Who Flees Conflict?

A Big-Data Approach to the

Determinants of Forced Migration[*]

Aidan Milliff [†] & Fotini Christia [‡]

May 22, 2023

Abstract

Every year, millions of people encounter political violence and some leave their homes in pursuit of safety. What factors influence whether or not a person migrates to escape conflict? Existing research focuses on a range of factors, from economic resources and political ideology to personality traits and violence intensity. The influence of social networks on migration decisions, however, is often overlooked because empirical social networks are very difficult to measure. We use data from 63.5 million anonymized, geo-located cell phone records to study the migration behavior of over 55,000 people during a 2011-2012 conflict between the Yemeni Government and Salafist insurgents. Our unique dataset allows us to reconstruct cell subscribers social networks, and show that the structure of individuals' social and physical networks are important predictors of migration. We show that not all social ties are equal: broad social networks vs. social networks with influential members have opposite marginal associations with probability of migration. We also show that network structure is correlated with important variables in the literature: Including measures of social network structure changes our understanding of how economic resources and violence exposure are related to migration potential. Our results demonstrate that social ties are important for understanding civilian migration and resilience during conflict.

Word Count: 3,896

[*]Thank you to Paolo Bertolotti, Michael Freedman, Adam Kaplan, Chagai Weiss and seminar participants at ESOC, MIT, and UC Davis for feedback. We appreciate HPC support from MIT Supercloud and further resources from the MIT Political Methodology Lab. This paper builds on results from a collaborative project with Michael Curry, Constantinos Daskalakis, Erik Demaine, John P. Dickerson, MohammadTaghi Hajiaghayi, Adam Hesterberg and Marina Knittel. We thank them all. The project is registered with the MIT Committee on the Use of Humans as Experimental Subjects both as protocol #1309005907 and exemption determination E-2274.

[†]Shorenstein Asia-Pacific Research Center, Stanford University. Contact: milliff.a@gmail.com.

[‡]Department of Political Science and IDSS, Massachusetts Institute of Technology. Contact: cfotini@mit.edu.

As of 2022, one out of every 78 people in the world was displaced from their home by conflict (United Nations High Commissioner for Refugees, 2023). Forced displacement represents a massive humanitarian challenge, and also promises to reshape economies (Taylor et al., 2016; Tumen, 2016), security (Salehyan, 2008; Couttenier et al., 2019), and socio-political life (Balcells, 2018; Rosenzweig and Zhou, 2021) in destination locations. So, how do people decide whether to flee their homes during conflict? People living in conflict zones often face circumstances in which both remaining and fleeing carry substantial, hard to measure risks. Previous empirical research on conflict-induced migration finds support for different explanations in different cases: more people leave areas where violence is more intense (Davenport, Moore and Poe, 2003; Moore and Shellman, 2004; Czaika and Kis-Katos, 2009; Blumenstock et al., 2020; Shaver, Krick, Blancaflor, Liu, Samara, Hu, Ku, Carreon, LIm, Raps, Velasquez, Angelo, de Melo and Zuo, 2022; Tai, Mehra and Blumenstock, 2022), people with more (liquid) economic resources are more likely to flee (Schon, 2019). Better social resources are sometimes associated with remaining (Adhikari, 2013; Marston, 2020) and sometimes with leaving (Schon, 2020). More satisfaction with governance predicts remaining (Revkin, 2019), and higher individual risk tolerance may either promote remaining (Mironova, Mrie and Whitt, 2019) or fleeing (Camarena et al., 2020).[1]

We model forced migration using a data source that provides a different perspective compared to survey- and interview-based studies that have been the main source of existing evidence (excepting Tai, Mehra and Blumenstock, 2022). We use over 63.5 million anonymized, time- and location-stamped Call Detail Records (CDRs, often called cell phone metadata), for over 55,000 cell subscribers in Yemen's Abyan governorate—roughly 10% of Abyan's total population—as they navigate violence and fluid territorial control in Abyan during 2011 and 2012. We use techniques developed in the applied computer science and economics literature to agument CDR data, creating measures of subscribers' migration behavior, social behavior, economic status, religiosity, and exposure to violence. We then test new and existing theories of migration using data that was collected in real time and is therefore not vulnerable to faulty memories or post-treatment interference. We focus on assessing the importance of social networks and mobility in understanding migration decisions, because the data structure allows us to know more than any person could recount about characteristics of their social network, or the exact times at which they were in particular locations.

In single-shot and time-series models of displacement, we find that previously un-measured factors like social network structure have large, robust associations with propensity to migrate. We also find

[1] We cite and discuss the existing literature more thoroughly in Table S1.

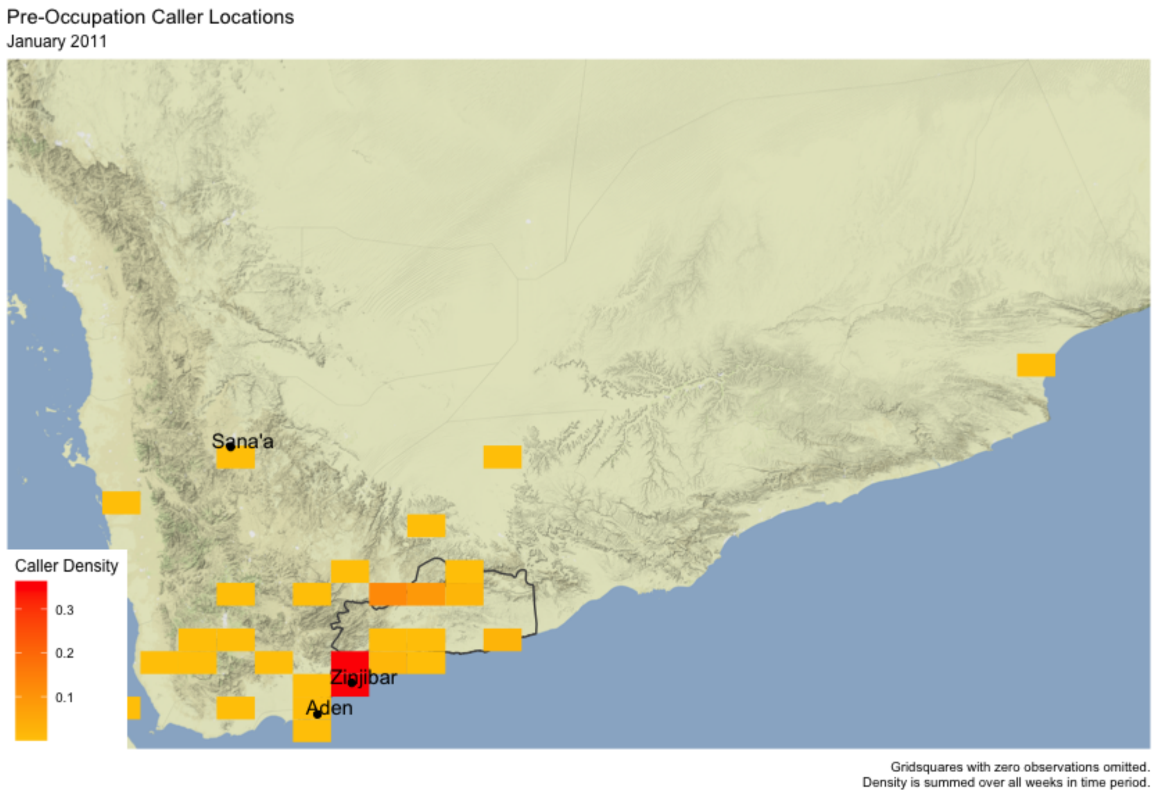
some support for existing theories that privilege the role of violence exposure and resources in determining migration decisions. Finally, we find much smaller robust associations between migration propensity and pre-migration mobility, another factor under-explored in previous literature. Including new social network factors changes the interpretation of variables like resource access and violence exposure, which are the main focus for existing theories of migration.

Social network structure and mobility matter in nuanced ways that deserve more attention (Munshi, 2020). We find large and opposite-signed marginal effects for different measures of network centrality: people who occupy more "influential" positions in pre-conflict communication networks are likelier to migrate during conflict. People who have overall broader networks, though, are more likely to remain. We also see that people who are on average more mobile *during* conflict, and those whose calling networks are geographically wider before conflict, are slightly more likely to migrate. These findings shed new light on the social dynamics of conflict migration. By showing that socially-central people are likely migrants, we provide new insight into the possible causes of the "tipping point" phenomenon in refugee crises: When influential people migrate, they might simultaneously a) inspire others in their networks, and b) decrease the potential resilience of the community they leave behind. In this article, we introduce our approach to modeling migration decisions, describe the data sources, and present results, in which we identify new precursors of forced migration.

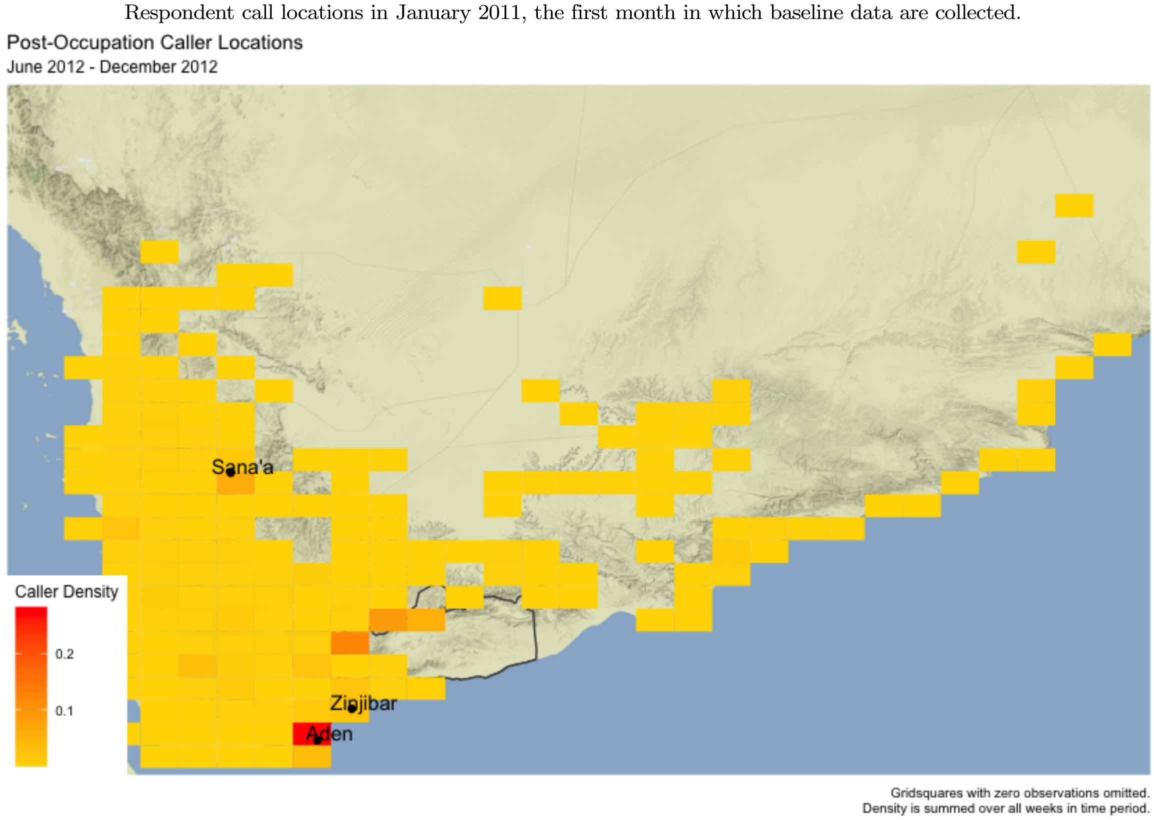
Yemen's War: 2010-2012

We use CDR data to study migration caused by a 2011-2012 conflict between the Yemen government and Al-Qaeda in the Arabian Peninsula (AQAP). The broader conflict stretches back into the late 1990s, but first erupted into serious violence with a 2009 government offensive against the newly-unified AQAP (Uppsala Conflict Data Program, N.d.). In March 2011, AQAP fighters captured major population centers in Abyan, a governorate in Southwestern Yemen just east of the port city of Aden (see Figure S1). Government and AQAP forces clashed repeatedly for the next year, causing over 2,500 recorded battle deaths (See Figure S2 for a heat map of violence events). Recent research suggests that this may under-estimate the severity of the conflict, because violence data originally gathered from news sources is often biased downward Shaver, Arriaga, Blackburn, de Kadt, Farrington, Felter, Hyneman, Leal Silva, Lorch, McLean, Prada, Tobia and Weintraub (2022).

Violence displaced as much as 45% of Abyan's pre-conflict population of 525,000 residents. Most moved to the adjacent Lahij and Aden governorates in Yemen (UNOCHA, 2013). In our data, the



Respondent call locations in January 2011, the first month in which baseline data are collected.



Respondent call locations after the end of the occupation, from June 2012- December 2012.

Figure 1: Subscriber locations are binned by gridsquares, which are shaded by density. Abyan residents were identified from a larger dataset of 6 million subscribers via their calling behavior in the "baseline month" of January 2011. Only subscribers with Abyan "home locations," i.e. the tower modal cell tower through which they connect, are tracked as residents in our study. After 15 months of conflict, pre-occupation Abyan residents are dispersed across the country, but many remain in Abyan.

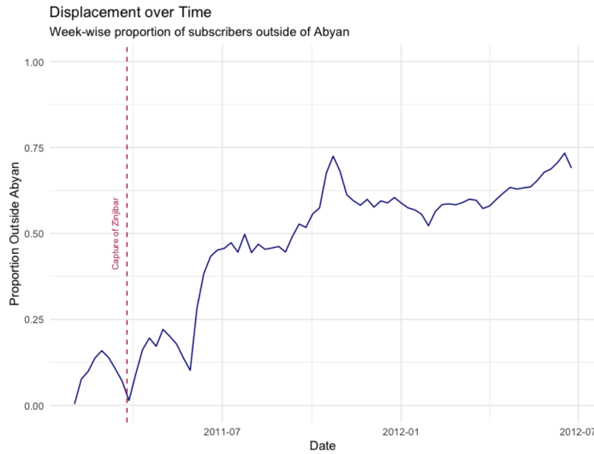


Figure 2: Weekly proportion of subscribers outside Abyan. Vertical, dashed line indicates start of conflict.

weekly displacement rate reaches as high as 70% at the height of the government offensive in 2012, but is not strictly increasing over time: Some people move in and out of Abyan repeatedly during the conflict (see Figure 2). Consistent with contemporaneous local reporting, our data show that large numbers of displaced people land in Aden, only about 65 kilometers away from Zinjibar, Abyan's capital (Compare the pre-violence distribution in the top image of Figure 1 to the post-violence distribution in the bottom image of Figure 1). Once in Aden, many IDPs found cramped temporary accommodations in school buildings, with multiple extended families sharing a single classroom. Housing IDPs in schools had knock-on effects that lasted beyond the AQAP-Government conflict in Abyan: Schooling in Aden was disrupted or irregular well into 2013 due to lack of space (IRIN News, 2013).

Social Networks and Forced Displacement

Social networks have been studied relatively little as a determinant of migration—likely due to the limitations of survey- and interview-based data that have yielded contradictory results in previous research (Adhikari, 2013; Schon, 2019). All the same, we have reason to believe that social ties generally are an important resource for civilians navigating conflict (Wood, 2008; Harpviken, 2009; Marston, 2020) just as they have been show to matter for decisions about economic migration (Munshi and Rosenzweig, 2016; Munshi, 2020).

Not all social relations are created equal. We posit that civilians facing violence need different types of social ties to facilitate fleeing from conflict vs. remaining. Leaving a conflict-affected area might require a single instance of help or a large favor that civilians source through their social networks. Remaining despite violence may draw on different social ties. Instead of a rolodex of important or useful contacts, remaining in place might depend on broader networks that facilitate the type of repeated mutual assistance necessary to survive conflict.

We argue that different measures of *centrality* within a given social network reflect these two types of "strength," which civilians leverage to flee or remain. First, people who are influential, or have influential connections in a network should have high eigencentrality or PageRank (Page et al., 1998). Having influential connections gives people a social resource that they can leverage to facilitate out-migration during conflict. Second, people whose social networks are strong in other ways—people who "know everyone," for example—should have high degree centrality. Wider networks should, in the strength of weak ties spirit, better support resilience in place (Finkel, 2017; Blumenstock, Chi and Tan, 2019; Marston, 2020), or may be an artifact of local leadership roles which separately disincentivize fleeing.

Empirical social networks are hard to accurately measure via common techniques like interviews or surveys. We use Call Detail Records (CDRs) from a major cell provider in Yemen to generate better measures of network structure and use those better measures to test our expectations about social structure and migration in the 2011-2012 Abyan conflict. We exploit the graph structure of the CDR data to calculate a variety of social network metrics which disaggregate different types of network "strength" including PageRank and degree centrality (See SI and Table S2). Individual-level cell data are somewhat novel for studying political violence (Freedman et al., 2021; Blumenstock et al., 2020; Tai, Mehra and Blumenstock, 2022),[2] but communication technology is widely recognized as a tool for measuring (and influencing) conflict behavior (Dafoe and Lyall, 2015; Christia et al., 2015; Walter, 2017). Beyond political science, researchers in economics and computer science have developed tools to *augment* the narrow set of attributes in raw communications data to study social and political phenomena (Christia et al., 2021, among others). Previous work has predicted individual economic status from CDRs, and validated predictions against direct measurement of wealth Blumenstock, Cadamuro and On (2015); Felbo et al. (2017). Other work has created a measure of religiosity specific to users in Muslim-majority countries Bozcaga et al. (2019). In conflict contexts, ongoing work uses CDRs to measure firm responses to violence Blumenstock et al. (2020). We draw on some of these results to augment CDR data in our models of migration.

[2]Though the effects of cell coverage have been studied (Shapiro and Weidmann, 2013; Shapiro and Siegel, 2015)

Table 1: Extract from raw Call Detail Record (CDR) data showing data format. The 63.5 million CDRs are provided with anonymous, unique, consistent subscriber IDs for callers (C) and recipients (R), call time and duration, and caller/recipient tower which can be used as a coarse measure of location.

Time	**Subscriber ID**		**Tower ID**		**Duration**
Start of Call	C	R	C	R	(seconds)
2012-06-13 7:57:12	#####	#####	3995	623	1139
2012-06-13 7:57:13	#####	#####	1373	1290	10
2012-06-13 7:57:13	#####	#####	3338	1401	32

Materials and Methods

Data

To test our network hypotheses, as well as pre-existing theories in the migration literature, we augment raw CDRs to generate data about individuals' migration status, exposure to violence, social milieu, socioeconomic status, religious observance, and surroundings in Abyan. We then combine these data with geo-located conflict event records and use these augmented data to model displacement from Abyan in the 2011-2012 conflict. CDRs are collected by all cell phone companies around the world, and their anonymized forms are increasingly used in social science research. Our data are provided with anonymous, unique, consistent subscriber IDs that allow us to follow particular subscribers over time.[3] In addition to timestamp, duration, and subscriber IDs for both caller and recipient (or texter and recipient), CDRs include tower IDs for both parties, which can be matched with tower location data to identify coarse caller/recipient locations at the time of the communication (See Table 1).[4]

We build our dataset by identifying residents of Abyan as of January 2011. For callers who connect to a tower in Abyan at least once in January 2011, we calculate four weekly "home locations." We discard subscribers whose home locations are outside of Abyan at any point in January, and use the remaining 56,170 subscribers as our sample of pre-conflict Abyan residents. Assuming a rough correspondence of one phone line to one person, our sample includes nearly 10% of the total pre-conflict population of Abyan.[5] We extract all call records associated with each resident (as either caller or recipient) and use those calls to build the rest of our data. Throughout the paper, individual units are always subscribers (people). Their activities are aggregated at the week level using appropriate summary statistics like modal tower visited, count of calls made, mean call duration, etc. Our spatial unit is always the weekly modal tower; all spatial covariates like

[3] See discussion in SI. To maintain anonymity, we never present data at the individual level.

[4] See SI for more detail and comparison to modern GPS-based location tracking.

[5] This already dwarfs coverage from even the largest surveys. Our sample could include members of up to half the households in Abyan under the assumption that cellphones are a good purchased by households, not individuals.

location of violent events, grid-square economic output, etc. are resolved to the relevant tower(s). Tables S2 and S3 describe data augmentation for economic status, violence exposure, religiosity, and other variables.

