Social capital mediates knowledge gaps in informing sexual and reproductive health behaviours across Africa

Till Koebe[1*], Theophilus Aidoo[1], Ridhi Kashyap[2, 3], Douglas R. Leasure[4, 2], Valentina Rotondi[5, 2], Ingmar Weber[1]

1 Saarland Informatics Campus, Saarland University, Saarbrücken, Germany

2 Nuffield College, University of Oxford, Oxford, United Kingdom

3 Department of Sociology, University of Oxford, Oxford, United Kingdom

4 Department of Population Health, University of Oxford, Oxford, United Kingdom

5 Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Italian Switzerland, Lugano, Switzerland

* Corresponding author: till.koebe@uni-saarland.de

Abstract

Advancing sexual and reproductive health is essential for promoting human rights and women's empowerment, and combating the HIV/AIDS epidemic. A large body of literature across the social sciences emphasizes the importance of social capital, generated through the strength of social networks, for shaping health behaviours. However, large-scale measurement of social capital and social networks remains elusive, especially in the context of low-income countries. Here we delve into the role of social capital dynamics, and in particular social connectedness across communities as measured through Facebook friendship links, in shaping knowledge diffusion and behavior related to sexual and reproductive health in 495 regions across 33 countries in Africa. Our findings demonstrate that regions with higher levels of social connectedness are more similar in their knowledge about contraception and HIV testing, as well as their adoption of these behaviours. We further observe that the mediating role of social connectedness becomes stronger when the knowledge gaps between regions are larger. In other words, regions are more similar in behaviours, despite knowledge gaps, when they are socially connected. These insights carry significant policy implications, especially for the design and targeting of public health campaigns. We highlight that social connectedness can serve both as a driver and a roadblock in behavior formation, underscoring the importance of understanding its influence on health-related outcomes.

1 Introduction

The concept of social capital has garnered significant attention in academic circles, spanning multiple disciplines such as sociology, political science, economics, education, and anthropology. Though not precisely defined, social capital generally refers to the features of social life that foster cooperation and coordination among individuals with shared goals [Fukuyama, 1995, Putnam, 2001]. Rooted in social networks, social capital emerges from individuals' connections and interactions with others, generating valuable resources for collective action [Bourdieu, 1986, Coleman, 1994, Lin et al., 2001].

The existence of social capital depends on the quality of networks, their ability to foster social trust [Sabatini, 2009], the actions individuals take to build social trust and reciprocity within and towards these networks, and the resources available within their connections [Portes, 2000]. Trust is often considered the cognitive component of social capital, while networks are viewed as its structural component [Burt, 2000]. Social capital's structural and cognitive components are intricately linked, positively or negatively [Sabatini, 2009]. Social trust, for example, can enhance cooperative behaviors that lead to the formation of networks, and these networks, in turn, strengthen trust and reciprocity. Conversely, certain types of networks can hinder trust by restricting external access [Woolcock, 2001].

In this context, social networks and social trust are valuable assets enabling individuals to build communities, establish commitments, and ultimately cooperate. However, cooperation brings benefits and costs, as it enhances the welfare of individuals within the group while potentially decreasing the welfare of non-members. These contrasting effects are commonly referred to as the positive or bright side and the negative or dark side of social capital, which have been recognized in the literature for a considerable time.

In the context of health, social capital plays a pivotal role, impacting health outcomes through direct and indirect pathways. Social support, derived from social capital, significantly influences an individual's health [Fiorillo and Sabatini, 2011, 2015, Tomioka et al., 2016]. On the other hand, social isolation leads to adverse health consequences, including increased stress levels, hypertension, and premature mortality [Cole et al., 2015, Luo et al., 2012, Cacioppo and Cacioppo, 2014]. Moreover, low social trust is associated with a higher prevalence of psychosomatic symptoms, musculoskeletal pain, and depression [Åslund et al., 2010]. However, often it is not merely the presence of social connections that yields health benefits; rather, it is the quality, content, and available resources within these connections that play a crucial role in the relationship between social capital and health outcomes [Moore et al., 2009].

Within the realm of sexual and reproductive health, social capital is of particular importance, especially concerning the spread of sexually transmitted infections like HIV/AIDS. Social trust has been identified as a critical factor facilitating timely HIV testing, leading to early diagnosis and appropriate care. This trust serves as an essential determinant in monitoring HIV care outcomes, particularly for vulnerable groups disproportionately affected by HIV [Ransome et al., 2017]. Strong social capital, characterized by high social cohesion, has shown to enhance HIV testing uptake in certain settings, demonstrating the effectiveness of building trust and solidarity within and between groups [Fonner et al., 2014, Grover et al., 2016].

However, social capital can also have negative effects on sexual behavior. For instance, research by Kalolo et al. [2019] revealed a suggestive, albeit statistically nonsignificant, association between social trust and engaging in multiple sexual partnerships. This finding suggests that adolescents who exhibit trust in others may be susceptible to the influence of behaviors endorsed by their social group. Consequently, interventions that leverage social networks and engage influential individuals can effectively address the multifaceted dynamics of sexual behavior and contribute to the prevention of sexually transmitted infections, including HIV. By targeting influential figures and leveraging social connections, interventions can promote positive sexual health outcomes and mitigate the potential negative impact of social capital on sexual behavior.

In this study, our focus lies on Sub-Saharan Africa (SSA), where combating HIV/AIDS remains an ongoing challenge due to unsafe sexual behavior among adolescents. Specifically, we seek to investigate the impacts of social capital on sexual and reproductive health outcomes. To address this, we leverage data on the Social Connectedness Index (SCI) from Facebook and its parent company Meta, which we combine with the Demographic and Health Surveys, to examine knowledge diffusion and behavior formation linked to sexual and reproductive health. The SCI, which assesses the likelihood of individuals in different regions being connected through Facebook friendship links, provides a novel and data-driven approach to studying social capital dynamics, particularly in Low- and Middle-Income Countries, where such data have been sparse.

Our contributions in this paper are twofold: First, we demonstrate that the SCI serves as a promising proxy for social capital in Africa, offering advantages in coverage, timeliness, and potentially frequency compared to traditional survey-based measures. Second, we show that social connectedness, as measured by the SCI, plays a mediating role in shaping health behavior and knowledge related to sexual and reproductive health. These findings carry crucial policy implications, as health information campaigns need to consider knowledge gaps among socially-connected regions. We highlight the significance of understanding how social connectedness can both drive and hinder behavior formation, making it a vital factor in designing effective public health interventions.

The paper is organized as follows: In Section 2, we provide a concise presentation of the conceptual background. Section 3 delves into the data used for this study, emphasizing the connection between survey results and network information obtained from the Social Connectedness Index (SCI). In Section 4, we outline the methodology employed to estimate the relationships between social connectedness and behavior, as well as knowledge related to sexual and reproductive health. Moving on to Section 5, we present the main findings, first establishing the SCI's validity as a proxy for social capital and then exploring its impact on sexual and reproductive health behavior and knowledge. Lastly, in Section 6, we discuss the policy implications arising from our research, with a particular focus on the role of social connectedness in shaping public health campaigns.

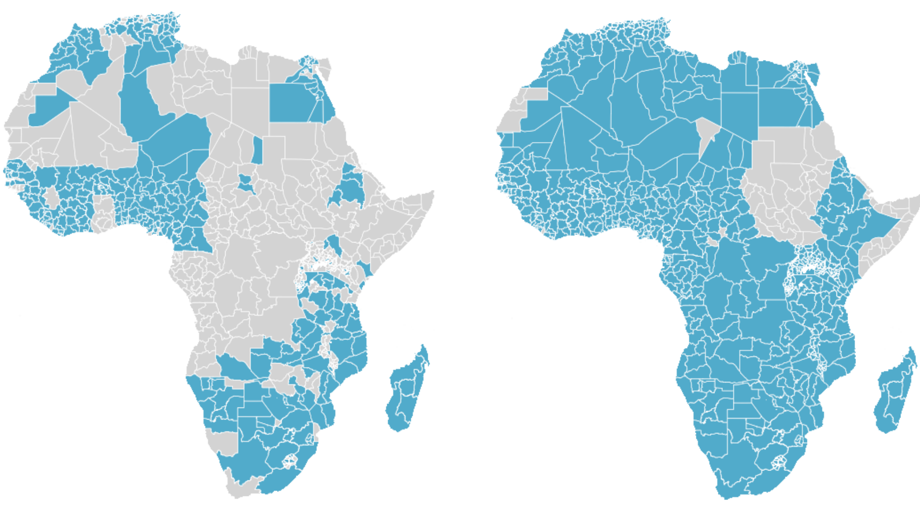
2 Conceptual Background

Existing social demographic literature has emphasized the significant role that social interactions, or in other words the connections between individuals and their social networks, play in the dissemination of ideas, behaviours, and preferences related to fertility and reproduction, particularly within linguistically or culturally homogeneous populations [Cleland and Wilson, 1987, Bongaarts and Watkins, 1996, Montgomery and Casterline, 1996, Entwisle et al., 1996, Kohler et al., 2001, Behrman et al., 2002]. This literature has argued that individual socioeconomomic characteristics or their access to institutions such as family planning programmes are insufficient in themselves for explaining changes in reproductive behaviours in low- and middle-income countries (LMIC), and as such, the pace of fertility decline is better explained through understanding the diffusion of new behaviours that are facilitated by social interactions in social networks. In other words, this literature highlights that social capital, generated through exchange and interaction within social networks, is capable of shifting behaviours.

