities with fellow residents would receive a score of 0; a province where everyone belonged to a community consisting entirely of residents of other provinces would receive a score of 1. (The precise formula can be found in Appendix L.)

Table 7 presents these National Integration scores for each province in Zambia for each network type as well as an average across all four networks. As anticipated, Western Province has an extremely low national integration score. Indeed, only Zambia’s two urban centers of Lusaka and Copperbelt are

¹⁶Again assuming each user lives at the centroid of their Second-Order Thiessen Polygon for simplicity.

¹⁷This is the case for two reasons. First, assuming all users are located at the centroid of their Thiessen Polygon amounts to assuming that population is uniformly distributed in space; in reality, users tend to cluster, and these clusters are often proximate to one another. And second, measurement error in geo-referencing will almost always lead to increases in the estimated distance between points.

¹⁸The author is indebted to Pierre Englebert for this suggestion.

Table 7: National integration scores by province.

	Avg Score	All-Days, >2	All-Days, >6	Weekend-Only, >2	Weekend-Only, >6
Copperbelt	0.214	0.148	0.243	0.245	0.222
Lusaka	0.303	0.222	0.337	0.338	0.315
Western	0.308	0.193	0.345	0.354	0.339
Eastern	0.312	0.214	0.347	0.352	0.337
Southern	0.315	0.202	0.353	0.359	0.345
Northwestern	0.343	0.223	0.385	0.390	0.372
Northern	0.402	0.286	0.442	0.449	0.433
Central	0.415	0.290	0.461	0.465	0.443
Luapula	0.440	0.292	0.493	0.498	0.476

more internally-oriented, despite Western Province being located relatively close to both these urban regions. As with the spatial distribution of network communities, this strongly suggests that network communities are capturing something closely related to what we intuitively think of as communities.

Network Fragmentation and ELF

Having established a measure of social network fragmentation in Section “Measuring Social Network Fragmentation” and validated the measure in Section “Measure Validation”, this section turns to the core substantive topic of this analysis: the relationship between ELF and network fragmentation.

To test the relationship between network fragmentation and ELF, network fragmentation ($NF_{i,p,\gamma}$) for constituency i in province p calculated with resolution parameter $\gamma \in (0, 1)$ is regressed on a set of constituency controls (X_i), province fixed-effects (ϕ_p) and ELF (ELF_i):

$$NF_{i,p,\gamma} = ELF_i\beta + X_i\delta + \phi_p + \epsilon_i \quad (2)$$

Constituency controls X_i consist of a set of controls to address sources of potential variation in the fidelity with which cell-phone meta-data captures the true structure of social networks, including population density, share of the constituency that is rural, share of residents who are subscribers with the Partner Telecom, and dummies for the urban centers of Lusaka and the Copperbelt.¹⁹ In addition, province fixed effects are included to ensure that

¹⁹More precisely, the specification includes log population, log area, and log number of subscribers, which, given additive separability of logged ratios, is equivalent to the described parameters. Lusaka and Copperbelt dummies include core urban constituencies and immediate neighbors. *Lusaka dummy* includes constituencies of Katuba, Kafue, Chilanga,

the variation examined exists among constituencies within each region and that results are not overly influenced by large regional differences in patterns of cell-phone usage or other unobservable factors, a distinct possibility in a country as regionally diverse as Zambia.

Measuring ELF

The measurement of ELF is necessarily sensitive to how ethnic groups are enumerated, and variations in enumeration choices can lead to substantial variations in measures (Posner, 2004). This is also complicated by variation in the salience of different ethnic identities over time in Zambia, as discussed in Appendix A. As data for this analysis comes from 2010, ELF is calculated with respect to the dimension of identity that has consistently been most political salient since the mid-1990s — what Posner (2005) terms *linguistic group* identity. As described in Appendix B, ELF is calculated according to a 5-fold *linguistic group* taxonomy using data from the Zambian census.

Note that *linguistic group* identity is not defined by one’s language. Most Zambians identify with one of ~72 tribes, each with its own language.²⁰ However, certain tribes — generally those whose languages share a similar linguistic heritage — tend to also co-identify.²¹ To differentiate between these two (nested) levels of ethnic identity, Posner uses the terms *tribal* identity and *linguistic group* identity. However, it is important to emphasize that these labels are academic constructions — both types of identity are referred to as “tribal” by Zambians (Posner, 2005, p. 115). Moreover, as suggested by the fact that most tribes have their own language, the terms are obviously imprecise. Indeed, two people who belong to the same linguistic group need not speak the same language — they need only identify with tribes whose languages have similar linguistic roots.

In addition, individuals who identify with different linguistic groups often share a language due to high levels of bilingualism. Many Zambians speak one language in the home (often their tribal language) and another language for commercial transactions and social interactions (Laitin, 1992). Indeed, in the 1990 Zambian census, fully 25% of people who spoke Bemba, Nyanji, Tonga, or Lozi spoke it as a second language (Posner, 2005, p. 60).

Chongwe, Chawama, Kabwata, Kanyama, Lusaka Central, Mandevu, Matero, and Munalu. *Copperbelt dummy* includes constituencies of Chililabombwe, Chingola, Nchanga, Kalulushi, Chimwemwe, Kamfinsa, Kwacha, Nkana, Wusakile, Luanshya, Roan, Kankoyo, Kantanshi, Mufurila, Kafulafuta, Masaiti, Bwana Mkubwa, Chifubu, and Ndola.

²⁰There is some dispute about whether each tribe has its own language or whether some are dialects of one another. As pointed out by Kashoki (2017b), however, the distinction is slippery at best.

²¹These identities are nested, like identifying as both Irish and White, or Chinese and Asian.

Inter-intelligibility across linguistic groups is further facilitated by the existence of a *lingua franca* in each region. These *linguae francae* are officially recognized languages, are commonly used in media and commerce, are approved for use by civil servants, and, until recently, primary school instruction was required to be in either English or each region's designated *lingua franca*.^{22,23}

As such, while *linguistic group* is a convenient label for an ethnic taxonomy that is well understood among Zambians, it is measuring an identity that is much deeper than language and only somewhat related mutual intelligibility.