Measuring Displacement

Our main measure of migration or displacement is a binary variable, calculated at the level of the subscriber-week, indicating whether or not the subscribers' weekly home location (modal tower used) is inside or outside of Abyan governorate. This is an appropriate and conservative proxy measurement for displacement for our study, even though crossing an administrative boundary is not a necessary component of internal displacement or conflict-induced migration. First, given the pattern of violence in Abyan in 2011-2012, most displaced persons did leave for a different governorate like Lahij or Aden where territorial control was not contested or changing. Second, setting a high bar for "migration" gives reasonable assurance that we are not over-counting migration by capturing people who are simply traveling or living mobile lives in Abyan. Supplementary mobility statistics—which we include as covariates to address known sources of measurement error in CDR data (Gonzalez, Hidalgo and Barabási, 2008)—are described in Supplementary Information. In Supplementary Information, we also include results using different functional forms like OLS (See Table S14, Figures S10–S11), and results that model "migration" using a hidden markov process rather than a binary, location-based variable (See Figures S12–S16) Both produce similar results.

Measuring Social Networks

We calculate many different measures of *social centrality*, focusing on two that have distinct substantive interpretations. One simple measure is degree centrality: Of all possible connections, what proportion of other people (nodes) do you share a personal connection (edge) with in the social network (graph)? Degree centrality measures network breadth—how many people a person "knows." It reflects popular intuition about well-networked individuals in a social network. Another measure, PageRank, captures influence by figuratively scoring nodes based on their own degree, and the degree of every node that has a directed edge *toward* them. Nodes have higher PageRank (are more influential) when they are referred-to by many other nodes that are themselves influential. High PageRank nodes, in other words, have "bridging power" (Berman, 1997) or influence over the broader network, not just their own connections. Previous work supports the choice of PageRank as a measure of "importance." PageRank expands on the intuition of betweenness centrality—how often a particular node appears on the shortest path between two randomly selected nodes—which historical research has shown to be a useful measure of social influence (Padgett and Ansell, 1993). More recent work has also shown that eigencentrality, a non-directed variant of PageRank, is associated with political and social importance in data from the Philippines (Cruz, Labonne and Querubín, 2017).

Centrality measures are calcuated once per subscriber, using a time-limited graph of *pre-occupation* calls and texts. In addition to centrality, we use software developed by Montjoye et al. (de Montjoye, Rocher and Pentland, 2016) to measure other attributes of a subscribers' social network and social behavior—including percent pareto interactions[6] and call count—on a weekly basis throughout the conflict.

Constructed Covariates

Beyond network statistics and mobility, we include a number of variables that existing migration literature identifies as important causes of displacement. To measure economic status, we rely on a suite of different proxies developed in the Call Detail Records literature Blumenstock, Shen and Eagle (2010); Blumenstock, Cadamuro and On (2015); Felbo et al. (2017). We focus on measures that are not socially mediated—i.e. do not measure the same things as our network centrality measures. First, we proxy wealth using call duration and percent initiated calls, because average call time and calling vs. receiving are positively associated with monetary resources in countries with pre-paid cell service. Second, we measure clustering coefficient—how close your network is to a complete graph—because CDR literature consistently identifies this graph statistic as negatively associated with wealth. We also use measures of local economic productivity, constructed from satellite data described in the SI.

Beyond economic variables, we use CDR proxies for subscriber religiosity (Bozcaga et al., 2019), a potentially important feature in a conflict where one major party espouses an extremely religiously conservative ideology called Salafism. Recent research on Iraq has shown that satisfaction with Salafist group governance is an important predictor of migration during conflict (Revkin, 2019). Again, the link between calling behavior and religiosity is described in SI. Finally, political violence is a key characteristic of Abyan during our period of study, and exposure to violence is a prominent explanation for migration behavior. We use the Uppsala Conflict Data Program Global Events Dataset (GED, Uppsala Conflict Data Program, N.d.) as a source of geo-located violence data. For the 400-plus GED events in Abyan in 2011-2012, we identify all cell towers within 10 miles of the GED event coordinates and label those towers as "affected" by the violent event in the week it occurred. We identify over 41,000 tower-event-week triplets. We also include latitude and longitude which, in a conventional conflict, are associated with violence exposure.

[6] Percentage of user's contacts accounting for 80% of interactions.

$$Y_i = \alpha + \xi K + \delta J + \beta \mathbf{X}_i \tag{1}$$

$$Y_{it} = \alpha_i + \gamma_t + \beta \mathbf{X}_{it} \tag{2}$$

Figure 3: **Above:** Functional form for time-invariant linear model with tower and district random effects, producing results shown in Figure 4.**Below:** Functional form for time-series model of subscriber location with subscriber and week random effects. Produces results shown in Figure 5.

Estimation

We use two different routines to model displacement from Abyan. First, we fit a time-invariant linear model with tower and district random effects (Equation 1). The dependent variable is the proportion of all weeks in the conflict that a subscriber spent outside Abyan. This serves as the best comparison to existing individual-level studies, which mostly model conflict migration as a "one-shot" process. More detail on model specification is in SI.

Second, we fit a time-series model to capture the effects of over-time variation in subscribers' location, behavior, and environment while also measuring stable network characteristics (see Equation 2). Here, the dependent variable is a binary indicator for whether the subscriber is inside or outside of Abyan in a given week. We specify a logit model at the subscriber-week level with subscriber and week random effects and use Stan's Hamiltonian Monte Carlo sampler to estimate the determinants of migration (See Figures S17 - S23 for information and diagnostics). Figure 5 shows first difference estimates from the random effects logit model. Tables S7-S11 shows raw parameter estimates. Additional results, including a model that only includes time-invariant network predictors for ease of interpretation, are also shown in the SI.

Computation

Results presented in Figures 4 and 5, as well as results in SI are estimated using the `Rstan` package, and other elements of the mc-stan modeling language ecosystem The Stan Development Team (2021). The main estimation was run on the MIT Supercloud high performance computing system Reuther et al. (2018). We used 18 cores on a single Intel Xeon Gold node with 384GB of RAM.

Results

Single-Shot Model

Results in Figure 4 shows that including rarely-measured network characteristics substantially alter canonical associations between migration behavior and factors like wealth and violence. Coefficients in Figure 4 estimate the association of standard deviation change in the predictor variable with the number of conflict weeks outside Abyan. The strongest predictors of migration are social: Pre-conflict network importance (Pagerank) and mobility (radius of gyration) predict migration, and degree centrality (network breadth) is strongly negatively associated with migration.[7] Results also show that respondents who have wider pre-conflict circles of contacts and higher numbers of total contacts are more likely to be outside of Abyan during the conflict.

Two proxies for wealth, a canonical predictor of migration, are insignificantly associated with time outside of Abyan—coefficients for local GDP where the subscriber resides and subscriber-level percent initiated interactions are noisy. A third individual-level measure of poverty—network clustering coefficient—has a small negative association with migration, in line with predictions from the literature. Including network metrics not only changes coefficients, but also explains more variation. Table S12 shows that the the fully-saturated model has a marginal R^2 that is 1.78 times higher than an otherwise similar specification without measures of PageRank, degree centrality, and clustering coefficient.

Time-Series Model

The time-series model shown in Figure 5 matches single-shot results in some but not all ways. First, many network statistics (calculated based on pre-occupation calling behavior) maintain their predictive power. Degree centrality is substantially negatively associated with migration, and PageRank is substantially positively associated (again, see discussion with Table S13). The 25th to 75th percentile difference in each measure is associated with roughly a 15% lower probability of being outside Abyan and a 4% higher probability, respectively. Interestingly, the weekly number of contacts a person calls pre-occupation is substantially associated with migration: While the measures are highly correlated, having a wide network (degree centrality), seems subtly different from keeping up with a wide network (number of contacts). Once over time variation is modeled, we recover canonical associations between wealth and migration—local GDP shows a large positive association, and clustering coefficient (a proxy measure of poverty) is

[7] These centrality measures are conceptually related and empirically correlated. Tables S12 and S13 compare models omitting each, and support our interpretation focused on "importance" vs. "breadth." Also see Figures S3 and S4.

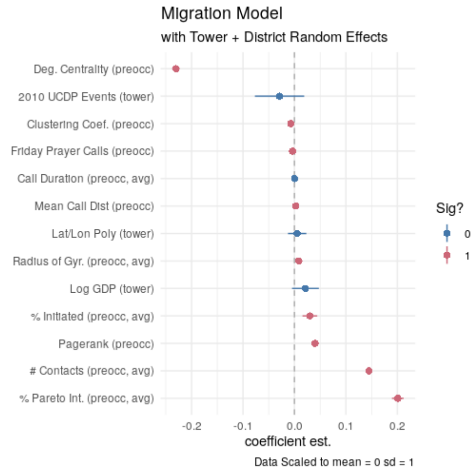


Figure 4: Coefficient estimates (95% CIs) from a time-invariant linear model with district and tower random effects. Data are scaled and demeaned, so estimates represent change in total time spent outside Abyan associated with a one standard-deviation change in a predictor. See Table S11 for un-scaled predictors. Error bars show 99% credible intervals from `stan`. Supplemental results in Figure S5 show consistent results from a less-saturated model that includes only the network centrality measures.

negatively associated.[8] We also see significant but substantively smaller associations between displacement and behavior *during* conflict. Migrants mobility is slightly but significantly higher. Finally, the model shows no substantial association between a (short) lagged measure of localized violence and migration.[9] Geographic patterns in migration (latitude/longitude coefficients) indicate more out-migration from areas in southwestern Abyan, which were more violent. Violence appears to play a large role in migration decisions, but people seem to respond to longer term trends in violence, not short-term shifts at the margin.

Discussion

Our CDR results show that social network structure—overlooked or coarsely measured in much of the existing literature on migration—plays a substantial role in forced displacement during conflict. Evidence from Yemen also adds interesting nuances to a number of canonical findings in the migration literature. We show, expectedly, that high-violence areas are the places that cell subscribers are most likely to leave (see Tai, Mehra and Blumenstock, 2022), but that after accounting for geographic location (highly correlated with long-term violence, in the Abyan conflict) weekly variation in violence has a vanishingly small association with migration propensity (see Fearon and Shaver, 2020).[10]

[8] Because the GDP covariate is a one-week lag, the association with *leaving Abyan* may be overstated if people tend to flee to prosperous areas, and stay once they get there (See SI).

[9] Violence events are rare, so first difference estimates for reasonable ranges are precisely zero. The raw odds ratio (see Figure S6) is significant, but minuscule compared to other associations.

[10] We support these interpretations with simpler models in the supplementary information, Table S14, Figures S10 and S11.

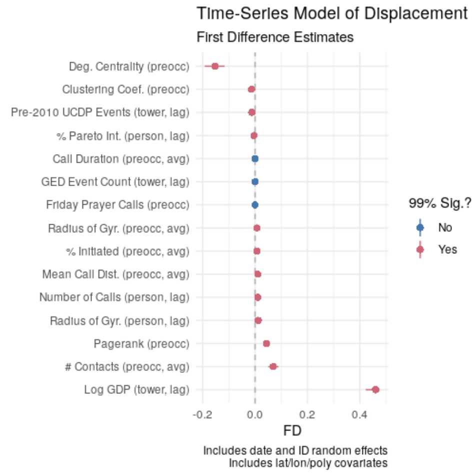


Figure 5: First-differences in weekly probability of "fleeing" in the time-series model with subscriber and week random effects, for the difference between a 25th- vs 75th-percentile value of the predictor. See Table S11 for un-scaled predictors. Error bars show 99% credible intervals from `stan`.

Time series models also show novel evidence of small "preparatory behaviors" that are leading indicators of a subscriber's appearance outside of Abyan. More mobility and making a greater number of calls overall are robustly associated with a slight increase in migration propensity. This may reflect people laying the groundwork to leave, perhaps reaching out to people outside of their geographic community to establish a "landing place." We find that having a geographically-broader social network (measured by pre-conflict call distance) is slightly associated with migration as well. While intuitive—a person is increasingly likely to leave in the strength of their ties to distant places—this finding is novel in quantitative literature. We argue that these results reflect plausible community-level social processes during conflict, and that the effects of social structure especially are a promising avenue of study for learning more about cascade dynamics in migration and other conflict-relevant behaviors (Kuran, 1989).

Overall, our results make three key contributions. First, we identify social network structure and network centrality as a new, important suite of explanations for migration during conflict. We extend the intuition that social ties shape civilian behavior during war (Finkel, 2017), and demonstrate a unique insight: that different types of "strong" social networks mean different things. People who are particularly influential as bridges/brokers in a network are more likely to leave, while people who, all else equal, have broadly-based social networks are more likely to remain. Second, demonstrating the importance of social network structure has broad implications for understanding the mechanisms that create "tipping points" in migration behavior arise during conflict. If the people likeliest and quickest to depart violence-affected communities are individuals with social influence, their behavior might likewise set an influential example, changing other people's social estimates of how safe it is to remain (Petersen, 2001). Third, our results demonstrate the potential for using "big data" to address humanitarian challenges. We show that a form of administrative data that is commonly available to governments and companies—but has only recently been used by researchers—can be used to identify patterns in civilian behavior during armed conflict. Some of the correlates of migration we identify, like social network structure and during-conflict mobility, should also be useful in a predictive framework, which would help practitioners and policymakers make better forecasts about the need for humanitarian aid.

Many questions about forced migration remain unanswered. Our findings, in conversation with recent work on related topics like bi-lateral flows (Azose and Raftery, 2019; Welch and Raftery, 2022), and destination "pull factors" add to that list: Why are different *types* of social status associated with different migration choices? What social signals—the information transmitted across social networks, which we do *not* measure—promote resilience versus fleeing (Blumenstock, Chi and Tan, 2019)? To what degree does the out-migration of influential individuals prompt cascading behavior by neighbors and contacts (Becker et al., 2021)? How do different social networks structures affect migrants' choice of destination (Steele, 2019; Munshi, 2020)? Addressing these questions will not only provide insight into critical decisions about migration, but also better knowledge for designing realistic, effective policy interventions that promote security, safety, and well-being for people caught in conflict.

References

Adhikari, Prakash. 2013. "Conflict-Induced Displacement, Understanding the Causes of Flight." *American Journal of Political Science* 57(1):82–89.

Azose, Jonathan J. and Adrian E. Raftery. 2019. "Estimation of emigration, return migration, and transit migration between all pairs of countries." *Proceedings of the National Academy of Sciences* 116(1):116–122.

URL: *https://www.pnas.org/doi/abs/10.1073/pnas.1722334116*

Balcells, Laia. 2018. "Dynamics of Internal Resettlement During Civil War: Evidence From Catalonia (1936-39)." *Journal of Peace Research* 55(2):236–251.

Becker, Sascha O., Volker Lindenthal, Sharun Mukand and Fabian Waldinger. 2021. Persecution and Escape: Professional Networks and High-Skilled Emigration from Nazi Germany. SoDa Laboratories Working Paper Series 2021-02 Monash University.

Berman, Sheri. 1997. "Civil Society and the Collapse of the Weimar Republic." *World Politics* 49(3):401–29.

Blumenstock, Joshua, Gabriel Cadamuro and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Data." *Science* 350(6264):1073–76.

Blumenstock, Joshua, Guanghua Chi and Xu Tan. 2019. Migration and the Value of Social Networks. Working paper U.C. Berkeley Berkeley, CA: .

URL: *http://www.jblumenstock.com/files/papers/migration$_n$etworks.pdf*

Blumenstock, Joshua, Tarek Ghani, Sylvan Herskowitz, Ethan Kapstein, Thomas Leo Sherer and Ott Toomet. 2020. How do Firms Respond to Insecurity? Evidence from Afghan Corporate Phone Records. Working paper University of California - Berkeley Berkeley, CA: .

Blumenstock, Joshua, Ye Shen and Nathan Eagle. 2010. A Method for Estimating the Relationship between Phone Use and Wealth. In *ICTD 2010*. ICT4 Collective London: .

Bozcaga, Tugba, Fotini Christia, Elizabeth Harwood, Constantinos Daskalakis and Christos Papademetriou. 2019. Syrian Refugee Integration in Turkey: Evidence from Call Detail Records. In *Guide to Mobile Data Analytics in Refugee Scenarios, The 'Data for Refugees Challenge' Study*, ed. Albert Ali Salah, Alex Pentland, Bruno Lepri and Emmanuel Letouzé. Springer pp. 223–249.

Camarena, Kara Ross, Sarah Claudy, Jijun Wang and Austin L. Wright. 2020. "Political and environmental risks influence migration and human smuggling across the Mediterranean Sea." *PLOS ONE* 15(7).

Christia, Fotini, Leon Yao, Stephen Wittels and Jure Leskovec. 2015. "Yemen Calling: Seven Things Cell Data Reveal about Life in the Republic." *Foreign Affairs* .

Christia, Fotini, Michael Curry, Constantinos Daskalakis, Erik Demaine, John P. Dickerson, Mohammad-Taghi Hajiaghayi, Adam Hesterberg, Marina Knittel and Aidan Milliff. 2021. Scalable Equilibrium Computation in Multi-agent Influence Games on Networks. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-21)*. Association for the Advancement of Artificial Intelligence Palo Alto: .

Couttenier, Mathieu, Veronica Petrencu, Dominic Rohner and Mathias Thoenig. 2019. "The Violent Legacy of Conflict: Evidence on Asylum Seekers, Crime, and Public Policy in Switzerland." *American Economic Review* 109(12):4378–4425.

Cruz, Cesi, Julien Labonne and Pablo Querubín. 2017. "Politician Family Networks and Electoral Outcomes: Evidence from the Philippines." *American Economic Review* 107(10):3006–37.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20150343*

Czaika, Mathias and Krisztina Kis-Katos. 2009. "Civil Conflict and Displacement: Village-Level Determinants of Forced Migration in Aceh." *Journal of Peace Research* 46(3):399–418.

Dafoe, Allan and Jason Lyall. 2015. "From cell phones to conflict? Reflections on the emerging ICT–political conflict research agenda." *Journal of Peace Research* 52(3):401–413.

Davenport, Christian, Will H. Moore and Steven Poe. 2003. "Sometimes You Just Have to Leave: Domestic Threats and Forced Migration, 1964-1989." *International Interactions* 29(1):27–55.

de Montjoye, Yves-Alexandre, Luc Rocher and Alex (Sandy) Pentland. 2016. "bandicoot: a Python Toolbox for Mobile Phone Metadata." *Journal of Machine Learning Research* 17(175):1–5.

Fearon, James D and Andrew Shaver. 2020. Civil War Violence and Refugee Outflows. Working paper, Stanford University Stanford, CA: .

Felbo, Bjarke, Pål Sundsøy, Alex 'Sandy' Pentland, Sune Lehmann and Yves-Alexandre de Montjoye. 2017. Modeling the Temporal Nature of Human Behavior for Demographics Prediction. Arxiv preprint MIT Media Lab Cambridge, MA: .

Finkel, Evgeny. 2017. *Ordinary Jews: Choice and Survival during the Holocaust.* Princeton, NJ: Princeton University Press.

Freedman, Michael, Fotini Christia, Spyros Zoumpoulis, Leo Yao and Ali Jadbabaei. 2021. "The Effect of Drone Strikes on Civilian Communication: Evidence from Yemen." *Political Science Research and Methods* .

Gonzalez, M.C., C.A. Hidalgo and Albert-László Barabási. 2008. "Understanding Individual Human Mobility Patterns." *Nature* 453:779–782.

Harpviken, Kristian Berg. 2009. *Social networks and migration in wartime Afghanistan.* London: Springer.

IRIN News. 2013. Schools start fresh after years of disruption in southern Yemen. Technical report The New Humanitarian.

Kuran, Timur. 1989. "Sparks and prairie fires: A theory of unanticipated political revolution." *Public Choice* 61(1):41–74.

Marston, Jerome. 2020. "Resisting Displacement: Leveraging Interpersonal Ties to Remain despite Criminal Violence in Medellin Colombia." *Comparative Political Studies* 53(13):1995–2028.

Mironova, Vera, Loubna Mrie and Sam Whitt. 2019. "Risk Tolerance During Conflict: Evidence from Aleppo, Syria." *Journal of Peace Research* 56(6):767–782.

Moore, Will H. and Stephen M. Shellman. 2004. "Fear of Persecution: Forced Migration, 1952-1995." *Journal of Conflict Resolution* 48(5):723–745.

Munshi, Kaivan. 2020. "Social Networks and Migration." *Annual Review of Economics* 12(1):503–524.

Munshi, Kaivan and Mark Rosenzweig. 2016. "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap." *American Economic Review* 106(1):46–98.

Padgett, John F. and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology* 98(6).

Page, Larry, Sergey Brin, Rajeev Motwani and Terry Winograd. 1998. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report SIDL-WP-1999-0120 Stanford University InfoLab Stanford, CA: .

URL: *http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf*

Petersen, Roger D. 2001. *Resistance and Rebellion: Lessons from Eastern Europe.* Cambridge: Cambridge University Press.

Reuther, Albert, Jeremy Kepner, Chansup Byun, Siddharth Samsi, William Arcand, David Bestor, Bill Bergeron, Vijay Gadepally, Michael Houle, Matthew Hubbell et al. 2018. Interactive supercomputing on 40,000 cores for machine learning and data analysis. In *2018 IEEE High Performance extreme Computing Conference (HPEC).* IEEE pp. 1–6.

Revkin, Mara. 2019. To Stay or to Leave? Displacement Decisions During Rebel Governance. Working paper Yale University New Haven, CT: .