Social networks, and the social capital they generate, can shift behaviours through two potential mechanisms – social learning or social influence [Montgomery and Casterline, 1996]. When it comes to adopting new behaviors, such as using modern contraceptive methods, individuals often face the challenge of embracing innovative practices in an environment characterized by high uncertainty. This is especially true in low-income country contexts, such as the African continent, where access to reproductive health services is still limited and unmet need for family planning is high [Cleland et al., 2006]. In this context, social capital becomes crucial as individuals rely on trusted sources of information to learn about and adopt these new behaviors (social learning). Social connections can help provide individuals with new information, shaping their knowledge, attitudes, and beliefs regarding reproductive health choices. On the other hand, social networks can also operate by exerting social influence, by shaping the normative context in which individuals alter their behaviour in response to the behaviour of others. Although large-scale data on social networks, social connectedness and social capital are rare, particularly in low- and middle-income countries, small-scale studies, drawing on specialized data collection on social networks, e.g. in Kenya [Behrman et al., 2002, Kohler et al., 2001], or Thailand [Entwisle et al., 1996], show the importance of social learning through social networks for contraceptive adoption. The lack of data on social capital in LMICs has meant that the role of social capital in the adoption of health-related behaviours, particularly in relation to sexual and reproductive health, has received limited empirical attention, despite the theoretical salience attributed to these processes.

The digital revolution, encompassing the spread of mobile phones and the internet, has provided researchers with unprecedented access to vast amounts of data, offering new insights into a wide range of socio-economic and population phenomena [Kashyap, 2021, Schmid et al., 2017]. The new data streams generated by the use of digital technologies have the potential to lend themselves for new types of social measurement, including phenomena for which large-scale social data are limited, such as social capital. The scarcity of large-scale, detailed, and comparable datasets on social capital poses a challenge for researchers in this field [Chetty et al., 2022]. To address this issue, researchers have turned to Facebook-generated social capital data, utilizing the Social Connectedness Index (SCI) developed by Meta. The SCI quantifies social connectedness by assessing the likelihood of Facebook users in different regions being connected through Facebook friendship links. This index has been instrumental in exploring the impact of social capital on various socio-economic outcomes, including economic prosperity, international trade, compliance with pandemic restrictions, and predicting COVID-19 cases and flood insurance decisions [Jahani et al., 2023, Bailey et al., 2021, Charoenwong et al., 2020, Vahedi et al., 2021, Hu, 2022]. However, in much of this work the SCI's effectiveness as a proxy for social capital has often been assumed rather than rigorously validated. Moreover, most of these studies have focused on high-income country contexts with high Facebook penetration rates, largely focusing on the US.

In our specific context, focusing on Africa and examining sub-national regions within countries across the continent, we can take advantage of the Afrobarometer, a nationally-representative public attitudes survey run in multiple countries in Africa, to assess how the SCI validates against other widely used measures of social capital operationalized in survey data. Details on the data can be found in Section 3.



(a) Afrobarometer's generalized trust indicator, 2011-13 (b) Meta's Social Connectedness Index, 2021

Figure 1: **Geographical coverage and timeliness of the latest data on social capital in Africa, by source.** (a) The latest data on the generalized trust indicator from the Afrobarometer covers 395 out of 859 African regions between 2011-13. (b) Meta's Social Connectedness Index (SCI) provides data for 710 out of 859 African regions for 2021.

Figure 1 compares the geographical coverage of the SCI from 2021 with coverage of the latest data on a commonly used measure used to capture social capital available for parts of the African continent – the generalized trust indicator from the years 2011-13 from the Afrobarometer surveys. Beyond the evident benefits of timely updates and potential for frequent data refreshes with the SCI, the vast geographical scope of the data is particularly striking. Within the 859 African regions listed under administrative level 1 in GADM v2.8 [GADM, 2015], data from the SCI encompasses 710 regions, in contrast to the Afrobarometer which covers 395. Consequently, the SCI presents a promising avenue for more expansive research into social capital dynamics, contingent upon its validation as a reliable proxy for measuring social capital.

3 Data

For our study, we mainly draw on three distinct data sources: the Afrobarometer, Meta's Social Connectedness Index, and the Demographic and Health Surveys (DHS).

Initiated in 1999, the Afrobarometer has consistently run individual-level surveys to gauge attitudes on a spectrum of political, economic, and social issues across Africa [BenYishay et al., 2017]. This study taps into the data from Afrobarometer Round 5, which spanned 27 sub-Saharan African countries between 2011 and 2013, offering the most current and comprehensive insight into survey-based measures of social capital within the continent. Leveraging the Afrobarometer's geolocalized data, this study incorporates information on social (generalized) trust and social participation, two widely-used indicators associated with social capital within the survey. Social trust is measured using a binary variable derived from respondents' answers to a question regarding trust in people within their country [Rosenberg, 1956], while social participation is assessed as a binary variable indicating whether respondents are members of any community or volunteer group.

The SCI developed by Facebook's parent company Meta quantifies social connectedness of two locations by assessing the probability of Facebook users in these locations being connected through Facebook friendship links [Bailey et al., 2018]. The latest SCI data is available for the year 2021 for large parts of the world on the sub-national level.

The DHS is a long-standing, large-scale household-survey programme of nationally representative surveys across low- and middle-income countries, funded by the United States Agency for International Development (USAID). We consider DHS surveys since 2010 onward in our study to allow for a large sample, thereby implicitly assuming that behavioural and knowledge information from earlier years still hold value for the analysis. A list of the DHS surveys used in the analysis can be found in Table 7 in the Appendix.

More specifically, in the first part of the paper, we look at survey outcomes related to social capital and trust covering 249 sub-national regions across Africa. For the second part, we look at survey outcomes related to sexual and reproductive health derived from DHS covering 497 regions across 33 African countries. While the DHS provides information on individuals – primarily women and men of reproductive ages – and the households they live in, and therefore follows a household survey structure, the SCI provides information on the connectedness of pairs of regions, thus representing a network structure. For each pair of the 497 regions, the SCI provides a measure of social connectedness, resulting in $497 \times 497 = 247,009$ regional edges. In order to analyze them jointly, aligning those two data sources requires either a) aggregating the SCI of a given region across its ties, thus leaving us with 497 observations or b) calculating the differences of aggregated survey information for pairs of regions, thus leaving us with 247,009 observations. As the correspondence between measures linked to social capital in Afrobarometer survey and SCI datasets are key for exploring the appropriateness of the SCI as a proxy of social capital, we opt for the first option, also called the *node-based* approach for the first part of the analysis. We exploit the second, also in the following called *edge-based* approach to identify whether differences in regional health outcomes can be explained by how socially connected those regions are.

Even though the SCI is available for a majority of regions in the world, we limit the extent of the network to the

247,009 edges of the 497 regions for which we have DHS information and/or social capital data. For linking the information on social capital from the Afrobarometer to the SCI, we average both the edge-level SCI across each region's ties and the social trust and social participation values for the regional level via Afrobarometer's cluster locations. In addition, we aggregate various groups of control variables to the regional level resulting in a joint sample of 99 to 249 sub-national regions across Africa for further analysis.

The SCI measures the relative probability that two Facebook users across two locations are friends on Facebook. We denote the SCI as

$$\text{SCI}_{a,b} = \frac{\text{FB\_Connections}_{a,b}}{\text{FB\_Users}_a \times \text{FB\_Users}_b}, \tag{1}$$

where $\text{FB\_Connections}_{a,b}$ is the number of Facebook connections measured as friendship links between location a and location b and FB_Users is the number of Facebook users in location a and b, respectively. The location of a user is either self-declared by the user on its profile or estimated from other network information. For public release, Meta provides a scaled version of the SCI with some additional privacy measures applied. For details on the SCI methodology, we refer to Bailey et al. [2018]. Without spelling out the privacy measures in detail, the SCI we use in this study is defined by:

$$\text{scaled\_SCI}_{a,b} = \frac{\text{SCI}_{a,b}}{\max_{i,j}(\text{SCI}_{i,j})}, \tag{2}$$

where $\max_{i,j}(\text{SCI}_{i,j})$ describes the highest regional-level SCI value available in the global SCI dataset provided by Meta. Thus, the SCI itself is an undirected graph, which means the edges between a pair of nodes (regions in our case) have identical SCI values. However, in order to align the survey structure with the SCI structure by taking the differences between regional-level statistical indicators, the direction starts to matter. Consequently, for the edge-based approach, from the 247,009 edges available between the 497 regions, we consider just one side of the difference matrix, namely when the differences in our outcomes of interest are positive ($\Delta >= 0$). This approach reduces the number of observations to roughly a half. Effects of this decision are further discussed in Section 5. In addition, as we expect the Facebook penetration rate to be an important control variable in subsequent analysis, we use data from the Facebook marketing API available for 495 of the 497 regions in our sample as described in Kashyap et al. [2020], reducing the final sample size to $n = 122,760$ edges. Table 1 provides an overview of the dataset we use for the edge-based analysis.