ELF and Network Fragmentation Results

Figure 2 shows the partial correlation of ELF and network fragmentation for the full sample of constituencies. Each estimate in the figure is the β coefficient from a regression of network fragmentation on ELF using Equation (2), where network fragmentation is calculated using the value of γ indicated on the x -axis. Moving from left to right along the x -axis corresponds to moving from a small number of large, loosely connected communities to lots of very small, dense communities.²⁴ Coefficients have been normalized so that point estimates can be interpreted as the impact on network fragmentation (in standard deviations) of a movement of ELF from 0 (no fractionalization) to 1 (full fractionalization).

Three features of the figure are worth noting. First, in nearly all networks and at all values of the resolution parameter, the point estimate for the partial correlation between ELF and network fragmentation is positive, as predicted by the network-proxy hypothesis. Second, these positive correlations are also statistically significant across a substantial range of resolution parameter values. Third, the estimated effect is also substantial — in the statistically significant estimates, a movement from non-fractionalized to fully-fractionalized constituencies appears to correspond to a relatively consistent 1–2 standard deviation increase in network fragmentation across all networks.²⁵

²²The one outlier in this pattern is Northwestern, which has three official *linguae francae* — Kaonde, Luvale, and Lunda. Other regions are restricted to instruction in a single *lingua franca*: Bemba in Copperbelt, Northern, Luapula, Kabwe (Urban), Mkushi, and Serenje districts; Nyanji in Eastern, and Lusaka region; Tonga in Southern, Kabwe (Rural) and Mumbwa District; and Lozi in Western, Livingstone (Urban) (Kashoki, 2017b).

²³In addition, while distinct, these *linguae francae* are closely related, as all belong to the Bantu language family. Moreover, in a test of mutual intelligibility among school children, Kashoki (2017a) found that Bemba, Tonga, Lozi, and Nyanja speaking children understood about 30% of the content of passages read in other languages. These languages were also found to have an overlap of about 30–45% with one another in basic vocabulary.

²⁴In the limit, moving a little to the left of the plot corresponds to placing all nodes in one large community and moving a little past the right side corresponds to placing every node in its own community.

²⁵Results without province fixed effects (shown in Appendix N) show statistically significant estimates of a *negative* correlation for very small communities sizes ($\gamma > -5$, where

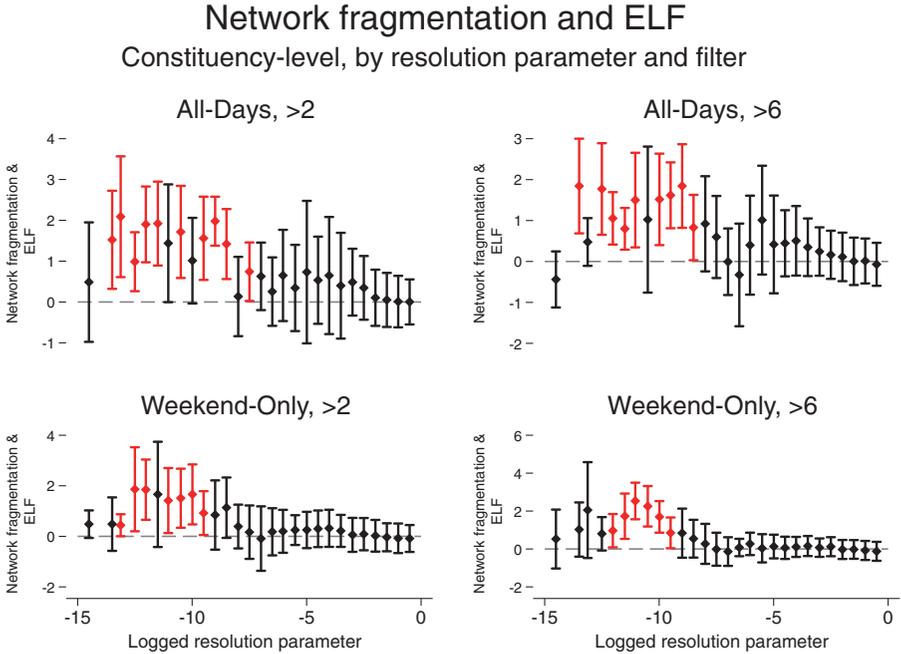


Figure 2: Partial correlation coefficients between ELF and network fragmentation plotted at different community detection resolutions, ranging from large, relatively inclusive network communities on the left end of the x -axis to small, dense communities on the right end of the x -axis. Point estimates in red are statistically significant at the 90% level. Coefficients have been normalized so point estimates can be interpreted as the impact on network fragmentation (in standard deviations) of a movement of ELF from 0 (no fractionalization) to 1 (full fractionalization). Standard errors are clustered at the province level.

Finally, the ranges of γ over which the relationship tends to be strongest correspond to network communities whose size comports well with our intuition of what scale of network structure might matter for political effectiveness. At $\gamma = -10$, average community sizes are approximately 2,000–30,000 people across network specifications, which roughly corresponds to the size of a large political rally or protest.²⁶ The question of whether these are the resolutions at which the network-proxy would predict the correlation to be strongest,

median network communities consist of less than 24–42 individuals). As previously noted, however, province fixed effects are theoretically preferable, given the regional heterogeneity of Zambia.

²⁶At $\gamma = -10$, population-weighted average community sizes are 29,838 for *All-Days, >2 Calls*, 17,008 for *All-Days, >6 Calls*, 11,736 *Weekend-Only, >2 Calls*, 2,091 for *Weekend-Only, >6 Calls*.

however, is addressed more systematically in Section “Public Goods and Voter Knowledge”.

It is also interesting to note that these community sizes are substantially smaller than the sizes of the ethnic groups in question, and are even smaller than the lower-level of ethnic identity groups in Zambia, tribes. This makes clear that while ethnically homogenous communities have less fragmented networks, this is not simply due to clean partitions along ethnic lines. Ethnic divisions clearly contribute to network fragmentation, but cannot alone explain network structure alone.