Rosenzweig, Leah R and Yang-Yang Zhou. 2021. "Team and nation: Sports, nationalism, and attitudes toward refugees." *Comparative Political Studies* 54(12):2123–2154.

Salehyan, Idean. 2008. "The Externalities of Civil Strife: Refugees as a Source of International Conflict." *American Journal of Political Science* 52(4):787–801.

Schon, Justin. 2019. "Motivation and Opportunity for Conflict-Induced Migration: An Analysis of Syrian Migration Timing." *Journal of Peace Research* 56(1):12–27.

Schon, Justin. 2020. *Surviving the War in Syria: Survival Strategies in a Time of Conflict*. Cambridge: Cambridge University Press.

Shapiro, Jacob N and David A Siegel. 2015. "Coordination and security." *Journal of Peace Research* 52(3):312–322.

Shapiro, Jacob N. and Nils B. Weidmann. 2013. "Is the Phone Mightier than the Sword? Cell Phones and Insurgent Violence in Iraq." *International Organization* 36(2).

Shaver, Andrew, Benjamin Krick, Judy Blancaflor, Xavier Liu, Ghassan Samara, Shengkuo Hu, Sarah Yein Ku, Martha Carreon, Trishia LIm, Rachel Raps, Alyssa Velasquez, Joshua Angelo, Sofia de Melo and Zhanyi Zuo. 2022. The Causes and Consequences of Refugee Flows: A Contemporary Re-Analysis. Working paper University of California Merced Merced, CA: .

Shaver, Andrew C, Kately Arriaga, Anne Blackburn, Daniel de Kadt, Bryn Farrington, Joseph Felter, Caterina Hyneman, Alexandra Leal Silva, Talia Lorch, Jack McLean, Mateo Prada, Marie Tobia and Michael Weintraub. 2022. News Media Reporting Patterns and our Biased Understanding of Global Unrest. Working paper, University of California Merced Merced, CA: .

Steele, Abbey. 2019. "Civilian resettlement patterns in civil war." *Journal of Peace Research* 56(1):28–41.

Tai, Xiao Hui, Shikhar Mehra and Joshua E. Blumenstock. 2022. "Mobile phone data reveal the effects of violence on internal displacement in Afghanistan." *Nature Human Behaviour* 6(5):624–634.

Taylor, J. Edward, Mateusz J. Filipski, Mohamad Alloush, Anubhab Gupta, Ruben Irvin Rojas Valdes and Ernesto Gonzalez-Estrada. 2016. "Economic impact of refugees." *Proceedings of the National Academy of Sciences* 113(27):7449–7453.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.1604566113*

The Stan Development Team. 2021. "Stan Modeling Language Users Guide and Reference Manual,." https://mc-stan.org.

Tumen, Semih. 2016. "The Economic Impact of Syrian Refugees on Host Countries: Quasi-experimental

Evidence from Turkey." *American Economic Review* 106(5):456–60.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.p20161065*

United Nations High Commissioner for Refugees. 2023. Global Trends. Report United Nationsl high Commissioner for Refugees Geneva: .

UNOCHA. 2013. Yemen Humanitarian Response Plan 2013. Technical report United Nations Office for the Coordination of Humanitarian Affairs New York: .

Uppsala Conflict Data Program. N.d. Conflict Encyclopedia. Technical report Uppsala University.

Walter, Barbara. 2017. "The New New Civil Wars." *Annual Review of Political Science* 20:469–86.

Welch, Nathan G. and Adrian E. Raftery. 2022. "Probabilistic forecasts of international bilateral migration flows." *Proceedings of the National Academy of Sciences* 119(35):e2203822119.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.2203822119*

Wood, Elisabeth Jean. 2008. "The Social Processes of Civil War: The Wartime Transformation of Social Networks." *Annual Review of Political Science* 11(1):539–561.

Supporting Information for

Who Flees Conflict? A Big Data Approach to the Determinants of Forced Migration

Aidan Milliff and Fotini Christia

Address correspondence to Aidan Milliff.
E-mail: milliff.a@gmail.com

This PDF file includes:

Supporting text
Figs. S1 to S26
Tables S1 to S14
SI References

Supporting Information Text

Summary of Existing Migration Literature

Previous studies of migration during conflict have been carried out in conditions that are challenging both for research design and the practicalities of data collection. We summarize the findings of this literature in Table S1. Two issues are common across previous answers to the question of "who flees?" These issues may limit the strength of the conclusions that can be drawn from the impressive data collection efforts. First, an early wave of quantitative studies of forced migration—including theoretically generative studies by Davenport et al. (1) and Moore and Shellman (2)—primarily use time-series cross-national data to explain the causes of refugee *flows.* These important data are not necessarily appropriate for explaining differences in migration behavior at the micro-level, comparing settlement to settlement or neighbor to neighbor (3).

A second wave of studies address this lacuna and focus on individual behavior using structured interviews and household surveys. These studies are clearly better oriented to identify individual- and community-level migration determinants. At the same time, though, many are unable for practical reasons to observe and measure the determinants of choice *before or during* critical decision points. Many studies of migration, for example, only measure the behaviors and attitudes of refugees, choosing for understandable reasons of accessibility to explain outcomes like migration timing, rather than the choice to migrate or not. Some studies that surmount massive obstacles to collect data from "remainers" during or immediately after violent conflicts face another inferential challenge. Survey responses, behavioral game outcomes, or interview testimony that refugees and remainers give during or immediately after conflict might vary systematically as a result of experiences that would be downstream consequences of a decision to migrate or remain.

Table S1. Causes of Conflict-induced Migration in the Literature

Paper	Outcome	Key Cause	Method	Context	n
Weiner (1996) (4)	Refugee flows	Intra-state conflict ↑	Case comparison	Cross-national	N/A
Schmeidl (1997) (5)	Refugee flows	"Generalized violence" ↑	Pooled time series	Cross-national	N/A
Davenport et al. (2003)(1)	Refugee flows	"Personal integrity threats" ↑	Pooled time series	Cross-national	N/A
Moore and Shellman (2004) (2)	Refugee flows	Threat of violence ↑	Pooled time series	Cross-national	N/A
Engel and Ibañez (2007) (6)	Individual migration	Non-liquid assets ↓	1x survey	Colombia	363
Czaika and Kis-Katos (2009) (7)	Community-level migration	Violence events ↑	2x Census Data	Indonesia (Aceh)	5,200
Adhikari (2013) (8)	Individual migration	Social networks ↓	1x Survey	Nepal	1,424
Balcells and Steele (2016) (9)	Community-level migration	Ideological align. w/ combatants ↓	1x Electoral returns	Spain	654
Jampaklay et al. (2017) (10)	Individual migration	Violence events ↑	1x Survey	Thailand	1,009
Schon (2019) (11)	Individual migration *timing*	Wasta (money/connections) ↑	Interviews	Turkey (Syria)	178
Mironova et al. (2019) (12)	Individual migration	Risk preferences ↓	1x Behavioral game	Syria	232
Revkin (2019) (13)	Individual migration	Governance satisfaction ↓	1x Survey	Iraq	1,458
Alrababa'h et al. (2020) (14)	Individual return	Violence at home ↓	1x Survey	Lebanon (Syria)	3,003
Fearon and Shaver (2020) (15)	Refugee flows	Conflict deaths ↑	pooled time series	Cross-national	N/A
Camarena et al. (2020) (16)	Refugee flows	Violence ↑/Transit risk ↓	Event data	Italy (Libya)	307,000
Marston (2020) (17)	Individual migration	Ties to Local Leaders ↓	Survey	Colombia	618
Becker et al. (2021) (18)	Individual migration	Ties to earlier migrants ↑	Co-authorship Net.	Nazi Germany	1,129
Tai et al. (2022) (19)	Individual migration	Violence ↑	Call Detail Records	Afghanistan	10m
Milliff and Christia	Individual migration	Social Centrality ↓/↑ Violence ↑, Economic status ↑	CDR data in time-series	Yemen	56,000 units 2.6M obs. 63.5m calls

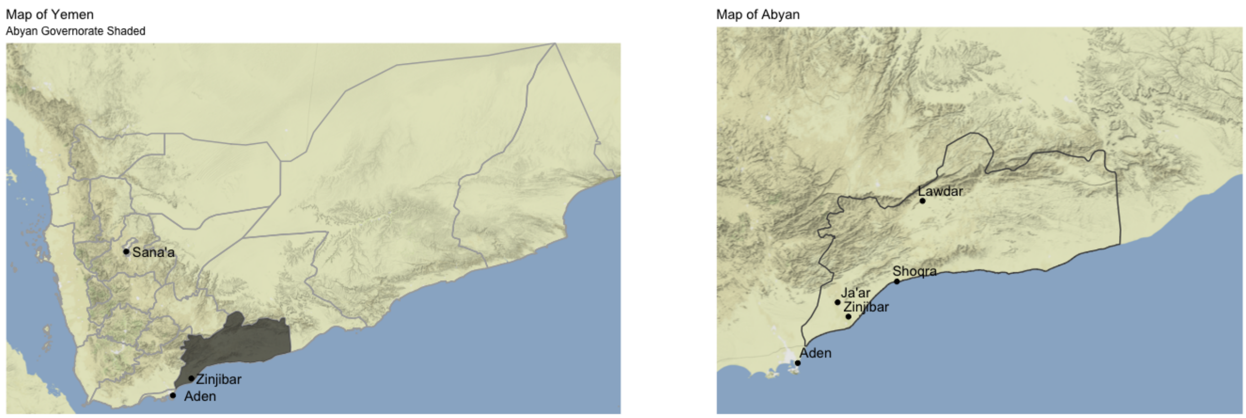


Fig. S1. Left: A map of Yemen, with Abyan Governorate shaded. In addition to the capital Sana'a, and the port city of Aden, the map identifies Zinjibar, the capital of Abyan governorate. **Right:** A map of Abyan governorate, identifying key population centers in and near the governorate including the capital Zinjibar and the major port city of Aden.

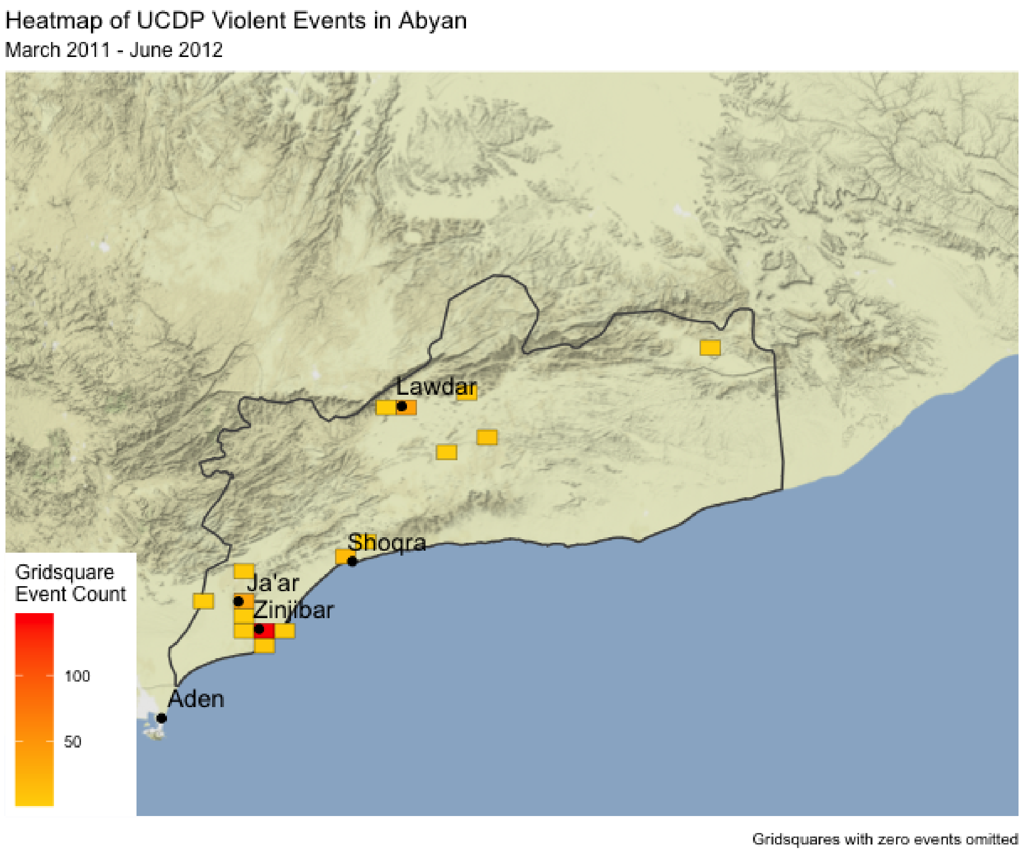


Fig. S2. Events in the UCDP Global Events Database for Abyan from March 2011 through June 2012. Violence is concentrated almost exclusively in settlements, with the greatest number of UCDP events occurring in and around Zinjibar, the capital of Abyan governorate. The map colors evenly-sized gridsquares according to the total count of separate events geolocated in the square. Squares in which no events occurred are not plotted.

Call Detail Record Literature

A substantial literature in computer science and economics use mobile communications data to study conflict relevant processes like the promulgation of ideas across networks (20–22), poverty and mobility (23–26), disaster response (27, 28), non-conflict migration (29), and the development/identification of social networks (30).

One output of this literature is a set of new, empirically validated techniques for agumenting CDR data such that it can be used as a proxy to measure other social, behavioral, and economic attributes of interest, even identity features like age and gender in some case (31). We draw on some of these results to augment CDR data in our models of migration. We caution though, that some models linking CDR attributes to wealth, age, and gender, for instance, seem to be culturally specific—Blumenstock (32) notes that models are "brittle" when applied to other social/cultural contexts. When using CDR attributes as proxies for wealth, we a) focus on attributes that the literature shows are consistently associated with wealth/income across a variety of contexts and b) often calculate multiple measures of a single underlying concept and look for them to be positively associated in our data (See Figure S3). Because age and gender prediction studies do not exist for contexts outside Western Europe, and because we expect the association between cell activity and age/gender to be highly culturally specific, we do not attempt to measure age and gender using CDRs in our data. We discuss the particulars of these measures in the Materials and Methods Section of the main paper, but provide an additional glossary in the next section.

Variable Descriptions/Glossary

Table S2. Variable descriptions and sources. Some variables are calculated differently across multiple time periods, see S3 for the time periods

Variable	Description	Citation/Source/ `python lib`
Call Duration	Grand mean of weekly average duration of calls	`bandicoot`
% Initiated Interactions	Weekly average percentage of calls initiated by user	`bandicoot`
# of Contacts	The number of contacts a user interacted with in a given week	`bandicoot`
Call Distance	Average distance between caller and recipient towers (km)	Author's calculation
Friday Prayer Calls	Count of calls made in the hour before Friday noon prayers, over entire pre-occupation window	Author's calculation & Bozcaga et al. (33)
Radius of Gyration	Measure of mobility: Equivalent distance of mass from center of gravity of weekly places visited	`bandicoot` & Gonzalez et al. (23)
PageRank	Scaled measure of "importance" in a network: node's PageRank is based on the PageRank of the nodes pointing toward it, i.e. it measures a node's importance through it's neighbors, and the independent importance of neighbors	`networkx` & Page et al. (34)
Clustering Coef	Measures density of a node's ties, or the degree to which a node's network approximates a "clique" in which each node shares an edge with every other node. Calculated as a ratio of existing edges to possible edges in a local graph.	`networkx` & Watts and Strogatz (35)
Degree Centrality	The fraction of nodes in a graph that a given node is connected to, normalized by the number of nodes in a graph	`networkx`
Latitude/Longitude	Coordinates of the *tower* to which user connects	Author's Data
Log GDP	Economic activity, measured at the 1km gridsquare level using DSMP nighttime light emissions and LandScan population data. We use the natural log of GDP associated with the km gridsquare that contains each tower.	Ghosh et al.(36) *via* Goodman et al. (37)
Pre-2010 GED Events	Sum of UCDP Georeferenced Events Dataset violence events that occur from 1989-2010 within 10 km of a given tower.	UCDP (38) *via* Goodman et al. (37)
GED Event Count	Count of UCDP GED events that occur within 10km of a tower in a given week in 2011-2012.	UCDP (38) *via* Goodman et al. (37)
# Records	Number of calls/texts a user conducts in a given week	Author's calculation
% Pareto Interactions	Proportion of user's contacts making up 80% of their interactions	`bandicoot`

Table S3. Time Period Descriptions for variables in all regression specifications.

Time Period	Description
“Preocc, avg”	Mean of weekly measures between 1 Jan 2011 and 25 March 2011
“Preocc”	Single statistic calculated from CDRs 1 Jan 2011 - 25 March 2011
“Tower, lag”	Characteristic calculated for user’s weekly modal tower (”home location”) in a given week from 26 March 2011 - June 2012. Lagged by one time period vs. migration status.
“Person, lag”	Characteristic calculated from a user’s weekly call behavior in a given week from 26 March 2011 - June 2012. Lagged by one time period vs. migration status.

Mobility Statistic Calculation. We calculate mobility metrics from CDRs using the `bandicoot` library (39). We calculate mobility metrics first for the months of pre-occupation baseline data (January-March 2011), and then for each week of data during the conflict (March 2011-June 2012). The key measure is radius of gyration (RoG, a measure of how wide an individual’s “typical” range of movement is over a period of time). We separately measure RoG to estimate an individual’s baseline, pre-conflict behavior, and again as a weekly statistic that mobility behavior during conflict. All mobility data is aggregated to the level of cell towers, which is more likely to capture coarse movements than quotidian movements within a home location. Subscribers can theoretically travel up to tens of miles (though usually less) without changing towers, and conversely, subscribers in an area covered by two towers might switch towers without moving. RoG, unlike other mobility measures, accounts for the second quirk.

Some CDR studies note an inverse relationship in the short term between mobility and cell activity, which makes it challenging to use CDRs for fine-grained location track information (23). Intuitively this relationship makes sense: people communicate more at rest than on the move, making their communication locations an incomplete record of the locations they may have visited between CDRs. Our data structure avoids some of this problem, though, by aggregating records to the week level, and calculating central tendencies and statistics at that level. We further address the possible confounding caused by the relationship between cellphone activity and mobility by including prior mobility as a covariate in our models.

Constructed Covariates + Violence Data. We rely on proxies developed in the CDR literature to measure economic status. Across a variety of contexts and different studies the literature identifies a few features as suitable proxies for relative wealth(24, 25, 31). We focus primarily on the proxies that are most closely connected to monetary resources and less on possible proxies that are mediated by social norms. 11 Wealth is positively associated with average call duration in countries (including Yemen) where pre-paid cell plans dominate. We measure pre-occupation average call duration to proxy wealth. Additionally, and less intuitively, wealth is negatively associated across varied societies with clustering coefficient: poorer people tend to have calling networks that are closer to being a clique (complete graph). We also use estimates of kilometer grid-square economic productivity (36, 37), which combines VIIRs nighttime light data and LandScan population density estimates to measure economic output. We link tower locations to the corresponding grid-square.

We use proxies for subscriber religiosity (more precisely, adherence to certain Islamic religious practices)as a final set of possible predictors. Religiosity is plausibly important in the Yemeni context because one of the major parties to the conflict, AQAP, espouses the extremely religiously conservative ideology of Salafism. Recent research on migration out of territory held by the Islamic State, a separate Salafist group, suggests that approval of ISIS governing practices influences choices to migrate (13). Regularity of religious practice is hardly a perfect proxy for approval of Salafismthe vast majority of observant Muslims around the world are not ideologically aligned with AQAP, but religiosity may factor into migration decisions when the occupying/governing power espouses an extreme religious ideology. We use a measure of religiosity developed by Bozcaga et al. (33), which focuses on calling behavior within specific temporal windows: Fridays in the hour before mid-day prayer, and during the days of religious holidays. We calculate the number and duration of calls made by each subscriber in the hour before Friday mid-day prayers. Scaling the count and duration metrics provides a proxy for religiosity vis a vis other callers—those who are calling more in these key times are, per Bozcaga et al. (33), relatively more religious.

Political violence is a key characteristic of Abyan during our period of study, and exposure to violence is a prominent explanation for migration behavior. We use the Uppsala Conflict Data Program Global Events Dataset (GED, 40) as a source of geo-located violence data. For the 400-plus GED events in Abyan in 2011-2012, we identify all cell towers within 10 miles of the GED event coordinates and label those towers as “affected” by the violent event in the week it occurred. We identify over 41,000 tower-event-week triplets.