Table 1: Final dataset characteristics of the edge-based approach

Final dataset coverage	
Individuals	684,928
Regions	495
Edges	122,760
Countries	33
% females (weighted)	70.5
% urban (weighted)	39.5

For both the node- and the edge-based part of the study, we focus on two important survey indicators, respectively. For the former, we use the indicators related to generalized trust and social participation from the Afrobarometer described above. For the latter, we focus on two important behavioural indicators from the DHS related to sexual

Region 1	Region 2	scaled_SCI	Region 1: Use of HIV test	Region 2: Use of HIV test	ΔUse of HIV test
A	B	0.3	0.4	0.2	0.4 - 0.2 = 0.2
A	C	0.7	0.4	0.6	0.4 - 0.6 = -0.2
B	A	0.3	0.2	0.4	0.2 - 0.4 = -0.2

Table 2: **Translating classical survey data into the edge-based setting.** Differencing can be done in two directions for a pair of regions, e.g. A-B and B-A. However, we only consider the absolute differences in health outcomes in the analysis as the SCI is direction-invariant.

and reproductive health: a) *Does the respondent use a modern method of contraception?* b) *Has the respondent ever been tested for HIV?*. As we are interested in the channels that explain these behaviours, we consider the SCI and corresponding knowledge indicators, specifically c) *Does the respondent know about modern methods of contraception?* and d) *Does the respondent have knowledge about HIV transmission?* calculated as a linear index across a set of HIV-related knowledge indicators asked within the DHS.

In order to further control for other factors related to social capital on one hand and for general levels of human development and physical connectedness on the other, we draw on additional datasets, namely satellite-derived covariates from WorldPop [Lloyd et al., 2018], sub-national scores from the Human Development Index Smits and Permanyer [2019], GlobalDataLab [2017-2021] and indicators on slave exports and explorer contact being key indicators of social trust in Africa as described in the seminal paper of Nunn and Wantchekon [2011]. Table 2 shows a fictitious example of how node-based survey data is differenced to align it with the network structure of the SCI for the edge-based approach.

4 Methods

Predictor variables were mean-centered and scaled by the standard deviation prior to analysis (i.e. SCI, WorldPop, and Afrobarometer), except for those that were already proportions (i.e. DHS and Facebook penetration; see Table 6). Cluster-robust standard errors on the level of regions are used to account for regional-level dependencies.

4.1 Linking social media friendships with broader social trust measures

We first investigate whether there is a correspondence between measures of social capital, as collected through social trust measures through household survey instruments and the SCI based on social media friendship connections, which can be more readily measured across space and through time (cf. Figure 1). To do this, we ran a simple linear regression with country-level fixed effects using validated measures of social capital from the Afrobarometer project as predictors of Meta's SCI ($Mean\_SCI_i$) at each location i:

$$Mean\_SCI_i = \alpha + \beta_1 trust_i + \beta_2 participation_i + \sum_{m=1}^{M} (\theta_m x_{m,i}) + \epsilon_i \qquad (3)$$

where $Mean\_SCI_i$ is the SCI of a specific region calculated as the average of its edge-level SCIs, $trust_i$ is a measure of generalised social trust and $participation_i$ is a measure of social participation. The regression also includes a set of M control variables $x_{m,i}$ and a Gaussian residual error term ϵ_i. We control for a bandwidth of other potentially relevant factors related to socio-demographic characteristics, education, mobile penetration and Facebook penetration, among others. A full list of control variables and their description can be found in Table 6 of the Appendix.

4.2 Social connectedness and health-related behaviour

We implement an edge-based simple linear model with country-level fixed effects to analyse how social connectedness may influence health behaviours related to modern contraception and HIV. The edge-based approach utilises *pairwise differences* in health behaviours between locations as the response variable (cf. Table 2). We hypothesise that these gaps in health behaviours correlate with knowledge gaps between locations and that this relationship may be mediated by the degree of social connectedness. In a second step, we further investigate whether social connectedness not only affects behaviour formation directly, but also indirectly by facilitating knowledge diffusion across regions.

To determine the weight of evidence for these hypotheses using our observational data, we perform four linear regressions. The response variables for the first two regressions are the *differences in the use* between pairs of locations i and j ($\Delta use_{i,j}$) of modern contraceptive methods and HIV testing, respectively. Turning to knowledge diffusion, the third and fourth response variables are the *differences in knowledge* between pairs of locations i and j ($\Delta know_{i,j}$) of modern contraceptive methods and HIV transmission, respectively. All responses are roughly normally distributed and centred on zero. The regressions on health behaviour took the following form:

$$\Delta use_{i,j} = \alpha + \beta_1(SCI_{i,j}) + \beta_2(\Delta know_{i,j}) + \beta_3(SCI_{i,j} \times \Delta know_{i,j}) + \sum_{m=1}^{M}(\beta_m \Delta x_{m,i,j}) + \boldsymbol{\beta F} + \epsilon_{i,j} \quad (4)$$

where $SCI_{i,j}$ is the scaled SCI defined in Eq. 2 which we show in later analysis to be a good proxy for social capital. $\Delta know_{i,j}$ is the difference in knowledge of modern contraception or difference in knowledge of HIV transmission for the first and second regressions, respectively. $\Delta x_{m,i,j}$ is a set of M control variables describing differences between location i and j. $\boldsymbol{\beta F}$ are the country-level fixed effects for the country of location i and the country of location j, respectively, and $\epsilon_{i,j}$ is a Gaussian residual error term. We use country-level fixed effects here to account for potential spatial dependencies created by national health care systems and their impact on health outcomes in general. We expect the SCI to reduce the difference in behaviours between locations (i.e. β_1 < 0). Where a knowledge gap exists between locations, we also expects to see a behaviour gap (i.e. β_2 > 0). The interaction between the SCI and knowledge (β_3) is of particular interest, because a negative value would suggest that social connectedness facilitates a spill-over of health behaviours (i.e. reduce Δuse) even when a knowledge gap remains.

To further explore the underlying mechanism driving any spillovers of health behaviours via social connectedness, we analyse the effects of SCI on differences in health knowledge between locations.

$$\Delta know_{i,j} = \alpha + \beta_1(SCI_{i,j}) + \sum_{m=1}^{M}(\theta_m \Delta x_{m,i,j}) + \boldsymbol{\beta F} + \epsilon_{i,j} \quad (5)$$

If knowledge transfers were the underlying mechanism for spillovers of health behaviours among socially connected locations, then we would expect a significant negative effect of social connectedness on differences in knowledge (i.e. $\beta_1 < 0$). The lack of an effect would indicate that our data do not provide evidence that knowledge exchange is the mechanism driving spillovers of health behaviours through a socially-connected network.

5 Results

5.1 Relationship of social connectedness with social capital

In this section, we present the results of our analysis aiming to assess the validity of the social connectedness index as a measure of social capital by comparing it against survey-based measures of social trust, and exploring its socioeconomic and historical correlates. Unlike previous studies that have primarily relied on Facebook indicators and assumed them to be measures of social capital (e.g. Bailey et al. [2020], Chetty et al. [2022]), we enhance the analysis by incorporating validated measures of social capital from the Afrobarometer. Our regression model estimates the relationship between the SCI and various demographic and socioeconomic factors, including those on social capital, as detailed in Eq. 3. The results of this exercise are depicted in Table 3.

The raw correlation coefficient between trust and social connectedness is 0.328 (p-value=0.000), indicating a significant positive relationship. While this correlation may appear moderate in magnitude, it is important to consider the 10-year time span between the measurement of social trust in 2011-2013 in the Afrobarometer survey and the Social Connectedness Index (SCI) in 2021. Moreover, social capital is a multidimensional concept, of which social trust is one component [Chetty et al., 2022, Portes, 2000]. Despite this time gap, the consistent and positive correlation between trust in the survey and the social connectedness index, as demonstrated in Table 3, reinforces the robustness of the relationship and highlights how the SCI is capturing broader social capital dynamics. Social trust consistently emerges across models 1–7 as a positive and statistically significant correlate of the SCI, even when controlling for a range of other socio-economic variables, underscoring its pivotal role in fostering social capital and strengthening social networks. It is worth noting that reverse causation is unlikely to occur in this analysis since social trust was measured before the SCI, further supporting the argument that the SCI is a reliable proxy for social capital.