One other aspect of these results worth noting is that the resolution parameter values at which ELF and network fragmentation are most correlated do *not* include the value of $\log(\gamma)$ that would be selected by the atheoretic criterion — approximately -5 in most graphs. Indeed, if one were to conduct this analysis using only the default value of γ from an out-of-the-box CPM package, one would erroneously conclude that ELF and network fragmentation are uncorrelated, illustrating the dangers of over-reliance on myopic statistical criteria when importing computational algorithms into the social sciences.

The positive association between ELF and network fragmentation is even more notable among rural constituencies. Figure 3 replicates Figure 2 for the 104 constituencies with at least 70% rural populations. The figure shows that in the rural sample, both the magnitude of the ELF–network fragmentation relationship is larger and also the coefficients are statistically significant across a larger range of resolution parameters.²⁷

This result appears relatively robust. As shown in Appendix O, tightening the sample restriction to constituencies that are at least 85% rural strengthens the correlation significantly despite the decrease in sample size. Similarly, a persistent positive correlation remains when dropping all controls except the log number of subscribers and province fixed effects. Results also do not appear to be driven by outliers, although they are not robust to exclusion of province fixed effects, which results in much noisier estimates and, as above, some estimates of negative correlations at very small network community sizes.

Two conclusions follow from these results. First, while additional work is needed to understand why the relationship only exists at certain resolutions — as addressed in Section “Public Goods and Voter Knowledge” — the results in this section are consistent with the network-proxy hypothesis.

Second, these results suggest that the relationship between ethnicity and network fragmentation is weaker in cities than rural communities. This speaks to a broader literature on the effects of urban migration and modernization on

²⁷This sub-sample analysis, like the analysis across resolutions, was not a part of the author’s original research design. As such, the statistical significance of this post-hoc exploratory analysis should be interpreted with a measure of caution (Casey *et al.*, 2012), though this limitation does not negate the value of the results (Laitin, 2013; Olken, 2015).

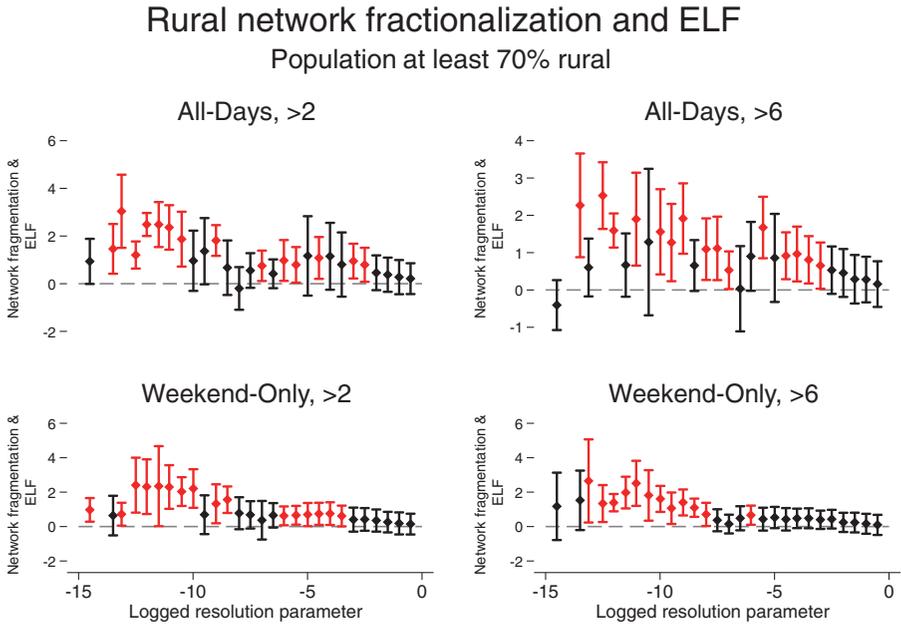


Figure 3: Partial correlation coefficients between ELF and network fragmentation for the 104 constituencies that are at least 70% rural plotted at different community detection resolutions. Network community sizes range from large, relatively inclusive network communities on the left end of the x -axis to small, dense communities on the right end of the x -axis. Point estimates in red are statistically significant at the 90% level. Coefficients have been normalized so point estimates can be interpreted as the impact on network fragmentation (in standard deviations) of a movement of ELF from 0 (no fractionalization) to 1 (full fractionalization). Standard errors are clustered at the province level.

ethnic identification, and on the question of whether urban migrants adopt class-based identities rather than ethnic identities. Much of this literature comes from the study of India, where evidence suggests that class-based cleavages have generally not been adopted (Thachil, 2015; Weiner, 1978). African cities have been less studied, but, to-date, results suggest a different dynamic. Melson (1971), for example, finds that Nigerian urban migrants seem to identify with *both* class and ethnicity. He finds a substantial share of workers support both a labor party and also an ethnic party at the same time. Similarly, in an audit survey of commercial interactions in an urban market in Lagos, Nigeria, Grossman and Honig (2015, p. 6) find “[n]on-coethnics who appear lower class are treated roughly the same as lower class coethnics,” while this is not the case for higher classes.

The findings from this analysis support the idea that urban migrants in African cities may be responding differently from those studied in India: while ethnicity is not unrelated to network structure in cities, the relationship does appear to attenuate significantly in urban constituencies.

Network Fragmentation and Non-salient Cleavages

The network-proxy hypothesis posits that ethnicity is political salient because the correspondence between ethnicity and network structure offers substantial organizational advantages to resultant political parties. A corollary of the hypothesis is that other cleavages — those which we do *not* see achieving political salience — must not offer this same advantage. To test for this, this section examines the relationship between network fragmentation and fragmentation along two other potential political cleavages that, despite many appealing attributes (like congruence with policy preferences), have not achieved salience in Zambia: employment sector and religious identification.