Summary Statistics

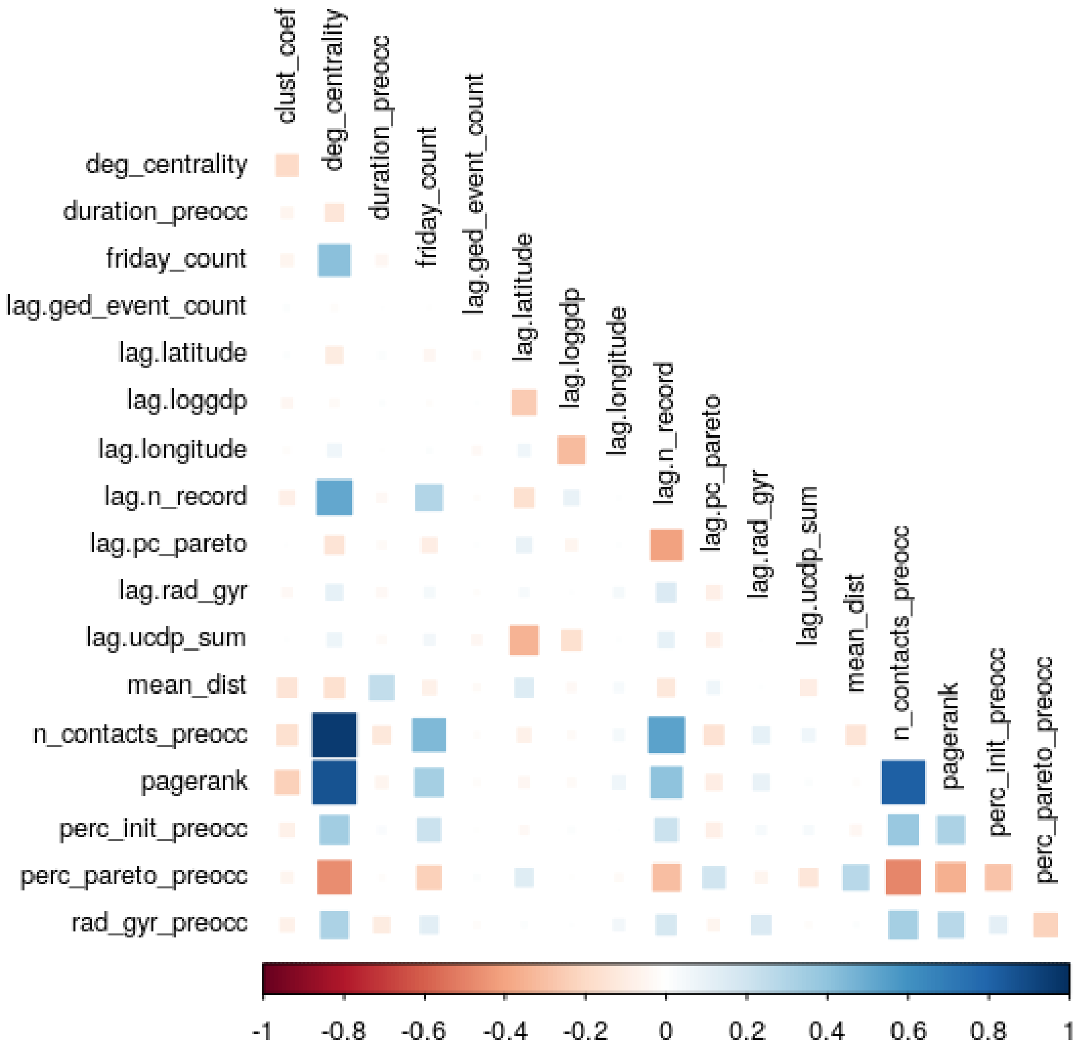


Fig. S3. Correlation of predictor variables (raw, un-transformed values). The highest bivariate correlation is 0.93.

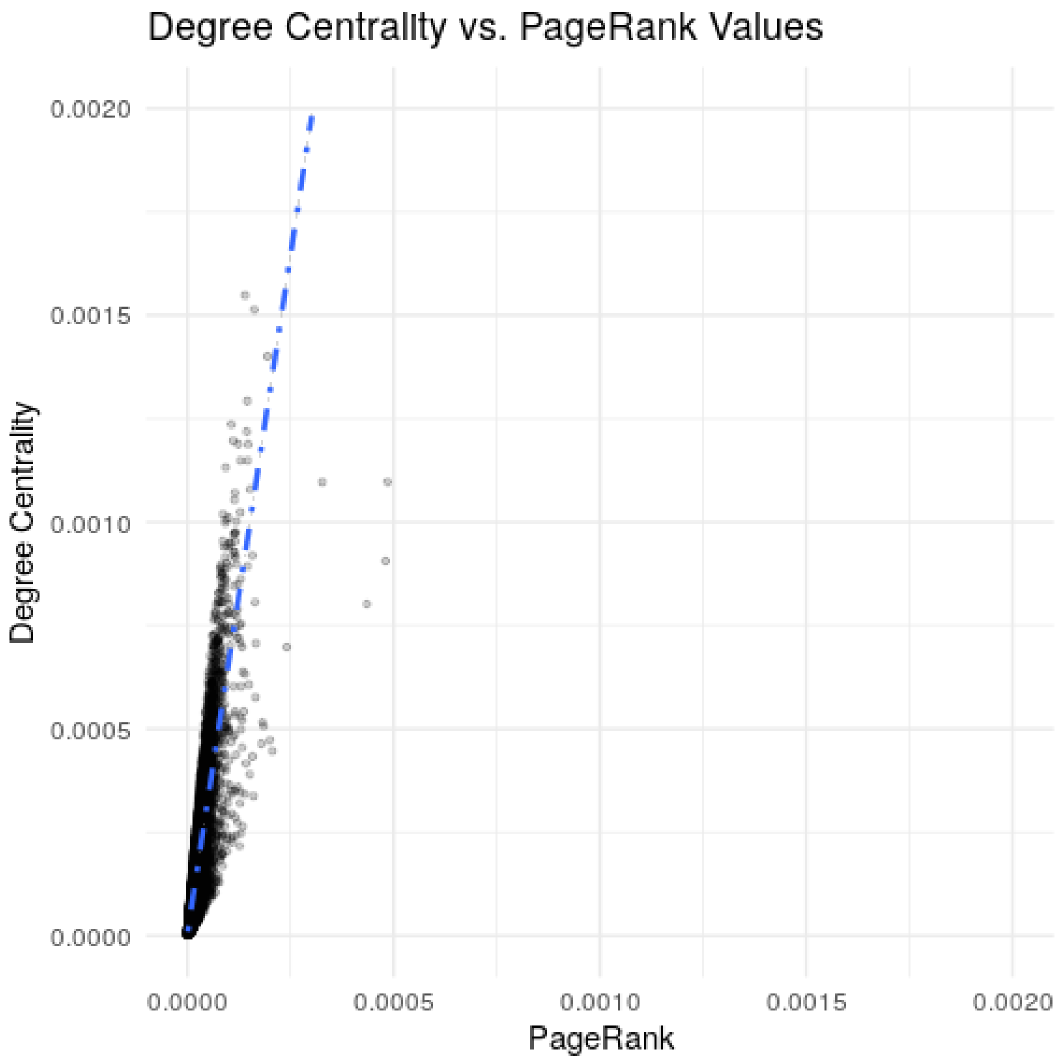


Fig. S4. Correlation of PageRank and Degree Centrality, two network measures which we juxtapose in the hypotheses and results. The measures are somewhat highly correlated $r = .83$, which we expect because they measure conceptually related quantities. The building blocks of Degree Centrality, in fact, are one component of PageRank. PageRank generally measures influence by the quantity of a node's out-directed edges (a main component of degree centrality), plus the quantity of the out-directed edges of all the original node's connections (a component of those first-degree connections' degree centrality). We expect that the two metrics are therefore correlated but not collinear. When we estimate models using both PageRank and Degree Centrality, we use the phrase "Influence" to describe the PageRank coefficient, because a node's influence (i.e. the embeddedness of that node's *neighbors*) is what the marginal association with PageRank measures after holding degree centrality constant.

Table S4. Summary statistics (and missingness) for predictor variables. Variables are de-meaned and scaled in model estimations. FD results presented in Figures 4 and 5 in the main text can be interpreted as change in probability of appearing in Abyan associated with moving from the value in the 25% column to the value in the 75% column.

Variable	Complete	Mean	SD	Min.	25%	Med.	75%	Max
Call Duration (preocc, avg)	0.959	166.095	178.714	2.000	58.769	99.388	200.806	1883.000
% Initiated (preocc, avg)	0.960	0.079	0.151	0.000	0.000	0.000	0.111	1.000
Call Distance km. (preocc, avg)	0.957	57.600	69.200	0.000	15.600	34.7000	73.100	908.000
Friday Prayer Calls (preocc, avg)	1.000	0.606	2.473	0.000	0.000	0.000	0.000	74.000
Radius of Gyr. (preocc, avg)	0.959	4.254	9.084	0.000	0.000	1.439	4.854	459.938
Pagerank (preocc)	0.960	1.32e-5	1.55e-5	9.33e-7	4.50e-6	8.90e-6	1.68e-5	0.002
Clust. Coef. (preocc)	0.960	0.093	0.141	0.000	0.013	0.051	0.111	1.000
Deg. Centrality (preocc)	0.960	9.48e-5	1.11e-4	4.34e-6	2.60e-5	6.07e-5	1.21e-4	0.004
Latitude (tower, lag)	0.889	13.554	0.663	12.685	13.134	13.359	13.883	17.796
Longitude (tower, lag)	0.889	45.279	0.851	42.675	44.995	45.298	45.382	53.095
Log GDP (tower, lag)	0.888	-0.822	2.056	-7.428	-1.878	-1.833	0.286	4.538
Pre-2010 GED Events (tower, lag)	0.888	73.421	105.303	0.000	0.000	16.000	214.000	408.000
GED Event Count (tower, lag)	0.979	0.109	0.537	0.000	0.000	0.000	0.000	11.000
# Records (person, lag)	0.961	24.835	43.605	0.000	1.000	9.000	29.000	1300.000
Radius of Gyr. (person, lag)	0.736	6.442	17.051	0.000	0.000	1.178	5.068	1179.088
% Pareto Int. (person, lag)	0.752	0.220	0.154	0.000	0.103	0.188	0.333	0.800

Table S5. Dependent variable summary for time series estimations.

Variable	Complete	Mean	Logical Vals.
In Abyan?	1.000	0.546	T: 1,454,388 F: 1,210,072

Table S6. Date summaries for time series estimation.

Variable	Complete	Min.	Max.	Med.	Unique Val.
Date	1.000	2011-01-30	2012-06-24	2011-10-16	74

Additional Model Figures/Tables

Model Specifications.

Linear Random Effects model (LMER) specification for results in Figure 4 (Main text):.

$$Y_i = \alpha + \xi K + \delta J + \beta \mathbf{X}_i$$

Where:

- α = Global intercept
- Y_i = Proportion of time outside Abyan
- K = Tower dummies
- J = District dummies
- X includes: Number of contacts, call duration, % initiated interactions, % pareto interactions, mean call distance, Friday prayer calls, radius of gyration, Pagerank, clustering coefficient, degree centrality, Latitude, Longitude, Lat/Lon polynomial, pre-conflict violence events, tower GDP.

Logit Random Effects model specification for results in Figure 5 (Main Text):.

$$Y_{it} = \alpha_i + \gamma_t + \beta \mathbf{X}_{it}$$

Where:

- Y_{it} = Binary indicator of weekly presence in Abyan
- i = Subscriber
- t = Time period
- X includes: Number of contacts, call duration, % initiated interactions, % pareto interactions, mean call distance, Friday prayer calls, radius of gyration, Pagerank, clustering coefficient, degree centrality, Latitude, Longitude, Lat/Lon polynomial, pre-conflict violence events, tower GDP.

Additional Results.

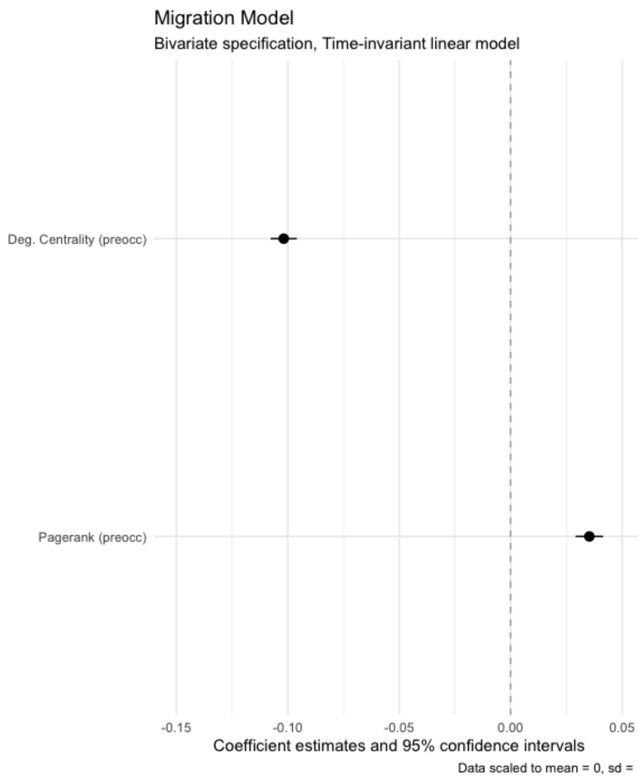


Fig. S5. Coefficient estimates (95% CIs) from a time-invariant linear model including only the key social network predictors. This corresponds to a less-saturated version of the model presented in Figure 4 (and Table S11), and shows results consistent with the fully-saturated version in the main body of the paper.

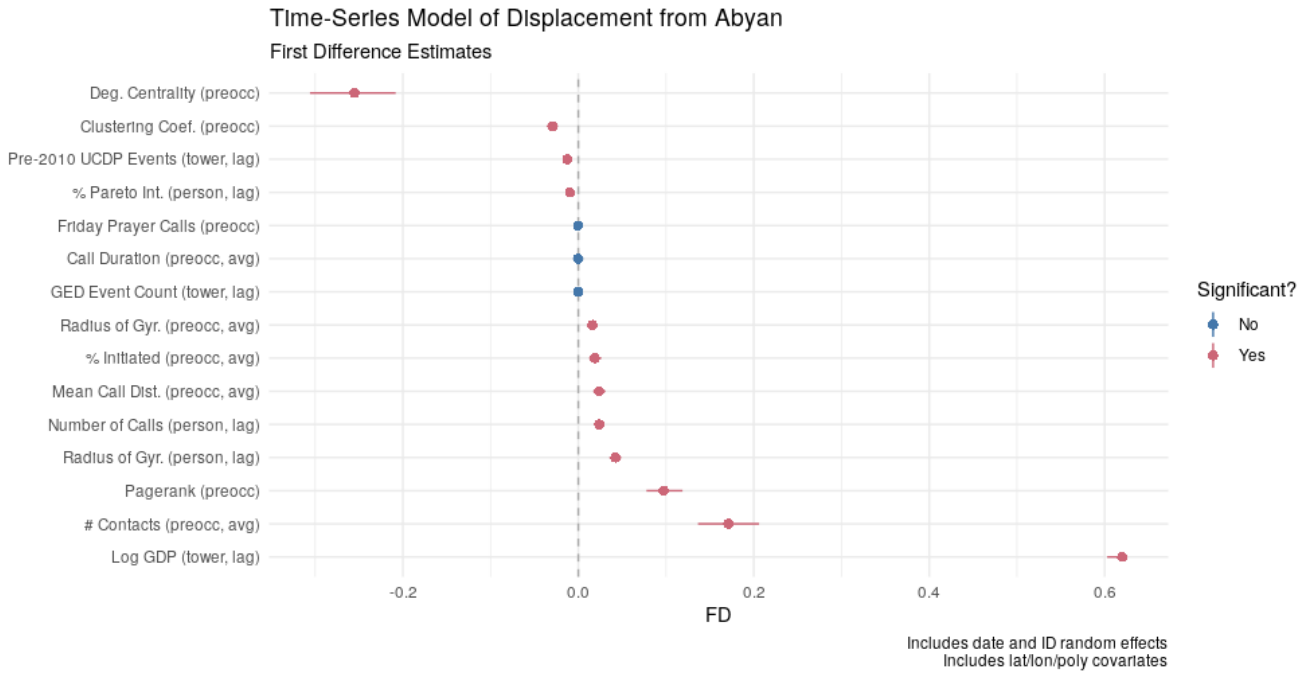


Fig. S6. Results of the HMC Estimation, presented as first-differences in probability of "fleeing" (i.e. being outside Abyan in a given week). Points show the estimated difference in probability of being outside Abyan associated with the difference between a 10th percentile value of the predictor versus a 90th percentile value. Error bars show 99% credible intervals.

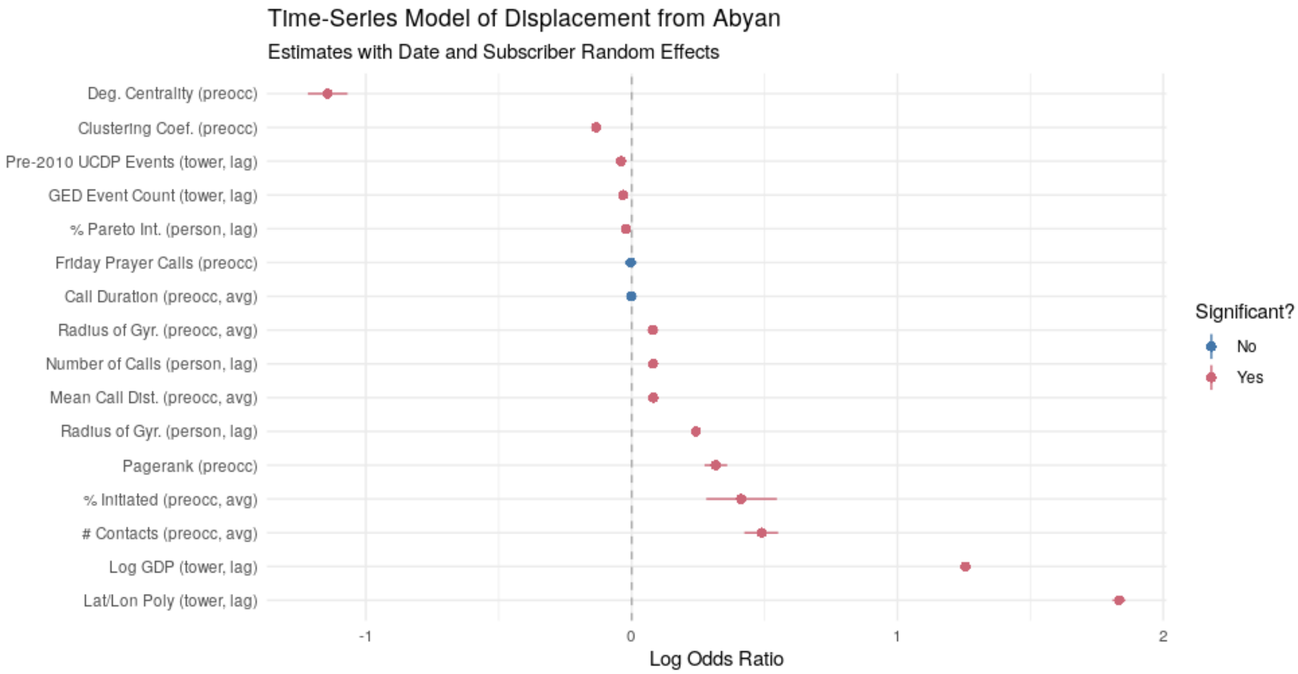


Fig. S7. Raw log odds ratio parameter estimates from the HMC Estimation. Log odds ratios here are associated with a one standard deviation change in the value of the predictor, as all numerical predictors (save latitude and longitude) are demeaned and scaled to facilitate better exploration of the posterior density. Error bars show 99% credible intervals, calculated from stan's sampling of the posterior density function.

Table S7. Parameter estimates and model diagnostics from Hamiltonian Monte-Carlo (HMC) estimation of a multilevel logistic regression with time- and unit- varying intercepts. Key diagnostics for exploration of the posterior—n_eff and Rhat—do not indicate issues with the model's exploration of the density function. 41 and 42 suggest not to interpret models with Rhat values > 1.05 or n_eff values below $5m$, where m is the number of chains after splitting.