Additionally, the inclusion of historical variables, such as the historical prevalence of slavery export and exposure to explorers, provides valuable insights into the long-lasting impact of historical events on contemporary social connectedness. As expected, regions with a higher historical prevalence of slavery export show a negative association with social connectedness, reflecting the enduring consequences of this historical legacy [Nunn and Wantchekon, 2011]. Conversely, exposure to explorers is positively associated with social connectedness, indicating the potential influence of historical exploration and cultural exchange on contemporary social capital, similar to the results shown in Enke [2023]. These consistent directions of the historical variables add to the robustness and credibility of the SCI as a proxy of social capital in this context.

5.2 The role of social connectedness in changing health behaviours

Now that we have shown that the SCI is a good proxy of social capital in general and of generalized social trust specifically, we further investigate the role of social connectedness in shaping health-related behaviour. For this study, we focus on the use of modern contraceptive methods and the use of HIV tests as key pillars of sexual and reproductive health. Consequently, we ask: Can social connectedness - as a proxy of social capital - shape health-related behaviour? And if yes, through which channels? Does it help spread knowledge which in turn shapes behavioural change or does it influence behaviour directly? To shed light on these questions, we exploit the network structure of Meta's SCI as described in Section 3. Table 4 summarizes the main results of this study.

As expected and in line with existent literature, knowledge is a major determinant of usage behaviour as shown across regression results indicating that knowledge differences have a positive and statistically significant effect on the differences in respective uses (see columns 1 and 2 in Table 4). By looking at the SCI across outcomes, a direct mediating effect of social connectedness becomes evident: socially better-connected regions show smaller differences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Regional SCI (averaged across ties)				
	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Trust: General (2011-2013)	1.807***	1.830***	1.441***	1.406***	1.406***	1.383***	1.674***
	(0.373)	(0.370)	(0.358)	(0.512)	(0.512)	(0.342)	(0.498)
Volunteer/Community members		0.556*	0.611**	0.855***	0.855***	0.930***	1.966***
		(0.315)	(0.289)	(0.307)	(0.307)	(0.299)	(0.597)
Wealth index: poorest			-0.086				
			(0.486)				
Wealth index: poorer			0.627				
			(0.811)				
Educ.: secondary or higher			-3.500**				
			(1.448)				
Mobile penetration rate			-2.884***	-1.190**	-1.190**	-2.039***	-3.248***
			(0.548)	(0.464)	(0.464)	(0.496)	(0.675)
Living in rural areas (in %)			-0.645**	-0.194	-0.194	-0.051	-0.908
			(0.299)	(0.252)	(0.252)	(0.281)	(0.610)
Age group 15-19 (in %)			-1.169	-2.507	-2.507	-2.050	-1.485
			(4.752)	(4.357)	(4.357)	(4.291)	(6.837)
Age group 20-24 (in %)			18.200***	28.331***	28.331***	23.760***	3.242
			(6.379)	(6.531)	(6.531)	(6.418)	(12.534)
Age group 25-29 (in %)			-14.594*	-24.367**	-24.367**	-17.152**	-5.057
			(7.663)	(10.227)	(10.227)	(8.029)	(14.514)
Age group 30-34 (in %)			-9.317	-13.543	-13.543	-9.925	-13.367
			(7.287)	(8.670)	(8.670)	(7.387)	(13.275)
Age group 35-39 (in %)			29.072**	18.564**	18.564**	18.629*	22.589
			(11.375)	(9.300)	(9.300)	(9.896)	(17.386)
Age group 40-44 (in %)			-2.587	25.408***	25.408***	14.859	8.673
			(8.944)	(9.274)	(9.274)	(9.709)	(18.448)
Age group 45-49 (in %)			9.601	3.776	3.776	4.568	3.620
			(6.608)	(6.331)	(6.331)	(6.579)	(14.369)
Facebook penetration rate			-0.335	-0.493*	-0.493*	-0.659**	-0.697
			(0.261)	(0.263)	(0.263)	(0.267)	(0.506)
HDI				-1.822***	-1.822***		
				(0.503)	(0.503)		
Night Lights						-0.095***	0.044
						(0.033)	(0.290)
Distance to major rd						-0.302	-1.097**
						(0.252)	(0.447)
Distance to inland water						0.325	1.050*
						(0.242)	(0.575)
Built settlement growth						-0.193***	-0.283*
						(0.051)	(0.163)
Local Slave Export (Log)							-0.503***
							(0.147)
District Ethnic Fractionalization							-0.539
							(0.544)
Explorer contact							0.816***
							(0.240)
Railway contact							-0.510**
							(0.240)
adj. R^2	0.108	0.120	0.433	0.359	0.359	0.386	0.717
N	249.	249	231	200	200	249	99

Note: Probit/OLS. SE clustered at the regional level in parentheses. Country FE included. * p<0.10, ** p<0.05, *** p<0.01

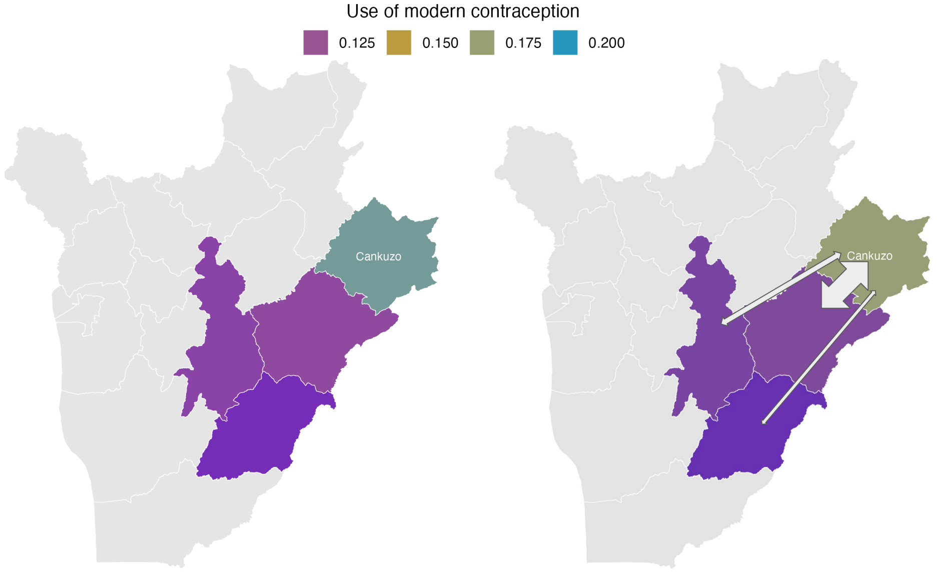
Table 3: **Relationship between social connectedness index (SCI) and social capital measures from Afrobarometer.**

	Dependent variable: Positive regional differences in...			
	...use of...		...knowledge about...	
	modern contraception	HIV tests	modern contraception	HIV
	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Constant	0.035***	0.037*	0.140***	0.062**
	(0.011)	(0.019)	(0.009)	(0.027)
SCI	−0.260***	−0.131***	−0.360***	−0.165***
	(0.040)	(0.027)	(0.050)	(0.032)
ΔContraceptive knowledge	0.217***			
	(0.011)			
SCI x ΔContraceptive knowledge	−14.395***			
	(5.257)			
ΔKnowledge about HIV		0.548***		
		(0.027)		
SCI x ΔKnowledge about HIV		−25.121***		
		(8.307)		
ΔControl variables (20)	Yes	Yes	Yes	Yes
Observations	122,760	122,760	122,760	122,760
R^2	0.726	0.932	0.697	0.932
Adjusted R^2	0.726	0.932	0.697	0.932

Note: OLS. SE clustered at the regional level. Country FE included. *p<0.1; **p<0.05; ***p<0.01

Table 4: **Effects of social connectedness and knowledge gaps on health behaviour.**

in both the use of (i.e. columns 1 and 2) and the knowledge about (i.e. columns 3 and 4) modern contraception and HIV between pairs of regions ($\beta_1 < 0$). Interestingly, we observe that social connectedness also helps to overcome knowledge gaps in determining differences in use by filling the void with social trust as demonstrated by the negative and significant interaction effect between the SCI and the regional differences in knowledge variables ($\beta_3 < 0$). In other words, the mediating role of social connectedness becomes stronger, the larger the knowledge gaps between regions. This also holds true when looking at the negative outcomes of a pairwise connection ($\Delta <= 0$, cf. Section 3) as shown in Table 8 in the Appendix. All effects remain almost identical, except for the SCI and the country-fixed effects, which both see a change in signs, but not in effect sizes as both effects are direction-invariant (cf. Table 2). Small changes in the effect sizes are due to the inclusion of cases where $\Delta = 0$. Using only non-zero differences in health outcomes would yield identical absolute effect sizes for either direction. Figure 2 gives an example of the SCI's mediating effect for the region of Cankuzo, Burundi, where both use of and knowledge levels about modern contraception are higher than in the three regions it is most socially connected to.