Figure 4 replicates the analysis from Section “Network Fragmentation and ELF” (Specification 2) with economic fragmentation and religious fragmentation substituted for ELF.²⁸ As the figure shows, there is no systematic relationship between network fragmentation and fragmentation along these alternate potential political cleavages, consistent with the network-proxy hypothesis.

These results raise an obvious question: *why* is ethnicity uniquely correlated with network structure? Though answering this question rigorously is beyond the scope of this analysis, the network-proxy hypothesis does offer one potential explanation: British intervention and subsequent positive reinforcement.

As in many colonies, British administrative policies contributed significantly to the emergence of ethnicity as a politically salient identity. Indeed, as documented by Posner (2005, p. 54), “[t]he reason Zambians identify themselves and others in tribal terms is because the institutions of the [British South African Company] and Colonial Office rule classified them in this manner and generated incentives for them to invest in these classifications.”

Once ethnicity achieved salience, subsequent positive reinforcement may be responsible for the persistence of the alignment of ethnicity and network structure. The British created an incentive for politicians to mobilize voters along ethnic lines, but these mobilization efforts likely just reinforced ethnic networks. As a result, over time parties with ethnic constituencies developed

²⁸Religious fragmentation is an Herfindahl index of the share of Constituency residents who identify as Protestant, Catholic, Muslim, Hindu, Buddhist, Bahai, Other, or None in the 2010 Zambian Census. Economic fragmentation is a measure of fragmentation across the categories of Manufacturing, Farming/Fishing/Forestry, Services, and Mining in the 2010 Zambian Census. Summary statistics can be found in Appendix U.

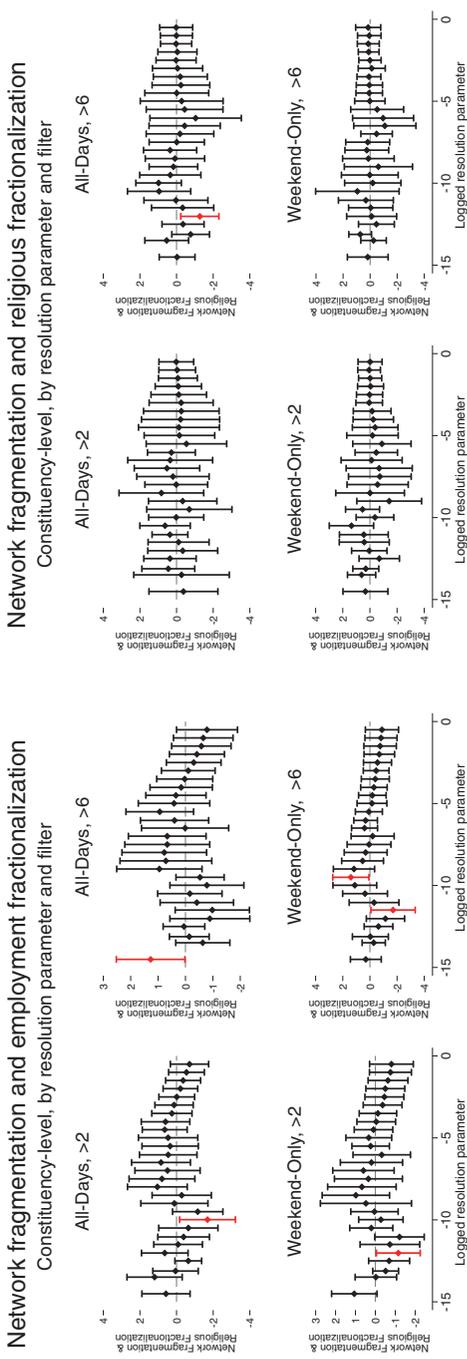


Figure 4: Partial correlation coefficients between alternate potential cleavages and network structure. Network community sizes range from large, relatively inclusive network communities on the left end of the x -axis to small, dense communities on the right end of the x -axis. Point estimates in red are statistically significant at the 90% level. Coefficients have been normalized so point estimates can be interpreted as the impact on network fragmentation (in standard deviations) of a movement of employment sector fragmentation (left) or religious fragmentation (right) from 0 (no fractionalization) to 1 (full fractionalization). Standard errors are clustered at the province level.

ever-greater organizational advantages over parties that could not take advantage of these ready-built networks. This would lead to more mobilization along ethnic lines, further reinforcing these networks in a cycle of positive reinforcement, eventually resulting in ethnicity's uniquely strong relationship with network structure today.

Public Goods and Voter Knowledge

The network-proxy hypothesis posits that ethnicity achieves political salience because it is correlated with network structure, and network structure matters politically because of its effect on social dynamics like information diffusion or the ability of citizens to hold politicians accountable. If true, then (a) corollary outcomes like voter knowledge and quality of public goods should also be negatively correlated with network fragmentation, and (b) they should be negatively correlated with network fragmentation *at the same resolutions* that network fragmentation and ELF are most correlated.²⁹

This section tests this by examining how the relationship among network fragmentation, voter knowledge, and the quality of public goods provision varies with the resolution parameter. It finds that (a) both voter knowledge and quality of public goods are generally negatively correlated with network fragmentation as predicted by theory, (b) these negative correlations are most statistically significant at resolutions where network fragmentation is positively correlated with ELF,³⁰ as predicted by the network-proxy hypotheses, and (c) both voter knowledge and quality of public goods are most strongly correlated with network fragmentation at approximately the same resolution, suggesting the existence of a scale of maximal political significance.

Public Goods and Knowledge Measurement

Public goods are calculated using data on a number of public services from the 2000 and 2010 Zambian National Census 10%-sample micro-data, which include data on four essential public services: electrification, access to a protected water supply, child enrollment, and infant mortality. These measures are combined into a single measure by measuring change in each good from 2000 to 2010 for rural and urban households in each constituency separately, calculating a weighted average of those values for each constituency, normalizing those values, and extracting the first component of a Principal Component Analysis (PCA). A lengthier discussion of data used, tests of index validity, and motivation for use of an index can be found in Appendix R.