Parameter	mean	mcse	sd	10%	50%	90%	n_eff	Rhat
(Intercept)	-1213.265	0.086	8.092	-1223.777	-1213.175	-1202.879	8764	1.000
Number of contacts (Preocc, avg)	-0.489	0.001	0.032	-0.530	-0.489	-0.447	1179	1.007
Call duration (Preocc, avg)	0.001	0.000	0.010	-0.012	0.001	0.013	1350	1.004
% Initiated Conv. (Preocc, avg)	-0.412	0.002	0.067	-0.499	-0.412	-0.328	1405	1.005
Average call distance (Preocc)	-0.081	0.000	0.010	-0.094	-0.081	-0.069	1736	1.002
Friday prayer calls (Preocc)	0.003	0.000	0.011	-0.011	0.003	0.018	1050	1.006
Radius of gyration (Preocc, avg)	-0.080	0.000	0.010	-0.093	-0.080	-0.066	1513	1.002
Pagerank (Preocc)	-0.316	0.001	0.022	-0.344	-0.316	-0.288	1341	1.006
Clust. Coef (Preocc)	0.134	0.000	0.009	0.122	0.134	0.146	1812	1.002
Degree Centrality (Preocc)	1.144	0.001	0.038	1.094	1.144	1.193	967	1.009
Latitude (tower,lag)	80.764	0.006	0.555	80.053	80.757	81.483	8855	1.000
Longitude (tower,lag)	27.476	0.002	0.181	27.244	27.473	27.711	8817	1.000
Lat*Lon Polynomial (tower,lag)	-1.834	0.000	0.012	-1.850	-1.833	-1.818	8856	1.000
Log GDP (tower,lag)	-1.256	0.000	0.005	-1.263	-1.256	-1.249	9813	1.000
Pre-2010 GED Events (tower, lag)	0.040	0.000	0.005	0.034	0.040	0.046	8977	1.000
GED Event count (tower, lag)	0.032	0.000	0.003	0.029	0.032	0.035	9854	0.999
Number of calls (lag)	-0.081	0.000	0.004	-0.086	-0.081	-0.076	9086	1.000
Radius of gyration (lag)	-0.242	0.000	0.003	-0.246	-0.242	-0.238	9433	1.000
Percent Pareto Interactions (lag)	0.021	0.000	0.003	0.017	0.021	0.025	10041	1.000

Table S8. First Difference estimates corresponding to Figure 6. Signs are reversed in the plot in order to show first differences in probability of leaving, not remaining (the dependent variable, shown in Table S5 takes a value of 1 for remaining.

Parameter	Median First Difference	1st Percentile Est.	99th Percentile Est.
Number of contacts (Preocc, avg)	-0.0691483	-0.0843806	-0.0547887
Call duration (Preocc, avg)	0.0000848	-0.0023765	0.0025607
% Initiated Conv. (Preocc, avg)	-0.0073333	-0.0102367	-0.0048960
Average call distance (Preocc)	-0.0101628	-0.0133024	-0.0073659
Friday prayer calls (Preocc)	0.0000000	0.0000000	0.0000000
Radius of gyration (Preocc, avg)	-0.0071395	-0.0094888	-0.0051076
Pagerank (Preocc)	-0.0436540	-0.0534913	-0.0346964
Clust. Coef (Preocc)	0.0135088	0.0107224	0.0166528
Degree Centrality (Preocc)	0.1529770	0.1253409	0.1821949
Latitude (tower,lag)	0.9999997	0.9999996	0.9999998
Longitude (tower,lag)	0.9753132	0.9681325	0.9807699
Lat*Lon Polynomial (tower,lag)	-1.0000000	-1.0000000	-1.0000000
Log GDP (tower,lag)	-0.4599450	-0.4733246	-0.4338020
Pre-2010 GED Events (tower, lag)	0.0126933	0.0093694	0.0164930
GED Event count (tower, lag)	0.0000000	0.0000000	0.0000000
Number of calls (lag)	-0.0101730	-0.0122323	-0.0083014
Radius of gyration (lag)	-0.0119649	-0.0139862	-0.0099782
Percent Pareto Interactions (lag)	0.0043684	0.0028920	0.0060301

Table S9. First Difference estimates corresponding to Figure S6. Signs are reversed in the plot in order to show first differences in probability of leaving, not remaining (the dependent variable, shown in Table S5 takes a value of 1 for remaining.

Parameter	Median First Difference	1st Percentile Est.	99th Percentile Est.
Number of contacts (Preocc, avg)	-0.1709156	-0.2060192	-0.1364214
Call duration (Preocc, avg)	0.0002027	-0.0057087	0.0060824
% Initiated Conv. (Preocc, avg)	-0.0187091	-0.0262197	-0.0124153
Average call distance (Preocc)	-0.0235968	-0.0309274	-0.0170763
Friday prayer calls (Preocc)	0.0004005	-0.0023460	0.0032693
Radius of gyration (Preocc, avg)	-0.0160472	-0.0213590	-0.0114574
Pagerank (Preocc)	-0.0973340	-0.1187690	-0.0776589
Clust. Coef (Preocc)	0.0292949	0.0232559	0.0361514
Degree Centrality (Preocc)	0.2551141	0.2082486	0.3058801
Latitude (tower,lag)	1.0000000	1.0000000	1.0000000
Longitude (tower,lag)	1.0000000	1.0000000	1.0000000
Lat*Lon Polynomial (tower,lag)	-1.0000000	-1.0000000	-1.0000000
Log GDP (tower,lag)	-0.6198577	-0.6258372	-0.6029100
Pre-2010 GED Events (tower, lag)	0.0126933	0.0093694	0.0164930
GED Event count (tower, lag)	0.0000000	0.0000000	0.0000000
Number of calls (lag)	-0.0238111	-0.0285587	-0.0194559
Radius of gyration (lag)	-0.0422357	-0.0489204	-0.0355227
Percent Pareto Interactions (lag)	0.0096161	0.0063764	0.0132629

Table S10. Raw odds ratios (with 95% credible intervals) corresponding to figure S7. Signs are reversed in the plot in order to show first differences in probability of leaving, not remaining (the dependent variable, shown in Table S5 takes a value of 1 for remaining.

Estimate	2.5%	97.5%	Significance?	Parameter
-1213.1747960	-1229.0921842	-1197.4177928	Yes	Intercept
-0.4889985	-0.5511139	-0.4241460	Yes	# Contacts (preocc, avg)
0.0006676	-0.0188245	0.0201217	No	Call Duration (preocc, avg)
-0.4116512	-0.5463700	-0.2807507	Yes	% Initiated (preocc, avg)
-0.0813945	-0.1013038	-0.0615684	Yes	Mean Call Dist. (preocc, avg)
0.0031660	-0.0181898	0.0247066	No	Friday Prayer Calls (preocc)
-0.0799217	-0.1000910	-0.0592208	Yes	Radius of Gyr. (preocc, avg)
-0.3156982	-0.3594177	-0.2741760	Yes	Pagerank (preocc)
0.1337839	0.1156159	0.1521835	Yes	Clustering Coef. (preocc)
1.1440058	1.0690886	1.2182049	Yes	Deg. Centrality (preocc)
80.7570660	79.6773096	81.8541257	Yes	Latitude (tower, lag)
27.4731782	27.1199930	27.8297528	Yes	Longitude (tower, lag)
-1.8334745	-1.8580053	-1.8092751	Yes	Lat/Lon Poly (tower, lag)
-1.2562643	-1.2669083	-1.2456920	Yes	Log GDP (tower, lag)
0.0401698	0.0313599	0.0490349	Yes	Pre-2010 UCDP Events (tower, lag)
0.0320105	0.0270127	0.0372811	Yes	GED Event Count (tower, lag)
-0.0810619	-0.0889503	-0.0731589	Yes	Number of Calls (person, lag)
-0.2418761	-0.2475830	-0.2362507	Yes	Radius of Gyr. (person, lag)
0.0210811	0.0144517	0.0276498	Yes	% Pareto Int. (person, lag)

Table S11. Raw coefficient estimates corresponding to the non time-series model in Figure 4. Signs are reversed in the plot in order to show marginal associations with time outside of Abyan, not Inside (the dependent variable, shown in Table S5 takes a value of 1 for remaining.

Estimate	Std. Err.	Z-Value	P-Value	2.5 %	97.5 %	Parameter
4.5149942	29.3463347	0.1538521	0.8777264	-53.0027648	62.0327533	Intercept
-0.0671430	0.0063389	-10.5922370	0.0000000	-0.0795670	-0.0547190	# Contacts (preocc, avg)
-0.0036961	0.0025880	-1.4281727	0.1532422	-0.0087684	0.0013763	Call Duration (preocc, avg)
0.0094660	0.0150928	0.6271850	0.5305380	-0.0201154	0.0390474	% Initiated (preocc, avg)
-0.2324781	0.0142842	-16.2751842	0.0000000	-0.2604747	-0.2044816	% Pareto Int. (preocc, avg)
-0.0028714	0.0031991	-0.8975649	0.3694176	-0.0091415	0.0033987	Mean Call Dist (preocc)
0.0029521	0.0019618	1.5048154	0.1323715	-0.0008929	0.0067971	Friday Prayer Calls (preocc)
-0.0051375	0.0024819	-2.0699734	0.0384548	-0.0100019	-0.0002730	Radius of Gyr. (preocc, avg)
-0.0098330	0.0047560	-2.0675134	0.0386858	-0.0191545	-0.0005115	Pagerank (preocc)
0.0092493	0.0032083	2.8829136	0.0039402	0.0029611	0.0155374	Clustering Coef. (preocc)
0.1086211	0.0072390	15.0050085	0.0000000	0.0944329	0.1228092	Deg. Centrality (preocc)
-0.3725637	2.0936730	-0.1779474	0.8587643	-4.4760874	3.7309600	Latitude
-0.1176713	0.6578136	-0.1788824	0.8580300	-1.4069622	1.1716196	Longitude
0.0105029	0.0469452	0.2237263	0.8229703	-0.0815081	0.1025139	Lat/Lon Poly (tower)
0.0560421	0.0030637	18.2924060	0.0000000	0.0500374	0.0620468	Pre-2010 UCDP Events (tower)
-0.0163719	0.0205024	-0.7985353	0.4245599	-0.0565559	0.0238121	Log GDP (tower)

Table S12. LMER results comparing models with and without centrality statistics. Note the 1.781493x increase in Marginal R^2 (Model 1 over Model 2) for the inclusion of the network stats. Including versus omitting centrality measures also meaningfully changes some other associations like the number of contacts, and call duration.

	Dependent variable:	
	Weeks Inside Abyan (proportion)	
	(1)	(2)
# Contacts (preocc, avg)	-0.145^{***}	0.039^{***}
	(0.003)	(0.001)
Call Duration (preocc, avg)	-0.0001	-0.004^{***}
	(0.001)	(0.001)
% Initiated (preocc, avg)	-0.030^{***}	-0.071^{***}
	(0.007)	(0.008)
% Pareto Int. (preocc, avg)	-0.201^{***}	-0.206^{***}
	(0.006)	(0.006)
Mean Call Dist (preocc)	-0.003^{**}	-0.011^{***}
	(0.001)	(0.001)
Friday Prayer Calls (preocc)	0.004^{***}	0.006^{***}
	(0.001)	(0.001)
Radius of Gyr. (preocc, avg)	-0.008^{***}	-0.013^{***}
	(0.001)	(0.001)
Latitude	0.283	-0.095
	(0.409)	(0.428)
Longitude	0.071	-0.039
	(0.135)	(0.141)
Lat/Lon Poly (tower)	-0.005	0.003
	(0.009)	(0.010)
Pre-conflict UCDP Events (tower)	0.029	0.038
	(0.024)	(0.027)
Log GDP (tower)	-0.021	-0.018
	(0.013)	(0.014)
Pagerank (preocc)	-0.040^{***}	
	(0.002)	
Clustering Coef. (preocc)	0.007^{***}	
	(0.001)	
Deg. Centrality (preocc)	0.230^{***}	
	(0.004)	
Constant	-3.261	1.897
	(6.080)	(6.355)
Observations	51,835	51,835
Log Likelihood	$-1,399.747$	$-3,351.148$
Akaike Inf. Crit.	2,837.493	6,734.297
Bayesian Inf. Crit.	3,005.754	6,875.990
Marginal R^2	0.130224	0.07309827
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table S13. These linear model results in columns 1-3 show the effects of leaving out individual centrality measures. As expected, degree centrality and pagerank are somewhat highly correlated ($cor = 0.83$) in the data because they measure related concepts. Comparing the models, we can interpret the models as showing that the part of Pagerank that is related to degree centrality (first degree network breadth) is a reasonably large part of Pagerank's total effect, and is associated with *remaining*. The elements of Pagerank that are not captured in degree centrality—the strength of the networks of a node's contacts—are substantively meaningful even after controlling for degree centrality, and predict leaving. This motivates our decision in the main text to interpret page rank as "influence" and degree centrality as "breadth." Column 4 shows the same results as the fully saturated model in Table S12 .

	Dependent variable:			
	Weeks Inside Abyan (proportion)			
	(1)	(2)	(3)	(4)
# Contacts (preocc, avg)	-0.136^{***}	0.013^{***}	-0.145^{***}	-0.145^{***}
	(0.003)	(0.002)	(0.003)	(0.003)
Call Duration (preocc, avg)	-0.002	-0.004^{***}	-0.001	-0.0001
	(0.001)	(0.001)	(0.001)	(0.001)
% Initiated (preocc, avg)	-0.037^{***}	-0.070^{***}	-0.031^{***}	-0.030^{***}
	(0.007)	(0.008)	(0.007)	(0.007)
% Pareto Int. (preocc, avg)	-0.209^{***}	-0.209^{***}	-0.206^{***}	-0.201^{***}
	(0.006)	(0.006)	(0.006)	(0.006)
Mean Call Dist (preocc)	-0.008^{***}	-0.013^{***}	-0.004^{***}	-0.003^{**}
	(0.001)	(0.001)	(0.001)	(0.001)
Friday Prayer Calls (preocc)	0.005^{***}	0.006^{***}	0.004^{***}	0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Radius of Gyr. (preocc, avg)	-0.008^{***}	-0.012^{***}	-0.009^{***}	-0.008^{***}
	(0.001)	(0.001)	(0.001)	(0.001)
Latitude	0.252	-0.055	0.286	0.283
	(0.412)	(0.426)	(0.410)	(0.409)
Longitude	0.063	-0.028	0.073	0.071
	(0.136)	(0.141)	(0.135)	(0.135)
Lat/Lon Poly (tower)	-0.005	0.002	-0.005	-0.005
	(0.009)	(0.009)	(0.009)	(0.009)
Pre-treatment UCDP Events (tower)	0.030	0.038	0.029	0.029
	(0.025)	(0.027)	(0.025)	(0.024)
Log GDP (tower)	-0.021	-0.018	-0.021	-0.021
	(0.013)	(0.014)	(0.013)	(0.013)
Deg. Centrality (preocc)	0.185^{***}		0.229^{***}	0.230^{***}
	(0.003)		(0.004)	(0.004)
Clustering Coef. (preocc)				0.007^{***}
				(0.001)
Pagerank (preocc)		0.033^{***}	-0.040^{***}	-0.040^{***}
		(0.002)	(0.002)	(0.002)
Constant	-2.847	1.375	-3.320	-3.261
	(6.114)	(6.333)	(6.085)	(6.080)
Observations	51,835	51,835	51,835	51,835
Log Likelihood	$-1,598.931$	$-3,176.096$	$-1,414.018$	$-1,399.747$
Akaike Inf. Crit.	3,231.863	6,386.193	2,864.037	2,837.493
Bayesian Inf. Crit.	3,382.412	6,536.742	3,023.442	3,005.754
Note:				*p<0.1; **p<0.05; ***p<0.01

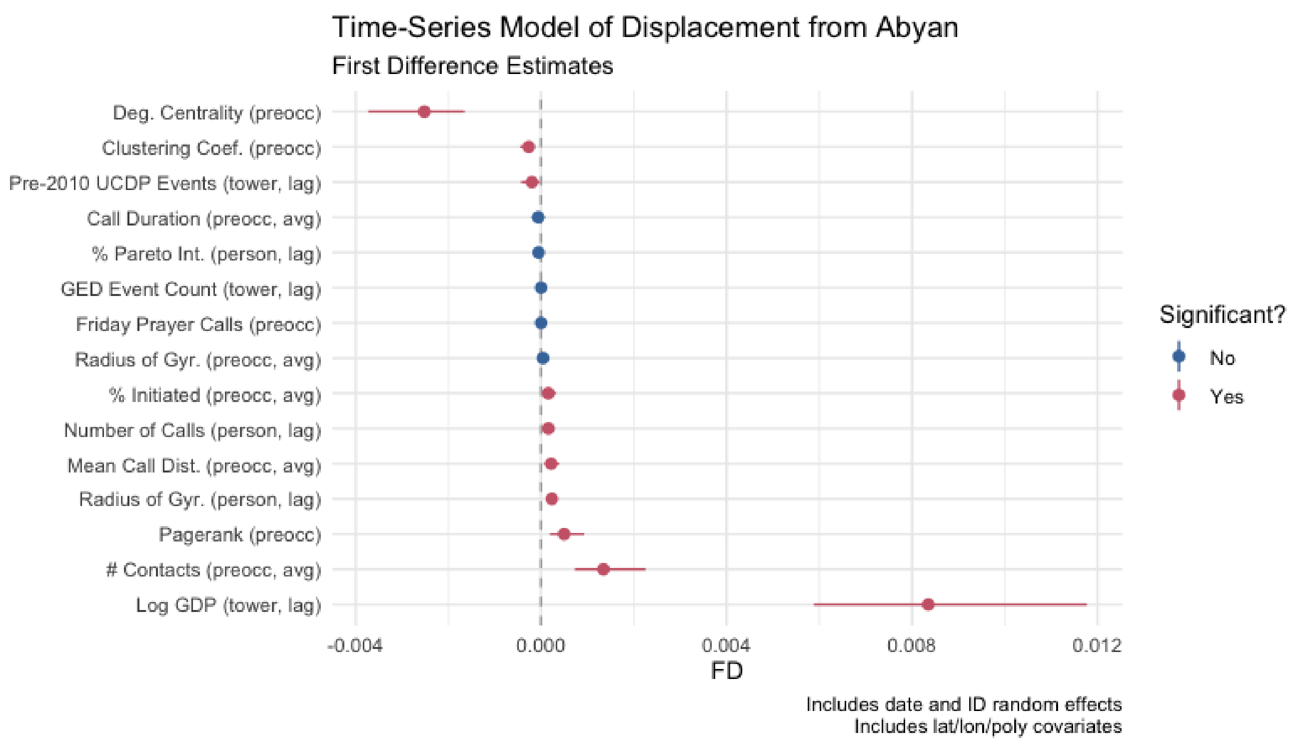


Fig. S8. First difference results (similar to Figure 5), but for a separate `stan` model estimation using data for 5,000 randomly-selected subscribers (roughly 10% of the data). These results are consistent with results from the full data, but are somewhat less precisely estimated with fewer observations.

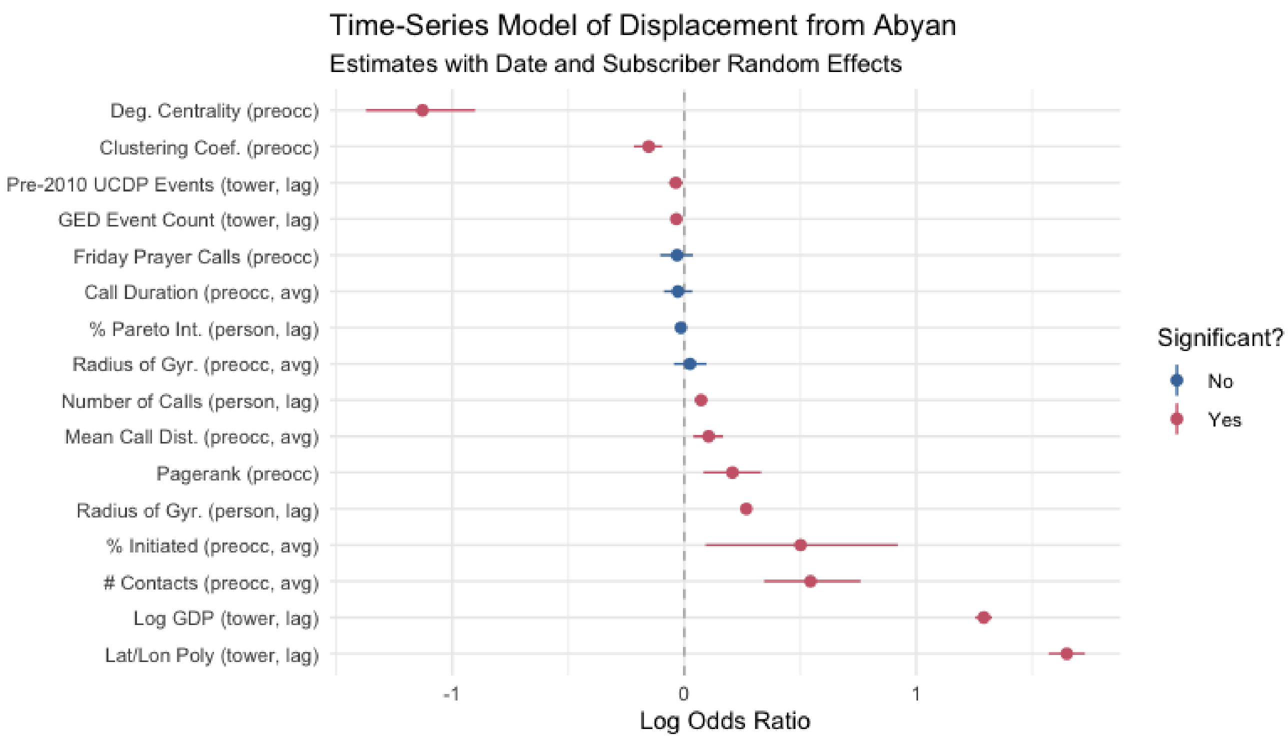


Fig. S9. Raw log odds ratios from the estimation with 5,000 randomly selected subscribers.