(a) Without social connectedness

(b) With social connectedness

Figure 2: **The effect of being socially connected.** Both Figure 2b and 2a map the predicted and actual use of modern contraception rates in Cankuzo, Burundi and the three regions Cankuzo is most socially connected to, respectively. In Figure 2b, the predicted use of modern contraception rate in Cankuzo takes social connectedness as measured per the SCI into account. The thickness of the white arrows indicates the strengths of social connectedness with the other three regions. Figure 2a shows the hypothetical setting of no social connectedness of Cankuzo, keeping all other things equal.

In Figure 2, we predict differences in modern contraceptive use for Cankuzo, Burundi for two distinct scenarios: (a) assuming no social connectedness, i.e. $\beta_1 = \beta_3 = 0$, and (b) accounting for social connectedness as measured per SCI. We then estimate the use level of Cankuzo for those two scenarios by adding the estimated differences in use to the actual use levels in the three others regions and averaging them. We observe that Cankuzo loses from being socially connected as its predicted use of modern contraception rate (17.6%) is below the rate expected in a setting without social connectedness (18.7%), but still higher than the average use across its three strongest ties (11.7%), assuming all other things equal. This nicely demonstrates the mediating effect of social connectedness: regions with comparatively higher use levels vis-à-vis its strongest ties lose, while regions with comparatively lower use levels benefit and the effect is stronger the larger the knowledge gap is. Since we look at those edges with positive differences in use only (cf. Section 3), the overall effect of the SCI on use levels is negative as shown in Table 5.

The regional-level effects are calculated as the average of the estimated tie-specific effects of the SCI and its interaction effect with knowledge differences in percentage points defined as $\widehat{\Delta use}_i = \frac{1}{J}\sum_{j=1}^{J} \hat{\beta}_1(SCI_{i,j}) \times 100$ and $\widehat{\Delta use}_i = \frac{1}{J}\sum_{j=1}^{J} \hat{\beta}_3(SCI_{i,j} \times \Delta know_{i,j}) \times 100$, respectively, where J is the total number of regions in our sample. We see that the median effect of social connectedness on health behaviour is small. In comparison, knowledge gaps increase differences in health behaviour on average by 1.7%-points and 12.5%-points for modern contraception and HIV testing, respectively.

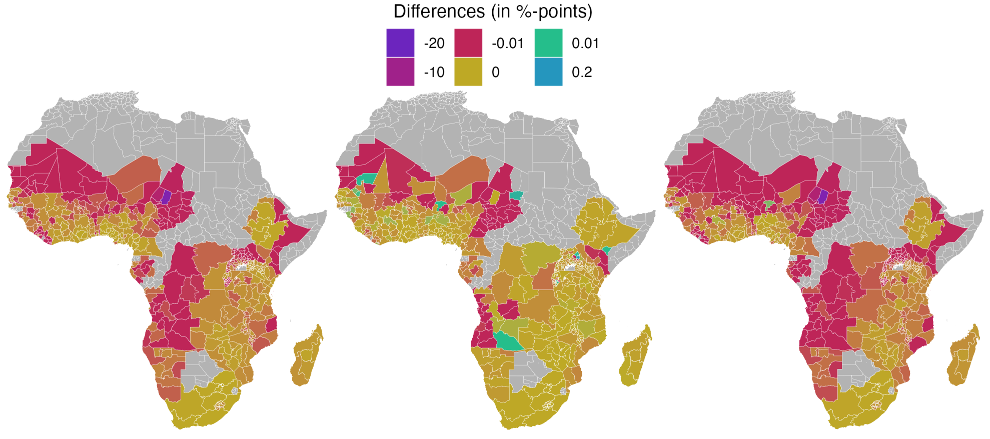
One potential reason why the effect of social connectedness on health-related indicators is surprisingly low is that

Direction of effect	Indicator	Use of modern contraception			Use of HIV testing		
		Main	Interaction	Total	Main	Interaction	Total
+	% of regions	0	12.1	1.0	0	14.1	7.1
+	effect size (in %-points)	-	0.001	0.000	-	0.002	0.002
−	% of regions	100	87.9	99.0	100	85.9	92.9
−	effect size (in %-points)	−0.004	−0.001	−0.006	−0.002	−0.005	−0.009
Overall effect size (in %-points)		−0.004	−0.001	−0.006	−0.002	−0.004	−0.008

Table 5: **Share of regions and median effect sizes of the social connectedness and its interaction effect with knowledge gaps on differences in health behaviour, by direction of effect.**

social networks are predominantly locally anchored and that the administrative level used in our analysis (i.e. regions) only captures geographic variation across the longest-ranging ties of a social network. This is supported by the fact that the SCI of self-connections is on average 643 times higher than the SCI to other regions, which means that by far the largest proportions of friendship links remain within the same region. In order to investigate how more granular geographic disaggregation affects the SCI's impact, we repeat the analysis for the country with the smallest average area size per region in our sample, namely Burundi (1591km^2 vis-à-vis 51931km^2 across the remaining countries). We observe that the SCI of self-connections over connections to other regions reduces to 456 and the overall SCI effect becomes stronger (−0.01 vis-à-vis −0.006, cf. Table 5), further supporting that argument we capture just a fraction of the overall SCI effect through the longest-ranging ties. Thus, we argue that our estimated effects represent a lower bound of the impact of social connectedness on health behaviour and expect the true effect to be significantly stronger.

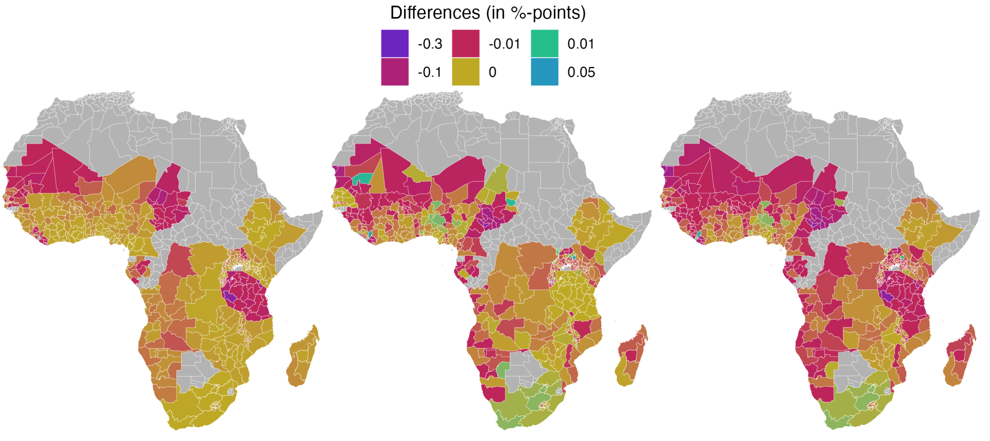
Figures 3 and 4 visualize the SCI-related effects on the regional level.



(a) Main effect (b) Interaction effect (c) Joint effect

Figure 3: **Regional-level effects of social connectedness on the use of modern contraception.** The effects of social connectedness in Figures 3a - 3c represent the median SCI per region across its respective ties for the 495 regions in the study sample.

Table 9 in the Appendix shows the results of our main analysis when excluding the extreme outlier region from our sample.[1] The resulting changes in coefficient sizes are negligible and neither change the direction nor the significance of the observed effects.



(a) Main effect (b) Interaction effect (c) Joint effect

Figure 4: **Regional-level effects of social connectedness on the use of HIV testing.** The effects of social connectedness in Figures 4a - 4c represent the median SCI per region across its respective ties for the 495 regions in the study sample.

In addition, by looking at Table 16 in the Appendix, we observe that the main coefficient of the SCI more than doubles for ties with a Facebook penetration rate above the median, as does the interaction effect between SCI and differences in knowledge. In other words, the mediating effect of social connectedness is stronger in regions with high Facebook penetration rates. In areas with higher Facebook penetration, we also expect Facebook ties to reflect social ties among a broader range of the population, rather than capturing only selective users. We further consider this as indication that Facebook is not only a good proxy for the cognitive component of social capital, i.e. trust, but also partly provides a medium for its structural component by facilitating the forging of a network between different communities.

Further, as we show in Tables 10 - 13 in the Appendix, the direction and significance of our effects of interest are robust across different model specifications. The same holds true for re-running the analysis on female- and male-specific DHS data, respectively (see Table 14 and Table 15) in the Appendix, thus further underscoring the robustness of our approach.