²⁹Discussion of the motivation for selection of voter knowledge and public goods quality can be found in Appendix P.

³⁰The most statistically significant is also the substantively largest in nearly all cases.

Political knowledge is measured using data from the 2009 Afrobarometer survey in which 1,200 Zambians were surveyed on a range of political issues in advance of the 2011 election. Political knowledge is operationalized as the share of political opinion questions to which respondents provided a response other than “Don’t Know/Haven’t Heard Enough.” For example, when asked “Now let’s speak about the present government of this country. How well or badly would you say the current government is handling the following matters, or haven’t you heard enough about them to say: Managing the economy?”, a response is coded **1** if the respondent provides an evaluation or refuses to answer, and is coded as **0** if the respondent answers “Don’t Know/Haven’t Heard Enough”. This coding is applied to 15 different questions (a full list of included questions can be found in Appendix Q), and a “Knowledge Index” for each respondent is computed as the share of responses coded as **1**. As the focus of this analysis is on network fragmentation measured at the level of electoral constituencies for the National Parliament, attention is restricted to questions pertaining directly to each constituency’s MP or views of national government policy over which the MP has influence. The measure has an average value of 94.5% and a standard deviation of 0.08.³¹

Public Goods and Knowledge Specifications

The relationship between public goods and network fragmentation is estimated as:

$$PublicGoods_{i,p,\gamma} = NF_{i,\gamma}\beta + X_i\delta + \phi_p + \epsilon_i \quad (3)$$

where i is an index of constituencies, ϕ_p is a vector of province fixed effects, and X_i is the set of Constituency-level controls used in Section “Network Fragmentation and ELF”. Dummies are also included for whether the Constituency was represented by an MP with the ruling party (MMD) in either the first or second parliament of the 2000s (although results are robust to exclusion of the MMD controls and Lusaka and Copperbelt dummies).

The relationship between voter knowledge and network fragmentation is estimated as:

$$PolKnowledgeIndex_{j,i,p,\gamma} = NF_{i,\gamma}\beta + X_i\delta + Z_j\gamma + \phi_p + \epsilon_j \quad (4)$$

where j is an index for individual survey respondents, ϕ_p is a vector of province fixed effects, Z_j are the individual-level controls including gender,

³¹The assumption underlying this measure is that exposure to information about government performance — either from direct experience or conversation — decreases the likelihood of a “Don’t know/Haven’t Heard Enough” response. The measure takes no position on the effects of information on attitude extremity or on the likelihood an individual will simply refuse to provide an answer (“Refuse to Answer” is a distinct code).

a PCA asset wealth index, urban/rural, and whether the respondent has completed primary school, and X_i are constituency-level controls used in Section “Network Fragmentation and ELF”.

In both regressions, the coefficient of interest is β — the partial correlation of network fragmentation and the corollary outcome.

Public Goods and Knowledge Results

Figure 5 plots normalized point estimates of β and the 90th percentile confidence intervals for each value of the resolution parameter $\gamma \in (0, 1)$. Recall that moving from left to right corresponds to moving from large, inclusive communities to smaller, more densely connected communities.

Several aspects of the figure are notable. First, consistent with theory, network fragmentation is broadly negatively correlated with both public goods and voter knowledge — that is, the point estimates are generally negative. For public goods, all statistically significant estimates are negative in three of the four networks; for voter knowledge, all statistically significant coefficients are negative.

Second, the resolutions at which political knowledge, voter knowledge, and network fragmentation are most closely correlated are quite distinct from those at which network communities are most unlikely in a random graph. These results show that we may be looking at the wrong resolution of networks when we rely on the atheoretic criterion, and point to the importance of thinking about the relevant level at which to measure fragmentation before making claims about how social dynamics affect political outcomes. (See Appendix T for further evaluation of differences between results at these most-correlated resolutions and resolutions chosen using an atheoretic criterion.)

Third, the resolution parameters that are most correlated are also very similar for the two social phenomena examined here (which are qualitatively different and come from different sources), as illustrated in Table 8. This strongly suggests that network fragmentation at these resolutions is robustly salient for social and political outcomes.

And finally, the resolutions at which public goods and voter knowledge are most correlated with network fragmentation are also close to the resolutions where ELF and network fragmentation have a positive and statistically significant correlation. Figure 6 replicates the plot of ELF and network fragmentation partial correlations from Section “Network Fragmentation and ELF” with the addition of a solid vertical line at the average of the resolutions most correlated with voter knowledge and public goods. The figure shows that regions where ELF and network fragmentation are especially well correlated are relatively close to the resolutions most correlated with public goods and voter knowledge. This suggests not only that ELF and network fragmentation