A Note on GDP. In Section 5, we note that the one-week lag of the GDP covariate might be mechanically inflated in the time series model if people leaving Abyan tend to a) flee to prosperous areas and b) remain in prosperous areas if they stay. This may anecdotally be true: People leaving Abyan tend to go toward Aden, which has much higher economic activity as a major port city and we know that at least some internally displaced people stayed in Aden through the end of the conflict. The lagged GDP value associated with their presence in Aden, then, is sometimes the GDP value associated with their presence in Aden—which we know to be higher than most of Abyan. Compare what we find in Figures 4 and 5, where the GDP value is for the subscriber's starting tower *prior* to the conflict. The GDP value associated with the initial tower—a qualitatively different measure, to be clear—is not significantly associated with migration after accounting for other factors, though the direction of the insignificant association is the same.

Alterative Models

We may wish to verify that the main results in Section 5 are not a product of the chosen functional forms of the models. To do this, we re-estimate many specifications (including the key results in Figure 4) with different functional forms. For tractability reasons, these "checks" are performed on a subset of the data representing 2,000 randomly selected individuals in the population of interest (about 3.5% of the total sample). Results below show that these new estimations are highly consistent with results in the body of the paper.

OLS Estimation. To provide additional assurance that the results in Figure 4 are not driven by the choice of a fixed effects logit model, we re-estimate the time-varying model (Figure 4) as an ordinary least squares model. Column 5 in table **??** shows the equivalent OLS model for the main specification, with the addition of respondent-clustered standard errors. Results are largely consistent with Logit/STAN results presented in the main text. Columns 1-4 show various reduced-form regression specifications focusing on single classes of explanations. Per Aronow and Samii (43), highly saturated multivariate regressions can produce marginal effects comparisons in populations that are dramatically un-representative of the population of interest, so it is useful to check sparser specifications. Three such specifications of interest are represented in Figures S10 and S11, which show simple bivariate associations (in the random sub-sample) for key implications and take-aways highlighted in the discussion. In Figure S10, we see that stripped-down models of the social network predictors of migration are consistent with the effects described in the main paper. Figure S11 shows the same for the effect of violence in AQ areas and pre-migration mobility.

Migration as a Markov Process. In the paper's main specifications, we define "migration" as a change in home location from a point *inside* Abyan to a point *outside* Abyan. While we believe this definition is useful (especially compared to existing literature, which largely defines migration by registration with UNHCR or similar organizations), it is also coarse. Defining "migration" as a shift in home location outside the borders of Abyan could both a) miss individuals who are displaced from their homes but move more locally and b) pick up individuals who are traveling, not migrating.

To relax the definition of "migration," we model migration as a two-state hidden markov process. Here, we assume that the time-series location data we gather from individual cell subscribers can be classified into different latent "modes" or "states" of behavior, which generate observed characteristics like the distance traveled between location measurements, or the "turning" angle described by any three location measurements. These hidden states, in turn, are modeled as a hidden markov chain following a common technique in *animal* migration studies (44).

We use an Rstan implementation called `moveHMM` (45) to model migration as a hidden markov process, because it allows for the inclusion of coavariates such that we can model the determinants of state-shifting, i.e. the determinants of going from stationary to migratory. Fitting the model in Stan returns two "states" defined by different distributions of travel distance and turning angle, which the investigator then labels as "migratory" or "stationary." Figures S12 and S13 show density plots of the travel distances and turn angles (respectively) for the two states. Intuitively, the state with more long travel distance segments and more turn angles near zero is labeled "migratory."

Having established the two states, the next step is to examine the determinants of moving between them—these would be the determinants of initiating or continuing migration. Figures S14, S15, and S16 show simulated posterior probabilities of transitioning between states associated with various values of Pagerank and Radius of Gyration, two of the main coefficients of interest for our conclusions about the importance of social networks and pre-migration mobility. Figures S14, S15, and S16 are marginal associations, accounting for the same range of covariates as included in the main Logit models (see Figure 4 and S6).

Results from the state transition curves generally support the interpretation of the main logit specifications but add some interesting nuance. First, the positive associations between displacement and both Pagerank and pre-migration mobility seem to obtain because individuals with higher Pagerank and more mobility are more *persistent* in migratory behavior (more likely to remain in State 2), not more likely to initiate migration. The results for Degree Centrality, however, are essentially as expected based on the logit results.

Stan diagnostics for the `moveHMM` model are available below in Figures S24, S26, and S25.

	(1) Violence	(2) Social Status	(3) Mobility	(4) Forecast	(5) Combination
Origin: Hist. UCDP	-0.040***				-0.032***
	(0.006)				(0.008)
Origin: AQ Occupation	-0.296***				-0.275***
	(0.011)				(0.017)
Origin: Prev. week violence	-0.007***				-0.034***
	(0.001)				(0.003)
Violence in AQ Occ. Area	0.014**				0.040***
	(0.005)				(0.009)
Pagerank		0.079***			0.107***
		(0.021)			(0.021)
Deg. Centrality		-0.092***			-0.115***
		(0.018)			(0.028)
Clustering Coef.		-0.022***			-0.025**
		(0.007)			(0.008)
Prev. week mobility			0.027***	0.023***	0.026***
			(0.003)	(0.003)	(0.003)
Prev. week calls				0.017***	0.029***
				(0.004)	(0.006)
Prev. week PercPare				-0.001	-0.014***
				(0.002)	(0.004)
R2 Adj.	0.596	0.300	0.631	0.637	0.410
RMSE	0.84	0.69	0.86	0.86	0.75
Std. Errors by:	ID	ID	ID	ID	ID
FEs	Date, ID	ID	Date, ID	Date, ID	ID
Geo. Poly	✓	✓	✓	✓	✓

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S14. OLS results for main and reduced-form models.

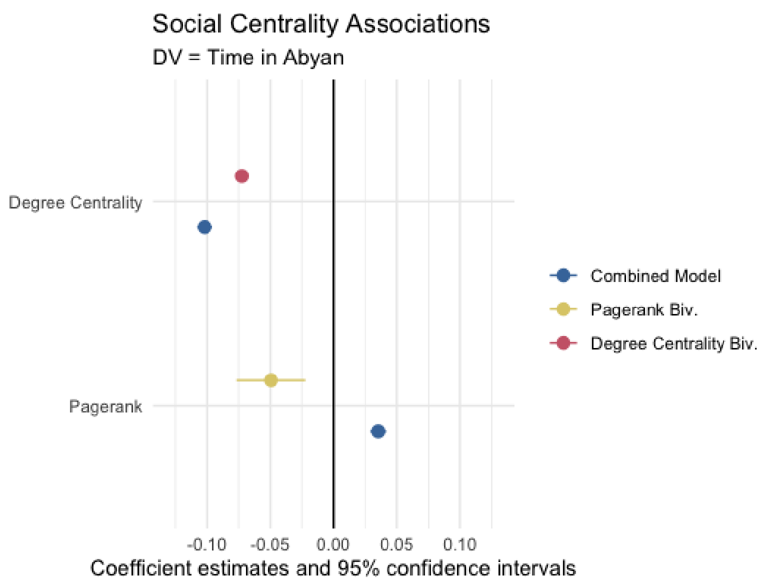


Fig. S10. Associations between migration and social centrality, with two way fixed effects and respondent clustered errors, but no other controls. The more basic specification represented here accords with the results as described in the paper's discussion.

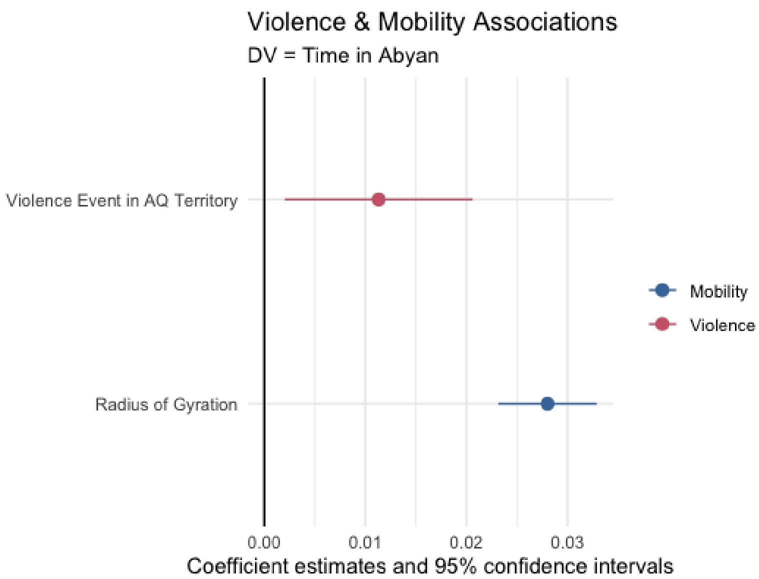


Fig. S11. Associations between migration and mobility, as well as migration adn violence, with two way fixed effects and respondent clustered errors, but no other controls. The more basic specification represented here accords with the results as described in the paper's discussion.

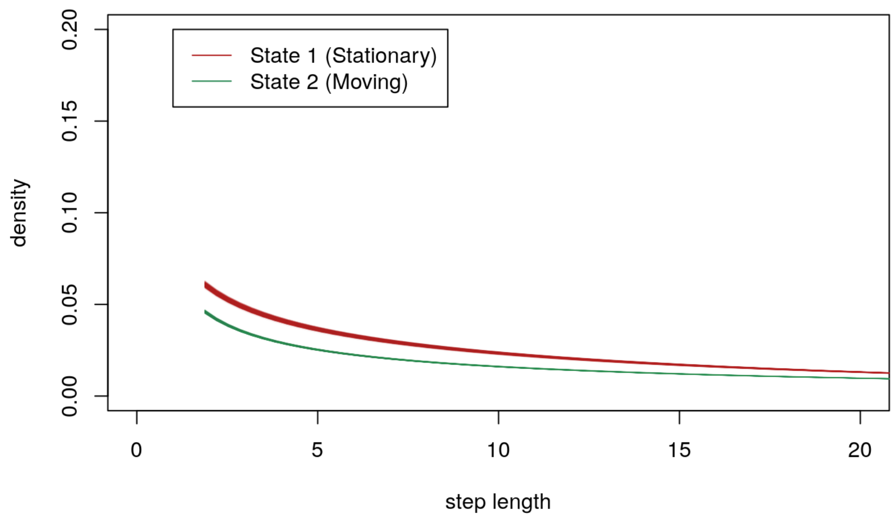


Fig. S12. Density plot of step lengths for the two hidden states from the Markov process model of migration. The state with the longer average step length (lower density of short steps) is labeled as "migratory."

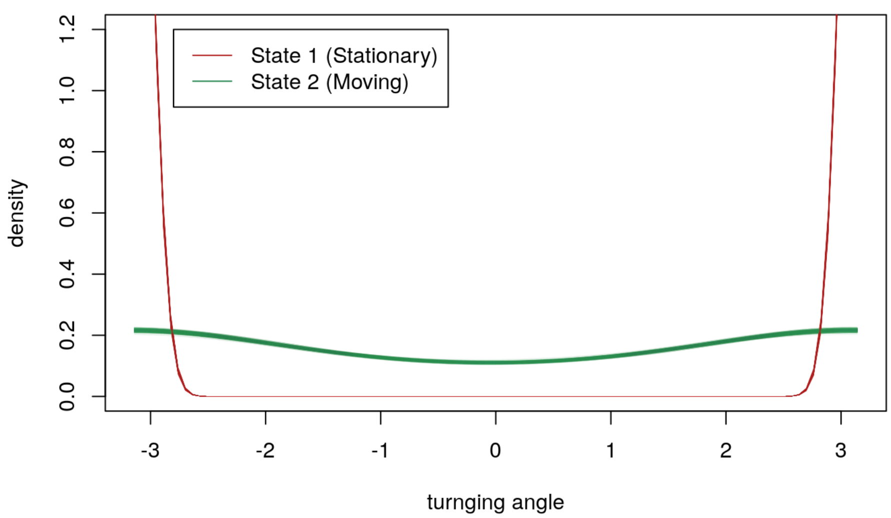


Fig. S13. Density plot of turn angles for the two hidden states from the Markov process model of migration. The state with a higher density of zero-angle turns (more straight-line movement) is labeled as "migratory."

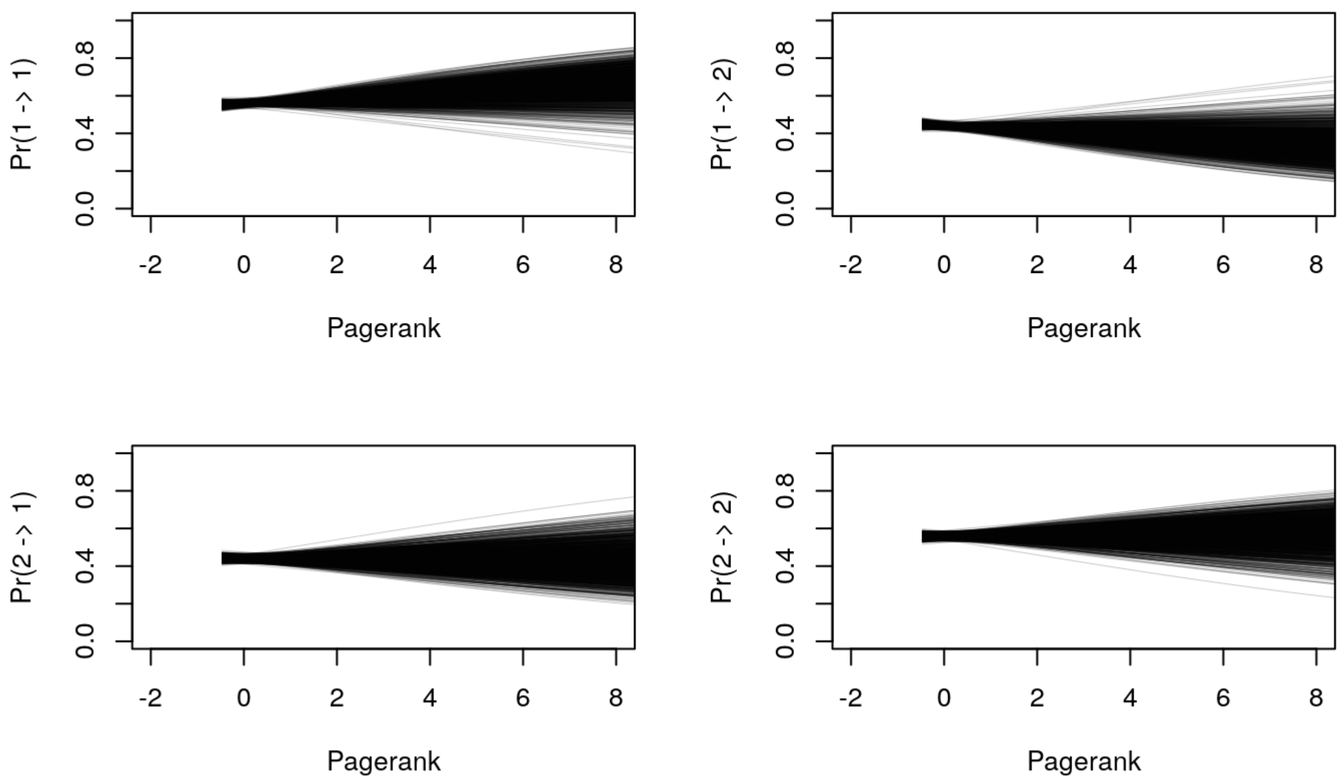


Fig. S14. State transition probabilities associated with a range of Pagerank values. Recall that State 1 is "stationary" and State 2 is "migratory." The probability of initiating migration does not appear to be increasing in pagerank, but the probability of sustaining migration does. This is consistent with our main finding that individuals with higher Pagerank are more likely, on balance, to end up outside of Abyan as a result of the conflict.

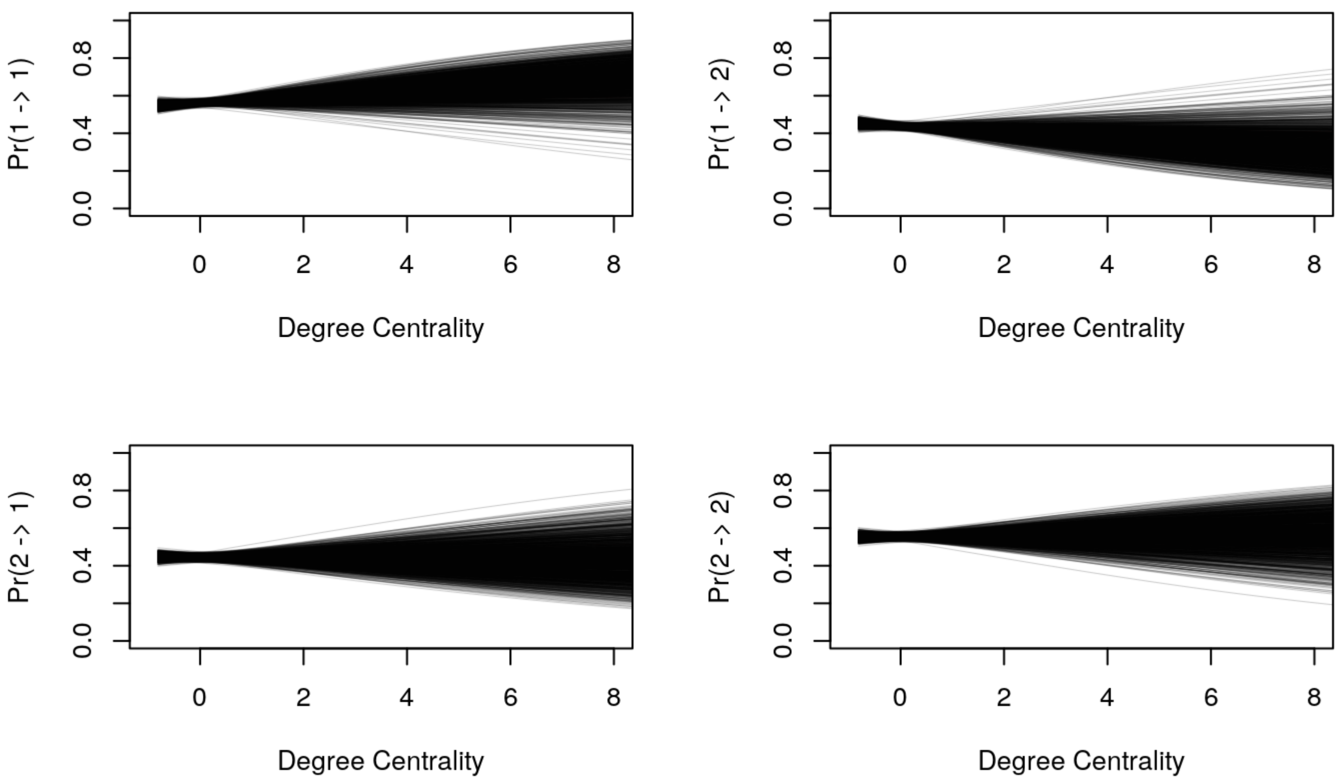


Fig. S15. State transition probabilities associated with a range of Degree Centrality values. Recall that State 1 is "stationary" and State 2 is "migratory." The probability of initiating migration is distinctly decreasing in degree centrality, and the probability of remaining stationary is distinctly increasing. This is consistent with our main finding that individuals with higher degree centrality are more likely, on balance, to remain resilient in place throughout conflict.