6 Conclusion

While a large body of research across the social sciences argues for the importance of social capital for shaping health outcomes, existing research has often faced difficulties in operationalizing social capital at scale, especially in low-

[1]As shown in Figure 3, Bahr el Gazel in Chad acts as the most notable outlier with an overall regional-level effect of -17.1%-points (cf. Table 5). With an SCI of 0.65, Bahr el Gazel has the highest SCI value among all ties in the sample and the fourth highest SCI value globally. This might be due to the fact that both the estimated Facebook penetration (0.2%) and the estimated population count of about 300,000 is comparatively low, leading to an estimated Facebook user count of 600 in Bahr el Gazel. This hints at a very well-connected few that use Facebook in this region.

and middle-income country contexts. Through the integration of Demographic and Health Survey data with a novel measure of social capital, as proxied by social connectedness of regions through Facebook friendship links between them, we provide insights into how social connectedness shapes the diffusion of health knowledge and behaviours. These findings provide empirical evidence to a large theoretical literature on social capital and its impacts. They further underscore crucial implications for the structuring of health information campaigns. Firstly, our findings show the profound influence of knowledge on health behaviours linked to the use of modern contraception and HIV testing. However, we also show that the effectiveness of an information initiative in a specific region isn't solely anchored in its inherent knowledge base; it is also intricately linked to its social connections with other regions and their respective health behaviours. In navigating these complex interrelationships, Meta's Social Connectedness Index, which taps into Facebook friendships as a representation of social capital, emerges as a valuable measure. It can pinpoint regions outside the primary focus area that might significantly sway the outcome of the campaign.

Furthermore, our findings indicate that information campaigns in regions strongly connected to areas with pronounced disparities in health knowledge might not be as effective as those in areas linked to regions with more similar knowledge outcomes. The Social Connectedness Index (SCI) then emerges as a pivotal tool, shedding light on the trajectory and intensity of potential behavioral ripple effects in sexual and reproductive health campaigns. Harnessing the SCI allows policymakers and health experts to delve deeper into the social intricacies influencing behavior, thus equipping them with the insights needed to craft more precise and potent health strategies.

Nevertheless, this study comes with limitations: Firstly, as mentioned before, the geographical granularity of the SCI in Africa as currently provided by Meta misses out on the majority of spatial variation. While for both the US and most of Europe finer geographical resolutions are available, the rest of the world including Africa falls short of the opportunity to leverage the full potential of an alternative measure of social capital, given this coarser geographical resolution. Secondly, working with proprietary datasets, derived from social media such as Meta's SCI, introduces several potential biases that need to be carefully considered. For instance, these data may structurally underrepresent certain segments of the population, such as women, children, elderly individuals, and the very poor, who may have limited access to digital platforms. Consequently, any population-level inferences drawn from such data should be interpreted with caution. Nevertheless, we implement several analyses to test the robustness of the SCI by comparing against survey-validated measures, and also examine the sensitivity of our results to the levels of Facebook penetration within a region, to more deeply examine the validity of the measure and its impacts.

Moreover, when examining the relationship between node-level SCI averages and the Facebook penetration rate in specific regions, we observe a non-linear pattern. Regions with low levels of Facebook penetration, typically below 10%, appear to exhibit higher levels of social connectedness than other regions. This suggests that the subset of Facebook users in these regions is a homophilic and well-connected group, which may not be entirely representative of the overall population. This phenomenon, where areas with low coverage can reflect a selected or distinctive user base, has also been noted in other analyses of online platforms, like LinkedIn [Kashyap and Verkroost, 2021] or Google+ [Magno and Weber, 2014]. However, it is important to note that this bias primarily affects grouped data, such as regional SCI averages, and is not evident at the individual or edge level. This observation aligns with Simpson's paradox, which hints at a potential confounding factor in the underlying data. To address this issue, we conduct a thorough analysis to ensure that we appropriately account for the non-linear relationship between SCI and Facebook penetration rate. We test this by examining the correlation between the regional Facebook penetration rate and node-level residuals derived from regressing the edge-level SCI on a set of control variables used in our preceding analysis. We find that the node-level correlation largely disappears, providing us with confidence that we have adequately captured the non-linear relationship between SCI and Facebook penetration rate in our analysis. Despite these challenges, and through these additional checks, we believe that leveraging the SCI data offers valuable insights, and our rigorous approach allows us to overcome potential biases effectively.

In sum, this research underlines the pivotal influence of social connectedness in determining the efficacy of public health initiatives, especially concerning sexual and reproductive health in Sub-Saharan Africa. The insights here show how aggregate data from social media on social connectedness can help tap into the intricate web of social capital dynamics, thereby enabling health professionals and policymakers to develop nuanced strategies that can bolster sexual and reproductive health outcomes in Sub-Saharan Africa. It is, however, indispensable to recognize that these strides hinge on the provisos of data integrity and accessibility. Navigating and rectifying the constraints tethered to non-standard, proprietary data reservoirs, like Facebook's Social Connectedness Index, is thus paramount to enable these insights to be applied to the development of health campaigns.

References

Cecilia Åslund, Bengt Starrin, and Kent W Nilsson. Social capital in relation to depression, musculoskeletal pain, and psychosomatic symptoms: a cross-sectional study of a large population-based cohort of swedish adolescents. BMC Public Health, 10(1):715, 2010.

Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Social connectedness: Measurement, determinants, and effects. Journal of Economic Perspectives, 32(3):259–80, 2018.

Michael Bailey, Drew Johnston, Theresa Kuchler, Dominic Russel, Bogdan State, and Johannes Stroebel. The determinants of social connectedness in europe. In Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12, pages 1–14. Springer, 2020.

Michael Bailey, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. International trade and social connectedness. Journal of International Economics, 129:103418, 2021.

Jere R Behrman, Hans-Peter Kohler, and Susan Cotts Watkins. Social networks and changes in contraceptive use over time: Evidence from a longitudinal study in rural kenya. Demography, 39(4):713–738, 2002.

Ariel BenYishay, Renee Rotberg, Jessica Wells, Zhonghui Lv, Seth Goodman, Lidia Kovacevic, and Dan Runfola. Geocoding afrobarometer rounds 1-6: Methodology & data quality. AidData. Available online at http://geo. aiddata. org, 2017.

John Bongaarts and Susan Cotts Watkins. Social interactions and contemporary fertility transitions. Population and Development Review, pages 639–682, 1996.

Pierre Bourdieu. The Forms of Capital. New York: Greenwood Press, 1986.

Ronald S Burt. The network structure of social capital. Research in organizational behavior, 22:345–423, 2000.

John T. Cacioppo and Stephanie Cacioppo. Social relationships and health: The toxic effects of perceived social isolation. Social and Personality Psychology Compass, 8(2):58–72, 2014. ISSN 1751-9004. doi: 10.1111/spc3.12087. URL http://dx.doi.org/10.1111/spc3.12087. SPCO-0623.R2.

Ben Charoenwong, Alan Kwan, and Vesa Pursiainen. Social connections with covid-19–affected areas increase compliance with mobility restrictions. Science advances, 6(47):eabc3054, 2020.

Raj Chetty, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. Social capital ii: determinants of economic connectedness. Nature, 608(7921):122–134, 2022.

John Cleland and Christopher Wilson. Demand theories of the fertility transition: An iconoclastic view. Population studies, 41(1):5–30, 1987.

John Cleland, Stan Bernstein, Alex Ezeh, Anibal Faundes, Anna Glasier, and Jolene Innis. Family planning: the unfinished agenda. The lancet, 368(9549):1810–1827, 2006.

Steven W. Cole, John P. Capitanio, Katie Chun, Jesusa M. G. Arevalo, Jeffrey Ma, and John T. Cacioppo. Myeloid differentiation architecture of leukocyte transcriptome dynamics in perceived social isolation. Proceedings of the National Academy of Sciences, 112(49):15142–15147, 2015. doi: 10.1073/pnas.1514249112.

James S Coleman. Foundations of social theory. Harvard university press, 1994.

Benjamin Enke. Market exposure and human morality. Nature Human Behaviour, 7(1):134–141, 2023.

Barbara Entwisle, Ronald R Rindfuss, David K Guilkey, Aphichat Chamratrithirong, Sara R Curran, and Yothin Sawangdee. Community and contraceptive choice in rural thailand: A case study of nang rong. Demography, 33(1):1–11, 1996.

Damiano Fiorillo and Fabio Sabatini. Quality and quantity: the role of social interactions in self-reported individual health. Social Science & Medicine, 73(11):1644–1652, 2011.

Damiano Fiorillo and Fabio Sabatini. Structural social capital and health in italy. Economics & Human Biology, 17: 129–142, 2015.

Virginia A Fonner, Deanna Kerrigan, Zandile Mnisi, Sosthenes Ketende, Caitlin E Kennedy, and Stefan Baral. Social cohesion, social participation, and hiv related risk among female sex workers in swaziland. PloS one, 9(1):e87527, 2014.

Francis Fukuyama. Trust: The social virtues and the creation of prosperity. World and I, 10:264–268, 1995.

GADM. Global administrative areas (gadm), 2015. data retrieved from GADM database, `https://gadm.org/download_country_v2.html`.