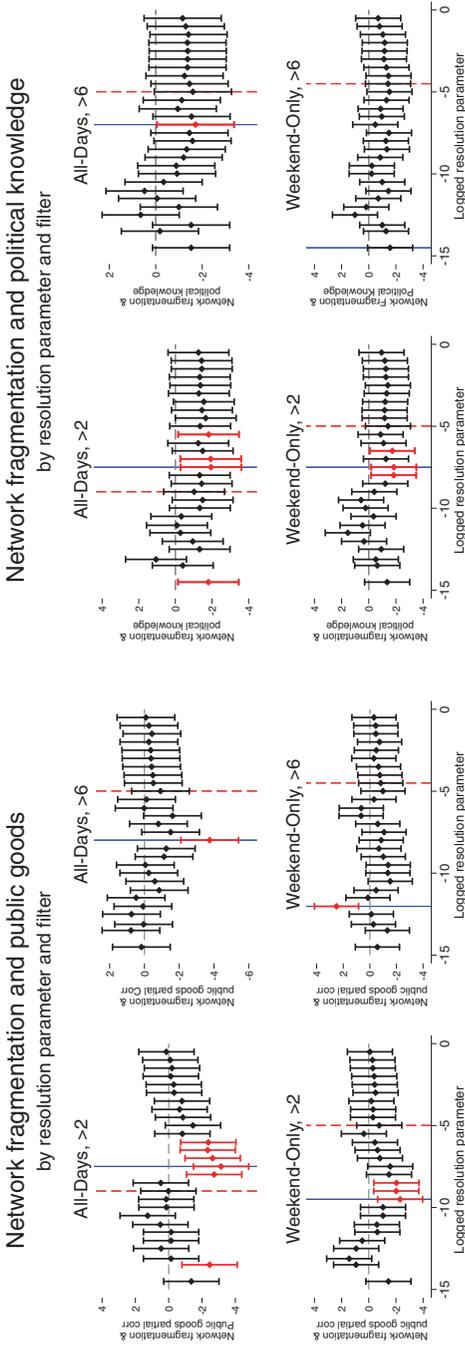


Figure 5: Partial correlation coefficients between the public goods index (left)/voter knowledge (right) and network fragmentation plotted at different community detection resolutions, ranging from large, relatively inclusive network communities on the left end of the x -axis to small, dense communities on the right end of the x -axis. Intervals are 90th percentile confidence intervals. The vertical dashed lines indicate the resolution selected by the atheoretic criterion. The vertical solid lines denote the resolutions at which the partial correlation coefficients are most significant for each specification. For readability, coefficients are normalized by their standard errors prior to plotting; as such, values correspond the t -statistic associated with each estimate. Plots with substantive magnitudes can be found in Appendix S. Standard errors are clustered at the province level for public goods and survey-site level for voter knowledge.

Table 8: Most significant (logged) resolution parameters.

	Pol Knowledge	Public goods	Average
All-Days >2	-7.50	-7.50	-7.50
All-Days >6	-6.99	-7.99	-7.49
Weekend-Only >2	-7.50	-9.50	-8.50
Weekend-Only >6	-14.51	-12.02	-13.27

Source: Correlation: .91

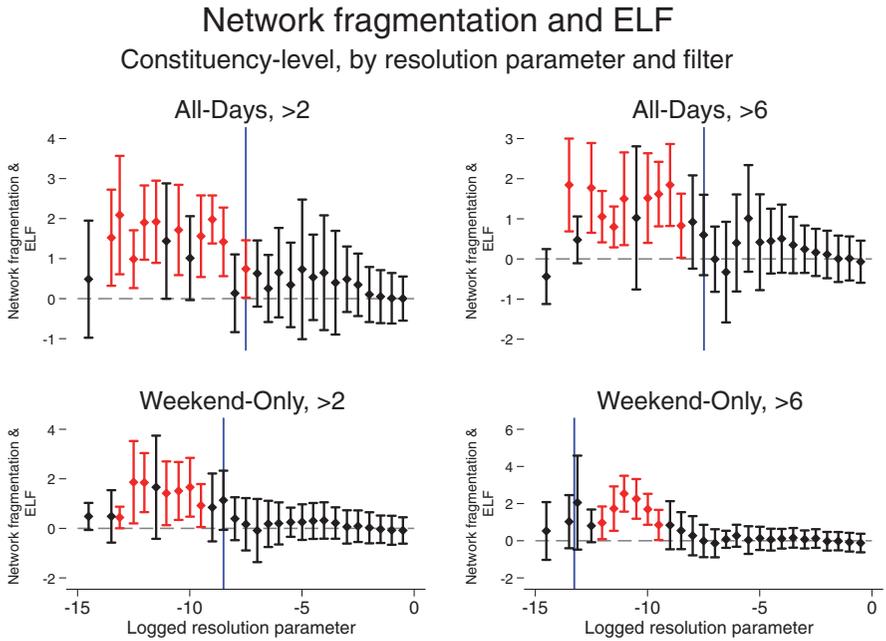


Figure 6: This figure plots partial correlation coefficients between ELF and network fragmentation at different community detection resolutions, ranging from large, relatively inclusive network communities on the left end of the x -axis to small, dense communities on the right end of the x -axis. 90th percentile confidence intervals. The vertical solid lines denote the resolution at which the partial correlation coefficients between network fragmentation and public goods/voter knowledge are most significant.

are consistently correlated at many resolutions — as illustrated in Section “Network Fragmentation and ELF” — but also that these correlations occur at politically relevant resolutions, providing further support for the network-proxy hypothesis.

Conclusion

This work makes several important contributions to our understanding of the role of ethnicity and social networks in the politics of developing democracies, and also to the study of civil society more broadly. Substantively, it provides clear evidence in support of the assumption underlying the network-proxy hypothesis: that ethnically fragmented communities have more fragmented social networks, and that this is not the case for alternate political cleavages that, tellingly, have not achieved political salience. In addition, it shows that voter knowledge and public goods tend to be negatively correlated with network fragmentation, providing further support for the network-proxy hypothesis.

The dynamics of ethnic politics in Zambian are very similar to those in other African countries, limiting concerns about external validity. The largest potential scope condition is that the dimension of ethnicity used in this analysis is related to language, as discussed in Section “Network Fragmentation and ELF”. As such, this correlation between network fragmentation and ELF may be more tenuous in communities where ethnicity is more divorced from language. As ethnicity and language are often intrinsically related, however, especially in Africa, even if this is the case, these findings are still likely to be informative for a large number of contexts.³²

These results also shed light on the results of other studies. For example, they are consistent with the Dionne’s (2015) finding that trust in behavioral games is higher among strongly connected individuals, regardless of whether they are co-ethnics (suggesting it is networks and not ethnicity that drives trust dynamics). The finding from Dionne (2015) that trust is higher among *non-co-ethnics* when partners are not close friends is harder to rationalize with the network-proxy hypothesis, and may be due to either something idiosyncratic about the social network within the village studied,³³ or as noted by Dionne (2015) it may suggest that intra-village dynamics diverge from social dynamics at larger scales. This analysis is also consistent with the empirical finding from Larson and Lewis (2017) that ethnically fragmented villages diffuse information less efficiently, although less consistent with the finding that ethnically fragmented communities have higher link density. Further research is needed to understand whether this divergence is due to the small sample size (two villages) in Larson and Lewis (2017), the scale of network structure being measured in Larson and Lewis (2017) (they look only at very local network properties and not at meso-scale properties like community fragmentation

³²The very fact that our standard measure of ethnic fractionalization is a measure of *ethno-linguistic* fractionalization underlines how intertwined these two dimensions are in our study of ethnic politics.