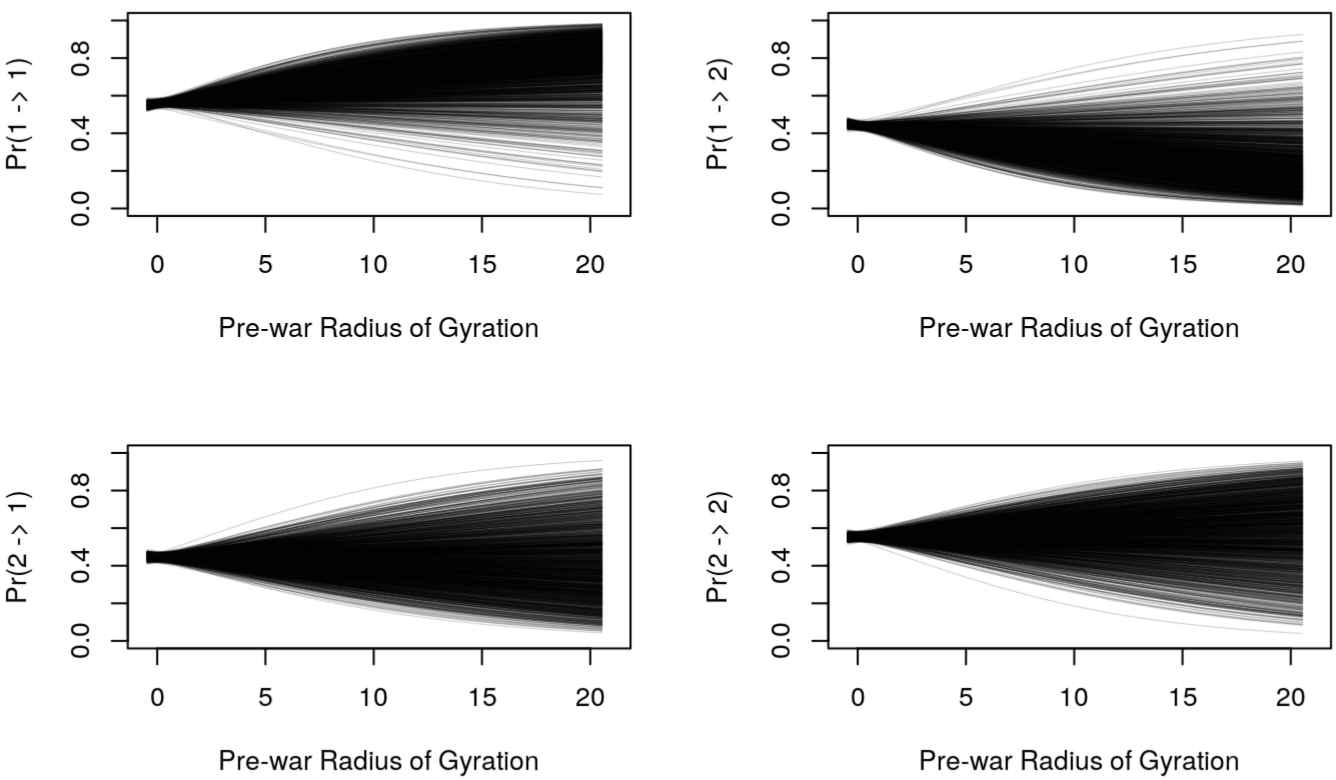


Fig. S16. State transition probabilities associated with a range of Radius of Gyration (mobility) values. Recall that State 1 is "stationary" and State 2 is "migratory." Interestingly, like pagerank, the linear association between mobility and ending up outside of Abyan appears to be largely driven by migration persistence (see the bottom right-hand pane) rather than a higher propensity to initiate migration. This too is consistent with our main findings, but adds interesting nuance.

Stan Diagnostics

Results presented in Figures 4, S6, S7, S8 and S9 are all estimated using the `Rstan` package, and other elements of the `mc-stan` modeling language ecosystem (42). The main estimation was run on the MIT Supercloud high performance computing system (46). We used 18 cores on a single Intel Xeon Gold node with 384GB of RAM.

Supplementary Hidden Markov results presented in Figures S14, S15, and S16 are all estimated using the `moveHMM` package, and other elements of the `mc-stan` modeling language ecosystem (42). The main estimation was run on the `xvii` high performance computing system in the MIT Department of Political Science. We used 4 cores on a single Intel Xeon E5-2680 node with 128GB of RAM.

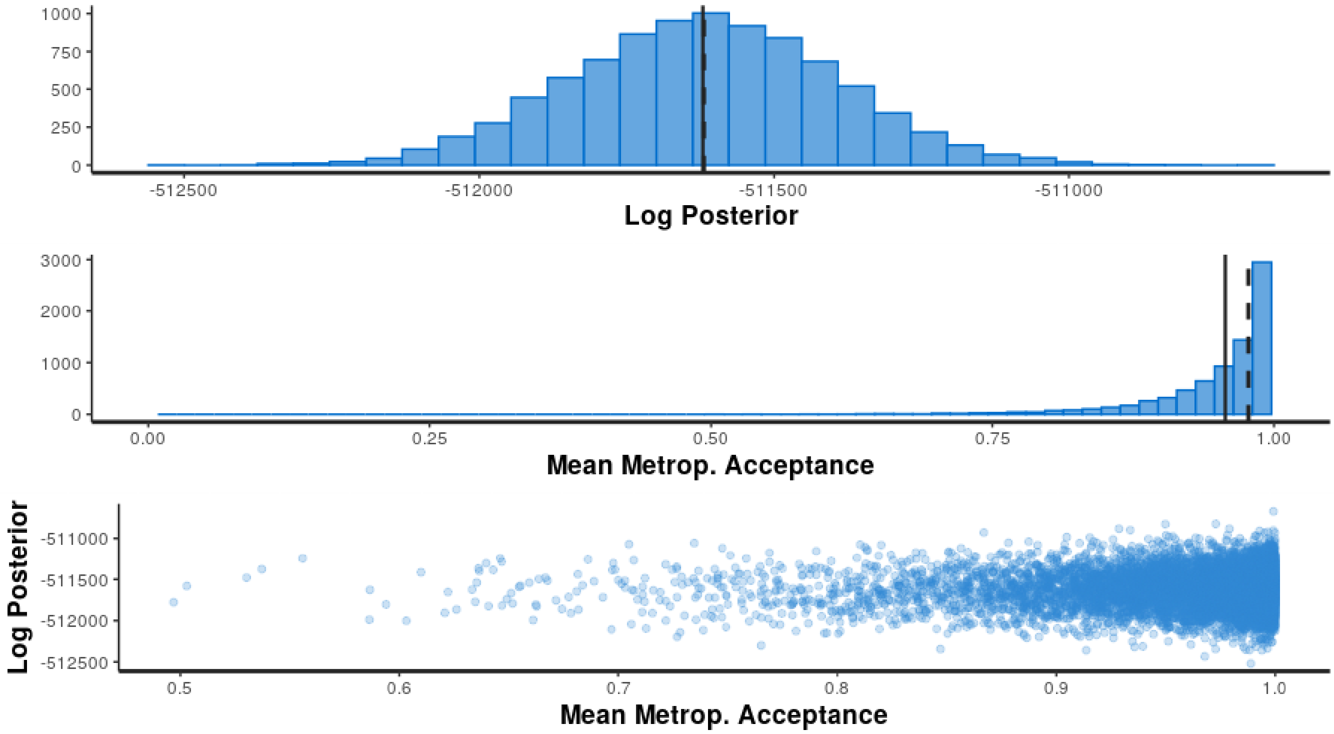


Fig. S17. Diagnostic plots from the Stan estimation we interpret in Figures S7, 4, and S6.

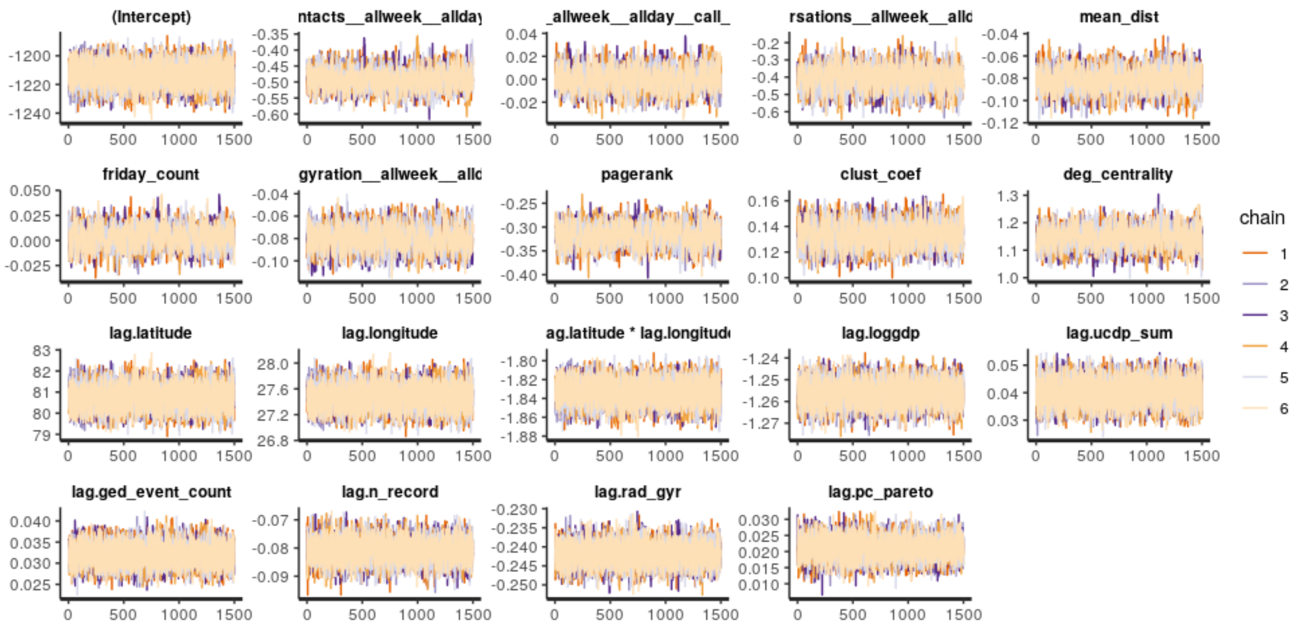


Fig. S18. Trace plots for estimation of fixed effects reported in Figures S7, 4, and S6.

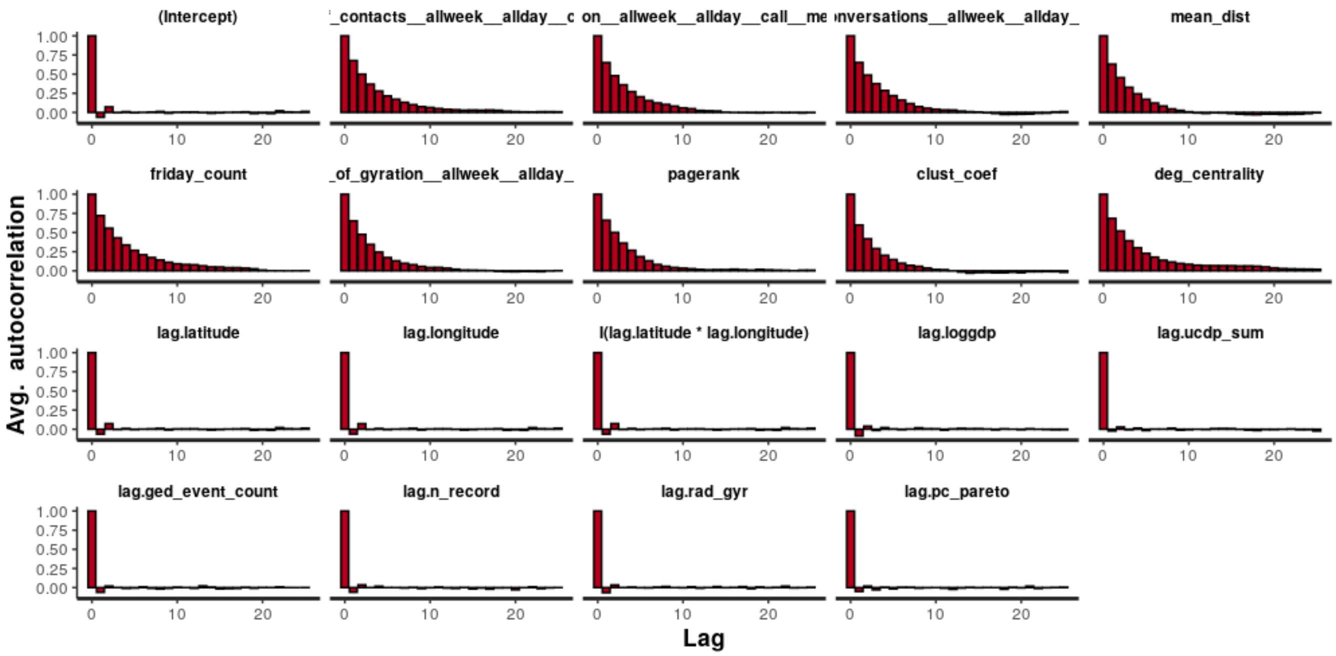


Fig. S19. Autocorrelation plots for estimation of fixed effects reported in Figures S7, 4, and S6.

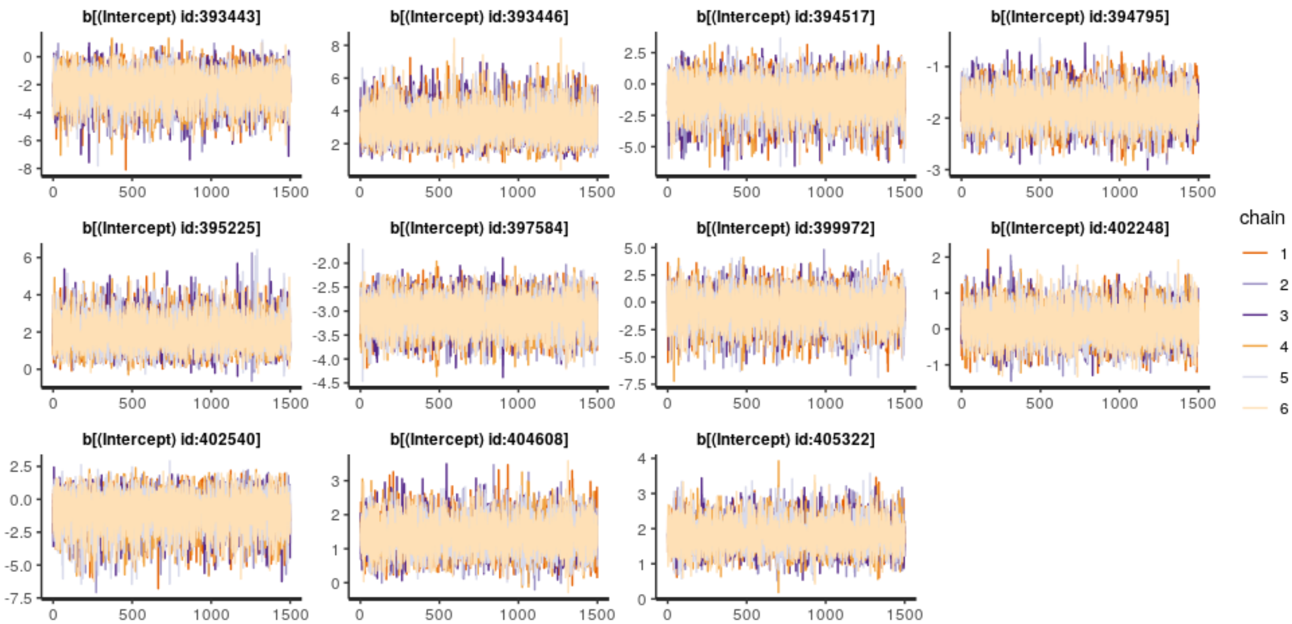


Fig. S20. Trace plots for estimation of respondent (ID) random effects from the models reported in Figures S7, 5, and S6. These diagnostics are for a randomly selected chunk of respondent (ID) random effects. Respondent random effects are not interpreted, but it's good to know if they are estimated appropriately.

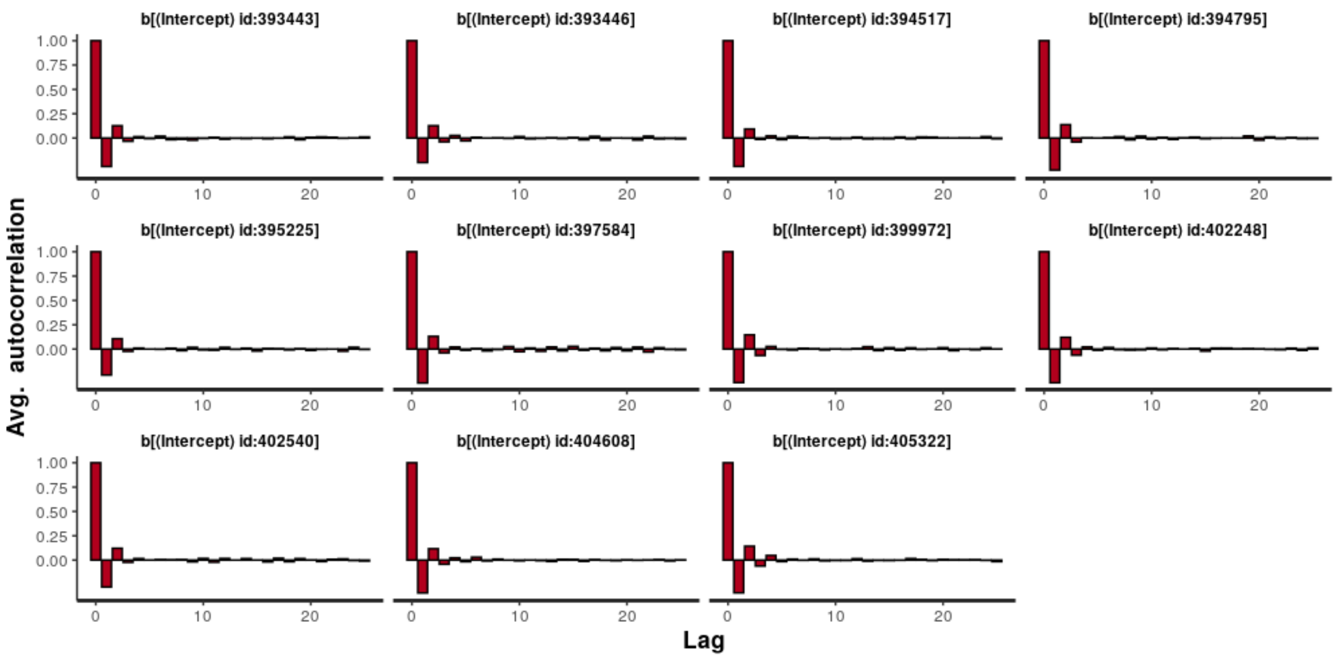


Fig. S21. Autocorrelation plots for estimation of respondent (ID) random effects from the models reported in Figures S7, 5, and S6. These diagnostics are for a randomly selected chunk of respondent (ID) random effects. Respondent random effects are not interpreted, but it's good to know if they are estimated appropriately

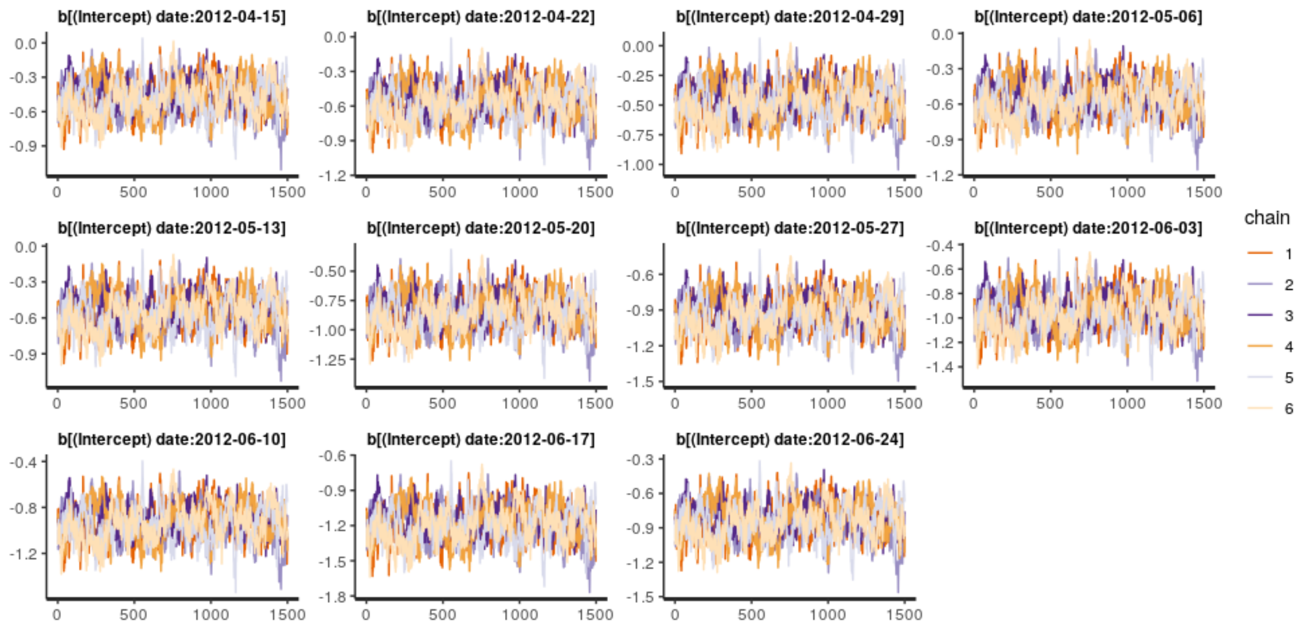


Fig. S22. Trace plots for estimation of date random effects from the models reported in Figures S7,5, and S6. Date random effects are not interpreted, but it's good to know if they are estimated appropriately. The trace plots show good mixing.