GlobalDataLab. Human development index, 2017-2021. data retrieved from the Subnational HDI Database of the Global Data Lab, `https://globaldatalab.org/shdi/`.

Elise Grover, Ashley Grosso, Sosthenes Ketende, Caitlin Kennedy, Virginia Fonner, Darrin Adams, Bhekie Sithole, Zandile Mnisi, Sibusiso Lulu Maziya, and Stefan Baral. Social cohesion, social participation and hiv testing among men who have sex with men in swaziland. AIDS care, 28(6):795–804, 2016.

Zhongchen Hu. Social interactions and households' flood insurance decisions. Journal of Financial Economics, 144 (2):414–432, 2022.

Eaman Jahani, Samuel P Fraiberger, Michael Bailey, and Dean Eckles. Long ties, disruptive life events, and economic prosperity. Proceedings of the National Academy of Sciences, 120(28):e2211062120, 2023.

Albino Kalolo, Jacob Mazalale, Anja Krumeich, and Michelene Chenault. Social cohesion, social trust, social participation and sexual behaviors of adolescents in rural tanzania. BMC public health, 19(1):1–9, 2019.

Ridhi Kashyap. Has demography witnessed a data revolution? promises and pitfalls of a changing data ecosystem. Population Studies, 75(sup1):47–75, 2021.

Ridhi Kashyap and Florianne CJ Verkroost. Analysing global professional gender gaps using linkedin advertising data. EPJ Data Science, 10(1):39, 2021.

Ridhi Kashyap, Masoomali Fatehkia, Reham Al Tamime, and Ingmar Weber. Monitoring global digital gender inequality using the online populations of facebook and google. Demographic Research, 43:779–816, 2020.

Hans-Peter Kohler, Jere R Behrman, and Susan C Watkins. The density of social networks and fertility decisions: Evidence from south nyanza district, kenya. Demography, 38(1):43–58, 2001.

Nan Lin, Karen S Cook, and Ronald S Burt. Social capital: Theory and research. Transaction Publishers, 2001.

Christopher Lloyd, Maksym Bondarenko, et al. Global 100m covariates. 2018.

Ye Luo, Louise C. Hawkley, Linda J. Waite, and John T. Cacioppo. Loneliness, health, and mortality in old age: A national longitudinal study. Social Science & Medicine, 74(6):907 – 914, 2012. doi: http://dx.doi.org/10.1016/j.socscimed.2011.11.028.

Gabriel Magno and Ingmar Weber. International gender differences and gaps in online social networks. In Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings 6, pages 121–138. Springer, 2014.

Mark R Montgomery and John B Casterline. Social learning, social influence, and new models of fertility. Population and development review, 22:151–175, 1996.

Spencer Moore, Mark Daniel, Lise Gauvin, and Laurette Dubé. Not all social capital is good capital. Health & place, 15(4):1071–1077, 2009.

Nathan Nunn and Leonard Wantchekon. The slave trade and the origins of mistrust in africa. American economic review, 101(7):3221–3252, 2011.

Alejandro Portes. Social capital: Its origins and applications in modern sociology. LESSER, Eric L. Knowledge and Social Capital. Boston: Butterworth-Heinemann, pages 43–67, 2000.

Robert D Putnam. Bowling alone: The collapse and revival of American community. Simon and Schuster, 2001.

Yusuf Ransome, Ashley Batson, Sandro Galea, Ichiro Kawachi, Denis Nash, and Kenneth H Mayer. The relationship between higher social trust and lower late hiv diagnosis and mortality differs by race/ethnicity: results from a state-level analysis. Journal of the International AIDS Society, 20(1):21442, 2017.

Morris Rosenberg. Misanthropy and political ideology. American sociological review, pages 690–695, 1956.

Fabio Sabatini. The relationship between trust and networks. an exploratory empirical analysis. Economics Bulletin, 29(2):661–672, 2009.

Timo Schmid, Fabian Bruckschen, Nicola Salvati, and Till Zbiranski. Constructing Sociodemographic Indicators for National Statistical Institutes by Using Mobile Phone Data: Estimating Literacy Rates in Senegal. Journal of the Royal Statistical Society Series A: Statistics in Society, 180(4):1163–1190, 08 2017. ISSN 0964-1998. doi: 10.1111/rssa.12305. URL https://doi.org/10.1111/rssa.12305.

Jeroen Smits and Iñaki Permanyer. The subnational human development database. Scientific data, 6(1):1–15, 2019.

Kimiko Tomioka, Norio Kurumatani, and Hiroshi Hosoi. Association between social participation and 3-year change in instrumental activities of daily living in community-dwelling elderly adults. Journal of the American Geriatrics Society, pages n/a–n/a, 2016. ISSN 1532-5415. doi: 10.1111/jgs.14447. URL http://dx.doi.org/10.1111/jgs.14447.

Behzad Vahedi, Morteza Karimzadeh, and Hamidreza Zoraghein. Spatiotemporal prediction of covid-19 cases using inter-and intra-county proxies of human interactions. Nature communications, 12(1):6440, 2021.

Michael Woolcock. The place of social capital in understanding social and economic outcomes. Canadian journal of policy research, 2(1):11–17, 2001.

Appendix

List of control variables

List of DHS survey used

Effects of social connectedness and knowledge gaps on health behaviour, for negative differences in health outcomes

Regression results with Bahr el Gazel

Robustness of model specification

Health outcomes by gender

Health outcomes by Facebook penetration rate

Indicator name	Description	Source
	Basic controls	
Wealth index: poorest	Share of respondents classified as poorest by DHS wealth index quintile	DHS
Wealth index: poorer	Share of respondents classified poorer by DHS wealth index quintile	DHS
Wealth index: richer	Share of respondents classified richer by DHS wealth index quintile	DHS
Wealth index: richest	Share of respondents classified richest by DHS wealth index quintile	DHS
male	Share of male respondents in the DHS	DHS
Educ.: secondary or higher	Share of respondents that have completed at least secondary education	DHS
Mobile penetration rate	Share of respondents that own a mobile phone	DHS
Living in rural areas (in %)	Share of respondents living in rural areas	DHS
Age group 15-19 (in %)	Share of respondents in the respective age group	DHS
Age group 20-24 (in %)	Share of respondents in the respective age group	DHS
Age group 25-29 (in %)	Share of respondents in the respective age group	DHS
Age group 30-34 (in %)	Share of respondents in the respective age group	DHS
Age group 35-39 (in %)	Share of respondents in the respective age group	DHS
Age group 40-44 (in %)	Share of respondents in the respective age group	DHS
Age group 45-49 (in %)	Share of respondents in the respective age group	DHS
	Additional controls	
Nights lights	Mean night-time light intensity (aggregated from grid)	WorldPop
Distance to major rd	Mean distance to major road (aggregated from grid)	WorldPop
Built settlement growth	Average space used by buildings (aggregated from grid)	WorldPop
FB_pntr_15to49_all	Facebook penetration rate	Kashyap et al. [2020]
FB_pntr_15to49_all:SCI	Interaction term with SCI to account for non-linear relationship	Kashyap et al. [2020]
	Other controls	
HDI	Sub-national HDI score (v7.0)	Global Data Lab
Distance to inland water	Mean distance to inland water (aggregated from grid)	WorldPop
	Historic controls	
Local Slave Export (Log)	Log of the number of slaves taken from a location (normalized by land area)	Nunn and Wantchekon [2011]
District Ethnic Fractionalization	Measure of ethnic heterogeneity within a region	Nunn and Wantchekon [2011]
Explorer contact	Dummy whether a European explorer traveled through land previously occupied by the ethnic group	Nunn and Wantchekon [2011]
Railway contract	Region historically linked to colonial railway networks	Nunn and Wantchekon [2011]

Table 6: **List of control variables used in this study.**

Country	Year	Country	Year
Angola	2015-16	Mauritania	2019-21
Burkina Faso	2010	Malawi	2015-16
Benin	2017-18	Mozambique	2011
Burundi	2016-17	Nigeria	2018
Congo (the Democratic Republic of the)	2013-14	Niger	2012
Côte d'Ivoire	2011-12	Namibia	2013
Cameroon	2018	Rwanda	2019-20
Ethiopia	2016	Sierra Leone	2019
Gabon	2012	Senegal	2019
Ghana	2014	Chad	2014-15
Gambia (the)	2019-20	Togo	2013-14
Guinea	2018	Tanzania, United Republic of	2015-16
Kenya	2014	Uganda	2016
Lesotho	2014	South Africa	2016
Liberia	2019-20	Zambia	2018
Madagascar	2021	Zimbabwe	2015
Mali	2018		