³³Suggesting this may be the case, Dionne (2015) found no evidence that reported close friends were any more likely to be co-ethnics than non-co-ethnics, an unusual finding given pervasive findings of homophily across a broad range of contexts (McPherson *et al.*, 2001).

as analyzed in this paper), or whether the social process that the analysis measures is distinct from the processes inherent to political mobilization.

Methodologically, this analysis also offers several important advancements in how to address the problem of scale in a social science setting. As an analogous problem exists in many inductive algorithms, including all clustering algorithms (like text-as-data Topic Models), these innovations have implications beyond network analysis.

Finally, this analysis points to a number of potential avenues for future research. For example, the finding that the network–ethnicity relationship is weaker in urban communities than rural communities begs for further investigation, as an improved understanding of the mechanisms behind this attenuation may be of value to policymakers seeking to decrease the political salience of ethnicity. And more broadly, this analysis illustrates how new sources of *big data* on the day-to-day interactions of citizens can be leveraged to rigorously answer questions about social organization, civil society, and political outcomes. Since at least the works of Putnam *et al.* (1993) and Almond and Verba (1989), scholars have posited that social capital and an effective civil society are crucial to good democratic governance because of their role in holding politicians to account. Yet, most research in this area has been unable to systematically measure social capital. Numbers of NGOs or membership in formal organizations are often-used measures that fail to capture the richness of the theories that they aim to test. This analysis illustrates how newly available data like cell-phone meta-data can be leveraged to develop empirical measures that are as rich as the theoretical literatures we aim to test.

References

- Afrobarometer. 2012. *Afrobarometer Data, Round 5, Zambia*.
- Alesina, A., R. Baqir, and W. Easterly. 1999. “Public Goods and Ethnic Divisions”. *The Quarterly Journal of Economics* 114(4): 1243.
- Alesina, A. and E. La Ferrara. 2005. “Ethnic Diversity and Economic Performance”. *Journal of Economic Literature* 43(3): 762–800.
- Almond, G. A. and S. Verba. 1989. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. London: SAGE Publications.
- Banerjee, A., A. Chandrasekhar, E. Duflo, and M. O. Jackson. 2014. “Gossip: Identifying Central Individuals in a Social Network”. *SSRN Working Paper Series* 20422.
- Barthélemy, M. 2011. “Spatial Networks”. *Physics Reports* 499(1–3): 1–101.
- Basedau, M., G. Erdmann, J. Lay, and A. Stroh. 2011. “Ethnicity and Party Preference in Sub-Saharan Africa”. *Democratization* 18(2): 462–89.

- Berge, L. I. O., K. Bjorvatn, S. Galle, E. Miguel, D. N. Posner, B. Tungodden, and K. Zhang. 2015. "How Strong Are Ethnic Preferences?" *SSRN Working Paper Series* 21715.
- Besley, T. 2007. *Principled Agents? The Political Economy of Good Government*. USA: Oxford University Press.
- Björkegren, D. 2015. "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda". *SSRN Working Paper Series*.
- Björkegren, D. and D. Grissen. 2015. "Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment". *World Politics*.
- Blondel, V., G. Krings, and I. Thomas. 2010. "Regions and Borders of Mobile Telephony in Belgium and in the Brussels Metropolitan Zone". *Brussels Studies*.
- Blumenstock, J. E. 2012. "Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda". *Information Technology for Development* 18(2): 107–25.
- Blumenstock, J., G. Cadamuro, and R. On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata". *Science* 350(6264): 1073–6.
- Casey, K. 2016. "Snap Judgments: Forecasting Politician Productivity under Limited Information". *NBER Working Paper Series* 3360 (October).
- Casey, K., R. Glennerster, and E. Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan". *Quarterly Journal of Economics* 127(4): 1755–812.
- Chandra, K. 2007. *Why Ethnic Parties Succeed. Patronage and Ethnic Head Counts in India*. Cambridge University Press.
- Charness, G. and M. Rabin. 2002. "Understanding Social Preferences with Simple Tests". *The Quarterly Journal of Economics* 117(3): 817–69.
- Chen, Y. and S. X. Li. 2009. "Group Identity and Social Preferences". *The American Economic Review* 99(1): 431–57.
- Chikulo, B. 1988. "The Impact of Elections in Zambia's One-Party Second Republic". *Africa Today* (2nd Quarter): 37–49.
- Cox, G. W. 2015. "Electoral Rules, Mobilization, and Turnout". *Annual Review of Political Science* 18(1): 49–68.
- Desmet, K., I. Ortuno-Ortin, and R. Wacziarg. 2015. "Culture, Ethnicity and Diversity". *NBER Working Paper Series* 20989.
- Dionne, K. Y. 2015. "Social Networks, Ethnic Diversity, and Cooperative Behavior in Rural Malawi". *Journal of Theoretical Politics* 27(4): 522–43.
- Easterly, W. and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions". *Quarterly Journal of Economics* 112(4): 1203–50.
- Ejdemyr, S., E. Kramon, and A. L. Robinson. 2015. "Segregation, Ethnic Favoritism, and the Strategic Targeting of Local Public Goods". <http://stanford.edu/~ejdemyr/docs/EKR-20151103.pdf>.