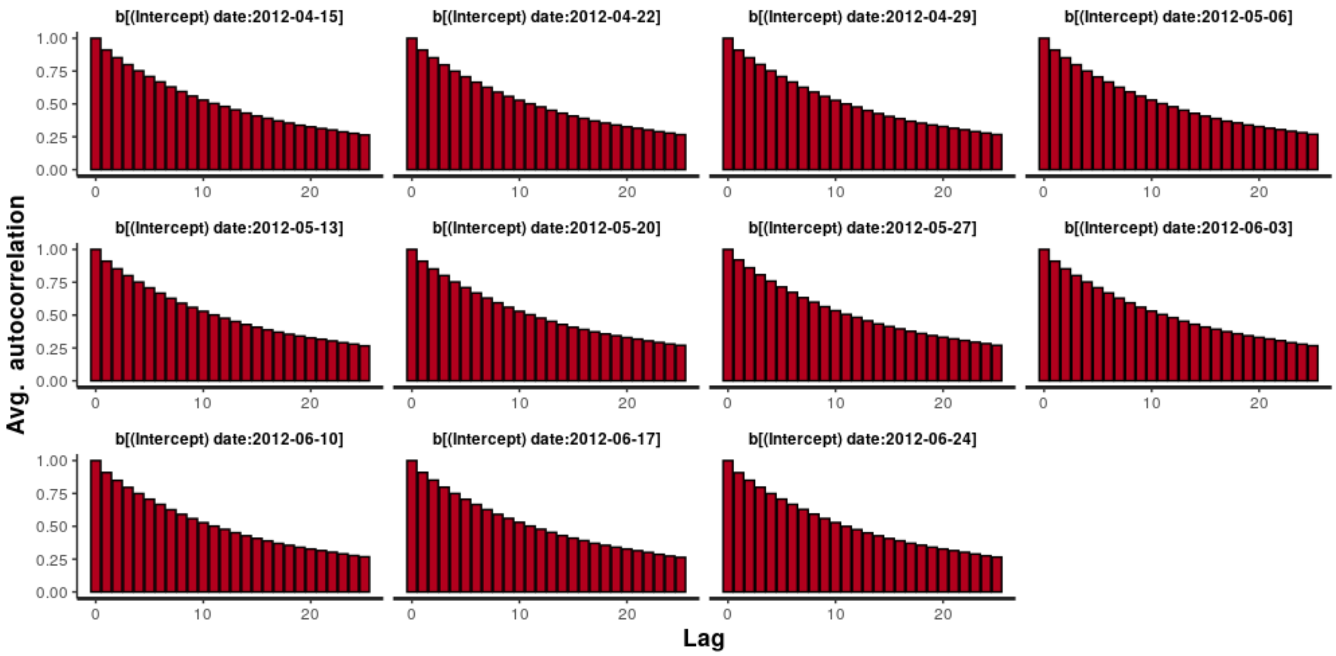


Fig. S23. Autocorrelation plots for estimation of date random effects from the models reported in Figures S7, 5, and S6. Autocorrelation over the burn-in period is still clearly higher than for IDs and for most fixed effects. There are two reasons we are not particularly concerned about this. First, the estimates produced by this model are substantially similar to the estimates from other models (not reported here for redundancy) with different initialization. We do not interpret the date random effects, and the coefficients we do interpret do not appear to depend on one or another initialization of the date random effects. Second, per Gelman (41, p. 284), our estimates of the date random effect parameters should not be particularly problematic. Our number of effective draws (`n_eff`) on these parameters is lower than for the ID random effects or the coefficients, but it substantially exceeds the recommended threshold of $5m$ where m is the number of chains after splitting, and the $\hat{R}$ values for the same parameters are all extremely close to the recommended value of 1.00. Estimates of these parameters may not be as precise as the fixed effects, but they do not appear problematic for interpretation of the model.

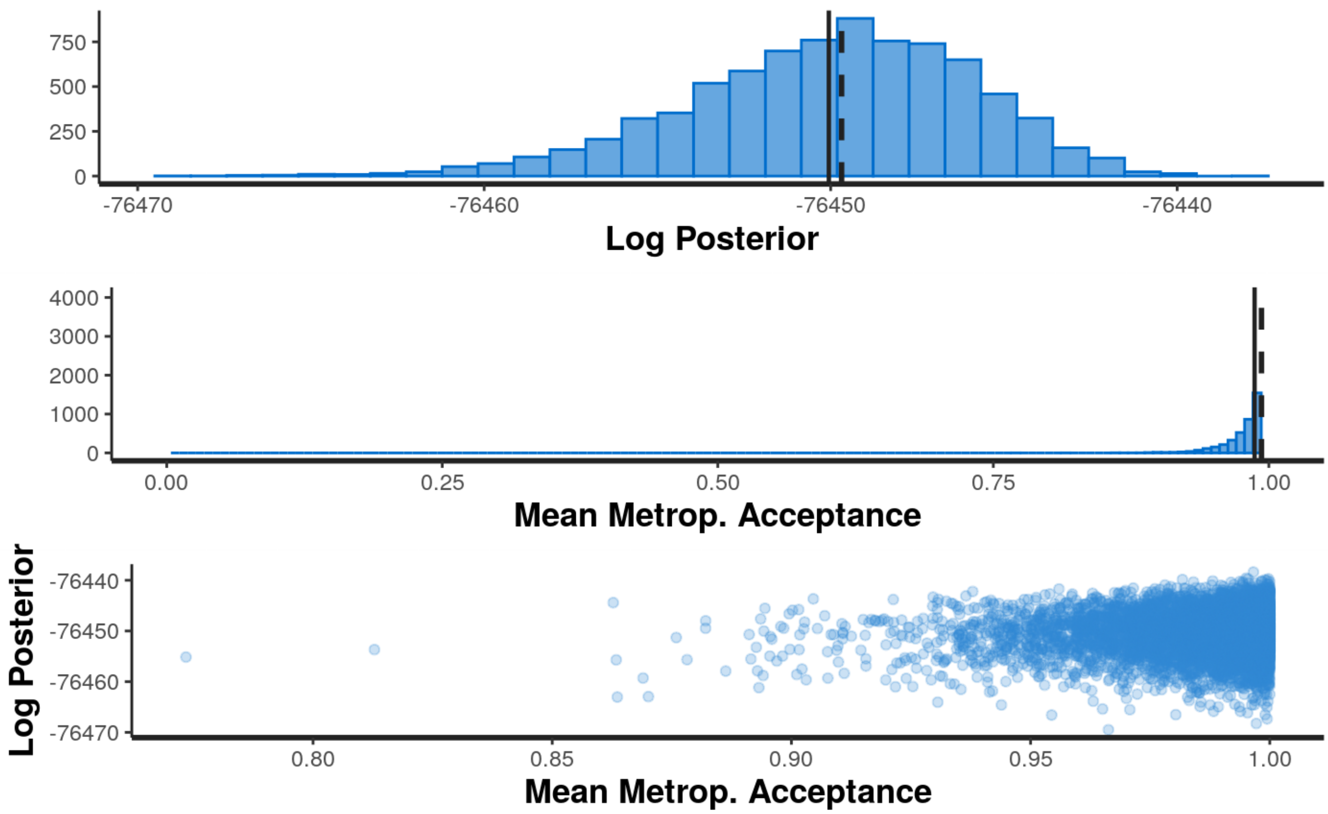


Fig. S24. Diagnostic plots from the Hidden Markov Model Stan estimation we interpret in Figures S14, S15, and S16.

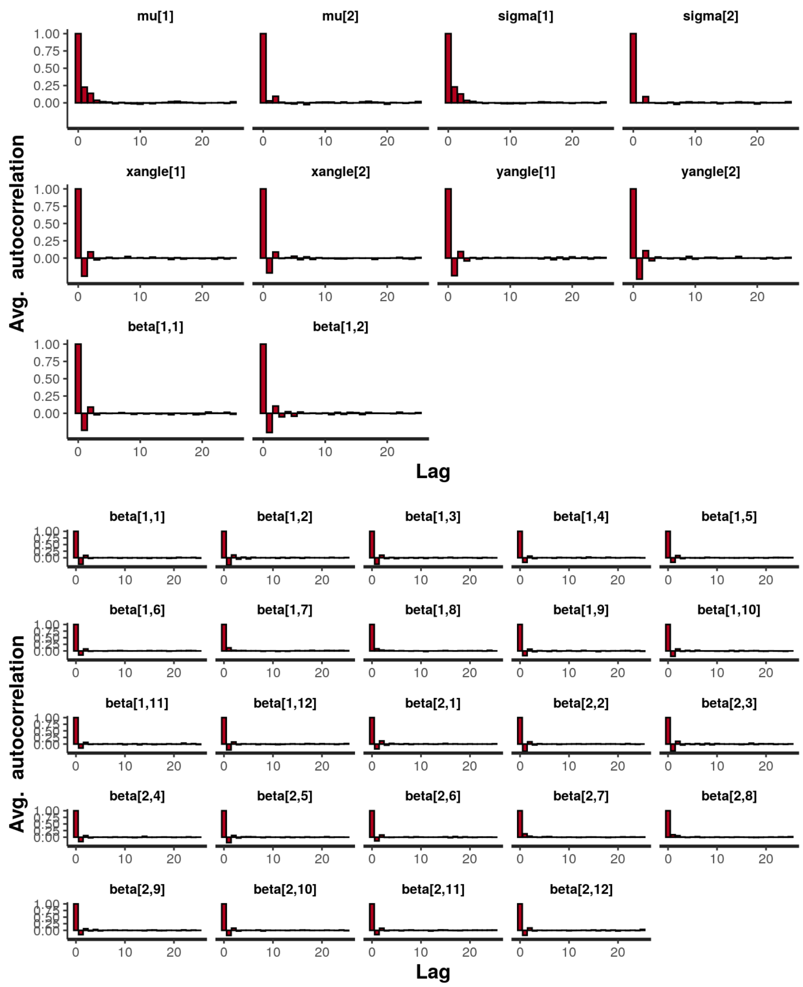


Fig. S25. Autocorrelation plots from the Hidden Markov Model Stan estimation we interpret in Figures S14, S15, and S16.

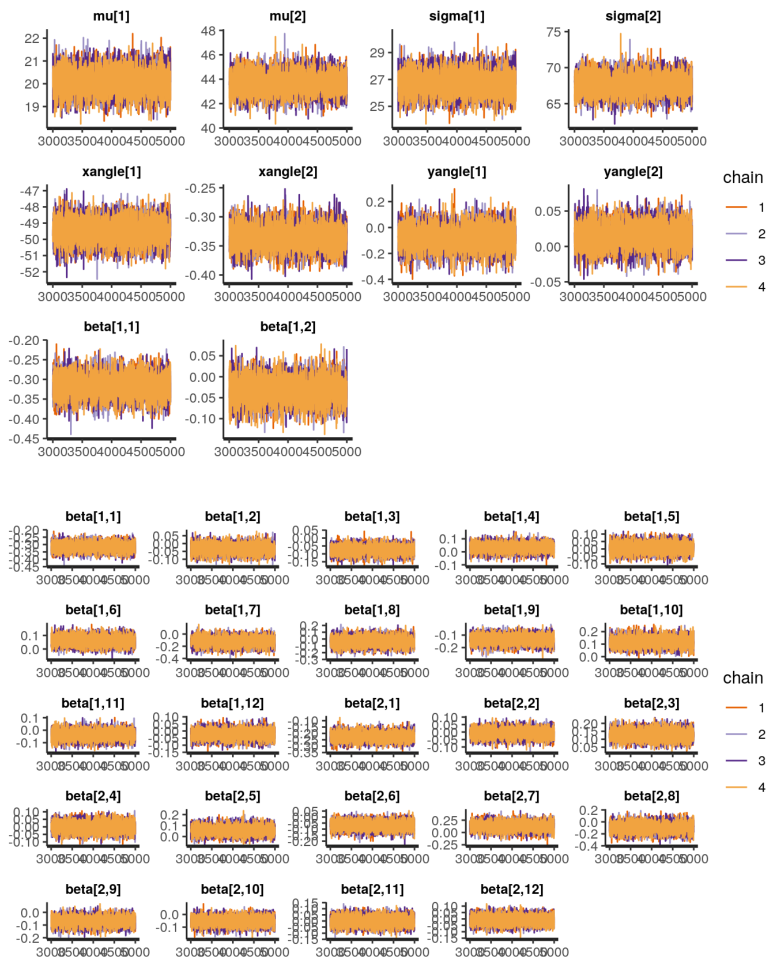


Fig. S26. Trace plots from the Hidden Markov Model Stan estimation we interpret in Figures S14, S15, and S16.

References

1. Christian Davenport, Will H. Moore, and Steven Poe. Sometimes you just have to leave: Domestic threats and forced migration, 1964-1989. *International Interactions*, 29(1):27–55, 2003.
2. Will H. Moore and Stephen M. Shellman. Fear of persecution: Forced migration, 1952-1995. *Journal of Conflict Resolution*, 48(5):723–745, 2004.
3. W. S. Robinson. Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3):351–357, 1950.
4. Myron Weiner. Bad neighbors, bad neighborhoods: An inquiry into the causes of refugee flows. *International Security*, 21 (1):5–42, 1996.
5. Susanne Schmeidl. Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly*, pages 284–308, 1997.
6. Stefanie Engel and Ana María Ibáñez. Displacement due to violence in colombia: A household-level analysis. *Economic Development and Cultural Change*, 55(2):335–365, Jan 2007.
7. Mathias Czaika and Krisztina Kis-Katos. Civil conflict and displacement: Village-level determinants of forced migration in aceh. *Journal of Peace Research*, 46(3):399–418, 2009.
8. Prakash Adhikari. Conflict-induced displacement, understanding the causes of flight. *American Journal of Political Science*, 57(1):82–89, 2013.
9. Laia Balcells and Abbey Steele. Warfare, political identities, and displacement in Spain and Colombia. *Political Geography*, 51:15–29, 2016.
10. Aree Jampaklay, Kathleen Ford, and Aphichat Chamratrithirong. How does unrest affect migration? evidence from the three southernmost provinces of thailand. *Demographic Research*, 37:25–52, Jul 2017.
11. Justin Schon. Motivation and opportunity for conflict-induced migration: An analysis of syrian migration timing. *Journal of Peace Research*, 56(1):12–27, 2019.
12. Vera Mironova, Loubna Mrie, and Sam Whitt. Risk tolerance during conflict: Evidence from aleppo, syria. *Journal of Peace Research*, 56(6):767–782, 2019.
13. Mara Revkin. To stay or to leave? displacement decisions during rebel governance. Working paper, Yale University, New Haven, CT, 2019.
14. Ala' Alrababa'h, Daniel Masterson, Marine Casalis, Dominik Hangartner, and Jeremy M. Weinstein. The dynamics of refugee return: Syrian refugees and their migration intentions. Working paper, 20-08, Stanford Immigration Policy Lab, Stanford, CA, 2020.
15. James D Fearon and Andrew Shaver. Civil war violence and refugee outflows. Working paper,, Stanford University, Stanford, CA, 2020.
16. Kara Ross Camarena, Sarah Claudy, Jijun Wang, and Austin L. Wright. Political and environmental risks influence migration and human smuggling across the mediterranean sea. *PLOS ONE*, 15(7), 07 2020.
17. Jerome Marston. Resisting displacement: Leveraging interpersonal ties to remain despite criminal violence in medellin colombia. *Comparative Political Studies*, 53(13):1995–2028, 2020.
18. Sascha O. Becker, Volker Lindenthal, Sharun Mukand, and Fabian Waldinger. Persecution and Escape: Professional Networks and High-Skilled Emigration from Nazi Germany. SoDa Laboratories Working Paper Series 2021-02, Monash University, 2021.
19. Xiao Hui Tai, Shikhar Mehra, and Joshua E. Blumenstock. Mobile phone data reveal the effects of violence on internal displacement in afghanistan. *Nature Human Behaviour*, 6(5):624–634, 2022.
20. Paolo Bertolotti, Fotini Christia, and Ali Jadbabaie. The social network effects of drone strikes. Technical report, Massachusetts Institute of Technology, Cambridge, MA, 2019.
21. Paolo Bertolotti, A. Jadbabaie, and Fotini Christia. Tests for network cascades via branching processes. *IEEE Transactions on Network Science and Engineering*, 2020.
22. Fotini Christia, Michael Curry, Constantinos Daskalakis, Erik Demaine, John P. Dickerson, MohammadTaghi Hajiaghayi, Adam Hesterberg, Marina Knittel, and Aidan Milliff. Scalable equilibrium computation in multi-agent influence games on networks. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-21)*, Palo Alto, 2021. Association for the Advancement of Artificial Intelligence.
23. M.C. Gonzalez, C.A. Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.
24. Joshua Blumenstock, Ye Shen, and Nathan Eagle. A method for estimating the relationship between phone use and wealth. In *ICTD 2010*, London, 2010. ICT4 Collective.
25. Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone data. *Science*, 350(6264):1073–76, 2015.
26. Gabriel E. Kreindler and Yuhei Miyauchi. Measuring Commuting and Economic Activity inside Cities with Cell Phone Records. Boston University - Department of Economics - Working Papers Series WP2020-006, Boston University - Department of Economics, February 2019. URL https://ideas.repec.org/p/bos/wpaper/wp2020-006.html.
27. James P. Bagrow, Dashun Wang, and Albert-László Barabási. Collective response of human populations to large-scale emergencies. *PLoS ONE*, 6(3):e17680, 2011.
28. Xin Lu, Linus Bengtsson, and Petter Holme. Predictability of population displacement after the 2010 haiti earthquake.

Proceedings of the National Academy of Sciences, 109(29):11576–11581, 2012.

29. Joshua Blumenstock, Guanghua Chi, and Xu Tan. Migration and the value of social networks. Working paper, U.C. Berkeley, Berkeley, CA, 2019. URL http://www.jblumenstock.com/files/papers/migration_networks.pdf.
30. Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.
31. Bjarke Felbo, Pål Sundsøy, Alex 'Sandy' Pentland, Sune Lehmann, and Yves-Alexandre de Montjoye. Modeling the temporal nature of human behavior for demographics prediction. Arxiv preprint, MIT Media Lab, Cambridge, MA, 2017.
32. Joshua E. Blumenstock. Estimating economic characteristics with phone data. *AEA Papers and Proceedings*, 108:72–76, May 2018. . URL https://www.aeaweb.org/articles?id=10.1257/pandp.20181033.
33. Tugba Bozcaga, Fotini Christia, Elizabeth Harwood, Constantinos Daskalakis, and Christos Papademetriou. Syrian refugee integration in turkey: Evidence from call detail records. In Albert Ali Salah, Alex Pentland, Bruno Lepri, and Emmanuel Letouzé, editors, *Guide to Mobile Data Analytics in Refugee Scenarios, The 'Data for Refugees Challenge' Study*, pages 223–249. Springer, 2019.
34. Larry Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report SIDL-WP-1999-0120, Stanford University InfoLab, Stanford, CA, 1998. URL http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf.
35. Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440–442, 1998.
36. Tilottama Ghosh, Rebecca Powell, Christopher Elvidge, Kimberly Baugh, Paul Sutton, and Sharolyn Anderson. Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3:147–160, 2010.
37. Seth Goodman, Ariel BenYishay, Miranda Lv, and Dan Runfola. Geoquery: Integrating hpc systems and public web-based geospatial data tools. *Computers and Geosciences*, 122:103–112, 2019.
38. Uppsala Conflict Data Program. conflict encyclopedia. Technical report, Uppsala University.
39. Yves-Alexandre de Montjoye, Luc Rocher, and Alex (Sandy) Pentland. bandicoot: a python toolbox for mobile phone metadata. *Journal of Machine Learning Research*, 17(175):1–5, 2016.
40. Therese Pettersson and Magnus Oberg. Organized violence, 1989-2019. *Journal of peace Research*, 57(4):597–613, 2020.
41. Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis, Third Edition.* CRC Press, Nov 2013.
42. The Stan Development Team. Stan modeling language users guide and reference manual,. https://mc-stan.org, February 2021.
43. P. M. Aronow and Cyrus Samii. Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1):250–267, 2016. . URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12185.
44. Toby A. Patterson, Alison Parton, Roland Langrock, Paul G. Blackwell, Len Thomas, and Ruth King. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438, 2017. . URL https://doi.org/10.1007/s10182-017-0302-7.
45. Théo Michelot, Roland Langrock, and Toby A. Patterson. movehmm: an r package for the statistical modelling of animal movement data using hidden markov models. *Methods in Ecology and Evolution*, 7(11):1308–1315, 2016. . URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12578.
46. Albert Reuther, Jeremy Kepner, Chansup Byun, Siddharth Samsi, William Arcand, David Bestor, Bill Bergeron, Vijay Gadepally, Michael Houle, Matthew Hubbell, et al. Interactive supercomputing on 40,000 cores for machine learning and data analysis. In *2018 IEEE High Performance extreme Computing Conference (HPEC)*, pages 1–6. IEEE, 2018.