Table 7: **List of DHS surveys used for this study.**

	Dependent variable: Negative regional differences in...			
	...use of...		...knowledge about...	
	modern contraception	HIV tests	modern contraception	HIV
	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Constant	−0.035***	−0.037**	−0.138***	−0.062***
	(0.012)	(0.015)	(0.029)	(0.019)
SCI	0.292***	0.149***	0.353***	0.184***
	(0.037)	(0.027)	(0.059)	(0.027)
ΔContraceptive knowledge	0.220***			
	(0.031)			
SCI x ΔContraceptive knowledge	−14.781***			
	(5.172)			
ΔKnowledge about HIV		0.549***		
		(0.043)		
SCI x ΔKnowledge about HIV		−24.970***		
		(9.152)		
ΔControl variables (20)	Yes	Yes	Yes	Yes
Observations	122,760	122,760	122,760	122,760
R^2	0.726	0.932	0.697	0.932
Adjusted R^2	0.726	0.932	0.697	0.932

Note: OLS. SE clustered at the regional level. Country FE included. *p<0.1; **p<0.05; ***p<0.01

Table 8: **Effects of social connectedness and knowledge gaps on health behaviour for negative differences in health outcomes ($\Delta <= 0$).**

	Dependent variable: Positive regional differences in...			
	...use of...		...knowledge about...	
	modern contraception	HIV tests	modern contraception	HIV
	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Constant	0.035***	0.037*	0.138***	0.062**
	(0.011)	(0.019)	(0.008)	(0.026)
SCI	−0.292***	−0.147***	−0.353***	−0.183***
	(0.035)	(0.026)	(0.053)	(0.028)
ΔContraceptive knowledge	0.220***			
	(0.011)			
SCI x ΔContraceptive knowledge	−14.781***			
	(5.461)			
ΔKnowledge about HIV		0.549***		
		(0.027)		
SCI x ΔKnowledge about HIV		−25.064***		
		(8.377)		
ΔControl variables (20)	Yes	Yes	Yes	Yes
Observations	122,760	122,760	122,760	122,760
R^2	0.726	0.932	0.697	0.932
Adjusted R^2	0.726	0.932	0.697	0.932

Note: OLS. SE clustered at the regional level. Country FE included. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 9: **Effects of social connectedness and knowledge gaps on health behaviour with the region Bahr el Gazel in Chad.**

	Dependent variable: Differences in the use of...		
	...modern contraception		
	(1)	(2)	(3)
Constant	0.125***	0.121***	0.035***
	(0.004)	(0.004)	(0.011)
SCI	−0.930***	−0.889***	−0.260***
	(0.108)	(0.102)	(0.040)
ΔContraceptive knowledge	0.458***	0.470***	0.278***
	(0.023)	(0.025)	(0.031)
SCI x ΔContraceptive knowledge	−86.464***	−72.339***	−14.395***
	(17.989)	(16.791)	(5.257)
ΔControl variables	Basic	Basic + Additional	Basic + Additional
Country-level fixed effects	No	No	Yes
Observations	122,760	122,760	122,760
R^2	0.265	0.292	0.726
Adjusted R^2	0.265	0.292	0.726

Note: OLS. SE clustered at the regional-level. *p<0.1; **p<0.05; ***p<0.01

Table 10: **Model robustness across specifications: Use of modern contraception.**

	Dependent variable: Differences in the use of...		
	...HIV tests		
	(1)	(2)	(3)
Constant	0.243***	0.235***	0.037*
	(0.011)	(0.010)	(0.019)
SCI	−1.839***	−1.762***	−0.131***
	(0.227)	(0.213)	(0.027)
ΔKnowledge about HIV	0.353***	0.350***	0.548***
	(0.034)	(0.030)	(0.027)
SCI x ΔKnowledge about HIV	−168.521***	−121.411***	−25.121***
	(39.414)	(33.871)	(8.307)
ΔControl variables	Basic	Basic + Additional	Basic + Additional
Country-level fixed effects	No	No	Yes
Observations	122,760	122,760	122,760
R^2	0.351	0.379	0.932
Adjusted R^2	0.350	0.379	0.932

Note: OLS. SE clustered at the regional-level. *p<0.1; **p<0.05; ***p<0.01

Table 11: **Model robustness across specifications: Use of HIV tests.**

	Dependent variable: Differences in the knowledge about...		
	...modern contraception		
	(1)	(2)	(3)
Constant	0.069***	0.061***	0.140***
	(0.0004)	(0.0004)	(0.002)
SCI	−0.481***	−0.423***	−0.360***
	(0.062)	(0.058)	(0.037)
ΔControl variables	Basic	Basic + Additional	Basic + Additional
Country-level fixed effects	No	No	Yes
Observations	122,760	122,760	122,760
R^2	0.120	0.238	0.697
Adjusted R^2	0.120	0.238	0.697

Note: OLS. SE clustered at the regional-level. *p<0.1; **p<0.05; ***p<0.01

Table 12: **Model robustness across specifications: Knowledge about modern contraception.**

	Dependent variable: Differences in the knowledge about...		
	...HIV transmission		
	(1)	(2)	(3)
Constant	0.248***	0.242***	0.062**
	(0.004)	(0.004)	(0.027)
SCI	−1.921***	−1.842***	−0.165***
	(0.229)	(0.218)	(0.032)
ΔControl variables	Basic	Basic + Additional	Basic + Additional
Country-level fixed effects	No	No	Yes
Observations	122,760	122,760	122,760
R^2	0.193	0.223	0.932
Adjusted R^2	0.193	0.223	0.932

Note: OLS. SE clustered at the regional-level. *p<0.1; **p<0.05; ***p<0.01

Table 13: **Model robustness across specifications: Knowledge about HIV transmission.**

	Dependent variable: Regional differences in...			
	...use of...		...knowledge about...	
	modern contraception	HIV tests	modern contraception	HIV
	(1)	(2)	(3)	(4)
Constant	0.035***	0.044**	0.191***	0.077**
	(0.010)	(0.021)	(0.011)	(0.031)
SCI	−0.331***	−0.112***	−0.424***	−0.196***
	(0.047)	(0.025)	(0.056)	(0.037)
$\triangle$Contraceptive knowledge	0.199***			
	(0.010)			
SCI x $\triangle$Contraceptive knowledge	−16.353***			
	(4.754)			
$\triangle$Knowledge about HIV		0.541***		
		(0.025)		
SCI x $\triangle$Knowledge about HIV		−16.567*		
		(8.526)		
$\triangle$Control variables (18)	Yes	Yes	Yes	Yes
Observations	122,760	122,760	122,760	122,760
R^2	0.692	0.934	0.667	0.918
Adjusted R^2	0.692	0.934	0.666	0.918

Note: OLS. SE clustered at the regional level. Country FE included. *p<0.1; **p<0.05; ***p<0.01

Table 14: **Health outcomes for women.**

	Dependent variable: Regional differences in...			
	...use of...		...knowledge about...	
	modern contraception	HIV tests	modern contraception	HIV
	(1)	(2)	(3)	(4)
Constant	0.092***	0.043***	0.062***	0.055***
	(0.002)	(0.002)	(0.002)	(0.002)
SCI	−0.416***	−0.315***	−0.364***	−0.336***
	(0.047)	(0.047)	(0.033)	(0.040)
$\triangle$Contraceptive knowledge	0.212***			
	(0.004)			
SCI x $\triangle$Contraceptive knowledge	−24.352***			
	(4.419)			
$\triangle$Knowledge about HIV		0.309***		
		(0.003)		
SCI x $\triangle$Knowledge about HIV		−3.327		
		(3.988)		
$\triangle$Control variables (18)	Yes	Yes	Yes	Yes
Observations	116,403	122,759	122,759	122,759
R^2	0.729	0.855	0.677	0.911
Adjusted R^2	0.729	0.855	0.677	0.911

Note: OLS. SE clustered at the regional level. Country FE included. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 15: **Health outcomes for men.**

	Dependent variable: Regional differences in use of...			
	...modern contraception w/...		...HIV testing w/...	
	...low FB penetration	...high FB penetration	...low FB penetration	...high FB penetration
	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Constant	0.030**	0.039***	0.037	0.034**
	(0.013)	(0.010)	(0.024)	(0.016)
SCI	−0.220***	−0.738***	−0.131***	−0.262***
	(0.037)	(0.182)	(0.027)	(0.089)
△Contraceptive knowledge	0.216***	0.238***		
	(0.011)	(0.018)		
SCI x △Contraceptive knowledge	−13.166***	−40.980***		
	(5.040)	(12.636)		
△Knowledge about HIV			0.597***	0.469***
			(0.034)	(0.027)
SCI x △Knowledge about HIV			−26.123***	−21.334
			(8.574)	(16.412)
△Control variables (20)	Yes	Yes	Yes	Yes
Observations	37,215	85,545	38,701	85,319
R^2	0.672	0.720	0.905	0.939
Adjusted R^2	0.672	0.720	0.905	0.938

Note: OLS. SE clustered at the regional level. Country FE included. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 16: **Effects of social connectedness and Facebook penetration on health behaviour disaggregated by Facebook penetration rate below and above the median.**