- Erdmann, G. 2007. "Ethnicity, Voter Alignment and Political Party Affiliation — An African Case: Zambia". *German Institute of Global and Area Studies Working Papers* 45.
- Fafchamps, M. and P. C. Vicente. 2013. "Political Violence and Social Networks: Experimental Evidence from a Nigerian Election". *Journal of Development Economics* 101: 27–48.
- Fearon, J. D. 1999. "Why Ethnic Politics and "pork" Tend to Go Together". In: *Conference on Ethnic Politics and Democratic Stability*.
- Fearon, J. D. and D. Laitin. 1996. "Explaining Interethnic Cooperation". *American Political Science Review* 90(4): 715–35.
- Ferraz, C. and F. Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes". *The Quarterly Journal of Economics* 123(2): 703.
- Gerber, A. S., D. P. Green, and C. W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment". *American Political Science Review* 102(1): 33–48.
- Granovetter, M. S. 1973. "The Strength of Weak Ties". *American Journal of Sociology* 78(6): 1360–80.
- Grossman, S. and D. Honig. 2015. "These Are My People: Ethnic and Class-based Discrimination in Lagos". http://danhonig.info/sites/default/files/these_are_my_people_-_grossman_and_honig_2015.pdf.
- Habyarimana, J., M. Humphreys, D. N. Posner, and J. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101(4): 709.
- Habyarimana, J., M. Humphreys, D. N. Posner, and J. Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. New York: Russell Sage Foundation.
- Harris, A. S. and M. G. Findley. 2012. "Is Ethnicity Identifiable?" *Journal of Conflict Resolution* 58(1): 4–33.
- Horowitz, D. L. 1985. *Ethnic Groups in Conflict*. University of California Press.
- Ioannidis, J. P. 2005. "Why Most Published Research Findings Are False". *PLoS Medicine* 2(8): e124.
- Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan. 2012. "Social Capital and Social Quilts: Network Patterns of Favor Exchange". *The American Economic Review* 102(5): 1857–97.
- Jackson, M. O., B. W. Rogers, and Y. Zenou. 2017. "The Economic Consequences of Social-Network Structure". *Journal of Economic Literature* 55(1): 49–95.
- Kashoki, M. E. 2017a. "Between-Language Communication in Zambia". In: *Linguistic Surveys of Africa: Language in Zambia*. Ed. S. Ohannessian and M. E. Kashoki. New York: Routledge.

- Kashoki, M. E. 2017b. "The Language Situation in Zambia". In: *Linguistic Surveys of Africa: Language in Zambia*. Ed. S. Ohannessian and M. E. Kashoki. New York: Routledge, 9–46.
- Keefer, P. 2010. "The Ethnicity Distraction: Political Credibility, Clientelism and Partisan Preferences in Africa". *World Bank Policy Research Working Paper Series* 5236.
- Khwaja, A. I. 2009. "Can Good Projects Succeed in Bad Communities?" *Journal of Public Economics* 93(7): 899–916.
- Kolaczyk, E. D. 2005. *Statistical Analysis of Network Data*. New York, NY: Springer Science and Business Media.
- Laitin, D. 1992. *Language Repertoires and State Construction in Africa*. New York: Cambridge University Press.
- Laitin, D. 1998. *Identity in Formation: The Russian-Speaking Populations in the Near Abroad*. Ithaca, New York: Cornell University Press.
- Laitin, D. 2013. "Fisheries Management". *Political Analysis* 21(1): 42–7.
- Larson, J. M. 2017a. "Networks and Interethnic Cooperation". *Journal of Politics* 79(2): 546–59.
- Larson, J. M. 2017b. "Why the West Became Wild". *World Politics* 69(4): 713–49.
- Larson, J. M. and J. I. Lewis. 2017. "Ethnic Networks". *American Journal of Political Science* 61(2): 350–64.
- Leskovec, J. and C. Faloutsos. 2006. "Sampling from Large Graphs": 631–6.
- Lieberman, E. S. and G. H. McClendon. 2012. "The Ethnicity–Policy Preference Link in Sub-Saharan Africa" 46(5): 574–602.
- McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology* 27(1): 415–44.
- Melson, R. 1971. "Ideology and Inconsistency: The "Cross-Pressured" Nigerian Worker". *American Political Science Review* 65(1): 161–71.
- Miguel, E. and M. K. Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya". *Journal of Public Economics* 89(11–12): 2325–68.
- Olken, B. 2015. "Promises and Perils of Pre-analysis Plans". *The Journal of Economic Perspectives* 29(3): 61–80.
- Onnela, J.-P., J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. 2007. "Structure and Tie Strengths in Mobile Communication Networks". *Proceedings of the National Academy of Sciences* 104(18): 7332–6.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science". *Science* 349(6251): aac4716–aac4716.
- Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.
- Posner, D. N. 2004. "Measuring Ethnic Fractionalization in Africa". *American Journal of Political Science* 48(4): 849–63.

- Posner, D. N. 2005. *Institutions and Ethnic Politics in Africa*. Cambridge: Cambridge University Press.
- Posner, D. N. and D. Simon. 2002. "Economic Conditions and Incumbent Support in Africa's New Democracies: Evidence from Zambia". *Comparative Political Studies* 35(3): 313–36.
- Putnam, R. D., R. Leonardi, and R. Nanetti. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, New Jersey: Princeton University Press.
- Reinikka, R. and J. Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda" 119(2): 679.
- Research ICT Africa. 2008. *ICT Household Survey*.
- Rojo, G. and E. Wibbels. 2014. "Clientelism, Public Services and Elections in the Slums of Jaipur and Udaipur". *Presented at Stanford Comparative Politics Seminar*.
- Siegel, D. A. 2009. "Social Networks and Collective Action". *American Journal of Political Science* 53(1): 122–38.
- Thachil, T. 2015. "Do Internal Migrants Divide or Unite across Ethnic Lines? Ethnographic and Experimental Evidence from Urban India". *SSRN Working Paper Series* 2455564.
- Weiner, M. 1978. *Sons of the Soil: Migration and Ethnic Conflict in India*. Princeton: Princeton University Press.
- Wesolowski, A., N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee. 2012. "Quantifying the Impact of Human Mobility on Malaria". *Science* 338(6104): 267–70.
- Wolitzky, A. 2012. "Cooperation with Network Monitoring". *The Review of Economic Studies* 80(1): 395–427.
- Zambian Central Statistics Office. 2013. *Migration and Urbanisation Analytical Report*. Technical report.